Faculty of Engineering
University of Porto

Minimizing Municipal Solid Waste through Food Waste Prevention: Comprehensive Analysis and Practical Tool

Miguel Nogueira Rodrigues

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

A thesis submitted to the Faculty of Engineering of the University of Porto for the Doctoral Degree in Industrial Engineering and Management

Supervisor: Vera Lucia Miguéis Oliveira e Silva

March 3, 2023

Minimizing Municipal Solid Waste through Food Waste Prevention: Comprehensive Analysis and Practical Tool

Miguel Nogueira Rodrigues

A thesis submitted to the Faculty of Engineering of the University of Porto for the Doctoral Degree in Industrial Engineering and Management

Approved by:

President: Maria Antónia da Silpa Lopes e Carravilha
Referee: Dries Benoit
Referee: Sara Sofia Baltazar Martins Correia
Referee: José Luís Cabral Moura Borges
Supervisor: Vera Lucia Miguéis Oliveira e Silva

March 3, 2023

To my sister and parents,
who are an inspiration and provided my education.
They have shaped me and my path.

To my family,
who have supported my journey.

And to my friends,
who have always been present.

"For scientists such as ourselves, perfection only brings despair.
It is our job to create things more wonderful than anything before them,
but never to obtain perfection.
A scientist must be a person who finds ecstasy
while suffering from that antimony."

Noriaki Kubo (1977—), Japanese manga artist.

Declaration

- Chapter 3 is based on the paper "Municipal Solid Waste: Topic Modeling and Content Analysis", under revision in *Journal of Cleaner Production*. The first author of this paper is Miguel Rodrigues and is co-authored with Vera Miguéis.

- Chapter 4 is based on the paper "A Literature Review on the Quantitative Approaches to Food Waste: Descriptive, Predictive, and Prescriptive Analyses", under revision in *Environmental Science and Pollution Research*. The first author of this paper is Miguel Rodrigues and is co-authored with Vera Miguéis.

- Chapter 5 is based on the paper "Machine Learning Models for Short-Term Demand Forecasting in Food Catering Services: A Solution to Reduce Food Waste", published in *Journal of Cleaner Production*. The first author of this paper is Miguel Rodrigues and is co-authored with Vera Miguéis, Susana Freitas, and Telmo Machado.

Abstract

The goal of this thesis is to comprehensively address and explore the intricate challenges rooted in municipal solid waste and its core component, i.e., food waste, with the ultimate aim of formulating effective strategies for sustainable waste management. Using a variety of quantitative and qualitative methods, such as Latent Dirichlet Allocation, Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement methodology, and machine learning algorithms, this research seeks to uncover the multifaceted dimensions of municipal solid waste and food waste mitigation. The thesis unfolds into three pivotal research works, each alluding to a different perspective of waste management.

The first work revolves around topic modeling, exploring the most common topics on municipal solid waste that concern the scientific community and the European Union policymakers. For each of these two spheres, a Latent Dirichlet Allocation model is developed, and the resulting topics are identified, labeled, and grouped into larger topic groups. The topics are presented and discussed to provide an overview of municipal solid waste. In addition, the similarity between the content generated by the scientific community and projects funded by the European Union is evaluated.

The second work explores quantitative food waste approaches through a systematic literature review using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement methodology. This analysis dissects analytical dimensions and culminates in a meticulous critical review. This work navigates through key features of descriptive, predictive, and prescriptive analysis, providing a panoramic view of food waste approaches within the waste management landscape.

The third work focuses on food waste minimization by deploying food demand machine learning predictive algorithms across three food services. This work assesses waste reduction effectiveness and quantifies the potential benefits by comparing the predictions of the deployed models to the baseline models, which aim to mimic how the food catering services perform their own predictions. The amount of food waste that can be avoided and the number of meals that can be served are quantified, emphasizing the importance of reliable demand forecasting models within food services operations.

Together, these three research works aim to bridge the gap between academic insights and actionable strategies for sustainable waste reduction. At the core of this research is the recognition of the staggering challenges posed by municipal solid waste and food waste across economic, environmental, and social domains. This research meticulously dissects prevailing waste management paradigms while advocating for a more cohesive, integrated approach. In particular, it highlights the crucial role of food waste minimization and robust demand forecasting models within the food service sector, urging a concerted effort between the research and policy communities. Through the use of various quantitative analyses, this thesis briefs a course towards sustainable waste reduction practices, with the aim of bridge the gap between academic insights and practical policy implementation.

***keywords*:** Cosine Similarity, Data Analytics, Demand Forecast, Food Catering Services, Food Waste, Frequency Analysis, Latent Dirichlet Allocation, Machine Learning, Municipal Solid Waste, Quantitative Analysis, Sustainability, Systematic Literature Review, Text Mining, Topic Modeling, Waste Reduction.

Resumo

O objetivo desta tese é abordar e explorar de forma abrangente os desafios intrincados enraizados nos resíduos sólidos urbanos e na sua componente principal, os resíduos alimentares, com o objetivo final de formular estratégias eficazes para uma gestão sustentável dos resíduos. Recorrendo a uma variedade de métodos quantitativos e qualitativos, tais como a *Latent Dirichlet Allocation*, a metodologia *Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement* e algoritmos de *machine learning*, esta investigação procura descobrir as dimensões multifacetadas dos resíduos sólidos urbanos e da mitigação dos resíduos alimentares. A tese desdobra-se em três trabalhos de investigação fundamentais, cada um deles aludindo a uma perspetiva diferente da gestão de resíduos.

O primeiro trabalho gira em torno da modelação de tópicos, explorando os tópicos mais comuns sobre resíduos sólidos urbanos que preocupam a comunidade científica e os decisores políticos da União Europeia. Utilizando dois modelos de *Latent Dirichlet Allocation*, os tópicos mais importantes podem ser definidos, nomeados e agrupados em grupos de tópicos maiores para as duas fontes mencionadas. Os tópicos são apresentados e discutidos de modo a proporcionar uma visão mais alargada dos resíduos sólidos urbanos. Além disso, a correlação entre as duas fontes é avaliada pela semelhança de cosseno.

O segundo trabalho analisa as abordagens quantitativas do desperdício alimentar através de uma revisão sistemática da literatura, utilizando a metodologia *Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement*. Esta análise disseca dimensões analíticas e culmina numa revisão crítica meticulosa. Este segmento navega através das principais características da análise descritiva, preditiva e prescritiva, proporcionando uma visão panorâmica do desperdício alimentar no panorama da gestão de resíduos.

O terceiro trabalho centra-se na minimização do desperdício alimentar através da aplicação de algoritmos de *machine learning* em três serviços alimentares. Este segmento avalia a eficácia da redução de resíduos e quantifica os potenciais benefícios, comparando as previsões dos modelos implementados com os modelos de referência, que visam imitar a forma como os serviços de restauração efectuam as suas próprias previsões. A quantidade de desperdício alimentar que pode ser evitada e o número de refeições que podem ser servidas são quantificados, realçando a importância de modelos fiáveis de previsão da procura nas operações de serviços alimentares.

Em conjunto, estes três trabalhos de investigação pretendem colmatar a lacuna entre os conhecimentos académicos e as estratégias praticadas para a redução sustentável dos resíduos. No centro desta investigação está o reconhecimento dos enormes desafios colocados pelos resíduos sólidos urbanos e pelos resíduos alimentares nos domínios económico, ambiental e social. Esta investigação disseca meticulosamente os paradigmas de gestão de resíduos prevalecentes, defendendo simultaneamente uma abordagem mais coesa e integrada. Em particular, destaca o papel central da minimização do desperdício alimentar e de modelos robustos de previsão da procura no sector dos serviços alimentares, apelando a um esforço concertado entre a investigação e as comunidades políticas. Através da utilização de várias análises quantitativas, esta tese traça um rumo para práticas sustentáveis de redução de resíduos, com o objetivo de colmatar a lacuna entre os conhecimentos académicos e a implementação prática de políticas.

***Palavras-chave*:** Similaridade de cosseno, Análise de Dados, Previsão da Procura, Serviços de Restauração, Desperdício Alimentar, Análise de Frequência, Latent Dirichlet Allocation, Machine Learning, Resíduos Sólidos Urbanos, Análise Quantitativa, Sustentabilidade, Revisão Sistemática da Literatura, Extração de Texto, Modelação de Tópicos, Redução de Resíduos.

Acknowledgments

I am immensely grateful to acknowledge the collaborative efforts and support extended by numerous individuals and institutions throughout the development of this work.

Acknowledgments to people

- I am deeply grateful to my supervisor **Professor Vera Miguéis** for the opportunity to work with her. Her insights, kindness, and ever-lasting motivation were the cornerstone of this work. Truly an inspiration.
- I must thank **Susana Freitas** and **Telmo Machado** for their cooperation and intuition.
- I also extend my appreciation to **Soledade Medeiros**, **Isabel Ribeiro**, **Vânia Magalhães** and **Mónica Silva** for their kindness and availability.
- Finally, I would like to thank everyone who has supported me along the way, both family and friends, who have given me strength and with whom I share my accomplishments.

Acknowledgments to institutions

I also acknowledge the following institutions that supported this research throughout its development.

- I must thank **Faculty of Engineering of the University of Porto (FEUP)** and the **Doctoral Program in Industrial Engineering and Management (PRODEGI)** for taking me as their student and providing the necessary tools to develop this thesis.
- I am thankful to **LIPOR** for their collaboration and funding, providing relevant insights, and sharing the data that was pivotal for the development of this work.
- I am also grateful to **INESC TEC: Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência** for contributing to the funding of this work.

Miguel Rodrigues

Contents

List of Figures

List of Tables

Acronyms

ARIMA	Autoregressive Integrated Moving Average
ARIMAX	Autoregressive Integrated Moving Average with Exogenous variables
ETS	Exponential smoothing
EU	European Union
FAO	Food and Agriculture Organization
FCS	Food Catering Service
FL	Food Loss
FSC	Food Supply Chain
FW	Food Waste
GAB	Guggenheim Anderson–de Boer
IQR	Interquartile Range
LDA	Latent Dirichlet Allocation
LightGBM	Light Gradient-Boosting Machine
LSTM	Long Short-Term Memory
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Average Percentage Error
ME	Mean Error
MLP	Multi-Layer Perceptron
MSE	Mean Squared Error
MSW	Municipal Solid Waste
NARXNN	Nonlinear Autoregressive Exogenous model in combination with Neural Networks
NN	Neural Networks
OWTF	Organic Waste Treatment Facility
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
R2	Coefficient of Determination
RDC	Distribution/Collection Center
RF	Random Forest
RMSE	Root Mean Square Error
SARIMA	Seasonal ARIMA
SL	Shelf-Life
TBATS	Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend and Seasonal components
TF-IDF	Term Frequency–Inverse Document Frequency
wMAPE	weighted Mean Absolute Percentage Error

CHAPTER 1

Introduction

The present chapter aims to contextualise the problems the thesis covers. The chapter starts with Section 1.1, presenting the motivation to cover municipal solid waste and food waste, while Section 1.2 introduces the research goals of the thesis. Section 1.3 identifies the main contributions of this work, and Section 1.4 presents how the thesis is organized.

1.1 Motivation

More than 2 billion tons of Municipal Solid Waste (MSW) is estimated to be produced each year, with this number growing to 3.40 billion tons per year in the next three decades (World Bank, 2021). This corresponds to all waste produced in urban areas, from mega-cities to small towns (Diaz-Barriga-Fernandez et al., 2018). Although its exact composition changes depending on the environment in which it is generated, i.e. the content changes from city to city (Adhikari et al., 2018; Ayeleru et al., 2018; Sondh et al., 2022), its maladies are universal. Considering its enormous volume, its consequences ripple through the economic, environmental, and social spheres (Gao et al., 2021; Girotto et al., 2015; Kamble et al., 2020).

These consequences will only increase as MSW is highly correlated with the size of the world's population, which is estimated to continue growing (Markande et al., 2021; Chew et al., 2019; Przydatek and Kanownik, 2019). This is also confirmed by the World Bank, which predicts a 70 % increase in MSW over the next three decades (World Bank, 2021). Thus, the MSW problem has naturally attracted the attention of the European Commission and other international organizations (The European Commission, 2003; Mazzanti and Zoboli, 2008).

At the forefront of addressing the challenges MSW poses is the crucial role of policymakers and governing bodies. Entities such as the European Union and the United Nations wield significant tools and resources to formulate and advocate for measures aimed at mitigating the effects of the blooming MSW generation and waste production in general. These tools and resources appear in the form of: regulations and policies; economic instruments; education and awareness campaigns; research and development; and international cooperation (European Commission, 2021; United Nations Environment Programme, 2021). However, understanding policymakers' key initiatives and objectives is paramount to understanding strategies aimed at reducing MSW and its adverse effects.

Extensive studies have delved into the realm of policies and legal norms regarding waste management, with some focusing on qualitative assessments (Giordano et al., 2020; Thyberg and Tonjes, 2016). In contrast, others adopt quantitative methodologies (Yue et al., 2021; Shirota et al., 2016) to discern prevalent themes within policy documents. However, there is a significant gap in understanding the underlying themes of projects funded by the EU budget and the relationship between the objectives of these projects and those addressed in the scientific literature. This gap highlights the need to analyze the harmony between the policymakers' goals and the scientific efforts to address the challenges MSW poses.

A way to deal with MSW is to segment and tackle a portion of it, targeting the actions that need to be taken. Taking MSW as a whole, its most significant slice is organic waste, with its majority being Food Waste (FW) (Gu et al., 2017; Ayeleru et al., 2018; Shapiro-Bengtsen et al., 2020). In relative terms, this means that 50 % of all MSW is FW (FAO, 2011). Therefore, focusing on FW is a solid starting point for mitigating MSW (Beheshti and Heydari, 2023; Pérez et al., 2023).

With its notorious status, FW is also a mature field in the literature. In fact, there are two main branches concerning FW, i.e., quantitative and qualitative. The former is concerned with quantifying and sometimes identifying the content of FW (Dhir et al., 2020; Pirani and Arafat, 2016; Filimonau and De Coteau, 2019), while the latter generally aims to find the drivers of FW, understand how to avoid it and better allocate the waste produced (Dhir et al., 2020; Canali et al., 2017; Hebrok and Boks, 2017). Of these two branches, the quantitative branch is responsible for presenting the quantitative methods that allow FW reduction. However, the reviews of this branch do not provide a detailed picture of the methods and data used in the different studies or their results. In addition, they address only the descriptive dimension of the analysis, neglecting the potential of predictive and prescriptive approaches.

Also, it is essential to continue to develop predictive approaches to support the decision-making of food catering services, mainly considering that the common cause for FW generation is food overproduction that comes from poor demand forecasts (Filimonau and De Coteau, 2019; Parfitt et al., 2010; Pirani and Arafat, 2016). Moreover, most hospitality companies cook 10 % more food than they need to avoid stockouts because they do not trust the accuracy of their forecasts (Pirani and Arafat, 2016). Hence, better and more accurate forecast approaches are more suited to fight against FW generation.

1.2 Research goals

This thesis appears as a top-down view of MSW, exploring the field, understanding its various nuances, and motivating its minimization. This is done by first analyzing the field of MSW in two different spheres - the scientific literature and the projects funded by the EU - to understand better the core themes addressed by both. Then, the focus is put on FW, the largest part of MSW. From there, the quantitative branch of FW is studied, and a tool capable of helping reduce FW generation is developed.

The management of MSW is of great importance as its volume is expected to increase every year. Given its economic, environmental, and social consequences, it is imperative to fight for the minimization of MSW. This can be done in two ways. On the one hand, by better understanding MSW and identifying solutions that minimize its largest consequences. On the other hand, by managing the contents of MSW in a way that minimizes their generation. Considering the above, three specific research objectives were developed and are presented below.

The first research goal is to capture the main themes explored in the field of MSW with respect to the scientific community and the projects funded by the EU and understand how the two are related. This consists of identifying the main topics and the overarching themes explored in each of the two mentioned spheres, as well as investigating the similarities between them.

The second research goal focuses on exploring the quantitative branch of FW in the literature. The goal is to focus on the three types of analytics - descriptive, predictive, and prescriptive - and to extract and summarize the most relevant features of the quantitative studies.

The third research objective is to develop a future demand forecasting model capable of informing food catering services managers. These predictions should aid short-term decision-making and reduce FW by motivating better provisioning.

To achieve the mentioned goals, two quantitative and one qualitative approaches are taken. To extract the main themes of MSW presented by the scientific community and EU-funded projects, two Latent Dirichlet Allocation (LDA) models were employed. The results of both were compared to investigate their similarities and differences, with the cosine measure being used to quantify their similarity.

Then, a systematic literature review was performed based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement methodology to explore the quantitative branch of FW. Different keywords were used to extract relevant articles from different databases, and eligibility criteria were also defined to help filter the relevant articles.

Focusing on FW, four different machine learning algorithms were applied to perform future food demand forecasting: Random Forests (RF), Long Short-Term Memory (LSTM) neural networks, Light Gradient Boosting Machines (LightGBM), and transformer neural networks. The results of these four algorithms were also compared to a baseline model designed to mimic how food catering service operators typically estimate future demand.

1.3 Main contributions

The response to the previously defined research goals aims to complement existing knowledge in the scientific community and advance current research. The contributions and the data collection process of each chapter are presented in the following paragraphs. In addition, a comprehensive description is also present in their corresponding chapter.

Chapter 3 presents a detailed analysis of the MSW field, taking into account both the scientific community and the projects funded by the EU simultaneously. It takes a novel approach to policy analysis by employing an LDA model and complementing it further with a cosine similarity analysis. It leverages over 25,000 scientific papers and 140 funded projects from 1983 to 2022 to extract 30 and 25 topics, respectively. The topics are then grouped into broader themes to provide a better overview of what has been addressed in the two spheres. Then, a comparison between the two spheres is performed to understand if the topics addressed by the scientific community are the same as the ones tackled by the projects funded by the EU, or if the EU-funded projects tend to favor specific domains.

In Chapter 4, a systematic literature review is performed, considering the quantitative branch of the literature on FW, based on the PRISMA statement methodology. A set of keywords was defined, and 499 unique papers were extracted from the Elsevier SCOPUS and Web of Science databases. Of these, 65 papers were considered and classified according to the analysis performed, i.e. descriptive, predictive, and/or prescriptive. Key characteristics were extracted for each paper, and other general conclusions such as the most popular analyses, settings, and methods were used. This chapter contributes to the literature by providing a critical analysis of the field, highlighting areas for improvement and guiding future research.

Chapter 5 presents a set of forecasting models that can reliably assist food service decision-makers in reducing FW generation. The developed future demand forecast models include two causal models - one based on an RF algorithm and the other based on a LightGBM algorithm - and two-time series models - one based on an LSTM neural network and the other based on a transformer neural network. These were trained, validated, and tested in three different settings: two school canteens and a company canteen. A baseline was then developed to reflect the process typically used by food catering services to estimate future demand. The added value of reliable predictions is quantified by comparing the predictions of the developed models with the predictions of a baseline. In fact, this methodology allows food catering service managers to measure the amount of FW that could have been avoided, while also quantifying the number of additional meals that could have been served with the support of the models proposed.

1.4 Thesis outline

In addition to this chapter, this thesis consists of five other chapters. Chapter 2 acts as an introductory contextualization, providing an explanation of terms and other aspects for a more seamless reading. Various literature reviews have been considered for a better exposition of the relevant issues, going into detail on both MSW and FW characteristics.

Chapters 3, 4, and 5 refer to the studies that are committed to the previously defined research goals and contributions. These studies have been submitted to prestigious journals with the aim of contributing to and complementing the existing work in the scientific community. The work presented in Chapter 5 has already been accepted and published in the Journal of Cleaner Production.

Finally, Chapter 6 raises the main conclusions of this thesis, going through the contributions that it has produced, as well as the limitations that have been faced. In addition, possible future research is considered, discussing possible paths for it.

CHAPTER 2

Challenges and trends in municipal solid waste

This chapter presents a comprehensive contextualization, introducing the issues related to MSW and its content. It highlights the most discussed domains of MSW, alluding to the seriousness of MSW in the past, present, and near future. This chapter also comprehensively explains its current status, typical composition, and innovative approaches to minimize its impact. In addition, relevance is also given to FW since it is a large portion of MSW.

2.1 Introduction

"Waste" can be defined as "an object the holder discards, intends to discard or is required to discard" according to the European Union, under the Waste Framework Directive 2008/98/EC, Article 3(1) (European Union, 2008).

Waste can also be classified according to its form and origin. In terms of its form, waste can be liquid, solid, or gaseous (Taylor and Allen, 2006), with solid waste being the most popular of the three. Solid waste can refer to any unwanted, discarded, or disposed material (garbage, refuse, or trash) that results from daily activities of the community and includes domestic, commercial, and industrial waste (Sreedevi, 2015; Girish et al., 2019).

Concerning its origin, waste can be classified as municipal, industrial, construction and demolition, mining, and agricultural (Environment Agency, 2014; European Environment Agency, 2004). Considering both the form and origin of the waste, MSW is the broadest and most notorious (Tian et al., 2012), being produced in massive quantities and leading to severe consequences (Diaz-Barriga-Fernandez et al., 2018).

MSW can be referred to as commercial and domestic wastes (in a solid or semi-solid form) generated in municipal or notified areas by various society activities (Hiremath and Goel, 2017; Girish et al., 2019). MSW includes organic and inorganic treated bio-medical waste but excludes industrial hazardous waste.

The composition of MSW is highly diverse, as it is generated by households (Chu et al., 2016; Fadhullah et al., 2022), schools (Ugwu et al., 2020), and commercial establishments (offices, hotels and shops) (Economic & Social Commission for Asia & the Pacific, 2000). Since the proportion of the materials found in MSW fluctuates not only over time but also across different geographical locations, its exact characterization becomes difficult (Yang et al., 2018; Sondh et al., 2022; Adhikari et al., 2018). Nevertheless, its most significant portion usually corresponds to organic waste, which in turn is mainly composed of FW (Gu et al., 2017; Ayeleru et al., 2018; Shapiro-Bengtsen et al., 2020).

2.2 Global perspective on MSW

MSW is the waste generated in urban areas, ranging from megacities to small towns (Diaz-Barriga-Fernandez et al., 2018). According to the World Bank, it is estimated that approximately 2.01 billion tons of MSW are generated annually worldwide, with this number increasing to 3.40 billion tons per year by 2050 (World Bank, 2021). Given that the problems caused by MSW are already alarming (The European Commission, 2003; Mazzanti and Zoboli, 2008), the expected 70 % increase in the amount produced over the next three decades will only exacerbate existing problems.

In addition to the World Bank's estimates, other studies also support the growth of MSW in the coming years. This is the case of the following two studies that analyze MSW in India. The estimation of Ramachandra et al. (2018) suggests an increase of MSW by 1.33 % per year, which means an increase of more than 5,000 % in 30 years, while Joshi and Ahmed (2016) states that MSW would increase by about 300 % between 2021 and 2051.

The countries that generated more MSW in 2020 were the USA, China, India, Brazil, and Indonesia, with estimates for 2050 putting China well ahead of any other country (Nanda and Berruti, 2021). This is consistent with the commonly reported correlation between population growth and MSW generation (Vyas et al., 2022; Markande et al., 2021), with many papers pointing to India and especially China as the main contributors to current and future MSW generation (Liu and wu Wu, 2011; Singh et al., 2011).

MSW has serious environmental, economic, and social consequences (Istrate et al., 2020; Shindhal et al., 2021; Xiao et al., 2020). Of the three, environmental problems are the most notorious (Shah et al., 2021; Khoiron et al., 2020b; Vaverková et al., 2018), usually related to sustainability issues (Iyamu et al., 2020) and highlighting air, soil, and water pollution (Khoiron et al., 2020a).

Air pollution is usually caused by greenhouse gases (Rajaeifar et al., 2017) that come from the disposal of non-hazardous waste through incinerators and burners (U.S. Environmental Protection Agency, 2023). In contrast, soil pollution is a product of landfill contamination (Nanda, 2011), and water pollution comes from groundwater contamination (Al-Salem et al., 2017). Nevertheless, their origin usually comes from the mismanagement of residuals (Ramachandra et al., 2018), with landfills being the place where the pollution/contamination occurs (Shen et al., 2015).

The economic consequences of MSW revolve around the cost of maintaining landfills, which includes the cost of dealing with waste in general (Guoyan et al., 2022). Although both have costs, a waste-to-energy system is more expensive than a resource recovery system (Liang et al., 2022).

In addition, MSW also has consequences for the overall well-being of the population. Some specific consequences are: 1) adverse public health effects on the population such as skin, eye, and nasal irritation, allergies, and a non-zero risk of cancer (Khoiron et al., 2020a); 2) an increased risk of mortality and respiratory diseases, and adverse mental health effects (Vinti et al., 2021); and 3) serious health problems caused by diseases spread by insects and pests that come from the natural decomposition process of organic waste (Sharholy et al., 2008).

Nevertheless, these are all interrelated, and the mismanagement of landfills leads to a cascade of consequences. Landfills, as places where waste is stored and managed after it has been collected, are at the heart of waste management (Renou et al., 2008; Kjeldsen et al., 2002) where waste left unattended in them contaminates the soil (environmental consequence) and contributes to the deterioration of public health (social consequence). Hence, tackling the consequences of MSW becomes an arduous process with many moving parts, where addressing one problem does not necessarily solve another (Filimonau and De Coteau, 2019).

In the field of MSW, a lot of attention is given to waste management, with a significant focus on landfills, although most of them are not adequately managed (Sharholy et al., 2008). Nonetheless, this is just one of the destinations that MSW can have. More specifically, MSW can have three different fates: 1) disposal of the waste while it decomposes naturally and passively; 2) incineration, usually of non-hazardous waste only; and 3) extraction of value from the waste through recycling and energy conversion.

The first option is the least desirable because it creates the worst problems. Although the incineration of waste is a better option than the previous, it is also not the best one (Abdel-Shafy and Mansour, 2018a). On the one hand, it significantly reduces the volume and mass of waste, completely disinfecting it (Li et al., 2003). On the other hand, it is a costly process that produces toxic emissions (Zhang et al., 2010).

Extracting value from waste is then the most desirable option. It minimizes toxic emissions, reduces the mass and volume of waste in landfills, and is a cheaper process. This, in turn, can be achieved through recycling and waste-to-energy, both of which promote circular economy (Siddique et al., 2008; Sharholy et al., 2008; Abdel-Shafy and Mansour, 2018a). Recycling benefits agricultural activities and allows waste transformation into other materials (Abdel-Shafy and Mansour, 2018b). At the same time, waste-to-energy methods work with waste, typically organic waste, to turn it into a source of energy (Zhao et al., 2014; Cheng and Hu, 2010).

Despite the desirability of extracting value from waste, issues around landfills are usually the responsibility of government agencies or similar organizations (United Nations Development Programme (UNDP), 2016; World Bank Group, 2018). Thus, despite the findings and statements of the scientific community, there is no indication that the former support the policymakers' decision process, making it crucial to understand whether the issues and topics brought up by the scientific community and policymakers are coordinated.

Work on policy analysis includes both quantitative and qualitative approaches (Mayer et al., 2004) and examines a range of policies covering diverse areas such as climate change (Clark et al., 2016), migration (Castles, 2004), public health (Wing et al., 2018), renewable energies (Lo, 2014), biodiversity (Kleijn and Sutherland, 2003), food waste (Giordano et al., 2020; Thyberg and Tonjes, 2016), and monetary policies (Shirota et al., 2016). Although these analyses usually focus only on the policies themselves, some papers directly compare them with other research, as in the case of Yue et al. (2021). While policy analysis extensively covers various domains using diverse methodologies, its application within the realm of MSW remains notably scattered. There is a lack of studies employing this analytical approach specifically within the context of MSW.

The EU is one of the larger organizations with the power to implement changes in MSW management. In fact, the EU has recognized MSW as a significant policy concern for many years (The European Commission, 2003; Mazzanti and Zoboli, 2008). There are studies in the literature that specifically address EU policies. Some provide a comprehensive overview of EU policy-making, focusing on the impact of recent expansions and economic crises (Wallace and Wallace, 1998; Cram, 2020). In contrast, others analyze the progress of the policies of the EU as a whole (Versluis et al., 2010; Pattyn et al., 2018). In addition, some works specify a domain of policies, such as circular economy (Friant et al., 2021), water management (Hering et al., 2010), agriculture (Barnes et al., 2019).

2.3 Food waste analysis

To minimize MSW, it is of utmost importance to understand its contents. MSW is composed of different types of materials. Some of these are ceramics, rubber, leather, wood, metal, glass, textiles, paper, plastics, food waste, clothing, disposable tableware, garden waste, and cans, among many others (Gu et al., 2017; Ayeleru et al., 2018; Zhu et al., 2020). Although the composition of MSW varies through time and location, there is a tendency for organic waste to make up the most significant portion.

The work of (Karak et al., 2012) analyzed several EU countries and presented the percentage of the physical composition of MSW in each of them in different years. Figure 2.1 shows the percentage of organic waste over time. In some years, more than one country reported its percentage of organic waste. In these cases, the maximum and minimum values are given instead.

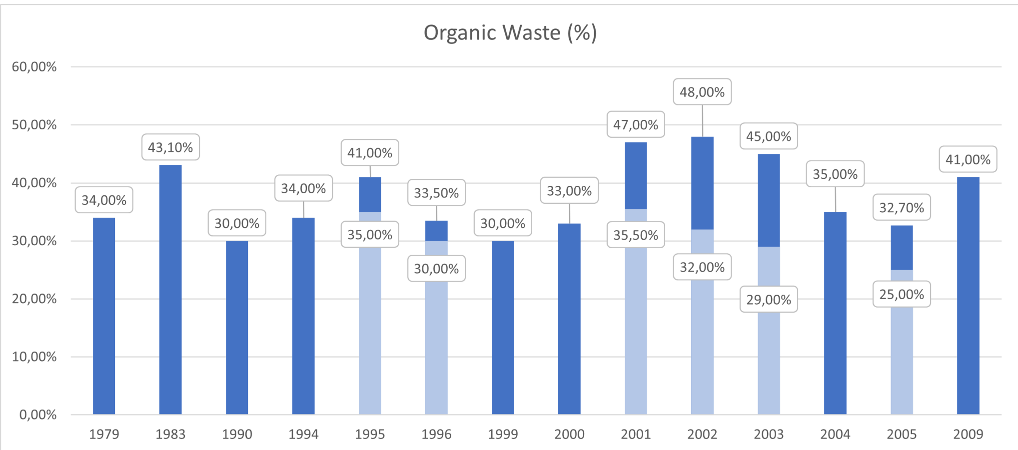

Figure 2.1: Minimum and maximum amounts of organic Waste (%) per year (Source: Karak et al. (2012))

Following the values presented by Karak et al. (2012), it can be seen that organic waste fluctuates between 25 and 48 % of MSW. Other works that provide the percentage of organic waste in MSW are Ugwu et al. (2020), which presents between 13.49 and 48.01 % in 2018. Considering the amount of FW in organic waste - between 41.3 and 60.2 % (Srivastava and Nema, 2008; Shapiro-Bengtsen et al., 2020; Gu et al., 2017) - this means that FW makes up the majority of MSW.

However, FW is a complex term that is sometimes incorrectly used (Thyberg and Tonjes, 2016; Girotto et al., 2015). Throughout the Food Supply Chain (FSC), food is produced, some of which is not being consumed and is discarded (Do et al., 2021). Depending on the FSC stage, this discarded food can have two classifications: FW or food loss (FL) (Li et al., 2022). If the waste occurs at the beginning of the FSC, it is usually called FL (Feukam Nzudie et al., 2020), while if it occurs at the end of the FSC, it is called FW (Priefer et al., 2016).

Furthermore, depending on its characteristics and condition, FW can have different classifications. Depending on its consumable properties, FW can be edible or inedible (Teigiserova et al., 2020a). Edible FW can be a product of incorrectly prepared food, expired date products, overproduction, food left from buffets, and leftover food from customers on the plate. At the same time, inedible FW can correspond to inedible parts of vegetables and bones (Silvennoinen et al., 2015).

FW can also be addressed as avoidable and unavoidable (Shaw et al., 2018). Avoidable FW refers to both discarded food that could have been eaten and food that went bad, while unavoidable refers to waste such as inedible peels and bones (Schott and Andersson, 2015). In addition, food that only some people eat, such as bread crusts, or that can only be eaten when prepared in a certain way, such as potato skins, can also be classified as possibly avoidable FW (Parfitt et al., 2010). These two types of classifications tend to overlap, with avoidable FW being edible and unavoidable FW being inedible.

By definition, edible/avoidable FW is worse than its counterpart because it represents food that could have been consumed. Although most studies do not quantify the proportion of avoidable FW (Pirani and Arafat, 2016), some reports show that up to 56 % of FW in households is avoidable (Edjabou et al., 2016), confirming the problematic nature of FW.

Regardless of its categorization, several different drivers can lead to the generation of FW. Although the exact number may vary, the identified drivers are somewhat consistent across studies. In general terms, it is accepted that social-demographic characteristics play a role in generating FW (Chia et al., 2023; Vittuari et al., 2023; Aloysius et al., 2023).

In terms of FW generated in households, the perception, awareness, and attitude of consumers towards FW are also influential (Diana et al., 2023), as well as how they deal with leftovers (Stancu et al., 2016), and their interaction with activities related to planning, shopping, storing, preparing and consuming food (Pearson and Minehan, 2013). Essentially, the study of consumer-level drivers is classified as behavioral studies that focus on understanding consumers' attitudes and, by extension, household members (Stancu et al., 2016; dos Santos et al., 2022).

Regarding the FW generated in the food services sector, this sector tends to overproduce and oversupply due to inaccurate forecasts (Filimonau and De Coteau, 2019; Parfitt et al., 2010; Pirani and Arafat, 2016). These inaccuracies usually result from rudimentary methods that cannot encompass the complexity and unpredictability nature of demand in the field. In addition, there are also other notable drivers of FW generation, such as portion size of the dish, menu variety, production process, and the use of pre-prepared foods (instead of whole foods) (Dhir et al., 2020). Overall, the drivers of FW can be classified into technological, institutional (business management and economics, and legislation and policy), and social (Canali et al., 2017).

Understanding the drivers behind FW is crucial due to the variety of problems it causes. Identifying the primary drivers of FW generation enables proactive measures to be taken to address its consequential impacts. These negatively impact the environmental, social, and economic spheres, manifesting through greenhouse gas emissions, global challenges to food security, and the substantial costs associated with wasted food (Gao et al., 2021; Beheshti et al., 2022). These problems are exacerbated the later the FW occurs in the FSC as food may have already been processed, packaged, stored, and even distributed (San Martin et al., 2021; Mosna et al., 2016). For this reason, developed countries are more harmful, as they have higher FW rates while developing countries have higher FL rates (Girotto et al., 2015; Parfitt et al., 2010). In any case, the food wasted throughout the FSC is usually not addressed as a multidimensional dilemma, as looking at it as an environmental and sustainability issue is quite different from looking at it as an economic and management issue (Filimonau and De Coteau, 2019).

Considering all this, it becomes even more imperative to minimize the issues raised by the FW generation. The options to achieve this are, from most to least desirable, prevention, preparation for reuse, recycling, recovery, and disposal (European Commission, 2021; Papargyropoulou et al., 2014). Therefore, priority should be given to preventive methods to avoid the generation of FW in the first place, while all other options focus on managing existing waste.

Notwithstanding, methods to reuse, recycle, and recover FW can be seen as circular economy-driven (Teigiserova et al., 2020b) as they focus on transforming and extracting value from FW (Siddique et al., 2008; Sharholy et al., 2008; Abdel-Shafy and Mansour, 2018a). For instance, a common approach is a waste-to-energy conversion, namely by producing biofuel and other derivatives (Zhao et al., 2014).

Notwithstanding, technological solutions, and food redistribution and recovery strategies also contribute to preventing and minimizing the impact of FW on society. Technological solutions may be relevant, such as packaging technologies (e.g., innovative packaging, biodegradable materials) (Poyatos-Racionero et al., 2018; Brennan et al., 2021) and apps that help consumers manage their food inventory and waste (Liegeard and Manning, 2020).

In order to prevent the generation of FW, it is important to understand not only why it happens but also its characteristics for the setting that is being studied to devise appropriate strategies. Some studies quantify and report the amounts of FW reported in different settings (Filimonau and De Coteau, 2019; Dhir et al., 2020). However, the number of studies is insufficient to provide significant insight and generalization since they are recorded in different settings, locations, and periods of time. Not only that, the way in which the amounts of FW are presented is also usually not consistent across studies. For example, some studies show the amount of FW as an absolute quantity (Malefors et al., 2022; Pirani and Arafat, 2016), while others present the amount of FW as a portion of the meals served or food produced (Conrad et al., 2018; Kandemir et al., 2022). Additionally, the amount of FW recorded can also be presented as total amounts or per meal/individual.

Despite all this, there are a few cases where enough information is given to compute values for comparisons, such as disclosing the number of individuals, time horizon, and total amount of food produced. For instance, Thanh et al. (2010) records the amount of FW per household for over a month, which translates into an estimate of over 6 tons of FW per year, or over 60kg of FW per year per person, while Adelodun and Choi (2020) presents the amount of FW in Korea to be over 4 million tons or over 180kg per person.

After understanding the massive amounts of FW produced in food services, it is necessary to invest in preventive measures to reduce the amount of FW produced. In turn, these preventive measures should focus on addressing the main drivers of FW generation. In the hospitality sector, FW is typically a byproduct of overproduction (Filimonau and De Coteau, 2019; Parfitt et al., 2010; Pirani and Arafat, 2016). Since companies do not have robust forecasting systems, they cannot rely on their forecasts. There are even reported cases where companies, due to their unreliable forecasts and to avoid stock-outs, choose to cook 10 % above their estimates (Pirani and Arafat, 2016). Thus, overproduction and overprovision seem to be a fundamental problem to be solved (Principato et al., 2021; Felix, 2018; Carbonneau et al., 2008). To this effect, more robust and reliable demand forecasting systems are needed (Aamer et al., 2021), where a reduction in forecasting error can also lead to higher revenues (Pereira and Cerqueira, 2021).

Although there are many areas where overproduction and overprovision occur, such as retail and households (Salinas et al., 2020; Kandemir et al., 2020), studies usually focus on the hospitality / HoReCa sector. In this area, most case studies revolve around school and university canteens that use data with daily periodicity and employ neural networks for their predictions (Faezirad et al., 2021; Malefors et al., 2021a, 2022). These applications can have a significant impact, reducing the amounts of FW produced by almost 80 % (Faezirad et al., 2021) and costs by 60 % (Malefors et al., 2021a).

2.4 Conclusion

In essence, this comprehensive contextualization provides a panoramic view of the multifaceted challenges posed by MSW and the pivotal role of FW within this landscape. The exploration goes beyond the historical gravity of MSW issues, underlining their persistence and amplification in current times, along with future estimations. Identifying the contents of MSW, its contributors, and its consequential impacts highlights the urgency of addressing this prevalent issue. At the same time, the attention directed towards FW, a substantial component of MSW, accentuates its diverse facets, drivers, and the necessity for appropriate strategies in its management. The lack of standardization in quantifying FW in various settings emphasises the complexity of this type of waste and the challenges in devising uniform approaches to its mitigation.

Furthermore, the clarity of the relationship between poor forecasts and FW generation highlights a critical link often overlooked in the discourse on waste management. By identifying reliable demand forecasting as a potential solution to mitigate FW, this contextualization sets the stage for innovative approaches to waste reduction. The integration of MSW and FW research within this chapter converges toward identifying optimal strategies to grasp these interconnected waste streams. This comprehensive understanding, which encompasses the historical, present, and foreseeable dimensions of MSW and FW, lays a solid foundation for the following exploration and analysis within this thesis.

In summation, this chapter not only delineates the severity of the MSW issues and the prominence of FW within it but also paves the way for a refined exploration of strategies, emphasizing the critical need for cohesive and suitable approaches in addressing these pressing waste management challenges.

CHAPTER 3

Municipal Solid Waste: Topic Modeling and Content Analysis

Municipal solid waste (MSW) is a very significant contemporary issue that poses monumental economic, environmental, and social challenges. To address these challenges sustainably, it is essential to first identify and analyze the prevalent themes, patterns, and emerging trends in this domain. In this context, advanced text analysis can help reveal the topics related to MSW that have been addressed by the scientific community and by policymakers through funded projects. Thus, by examining and comparing the content that has been produced by both spheres, this chapter aims to clarify what have been the concerns of the former in relation to the priorities of the latter in terms of reducing MSW and whether they are aligned. To achieve this, the frequency of the most popular terms in the corpora of both mentioned spheres was analyzed. Then, two latent Dirichlet allocation (LDA) models were developed to extract the most relevant topics from both corpora. Furthermore, this analysis was complemented with a quantitative estimation of the similarity between the two corpora. The findings of this work suggest that the two spheres are aligned, with the scientific community being a very mature actor in the field, exploring a wide variety of issues on MSW over the years. Furthermore, the popularity of the topics extracted from the European Union-funded projects fluctuated significantly over time. This shows how priorities change to accommodate the actions needed to minimize MSW at any given time. Finally, the projects funded by the European Union were found to have a clear tendency to deal with biowaste and waste management processes.

