MESG
Mestrado em Engenharia
de Serviços e Gestão

Exploring and Preventing Churn in Gyms

Ana Raquel dos Santos Costa Oliveira

Master Thesis

FEUP Supervisor Professor Vera Miguéis
Co-supervisor Professor Ivo Pereira

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

2024-07-31

To my family ♥ ♥ ♥

Abstract

The fitness industry has experienced significant growth in recent years, highlighting the need for effective strategies to retain members and reduce churn. This thesis focuses on exploring the reasons behind gym membership cancellations and developing predictive models to anticipate potential churn. Considering variables such as demographic, behavioral and member-related variables, the study aims to provide actionable insights for gym management to improve member retention and satisfaction. The research uses machine learning techniques to predict churn and proposes strategic interventions based on the findings. Ultimately, the goal is to help health clubs maintain a stable and loyal customer base to ensure sustainable growth and profitability.

Keywords

Industry of Fitness/Gym, Churn, Retention, Predictive Models, Machine Learning

Acknowledgments

I would like to express my deepest gratitude to all those who have given me invaluable support throughout this project.

I would like to start by expressing my sincerest gratitude to my supervisors, Professor Vera Miguéis and Professor Ivo Pereira, for their guidance and support. Their invaluable guidance, precise feedback and encouragement were fundamental to the realization of this work.

In addition, I would like to thank my colleagues at Life Club and Dr Carlos Ribeiro, as CEO, for their support and flexibility.

I would like to extend my gratitude to my Mum and Dad, who has constantly supported me and without whom this would not have been possible.

I would also like to express my gratitude to my family (my husband, Daniel, and my sons, Gustavo and Mateus), who have been a source of strength, inspiration and joy throughout this journey. Your unwavering support and affection have been fundamental in enabling me to reach this milestone.

Contents

List of tables

List of figures

List of abbreviations

AUC	Area under the curve
CEQ	Customer Experience Quality
CRISP-DM	Cross-Industry Standard Process for Data Mining
CRM	Customer Relationship Management
IHRSA	International Health Racquet and Sportsclub Association
K-NN	K-Nearest Neighbors
SVM	Support Vector Machine
UX	Customer Experience
WHO	World Health Organization

1. Introduction

1.1. Project Background

The World Health Organization (WHO 2018) emphasizes the importance of regular physical activity in preventing and treating several chronic non-communicable diseases. In line with this, the WHO's Global Action Plan on Physical Activity 2018-2030 aims to reduce insufficient physical activity by 15% among adolescents by 2030. Additionally, the European Physical Activity Strategy 2016-2025 encourages physical activity and aims to make society aware of its benefits. The WHO defines health as a state of extensive physical, mental, and social well-being. The prevalence of non-communicable diseases is alarming, with these diseases responsible for 70% of deaths worldwide by 2020. Moreover, WHO data indicates that one in ten people suffer from a mental illness during their lifetime (WHO 2018).

The relationship between health and fitness is multifaceted and deeply interconnected. Fitness is essential for health because it encompasses the ability to execute daily functional activities with optimal performance, endurance, and strength, which helps manage disease, fatigue, stress, and reduces sedentary behavior (Kapoor et al. 2022). Regular physical activity has been shown to reduce the negative effects of aging and slow the development and progression of chronic diseases in older adults. In this way fitness contributes to health by improving cardiorespiratory and muscular fitness, bone and cardiometabolic health, and weight status, while also boosting mental and social health (Kapoor et al. 2022; Liusnea 2022; Tokarski, Tosounidis, and Zarotis 2023). The growing awareness of health and well-being leads to people becoming more informed about the benefits of regular physical exercise and this reflets in the increase of member in gyms and fitness centers. As can be seen in Figure 1, which represents the estimates for the year in Europe, there is a growing increase in gym membership, which has only been slowed down by the Covid-19 pandemic. In Portugal (Figure 2), the same trend is being followed, although it should be borne in mind that the figures presented for Portugal are not estimates but actual figures for a given year.

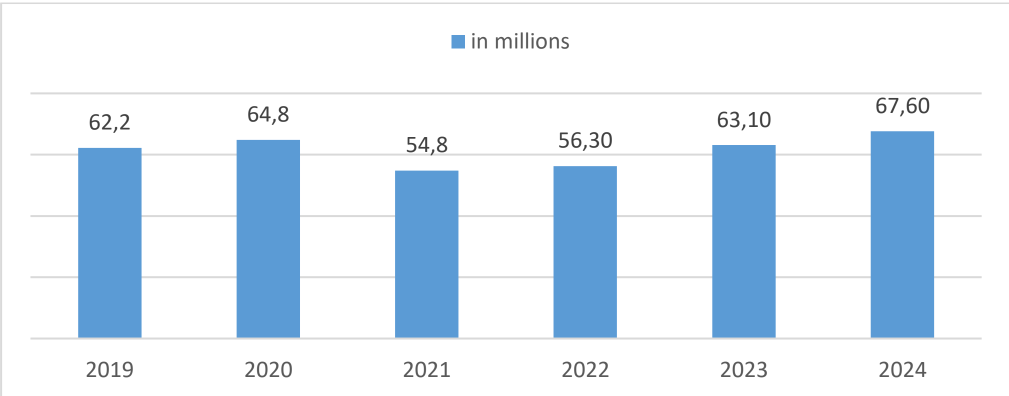


Figure 1 - Number of fitness club member in Europe (adapted from: Rutgers et al. ((Rutgers et al. 2021, 2020, 2022, 2019, 2023, 2024))

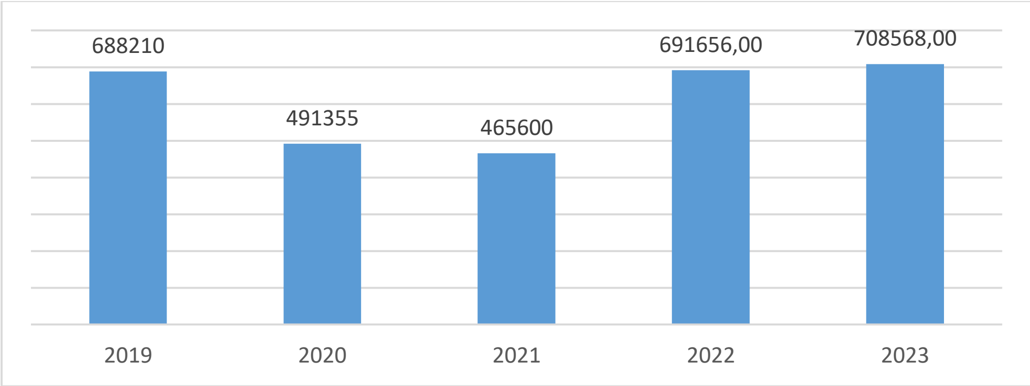


Figure 2 - Number of fitness club member in Portugal (adapted from: (Rutgers et al. 2021, 2020, 2022, 2019, 2023, 2024))

The increase in the number of people interested in exercising has led to a significant growth in investment in gyms. With more individuals seeking a healthy lifestyle, investors see a lucrative opportunity (APPENDIX A: Revenue in Europe and Portugal (2020 – 2024) to open new gyms or improve existing ones to attract and retain clients

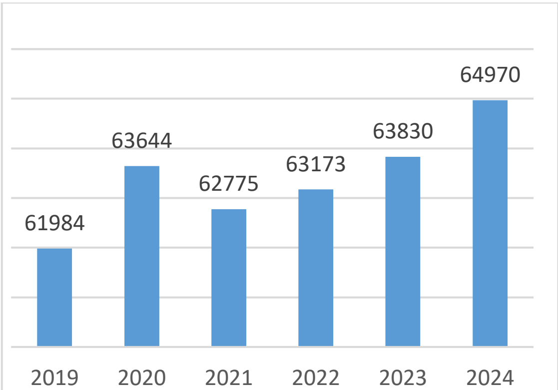


Figure 3 - Number of clubs in Europe (adapted from:(Rutgers et al. 2019, 2020, 2021, 2022, 2023, 2024)

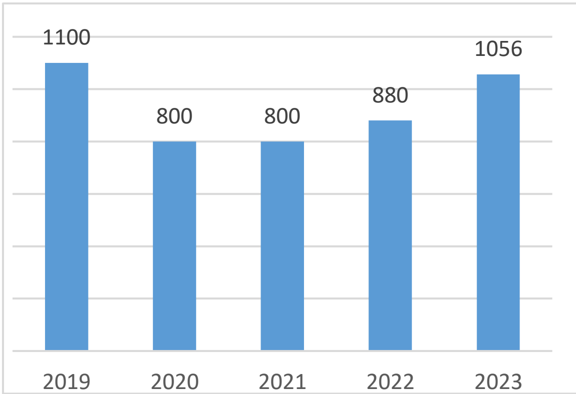


Figure 4 - Number of clubs in Portugal (adapted from: (Rutgers et al. 2019, 2020, 2021, 2022, 2023, 2024)

Between 2019 and 2024, the number of member and consequently the revenue trends (APPENDIX A: Revenue in Europe and Portugal (2020 – 2024) in Portugal and Europe faced considerable turbulence due to the COVID-19 pandemic, which began impacting global economies in 2020. This period saw varied responses in actual revenues in Portugal and projected revenues across Europe, each reflecting the timing and extent of the pandemic's economic disruptions. The initial sharp declines reflect the immediate disruptions caused by the pandemic, while the later growth indicates the resilience of economic systems and the effectiveness of recovery strategies employed to mitigate the pandemic's effects and stimulate economic growth.

The growing demand, after COVID-19 pandemic, encourages investment which in turn led to the proliferation of gyms and consequently the constant search for space in the fitness market. The variety of offers and types has made consumers increasingly sensitive to issues such as price, variety of services and their quality, leading them to switch gyms more easily. Currently, customer retention, loyalty and prevention of churn are crucial to ensuring a steady cash flow.

and to achieve this it is necessary to keep customers motivated and engaged through strategies such as loyalty programs and personalized service. In this context, sustainable financial growth in the fitness industry depends directly on how well its managers manage to keep customers engaged. Understanding the factors that drive customer engagement in health and fitness clubs is critical for managers. Retention in this context refers to the act of keeping customers from switching to another service provider, which is crucial in industries with intense competition like the fitness industry (Aldosary and Alrashdan 2021). This is one of the biggest challenges facing health and fitness club managers today and to achieve this goal, managers need to ensure customer satisfaction by providing quality services. The satisfaction of customers will foster loyalty and ensure continued sponsorship.

1.2. Problem Description

Fitness services are an important part of the health and activity industry, which has experienced significant growth in recent years (Pedragosa; and Ferreira; 2023). The fitness industry grapples with a concerning phenomenon for gym managers and owners: churn, referring to the rate at which customers discontinue their memberships over time (Clavel San Emeterio et al. 2020). Understanding the reasons why gym-goers abandon their workout routines can be crucial for the long-term sustainability and success of these establishments.

To combat churn, gym managers must be responsive to their clients' changing needs, with predicting gym member churn being crucial for proactive decision-making and implementing preventative strategies. Exploring predictive machine learning models, comparing various classifiers, and identifying key features for churn prediction constitute a holistic approach to enhance member retention in gyms. This not only provides fresh insights into undertaking churn but also fosters more informed decision-making and strategic development in the fitness industry. Therefore, strategies such as loyalty programs, personalization of services and investment in innovation can be used in a preventative rather than reactive way.

1.3. Project Description

The project was developed in Life Club and centers around understanding and mitigating customer churn. Churn prediction and the development of strategies to retain customers are critical aspects of this project. In the last years, as seen a significant growth and an increase in active users, however, to sustain this growth, it is essential to understand why customers leave (churn) and implement strategies to retain them.

The Life Club project can be structured in four stages:

1. **Identify and understand** why former members' churn: in a closer approach to find out from former members why they cancelled
2. **Characterization and analysis of the data**: Understand the Life Club's data and which variables can be explored.

3. **Identification of the intention of Churn**: Analyze different algorithms to identify patterns and reasons for churn. Identify the variables that influence the intention of churn.
4. **Formulation and Analysis of Solutions**: Develop strategies to address the identified causes of churn. This might include personalized engagement tactics, loyalty programs, or improved customer support.

With the following a structured approach, the Life Club can effectively predict churn and implement strategies to enhance customer retention, thereby maintaining its growth trajectory and competitive edge in the market.

1.4. Research Questions

The fitness industry, characterized by its dynamic and competitive nature, continuously seeks innovative strategies to retain members and sustain growth. In this context, the application of predictive machine learning models to understand and anticipate gym member churn emerges as a potentially transformative approach. This concept leverages the power of data analytics to extract actionable insights from member behavior, aiming to enhance member retention and foster loyalty. For dressing the problem, this research propose answer to these three questions:

- RQ1. Can the use of predictive machine learning models applied to the context of gyms and member churn provide a state-of-the-art approach to address churn challenge?

- RQ2. Can the identification of relevant features for predicting gym member churn provide valuable insights for developing effective retention strategies to reduce churn and increase customer loyalty?

The first question, the utilization of predictive machine learning models in gyms to predict member churn represents a state-of-the-art and data-driven strategy (Macon, 2020). Analyzing historical data and identifying patterns correlated with member drop-offs, gyms can proactively intervene, offering personalized experiences or incentives to at-risk members, thus potentially reversing the churn trend.

The second question highlights the importance of identifying relevant features that contribute to predicting gym member churn. Understanding these key predictors can provide gym owners and managers with profound insights into member preferences and pain points. This knowledge, in turn, can inform the development of targeted retention strategies, such as personalized workout plans, flexible membership options, or improved gym facilities, all aimed at addressing the specific reasons behind member dissatisfaction and churn.

1.5. Research Gaps

One of the notable shortcomings of the scientific literature is that most articles on churn predominantly focus on the telecommunications and banking sectors (Sundararajan and Gursoy 2020); (Ullah et al. 2019). This limited scope can result in a biased understanding of churn, as it does not encompass the full spectrum of industries where churn is a critical issue. Telecommunications and banking, while vital and relevant, represent only a portion of the broader economic and consumer landscape. Consequently, the insights and conclusions derived from these sectors may not be universally applicable or reflective of other industries. Focusing mainly on telecommunications and banking means that the research might miss out on the unique churn dynamics present in other fields such as the one addressed in this thesis. Each industry has its specific customer behavior patterns, market conditions, and competitive environments, which can significantly influence churn rates and the effectiveness of retention strategies. When neglecting these diverse contexts, the research risks providing a skewed perspective that limits the generalizability of its findings.

Another point is the limited research on mapping the entire customer journey in the fitness industry. Identify and understand each touchpoint and potential pain points throughout the member lifecycle could help identify critical areas for intervention to reduce churn (Kwon and Nam 2022; Jang and Baek 2024).

1.6. Company characterization

The Life Club in Vila Nova de Gaia, is a Health Club with diverse infrastructure, catering to a wide range of user needs. Covering an area of approximately 3000 m², the gym stands out for its spacious and naturally lit environment, contributing to a welcoming and motivating atmosphere.

