Unlocking the Valuation Puzzle: Harnessing the Power of Interpretable Machine Learning for Enhanced Company Valuation

Luís Baltazar Blanquet

Master's Dissertation

FEUP Supervisor: Miguel Alves Pereira
Company Supervisor: Stefan Petrov



Master in Industrial Engineering and Management

2023-06-26

Abstract

Valuing private enterprises, especially early-stage firms, has posed challenges, resulting in an average overvaluation of 50%. This has adversely affected the venture capital market, diminishing investor returns and eroding industry trust. Artificial intelligence offers a triple benefit by improving valuation accuracy, enabling automation, and ensuring impartiality. This thesis addresses these issues by introducing robust and generalizable models capable of handling the complex valuation process. It also proposes an innovative framework that closely resembles the real-world valuation process. The study fills a research gap by examining significant determinants affecting firm value and exploring the unique contributions of valuation methodologies at different company maturity stages. Furthermore, it advocates for adopting unconventional valuation methods to enhance current practices. Four machine learning frameworks are developed, including a novel one that outperforms similar neural network frameworks by closely emulating the entire valuation process. Notably, the top-performing framework achieves a Mean Absolute Percentage Error of 34.90%, representing a significant 32% improvement over the industry average. The evaluation of valuation methodologies reveals the Total Cost of Ownership as the most critical factor, with nine out of twelve methodologies showing statistical significance. Financial metrics contribute significantly (83%) to the model's performance, while non-financial factors complement and enhance the overall valuation process. Moreover, the study unveils the nuanced effects of individual characteristics on valuation structure, considering specific contexts and methodologies. Financial metrics demonstrate greater relevance in predicting firm value as the firm's value increases. Ultimately, this work harnesses the full potential of data mining techniques, offering a reliable and comprehensive approach to valuing private enterprises in the venture capital landscape. Future research can explore similar challenges in emerging countries, applying these techniques and utilizing primary data sources to gain a deeper understanding of the determinants of company valuation.

Keywords: Startup Valuation; Machine Learning; Venture Capital; Artificial Intelligence; Enterprise Value; Regression

Resumo

A avaliação de empresas privadas, especialmente de startups, é complexa e desafiante, o que resulta numa sobreavaliação média destas em 50%. Isto afeta negativamente o mercado de capital de risco, diminuindo os retornos dos investidores e criando desconfiança na indústria. A inteligência artificial oferece um triplo benefício, melhorando a precisão da avaliação, permitindo a automação e garantindo a imparcialidade. Esta tese aborda estas questões através da introdução de modelos robustos e generalizáveis capazes de lidar com o complexo processo de avaliação de empresas. Propõe também um modelo inovador que se assemelha ao processo de avaliação tradicional. O estudo preenche uma lacuna no campo de investigação ao examinar os determinantes significativos que afectam o valor da empresa e ao explorar os contributos únicos das metodologias de avaliação em diferentes fases de maturidade da empresa. Além disso, defende a adoção de métodos de avaliação não convencionais para melhorar as práticas atuais. São desenvolvidas quatro modelos de ML, incluindo um novo, que se aproxima do processo de avaliação, que supera modelos de redes neurais comparáveis. O modelo com melhor desempenho atinge um erro percentual absoluto médio de 34,90%, representando uma melhoria significativa de 32% em relação à média da indústria. A avaliação das metodologias de avaliação revela o Total Cost of Ownership como o fator mais importante, com nove das doze metodologias a mostrarem significância estatística. As métricas financeiras contribuem significativamente (83%) para o desempenho do modelo, enquanto os fatores não financeiros complementam e melhoram o processo de avaliação global. Além disso, o estudo revela os efeitos heterogéneos das características individuais na estrutura de avaliação, considerando contextos e metodologias específicos. As métricas financeiras demonstram maior relevância na previsão do valor da empresa à medida que o valor da empresa aumenta. Em última análise, este trabalho aproveita todo o potencial das técnicas de DM, oferecendo uma abordagem fiável e abrangente para a avaliação de empresas privadas no panorama do capital de risco. A investigação futura pode explorar desafios semelhantes em países emergentes, aplicando estas técnicas e utilizando fontes de dados primárias para obter uma compreensão mais profunda dos fatores determinantes para a avaliação de empresas.

Palavras-chave: Avaliação de startups; Aprendizagem automática; Capital de risco; Inteligência artificial; Valor da empresa; Regressão

Acknowledgements

The development of this thesis has been a great test of my ability to self-organize, work independently, and push forward with an idea that lacked a defined structure. It was an opportunity to prove that conformity is unnecessary to achieve success. With just an idea in mind and the support of the right people, one can pursue their passions and make a meaningful impact by doing what they do best.

First and foremost, I want to extend my heartfelt thanks to Miguel Alves Pereira. Despite not being directly involved in the fields of study, you have always been there, eager to learn and lend a helping hand. Even when I bombarded you with daily 15-minute updates and existential doubts, your advice and support meant the world to me. You never stopped believing in my capabilities; I am genuinely grateful.

I cannot express enough gratitude to Stefan Petrov. Your genuine interest in my thesis and dedication to supporting me on the company side has been instrumental in its success. You went above and beyond, connecting me with the right people, providing essential documentation, and helping me refine my vision. I am so grateful for your continuous support, even when time was limited.

Thank you to Teacher José Luís Moura Borges for your support and technical guidance, even though you were not directly involved in the project.

To my amazing friends, I want to express my deepest gratitude. Your presence throughout this journey has meant more to me than words can convey. Your funny personality and unwavering support have made my challenges seem small and have reminded me of the importance of pursuing my dreams. Thank you for always being there and believing in me.

Maria João, I cannot express my gratitude for your support. You have been there through all my existential doubts, offering emotional and technical guidance. Whether discussing ideas or simply venting, your presence and understanding have been a source of strength for me. I am grateful to have you by my side.

And to my incredible family, with a special note to my parents, I owe everything to you. You have taught me the values of dedication and effort, which have been the foundation of my journey. Your unwavering support and the knowledge that you would catch me if I fell have given me the courage to aim high and pursue my passions. Thank you for always believing in me.

To my sister, you have been my rock. Our conversations may not always be extensive, but our understanding is beyond words. Your acceptance and encouragement to be true to me mean the world. Thank you for reminding me to find happiness in everything I do.

To everyone mentioned above and those who have supported me in ways big and small, I extend my warmest and most personal gratitude. Your belief in me, guidance, and unwavering support have shaped this thesis and my personal growth. Thank you for being an integral part of my journey and for impacting my life.

"Muchas gracias afición, esto es para vosotros: SIUUU"

Cristiano Ronaldo (2018)

Contents

1 Introduction **1**
- 1.1 Problem description . . . 1
- 1.2 Context . . . 2
- 1.3 Research objectives . . . 3
- 1.4 Outline . . . 4

2 Theoretical framework **5**
- 2.1 Company valuation . . . 5
  - 2.1.1 Financial statements-based company valuation methodologies . . . 5
  - 2.1.2 Implicit intangible assets-based valuation methodologies . . . 13
- 2.2 Data mining applied to the industry . . . 19
  - 2.2.1 Applications in the financial industry . . . 19
  - 2.2.2 Applications to firm valuation . . . 20
- 2.3 Knowledge gap . . . 22

3 Methodological framework **25**
- 3.1 Machine learning process . . . 26
  - 3.1.1 Data understanding . . . 27
  - 3.1.2 Data preparation . . . 30
  - 3.1.3 Modelling . . . 37
  - 3.1.4 Evaluation . . . 38
- 3.2 Proposed valuation models . . . 39
  - 3.2.1 Multilayered integrated model with industry restrictions . . . 40
  - 3.2.2 Comparison of three alternative frameworks . . . 44

4 Results and discussion **45**
- 4.1 Framework validation . . . 45
- 4.2 Valuation methodology importance analysis . . . 48
- 4.3 Company characteristics importance analysis . . . 50
- 4.4 Valuation methodology and predictive factors importance relation analysis . . . 54

5 Conclusion **57**
- 5.1 Critical analysis . . . 58
- 5.2 Future research . . . 60

References **61**

A Results plots **69**

Acronyms and Symbols

AI	Artificial Intelligence
ANN	Artificial Neural Networks
APV	Adjusted Present Value
BAIT	Business Acquisitions and Investments Test
BSC	Balanced Scorecard
BV	Book Value
CF	Cash Flow
CFROI	Cash Flow Return On Investment
CPI	Customer Price Index
CRISP-DM	Cross-Industry Standard Process for Data Mining
CIV	Composite Indicator of Value
DCCF	Discounted Cash Flow
DM	Data Mining
DSA	Data Staging Area
EBIT	Earnings Before Interest and Taxes
EBITDA	Earnings Before Interest, Taxes, Depreciation, and Amortization
EPS	Earnings Per Share
EVA	Economic Value Added
EV/EBITDA	Enterprise Value-to-Earnings Before Interest, Taxes, Depreciation, and Amortization
EV/S	Enterprise Value-to-Sales
FCF	Free Cash Flow
GAAP	Generally Accepted Accounting Principles
GDP	Gross Domestic Product
GPR	Gaussian Process Regression
HRC	Human Resource Costing
IC	Intellectual Capital
IPO	Initial Public Offering
IV	Intrinsic Value
IVM	Inclusive Valuation Methodology
LIME	Local Interpretable Model-Agnostic Explanations
KNN	K-Nearest Neighbours
KPI	Key Performance Indicator
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Mean Square Error
ML	Machine Learning
MLP	Multilayer Perceptron
NN	Neural Networks

OAMV	Owner Assigned Market Value
P/CF	Price-to-Cash Flow
P/E	Price-to-Earnings
P/EBITDA	Price-to-Earnings Before Interest, Taxes, Depreciation, and Amortization
P/S	Price-to-Sales
RIM	Residual Income Model
RIR	Real Interest Rate
RMSE	Root Mean Square Error
RR	Ridge Regression
SHAP	SHapley Additive exPlanations
SVM	Support Vector Machines
SVR	Support Vector Regression
VC	Venture Capital
VAIC	Value Added Intellectual Coefficient
WACC	Weighted Average Cost of Capital
XGB	XGBoost

List of Figures

2.1 Diagram of the value distinction tree for intellectual capital. Source: Roos, Roos, Dragonetti, and Edvinsson (1998) . . . 14
2.2 Diagram of relationships in contributions of the study - interconnectedness between firm value, valuation methodologies and company characteristics . . . 23

3.1 CRISP-DM process diagram . . . 25
3.2 Data mining methodology and frameworks interconnectedness diagram (interplay between CRISP-DM and four generic frameworks) . . . 26
3.3 Data warehouse architecture with data staging area diagram . . . 27
3.4 Data warehouse data model before JSON files transformation (Star model) . . . 28
3.5 Final data warehouse data model (Snowflake model) . . . 29
3.6 Data preparation pipeline and its effect on the dataset composition . . . 30
3.7 Distribution of missing values as a % of each characteristic and of each company 32
3.8 Sensitivity matrix of the thresholds for eliminating missing values . . . 32
3.9 Sensitivity matrix of the thresholds for data augmentation . . . 34
3.10 Distribution of intermediate valuations compared to real valuation line chart . . . 36
3.11 Flowchart of the multilayer integrated model with industry restrictions . . . 42
3.12 Diagram of the multilayer integrated neural network model . . . 43

4.1 Variable importance plot of significant valuation methodologies (%IncMSE) . . . 49
4.2 Correlation between valuation methodologies importance and real valuation . . . 50
4.3 Category importance plot of significant factors across categories (%IncMSE) . . . 52
4.4 Correlation between most important predictive factors importance and real valuation 53
4.5 Correlation between most correlated predictive factors importance and real valuation 54

A.1 Regression model with intermediate valuations results distribution . . . 69
A.2 Regression model with company characteristics results distribution . . . 69
A.3 Enhanced regression model results distribution . . . 70
A.4 Frameworks' results distribution . . . 70

List of Tables

2.1 Compilation of company valuation methods based on financial statements. Adapted from Vayas-Ortega, Soguero Ruiz, Rojo-Álvarez, and Gimeno (2020) 13
2.2 Intellectual capital measurement methods. Adapted from Sitar (2005) 16
2.3 Overview of ML Applications in Firm Valuations: Research Summary 21

3.1 Summary of relevant company characteristics used in comparable literature . . . 29
3.2 Company characteristics categories and their representativity in the dataset after data collection 31
3.3 Company characteristics categories and their representativity in the dataset after data preparation 35
3.4 Statistical summary of the intermediate valuations dataset after data preparation . 36
3.5 Summary of the ML algorithms and their usage in comparable studies 37
3.6 Definition and critical analysis of regression algorithms 38
3.7 Summary of the frameworks tested by algorithm, inputs and usage 40

4.1 Regression with intermediate valuations framework results 46
4.2 Regression with company characteristics framework results 46
4.3 Enhanced regression model framework results 47
4.4 Comparison of the frameworks results 47
4.5 Summary of the valuation methodologies importance 48
4.6 Summary of the predictive factors importance 51
4.7 Summary of company characteristics categories' significance and representativity 51
4.8 Correlation matrix of categories importance with valuation methodologies importance (Mean absolute correlation) 55

Chapter 1

Introduction

Valuing companies is a crucial task in the financial industry, aiding investors in making informed decisions. However, the complexity and numerous variables involved in valuation make it challenging and prone to inaccuracy. Notably, venture capitalists have been found to overvalue companies with high diversifiable risk by an average of 50% (Ewens, Jones, & Rhodes-Kropf, 2013).

There is a knowledge gap in integrating valuation methodologies into the day-to-day operations of the VC industry. It is crucial to understand when to use each methodology according to company specificities and its overall relevance. Bridging this gap will provide investors with better information, empowering them to make informed decisions and increasing the success and sustainability of their investments (Silva, Fantin, Fukui, & Juca, 2021).

The current project aims to address this issue by proposing an accurate and tailored model for company valuation. This model enables the study of the most relevant valuation methodologies, company characteristics, and their interrelationships.

Section 1.2 details the scope of the project, Section 1.2 provides a brief description of the project's background and context in which it is integrated, Section 1.3 outlines the goals and phases of the project, and Section 1.4 presents the structure of the work.

1.1 Problem description

Valuing a company is a complex task that involves considering multiple factors impacting its value. However, this process often leads to inaccuracies, resulting in inflated valuations and potential investor losses. The resulting disparity has diminished investors' trust in the venture capital (VC) market, leading to its contraction and posing challenges for startups in raising funds (Silva et al., 2021).

Determining a company's worth is crucial for various purposes, including investment decision-making, mergers and acquisitions, and financial reporting. However, it is challenging for public and private companies due to information asymmetry, uncertainty in methodologies, and the need for additional financial data.

In the field of company valuation, it has been customary for valuers to rely on a single valuation methodology. However, a shift is occurring within the industry towards conducting comprehensive valuations, which involve the integration of multiple valuation methodologies to derive a final valuation estimate. While some valuers may adopt a simplistic approach by averaging the intermediate valuations, a more precise and refined technique involves assigning distinct weights to individual valuation methodologies based on the specific attributes and characteristics of the company under evaluation. This rationale underlies the framework developed in Subsection 3.2.1. By incorporating the concept of weighted valuation methodologies, a more accurate and sophisticated valuation approach can be achieved, enhancing the reliability and credibility of the valuation process.

Valuation methodologies vary depending on the actors involved and the circumstances. These methodologies aim to determine a company's intrinsic value and can be broadly categorised into traditional accounting-based approaches and non-traditional methods. In the startup environment, valuation becomes even more uncertain due to information discrepancies, limited historical financial data, and the complexities of assessing intellectual capital (Gavious & Schwartz, 2011).

Valuing public companies is relatively easier as their market value can be determined through stock prices. However, private companies, especially early-stage startups, present significant challenges (Audretsch & Link, 2012). Early-stage firms often lack revenues, and their primary value driver may lie in their intellectual capital, such as a unique business model or cutting-edge technology, which is difficult to evaluate. Additionally, traditional accounting-based valuation methodologies may not be sufficient for such firms, further complicating the process.

The VC industry has developed new valuation methodologies to account for these unique characteristics. However, the multitude of available methods and the lack of clarity on their appropriate usage make it challenging for practitioners to choose the most suitable approach. The impact of specific company characteristics on valuation and their influence on the valuation structure still require further exploration. Alternatively, statistical and machine learning techniques can be used for company valuation, but their applicability and interpretability are limited, and they are not widely adopted in the industry due to their complexity.

1.2 Context

Apollonian, a Luxembourg-based software-as-a-service company specialising in company analysis and valuation solutions, identified information gaps that hinder investment decision-making in the private equity market. The project aims to develop an artificial intelligence (AI)-powered tool called Insight to enhance Apollonian's flagship platform. This tool will leverage multiple valuation methodologies and company characteristics to provide accurate valuations and extract valuable insights. Currently, the valuations are performed manually by industry experts, which is labour-intensive and prone to human error, despite having a non-exhaustive econometric model to support decision-making.

A business analyst at Apollonian leads the study in collaboration with valuation experts and a project manager to improve the valuation process. The project has partnered with the Faculty of Engineering at the University of Porto to develop a model that determines the importance of different valuation methodologies. The results will provide Apollonian with a more secure and reliable valuation process. Additionally, the project's outcomes and methodology offer insights for improving the accuracy of valuations in the venture capital industry, contributing to the broader field of company valuation.

This thesis conducts an in-depth analysis of machine learning's impact on company valuation, using Apollonian as a case study. It highlights the transformative potential of machine learning in the VC industry and its role in informing valuation decision-making.

1.3 Research objectives

The primary objective of this research is to enhance current company valuations by surpassing industry benchmarks, exploring the impact of various valuation methodologies and company characteristics on firm valuation and analysing how the valuation structure (contributions of different valuation methodologies) adapts to unique company attributes. A key aim is to develop a universally applicable system that can be widely adopted across the industry.

To achieve these objectives, the following research questions are addressed:

- What are the most useful valuation methodologies for the VC industry?
- What factors are the most influential when valuing startups?
- How do individual factors change valuation structure?

This study introduces an innovative approach to the valuation problem by incorporating advanced machine learning (ML) techniques. A dataset of companies seeking or recently receiving their first round of investments is utilised. Four regression models are established, including a novel framework resembling the actual valuation process. The objective is to validate these models by surpassing industry performance benchmarks. The study employs Local Interpretable Model-Agnostic Explanations (LIME), a technique that provides insights into the most relevant valuation methodologies and company characteristics for firm valuation and how they change with the latter. Furthermore, the analysis explores the correlation between changes in company characteristics and the evolution of the valuation structure.

The proposed solution is a groundbreaking decision support system with high interpretability. It determines the weights of valuation methodologies in a comprehensive valuation process, resulting in a more accurate company valuation. The system incorporates multiple valuation methodologies and considers their contributions to the final valuation. The obtained results empower practitioners and investors to make data-driven investment decisions, reducing reliance on expertise or intuition. Additionally, the system facilitates mergers and acquisitions by providing a precise and objective valuation. Its user-friendly nature and adaptability to the industry enhance its appeal and minimise resistance to change.

1.4 Outline

The remainder of the thesis is as follows. Chapter 2 summarises the most relevant valuation methodologies and a review of ML applications in finance and company, emphasising the most relevant contributions to this work. Chapter 3 discusses the methodology, going over the data mining (DM) methodology and the proposed ML frameworks. Furthermore, Chapter 4 analyses the study results by validating the frameworks and discussing the importance and relations of the valuations methodologies and company characteristics. Lastly, Chapter 5 presents the main conclusions and alludes to the work plan intended to follow this study.

Chapter 2

Theoretical framework

The theoretical framework presented in this chapter is paramount for the reader to grasp state of the art in machine learning applications, particularly in finance and firm valuation. Section 2.1 provides a comprehensive overview of valuation methodologies, highlighting their significance. Through Section 2.2, the foundation is laid for understanding the context in which this research operates. Finally, Section 2.3 identifies the existing knowledge gap and articulates the purpose of the study in addressing it. Together, these sections set the stage for the subsequent chapters, paving the way for a comprehensive and insightful exploration of firm valuation in the context of machine learning.

2.1 Company valuation

This section examines various methodologies employed to assess the value of a company. It is divided into two primary approaches: Subsection 2.1.1 focuses on financial statements-based company valuation methodologies, and Subsection 2.1.2 reports on implicit intangible assets-based valuation methodologies. The former scrutinises financial statements and employs discounted cash flow (DCF) analysis and market multiples. Conversely, the latter considers intangible assets that may not be explicitly reflected in financial statements. This section comprehensively analyses these methodologies' strengths, limitations, and applications, contributing to a deeper understanding of company valuation within the academic discourse.

2.1.1 Financial statements-based company valuation methodologies

Financial valuation methodologies are critical in various business and financial activities. Whether it is assessing the value of a corporation, evaluating investment opportunities, or setting incentive pay for investors, accurate valuations are crucial to making informed decisions (Koller, Goedhart, & Wessels, 2020; Shapiro, Mackmin, & Sams, 2019). Fair Value is at the heart of these valuations: the price at which an asset would be sold in an orderly transaction between market participants at the measurement date (Pinto, Henry, Robinson, & Stowe, 2010; Weil, Schipper, & Francis, 2012).

As market conditions change and new information becomes available, valuations can fluctuate. Still, with the right tools and techniques, valuers can provide well-founded and generally accepted assessments that reflect current market conditions and participant assumptions. Using current market data and careful judgement and estimation can help ensure valuations are as accurate as possible (Pinto et al., 2010; Fabozzi, Focardi, & Jonas, 2017).

Valuers can use primary financial statements (balance sheet, income statement, and cash flow statement) to categorise various valuation techniques based on financial statements. The latter financial statement category can be further broken down into discounted cash flow, multiples, real options and value creation methods (Vayas-Ortega, Soguero Ruiz, et al., 2020). The methodologies belonging to each category are summarised in Table 2.1.

Valuers must choose the technique or techniques best suited to the nature, facts, and circumstances of the investment being valued. There are uncertainties involved in estimating fair value, so valuers should take into account all information that is known or knowable as of the assessment date - with careful analysis and attention to detail, valuers can provide valuable insights that help guide financial decision-making in a variety of contexts (Koller et al., 2020).

2.1.1.1 Balance sheet-based methods

Balance sheet methods value a company using pre-existing accounting information from its records. They provide an accurate financial position by considering contextual factors and are widely used by analysts and investors in finance (Glaveckaite, 2020). One of the key benefits of balance sheet-based methods is that they allow analysts to use actual financial data to make informed decisions about a company's value. This information is typically readily available in company records, making it easy to access and use for valuation purposes. Additionally, these methods often incorporate contextual factors such as future taxes, goodwill, unrealised capital gains, or project finance-associated variables, providing a more comprehensive understanding of a company's financial position (Vayas-Ortega, Soguero Ruiz, et al., 2020)

Valuation methodologies such as Adjusted Book Value, Liquidation Value, Substantial Value, and Net Assets Value determine a company's or its assets' worth. Adjusted Book Value adjusts the company's book value by considering the fair market value of its assets and liabilities. At the same time, Liquidation Value estimates the amount realised if a company's assets were sold individually in a forced or orderly liquidation process. Substantial value considers the significant investments made in a company or its assets, and Net Assets Value calculates the value of a company by subtracting its total liabilities from its total assets.

Book Value (BV) valuation methodology is a financial analysis technique that determines the value of a company based on its net assets.

$$Book\ Value = Total\ Assets - Total\ Liabilities$$

Where:

- *Total Assets* represents the sum of all assets owned by the company, including tangible and intangible assets.

- *Total Liabilities* is the sum of all debts and obligations owed by the company.

The rationale behind this methodology is that the value of a company is equal to the sum of its assets minus the sum of its liabilities, which results in the net asset value. It benefits companies with significant tangible assets, such as manufacturing or real estate. BV provides a straightforward and objective approach to valuing a company based on its historical financial records, making it easy to understand and calculate. However, it does not consider factors such as market conditions, future earnings potential, or intangible assets, which may result in an undervaluation or overvaluation of the company's true worth.

2.1.1.2 Income statement-based methods

The income statement is a fundamental document for understanding a company's financial performance, as it summarises revenues and expenses over a specific period. Income statement-based methods utilise this information to assess a company's value. This approach is beneficial when valuing companies with strong revenue streams and a history of stable earnings (Vayas-Ortega, Soguero Ruiz, et al., 2020).

One of the most common income statement-based methods is the Price-to-Earnings Ratio (P/E), which measures how much investors are willing to pay per dollar of earnings. Another method is Sales Multiple, which compares a company's sales to those of similar companies to determine its value. The Price-to-EBITDA Ratio (P/EBITDA) is also widely used to assess a company's earnings before interest, taxes, depreciation, and amortisation. Finally, the Business Acquisitions and Investments Test (BAIT) is a comprehensive approach that considers a range of factors, such as market conditions and industry trends, to determine a company's value (Mnejja & Sahut, 2010; Vidal & Ribal, 2017; Liu, Nissim, & Thomas, 2002).