3.1 Introduction

There are several types of waste and each type has unique features and consequences that vary according to how, when, and where it was generated. The four main universal types of waste are municipal solid waste (MSW), industrial solid waste, medical waste, and construction waste (Hou et al., 2012; Tian et al., 2012). MSW consists of waste that is generated in urban areas, which can range from megacities to small towns (Diaz-Barriga-Fernandez et al., 2018), and is the most significant out of the four types. Every year, it alone represents 2.01 billion tonnes of generated waste (World Bank, 2021).

MSW can be divided into institutional, domestic, and commercial waste (Mukherjee et al., 2020; Joseph and Prasad, 2020; Ravi et al., 2020). It is produced by households, offices, hotels, shops, schools, and other institutions (Economic & Social Commission for Asia & the Pacific, 2000). Examples of MSW include ceramics, rubber, leather, wood, metal, glass, textiles, paper, plastics, food, clothing, disposable tableware, yard trimmings, and cans (Gu et al., 2017; Adhikari et al., 2018; Ayeleru et al., 2018; Sondh et al., 2022). Although the composition of MSW depends on the context in which it is produced and collected, it is mainly composed of organic waste, which in turn consists mostly of food waste (Gu et al., 2017; Ayeleru et al., 2018; Shapiro-Bengtsen et al., 2020).

All of this waste generation, but in particular MSW generation, leads to severe environmental, social, and economic issues (Gao et al., 2021; Girotto et al., 2015; Kamble et al., 2020; Filimonau and De Coteau, 2019). Furthermore, MSW is highly correlated with the number of people in the world, which is constantly increasing (Vyas et al., 2022; Markande et al., 2021; Chew et al., 2019; Przydatek and Kanownik, 2019). This means that the amount of MSW is also increasing, and its consequences are becoming more severe. Moreover, the amount of MSW is expected to increase by around 70% in the next 30 years (World Bank, 2021). Therefore, it is increasingly important to develop measures to reduce MSW generation.

MSW has become such a large problem that the European Commission has recognized it as a major concern and area of policy focus (The European Commission, 2003; Mazzanti and Zoboli, 2008). The European Union (EU) has been a major supporter of numerous projects in a wide range of fields, including MSW, promoting advances in technology, innovation, sustainability, health, and social sciences. As the EU continues to fund research and development initiatives related to MSW, it becomes relevant to understand whether the scientific community and EU policymakers are aligned and pushing toward similar goals in the context of EU-funded projects, or if there is some type of misalignment regarding some specific dimensions of MSW.

Thus, in the present work, a text mining approach is proposed to face this challenge. The first step in this approach is to perform a separate frequency analysis on two different spheres: EU-funded projects and scientific research. Then, the second step is to extract the topics that best characterize each of these spheres by using latent Dirichlet allocation (LDA). The main objective is to identify the similarities between these two spheres, focusing on the themes that both parties deem crucial while also highlighting any research gaps that the scientific community is working on and the issues that policymakers prioritize in their funding decisions. A quantitative approach that considers the content of the analyzed documents is also proposed with the goal of computing the similarities between the themes addressed by the EU-funded projects and the work being carried out within the scientific community.

The contributions of the present chapter should be manifold. First, the results obtained should be able to guide future research in two different ways. On the one hand, they should help researchers in addressing real-world challenges and contributing more directly to policymaking. On the other hand, future research could focus on issues that are found to be under-researched but highly funded. Additionally, the present work should provide insights into the strengths and limitations of the EU funding landscape and research community and thus motivate further joint efforts. Ultimately, the purpose of the present work is to enable the refinement of strategic decision-making processes within the EU funding landscape, allowing a more focused and synergistic approach to advancing scientific knowledge and addressing social challenges.

3.2 Literature review

The topic of MSW is quite mature within the scientific community. Out of the many different general types of waste, MSW can be considered the most relevant (Tian et al., 2012), since it is generated in large quantities and has serious consequences (Diaz-Barriga-Fernandez et al., 2018). The literature on MSW can be divided into three broad themes. The first relates to understanding the composition of MSW (Renou et al., 2008; Kjeldsen et al., 2002; Kulikowska and Klimiuk, 2008). This includes studying the absolute and relative amounts of content in MSW in different locations and how its characteristics change depending on the environment where it is generated (Shapiro-Bengtsen et al., 2020; Gu et al., 2017; Ugwu et al., 2020). Additionally, some studies focus on analyzing the types of waste that make up MSW (Campuzano and González-Martínez, 2016; Karak et al., 2012; Sondh et al., 2022).

The second major theme in the MSW literature concerns how to manage MSW and how to extract value from it (Al-Salem et al., 2009; Sánchez, 2009; Mao et al., 2015). There are two main research directions within this theme. The first is waste valorization, where value is extracted from existing MSW by recycling some of its components, thus reusing resources and contributing to the circular economy (Siddique et al., 2008; Sharholy et al., 2008; Abdel-Shafy and Mansour, 2018a). A very specific but mature approach to this is waste-to-energy conversion, wherein MSW is converted into energy (Ramos et al., 2018; Chand Malav et al., 2020; Scarlat et al., 2019). Usually, this process uses biomass to produce biofuel and other derivatives (Zhao et al., 2014).

The second research direction within waste management is the disposal of existing waste (Abdel-Shafy and Mansour, 2018a; Zhang et al., 2010; Zacarias-Farah and Geyer-Allély, 2003). This usually happens in landfills (Renou et al., 2008; Kjeldsen et al., 2002), causing several issues related to hazardous substances, waste streams, and chemical releases, which have also been explored in the literature (Abdel-Shafy and Mansour, 2017; Abdel-Shafy, 2015; Blundell, 2004). Other disposal processes explored in the literature include waste collection (Troschinetz and Mihelcic, 2009; Rowe, 2005) and incineration (Cherubini et al., 2009; Cheng and Hu, 2010; McKay, 2002).

The third and last major theme in the MSW literature concerns the consequences of MSW generation. Most of the studies in the literature mention these consequences, even if they are not their main focus. These consequences are generally classified as economic, environmental, and social (Istrate et al., 2020; Shindhal et al., 2021; Xiao et al., 2020). Out of these three, most studies focus on the environmental consequences (Shah et al., 2021; Khoiron et al., 2020b; Vaverková et al., 2018), which are deeply connected with sustainability issues (Iyamu et al., 2020).

Overall, the problem of MSW generation has interested not only the scientific community but also policymakers. More specifically, the EU has, for many years, acknowledged MSW generation as a significant concern and area of policy focus (The European Commission, 2003; Mazzanti and Zoboli, 2008). However, there is no indication that the scientific community and EU policymakers are aligned within this topic. It is therefore crucial to understand if the issues and topics studied by the scientific community are aligned with the EU-funded projects. This information is key for both spheres to continue to develop further.

The scientific community has published many articles on the topic of MSW, making it extremely hard to cover all the content. Thus, applying topic modeling to these scientific articles to extract intrinsic knowledge from them is a reasonable option (Blei and McAuliffe, 2008). One of the most commonly adopted and well-known topic modeling methods is LDA (Blei et al., 2003). LDA enables the identification of topics covered in text documents and determines the terms that best encapsulate each topic (Maier et al., 2018).

The application of LDA is well-documented, appearing in a panoply of different reviews that tackle distinct subjects, such as sustainability (D'Amato et al., 2017; Bickel, 2019; Wu et al., 2021a), health (Shatte et al., 2019; Zhang et al., 2017), virtual reality (Loureiro et al., 2019), quality management (Sánchez-Franco et al., 2023), service quality (Korfiatis et al., 2019), and even coronaviruses (Tran et al., 2020). In addition, LDA has been widely used to perform sentiment analysis (Guo et al., 2017; Mäntylä et al., 2018; Wu et al., 2021b).

Conversely, the use of LDA to perform policy analysis is less popular, with only a few studies covering this approach. One such example is the work of Yue et al. (2021), who used LDA to analyze 115 coal-related policy planning documents issued by Chinese authorities and 4,019 titles of coal-related research projects, extracting 12 and 30 topics, respectively. Another exception is the work of Shirota et al. (2016), who studied the Bank of Japan's monetary policy in 1998. There are also papers on the study of policies that did not use LDA (Giordano et al., 2020; Thyberg and Tonjes, 2016). However, to the best of the authors' knowledge, no studies in the field of MSW have used LDA to analyze EU-funded projects.

Hence, the present work provides three major contributions to the literature. First, by analyzing EU-funded projects in the field of MSW through the use of LDA. Second, by comparing the topics covered by EU-funded projects with the topics discussed in the scientific literature. And finally, by employing quantitative methods to estimate the similarity between EU-funded projects and scientific articles.

3.3 Methodology

In order to identify and compare the topics addressed by the scientific community and EU policy-makers in the context of EU-funded projects on MSW, two distinct LDA models were developed. In addition, cosine similarity was used to quantify the similarity of the content produced by each of these two spheres.

Carrying out the proposed methodology involved gathering the necessary data to be used as input in the models, preprocessing the data extracted from the texts, deciding the number of topics to generate, and then analyzing the output. As a complement to the LDA analysis, the cosine similarity measure was used to determine the similarity of the two spheres of discussion. With the exception of the first stage, all others were conducted using the R programming language. A representation of the methodology is shown in Figure 3.1.

The aforementioned stages are described in detail in the following sections.

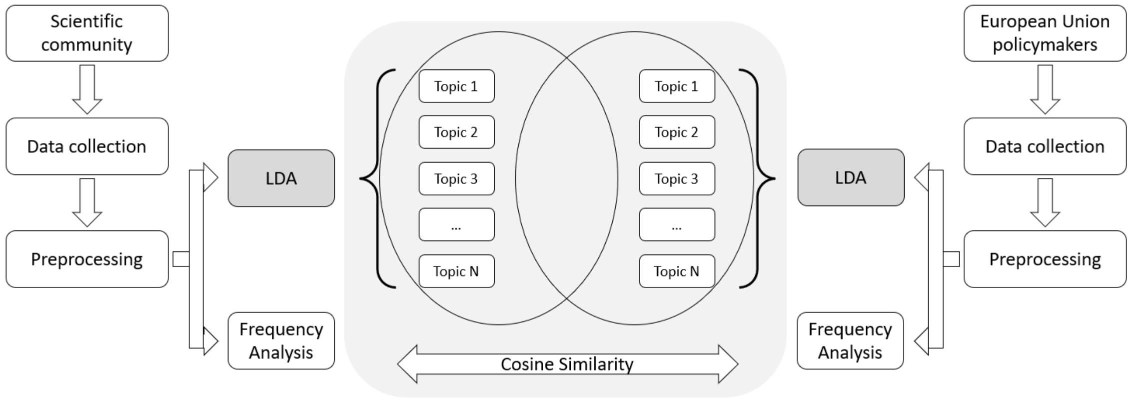

Figure 3.1: Visual representation of the methodology adopted

3.3.1 Data collection

The data collection was a two-step process. In the case of the scientific articles, the text data, i.e., the title, abstract, and author keywords, were extracted from Elsevier's Scopus database. In the case of the EU policymakers, it was considered that EU-funded projects corresponded to all documents with the classification of "project" and "programme". From these, the title, objective, and type of program were extracted from the Community Research and Development Information Service (CORDIS) website.

In both cases, the keywords "municipal solid waste" and its abbreviation "msw" were adopted. Although the adoption of such a broad concept allied with a reduced number of keywords can result in a large number of documents at the same time, this was necessary to allow a full review of the topic from both fronts.

3.3.2 Preprocessing

After extracting the relevant text data, it needs to be preprocessed. This stage involved several steps, starting with the removal of the entries that did not have complete data. In the case of the scientific articles, this referred to the title, abstract, and/or author keywords. This was the only step where the number of documents to be used was reduced.

The following steps focused on the content of the remaining documents and consisted of removing and manipulating the terms contained in them. This included removing words that started or ended with numbers, replacing "." and "," with a space, removing all punctuation and suppressing hyphens between words, removing non-alphanumeric characters, removing words with two or fewer characters, replacing "municipal solid waste" with "msw", replacing "ghg" with "greenhouse gas", stemming words to a common base, and removing stop words. For the latter, a set of 1,142 stop words extracted from the "onix", "SMART", and "snowball" lexicons were considered. Additionally, the words "Elsevier", "Ltd", "springer", and "springerverlag" were also removed. These words appeared often at the end of the abstracts, even though they were not related to the content, but rather to publishing conditions.

3.3.3 Parameter tuning: number of topics

Before performing the LDA analysis, the number of topics to use in each model needs to be defined. This process is done by estimating one or more performance metrics for a set of values representing the number of topics to extract. In the literature, this range is usually set to be between 2 and 100 (Karami et al., 2020) or a sub-range of it. For instance, Sutherland and Kiatkawsin (2020) considered a range of 5 to 75 topics, Chu et al. (2020) considered a range of 20 to 80 topics, and Lee and Yu (2018) considered a range of 15 to 25 topics.

For the present work, a range of 2 to 100 topics was adopted. A model for each possible number of topics was created and its performance was assessed through four metrics: Griffiths2004, CaoJuan2009, Arun2010, and Deveaud2014. These metrics were developed by Griffiths and Steyvers (2004), Cao et al. (2009), Arun et al. (2010), and Deveaud et al. (2014), respectively. To determine how many topics should be extracted Griffiths2004 and Deveaud2014 should be maximized, while CaoJuan2009 and Arun2010 should be minimized.

It is normal for the suggested number of topics to differ from metric to metric, as each metric measures different things. Deveaud2014 measures topic coherence, i.e., the degree of semantic similarity between the top words in a topic; Griffiths2004 measures perplexity, i.e., how well a model can predict new data; CaoJuan2009 measures document coherence, i.e., how well a topic model can group similar documents together; and Arun2010 measures topic exclusivity, i.e., how distinct each topic is from all other topics in the model (Sarkar et al., 2020).

In addition, when choosing the number of topics to adopt, it is relevant to consider the interpretability of the results (Buenano-Fernandez et al., 2020). In other words, a certain level of topic readability should be maintained (Debnath and Bardhan, 2020; Chu et al., 2020).

After defining the number of topics to use, the final LDA models were created.

3.3.4 Latent Dirichlet allocation

After preprocessing the data, it became a viable input for the LDA model. LDA models operate as three-level hierarchical Bayesian models that aim to cluster unlabeled documents into coherent and representative topics (Blei et al., 2003). At their core, LDA models hypothesize that a collection of documents (corpus) contains latent themes (topics), with each theme being comprised of terms found across the documents (Chu et al., 2020).

There are three key steps to generating an LDA model. Initially, the number of topics needs to be defined, and then a topic has to be assigned to each term in every document. Next, using the Gibbs sampling algorithm, new topics are assigned to each term iteratively until convergence is achieved. Finally, the generation of the topic-word co-occurrence frequency matrix summarizes the learned relationships between topics and words, serving as the finalized LDA model.

As a result, LDA yields probabilities for words within topics and probabilities for topics within documents. These probabilities form distributions that enable the identification of the most representative words for each topic and assess the likelihood of documents aligning with specific topics

(Karami et al., 2020). Moreover, LDA facilitates the extraction of prevailing topics across an entire corpus, offering insights into the dominant themes permeating the collection of documents.

3.3.5 Similarity estimation

To complement the LDA analysis, cosine similarity was used to give a measure of the approximation between the documents. This method provides a quantitative analysis that can complement other analyses (Chen et al., 2015; Patra et al., 2020). In fact, some researchers have already used cosine similarity in conjunction with LDA to achieve a more comprehensive analysis of different subjects (Lewis and Young, 2019; Blair et al., 2020; Rieger et al., 2021; Ramya et al., 2018).

Cosine similarity is a measure of similarity between two non-zero vectors. It is a mathematical metric used to determine the cosine of the angle between two vectors and not the magnitude of the vectors themselves. It ranges from -1 (perfect dissimilarity) to 1 (perfect similarity) (Manning et al., 2008).

Extending cosine similarity to the field of text translates into measuring the similarity between the absolute term frequency, or the term frequency–inverse document frequency (TF-IDF) weights of words, in two or more documents.

Before computing the cosine similarity between the documents in both corpora, the approach presented by Singh et al. (2013) was followed and the terms that only appeared once in all the documents combined were removed. This was done to reduce the dimensionality of the feature space. However, since less frequent words may have more information (Liang et al., 2017), the term frequency of the documents was transformed into TF-IDF weights. Given the weights adopted, the cosine similarity could only range from 0 (decorrelation) to 1 (perfect similarity), since term frequencies cannot be negative (Jurafsky and Martin, 2020; Salton and McGill, 1986).

The formula used to calculate the cosine similarity is as follows:

$$\text{Cosine Similarity } (\cos(\theta)) = \frac{A \cdot B}{\|A\| \cdot \|B\|} \tag{3.1}$$

Wherein the cosine similarity is represented as $\cos(\theta)$, and θ refers to the angle between the two vectors being compared. These two vectors are denoted as A and B, and they correspond to TF-IDF vectors from two documents, where each component of the vector corresponds to a unique term in those documents. In other words, each of the two vectors refers to two different documents populated with the TF-IDF of every word present in all documents. $A \cdot B$ is the dot product of vectors A and B, and $|A| \cdot |B|$ represents the magnitudes or norms of vectors A and B, respectively.

The cosine similarity analysis explored the degree of similarity within and between two distinct corpora, this is, between the scientific articles and EU-funded projects. Initially, the similarity between these corpora was assessed by computing the maximum cosine similarity score between each scientific article and all EU-funded projects, as well as between each EU-funded project and all scientific articles. This step allowed the capture of the highest level of similarity between any pair of documents from different sources.

Subsequently, the internal consistency of each corpus was examined by determining the maximum cosine similarity within each set of documents. Specifically, the maximum similarity value among scientific articles and among EU-funded projects was estimated. This approach facilitated the identification of the similarity of the most closely related documents within each corpus independently.

By adopting this methodology, the study aimed to comprehensively analyze the similarities both across and within the scientific articles and EU-funded projects, thereby providing insights into the thematic coherence and potential intersections between these two fields.

3.3.6 Results and analysis

A frequency analysis was conducted for both corpora, which consisted of extracting the distribution of the word frequency as a whole, determining the most common words, and representing them in a word cloud. This provided us with an overview of the corpus properties, such as the vocabulary diversity of the documents.

Additionally, the topics extracted from both corpora were analyzed. This included naming every topic according to the most common terms and categorizing sets of topics into broader topic groups, i.e., the extracted topics corresponded to subgroups of larger clusters. This provided a good overview of the contents raised by the scientific community and the EU-funded projects.

The frequency of each topic was also considered in the analysis. This information was made available by comparing the gamma of each topic, i.e., the topic distribution for each document in the corpus, to the sum of the gammas of every topic. Then, the frequency of the different topics was grouped according to the topic group they corresponded to.

The similarity of the content generated by the two explored spheres is presented in the next section.

3.4 Results

In the following subsections, the results obtained from the application of the proposed methodology are presented. The nomenclature adopted for the different objects was ’topics’, i.e., latent themes extracted by the LDA models; ’topic groups’, i.e., a group of topics; and ’corpus’, i.e., the documents corresponding to either the scientific articles or the EU-funded projects. A visual representation of this can be seen in Figure 3.2.

3.4.1 Data collection and preprocessing

The document extraction query was performed in July of 2023. This query yielded 34,689 unique articles from SCOPUS and 140 EU-funded projects from the CORDIS website. Figure 3.3 shows how many EU documents are available for each year. Since the time horizon covered by the EU documents was more limited, the time horizon of the analysis was framed by that period, i.e., from 1983 to 2022. This meant that 550 scientific articles were excluded with 34,139 left to use.

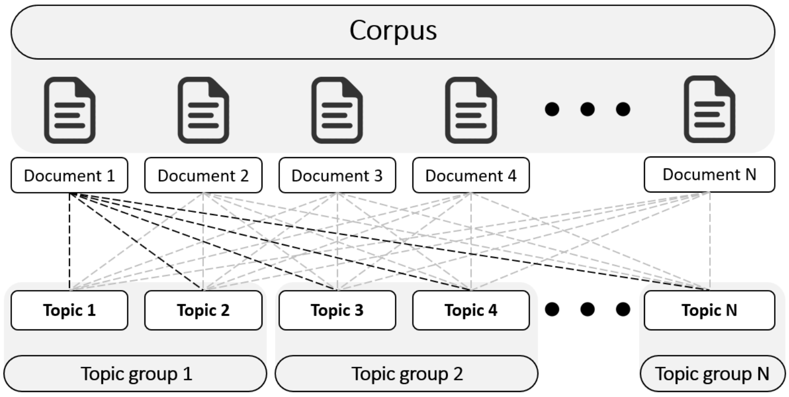

Figure 3.2: Nomenclature

Then, all entries with incomplete data were removed. This resulted in the exclusion of 9,018 scientific articles, leaving us with 25,121. No documents related to the EU-funded projects had to be excluded since all of the relevant fields were populated.

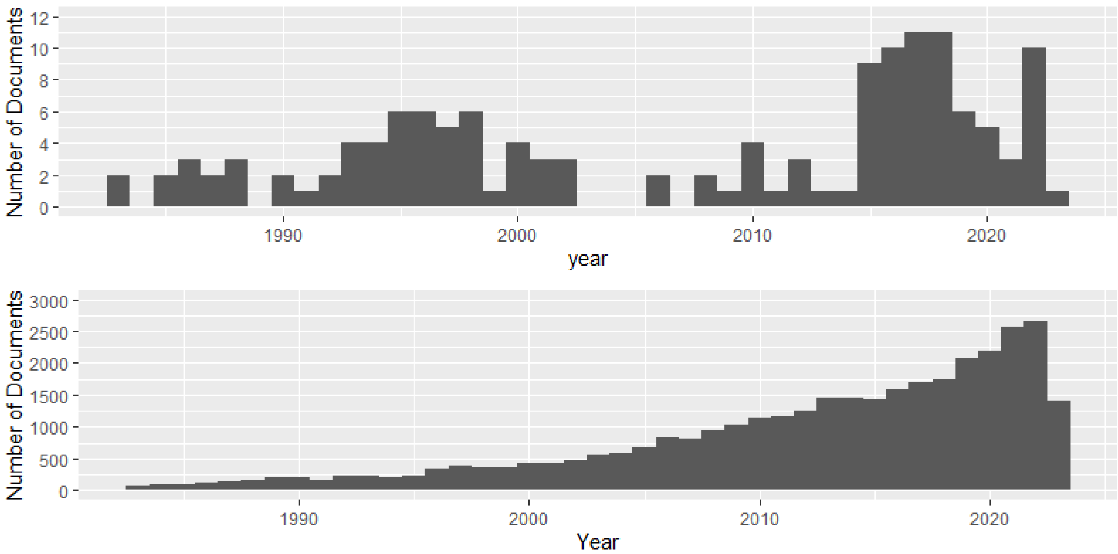

Figure 3.3: Number of documents related to MSW per year. The top graph corresponds to scientific community articles, and the bottom graph to the EU-funded projects.

3.4.2 Scientific articles

The first corpus analyzed was the one related to the abstracts of the extracted scientific articles. The corpus was investigated, presenting its relevant characteristics, followed by the development of the LDA model to extract its most prominent topics.

3.4.2.1 Frequency analysis

To improve interpretability, the lemma of the words was used instead of the stem. In other words, the base or dictionary form of the words was used instead of the part of the words that remains after removing any prefixes or suffixes. For instance, the stem form of the words *waste*, *wasting*, and *wasted* is "wast". However, this stem form is a word that does not exist; hence, it was converted to its *waste* lemma. This process is called lemmatization.

The most frequent words and the number of times they appear are presented in Figure 3.4. Although the keyword used to constitute the corpus was "municipal solid waste" (which was then replaced with *msw*), this is only the second most frequent word.

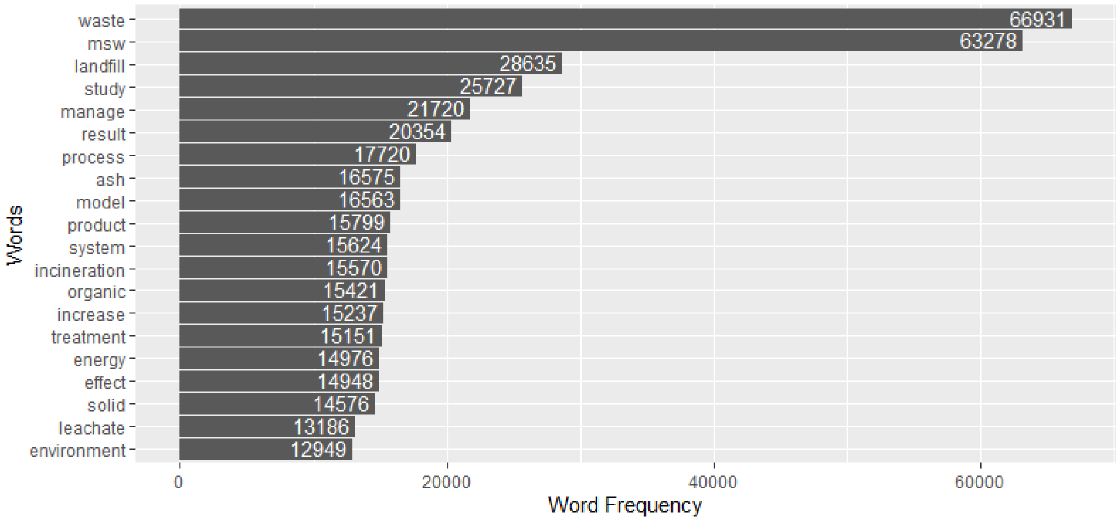

Figure 3.4: Frequency of the 20 most frequent words adopted by the scientific community

The word that appears the most is *waste*. Its stem form can come from the words "waste", "wasted", "wasting", and "wastes". This was to be expected, as "waste" can be used as a noun (this happens commonly in different contexts such as food waste), as an adjective, or even as a verb. Hence, the versatility of this stem form in the context of MSW justifies the frequency of the word, which appears almost 67,000 times throughout the 34,139 extracted abstracts.

The second most frequent word (*msw*) appears 63,278 times, while the third (*landfill*) only appears 29,000 times. This is a very significant drop in frequency (over 50%). The frequency drop for the next words is not as severe. For instance, the 20^{th} most frequent word (*environment*) appears almost 13,000 times.

Out of the 63,621 unique words present in the corpus, 29,156 only appear once. This means that around 46% of the words were not relevant to this analysis. Considering solely words that appear 100 or fewer times in all abstracts as a whole, the number of relevant unique words decreases to 2,800.

The 50 most frequent words are shown in a word cloud in 3.5. In this word cloud, words are grouped by their frequency level, with each group having its own color. The middle of this word cloud is populated by the words *waste* and *msw*, followed by *study* and *landfill*. Other less frequent words are *anaerobic*, *paper*, and *disposal*.

Figure 3.5: Word cloud of the 50 most frequent words adopted by the scientific community

3.4.2.2 Number of topics

Different metrics were adopted to determine how many topics should be extracted from the corpus, but first, a range of possible topics had to be defined and the corresponding LDA models generated. A range of 2 to 100 topics was considered, meaning 99 different LDA models were generated and compared according to the four different metrics.

Figure 3.6 shows the disparity in the number of topics suggested by each metric. Specifically, Deveaud2014 and Griffiths2004 suggested the adoption of 30 and 99 topics, respectively, while both CaoJuan2009 and Arun2010 suggested 95 topics.

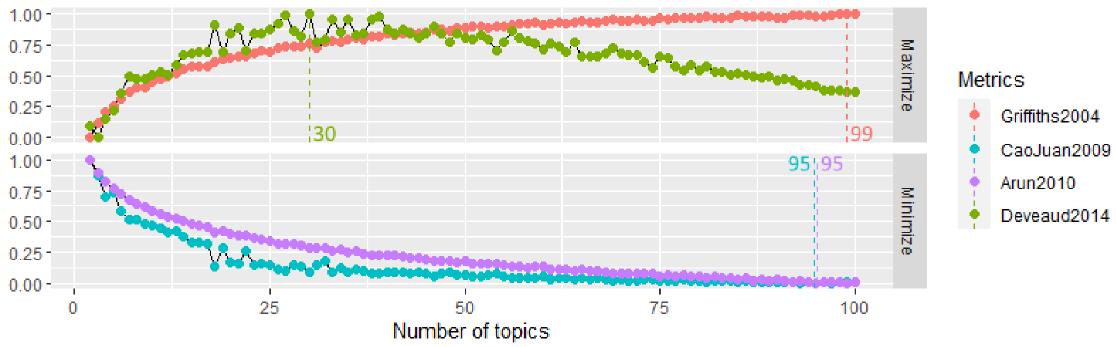

Figure 3.6: Number of topics to extract from the scientific community corpus

Nevertheless, interpretability and readability are essential features when analyzing LDA results. Thus, defining fewer topics should enable a better and easier presentation of the available data (Chu et al., 2020; Debnath and Bardhan, 2020; Buenano-Fernandez et al., 2020). In this sense, the number of topics suggested by the Deveaud2014 metric was the most appropriate. This metric maximizes the information divergence of all topic pairs or, in other words, maximizes differences between the topics (Sutherland and Kiatkawsin, 2020) and is commonly used to determine the number of topics in other LDA applications (Debnath and Bardhan, 2020; Buenano-Fernandez et al., 2020; Sutherland and Kiatkawsin, 2020).

Additionally, the other metrics presented a constant increase or decrease regarding the number of topics, which may indicate that their outcomes are uninformative (Hussain et al., 2023). Although this number could stabilize if the range of topics tested was wider, a higher number of topics would jeopardize interpretability. Hence, the LDA model was developed considering a total of 30 topics.

3.4.2.3 Topics

The final model offered a panoply of information to understand the most relevant themes on the subject of MSW. For example, it revealed the most relevant words for each topic, making it easier to discern precisely what topic they related to, since the LDA model by itself did not name topics.

Although naming 30 different topics is an excellent start to understanding the academic themes around MSW, especially considering that this output originated from over 25,000 distinct articles, it is still a high number to grasp simultaneously.

Therefore, the topics were unified into broader groups (topic groups). This was achieved by examining the most frequent words of each topic and grouping those that were alike. These topic groups were also named to make it easier for readers to understand the MSW dimensions discussed in the present work.

Table 3.1 shows the 30 referenced topics and their corresponding topic group. Additionally, the 10 most relevant words for each topic are presented. These words justify the topic denomination and, therefore, the name of the corresponding topic group.

Table 3.2 presents the frequency of the topics, along with the frequency of each topic group. Topic *1. MSW characterization and analysis* is the most popular topic, as it is the topic that has a greater presence among all the topics, while the topic *7. Health risks and population demographic characteristics* is the least popular when considering the total frequency of the words included in the topics.

The most common topic group on the subject of MSW is *G) Environmental impact assessment and sustainability*, which is to be expected, since it is one of the topic groups with more topics (five). The prevalence of *G) Environmental impact assessment and sustainability* suggests a strong emphasis on the impact of MSW in cities and on efforts to minimize its impact/reuse it in a sustainable manner. At the other end of the spectrum, *E) Leaching and extraction of heavy metals* is the least popular topic group, possibly because there are only two topics.

One of the most popular topics - *22. Site selection methods and criteria for MSW facilities* - is also part of one of the least popular topic groups - *F) Optimization and site selection for MSW systems*. Another similarly popular topic is *9. Usage of ash and slag from incineration in cement and concrete*, which focuses on converting MSW, followed by *15. Anaerobic digestion of the organic fraction of MSW and biogas production*. On the other hand, the least frequent topics are *9. Microbial and bacterial diversity and resistance* and *7. Health risks and population demographics*.

The evolution of the topic group frequency of the scientific article corpus over the years was also analyzed, as graphed in Figure 3.7. The graph shows that the popularity of the topic groups is volatile at the start, but becomes more stable as time passes. The topic group with the highest volatility is *G) Environmental impact assessment and sustainability*. It has a frequency of over 25% in the first years, which then drops by more than half, and increases slightly afterward. The opposite is true for topic group *B) Pollution and health risks*, which has a big increase in frequency during the first years before starting to lose popularity.

Overall, the most stable topic groups are *A) MSW properties* and *E) Leaching and extraction of heavy metals*. The former is less frequent in the first decade but then gains popularity. This is relevant as research and exploration of MSW properties are at the core of the topic group itself. On the other hand, the increase in popularity of topic group *4. Biological treatment processes* is practically constant over time.

Table 3.1: Topic summary for the scientific community

Topic group name	Topic name	Top 10 words
A) MSW properties	1. MSW characterization and analysis	msw; china; study; reserve; analyze; paper; effect; method; main; disposal
	2. Statistical analysis of factors (Academic research/Research and analysis)	analysis; study; factor; indication; result; influence; level; data; significance; author
	3. Characteristics of neutrino and its effects	neutrino; effect; mass; dive; measure; magnet; observe; time; perform; frequency
	4. Physical characterization and moisture content of MSW fractions	content; composition; value; msw; sample; particle; characteristic; size; dry; moisture
B) Pollution and health risks	5. Water quality and pollution	water; pollution; contamination; quality; site; groundwater; source; dump; season; surface
	6. Air pollution from the incineration of organic compounds	concentration; incineration; sample; compound; dioxin; air; pcddfs; pollution; level; source
	7. Health risks and population demographics	risk; health; human; study; association; exposure; assess; report; worker; sex
	8. Control and emissions of greenhouse gas from MSW incineration	gas; emission; greenhouse; incineration; carbon; combust; reduction; air; msw; reduce
	9. Microbial and bacterial diversity and resistance	active; microbiology; community; bacteria; resist; isolation; divers; study; abundant; gene
C) MSW and biomass conversion	10. Usage of ash and slag from incineration in cement and concrete	ash; fly; incineration; bottom; cement; material; strength; concrete; property; slag
	11. Waste sorting and recycling of materials	waste; recycle; material; plastic; separation; food; paper; recovery; product; sort
	12. Thermal conversion of MSW and biomass into fuel and other products	fuel; product; process; biomass; gasification; pyrolysis; heat; thermal; produce; combust
	13. Waste-to-energy technologies for energy generation (electricity and heat)	energy; plant; generate; electric; power; potential; recovery; technology; economic; renew
D) Biological treatment processes	14. Biological treatment of municipal wastewater and sludge	treatment; sludge; remove; wastewater; municipal; sewage; process; code; treat; efficiency
	15. Aerobic composting of the organic fraction of MSW	organic; compost; process; degradable; matter; biodegradable; during; biology; fraction; aerobic
	16. Bioreactor landfill operation and leachate recirculation	landfill; leachate; msw; bioreactor; recirculate; operate; study; cover; sanitary; refuse
	17. Anaerobic digestion of the organic fraction of MSW and biogas production	anaerobic; digest; product; organic; biogas; methane; fraction; ofmsw; pretreat; reactor
E) Leaching and extraction of heavy metals	18. Adsorption and oxidation of chlorine by biochar and other carbon-based materials	carbon; active; adsorption; reaction; effect; oxidation; surface; mechanic; biochar; format
	19. Leaching and extraction of heavy metals from residues	metal; heavy; leach; concentrate; extract; residual; toxic; element; lead; acid
F) Optimization and site selection for MSW systems	20. Optimization of waste collection and transportation systems	system; collect; optimal; cost; manage; design; transport; propose; waste; facility
	21. Data-based modeling and simulation of MSW processes	model; predict; data; simulation; base; parameter; estimation; develop; propose; network
	22. Site selection methods and criteria for MSW facilities	method; select; site; study; evaluate; suitable; process; analysis; base; decision
G) Environmental impact assessment and sustainability	23. Research of social practices and education in waste management	social; practice; student; education; research; program; field; knowledge; learn; experience
	24. Life cycle assessment and environmental impact of MSW	environment; impact; assess; scenario; cycle; life; manage; result; study; treatment
	25. Sustainable waste management policies and practices in developing countries	manage; municipal; policy; develop; waste; country; local; sustain; implement; region
	26. Generation and management of MSW in urban areas (cities)	waste; solid; generate; manage; dispose; city; urban; household; municipal; population
	27. Circular economy technologies for industrial waste management and environmental sustainability	technology; develop; sustain; industry; review; research; resource; environmental; discuss; application
H) Effects of waste treatment on the soil and plants	28. Properties and experimentation of liners for landfills	test; conduct; pressure; shear; msw; strength; liner; compress; property; settlement
	29. Effects of compost and organic fertilizers on the soil and plant growth	soil; compost; plant; organic; application; fertile; amend; agriculture; growth; increase
	30. Effects of temperature and additives on MSW degradation rate	increase; rate; effect; temperature; decrease; time; during; ratio; result; condition

Table 3.2: Topic frequency for the scientific community

Topic #	Individual topic frequency	Topic group frequency
1.	3.42%	12.86%
2.	3.13%	
3.	3.13%	
4.	3.18%	
5.	3.35%	16.30%
6.	3.89%	
7.	2.81%	
8.	3.11%	
9.	3.14%	
10.	2.84%	13.11%
11.	3.49%	
12.	3.25%	
13.	3.53%	
14.	3.35%	12.87%
15.	3.08%	
16.	3.12%	
17.	3.33%	
18.	3.30%	6.14%
19.	2.83%	
20.	3.17%	10.57%
21.	3.42%	
22.	3.97%	
23.	4.09%	17.31%
24.	3.17%	
25.	3.60%	
26.	3.33%	
27.	3.13%	
28.	3.80%	10.84%
29.	3.79%	
30.	3.26%	

3.4.3 European Union

This section focuses on the EU-funded projects. As in the previous section, an overview of the corpus extracted from the CORDIS website is provided before presenting detailed information about the creation of the LDA model.

3.4.3.1 Frequency analysis

Considering the corpus as a whole, the number of times any given word appears in a text may be indicative of its content. Thus, the 20 most frequently used words were examined using their respective lemma instead of their stem. As with the scientific articles, *msw* is only the second most used word in the EU corpus, even though it was the keyword used for the extraction. *Waste* is yet again the most used word, as seen in Figure 3.8.

The difference between the word frequencies was not as significant as the one verified in the previous analysis, which could be due to the size difference between both corpora. In this case, there are 3,586 unique words, compared to the 63,621 unique words present in the scientific articles. An interesting observation is the fact that *product* appears in third place in regards to term frequency, just barely behind *msw*. Other relevant words depicted in the graph are commonly related to energy, such as *energy*, *industry*, *fuel*, and *gas*. Additionally, words like *technology* and *plant* may also be related to the concept of energy, depending on the context.

Figure 3.9 draws a word cloud of the 50 most frequent words. As seen before, the words in the middle are *waste*, *msw*, and *product*. Then, the cloud shows the words *process* and *project*, along with words such as *system* and *develop*. Other examples of less common words found in the cloud are *increase*, *aim*, and *fraction*. The appearance of these words makes sense in the context of the EU-funded projects, as they can be used to describe project information.

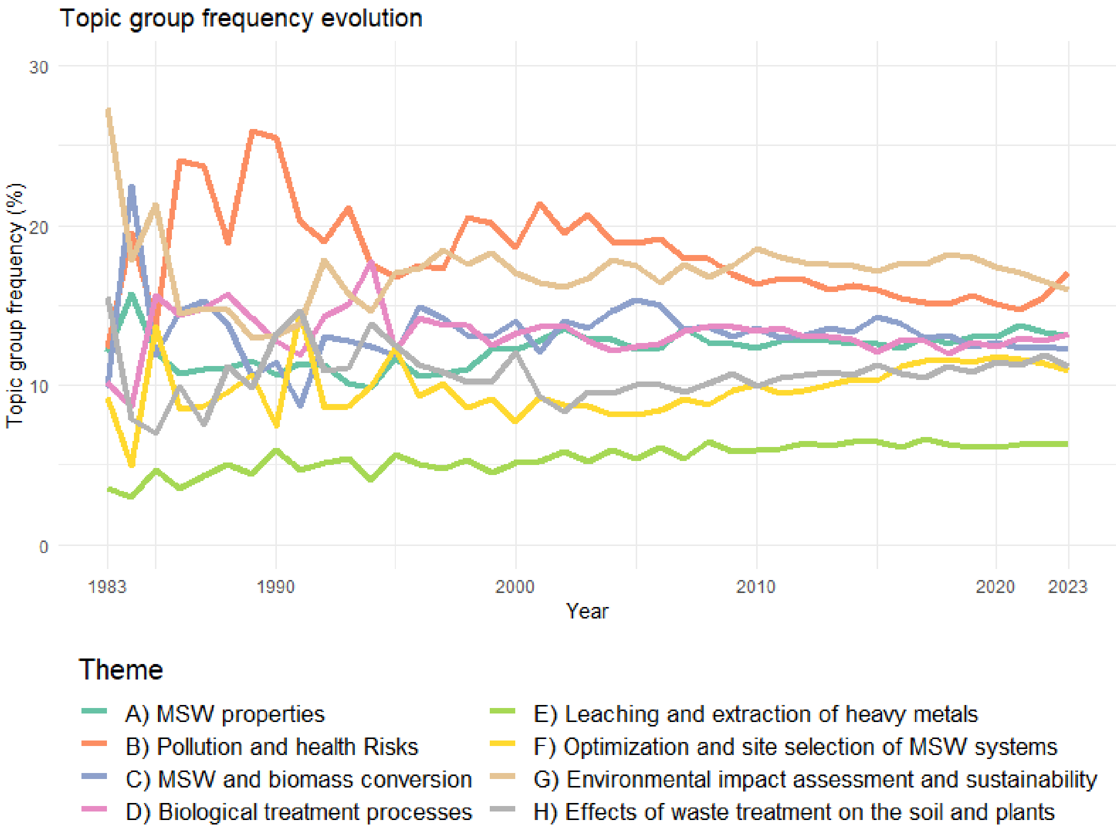

Figure 3.7: Topic group frequency evolution in the scientific community

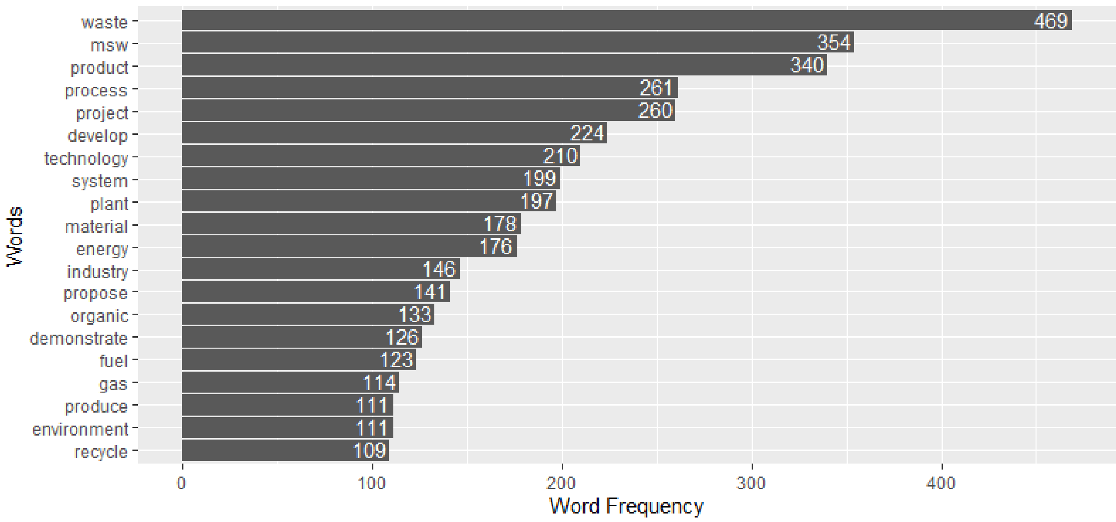

Figure 3.8: Frequency of the 20 most frequent words adopted by the European Union-funded projects

Figure 3.9: Word cloud of the 50 most frequent words adopted by the European Union-funded projects

3.4.3.2 Number of topics

For the EU corpus, the number of topics that should be extracted was determined in a similar way to that described above, relying on different metrics, i.e., Deveaud2014, Griffiths2004, Cao-Juan2009, and Arun2010. The range of possible topics was also set to be between 2 and 100 and standardized the values for each metric to be between 0 and 1. Figure 3.10 shows these metrics.