Life Club is equipped with a variety of facilities and services designed to promote health and well-being, including:

- Strength and Cardio Fitness Area: Features modern and diverse equipment for strength training and cardiovascular workouts, with personalized training plans and support from personal trainers.
- Swimming Pool: Offers aquatic activities such as AquaBike, AquaJump Circuit, water aerobics, swimming lessons for all ages, and personalized aquatic training, catering to both beginners and advanced swimmers.
- SPA and Wellness: Includes services like therapeutic and relaxation massages, hydrojet, jacuzzi, Turkish bath, and sauna, providing a comprehensive environment for recovery and relaxation.
- Classes: Offers a wide range of group classes, including Pilates, yoga, Zumba, HIIT, cross training, fit boxing, stretching, Latin rhythms, among others, catering to different preferences and fitness goals.

Life Club is renowned for its family-friendly and welcoming environment, with close and personalized service.

1.7. Document Structure

This document is structured to provide a comprehensive understanding of the study on preventing and mitigating churn in gyms. The structure follows a logical sequence that facilitates reading and understanding of the topics covered, starting with the “Introduction” section with the "Project Background," which contextualizes the project by highlighting the importance of regular physical activity and global initiatives to promote exercise. It also addresses the growth in gym memberships and the impact of the COVID-19 pandemic on the fitness industry. The "Problem Description" that defines the issue of churn, explaining why understanding membership cancellations is crucial for the long-term sustainability of gyms. The "Project Description" outlines the objective of the project developed at Life Club, aiming to identify the reasons behind membership cancellations and develop predictive models to anticipate churn. The "Research Questions" section presents the key research questions, focusing on the application of machine learning models for churn prediction and identifying relevant features. The "Research Gaps" section discusses the limitations of existing research, need for studies addressing churn in the fitness industry. The "Company Characterization" provides an overview of Life Club, including its facilities, services, and customer base, setting the context for the study.

The “Literature Review” section explores various facets of the fitness industry. "The Fitness Industry" subsection examines growth trends both globally and in Portugal, the "Retention" subsection analyzes strategies for retaining gym members, emphasizing the economic importance of customer retention and various approaches documented in the literature. The "Churn" subsection provides a detailed examination of churn, its implications for the fitness industry, and the factors influencing gym membership cancellations. The "Comparative Analysis of Existing Approaches" reviews different machine learning techniques used in churn prediction and their applicability to the fitness industry. Finally, the "Marketing Strategies" subsection presents effective marketing strategies, including personalized services, CRM systems, loyalty programs, and customer experience enhancements.

The “Methodology” section describes the CRISP-DM model adopted for this research, detailing its phases and their practical application in the study.

The “Data and Results” section is divided into three parts. "Data Distribution from the Interviews" analyzes data collected from interviews with former gym members, providing insights into their reasons for cancellation and suggestions for improvement. "Data Distribution from the Database" presents data extracted from the gym's database, highlighting key variables and their distributions. The "Machine Learning Model Results" subsection compares different predictive models, evaluating their performance based on various metrics and selecting the most effective model for churn prediction.

And finally, the “Conclusion” section summarizes the main findings of the study, underscoring the importance of understanding and preventing gym member churn. It discusses the practical implications of the results and offers recommendations for future retention strategies.

2. Literature Review

This section provides an in-depth exploration of the existing body of knowledge related to the fitness industry, retention, churn and marketing strategies.

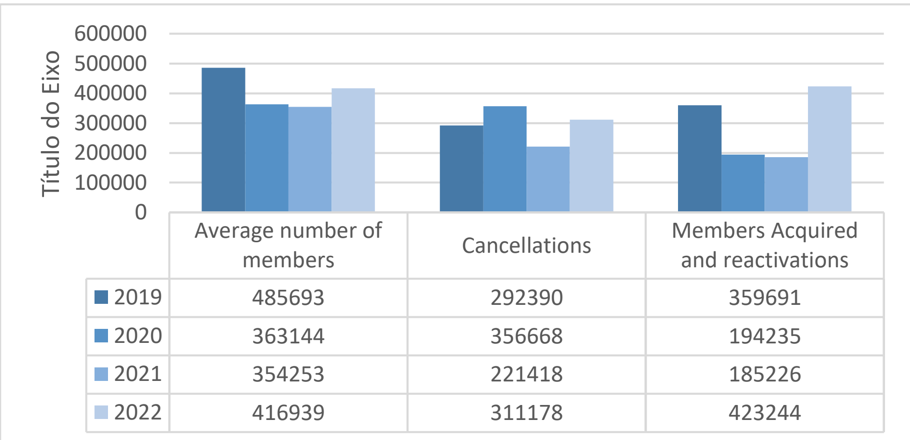
2.1. The Fitness Industry

The fitness industry has shown resilience adapting to the changing market dynamics and consumer preferences. Notably, the fitness market has consistently grown since the downturn in 2020, driven by factors like rising obesity levels, ageing populations, and growing consumer interest in health and wellness.

Globally, the fitness industry is also experiencing positive trends. According to the IHRSA Global Report (2023), about 80% of health club, gym, and studio operators expect membership and revenue to grow by more than 5% in 2023. This optimism is fueled by an increased consumer desire to improve physical health and well-being, with 86% of consumers in more than 50 countries and territories motivated to take more action towards their health today than in the past.

In Europe, the fitness market displays impressive figures, with an industry revenue of €28 billion and 63.1 million members attending over 63,830 clubs. In the Portuguese market, from 2017 onwards, the gym's performance has remained consistent, peaking in 2019 at 289 million euros in sales, 1100 clubs, 688210 subscribers and 6.7 per cent market penetration. These indicators stabilized in 2021 compared to 2020, after a sharp decline in 2020 due to the external impact of COVID-19. In 2022, there was already an upturn in the sector, with club revenues of around 230 million euros, 880 clubs, more than 691,000 members and a market penetration rate of 6.7% (similar to 2019) (Pedragosa 2023).

Although this growth pattern, based on the information from the 2022 barometer, the cancellation rate of members in gyms increased from 2021 to 2022 by 12 percentage points, to 75%. This means that on average 75 people cancel their membership out of every 100 active members.



	Average number of members	Cancellations	Members Acquired and reactivations
2019	485693	292390	359691
2020	363144	356668	194235
2021	354253	221418	185226
2022	416939	311178	423244

Figure 5 - Average number of Members, Cancellations and Members Acquired and reactivations (adapted from: (Pedragosa 2023))

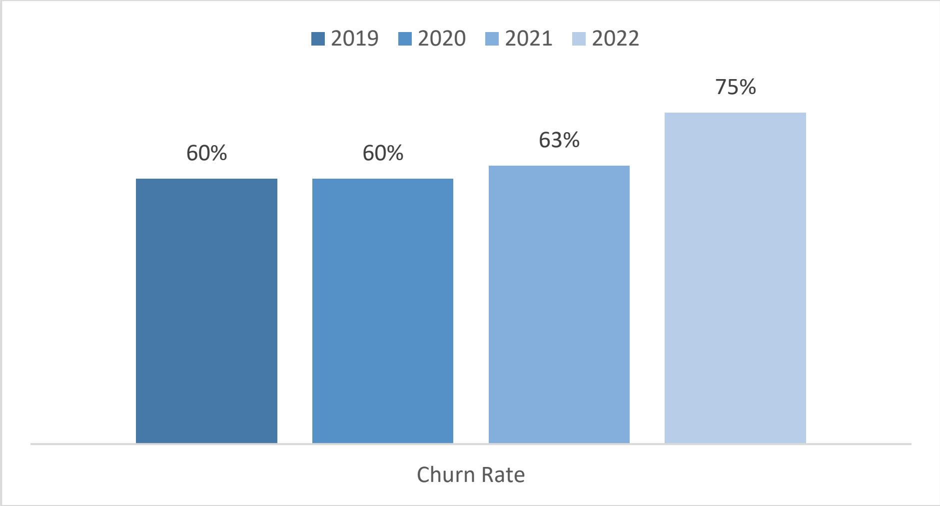


Figure 6 - Churn Rate (adapted from: (Pedragosa 2023))

Based on the data for the years 2019 to 2022, it is possible to see the average fluctuation in the number of members over the years. There was a significant drop from 2019 to 2021, with the lowest number in 2021 at 354,253 members. However, there is a notable increase in 2022, reaching 416,939 members. The number of cancellations peaked in 2020 with 356,668. There was a decrease in 2021 with 221,418 cancellations, but an increase again in 2022 to 311,178. This suggests a level of volatility in member retention across the years.

Related to the members acquired and reactivations there was a sharp decline from 2019 to 2021, with the lowest acquisitions and reactivations in 2021 at 185,226. However, there is a significant rebound in 2022, with the highest number across the four years at 423,244.

The churn rate has been increasing each year. It remained steady at 60% for 2019 and 2020, increased to 63% in 2021, and then jumped to 75% in 2022. This indicates that while the number of members acquired and reactivated in 2022 is the highest, the churn rate also increased significantly, which is a concerning trend for the organization. Despite acquiring more members in 2022, the churn rate has increased, indicating that the problem may not be solely with acquisition but also with retention. The peak in cancellations in 2020 might be due to external factors (such as the COVID-19 pandemic) that could have influenced member behavior.

This data reveals that while more members are joining or reactivating, a larger proportion is leaving, which could be indicative of deeper issues with member satisfaction or engagement. The steady churn rate in 2019 and 2020 despite different membership numbers suggests that there were issues already existing with retention that have not been addressed effectively.

The data also suggests a critical need for the organization to examine their retention strategies and implement targeted measures to improve member satisfaction and engagement, to reduce the churn rate. According to Brinks (2020), acquiring new customers can be notably more expensive, five times more, than retaining existing ones. This underscores the importance of focusing on customer retention. Moreover, Gallo (2014) refers another layer to this understanding. It states that even a modest increase in customer retention (5%) can lead to a substantial increase in profits, ranging from 25% to 95%. This increase can be attributed to a range of factors, including the reduced marketing costs of reaching out to existing customers, the increased purchases over time from loyal customers, and the positive word-of-mouth that satisfied customers can generate.

These statistics collectively show the importance of businesses investing in customer relationship management and retention strategies. They suggest that focusing on keeping the current customer base satisfied is not just beneficial but also highly cost-effective compared to the efforts and resources required to acquire new customers.

2.2. Retention

The insights from Sobreiro et al. (2021), Aldosary and Alrashdan (2021), Srinivasan, Rajeswari, and Elangovan (2023), Rahman Rahman, Alam, and Hosen (2022), Koutsoubakis (2020) and Artha et al. (2022) offers a comprehensive understanding of retention, particularly in the context of fitness centers, but also extending to broader business models. These perspectives collectively underscore the multifaceted nature of retention strategies, the importance of understanding member behavior, and the economic implications of maintaining a loyal customer base.

Sobreiro et al. (2021) sets the stage by emphasizing the economic importance of retention in fitness centers, highlighting the cost-effectiveness of maintaining existing members compared to the acquisition of new ones. This introduces the concept of retention as a key factor in the profitability and sustainability of fitness businesses. For example, Macon (2020) highlights that customer retention is crucial for the financial success of gyms, as acquiring new customers can be costly and time-consuming. The research shows that about 50% of customers abandon the gym within 90 days of joining, and only about 3.7% continue their gym activities for more than 12 consecutive months. To increase retention, gym managers should focus on engagement and communication strategies with customers, as well as offering high-quality and personalized services.

Aldosary and Alrashdan (2021) and Srinivasan, Rajeswari, and Elangovan (2023) explore into the specific strategies and efforts implemented to retain customers. Aldosary and Alrashdan (2021) focuses on a wide array of factors including member behavior, habit formation, satisfaction, and perceived service quality, among others, providing a nuanced understanding of the factors influencing retention in fitness clubs. Srinivasan, Rajeswari, and Elangovan (2023) discusses the overarching strategies to prevent customer churn in subscriber-based service models, emphasizing the financial viability and revenue retention. Rahman, Alam, and Hosen (2022) further broadens this perspective by considering the analysis of customer actions, revenue impact, and measures to reduce churn rate, thereby highlighting the strategic importance of retention in ensuring business continuity and profitability.

The empirical studies by Koutsoubakis (2020) contribute by highlighting specific strategies and outcomes of retention efforts in the fitness industry. Koutsoubakis (2020) demonstrates the effectiveness of orientation programs in improving member integration and retention.

Artha et al. (2022) synthesizes these viewpoints by framing customer retention as a dynamic and complex process influenced by several factors. The emphasis is on building a loyal customer base through continuous value provision, trust maintenance, and satisfaction ensuring. This perspective encapsulates the essence of the preceding discussions, highlighting the necessity of a comprehensive, strategic approach to retention that encompasses understanding and influencing customer behavior, service quality enhancement, effective use of analytics, and the strategic alignment of various retention efforts.

The scientific literature presents a multifaceted and detailed picture of retention, emphasizing its significance in the economic viability of fitness centers and other businesses. The insights suggest that successful retention strategies require a holistic understanding of customer behavior, a focus on providing continuous value and satisfaction, and the strategic alignment of various retention-related efforts.

2.3. Churn

The phenomenon of churn, which Ahn et al. (2020) defines as the tendency of customers to cease their business relationships with a company over a specific period or contractual term, is increasingly garnering attention across various sectors. This concept, often termed the cancellation, manifests mainly in industries such as telecommunications and banking, as highlighted in the research of Ascarza et al. (2018) and Lalwani et al. (2022). Orina, Rimiru, and Mwangi (2023) underscores the importance of these sectors in modern research, emphasizing the need for a comprehensive analysis.

In the fitness industry, the issue of churn is becoming particularly pronounced. Gyms are beginning to recognize the essential role of member retention in ensuring the sustainability and growth of their businesses. As Colgate and Danaher (2000) suggests, understanding the underlying patterns of churn is fundamental for businesses to predict and counteract customer losses effectively. This understanding prompts the adoption of proactive and preventive strategies, especially in sectors dependent on subscriptions or memberships. Given the competitive nature of the fitness industry, retaining members emerges as a challenge, requiring proactive measures to curtail churn and elevate customer satisfaction. Strategies such as loyalty programs, personalized training, regular progress tracking, and fostering an engaged community are crucial, as Semrl and Matei (2017) notes.

Moreover, the ongoing analysis of churn data and the refinement of predictive models are vital to ensure that these strategies attuned to evolving member preferences and behaviors. Azzouz, Boukhedouma, and Alimazighi (2022) contends that a predictive approach serves as a potent marketing tool in combating churn, enabling the anticipation of customer behavior and the assertive implementation of retention strategies. This is particularly pertinent in customer-centric commercial enterprises where competition is intense, and customers are in pursuit of competitive pricing, value for money, and superior service quality. In their studies, various authors have mentioned the variables they applied to their models, some of which are: Age and Gender (Clavel San Emeterio et al. 2020; Azzam et al. 2023; Semrl and Matei 2017; Lalwani et al. 2022; Sobreiro et al. 2021; Sobreiro et al. 2022); Payments and Member attendance frequencies (Semrl and Matei 2017; Azzouz, Boukhedouma, and Alimazighi 2022); Contract and Membership length (Clavel San Emeterio et al. 2020; Azzam et al. 2023; Semrl and Matei 2017; Lalwani et al. 2022; Sobreiro et al. 2021; Sobreiro et al. 2022).