While income statement-based methods provide a valuable framework for company valuation, they have limitations. For example, they may not be appropriate for companies with volatile earnings or those in rapidly changing industries (which is generally the case for startups). When used with other methods, such as balance sheet-based methods and discounted cash flow analysis, these can provide a comprehensive and robust company valuation.

Graham's Intrinsic Value (Graham's IV) methodology is a financial analysis technique that determines a company's intrinsic value based on its fundamental data. The rationale behind this methodology is that a company's intrinsic value equals the present value of its future earnings.

$$\textit{Intrinsic Value} = (\textit{EPS} \times (8.5 + (2 \times \textit{long-term growth rate of the company})) \times 4.4) \div \textit{AA Corporate Yield}$$

Where:

- *EPS* represents the company's earnings per share, indicating its profitability.
- *long-term growth rate of the company* is the expected long-term average annual growth rate.
- *AA Corporate Yield* is the yield on AA-rated corporate bonds.

For the intrinsic value of a whole company, the formula would be:

$$\textit{Intrinsic Value} = \textit{Graham's formula} \times \textit{Shares outstanding}$$

$$= (\textit{Earnings} \times (8.5 + (2 \times \textit{CAGR})) \times 4.4) \div \textit{AA Corporate Yield}$$

Where:

- *Earnings* represents the company's earnings, indicating its profitability.

- *CAGR* is the Compound Annual Growth Rate of the company.

- *AA Corporate Yield* is the yield on AA-rated corporate bonds.

Calculating intrinsic value involves estimating the company's future earnings and discounting them back to their present value using a discount rate. The discount rate is typically the risk-free rate plus a premium for the company's risk. The future earnings estimate is based on the company's historical earnings growth rate and expected future growth rate. The expected growth rate is based on the company's industry and competitive position. It is beneficial for companies with stable earnings and predictable growth rates. It also offers advantages in its focus on fundamental analysis and simplicity. However, it has limitations, relying on historical data that may not reflect future performance or intangible assets' value. It assumes market efficiency, which may not always hold.

2.1.1.3 Discounted cash flow analysis

Valuing a company accurately is crucial for investors, and the Discounted Cash Flow (DCF) analysis is one of the most commonly used methods for large-company valuation (Williams, 1997; Bruner, Eades, Harris, & Higgins, 1998; Liu, Nissim, & Thomas, 2007). This method involves discounting the expected future cash flows, specifically the Free Cash Flow (FCF), to arrive at the company's present value. The Residual Income Model (RIM) is another popular method within this set, which considers net future residual income plus the firm's capital to estimate the company's value (Dechow, Hutton, & Sloan, 1999; Francis, Olsson, & Oswald, 2001).

To estimate the fair value of an investment using the DCF method, the enterprise value is derived by projecting future cash flows or earnings and calculating the terminal value. This enterprise value is then discounted to the present value using the Weighted Average Cost of Capital (WACC), which reflects the company's debt and equity costs. The valuer adjusts the enterprise value by considering surplus assets, non-operating assets, and excess liabilities. Additionally, they account for financial instruments that rank higher than the firm's instrument in a liquidation scenario and assess the impact of potential dilution on the investment. Finally, the attributable enterprise value is allocated among the relevant financial instruments based on market participant perspectives (Pablo, 2002; Damodaran, 2014).

$$\textit{DCF Value} = \frac{\textit{Cash Flow Year 1}}{(1+\textit{Discount Rate})^1} + \frac{\textit{Cash Flow Year 2}}{(1+\textit{Discount Rate})^2} + \ldots + \frac{\textit{Cash Flow Year N}}{(1+\textit{Discount Rate})^N}$$

Where:

- *Cash Flow Year i* refers to the expected cash flow in year *i*.

- *Discount Rate* is the rate used to discount the future cash flows. It represents the investor's required rate of return and considers the risk associated with the investment.

The DCF method, while valuable, comes with specific considerations. It relies on detailed cash flow forecasts and estimates the terminal value and an appropriate risk-adjusted discount rate. These inputs involve subjective judgments and even slight variations can substantially impact the calculated present value (Damodaran, 2016). However, the DCF method offers flexibility, making it applicable to various cash flow or earnings streams. It is particularly advantageous for companies with stable cash flows, predictable growth rates, and situations where other valuation techniques may not be suitable (Olbrich, Quill, & Rapp, 2015).

The terminal growth methodology is another widely used approach for estimating a company's value. It applies a terminal growth rate to account for the expected cash flows beyond the explicit forecast period. This methodology assumes that companies will continue to generate cash flows in the long term. The free cash flows beyond the explicit forecast period are projected to calculate the terminal value, and a terminal growth rate is applied. Typically, the terminal growth rate is based on the expected long-term growth rate of the company's industry. The resulting terminal value is then discounted to the present value using the WACC as the discount rate. Adding the present value of the explicit forecast period cash flows to the terminal value provides the company's total enterprise value.

$$Terminal\ Value = \frac{Cash\ Flow\ Year\ N \times (1 + Terminal\ Growth\ Rate)}{Discount\ Rate - Terminal\ Growth\ Rate}$$

Where:

- *Cash Flow Year N* refers to the expected cash flow in the last projected year.
- *Terminal Growth Rate* represents the assumed growth rate for cash flows beyond the projection period.
- *Discount Rate* is the rate used to discount the terminal value back to its present value.

The methodology captures the long-term value of a company and is helpful for companies with stable cash flows. However, it relies on subjective assumptions and may not predict indefinite cash flow generation accurately. Selecting the terminal growth rate and estimating cash flows beyond the forecast period is crucial, and small changes can significantly impact the valuation.

2.1.1.4 Multiples valuation

Valuation techniques often involve multiples, which can be applied to earnings or revenue to estimate the fair value of an enterprise. The valuer should select an appropriate and reasonable multiple based on the company's size, risk profile, and earnings growth prospects. They should then adjust the enterprise value for surplus or non-operating assets, liabilities, and other contingencies to derive an adjusted enterprise value. The valuer should deduct any financial instruments that rank ahead of the highest-ranking instrument of the fund in a liquidation scenario and then apportion the attributable enterprise value appropriately between relevant financial instruments (*International Private Equity and Venture Capital Valuation guidelines*, 2022).

Several earnings multiples or ratios are commonly used, including Enterprise Value/Earnings Before Interest and Tax (EV/EBIT), Amortisation (EV/EBITA) and Depreciation (EV/EBITDA).

A revenue multiple may be appropriate for enterprises that have not yet obtained sustainable profitability. Industry benchmarks may be helpful as a sanity check of values produced using other techniques. Some industries have industry-specific valuation benchmarks, such as "price per barrel of oil equivalent (BOE)" for oil and gas exploration and production companies and "price per ton" for steel and other commodity-based industries where the profitability of businesses in the industry does not vary much (Damodaran, 2012).

The use of multiples has some limitations that should be taken into consideration: one of the main limitations is that it relies heavily on assumptions and judgments made by market participants regarding the future earnings potential of the company being valued. In addition, multiples can be affected by factors such as changes in market conditions or industry trends, making it challenging to compare multiples over time. Another limitation is that multiples can be influenced by non-operating items such as one-time gains or losses, making it challenging to determine the actual underlying earnings of a company. Finally, multiples do not consider the risks associated with the company's value, such as regulatory risks, which can impact its future earnings potential[1].

The benchmark valuation methodology compares a company to similar firms in the market. It involves identifying comparable companies, gathering their financial data, and applying their average or median multiples to the valued company. This approach is simple, market-based, and helpful for assessing relative valuation. However, it relies on accurate data, assumes market accuracy, and may oversimplify or overlook unique company characteristics. It serves as a reference point when used alongside other valuation methods.

$$\textit{Company Value} = \textit{Benchmark Multiple} \times \textit{Financial Metric}$$

Where:

- *Benchmark Multiple* refers to the valuation multiple derived from similar companies or industry standards.

- *Financial Metric* represents the relevant financial metric of the company being valued, such as earnings, sales, or EBITDA.

2.1.1.5 Value creation-based methods

The production of cash flows and value over time is the main scope of value-based valuation approaches. These techniques are beneficial for assessing businesses that have invested significantly in intangible assets like intellectual property or brand equity. These methodologies offer a more thorough and long-term perspective of the company's worth by examining the cash flows produced by a company's operations and comparing them to the expenses of capital (Koller et al., 2020).

Value creation-based methods include Cash Flow Return on Investment (CFROI) and Economic Profit. CFROI measures cash flow relative to invested capital, while Economic Profit considers the cost of capital. They provide comprehensive insights into a company's performance,

[1] Available at Damodaran Online (`https://pages.stern.nyu.edu/~adamodar/`), the webpage of Prof. Aswath Damodaran, a scholar of the Stern School of Business at New York University.

including expenses and investments. These methods aid investment decisions, particularly for long-term evaluation. Institutional money managers widely use CFROI and Economic Profit captures profitability with the cost of capital in mind. They differ from traditional methods by incorporating the cost of capital and time value of money (Pablo, 2002).

Value creation-based approaches have significant limits, despite offering a more thorough picture of a company's value and potential for future growth. They depend, for instance, on subjective calculations of the cost of capital, which are highly variable depending on the state of the market and investor opinions. Furthermore, it can be challenging to compare businesses when non Generally Accepted Accounting Principles (GAAP) measurements like Economic Value Added (EVA) and Economic Profit are used. Nonetheless, these methods continue to be a valuable resource for investors and analysts looking better to understand a company's potential for wealth creation (Damodaran, 2009).

The EVA methodology is a financial analysis technique that measures the value created by a company for its shareholders. The rationale behind this methodology is that traditional accounting profits do not fully reflect the actual value created by a company, and therefore, a more comprehensive measure is needed. EVA is calculated by subtracting the cost of capital from the after-tax operating profit of a company.

$$EVA = Net\ Operating\ Profit\ After\ Tax\ (NOPAT) - Capital\ Charge$$

Where:

- *NOPAT* refers to the company's operating profit after deducting taxes but before deducting the cost of capital.

- *Capital Charge* represents the cost of the capital employed in the business. It is calculated by multiplying the invested capital by the company's cost of capital, often expressed as a percentage.

The resulting EVA represents the value created by the company above and beyond the cost of capital. A positive EVA indicates that the company is creating value for its shareholders, while a negative EVA indicates that it is destroying value. It measures value creation and aligns shareholder and management interests. However, EVA relies on accounting data and can be complex to calculate.

The Cash Value Added (CVA) Valuation Methodology is a financial analysis technique that measures the value created by a company for its shareholders. The CVA methodology is similar to the EVA methodology. Still, instead of subtracting the cost of capital from the after-tax operating profit of a company, it subtracts the cash cost of capital from the after-tax operating cash flow of a company.

$$CVA = Operating\ Cash\ Flow - Cost\ of\ Capital$$

Where:

- *Operating Cash Flow* represents the cash flow generated by the company from its operations, including revenues and cash expenses.

- *Cost of Capital* refers to the opportunity cost of the capital employed in the business. It is typically determined by applying a reasonable discount rate to the company's invested capital.

It directly measures a company's ability to generate cash flows and is particularly useful for assessing liquidity and cash flow management. However, CVA may overlook non-cash items and does not account for the time value of money. It is essential to consider cash and non-cash factors when evaluating a company's financial performance.

The Total Cost of Ownership (TCO) is a financial analysis technique that measures the total cost of owning and operating an asset over its entire life cycle. The rationale behind this methodology is that traditional valuation methods, such as DCF, do not fully capture the total cost of owning and operating an asset. The TCO methodology considers all costs associated with owning and operating an asset, including acquisition, maintenance, operating, and disposal costs. The TCO is calculated by adding all these costs over the asset's life cycle and discounting them to their present value.

$$TCO = Acquisition\ Cost + Operating\ Costs + Maintenance\ Costs + Disposal\ Costs$$

Where:

- *Acquisition Cost* refers to the initial cost of acquiring the asset or system, which includes purchase price, installation costs, and any necessary modifications.

- *Operating Costs* include expenses incurred during the operational phase, such as energy costs, consumables, labour, training, and insurance.

- *Maintenance Costs* encompass expenses associated with regular maintenance, repairs, and replacements required to keep the asset or system functioning optimally.

- *Disposal Costs* represent the expenses related to the end-of-life phase, including decommissioning, disposal, and any associated environmental or regulatory compliance costs.

One advantage of TCO is that it provides a comprehensive view of the actual cost of ownership, enabling better decision-making regarding investments. It helps identify cost-saving opportunities and facilitates comparison between different options. However, TCO requires accurate data and assumptions, which can be challenging. It may also overlook intangible benefits and long-term value creation.

2.1.1.6 Real options analysis

Real options analysis is a financial technique that allows decision-makers to evaluate and analyse the potential value of investment opportunities by considering the flexibility they offer. It is based on the idea that investment projects are not static but involve options that can be exercised or abandoned, depending on how the project develops over time. This approach is useful when dealing with projects with significant uncertainty and flexibility, such as R&D projects, real estate development, and oil exploration projects, among others (Vayas-Ortega, Soguero Ruiz, et al., 2020).

Real options analysis encompasses various valuation techniques, including the Black-Scholes model, investment options, expand/delay project valuation, Monte Carlo simulation, and Option Pricing valuation. The Black-Scholes model calculates the theoretical value of European-style options by assuming a log-normal distribution for the underlying asset's price and exercise only at

Table 2.1: Compilation of company valuation methods based on financial statements. Adapted from Vayas-Ortega, Soguero Ruiz, et al. (2020)

Balance Sheet	***Income Stat.***	***Cash Flow***			
		Multiples	***DCF***	***Options***	***Value Creation***
Book Value	Graham's IV	P/E	Free CF	Black and Scholes	CFROI
Adjusted BV	P/Sales	EV/Sales	Equity CF	Investment Options	EVA
Liquidation Value	P/EBITDA	EV/EBITDA	Capital CF	Expand/Delay project	Economic Profit
Substantial Value	EPS	P/CF	RIM	Monte Carlo	Cash Value Added
Net Assets Value	BAIT	Industry	Terminal Growth	Option Pricing	TCO

expiration. Investment options involve the flexibility to adjust project elements based on the resolution of uncertainties. Accurate options valuation treats a capital investment as a series of options that provide future flexibility to delay, abandon, or expand project elements. The expand/delay project methodology values investments with the ability to adjust project elements based on uncertainty resolution. Monte Carlo simulation employs statistical methods to model the probability of outcomes in unpredictable processes influenced by random variables. Option Pricing valuation is a technique used to determine the value of financial options based on factors such as underlying asset price, strike price, time to expiration, volatility, and interest rates (Trigeorgis, 1996).

It is important to note that accurate options analysis has its limitations. The approach relies heavily on assumptions about the behaviour of markets and the underlying asset, which may not always hold. Additionally, the valuation of investment options depends on the accuracy of the underlying assumptions and estimates used in the analysis. Real options analysis is a valuable tool for evaluating investment opportunities and identifying sources of value that would otherwise be overlooked (Dixit & Pindyck, 1994).

Table 2.1 summarises the valuation methodologies mentioned above and their categories.

2.1.2 Implicit intangible assets-based valuation methodologies

The importance and integration of comprehensive company valuation cannot be overstated in today's business environment. Various corporate operations such as mergers and acquisitions, alliances and joint ventures, corporate bond issues, IPOs, banking credit, portfolio management, risk assessment, and even the top-management bonus require fair valuations that comply with well-founded and generally accepted systematic reviews (Koller et al., 2020; Fabozzi et al., 2017; Weil et al., 2012; Olbrich et al., 2015). Traditional financial valuation methods may not always provide accurate results. This is especially true for startup companies that lack complete economic indicators to evaluate potential startup value. A company's intrinsic value is also related to its non-financial information, such as its technical level, human resources, reputation, and intangible assets (Sievers, Mokwa, & Keienburg, 2013).

The recent investment frenzy in the startup ecosystem has resulted in crazy valuations utterly detached from fundamentals, and the number of seed deals is shrinking. The related uncertainty makes many of those deals hard to finalise, while a good opportunity still exists for investments in

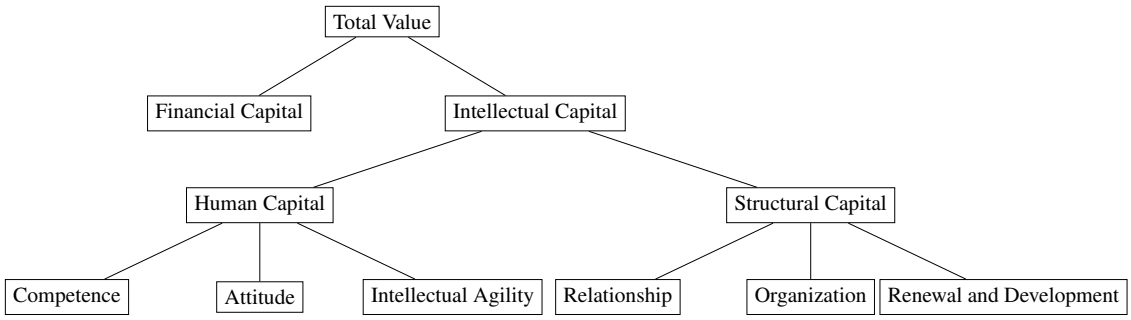


Figure 2.1: Diagram of the value distinction tree for intellectual capital. Source: Roos et al. (1998)

early-stage projects (Arroyo, Corea, Jimenez-Diaz, & Recio-García, 2019). Thus, assigning value to startups is challenging due to inaccuracies, making this even more relevant. Startups typically experience negative initial results, making their evolution to scaleups dependent on strategic decisions such as supplier choices, incentives to attract investors, and a qualified management team (Silva et al., 2021). Startups share several characteristics despite diversity, including lack of financial or accounting historical reference, equity dependence, high failure probability in their early stages, low liquidity, a high concentration of intangible assets, uniqueness in comparison to other companies, and a focus on management quality (Gavious & Schwartz, 2011; Glaveckaite, 2020). Furthermore, traditional company valuation methods are unsuitable for valuing startups due to factors such as sparse financial data, small scale, higher risk and uncertainty in value assessment, and ignoring non-financial information, which is critical for startups (W.-B. Huang, Liu, Bai, & Zhang, 2020). Valuing startups is also challenging as it involves subjective judgements over financial features (Mirzanti, Sinaga, & Soekarno, 2019). This is because the value of an entrepreneurial project is typically based on intangible assets during the first few years of a new venture, and classical financial due diligence for valuation purposes is often limited.

Intellectual capital, also known as intangible assets, is an essential element for businesses in the modern economy; however, there is no consensus on its definition, and various authors have different interpretations: for instance, Bontis, Dragonetti, Jacobsen, and Roos (1999) defined intellectual capital as "the collection of intangible resources and their flows". Similarly, Edvinsson (1997) saw it as a "source of intangible (hidden) assets that often do not appear on the balance sheet". Harrison and Sullivan (2000) defined it as "knowledge that can be converted into profit". In contrast, Roos et al. (1998) viewed it as "the sum of knowledge of a company's members and practical translation of this knowledge like trademarks, patents, and brands". Finally, Stewart (1999) defined it as "intellectual material - knowledge, information, intellectual property, experience - that can be used to create wealth". Furthermore, Roos et al. (1998) suggested a Value Distinction Tree, shown in Figure 2.1. These different definitions highlight the diversity of perspectives regarding intellectual capital and suggest that implicit intangible asset valuation may be necessary to evaluate a company's potential value.

The reasons why a company decides to start evaluating its intellectual capital can vary, but they can be categorised into two groups: internal and external. External factors, such as a better

public image, an increase in market value, a decrease in the difference between market and book value, more information for potential investors, and the overall market, frequently take precedence over internal ones when considering their impact on decision-making, overall business success, the connection between investments in intangibles and business goals, as well as the necessity to manage them (Sitar, 2005).

The market value of a firm's shares ultimately reflects the value of all its net assets, including intangible assets, which represent future growth opportunities and profitability that go towards increasing firm value (Chan, Lakonishok, & Sougiannis, 2000; Eckstein, 2004; Gleason & Klock, 2006). The valuation of early-stage companies can be challenging due to not having public shares, their high volume of intangible assets and a lack of adequate financial information. Additionally, startups are in the early life cycle stage, which makes estimates of cash flow, discount rate, and growth results distorted (Audretsch & Link, 2012; Yoo, Yang, Kim, & Heo, 2012).

Non-financial information can help investigate a company's value, and venture capital-backed firms are likely to combine non-financial and financial information on startup company valuation. There are not any established non-financial valuation methods in the intellectual capital market. This means individual studies must be analysed to grasp the state of the art.

According to research, venture capital-backed companies are more inclined to incorporate non-financial and financial information when valuing startup companies (Sievers et al., 2013). Research that considers non-financial elements frequently employs two points of view. The first stage in decomposing corporate value into a collection of different values is to build a qualitative evaluation methodology. The second phase examines the connection between corporate value and internal or external factors affecting the company's worth. Sievers et al. (2013) establishes the soundness of the three-factor model in his analysis of the correlations between financial and non-financial data, investor qualities, and business value. Mills (1995) analysed the company value from seven viewpoints, including employee personal development, service standards, employee tolerance of different ideas, entrepreneurship, social responsibility, level of the leadership team, and other behaviours connected to corporate values. Simpson (2008) looked into the relationship between the persistence of non-financial disclosures, including client acquisition costs, average revenue per user, and customer count and earnings forecasting errors. Moreover, data mining regression results show an alarming average correlation between diversity and intangible assets. There is conflicting data from the studies on the various elements that affect intangible assets (Lins, 2003; Wiwattanakantang, 2001).

It is essential to note that methodologies tend to split the latter into sections for measuring intellectual capital because this division is not consensual. Overall, intellectual capital can be divided into human capital and structural capital. Structural capital can be further subdivided into internal and external means. Kaplan and Norton (2007) divided intellectual capital into three perspectives: Learning and growth, Internal business processes, and Customer perspectives. Roos et al. (1998)'s model was based on Human capital, Organisation Renewal and development, and Relationships. Sveiby (2018) divided intellectual capital into Human competence, Internal structure, and External structure. Brooking (1996) modelled intellectual capital as the composition

Table 2.2: Intellectual capital measurement methods. Adapted from Sitar (2005)

Scorecard	***Direct Intellectual Capital***	***Market Capitalization***	***Return on Investments***
Human Capital Intelligence	Technology Broker	Invisible Balance Sheet	EVA
Balanced Scorecard	Citation-Weighted Patents	Market-to-Book Value	VAIC
Skandia Navigator	HRC	OAMV	CIV
Checklist	IVM	Tobin's Q	Knowledge Capital Earnings

of Human-centered assets, Infrastructure assets, Intellectual property assets, and Market assets. Edvinsson (1997) simplified Brooking's model to only three components: Human capital, Organisational capital, and Customer capital. Different classifications are used for different valuation methods; some are specific to one section, while others value intellectual capital.

The theoretical advancement in the field and the required adjustments in the actual implementation frequently lead to the presentation of new measurement methods or, at the very most, minor variations of those already known. Sveiby (2018), however, developed a way of presenting the plans that were the most structured. Based on the measurement level and the evaluation technique, the methods are divided into four groups: Scorecard, Direct Intellectual Capital Methods, Market Capitalization, and Return on Investments. These groups and their main valuation methodologies are summarised in Table 2.2.

Market Capitalisation methods calculate the total monetary worth of intangible assets. They primarily gauge the total amount of intellectual capital by the discrepancy between book worth and market value. The Invisible Balance Sheet, Market-to-Book Value, Owner Assigned Market Value, and Tobin's Q are a few techniques used in this area.

The second group, Return on Investments methods, evaluates intangibles with conventional financial measures on an overall organisational level, just like the first group. Unlike market capitalisation methods, return on assets methods do not attempt to calculate the total monetary worth of intangible assets. They focus on evaluating the impact of these intangibles on a company's overall financial performance, such as by measuring the return on investment (ROI) generated by these assets. The most widely used techniques in this category are Knowledge Capital Earnings, Economic Value Added, Value Added Intellectual Coefficient, and Calculated Intangible Value.

Scorecard valuation methods involve assessing a company's value by assigning scores or weights to qualitative and quantitative factors. These methods use a structured evaluation framework to consider criteria such as financial performance, market position, management quality, industry trends, and competitive advantage. They provide a systematic approach to compare and assess companies for investment decision-making. Only the Scorecard method uses distinct components and various indicators to determine intangible assets in non-financial terms. Forms of this group are Human Capital Intelligence, Balanced Scorecard, Skandia Navigator, Index, Intangible Asset Monitor, Knowledge Audit Cycle, Value Chain Scoreboard, Meritum Guidelines, Danish Guidelines, Topplinjen/Business, Holistic Value Approach, Rating, Measuring and Accounting and Knowledge Assets Map.