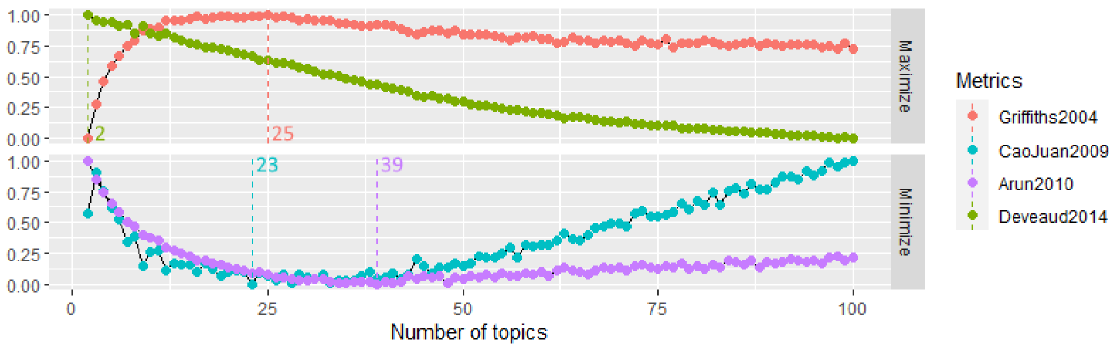

Figure 3.10: Number of topics to extract from the European Union-funded projects

In this context, the previously prioritized metric, i.e., Deveaud2014, suggested only two topics, which would not extract much information from the corpus in question. Thus, other metrics were considered. Arun2010 suggested a high number of topics (39), which could harm the interpretability and readability of the results. Hence, the most sensible option was to follow either Griffiths2004 or CaoJuan2009, which suggested 25 and 23 topics, respectively. In the end, the number of topics was set to 25 since this number is closer to the one adopted in the first LDA model, allowing for a better comparison between the two.

In opposition to the first LDA model, most metrics in this one presented either a valley (CaoJuan2009 and Arun2010) or a peak (Griffiths2004) instead of a constant increase or decrease. The only exception was Deveaud2014, which had a constant decrease, whereas in the previous analysis, it was the only metric that did not present this behavior. This further suggests that this metric was not adequate for this context.

3.4.3.3 Topics

A total of 25 topics were extracted from the EU corpus by employing the LDA model. Although these are not as many as the ones identified in the scientific article corpus, the topics are still sufficiently different between themselves. These topics were also named, and there was an attempt to keep the same nomenclature adopted when analyzing the scientific articles to allow a good comparison with the topics of the previous corpus.

The topics were also unified into groups to facilitate their analysis and subsequent overview. The topics, their corresponding groups, and their most frequent words are represented in Table 3.3.

Table 3.3: Topic summary for European Union-funded projects

Topic group name	Topic name	Top 10 words
A) Energy conversion	1. Conversion of biomass into energy, fuel, and other products	plant; energy; biomass; demonstrate; efficient; product; fuel; project; design; result
	2. Anaerobic digestion of agricultural waste	agriculture; reactor; anaerobic; effluent; carbon; concentrate; residual; cellulosome; optimal; carry
	3. Energy generation from MSW	plant; energy; msw; fuel; heat; electric; combust; boiler; steam; generate
	4. Landfill value extraction	landfill; system; product; increase; content; include; sever; gas; measure; extract
B) Value extraction and waste valorization	5. Waste treatment and value extraction technologies	technology; innovate; advance; research; treatment; develop; value; project; joint; universe
	6. Recycling and product quality	recycle; cost; materi; develop; recovery; separate; increase; product; rate; stream
	7. Value extraction from MSW	product; propose; organic; msw; fraction; value; feedstock; technology; chain; market
	8. Plastic recycling technologies	plastic; technology; industry; economy; europe; achieve; large; repair; advantage; circular
C) Biological treatment processes	9. Biological waste management and innovation	waste; urban; biobase; stream; include; biowaste; process; valorise; wastewater; innovate
	10. Organic waste treatment processes	process; biogas; treatment; organic; fraction; digest; sort; separate; compost; plant
D) Waste management processes	11. Industrial processes for MSW management	process; product; industry; msw; integrate; acid; cost; yield; chemical; sugar
	12. Urban waste collection management	project; truck; fuel; urban; implement; technology; require; city; cell; operate
	13. Waste collection and disposal	waste; collect; active; addition; direct; dispose; amount; base; involve; municipal
	14. Industrial waste management	material; industry; waste; raw; sector; tonne; million; construct; produce; billion
E) Projects and programs	15. Waste management project proposal	project; waste; develop; manage; region; european; propose; application; solution; contribute
	16. Municipal programs	local; demonstrate; object; plan; municipal; author; programm; msw; relate; support
F) MSW characterization and policies	17. MSW composition and material handling	material; ash; msw; glass; recycle; incinerate; obtain; environment; process; composition
	18. Scientific research and policy development	research; network; h2020; assess; scientific; train; institute; model; mine; approach
	19. Public environmental policies	public; environment; effect; tax; achieve; include; energy; wide; spain; handbook
G) Sustainability and pollution	20. Sustainable water management systems	water; sustain; system; integrate; economy; approach; solution; manage; set; social
	21. Air pollution reduction	reduce; air; bubble; significance; produce; because; commercial; stage; aim; practice
	22. Circular economy awareness	circular; service; environment; provide; eufund; develop; product; reuse; platform; social
H) Practical experimentation test systems	23. Hydrogen experimentations	dive; msw; diver; decompress; hydrogen; depth; phase; pressure; during; experience
	24. Clean gas and combustion systems	gas; process; system; msw; clean; test; gasification; filter; coal; engine
	25. Welding and equipment performance	weld; system; develop; test; control; technique; wet; equip; improve; quality

Apart from the words akin to *waste* and *msw* that appear frequently in the corpus, other words such as *product*, *develop*, and *project* are also trendy.

Regarding topic frequency, there seems to be an even distribution, as depicted in Table 3.4. The only exception is topic *18. Scientific research and policy development*, which has a frequency of over 7%, while the second most frequent topic (*12. Urban waste collection management*) has a frequency of 5.38%. It is clear that a large part of the EU corpus deals with policy development and research, as well as issues related to MSW recovery and valorization. On the other spectrum, topic *21. Air pollution reduction* is the least significant.

The group frequency is also relatively stable, even when considering size discrepancies. However, topic group *D) Waste management processes* seems to be the one that gets the most attention from EU projects, closely behind *A) Energy conversion* and *B) Value extraction and waste valorization*. These are topic groups wherein the three most popular topics also reside, revealing that waste recovery and valorization are at the core of the EU-funded policies.

Table 3.4: Topic frequency for European Union-funded projects

Topic #	Individual topic frequency	Topic group frequency
1.	5.15%	15.66%
2.	2.78%	
3.	3.43%	
4.	4.31%	
5.	3.71%	15.95%
6.	4.34%	
7.	3.98%	
8.	3.91%	
9.	4.45%	7.81%
10.	3.35%	
11.	3.57%	16.46%
12.	5.38%	
13.	3.62%	
14.	3.89%	
15.	3.24%	7.05%
16.	3.81%	
17.	3.71%	14.33%
18.	7.57%	
19.	3.05%	
20.	5.23%	11.66%
21.	2.39%	
22.	4.04%	
23.	4.13%	11.08%
24.	3.18%	
25.	3.77%	

Within topic group *D) Waste management processes*, four different topics focus on the collection and disposal of waste as a whole while also having a strong industry focus. As far as the other two major topic groups are concerned, there is a strong emphasis on trying to reuse and draw some value from produced waste, either by recycling it or by producing energy from biomass.

Topic group *C) Biological waste management and innovation* is one of the least popular, probably due to having only two topics. Another unpopular topic group is *D) Waste management processes*, which deals with treating biological waste when it occurs and cannot be converted into energy. Still, the least popular topic group is *E) Projects and programs*, which mostly revolves around the call to action present in every project description.

There is also a considerable focus on sustainability and pollution issues, which demonstrates that the aim and priority of these EU projects is to address both the current and the future environmental impact of MSW in Europe.

However, looking at the topic frequency throughout the years, each topic group has a radically different importance. However, it is important to analyze this information critically, in the sense that the corpus presents an average of three to four projects per year. In other words, certain topics might have been addressed at very specific times.

Regardless, looking at Figure 3.11, the prevalence of each group is very inconsistent. The trend throughout the years has been for a specific topic group to represent more than 25% of all groups, while all the other topic groups share similar values. For example, most of the corpus from 1983 refers to the topic group *H) Practical experimentation test systems*, but afterward, this topic group quickly returns to a frequency more closely related to its counterparts. On the other hand, *D) Waste management processes* is the most relevant topic group from 1990 to 1993 and again from 1996 to 1999.

However, in the last two decades, the tendency seems to be for *A) Energy conversion* and *F) MSW characterization and policies* to gain traction and steal the spotlight. The only exception can be found in 2009, when topic group *B) Value extraction and waste valorization* has a frequency greater than 50%.

Overall, topic group *C) Biological treatment processes* is the most stable throughout the years. On the other hand, *D) Waste management processes* is the most unpredictable.

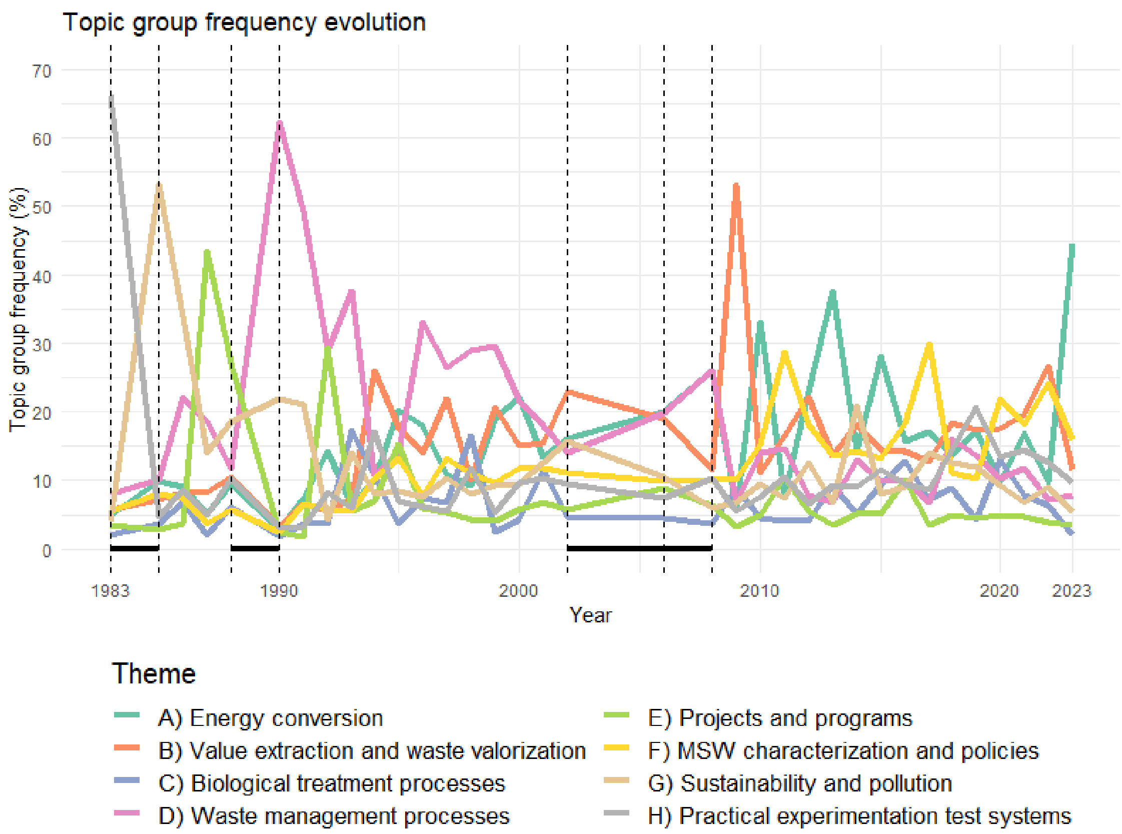

Figure 3.11: Topic group frequency evolution in the European Union-funded projects

3.4.4 Similarity estimation

The cosine measure was used to quantify the maximum similarity observed in the documents of both corpora (scientific community and EU-funded projects).

Figure 3.12 presents the results, showcasing two notable characteristics: the difference in the maximum similarity between the two corpora and the number of outliers in both plots.

The results show a significant discrepancy in the similarity levels between the two corpora. The quartile values suggest that the maximum cosine similarity between the scientific articles and the descriptions of the EU-funded projects tends to be relatively low, with most comparisons yielding values below 0.1562. This indicates a moderate degree of similarity, and that most pairs of documents have a moderate degree of commonality or thematic relevance.

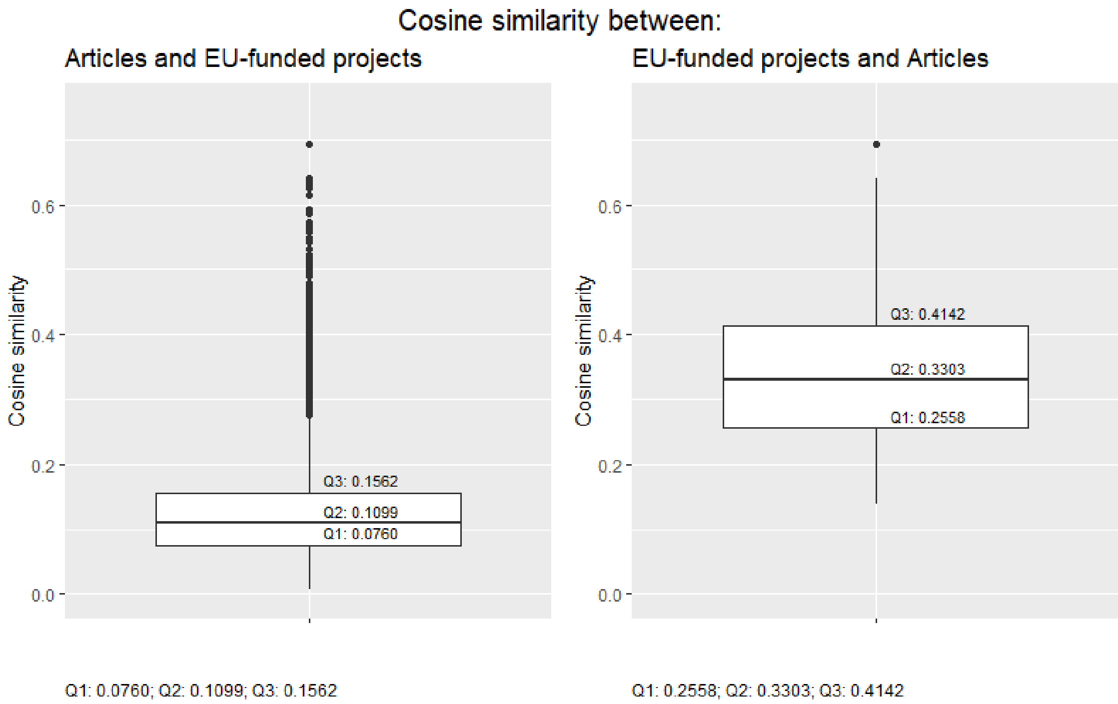

Figure 3.12: Cosine similarity between the two corpora

However, when examining the maximum cosine similarity obtained from the description of the EU-funded projects in relation to the scientific articles, as seen in the right-hand side of Figure 3.12, the quartile values increase significantly, with most comparisons yielding values below 0.4142. This indicates a comparatively higher degree of similarity between scientific articles and the EU-funded projects, suggesting that a significant proportion of document pairs have a high degree of overlap in thematic content across a significant portion of document pairs.

In summary, the observed values indicate a moderate degree of similarity between the scientific articles and the EU-funded projects, but a significantly higher degree of similarity in the opposite direction.

To complete this analysis, the similarity between documents of the same corpus were also examined. Figure 3.13 draws two boxplots, one for each case. The left boxplot corresponds to the maximum similarity between scientific articles and the right boxplot corresponds to the similarity between the EU-funded projects. In this regard, the scientific articles are more similar between themselves than the EU-funded projects. It should also be noted that the similarity between the scientific articles has more outliers, with some of them being close to 1, indicating very similar works, while the outliers for the EU-funded projects are almost non-existent.

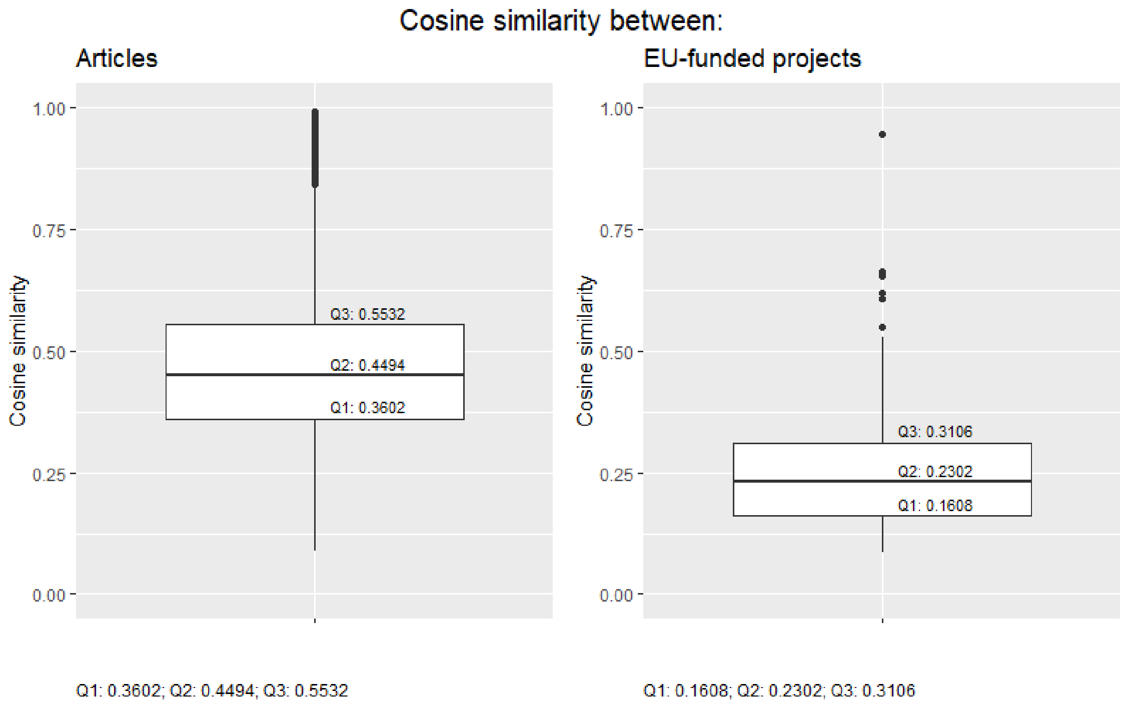

Figure 3.13: Cosine similarity between the two corpora

3.5 Discussion

The two analyzed corpora have different inherent characteristics. For instance, their size is very distinct, with the corpus of the scientific articles having a total of 3,233,738 words, while the corpus of the EU-funded projects only has 26,823. Not only that, but the scientific article corpus, likely due to its size, has 63,621 unique words, which is not only more than the 3,586 unique words in the corpus of the EU-funded projects but also more than its total number of words.

While 140 EU-funded projects are a considerable number, they are shrouded by the enormous number of scientific articles in the literature. This leads to two issues. First, the scientific community has a bigger portfolio of documents and thus a better opportunity to expand on all topic groups related to MSW. Second, because the EU-funded projects have a reduced portfolio, disparities between topics are more noticeable.

Nonetheless, when comparing the word frequency in both corpora, they seem to be somewhat aligned. The 20 most frequently used words in both cases seem to relate to the same problems, even if they are not an exact match. The top words present in the abstracts of the scinetific articles are usually not related between themselves, suggesting that several themes of MSW are being considered simultaneously. Conversely, the key words in the EU-funded projects are more closely related to more practical applications, which was to be expected.

The scientific side of MSW is very broad. This is demonstrated by the panoply of words used and by each topic having a different set of 10 most representative terms. In fact, the same top-10 term rarely appears in more than one topic. The works in the field of MSW cover the subject from different perspectives, which are linked to different themes. There are works on analyzing and understanding the composition of MSW, repurposing waste for value extraction and energy generation, determining optimal waste management approaches, and evaluating the impact of MSW on the economy, environment, and society.

The MSW field is extremely mature in the literature, with a deep understanding of its nuances even when considering that MSW characteristics differ from city to city. There is also a big concern with figuring out practical applications to reduce, reuse, recycle, recover, and dispose of existing waste. Nevertheless, the covered themes have been relatively stable throughout the years. In other words, the literature seems to keep complementing each theme, even if at times some nuances are more popular than others.

Regarding the EU-funded projects on MSW, it is hard to find a pattern. The projects and programs are clearly aimed at motivating countries to do better and fight against the consequences of MSW. It is therefore natural that, over the years, more emphasis has been placed on the present and/or near-future problems associated with MSW. This priority can be seen in the oscillation in the frequency of topics over the years. There is less focus on the characterization of MSW and its composition, and more on its management and consequences. These consequences are linked to sustainability issues, and several projects have explored how the ideology of the circular economy is an asset to foster sustainability.

Although there is a hierarchy in the options available to reduce the consequences of MSW - reduction, reuse, recycling, recovery, and disposal (Papargyropoulou et al., 2014) -, the EU-funded projects have primarily focused on the last four, in four different ways. First, by trying to extract value from existing waste and recycling it whenever possible. Second, by recovering biowaste and trying to convert it into energy. Third, by treating and dealing with existing waste at the waste management level, including biological waste. Fourth, by experimenting with new methods and systems that are more efficient, produce less waste, and may also be less polluting.

Nonetheless, there is also a great concern regarding setting policies to enable the above options. One of the 25 extracted topics deals with proposing waste management plans and another deals with municipal programs. Together, these two topics have a frequency of 7.81%.

In terms of direct comparisons and contrasts between the two corpora explored in this chapter, some aspects stand out. For instance, the EU-funded project corpus is very focused on everything related to biological waste and waste management, such as waste collection and disposal, but not as focused on other areas of MSW that the scientific community has studied and documented in detail. Considering that EU-funded projects focus on contemporary issues, there is an indication that biowaste and waste management in European cities are two very concerning issues.

The analysis of the cosine similarity between the two corpora sheds light on the similarities and differences between the two. Through this measure, it is possible to verify that the EU-funded projects seem to be much more closely related to the works of the scientific community than the other way around. In particular, there are cases where the similarity between the scientific articles and the EU-funded projects is negligible or even almost 0, while this does not happen the other way around.

The differences in similarity may stem from different factors. The scientific community may cover a wider range of topics compared to the EU-funded projects, leading to more varied content and potentially lower similarity. Another reason for this may be that the EU-funded projects are more focused or specific in nature, causing higher similarity values due to having shared themes or characteristics. However, it is also true that the scientific articles are more similar between themselves than the EU-funded projects. This may be due to the large disparity in the sample size of both corpora, i.e., 25,121 documents against 140 documents, with the EU-funded projects displaying more concentrated and unique patterns.

Another reason for these discrepancies may be that the EU-funded projects rely either on the work of the scientific community or on information provided by organizations such as FAO and the World Bank. This way, policymakers can rely on already available information and focus their attention on practical solutions, while the scientific community seeks to generate its own theoretical background.

3.6 Conclusion

The topic of MSW has gained traction over the years. Considering the amount of MSW generated every year, it is easy to understand how severe its consequences are and why the scientific community has focused so much on this topic. In fact, this is a very mature branch of the literature on waste, and the characterization and composition of MSW, its consequences, and how it can be dealt with have been well explored.

However, to understand if this focus of the scientific community is matched by actual actions to minimize the consequences of MSW, a text mining approach was employed to compare what has been addressed by the scientific community with what has been addressed by EU-funded projects. For this purpose, the corpora representing each sphere were compared, their word frequency was analyzed, and then the relevant topics were extracted from the content of the documents by creating and using two LDA models.

The LDA models gave an overview of the major themes explored in the two settings mentioned above. On the one hand, the scientific body of work surrounding the topic of MSW is, as expected, extremely mature, and there is a solid understanding of all aspects of MSW. Despite the growth in the number of publications throughout the years, the proportion of studies devoted to each topic has remained constant, with no one topic dominating the others. On the other hand, the EU-funded projects focus heavily on some specific subjects, neglecting others.

The EU-funded project corpus barely mentions waste reduction options but focuses heavily on waste reuse, recycling, recovery, and disposal. In the EU-funded projects, the most common themes are biowaste, with a special focus on procedures that enable its conversion into energy, and waste management, which includes the collection and disposal of waste. At the same time, the frequency with which the topics are covered changes from year to year. This supports the view that the projects are designed to solve ever-changing problems with present or near-future consequences.

This information suggests that despite their flexibility and adaptability to the current reality, the EU-funded projects focus mostly on dealing with existing MSW and not much on reducing MSW generation. This leads to other issues such as landfills and the effects of MSW and waste-based fertilizers on the soil and plants, which have been extensively studied in the literature but have not received much attention in the EU-funded projects. Although the EU-funded projects do mention other topics such as the characterization of MSW, its overall consequences, and sustainability-related issues, the scientific community seems to have a much better understanding of these topics.

To complement this information, a cosine similarity analysis was conducted to measure the similarity between the two corpora. This analysis revealed that the scientific articles have a higher internal similarity than the EU-funded projects. In addition, the EU-funded projects are more similar to the scientific articles than the other way around. This difference between the two corpora may be due to the broader range of topics explored by the scientific community and the more focused nature of the EU-funded projects.

In summary, the two corpora appear to be mostly aligned. Regardless, the scientific community has a larger portfolio of information, which provides a much broader understanding and knowledge of the MSW issue. The EU-funded projects, on the other hand, do not focus so much on the characterization and consequences of MSW, but rather on what are the best options to deal with current and near-future problems. Nevertheless, there is a clear trend toward biowaste and waste management solutions.

One of the limitations of this work was the small number of documents related to EU-funded projects in comparison to the scientific articles, making it more difficult to draw conclusions about the latter. Although this is not as noticeable when using an LDA analysis, this reduced number of documents became more apparent when calculating the similarity between the two corpora. Thus, it could be interesting for future research to test quantitative tools other than LDA and the cosine similarity to extract information from smaller corpora, or to collect a larger corpus with a higher number of documents representing the policies adopted at the EU level. Future research could also study other features from both corpora, such as the country of origin of the participants or researchers and their number.

CHAPTER 4

A Literature Review on the Quantitative Approaches to Food Waste: Descriptive, Predictive, and Prescriptive Analyses

Food waste generated throughout the food supply chain raises several environmental, social, and economic issues. Quantitative methods can aid in managing food waste by describing current contexts, predicting future scenarios, and improving related operations. However, a literature review on the use of quantitative methods, namely on the descriptive, predictive, and prescriptive dimensions, to assess and prevent food waste is lacking. This chapter aims to explore and categorize the quantitative studies that perform descriptive, predictive, and prescriptive analysis concerning food waste in order to identify gaps and future research directions. For this purpose, we developed a systematic literature review based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement methodology, supported by 65 papers. We identified the key features of each data analytics approach, with a particular focus on i) food waste quantification methods, ii) demand, food waste, and shelf-life forecasting algorithms, and iii) optimization approaches. Additionally, the context in which each of these studies are focused is also explored. We found that predictive analysis is the most prominent among the data analytics approaches, followed by descriptive and prescriptive systems, respectively. Moreover, the setting more explored is the hospitality sector, being the only context in which all descriptive, predictive, and prescriptive approaches can be found. The algorithms and models adopted in the studies vary, and there is still room for adopting more recent or advanced methods. We believe this chapter can contribute to setting the ground for more oriented and structured quantitative research in the food waste field.

4.1 Introduction

The food supply chain (FSC) is the culmination of several processes dealing with the production and supply of food and involves all the distinct stages necessary for food to reach its final consumer. Food waste (FW) and food loss (FL) can be generated at all stages of the FSC Li et al. (2022). If the waste is produced at the start of the FSC, i.e. at the production, postharvest, and processing stages, it is usually called FL (Feukam Nzudie et al., 2020). If the waste happens at the end of the FSC, i.e. at the retail and consumer stages, it is deemed FW Thyberg and Tonjes (2016); Do et al. (2021). Despite this distinction, some studies use these two concepts as synonyms (Thyberg and Tonjes, 2016; Girotto et al., 2015). In fact, this distinction can be ambiguous, especially when defining the waste generated in the middle stages of the FSC. Nevertheless, most of the wasted food is generated either at the start of the FSC (agriculture) or at the end (consumption), with the sum of both accounting for more than half of the waste in the whole FSC (Priefer et al., 2016; Salina Md Salim et al., 2016). In the present work, we distinguish between FL and FW and use the term 'wasted food' as a broader concept to refer to all the waste produced in the human-oriented FSC. In other words, we consider wasted food to be the combination of FL and FW.

All types of wasted food have an impact on the environmental, social, and economic spheres through the production of greenhouse gases, problems associated with global food security, and the costs related to wasted food (Gao et al., 2021; Shams et al., 2017; Beheshti et al., 2022; Masud et al., 2020). Despite this fact, wasted food is not usually addressed as a multi-dimensional dilemma, given that looking at it as an environmental and sustainability problem is entirely different from looking at it as an economic and management issue (Filimonau and De Coteau, 2019).

The same amount of wasted food can have different impacts at different stages of the FSC (Priefer et al., 2016). The later the wasted food arises, the higher the environmental and economic cost since the food has already been processed, packaged, stored, or even distributed (San Martin et al., 2021; Mosna et al., 2016). For this reason, developed countries are usually more harmful in this regard, as they have higher FW rates while developing countries have higher FL rates (Girotto et al., 2015; Parfitt et al., 2010). In short, the same amount of wasted food can have different consequences depending on the type of food in question and the stage of the FSC where it happens. This justifies the need for transparency not only in the quantification of wasted food at each stage of the FSC but also in the identification of the kind of waste that is generated. Although the quantities of FL are higher than the quantities of FW, it is the latter that leads to worse environmental, social, and economic issues (Praneetrattananon et al., 2005; Feukam Nzudie et al., 2020; Gao et al., 2021; Girotto et al., 2015; Kamble et al., 2020). Hence, the present work focuses on FW, i.e. wasted food generated in retail, households, and food services, as reducing FW in these settings enables to minimize environmental, social, and economic issues.

The literature on FW has adopted two main approaches: quantitative and qualitative. The former relates to the quantification of FW and, sometimes, its discrimination. The latter usually aims at finding the reasons/drivers of FW and at understanding how it can be avoided and how to better allocate the waste that is produced.

Regarding the quantitative approach, specifically with regard to descriptive analysis, by 2010 there were only a few publications quantifying FW around the globe (Parfitt et al., 2010), and most of their data referred back to the 1980s. In 2013, a review on FW prevention found only four publications with data related to FW in the hospitality sector (Schneider, 2013). In 2016, Pirani and Arafat (2016) concluded that although some reports already quantified FW, many did not mention the quantities of avoidable FW. A more recent literature review (Filimonau and De Coteau, 2019) presented publications that quantify FW in the hospitality sector (Chalak et al., 2018; Principato et al., 2018). The authors found that, at the time, the number of publications quantifying FW in the last five years (2015-2019) was greater than in the previous 20 years (1995-2014). In 2020, Dhir et al. (2020) found 24 quantitative studies and a further 17 mixed methods (between quantitative and qualitative) tackling FW generation in the profit sector of hospitality and food services. These literature reviews are evidence of the current popularity of FW quantification, namely after 2015. However, these reviews do not provide a detailed picture of the methods and data used in the different studies or their results. Moreover, they only address the descriptive dimension of analytics, neglecting the potential of the predictive and prescriptive approaches.

Conversely, the qualitative approach to FW has been intensively explored (Batista et al., 2021). In 2017, there were already a considerable number of papers portraying the drivers that lead to wasted food. Hebrok and Boks (2017) identified 112 scientific sources between 2010 and 2015 that explore the link between FW and consumer behavior (households), pinpointing nine main drivers. Canali et al. (2017) located 28 studies on the subject of FW and revealed that most of them do not classify FW causes or drivers. The authors then classified and enumerated the FW drivers into three groups - technological, institutional, and social. Dhir et al. (2020) found 18 qualitative studies and a further 17 mixed methods (between quantitative and qualitative) in the field of FW generation in the profit sector of hospitality and food services. The authors identified several FW drivers and concluded that the main ones are the nature of the food menu, the production procedure, and the use of pre-prepared food products (instead of whole).

Summing up, even though the topic of FW has become more popular in recent years and its qualitative analysis stream is already mature, the quantitative analysis stream is still developing, and some gaps must be filled. In particular, there are no literature reviews on the predictive and prescriptive analysis of FW. Furthermore, the literature reviews on the descriptive analysis of FW did not fully display the relevant features of these analyses, such as the methods used to measure FW and the amount of FW estimated within each specific context.

Thus, we propose a systematic literature review of the quantitative approaches to FW, referencing and discriminating the papers by descriptive, predictive, and prescriptive analytics and showcasing their features. For each of these three dimensions, we present a summary of the relevant studies.

The need for such a review is threefold. First, research on FW has focused more on qualitative approaches, thoroughly exploring the main drivers that enable FW generation. On the other hand, and to the best of our knowledge, no previous review was conducted on the quantitative approach to FW, which simultaneously aggregates descriptive, predictive, and prescriptive approaches. Moreover, the need to reduce FW calls for advanced models supported by the increasing volume of data collected by organizations, given the increasingly complex contexts of food services. Therefore, this chapter aims to approach FW, focusing on the models adopted to explore it in the different FW settings. We identify the modeling approaches in each setting, summarizing the relevant decisions considered by past researchers and further presenting key features that promote the future adoption of related models.

4.2 Systematic Literature Methodology

The methodology adopted to conduct this literature review is based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement methodology (Page et al., 2021). The process of developing the literature review is divided into five steps: objectives, searching strategy, eligibility criteria, study selection, and data collection.

4.2.1 Objective

Studies dealing with the quantitative analysis stream - descriptive, predictive, and prescriptive - of FW have been increasing in both popularity and number. However, no literature review has simultaneously presented the different types of data analytics in the quantitative branch.

It is important to reference and succinctly present these works to bring more attention to the subject. This includes presenting what has been done, how it was done, in what environment it was done, and what was its performance/the benefit of adopting it. This analysis should enable newcomers to quickly pick up on the current state of the research on this subject and display what they can test in their own environment and how to do it.

To objectively address these issues and shed light on the important subjects, this chapter aims to address the following research questions:

- What is the most popular data analytics approach among quantitative studies?
- What are the most explored settings?
- How is FW measured and described?
- What are the main forecast/optimization methods adopted to prevent FW in each sector (retail, hospitality sector, final consumers, country, and world population)?
- What are the main factors used in these forecasts/optimizations?
- How much improvement do the quantitative models provide compared to the normal operation (considering the experience of the managers or simpler models)?

4.2.2 Searching strategy

We built the frame of this work by extracting the literature content of the Elsevier SCOPUS and Web of Science databases. Then, with the knowledge obtained from analyzing the gathered literature, we addressed the previous research questions in the quantitative sphere of FW. To do so, we defined a set of keywords, which can be divided into two main groups, as seen in Table 4.1.

We reserved the first set of keywords for terms related to the overall subject of FW and used the second group for terms concerning quantitative analytics.

We searched for the studies in the above-mentioned databases published up to and including 2022, ensuring that the netted papers contained at least one term from each of the two groups in their title or abstract. Note that although our focus is food waste, we have included “food loss*” in the query to gather the studies that use this terminology to refer to waste generated in the latter stages of the supply chain. Additionally, the query was performed only to retrieve articles and review articles - excluding other document types such as conference papers - and studies written and published in English.

Table 4.1: Group of keywords in the query

Food waste		Quantitative analytics
food wast* food loss* wast* of food food AND ("wast* reduction") food AND ("wast* prevention")	AND	data analytics data analysis big data descriptive anal* explanatory analysis data mining statistical model predictive anal* prediction forecast* machine learning prescriptive anal* optimization model simulation model mathematical model mathematical program*

Although the set of keywords used passed through several iterations and discussions, it is still bound to have its own limitations. Regarding the keywords related to food waste, there might be other terminology and terms that authors may use to refer to it. One such case may be studies that address specific food items, e.g. banana waste, rather than the broad categorization of food waste. As for the quantitative analytics keywords, specific terms can be adopted. Even though some of the keywords employed are generic to consider all articles in this sphere, some can still not be captured using this query.

4.2.3 Eligibility Criteria

To ensure that the extracted articles are aligned with the scope of this work, a set of inclusion and exclusion criteria were noted and used in the screening process.

The following items are part of the inclusion criteria where any paper that respects any of the following items was considered in this literature:

- Measures or estimates the amounts of food waste produced in a given setting;
- Forecasts future amounts of food waste that will be produced;
- Develops a forecasting tool to mitigate food waste generation (this includes forecasting demand and/or shelf-life of products);
- Employs optimization processes to minimize food waste generation.

Notwithstanding the inclusion criteria, if a paper fits in any of the exclusion criteria it is not considered in this review. This boosts the precision of the studies to be extracted from the databases, further refining the objective and elevating the quality of the information to be presented. Hence, the exclusion criteria are as follows:

- Covers wasted food that was generated at the earlier stages of the FSC, i.e. all stages of the FSC excluding retail and consumer stages;
- Uses FW as a factor to predict other variables instead of forecasting FW;
- Uses predictive methods to study the conversion of FW into energy (e.g., anaerobic digestion);
- Uses qualitative data from interviews.

4.2.4 Study Selection

After performing the query and defining the eligibility criteria of the studies to consider, duplicate studies retrieved from both databases were removed. Only then did the screening process take place. The screening process was twofold. The first phase was performed by both authors, where the title and abstract were analyzed. The authors evaluated whether the studies should be included in this systematic review or not, considering the eligibility criteria. In case of doubt, the study was passed through to the next phase.

Although this first stage of the screening was performed by two people, there was no communication between the two parties before all studies were classified. Only after finishing the first phase did both parties have access to each other evaluation, and discussed in case of conflict until getting a consensus on the decision.

After reaching a consensus among the authors on the relevant studies, studies were used in the second phase of the screening. In this phase, the studies passed through a full reading and were classified by the type of analysis they performed. The screening process was streamlined through the usage of the Rayyan platform (Ouzzani et al., 2016).

4.2.5 Result presentation

The studies that come from the two phases of the screening process will be presented and further explored in the following four sections of this chapter. In section 4.3 the results of the methodology adopted are shown, presenting the number of articles first extracted from the set of keywords and going through the different stages until the final batch of studies is consolidated.