Furthermore, Clavel San Emeterio et al. (2020) and Orina, Rimiru, and Mwangi (2023) associate churn specifically with the withdrawal of members from leisure centers, noting demographic and behavioral patterns like younger age, shorter membership duration, infrequent visits, reduced time per visit, and lower expenditure at the center. These insights are invaluable for comprehending customer behavior and devising more effective retention strategies.

Authors such as Artha et al. (2022), Macon (2020), Ghadiri et al. (2021), Muñoz et al. (2022), and Koutsoubakis (2020) delve deeper into churn, discussing proactive management, service performance enhancement, the implementation of service improvement plans, and the optimization of expenditure strategies. In particular, Koutsoubakis (2020) spotlights specific strategies aimed at mitigating churn and bolstering member retention in commercial fitness organizations, such as the Gym Orientation Program, which aids new members in adapting to the gym environment, thereby fostering member retention.

Liang (2023) addresses churn within the broader context of highly competitive markets, emphasizing the importance of churn prediction in industries where acquiring new customers is costlier than retaining existing ones. These industries include mobile telecommunications, banking, insurance, retail markets, gaming, logistics, e-commerce, and media. Building upon this, Ahn et al. (2020) discusses the advent of churn analysis via deep learning algorithms, while Rahman, Alam, and Hosen (2022), Aldosary and Alrashdan (2021), Soumi, P, and Paulose (2021), Srinivasan, Rajeswari, and Elangovan (2023), and Sobreiro et al. (2021) underscore the significance of predicting churn, comprehending its revenue impact, reducing churn rates, and scrutinizing dropout reasons. They advocate using sophisticated datasets and algorithms to probe customer marketing trends from extensive databases.

Integrating these perspectives reveals that addressing churn involves a multifaceted approach, spanning customer behavior and demographic analysis, proactive strategy implementation, and leveraging advanced technologies for prediction and retention. This field demands constant innovation and adaptability in response to evolving customer expectations and market dynamics. The concept of churn, particularly within the fitness sector, transcends the mere cessation of customer relationships with a company. As Clavel San Emeterio et al. (2020) illustrates, churn behavior stems from a complex interplay of internal factors (such as physical and psychological aspects) and external factors (including environmental features, social interactions, institutional policies and norms, and familial influences). This intricate behavioral system implies that addressing churn necessitates a thorough examination of the behavioral characteristics and predictive elements of dropout tendencies.

Clavel San Emeterio et al. (2020) underscores the practical aspects of churn management, highlighting the comparative ease of retaining a member contemplating departure over re-engaging one who has already left. This observation is pivotal for fitness establishments, stressing the importance of identifying and addressing the concerns of at-risk members proactively, prior to their actual departure. Such an approach harmonizes with the strategies supported by Azzouz, Boukhedouma, and Alimazighi (2022) and Koutsoubakis (2020), who defend predictive methodologies and early intervention tactics to combat churn effectively.

Understanding the root causes behind gym members' decision to terminate their memberships is not merely reactive; it's vital for the long-term viability and success of these fitness establishments. The complex nature of churn, as detailed by Clavel San Emeterio et al. (2020), calls for a comprehensive strategy that encompasses the analysis of behavioral patterns, the crafting of tailored retention strategies, and the continual innovation in engagement methods. This holistic approach not only addresses immediate churn-related issues but also fosters a robust foundation for customer loyalty and sustainable business growth in the fiercely competitive landscape of the fitness industry.

Eskiler and Safak (2022) also emphasizes the importance of customer experience quality (CEQ) in customer loyalty in fitness services. The research shows that service outcome quality, customer-employee interaction quality, and customer-customer interaction quality are critical factors affecting CEQ, which in turn influences customer loyalty. Improving these interactions can help reduce churn, as satisfied and loyal customers are less likely to leave the gym. Therefore, to reduce churn in gyms, it is essential to focus on service quality, customer satisfaction and loyalty, and adapting to new technological and behavioral trends among consumers.

2.4. Comparative Analysis of Existing Approaches

A comprehensive overview of various machine learning techniques employed in different studies in the fitness sector for predictive modeling and data analysis tasks, such as churn prediction, dropout behavior analysis, and other related predictive tasks, as seen in Table 1. Each technique is summarized, provides a brief description and lists the authors who have utilized these methods in their research. This consolidation aims to offer a clear perspective on the diversity and applicability of machine learning techniques in tackling complex analytical challenges across various domains.

Table 1 - Machine learning techniques in different studies (own construction)

Technique	Description	Authors
Logistic Regression	Used for binary classification problems to predict the likelihood of an event. Known for its simplicity and efficiency.	Muñoz et al. (2022), Azzouz, Boukhedouma, and Alimazighi (2022), Clavel San Emeterio et al. (2020), Orina, Rimiru, and Mwangi (2023) Lalwani et al. (2022), Muñoz et al. (2022), Liang (2023), Rahman, Alam, and Hosen (2022), Soumi, P, and Paulose (2021), Srinivasan, Rajeswari, and Elangovan (2023), Sobreiro et al. (2021), Azzam et al. (2023).
Decision Trees	Non-parametric supervised learning method used for classification and regression, capturing complex patterns by dividing the dataset based on feature value conditions.	Höppner et al. (2020), Azzouz, Boukhedouma, and Alimazighi (2022), Orina, Rimiru, and Mwangi (2023), Lalwani et al. (2022), Muñoz et al. (2022), Liang (2023), Rahman, Alam, and Hosen (2022), Soumi, P, and Paulose (2021), Srinivasan, Rajeswari, and Elangovan (2023), Sobreiro et al. (2021).
Random Forest	An ensemble learning method constructing multiple decision trees at training time,	Höppner et al. (2020), Azzouz, Boukhedouma, and Alimazighi (2022),Orina, Rimiru, and Mwangi (2023), Lalwani et al. (2022), Liang (2023), Aldosary

	outputting the class that is the mode of the classes of individual trees.	and Alrashdan (2021), Soumi, P, and Paulose (2021), Sobreiro et al. (2021).
Support Vector Machine (SVM)	Powerful for a wide range of classification problems, useful for both linear and non-linear classification.	Vafeiadis et al. (2015), Azzouz, Boukhedouma, and Alimazighi (2022), Orina, Rimiru, and Mwangi (2023), Lalwani et al. (2022), Muñoz et al. (2022), Liang (2023), Srinivasan, Rajeswari, and Elangovan (2023), Sobreiro et al. (2021),Azzam et al. (2023).
Neural Networks / Deep Learning	Capable of uncovering intricate patterns in large datasets, suitable for complex prediction tasks like churn prediction.	Aldosary and Alrashdan (2021), Azzouz, Boukhedouma, and Alimazighi (2022), Liang (2023), Rahman, Alam, and Hosen (2022), Soumi, P, and Paulose (2021)
K-Nearest Neighbors (KNN)	A simple, versatile algorithm used for classification and regression, predicting the value of new data points based on most of its 'K' nearest neighbors.	Vafeiadis et al. (2015), Liang (2023), Azzam et al. (2023)
Naïve Bayes	Simple probabilistic classifiers based on applying Bayes' theorem with strong independence assumptions between the features.	Azzouz, Boukhedouma, and Alimazighi (2022), Lalwani et al. (2022), Liang (2023), Soumi, P, and Paulose (2021), Srinivasan, Rajeswari, and Elangovan (2023), Azzam et al. (2023).

2.4.1. Other Techniques

Various other machine learning techniques and approaches have been mentioned across studies, including:

- Gradient Boosting Classifier (Sobreiro et al. 2021)
- CatBoost (Lalwani et al. 2022)
- Extra Tree Classifier (Lalwani et al. 2022)
- Multi-layer perceptron (MLP) (Aldosary and Alrashdan 2021)
- Feature selection (Aldosary and Alrashdan 2021)
- SMOTE-ENN, Recursive Feature Elimination (Srinivasan, Rajeswari, and Elangovan 2023)
- Dual-ANN, SOM+ANN, Ant-Miner, ALBA (Srinivasan, Rajeswari, and Elangovan 2023)
- Multilayer sensors, CNN-derived model (Srinivasan, Rajeswari, and Elangovan 2023)

2.5. Marketing Strategies

Recent studies have explored marketing strategies for gyms and fitness centers, highlighting the importance of targeted approaches to attract and retain customers. Key strategies include utilizing social media for promotion (Tsitskari and Batrakoulis 2021), offering attractive pricing and maintaining facilities (Tsitskari and Batrakoulis 2021), and implementing digital marketing techniques (Garcia-Feo and Rodríguez-Rodríguez 2020). Some gyms have focused on niche markets, such as female-only facilities (Wang, Hsiao, and Hsiung 2022). The marketing strategies found in the literature encompass various approaches to customer retention, loyalty, and satisfaction. One of the strategies are Relationship Marketing that involves creating, maintaining, and enhancing long-term relationships with customers and stakeholders to meet their needs and expectations (Albérico and Joaquim A. 2023), (Della 2022). This strategy includes developing a core service, personalizing relationships, offering extra benefits, pricing to stimulate loyalty, and communicating effectively with employees to enhance their performance with customers. For gyms, Della (2022) suggests implementing the offer of personalized fitness plans, hosting member appreciation events, and maintaining regular check-ins to ensure customer satisfaction. Benefits include increased customer retention and loyalty, higher profitability per customer due to lower attraction costs, and reduced sales costs.

The Customer Relationship Management (CRM) is used to manage business relationships with customers, aiming for prosperous long-term relationships with loyal and profitable customers (Almohaimmeed 2019). CRM systems can manage communication, analyze customer data, and improve process efficiency (Almohaimmeed 2019). A CRM system can track member attendance, preferences, and feedback, allowing personalized communication and targeted marketing campaigns to increase satisfaction and loyalty. Almohaimmeed (2019) highlights the positive CRM positively impacts customer satisfaction, loyalty, profitability, and organizational performance. Moreover, the Quality of Products and Services is crucial for generating positive word-of-mouth, customer loyalty, and satisfaction (Nursanti and Tomoliyus 2021). High-quality gym equipment, clean facilities, and professional trainers can make a significant difference. One of the essential points are ensuring that the gym environment is safe, welcoming, and well-maintained encourages members to stay loyal and recommend the gym to others. The creation of Loyalty Programs is also popular, aiming to support the design of marketing actions to make customers more loyal, by offering rewards such as redeemable points, cash, and free products in exchange for their business (Mujayana 2020). In the case of redeemable points, gyms can implement loyalty programs where members earn points for attending classes, referring friends, or achieving fitness milestones, which can be redeemed for discounts on memberships, merchandise, or personal training sessions. These programs increase purchase frequency and customer trust in the brand (Mujayana 2020).

The Customer Experience (UX) is crucial for customer satisfaction and retention. An intuitive design and easy-to-use interface are important for mobile fitness apps, for instance (Nursanti and Tomoliyus 2021). In fact, Gyms can enhance UX by providing user-friendly apps that allow members to book classes, track their progress, and access fitness resources seamlessly. Improving usability and providing accurate information are essential for increasing user satisfaction (Mujayana 2020). According to Nalluri and Chen (2022) the marketing experience involves creating memorable experiences for customers, which can increase loyalty and purchase intention. Gyms can create unique fitness events, workshops, or challenges that engage members emotionally and physically. These experiences can include themed workout sessions, community-building activities, or exclusive access to new fitness trends. Components such as sensory, emotional, cognitive, and behavioral aspects influence customers' purchase intentions (Nalluri and Chen 2022).

Additionally, the Communication Strategies are fundamental for building trust and ensuring that products reach customers (Kotler and Keller 2021). An effective communication can include regular newsletters with fitness tips, updates on new classes, and motivational messages. Promotion methods such as advertising, public relations, and personal selling can help in promoting the gym, its services, and its community (Kotler and Keller 2021) These marketing strategies are essential for customer retention and loyalty, ensuring satisfaction, which in turn contributes to the success and sustainable growth of gyms. Betting in personalizing services, using CRM systems, ensuring high quality, implementing loyalty programs, enhancing customer experience, creating memorable experiences, and maintaining effective communication, gyms can foster a loyal customer base and thrive in a competitive market.

3. Methodology

This section outlines the systematic approach adopted to investigate and address the issue of gym member churn. This research utilizes the Cross-Industry Standard Process for Data Mining (CRISP-DM) model, a proven and structured framework for data mining projects.

3.1. Method Used in the Thesis

Data Mining has become an important tool used in various academic and industrial domains to extract valuable knowledge from large datasets. In the context of this thesis, the adoption of the Cross-Industry Standard Process for Data Mining (CRISP-DM)(Chapman 2000) model represents a structured and proven approach to effectively and systematically conduct data mining projects.

The choice of CRISP-DM as the reference model for this thesis is motivated by its ability to guide the process from the initial understanding of the requirements of the problem to the practical implementation of the proposed solutions. This model provides a flexible and adaptable framework that allows it to be applied in a variety of contexts and disciplines. Using CRISP-DM, a consistent methodological approach to the exploration and analysis of the data that is adopted. The different phases of the model - from business and data understanding to modelling, evaluation and deployment - will provide a solid foundation for developing and validating the hypotheses and conclusions of the thesis.

In addition, the iterative nature of CRISP-DM will allow for adjustments to be made as new insights are gained throughout the research process, thus ensuring a robust and improved analysis of the data. It is expected that the use of this model will contribute significantly to achieving the objectives outlined in my thesis. It will also promote a rigorous and efficient methodological approach to data analysis.

The methodology unfolds through six meticulously structured steps, each playing a pivotal role in distilling valuable information from member data.

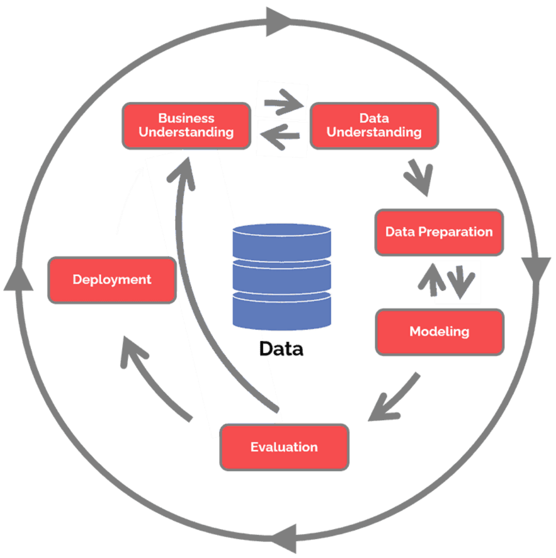


Figure 7 - Methodology CRISP-DM (Chapman 2000)

The first phase is Business Understanding, where the objective is to understand the thesis objectives and requirements from a business perspective. This knowledge was then converted into a data mining problem definition and a preliminary plan. Until the start of the thesis, the data was only used to identify the number of new members and the number of cancellations, so that this gap could be filled in the next month. With the start of the study, the organization aims to understand and anticipate the cancellation of subscriptions, thus achieving greater retention and a lower churn rate.