Direct Intellectual Capital methods similarly assess various categories of intellectual capital

individually, but once more in financial terms, and then calculate the total value by summing them up (Sveiby, 2018). These methods value a company's intangible assets, like patents and trademarks, by quantifying their worth and assessing their contribution to overall value. These methods consider uniqueness, earnings potential, market demand, and competitive advantage. Methods in this group are Technology Broker, Citation-Weighted Patents, Human Resource Costing & Accounting, Inclusive Valuation Methodology, Accounting for the Future, Statement, The Value Explorer, Intellectual Asset Valuation, Total Value Creation and Financial Method of Intangible Assets Measuring.

Based on the industry's popularity and frequency of use, 3 of the most popular intellectual capital valuation methods are Balanced Scorecard, Checklist Method and VAIC. These will be further explored next.

2.1.2.1 Balanced Scorecard

The Balanced Scorecard (BSC) is a management technique for developing, implementing, and managing business strategies. It provides a framework for performance measurement, transparency, and accountability. The BSC translates mission and method into objectives and measures across financial, customer, internal processes, and learning and growth perspectives. The BSC is calculated by defining key performance indicators (KPIs) across four perspectives: financial, customer, internal processes, and learning and growth. Each view represents a specific aspect of the organisation's performance critical to its success. For each perspective, specific objectives and corresponding measures are identified. Depending on the perspective, these measures can be financial and non-financial and are usually quantitative, allowing for easy tracking and comparison over time.

$$Valuation = Weight\ 1 \times Score\ 1 + Weight\ 2 \times Score\ 2 + \ldots + Weight\ N \times Score\ N$$

Where:

- *Weight i* represents the weight assigned to the *i*-th criterion.

- *Score i* represents the score assigned to the *i*-th criterion.

The weighted score of each criterion is calculated by multiplying the weight assigned to that criterion by the score given to it. This allows for the consideration of the relative importance of each criterion in the overall valuation.

It is widely used to evaluate performance in universities and corporate organisations, enhancing strategic management and fostering transparency. The BSC supports decision-making, benchmarking, and continuous improvement by linking non-financial and financial measures. However, limitations include the potential neglect of risks, environmental factors, sustainability, and stakeholder concerns.

2.1.2.2 Checklist methodology

The Checklist methodology is a qualitative method used to value a business. This method assumes a fixed maximum valuation based on the region and assigns the company a score for each of the

five criteria based on the answers to a set of questions. The five criteria are management, market, competition, technology, and financials. Each criterion is rated on a scale of 0 to 5, with 5 being the highest score. The sum of each element's values then becomes the business's valuation. The method treats qualitative traits as intangible building blocks that sum up the assumed maximum pre-money valuation. The process unfolds in three steps: (1) identifying the criteria, (2) scoring each criterion, and (3) calculating the total score. The Checklist Method Valuation Methodology is a straightforward model solely based on qualitative aspects, making it a traditional way to value pre-revenue startups.

The advantages of this approach include simplicity, ease of use, and the ability to evaluate multiple factors quickly. It provides a structured and systematic way to analyse various aspects of a business and can be helpful for novice investors or when time is limited. However, checklist valuation may oversimplify complex situations and overlook essential qualitative elements. It relies heavily on predetermined criteria, which may not fully capture a specific investment's unique characteristics or circumstances. Additionally, checklist valuation may not account for changing market conditions or incorporate dynamic factors that can impact the value of a company.

2.1.2.3 Value Added Intellectual Coefficient

The Value Added Intellectual Coefficient (VAIC) is a measurement tool used to evaluate a company's intellectual capital (IC) performance. The metric was developed by Ante Pulic, a Croatian researcher, in 1998. VAIC aims to determine how much value a company creates through the use of its intellectual capital. It uses information from formal financial statements to evaluate intellectual capital and, most importantly, its efficiency. Thereby it indicates the organisational intellectual ability (Sitar, 2005).

To calculate VAIC, the formula multiplies a company's physical capital (PC) by its human capital (HC) and then divides that product by its structural capital (SC). The resulting value is multiplied by a weighting coefficient (Pulic, 2000, 2004, 2005). The equation is as follows:

$$VAIC = \frac{TotalValueAdded}{Revenue}(HC + SC + CE)$$

Where:

- Total Value Added = Operating profit + depreciation + amortisation

- Revenue = Total sales revenue

- Human Capital Coefficient = (Employee satisfaction index + Human Capital Investment rate) / 2

- Structural Capital Coefficient = (Proportion of workforce with computer access + R&D expenditure / Total revenue) / 2

- Capital Employed Coefficient = (Value of physical assets / Total revenue) / 2

The rationale behind VAIC is that companies with higher intellectual capital, including knowledge, expertise, patents, trademarks, and other intangible assets, are more likely to generate higher profits and value for their shareholders. VAIC allows investors and stakeholders to measure the efficiency and effectiveness of a company's intellectual capital.

However, several studies have questioned the accuracy and reliability of VAIC as a measurement tool. Critics argue that VAIC does not adequately capture the value of a company's intellectual capital and that the weighting coefficient used in the formula is arbitrary and lacks a theoretical basis. Additionally, VAIC does not account for differences in industry, size, or market conditions (Clarke, Seng, & Whiting, 2011), which can affect a company's intellectual capital performance. VAIC has limitations and should be used cautiously as a single metric for evaluating a company's intellectual capital performance (Ståhle, Ståhle, & Aho, 2011).

2.2 Data mining applied to the industry

This section delves into the remarkable developments within the financial industry, particularly in the realm of AI and machine learning, as discussed in Section 2.1. The comprehensive exploration of these techniques reveals their broad range of applications and their swift integration into real-world practices. Section 2.2 takes a focused approach, delving deep into the specific landscape relevant to this study: company valuation. This area is intricately intertwined with investment criteria and plays a pivotal role in the assessment of investment opportunities.

2.2.1 Applications in the financial industry

In recent years, ML techniques have been increasingly utilised in the financial field (Fischer & Krauss, 2018). The usefulness of ML techniques in the financial sector has been shown in several studies (Alpaydin, 2010; Gan, 2013; Park, Kim, & Lee, 2014; Gunn, 1998), including the prediction of bankruptcy (Krauss, Do, & Huck, 2017; Aruldoss, Manjula, & Venkatesan, 2011), the classification of debt and credit ratings, and follow-up analysis and pattern detection in stock-quote series (Moreno Gutiérrez & Melo-Velandia, 2011; W. Huang, Nakamori, & Wang, 2005).

One of the primary issues in the literature on time series and machine learning is financial time series forecasting, studied by Tay and Cao (2001), which has shown some success with the application of machine learning techniques (Lee, 2009). Financial reports (Wang, Huang, & Wang, 2012), financial news (Groth & Muntermann, 2011; Schumaker, Zhang, Huang, & Chen, 2012), and even information from microblogs (Ruiz, Hristidis, Castillo, Gionis, & Jaimes, 2012) are all regarded as relevant sources of information in this area of research for forecasting future market behaviour.

An empirical approach using Artificial Neural Networks was put forth by Vanstone and Finnie (2009) to develop automated trading systems. The authors of that article analysed some research that builds intelligent trading systems using fundamental and technical analysis. Data mining techniques have been utilised in commercial applications like auditing and accounting, credit scoring, financial analysis, inflation, and marketing, among others, according to a relatively recent survey (Tkáč & Verner, 2016). Aguilar-Rivera, Valenzuela-Rendón, and Rodríguez-Ortiz (2015) explored evolutionary computation techniques to address financial issues, including genetic algorithms, genetic programming, and evolutionary algorithms.

To identify organisations in financial difficulty, W.-S. Chen and Du (2009) used MLP to learn some economic statistics and balance sheets of enterprises. Xie (2011) used Support Vector Regression to estimate stock closing prices for the third day in advance. When building an earnings prediction model utilising financial statement data, Ou and Penman (1989) used stepwise logistic regression analysis. Hunter, Saini, and Zaman (2018) suggested using a greedy strategy to address the "picking winners" problem starting from a Brownian motion model. In Bonaventura et al. (2019), the authors examine the task of predicting startup success within the framework of a global startup network. Similar research shows that different fundraising event stages result in various success factors (Gastaud, Carniel, & Dalle, 2019).

Although the advantages provided by data science, ML studies in entrepreneurship research are still in their infancy (Obschonka & Audretsch, 2020), new data analytics techniques present several problems for entrepreneurship researchers by changing the research methodology (George, Osinga, Lavie, & Scott, 2016; Kitchin, 2014; Sivarajah, Kamal, Irani, & Weerakkody, 2017), highlighting the value of interdisciplinary cooperation (Jin, Wah, Cheng, & Wang, 2015). ML has been applied to several other fields, such as the survival of new ventures (Antretter, Blohm, Grichnik, & Wincent, 2019), entrepreneurial skills (Prüfer & Prufer, 2020), or high-growth firms (Coad & Srhoj, 2019). Lin, Liang, Yeh, and Huang (2014) took on the financial distress prediction challenge, incorporating expert knowledge to choose the best features.

2.2.2 Applications to firm valuation

Machine learning techniques offer several advantages for company valuation, including flexibility, adaptability to specific company cases, and the ability to avoid outdated variables by consolidating historical information with real-time market trends and actual general economy drifts (Vayas-Ortega, Soguero-Ruiz, Rodríguez-Ibáñez, Rojo-Álvarez, & Gimeno-Blanes, 2020). Due to the complexity and numerous variables involved, few studies have successfully applied these methodologies to firm valuation (Pinto et al., 2010; Visconti, Montesi, & Papiro, 2018).

Despite these challenges, a few studies have successfully applied data science methods to startup investment and valuation, including predictive modelling to predict funding events (Sharchilev et al., 2018), ML investment performance (Blohm, Antretter, Sirén, Grichnik, & Wincent, 2020) and undisclosed startup valuation corresponding to a funding round for which only the funding amount was disclosed (Garkavenko, Mirisaee, Gaussier, Guerraz, & Lagnier, 2021). In a relevant work, Littunen and Niittykangas (2010) used financial and non-financial variables to study the relation of growth in 200 Finnish firms. Furthermore, numerous studies have utilised machine learning techniques to predict the outcomes of mergers and acquisitions or the success of startups. A rule-based expert system that forecasts the purchase of companies is presented in one of the earlier publications in this field (Ragothaman, Naik, & Ramakrishnan, 2003) and, despite just using a dataset of 200 companies for evaluation, this method has a 70% success rate. Using a vast dataset, Arroyo et al. (2019) accurately predicted the future outcomes of companies, including acquisitions, closures, IPOs, funding rounds, or no events. Similarly, Haralampiev, Yankov, and Ruskov (2014) examined various machine learning techniques to forecast the success of 142 Bulgarian

companies using information from a survey. For the prediction of business models with better growth expectations and survival prospects, a more advanced method combines supervised (Support Vector Machines) and unsupervised learning (clustering) (Haddad, Weking, Hermes, Böhm, & Krcmar, 2020). When forecasting the outcomes of early-stage firms, compare how well human experts and machine learning systems perform (McKenzie & Sansone, 2017) (Machine learning techniques do not significantly outperform human expertise). Machine learning was used by Futagami, Fukazawa, Kapoor, and Kito (2021) to forecast acquisitions for businesses created between 1970 and 2007. Machine learning with big data has been used to remove psychological bias in asset prices for company valuation by using observable asset prices as objective variables, and panel data is used to increase sample size (Tsukahara, Tanaka, & Suzuki, 2022). Interpretability is questioned, but the ability to enhance versatility and expressiveness is an advantage. Traditional models can be built to mitigate the black box problem rather than introducing new machine learning techniques from scratch.

More recently, Quintero (2019) examined the relationship between startup valuation at various stages and the amount of funding. They suggest a linear (in log-log space) model for this objective. Only the financing amount and the round funding series are considered in their proposed method.

Table 2.3 provides a comprehensive overview of ML applications in firm valuation as documented in the literature.

Table 2.3: Overview of ML Applications in Firm Valuations: Research Summary

Reference	**Purpose**	**ML Technique(s)**	**Key Findings**
Blohm et al. (2020)	Predicting new venture survival	Gradient Boosted Decision Trees	ML algorithms consistently generate higher investment returns than angel investors on angel investment platforms.
Futagami et al. (2021)	Predicting future pairs of the acquirer and acquiree companies	SHAP value interpretation	Newly considered network and company relation features demonstrate significant importance, with 10 out of the top 22 highly influential features.
Garkavenko et al. (2021)	Inferring startup valuation corresponding to announced funding rounds	Supervised, unsupervised and semi-supervised domain adaptation	The proposed method outperforms previous approaches in the literature by a significant margin. Domain Adaptation framework proves suitable for the task.
Haddad et al. (2020)	Predicting business models with growth potential and survival prospects	Support Vector Machines (SVM) and clustering	Some business models exhibit strong associations with different variables affecting startup success.
Haralampiev et al. (2014)	Predicting startup success	Trees, Bayesian and Rules-based algorithms	Start-up success is predicted with 83.76% accuracy.
Littunen and Niittykangas (2010)	Examining factors influencing high growth in new firms	Logistic regression analysis	Clear connection between entrepreneur's know-how and firms' high growth. External networks also accelerate growth.
McKenzie and Sansone (2017)	Predicting winning and non-winning companies	LASSO, SVM, Boosted Regression	Modern machine learning methods do not significantly improve predictions, indicating the challenge of selecting competition winners.
Quintero (2019)	Examining the relationship between startup valuation and funding amount	Linear (log-log space) model and Bayesian model	Statistically significant results provide practitioners with a tool to estimate post-money valuations
Ragothaman et al. (2003)	Classifying firms as acquisition candidates based on financial and other data	IXL, a rule induction technique	A rule induction technique aids financial analysts and investors in buy/sell decisions
Sharchilev et al. (2018)	Predicting early-stage startup success	Gradient boosting (GB)	Utilising companies' web mentions significantly enhances predictive performance compared to using structured data alone.
Tsukahara et al. (2022)	Separating intrinsic equity value from mispricing caused by irrational investor sentiments	Drift Detection Method, SVM and Recurrent Inference Machines	The approach confirms its validity in separating intrinsic value from sentiment-driven mispricing.
Vayas-Ortega, Soguero-Ruiz, et al. (2020)	Pricing and benchmarking Enterprise Value for valuation and credit risk assessment	Decision Trees (DT), Bagging Trees, SVM, Gaussian Process Regression	Features related to value creation (DCF) are crucial while incorporating industry/country-specific variables improves ML performance.
Arroyo et al. (2019)	Predicting the next outcome of a company: acquisition, closure, IPO, funding round, or no event	SVM, DT, Random Forests (RF), Extremely Randomised Trees (ERT) and GB.	Machine learning aids venture investors in decision-making, uncovering opportunities and assessing investment risks more effectively.

2.3 Knowledge gap

Valuing companies is an important task in the financial industry, as it helps investors make informed decisions about new or ongoing investments since data mining and machine learning techniques have been increasingly used. Some limitations and gaps still need to be addressed in future research.

First, there is a lack of systematic reviews that evaluate the performance of data mining and machine learning techniques for public and private equity companies (Pinto et al., 2010). As stated in Subsection 2.2.2, some studies have been conducted on specific aspects of company valuation, such as predicting mergers and acquisitions or IPOs. However, a comprehensive evaluation of these techniques for all companies is still missing. This study seeks to improve startup valuation techniques by applying personalised and comprehensive DM and ML approaches. By analysing the startup environment, the study aims to determine the effectiveness of these techniques in predicting the valuation of these types of firms.

Second, the complexity and many variables involved in company valuation make it challenging for data mining and machine learning algorithms (Melgarejo-Meseguer et al., 2019). As a result, most of the works presented so far have either worked with a small dataset or with data retrieved ad hoc, which limits the generalisation of their results (Arroyo et al., 2019). Therefore, there is a need for the development of more robust and scalable algorithms that can effectively handle the complexity of company valuation. Although the study features fewer companies in the dataset, it can still provide a comprehensive analysis using various exhaustive data points. By basing the predictor selection on previous literature, a robust, reproducible and reliable set of mutually exclusive and thorough predictors was developed, which can be used in future research.

Third, there is a lack of research on the determinants that influence firm value at different stages of maturity, such as early stage and scaleup. This study employs an exhaustive set of data points to develop interpretable and reproducible predictors for startup valuation. By examining the influence of different predictors, such as business maturity, on various valuation values, the study sheds light on the factors most important in determining startup value and how they vary depending on valuation. Additionally, the study reflects on the varying levels of importance assigned to different valuation methodologies based on company characteristics.

Fourth, there is a need to adopt unconventional valuation methods, namely non-financial methods mentioned in Subsection 2.1.2. These methods can complement traditional accounting-based methods and provide a more comprehensive and accurate assessment of company valuation. This study places particular emphasis on the implicit intangible assets of the startup and identifies variables that reflect this based on relevant literature. By analysing the importance of these variables for startup valuation, the study aims to provide insight into the critical implicit intangible asset methodologies that affect firm valuation.

Finally, using primary data through surveys and class associations can provide a more complete and accurate picture of company valuation. These methods can help overcome the limitations of secondary data sources and provide a more nuanced understanding of the determinants of

company valuation. It is essential to take a closer look at emerging countries, as they are often overlooked in the literature despite their potential and growing startup ecosystem. These countries have specific characteristics and nuances that have yet to be thoroughly analysed, making it necessary to explore this area in more detail. This study has a few limitations regarding reducing the research gap in the field. Firstly, the lack of primary data usage and the need to rely on secondary sources limit the depth of the analysis. Additionally, although the study aims to explore the potential of emerging countries, the limited number of observations in these regions means there is still much to be discovered regarding their startup ecosystems. Finally, the low number of observations across the dataset is a significant limitation, as it limits the ability to develop accurate models and derive meaningful results.

The study significantly contributes to evaluating private enterprises within the venture capital landscape by offering a comprehensive and reliable approach. By integrating diverse valuation methodologies and leveraging ML to determine the weights assigned to each method, this research enables valuers to adopt a data-driven approach to company valuation. Furthermore, the study investigates the impact of company characteristics, serving as generalisable predictors, on the valuation process, and explores how this influence evolves with firm value. Another noteworthy contribution of this research lies in developing a novel ML framework that closely emulates the valuation process. By striving to achieve superior valuations compared to industry benchmarks, the frameworks tested aim to enhance the accuracy and effectiveness of valuations. Overall, the findings of this study provide valuable insights into the factors and valuation methodologies that contribute to startup valuation. These insights serve as a valuable resource for researchers and practitioners. The study's contributions extend beyond interpretability and encompass testing and validating four distinct machine learning frameworks, including a disruptive framework, always considering predictive performance. The relationships tackled in the present study are reflected in Figure 2.2.

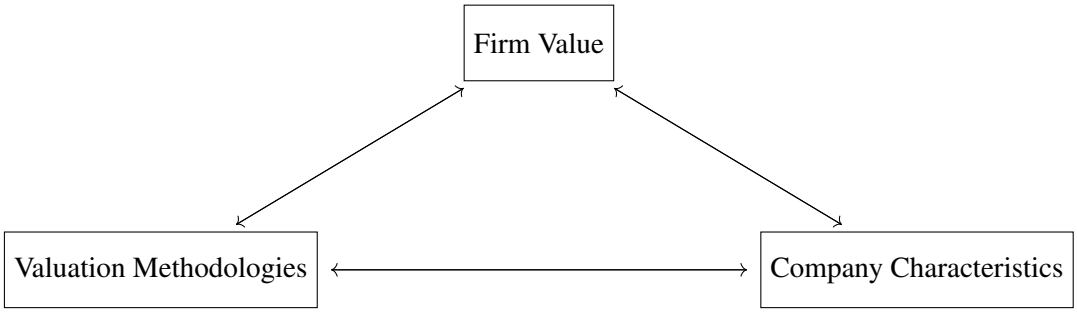


Figure 2.2: Diagram of relationships in contributions of the study - interconnectedness between firm value, valuation methodologies and company characteristics

Chapter 3

Methodological framework

This chapter presents the methodology employed in this study, which focuses on two key aspects: the data mining process and the approach to company valuation.

It is imperative to delve into the comprehensive valuation process, previously outlined in Section 1.1, to understand the methodology employed in this study. This process entails valuers computing multiple intermediate valuations for the same company, employing diverse valuation methodologies such as those elucidated in Section 2.1. Subsequently, the valuer assigns arbitrary weights to each intermediate valuation and calculates a weighted average, which yields the final company valuation. These weights are typically based on the valuer's experience and occasionally take into consideration specific company characteristics.

Section 3.1 focuses exclusively on the data mining process, a critical component applicable across diverse models. This section adheres to the well-established Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology (Figure 3.1), tailored to the specific datasets.

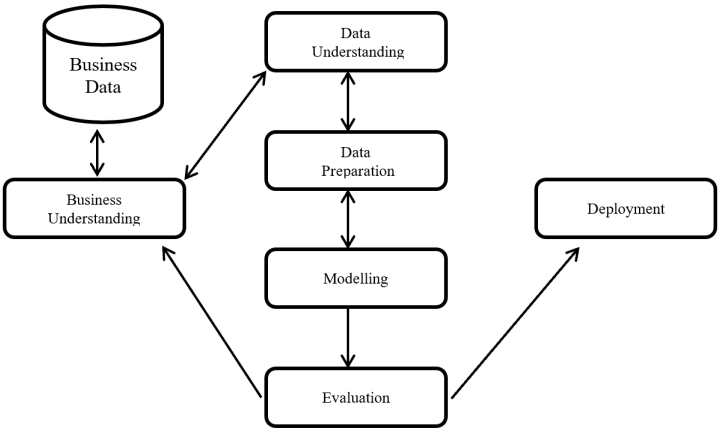


Figure 3.1: CRISP-DM process diagram

On the other hand, Section 3.2 elaborates on the construction of the model frameworks themselves, delineating the datasets selected and the architecture implemented in each one to fulfil different needs. Specifically, Subsection 3.2.1 introduces an innovative framework designed to

emulate the company valuation model, while Subsection 3.2.2 focuses on alternative models employing a conventional regression approach with varied datasets.

The interplay between these elements is visually represented in Figure 3.2, elucidating the interconnectedness of the data mining process. The green highlighted area indicates the scope of the frameworks, where A, B, C, and D represent four distinct generic frameworks. The white sections depict the processes aligned with the CRISP-DM/data mining methodology. The data mining process acts as an input to the models after the data treatment stage, while the modelling and evaluation stages are inherently inside each model.

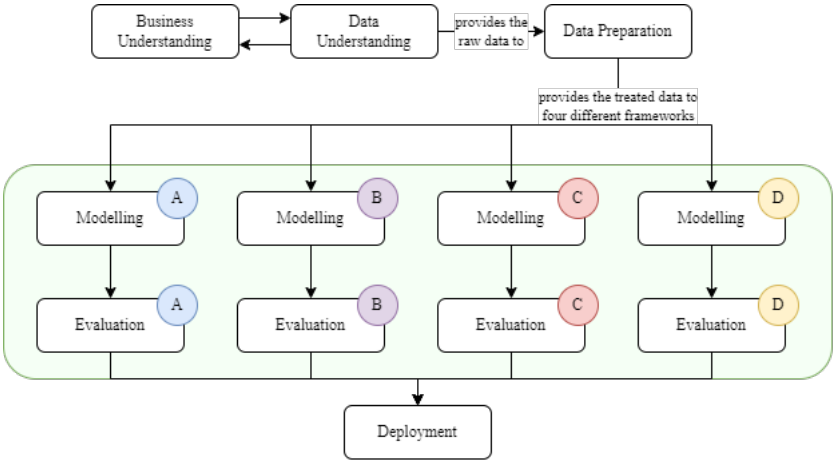


Figure 3.2: Data mining methodology and frameworks interconnectedness diagram (interplay between CRISP-DM and four generic frameworks)

3.1 Machine learning process

The widely adopted CRISP-DM approach is followed in this endeavour. CRISP-DM divides the process into six distinct sections, each serving a crucial purpose in the workflow. These include business understanding, data understanding, data preparation, modelling, evaluation, and deployment. By following this structured methodology, the author ensures that all essential aspects of the machine learning process are covered, leading to effective and reliable results. Figure 3.1 illustrates the CRISP-DM process.

The CRISP-DM is thoroughly applied, encompassing the following sections:

- **Business Understanding:** Chapter 1 provides an overview of the industry context and business and the objectives determined from a business and scientific standpoint.
- **Data understanding:** Subsection 3.1.1 outlines the processes required to collect the data while also understanding it thoroughly through assessing the company characteristics relevancy based on literature review and computing the intermediate valuations.
- **Data preparation:** In Subsection 3.1.2, the main characteristics of the dataset are thoroughly analysed, and tailored techniques are applied to address its unique properties. This

stage involves data preprocessing, cleaning, normalisation, and other relevant transformations to ensure the dataset is suitable for further analysis.

– **Modelling:** Subsection 3.1.3 provides an overview of the possible algorithms, explaining their theoretical foundations and applicability to the problem.

– **Model evaluation:** Subsection 3.1.4 delves into the tools to validate and compare model performance. It also examines the importance of the selected features in predicting the desired outcomes.