In section 4.4, we divide the papers depending on the type of analysis performed, e.g. descriptive, predictive, or prescriptive. Descriptive analysis account for the quantification of the amount of FW in different settings. The predictive analysis focuses on the application of any forecasting method to predict future demand, shelf-life, or the amount of FW directly. Lastly, prescriptive analysis involves optimization models to reach an optimal result that allows the minimization of FW.

After this categorization, a summary of each quantitative approach will be produced in their own table to systematize their main characteristics at once. These tables share the same structure, with slight adaptations. By categorizing the papers by their data analytics approach, it highlights the techniques, models, and underlying features adopted. These tables also present an overview of the subject, exposing the amounts of FW generated along with how to reduce it, and what type of results can be expected.

In section 4.5, we divide the studies according to their context, e.g. retail, restaurant, and canteens. This way, the results are bestowed following another compartmentalization that makes it easier to extract information and draw conclusions regarding only a specific context. This is important as different contexts tend to have prevalent features.

Lastly, section 4.6 is meant to discuss the different findings on the final batch of papers as a whole. For this, the years of publication are also presented to highlight the increasing popularity of the subject throughout the years.

4.3 Results

According to the aforementioned PRISMA methodology, the query performed on the Elsevier SCOPUS and Web of Science databases returned 414 and 330 papers, respectively. After removing duplicate studies, we extracted a total of 499 unique papers to be used in the screening process. The above and the further steps are drawn in Figure 4.1 through a flow diagram, identifying the number of articles in each of them.

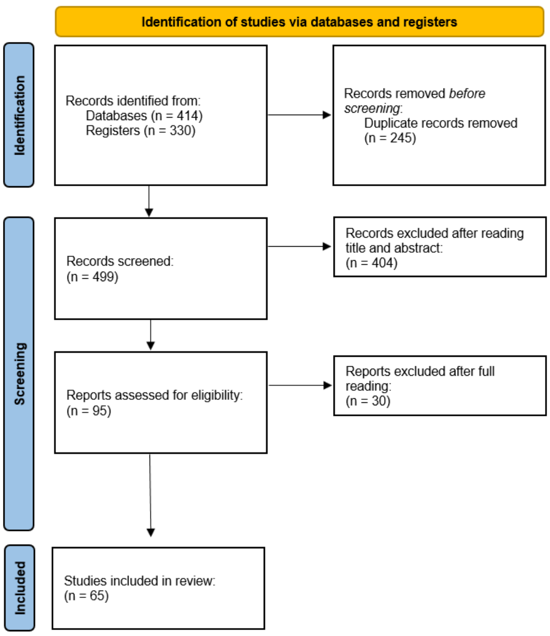

Figure 4.1: PRISMA flow diagram showing study selection applied to quantitative approaches regarding food waste.

In the first screening phase, the title and abstract of the 499 papers were read by both authors. After reading and discussing conflict decisions, 404 papers were disfavored, leaving 95 papers for the second screening phase.

The 95 papers were fully read to assess which ones should be acknowledged in this systematic literature review. This resulted in a final batch of 65 papers that were categorized according to the type of analysis performed and the setting they focused on.

4.4 Data Analytics Approaches

In this section, we organize the gathered studies by the type of analytics performed. Furthermore, we dedicate a sub-section to each type of analysis that includes a table with the key features of the studies, allowing for a better comparison between them. Finally, we report and discuss the relevant findings for each category.

4.4.1 Descriptive analysis

Uneasiness toward the increasing amount of FW has been growing in recent years. Organizations like the Food and Agriculture Organization of the United Nations (FAO) have made important contributions by providing figures on the matter. More specifically, their 2011 report is very enlightening, stating that around one-third of all the food produced for human consumption ends up being wasted (FAO, 2011). This value has been referenced several times and is usually the quotation used to report the global FW amount.

The specific amount of FW disclosed in the FAO report was calculated considering some assumptions. Hence, the real amount for each food service provider and/or consumer may be different. This highlights the importance of having a quantification of FW in the different types of food services around the world and at different moments. It is also equally important to know the methodology adopted to quantify the FW so that the measurement methods can be replicated.

There are over a dozen scientific studies that quantify FW. Most of these are from the last decade, which emphasizes the increasing relevance of the topic and the need to know the amount that is, in fact, wasted. Table 4.2 presents the most relevant features of the studies that follow a descriptive approach to FW. This table provides information such as the measurement's context, methodology of the measurement, periodicity adopted in the quantification, for how long the measurement was taken, and other meaningful factors. The goal of this table is to present a concise summary of the key features regarding FW quantification, which should allow the reader to get an overall knowledge of the subject quickly.

Regarding the context of the descriptive analysis, most papers focus on FW in the hospitality sector, with schools as the highlight (Endres et al., 2003; Ferreira et al., 2013; Chapman et al., 2017; Lubura et al., 2022; Malefors et al., 2022; Persson Osowski et al., 2022). The quantification of FW produced by consumers/households is closely behind (Conrad et al., 2018; Grainger et al., 2018; Fanelli, 2019; Vittuari et al., 2020; Afzal et al., 2022; Kandemir et al., 2022), followed by the quantification on a city (Thanh et al., 2010; Al-Maliky and ElKhayat, 2012) and country scale (Adelodun and Choi, 2020; Alshabanat et al., 2021).

Respecting the type of data used for the quantification, 12 of the 17 studies use primary data, i.e. data gathered by the research team (Endres et al., 2003; Thanh et al., 2010; Al-Maliky and ElKhayat, 2012; Ferreira et al., 2013; Pirani and Arafat, 2016; Chapman et al., 2017; Grainger et al., 2018; Fanelli, 2019; Afzal et al., 2022; Lubura et al., 2022; Malefors et al., 2022; Persson Osowski et al., 2022), and the remaining five use secondary data (Conrad et al., 2018; Adelodun and Choi, 2020; Vittuari et al., 2020; Alshabanat et al., 2021; Kandemir et al., 2022), i.e., data gathered by a third-party.

Concerning the FW measures adopted, most studies quantify FW either as an absolute quantity (Al-Maliky and ElKhayat, 2012; Ferreira et al., 2013; Pirani and Arafat, 2016) or as a mean amount per meal or individual (Endres et al., 2003; Thanh et al., 2010; Pirani and Arafat, 2016; Conrad et al., 2018; Grainger et al., 2018; Fanelli, 2019; Alshabanat et al., 2021; Afzal et al., 2022; Malefors et al., 2022; Persson Osowski et al., 2022). Only a few studies present the percentage of FW in reference to the total amount of waste/food produced (Pirani and Arafat, 2016; Conrad et al., 2018; Alshabanat et al., 2021) or the average percentage of FW per meal (Chapman et al., 2017; Kandemir et al., 2022; Lubura et al., 2022). It should be noted that some studies present their results using different metrics.

Table 4.2: Summary table for the descriptive analysis

Authors	Context	Type of data	Data collection method	Factors collected	Data	Periodicity	Observation period
Endres et al. (2003)	School	Primary	Weighing 10 samples randomly selected from plates ready for the serving cart	Amount served, amount consumed	Absolute per plate	Daily	3 weeks
Thanh et al. (2010)	City	Primary	Weighing biodegradable and non-biodegradable wastes	Kitchen waste	Absolute per household	Daily	1 month (Fev-Mar 2009), 2 weeks (Oct 2009)
Al-Maliky and ElKhayat (2012)	City	Primary	FW and purchases weighed in grams. Questionnaire on the no. of family members and income	Kitchen FW, no. of family members, food purchases	Absolute total	Weekly	8 months (Nov 2011-Jun 2012)
Ferreira et al. (2013)	University canteen	Primary	Weighing raw and cooked food available for distribution	Ratio between plate waste and the quantity of food distributed, subtracting the weight of leftovers	Absolute total	Daily	4 weeks (Oct-Nov 2010)
Pirani and Arafat (2016)	Hotel	Primary	Weighing the ingredients used for the meal and FW (material flow analysis)	Preparation waste, plate waste, waste from serving dishes	Absolute total, absolute per guest, percentage of total input of cooked food	Daily	2 years (2012-2013)
Chapman et al. (2017)	School	Primary	Weighing 10 trays before serving for reference and weighing all trays after the meal	Tray waste (entrée, fruit, vegetable, milk) before and after reverse recess	Percentage of lunch consumed per student	Daily	8 weeks (Jan-Apr 2014)
Conrad et al. (2018)	Consumer	Secondary	Estimation of food loss through loss factors taken from published studies and discussions with commodity experts. Dietary and composition of mixed dishes data taken from surveys	FW items, FL items	Absolute per capita, FW percentage	Yearly	8 years (2007-2014)
Grainger et al. (2018)	Household	Primary	Weighing household waste collected from the kerbside	Avoidable waste per household	Absolute per household	-	1 year (2012)
Fanelli (2019)	Household	Primary	Survey	Food waste	Absolute per capita	Daily	April 2017 and July 2018
Adelodun and Choi (2020)	Country	Secondary	Top-down mass flow analysis	FW by food commodity type and by FSC stage	Absolute per food commodity type	-	11 years (2007-2017)
Vittuari et al. (2020)	Consumer	Secondary	Estimation of food loss through loss factors taken from published studies and discussions with commodity experts.	Refrigerated FW, non-refrigerated FW, home FW, out of home FW	Absolute total	Yearly	8 years (2005-2012)
Alshabanat et al. (2021)	Country	Secondary	Mass flow analysis	Amount of food loss and food waste of 19 food products	Percentage per food group, absolute per capita	Yearly	1 year (2016)
Afzal et al. (2022)	Hospitality	Primary	Questionnaire surveys (restaurants, hotels and wedding caterers); live-tracking: weighing machine (restaurants)	Food waste	Waste per customer, percentage waste	Daily	Questionnaire surveys: 4 months (Nov 2010 - Fev 2021); Live-tracking: 7 days
Kandemir et al. (2022)	Household	Secondary	HHSM built using ARENA Simulation Software with 7 different scenarios	Milk wasted	Percentage per household	Daily	50 years, 30 replications
Lubura et al. (2022)	School	Primary	Employment of a convolutional neural network (CNN) to classify % of food waste through images	Food waste	Percentage per portion	-	-
Malefors et al. (2022)	School	Primary	Weighing the mass of waste	Kitchen waste; serving waste, plate waste	Absolute per portion	Daily (with some Nas)	7 years (spring 2014 - autumn 2020)
Persson Osowski et al. (2022)	School	Primary	Performed by the kitchen staff	Kitchen waste, serving waste, plate waste	Absolute per guest	Daily (with some Nas)	7 years (spring 2014 - autumn 2020), 9 years (2012-2020)

Three studies do not mention the periodicity of the waste quantification (Grainger et al., 2018; Adelodun and Choi, 2020; Lubura et al., 2022). Of the 14 that do, 11 measure FW daily (Endres et al., 2003; Thanh et al., 2010; Ferreira et al., 2013; Pirani and Arafat, 2016; Chapman et al., 2017; Fanelli, 2019; Afzal et al., 2022; Kandemir et al., 2022; Malefors et al., 2022; Persson Osowski et al., 2022), one measures it weekly (Al-Maliky and ElKhayat, 2012), and three measures FW yearly (Conrad et al., 2018; Vittuari et al., 2020; Alshabanat et al., 2021).

It is vital to examine the analysis period in conjunction with the periodicity. The studies that measure FW daily have an observation period that ranges from three weeks to eight weeks (Endres et al., 2003; Thanh et al., 2010; Ferreira et al., 2013; Chapman et al., 2017; Conrad et al., 2018; Kandemir et al., 2022; Malefors et al., 2022; Persson Osowski et al., 2022), except for Malefors et al. (2022) that collected data for seven years, and Persson Osowski et al. (2022) that collected data for seven and nine years. This range roughly translates to about 15 to 40 weekdays (observations), as FW on weekends is usually not measured.

Although Pirani and Arafat (2016), Fanelli (2019), and Afzal et al. (2022) present a daily periodicity of measurement, they were not considered in this analysis period evaluation since the first was performed throughout two years but considered ten different events that happened over a single day, and the former two because the data presented is a product of surveys that do not mention for what period should the participants answer to.

In contrast, Al-Maliky and ElKhayat (2012) measures FW weekly during eight months (around 34 observations) (Al-Maliky and ElKhayat, 2012), while studies that adopt a yearly periodicity measure FW in one (Alshabanat et al., 2021) to eight years (Conrad et al., 2018; Vittuari et al., 2020).

Depending on the information disclosed, it was possible to extrapolate the amount of FW expected in a year and attain a value that can be easily compared between studies. We assumed that a year was compromised of 260 weekdays (52 weeks) and computed, whenever possible, the yearly FW (260 days) for the total number of individuals considered in each study.

Table 4.3: Amount of FW in the descriptive analysis

Authors	Context	Data	No. of individuals	Average yearly amount of FW (kg) for the total no. of individuals	Average yearly amount of FW per individual (kg)
Endres et al. (2003)	School	Absolute per plate	39 students children	2352.48; 1733.94[1]	60.32; 44.46[1]
Thanh et al. (2010)	City	Absolute per household	100 households in first stage; 60 in second (average 4,41 residents per household)	6196.09 - 6360.12	61.96 - 63.60
Al-Maliky and ElKhayat (2012)	City	Absolute total	61 adults, 44 children	272.46	2.59
Ferreira et al. (2013)	University	Absolute total	988 individuals	20,540	20.80
Pirani and Arafat (2016)	Hotel	Absolute total, absolute per guest, percentage of total input of cooked food	1,202 guests	337,818	281.05
Chapman et al. (2017)	School	Percentage of lunch consumed per student	20,183 students	-	-
Conrad et al. (2018)	Consumer	Absolute per capita, percentage of FW	35,507 consumers	3,895,828	109.72
Grainger et al. (2018)	Household	Absolute per household	1,770 individuals	433,680	245.02
Fanelli (2019)	Household	Absolute per capita	1008 households[2]	24.07[2]	23.90[2]
Adelodun and Choi (2020)	Country	Absolute per type of food commodity	49,170,125 individuals	4,682,616,000	95.23
Vittuari et al. (2020)	Consumer	Absolute total	-	58,000	-
Alshabanat et al. (2021)	Country	Loss and Waste rates; FLW rates; FLW per capita	31,742,580 people	5,840,635,000 tones	184
Afzal et al. (2022)	Restaurant (n=3)	Absolute total, Absolute per customer, Percentage of total food	-	107,250	39.9
Kandemir et al. (2022)	Households / country	Percentage of milk purchased per household	UK population	-	-
Lubura et al. (2022)	School	Percentage per portion	30 students	-	-
Malefors et al. (2022)	School	Absolute per portion	3227	56,047 - 62,255	17.40 - 19.30
Persson Osowski et al. (2022)	School (n=3)	Absolute per guest	B3:700; A1:400; B1:500[3]	8190; 10192; 8320[3]	11.70; 25.50; 16.60[3]

[1] The same group of students are presented with traditional and soy-enhanced menus in different time periods
[2] Households have between 1 and 7+ members, thus estimations were performed per household instead of per individual
[3] Three different schools inspected in different time periods

For studies that quantify FW in several food service establishments simultaneously, the yearly amount of FW is estimated for each food service establishment and only then aggregated into a single value, i.e. the sum of the values obtained for each one. Table 4.3 displays the extrapolated amounts of yearly FW for the sample of individuals considered in each study and per individual.

The sample of individuals considered in the studies varies considerably, going from 39 individuals (Endres et al., 2003) to an entire country (Alshabanat et al., 2021). The yearly waste can range from 2.59 kg per individual (Al-Maliky and ElKhayat, 2012) to 281.05 kg (Pirani and Arafat, 2016). Please note that these values should not be compared since the conditions behind the waste quantification were not the same, i.e. they happened in different settings, different locations, and the yearly estimates may be based on data collected in different periods of the year. Additionally, extrapolation of yearly amounts can neglect or propel certain FW behaviors that concern only certain months of the year - seasonal cycles.

4.4.2 Predictive analysis

Regarding predictive analysis, we identified 39 studies that focus on forecasting. These papers can be distinguished based on the prediction targets. The literature presents studies aiming at predicting: a) future demand (or a proxy of it) in order to minimize future FW, b) amount of FW, or c) SL or expiration date of perishable food products.

The first group of studies is more in line with the objective of developing tools to avoid FW generation, which has a more practical usage in the field. On the other hand, the second group of studies enables an understanding of the factors that impact FW and estimate the amount of FW that may be generated in the future if there is no change in the current trend. The third group of papers has a purpose similar to future demand studies in the sense that its results may help food services manage their inventory and consequently prevent FW generation.

4.4.2.1 Demand

Only 14 of the above-mentioned 39 studies predict future demand. These studies use mostly causal models, with some employing models based on time series to forecast future demand or a proxy of it. This and additional information on the 39 studies are presented in Tables 4.4 and 4.5. In relation to causal models, the most predominant algorithms are the neural networks (NN) (Arunraj and Ahrens, 2015; Sakoda et al., 2019; Lee et al., 2020; Faezirad et al., 2021; Malefors et al., 2021b; Lutoslawski et al., 2021; Miguéis et al., 2022; Malefors et al., 2022) and Random Forests (RF) (Malefors et al., 2021a; Miguéis et al., 2022). Models based on time series are mainly supported by algorithms such as ARIMA (autoregressive integrated moving average) and its variations (Arunraj and Ahrens, 2015; Møller Christensen et al., 2021; Posch et al., 2021; Sakoda et al., 2019), TBATS (Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend and Seasonal components) (Dharmawardane et al., 2021; Posch et al., 2021), naïve (Malefors et al., 2021b; Møller Christensen et al., 2021; Posch et al., 2021), and Prophet (Malefors et al., 2021b; Posch et al., 2021).

The external variables most frequently used to explain the demand in the causal models are holidays, day, month, year, and reservations. Apart from these, there are three other factors that appear in the studies, namely Covid-19 cases (Malefors et al., 2021a), which highlights the impact of this virus on society, weather conditions (Arunraj and Ahrens, 2015; Miguéis et al., 2022), and the menu served (Faezirad et al., 2021).

The data granularity in all but three studies (Noone and Coulter, 2012; Lee et al., 2020; Mobaseri et al., 2021) is daily. This aligns with the forecasting granularity, which is also mostly daily. Although there is a clear trend in the data periodicity, this is not the case for the time horizon. The time horizon of the data collected to support these predictions ranges from 90 days to 11 years, with no clear trend on what is most popular. For instance, Mobaseri et al. (2021) collected data for 11 years but with a yearly periodicity translating into 11 observations. In contrast, Malefors et al. (2021a) and Malefors et al. (2021b) have a time horizon of 10 years and a daily periodicity translating to around 2600 weekdays worth of observations (260x10).

Regarding model performance evaluation, in general, the data used can be divided into training, test, and validation periods. The 14 studies rarely mention the validation period, but most refer to having a test period. Most studies use either a rolling window or a moving window (Sakoda et al., 2019; Lee et al., 2020; Malefors et al., 2021b; Posch et al., 2021; Lutoslawski et al., 2021; Miguéis et al., 2022), even when causal models are employed. Four studies use k-fold cross-validation (Arunraj and Ahrens, 2015; Faezirad et al., 2021; Malefors et al., 2021a; Møller Christensen et al., 2021).

The most used forecasting performance measures are the Mean Square Error (MSE) and the Mean Average Error (MAE) and their variations, e.g., the Root Mean Square Error (RMSE), weighted Mean Absolute Percentage Error (wMAPE), Mean Error (ME), and Mean Absolute Percentage Error (MAPE). Although the R2 error is a well-known metric in the field of machine learning/forecasting, only two of the studies use it (Faezirad et al., 2021; Miguéis et al., 2022). Other less frequently used measures are the Bias coefficient (Dharmawardane et al., 2021) and F1 score (Lee et al., 2020).

Finally, in general, the performance of the models needs to be benchmarked in order to determine their potential impact. When previous estimates are available, these can be selected to evaluate the actual impact of the models. When demand estimates are unavailable, it is possible to create a baseline model to mimic a potential forecasting model. This baseline model can be as simple as a naïve model replicating how a manager would estimate future demand. Then, the estimated values produced by the proposed models can be compared with those from the baseline model. In the set of studies analyzed in this review, four studies introduce a baseline model and compare the forecasts of the proposed models and the baseline (Arunraj and Ahrens, 2015; Lee et al., 2020; Dharmawardane et al., 2021; Møller Christensen et al., 2021).

One of these studies uses the Holt-Winters method as a baseline against the developed TBATS model (Dharmawardane et al., 2021). These authors tested two different settings, wherein the first setting the developed model was better than the baseline with a lower RMSE (3% to 26% reduction), MAE (2% to 30% reduction), ME (51% to 89% reduction), and forecast bias (28% to 57% reduction). In the second setting, however, the model proposed presented worse results with a higher RMSE (3% to 13% reduction) and MAE (4% to 35% reduction), a ME ranging from -71% up to +100%, and a forecast bias ranging from -41% up to +12% when compared to the baseline. The forecasting LSTM-based model developed in Lee et al. (2020) led to a higher F1 score of 10% to 21% over the collaborative filter-based model (baseline).

As for Arunraj and Ahrens (2015), SARIMA (seasonal ARIMA) variants (SARIMA-MLR, SARIMA-QR, SARIMA) and a neural network (MLPNN) were tested, achieving a lower MAPE (decrease between 8.75 and 11.81 percentage points) and RMSE (decrease between 3.73 and 4.19) when compared against the seasonal naïve model used as the baseline. Lastly, Miguéis et al. (2022) adopted a panoply of algorithms - LSTM, Feedforward NN, SVM, RF, and Holt-Winters - and tested their performance against a baseline developed by employing a naïve approach. The best prediction output was able to beat the baseline forecasts in all metrics, decreasing the RMSE by 8.14 (from 35.96 to 27.82), the MAE by 3.38% (24.01%-20.63%), the MPE by 6.71% (24.57%-17.86%), and the MNE by 3.04% (23.52%-20.48%).

Additionally, three studies mention an improvement but not the technique used as a baseline. From these, Malefors et al. (2021a) reduced costs by 17% and 63% depending on the kitchen (a total reduction of 30%), Malefors et al. (2021b) reduced the MAPE from 20-40% to 2-3%, and Faezirad et al. (2021) reduced FW by 79.66%.

Table 4.4: Summary table for the predictive analysis of demand - part 1

Authors	Context	Algorithm adopted	Dependent variable adopted	Independent variables adopted	Error / performance measure	Perio-dicity	Input period
Noone and Coulter (2012)	Restaurant	Cook-decision algorithms based on Zaxby's specific operating rules and procedures	Demand of individual products (no. of unmet orders)	Guest arrival	-	-	90 days
Arunraj and Ahrens (2015)	Retail	SARIMA-MLR, SARIMA-QR, SARIMA, MLPNN	Banana sales	Day, month, holiday (before and after), promotional effect, discount effect (discounted sales), weather (relative humidity, temperature, sunshine duration, fresh snow depth, and snow depth)	MAPE, RMSE	Daily	5 years (2010-2014)
Sakoda et al. (2019)	Retail	ARIMA, NN, Exponentially weighted moving average	Demand of perishable items	-	-	Daily	90 days (2010)
Lee et al. (2020)	Retail	LSTM-based model	Set of items to be purchased next time by a consumer	Purchasing information and purchasing order	F1 score	-	1 year (2017)
Dharmawardane et al. (2021)	Retail	TBATS	Sales	-	RMSE, MAE, ME, Bias coefficient	Daily	136 weeks (2015-2018)
Faezirad et al. (2021)	University	NN with chance-constrained programming	Users' show rate (times no. of students = no. of meals served)	Weekdays, consecutive holidays, food price level, food name, no. of reservations by academic degree, no. of reservations by accommodation status, ratios of previous attendance	MSE, MAE, R2	Daily	7 years (2013-2019)
Lutoslawski et al. (2021)	Retail	Nonlinear autoregressive exogenous model in combination with neural networks (NARXNN)	Demand of products	-	MAE, MAPE, RMSE, and R2	Daily	2 years (2017-2019), 3.5 years (2016-2019)
Malefors et al. (2021a)	School	Random forests	No. of meal portions	Holiday, year, day of the year, quarter, month, trend, weekday, weekend, week. Specific variables for individual kitchens: no. of daily new Covid-19 cases in the municipality of the kitchen, identification of post March 19, 2020	MSE, MAPE, cMAPE, ARB	Daily	10 months (2019-2020)
Malefors et al. (2021b)	School	Benchmark scenario, last value, moving average, Prophet, NN	No. of portions served	No. of guests, dates for holidays and breaks with activity, enrolled students per school and school year as of October	MAPE	Daily	10 years (2010-2019), 5 years (2015-2019), 4 years (2016-2019)
Mobaseri et al. (2021)	Country	Linear regression (considers a constant increase afterwards)	Food demand			Yearly	11 years (2004–2015)
Møller Christensen et al. (2021)	Retail	Naïve, naïve with seasonality and moving average, ARIMA, Theta, Exponential smoothing (ETS), combination model (arithmetic mean value from ARIMA, Theta and ETS)	Demand of products	-	RMSE, wMAPE, wQL, wSLE	Daily	12 months
Posch et al. (2021)	Restaurant and staff canteen	Bayesian generalized additive (normal & negBinom), PhophetS, PhophetA, ARIMA, SARIMA, ExpSmooth, TBATS	Quantity of menu items sold	-	MAD, MSE, MAPE, MFE, WAPE, MAAPE, MASE1, MASE7, PICP, PINAW, MSIS1, MSIS7	Daily	20 months
Miguéis et al. (2022)	Retail	LSTM, Feedforward NN, SVR, Random Forests, Holt-Winters	Fish sales	Month, first week of the month, last week of the month, day of the week, summer period, unit price, discount percentage, sales, lent, public holidays, expected precipitation, expected temperature	RMSE, MAE, MPE, MNE	Daily	3 years (2017-2019)
Malefors et al. (2022)	School	Neural network	Number of guests	-	-	Daily	7 years (2014-2020)

Table 4.5: Summary table for the predictive analysis of demand - part 2

Authors	Context	Algorithm adopted	Validation method	FC period ahead	Improvement (%)	Algorithm in baseline	Dependent variable in baseline	Independent variables in baseline
Noone and Coulter (2012)	Restaurant	Cook-decision algorithms based on Zaxby's specific operating rules and procedures	-	-	-	-	-	-
Arunraj and Ahrens (2015)	Retail	SARIMA-MLR, SARIMA-QR, SARIMA, MLPNN	Training (70%) and validation (30%) sets	-	MAPE from 35.28 to 23.47-26.53, RMSE from 23.08 to 18.89-19.35, accuracy from 64.72% to 75.82%-76.53%	Seasonal naïve model (last week values without promotions and/or holidays)	Daily demand	-
Sakoda et al. (2019)	Retail	ARIMA, NN, Exponentially weighted moving average	Rolling window (63 days)	Next day	-	-	-	-
Lee et al. (2020)	Retail	LSTM-based model	Multi-period, moving window, rolling window	5 days, next day	Multi-period: F1 score 10-21% higher; Moving window: F1 score 17% higher	Collaborative filter-based model	Recommended set of items to be purchased next time	Purchasing history
Dharmawardane et al. (2021)	Retail	TBATS	Rolling and moving window (132 days)	4 weeks, 18 weeks	Fast moving: RMSE 3%-26% lower, MAE 2%-30% lower, ME 51%-89% lower, forecast bias 28%-57% lower; Highly intermittent: RMSE 3%-13% higher, MAE 4%-35% higher, ME 71% lower up to 100% higher, forecast bias 41% lower up to 12% higher	Holt-Winters (rolling window every 4 weeks)	Sales	-
Faezirad et al. (2021)	University	NN with chance-constrained programming	Training (70%), validation (15%), and testing (15%) sets	1 year	FW reduced by 79.66%	-	-	-
Lutoslawski et al. (2021)	Retail	Nonlinear autoregressive exogenous model in combination with neural networks (NARXNN)	Rolling window (last 4% and 10% of observations)	Between 1-8 days	-	-	-	-
Malefors et al. (2021a)	School	Random forests	Holdout cross-validation (2 weeks for kitchen 1, 3 weeks for kitchen 2)	All at once	Total: -30% costs against no plan, and +5% against actual data; Kitchen 1: -17% costs against no plan, and +0.2% against actual data; Kitchen 2: -63% costs against no plan, and +45% against actual data	-	-	-
Malefors et al. (2021b)	School	Benchmark scenario, last value, moving average, Prophet, NN	Rolling window (1 year)	Next day	MAPE decrease from 20%-40% to 2%-3%	-	-	-
Mobaseri et al. (2021)	Country	Linear regression (considers a constant increase afterwards)	-	Next year				
Møller Christensen et al. (2021)	Retail	Naïve, naïve with seasonality and moving average, ARIMA, Theta, ETS, combination model (arithmetic mean value from ARIMA, Theta and ETS)	Rolling window (1 year)	Next day	-	Weekday pattern analysis	Total demand	-
Posch et al. (2021)	Restaurant and staff canteen	Bayesian generalized additive (normal & negBinom), PhophetS, PhophetA, ARIMA, SARIMA, ExpSmooth, TBATS	Reverse rolling window with 15 k-folds (14 testing sets)	-	-	-	-	-
Miguéis et al. (2022)	Retail	LSTM, Feedforward NN, SVR, Random Forests, Holt-Winters	Rolling window: validation (6 months), test (1 year)	Two days after	RMSE from 35.96 to 27,82-33,13, MAE from 24,01% to 20,63%-25,07%, MPE from 24,57% to 17,86%-31,08%, MNE from 23,52% to 20,48%-21,81%			
Malefors et al. (2022)	School	Neural network	-	-	-	-	-	-

4.4.2.2 Food waste

Concerning FW prediction, 13 studies were found with all its identified features displayed in Tables 4.6 and 4.7. Mobaseri et al. (2021) was also considered in this category along with the demand forecast category, as it does both demand and FW forecast. Unlike the future demand forecasts, most studies on FW prediction consider a yearly periodicity. Moreover, these studies tend to focus on the waste of an entire city/municipality instead of just a food service establishment. Only three papers deviate from this trend and focus on a food company instead (Kazancoglu et al., 2018; Garre et al., 2020; Kodors et al., 2022). They all present a different periodicity, namely a daily, weekly, and yearly granularity.

In terms of the forecasting models adopted, most studies use causal models (Shapiro-Bengtsen et al., 2020; Garre et al., 2020; Ayeleru et al., 2018; Hall et al., 2009; Srivastava and Nema, 2008; Ozkan et al., 2021; Melikoglu, 2020; Mobaseri et al., 2021). All of these eight studies use regression algorithms (linear regression, multi-linear regression, fuzzy regression), with Garre et al. (2020) also employing random forests and gradient boosting. The other studies adopt a GM (1,1) grey rolling model (Kazancoglu et al., 2018), a simple percentage growth each year (Lee et al., 2019), a Q-learning (Q-table) based reinforcement learning (Dey et al., 2022), an ANN with Bayesian hyperparameter optimization (Hoy et al., 2022), and a Monte Carlo simulation (Kodors et al., 2022).

In terms of variables, the dependent variable adopted in most studies is the amount of FW (Garre et al., 2020; Srivastava and Nema, 2008; Ozkan et al., 2021; Melikoglu, 2020; Mobaseri et al., 2021; Hoy et al., 2022; Lee et al., 2019). The others forecast municipal solid waste (MSW) (Shapiro-Bengtsen et al., 2020; Ayeleru et al., 2018), milk production (Kazancoglu et al., 2018), food intake and assume that a fixed percentage of these variables is FW (Hall et al., 2009), and plate waste (Kodors et al., 2022).

Table 4.6: Summary table for the predictive analysis of FW - part 1

Authors	Context	Algorithm adopted	Dependent variable adopted	Independent variables adopted	Error / performance measure	Periodicity	Input period
Srivastava and Nema (2008)	City	Fuzzy regression	Percentage of components of the total waste (paper, plastic, metal, glass, rags, inert, food)	Socioeconomic conditions (per capita income, GDP, individuals per household, total population, population density), solid waste composition (paper, plastics, food, metal, glass, other wastes)	-	Yearly	3 years (1989, 2000, 2006)
Hall et al. (2009)	Country	Linear regression	FW	Average food intake (in reference to the average body weight), food supply	-	Yearly	31 years (1974-2004)
Ayeleru et al. (2018)	City	Linear regression	MSW (uses historical FW % to estimate current FW)	Population, no. of households	MAPE, MAD, MSD, R2	Yearly	20 years (1996-2015)
Kazancoglu et al. (2018)	Food company	Grey rolling model	Milk production (uses historical waste % to estimate current waste)	-	Absolute error, traditional error analysis	Yearly	7 years (2007-2013)
Lee et al. (2019)	City	Scenario analysis (utilization rate of the Organic Waste Treatment Facility (OWTF) = 0,1; 0,47; and 0,9)	Annual amount of food waste sent to landfills	Forecast values of annual waste sent to landfills, utilization rate of the OWTF	-	Yearly	-
Garre et al. (2020)	Food company	Regression (linear with stepwise selection, ridge, lasso, spline), regression tree, bagged tree, random forests, gradient boosting, and elastic net	Production losses	Total input weight used for production, filler used, no. of mixers required, type of container, main component of the recipe, type of product, type of filler used, homogenizer used, minimum and maximum Brix degree values of ingredients, minimum and maximum water activity of ingredients, minimum and maximum pH of ingredients	RMSE, MAE, R2	Daily	10 months
Melikoglu (2020)	Hospital	Linear regression	Yearly FW generation	No. of hospital beds, FW generation per bed per day, average yearly bed occupancy rate	-	Yearly	16 years (2003-2018)
Shapiro-Bengtsen et al. (2020)	Country	Multi-linear regression	MSW quantities (assumes 56% is FW)	Urban population, real disposable household income	R2, DW	Yearly	15 years (2003-2017)
Mobaseri et al. (2021)	Country	Linear regression (FW = 30%–40% of food demand) and scenario analysis (FW reduction by 10, 25 and 35%)	FW	Forecast food demand	-	Yearly	11 years (2004–2015)
Ozkan et al. (2021)	City	Regression (linear, exponential, polynomial)	FW	Income, population	SSE, RMSE	-	-
Dey et al. (2022)	Households	Q-learning (Q-table) based reinforcement learning	Amount of food that should be reduced or increased in the food surplus	-	Food surplus and food security	Weekly	1 year (Dec 2020 - Nov 2021)
Hoy et al. (2022)	Country	ANN with Bayesian hyperparameter optimization	FW	Social (e.g., population, life expectancy, and fertility rate), economic (e.g., gross domestic product, gross national income, and gross national expenditure), and environmental (e.g., CO2 emissions, energy use, and electricity consumption) indicators	RMSE, MAE	Yearly	35 years (1985–2020)
Kodors et al. (2022)	School	Monte Carlo	Plate waste	Portion size, eating rate, eating habits, children's preferences	RMSE, MAPE	Weekly	1 month (June-July 2021)

Table 4.7: Summary table for the predictive analysis of FW - part 2

Authors	Context	Algorithm adopted	Validation method	FC period ahead	Improvement (%)	Algorithm in baseline	Dependent variable in baseline	Independent variables in baseline
Srivastava and Nema (2008)	City	Fuzzy regression	-	18 years (2007-2024)	-	-	-	-
Hall et al. (2009)	Country	Linear regression	-	-	-	-	-	-
Ayeleru et al. (2018)	City	Linear regression	Testing set (100%)	10 years (2016-2025)	-	-	-	-
Kazancoglu et al. (2018)	Food company	Grey rolling model	Rolling window (3 years)	Next day	-	-	-	-
Lee et al. (2019)	City	Scenario analysis (utilization rate of the Organic Waste Treatment Facility (OWTF) = 0,1; 0,47; and 0,9)	-	4 years (2011-2014)	-	-	-	-
Garre et al. (2020)	Food company	Regression (linear with stepwise selection, ridge, lasso, spline), regression tree, bagged tree, random forests, gradient boosting, and elastic net	Training and validation set (70% of the data) where a 10 k-fold is applied, and a testing set (remaining 30% of the data)	-	-	-	Production planning	Number of tanks to be used, final packaging format
Melikoglu (2020)	Hospital	Linear regression	-	-	-	-	-	-
Shapiro-Bengtsen et al. (2020)	Country	Multi-linear regression	Testing set (100%)	31 years (2020-2050)	-	-	-	-
Mobaseri et al. (2021)	Country	Linear regression (FW = 30%–40% of food demand) and scenario analysis (FW reduction by 10, 25 and 35%)	-	11 years (2004–2015)	-	-	-	-
Ozkan et al. (2021)	City	Regression (linear, exponential, polynomial)	K-fold cross-validation	-	-	-	-	-
Dey et al. (2022)	Households	Q-learning (Q-table) based reinforcement learning	-	-	-	-	-	-
Hoy et al. (2022)	Country	ANN with Bayesian hyperparameter optimization	Training (70%) and validation (30%) sets	-	-	-	-	-
Kodors et al. (2022)	School	Monte Carlo	-	1 Week	-	-	-	-

Concerning the predictor variables, there is no consensus on the variables adopted other than the population (Shapiro-Bengtsen et al., 2020; Ayeleru et al., 2018; Ozkan et al., 2021; Hoy et al., 2022) and income/GDP (Shapiro-Bengtsen et al., 2020; Srivastava and Nema, 2008; Ozkan et al., 2021; Hoy et al., 2022), which appear a handful of times. The other context-specific variables adopted are food intake and food supply (Hall et al., 2009), the number of formally employed and unemployed people (Ayeleru et al., 2018), production characteristics (Garre et al., 2020), the number of hospital beds and bed occupancy rate (Melikoglu, 2020), environmental indicators (Hoy et al., 2022), and eating rate and habits (Kodors et al., 2022). It is worth noting that some of these variables are calculated based on others. For instance, Hall et al. (2009) calculates food intake by considering the average body weight, and Adhikari et al. (2006) calculates GDP and population growth by assuming a fixed year-to-year growth.

Even though the papers in this sub-category tend to forecast the amount of FW, they do not compare these predictions with other known forecasts. Only Garre et al. (2020) mentions a baseline for comparison, but it does not disclose the values generated by its baseline nor does it quantify the improvement in any way.

Lastly, regarding the error measures adopted, only half of the papers present an error measure. In these cases, the RMSE (Shapiro-Bengtsen et al., 2020; Ozkan et al., 2021; Hoy et al., 2022; Kodors et al., 2022) is the most popular measure, followed by the R2 (Shapiro-Bengtsen et al., 2020; Garre et al., 2020; Ayeleru et al., 2018), MAPE (Ayeleru et al., 2018; Kodors et al., 2022), MAE (Hoy et al., 2022) and absolute error (Kazancoglu et al., 2018).

4.4.2.3 Shelf-life

In what concerns the forecast of the shelf-life (SL) of products, 13 different studies were identified with their key features presented in Tables 4.8 and 4.9. From these, almost all (11 studies) operate in the retail settings (Annese et al., 2014; García et al., 2015; Sciortino et al., 2016; Göransson et al., 2018; Albrecht et al., 2021; Ktenioudaki et al., 2022; Logan et al., 2021; Nair et al., 2021; Pina et al., 2021; Nugraha et al., 2022; Zou et al., 2022), with the remaining two focusing on households (Orfanou et al., 2019; Richardson et al., 2019). This disparity comes from SL being a more retail-oriented problem where retailers need to know the SL, or a proxy of it, to manage and optimize their inventory.

In terms of the mathematical models adopted to forecast SL, these can be divided into machine learning and empirical models. The latter is represented by 11 studies that employ the Arrhenius equation (Arrhenius Law) (Annese et al., 2014; Sciortino et al., 2016; Ktenioudaki et al., 2022; Zou et al., 2022), regression models - linear, partial least squares, and logistic - (Sciortino et al., 2016; Göransson et al., 2018; Richardson et al., 2019; Nair et al., 2021; Nugraha et al., 2022), physical model (Göransson et al., 2018), Guggenheim Anderson–de Boer (GAB) model and Weibull hazard analysis (Orfanou et al., 2019), and ComBase growth model (Pina et al., 2021).

On the other side, four studies employ machine learning algorithms, namely support vector machine (Richardson et al., 2019; Nair et al., 2021), Levenberg–Marquardt algorithm (Albrecht et al., 2021), convolutional neural network (Logan et al., 2021), decision tree classifier, k-nearest neighbors, and Gaussian Naïve Bayes (Nair et al., 2021).

Regardless of the mathematical model, SL forecasts can be achieved differently. For instance, Annese et al. (2014) attains SL through the forecast of the degradation rate, while Göransson et al. (2018) does so through the quality sensory index, Orfanou et al. (2019) through the water sorption isotherms, Albrecht et al. (2021) through the bacterial count, and Pina et al. (2021) through microbial growth status. The other proxies of SL identified in other articles were the expiration date (Nugraha et al., 2022; Zou et al., 2022), the quality of the product (García et al., 2015; Sciortino et al., 2016), the spoilage (Richardson et al., 2019; Logan et al., 2021), and the ripeness (Nair et al., 2021).

Concerning the independent variables adopted to forecast SL or its proxy, the most frequent one is temperature (Annese et al., 2014; Sciortino et al., 2016; Göransson et al., 2018; Orfanou et al., 2019; Nair et al., 2021; Pina et al., 2021; Nugraha et al., 2022; Zou et al., 2022) followed by humidity (Annese et al., 2014; Sciortino et al., 2016; Nair et al., 2021; Nugraha et al., 2022), and light related characteristics (Annese et al., 2014; Richardson et al., 2019; Ktenioudaki et al., 2022). Additionally, there are also other relevant factors used in these forecasts, such as bacterial / organism-related characteristics (García et al., 2015; Pina et al., 2021), CO2 rates (Sciortino et al., 2016; Albrecht et al., 2021), and PH levels (Albrecht et al., 2021; Pina et al., 2021).