One of the first steps was to access the cancellations and find out why they had occurred. To identify and understand the main reasons why former members cancelled their gym memberships, interviews were conducted. The sample selection was made up of former members who cancelled their subscriptions in the year 2023, ensuring that the experiences were recent and relevant. A total of 20 former members were randomly selected from the cancellation forms handed in at the time of cancellation. With the defined sample, a semi-structured interview was developed (APPENDIX B: semi-structured interview with gym former members), with 10 questions, including open questions and in some questions a scale (from 0 to 10) was used to complete the response. The open questions allowed for more detailed insights, while the closed questions facilitated quantitative analysis. The script of interview was structured in 9 parts to touch on all the essential points, such as the gym experience (enrolment history, general assessment, services used), observations over the time (changes of quality), communication and support, positive points, reasons for quitting, decision moment, suggestions for improvement and final comments. The ex-members were then contacted by telephone and the invitation to take part in the research was explained in detail. The invitation explained the purpose of the research, the importance of their participation and guaranteed the confidentiality of their responses. The interviews were conducted over the phone and lasted an average of 15 minutes. The interviews were carried out until the responses were saturated.

By finding out through interviews the main reasons why they had cancelled their subscriptions, it was easier to determining the data mining objectives. The interviews revealed that certain variables were pivotal in the cancellation process. These included the frequency of visits to the gym, which was identified as a key factor. Additionally, the interviewees noted a consistent decline in the number of visits until the cancellation, which was attributed to a lack of payment. Furthermore, the type of activity undertaken during visits, whether collective or individual, was identified as a significant aspect. This is because the presence or absence of guidance and support in these activities could potentially lead to increased demotivation and subsequent churn.

The second phase, Data Understanding, involves collecting initial data, familiarizing oneself with the data, identifying data quality issues, and discovering initial insights. In this thesis, data was collected from the gym's database, covering the period from January 2022 to February 2024. The dataset consists of a sample of 3,217 gym members. Tasks here include collecting initial data from CRM, describing the data's properties, performing exploratory data analysis, and verifying data quality. This phase helps in understanding the nature of the data and its suitability for addressing the business objectives. Relevant data variables include membership duration (reflected in the payment), frequency of attendance, types of classes attended, maximum frequency for membership type, member profile, contract, frequency ratio, difference in frequency between $n-(n-1)$, number of months of contracted membership, number of hours spend at the gym, mode of the time of day for training, mode of the activity and demographic information.

To explain the variables used in predicting customer churn for a gym, here is a brief description of each:

Table 2 - Description of variables

Variable	Description	Type
Gender	Gender of the member: Male or Female	Text 0 = Female 1 = Male
Age	Age of the member in years	Number
Payments	Member´s payment history. It indicates if the member churned or not. Used as a target variable	Binary 1 = with payment 0 = first month without payment, moment of churn
Member attendance frequencies	The number of times the member visits the gym in a month	Number
Maximum Frequency for membership type	The maximum allowed visits according to the membership type (e.g., a basic plan may allow 9 visits per month, while a premium plan may allow 30 visits). This is compared to the current frequency to identify if the member is fully utilizing their plan.	Number
Member Profile	Type of membership the member has (basic, premium, etc.).	Text
Contract	Indicates whether the member has a membership with a contract or not. Members with a contract may be less likely to cancel in the first six months, the duration of the contract	Binary 1 = with contract 0 = without contract
Frequency Ratio	The ratio between the current frequency and the maximum frequency allowed by the plan. This can indicate the member's level of engagement with the gym.	Number
Difference in Frequency between $\boldsymbol{n-(n-1)}$	The change in visit frequency between two consecutive months. A significant drop can signal increasing disinterest and potential churn.	Number
Number of Months of Contracted Membership	The total time the member has stayed with the gym under a contract. Can stay the time of contract (six month) or extend for more months	Number
Number of Hours Spent at the Gym	The total hours the member has spent at the gym.	Number
Mode of the Time of Day for Training	The most frequent time of day the member chooses to train. This can indicate the member's habits and preferences, helping to better understand their behavior.	Number 0 = Without payment and frequency 1 = Morning (7am – 12am)

		2 = Lunch (12am – 3pm) 3 = Afternoon (3pm – 5pm) 4 = Night (6pm – 10pm)
Mode of the type Activity done	The most frequent activity done. If it was collective or singular.	Number 0 = without information 1 = individual 2 = group

These variables provide a comprehensive overview of customer behavior, allowing for a detailed analysis of patterns that may indicate a higher or lower likelihood of churn

The analysis of the variables chosen to predict churn in the gym is essential to understand the factors influencing members' decisions to continue or discontinue their memberships. The variable "Payments" is crucial as it tracks members' payment history, indicating whether they made payments and the first month without payment, which is a direct indicator of churn. The "Member attendance frequencies" variable, which measures the number of times a member visits the gym in a month, reflects member engagement, as those who visit the gym regularly are less likely to churn. The "Maximum Frequency for membership type" variable allows for comparing the actual usage frequency with the maximum allowed by the member's plan. If a member is not fully utilizing their membership, it may indicate disinterest. The "Member Profile," describing the type of membership (basic, premium, etc.), is important because different member profiles may exhibit different behaviors. The "Contract" variable indicates whether a member has a membership contract, with this the main objective is see if member is likely to churn with contract (even if they must stay for 6 months) or if not having this obligation makes members stay longer. The "Frequency Ratio" measures the ratio between the current frequency and the maximum allowed frequency by the plan, indicating how much a member is utilizing their plan. The "Difference in Frequency $n-(n-1)$" variable assesses the change in visit frequency between two consecutive months. A significant drop in frequency can signal increasing disinterest and potential churn. The "Number of Months of Contracted Membership" measures the total number of months a member has been under contract, suggesting that members who have been contracted for longer may be more resistant to churn, while newer members might be more prone to quitting. The "Number of Hours Spent at the Gym" variable indicates the total hours a member has spent at the gym, reflecting their level of engagement and satisfaction with the services provided. The "Mode of the Time of Day for Training" variable shows the most frequent time of day the member chooses to train, helping to identify if the gym meets the members' needs at specific times and if the time of the day influence the churn of the member. Lastly "Mode of Activity done" identify the activity (singular or group) the member attend more.

In the third phase, Data Preparation, the objective is to construct the final dataset from the initial raw data. At this stage, the CRM database was extracted from all member entries between January 1, 2022, and February 28, 2024. For this set of members, the existing payments were checked and the first and last payments were identified, with the month following the last payment marked as the month of abandonment. Frequencies per month were calculated for each member. The time spent at the gym was calculated using the entry and exit time records. One of the problems encountered was the lack of information about the time members left, many of which were blank. To make up for this lack of data, the median time spent in the gym was calculated from the existing records.

To understand the time of day when the members went to train, considering the time they entered the gym, the following conditions were considered: between >=7am and <12am "Morning" period >=12am and <3pm 'Lunch' period, between >=3pm and <5pm "Afternoon" period and between >=5pm and 9.30pm "Evening" period. This division by time of day is an attempt to understand which period of the day they train most often and to calculate the mode of the period of the day in which they train for each month. Conditions were also created for the type of sport they practice, in this case dividing the activities by type, either individual or group. In this case, as there were many members who did not record the type of activity, the value "0" was assigned to no record, "1" individual activity and "2" group activity. New Feature engineering was performed to create new variables such as average monthly attendance, compare attendance between n and n-1. This phase ensures the dataset is clean, relevant, and ready for the modeling phase, thus improving the accuracy and reliability of the predictions.

The fourth phase, Modeling, involves selecting and applying various modeling techniques and calibrating their parameters to optimal values. The tasks include selecting appropriate modeling techniques, creating a plan for training and testing the models, building the models using the prepared data, and assessing the models' performance using metrics like accuracy, precision, recall, and AUC-ROC. For this thesis, based on the literature review, techniques such as logistic regression, decision trees, random forest, Support Vector Machine (SVM), Neural Networks, K-Nearest Neighbors (KNN) and Naïve Bayes were employed to predict customer churn. In the way to facilitate this, RapidMiner, a powerful data science platform, was utilized to test various models. RapidMiner's intuitive interface and comprehensive tools allowed for efficient model building and evaluation. This phase helps in creating predictive models that can accurately identify customers likely to churn.

The process showed in the RapidMiner figures represents the process created for building, training, validating, and testing the machine learning models. The process involves reading data from an Excel file, preparing the data, training a model using different machine leaning algorithms, and then validating the model's performance. This approach ensures that the machine learning model is both accurate and reliable, enabling informed decision-making based on the data.

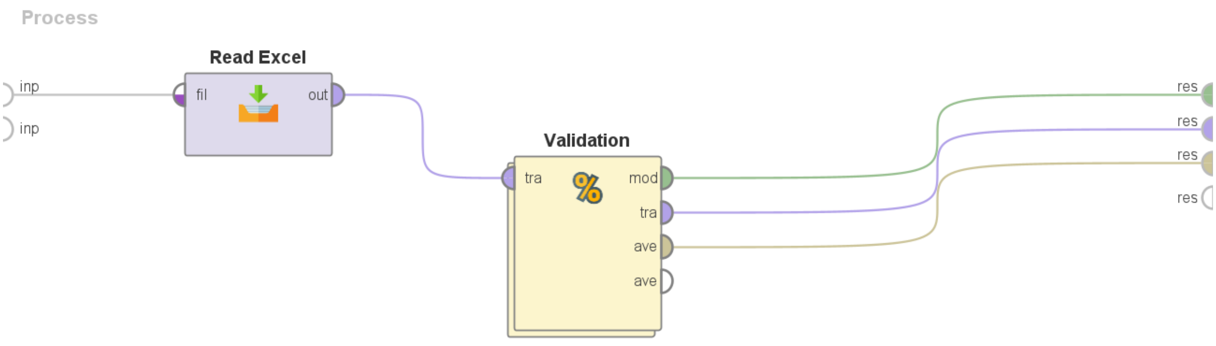


Figure 8 - RapidMiner Complete Process

The operator "Read Excel" is used to load the dataset from an Excel file into RapidMiner. The data read from the Excel file serves as the starting point for the entire process. The validation phase is critical as it splits the data into training and testing sets, applies the machine learning model, and evaluates its performance. This ensures that the model is evaluated on data it has not seen during training, providing an unbiased estimate of its performance. (Figure 8 - RapidMiner Complete Process).

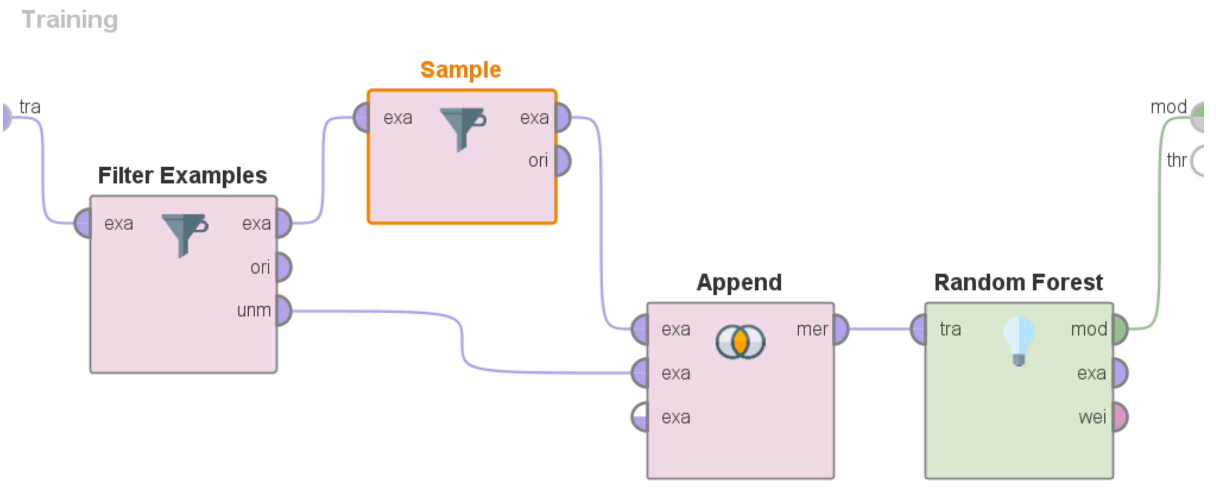


Figure 9 - The training process inside the validation

The Filter Examples operator selects subsets of the data based on predefined conditions, helping to exclude irrelevant or anomalous data points. This step is crucial for ensuring that the training data is of high quality. The Sample operator creates a smaller, manageable subset of the dataset for training purposes. This helps in speeding up the training process while maintaining the representativeness of the sample. The Append operator merges multiple datasets into one, allowing for a comprehensive training dataset that includes all relevant data points. The Apply Model operator uses the different models to make predictions on the test data, generating output that includes both the predictions and the original test data (Figure 9 - The training process inside the validation)

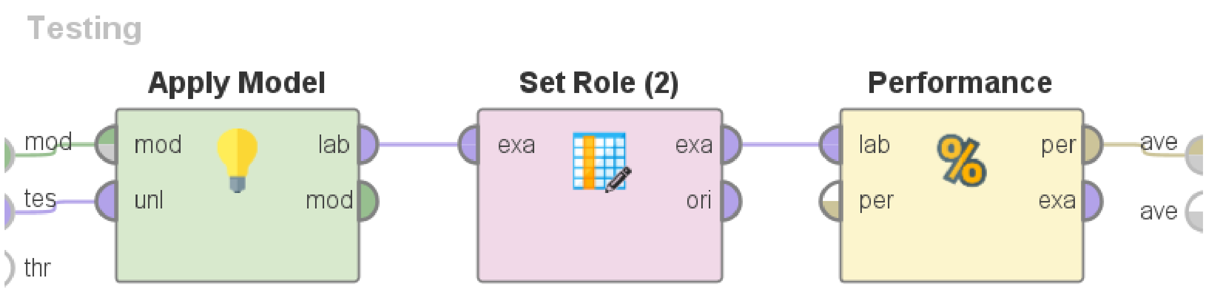


Figure 10 - The testing process inside the validation

The Apply Model operator uses the different models to make predictions on the test data, generating output that includes both the predictions and the original test data. The Set Role operator assigns specific roles to the attributes in the dataset, such as designating which attribute represents the predicted label, in this case "Payments". Finally, the Performance operator assesses the model's effectiveness using a variety of metrics, including accuracy, precision, recall, AUC (area under the curve) (optimistic), AUC and AUC (pessimistic) (Figure 10 - The testing process inside the validation)

The fifth phase is Evaluation, building the models using the prepared data, and assessing the models' performance using metrics like accuracy, precision, recall, specificity, negative predictive values and AUC and F1-score.