– **Deployment:** While not explored in this thesis, the model is put into practice in Apollonian's internal systems in beta testing by executing the R script in VBA inserted in an Excel sheet standardly used to perform the valuation.

3.1.1 Data understanding

This section discusses the data collection process for two datasets: company characteristics and intermediate valuations. The collection of these datasets necessitated the implementation of different methodologies due to their unique requirements.

Apollonian's operational database contains the company characteristics data. However, using this data directly is risky due to the potential for corruption. The operational database prioritises real-time transactional processing and lacks the necessary data structure, historical information, and analytical capabilities for machine learning tasks.

Instead, Data Warehouses (DWs) are purpose-built for analytical tasks. They offer a denormalised or dimensional data structure, which allows for quick querying, aggregation, and exploration of large volumes of data. These attributes make DWs more suitable for machine learning tasks that require in-depth analysis, scalability, and the ability to train models on comprehensive datasets. Therefore, establishing a DW is essential to meet these requirements effectively. A Data Staging Area (DSA) is constructed to ensure data safety, containing a copy of the relevant information from the current operational database. Figure 3.3 demonstrates the architecture behind the implemented system.

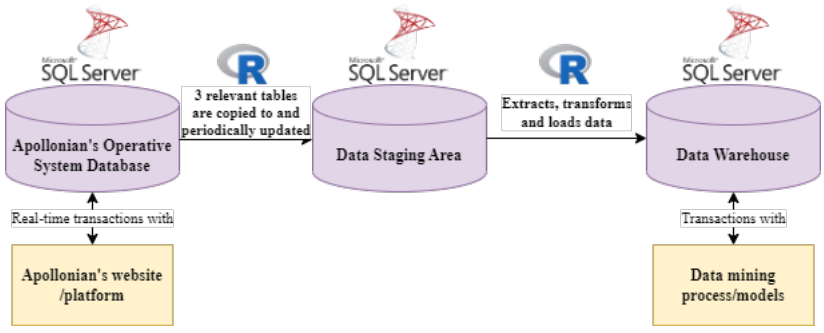


Figure 3.3: Data warehouse architecture with data staging area diagram

The data source provided by Apollonian is a SQL Server used to power the company's web platform. SQL Server Management Studio is used to access this source. Only three tables from this source are useful, and for the DSA, three similar auxiliary tables are created to manipulate the data without jeopardising the safety of the operational database. These tables' specific fields and names cannot be disclosed due to privacy concerns.

The star schema, a common model in data warehousing, is applied in this case. The DSA feeds this data model, as shown in Figure 3.3. The star schema organises data into a central fact table surrounded by multiple dimension tables. The fact table serves as a common denominator, while the dimension tables provide descriptive attributes that are logically linked or categorised. In this instance, the fact table revolves around each company, and the linked dimensions represent client characteristics logically grouped into divisions such as financial or core sections, which Apollonian already used in other circumstances. R is used to create and update the DSA and DW, ensuring the data is up-to-date and accurate whenever significant changes occur in the operational database. Figure 3.4 reports on the intermediate data model (before the JSON files treatment), illustrating a good example of the star model.

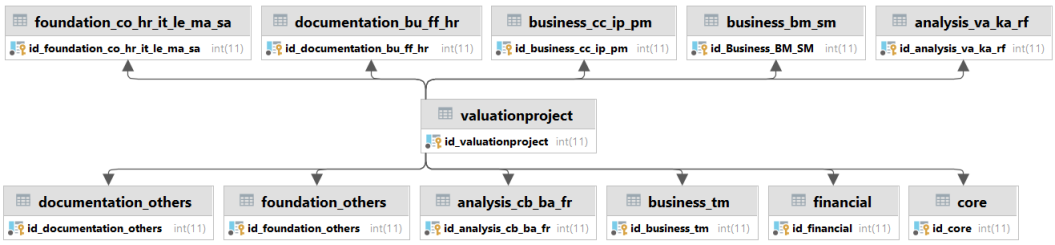


Figure 3.4: Data warehouse data model before JSON files transformation (Star model)

During the data collection phase, JSON files are encountered as observations within the DW. An automated R Script is developed to handle these files efficiently. The script decodes the JSON files and creates new tables linked to specific fields based on their structure, resulting in a snowflake-like data model. This model involves normalising the dimension tables by splitting them into additional tables, which leads to a more complex structure. By analysing the JSON files, the script identifies common column names derived from the first set of keys and populates the tables with the remaining data. Figure 3.5 illustrates the final data model, which includes a fact sheet linked to 11 other tables, as the star model. Due to the JSON files, two of these 11 tables are linked and dependent on 12 and 19 tables generated from those files.

A direct connection is established to the SQL Server hosting the data warehouse to ensure the freshness of the data for the subsequent data treatment phase. This connection eliminates reliance on local files and allows the script to access updated data whenever loaded. A sample of the company database is taken as the working dataset for company characteristics, consisting of 73 observations and 1308 variables. This composition reflects the significant dimensionality of the dataset despite the relatively small number of observations, and it was selected as a representative sample due to confidentiality restrictions.

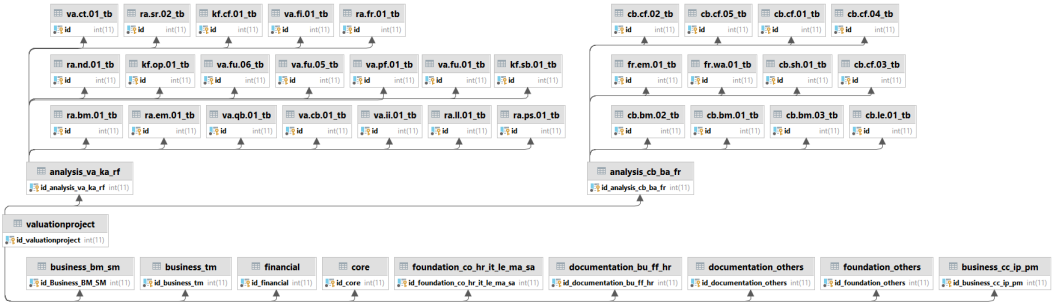


Figure 3.5: Final data warehouse data model (Snowflake model)

When analysing the data, a cautious approach is adopted to address the confidentiality requirements imposed by the company and maintain the privacy of proprietary data. Instead of directly revealing and characterising the company characteristics, only the categories they belong to are presented. This approach prevents the easy inference of the proprietary data collection methods while still providing valuable insights. Moreover, an extensive list of categories increases granularity and maintains comprehensiveness, ensuring a more detailed representation of the predictive factors while respecting confidentiality requirements.

In order to establish well-defined categories and focus solely on pertinent company characteristics, a comprehensive literature review was conducted to identify factors discussed in previous research about firm or asset valuation and investment decision-making. The findings of this extensive search are presented in Table 3.1, which outlines the company characteristics deemed significant based on the literature.

Table 3.1: Summary of relevant company characteristics used in comparable literature

Study	Scope	Relevant company characteristics
Miloud et al. (2012)	Firm valuation	Financial capital and human capital, industry and market, product and technology, attractive industries, high-quality founders and top management teams, strong external relationships
Tumasjan et al. (2021)	Firm valuation	Twitter sentiment
Zheng et al. (2010)	Firm valuation	Network, knowledge sharing
Vayas-Ortega, Soguero Ruiz, et al. (2020)	Firm valuation	Real interest rate (RIR), consumer price index (CPI), gross domestic product (GDP), weighted average cost of capital (WACC)
Dhochak and Doliya (2020)	Firm valuation	Internal resources, industry resources, network-based resources
Garkavenko et al. (2021)	Firm valuation	Age, country, founders, employees, news, funding round, financial history, social network
Blohm et al. (2020)	Firm valuation	Online legitimacy, initial funding and legal form, market timing, founder experiences, industry, founder education
Sharchilev et al. (2018)	Firm valuation	Age, industry, websites, offices, description, products, investor-related factors, team-related factors
Block et al. (2019)	Investment decision-making	Human capital, financial capital, market timing, industry timing, online legitimacy
Mason and Stark (2004)	Investment decision-making	Marketability, market potential, industry features
Tsai et al. (2012)	Intellectual asset valuation	R&D intensity, industry, advertising intensity, family, cash flow rights, cash flow participation, cash flow non-participation, business group, board independence, blockholder, sale growth, size, leverage, capital intensity, dividend, profitability, age, export, analyst following, concentration, market share
Block et al. (2019)	Investment decision-making	Revenue growth, product/service value-added, management team track record, profitability
Maxwell et al. (2011)	Investment decision-making	Product adoption, status, protectability, customer engagement, route to market, market potential, relevant experience, financial model
Csaszar et al. (2006)	Investment decision-making	Strategy, team and finance

These company characteristics are ordered, and logical categories can be inferred. The 1308 variables fit into these categories, resulting in the distribution shown in Table 3.2. This provides a

clear reference and understanding of the relationship between the literature-based variables, their corresponding categories, and their representation in the initial dataset.

Intermediate valuations are derived from the company characteristics to obtain a comprehensive understanding of company valuations. The calculations outlined in Section 2.1 are implemented on the company characteristics dataset to compute the intermediate valuations. The chosen valuation methodologies include the Checklist Method, Benchmark, Book Value, DCF, Terminal Growth, VAIC, Balanced Scorecard, Total Cost of Ownership, Economic Value Added, Cash Value Added and Graham's IV. These methodologies are selected based on the availability of required variables in the dataset, their relevance (common usage and scientific approval), and to ensure diversity in the valuation approaches. Each company has a unique combination of valuation methodologies since some were impossible to compute for some observations due to the absence of necessary variables. This dataset features 11 variables (number of intermediate valuations) and 73 observations (one for each company).

3.1.2 Data preparation

In this section, the data is processed and transformed to facilitate the execution of the models. The main dataset (company characteristics) presented several challenging aspects that require careful treatment. Firstly, the dataset exhibits high sparsity, necessitating the handling of missing values as a top priority in this phase. Additionally, the high dimensionality of the dataset is another important characteristic addressed in the subsequent subsection. Another aspect is the low number of observations, which had the potential to impact the quality and performance of the models. The data treatment process and its effects on the dataset composition can be consulted in Figure 3.6, in which characteristics refer to company characteristics, which are the variables, and the observations refer to the companies themselves.

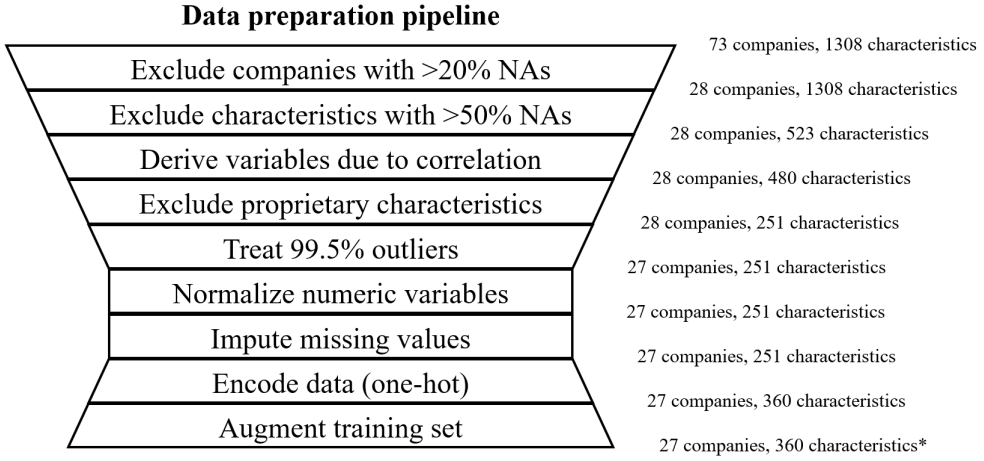


Figure 3.6: Data preparation pipeline and its effect on the dataset composition

Referring to the first aspect, Figure 3.7 illustrates the data sparsity by displaying the distribution of missing values as a percentage for each characteristic (variable) and company (observation).

Table 3.2: Company characteristics categories and their representativity in the dataset after data collection

Categories deducted from literature	**Factors present in literature**	**Nº factors in dataset**
Financial Metrics	Profitability; Revenue growth; Assets under management; Cash-on-cash multiples	228
Market and Industry Factors	Market timing; Industry timing; Market potential; Product and technology; Concentration; Market share; Industry structure Differentiation; Industry growth; Google Trends Index industry; Google Trends Index product	53
Team and Management	Management team track record; Internal resources; Industry resources; Network-based resources; Strategy; Founder experiences; Founder education; Founder/TMT Industrial experience; Management experience; Startup experience; Solo/team dummy; TMT completeness	53
Venture Characteristics	Developmental stage; Legal form; Year of founding; B2C/B2B; Industry specialisation; Funding round	282
Education and Experience	MBA degree; Bachelor's degree; Master's degree; Technical education; PhD degree; Managerial experience; Technical experience; Entrepreneurial experience; Consulting experience; Investment banking experience	43
Founder and Team Attributes	Entrepreneur's personality; Entrepreneur's experience; Demonstrated leadership in past; Evaluates and reacts to risk well; At least ten times return in 5-10 years; Articulates venture well	17
Funding and Financial History	Equity capital before BA funding; Financial history	18
Legal and Regulatory Factors	Legal issues; Initial funding and legal form; Government; Foreign investor	8
Market and Competitive Advantage	Size of the market; Sustainability of competitive advantage; Value added of product/service; International scalability	45
Investor and Funding Factors	Investors; Capital availability; Motivation; Sensitivity of scenarios; Investor Fit; Business Plan	32
Company and Employee Factors	Country; Employees; Current investors; Operations; Product/Service; Business Model; Cash flow rights; Cash flow participation; Cash flow nonparticipation; Business group; Board independence; Blockholder; Multi control; Foreign listing	173
Corporate Governance	Board size; Board independence	39
Founder and Ownership Structure	Management owners; Pyramids	65
Risk and Performance Indicators	Leverage; Capital intensity; Dividend; Age	5
Market and Customer Factors	Export; Analyst Following; Market share; Twitter sentiment; Social media presence; No. of tweets; No. of retweets; Negative emotions; Positive emotions; No. of tweet languages; Social network Network size	44
Product and Service Factors	Value offer; Other assets	90
Founder and Team Education	Business education; Uni top business; Uni top tech	6
Intellectual Property	Protectability; Proprietary protection	38
Investor Perception	Route to market; Relevant experience; Financial model	7
Economic Factors	Real interest rate; Consumer price index; Gross domestic product	63
Total		**1308**

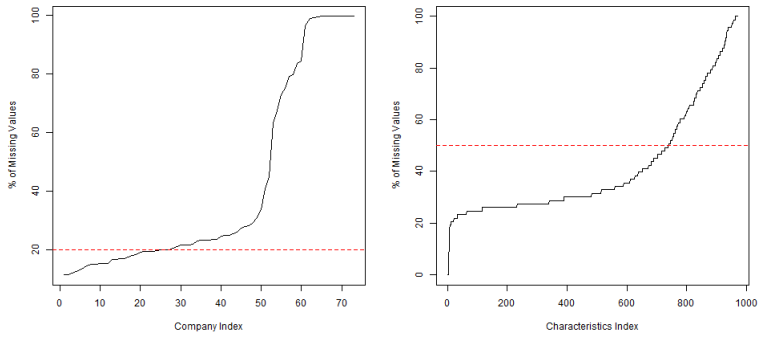


Figure 3.7: Distribution of missing values as a % of each characteristic and of each company

A two-step approach is adopted to address data sparsity. Initially, companies with more than 20% missing values are eliminated from the dataset as most of them are either experiments that were done in the server or companies from an older platform that does not share the same fields and cannot be used. Subsequently, variables with more than 50% missing values were also excluded to ensure consistency and account for clear differences in data quality. The analysis of Figure 3.7 suggests the possibility of setting a more reasonable threshold of around 30% for eliminating companies. However, if the column threshold is maintained, it would lead to the elimination of crucial (financial) variables. On the other hand, lowering the second threshold to accommodate missing values in the data would negatively impact the model's performance, as it can be seen in the sensitivity analysis presented in Figure 3.8, which relates the two thresholds looking for the combination that best enhances performance (lower MAPE value is preferred).

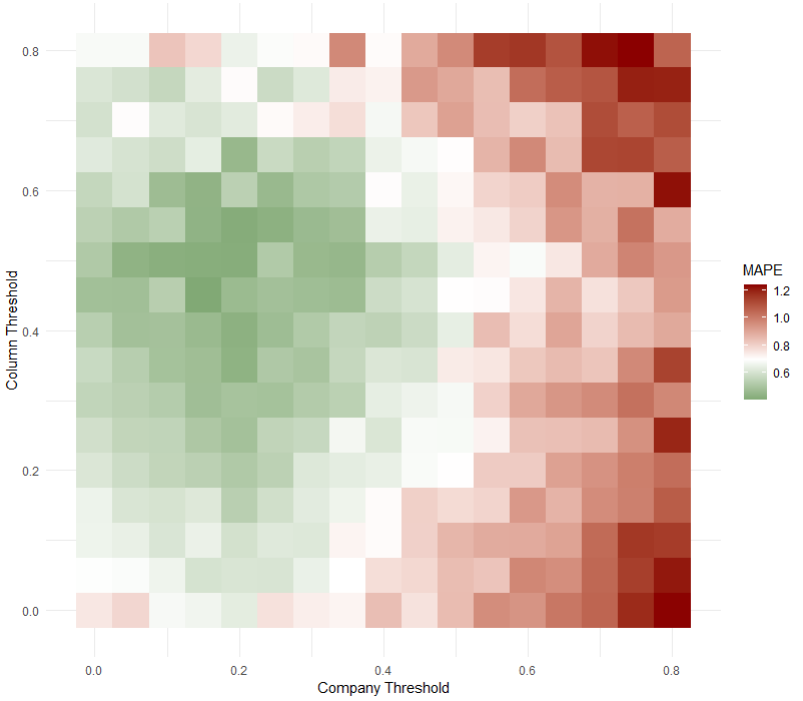


Figure 3.8: Sensitivity matrix of the thresholds for eliminating missing values

After these steps, only high-quality columns are retained, considering the high dimensionality of the dataset. As a result of prioritising data quality, the already limited number of observations was further reduced from 73 to 28 companies, and the number of company characteristics was reduced from 1308 to 523.

Correlation analysis uses the Kendall rank correlation coefficient for comparing ordered variables and numerical variables; the correlation coefficient is used for comparing numerical variables; Cramér's V is used to compare nominal variables; and Polychoric correlation is used for comparing ordered variables techniques. However, since the chosen algorithm incorporates strong feature selection, excluding variables based on correlation is deemed unnecessary. Nevertheless, some variables are merged when they exhibit similar scales and complementary meanings for the same end. This allowed for the removal of 43 variables, going from 523 to 480. Moreover, considering that certain data points are derived using proprietary formulas by the company and others are intermediary valuations, these variables are likewise excluded from the dataset. This reduced the number of variables from 480 to 251.

Outlier treatment is implemented using a 99.5% threshold to avoid excessively manipulating the values while addressing extreme outliers, causing the removal of one observation, which exhibited a significant 7-sigma deviation from the mean of the target variable. Depending on their distribution, numeric variables are normalised using either z-score (outperformed min-max normalisation) or logarithmic normalisation. After this, the remaining missing values are imputed using the MissForest, an imputation method that utilises a random forest algorithm to estimate missing values in a dataset by using the observed values and other variables as predictors. Alternative approaches, such as imputing with the mode (categorical variables), 0 and median (numerical variables), or via k-nearest neighbours (KNN), are attempted but yield inferior performance, namely, KNN, mode and median, and mode and 0 yielded an 18%, 34% and 38% superior error, respectively, on average, compared to the best model's MAPE.

One-hot encoding is applied to the categorical variables, as it outperformed label encoding in terms of model performance on average by 54%, compared to the best model. Although ordered factors are initially considered for variables represented as scales, they are helpful for only one algorithm, which is ultimately outperformed, leading to the abandonment of this idea.

Recognising the limitations imposed by the low number of observations in the dataset, various techniques are employed to mitigate its impact. While L2 regularisation techniques, specifically ridge regression, are explored, they do not yield significant improvements for the dataset, as it is possible to analyse in Section 4.1. As an alternative approach, data augmentation is implemented on the training set by introducing variations in the numerical values of existing observations. This technique proved effective and resulted in meaningful outcomes. The training set is directly augmented during the modelling phase by doubling the number of observations. This is achieved by adding a random number, following a distribution with a mean of 0 and a standard deviation of 0.1 times the original standard deviation, to each existing observation. This process introduces noise and creates new observations. It is worth noting that no changes are made to categorical variables, which represent approximately 42% of the dataset's variables, as applying this technique to such

variables is generally not recommended. The number of observations to be added and the standard deviation multiple to be used was decided based on a sensitivity analysis, presented in Figure 3.9.

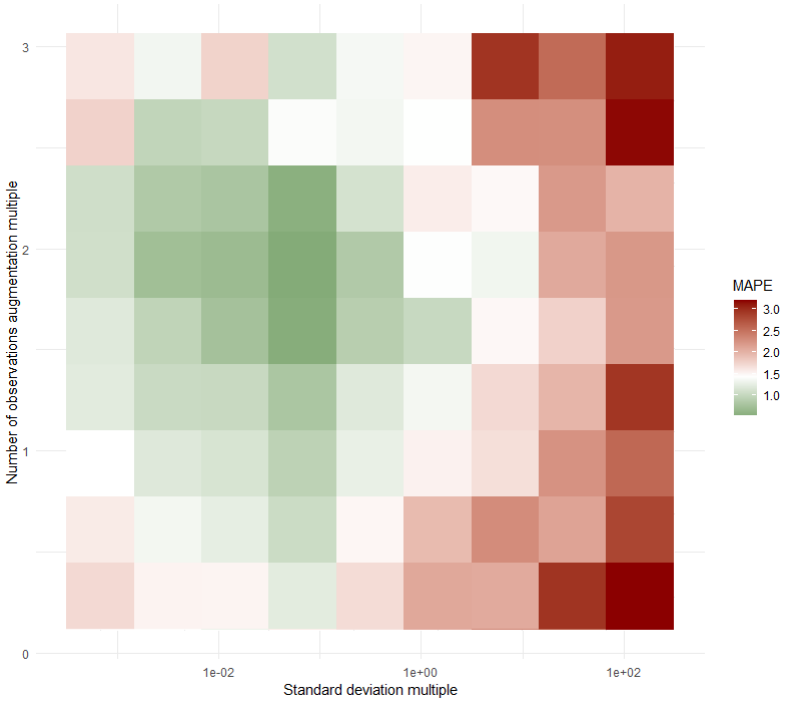


Figure 3.9: Sensitivity matrix of the thresholds for data augmentation

Ultimately, the dataset comprises 27 observations and 360 variables - 251 if not accounting for one-hot encoding (250 company characteristics). Table 3.3 reports on the used variables after data treatment.

Valuation methodologies are also influenced by data treatment, being that it is necessary only to use the observations that are used in the company characteristics dataset (since the latter comprises the target variable). Furthermore, as reported in Subsection 3.1.1, this dataset is sparse, and Table 3.4 provides a visual representation of the sparsity and distribution of the valuation methodologies illustrating the frequency and category of each methodology.

Figure 3.10 illustrates the distribution of intermediate valuations compared to the actual valuation after data treatment using a logarithmic scale. This figure effectively displays the significant variability in results obtained from different valuation methodologies, emphasising the necessity for comprehensive company valuations. Furthermore, it is noticeable that most valuation methodologies tend to undervalue the company's value, which may be due to unfit methodologies or inflated valuations. Solely Graham's IV and Cash Value Added overvalue, at times, the company.

For the missing values, in order for the model to recognise a non-usable valuation methodology, an absurd number (-9×10^{13}) is imputed. This proved more beneficial than imputing 0 since this value was possible in certain algorithms, or the median, which proved a worse performance by 76% and 103%, respectively, compared to the best model.

