University of Bari Aldo Moro

Department of Computer Science

PhD Programme in Computer Science and Mathematics
CYCLE XXXVII

Scientific Disciplinary Sector: IINF-05/A

Data Science approaches to support the design of crowdfunding campaigns

ROSA PORRO
14/10/2025 15:22:13

Rosa Porro

Rosa Porro

Supervisor:
Prof. Corrado Loglisci

CORRADO
LOGLISCI
16.10.2025
10:49:54
GMT+02:00

Co-supervisor:
Dr. Gennaro Vessio

Gennaro Vessio
15.10.2025 21:14:1
GMT+01:00

Coordinator:
Prof. Francesca Mazzia

Francesca Mazzia

FRANCESCA
MAZZIA
22.10.2025 17:08:39
GMT+02:00

2025

Abstract

Crowdfunding has become a powerful mechanism for collective financing, offering a global reach and versatile applications across various sectors. It challenges traditional funding sources like bank loans, venture capital, and private equity. This work delves into the complex dynamics of crowdfunding platforms, focusing on investor behaviour and investment patterns within equity and lending campaigns.

By employing advanced machine learning techniques, such as XGBoost and LSTM networks, I have developed predictive models that analyse real-time and historical data to accurately forecast the success or failure of crowdfunding campaigns. A distinctive aspect of this research is the introduction of a "pre and post-launch" analysis framework, which examines the factors influencing campaign success both before the launch (e.g., marketing strategies, initial investor interest) and after (e.g., ongoing investor engagement and funding momentum). This dual-phase perspective addresses a critical gap in existing studies, offering a more comprehensive understanding of campaign dynamics.

To further enhance crowdfunding analysis tools, I introduce two novel datasets, one for equity crowdfunding and another for lending, tailored to capture the unique temporal and behavioural patterns within Italian crowdfunding platforms. Additionally, my approach moves beyond traditional binary success metrics, proposing innovative measures that better capture campaign outcomes, such as funding ratios, overfunding levels, and long-term visibility metrics.

The insights gained from this study can significantly enhance crowdfunding strategies, improving project selection, pre-launch preparations, and post-launch promotional tactics on platforms. By refining decision-making processes and offering forward-looking guidance to investors, our computational model empowers both campaign creators and platform administrators. Ultimately, this research contributes to increasing the overall efficacy and sustainability of crowdfunding as a financing tool.

Contents

1 Introduction **1**
1.1 Objectives and Purposes 1
1.2 Overview of the Thesis 3

2 State of the Art **7**
2.1 Crowdfunding and Digital Transformation 7
2.2 Regulatory Framework 9
2.3 Equity Crowdfunding 11
2.4 Lending Crowdfunding 14
2.5 Research Opportunities 16

3 Modelling for Crowdfunding Campaigns **19**
3.1 Data Preparation for predictive modelling 19
3.2 Input Data and Predictive Modeling in Crowdfunding Dynamics 20
3.2.1 Resampling Techniques for Crowdfunding Data 22
3.2.2 Machine Learning Models 22
3.2.3 Temporal Feature Integration and Modelling Strategy 24
3.3 Performance Evaluation and Model Assessment 25

4 Experimental framework **27**
4.1 Data preparation for predictive modeling 28
4.1.1 Pre- and Post-Launch Features 28
4.1.2 Dataset Description 29
4.1.3 Data Preprocessing 33
4.1.4 Standardize Time Series 35
4.1.5 Encode Categorical Variables 37
4.1.6 Create new feature 39
4.2 Output data and predictive modeling 41
4.2.1 Handle imbalance data: SMOTE AND SMOGN 41
4.2.2 Model training 42
4.2.3 Hyperparameter Optimization 43
4.2.4 Results 45
4.2.5 Benchmarking of Predictive Models in Equity Crowdfunding . . 51

4.2.6 Model Interpretability with SHAP 52
4.3 Key Findings and Advancements Over Previous Research 55
4.4 Key Drivers of Crowdfunding Success 58
4.4.1 Applicability of the Model to European Platforms 59
4.5 U.S. dataset analysis and international comparison 60
4.6 Deployment, Scalability and Integration 63

Appendices **75**

A Equity Results **77**
A.1 Equity L10 results: Success 77
A.2 Equity L10 results: Funding Ratio 81
A.3 Equity L10 results: Overfunding 85
A.4 Equity L14 results: Success 89
A.5 Equity L14 results: Funding Ratio 92
A.6 Equity L14 results: Overfunding 95

B Lending Results **99**
B.1 Lending L10 results: Success 99
B.2 Lending L10 results: Funding Ratio 103
B.3 Lending L10 results: Overfunding 107
B.4 Lending L14 results: Success 111
B.5 Lending L14 results: Funding Ratio 114
B.6 Lending L14 results: Overfunding 117