Table 4.8: Summary table for the predictive analysis of shelf-life - part 1

Authors	Context	Algorithm adopted	Dependent variable adopted	Independent variables adopted	Error / performance measure	Periodicity	Input period
Annese et al. (2014)	Retail	Arrhenius equation	Remaining SL through degradation rate	Temperature, ambient light, humidity	-	Hourly	24 hours
García et al. (2015)	Retail	Regression	SL through quality sensory index	Growth of specific spoilage organisms: concentration of Pseudomonas, concentration of Shewanella	-	Daily	Between 5 to 12 days
Sciortino et al. (2016)	Retail	Arrhenius equation	Quality of the product	CO2 respiration rate, temperature, humidity rate, volatile organic compounds	-	Every 10 minutes	48 hours
Göransson et al. (2018)	Retail	Linear regression, physical model	Loss of dynamically predicted SL; Loss of static SL	Temperature	-	Every 10 minutes	11.6-17.6 days, 5 days
Orfanou et al. (2019)	Households	Guggenheim Anderson–de Boer (GAB) model, Weibull hazard analysis	Secondary SL through water sorption isotherms	Storage temperature, product water activity	-	Daily	Between 6 to 127 days
Richardson et al. (2019)	Households	Partial least squares regression, support vector machine	Age of milk through spoilage	Spectral profile of the milk	RMSE, R2	Every 6 hours	Ten cartons of milk for 102 hours
Albrecht et al. (2021)	Retail	Levenberg–Marquardt algorithm	SL of pork through bacterial count	PH, CO2 concentration, O2 concentration, color, texture (elasticity, cohesion, chewiness, hardness, gumminess, adhesive strength, springiness), odor	RMSE, Bf, Af, R2	Every 10 minutes	250, 1400 hours
Ktenioudaki et al. (2022)	Retail	Arrhenius equation	SL through strawberry appearance	Wavelength bands of the spectra captured by the hyperspectral imaging technology	RMSE, RPD, R	-	385 samples during 9 days (day 0, 1, 3, 5, 7, 9)
Logan et al. (2021)	Retail	Convolutional neural network	Freshness classification, age estimation	Spectral reflectance (900 spatial pixel x 300 spectral pixel image)	-	Daily	-
Nair et al. (2021)	Retail	Logistic regression, decision tree classifier, support vector classifier, K-nearest neighbors classifier, gaussian naïve bayes	Ripeness / overripe	Gas level concentrations (mainly ethylene), temperature, humidity	Accuracy, ROC AUC	Every 30 seconds	2 days
Pina et al. (2021)	Retail	ComBase growth model	SL through microbial growth status	Water activity, PH, temperature, initial physiological state, initial bacterial dose	-	Every 12 minutes	8 days (199 hours)
Nugraha et al. (2022)	Retail	Polynomial regression	Chicken meat expiration date	Temperature, humidity, NH3, sensor output, timestamp, time difference	RMSE, R2	Daily	20 days
Zou et al. (2022)	Retail	Arrhenius equation	SL through expiration date	Temperature, time	Average remaining SL rate, food waste rate, customer demand satisfaction rate, expired food sold rate	Every minute	1 and 2 days

Table 4.9: Summary table for the predictive analysis of shelf-life - part 2

Authors	Context	Algorithm adopted	Validation method	Improvement (%)	Algorithm in baseline	Dependent variable in baseline	Independent variables in baseline
Annese et al. (2014)	Retail	Arrhenius equation	-	-	-	-	-
García et al. (2015)	Retail	Regression	-	-	-	-	-
Sciortino et al. (2016)	Retail	Arrhenius equation	-	-	Microbiological SL models	-	-
Göransson et al. (2018)	Retail	Linear regression, physical model	-	-	-	-	-
Orfanou et al. (2019)	Households	Guggenheim Anderson–de Boer (GAB) model, Weibull hazard analysis	-	-	-	-	-
Richardson et al. (2019)	Households	Partial least squares regression, support vector machine	-	-	-	-	-
Albrecht et al. (2021)	Retail	Levenberg–Marquardt algorithm	-	-	-	-	-
Ktenioudaki et al. (2022)	Retail	Arrhenius equation	Training set 70%, validation 15%, testing 15%	-	-	-	-
Logan et al. (2021)	Retail	Convolutional neural network	10-fold	-	-	-	-
Nair et al. (2021)	Retail	Logistic regression, decision tree classifier, support vector classifier, K-nearest neighbours classifier, gaussian naïve bayes	Training set 75%, testing set 25%	-	-	-	-
Pina et al. (2021)	Retail	ComBase growth model	-	-	-	-	-
Nugraha et al. (2022)	Retail	Polynomial regression	7 days	-	Monitoring System	Chicken meat expiration date	RFID tag, temperature sensor, humidity sensor, gas sensor, RFID reader
Zou et al. (2022)	Retail	Arrhenius equation	-	-	-	-	-

Although these studies do not usually employ any error measure to quantify the SL forecast performance, it is common for them to check if the real freshness and edibility of the products match with the forecast. The periodicity of SL forecasts is usually much higher than the ones presented in the FW forecasts and even demand forecasts. The lowest periodicity is daily (García et al., 2015; Orfanou et al., 2019; Logan et al., 2021; Nugraha et al., 2022), while the highest is every 30 seconds (Nair et al., 2021). The most common periodicity is daily, followed by every 10 minutes (Sciortino et al., 2016; Göransson et al., 2018; Albrecht et al., 2021). Considering the highest periodicity, the input periods are also smaller ranging from 1 to 127 days. However, these studies consider the same number or more observations than those recorded in both demand and FW forecasts.

It is unclear if most of the studies have a testing and validation period, and if they have, how they perform their validation. Only three of the aforementioned 13 studies disclose their validation method. Ktenioudaki et al. (2022) randomly divided its samples and used 30% of them as calibration and validation sets. Similarly, Nair et al. (2021) used 75% of the data available for training and the remaining 25% for testing, Pina et al. (2021) used the first 20 days for training and the following seven days for testing.

4.4.3 Prescriptive analysis

A total of ten studies use optimization models in order to either minimize FW or maximize the SL of products, adopting single-objective linear programming, multi-objective non-linear programming, or something in between. These and other features are shown in Table 4.10.

Most studies embrace a single-objective optimization problem (Vilas et al., 2020; Birisci and McGarvey, 2022; Kabadurmus et al., 2022; van der Walt and Bean, 2022), using linear, non-linear, or mixed-integer programming. On the other hand, three studies adopt a multi-objective optimization problem (Birisci and McGarvey, 2018; Cristóbal et al., 2018; Sedehzadeh and Seifbarghy, 2021). Additionally, two other studies use simulation-based optimization (Buisman et al., 2019; Kayikci et al., 2022) and one that adopts a system dynamics approach (Ching et al., 2019). From the single-objective studies, Vilas et al. (2020) employs a metaheuristic - genetic algorithm - to solve the optimization problem, and Birisci and McGarvey (2022) uses the Hooke-Jeeves algorithm in their problem.

Table 4.10: Summary table for the prescriptive analysis

Authors	Context	Modeling approach	Objective function	Decision variables	Constraints
Birisci and Mc-Garvey (2018)	Food service provider	Multi-objective non-linear programming (derivative-free pattern search algorithm)	Min. waste weight (wasted pre-production inventory of ingredients, wasted post-production inventory of menu items) and shortfalls, min. GHG emission (CO2-eq emissions) and shortfalls	Adjustment to the forecast production level, minimum forecast level for which an item can potentially allow for substitution by leftovers, percentage of forecast demand for a menu item for which leftovers will be substituted	Inventory available, ingredient inventory (pre-production inventory), inventory of prepared menu items (post-production inventory, aka "leftovers"), leftover utilization threshold
Cristóbal et al. (2018)	City	Multi-objective optimization integer linear programming	Max. TEIA (total environmental impact avoided) and make TFC (total annual financial cost) less than or equal to a given target values (different budgets)	Different options for food waste prevention and management to be implemented	Budget
Buisman et al. (2019)	Retail	Simulation-based optimization (mathematical or dynamic programming)	Max. Profit	Safety factor, ordering level, profit, waste (%), shortages (%) in terms of waste, profit, shortages and product quality	Discount levels, replenishment quantities, SL (fixed or dynamic) days, LEFO/FEFO ratio
Ching et al. (2019)	Retail	System dynamics	Min. time a unit spends in the system from production to sale	Inventory policies (FIFO and LIFO) and the quality and sustainability implications of these policies	-
Vilas et al. (2020)	Retail	Mixed-integer programming with a metaheuristic (differential evolution) for global optimization, and interior-point method for local search	Max. SL (use-by date and best-before date)	Film thickness, initial carvacrol concentration in each layer	Safety constraints (Listeria concentration at final time), design constraints (concentration of carvacrol in the food at all times); quality constraints (freshness indicator value at final time)
Sedehzadeh and Seifbarghy (2021)	City	Multi-objective mixed-integer linear programming	Min. economic, environmental and social (unsatisfied food banks demand) value	If a new distribution/collection center (RDC) is opened at a location in a certain period, if capacity is installed at a location for a product in a certain period, quantity of a product transported in a certain period from origin to destination, quantity of a usable returned product transported in a certain period between nodes, number of fleets transported in a certain period between nodes via a mode and capacity level, quantity of a product manufactured at a production/recovery center in a certain period, shortage of a product at a distribution/collection center in a certain period, unused capacity of a fleet mode between pair in a certain period	Changes in the status of RDCs, capacity level that can be installed at a RDC for a product, storage capacity, open state of existing centers, threshold for unsatisfied demand, production amount, balance of product flow in RDC, balance of returned products, capacity restrictions, fleet capacity, total fleet weight, budget
Birisci and Mc-Garvey (2022)	Food service provider	Hooke-Jeeves algorithm (single objective non linear)	Min. FW (in kg or in kg Co2)	184 decision variables: Amount to be produced at the beginning of a day for a menu item, production adjustment to the forecast, threshold for a menu item, substitution percent for a menu item, target shortfall probability, available leftover inventory of a menu item of a certain age at the end of a day, portions served from leftovers of an item of a certain age served during a day	Production adjustment to the forecast, leftover substitution decision, portions of a menu item served from new production during a day, inventory availability, usage of leftovers over new production, amount of food waste generated
Kabadurmus et al. (2022)	Municipalities	Mixed-integer linear programming	Min. total cost of the circular food supply chain network	The amount of food waste type transported through the edge; the amount of food waste type transferred to a certain node; the amount of food waste type treated at a certain treatment node with a certain treatment technology; if a collection, reuse, landfill, incineration facility is established at a certain node; if a treatment center is established at a certain node with a certain treatment technology	-
Kayikci et al. (2022)	Retail	Simulation-based optimization (multi-stage dynamic programming)	Max. profit (price), Min. FW	Pricing strategy	-
van der Walt and Bean (2022)	Food service provider (in-flight)	Stochastic mixed-integer programming (single objective: multi-objective programming through weighted goal programming)	Min. number of excess meals on a flight, max. flight's passenger satisfaction level	Suggested order and production quantity of a meal option for the particular flight instance, number of passengers who will receive their first-choice meal option, number of passengers that are assigned to receive a meal option as a substitute of another meal option, number of meal options available	Passengers' substitution behavior, available stock of meal option, meal quantities are restricted to integer values only, minimum requirement for flight's passenger satisfaction level

There are three main themes that the articles focus on, namely retail (Buisman et al., 2019; Ching et al., 2019; Vilas et al., 2020; Kayikci et al., 2022), food services (Birisci and McGarvey, 2022, 2018; van der Walt and Bean, 2022), and cities (Cristóbal et al., 2018; Sedehzadeh and Seifbarghy, 2021; Kabadurmus et al., 2022). From the four studies that tackle retail, two of them adopt simulation-base optimizations with the objective function being maximizing the profit (Buisman et al., 2019; Kayikci et al., 2022), while the other two focus instead on minimizing the time a unit spends in the system from production to sale (Ching et al., 2019) and maximizing the SL of products (Vilas et al., 2020).

In food services context, the focus is on reducing the amount of food waste generated. The objective function of two studies identified is to minimize the amount in kg of FW produced (Birisci and McGarvey, 2022, 2018), while the objective of the third study is to both minimize the number of excess meals to be served and also maximize client satisfaction (van der Walt and Bean, 2022).

Finally, concerning the papers that study cities, they focus more heavenly on the economic, environmental, and social spheres of waste. For instance, Sedehzadeh and Seifbarghy (2021) uses an objective function that tackles all three simultaneously. Meanwhile, Cristóbal et al. (2018) focuses on the economic and environmental sphere, and Kabadurmus et al. (2022) solely on minimizing the total cost of the circular food supply chain network on a set of municipalities.

The decision variables generally adopted in the studies are usually very content-specific and are not repeatedly used in different studies. Despite that, if we would group them into broader categories, the most dominant category of decision variables are inventory management-related (Birisci and McGarvey, 2018; Buisman et al., 2019; Ching et al., 2019; Birisci and McGarvey, 2022).

The way that the decision variables are presented is also different between studies. Some studies are very detailed in how they present them, e.g. Birisci and McGarvey (2022) that goes into detail regarding their 184 decision variables, while others do not present much detail. For example, Kayikci et al. (2022) only states they are optimizing the pricing strategy.

4.5 Context of the studies

A total of 65 papers in the quantitative stream of the literature involve FW. These studies have different focuses, namely the retail and hospitality sectors, households in small areas, and entire countries/the entire world. A small number of studies also tackle food companies (Garre et al., 2020; Kazancoglu et al., 2018) and hospitals (Melikoglu, 2020). Therefore, when describing these studies, it is vital to consider not only the type of analysis performed but also the context.

4.5.1 Retail sector

Most of the studies explored within the quantitative stream of the literature on FW are set in the retail sector, which is a part of the food supply chain. This is the most explored setting, with a total of 22 papers. However, there are no descriptive analyses in the retail sector, with most studies performing predictive analysis - 7 on demand forecast, and 11 on SL - and only four on prescriptive analysis.

The predictive analyses that tackle demand forecasting consider a large time horizon, ranging from 90 days to five years with a daily periodicity. As for SL prediction, although it is supported by a short time horizon, i.e. a few days at best, the number of observations used for this purpose is considerable due to the high periodicity of data collection, i.e. every 30 seconds to daily data. On the other hand, the prescriptive analysis performed within the retail context focuses on the minimization of profit, maximization of SL, and minimization of time that a food product spends in the system from production to sale.

4.5.2 Hospitality sector

The hospitality sector is the second most popular setting with 14 studies, which includes canteens, restaurants, and hotels. Among these establishments, canteens are the most frequently explored (8 studies). Although most of the studies explore a single setting at a time, some articles address more than one simultaneously, as is the case of Posch et al. (2021), which focuses on a staff canteen and a restaurant.

Out of the three types of analysis - descriptive, predictive, and prescriptive -, the most prominent in the hospitality sector is the descriptive, with eight studies, followed by the predictive with seven studies, and the prescriptive with two. The time horizon of the data collected by these scientific studies varies, ranging from a couple of days to several years. In general, predictive analysis entails a longer time horizon than descriptive and prescriptive analysis. However, it should be noted that the data used for forecasting in the predictive analysis is usually extracted from external sources and does not result from direct measurements.

4.5.3 Final consumers

There is a group of 18 studies that explore several contexts related to the final links in the FSC, including households and consumers. These studies address settings that vary from a small number of people to entire cities. In this category, nine studies implement a descriptive analysis, seven perform a predictive analysis, and only three of these perform a prescriptive analysis.

Nine studies focus on an entire city. From these, four predict what will be the amount of FW produced by employing regression or scenario analysis. Three fit into the descriptive analysis wherein two of these studies resort to primary data to collect the absolute total of kitchen FW. Only two studies carried out a prescriptive analysis focusing on minimizing the economic and environmental consequences of FW generation.

From the nine other studies that have a smaller scope, six measure household waste (Orfanou et al., 2019; Richardson et al., 2019; Dey et al., 2022; Grainger et al., 2018; Fanelli, 2019; Kandemir et al., 2020), two focus on consumer waste (Vittuari et al., 2020; Conrad et al., 2018), and only one measures the waste generated in several municipalities (Kabadurmus et al., 2022).

4.5.4 Country and world population

A number of studies focus on the waste produced by entire countries or even the entire world. From these six studies, there are four that perform predictive analysis of yearly FW amount, where most of them employ regression models (Shapiro-Bengtsen et al., 2020; Hall et al., 2009; Mobaseri et al., 2021). The other study (Hoy et al., 2022) forecasts future FW by employing an ANN. Additionally, there are also two studies that measure FW on a country scale, where both take advantage of secondary data for their analysis (Adelodun and Choi, 2020; Alshabanat et al., 2021).

4.6 Discussion

The topic of FW has been increasing in popularity, as shown by the increasing number of scientific papers on the subject published in the last few years. The number of studies published between 2018 and 2022 is far greater than the number of papers published in the previous years, as seen in Figure 4.2. In fact, throughout these last five years, 52 papers were published, while the combined number of articles published before is 13.

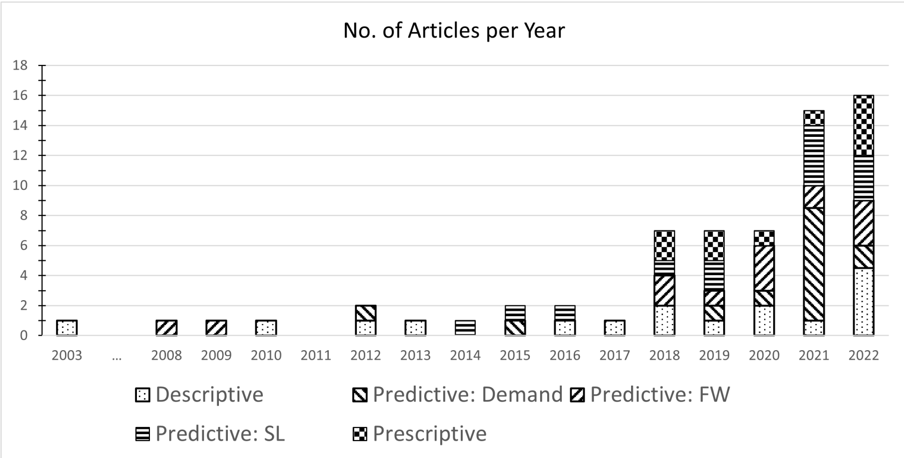

Figure 4.2: Number of quantitative articles throughout the years

Although this trend is valid for all types of quantitative analysis, some types of data analytics studies have a more prominent growth. From the 65 extracted papers, predictive analytics is the most popular approach (demand, FW, and SL forecasts), followed by descriptive and prescriptive ones. In parallel, the most scrutinized setting is the retail sector, followed by the hospitality and final consumer analysis. Although the retail sector is the most common setting, it is the hospitality sector that encapsulates the three different types of data analytics approaches.

Regarding the descriptive analysis, despite the large number of studies published in the last years, this growth has not been as significant as in the other types of approaches. Throughout the years, these studies moved from mainly collecting their own data to retrieving it from a third party. Although in both cases, the amount of FW is gathered through weighting, the way the data is presented can differ. The amount of FW is estimated by weighting the FW of every meal, as seen in case studies focusing on canteens, or by extrapolating the amount of FW from samples of waste for a bigger sample of individuals, as seen in case studies focusing on households.

Moreover, there is no clear consensus on the metrics used to describe and quantify FW, with some studies displaying the absolute amount of FW generated in a specific context, and others presenting a percentage of FW (for example, in a meal). This dissonance, along with others related to the way FW is measured or estimated, makes it difficult to compile and compare different measurements, especially if the information disclosed is insufficient to compute a comparable metric. It is also important to note that these studies address different time horizons when it comes to measurements, as well as different contexts. Thus, the reported values of FW vary greatly, ranging from a few hundred kilograms to millions of kilograms.

Regarding the predictive approach, studies that forecast demand with an integrated FW perspective have only started to appear after 2012. However, the popularity of this specific subject has grown exponentially in recent years, with more scientific papers published in 2021 and 2022 than in all other years combined. The data used in this setting usually has a daily periodicity, and the time horizon that supports the analysis ranges from 90 days to 11 years. Many types of algorithms are adopted, including several causal models and time series models. In the retail setting, the most common algorithms adopted for forecasting are ARIMA (and its variations) and neural networks. In contrast, in the hospitality sector, the most common algorithms are neural networks, ARIMA, TBATS, random forests, and naïve models.

For causal models as a whole, the most popular external variables considered are holidays, day, month, year, and reservations. Different error measures are also employed to test forecast accuracy, with the most common being the MSE and MAE and their variations. Some studies also compare the performance of the developed models with some baseline models to show that more complex algorithms can have real positive effects on the operations of food establishments. In these cases, the results of the forecast models are 3% to 63% better than the ones provided by the baseline models.

Predictive approaches have been applied to the FW forecast setting ever since 2008. However, there are only two studies published before 2017, with the remaining 11 studies being published between 2018 and 2022. The papers dedicated to predicting FW usually handle yearly data and, thus, provide forecasts with a yearly periodicity. Generally, these studies estimate MSW or urban waste and then consider a percentage of it as FW, based on historical data. Moreover, even the studies that forecast FW directly usually employ regression analysis algorithms. Nevertheless, there are also a couple of studies that adopt other types of forecasting techniques, such as random forests and grey rolling models. These forecasting models often use the population size and the income/GDP as predictors. The performance of these models is then commonly evaluated using R2, RMSE, and MAPE measures.

As for the SL forecasts, this is a setting where the first article in this scene appears in 2014. Throughout the remaining years, it has been observed an increase in popularity, the same as the remaining quantitative approaches. Notwithstanding, this popularity has been mainly marked by the retail sector, where almost half of the studies are food supply chain focused, more aimed at inventory management. Both machine learning and empirical models are employed, resulting mainly in the use of support vector machines and the Arrhenius equation, respectively. The most commonly used independent variables in this domain are temperature, humidity, and light-related characteristics. The data collection periods to perform SL forecasts are usually very small, yet their periodicity is also analogously high.

In regard to the prescriptive dimension, optimization models are applied in retail, food service providers, and city settings. These studies have only started to emerge after 2018 and focus on minimizing the amount of FW produced, maximizing the SL / expiration date, or minimizing the consequences produced by the generation of FW. Different types of programming models are considered, such as single-objective linear programming and multi-objective non-linear programming. There are no standard decision variables among the prescriptive analysis studies, these being usually very content-specific. However, they usually fall under the umbrella of inventory management variables. None of the prescriptive studies compare their results to a baseline.

4.7 Conclusions and Future Research

The aim of this chapter is to summarize and synthesize all scientific studies that adopted a quantitative approach to food waste. It should allow for an overall understanding of what different food waste producers can improve in their own environments, how to do it, and what type of results can be expected.

A total of 65 unique studies were explored. Among these, the most prominent data analytics approach is predictive, followed by descriptive and prescriptive analytics respectively. In addition, considering the setting the papers explore, the most popular is the retail sector, tailed by both hospitality and final consumer analysis.

Regarding the descriptive analysis, most of the studies available in the literature are supported by data resulting from food waste weighting. These food waste measurements are not comparable since the contexts and measurement conditions are different. Moreover, some studies present food waste quantification in absolute terms and others in relative terms.

With respect to predictive analysis, the literature presents three avenues: forecasting future demand as a way to prevent FW, forecasting food waste directly, and forecasting the shelf-life of products to avoid FW generation. The data used to forecast future demand generally has a daily periodicity, which is then commonly fed into artificial neural networks and ARIMA models. When external factors are considered, the most notorious external variables are holidays, date-related variables (day, month, year), and reservations (in restaurants and schools). The results of the forecast models point to 3% to 63% better results when compared to a baseline. In relation to forecasting food waste directly, the data has usually a yearly periodicity and regression analysis is the primary technique employed. In this case, population size and income/GDP are the most common predictors adopted. Finally, for shelf-life prediction, the Arrhenius Law, regression models, and support vector machines are the most popular methods adopted. These forecasts are done by leveraging temperature, humidity, and light-related characteristics.

Regarding prescriptive analysis, most studies focus on minimizing the amount of FW produced, maximizing the shelf-life / expiration date, or minimizing the consequences produced by the generation of FW. This is achieved by both single-objective and multi-objective optimization, and also simulation-based approaches. The decision variables used are usually very intertwined with the context in which they are applied, being connected to inventory management.

In conclusion, we believe that this chapter can provide researchers and managers with the necessary background to further develop the quantitative analysis regarding food waste. There are opportunities in each of the three data analytics approaches that can be considered in future research. For example, in the descriptive field, there is space for a framework to standardize food waste measurements which should allow for better and easier comparison between studies. Regarding the predictive sphere, two fronts can be further explored. Firstly, in the case of explanatory models, more context-specific independent variables can be explored. Secondly, exploring more recent machine learning forecast techniques like transformers could also be beneficial to the field, especially when forecasting food waste directly. Lastly, in the prescriptive department, there could be merit in designing an optimization system capable of deriving menus able to minimize the amount of food waste produced. This system could also be designed to maximize the number of meals served and resource allocation.

CHAPTER 5

Machine Learning Models for Short-Term Demand Forecasting in Food Catering Services: A Solution to Reduce Food Waste

Food waste is responsible for severe environmental, social, and economic issues and therefore it is imperative to prevent or at least minimize its generation. The main cause of food waste is poor demand forecasting and so it is essential to improve the accuracy of the tools tasked with these forecasts. The present work proposes four models meant to help food catering services predict food demand accurately and thus avoid overproducing or underproducing. Each model is based on a different machine learning technique. Two baseline models are also proposed to mimic how food catering services estimate future demand and to infer the added value of employing machine learning in this context. To verify the impact of the proposed models, they were tested on data from the three different canteens chosen as case studies. The results show that the models based on the random forest algorithm and the long short-term memory neural network produced the best forecasts, which would lead to a 14 % to 52 % reduction in the number of wasted meals. Furthermore, by basing their decisions on these forecasts, the food catering services would be able to reduce unmet demand by 3 % to 16 % when compared with the forecasts of the baseline models. Thus, employing machine learning to forecast future demand can be very beneficial to food catering services. These forecasts can increase the service level of food services and reduce food waste, mitigating its environmental, social, and economic consequences.

5.1 Introduction

Food loss and waste have severe environmental, social, and economic impacts, such as the production of greenhouse gases, global food security problems, and related costs, respectively (Girotto et al., 2015; Kamble et al., 2020; Martin-Rios et al., 2018). Considering that, in 2011, it was estimated that one-third of all food produced is either lost or wasted, which translates into more than a billion tonnes of food, the impact of food loss and waste is immeasurable (FAO, 2011).

Food loss and waste occur throughout the entire food supply chain. Food loss (FL) happens at an early stage of the food supply chain (agricultural production), and food waste (FW) happens at the final consumer stage (Priefer et al., 2016; Messner et al., 2020). FW has the worst consequences of the two, since at this stage, the food has already been processed, packaged, stored, and possibly distributed (San Martin et al., 2021; Martin-Rios et al., 2021). However, food service companies are not actively innovating in the waste domain (Martin-Rios et al., 2018), despite their owners, chefs, and kitchen managers demonstrating a general willingness to reduce FW (Henchen, 2019). It is, therefore, both urgent and imperative to find ways to reduce FW and to extract value from FW already generated, and then provide this information to such stakeholders.

There are several options when it comes to extracting value from FW. For example, this can be done by motivating and providing the means for customers to take what remains of their meals home (DRAAF, 2014). This can also be achieved by transforming FW into a valuable resource, e.g., biofuel, bioenergy, and biomaterials (Filimonau and De Coteau, 2019; Girotto et al., 2015; Priefer et al., 2016). The most popular techniques to achieve this are biogas production (De Clercq et al., 2019; Curry and Pillay, 2012) and anaerobic digestion (Dai et al., 2013; Li et al., 2018).

Although being able to extract value from FW is essential, preventing its generation is still the most desirable approach to minimize its consequences (Papargyropoulou et al., 2014; Messner et al., 2020). The literature presents different approaches to prevent FW, such as through the use of packaging technologies (Brennan et al., 2021) and data gathering technologies (Martin-Rios et al., 2021). Nevertheless, FW generation is usually caused by food overproduction, which in turn is mostly due to poor demand forecasts (Filimonau and De Coteau, 2019; Parfitt et al., 2010; Pirani and Arafat, 2016). Most companies in the hospitality sector do not trust the accuracy of their forecasts, and therefore cook 10 % more food than their estimated needs in order to avoid stock-outs (Pirani and Arafat, 2016). Therefore, a major effort should be made to develop more accurate and reliable demand forecasting models.

Food demand forecasting models based on machine learning algorithms (such as the ones proposed in the present chapter) have the potential to improve waste management in the HoReCa (Hotel, Restaurant, and Café) sector with their technological innovation (Martin-Rios et al., 2021), providing businesses with a tool capable of contributing to the minimization of FW generation while also preventing underproduction. By knowing the quantities of food they need to produce/serve in advance, businesses can minimize overproduction and thus FW.

The present chapter proposes four food demand forecasting models that can be divided into two categories. The first category comprises two causal models, one based on a random forest (RF) algorithm and the other based on the light gradient-boosting machine (LightGBM) algorithm. The second category comprises two time series models, one based on a long short-term memory (LSTM) neural network and the other based on a transformer neural network. Given the topic explored in the present chapter, the causal models should outperform the time series models.

Causal models adopt a set of independent variables in order to better understand the environment at hand and increase their forecasting capabilities. Within the context of the present chapter, these variables could be the menu, the number of reservations, and date-related variables such as the day of the week, day of the month, and month of the year. On the other hand, most time series models rely only on the previous values of a target variable and its patterns, trends, and cyclicality to forecast future values. Some examples of these models are the ones based on transformer neural networks or the Holt-Winters algorithm. There are however some exceptions that also take advantage of external variables to help forecast future values, such as models based on LSTM neural networks and autoregressive integrated moving average with exogenous variables (ARIMAX) models.

The performance of food demand forecasting models is usually measured and analyzed in relation to the quality of their forecasts. In theory, the closer the forecast demand values are to the real demand, the better the model. It is important to note that in most cases, these models not only seek to minimize FW but also avoid underproduction so as to help the catering services meet their financial objectives (Messner et al., 2020).

This chapter is structured as follows. First, a review of the literature on FW is presented, exploring both the quantitative and predictive branches. Second, the methods adopted in the present chapter are described, and the data collected from each FCS is presented. Then, using data from each FCS, the machine learning-based forecasting models are compared with the baseline models to determine how much FW could have been prevented. The key conclusions from the results are then pointed out and discussed, providing insights and guidelines for future work.

5.2 Literature Review

The literature on FW can be divided according to two types of approach: qualitative and quantitative. As the names suggest, the former focuses on quantifying FW, while the latter highlights the drivers of FW, how to prevent it, and how to allocate it when it occurs.

The qualitative branch is the most popular of the two in terms of the number of studies, which have remained consistent throughout the years (Hebrok and Boks, 2017; Dhir et al., 2020). In 2010, Parfitt et al. (2010) reported a lack of descriptive analysis studies and pointed out that the few that existed used data mainly from the 1980s. In terms of quantitative studies, Schneider (2013) identified four (focusing on the hospitality sector) in 2013, Filimonau and De Coteau (2019) identified 47 in 2019, Dhir et al. (2020) identified 41 in 2020, and Rodrigues et al. (2023) identified 65 in 2022.

Although it is hard to quantify the exact number of existing works, the increasing popularity of the quantitative approach to FW is clear. Rodrigues et al. (2023) studied the methods and data used in the studies published in the field and classified the studies as descriptive, predictive, or prescriptive. Of these three types of analysis, predictive analysis is particularly important when it comes to preventing FW generation. Accurate forecasts mean that FCSs can predict future demand and adapt their production accordingly. When poor demand forecasts are produced, they motivate overproduction, leading to FW (Filimonau and De Coteau, 2019; Parfitt et al., 2010; Pirani and Arafat, 2016).

Demand forecasting can be considered one of many means to prevent FW. Martin-Rios et al. (2021) explored FW management innovations in the food service industry and highlights the focus on providing faster responses to market/customer demands. Innovation theory is just one of the angles from which to tackle FW management, be it by preventing, reducing, or recycling it (Martin-Rios et al., 2018). Another approach to minimizing FW is practice theory. By focusing on the knowledge/competence component, it is possible to identify FW prevention related issues and try to derive alternatives to reduce FW(Hennchen, 2019). Messner et al. (2020) explored the issue from a different angle and presented a paradoxical idea of FW prevention. While the most promising option to reduce the impact of FW is its prevention (Papargyropoulou et al., 2014), FW prevention actions taken by industry and governments can sometimes be distorted to mean management rather than prevention, thus promoting structures that manage rather than reduce food waste. (Messner et al., 2020).

Most of the studies tackling future demand forecasts adopt time series models. These models identify seasonalities and trends to estimate future demand. The algorithms most commonly used in these models are the autoregressive integrated moving average (ARIMA) algorithm and its variations (Møller Christensen et al., 2021; Sakoda et al., 2019), the trigonometric seasonality, Box-Cox transformation, ARMA errors, trend and seasonal components (TBATS) algorithm (Dharmawardane et al., 2021; Posch et al., 2021), naïve algorithms (Møller Christensen et al., 2021; Malefors et al., 2021b), and the Prophet algorithm (Posch et al., 2021; Malefors et al., 2021b). These kinds of models usually benefit from long data collection periods, such as the ones available to Arunraj and Ahrens (2015); Dharmawardane et al. (2021), which ranged from four to ten years.

In addition to studies using time series models, some studies employ causal models that use external variables to better understand the environment at hand and, in theory, produce better forecasts. Different algorithms are used in this context, with neural networks being the most popular (Lee et al., 2020; Faezirad et al., 2021; Sakoda et al., 2019). Other, less popular algorithms also adopted in this field are the RF algorithm (Malefors et al., 2021a) and the extreme learning machine algorithm (Gružauskas et al., 2019). Regarding the external variables adopted in the causal models, they mostly concern the number of reservations and time-related variables (day, month, year, and holidays). Other, less common external variables include the number of Covid-19 infections (Malefors et al., 2021a), weather-related variables (Arunraj and Ahrens, 2015), and the menu served on a given day (Faezirad et al., 2021). It should be noted that food demand forecasting studies generally estimate demand daily, with most performing a next-day forecast (Sakoda et al., 2019; Malefors et al., 2021b; Møller Christensen et al., 2021).

Most of the studies available in the field do not evaluate how the proposed forecasting models impact FW reduction. In fact, to the best of the authors' knowledge, only Dharmawardane et al. (2021); Lee et al. (2020); Arunraj and Ahrens (2015) verified the added value of the proposed models by comparing them with a baseline model. The first study compared a baseline model based on the Holt-Winters algorithm with the developed TBATS model (Dharmawardane et al., 2021) using data from two different settings. The developed model beat the baseline model in the first setting but lost in the second. The second study adopted an LSTM-based model that achieved a higher F1 score than a baseline model based on the collaborative filtering technique (Lee et al., 2020). Lastly, the third study employed four models. Three ARIMA variants (SARIMA-MLR, SARIMA-QR, and SARIMA) and a multi-layer perceptron (MLP) neural network that produced a lower root mean square error (RMSE) and mean average percentage error (MAPE) than a baseline model based on a seasonal naïve algorithm (Arunraj and Ahrens, 2015). Although there are other studies that mention improvements over a baseline model (Malefors et al., 2021a,b; Gružauskas et al., 2019; Faezirad et al., 2021), even quantifying these improvements (albeit not always via error measures), they do not mention the techniques used to produce these baseline models.

In short, the use of forecasting models (such as those based on machine learning algorithms) to reduce FW is a topic that is still in an early stage of development. There is a need for further studies on this topic, particularly with a focus on causal models that include more diverse variables and on exploring the potential of advanced machine learning algorithms. Moreover, more studies should evaluate how the proposed models impact FW reduction in different settings. The present chapter proposes to do exactly that. It contributes to the literature by proposing a set of food demand forecasting models that include a diverse set of variables and are based on advanced machine learning algorithms that have yet to be explored in this context, and it innovates by testing the added value of the proposed models on data from three different FCSs.

5.3 Methodology and Data

The present chapter seeks to quantify how beneficial food demand forecasting can be for the food catering sector. To that end, four food demand forecasting models were developed, i.e. two causal models and two time series models. Each model was based on a different machine learning algorithm, and all models were designed to predict demand in the short future (next-day forecasts). The forecasts produced by these models should aid managers in determining the quantity of food to prepare and the resources that should be available/allocated to meet the estimated demand.

To validate the applicability and potential of the proposed food demand forecasting models, these were tested on data from three food catering services (FCSs): two student canteens and a company canteen. Since the three units have different inherent features, the goal is also to explore how the different algorithms may impact the quality of the forecasts. The data gathered from these FCSs allow the use of an entire year to test the models' ability to forecast the food demand and consequently reduce FW. To quantify the theoretical FW reduction provided by the proposed models and thus evaluate their added value, they were compared with two baseline models designed to mimic how FCSs typically estimate future demand.

In the case of the causal models and the model based on the LSTM neural network, external variables were specified. These variables should be able to represent the environment of the different FCSs and the day for which the demand is being forecast. Some of these external variables were collected by the FCSs directly, e.g., the number of reservations and the menu for the day. In contrast, other variables such as weather data were extracted from alternative databases. The variables considered for each FCS are detailed in the following section. As for the model based on the transformer neural network, it was not necessary to specify any variables other than the demand.

The hyperparameters of the different employed algorithms were optimized through nested cross-validation. Additionally, by comparing the predicted demand with the actual demand recorded by the FCSs, it was possible to compute different metrics that quantify the forecasting performance of the proposed models. Then, by analyzing these metrics, it was possible to determine the model with the best performance, i.e., that can lead to the highest reduction in FW.

Nevertheless, to assess the impact and infer the added value of the proposed models, the produced demand forecasts should be compared with the demand estimated by the decision-makers at the FCSs used as case studies. However, since these decision-makers did not make such registers, two baseline models designed to mimic how they might have forecast the demand were used instead.

The following subsections provide a detailed description of the methodology and data used. First, a more in-depth description of the data available for each FCS is provided, indicating the time horizon and the variables adopted in the models. Next, the models proposed are introduced, the implementation of these models is explained, and the hyperparameter tuning process is presented. Lastly, the baseline models proposed are described in detail.

5.3.1 Data available

Each FCS provided data related to a different set of factors that were unique to their environments. It was also possible to extract and generate other variables that were not directly related to the FCS in question. Some examples of the latter include variables related to the date for which the demand is being forecast, such as the day of the week, the month, whether it is a holiday, or whether it is during the Queima week (a week-long traditional academic festival that marks the end of the academic year for university students and during which there are no classes), and variables related to the weather predicted for the same day, such as temperature and precipitation. The weather-related variables were extracted from the Visual Crossing database (Crossing, 2003).

5.3.1.1 Food Catering Service 1 - Student canteen

One of the FCSs used as case studies operates a student canteen and will henceforward be referred to as FCS1. This canteen serves an engineering school with over 8,000 students. Since this school is quite large, besides the canteen operated by FCS1, two other canteens serve the students to ensure their needs are met. Nonetheless, FCS1 still records an average of 400 daily meals served.

The data available for FCS1 includes historical demand by dish typology (such as meat, fish, and vegetarian) and the daily number of students who have classes only in the morning, only in the afternoon, or in both periods. This data refers to the period from January 2017 to December 2019. The number of daily meals served by dish typology is displayed in Figure 5.1. Looking at the y-axis, it is apparent that the popularity of the three dishes is very different.

However, their distribution throughout the year seems to be relatively similar and constant. By plotting the sum of the demand of the three dishes, as seen in Figure 5.2, this distribution becomes clearer. First, a constant number of meals served throughout the year is depicted via the blue line, as the demand was fitted to it, showing almost no slope. Second, there is a clear drop in the number of meals served at the end of June, only to return to normal levels in October. This is due to students no longer having classes at the end of May, even though some students still have exams and therefore still attend the university. Additionally, this is also a time when both students and university staff are on vacation, which further drives the demand down. Although students do not have classes at the end of the year, this period is much shorter and its effects are not felt as much, except for the last two weeks of December due to the Christmas and New Year's weeks. Finally, the drop in demand around the first week of May is due to this being the Queima week.