Table 3.3: Company characteristics categories and their representativity in the dataset after data preparation

Categories deducted from literature	**Factors present in literature**	**Nº factors in dataset**
Financial Metrics	Profitability; Revenue growth; Assets under management; Cash-on-cash multiples	105
Market and Industry Factors	Market timing; Industry timing; Market potential; Product and technology; Concentration; Market share; Industry structure Differentiation; Industry growth; Google Trends Index industry; Google Trends Index product	12
Team and Management	Management team track record; Internal resources; Industry resources; Network-based resources; Strategy; Founder experiences; Founder education; Founder/TMT Industrial experience; Management experience; Startup experience; Solo/team dummy; TMT completeness	9
Venture Characteristics	Developmental stage; Legal form; Year of founding; B2C/B2B; Industry specialisation; Funding round	10
Education and Experience	MBA degree; Bachelor's degree; Master's degree; Technical education; PhD degree; Managerial experience; Technical experience; Entrepreneurial experience; Consulting experience; Investment banking experience	3
Founder and Team Attributes	Entrepreneur's personality; Entrepreneur's experience; Demonstrated leadership in past; Evaluates and reacts to risk well; At least ten times return in 5-10 years; Articulates venture well	5
Funding and Financial History	Equity capital before BA funding; Financial history	13
Legal and Regulatory Factors	Legal issues; Initial funding and legal form; Government; Foreign investor	5
Market and Competitive Advantage	Size of the market; Sustainability of competitive advantage; Value added of product/service; International scalability	5
Investor and Funding Factors	Investors; Capital availability; Motivation; Sensitivity of scenarios; Investor Fit; Business Plan	6
Company and Employee Factors	Country; Employees; Current investors; Operations; Product/Service; Business Model; Cash flow rights; Cash flow participation; Cash flow nonparticipation; Business group; Board independence; Blockholder; Multi control; Foreign listing	27
Corporate Governance	Board size; Board independence	7
Founder and Ownership Structure	Management owners; Pyramids	15
Risk and Performance Indicators	Leverage; Capital intensity; Dividend; Age	1
Market and Customer Factors	Export; Analyst Following; Market share; Twitter sentiment; Social media presence; No. of tweets; No. of retweets; Negative emotions; Positive emotions; No. of tweet languages; Social network Network size	9
Product and Service Factors	Value offer; Other assets	8
Founder and Team Education	Business education; Uni top business; Uni top tech	1
Intellectual Property	Protectability; Proprietary protection	5
Investor Perception	Route to market; Relevant experience; Financial model	2
Economic Factors	Real interest rate; Consumer price index; Gross domestic product	2
Total		**250**

Table 3.4: Statistical summary of the intermediate valuations dataset after data preparation

Category	Valuation Methodology	Median (millions)	σ (millions)	% NAs
Balance Sheet	Book Value	0.34	25.40	3.70%
Income Statement	Graham's Intrinsic Value	(1,975.85)	9,817.60	18.52%
Multiples	Benchmark	334.16	8,033.17	62.96%
DCF	Discounted Cash Flows	16.31	93.19	44.44%
	Terminal Growth	(34.77)	95.26	77.78%
Value Creation	Economic Value Added	(0.38)	14.01	66.67%
	Cash Value Added	16.94	284.90	81,49%
	Total Cost of Ownership	14.64	40.55	0.00%
Scorecard	Balance Scorecard	7.82	201.28	7.40%
	Checklist Method	0.81	14.24	7.40%
Direct Intellectual Capital	VAIC	60.14	121.60	0.00%
Real Valuation		211.02	2171.42	0.00%

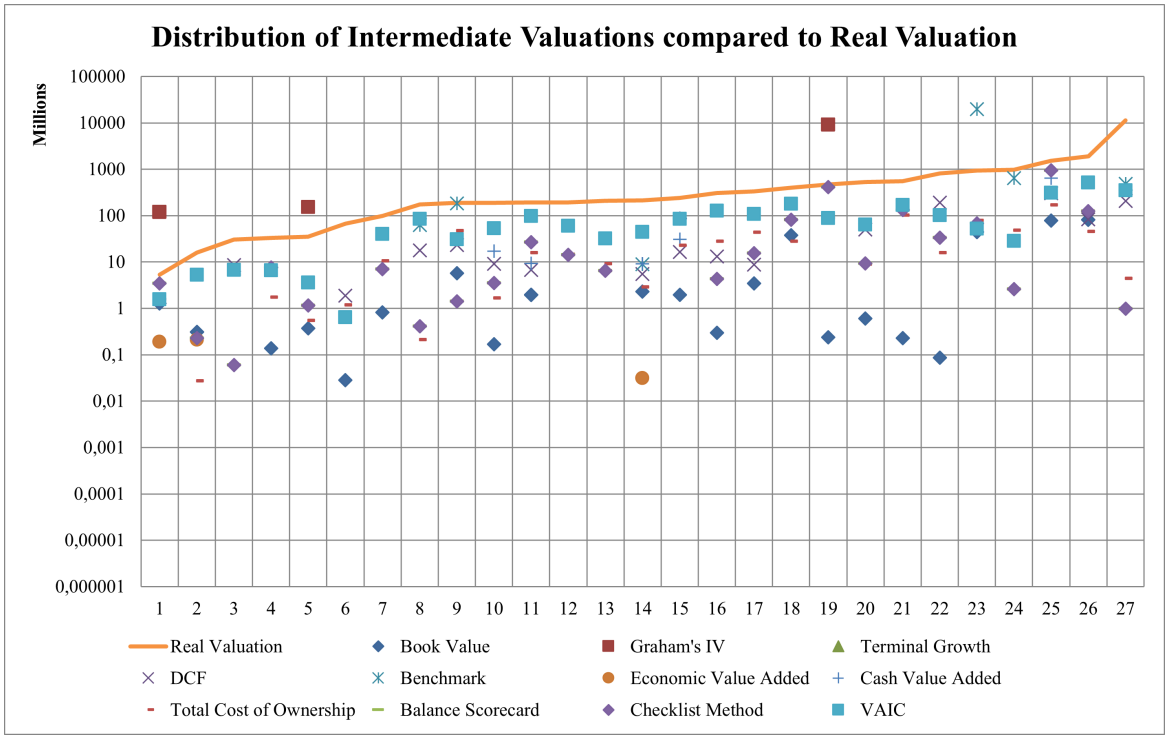


Figure 3.10: Distribution of intermediate valuations compared to real valuation line chart

3.1.3 Modelling

Many ML algorithms have been applied in the finance/business industry. The scope of the research has a significant impact on choosing the appropriate algorithm.

Table 3.5: Summary of the ML algorithms and their usage in comparable studies

Algorithm	Studies
Decision trees-based	Haralampiev et al. (2014), Futagami et al. (2021),Vayas-Ortega, Soguero Ruiz, et al. (2020), Friedman (2000), T. Chen and Guestrin (2016), Ke et al. (2017), Dorogush et al. (2017)
Bayesian-based	Futagami et al. (2021)
Neural Network-based	Lu, Lee, and Chiu (2009), Tay and Cao (2001), Vanstone and Finnie (2009), W.-S. Chen and Du (2009), Amorim Neto, Tavares, Alves, Cavalcanti, and Ing Ren (2010)
Generalized Linear Models	Ou and Penman (1989), Littunen and Niittykangas (2010)

In the modeling stage, the data is partitioned using K-Fold Cross Validation, adhering to the standard practice of utilising ten folds, despite the challenge of a limited number of observations. While this may lead to a test set with a small number of observations per fold, the use of multiple folds helps to mitigate the volatility caused by individual poorly calculated observations. By adopting this approach, the model's performance is evaluated more robustly, effectively addressing the impact of limited data.

The performance of various commonly used algorithms in the field is evaluated, with an overview of these algorithms and their corresponding studies provided in Table 3.5. The analysis reveals that decision tree-based and neural network-based algorithms are prevalent in the literature. Therefore, the first category of algorithms, comprising random forest, C5.0, and XGBoost (XGB), will be employed, while the latter category will be represented by neural networks, which are also used in the innovative framework described in Subsection 3.2.1. Additionally, kernel, regularisation, and instance-based methods that were not present in comparative literature are explored through the use of support vector machines (SVM), ridge regression (RR), neural networks (NN), and K-nearest neighbours (KNN).

In order to better understand the capabilities and fit of these algorithms, Table 3.6 describes them thoroughly and elaborates on how they offer diverse approaches to address different types of data and modelling requirements.

During the modelling phase, selecting appropriate algorithms to assess variable contributions and enhance explainability is crucial. For this reason, the LIME approach was employed over the SHapley Additive exPlanations (SHAP) technique. Only these two were considered due to being highly applied in ML, given that Belle and Papantonis (2021) concluded that these were the most used model-agnostic techniques for explainability problems.

SHAP and LIME are renowned methodologies for interpreting predictions derived from machine learning models and understanding the significance of variables within those predictions.

SHAP, based on cooperative game theory, establishes a comprehensive framework for explaining the output of any machine learning model. It quantifies the individual contribution of each feature in a prediction using Shapley values derived from cooperative game theory. SHAP

Table 3.6: Definition and critical analysis of regression algorithms

Algorithm	Description	Strengths	Weaknesses	Capabilities
SVM	A powerful regression algorithm that finds an optimal hyperplane to minimise the error between predicted and actual values. It effectively handles both linear and nonlinear relationships in regression tasks.	Handles high-dimensional data, robust to outliers, suitable for both linear and nonlinear relationships.	Requires careful hyperparameter tuning, which can be computationally expensive for large datasets.	Deals with high-dimensional data and complex relationships.
Ridge Regression	A linear regression technique that adds an L2 regularisation term to the cost function, reducing multicollinearity impact and preventing overfitting.	Works well with correlated features and stabilises model coefficients. Features intrinsic interpretability.	Assumes linear relationships may not perform well with nonlinear relationships. The choice of regularisation parameter is important.	Suitable for scenarios with correlated features and helps prevent overfitting.
Neural Network	Algorithm inspired by the structure and function of the human brain, capable of learning complex patterns and relationships	Captures nonlinearities and interactions, high flexibility.	Requires a large amount of data, computationally intensive, and can overfit with limited data.	Captures complex nonlinear relationships, often used with large amounts of data.
KNN	A non-parametric regression algorithm that predicts the value of a data point based on the average of its k-nearest neighbours.	Simple to implement, works well with small and noisy datasets. Features intrinsic interpretability.	Computationally expensive for large datasets, choice of k and distance metric impacts performance.	Performs well with small and noisy datasets and captures local patterns.
Random Forest	An ensemble learning method that combines multiple decision trees to make predictions in regression tasks.	Handles high-dimensional data, outliers, and missing values. Robust to overfitting.	Slow for large datasets, may not perform well with highly imbalanced data. Lower interpretability compared to single decision trees.	Provides robustness to outliers and missing values.
C5.0	A regression algorithm that constructs a decision tree by recursively splitting the data based on features.	Handles missing values and captures complex interactions between features. Features intrinsic interpretability.	Can overfit the training data, complex tree structure difficult to interpret.	Offers interpretability, handles numerical and categorical features.
XGBoost	An ensemble learning method that combines multiple weak learners (decision trees) to make predictions in regression tasks.	Highly efficient, handles large datasets with high-dimensional features. Provides variable importance measures.	Requires careful hyperparameter tuning, which can be computationally intensive for large datasets.	Suitable for large-scale regression tasks, combines the power of ensemble learning with speed and efficiency.

values measure the influence of a feature's value on the discrepancy between the predicted output and the model's average output.

In contrast, LIME approximates the behaviour of complex models by fitting a locally interpretable and simplified model around a specific instance of interest. This involves generating "perturbed" versions of the instance through random feature modifications. A simplified model is then trained using these perturbed instances to provide explanations based on the introduced perturbations.

The selection of LIME over SHAP was driven by SHAP's limitation in providing localised explanations for individual predictions. LIME captures the nuanced behaviour of the model near a specific instance, unravelling the distinctive contributions of each feature to that particular prediction. Conversely, SHAP provides a global explanation by considering the average impact of each feature across all possible instances.

3.1.4 Evaluation

The reasoning behind the metrics employed in studies is generally not well understood in the literature. Notwithstanding this, in Vayas-Ortega, Soguero Ruiz, et al. (2020)'s study, the models' performance is assessed using the mean and Standard Deviation of the Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), and D-index of agreement.

The evaluation metrics selected for regression analysis in this study are R^2 (coefficient of determination) and MAPE (Mean Absolute Percentage Error). These metrics are chosen because

they provide relative measures, particularly valuable when assessing data with high absolute values. Moreover, these metrics differ in computation and have gained widespread acceptance in the literature.

Furthermore, MAPE is derived from MAE and offers an advantage over RMSE as it reduces the influence of outliers. This characteristic is important considering the high variance present in the target variable. Both R^2 and MAPE are scale-invariant, which is advantageous given the dataset's characteristics.

However, it is important to be cautious with the interpretation of R^2, as it tends to increase even with the addition of irrelevant independent variables. This issue highlights the need to consider the meaningfulness of any newly added independent variables carefully. On the other hand, MAPE may not perform well for small or near-zero values, but this is not the case in our dataset. K-fold cross-validation also helps prevent overfitting by providing a more realistic estimate of a model's performance on unseen data. It trains the model on multiple subsets of the data, ensuring better generalisation and model selection.

In terms of explainability, to assess the significance of features in a Machine Learning Model, two crucial metrics play a vital role in understanding the employed methods: %IncMSE and %IncNodePurity.

%IncMSE, considered the most robust and informative measure, quantifies the increase in MSE of predictions (estimated with out-of-bag cross-validation) resulting from permuting variable j, wherein the values are randomly shuffled. A higher %IncMSE value indicates greater importance of the variable in the model.

On the other hand, IncNodePurity is associated with selecting the best splits based on the loss function. In regression, the loss function is MSE, while in classification, it is gini impurity. More useful variables yield higher increases in node purity, indicating the identification of splits that exhibit substantial inter-node variance and minimal intra-node variance.

Since the target variable in this context is continuous, it is important to choose a suitable metric for assessing variable importance. While IncNodePurity may not handle continuous variables effectively, %IncMSE (Percentage Increase in Mean Squared Error) can serve as a reliable metric in this scenario. Variables will be considered statistically significant if their corresponding *p*-values derived from %IncMSE fall below the desired significance level.

3.2 Proposed valuation models

As discussed in Chapter 1, the venture capital (VC) industry faces challenges in implementing ML models, which are not yet commonly practised due to the model's complexity and lack of proven performance, as well as the industry's resistance to change. This study aims not only to develop models that outperform the industry average (50% of MAPE) but also to promote the adoption of these methods in the VC landscape. In Subsection 3.2.1, a distinctive model is presented that bridges the gap between the manual weight assignment process and the ML approach,

closely mimicking the actual process. Additionally, it provides the flexibility to impose restrictions on the weights assigned to each valuation methodology, allowing it to align with industry practices. Additionally, Subsection 3.2.2 introduces three different models, which only differ in the inputs used, serving three main purposes: validating the first model by providing direct and reliable performance references, competing to surpass the industry average potentially, and utilising explainability on logic datasets (i.e. limiting the inputs to intermediate valuations, the model avoids the "noise" from the company characteristics dataset when assessing the importance of intermediate valuations). Table 3.7 summarises the four frameworks for better understanding.

Table 3.7: Summary of the frameworks tested by algorithm, inputs and usage

Framework	**Algorithms tested**							**Datasets used**		**Main usage**		
	C5.0	**KNN**	**NN**	**RF**	**RR**	**SVM**	**XGB**	**Company Char.**	**Intermediate Vals.**	**Performance benchmark**	**Explainability**	**Industry process fit**
Regression with intermediate valuations	X	X	X	X	X	X	X		X	X	X	
Regression with company characteristics	X	X	X	X	X	X	X	X		X	X	
Enhanced regression model	X	X	X	X	X	X	X	X	X	X		
Multilayered integrated model with industry restrictions			X					X	X	X		X

3.2.1 Multilayered integrated model with industry restrictions

This section adopts a structured narrative approach to provide a comprehensive understanding of the underlying rationale behind developing this innovative model.

To develop a realistic predicting model architecture, it is essential to align the practical methodology with the architecture of the theoretical model. The valuation of companies involves three crucial dimensions that must be considered: company characteristics, intermediate valuations, and weight restrictions.

Company characteristics: Company characteristics encompass the unique combination of features that are specific to each company. These characteristics directly impact the company's valuation, contributing positively or negatively to the overall value.

Intermediate valuations: the values computed for a company using various valuation methodologies, as discussed in Section 3.1. These valuations can vary significantly and may be positive or negative. Not all valuation methodologies apply to every company; some may be irrelevant depending on the company's stage or industry, while others can be of utmost importance. For example, intangible asset valuation methods are particularly relevant for technology startups, while DCF is more suitable for mature, stable companies.

Weight restrictions: When conducting a comprehensive and multidimensional company valuation, it is important to impose restrictions on the weights assigned to different valuation methodologies. For instance, it is logical to expect that weight should be between 0% and

100%, and the sum of weights should equal 100%. These restrictions ensure a realistic and meaningful distribution of importance among the various valuation methodologies.

Based on these considerations, the framework should follow the normal pipeline of a comprehensive company valuation, starting by choosing the intermediate valuations to use and computing each one of them using company characteristics. Next, weights are assigned to each intermediate valuation based on empirical knowledge and domain expertise. These weights reflect the relative importance of each valuation methodology in the overall company valuation. Finally, the model should compute the final valuation by taking the scalar product of the weights and the intermediate valuations.

By incorporating these dimensions into the model architecture and following this structured pipeline, a tailored approach to company valuation is achieved, ensuring that the resulting valuations capture the complexity and multidimensionality of the real-world scenario.

In the field of data mining, the proposed model architecture aims to satisfy the following requirements:

- **Weight prediction:** Computing intermediate valuations and performing scalar products are objective processes that do not involve subjectivity. However, the determination of weights for these intermediate valuations is a different matter. This task is the scope of the model, being the main source of inaccuracies, which should be fed by the company's characteristics and aimed at predicting the appropriate weights.

- **Absence of training set for weights:** Since no existing training set is available for the vector of weights and is not explicitly known even to domain experts, the model cannot rely on training based solely on the vector of weights. Instead, it must train and make predictions based on the final valuation.

- **Weight restrictions:** To ensure that the model outputs a weight vector that aligns with logical and industry-practice guidelines, weight restrictions are enforced. These restrictions constrain the feasible range of values the weight vector can assume.

  - The weight vector components must satisfy the constraint of being greater than or equal to 0% and less than or equal to 100%.
  - The summation of all weight vector elements should equal 100%.
  - In cases where an intermediate valuation value is missing due to unavailable or incomplete data, the corresponding weight vector component should be assigned a value of 0%.
  - To ensure comprehensive valuation incorporating expert input from Apollonian, no weight in the vector should exceed 40%.

Figure 3.11 illustrates the flowchart of a multilayer integrated model that combines the necessary framework requirements. By adhering to these conditions, the proposed model architecture

aims to overcome the challenges associated with the absence of a training set for the vector of weights. Instead, it leverages intermediate weight valuations and their scalar product with the weight vector to infer the final valuation, thereby facilitating effective training and evaluation of the model. Additionally, weight restrictions add a layer of reliability to the model's outputs, ensuring that the predicted weight vector aligns with established industry practices.

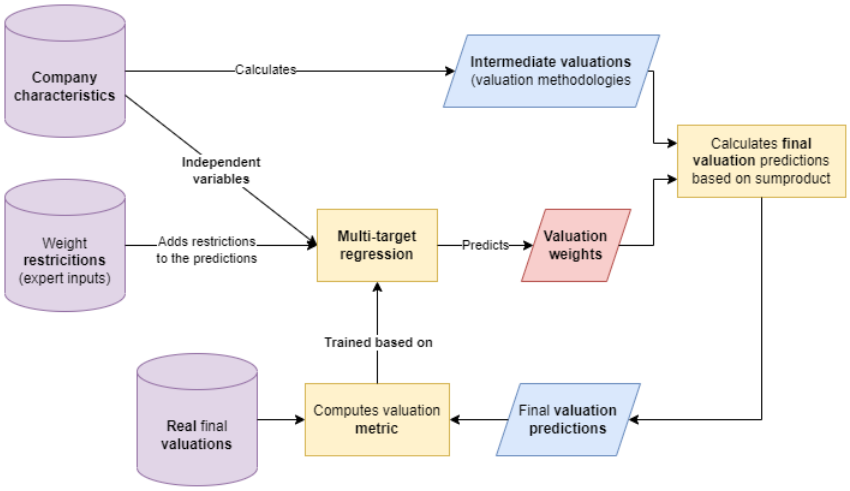


Figure 3.11: Flowchart of the multilayer integrated model with industry restrictions

This research explores two options for algorithms for regression: tree-based algorithms and neural networks.

The objective is to apply the above requirements to the algorithms, and it is found that tree-based algorithms are complex to manipulate to meet these conditions. Manually computing a known tree-based algorithm proved too complex for the given task. The rationale behind the initial consideration of tree-based algorithms is that their nodes could represent company characteristics, while their leaves could represent weight vectors. The algorithm continuously updates the leaf vectors and determines the nodes/combinations leading to these leaves. However, due to the complexity of this approach and the availability of a more suitable alternative, it is decided to explore neural networks instead.

Neural networks have been extensively studied and applied in various domains, including deep learning and meta-learning, which provide a solid foundation for this research. The architecture of the chosen neural network model is outlined as follows, and an illustrative diagram is showcased in Figure 3.12:

1. **Input layer:** The initial layer of the neural network consists of neurons representing the company characteristics featured in the model.

2. **Hidden layers (*2...n-2*):** Optionally, the model can incorporate one or more hidden layers with varying numbers of nodes. The selection and configuration of these hidden layers can be iterated based on the model's performance.

3. ***n-1* layers:**

- **Weight vector layer:** Following the hidden layers, a set of 11 nodes is derived from the combination of the hidden layers. These nodes represent the weight vector with a fixed length of 11 (number of intermediate valuations). In this layer, weight restrictions are imposed to constrain the range of values the weight vector can take.
- **Intermediate valuations layer:** Running in parallel but not directly connected to the neural network, there is a second layer of inputs consisting of 11 nodes representing already computed intermediate valuations.

4. **Output layer:** The output layer comprises a single node. The output is obtained through the scalar product of the last two layers: the 11 nodes representing the weight vector and the 11 nodes representing the intermediate valuations.

A diagram of the complete neural network is reported in Figure 3.12, being n the number of inputs, m the number of nodes in the first hidden layer and l the number of hidden layers.

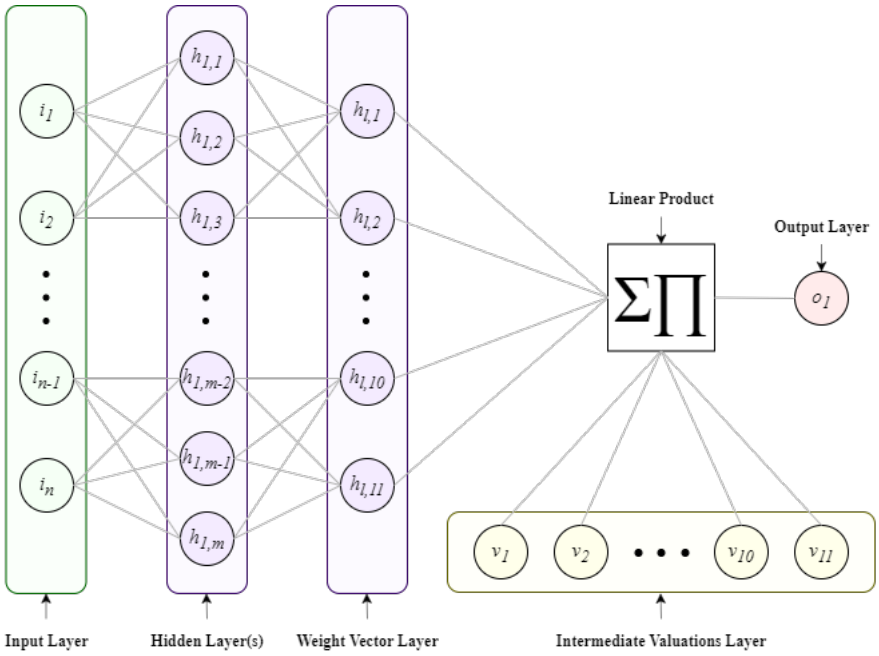


Figure 3.12: Diagram of the multilayer integrated neural network model

The framework presented above offers several distinct advantages that contribute to its overall effectiveness and potential. Firstly, its resemblance to reality reduces adoption resistance, as it aligns closely with real-world scenarios. Secondly, the framework allows for the incorporation of industry-specific restrictions, enabling tailored solutions to address unique challenges. Lastly, its inherent potential for future development ensures adaptability and growth, paving the way for ongoing enhancements and improvements.

While the framework possesses notable advantages, it is important to acknowledge certain limitations it presents. Primarily, the framework's high computational time arises from the intensive nature of training neural networks. Additionally, its data requirements entail the reliance

on large datasets to capture nonlinear relationships effectively. However, these drawbacks can be addressed through strategic mitigation strategies. Optimised usage, such as running the models periodically and utilising saved models for new predictions, can help alleviate the computational burden. Furthermore, implementing data augmentation and regularisation techniques has proven decisive in overcoming the challenges associated with data requirements, as demonstrated in this study.

3.2.2 Comparison of three alternative frameworks

This section introduces a set of regression models designed for specific purposes. The first model utilises intermediate valuations to predict firm value and assess their importance. The second model employs company characteristics to predict the same and evaluate their significance. The third model combines both intermediate valuations and company characteristics to predict value.

In the regression model with intermediate valuation, 11 intermediate valuations serve as predictors, and the final company valuation is the response variable. This framework allows for easily extracting insights into the impact and significance of each intermediate valuation on the final company valuation, independently of other factors. These insights contribute to understanding the overall (global) and company-specific (local) valuation methodologies.