In predictive modeling, especially in contexts like churn prediction where the objective is to identify customers who are likely to stop using a service, evaluating the performance of our models is crucial. Different metrics (Figure 11) provide various perspectives on how well our model is performing and help us understand the trade-offs involved in its predictions.

		Predicted Churn			
		Negative	Positive		
Actual Churn	Negative	True Negative (TN)	False Positive (FP)	Specificity $\frac{TN}{(TN+FP)}$	AUC $\frac{TP}{(TP+FN)}$ $\frac{FP}{(FP+TN)}$
	Positive	False Negative (FN)	True Positive (TP)	Recall $\frac{TP}{(TP+FN)}$	F1-Score $2x\,\frac{Precision\ x\ Recall}{Precision + Recall}$
		Negative Predictive Value $\frac{TN}{(TN+FN)}$	Precision $\frac{TP}{(TP+FP)}$	Accuracy $\frac{TP+TN}{(TP+TN+FP+FN)}$	

Figure 11 - Confusion matrix and performance metrics

The accuracy of the models was calculated as the proportion of correct predictions (both true positives and true negatives) out of the total of predictions. It provides an overall view of how well the model predicted churn. The specificity, or the ability to identify members who will not churn, is crucial to avoid false alarms and make sure that retention resources are not wasted on customers who are not at risk. The Recall or Sensitivity to correctly identify members who are churning. A high recall is essential to ensure that the maximum number of churning members are identified for the retention plan. The precision in our model indicates the proportion of true positives out of the predicted positives. In this case, high precision ensures that retention efforts are focused on members who are truly at risk of churning. The calculation of negative predictive value measures the proportion of members predicted not to churn who do not churn. A high negative predictive value ensures that customers who are predicted not to churn are not at risk. The F1 score combines precision and recall into a single metric and is useful for evaluation. The AUC is useful to measure the ability of a classifier to distinguish between positive and negative classes. For all the models considered to be relevant, the metrics described above have been calculated to check which model is best suited to the purpose of predicting churn. In addition to the traditional AUC, two specific variations have been calculated: AUC Optimistic and AUC Pessimistic. The AUC Optimistic refers to an optimistic estimate of the model's performance, often derived from testing the model on the same data it was trained on, which can result in an overestimation of its true performance. On the other hand, the AUC Pessimistic provides a more conservative estimate, typically obtained under more stringent testing conditions or from a dataset that presents greater challenges, thereby giving a lower bound on expected performance. Based on the model with the best performance, the variables that had the greatest impact on the results are analyzed and customer retention actions can be created.

The final phase, Deployment, involves deploying the model in the operational environment and monitoring its performance. Tasks include planning the deployment to implement the model in the business environment, continuously monitoring and maintaining the model’s performance, reporting the results, and conducting a post-mortem to review what went well and what could be improved. In this project, the model will be integrated into the gym's customer relationship management (CRM) system to provide real-time churn predictions and insights for retention strategies. This phase ensures the model will be effectively integrated into business processes and provides actionable insights to reduce churn.

4. Data and Results

This section presents a detailed analysis of the data collected and the outcomes derived from the application of various machine learning models. This section is divided into three key parts: data distribution from the interviews, data distribution from the database, and machine learning model results.

4.1. Data Distribution from the Interviews

The data from the interviews with former gym members reveal some clear trends. Most of the interviewees attended the gym for periods ranging from six months to a little over a year, with frequencies of one to five times per week. significant portion of gym members attended 2-3 times per week, with many also maintaining their membership for 6 to 12 months. Most respondents expressed high satisfaction, rating their experience between 8 and 10, while a smaller group gave ratings below 7. The pool, cardio facilities, and group classes were the most popular services among members. Most interviewees did not observe significant changes in service quality during their membership, and the vast majority rated communication and support as good or excellent. The primary reasons for leaving the gym included relocation, professional obligations, lack of time, budget constraints, and health-related issues such as pregnancy. Overall, the data indicate that despite high satisfaction with the gym's services and support, external and personal factors were the main reasons for leaving.

4.2. Data Distribution from the Database

The data taken from the customer database for the period from 01 January 2022 to 28 February 2024 is shown here in the Table 3 .

Table 3 - Main variables analyzed

Variable	Description	Mean	Median	Mode	Standard deviation	Minimum	Maximum
Age	years	31,24	28,00	8,00	22,57	11 months	93
Gender	0= Female 1= Male	0,49	-	0	-	-	-
Payments	Monthly	10,92	8,00	26,00	8,45	1	26,00
Payments Member	with contract (monthly)	10,97	8,00	26,00	8,46	1	26,00
Payments Member	without contract (monthly)	10,73	8,00	26,00	8,43	1	26,00
Member attendance frequencies	month	3,89	3	0	3,67	0	21,50
Member Profile	Type of service			Duo 19			
Contract	0 = without contract 1 = with contract	1	1	1			

Number of Hours Spent at the Gym	(not considering 0 of frequency)	6,86	5,30	1,41	5,90	0,05	84,24
Mode of the Time of Day for Training	0 = Without payment and frequency 1 = Morning (7am – 12am) 2 = Lunch (12am – 3pm) 3 = Afternoon (3pm – 5pm) 4 = Night (6pm – 10pm)	2	1	0	2		
Mode of the type Activity done	0 = without information 1 = individual 2 = group	1	0	1	0		

The analysis of the data (Table 3) offers a view of the members of a gym, for the overview we are focused on demographic characteristics, payment patterns, attendance frequencies, and training preferences. The average age of the members is approximately 31.24 years, with a median age of 28 years, suggesting that most gym-goers are in their thirties. However, the minimum recorded age is 11 months, and the maximum is 93 years, indicating a wide age range among members. The standard deviation of 22.57 years reinforces this heterogeneity. Interestingly, the mode is 8 years, pointing to a significant concentration of very young members because one of the most popular activities is swimming.

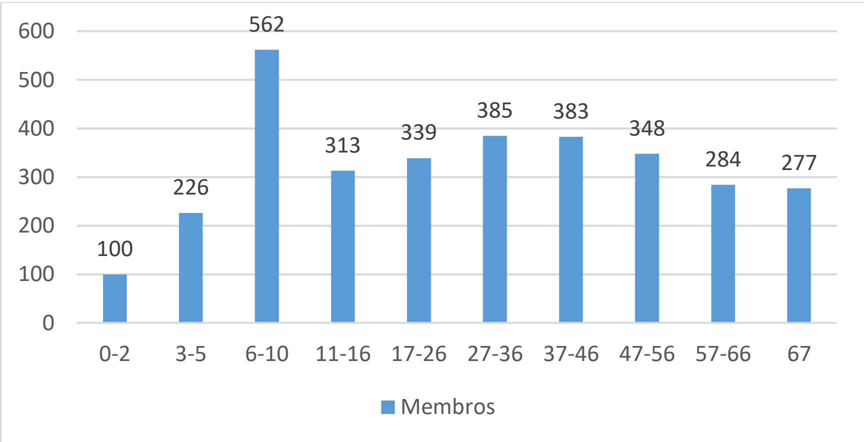


Figure 12 - Distribution of the members' age

Regarding gender, the distribution is quite balanced, with a mean of 0.4908, meaning that approximately 49% of the members are male and 51% are female. Despite this, the mode is 1, indicating a slight predominance of males when considering absolute frequency.

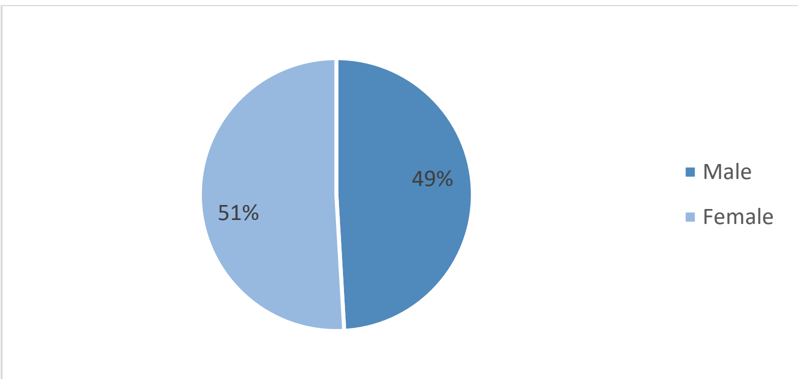


Figure 13 - Distribution of members' gender

The payment data reveals that the average number of payments made is 10.92 months, with a median of 8 months and a mode of 26 months, which is also the maximum number of payments recorded and the time of data collection. This suggests that while most members make around 8 months of payments, there is a significant group that makes the maximum possible number of payments. This pattern is repeated for both members with and without contracts, with slightly different means but similar modes and medians, indicating consistent payment behavior between the two groups. Attendance frequency at the gym has an average of 3.89 times per month, with a median of 3 times per month and a standard deviation of 3.67 months, indicating that many members attend the gym sporadically. The modal value of 0 visits suggests that a considerable number of members sign up but do not attend, which could be a point of interest for engagement and retention strategies. Regarding the time spent at the gym, the average is 6.86 hours per month, with a median of 5.30 hours and a standard deviation of 5.90 hours. This shows a large variation in the time members spend, from those who spend only a few minutes to those who dedicate more than 84 hours. Preferred training times show two distinct peaks: in the morning (7am to 12pm) and at lunchtime (12pm to 3pm). These times, respectively represented by the mode values of 1 and 2, are the most popular among members. Regarding the type of activity, the mode indicates that most members prefer individual activities over group activities.

Finally, most of the analyzed members have contracts, as indicated by the mean, median, and mode all being 1. This data points to an adherence policy that favors or requires contracts, reflecting a more formal commitment to the gym.

4.3. Machine Learning Model Results

This section provides a comprehensive evaluation of the predictive models applied to forecast gym member churn. This section is essential in identifying which model offers the best performance in predicting potential churners, thus enabling proactive retention strategies.

4.3.1. Comparison of the Different Models

Through the comparison of the different models, according to different metrics, it is possible to evaluate the one that is more adequate to the project (Table 4 & Table 5) . Logistic Regression shows a good balance between precision and recall with a high AUC of 0.9, indicating good distinguishing capability. Random Forest stands out with the best precision, negative predictive value, F1-Score, and an AUC of 0.947, making it highly reliable. The Decision Tree model demonstrates high accuracy and good balance, with an ideal AUC of 1 under optimal conditions but potential limitations in pessimistic scenarios with an AUC of 0.758. The K-nn model has good accuracy and precision but lower recall and F1-Score, reflected in an AUC of 0.894. Naïve Bayes exhibits high recall but low precision, leading to more false positives, and an AUC of 0.889. Deep Learning, on the other hand, provides an excellent balance across all metrics, with a high AUC of 0.917, making it ideal for comprehensive churn prediction.

Table 4 - Comparison of different models

	Accuracy	Precision	Recall	Specificity	Negative Predicted Value
Logistic Regression	98,53%	96,24%	73,19%	98,62%	99,85%
Random Forest	98,80%	100%	75,82%	98,53%	100,00%
Decision Tree	98,79%	99,71%	73,82%	98,75%	99,99%
K-nn	98,29%	96,56%	67,91%	98,35%	99,87%
Naïve Bayes	95,92%	56,78%	74,51%	98,65%	97,04%
Deep Learning	98,70%	99,70%	74,07%	98,66%	99,99%

Table 5 - Comparison of different models(cont.)

	AUC	AUC (optimistic)	AUC (pessimistic)	F1-Score
Logistic Regression	0,9	0,9	0,9	83,15%
Random Forest	0,947	0,947	0,947	86,25%
Decision Tree	0,5	1	0,758	86,14%
K-nn	0,894	0,986	0,802	79,74%
Naïve Bayes	0,889	0,889	0,889	64,45%
Deep Learning	0,917	0,917	0,917	84,99%

4.3.2. Logistic Regression Model

Logistic Regression shows a solid overall performance with an accuracy of 98.53% and an F1-Score of 83.15%. This model has a high precision of 96.24%, meaning it is good at making correct positive predictions, reducing the number of false positives. However, its recall is relatively low at 73.19%, indicating it might miss some actual positive cases (those who will churn). The specificity of 98.62% indicates it is highly effective at correctly identifying non-churners. This model's balance between precision and recall is good but not the best among the compared models. Its AUC is 0.9, which shows it has a good ability to distinguish between churn and non-churn cases.

Table 6 - Metric Performance - Logistic Regression Model

		Predicted Churn			
		Negative	Positive		
Actual Churn	Negative	8698	122	Specificity 98,62%	AUC 0,90
	Positive	13	333	Recall 73,19%	F1-Score 83,15%
		Negative Predictive Value 99,85%	Precision 96,24%	Accuracy 98,53%	

For Logistic Regression, the standard (Figure 14), optimistic, and pessimistic AUC (Index) values are all 0.9. This consistency indicates that the model's ability to distinguish between churn and non-churn cases is stable across different scenarios. Whether the conditions are favorable or not, Logistic Regression maintains a reliable performance level.

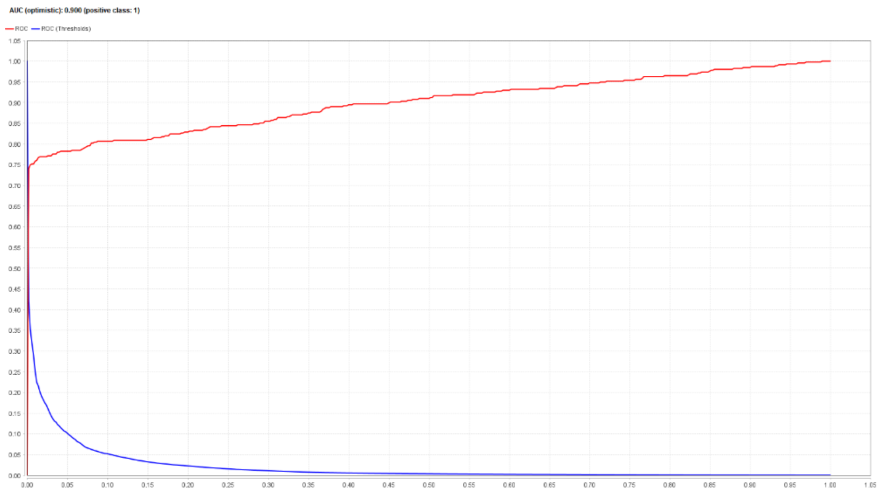


Figure 14 - AUC standard for Logistic Regression

4.3.3. Random Forest Model

Random Forest stands out with the highest precision (100%) and negative predictive value (100%), ensuring highly reliable predictions. It also has the highest F1-Score of 86.25% and an AUC of 0.947, indicating excellent performance in distinguishing between churn and non-churn. With an accuracy of 98.80% and a recall of 75.82%, this model effectively identifies most churn cases while maintaining no false positives. Specificity at 98.53% further supports its robust performance.