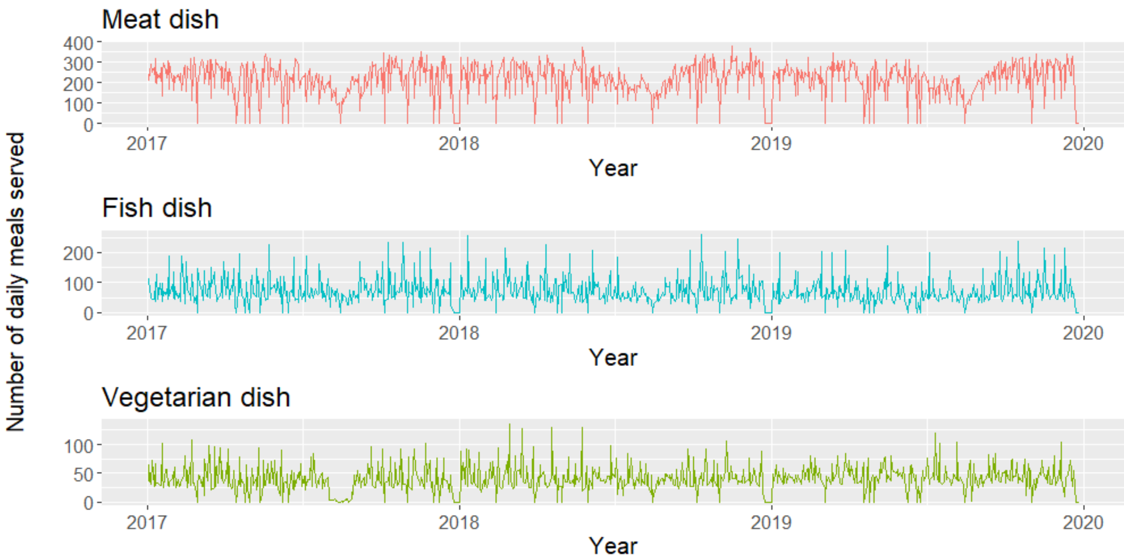

Figure 5.1: Number of daily meals served by FCS1 separated by dish

Regarding the number of daily meals served, these are better displayed in Figure 5.3, where the boxplots for each dish are plotted side by side. Although there are many types of dishes, it is apparent that the diet, extras, and salad dishes are not as popular as the rest. In fact, there were days when no such dishes were served at all. On the other hand, meat, fish, and vegetarian dishes are significantly more popular. In particular, the meat dish was served more times than the fish and vegetarian dishes combined. While the meat dish only has bottom outliers, i.e., values lower than 1.5 times the Interquartile Range (IQR) from the bottom hinge (the 25th percentile), the fish and vegetarian dishes have many upper outliers, i.e., values higher than 1.5 * IQR from the top hinge (the 75th percentile), alluding to a more irregular distribution of the demand for the latter.

In addition to the historical demand by dish typology and the daily number of students with classes, date and weather-related variables were also included in the models proposed for FCS1. In the case of holidays, not only was it noted whether the day was a holiday, but also if the next or previous three days were holidays as well. Table 5.1 provides more details about the available data.

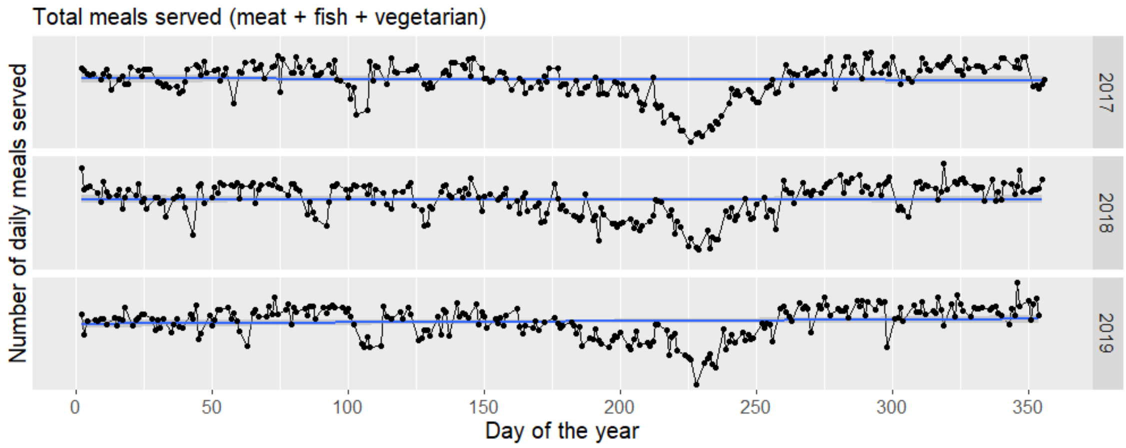

Figure 5.2: Number of daily meals served by FCS1

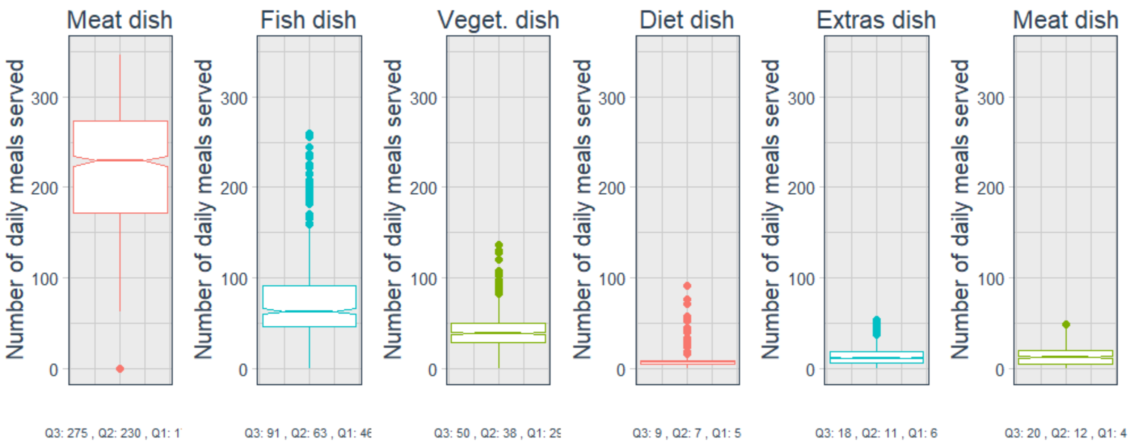

Figure 5.3: Number of daily meals served by FCS1 separated by dish

Table 5.1: Summary of the variables available for FCS1

Variable name	Type	Description
Dish_meat; Dish_fish; Dish_vegetarian; Dish_diet; Dish_extra; Dish_total; Dish_soups	Continuous	Daily number of meals served for the meat, fish, vegetarian, diet, extras, and soup dishes, respectively
Students_Morning; Students_Afternoon; Students_AllDay	Continuous	Daily number of students who have classes in the morning, afternoon, or all day
Day	Discrete	Day of the month
Week_Monday; Week_Tuesday; Week_Wednesday; Week_Thursday	Binary	Day of the week
Month_January; Month_February; ... ; Month_October; Month_November	Binary	Month of the year
Holiday	Binary	Holiday or not
HolidayAfter1; HolidayAfter2; HolidayAfter3	Binary	Holiday in the next day, in two days, or in three days, respectively
HolidayBefore1; HolidayBefore2; HolidayBefore3	Binary	Holiday in the previous day, two days ago, or three days ago, respectively
Temperature	Continuous	Average temperature for the day, measured in Celsius
Precipitation	Continuous	Average precipitation for the day, measured in millimeters
Weather_Condition_Partially_Cloudy; Weather_Condition_Rain; Weather_Condition_Rain_Overcast; Weather_Condition_Rain_Partially_Cloudy; Weather_Condition_Rain; Weather_Condition_Wind	Binary	Weather condition

Time horizon: January 2017 to December 2019

5.3.1.2 Food Catering Service 2 - Student canteen

The second FCS (FCS2) also operates a student canteen but in a different setting. In particular, it serves around 4,000 students from a medical school. Although it is the only canteen in its university, there are other FCSs nearby. This explains why FCS2 only serves around 225 daily meals, a small portion of the total population it could serve. The competition around FCS2 includes the canteen from a neighboring hospital and several restaurants (including fast food restaurants) in a nearby shopping mall.

The data provided by FCS2 includes the historical demand by dish typology (meat, fish, or vegetarian) and the menu for each day between July 2016 and December 2019. The menus were subsequently categorized according to the ingredients used, e.g., type of meat and type of fish, and how they were cooked, e.g., grilled and stewed.

The daily number of meals served varies according to the dish typology, as displayed in 5.4. The meat dish is also the most popular in the context of FCS2, although with a volatile distribution. When plotting the sum of the demand for the three dishes, the trend that is drawn is similar to that of FCS1's. However, the drop in demand from June to October is more accentuated. As with FCS1, this drop is likely due to students not having classes during this period and also because it is a vacation period for the university.

The meat dish represents more than half of all dishes served. However, this FCS operates on a smaller scale than the previous one. This, allied with a more volatile distribution, leads to the

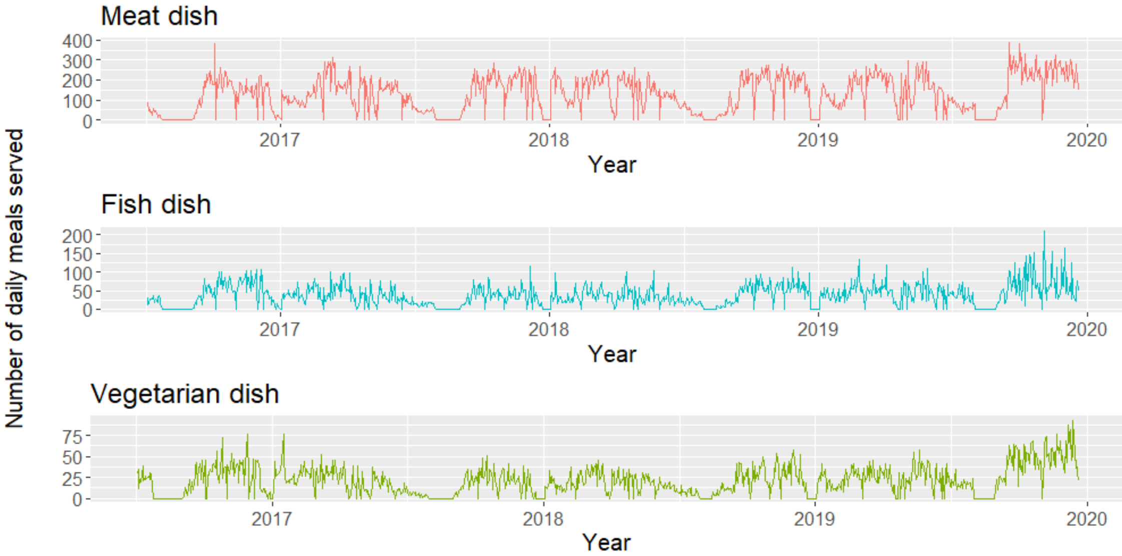

Figure 5.4: Number of daily meals served by FCS2 separated by dish

number of meat meals served not having any lower outliers when plotting its values in a boxplot (Figure 5.6). As for the fish and vegetarian dishes, these still have upper outliers.

Other information taken into account includes the day of the month, the day of the week, and the month of the day for each recorded observation. Additionally, it was noted whether that day and the three next or previous days were holidays. The temperature and precipitation were also considered to help forecast future demand values. Table 5.2 provides a more comprehensive overview of the available variables.

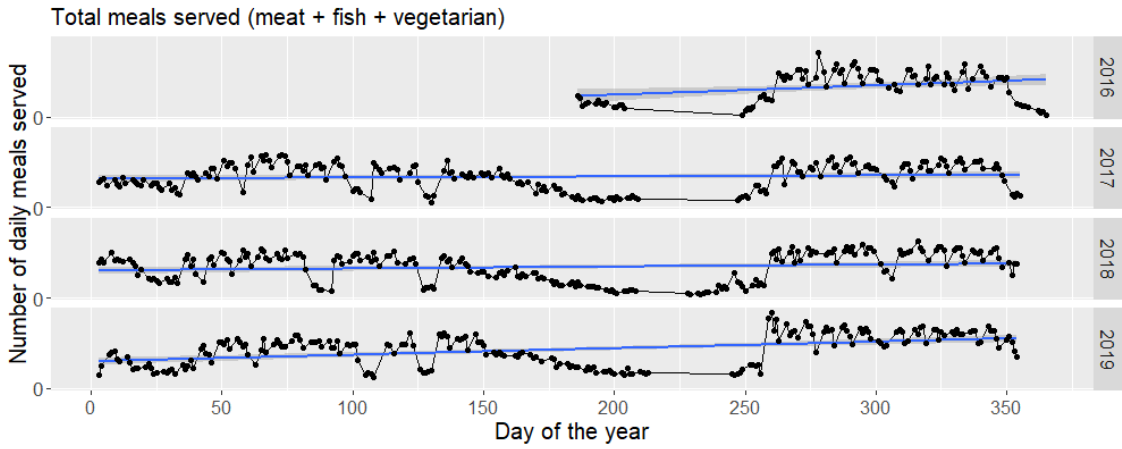

Figure 5.5: Number of daily meals served by FCS2

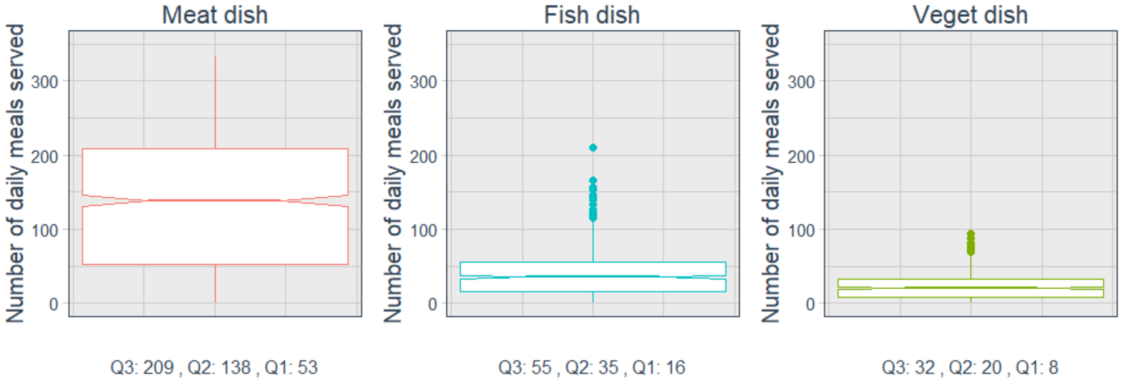

Figure 5.6: Number of daily meals served by FCS2 separated by dish

Table 5.2: Summary of the variables available for FCS2

Variable's name	Type	Description
Dish_meat; Dish_fish; Dish_vegetarian	Continuous	Daily number of meals served for the meat, fish, and vegetarian dishes, respectively
Meat_pork; Meat_beef; Meat_chicken; Meat_roast; Meat_grilled; Meat_stew; Meat_boiled; Meat_fried; Meat_others	Binary	Meat menu classification
Fish_forkbeard; Fish_tuna; Fish_codfish; Fish_horsemackerel; Fish_seafood; Fish_perch; Fish_hake; Fish_octopus_squid; Fish_sardine; Fish_salmon; Fish_flounder; Fish_redfish; Fish_swordfish; Fish_whiting; Fish_dogfish	Binary	Fish menu classification
Vegetarian_soya; Vegetarian_tofu; Vegetarian_seitan; Vegetarian_vegetables	Binary	Vegetarian menu classification
Day	Discrete	Day of the month
Week_Monday; Week_Tuesday; Week_Wednesday; Week_Thursday	Binary	Day of the week
Month_January; Month_February; … ; Month_October; Month_November	Binary	Month of the year
Holiday	Binary	Holiday or not
HolidayAfter1; HolidayAfter2; HolidayAfter3	Binary	Holiday in the next day, in two days, or in three days, respectively
HolidayBefore1; HolidayBefore2; HolidayBefore3	Binary	Holiday in the previous day, two days ago, or three days ago, respectively
Temperature	Continuous	Average temperature forecast for the day, measured in Celsius
Precipitation	Continuous	Average precipitation forecast for the day, measured in millimeters
Weather_Condition_Partially_Cloudy; Weather_Condition_Rain; Weather_Condition_Rain_Overcast; Weather_Condition_Rain_Partially_Cloudy; Weather_Condition_Rain; Weather_Condition_Wind	Binary	Weather condition

Time horizon: July 2016 to December 2019

5.3.1.3 Food Catering Service 3 - Company canteen

The final FCS (FCS3) used as case study is different from the other two because it serves the workers of a company instead of students. This is an engineering and technology company with around 3,700 employees in its headquarters. Unlike the other FCSs, this one operates a canteen that serves a significant portion of the company's workers, handling around 1,900 daily meals.

The data provided by FCS3 contains the number of meals served by type of dish (meat, fish, or diet), as well as reservations for each of the dishes, which are announced two days in advance. The available records refer to the period between September 2011 and December 2019. Therefore, this is the FCS with the most prolonged data collection horizon.

Additionally, the average number of meals served daily by FCS3 is significantly higher than the two other FCSs, as seen in Figure 5.7. In this figure, it is possible to see that there is a growing trend in the number of meat meals served throughout the years, while the same is not as obvious for the fish and diet dishes.

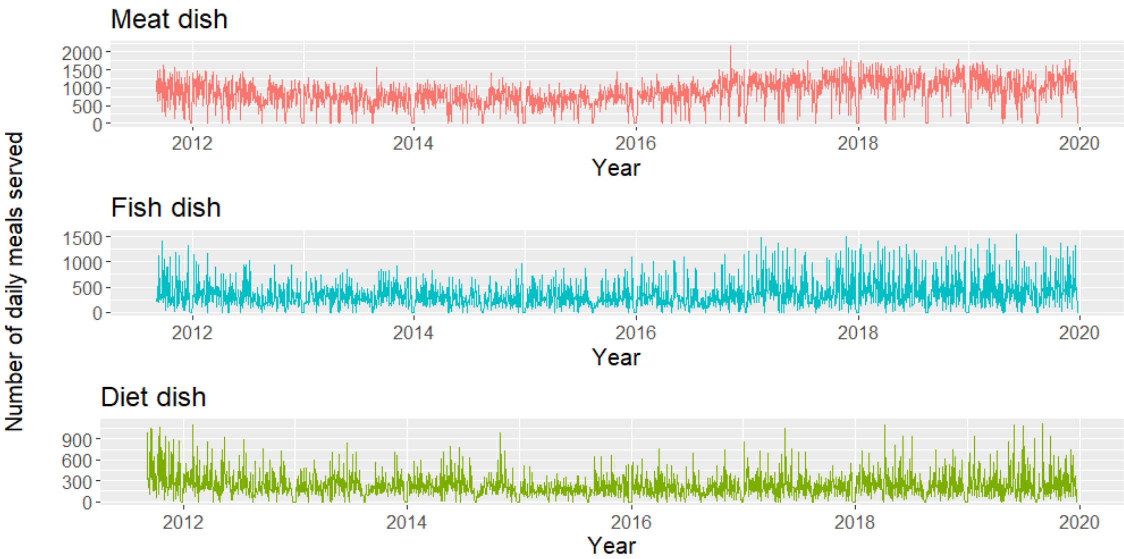

Figure 5.7: Number of daily meals served by FCS3 separated by dish

The overall demand trend of FCS3 becomes more evident when plotting the total number of meals served, as presented in Figure 5.8. In general, demand remained constant throughout the years, with some dips presumably due to holidays or other extraordinary occurrences. As with the other two school canteens, demand was lower in the summer. Although less accentuated, there were also days with no demand whatsoever, which may symbolize that the company, canteen, or both were closed for vacations.

Just like in the other two canteens, the meat dish was the most consumed, being served significantly more than the other dishes. Possibly due to its popularity, there is only a single outlier in the number of meat dishes served. This can be visualized in Figure 5.9. While the fish and diet dishes were served less often than the meat dish, there were days when demand for the former peaked very high. In the case of salad dishes, the median values of meals served were higher than the previous two dishes, but the demand for salads was more constant and did not reach such high peaks. Finally, the vegetarian option seems to be the least popular in this company.

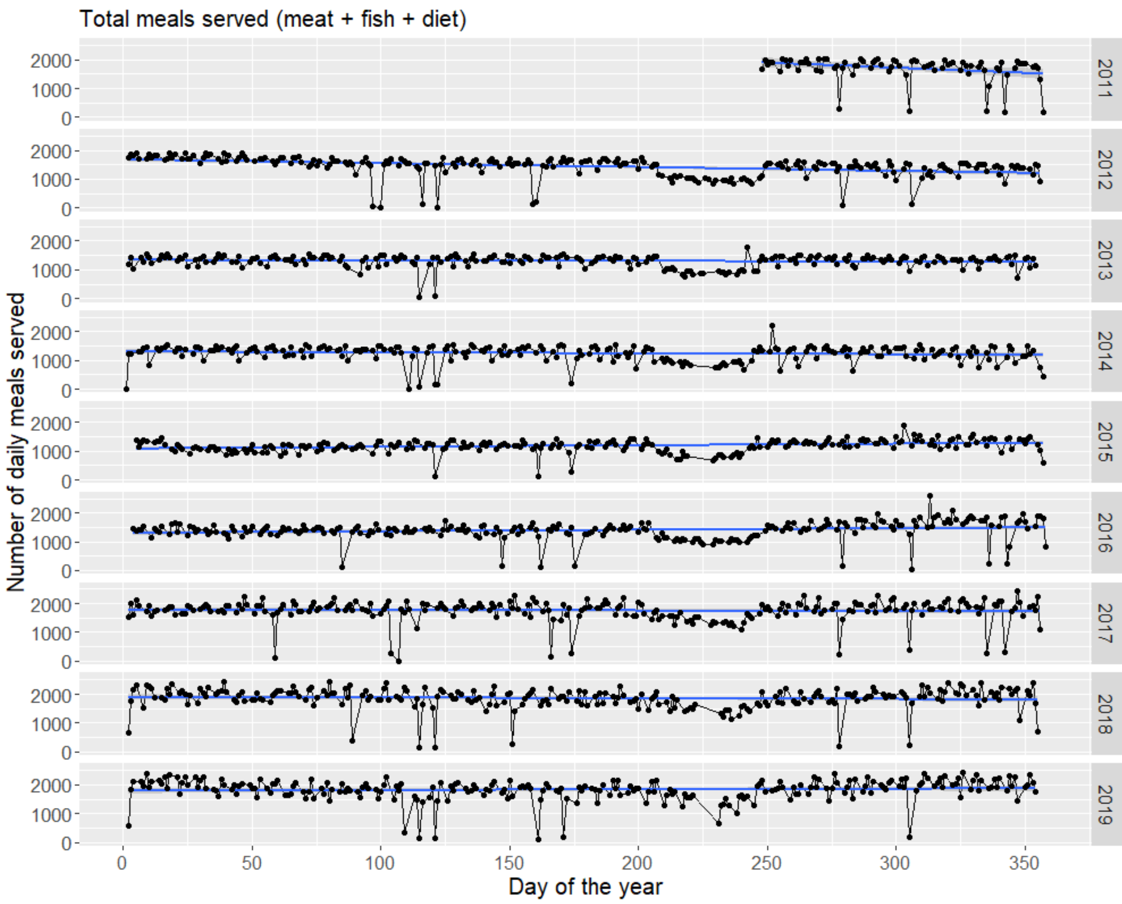

Figure 5.8: Number of daily meals served by FCS3

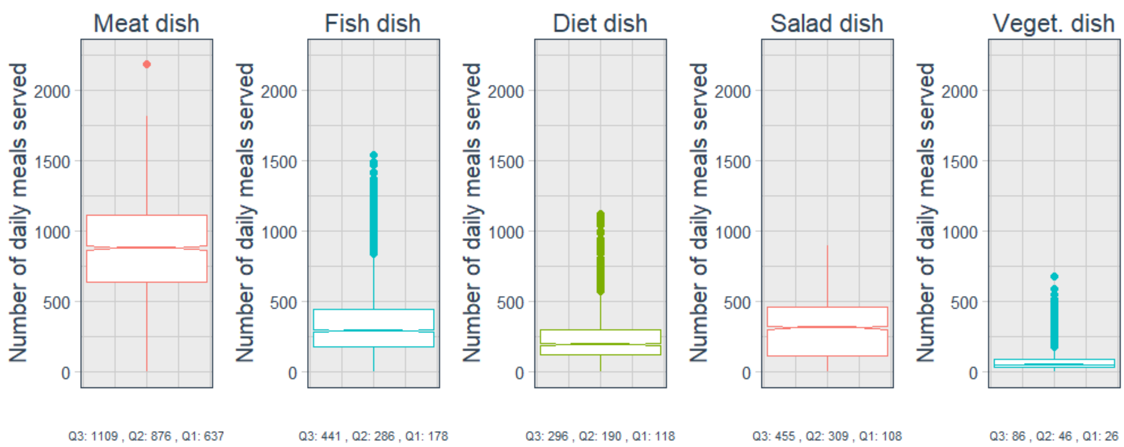

Figure 5.9: Number of daily meals served by FCS3 separated by dish

As with the previous FCSs, date and weather-related variables were also considered for FCS3. Additionally, the number of reservations for a given day was available in the case of FCS3 (the employees of this company need to show interest in the dish they want two days in advance, although they can still attend the canteen without a reservation). A summary of the variables available for FCS3 is displayed in Table 5.3

The available data relates to the period from September 2011 to December 2019.

Table 5.3: Summary of the variables available for FCS3

Variable name	Type	Description
Dish_meat; Dish_fish; Dish_diet; Dish_salads; Dish_vegetarian	Continuous	Daily number of meals served for the meat, fish, diet, salad, and vegetarian dishes, respectively
Reservation_meat; Reservation_fish; Reservation_diet; Reservation_salads; Reservation_vegetarian	Continuous	Daily number of reservations for the meat, fish, diet, salad, and vegetarian dishes, respectively
Day	Discrete	Day of the month
Week_Monday; Week_Tuesday; Week_Wednesday; Week_Thursday	Binary	Day of the week
Month_January; Month_February; ... ; Month_October; Month_November	Binary	Month of the year
Holiday	Binary	Holiday or not
HolidayAfter1; HolidayAfter2; HolidayAfter3	Binary	Holiday in the next day, in two days, or in three days, respectively
HolidayBefore1; HolidayBefore2; HolidayBefore3	Binary	Holiday in the previous day, two days ago, or three days ago, respectively
Temperature	Continuous	Average temperature for the day, measured in Celsius
Precipitation	Continuous	Average precipitation for the day, measured in millimeters
Weather_Condition_Partially_Cloudy; Weather_Condition_Rain; Weather_Condition_Rain_Overcast; Weather_Condition_Rain_Partially_Cloudy; Weather_Condition_Rain; Weather_Condition_Wind	Binary	Weather condition

Time horizon: September 2011 to December 2019

5.3.2 Models proposed

Four food demand forecasting models were developed and then refined (to enhance their predictive capabilities). These models were designed to improve the operations of FCSs by outputting reliable next-day forecasts. In other words, considering today to be moment t, i.e., the moment when the forecast is produced, the aim is to forecast demand for $t+1$, i.e., the next day. To better illustrate the potential of the developed models, the present chapter focused on forecasting the demand for the meat dish, as this was the type of dish served most often by all three FCSs. This means that the dependent variable of the developed models is the daily number of meat dishes served.

To further test the duality and performance of short-term demand forecasting, two types of forecasting models were employed and tested: time series and causal models. Time series models start by analyzing a time series and identifying patterns, trends, and cyclicalities. Then, they replicate the results to predict future values that are aligned with the setting in question. On the other hand, causal models employ exogenous variables to help categorize the environment in question and use these variables to predict a target variable. Both have their advantages and disadvantages and their performance depends on the setting at hand.

The causal models were based on the RF and LighGBM algorithms, while the time series models were based on LSTM and transformer neural networks. Considering the time horizon and the number of variables available for each FCS, the causal models should outperform the time series models in the case of FCS1 and FCS2. On the other hand, the model based on the LSTM neural network should outperform the causal models in the case of FCS3, since the time horizon is significantly wider. Lastly, the regression side of transformer neural networks has not been widely adopted in the literature yet. Hence, this is an opportunity to assess its performance in a real-life scenario considering both a small and a large time horizon.

causal models rely on external factors known as independent variables. In the present chapter, the variables considered were the number of meals served, the menu, date-related features, weather-related features, the number of students expected to be attending classes, and the number of reservations. These variables were used in both causal models and in the model based on the LSTM neural network. A more comprehensive description of all the independent variables used in the present chapter is available in Tables 5.1-5.3.

There are some variables whose values can be known before the day for which demand is being forecast, e.g., the number of students expected to attend classes on a given day, and others whose values can only be known after their respective occurrences, e.g., the number of meals served for a given dish. In other words, considering today to be moment t and supposing that the forecast is for the next day (moment $t+1$), the number of students expected to attend classes on $t+1$ and the number of meals served in moment t are used as independent variables. Moreover, for variables with a known $t+1$ value, the values for moments $t+1, t, t-1, t-2$, and $t-3$ are used as input. For example, the number of students expected to attend classes on the next day and the number of students who attended classes today and in the three previous days are used to forecast tomorrow's demand. On the other hand, for variables with a value that is only known in moment t, the values for moments $t, t-1, t-2, t-3$, and $t-4$ are used as input. This was the case for the independent variables, such as the number of meals of the three dishes that were served. The list of time-dependent variables adopted and their corresponding time horizon coverage is displayed in Table 5.4.

Table 5.4: Independent variables used in moment t to forecast the demand for moment t+1

Type	Variable	Corresponding days		Variable type
		From	To	
No. of meals served	Dish_meat; Dish_fish; Dish_vegetarian	t	t-4	Continuous
Menu classification	Meat_pork; Meat_beef; ...; Vegetarian_seitan; Vegetarian_vegetables	t+1	t-3	Binary
Date	Day	t+1	t-3	Discrete
	Week_Monday; ...; Week_Thursday; Month_January; ... ; Month_November; Holiday; HolidayAfter1; ...; HolidayBefore3	t+1	t-3	Binary
Weather	Temperature; Precipitation	t+1	t-3	Continuous
	Weather_Condition_Partially_Cloudy; ...; Weather_Condition_Wind	t+1	t-3	Binary

5.3.3 Hyperparameters and Model Validation

Several iterations of the developed models were put to the test (performing next-day forecasts), wherein each iteration had a different combination of hyperparameters. The platform used for these tests was Google Colab, which employs Python. The packages responsible for running the different algorithms were "tensorflow" (LSTM), "lightgbm" (LightGBM), "sklearn" (RF), and "darts" (transformer).

Different hyperparameters were tuned to maximize the algorithms' predictive power for a specific data set. For the RF, the tuned hyperparameters were the percentage of variables to use and number of trees. For the LightGBM algorithm, they were the percentage of variables to use, the learning rate, and the type of boosting. For the LSTM neural network, they were the number of neurons in the first and second layers. And finally, for the transformer neural network, they were the number of expected features in the encoder/decoder inputs and the number of encoder/decoder layers. Different combinations of hyperparameter values were defined for each algorithm and validated before doing another test on new data. The specific values tested for each hyperparameter are shown in Table 5.5.

Table 5.5: Hyperparameters considered for optimization

Algorithm	Hyperparameter	Values	Number of combinations
Random Forest	% of features used	0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1	100
	Number of trees	100; 200; 300; 400; 500; 600; 700; 800; 900; 1000	
LightGBM	Boosting	GBDT, DART, GOSS	300
	Learning rate	0,05; 0,10; 0,15; 0,20; 0,25; 0,30; 0,35; 0,40; 0,45; 0,50	
	% of features used	0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1	
LSTM	Number of neurons in the first layer	10; 50; 100	9
	Number of neurons in the second layer	10; 50; 100	
Transformer	Number of expected features in the encoder/decoder inputs	16; 32; 64	9
	Number of encoder/decoder layers	2; 4; 6	

The technique used in the development and testing of the machine learning models was a nested cross-validation with a rolling window, as depicted in Figure 5.10. This technique employs a sequential rolling window approach, initially dividing the dataset into training and test sets through an outer loop. Subsequently, within this training set, a secondary rolling window method is applied to partition it into training and validation subsets, constituting an inner loop. One of the advantages of this cross-validation technique is that it can be applied to both causal and time series models.

In the present chapter, the test period (outer loop) for the FCSs was considered to be the last year for which the data was available. Coincidentally, it was the year 2019 for all FCSs. Therefore, the test period corresponded to the 52 weeks/260 weekdays of 2019. As in a normal rolling window, the test period (2019) was divided into 52 folds with five observations each, i.e., each fold corresponded to a week. In other words, the models forecast future values in groups of weeks.

The validation period was part of the inner loop, meaning it was indexed to the test period. More specifically, the validation period corresponded to the 20 weekdays before the testing period. Since the test period was divided into folds, the validation period also moved with the fold considered at any given moment. For instance, for the first fold of the test period, i.e., the first week of 2019, the validation period corresponded to the last four weeks of 2018; and for the second fold of the test period, i.e., the second week of 2019, the validation period corresponded to the first week of 2019 and the last three weeks of 2018.

The rolling window technique was also applied to the inner loop. Then, the validation period, i.e., the last 20 weekdays before the test period fold, was divided into four folds, each composed of five weekdays. Thus, as with the outer loop, demand was forecast in groups of weeks, but this time for those 20 weekdays. A visual representation of the inner loop is displayed in Figure 5.10.

Considering the last weekday of 2018 to be moment t, and the first weekday of 2019 to be moment $t+1$, the first fold of the test period would be from moment $t+1$ to moment $t+5$, while the validation period in the inner loop for this specific fold would be from moment $t-19$ to moment t. However, the validation period would be further divided into four folds, e.g., from moment $t-19$ to moment $t-15$, from moment $t-14$ to moment $t-10$, from moment $t-9$ to moment $t-5$, and from moment $t-4$ to moment t.

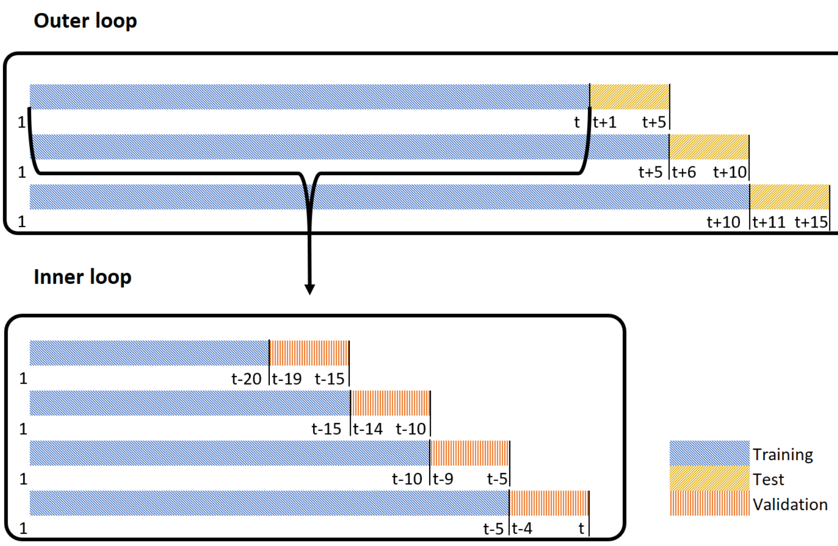

Figure 5.10: Nested cross-validation with a rolling window

The validation period was used to estimate the forecast performance of different hyperparameters throughout a set of 20 weekdays. The hyperparameters that achieved the best forecasts in that validation period were then used to forecast the values in the fold of the test period that the validation period was indexed to. Thus, each fold of the test period can have different hyperparameters. Nevertheless, the overall forecast performance of the whole test period was considered to be the potential performance of the model.

The forecast performance in the validation and test periods was assessed via error measures. Although each error measure has its benefits and demerits, the one chosen to determine the best set of hyperparameters in the present chapter was the RMSE. The RMSE was chosen over the commonly adopted MAPE since the former penalizes outliers more heavily than the latter. This is crucial in the context of the present chapter, where it is more important to be consistently close to the real demand than to be spot on some of the time and wildly inaccurate the rest of the time. Furthermore, FCSs can adapt their operations on the spot, and the more negligible the deviation between the forecast and the actual demand, the easier it is for them to do that adaptation. It should be noted that minimizing the RMSE can help prevent not only the overproduction of meals but also the underproduction of meals, which is an important factor for catering services, which are of course driven by economic goals.

Nevertheless, several other performance metrics were examined. Firstly, the difference between the number of meat meals estimated and the actual demand was calculated. This resulted in three values: the number of meals wasted, i.e., the forecast above the actual demand; the unmet demand, i.e., the forecast below the actual demand; and the absolute difference between the forecast and the actual values. These values allowed quantifying the deviation between the actual demand and the produced forecast. At the same time, error measures such as the RMSE, MAPE, and the coefficient of determination (R2) were computed and analyzed in order to obtain a metric that could be used to easily compare the different models not only when looking at the same FCSs, but also when looking at different FCSs.

5.3.4 Baseline models

To quantify the benefits provided by the proposed models, the estimates from the models should be compared with the estimates from the decision-makers at the FCSs. However, these FCSs did not have such records, and a baseline model had to be created instead. Ergo, two preliminary baseline models were designed to mimic how food demand estimates would have been made. One of the preliminary baseline models was based on a naïve approach and the other was based on a moving average (MA) approach.

The naïve approach estimates demand by considering that the demand will be the same as the demand observed in a previous period. For this approach, different look-back periods were considered: 1, 5, 10, 15, 20, 25, and 30 days. Thus, for instance, when forecasting the demand for $t+1$, the demand could be considered to be the same as in the previous day, i.e., moment t, or the same as 30 days ago, i.e., moment $t-29$. Then, the look-back period that produces the best results is chosen.

Concerning the MA approach, the demand estimated for a given day will be equal to the average demand of a set of previous days. The set of days considered corresponded to the 5, 10, 15, 20, 25, and 30 previous days, which is similar to the number of days contemplated for the naïve approach. This means that, for instance, when forecasting demand for $t+1$, the demand could be considered to be the same as the mean of the last five days, i.e., the average demand between moments $t-4$ and t, or the mean of the last 30 known days, i.e., the average demand between moment $t-29$ and t. Then, the period that bears the best results is chosen.

5.4 Results

The forecast results demonstrate the benefits of developing models that predict future demand in the context of preventing FW. Most of the proposed models were able to beat the baseline models responsible for mimicking how managers do their forecasts. This means that by employing more sophisticated forecasting techniques, it is possible to better predict future demand. Thus, these forecasting models constitute a valuable asset for managers to make better decisions and ultimately minimize FW generation and maximize the number of meals served.

The forecast performance (measured through forecast differences and error measures) of the best model for each algorithm is detailed in Tables 5.6-5.8. The model based on the RF algorithm produced the best forecasts for the first two FCSs, while the LSTM algorithm had the better results for FCS3, i.e., its results had the lowest total absolute difference between the forecast and actual values, which in turn means the overall lowest RMSE. In the case of the baseline models, the one based on the MA approach was always better than the one based on the simple naïve approach.

5.4.1 Food Catering Service 1 - Student canteen

For the first FCS, the error metrics and the absolute difference between the forecast and actual values are displayed in Table 5.6. The model based on the RF algorithm was the best out of all developed models in terms of error metrics, achieving a RMSE of 51.57. As for the baseline models, the one based on the MA approach was the best when considering the average of the last 15 days.

Table 5.6: Results of the forecasting models for FCS1

	Error measures			Forecast differences		
Model	RMSE	MAPE	R2	Wasted meals	Unmet demand	Total absolute difference
RF	51.57	0.2214	0.2805	4,671	5,276	9,947
LightGBM	57.12	0.2400	0.1819	4,867	6,322	11,189
LSTM	65.70	0.2638	0.1397	5,901	6,518	12,419
Transformer	57.60	0.2367	0.1702	3,869	6,577	10,446
Moving Average (15 days)	55.65	0.2460	0.1726	5,453	5,471	10,924
Naïve (5 days)	75.17	0.3053	0.0059	7,496	7,189	14,685

Note: Around 56,000 meat meals were served in 2019, which translates into roughly 215 daily meals

By basing their decisions on the forecasts provided by the RF algorithm instead of an MA of the last 15 known days, this FCS would be able to reduce the number of meals wasted in 2019 by 782, which translates into a 14 % reduction in the FW generated. At the same time, it would be able to reduce the unmet demand in the same period by 195 dishes. A visual representation of the actual demand and the forecast demand produced by the model based on the RF algorithm is displayed in Figure 5.11.