Similarly, in the regression model with company characteristics, 361 characteristics act as predictors for the same target variable. By analysing the contribution and importance of each characteristic to firm valuation, we gain insights into the various valuation methodologies' impact on the overall and company-specific valuation.

However, it is important to note that the enhanced regression model, which combines both company characteristics and intermediate valuations, presents a challenge due to their inherent collinearity. This model is not suitable for assessing the isolated contributions of pure intermediate valuations or company characteristics, as the intermediate valuations are derived from certain company characteristics.

These traditional models offer several advantages compared to the multilayer algorithm. Firstly, they provide a more straightforward implementation process with easy reproducibility, making them easily accessible for practical use. Additionally, two of these models prioritise interpretability, which is not a characteristic that can be directly compared to the multilayer model. However, it's important to note that the two models which do not account for merged datasets may encounter issues with incomplete information. Despite this limitation, the main criticism directed towards these three models is their lack of industry-specific adaptation and their limited ability to offer substantial advantages beyond performance when compared to existing models that have not yet been widely adopted. In order to bridge this gap, it becomes crucial to focus on enhancing contextual adaptation to ensure real-life applicability. By doing so, decision-makers can be brought closer to the tool, enabling more effective and informed decision-making processes.

Chapter 4

Results and discussion

To achieve the research objectives, in Section 4.1, the frameworks are validated by comparing the underlying algorithms and assessing their relative performance against the industry benchmark. This assessment ultimately determines the real-world applicability of the frameworks by highlighting their improved performance compared to existing approaches. This section also acts as a crucial bridge to ensure the composition and quality of the datasets, as the subsequent sections rely on the reliability of the data. If the performance of the models is good, it can be inferred that the data quality is sufficient.

Moving forward, Section 4.2 focuses on individually analysing the contributions of valuation methodologies. It identifies the significant and most important methodologies for the model. Additionally, it examines how these contributions change as the firm value increases or decreases.

Similarly, in Section 4.3, the analysis shifts towards assessing the importance of company characteristics. It explores the significant and absolute most important variables for the model, considering the variations in their importance based on changes in firm value.

Lastly, Section 4.4 delves into the correlation between the importance of company characteristics and the contribution of intermediate valuations. This analysis aims to understand how the importance of valuation methodologies changes with shifts in the importance of specific company characteristics and vice versa.

By systematically examining these aspects, the study aims to provide comprehensive insights into the frameworks, algorithms, data quality, and the interrelationships between different variables, contributing to a deeper understanding of their practical implications.

4.1 Framework validation

This section aims to validate the results obtained from the frameworks discussed in Section 3.2 to assess their viability for real-world implementation. The validation process primarily relies on evaluating the results using the machine learning metrics discussed in Subsection 3.1.4. The validation of a framework goes beyond confirming if the predicted values align with the actual

valuations; it also involves assessing the overall coherence and soundness of the framework's results, namely through assessing its consistency (σ).

This subsection compares the performance of various regression algorithms within three frameworks: the Regression with Intermediate Valuations, the Regression with Company Characteristics and the Enhanced Regression Model. The performance is assessed using two evaluation metrics: the R^2 and MAPE, as discussed in Subsection 3.1.4. Table 4.1 reports on the results of the Regression with Intermediate Valuations. Figure A.1 in Appendix A visually depicts these results.

Table 4.1: Regression with intermediate valuations framework results

Algorithm	**Median MAPE**	**σ MAPE**	**Median R^2**	**σ R^2**
C5.0	**0.456**	1.488	0.810	0.319
XGBoost	0.565	**0.492**	0.946	**0.290**
Random Forest	0.655	1.834	**0.988**	0.298
Ridge Regression	0.749	6.024	0.876	0.433
KNN	0.835	1.539	0.919	0.326
SVM	0.959	7.025	0.940	0.390
Neural Network	1.730	1.900	0.431	0.411

Among the algorithms tested in the Simple Regression Model with an Intermediate Valuations framework, the C5.0 algorithm has the lowest median MAPE. Despite this, it shows a high variance in values and a relatively low R^2 value. On the other hand, the XGBoost is the second best in median MAPE and is second best in median R^2, while showing the lowest σ of the models in the two metrics, suggesting consistent and stable performance while showing a remarkable ability on explaining the variance of the data. Therefore, XGBoost is the best model within this framework, considering its high R^2, low MAPE, and low σ on the two metrics. Overall, the model still under-performs the industry average error of 50% by 13%.

Table 4.2 reports on the Regression with Company Characteristics results. Figure A.2 in Appendix A visually depicts these results.

Table 4.2: Regression with company characteristics framework results

Algorithm	**Median MAPE**	**σ MAPE**	**Median R^2**	**σ R^2**
XGBoost	**0.347**	**0.950**	0.997	**0.104**
C5.0	0.534	9.628	0.844	0.451
KNN	0.705	3.485	0.993	0.371
Neural Network	0.843	2.671	0.980	0.163
SVM	0.868	2.917	0.949	0.313
Ridge Regression	0.952	1.911	0.863	0.400
Random Forest	1.130	3.255	**0.999**	0.310

Within the Simple Regression Model with Company Characteristics framework, the XGBoost consistently achieves the lowest MAPE value, indicating an excellent fit to the data, also featuring the lowest standard deviation. It also demonstrates the second lowest median value and the lowest

standard deviation for R^2, indicating stable and consistent performance. Despite Random Forest featuring the highest R^2, its MAPE is the highest, being an unreliable algorithm for the case. Therefore, the XGBoost algorithm is the best model within this framework, considering its high R^2, low MAPE, and low variability. Overall, the model outperforms the industry average error of 50% by 30%.

Table 4.3 reports on the results of the Enhanced Regression Model. Figure A.3 in Appendix A visually depicts these results.

Table 4.3: Enhanced regression model framework results

Algorithm	**Median MAPE**	**σ MAPE**	**Median R^2**	**σ R^2**
XGBoost	**0.489**	1.678	**0.996**	0.19
C5.0	0.699	1.632	0.985	0.297
SVM	0.703	3.193	*	*
KNN	0.759	**1.592**	0.943	0.454
Random Forest	0.94	2.252	0.95	**0.109**
Neural Network	1.266	1.709	0.811	0.367
Ridge Regression	1.277	4.423	0.935	0.392

Within the Enhanced Regression Model framework, the XGBoost consistently achieves the highest R^2 value, indicating an excellent fit to the data. It also demonstrates the lowest MAPE values and a low SD for R^2, indicating stable and consistent performance. Despite high variance in the MAPE, it is still comparable to other algorithms. Therefore, the XGBoost is the best model within this framework, considering its high R^2, low MAPE, and low standard deviation. Overall, the model outperforms the industry average error of 50% by 2%.

Table 4.4 compares the results of the four frameworks after choosing the best algorithm. Figure A.4 in Appendix A visually depicts these results.

Table 4.4: Comparison of the frameworks results

Framework	**Median MAPE**	**σ MAPE**	**Median R^2**	**σ R^2**
Regression with Intermediate Valuations	0.565	0.492	0.946	0.290
Regression with Company Characteristics	**0.347**	0.950	**0.997**	0.104
Enhanced Regression Model	0.489	1.678	0.996	0.190
Multilayer NN w/ Industry Restrictions	0.535	**0.074**	0.983	**0.002**

Among the four compared frameworks, the Regression with Company Characteristics exhibits superior performance in both metrics, demonstrating greater consistency across the board. However, it is important to note that no statistical evidence has been found to establish one framework's superiority over the others using the t-test (due to the small sample size).

In conclusion, the performance of the regression models in the four frameworks can be considered valid. This conclusion is supported by the achieved R^2 values exceeding 0.8, indicating that the models explain a substantial portion of the variance in the data. Although the MAPE

values vary among the algorithms, they fall within an acceptable range of the industry benchmark. Despite this, only the Regression with Company Characteristics and Enhanced Regression frameworks are better than the industry average of 50%.

The Multilayer Neural Network, while a validated and competitive model, does not exhibit performance improvements for this specific dataset. However, its performance is significantly better when compared to simpler and enhanced regression models with the NN algorithm, highlighting its potential for other datasets.

Considering these findings, the Regression with Company Characteristics model demonstrates a significant improvement of 30% compared to the industry benchmark. This notable performance enhancement makes it a compelling option for potential adoption within the industry. On the other hand, the Multilayer NN model exhibits immense potential for future research, testing, and improvement. Its strong alignment with industry requirements and flexibility for incorporating specific restrictions make it a promising avenue for further exploration and development, showing initial promising results.

4.2 Valuation methodology importance analysis

Building upon the elucidation of the metrics for assessing variable importance in Subsection 3.1.4, examining the overall variable importance of the valuation methodologies becomes paramount. Table 4.5 presents the %IncMSE and IncNodePurity values for each valuation methodology alongside their respective p-values.

Table 4.5: Summary of the valuation methodologies importance

Valuation Methodology	**%IncMSE**	***p*-value** (%IncMSE)	**IncNodePurity**	***p*-value** (IncNodePurity)
Total Cost of Ownership	**38.11**	0.70%	6.63E+17	22.08%
Book Value	31.12	1.40%	**1.01E+18**	5.39%
Checklist Method	29.00	2.20%	6.21E+17	16.78%
VAIC	27.09	2.60%	8.87E+17	9.99%
Graham's IV	26.67	2.70%	3.37E+17	82.02%
DCF	26.20	2.10%	6.31E+17	10.59%
Balanced Scorecard	25.15	5.00%	4.62E+17	62.64%
Benchmark	16.01	3.10%	2.08E+17	24.08%
Economic Value Added	6.30	19.68%	5.12E+16	99.30%
Terminal Growth	0.99	26.97%	4.61E+16	99.30%
Cash Value Added	-1.15	39.56%	8.26E+16	48.25%

Upon close observation, it becomes evident that the Total Cost of Ownership emerges as the most significant valuation methodology, as indicated by the %IncMSE metric, which is considered the primary indicator. Among the valuation methodologies, those deemed statistically significant at a 5% significance level are Total Cost of Ownership, Book Value, Checklist Method, VAIC, Graham's IV, DCF, Balanced Scorecard, and Benchmark. All valuation methodologies hold importance for the model, underscoring the necessity of a comprehensive and multidimensional

valuation approach. Notably, these methodologies' %IncMSE values do not significantly differ within the group. Furthermore, as the literature confirms, Benchmark does not contribute much to the model (Davila, Foster, & Jia, 2015; Dhochak & Doliya, 2020). Figure 4.1 illustrates the %IncMSE for each valuation methodology.

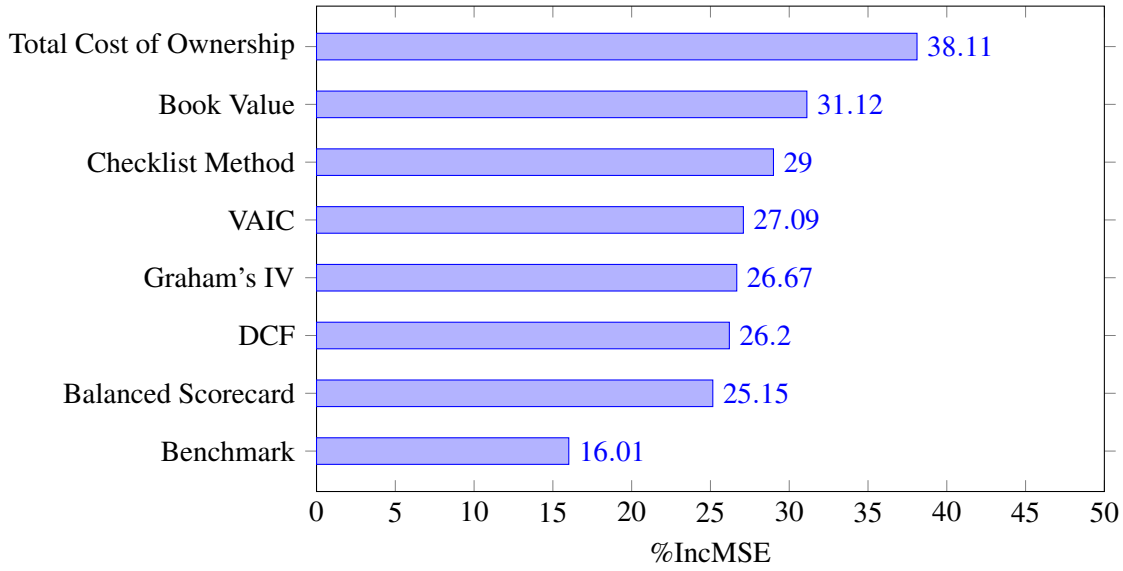


Figure 4.1: Variable importance plot of significant valuation methodologies (%IncMSE)

Conversely, the valuation methodologies that do not exhibit statistical significance for the model are Economic Value Added, Terminal Growth, and Cash Value Added. These methodologies share two common characteristics: they possess higher sparsity rates of 66.67%, 77.78%, and 81.49%, respectively, and they feature other valuation methodologies within the same category. For instance, EVA and CVA incorporate Total Cost of Ownership in their category, while TG incorporates DCF in its category. Such interdependence contributes to a high correlation between these methodologies, resulting in the exclusion of at least one of them.

Furthermore, it is worth noting that valuation methodologies based on implicit intangible assets account for 37% of the cumulative %IncMSE, while those based on financial statements contribute 63%. This observation underscores the substantial potential contribution of these non-traditional methods to company valuations.

It is noteworthy to investigate how the actual valuation of companies influences various valuation methodologies. The research examines whether different valuation levels impact the relative significance of valuation methodologies. Specifically, the study seeks to determine if certain valuation methodologies become more or less prominent in valuing companies as their value increases. To address this inquiry, a correlation analysis explores the relationship between the importance (weight) assigned to each valuation methodology and the company's actual valuation. Figure 4.2 refers to the correlation between the valuation methodologies' importance and real company valuation.

The correlation coefficient ranges from 0 to 1, where 0 represents no correlation, 1 represents

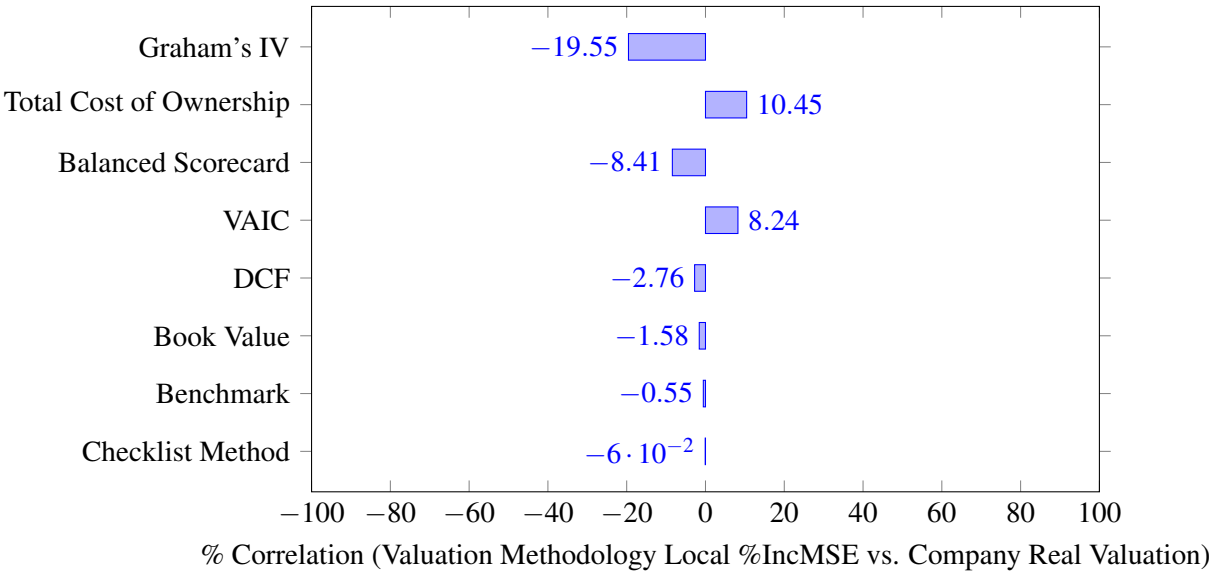


Figure 4.2: Correlation between valuation methodologies importance and real valuation

a perfect positive correlation, and -1 represents a perfect negative correlation. Typically, a correlation coefficient between 0.1 and 0.3 is considered weak, between 0.3 and 0.7 is moderate, and above 0.7 is considered strong. The correlation between all Valuation Methodologies Importance and Real Company Valuation reveals weak associations. Drawing definitive conclusions about their correlation would be inappropriate. Additionally, certain results seem counter-intuitive, such as the decreasing importance of Graham's IV for higher-value companies, which is unusual given that this method is commonly employed for large, publicly traded firms. In summary, this analysis does not yield significant findings, as the importance of all valuation methodologies exhibits only a weak correlation with company valuation.

4.3 Company characteristics importance analysis

Table 4.6 presents a comprehensive summary of each category's results, specifically focusing on the statistically significant predictive factors at a significance level of 5.00%. The table includes the count of significant factors per category, the Sum of %IncMSE and the Sum of IncNodePurity for these variables, and their corresponding *p*-values.

Table 4.6: Summary of the predictive factors importance

Category	**Nº factors**	**%IncMSE**	***p*-value**	**IncNodePurity**	***p*-value**
		Sum	*Mean | %IncMSE*	*Sum*	*Mean | IncNodePurity*
Financial Metrics	**37**	**433.08**	1.34%	**4.36E+18**	5.73%
Founder and Ownership Structure	4	18.17	1.92%	1.14E+16	28.30%
Market and Industry Factors	2	14.11	2.45%	2.57E+16	59.84%
Team and Management	2	9.49	4.40%	1.05E+17	10.89%
Corporate Governance	2	5.40	2.55%	7.95E+14	42.41%
Company and Employee Factors	2	4.27	4.50%	1.11E+15	59.69%
Funding and Financial History	2	4.11	4.25%	8.77E+14	54.40%
Risk and Performance Indicators	1	2.73	4.70%	7.21E+15	31.07%
Product and Service Factors	1	2.10	4.20%	2.99E+14	55.34%
Intellectual Property	1	1.90	4.90%	4.20E+13	100.00%

Out of the initial pool of 361 predictive factors, a notable total of 54 factors (14.96%) exhibited statistical significance about the model, as determined by %IncMSE. This outcome indicates that despite the high dimensionality of the dataset, the model successfully capitalised on many predictive factors.

An intriguing observation arises from the dominance of the Financial Metrics category, which emerges as the most important category when considering the %IncMSE metric. This finding aligns with the well-established notion in the literature that highlights the criticality of financial metrics for assessing various types of companies. Consequently, this outcome can be regarded as an expected result. Notably, the Financial Metrics category contributes a substantial 87.43% to the overall cumulative %IncMSE.

For a more comprehensive analysis, refer to the detailed information provided in Table 4.7.

Table 4.7: Summary of company characteristics categories' significance and representativity

Category	**Number of factors in dataset**	**Number (%) Significant Factors**	**Average %IncMSE**
Financial Metrics	**105**	**37** (35.24%)	**11.70**
Company and Employee Factors	27	2 (7.41%)	2.13
Founder and Ownership Structure	15	4 (26.67%)	4.54
Funding and Financial History	13	2 (15.38%)	2.05
Market and Industry Factors	12	2 (16.67%)	7.06
Market and Customer Factors	9	0 (0.00%)	0.00
Social Factors	9	0 (0.00%)	0.00
Team and Management	9	2 (22.22%)	4.75
Product and Service Factors	8	1 (12.5%)	2.10
Corporate Governance	7	2 (28.57%)	2.70
Financial Factors	6	0 (0.00%)	0.00
Market and Competitive Advantage	5	0 (0.00%)	0.00
Intellectual Property	5	1 (20.00%)	1.90
Founder and Team Attributes	5	0 (0.00%)	0.00
Legal and Regulatory Factors	5	0 (0.00%)	0.00
Venture Characteristics	4	0 (0.00%)	0.00
Investor Perception	2	0 (0.00%)	0.00
Economic Factors	2	0 (0.00%)	0.00
Risk and Performance Indicators	1	1 (100.00%)	2.73
Founder and Team Education	1	0 (0.00%)	0.00

Upon closer examination, it becomes evident that out of the initially established 22 categories, only ten are relevant to the model. Notably, the Financial Metrics category continues to stand

out, boasting the highest average %IncMSE. This signifies that, on average, individual financial metrics carry more importance than any other factor. Additionally, this category ranks second (due to the existence of a category that had one factor and that factor was selected) in the percentage of significant factors among the selected dataset. These two characteristics underscore the fundamental significance of financial metrics in determining the value of a company.

It is worth highlighting that the substantial number of factors chosen within the Financial Metrics category is influenced by repeated data definitions with only the year varying. For instance, the category incorporates revenues from five consecutive years, representing five distinct variables. This repetition contributes to the abundance of significant factors within this category.

Figure 4.3 reflects on the cumulative importance of each category.

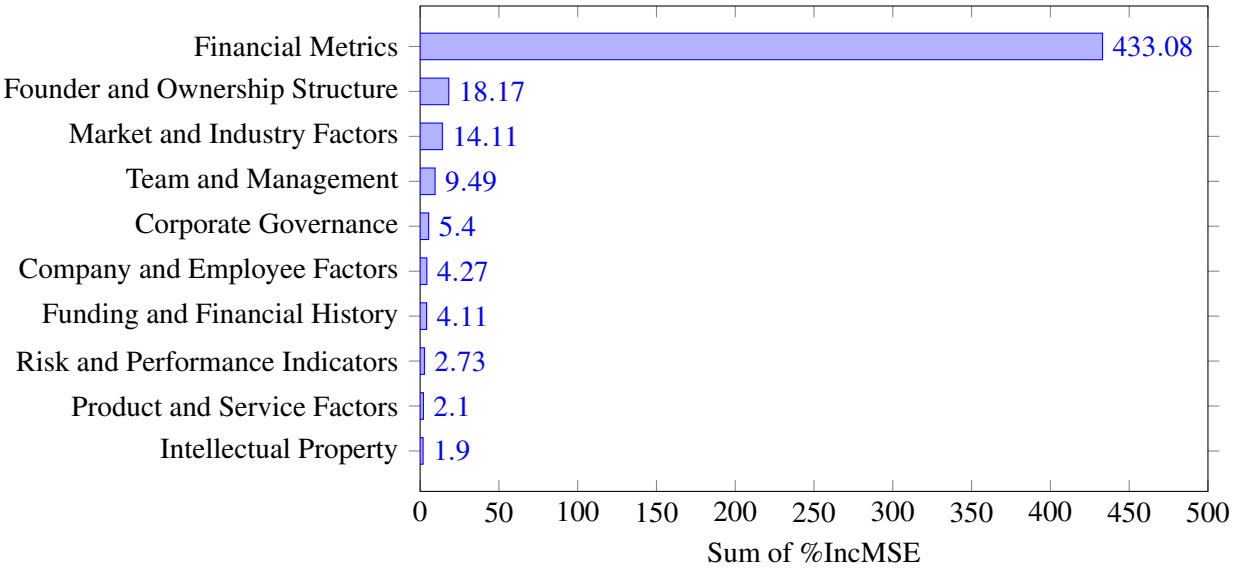


Figure 4.3: Category importance plot of significant factors across categories (%IncMSE)

The predominance of financial metrics across various categories is evident. However, it is essential to acknowledge the significant presence of diverse categories that have contributed to the model's performance, directly or indirectly associated with intellectual capital. While financial metrics such as Revenues, Operating Expenses, EBITDA, Net Operating Profit, and others constitute a substantial portion and individually outweigh any category other than Financial Metrics, it is noteworthy that the inclusive approach of considering non-financial factors plays a vital role in accounting for approximately 17% of the model's performance. This comprehensive perspective recognises the importance of incorporating various factors beyond the company's financials, emphasising their contribution to the model's overall performance.

The influence of actual company valuation on various company characteristics warrants investigation. This research explores whether different valuation levels affect the relative importance of these characteristics. Specifically, the study seeks to determine if certain company characteristics become more or less significant in valuing companies as their value increases. A correlation

analysis is performed to examine the relationship between each valuation methodology's assigned importance (weight) and the actual company valuation to address this research question.

It is important to note that utilising predefined categories for this analysis would not be suitable, as it would not provide a valid approach for studying the correlation between category importance and real company valuations. Restricting the analysis to categories would result in the loss of valuable insights.

Two analyses are conducted: Firstly, the study focuses on the 20 factors with the highest %IncMSE (above 10), aiming to understand how the contribution of these most impactful factors changes with varying valuations. Secondly, the study examines 2 of the eight factors with the highest absolute correlation (above 30%) between their importance and real valuation to derive insightful conclusions.