Table 7 - Metric Performance - Random Forest Model

		Predicted Churn			
		Negative	Positive		
Actual Churn	Negative	8711	130	Specificity 98,53%	AUC 0,947
	Positive	0	325	Recall 75,82%	F1-Score 86,25%
		Negative Predictive Value 100%	Precision 100%	Accuracy 98,80%	

Random Forest exhibits the highest and most consistent AUC across all scenarios (0.947). This suggests that the model is highly effective in distinguishing between churn and non-churn cases, regardless of the conditions. The consistency in AUC values further highlights its robustness and reliability in various situations.

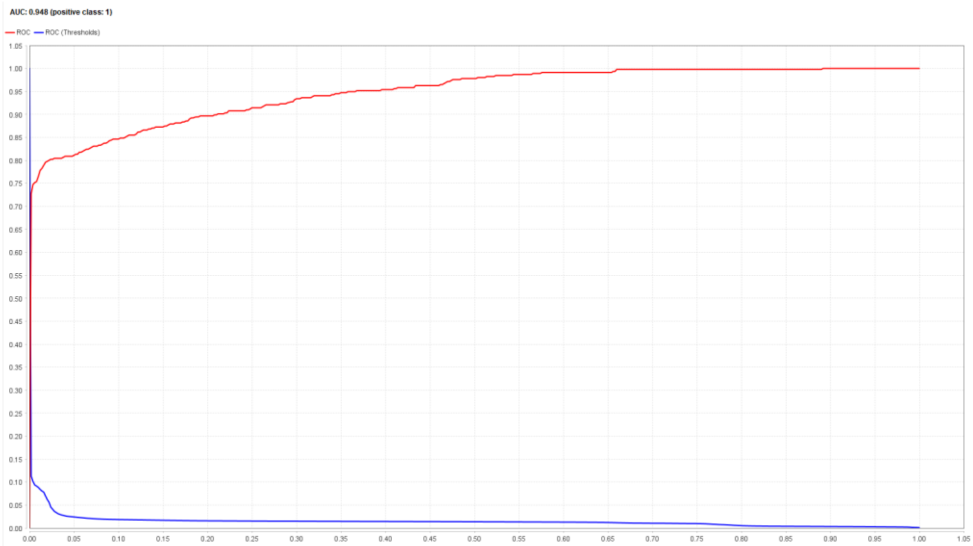


Figure 15 - AUC Model Random Forest

4.3.4. Decision Tree Model

The Decision Tree model demonstrates high accuracy (98.79%) and proper recall (75.82%), indicating it can correctly identify a significant portion of churners. Its precision is also high at 99.71%, resulting in few false positives. The F1-Score is 86.14%, reflecting a good balance between precision and recall. Specificity is 98.75%, showing its strong ability to correctly identify non-churners.

Table 8 - Metric Performance - Decision Tree Model

<table>
<tr><td></td><td></td><td colspan="2">Predicted Churn</td><td></td><td></td></tr>
<tr><td></td><td></td><td>Negative</td><td>Positive</td><td></td><td></td></tr>
<tr><td rowspan="2">Actual Churn</td><td>Negative</td><td>8710</td><td>110</td><td>Specificity
98,75%</td><td>AUC
0,947</td></tr>
<tr><td>Positive</td><td>1</td><td>345</td><td>Recall
73,82%</td><td>F1-Score
86,14%</td></tr>
<tr><td></td><td></td><td>Negative Predictive Value
99,99%</td><td>Precision
99,71%</td><td>Accuracy
98,79%</td><td></td></tr>
</table>

However, its AUC pessimistic of 0.758 suggests some limitations in more pessimistic scenarios. The AUC optimal value is 1, which is an ideal scenario showing the model can perfectly distinguish between classes under optimal conditions.

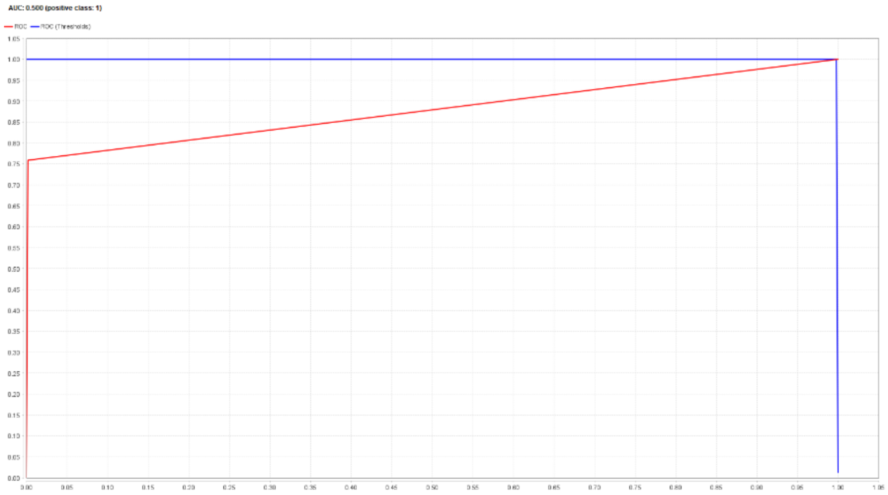


Figure 16 - AUC Decision Tree

4.3.5. K-nn Model

The K-nn model achieves good accuracy (98.29%) and precision (96.56%), indicating reliable positive predictions. However, it has the lowest F1-Score of 79.74% among the models, suggesting a less balanced performance. Its recall is 67.91%, meaning it may miss more churn cases compared to other models. The AUC of 0.894 suggests a reasonably good ability to distinguish between churners and non-churners, though not as high as some other models. The AUC pessimistic value is 0.986, indicating that in more pessimistic scenarios, the model's ability to distinguish between classes is still relatively strong. Specificity of 98.35% is reasonable, but not the best. Overall, K-nn is less balanced, with good accuracy but lower recall and F1-Score.

Table 9 - Metric Performance - K-nn Model

<table>
<tr><td></td><td></td><td colspan="2">Predicted Churn</td><td></td><td></td></tr>
<tr><td></td><td></td><td>Negative</td><td>Positive</td><td></td><td></td></tr>
<tr><td rowspan="2">Actual Churn</td><td>Negative</td><td>8700</td><td>110</td><td>Specificity
98,35%</td><td>AUC
0,894</td></tr>
<tr><td>Positive</td><td>11</td><td>146</td><td>Recall
67,91%</td><td>F1-Score
86,14%</td></tr>
<tr><td></td><td></td><td>Negative Predictive Value
99,87%</td><td>Precision
96,56%</td><td>Accuracy
98,29%</td><td></td></tr>
</table>

The AUC of 0.894 suggests a reasonably good ability to distinguish between churners and non-churners, though not as high as some other models. The AUC pessimistic value is 0.986, indicating that in more pessimistic scenarios, the model's ability to distinguish between classes is still relatively strong.

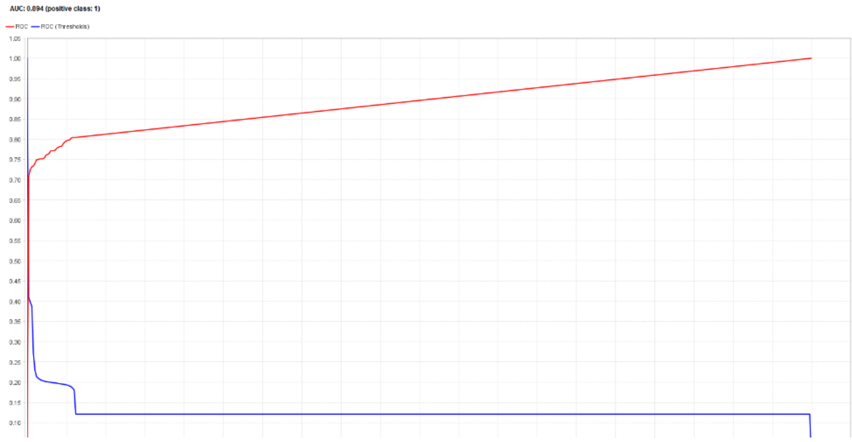


Figure 17 - AUC K-nn Model

4.3.6. Naïve Bayes Model

Naïve Bayes has the lowest accuracy (95.92%) and very low precision (56.78%), leading to many false positives. However, it has the highest recall (74.51%), indicating it can correctly identify most churners, though at the cost of precision. Its F1-Score is relatively low at 64.45%, showing a significant imbalance between precision and recall. The specificity is 98.65%, which is strong. The AUC of 0.889 indicates a fair ability to distinguish between churners and non-churners, but not as high as other models. The AUC pessimistic value is also 0.889, showing consistency in its performance under various conditions.

Table 10 - Metric Performance - Naive Bayes Model

<table>
<tr><td colspan="2"></td><td colspan="2">Predicted Churn</td><td colspan="2"></td></tr>
<tr><td colspan="2"></td><td>Negative</td><td>Positive</td><td colspan="2"></td></tr>
<tr><td rowspan="2">Actual Churn</td><td>Negative</td><td>8453</td><td>116</td><td>Specificity
98,65%</td><td>AUC
0,889</td></tr>
<tr><td>Positive</td><td>258</td><td>339</td><td>Recall
74,51%</td><td>F1-Score
64,45%</td></tr>
<tr><td colspan="2"></td><td>Negative Predictive Value
97,04%</td><td>Precision
56,78%</td><td>Accuracy
95,92%</td><td></td></tr>
</table>

The AUC of 0.889 indicates a fair ability to distinguish between churners and non-churners, but not as high as other models. The AUC pessimistic value is also 0.889, showing consistency in its performance under various conditions.

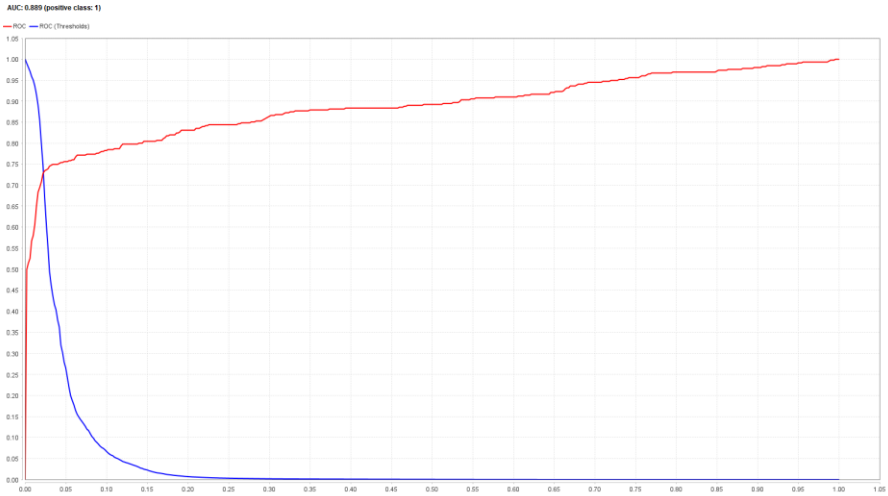


Figure 18 - AUC Naïve Bayes Model

4.3.7. Deep Learning Model

Table 11 - Metric Performance - Deep Learning Model

		Predicted Churn			
		Negative	Positive		
Actual Churn	Negative	8710	118	Specificity 98,66%	AUC 0,917
	Positive	1	337	Recall 74,07%	F1-Score 84,99%
		Negative Predictive Value 99,99%	Precision 99,70%	Accuracy 98,70%	

Deep Learning excels across all metrics, presenting an excellent balance with a high precision of 99.70% and recall of 74.07%. It achieves an F1-Score of 84.99% and an accuracy of 98.70%, indicating strong overall performance. The AUC is 0.917, showing a very good ability to distinguish between churn and non-churn cases. This high AUC value indicates that the model is effective in ranking predictions correctly. Specificity is 98.66%, ensuring effective identification of both churners and non-churners. Its negative predictive value is 99.99%, suggesting very reliable non-churn predictions. This model's balanced performance across all key metrics makes it highly suitable for churn prediction, combining high precision, recall, and overall accuracy.

The use of predictive machine learning models represents a state-of-the-art approach to addressing the challenge of churn in gyms. These models analyze historical data to identify patterns correlated with member churn, allowing gyms to proactively intervene. Among the analyzed models, logistic regression, decision trees, random forests, K-nn, Naïve Bayes Deep Learning were utilized to predict the likelihood of a member churning and to implement targeted retention strategies.

4.3.8. Main Conclusions and Actions

Among the tested models, the Random Forest stood out as the most effective. It presented the highest precision (100%) and negative predictive value (100%), ensuring highly reliable predictions. It also had the highest F1-Score of 86.25% and an AUC of 0.947, indicating excellent performance in distinguishing between churn and non-churn. With an accuracy of 98.80% and a recall of 75.82%, this model effectively identifies most churn cases while maintaining zero false positives. Specificity at 98.53% further supports its robust performance. Additionally, identifying key features for predicting gym member churn provides valuable insights for developing effective retention strategies. Relevant features typically include demographic data (age and gender), membership details (contract length, payment history), and behavioral metrics (attendance frequency, type of activities participated in).

The main features identified were (APPENDIX D: Random Forest Model results):

- **Demographic Information:** Age and gender influence churn rates, with different age groups and genders displaying distinct patterns in gym usage and attrition.
- **Payment History:** Regular and timely payments indicate a lower likelihood of churn, whereas lapses in payment can be an early indicator of potential churn.
- **Attendance Frequency:** A consistent reduction in gym visits often precedes membership cancellation. Tracking this metric allows for early identification of at-risk members.
- **Time of Day for Training:** The time of day for training can influence the retention of gym members for a few reasons. These include factors such as personal convenience, the number of members who attend, their energy levels, opportunities for social interaction and the availability of services or professionals.

Understanding these variables, it's possibly created retention strategies to address the specific reasons behind member dissatisfaction and churn. The gym should implement some measures to enhance client retention based on demographic data, payment history, frequency of attendance and the types of activities undertaken. The analysis of usage patterns by age and gender is a crucial step in understanding the behavior of different demographic groups. Based on the conclusions mentioned above, several structured actions are recommended to improve member retention at the gym. One of the main recommendations is the development of personalized programs tailored to different age groups and genders, as these factors directly influence members' behavior. For example, for young adults, dynamic programs focused on quick results, such as high-intensity interval training (HIIT) and friendly competitions, may be more effective in keeping them engaged. In contrast, for older individuals, the focus should be on wellness-centered activities, such as yoga and adapted pilates sessions that cater to their needs. Additionally, the introduction of women-only sessions could create a more comfortable environment, addressing the specific preferences of this demographic group.

Another key point is the implementation of a monitoring system that tracks members' payment histories. This system should be capable of alerting the support team when delays or missed payments are detected, which can be an early sign of churn. To prevent cancellations, it is advisable to send automatic reminders via email or SMS, along with offering incentives for those who pay on time, such as discounts or additional benefits, which could strengthen members' commitment to the gym.

Additionally, a decrease in visit frequency is one of the strongest indicators of churn, making it essential for the gym to develop an alert system for such cases. When a member begins to visit the gym less regularly, the system should send automatic notifications, suggesting personalized offers to encourage their return, such as a free training session or a consultation with a personal trainer. This type of personalized intervention can be crucial in reactivating members' interest. Promoting group activities has also proven to be an effective strategy for reducing churn, as these activities foster social interaction and create a sense of community among members. Expanding the offering of group classes, such as Zumba, spinning, or circuit training, can increase members' engagement. To complement this, offering discounts or promotions to those who regularly participate in these classes can further boost retention by encouraging greater adherence to these activities.