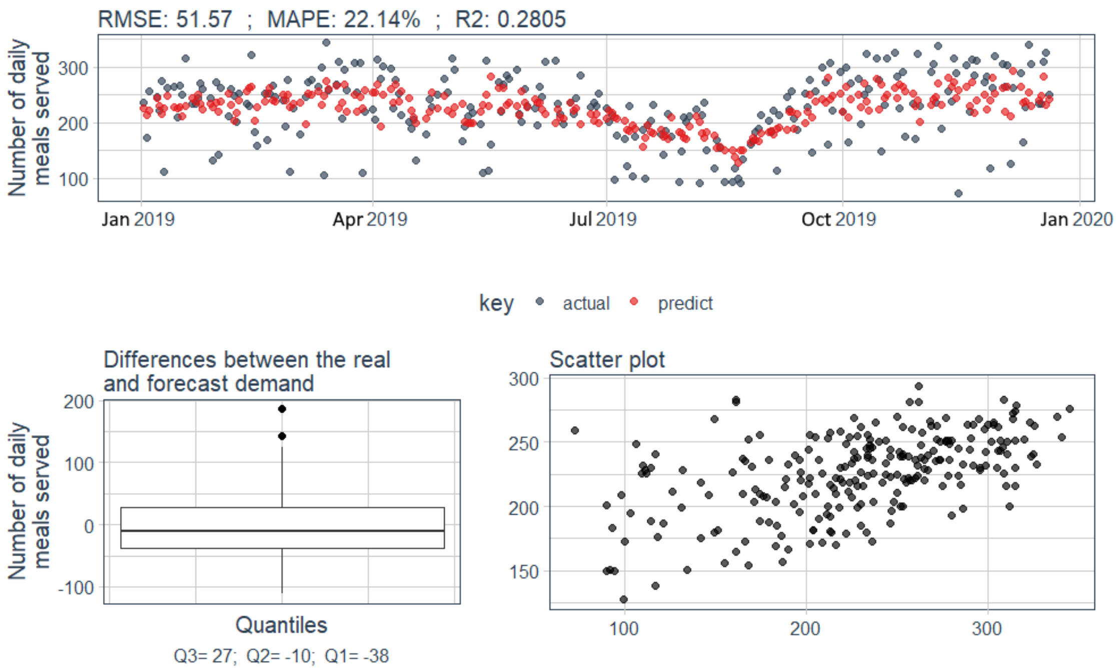

Figure 5.11: Visual representation of the real and forecast demand for FCS1

Although the other developed forecasting models performed worse than the model based on the MA approach, they were still better than the model based on the naïve approach when considering the quality of the predictions as a whole. However, the forecasts produced by the models based on the LightGBM algorithm and the transformer neural network would lead to fewer wasted meals when compared with the forecasts produced by the model based on the MA approach. When considering the model based on the naïve approach instead, every developed model would lead to fewer wasted meals. However, every developed model except the one based on the RF algorithm would lead to fewer meals being served (i.e., a higher unmet demand) when compared with the one based on the MA approach, and the inverse when compared with the one based on the naïve approach. In short, the best forecasting model is the one based on the RF algorithm, followed by the one based on the MA approach, the other developed forecasting models, and, lastly, the model based on the naïve approach.

However, if the goal was solely to minimize the number of wasted meals, the forecasts produced by the model based on the transformer neural network would be the most desirable. Otherwise, the forecasts produced by the model based on the RF algorithm are superior to every other prediction.

5.4.2 Food Catering Service 2 - Student canteen

For the second FCS, only the models based on the RF and LightGBM algorithms were able to beat the best baseline model, as highlighted by their lower RMSE, MAPE, and total absolute difference, and their higher R2. The performance of all forecasting models is presented in Table 5.7. Only the forecasts produced by the model based on the RF algorithm resulted in a lower amount of wasted meals and unmet demand when compared with the best baseline model. The forecasts of the model based on the RF algorithm would result in 3,268 meals being wasted in 2019, given the actual demand.

Table 5.7: Results of the forecasting models for FCS2

	Error measures			Forecast differences		
Model	RMSE	MAPE	R2	Wasted Meals	Unmet Demand	Total Absolute Difference
RF	47.01	0.2314	0.6806	3,268	4,391	7,659
LightGBM	51.06	0.2400	0.6344	3,200	5,047	8,257
LSTM	60.90	0.3061	0.5066	3,734	6,292	10,026
Transformer	76.53	0.3543	0.3614	2,521	10,487	13,008
Moving Average (5 days)	56.22	0.2765	0.5510	4,478	4,642	9,120
Naïve (1 day)	57.71	0.2533	0.5710	4,627	4,655	9,282

Note: Around 40,000 meat meals were served in 2019, which translates into roughly 150 daily meals

Although this is a considerable number of wasted meals, it is still 1,210 fewer wasted meals (or minus 27 % meals in relative terms) when compared with the baseline model based on the MA of the last five known days. Additionally, while the model based on the RF algorithm would lead the FCS to sell 4,391 fewer meals in 2019 (i.e., a higher unmet demand), this is still 251 more meals served when compared with the baseline model based on the MA approach. An overall view of the forecasts generated by the model based on the RF algorithm for FCS2 is provided in Figure 5.12.

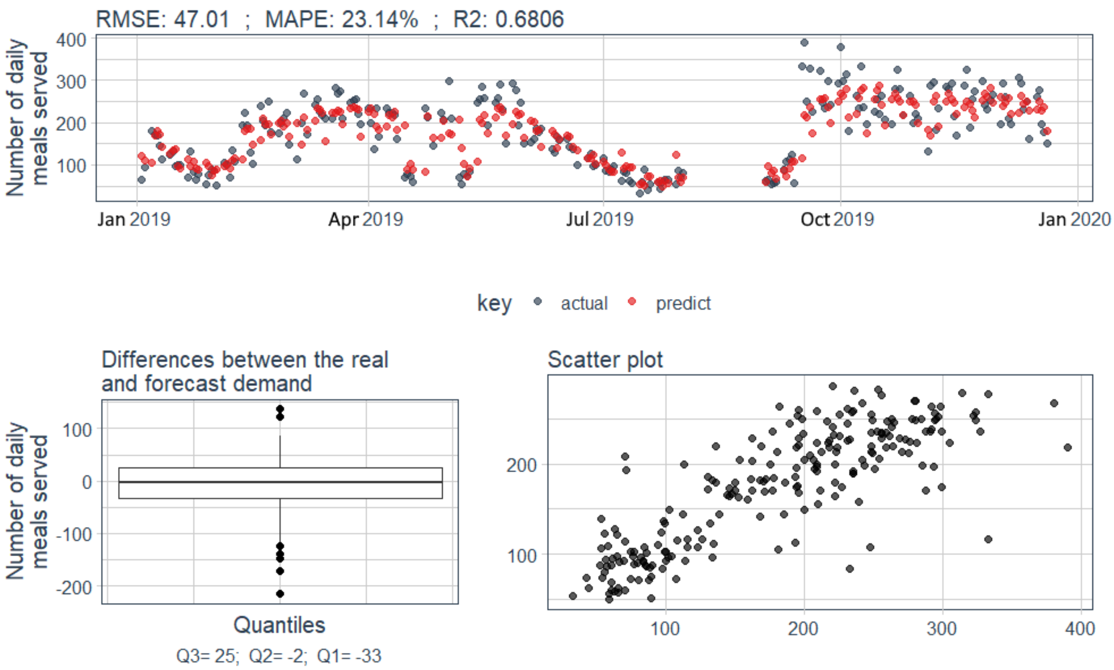

Figure 5.12: Visual representation of the real and forecast demand for FCS2

The forecasts produced model based on the LightGBM algorithm are also noteworthy. Although these would lead to a lower number of meals served (i.e., a higher unmet demand) when compared with the forecasts of the models based on the RF algorithm and the baseline models, it would reduce the amount of food wasted more significantly than the other two. While the model based on the RF algorithm and the baseline models would lead to 3,268 and 4,478 wasted meals, respectively, the model based on the LightGBM algorithm would lead to only 3,200 wasted meals.

The model based on the LSTM neural network would lead to more wasted meals than the models based on the RF and LightGBM algorithms. When compared with the baseline models, the model based on the LSTM neural network would lead to fewer wasted meals, but a considerably higher unmet demand. Ultimately, its error measures and total absolute difference metric are worse than those of both baseline models.

The model based on the transformer neural network produced the worst forecasts out of all the models in the context of FCS2. Although it would lead to the lowest number of wasted meals (2,521), this would be at the cost of a very high unmet demand.

In brief, the model based on the RF algorithm would grant FCS2 more reliable forecasts in comparison with the baseline models. The forecasts produced by the model based on the RF algorithm would allow FCS2 to generate considerably less FW while slightly increasing the number of meals served, i.e., the unmet demand would be lowered. However, if this FCS wanted to waste the lowest number of meals possible, regardless of the number of meals served, the model based on the transformer neural network would be a better choice than the one based on the RF algorithm.

5.4.3 Food Catering Service 3 - Company canteen

This FCS operates on a much larger scale than the other two other FCSs, having a considerably larger number of meals served in 2019. Still, even in this setting, the model based on the RF algorithm outperformed the best baseline model. Moreover, for RCS3, the model based on the LSTM neural network provided the most accurate forecasts, winning on every metric adopted, as seen in Figure 5.8.

Table 5.8: Results of the forecasting models for FCS3

	Error measures			Forecast differences		
Model	RMSE	MAPE	R2	Wasted meals	Unmet demand	Total absolute difference
RF	227.25	0.2062	0.5477	19,026	25,111	44,137
LightGBM	213.22	0.1941	0.5902	15,153	25,004	40,157
LSTM	173.36	0.1432	0.7325	15,151	17,356	32,507
Transformer	358.88	0.3813	0.0254	21,955	47,476	69,431
Moving Average (30 days)	323.83	0.3996	0.0381	31,782	30,059	61,841
Naïve (5 days)	432.13	0.4427	0.0259	39,685	43,767	83,452

Note: Around 270,000 meat meals were served in 2019, which translates into roughly 1,050 daily meals

By employing the model based on the LSTM neural network, it would be possible to waste 16,631 fewer meals in 2019 than when compared with the model based on the MA of the last 30 known days, which would lead to 31,782 wasted meals in the same period. This difference is considerable, translating into 52 % less FW in favor of the model based on the LSTM neural network. Not only that, but the forecasts provided by the model based on the LSTM neural network would also manage to minimize the amount of unmet demand, i.e., the number of meals not served due to underproduction. Although it would cause demand for 17,356 meals to go unmet in 2019, this model would still lead to 12,703 more meals being served than the baseline model based on the MA approach. A visual representation of the forecasts provided by the model based on the LSTM neural network and the real demand is drawn in Figure 5.13.

The models based on the RF and LightGBM algorithms also outperformed the baseline model based on the MA approach. These two forecasts would lead to fewer wasted meals and less unmet demand by a considerable margin, especially considering the number of wasted meals. In terms of unmet demand, these two models performed similarly, but the one based on the LightGBM algorithm would produce significantly fewer wasted meals, essentially matching the same number produced by the model based on the LSTM neural network.

The demand forecast produced by the model based on the transformer neural network would lead to the most wasted meals out of all the predictions, but fewer wasted meals than both baseline models. However, to achieve this, the model would lead to the fewest meals being served out of any model, causing the FCS to not meet the demand for 47,476 meals in 2019.

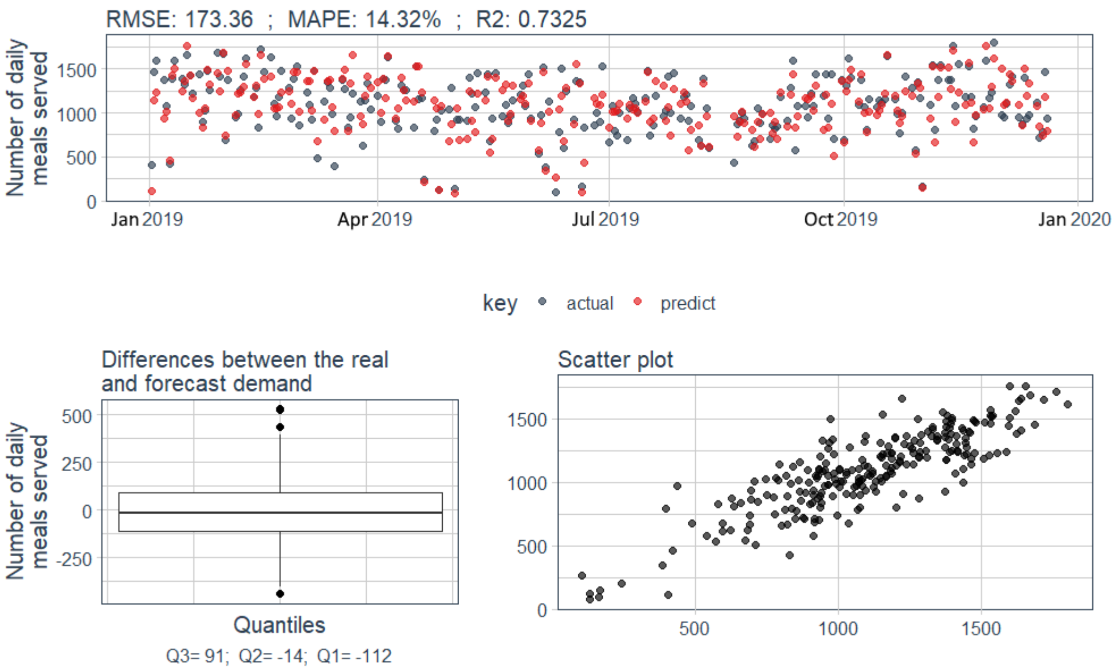

Figure 5.13: Visual representation of the real and forecast demand for FCS3

Contrary to the other two FCSs, for FCS3, the model based on the LSTM neural network produced the best forecasts. Not only that, but it would also minimize the number of wasted meals while providing better demand predictions to maximize the number of meals served. Thus, for FCS3, this model would be the preferred choice.

5.4.4 Overall results

As expected, each FCS had unique properties and, thereby, each developed model performed differently depending on the FCS. Nevertheless, it is clear that the proposed models would enable all the studied FCSs to reduce their FW.

The RMSE, being an absolute metric, was significantly higher for FCS3 than FCS1 and FCS2. On the other hand, relative measures like the MAPE and R2, which should not be impacted by the value range of the target variable, were also different for each FCS. At first glance, and given that a higher relative error measure would indicate a lower predictive power of the forecasting model, it is possible to verify that the developed models were the most accurate for FCS2, while the food demand in FCS1 was the hardest to forecast.

More specifically, the model based on the RF algorithm provided the best results for FCS1 and FCS2, i.e., the school canteens, especially in terms of minimizing the number of wasted meals in 2019. The predictions of the RF model were better than those of the baseline models in terms of both the number of meals wasted and the number of meals served. Although the model based on the transformer neural network provided forecasts that would minimize the number of meals wasted, this would happen due to an underestimation, and therefore at the cost of increasing the number of meals not provided. Thus, for FCS1 and FCS2, adopting the model based on the RF algorithm would be the best option. This is also reinforced by the RMSE, MAPE, and R2 scores attained.

For FCS3, which operates in a company setting, the best model was instead the one powered by the LSTM neural network, which performed significantly better than the rest. This model would be able to minimize the number of meals wasted while also minimizing the unmet demand. This model's superiority is also evident from its attained error measures, which were significantly better than the other models'. Nevertheless, the models based on the RF and LightGBM algorithms also produced excellent results, beating the baseline model based on the MA of the last 30 days by a significant amount. Like in the previous FCSs, the model based on the transformer neural network would lead to the least provisioned meals, but in this case, it would also cause the most wasted meals out of all the developed models. Additionally, the error measures of the model based on the transformer neural network were worse than the ones of the baseline model based on the MA approach, but better than the ones of the baseline model based on the naïve approach.

5.5 Discussion

Out of the four proposed food demand forecasting models, only two outperformed the others in a given setting. More specifically, the model based on the RF algorithm outperformed the rest in the school settings (FCS1 and FCS2), and the model based on the LSTM neural network outperformed the rest in the company setting (FCS3). While the setting may partially explain these results, it is also important to analyze the time horizon and the levels of demand of each FCS. The model based on the LSTM neural network excelled in a context with an extensive time horizon (nine years) and high demand (around 1,900 daily meals), while the model based on the RF algorithm stood out in two contexts with short time horizons (between three and four years) and a more modest number of daily meals served (between 200 and 400). Nonetheless, more experiments should be conducted before claiming any such correlations. Some other noteworthy performances are those of the model based on the LightGBM algorithm in the context of FCS2 and FCS3 and the model based on the transformer neural network in the context of FCS1.

Out of the two baseline models, the one based on the MA approach provided the best results in all studied FCSs. Furthermore, this model always managed to outperform one or more of the proposed forecasting models. The model based on the naïve approach produced the worst forecasts in the context of FCS1 and FCS3 but managed to outperform the model based on the LSTM and transformer neural networks in the context of FCS2.

The predictions produced by the developed models were not perfect. However, they were still significantly better than the best forecasts generated by the baseline models. For FCS1, which served 56,000 meat meals in 2019, the adoption of the model based on the RF algorithm would lead to 4,671 meals being wasted. Furthermore, FCS1 would not be able to meet the demand for 5,276 meals due to underproduction on some days. Although these numbers amount to a significant portion of the total number of meals served, they are still lower than what possibly happened (as measured by the baseline model). More specifically, in the test performed for 2019, adopting the predictions of the model based on the RF algorithm would lead to a 14 % reduction in the number of wasted meals, or 782 meals, while also serving 195 more meals.

The tests performed for the other FCSs had similar outcomes. In the case of FCS2, which served 56,000 meat meals in 2019, following the forecasts of the model based on the RF algorithm would lead to a 27 % decrease in FW (1,210 meals) over the moving average baseline model, while also being able to serve 251 more meals. As for FCS3, the differences were significantly larger, possibly due to the size of the canteen and the larger time horizon of the data used to train the models. Namely, in this setting, it would be possible to reduce FW by more than half, or 31,782 meals, whilst serving 12,703 more meals in the same period.

However, simply stating that applying these forecasting models throughout the test period (2019) would prevent 18.623 fewer meals from being wasted is disingenuous. Doing so is assuming that the FCSs 1) do their forecasts exactly as the moving average baseline model; 2) follow the forecasts blindly; and 3) have no measures in place to allow for some variance between the forecast and actual demand, e.g., produce more meals if the demand is above the one that was forecast, and recall meals produced or halt production if the demand is below the one that was forecast. Considering that FCSs can compensate for deviations between the forecast and actual demand, choosing the model with the lowest RMSE should be beneficial. The more consistently close the forecast values are to the actual demand, the easier it is for FCSs to perform minor adjustments on the day, depending on how it unfolds.

The present chapter has some further limitations. First, the three studied settings may not be representative of how most school or company canteens operate. Thus, the results obtained should not be generalized. Furthermore, since the data collected corresponds to the number of dishes served and not the quantity served, the demand estimates provided by the developed models may not be the most appropriate in terms of suggesting the amount of food to prepare. This is due to the fact that the amount served on each plate may vary slightly. Additionally, it is important to emphasize that although this chapter focuses on the waste generated by FCSs, the amount of waste generated by customers who do not finish their meals (i.e., plate waste) is very significant.

A further limitation is that the evaluation of the proposed models was supported by baseline models that were created to mimic the forecasts of the studied FCSs. It was necessary to design these baseline models because the data collected did not include the estimates made by these services, possibly because they were never registered. Nevertheless, the adopted baseline models may not accurately reflect the thought process of the managers of these FCSs and other approaches may be preferred. It is also important to highlight that the results obtained might have been affected by some data quality issues, namely the accuracy and accessibility of the data. The proposed models were limited by a lack of relevant information that could have been used to infer the potential demand for a particular dish, such as daily menu information (with the exception of FCS2) and the preferences of individual customers. However, even when daily menu information was recorded (FCS2), the relatively few days during which a specific menu was served during the analysis period implied a broad classification of the menus.

5.6 Conclusion and Future Research

Mass amounts of FW are produced annually, leading to immeasurable environmental, social, and economic consequences. One of the main reasons for FW generation is overproduction, which comes mainly from poor demand forecasts. Thus, the present work proposes and compares food demand forecasting models based on different algorithms to support the decision-makers at FCSs. By having more accurate and reliable forecasts, these decision-makers can make sure that their services are better prepared to meet customer demand.

To test the performance of the proposed models, data was collected from three different FCSs: two operating school canteens and one operating a company canteen. Among the tested models, the ones based on the RF algorithm and LSTM neural network provided the best results. Comparing the forecasts provided by these models with the forecasts of the baseline models, the former managed to reduce the amount of generated FW by between 14 % and 52 %.

The forecasts provided by the model based on the RF algorithm were more reliable in the context of the two school canteens, while the model based on the LSTM neural network proved to be better in the context of the company canteen. The model based on the RF algorithm produced better results in settings with three to four years of historic data and an average of 200 to 400 meals served daily, while the model based on the LSTM neural network outperformed the rest when the data available had a time horizon of more than nine years and around 1,900 meals served daily.

Overall, the LSTM models produce the greatest theoretical amount of FW reduction, both in absolute and relative terms. This fruitful performance comes mainly from the results of FCS3, where the results may be a consequence of the longer time horizon of the data, which allowed the prediction models to better understand and adapt to the setting.

The forecasts produced with the support of machine learning techniques proved to be more accurate and reliable than those that FCSs usually perform on their own. Even the best baseline model was always beaten by the models based on the RF algorithm or LSTM neural network, by a significant margin. That said, FCSs that do not rely on machine learning forecasts are better off relying on an MA approach, i.e., the average of the last known days, over a naïve approach, i.e., the same demand of a previous day.

Even though the machine learning forecasts beat the baseline models, enabling a theoretical reduction of FW, it is important to understand that in the setting at hand, there is no one-size-fits-all algorithm, with different algorithms capitalizing on different features and being more adequate to different settings. Additionally, this chapter only focused on forecasting demand for a single dish. Although this dish was the most popular in the case of all three studied FCSs, full menu predictions could also be beneficial for planning purposes. Moreover, it would be useful for future forecast works to explore different research avenues, such as understanding individual preferences (in cases where reservations are made) and/or suggesting future menus in order to motivate demand and minimize FW. Furthermore, it would be relevant to confirm whether larger time horizons of data in the field of FCSs are indexed to better forecasting results. Lastly, it would be very interesting to design a system that can be trained on data from different settings simultaneously in order to leverage pattern recognition.

CHAPTER 6
Conclusions

This chapter presents the main conclusions of this thesis. The reader will be able to understand how the research objectives referred to in Chapter 1 have been addressed and what the main contributions of this thesis are to the scientific community. The limitations of this work and how they may have affected the developed work are discussed later. This discussion bridges over to the suggestion of future research.

6.1 Fulfilment of the research objectives

The focus of this thesis was threefold: to study MSW, to determine what measures can be taken to minimize it, and to develop a predictive model for this effect. Each of these three objectives was addressed in a dedicated chapter, presenting unique methodologies tailored to achieve its goal.

First, a detailed description of the main MSW issues discussed in the scientific community and in EU-funded projects was provided, and the similarity between the content generated in the scope of the two spheres was estimated.

Second, an analysis of the quantitative methods applied to FW, the largest part of MSW, was carried out. For this purpose, three types of analytics were explored, namely the descriptive, predictive, and prescriptive approaches toward FW. The main characteristics of each type of analytics were presented, including their methodology and results. This involved introducing the amounts of FW produced in different environments, the forecasting models used, and the optimization systems developed in the field.

Finally, a series of demand forecasting models were designed to support existing advances in the field of FW minimization. These aimed to test the performance of different machine learning algorithms in different settings, and to evaluate their added value compared to the forecasts used by the food services themselves.

An article was written in the scope of each goal, which in turn are reflected in the chapters 3, 4, and 5. The methodologies used for each are also different given their different underlying designs. Their disclosure, as well as the data used and the results produced, can all be found in their respective chapters. Nevertheless, the following subsections will discuss the findings of these three works, addressing their results and how they respond to the defined research objectives.

6.1.1 Municipal Solid Waste domains

Chapter 3 explores the broad theme of MSW from the perspective of the scientific community and the EU policymakers through EU-funded projects.

By analyzing the articles published by the scientific community, the main topic groups are *MSW properties*, *Pollution and Health Risks*, *MSW and biomass conversion*, *Biological treatment processes*, *Leaching and extraction of heavy metals*, *Optimization and site selection of MSW systems*, *Environmental Impact Assessment and Sustainability*, and *Soil and Plant Effects of Waste Treatment*.

The analysis also revealed that the scientific side of MSW is very mature, with works covering the subject from different perspectives linked to other themes. This includes analysing and understanding the composition of MSW, repurposing waste for value and energy recovery, determining optimal waste management approaches, and the evaluation of their impacts in economic, environmental, and social dimensions. Although MSW characteristics vary from area to area, the literature still deeply understands its nuances. In addition, there is also a significant concern with practical applications in trying to reduce, reuse, recycle, recover, and dispose of the existing waste. Over the years, all the topic groups have shown similar importance, which means that there is a constant complementation of each domain.

On the other hand, the main topic groups in the projects funded by the EU are *Energy conversion*, *Value extraction and waste valorization*, *Biological treatment processes*, *Waste management processes*, *Projects and programs*, *MSW characterization and policies*, *Sustainability and pollution*, and *Practical experimentation test systems*.

The projects and programs are aimed at motivating countries to do better and to address the consequences of MSW, which means that over the years more emphasis has been placed on the present and near future problems associated with MSW. Consequently, the EU has placed significant emphasis on exploring options to mitigate the consequences of MSW. This highlights the need to address challenges in formulating waste management plans and other essential municipal programs, which are essential for promoting the development of effective strategies to mitigate the impact of MSW.

When comparing the description of the EU-funded projects and the articles of the scientific community, the former is more closely related to the latter than the other way around. Differences in similarity between the two may be due to the range of topics they cover. Nevertheless, it is also true that the articles are more similar to each other than to the EU-funded projects. This may be due to the significant difference in their sample sizes and how EU-funded projects can extract information from the scientific community or other organizations such as the FAO and the World Bank.

6.1.2 Quantitative approaches toward Food Waste

Chapter 4 analyzes the quantitative approaches applied in the field of FW in the literature by examining 65 unique studies. Key features of all three types of analysis (descriptive, predictive, and prescriptive) are evaluated.

Studies performing descriptive analytics have shown varying approaches to collecting and presenting FW data, ranging from direct collection to retrieval from third-party sources. There is a lack of consensus on metrics and measurement methods, making it difficult to compare data across studies. Reported FW volumes vary widely due to different time frames and contexts, ranging from a few hundred to millions of kilograms.

FW demand forecasting has gained considerable popularity, especially after 2012, with an exponential increase in related scientific papers. Different sectors use different algorithms. Retail tends to use ARIMA and neural networks, while hospitality favors neural networks, ARIMA, TBATS, random forests, and naive models. Studies typically consider daily periodicity and employ causal and time series models, using measures such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) to assess accuracy. Predictive approaches have been applied to shorter (daily) and longer (annual) time periods.

The field of prescriptive analytics has evolved significantly since 2018, focusing on optimizing scenarios to reduce FW generation, extend shelf life, or mitigate FW consequences. It covers retail, food service, and city settings and employs various programming models, including single-objective linear programming and multi-objective nonlinear programming. Decision variables are highly contextual but often revolve around inventory management. These studies lack baseline comparisons.

6.1.3 Short-term demand forecasting in food catering services

Chapter 5 focuses on the development of demand forecasting models to provide managers with accurate forecasts to aid in decision-making. Four models are developed based on RF, LightGBM, LSTM neural networks, and transformer neural network algorithms.

These four models were tested in three different settings: two school canteens and one company canteen. Among the four developed food demand forecasting models, the model based on the RF algorithm outperformed the others in the school setting. In contrast, the model based on the LSTM neural network outperformed the others in the company setting. This difference may be due to the inherent characteristics of the settings themselves, as well as the time horizon and level of demand for each food service.

Two baseline models were also created to mimic the predictions of the food services in question: one based on a naive approach and the other based on an MA approach. The one based on the MA approach provided the best results, outperforming one or more of the developed forecasting models.

However, the developed forecasting models were significantly better than the best forecasts generated by the baseline models. Considering each food service separately, the developed models would be able to reduce the number of wasted meals in 2019 by 14 %, 27 %, and 52 %, respectively, while also being able to serve more meals compared to the MA baseline model.

6.2 Research limitation

The research presented throughout this thesis has been able to achieve all the research goals it defined for itself. However, the thesis has some limitations that must be acknowledged in any case to provide context to the results presented. Although these have been mentioned in their respective chapter, a synopsis of the limitations will also be provided in the following paragraphs.

An overall limitation of this work is related to the availability and quality of the data used in all three works. The data used in Chapters 3 and 4 has been queried through databases using keywords. Although in both cases the keywords adopted had the purpose of retrieving the relevant works, it is possible that studies were captured but were not relevant to the theme in question, or relevant studies were not captured with the keywords used. Although the first case is not as relevant for the study in Chapter 4 since the studies were individually considered, this may be problematic for the study in Chapter 3 since there was no manual pruning done.

Further, the databases used - the Elsevier SCOPUS and CORDIS databases for the study in Chapter 3 and the Elsevier SCOPUS and Web of Science databases for the study in Chapter 4 - may not be representative of the scientific community and EU policymakers.

On the other hand, the data used in Chapter 5 may not depict how most schools or company canteens operate, difficulting the generalization of the results. Furthermore, since the collected data corresponds to the number of dishes served and not the quantity served, the demand estimates provided by the developed models may not be the most appropriate for suggesting the amount of food to prepare.

Another overarching limitation may be related to the methodologies adopted. In Chapter 3, LDA models are employed. However, the corpus size of the EU-funded projects description is not big, and this application may not be as fruitful. Regarding Chapter 4, although the PRISMA statement methodology was applied and is a prevalent and renowned methodology, the screening process can be somewhat subjective. Furthermore, the exclusion criteria used for the screening process affect the choice of studies to include and analyze in the review. Although necessary, it introduces additional subjectivity into the process that can inadvertently omit relevant studies or introduce bias into the review. Lastly, for Chapter 5, the baseline models adopted may not accurately reflect the thought process of managers. In addition, FCSs do not follow forecasts blindly. They can perform minor adjustments on the day, adapting their demand depending on how it unfolds.

Regardless, the above limitations should not hinder the conclusions taken from the mentioned works. In fact, they can be seen as an opportunity for future research, complementing the literature with more robust studies.

6.3 Future Research

This chapter has presented the main findings of the work carried out in this thesis and also discussed its limitations. Nevertheless, the limitations mentioned can also be seen as challenges to be overcome and motivate future research to complement the literature.

For example, when researching the topics covered by the EU policymakers in Chapter 3, the

EU-funded projects were used to represent them. However, this resulted in a corpus of 140 documents against over 25,000 articles representing the scientific community. Thus, there is an opportunity to query different sources that also represent EU policymakers in order to have a larger corpus and more robust conclusions. In addition, the application of other methods beyond LDA and cosine similarity may also provide more objective results.

On another note, the literature review of quantitative methods applied in the field of FW covered in Chapter 4 is heavily based on defined keywords to extract relevant articles. In this process, other sets of keywords can be further explored to represent better the area being studied, thus yielding more appropriate studies. To complement this, the exclusion criteria can also be further tinkered with to avoid bias and better focus on the relevant studies in the field of quantitative analysis of FW.

In addition, the work presented in Chapter 4 provides a basis for further development of quantitative FW analysis, offering opportunities for future research in each analytical approach. For example, a standardized measurement framework could improve study comparisons in the descriptive domain. Predictive analysis could benefit from context-specific variables in explanatory models and the exploration of advanced machine learning techniques, such as transformers for direct FW prediction. For prescriptive analytics, for example, an optimization system could be developed to design menus that minimize FW while maximizing meals served and resource efficiency, offering potential avenues for further exploration and development.

Finally, when conducting future forecasting similar to the work in Chapter 5, it could be fruitful to explore individual preferences and suggest future menus to motivate demand and minimize FW. Furthermore, it would be relevant to confirm whether longer time horizons of data in FCS are associated with better forecasting results. Finally, designing a system that can be trained simultaneously on data from different environments to exploit pattern recognition would be very interesting.

References

Aamer, A. M., Yani, L. P. E., and Priyatna, I. M. A. (2021). Data analytics in the supply chain management: Review of machine learning applications in demand forecasting. *Operations and Supply Chain Management*.

Abdel-Shafy, H. (2015). Chemical treatment for removal of heavy metals from industrial wastewater. *Egyptian Journal of Chemistry*.

Abdel-Shafy, H. I. and Mansour, M. S. (2018a). Solid waste issue: Sources, composition, disposal, recycling, and valorization. *Egyptian Journal of Petroleum*.

Abdel-Shafy, H. I. and Mansour, M. S. (2018b). Solid waste issue: Sources, composition, disposal, recycling, and valorization. *Egyptian journal of petroleum*, 27(4):1275–1290.

Abdel-Shafy, H. I. and Mansour, M. S. M. (2017). Treatment of pharmaceutical industrial wastewater via anaerobic /aerobic system for unrestricted reuse. *Journal of Scientific and Industrial Research*.

Adelodun, B. and Choi, K. (2020). Impact of food wastage on water resources and ghg emissions in korea: A trend-based prediction modeling study. *Journal of Cleaner Production*, 271.

Adhikari, B., Barrington, S., and Martinez, J. (2006). Predicted growth of world urban food waste and methane production. *Waste Management and Research*, 24(5):421–433.

Adhikari, S., Nam, H., and Chakraborty, J. P. (2018). *Chapter 8 - Conversion of Solid Wastes to Fuels and Chemicals Through Pyrolysis*. Thallada Bhaskar and Ashok Pandey and S. Venkata Mohan and Duu-Jong Lee and Samir Kumar Khanal.

Afzal, N., Basit, A., Daniel, A., Ilyas, N., Imran, A., Awan, Z., Papargyropoulou, E., Stringer, L., Hashem, M., Alamri, S., Bashir, M., Li, Y., and Roy, N. (2022). Quantifying food waste in the hospitality sector and exploring its underlying reasons—a case study of lahore, pakistan. *Sustainability (Switzerland)*.

Al-Maliky, S. and ElKhayat, Z. (2012). Kitchen food waste inventory for residential areas in baghdad city. *Modern Applied Science*, 6(8):45–51.

Al-Salem, S., Antelava, A., Constantinou, A., Manos, G., and Dutta, A. (2017). A review on thermal and catalytic pyrolysis of plastic solid waste (psw). *Journal of environmental management*, 197:177–198.

Al-Salem, S., Lettieri, P., and Baeyens, J. (2009). Recycling and recovery routes of plastic solid waste (psw): A review. *Waste Management*, 29(10):2625–2643.

Albrecht, A., Mittler, M., Hebel, M., Waldhans, C., Herbert, U., and Kreyenschmidt, J. (2021). A multi-model approach to implement a dynamic shelf life criterion in meat supply chains. *Foods*.

Aloysius, N., Ananda, J., Mitsis, A., and Pearson, D. (2023). Why people are bad at leftover food management? a systematic literature review and a framework to analyze household leftover food waste generation behavior. *Appetite*.

Alshabanat, Z., Alkhorayef, A., Ben Haddad, H., Mezghani, I., Gouider, A., Tlili, A., Allouche, M., and Gannouni, K. (2021). Quantifying food loss and waste in saudi arabia. *Sustainability (Switzerland)*, 13(16).

Annese, V., Biccario, G., Cipriani, S., and de Venuto, D. (2014). Organoleptic properties remote sensing and life-time prediction along the perishables goods supply-chain. *International Journal on Smart Sensing and Intelligent Systems*, 7(5).

Arun, R., Suresh, V., Veni Madhavan, C., and Narasimha Murthy, M. (2010). Finding the number of topics in latent dirichlet allocation (lda) with a stability-based method. In *Proceedings of the Eighth International Conference on Advances in Information Mining and Management*, pages 1–7.

Arunraj, N. and Ahrens, D. (2015). A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting. *International Journal of Production Economics*, 170:321–335.

Ayeleru, O., Okonta, F., and Ntuli, F. (2018). Municipal solid waste generation and characterization in the city of johannesburg: A pathway for the implementation of zero waste. *Waste Management*, 79:87–97.

Barnes, A., Soto, I., Eory, V., Beck, B., Balafoutis, A., Sánchez, B., Vangeyte, J., Fountas, S., van der Wal, T., and Gómez-Barbero, M. (2019). Exploring the adoption of precision agricultural technologies: A cross regional study of eu farmers. *Land Use Policy*.

Batista, L., Dora, M., Garza-Reyes, J. A., and Kumar, V. (2021). Improving the sustainability of food supply chains through circular economy practices – a qualitative mapping approach. *Management of Environmental Quality: An International Journal*, 32(4):752–767.

Beheshti, S. and Heydari, J. (2023). Municipal food waste recycling as an environmental strategy: a game-theoretical approach. *Environmental Science and Pollution Research*, 30(13):36744–36768.

Beheshti, S., Heydari, J., and Sazvar, Z. (2022). Food waste recycling closed loop supply chain optimization through renting waste recycling facilities. *Sustainable Cities and Society*, 78:103644.

Bickel, M. W. (2019). Reflecting trends in the academic landscape of sustainable energy using probabilistic topic modeling. *Energy, Sustainability and Society*.

Birisci, E. and McGarvey, R. (2018). Optimal production planning utilizing leftovers for an all-you-care-to-eat food service operation. *Journal of Cleaner Production*, 171:984–994.

Birisci, E. and McGarvey, R. G. (2022). Cost-versus environmentally-optimal production in institutional food service operations. *Socio-Economic Planning Sciences*.

Blair, S. J., Bi, Y., and Mulvenna, M. D. (2020). Aggregated topic models for increasing social media topic coherence. *Applied Intelligence*.

Blei, D. M. and McAuliffe, J. D. (2008). Supervised topic models.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*.

Blundell, T. (2004). Chemicals in products-safeguarding the environment and human health. *Science in Parliament*, 61(3):5.

Brennan, L., Langley, S., Verghese, K., Lockrey, S., Ryder, M., Francis, C., Phan-Le, N. T., and Hill, A. (2021). The role of packaging in fighting food waste: A systematised review of consumer perceptions of packaging. *Journal of Cleaner Production*, 281:125276.

Buenano-Fernandez, D., Gonzalez, M., Gil, D., and Lujan-Mora, S. (2020). Text mining of open-ended questions in self-assessment of university teachers: An lda topic modeling approach. *IEEE Access*.

Buisman, M., Haijema, R., and Bloemhof-Ruwaard, J. (2019). Discounting and dynamic shelf life to reduce fresh food waste at retailers. *International Journal of Production Economics*, 209:274–284.

Campuzano, R. and González-Martínez, S. (2016). Characteristics of the organic fraction of municipal solid waste and methane production: A review. *Waste Management*, 54:3–12.

Canali, M., Amani, P., Aramyan, L., Gheoldus, M., Moates, G., Östergren, K., Silvennoinen, K., Waldron, K., and Vittuari, M. (2017). Food waste drivers in europe, from identification to possible interventions. *Sustainability (Switzerland)*, 9(1).

Cao, J., Xia, T., Li, J., and Zhang, Y. (2009). Density-based clustering based on hierarchical density estimates. *Proceedings of the 2009 SIAM International Conference on Data Mining*, pages 246–257.

Carbonneau, R. A., Laframboise, K., and Vahidov, R. M. (2008). Application of machine learning techniques for supply chain demand forecasting. *Eur. J. Oper. Res.*, 184:1140–1154.

Castles, S. (2004). The factors that make and unmake migration policies. *International Migration Review*.

Chalak, A., Abou-Daher, C., and Abiad, M. (2018). Generation of food waste in the hospitality and food retail and wholesale sectors: lessons from developed economies. *Food Security*, 10(5):1279–1290.

Chand Malav, L., Yadav, K. K., Gupta, N., Kumar, S., Sharma, G. K., Krishnan, S., Rezania, S., Kamyab, H., Pham, Q. B., Yadav, S., Bhattacharyya, S., Yadav, V. K., and Bach, Q.-V. (2020). A review on municipal solid waste as a renewable source for waste-to-energy project in india: Current practices, challenges, and future opportunities. *Journal of Cleaner Production*.

Chapman, L., Cohen, J., Canterberry, M., and Carton, T. (2017). Factors associated with school lunch consumption: Reverse recess and school "brunch". *Journal of the Academy of Nutrition and Dietetics*, 117(9):1413–1418.

Chen, Y., Ghosh, J., Bejan, C. A., Gunter, C. A., Gupta, S., Kho, A., Liebovitz, D., Sun, J., Denny, J., and Malin, B. (2015). Building bridges across electronic health record systems through inferred phenotypic topics. *Journal of Biomedical Informatics*.

Cheng, H. and Hu, Y. (2010). Municipal solid waste (msw) as a renewable source of energy: Current and future practices in china. *Bioresource Technology*.

Cherubini, F., Bargigli, S., and Ulgiati, S. (2009). Life cycle assessment (lca) of waste management strategies: Landfilling, sorting plant and incineration. *Energy*.

Chew, K., Chia, S., Yen, H.-W., Nomanbhay, S., Ho, Y.-C., and Show, P. (2019). Transformation of biomass waste into sustainable organic fertilizers. *Sustainability (Switzerland)*, 11(8).

Chia, D., Yap, C. C., Wu, S. L., Berezina, E., Aroua, M. K., and Gew, L. T. (2023). A systematic review of country-specific drivers and barriers to household food waste reduction and prevention. *Waste Management and Research*.

Ching, P., Mutuc, J., and Jose, J. (2019). Assessment of the quality and sustainability implications of fifo and lifo inventory policies through system dynamics. *Advances in Science, Technology and Engineering Systems*, 4(5):69–81.

Chu, C.-Y., Park, K., and Kremer, G. E. (2020). A global supply chain risk management framework: An application of text-mining to identify region-specific supply chain risks. *Advanced Engineering Informatics*.

Chu, Z., Wang, W., Wang, B., and Zhuang, J. (2016). Research on factors influencing municipal household solid waste separate collection: Bayesian belief networks. *Sustainability*, 8(2).

Clark, P. U., Shakun, J. D., Marcott, S. A., Mix, A. C., Eby, M., Kulp, S., Levermann, A., Milne, G. A., Pfister, P. L., Santer, B. D., Schrag, D. P., Solomon, S., Stocker, T. F., Strauss, B. H., Weaver, A. J., Winkelmann, R., Archer, D., Bard, E., Goldner, A., Lambeck, K., Pierrehumbert, R. T., and Plattner, G.-K. (2016). Consequences of twenty-first-century policy for multi-millennial climate and sea-level change. *Nature Climate Change*.