Figure 4.4 depicts the correlation between the importance of the 20 most impactful predictive factors and the real company valuation. The factors are ordered by decreasing importance, being t the year of the analysis.

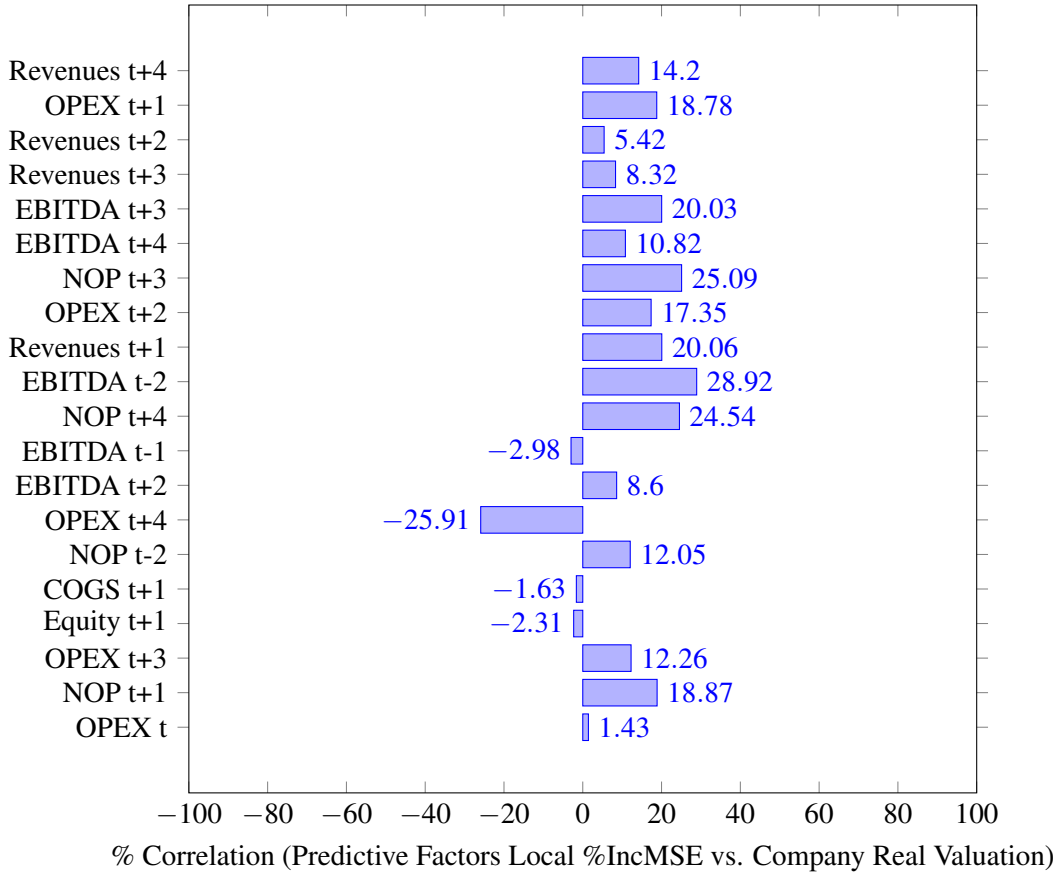


Figure 4.4: Correlation between most important predictive factors importance and real valuation

Upon thorough examination, the correlation analysis of the 20 most significant predictive factors and Real Company Valuation reveals relatively weak associations (<30%). Drawing definitive

conclusions regarding their correlation would be premature and unwarranted. However, it is noteworthy that since all these factors pertain to financial metrics, there is a prevailing tendency for a positive correlation. This observation strengthens the argument that financial metrics assume greater importance as companies expand. At the same time, non-financial factors gain relatively more significance for smaller companies in their early stages of development. To summarise, this analysis fails to yield relevant findings, as the significance of all valuation methodologies demonstrates only a weak correlation with company valuation.

Figure 4.5 depicts the correlation between the importance of two of the eight factors with higher importance correlation with the real company valuation (only two will be analysed, given the confidentiality of the data points).

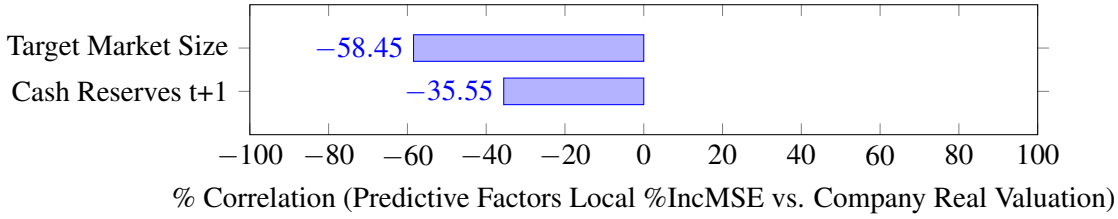


Figure 4.5: Correlation between most correlated predictive factors importance and real valuation

First and foremost, it is pertinent to note that only the eight most highly correlated variables were first considered for analysis, as they displayed a reasonably acceptable level of correlation (above 30%). No strong correlations are observed. As said, only two will be analysed, given the confidentiality of the data points (related to intellectual capital).

As the company grows, the target market size and cash reserves appear to diminish in significance for company valuation. This suggests that these factors hold greater importance during the early stages of a company's life cycle.

4.4 Valuation methodology and predictive factors importance relation analysis

Exploring the relationship between changes in valuation methodology importance and corresponding changes in predictive factor importance adds an intriguing dimension to the analysis. This investigation seeks to unveil the direct and indirect impacts of shifts in valuation importance on the significance of different characteristics. It is important to note that this analysis focuses on indirect, non-linear relationships rather than direct correlations between valuation methodologies and characteristics (as direct correlations can introduce biases due to shared computation elements).

The current section utilises categories instead of individual characteristics to extract meaningful insights. This approach enhances the interpretability of the findings, as the relationships identified at the category level offer a broader perspective. Out of the 88 individual characteristics assessed, moderate correlations with valuation methodology importance are observed, while

no strong correlations are found. This further underscores the complexity and intricacy of the analysis.

Through this analysis, valuable insights can be derived, shedding light on how changes in specific valuation methodologies may correspond to fluctuations in the significance of different factors. The objective is to identify patterns such as "when valuation methodology X tends to increase, factor Y also exhibits a similar trend." By investigating these correlations, this section aims to uncover deeper insights into the interplay between valuation methodologies and predictive factors, providing a more comprehensive understanding of how different aspects contribute to overall company valuation.

Table 4.8 presents the mean absolute correlation between the contribution of various valuation methodologies and the importance of predictive factor categories.

Table 4.8: Correlation matrix of categories importance with valuation methodologies importance (Mean absolute correlation)

Category	DCF	Checklist Method	TCO	Balanced Scorecard	Book Value	VAIC	Graham's IV	Benchmark
Company and Employee Factors	13.27%	13.97%	15.45%	6.63%	17.04%	10.39%	27.85%	20.71%
Corporate Governance	5.00%	4.39%	20.78%	23.44%	13.09%	2.96%	23.11%	9.99%
Financial Metrics	21.15%	27.24%	19.14%	21.73%	19.12%	21.79%	17.74%	13.49%
Founder and Ownership Structure	18.92%	11.21%	19.95%	19.17%	17.80%	10.10%	27.31%	9.41%
Funding and Financial History	22.49%	10.13%	6.63%	15.34%	27.75%	14.52%	21.02%	12.91%
Intellectual Property	**33.68%**	4.23%	**33.28%**	12.79%	14.47%	8.13%	29.57%	19.78%
Market and Industry Factors	18.79%	24.00%	24.67%	10.47%	8.34%	6.47%	**36.35%**	18.87%
Product and Service Factors	16.83%	22.13%	10.83%	18.26%	**37.87%**	7.90%	29.02%	11.99%
Risk and Performance Indicators	25.04%	21.29%	22.00%	16.73%	3.96%	**44.31%**	**31.83%**	19.43%
Team and Management	15.81%	24.52%	21.01%	17.06%	10.00%	8.18%	14.59%	29.06%

As evident from the analysis, weak correlations dominate the findings; however, six relationships stand out with moderate correlation, warranting further investigation.

The correlation between the importance of the DCF method and Intellectual Property reveals a moderate positive relationship. This counteracts the notion that as firms grow larger, Intellectual Capital tends to diminish in value while financial metrics gain prominence. It is worth noting that this category primarily hinges on a single significant factor: the company's adequate protection of its trademarks. Consequently, as companies expand, financial metrics and the need to safeguard intellectual assets are important, as competition and value propositions rely heavily on them.

Similarly, the same category and predictive factor moderately correlate with the Total Cost of Ownership. A closer examination reveals a negative correlation, indicating that the Total Cost of Ownership approach tends to undervalue companies with a significant emphasis on intellectual assets. This is reasonable as cost-based methods are less effective for companies whose intrinsic value lies in their capacity for value creation. Instead, methods emphasising metrics like revenue creation become more relevant for these intellectual asset-intensive companies.

Regarding the Book Value methodology, its contribution moderately correlates with the importance of Product and Service Factors. Upon closer scrutiny, a negative correlation is observed. This can be attributed to several factors. Firstly, subscription-based companies possess substantial intangible assets, such as customer relationships and brand value, which the book value fails to capture adequately. Secondly, the recognition of deferred revenue over the subscription period,

a key aspect for these companies, is omitted in the book value, leading to an incomplete valuation. Thirdly, the book value does not account for the growth potential of subscription-based companies that rely heavily on recurring revenue and expanding customer bases. Additionally, the book value does not provide insights into the subscriber base or the financial performance of these companies, rendering it an inadequate valuation method that diminishes in importance as the subscription-based business model gains significance (and vice versa).

The VAIC methodology exhibits a moderate negative correlation with Risk and Performance Indicators. When the importance of VAIC decreases, the significance of risk and performance indicators tends to increase. This negative correlation is influenced by the focus and factors associated with internal audit and operational controls. The negative correlation can be attributed to several factors. The VAIC methodology emphasises the assessment of value generated by intellectual capital, prioritising intangible assets and knowledge-based resources. In contrast, internal audit and operational controls primarily concentrate on tangible and process-oriented aspects like risk management, compliance, and operational efficiency. Furthermore, the VAIC methodology emphasises the broader strategic aspects of intellectual capital, such as innovation, knowledge sharing, and customer relationships. On the other hand, internal audit and operational controls have a narrower focus, specifically targeting internal controls, risk mitigation, and compliance. Companies may allocate resources differently between enhancing intellectual capital and strengthening internal audit and operational controls.

Lastly, Graham's Intrinsic Value moderately correlates with Market and Industry Factors and Risk and Performance Indicators. Within these categories, all the factors demonstrate positive values, indicating that as Graham's IV gains importance in company valuation, the significance of internal audit and operational controls, the size of the current target market, and the quality of key partners also increases. According to Graham's IV methodology, this observation highlights that companies with robust internal controls and streamlined operational processes exhibit better risk management practices, leading to a higher intrinsic value. Moreover, Graham's IV methodology recognises the importance of market potential in assessing a company's value. It acknowledges that the size and potential of the target market play a crucial role in determining a company's intrinsic value. Based on Graham's IV methodology, companies operating in large and promising markets are likelier to have higher intrinsic values. Similarly, the methodology acknowledges the impact of strong and reliable partnerships on a company's intrinsic value. By emphasising the significance of quality key partners, the methodology acknowledges that strategic partnerships can enhance a company's competitiveness, provide access to broader markets, and improve operational efficiency. This suggests, according to Graham's IV methodology, established and dependable partnerships are more likely to possess higher intrinsic values.

Chapter 5

Conclusion

The study effectively accomplishes its research objectives by introducing an advanced model surpassing the industry benchmark by a significant margin of 30%. Moreover, the study extracts valuable insights into the valuation structure, specifically in relation to different company characteristics. It sheds light on the factors that hold the greatest influence on overall valuation within the context of a company's specific profile. In addition to these contributions, the study also proposes a more accessible and industry-tailored framework. This system is designed to address the specific needs and requirements of the industry, making it easier for stakeholders to adopt and integrate into their valuation processes.

Throughout the thesis, the research comprehensively addresses the research questions. Firstly, the study identifies the most useful valuation methodologies for the venture capital industry, considering the complexities of valuing startups. By leveraging both traditional and non-traditional valuation methodologies, the research provides a sophisticated approach to company valuation, accounting for the unique characteristics of startups, such as limited financial data and the value of intrinsic intellectual capital. This allows the response to the first research question, "What are the most useful valuation methodologies for the VC industry?" with the following statements:

- Total Cost of Ownership is considered the most important valuation methodology.
- Nine out of twelve valuation methods are statistically significant.
- A combination of financial and non-financial, as well as traditional and non-traditional methodologies, is highly beneficial for valuing companies in the VC industry.

Secondly, the study investigates the most influential factors when valuing startups. The research examines the significance of various company characteristics in the valuation process by analysing a dataset of companies seeking or recently receiving their first round of investments. This analysis sheds light on the key factors that shape the valuation structure and provides valuable insights for better decision-making in the VC industry. This allows the response to the second research question, "What factors are the most influential when valuing startups?" with the following statements:

– Financial metrics are identified as the most influential factors for company valuation, outweighing the contribution of non-financial factors.

– Non-financial factors contribute significantly (17%) to the company valuation model. While financial metrics play a dominant role, it is essential to consider both financial and non-financial aspects when valuing startups.

Lastly, the research explores how individual factors change valuation structure. By considering different company characteristics, the study examines the dynamic relationship between these factors and the overall valuation of startups. This analysis deepens the understanding of how specific elements impact the valuation process, enhancing the accuracy and reliability of company valuations. This allows the response to the third research question, "How do individual factors change valuation structure?" with the following statements:

– There are varying degrees of correlation between company characteristics' importance and intermediate valuations contribution, being mostly below 30%.

– Nine reasonable correlations are found, revealing the relationships between valuation methodologies' rationale and their corresponding characteristics.

– The impact of individual factors on valuation structure is nuanced and depends on the specific context and methodology employed.

5.1 Critical analysis

In greater detail, the findings of this thesis contribute valuable insights to the field of comprehensive company valuation. The research successfully validated all framework models, achieving high R^2 values above 80% and low MAPE below or similar to the industry average of overvaluing companies by 50%. Although there are no significant differences in model performance among the four frameworks employed, each framework played a distinct role in the research. Notably, XGBoost consistently outperformed other algorithms in the frameworks that involved iterative algorithms.

The complex neural network model demonstrated that different frameworks, which closely resemble the company valuation process, can be used effectively. This finding paves the way for future research employing more sophisticated machine learning techniques to explore the true potential of company valuation. By allowing specific weight restrictions, the customisable approach balances the lack of adaptability of machine learning applied to company valuation and the subjectivity of traditional valuation processes.

The regression model incorporating both intermediate valuations and company characteristics yielded high performance. However, this model does not outperform the other frameworks due to high collinearity and dimensionality. This finding challenges the common belief that regression models with intermediate valuations and company characteristics are the most effective for company valuation.

The regression model with intermediate valuations enabled the interpretation of the importance of each valuation methodology on the actual valuation and validated that high performance can be achieved even with a limited number of data points. Similarly, the regression model with company characteristics allowed for interpreting the importance of different categories and individual factors, demonstrating that accurate valuation methodologies can be developed without relying solely on established methods.

Regarding the importance of valuation methodologies, the Total Cost of Ownership emerged as the most significant methodology, with nine out of twelve statistically significant valuation methods. The distribution of methodology importance varied across the significant valuation methods, emphasising the benefits of combining financial and non-financial and traditional and non-traditional methodologies within the valuation model. However, the overall correlation between the importance of valuation methodologies and company valuation is weak, indicating no significant findings.

Financial metrics are the most important factors for company valuation, overshadowing the contribution of non-financial factors, which accounted for only 17% of the valuation model. Nevertheless, the study derived specific, intuitive, and insightful conclusions from reasonably correlated factors.

Overall, the study revealed weak correlations between the importance of most categories and valuation methodologies. Despite this, nine reasonable correlations are identified, illustrating the relationships between valuation methodologies' rationale and their corresponding characteristics. These findings highlight the nuanced nature of company valuation and emphasise the need to consider multiple factors and methodologies to achieve accurate valuations.

This study has important implications for the field of company valuation and the application of data mining and machine learning techniques.

Developing robust models is essential in navigating the complexities of company valuation. This study utilises an extensive dataset to construct models, one of which surpasses the industry benchmark performance by an impressive 30%, a result which has not been previously achieved.

The study highlights the importance of non-financial valuation methods to complement traditional approaches, ones that are normally used in literature, underscoring the significance of these methodologies in accurately valuing startups.

The research examines the influence of determinants of firm value, namely company characteristics and valuation methodologies, and their variations across different valuation contexts, enhancing the field's understanding of firm value dynamics. It also sheds light on the intricate relationship between valuation methodologies, company characteristics, and overall valuation outcomes, aiding decision-making by practitioners in the field.

At last, it proposes an innovative framework which mimics the valuation process, targeting an unexplored research focus: developing tailored ML solutions for the finance landscape to reduce the industry's resistance to change.

5.2 Future research

The study is still subject to certain limitations that warrant consideration. First, the dataset utilised in the analysis is characterised by a relatively small number of observations, which may compromise the reliability and generalisability of the findings. Moreover, the focus on startup companies, primarily in the Nordic region, restricts the applicability of the results to a narrower context. Additionally, high sparsity in various characteristics and companies necessitated their exclusion, potentially omitting important variables that could have influenced the outcomes.

Second, the research does not address the gap in the literature regarding the analysis of companies in emerging countries. This omission hinders a comprehensive understanding of valuation dynamics within these contexts and leaves room for further exploration in future studies.

Third, the study solely relies on secondary data sources, such as existing databases, without incorporating primary data collection methods like surveys or class associations. This absence of primary data limits the depth of analysis and restricts the ability to capture nuanced insights that can only be obtained through direct data collection.

Lastly, privacy concerns about company characteristics may have imposed constraints on the level of detail and disclosure during the analysis. As a result, certain aspects of the valuation process may not have been fully explored, potentially reducing the comprehensiveness of the findings.

These limitations underscore the need for future research endeavours to address these shortcomings. This includes expanding the dataset to encompass a larger and more diverse sample of companies, considering the unique dynamics of emerging countries, incorporating primary data collection methods to augment the analysis, and addressing privacy concerns to facilitate a more comprehensive examination of a company valuation.

Several recommendations for future research can be made based on the identified limitations. Firstly, expanding the dataset to include a larger and more diverse sample of companies would enhance the reliability and generalisability of the findings. Additionally, conducting studies focusing on analysing companies in emerging countries would provide valuable insights into valuation dynamics within these contexts. Incorporating primary data collection methods, such as surveys and class associations, would allow for a deeper analysis and capture nuanced insights that cannot be obtained solely from secondary data sources. Additionally, addressing privacy concerns to facilitate a more detailed analysis of company characteristics would lead to a more comprehensive examination of company valuation. Finally, apply deep learning techniques and improvements to the new framework to understand its potential.

References

Aguilar-Rivera, A., Valenzuela-Rendón, M., & Rodríguez-Ortiz, J. (2015, 11). Genetic algorithms and darwinian approaches in financial applications: A survey. *Expert Systems with Applications*, *42*, 7684–7697. doi: 10.1016/j.eswa.2015.06.001

Alpaydin, E. (2010). *Introduction to machine learning* (2nd ed.). London, England: MIT Press.

Amorim Neto, M. C., Tavares, G., Alves, V. M. O., Cavalcanti, G. D. C., & Ing Ren, T. (2010). Improving financial time series prediction using exogenous series and neural networks committees. In *The 2010 international joint conference on neural networks (ijcnn)* (p. 1-8). doi: 10.1109/IJCNN.2010.5596911

Antretter, T., Blohm, I., Grichnik, D., & Wincent, J. (2019). Predicting new venture survival: A twitter-based machine learning approach to measuring online legitimacy. *Journal of Business Venturing Insights*, *11*, e00109. Retrieved from https://www.sciencedirect.com/science/article/pii/S2352673418301197 doi: https://doi.org/10.1016/j.jbvi.2018.e00109

Arroyo, J., Corea, F., Jimenez-Diaz, G., & Recio-García, J. (2019, 08). Assessment of machine learning performance for decision support in venture capital investments. *IEEE Access*, *PP*, 1-1. doi: 10.1109/ACCESS.2019.2938659

Aruldoss, M., Manjula, M., & Venkatesan, V. (2011, 09). A business intelligence model to predict bankruptcy using financial domain ontology with association rule mining algorithm. *International Journal of Computer Science Issues*, *8*.

Audretsch, D. B., & Link, A. N. (2012, Feb 01). Valuing an entrepreneurial enterprise. *Small Business Economics*, *38*(2), 139-145. Retrieved from https://doi.org/10.1007/s11187-011-9409-5 doi: 10.1007/s11187-011-9409-5

Belle, V., & Papantonis, I. (2021). Principles and practice of explainable machine learning. *Frontiers in Big Data*, *4*. Retrieved from https://www.frontiersin.org/articles/10.3389/fdata.2021.688969 doi: 10.3389/fdata.2021.688969

Block, J., Fisch, C., Vismara, S., & Andres, R. (2019). Private equity investment criteria: An experimental conjoint analysis of venture capital, business angels, and family offices. *Journal of Corporate Finance*, *58*, 329-352. Retrieved from https://www.sciencedirect.com/science/article/pii/S0929119918307132 doi: https://doi.org/10.1016/j.jcorpfin.2019.05.009

Blohm, I., Antretter, T., Sirén, C., Grichnik, D., & Wincent, J. (2020, 09). It's a peoples game, isn't it?! a comparison between the investment returns of business angels and machine learning algorithms. *Entrepreneurship: Theory and Practice*. doi: 10.1177/1042258720945206

Bonaventura, M., Ciotti, V., Panzarasa, P., Liverani, S., Lacasa, L., & Latora, V. (2019). *Predicting success in the worldwide start-up network.*

Bontis, N., Dragonetti, N. C., Jacobsen, K., & Roos, G. (1999). The knowledge toolbox: A review of the tools available to measure and manage intangible resources. *European Management Journal*, *17*, 391-402.

Brooking, A. (1996). *Intellectual capital*. Andover, England: Cengage Learning EMEA.

Bruner, R., Eades, K. M., Harris, R. S., & Higgins, R. F. (1998). *Best practices in estimating the cost of capital: Survey and synthesis.* Harvard Business School. Retrieved from https://www.hbs.edu/faculty/Pages/item.aspx?num=50378

Chan, L. K., Lakonishok, J., & Sougiannis, T. (2000, Jun). *The stock market valuation of research and development expenditures.* NBER Working Paper No. w7223. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=227564

Chen, T., & Guestrin, C. (2016, 08). Xgboost: A scalable tree boosting system. In (p. 785-794). doi: 10.1145/2939672.2939785

Chen, W.-S., & Du, Y.-K. (2009). Using neural networks and data mining techniques for the financial distress prediction model. *Expert Systems with Applications*, *36*(2, Part 2), 4075-4086. Retrieved from https://www.sciencedirect.com/science/article/pii/S0957417408001954 doi: https://doi.org/10.1016/j.eswa.2008.03.020

Clarke, M., Seng, D., & Whiting, R. H. (2011, Jan 01). Intellectual capital and firm performance in australia. *Journal of Intellectual Capital*, *12*(4), 505-530. Retrieved from https://doi.org/10.1108/14691931111181706 doi: 10.1108/14691931111181706

Coad, A., & Srhoj, S. (2019, 01). Catching gazelles with a lasso: Big data techniques for the prediction of high-growth firms. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3395527

Csaszar, F., Nussbaum, M., & Sepulveda, M. (2006). Strategic and cognitive criteria for the selection of startups. *Technovation*, *26*(2), 151-161. Retrieved from https://www.sciencedirect.com/science/article/pii/S0166497205000453 doi: https://doi.org/10.1016/j.technovation.2005.01.010

Damodaran, A. (2009). *The dark side of valuation: Valuing young, distressed, and complex businesses*. Upper Saddle River, NJ: Financial Times Prentice Hall.

Damodaran, A. (2012). *Investment valuation: Tools and techniques for determining the value of any asset* (3rd ed.). Nashville, TN: John Wiley & Sons.

Damodaran, A. (2014). *Applied corporate finance* (4th ed.). Nashville, TN: John Wiley & Sons.

Damodaran, A. (2016). *Damodaran on valuation: Security analysis for investment and corporate finance* (2nd ed.; A. Damodaran, Ed.). Nashville, TN: John Wiley & Sons.

Davila, A., Foster, G., & Jia, N. (2015). The valuation of management control systems in start-up companies: International field-based evidence. *European Accounting Review*, *24*(2), 207-239. Retrieved from https://doi.org/10.1080/09638180.2014.965720 doi: 10.1080/09638180.2014.965720

Dechow, P. M., Hutton, A. P., & Sloan, R. G. (1999). An empirical assessment of the residual income valuation model 1 this paper has benefited from the c. *Journal of Accounting and Economics*.