During off-peak hours, members tend to feel less engaged, which can increase the risk of churn. To counteract this, the gym should ensure that professionals are always available to provide support and guidance, regardless of the time of day. Moreover, organizing special events or exclusive classes during these less busy times could be advantageous in attracting more members and promoting greater interaction among them.

Creating a loyalty program is another measure that can reinforce members' commitment to the gym. This program could reward members for their consistency and loyalty, offering them discounts on services or products, access to exclusive events, or additional benefits for those who maintain a high frequency of visits. These rewards can not only encourage the continuation of memberships but also increase long-term member engagement. Regular and personalized communication with members is crucial for retention. The gym should adopt a customer relationship management (CRM) system that allows for sending personalized communications based on members' behavior. These communications could include class reminders, information on new offers or promotions, or even motivational messages to keep members engaged. At the same time, conducting periodic satisfaction surveys can help identify areas for improvement in the services provided, contributing to a better understanding of members' needs.

Finally, continuous follow-up and regular feedback collection are essential to ensuring the effectiveness of the retention strategies implemented. Regular surveys on members' satisfaction with classes, customer service, and the overall gym environment can provide valuable insights for the continuous improvement of services. Additionally, evaluating the impact of the actions implemented and adjusting strategies based on these results is crucial to ensuring effective and sustainable member retention.

5. Conclusion and Future Research

The main aim of this study was to explore and prevent the phenomenon of churn in gyms, using the Life Club gym as a case study. Machine learning techniques were used to identify the main factors that lead members to cancel their memberships. The dissertation was developed around four main pillars: the introduction (understanding the background, describing the problem, research gaps and characterizing the company), the literature review, the methodology used, and the results obtained from implementing the predictive models. The aim was to propose practical solutions to mitigate the problem of churn, based on the results generated by the models, in a gym management context. The development of this study allowed an in-depth analysis of various aspects of the problem, bringing relevant contributions to gym management.

The literature review allowed the problem to be contextualized within the fitness sector, which has grown considerably in recent years. However, alongside this growth, we have seen a worrying increase in churn rates, which has led gyms to invest more in customer retention strategies. The literature that supported this study showed that the cost of acquiring new members is significantly higher than the cost of retaining existing members, making retention a strategic priority for the sustainability of gyms. In the European and Portuguese context, it was possible to observe that although the fitness sector has expanded, especially with the emergence of low-cost gyms, the volatility of subscriptions has intensified, largely due to changes in consumer preferences and the emergence of new digital business models, such as online training platforms.

The literature review also highlighted the relevance of understanding customer behavior in the fitness sector. Studies from other sectors, such as telecommunications and banking, have provided a solid basis on how machine learning techniques can be used to predict cancellation behavior, but there is a scarcity of studies focused specifically on churn in gyms. The literature also highlighted the need to integrate behavioral, demographic and contextual data for a better understanding of the factors that influence churn. The importance of a personalized approach to member retention was also discussed, as opposed to homogeneous retention strategies, which often fail to meet the specific needs of each user.

The state of the art allowed us to understand the technological approaches that have been used to predict churn, with a particular focus on machine learning techniques. Through the analysis of existing studies, it was possible to identify those models such as logistic regression, decision trees, random forest and deep learning have been widely and successfully used in different sectors to predict churn. However, a clear gap was identified: the use of these models in the fitness sector is still limited, which opens a significant opportunity for the application of these techniques. The review of the state of the art showed that, while machine learning offers excellent predictive capabilities, its effectiveness depends on the quality of the data and the way in which the predictive variables are chosen and treated. Thus, the success of churn prediction is not only related to the choice of model, but also to the suitability of the data used.

In methodological terms, this study followed the CRISP-DM model (Cross-Industry Standard Process for Data Mining), a structured and effective approach for data mining projects. The application of CRISP-DM in this study involved several critical stages, from understanding the business and defining the churn problem in the specific context of gyms, to collecting, preparing and processing the data, to modeling and validating the predictive models. The data collection phase involved gathering detailed information on gym members, including frequency of use data, demographic characteristics, payment history and member feedback. To ensure the quality of the data, rigorous cleaning was carried out, removing outliers and filling in missing values.

The modeling involved the use of six algorithms: logistic regression, random forest, decision trees, K-nn, Naives Bayes and deep learning. Each of these algorithms was chosen based on the literature review, their ability to deal with different types of data and non-linear relationships between predictive variables. The validation process was conducted using techniques such as cross-validation and performance metrics such as accuracy, precision, recall and the area under the ROC curve. These indicators made it possible to robustly assess the predictive capacity of each model, ensuring that the results were reliable and applicable in the practical context of gyms.

The results obtained by applying the predictive models showed that the random forest was the best performing model, with high accuracy in identifying members who were at risk of canceling their memberships. The model was able to effectively capture the complexity of the data, identifying subtle interactions between variables that other models, such as logistic regression, could not capture as effectively. The main variables identified as predictive of churn included frequency of gym use, age of members, length of membership and the time of day they train. It was found that members who attended the gym less regularly, especially over extended periods, were more likely to cancel their memberships. Age also proved to be a significant factor, with younger members showing a higher churn rate, which can be explained by their more volatile routines and changes in interest over time.

In addition to the quantitative results, the qualitative analysis, carried out through interviews with former members, provided additional insights into the reasons for churn. Many former members mentioned lack of time, demotivation and the absence of personalized services as the main reasons for cancellation. This highlights the need for gyms to adopt a more customer-centric approach, providing personalized experiences that meet members' individual needs and expectations. Offering personalized workouts, creating incentive programs and providing regular follow-up services can help reduce churn by providing members with a more engaging and motivating experience.

Based on the results obtained, it is possible to prescribe a set of strategies for gyms that want to mitigate churn. Firstly, it is recommended to implement CRM systems that integrate predictive churn models, allowing gyms to identify members who are at risk of canceling in good time. These systems should be able to monitor member behavior, analyzing attendance patterns, interaction with services and satisfaction levels. When signs of disengagement are detected, gyms can act proactively by contacting members and offering incentives to resume attendance, such as personalized training sessions, discounts or exclusive promotions.

Another recommendation is to create loyalty programs that reward member consistency and involvement. These programs can include additional benefits for members who maintain high attendance over time, such as access to exclusive events, special classes or discounts on gym products and services. In addition, regular communication with members, whether through newsletters, personalized messages or invitations to events, can help keep members engaged and informed about new offers and training opportunities.

Finally, it is suggested that gyms should create a community atmosphere, fostering a sense of belonging among members. Organizing social events, fitness challenges and group activities can help strengthen the bonds between members and the gym, creating an atmosphere of support and motivation. These events not only encourage social interaction, but also help to create a more enjoyable and engaging experience, which can increase the likelihood that members will maintain their membership in the long term.

In short, this study has made a significant contribution to the understanding of churn, specifically as used in the study, providing valuable insights for management practice. Although the results are promising, there is still room for improvement and future research. It is suggested that future studies explore the inclusion of more detailed psychographic and behavioral variables, as well as evaluating the effectiveness of proposed retention strategies when implemented in the field. The combination of quantitative and qualitative approaches proved particularly effective in this study, and it is hoped that this methodology will continue to be applied in future research, with a view to further enhancing the understanding of this phenomenon and finding effective solutions for its mitigation.

References

Ahn, Jaehyun, Junsik Hwang, Doyoung Kim, Hyukgeun Choi, and Shin-sung Kang. 2020. "A Survey on Churn Analysis in Various Business Domains." *IEEE Access* 8: 220816-220839.

Albérico, Rosário, and Casaca Joaquim A. 2023. "Relationship Marketing and Customer Retention - A Systematic Literature Review." *Studies in Business and Economics* 18: 44 - 66.

Aldosary, May, and Abdalla Alrashdan. 2021. "Churn Prediction for Gym Members Using Artificial Neural Networks Assisted with The Psychological Concept of Habit Formation in The Fitness Industry." *Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management*.

Almohaimmeed, Bader M. A. 2019. "Pillars of customer retention: An empirical study on the influence of customer satisfaction, customer loyalty, customer profitability on customer retention." *Serbian Journal of Management* 14 (2): 421 - 435 https://doi.org/10.5937/sjm14-15517.

Artha, Bhenu, Inna Zahara, Bahri, and Niken Permata Sari. 2022. "Customer Retention: A Literature Review." *Social Science Studies* Vol.2 No. 1 030 - 045. https://doi.org/10.47153/sss21.2952022

Ascarza, Eva, Scott A Neslin, Oded Netzer, Zachery Anderson, Peter S Fader, Sunil Gupta, Bruce GS Hardie, Aurélie Lemmens, Barak Libai, and David Neal. 2018. "In pursuit of enhanced customer retention management: Review, key issues, and future directions." *Customer Needs and Solutions* 5: 65-81. https://doi.org/10.1007/s40547-017-0080-0.

Azzam, Diaa, Manar Hamed, Nora Kasiem, Yomna Eid, and Walaa Medhat. 2023. "Customer Churn Prediction Using Apriori Algorithm and Ensemble Learning." *2023 5th Novel Intelligent and Leading Emerging Sciences Conference (NILES)*: 377-381. https://doi.org/0.1109/NILES59815.2023.10296608.

Azzouz, Mahdia, Saïda Boukhedouma, and Zaia Alimazighi. 2022. "An approach with flexible choice of model for customer churn prediction and retention help." *2022 5th International Symposium on Informatics and its Applications (ISIA)*: 1-6.

Brinks, S. 2020. "Improving Customer Experience And Revenue Starts With The App Portfolio." *A Forrester Consulting Thought Leadership Paper*.

Chapman, Peter; Clinton, Julian; Kerber, Randy , Khabaza,Thomas, Reinartz, Thomas , Shearer, Colin and Wirth, Rüdiger 2000. *CRISP-DM 1.0: Step-by-step Data Mining Guide*. edited by SPSS.

Clavel San Emeterio, Iván, Eliseo Iglesias-Soler, Jorge García-Unanue, Leonor Gallardo, and Xián Mayo. 2020. "A model for predicting dropouts from physical activity interventions in leisure centres." *Sport Sciences for Health* 16: 465 - 472.

Colgate, Mark R, and Peter J Danaher. 2000. "Implementing a customer relationship strategy: The asymmetric impact of poor versus excellent execution." *Journal of the Academy of marketing Science* 28 (3): 375-387.

Della, Erick Sierra Santos Ribeiro Moreira. 2022. "Fidelização de clientes: como o marketing e as suas vertentes podem ser aplicadas no relacionamento entre consumidores e organizações." *Revista Científica Semana Acadêmica* EDIÇÃO 216. V.10 https://doi.org/10.35265/2236-6717-216-09159.

Eskiler, Ersin, and Furkan Safak. 2022. "Effect of Customer Experience Quality on Loyalty in Fitness Services." *Physical Culture and Sport. Studies and Research* 94: 21 - 34.

Gallo, Amy. 2014. "The Value of Keeping the Right Customers." *Harvard Business Review*.

Garcia-Feo, Matt, and Mónica Tatiana Rodríguez-Rodríguez. 2020. "Estrategias de promoción en la prestación del servicio para el Gimnasio Bogotá Fitness Club."

Ghadiri, MohammadSadegh, Ali Fahiminezhad, Hadi Bagheri, and Seyed Mostafa Tayebi. 2021. "Market-Retention in Fitness Centers: Identifying and Ranking the Factors Affecting Customer Retention." *Journal of New Studies in Sport Management* 2 (1): 95-110. https://doi.org/10.22103/jnssm.2020.16749.1012. https://jnssm.uk.ac.ir/article_2797_0265deec18ec68e7a87e94ee68ec46ef.pdf.

Höppner, Sebastiaan, Eugen Stripling, Bart Baesens, Seppe vanden Broucke, and Tim Verdonck. 2020. "Profit driven decision trees for churn prediction." *European journal of operational research* 284 (3): 920-933.

Jang, Won yong, and Sang-Yeol Baek. 2024. "The relative importance of servicescape in fitness center for facility improvement." *Heliyon* 10.

Kapoor, Gaurav, Priyanshi Chauhan, Gurjant Singh, Nitesh Malhotra, and Aksh Chahal. 2022. "Physical Activity for Health and Fitness: Past, Present and Future." *Journal of Lifestyle Medicine* 12: 9 - 14.

Kotler, P., and K.L. Keller. 2021. *Marketing Management, Global Edition*. Pearson Education.

Koutsoubakis, Dimitri. 2020. " Testing the Effectiveness of a Gym Orientation Program." *Journal of Human Sciences* 17, no. no. 2: 469-84. https://doi.org/https://doi.org/10.14687/jhs.v17i2.5980.

Kwon, Jaeyoon, and Sang Back Nam. 2022. "Analysis of Customer Journey Map of Fitness Center Centered on Touch Point." *Journal of Sport and Leisure Studies* 90: 87 - 100. https://doi.org/ 10.51979/KSSLS.2022.10.90.87.

Lalwani, Praveen, Manas Kumar Mishra, Jasroop Singh Chadha, and Pratyush Sethi. 2022. "Customer churn prediction system: a machine learning approach." *Computing* 104 (2): 271-294. https://doi.org/10.1007/s00607-021-00908-y. https://doi.org/10.1007/s00607-021-00908-y.

Liang, Zhiqing. 2023. "Predict Customer Churn based on Machine Learning Algorithms." *Highlights in Business, Economics and Management* 10: 270 - 275. https://doi.org/10.54097/hbem.v10i.8051.

Liusnea, Cristian Stefan. 2022. "Relationship between Fitness and Healthy Lifestyle." *Balneo and PRM Research Journal* 13(4): 521-527. https://doi.org/10.12680/balneo.2022.521.

Macon, R. Scott. 2020. "Customer Retention Strategies in the Fitness Industry."

Mujayana, Marya. 2020. "Client Retention Strategies and customer loyalty development." *Journal of Applied Management and Business (JAMB)* 1(1): 16-19. https://doi.org/10.37802/jamb.v1i1.60.

Muñoz, Joaquin Moreno, Miguel Ramirez Osorio, Ciro Rodriguez, Diego Rodriguez, Pedro Lezama, and Yuri Pomachagua. 2022. "Gyms Customer Loyalty Using the Logistic Regression Algorithm Technique." 14th International Conference on Computational Intelligence and Communication Networks (CICN).

Nalluri, Venkateswarlu, and Long-Sheng Chen. 2022. "Exploring the Relationship Among Experience Marketing, Customer Loyalty on Purchase Intention - A Case Study of Banking Sector." *2022 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*: 1149-1153.

Nursanti, Metasari Dian, and Tomoliyus Tomoliyus. 2021. "Factors Affecting Satisfaction and Loyalty of Health and Fitness Club Customers." *KnE Social Sciences*. https://doi.org/10.18502/kss.v6i2.10002.