Conrad, Z., Niles, M., Neher, D., Roy, E., Tichenor, N., and Jahns, L. (2018). Relationship between food waste, diet quality, and environmental sustainability. *PLoS ONE*, 13(4).

Cram, L. (2020). Policy-making in the european union.

Cristóbal, J., Castellani, V., Manfredi, S., and Sala, S. (2018). Prioritizing and optimizing sustainable measures for food waste prevention and management. *Waste Management*.

Crossing, V. (2003). Weather condition databse. Online.

Curry, N. and Pillay, P. (2012). Biogas prediction and design of a food waste to energy system for the urban environment. *Renewable Energy*, 41:200–209.

Dai, X., Duan, N., Dong, B., and Dai, L. (2013). High-solids anaerobic co-digestion of sewage sludge and food waste in comparison with mono digestions: Stability and performance. *Waste Management*, 33(2):308–316.

D'Amato, D., Droste, N., Allen, B., Kettunen, M., Lähtinen, K., Korhonen, J., Leskinen, P., Matthies, B., and Toppinen, A. (2017). Green, circular, bio economy: A comparative analysis of sustainability avenues. *Journal of Cleaner Production*.

De Clercq, D., Jalota, D., Shang, R., Ni, K., Zhang, Z., Khan, A., Wen, Z., Caicedo, L., and Yuan, K. (2019). Machine learning powered software for accurate prediction of biogas production: A case study on industrial-scale chinese production data. *Journal of Cleaner Production*, 218:390–399.

Debnath, R. and Bardhan, R. (2020). India nudges to contain covid-19 pandemic: A reactive public policy analysis using machine-learning based topic modelling. *PLoS ONE*.

Deveaud, R., SanJuan, E., and Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 1009–1018.

Dey, S., Saha, S., Singh, A. K., and McDonald-Maier, K. (2022). Smartnoshwaste: Using blockchain, machine learning, cloud computing and qr code to reduce food waste in decentralized web 3.0 enabled smart cities. *Smart Cities*.

Dharmawardane, C., Sillanpää, V., and Holmström, J. (2021). High-frequency forecasting for grocery point-of-sales: intervention in practice and theoretical implications for operational design. *Operations Management Research*, 14(1-2):38–60.

Dhir, A., Talwar, S., Kaur, P., and Malibari, A. (2020). Food waste in hospitality and food services: A systematic literature review and framework development approach. *Journal of Cleaner Production*, 270:122861.

Diana, R., Martianto, D., Baliwati, Y. F., Sukandar, D., and Hendriadi, A. (2023). Determinants of household food waste in southeast asia: A systematic review. *Journal of Hunger and Environmental Nutrition*.

Diaz-Barriga-Fernandez, A. D., Santibañez-Aguilar, J. E., Betzabe González-Campos, J., Nápoles-Rivera, F., Ponce-Ortega, J. M., and El-Halwagi, M. M. (2018). Strategic planning for managing municipal solid wastes with consideration of multiple stakeholders. *Computer Aided Chemical Engineering*, 44:1597–1602.

Do, Q., Ramudhin, A., Colicchia, C., Creazza, A., and Li, D. (2021). A systematic review of research on food loss and waste prevention and management for the circular economy. *International Journal of Production Economics*, 239.

dos Santos, J. I. A. S., da Silveira, D. S., da Costa, M. F., and Duarte, R. B. (2022). Consumer behaviour in relation to food waste: a systematic literature review. *British Food Journal*.

DRAAF (2014). Proposer une boîte à emporter. Note réglementaire.

Economic & Social Commission for Asia & the Pacific (2000).

Edjabou, M. E., Petersen, C., Scheutz, C., and Astrup, T. F. (2016). Food waste from danish households: Generation and composition. *Waste management*, 52:256–268.

Endres, J., Barter, S., Theodora, P., and Welch, P. (2003). Soy-enhanced lunch acceptance by preschoolers. *Journal of the American Dietetic Association*, 103(3):346–351.

Environment Agency (2014). Technical guidance wm3: Waste classification - guidance on the classification and assessment of waste.

European Commission (2021). Waste framework directive. `https://environment.ec.europa.eu/topics/waste-and-recycling/waste-framework-directive_en`. Accessed: 14-Dec-2023.

European Environment Agency (2004). European waste catalogue.

European Union (2008). Waste framework directive 2008/98/ec, art. 3(1).

Fadhullah, W., Imran, N. I. N., Ismail, S. N. S., Jaafar, M. H., and Abdullah, H. (2022). Household solid waste management practices and perceptions among residents in the east coast of malaysia. *BMC public health*, 22(1):1–20.

Faezirad, M., Pooya, A., Naji-Azimi, Z., and Amir Haeri, M. (2021). Preventing food waste in subsidy-based university dining systems: An artificial neural network-aided model under uncertainty. *Waste Management and Research*, 39(8):1027–1038.

Fanelli, R. (2019). Using causal maps to analyse the major root causes of household food waste: Results of a survey among people from central and southern italy. *Sustainability (Switzerland)*, 11(4).

FAO (2011). *Global food losses and food waste - Extent, causes and prevention*. Rome.

Felix, P. (2018). Data-driven retail food waste reduction: a comparison of demand forecasting techniques and dynamic pricing strategies. Master's thesis, University of Twente.

Ferreira, M., Martins, M., and Rocha, A. (2013). Food waste as an index of foodservice quality. *British Food Journal*, 115(11):1628–1637.

Feukam Nzudie, H., Zhao, X., Tillotson, M., Zhang, F., and Li, Y. (2020). Modelling and forecasting roots & tubers losses and resulting water losses in sub-saharan africa considering climate variables. *Physics and Chemistry of the Earth*, 120.

Filimonau, V. and De Coteau, D. A. (2019). Food waste management in hospitality operations: A critical review. *Tourism Management*, 71(October 2018):234–245.

Friant, M. C., Vermeulen, W. J., and Salomone, R. (2021). Analysing european union circular economy policies: Words versus actions. *Sustainable Production and Consumption*, 27:337–353.

Gao, S., Bao, J., Li, R., Liu, X., and Wu, C. (2021). Drivers and reduction solutions of food waste in the Chinese food service business. *Sustainable Production and Consumption*, 26:78–88.

García, M., Vilas, C., Herrera, J., Bernárdez, M., Balsa-Canto, E., and Alonso, A. (2015). Quality and shelf-life prediction for retail fresh hake (merluccius merluccius). *International Journal of Food Microbiology*, 208:65–74.

Garre, A., Ruiz, M., and Hontoria, E. (2020). Application of machine learning to support production planning of a food industry in the context of waste generation under uncertainty. *Operations Research Perspectives*, 7.

Giordano, C., Falasconi, L., Cicatiello, C., and Pancino, B. (2020). The role of food waste hierarchy in addressing policy and research: A comparative analysis. *Journal of Cleaner Production*, 252:119617.

Girish, Amin, A. A., Huns, A. S., Hafeez, H. M., and SCPrabhu (2019). A review on solid waste management. *International Journal of Advance Research and Innovative Ideas in Education*, 5:2181–2187.

Girotto, F., Alibardi, L., and Cossu, R. (2015). Food waste generation and industrial uses: A review. *Waste Management*, 45:32–41.

Grainger, M., Aramyan, L., Piras, S., Quested, T., Righi, S., Setti, M., Vittuari, M., and Stewart, G. (2018). Model selection and averaging in the assessment of the drivers of household food waste to reduce the probability of false positives. *PLoS ONE*, 13(2).

Griffiths, T. L. and Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl 1):5228–5235.

Gružauskas, V., Gimžauskienė, E., and Navickas, V. (2019). Forecasting accuracy influence on logistics clusters activities: The case of the food industry. *Journal of Cleaner Production*, 240.

Gu, B., Jiang, S., Wang, H., Wang, Z., Jia, R., Yang, J., He, S., and Cheng, R. (2017). Characterization, quantification and management of china's municipal solid waste in spatiotemporal distributions: A review. *Waste Management*, 61:67–77.

Guo, Y., Barnes, S. J., and Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*.

Guoyan, S., Khaskheli, A., Raza, S. A., and Ahmed, M. (2022). Nonlinear impact of municipal solid waste recycling and energy efficiency on environmental performance and economic growth: evidence from non-parametric causality-in-quantiles. *Environmental Science and Pollution Research*.

Göransson, M., Nilsson, F., and Jevinger, A. (2018). Temperature performance and food shelf-life accuracy in cold food supply chains – insights from multiple field studies. *Food Control*, 86:332–341.

Hall, K., Guo, J., Dore, M., and Chow, C. (2009). The progressive increase of food waste in america and its environmental impact. *PLoS ONE*, 4(11).

Hebrok, M. and Boks, C. (2017). Household food waste: Drivers and potential intervention points for design – an extensive review. *Journal of Cleaner Production*, 151:380–392.

Hennchen, B. (2019). Knowing the kitchen: Applying practice theory to issues of food waste in the food service sector. *Journal of Cleaner Production*.

Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C. K., Heiskanen, A.-S., Johnson, R. K., Moe, J., Pont, D., Solheim, A. L., and de Bund, W. v. (2010). The european water framework directive at the age of 10: A critical review of the achievements with recommendations for the future. *Science of the Total Environment*.

Hiremath, B. and Goel, S. (2017). Survey of municipal solid waste (msw) treatment methods and compost samples.

Hou, D., Al-Tabbaa, A., Guthrie, P., and Watanabe, K. (2012). Sustainable waste and materials management: National policy and global perspective. *Environmental Science and Technology*, 46(5):2494–2495.

Hoy, Z., Woon, K., Chin, W., Hashim, H., and Fan, Y. (2022). Forecasting heterogeneous municipal solid waste generation via bayesian-optimised neural network with ensemble learning for improved generalisation. *Computers and Chemical Engineering*.

Hussain, A., Hannan, A., and Shafiq, M. (2023). Exploring mobile banking service quality dimensions in pakistan: a text mining approach. *International Journal of Bank Marketing*.

Istrate, I.-R., Iribarren, D., Gálvez-Martos, J.-L., and Dufour, J. (2020). Review of life-cycle environmental consequences of waste-to-energy solutions on the municipal solid waste management system. *Resources, Conservation and Recycling*.

Iyamu, H., Anda, M., and Ho, G. (2020). A review of municipal solid waste management in the bric and high-income countries: A thematic framework for low-income countries. *Habitat International*, 95:102097.

Joseph, L. P. and Prasad, R. (2020). Assessing the sustainable municipal solid waste (msw) to electricity generation potentials in selected pacific small island developing states (psids). *Journal of Cleaner Production*, 248:119222.

Joshi, R. and Ahmed, S. (2016). Status and challenges of municipal solid waste management in india: A review. *Cogent Environmental Science*.

Jurafsky, D. and Martin, J. H. (2020). *Speech and Language Processing*. 3rd edition.

Kabadurmus, O., Kazançoğlu, Y., Yüksel, D., and Pala, M. O. (2022). A circular food supply chain network model to reduce food waste. *Annals of Operations Research*.

Kamble, S., Gunasekaran, A., and Gawankar, S. (2020). Achieving sustainable performance in a data-driven agriculture supply chain: A review for research and applications. *International Journal of Production Economics*, 219:179–194.

Kandemir, C., Reynolds, C., Tom, Q., Fisher, K., Devine, R., Herszenhorn, E., Koh, S., and Evans, D. (2020). Using discrete event simulation to explore food wasted in the home. *Journal of Simulation*.

Kandemir, C., Reynolds, C., Tom, Q., Fisher, K., Devine, R., Herszenhorn, E., Koh, S., and Evans, D. (2022). Using discrete event simulation to explore food wasted in the home. *Journal of Simulation*.

Karak, T., Bhagat, R., and Bhattacharyya, P. (2012). Municipal solid waste generation, composition, and management: The world scenario. *Critical Reviews in Environmental Science and Technology*, 42(15):1509–1630.

Karami, A., Lundy, M., Webb, F., and Dwivedi, Y. K. (2020). Twitter and research: A systematic literature review through text mining. *IEEE Access*.

Kayikci, Y., Demir, S., Mangla, S. K., Subramanian, N., and Koc, B. (2022). Data-driven optimal dynamic pricing strategy for reducing perishable food waste at retailers. *Journal of Cleaner Production*.

Kazancoglu, Y., Ozkan-Ozen, Y., and Ozbiltekin, M. (2018). Minimizing losses in milk supply chain with sustainability: An example from an emerging economy. *Resources, Conservation and Recycling*, 139:270–279.

Khoiron, K., Probandari, A., Setyaningsih, W., Kasjono, H. S., Setyobudi, R. H., and Anne, O. (2020a). A review of environmental health impact from municipal solid waste (msw) landfill. *Annals of Tropical Medicine and Public Health*, 23:152–159.

Khoiron, K., Probandari, A. N., Setyaningsih, W., Kasjono, H. S., Setyobudi, R. H., and Anne, O. (2020b). A review of environmental health impact from municipal solid waste (msw) landfill. *Annals of tropical medicine and public health*, 23(3):60–67.

Kjeldsen, P., Barlaz, M. A., Rooker, A. P., Baun, A., Ledin, A., and Christensen, T. H. (2002). Present and long-term composition of msw landfill leachate: A review. *Critical Reviews in Environmental Science and Technology*, 32(4):297–336.

Kleijn, D. and Sutherland, W. J. (2003). How effective are european agri-environment schemes in conserving and promoting biodiversity? *Journal of Applied Ecology*.

Kodors, S., Zvaigzne, A., Litavniece, L., Lonska, J., Silicka, I., Kotane, I., and Deksne, J. (2022). Plate waste forecasting using the monte carlo method for effective decision making in latvian schools. *Nutrients*.

Korfiatis, N., Stamolampros, P., Kourouthanassis, P., and Sagiadinos, V. (2019). Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews. *Expert Systems with Applications*.

Ktenioudaki, A., Esquerre, C. A., Do Nascimento Nunes, C. M., and O'Donnell, C. P. (2022). A decision support tool for shelf-life determination of strawberries using hyperspectral imaging technology. *Biosystems Engineering*.

Kulikowska, D. and Klimiuk, E. (2008). The effect of landfill age on municipal leachate composition. *Bioresource Technology*, 99(13):5981–5985.

Lee, C., Ng, K., Kwong, C., and Tay, S. (2019). A system dynamics model for evaluating food waste management in hong kong, china. *Journal of Material Cycles and Waste Management*, 21(3):433–456.

Lee, H., Choi, I., Moon, H., and Kim, J. (2020). A multi-period product recommender system in online food market based on recurrent neural networks. *Sustainability (Switzerland)*, 12(3).

Lee, K. and Yu, C. (2018). Assessment of airport service quality: A complementary approach to measure perceived service quality based on google reviews. *Journal of Air Transport Management*.

Lewis, C. and Young, S. (2019). Fad or future? automated analysis of financial text and its implications for corporate reporting. *Accounting and Business Research*.

Li, C., Bremer, P., Harder, M. K., Lee, M. S., Parker, K., Gaugler, E. C., and Mirosa, M. (2022). A systematic review of food loss and waste in China: Quantity, impacts and mediators. *Journal of Environmental Management*, 303:114092.

Li, M., Hu, S., Xiang, J., Sun, L.-S., Li, P.-S., Su, S., and Sun, X.-X. (2003). Characterization of fly ashes from two chinese municipal solid waste incinerators. *energy & fuels*, 17(6):1487–1491.

Li, Y., Jin, Y., Li, H., Borrion, A., Yu, Z., and Li, J. (2018). Kinetic studies on organic degradation and its impacts on improving methane production during anaerobic digestion of food waste. *Applied Energy*, 213:136–147.

Liang, H., Sun, X., Sun, Y., and Gao, Y. (2017). Text feature extraction based on deep learning: a review. *Eurasip Journal on Wireless Communications and Networking*.

Liang, Z., Luo, Z., Yuan, J., Li, M., Xia, Y., Che, T., Huang, L., and Liu, J. (2022). Evaluating the environmental and economic performance of municipal solid waste disposal by all-component resource recovery. *Sustainability*.

Liegeard, J. and Manning, L. (2020). Use of intelligent applications to reduce household food waste. *Critical Reviews in Food Science and Nutrition*.

Liu, C. and wu Wu, X. (2011). Factors influencing municipal solid waste generation in china: A multiple statistical analysis study. *Waste Management & Research*, 29:371 – 378.

Lo, K. (2014). A critical review of china's rapidly developing renewable energy and energy efficiency policies. *Renewable and Sustainable Energy Reviews*.

Logan, R., Scherrer, B., Senecal, J., Walton, N., Peerlinck, A., Sheppard, J., and Shaw, J. (2021). Assessing produce freshness using hyperspectral imaging and machine learning. *Journal of Applied Remote Sensing*, 15(3).

Loureiro, S. M. C., Guerreiro, J., Eloy, S., Langaro, D., and Panchapakesan, P. (2019). Understanding the use of virtual reality in marketing: A text mining-based review. *Journal of Business Research*.

Lubura, J., Pezo, L., Sandu, M., Voronova, V., Donsì, F., Sic Zlabur, J., Ribic, B., Peter, A., Suric, J., Brandic, I., Klõga, M., Ostojic, S., Pataro, G., Virsta, A., Oros, A., Micic, D., Durović, S., De Feo, G., Procentese, A., and Voca, N. (2022). Food recognition and food waste estimation using convolutional neural network. *Electronics (Switzerland)*.

Lutoslawski, K., Hernes, M., Radomska, J., Hajdas, M., Walaszczyk, E., and Kozina, A. (2021). Food demand prediction using the nonlinear autoregressive exogenous neural network. *IEEE Access*.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., and Adam, S. (2018). Applying lda topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*.

Malefors, C., Secondi, L., Marchetti, S., and Eriksson, M. (2021a). Food waste reduction and economic savings in times of crisis: The potential of machine learning methods to plan guest attendance in swedish public catering during the covid-19 pandemic. *Socio-Economic Planning Sciences*.

Malefors, C., Strid, I., Hansson, P.-A., and Eriksson, M. (2021b). Potential for using guest attendance forecasting in swedish public catering to reduce overcatering. *Sustainable Production and Consumption*, 25:162–172.

Malefors, C., Sundin, N., Tromp, M., and Eriksson, M. (2022). Testing interventions to reduce food waste in school catering. *Resources, Conservation and Recycling*, 177.

Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to information retrieval*.

Mao, C., Feng, Y., Wang, X., and Ren, G. (2015). Review on research achievements of biogas from anaerobic digestion. *Renewable and Sustainable Energy Reviews*, 45:540–555.

Markande, A. R., Patel, D., and Varjani, S. (2021). A review on biosurfactants: properties, applications and current developments. *Bioresource Technology*, 330:124963.

Martin-Rios, C., Demen-Meier, C., Gössling, S., and Cornuz, C. (2018). Food waste management innovations in the foodservice industry. *Waste Management*.

Martin-Rios, C., Hofmann, A., and Mackenzie, N. (2021). Sustainability-oriented innovations in food waste management technology. *Sustainability (Switzerland)*.

Masud, M. H., Ananno, A. A., Ahmed, N., Dabnichki, P., and Salehin, K. (2020). Experimental investigation of a novel waste heat based food drying system. *Journal of Food Engineering*.

Mayer, I. S., Van Daalen, C. E., and Bots, P. W. (2004). Perspectives on policy analyses: A framework for understanding and design. *International Journal of Technology, Policy and Management*.

Mazzanti, M. and Zoboli, R. (2008). Waste generation, waste disposal and policy effectiveness: Evidence on decoupling from the european union. *Resources, Conservation and Recycling*, 52(10):1221–1234.

McKay, G. (2002). Dioxin characterisation, formation and minimisation during municipal solid waste (msw) incineration: Review. *Chemical Engineering Journal*.

Melikoglu, M. (2020). Appraising food waste generation and forecasting food waste to energy potentials of hospitals in Turkey: A global to local analysis. *Sustainable Production and Consumption*, 24:292–297.

Messner, R., Richards, C., and Johnson, H. (2020). The “prevention paradox”: food waste prevention and the quandary of systemic surplus production. *Agriculture and Human Values*.

Miguéis, V. L., Pereira, A., Pereira, J., and Figueira, G. (2022). Reducing fresh fish waste while ensuring availability: Demand forecast using censored data and machine learning. *Journal of Cleaner Production*.

Mobaseri, M., Mousavi, S. N., and Haghighi, M. H. M. (2021). Environmental effects of food waste reduction using system dynamics approach. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*.

Mosna, D., Vignali, G., Bottani, E., and Montanari, R. (2016). Life cycle assessment of a new feed production obtained by wasted flour food collected from the distribution and retail phases. *International Journal of Food Engineering*.

Mukherjee, C., Denney, J., Mbonimpa, E., Slagley, J., and Bhowmik, R. (2020). A review on municipal solid waste-to-energy trends in the usa. *Renewable and Sustainable Energy Reviews*, 119:109512.

Mäntylä, M. V., Graziotin, D., and Kuutila, M. (2018). The evolution of sentiment analysis—a review of research topics, venues, and top cited papers. *Computer Science Review*.

Møller Christensen, F., Solheim-Bojer, C., Dukovska-Popovska, I., and Steger-Jensen, K. (2021). Developing new forecasting accuracy measure considering product’s shelf life: Effect on availability and waste. *Journal of Cleaner Production*, 288.

Nair, K., Sekhani, B., Shah, K., and Karamchandani, S. (2021). Expiry prediction and reducing food wastage using iot and ml. *International Journal of Electrical and Computer Engineering Systems*, 12(3):155–162.

Nanda, H. S. (2011). Impact of municipal solid waste disposal on geotechnical properties of soil.

Nanda, S. and Berruti, F. (2021). Municipal solid waste management and landfilling technologies: a review. *Environmental Chemistry Letters*.

Noone, B. and Coulter, R. (2012). Applying modern robotics technologies to demand prediction and production management in the quick-service restaurant sector. *Cornell Hospitality Quarterly*, 53(2):122–133.

Nugraha, I. G. D., Wijaya, G. T., and Ramli, K. (2022). Improving meat expiration time prediction using the internet of things and polynomial regression. *ASEAN Engineering Journal*.

Orfanou, F., Dermesonlouoglou, E., and Taoukis, P. (2019). Greek coffee quality loss during home storage: Modeling the effect of temperature and water activity. *Journal of Food Science*, 84(10):2983–2994.

Ouzzani, M., Hammady, H., Fedorowicz, Z., and Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews*.

Ozkan, A., Ozkan, K., Isik, S., and Banar, M. (2021). Evaluation of different curve fitting models for prediction of municipal solid waste composition. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 27(3):385–393.

Page, M., McKenzie, J., Bossuyt, P., Boutron, I., Hoffmann, T., Mulrow, C., Shamseer, L., Tetzlaff, J., Akl, E., Brennan, S., Chou, R., Glanville, J., Grimshaw, J., Hróbjartsson, A., Lalu, M., Li, T., Loder, E., Mayo-Wilson, E., McDonald, S., McGuinness, L., Stewart, L., Thomas, J., Tricco, A., Welch, V., Whiting, P., and Moher, D. (2021). The prisma 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ*.

Papargyropoulou, E., Lozano, R., K. Steinberger, J., Wright, N., and Ujang, Z. B. (2014). The food waste hierarchy as a framework for the management of food surplus and food waste. *Journal of Cleaner Production*, 76:106–115.

Parfitt, J., Barthel, M., and MacNaughton, S. (2010). Food waste within food supply chains: Quantification and potential for change to 2050. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1554):3065–3081.

Patra, B. G., Maroufy, V., Soltanalizadeh, B., Deng, N., Zheng, W. J., Roberts, K., and Wu, H. (2020). A content-based literature recommendation system for datasets to improve data reusability – a case study on gene expression omnibus (geo) datasets. *Journal of Biomedical Informatics*.

Pattyn, V., Van Voorst, S., Mastenbroek, E., and Dunlop, C. A. (2018). Policy evaluation in europe. *The Palgrave handbook of public administration and management in Europe*, pages 577–593.

Pearson, D. and Minehan, M. (2013). Food waste in australian households: Why does it occur?

Pereira, L. N. and Cerqueira, V. L. (2021). Forecasting hotel demand for revenue management using machine learning regression methods. *Current Issues in Tourism*, 25:2733 – 2750.

Pérez, T., Vergara, S. E., and Silver, W. L. (2023). Assessing the climate change mitigation potential from food waste composting. *Scientific reports*, 13(1):7608.

Persson Osowski, C., Osowski, D., Johansson, K., Sundin, N., Malefors, C., and Eriksson, M. (2022). From old habits to new routines—a case study of food waste generation and reduction in four swedish schools. *Resources*.

Pina, M., Gaspar, P. D., and Lima, T. M. (2021). Decision support system in dynamic pricing of horticultural products based on the quality decline due to bacterial growth. *Applied System Innovation*.

Pirani, S. and Arafat, H. (2016). Reduction of food waste generation in the hospitality industry. *Journal of Cleaner Production*, 132:129–145.

Posch, K., Truden, C., Hungerländer, P., and Pilz, J. (2021). A bayesian approach for predicting food and beverage sales in staff canteens and restaurants. *International Journal of Forecasting*.

Poyatos-Racionero, E., Ros-Lis, J. V., Vivancos, J.-L., and Martínez-Máñez, R. (2018). Recent advances on intelligent packaging as tools to reduce food waste. *Journal of Cleaner Production*, 172:3398–3409.

Praneetrattananon, S., Wakisakai, M., Shirai, Y., and Mtpreechavanich, V. (2005). Kitchen refuse : A novel substrate for l (+) -lactic acid production by rhizopus oyyzae in submerged fermentation. *Japan Journal of Food Engineering*.

Priefer, C., Jörissen, J., and Bräutigam, K. R. (2016). Food waste prevention in Europe - A cause-driven approach to identify the most relevant leverage points for action. *Resources, Conservation and Recycling*, 109:155–165.

Principato, L., Mattia, G., Di Leo, A., and Pratesi, C. A. (2021). The household wasteful behaviour framework: A systematic review of consumer food waste. *Industrial Marketing Management*.

Principato, L., Pratesi, C., and Secondi, L. (2018). Towards zero waste: an exploratory study on restaurant managers. *International Journal of Hospitality Management*, 74:130–137.

Przydatek, G. and Kanownik, W. (2019). Impact of small municipal solid waste landfill on groundwater quality. *Environmental Monitoring and Assessment*, 191.

Rajaeifar, M. A., Ghanavati, H., Dashti, B. B., Heijungs, R., Aghbashlo, M., and Tabatabaei, M. (2017). Electricity generation and ghg emission reduction potentials through different municipal solid waste management technologies: A comparative review. *Renewable and Sustainable Energy Reviews*.

Ramachandra, T., Bharath, H., Kulkarni, G., and Han, S. S. (2018). Municipal solid waste: Generation, composition and ghg emissions in bangalore, india. *Renewable and Sustainable Energy Reviews*.

Ramos, A., Monteiro, E., Silva, V., and Rouboa, A. (2018). Co-gasification and recent developments on waste-to-energy conversion: A review. *Renewable and Sustainable Energy Reviews*.

Ramya, R., Ganesh Singh, T., Sejal, D., Venugopal, K., Iyengar, S., and Patnaik, L. (2018). Drdlc: Discovering relevant documents using latent dirichlet allocation and cosine similarity.

Ravi, A., Gurunathan, B., Rajendiran, N., Varjani, S., Gnansounou, E., Pandey, A., You, S., Raman, J. K., and Ramanujam, P. (2020). Contemporary approaches towards augmentation of distinctive heterogeneous catalyst for sustainable biodiesel production. *Environmental Technology & Innovation*, 19:100906.

Renou, S., Givaudan, J., Poulain, S., Dirassouyan, F., and Moulin, P. (2008). Landfill leachate treatment: Review and opportunity. *Journal of Hazardous Materials*, 150(3):468–493.

Richardson, Z., Perez-Guaita, D., Kochan, K., and Wood, B. (2019). Determining the age of spoiled milk from dried films using attenuated reflection fourier transform infrared (atr ft-ir) spectroscopy. *Applied Spectroscopy*, 73(9):1041–1050.

Rieger, J., Jentsch, C., and Rahnenführer, J. (2021). Rollinglda: An update algorithm of latent dirichlet allocation to construct consistent time series from textual data.

Rodrigues, M., Miguéis, V., and Freitas, S. (2023). A literature review on the quantitative approaches to food waste: descriptive, predictive, and prescriptive analyses. *Environmental Science and Pollution Research*.

Rowe, R. (2005). Long-term performance of contaminant barrier systems. *Geotechnique*.

Sakoda, G., Takayasu, H., and Takayasu, M. (2019). Data science solutions for retail strategy to reduce waste keeping high profit. *Sustainability (Switzerland)*, 11(13).

Salina Md Salim, N., Kudakasseril Kurian, J., Gariepy, Y., and Raghavan, V. (2016). Application and the techno-economical aspects of integrated microwave drying systems for development of dehydrated food products. *Japan Journal of Food Engineering*.

Salinas, D., Flunkert, V., Gasthaus, J., and Januschowski, T. (2020). Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*.

Salton, G. and McGill, M. J. (1986). *Introduction to Modern Information Retrieval*.

San Martin, D., Orive, M., Martínez, E., Iñarra, B., Ramos, S., González, N., Guinea de Salas, A., Vázquez, L., and Zufía, J. (2021). Multi-criteria assessment of the viability of valorising vegetable by-products from the distribution as secondary raw material for animal feed. *Environmental Science and Pollution Research*, 28(13):15716–15730.

Sarkar, S., Pramanik, A., Maiti, J., and Reniers, G. (2020). Predicting and analyzing injury severity: A machine learning-based approach using class-imbalanced proactive and reactive data. *Safety Science*.

Scarlat, N., Fahl, F., and Dallemand, J.-F. (2019). Status and opportunities for energy recovery from municipal solid waste in europe. *Waste and Biomass Valorization*.

Schneider, F. (2013). Review of food waste prevention on an international level. *Proceedings of Institution of Civil Engineers: Waste and Resource Management*, 166(4):187–203.

Schott, A. B. S. and Andersson, T. (2015). Food waste minimization from a life-cycle perspective. *Journal of Environmental Management*, 147:219–226.

Sciortino, R., Micale, R., Enea, M., and La Scalia, G. (2016). A webgis-based system for real time shelf life prediction. *Computers and Electronics in Agriculture*, 127:451–459.

Sedehzadeh, S. and Seifbarghy, M. (2021). Redesigning a closed loop food supply chain network considering sustainability and food banks with different returns. *Economic Computation and Economic Cybernetics Studies and Research*.

Shah, A. V., Srivastava, V. K., Mohanty, S. S., and Varjani, S. (2021). Municipal solid waste as a sustainable resource for energy production: State-of-the-art review. *Journal of Environmental Chemical Engineering*, 9(4):105717.

Shams, S., Sahu, J. N., Rahman, S. M. S., and Ahsan, A. (2017). Sustainable waste management policy in Bangladesh for reduction of greenhouse gases. *Sustainable Cities and Society*, 33:18–26.

Shapiro-Bengtsen, S., Andersen, F., Münster, M., and Zou, L. (2020). Municipal solid waste available to the chinese energy sector – provincial projections to 2050. *Waste Management*, 112:52–65.

Sharholy, M., Ahmad, K., Mahmood, G., and Trivedi, R. (2008). Municipal solid waste management in indian cities - a review. *Waste Management*.

Shatte, A. B. R., Hutchinson, D. M., and Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine*.

Shaw, P. J., Smith, M. M., and Williams, I. D. (2018). On the prevention of avoidable food waste from domestic households. *Recycling*, 3(2):24.

Shen, Y., Linville, J. L., Urgun-Demirtas, M., Mintz, M. M., and Snyder, S. W. (2015). An overview of biogas production and utilization at full-scale wastewater treatment plants (wwtps) in the united states: Challenges and opportunities towards energy-neutral wwtps. *Renewable and Sustainable Energy Reviews*.

Shindhal, T., Rakholiya, P., Varjani, S., Pandey, A., Ngo, H. H., Guo, W., Ng, H. Y., and Taherzadeh, M. J. (2021). A critical review on advances in the practices and perspectives for the treatment of dye industry wastewater. *Bioengineered*, 12(1):70–87.

Shirota, Y., Hashimoto, T., Sakura, T., and Chakraborty, B. (2016). Monetary policy topic extraction by using lda:: Termination of asian financial crisis. *Frontiers in Artificial Intelligence and Applications*.

Siddique, R., Khatib, J., and Kaur, I. (2008). Use of recycled plastic in concrete: A review. *Waste Management*.

Silvennoinen, K., Heikkilä, L., Katajajuuri, J.-M., and Reinikainen, A. (2015). Food waste volume and origin: Case studies in the finnish food service sector. *Waste management*, 46:140–145.

Singh, R. P., Tyagi, V. V., Allen, T., Ibrahim, M. H., and Kothari, R. (2011). An overview for exploring the possibilities of energy generation from municipal solid waste (msw) in indian scenario. *Renewable & Sustainable Energy Reviews*, 15:4797–4808.

Singh, V., Kumar, B., and Patnaik, T. (2013). Feature extraction techniques for handwritten text in various scripts: a survey. *International Journal of Soft Computing and Engineering (IJSCE)*, 3(1):238–241.

Sondh, S., Upadhyay, D. S., Patel, S., and Patel, R. N. (2022). A strategic review on municipal solid waste (living solid waste) management system focusing on policies, selection criteria and techniques for waste-to-value. *Journal of Cleaner Production*, 356:131908.

Sreedevi, S. (2015). Solid waste generation and its management-a case study.

Srivastava, A. and Nema, A. (2008). Forecasting of solid waste composition using fuzzy regression approach: A case of delhi. *International Journal of Environment and Waste Management*, 2(1-2):65–74.

Stancu, V., Haugaard, P., and Lähteenmäki, L. (2016). Determinants of consumer food waste behaviour: Two routes to food waste. *Appetite*, 96:7–17.

Sutherland, I. and Kiatkawsin, K. (2020). Determinants of guest experience in airbnb: A topic modeling approach using lda. *Sustainability (Switzerland)*.

Sánchez, C. (2009). Lignocellulosic residues: Biodegradation and bioconversion by fungi. *Biotechnology Advances*, 27(2):185–194.

Sánchez-Franco, M. J., Calvo-Mora, A., and Periáñez-Cristobal, R. (2023). Clustering abstracts from the literature on quality management (1980–2020). *Total Quality Management and Business Excellence*.

Taylor, R. G. and Allen, A. R. (2006). Waste disposal and landfill: information needs.

Teigiserova, D. A., Hamelin, L., and Thomsen, M. (2020a). Towards transparent valorization of food surplus, waste and loss: Clarifying definitions, food waste hierarchy, and role in the circular economy. *Science of the Total Environment*, 706:136033.

Teigiserova, D. A., Hamelin, L., and Thomsen, M. (2020b). Towards transparent valorization of food surplus, waste and loss: Clarifying definitions, food waste hierarchy, and role in the circular economy. *Science of the Total Environment*.

Thanh, N., Matsui, Y., and Fujiwara, T. (2010). Household solid waste generation and characteristic in a mekong delta city, vietnam. *Journal of Environmental Management*.

The European Commission (2003).

Thyberg, K. and Tonjes, D. (2016). Drivers of food waste and their implications for sustainable policy development. *Resources, Conservation and Recycling*, 106:110–123.

Tian, H., Gao, J., Lu, L., Zhao, D., Cheng, K., and Qiu, P. (2012). Temporal trends and spatial variation characteristics of hazardous air pollutant emission inventory from municipal solid waste incineration in china. *Environmental Science and Technology*, 46:10364–10371.

Tran, B. X., Ha, G. H., Nguyen, L. H., Vu, G. T., Hoang, M. T., Le, H. T., Latkin, C. A., Ho, C. S., and Ho, R. C. (2020). Studies of novel coronavirus disease 19 (covid-19) pandemic: A global analysis of literature. *International Journal of Environmental Research and Public Health*.

Troschinetz, A. M. and Mihelcic, J. R. (2009). Sustainable recycling of municipal solid waste in developing countries. *Waste Management*.

Ugwu, C. O., Ozoegwu, C. G., and Ozor, P. A. (2020). Solid waste quantification and characterization in university of nigeria, nsukka campus, and recommendations for sustainable management. *Heliyon*, 6(6):e04255.

United Nations Development Programme (UNDP) (2016). Municipal solid waste management in crisis and post-crisis settings. `https://www.undp.org/publications/municipal-solid-waste-management-crisis-and-post-crisis-settings`.

United Nations Environment Programme (2021). International environmental technology centre. `https://www.unenvironment.org/ietc`. Accessed: 14-Dec-2023.

U.S. Environmental Protection Agency (2023). Greenhouse gas reporting program (ghgrp) waste. https://www.epa.gov/ghgreporting/ghgrp-waste. Accessed on 29 Nov 2023.

van der Walt, A. and Bean, W. L. (2022). Inventory management for the in-flight catering industry: A case of uncertain demand and product substitutability. *Computers and Industrial Engineering*.

Vaverková, M. D., Adamcová, D., Zloch, J., Radziemska, M., Berg, A. B., Voberková, S., and Maxianová, A. (2018). Impact of municipal solid waste landfill on environment - a case study. *Journal of Ecological Engineering*.

Versluis, E., Van Keulen, M., and Stephenson, P. (2010). *Analyzing the European Union policy process*. Bloomsbury Publishing.

Vilas, C., Mauricio-Iglesias, M., and García, M. (2020). Model-based design of smart active packaging systems with antimicrobial activity. *Food Packaging and Shelf Life*, 24.

Vinti, G., Bauza, V., Clasen, T. F., Medlicott, K. O., Tudor, T., Zurbrügg, C., and Vaccari, M. (2021). Municipal solid waste management and adverse health outcomes: A systematic review. *International Journal of Environmental Research and Public Health*, 18.

Vittuari, M., Garcia Herrero, L., Masotti, M., Iori, E., Caldeira, C., Qian, Z., Bruns, H., van Herpen, E., Obersteiner, G., Kaptan, G., Liu, G., Mikkelsen, B. E., Swannell, R., Kasza, G., Nohlen, H., and Sala, S. (2023). How to reduce consumer food waste at household level: A literature review on drivers and levers for behavioural change. *Sustainable Production and Consumption*.

Vittuari, M., Pagani, M., Johnson, T., and De Menna, F. (2020). Impacts and costs of embodied and nutritional energy of food waste in the us food system: Distribution and consumption (part b). *Journal of Cleaner Production*, 252.

Vyas, S., Prajapati, P., Shah, A. V., and Varjani, S. (2022). Municipal solid waste management: Dynamics, risk assessment, ecological influence, advancements, constraints and perspectives. *Science of The Total Environment*, 814:152802.

Wallace, W. and Wallace, H. (1998). Policy-making in the european union. *CrossRef Listing of Deleted DOIs*, 28:157.

Wing, C., Simon, K., and Bello-Gomez, R. A. (2018). Designing difference in difference studies: Best practices for public health policy research. *Annual Review of Public Health*.

World Bank (2021). Trends in solid waste management. `https://datatopics.worldbank.org/what-a-waste/trends_in_solid_waste_management.html`.

World Bank Group (2018). What a waste 2.0: A global snapshot of solid waste management to 2050. Technical report.

Wu, Z., He, Q., Chen, Q., Xue, H., and Li, S. (2021a). A topical network based analysis and visualization of global research trends on green building from 1990 to 2020. *Journal of Cleaner Production*.

Wu, Z., Zhang, Y., Chen, Q., and Wang, H. (2021b). Attitude of chinese public towards municipal solid waste sorting policy: A text mining study. *Science of the Total Environment*.

Xiao, S., Dong, H., Geng, Y., Francisco, M.-J., Pan, H., and Wu, F. (2020). An overview of the municipal solid waste management modes and innovations in shanghai, china. *Environmental Science and Pollution Research*.

Yang, R., Xu, Z., and Chai, J. (2018). A review of characteristics of landfilled municipal solid waste in several countries: Physical composition, unit weight, and permeability coefficient. *Polish Journal of Environmental Studies*.

Yue, F., Guo, K., Yue, M., and Chen, W. (2021). Topic mining and effectiveness evaluation of china's coal-related policy based on lda model. *Lecture Notes in Electrical Engineering*.

Zacarias-Farah, A. and Geyer-Allély, E. (2003). Household consumption patterns in oecd countries: Trends and figures. *Journal of Cleaner Production*.

Zhang, D. Q., Tan, S. K., and Gersberg, R. M. (2010). Municipal solid waste management in china: Status, problems and challenges. *Journal of Environmental Management*.

Zhang, Y., Chen, M., Huang, D., Wu, D., and Li, Y. (2017). idoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization. *Future Generation Computer Systems*.

Zhao, P., Shen, Y., Ge, S., and Yoshikawa, K. (2014). Energy recycling from sewage sludge by producing solid biofuel with hydrothermal carbonization. *Energy Conversion and Management*.

Zhu, Y., Zhang, Y., Luo, D., Chong, Z., Li, E., and Kong, X. (2020). A review of municipal solid waste in china: characteristics, compositions, influential factors and treatment technologies. *Environment, Development and Sustainability*, 23:6603 – 6622.

Zou, Y., Wu, J., Liu, G., Piron, M., Fedele, A., Antonio, S., and Manzardo, A. (2022). Examining the trade-offs in potential retail benefits of different expiration date modes: Insights into multidimensional scenarios. *Resources, Conservation and Recycling*.