Dhochak, M., & Doliya, P. (2020). Valuation of a startup: Moving towards strategic approaches. *Journal of Multi-Criteria Decision Analysis*, *27*(1-2), 39-49. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/mcda.1703 doi: https://doi.org/10.1002/mcda.1703

Dixit, A. K., & Pindyck, R. S. (1994). *Investment under uncertainty*. Princeton, NJ: Princeton University Press.

Dorogush, A. V., Gulin, A., Gusev, G., Kazeev, N., Prokhorenkova, L. O., & Vorobev, A. (2017). Fighting biases with dynamic boosting. *CoRR*, *abs/1706.09516*. Retrieved from http://arxiv.org/abs/1706.09516

Eckstein, C. (2004, 06). The measurement and recognition of intangible assets: Then and now. *Accounting Forum*, *28*, 139-158. doi: 10.1016/j.accfor.2004.02.001

Edvinsson, L. (1997). Developing intellectual capital at skandia. *Long Range Planning*, *30*(3),

366-373. Retrieved from https://www.sciencedirect.com/science/article/pii/S002463019790248X doi: https://doi.org/10.1016/S0024-6301(97)90248-X

Ewens, M., Jones, C. M., & Rhodes-Kropf, M. (2013). The price of diversifiable risk in venture capital and private equity. *The Review of Financial Studies*, *26*(8), 1853–1889. Retrieved 2023-04-19, from http://www.jstor.org/stable/23470211

Fabozzi, F. J., Focardi, S. M., & Jonas, C. (2017). *Equity valuation: Science, art, or craft?* CFA Institute Research Foundation.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654-669. Retrieved from https://www.sciencedirect.com/science/article/pii/S0377221717310652 doi: https://doi.org/10.1016/j.ejor.2017.11.054

Francis, J., Olsson, P., & Oswald, D. (2001, 09). Comparing the accuracy and explainability of dividend, free cash flow, and abnormal earnings equity value estimates. *Journal of Accounting Research*, *38*, 45. doi: 10.2307/2672922

Friedman, J. (2000, 11). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*. doi: 10.1214/aos/1013203451

Futagami, K., Fukazawa, Y., Kapoor, N., & Kito, T. (2021). Pairwise acquisition prediction with shap value interpretation. *The Journal of Finance and Data Science*, *7*, 22-44. Retrieved from https://www.sciencedirect.com/science/article/pii/S2405918821000015 doi: https://doi.org/10.1016/j.jfds.2021.02.001

Gan, G. (2013, 11). Application of data clustering and machine learning in variable annuity valuation. *Insurance: Mathematics and Economics*, *53*, 795–801. doi: 10.1016/j.insmatheco.2013.09.021

Garkavenko, M., Mirisaee, H., Gaussier, É., Guerraz, A., & Lagnier, C. (2021). Valuation of startups: A machine learning perspective. In *European conference on information retrieval.*

Gastaud, C., Carniel, T., & Dalle, J.-M. (2019). *The varying importance of extrinsic factors in the success of startup fundraising: competition at early-stage and networks at growth-stage.*

Gavious, I., & Schwartz, D. (2011). Market valuations of start-up ventures around the technology bubble. *International Small Business Journal*, *29*(4), 399-415. Retrieved from https://doi.org/10.1177/0266242610369750 doi: 10.1177/0266242610369750

George, G., Osinga, E., Lavie, D., & Scott, B. (2016, 10). Big data and data science methods for management research. *The Academy of Management Journal*, *59*, 1493-1507. doi: 10.5465/amj.2016.4005

Glaveckaite, Z. (2020). The development process of the right team in early stage start-ups. *Entrepreneurship and sustainability issues*. Retrieved from https://jssidoi.org/jesi/article/670

Gleason, K. I., & Klock, M. (2006). Intangible capital in the pharmaceutical and chemical industry. *The Quarterly Review of Economics and Finance*, *46*(2), 300-314. Retrieved from https://www.sciencedirect.com/science/article/pii/S1062976906000238 doi: https://doi.org/10.1016/j.qref.2005.03.001

Groth, S., & Muntermann, J. (2011, 03). An intraday market risk management approach based on textual analysis. *Decision Support Systems*, *50*, 680-691. doi: 10.1016/j.dss.2010.08.019

Gunn, S. R. (1998). Support vector machines for classification and regression..

Haddad, H., Weking, J., Hermes, S., Böhm, M., & Krcmar, H. (2020, 03). Business model choice matters: How business models impact different performance measures of startups. In *Proceedings of the 15th international conference on wirtschaftsinformatikat.*

Haralampiev, K., Yankov, B., & Ruskov, P. (2014). Models and tools for technology start-up companies success analysis. *Economic Alternatives*, 15-24.

Harrison, S. S., & Sullivan, P. H. (2000). Profiting from intellectual capital: learning from leading companies. *Industrial and Commercial Training*.

Huang, W., Nakamori, Y., & Wang, S.-Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, *32*(10), 2513-2522. Retrieved from https://www.sciencedirect.com/science/article/pii/S0305054804000681 (Applications of Neural Networks) doi: https://doi.org/10.1016/j.cor.2004.03.016

Huang, W.-B., Liu, J., Bai, H., & Zhang, P. (2020). Value assessment of companies by using an enterprise value assessment system based on their public transfer specification. *Information Processing & Management*, *57*(5), 102254. Retrieved from https://www.sciencedirect.com/science/article/pii/S0306457319307976 doi: https://doi.org/10.1016/j.ipm.2020.102254

Hunter, D. S., Saini, A., & Zaman, T. (2018). *Picking winners: A data driven approach to evaluating the quality of startup companies.*

International private equity and venture capital valuation guidelines. (2022). International Private Equity and Venture Capital. Retrieved from https://www.privateequityvaluation.com/valuation-guidelines/4588034291

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research*, *2*(2), 59-64. Retrieved from https://www.sciencedirect.com/science/article/pii/S2214579615000076 (Visions on Big Data) doi: https://doi.org/10.1016/j.bdr.2015.01.006

Kaplan, R., & Norton, D. (2007, 07). Using the balanced scorecard as a strategic management system. *Harvard Business Review*, *85*.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Nips*.

Kitchin, R. (2014, 04). Big data, new epistemologies and paradigm shift. *Big Data & Society*, *1*, 1-12. doi: 10.1177/2053951714528481

Koller, T., Goedhart, M., & Wessels, D. (2020). *Valuation - measuring and managing the value of companies, seventh edition* (7th ed.). Nashville, TN: John Wiley & Sons.

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the s&p 500. *European Journal of Operational Research*, *259*(2), 689-702. Retrieved from https://www.sciencedirect.com/science/article/pii/S0377221716308657 doi: https://doi.org/10.1016/j.ejor.2016.10.031

Lee, M.-C. (2009). Using support vector machine with a hybrid feature selection method to the stock trend prediction. *Expert Syst. Appl.*, *36*, 10896-10904.

Lin, F., Liang, D., Yeh, C.-C., & Huang, J.-C. (2014). Novel feature selection methods to financial distress prediction. *Expert Systems with Applications*, *41*(5), 2472-2483. Retrieved from https://www.sciencedirect.com/science/article/pii/S0957417413008014 doi: https://doi.org/10.1016/j.eswa.2013.09.047

Lins, K. (2003, 03). Equity ownership and firm value in emerging markets. *Journal of Financial and Quantitative Analysis*, *38*, 159-184.

Littunen, H., & Niittykangas, H. (2010, 02). The rapid growth of young firms during various stages of entrepreneurship. *Journal of Small Business and Enterprise Development*, *17*, 8-31. doi: 10.1108/14626001011019107

Liu, J., Nissim, D., & Thomas, J. (2002, 03). Equity valuation using multiples. *Journal of Accounting Research*, *40*, 135 - 172. doi: 10.1111/1475-679X.00042

Liu, J., Nissim, D., & Thomas, J. (2007). Is cash flow king in valuations? *Financial Analysts Journal*, *63*(2), 56-68. Retrieved from https://doi.org/10.2469/faj.v63.n2.4522

doi: 10.2469/faj.v63.n2.4522

Lu, C.-J., Lee, T.-S., & Chiu, C.-C. (2009). Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems*, *47*(2), 115-125. Retrieved from https://www.sciencedirect.com/science/article/pii/S0167923609000323 doi: https://doi.org/10.1016/j.dss.2009.02.001

Mason, C., & Stark, M. (2004, 06). What do investors look for in a business plan?: A comparison of the investment criteria of bankers, venture capitalists and business angels. *International Small Business Journal - INT SMALL BUS J*, *22*, 227-248. doi: 10.1177/0266242604042377

Maxwell, A. L., Jeffrey, S. A., & Lévesque, M. (2011). Business angel early stage decision making. *Journal of Business Venturing*, *26*(2), 212-225. Retrieved from https://www.sciencedirect.com/science/article/pii/S0883902609000974 doi: https://doi.org/10.1016/j.jbusvent.2009.09.002

McKenzie, D., & Sansone, D. (2017, Dec). *Man vs. machine in predicting successful entrepreneurs: Evidence from a business plan competition in nigeria.* World Bank, Washington, DC. Retrieved from https://openknowledge.worldbank.org/entities/publication/54174a6b-a514-5647-ad73-0301bac3b056

Melgarejo-Meseguer, F. M., Gimeno-Blanes, F. J., Salar-Alcaraz, M.-E., Gimeno-Blanes, J. R., Martínez-Sánchez, J., García-Alberola, A., & Rojo-álvarez, J. L. (2019). Electrocardiographic fragmented activity (ii): A machine learning approach to detection. *Applied Sciences*.

Mills, R. (1995). Strategic value analysis: Linking finance and strategy [Article]. *Management Accounting: Magazine for Chartered Management Accountants*, *73*(4), 42 – 44. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-0345859939&partnerID=40&md5=80f6e3384d9de0f68547f46ff0ef2392 (Cited by: 5)

Miloud, T., Aspelund, A., & Cabrol, M. (2012, 04). Startup valuation by venture capitalists: An empirical study. *Venture Capital: An International Journal of Entrepreneurial Finance*, *14*, 1-24. doi: 10.1080/13691066.2012.667907

Mirzanti, I. R., Sinaga, E. E. M., & Soekarno, S. (2019). Determinant factors in digital start-up valuation for agro seed funding. *International Journal of Agricultural Resources, Governance and Ecology*, *15*(4), 338-357. Retrieved from https://www.inderscienceonline.com/doi/abs/10.1504/IJARGE.2019.104202 doi: 10.1504/IJARGE.2019.104202

Mnejja, A., & Sahut, J.-M. (2010, 09). The discount premium between private m&a and lbo transactions. *SSRN Electronic Journal*. doi: 10.2139/ssrn.1676139

Moreno Gutiérrez, J., & Melo-Velandia, L. (2011, 01). Pronóstico de incumplimientos de pago mediante máquinas de vectores de soporte: una aproximación inicial a la gestión del riesgo de crédito. *Borradores de Economía*.

Obschonka, M., & Audretsch, D. (2020, 10). Artificial intelligence and big data in entrepreneurship: A new era has begun. *Small Business Economics*, *55*. doi: 10.1007/s11187-019-00202-4

Olbrich, M., Quill, T., & Rapp, D. (2015, 01). Business valuation inspired by the austrian school. *Journal of Business Valuation and Economic Loss Analysis*. doi: 10.1515/jbvela-2014-0001

Ou, J. A., & Penman, S. H. (1989). Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics*, *11*(4), 295-329. Retrieved from https://www.sciencedirect.com/science/article/pii/0165410189900177 doi: https://

doi.org/10.1016/0165-4101(89)90017-7

Pablo, F. (2002). *Valuation methods and shareholder value creation*. Academic Press.

Park, H., Kim, N., & Lee, J. (2014). Parametric models and non-parametric machine learning models for predicting option prices: Empirical comparison study over kospi 200 index options. *Expert Syst. Appl.*, *41*, 5227-5237.

Pinto, J. E., Henry, E., Robinson, T. R., & Stowe, J. D. (2010). *Equity asset valuation* (2nd ed.). Chichester, England: John Wiley & Sons.

Prüfer, J., & Prufer, P. (2020, 10). Data science for entrepreneurship research: studying demand dynamics for entrepreneurial skills in the netherlands. *Small Business Economics*, *55*. doi: 10.1007/s11187-019-00208-y

Pulic, A. (2000). Vaic™ – an accounting tool for ic management. *International Journal of Technology Management*, *20*(5-8), 702-714. Retrieved from https://www.inderscienceonline.com/doi/abs/10.1504/IJTM.2000.002891 doi: 10.1504/IJTM.2000.002891

Pulic, A. (2004). Intellectual capital – does it create or destroy value? *Measuring Business Excellence*, *8*, 62-68.

Pulic, A. (2005, 12). Value creation efficiency at national and regional levels: Case study—croatia and the european union. In (p. 197-211). doi: 10.1016/B978-0-7506-7773-8.50015-7

Quintero, S. (2019). An empirical perspective on startup valuations..

Ragothaman, S., Naik, B., & Ramakrishnan, K. (2003, 12). Predicting corporate acquisitions: An application of uncertain reasoning using rule induction. *Information Systems Frontiers*, *5*, 401-412. doi: 10.1023/B:ISFI.0000005653.53641.b3

Roos, J., Roos, G., Dragonetti, N. C., & Edvinsson, L. (1998). *Intellectual capital*. New York, NY: New York University Press.

Ruiz, E. J., Hristidis, V., Castillo, C., Gionis, A., & Jaimes, A. (2012). Correlating financial time series with micro-blogging activity. In *Web search and data mining*.

Schumaker, R. P., Zhang, Y., Huang, C.-N., & Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems*, *53*(3), 458-464. Retrieved from https://www.sciencedirect.com/science/article/pii/S0167923612000875 doi: https://doi.org/10.1016/j.dss.2012.03.001

Shapiro, E., Mackmin, D., & Sams, G. (2019). *Modern methods of valuation* (12th ed.). London, England: CRC Press.

Sharchilev, B., Roizner, M., Rumyantsev, A. Y., Ozornin, D., Serdyukov, P., & de Rijke, M. (2018). Web-based startup success prediction. *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*.

Sievers, S., Mokwa, C. F., & Keienburg, G. (2013). The relevance of financial versus non-financial information for the valuation of venture capital-backed firms. *European Accounting Review*, *22*(3), 467-511. Retrieved from https://doi.org/10.1080/09638180.2012.741051 doi: 10.1080/09638180.2012.741051

Silva, W., Fantin, C., Fukui, M., & Juca, M. (2021, 01). Startups' valuation: A bibliometric analysis and systematic literature review. *Open Journal of Business and Management*, *09*, 1647-1670. doi: 10.4236/ojbm.2021.94090

Simpson, A. V. (2008, May). *Analysts' use of non-financial information disclosures.* Contemporary Accounting Research. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1133182

Sitar, A. S. (2005). Measuring intellectual capital : Lessons learned from a practical implementation..

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis

of big data challenges and analytical methods. *Journal of Business Research*, *70*, 263-286. Retrieved from https://www.sciencedirect.com/science/article/pii/S014829631630488X doi: https://doi.org/10.1016/j.jbusres.2016.08.001

Ståhle, P., Ståhle, S., & Aho, S. (2011, Jan 01). Value added intellectual coefficient (vaic): a critical analysis. *Journal of Intellectual Capital*, *12*(4), 531-551. Retrieved from https://doi.org/10.1108/14691931111181715 doi: 10.1108/14691931111181715

Stewart, T. A. (1999). *Intellectual capital: The new wealth of organizations*. New York, NY: Bantam Doubleday Dell Publishing Group.

Sveiby, ©.-E. (2018). Measuring intangibles and intellectual capital-an emerging first standard..

Tay, F. E., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, *29*(4), 309-317. Retrieved from https://www.sciencedirect.com/science/article/pii/S0305048301000263 doi: https://doi.org/10.1016/S0305-0483(01)00026-3

Tkáč, M., & Verner, R. (2016). Artificial neural networks in business: Two decades of research. *Applied Soft Computing*, *38*, 788-804. Retrieved from https://www.sciencedirect.com/science/article/pii/S1568494615006122 doi: https://doi.org/10.1016/j.asoc.2015.09.040

Trigeorgis, L. (1996). *Real options*. London, England: MIT Press.

Tsai, C.-F., Lu, Y.-H., & Yen, D. C. (2012). Determinants of intangible assets value: The data mining approach. *Knowledge-Based Systems*, *31*, 67-77. Retrieved from https://www.sciencedirect.com/science/article/pii/S0950705112000470 doi: https://doi.org/10.1016/j.knosys.2012.02.007

Tsukahara, Y., Tanaka, R., & Suzuki, T. (2022). Nonlinear modeling for equity valuation by machine learning. *Nonlinear Theory and Its Applications, IEICE*, *13*(2), 246-251. doi: 10.1587/nolta.13.246

Tumasjan, A., Braun, R., & Stolz, B. A. (2021). Twitter sentiment as a weak signal in venture capital financing. *Journal of Business Venturing*, *36*, 106062.

Vanstone, B., & Finnie, G. (2009, 04). An empirical methodology for developing stockmarket trading systems using artificial neural networks. *Information Technology papers*, *36*. doi: 10.1016/j.eswa.2008.08.019

Vayas-Ortega, G., Soguero-Ruiz, C., Rodríguez-Ibáñez, M., Rojo-Álvarez, J.-L., & Gimeno-Blanes, F.-J. (2020, Aug). On the differential analysis of enterprise valuation methods as a guideline for unlisted companies assessment (ii): Applying machine-learning techniques for unbiased enterprise value assessment. *Applied Sciences*, *10*(15), 5334. Retrieved from http://dx.doi.org/10.3390/app10155334 doi: 10.3390/app10155334

Vayas-Ortega, G., Soguero Ruiz, C., Rojo-Álvarez, J. L., & Gimeno, J. (2020, 08). On the differential analysis of enterprise valuation methods as a guideline for unlisted companies assessment (i): Empowering discounted cash flow valuation. *Applied Sciences*, *10*, 5875. doi: 10.3390/app10175875

Vidal, R., & Ribal, J. (2017, 10). Valuation of agrifood smes. lessons to be learnt from the stock market. *Spanish Journal of Agricultural Research*, *15*, e0118. doi: 10.5424/sjar/2017154-11668

Visconti, R. M., Montesi, G., & Papiro, G. (2018). Big data-driven stochastic business planning and corporate valuation. *DecisionSciRN: Business Analytics (Topic)*.

Wang, B., Huang, H., & Wang, X. (2012). A novel text mining approach to financial time series forecasting. *Neurocomputing*, *83*, 136-145.

Weil, R. L., Schipper, K., & Francis, J. (2012). *Financial accounting* (14th ed.). Belmont, CA: Wadsworth Publishing.

Williams, J. B. (1997). *The theory of investment value*. Fraser Pub.

Wiwattanakantang, Y. (2001). Controlling shareholders and corporate value: Evidence from thailand. *Pacific-Basin Finance Journal*, *9*(4), 323-362. Retrieved from https://www.sciencedirect.com/science/article/pii/S0927538X01000221 doi: https://doi.org/10.1016/S0927-538X(01)00022-1

Xie, G.-q. (2011, 07). The optimization of share price prediction model based on support vector machine. In *2011 international conference on control, automation and systems engineering.* IEEE. doi: 10.1109/ICCASE.2011.5997714

Yoo, C., Yang, D., Kim, H., & Heo, E. (2012). Key value drivers of startup companies in the new media industry—the case of online games in korea. *Journal of Media Economics*, *25*(4), 244-260. Retrieved from https://doi.org/10.1080/08997764.2012.729546 doi: 10.1080/08997764.2012.729546

Zheng, Y., Liu, J., & George, G. (2010, 11). The dynamic impact of innovative capability and inter-firm network on firm valuation: A longitudinal study of biotechnology start-ups. *Journal of Business Venturing*, *25*, 593-609. doi: 10.1016/j.jbusvent.2009.02.001

Appendix A

Results plots

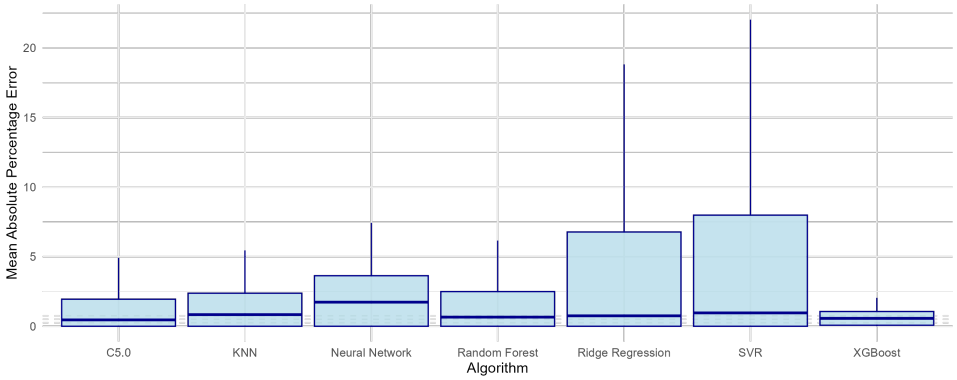


Figure A.1: Regression model with intermediate valuations results distribution

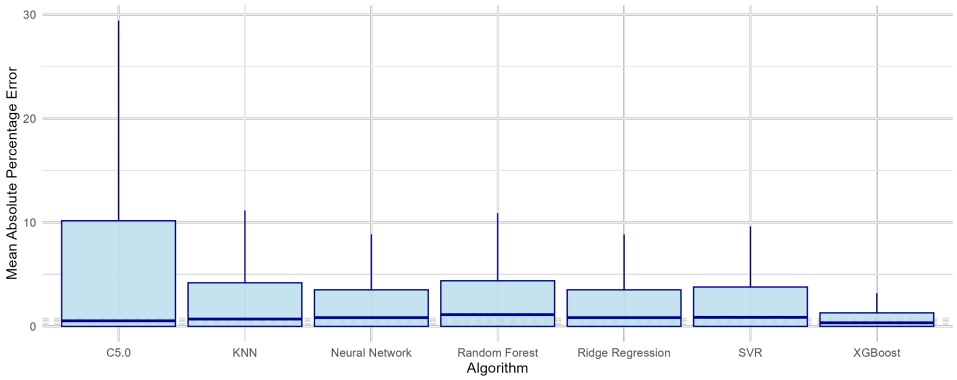


Figure A.2: Regression model with company characteristics results distribution

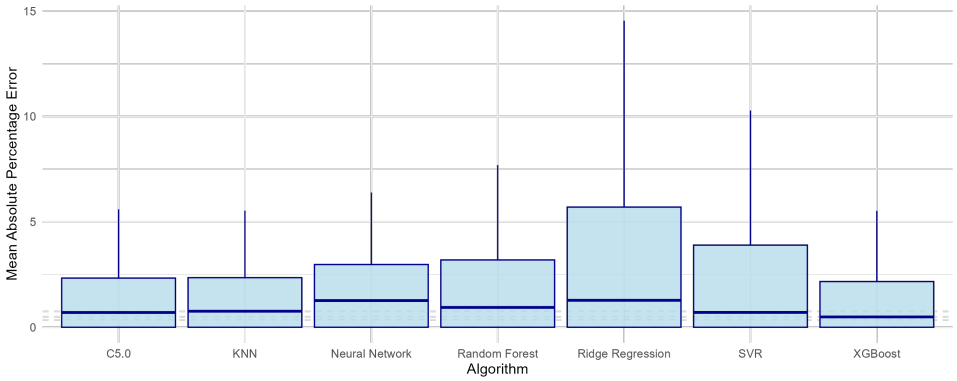


Figure A.3: Enhanced regression model results distribution

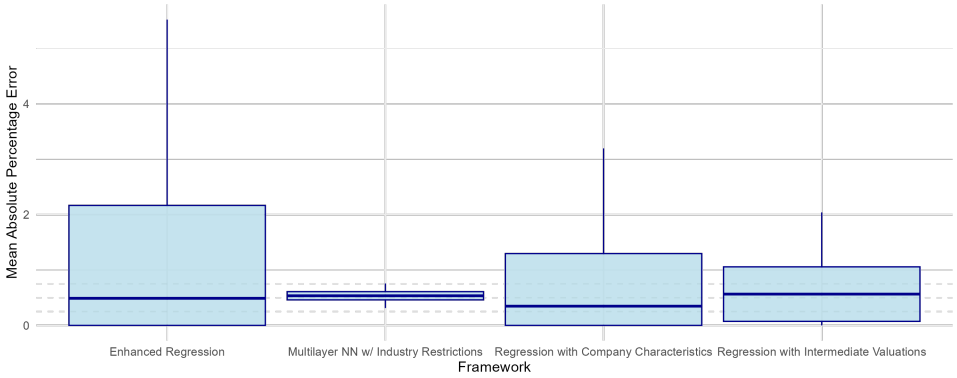


Figure A.4: Frameworks' results distribution