Orina, Denis Ong'Uti, Richard Rimiru, and Waweru Mwangi. 2023. "A Comparative Study of Predictive Data Mining Techniques for Customer Churn in the Banking Industry." 2023 Intelligent Methods, Systems, and Applications (IMSA).

Pedragosa, V., Cardadeiro, E., & Santos, A. 2023. *Barómetro do fitness em Portugal 2022.* AGAP.

Pedragosa;, Vera, and Allan Ferreira;. 2024 2023. *Barómetro do Fitness.* Portugal Ativo.

Rahman, M D Sayedur, M D Sahrial Alam, and Md. Mofazzal Hosen. 2022. "To Predict Customer Churn By Using Different Algorithms." *2022 International Conference on Decision Aid Sciences and Applications (DASA)*: 601-604. https://doi.org/10.1109/DASA54658.2022.9765155.

Rutgers, Herman; , Karsten; Hollasch, Stefan; Ludwig, Steffen; Gausselmann, Christian; Rump, and Jakob Papenbrock. 2019. *European Health & Fitness Market Report 2019.* (EuropeActive).

---. 2020. *European Health & Fitness Market Report 2020.* (EuropeActive).

---. 2021. *European Health & Fitness Market Report 2021.* (EuropeActive).

---. 2022. *European Health & Fitness Market Report 2022.* (EuropeActive).

---. 2023. *European Health & Fitness Market Report 2023.* (EuropeActive).

---. 2024. *European Health & Fitness Market Report 2024.* (EuropeActive).

Semrl, Jas, and Alexandru Matei. 2017. "Churn prediction model for effective gym customer retention." 2017 International Conference on Behavioral, Economic, Socio-cultural Computing (BESC).

Sobreiro, Pedro, Pedro Guedes-Carvalho, Abel Santos, Paulo Pinheiro, and Celina Gonçalves. 2021. "Predicting Fitness Centre Dropout." *International Journal of Environmental Research and Public Health* 18.

Sobreiro, Pedro, Domingos Martinho, José Manuel García Alonso, and Javier Berrocal. 2022. "A SLR on Customer Dropout Prediction." *IEEE Access* 10: 14529-14547. https://doi.org/10.3390/ijerph181910465.

Soumi, De, Prabu P, and Joy Paulose. 2021. "Effective ML Techniques to Predict Customer Churn." *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*: 895-902. https://doi.org/0.1109/ICIRCA51532.2021.9544785.

Srinivasan, R., D. N. Rajeswari, and G. Elangovan. 2023. "Customer Churn Prediction Using Machine Learning Approaches." *2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF)*: 1-6. https://doi.org/10.1109/ICECONF57129.2023.10083813.

Sundararajan, Abhijit, and Kemal Gursoy. 2020. "Telecom Customer Churn Prediction." *International Journal for Research in Applied Science and Engineering Technology*.

Tokarski, Walter, Antonis Tosounidis, and Georgios F. Zarotis. 2023. "Need for Health and Prevention through Fitness Sports." *European Journal of Sport Sciences* 2(4): 34-42. https://doi.org/10.24018/ejsport.2023.2.4.91.

Tsitskari, Efi, and Alexios Batrakoulis. 2021. "Investigating the marketing strategies of the health and fitness sector in Greece; the managers' perspective (Investigar las estrategias de marketing del sector de la salud y el fitness en Grecia; la perspectiva de los gerentes)." *Retos* 43: 264-273. https://doi.org/10.47197/retos.v43i0.88880.

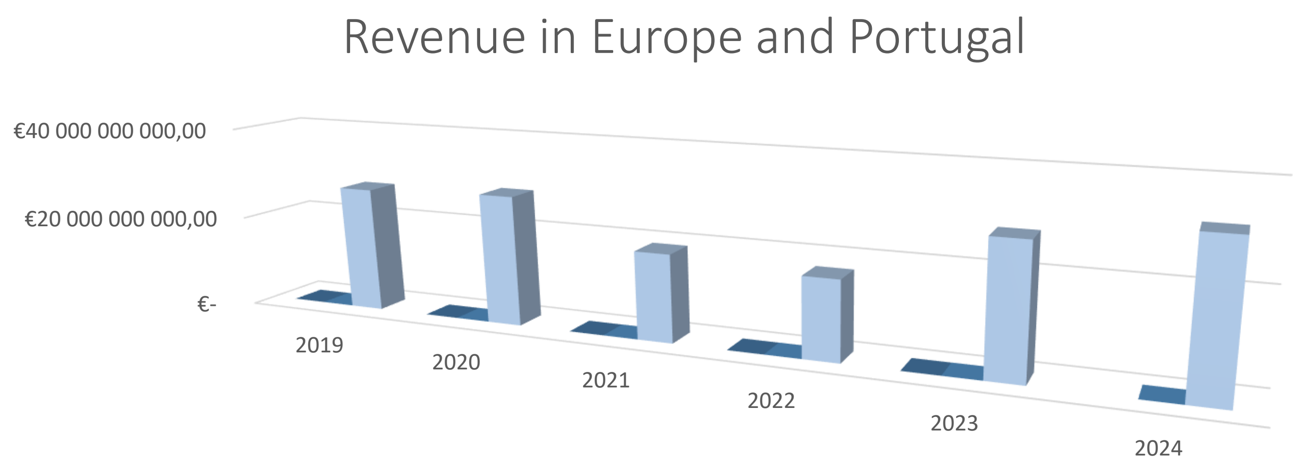
Ullah, Irfan, Basit Raza, Ahmad Kamran Malik, Muhammad Imran, Saif ul Islam, and Sung Won Kim. 2019. "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector." *IEEE Access* 7: 60134-60149.

Vafeiadis, Thanasis, Konstantinos I Diamantaras, George Sarigiannidis, and K Ch Chatzisavvas. 2015. "A comparison of machine learning techniques for customer churn prediction." *Simulation Modelling Practice and Theory* 55: 1-9.

Wang, Fong-Jia, Chia-Huei Hsiao, and Tao-Tien Hsiung. 2022. "Marketing strategies of the female-only gym industry: A case-based industry perspective." *Frontiers in Psychology* 13. https://doi.org/10.3389/fpsyg.2022.928882.

WHO. 2018. *Global action plan on physical activity 2018–2030: more active people for a healthier world.* World Health Organization.

APPENDIX A: Revenue in Europe and Portugal (2020 – 2024)



	2019	2020	2021	2022	2023	2024
Revenue in Portugal	€289 371,00	€167 408,00	€165 364,00	229 218,00 €	270 477,24 €	
Revenue of top 20 operators in europe (Projections for the year)	€-	€17 200 000,00	€12 300 000,00	13100 000,00 €	14800 000,00 €	17400 000,00 €
Total revenue in europe (Projections for the year)	€27 200 000 000,00	€28 200 000 000,00	€18 900 000 000,00	17100000 000,00 €	28000000 000,00 €	31800000 000,00 €

Figure 19 - Revenue from gyms in Europe and Portugal (adapted: (Rutgers et al. 2021))

APPENDIX B: semi-structured interview with gym former members

Introduction Welcome and Presentation of the Study: Good morning. [My introduction]. Thank you in advance for your willingness to take part in this interview, which is part of my master's thesis. The aim is to understand the reasons why members leave the gym. I guarantee the total anonymity of your answers, which will be used exclusively for academic purposes.

Informed Consent: You confirm that you understand the objectives of this study and that you consent to take part in this interview, knowing that your answers will be treated anonymously and confidentially?

Enrolment history: 1. Could you indicate how long you were a member of the gym and how often you attended?

General assessment: 2. How would you rate your overall satisfaction with the gym over the time you've been a member? (Scale from 0 to 10)

Services used: 3. Which services or facilities did you use the most while you were at the gym?

Changes in quality: 4. Did you notice any changes in the quality of the services, equipment or atmosphere of the gym that influenced your decision?

Communication and Support: 5. How do you rate the gym's communication and support? Are there any areas that you think could be improved?

Positive points: 6. Are there any aspects or services of the gym that you had a particularly positive experience with and would like to emphasize?

Reasons for Quitting: 7. What was the main reason you gave up your gym membership?

Decisive moment: 8. Was there a specific event or situation that precipitated your decision to leave the gym? Can you describe it?

Suggestions for improvement: 9. Is there anything the gym could have done differently that would have made you reconsider your decision?

Final Comments: Are there any other observations or suggestions you'd like to share that you think are relevant to this study?

Acknowledgement and Closing: Thank you very much for your participation and the valuable information you have provided me with. Your contribution is extremely important for the development of my master's project. Thank you very much.

APPENDIX C: answers to the semi-structured interview with gym members who cancelled

Interview number	**1**

Gender	M
Age	51

Attendance	12	7	1	3	1	1																				
Question 1	6 meses - 2x semana																									
Question 2	5/6																									
Question 3	Piscina																									
Question 4	não, apenas pela questão da bebe																									
Question 5	Boa - 8																									
Question 6	interação dos professores																									
Question 7	facto de a menina não poder frequentar																									
Question 8	indicação médica																									
Question 9	não se prende com o ginásio																									

Interview number	**2**

Gender	M
Age	34

Attendance	5	4	1	2	3	0	0	1	5	4	1	2														
Question 1	11 meses; 1 a 2x semana																									
Question 2	10																									
Question 3	cardio musculação, piscina, sauna																									
Question 4	não																									
Question 5	10																									
Question 6	Mudança de local de residência																									
Question 7	Atenção dos colaboradores																									
Question 8	Mudança de residência																									
Question 9	não																									

Interview number	**3**

Gender	F
Age	43

Attendance	2	0	3	3	1	4	3	2	1	2	0	0	0	0	2	0	0	0	0				

Question 1	14 meses - 2x
Question 2	9
Question 3	piscina, pilates
Question 4	Não
Question 5	10
Question 6	Logística – vive distante pagamento de parque boas instalações deixaram de dar toalha
Question 7	Logística, estacionamento
Question 8	quando começou a perceber que não vinha, devido logística
Question 9	logística, estacionamento

Interview number	**4**

Gender	M
Age	20

Attendance	10	13	11	12	17	7	6	3	3	6	7	6	0										

Question 1	13 meses - livre
Question 2	5
Question 3	Cardio e musculação
Question 4	Desgaste de algumas máquinas
Question 5	7, questão da manutenção
Question 6	desinteresse
Question 7	desinteresse
Question 8	não
Question 9	-

Interview number	**5**

Gender	F
Age	19

Attendance	10	13	11	12	17	7	6	3	3	6	7	6	0												

Question 1	10 meses - livre
Question 2	10
Question 3	Cardio e musculação
Question 4	Nada a considerar
Question 5	9
Question 6	Familiar, professores ajudavam, amigáveis
Question 7	motivos externos ao ginásio trabalho e faculdade
Question 8	não
Question 9	-

Interview number	**6**

Gender	F
Age	36

Attendance	9	9	9	8	7	1													

Question 1	6 meses – 2x
Question 2	10
Question 3	Piscina, hidroginástica, hidrobike
Question 4	Não
Question 5	8, nenhum comentário
Question 6	gravidez
Question 7	gravidez
Question 8	não
Question 9	-

Interview number	**7**

Gender	M
Age	55

Attendance	3	3	9	4	0																				
Question 1	Livre 3x																								
Question 2	8																								
Question 3	Piscina, h2bola, natação, cardio e musculação																								
Question 4	Não																								
Question 5	8, nenhum comentário																								
Question 6	As pessoas, os profissionais																								
Question 7	Orçamento familiar																								
Question 8	Necessidade de ajuste																								
Question 9	neste caso não. dinâmica da aula de natação.																								

Interview number	**8**

Gender	F
Age	56

Attendance	11	10	1	1	10	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0
Question 1	21 meses – 2x																			
Question 2	8																			
Question 3	Cardio e musculação																			
Question 4	nao houve alteraçao acompanham por parte dos profissionais																			
Question 5	8, acompanhamento																			
Question 6	Falta de tempo, custo																			
Question 7	Fazer contas e ver que não estava a utlizar																			
Question 8	Acompanhamento mais personalizado																			
Question 9	-																			

Interview number	**9**

Gender	F
Age	24

Attendance	2	1	0	0	0	0																			
Question 1	7 meses – 2x																								
Question 2	9/10																								
Question 3	Piscina, cardio e musculação, aulas de grupo																								
Question 4	não																								
Question 5	8																								
Question 6	Staff simpático																								
Question 7	falta de disponibilidade																								
Question 8	Verificando a frequencia																								
Question 9	não																								

Interview number	**10**

Gender	F
Age	30

Attendance	2	0	1	0	0	0																			
Question 1	7 meses – 2x																								
Question 2	10																								
Question 3	Piscina, aulas de grupo																								
Question 4	não																								
Question 5	8																								
Question 6	receção																								
Question 7	falta de motivação																								
Question 8	não																								
Question 9	não acha que houvesse.																								

APPENDIX D: Random Forest Model results

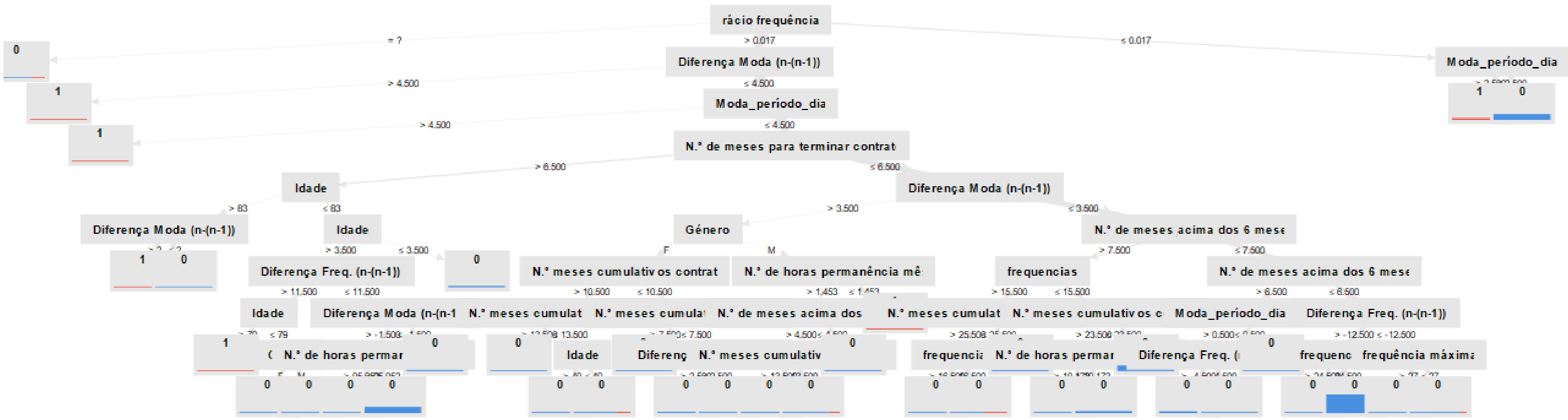


Figure 20 - Random Forest Model results