List of Tables

2.1 State-of-the-art on predicting the success of crowdfunding campaigns. . . 18

4.1 Calculation of daily returns r_j for a 10-day crowdfunding campaign . . 36
4.2 Summary table of variable definitions used in the subsequent analysis . 38
4.3 Summary of Hyperparameter Optimization for Key Models 44
4.4 Performance of AdaBoost on equity campaigns over a 10-day time series for predicting Success. 45
4.5 Performance of Histogram Gradient Boosting on equity campaigns over a 10-day time series for predicting Funding Ratio. 45
4.6 Performance of XGBoost on equity campaigns over a 10-day time series for predicting Overfunding. 46
4.7 Performance of Random Forest on lending campaigns over a 10-day time series for predicting Success. 46
4.8 Performance of LSTM on lending campaigns over a 10-day time series for predicting Funding Ratio. 46
4.9 Performance of Bagging on lending campaigns over a 10-day time series for predicting Overfunding. 47
4.10 Performance of Histogram Gradient Boosting on equity campaigns over a 14-day time series for predicting Success. 47
4.11 Performance of Histogram Gradient Boosting on equity campaigns over a 14-day time series for predicting Funding Ratio. 47
4.12 Performance of LSTM on equity campaigns over a 14-day time series for predicting Overfunding. 48
4.13 Performance of Histogram Gradient Boosting on lending campaigns over a 14-day time series for predicting Success. 48
4.14 Performance of LSTM on lending campaigns over a 14-day time series for predicting Funding Ratio. 48
4.15 Performance of Bagging on lending campaigns over a 14-day time series for predicting Overfunding. 49
4.16 TN, FP, FN and TP values for confusion matrices on a set of 10 success lending observations L10 for random forest 49
4.17 Summary of model benchmarking results 52
4.18 Example of Log-Transformed Variables 53

4.19 Model Performance Metrics 54
4.20 Comparison of equity crowdfunding campaigns in Italy and the USA . . 62

A.1 Performance of Decision Tree on equity campaigns over a 10-day time series for predicting Success. 77
A.2 Performance of k-NN on equity campaigns over a 10-day time series for predicting Success. 77
A.3 Performance of Logistic Regression on equity campaigns over a 10-day time series for predicting Success. 78
A.4 Performance of Bagging on equity campaigns over a 10-day time series for predicting Success. 78
A.5 Performance of Random Forest on equity campaigns over a 10-day time series for predicting Success. 78
A.6 Performance of AdaBoost on equity campaigns over a 10-day time series for predicting Success. 79
A.7 Performance of XGBoost on equity campaigns over a 10-day time series for predicting Success. 79
A.8 Performance of Histogram Gradient Boosting on equity campaigns over a 10-day time series for predicting Success. 79
A.9 Performance of MLP on equity campaigns over a 10-day time series for predicting Success. 80
A.10 Performance of LSTM on equity campaigns over a 10-day time series for predicting Success. 80
A.11 Performance of Decision Tree on equity campaigns over a 10-day time series for predicting Funding Ratio. 81
A.12 Performance of k-NN on equity campaigns over a 10-day time series for predicting Funding Ratio. 81
A.13 Performance of Ridge Regression on equity campaigns over a 10-day time series for predicting Funding Ratio. 81
A.14 Performance of Bagging on equity campaigns over a 10-day time series for predicting Funding Ratio. 82
A.15 Performance of Random Forest on equity campaigns over a 10-day time series for predicting Funding Ratio. 82
A.16 Performance of AdaBoost on equity campaigns over a 10-day time series for predicting Funding Ratio. 82
A.17 Performance of XGBoost on equity campaigns over a 10-day time series for predicting Funding Ratio. 83
A.18 Performance of Histogram Gradient Boosting on equity campaigns over a 10-day time series for predicting Funding Ratio. 83
A.19 Performance of MLP on equity campaigns over a 10-day time series for predicting Funding Ratio. 83
A.20 Performance of LSTM on equity campaigns over a 10-day time series for predicting Funding Ratio. 84

A.21 Performance of Decision Tree on equity campaigns over a 10-day time series for predicting Overfunding. 85
A.22 Performance of k-NN on equity campaigns over a 10-day time series for predicting Overfunding. 85
A.23 Performance of Ridge Regression on equity campaigns over a 10-day time series for predicting Overfunding. 85
A.24 Performance of Bagging on equity campaigns over a 10-day time series for predicting Overfunding. 86
A.25 Performance of Random Forest on equity campaigns over a 10-day time series for predicting Overfunding. 86
A.26 Performance of AdaBoost on equity campaigns over a 10-day time series for predicting Overfunding. 86
A.27 Performance of XGBoost on equity campaigns over a 10-day time series for predicting Overfunding. 87
A.28 Performance of Histogram Gradient Boosting on equity campaigns over a 10-day time series for predicting Overfunding. 87
A.29 Performance of MLP on equity campaigns over a 10-day time series for predicting Overfunding. 87
A.30 Performance of LSTM on equity campaigns over a 10-day time series for predicting Overfunding. 88
A.31 Performance of Decision Tree on equity campaigns over a 14-day time series for predicting Success. 89
A.32 Performance of k-NN on equity campaigns over a 14-day time series for predicting Success. 89
A.33 Performance of Logistic Regression on equity campaigns over a 14-day time series for predicting Success. 89
A.34 Performance of Bagging on equity campaigns over a 14-day time series for predicting Success. 90
A.35 Performance of Random Forest on equity campaigns over a 14-day time series for predicting Success. 90
A.36 Performance of AdaBoost on equity campaigns over a 14-day time series for predicting Success. 90
A.37 Performance of XGBoost on equity campaigns over a 14-day time series for predicting Success. 90
A.38 Performance of Histogram Gradient Boosting on equity campaigns over a 14-day time series for predicting Success. 91
A.39 Performance of MLP on equity campaigns over a 14-day time series for predicting Success. 91
A.40 Performance of LSTM on equity campaigns over a 14-day time series for predicting Success. 91
A.41 Performance of Decision Tree on equity campaigns over a 14-day time series for predicting Funding Ratio. 92

A.42 Performance of k-NN on equity campaigns over a 14-day time series for predicting Funding Ratio. 92
A.43 Performance of Ridge Regression on equity campaigns over a 14-day time series for predicting Funding Ratio. 92
A.44 Performance of Bagging on equity campaigns over a 14-day time series for predicting Funding Ratio. 93
A.45 Performance of Random Forest on equity campaigns over a 14-day time series for predicting Funding Ratio. 93
A.46 Performance of AdaBoost on equity campaigns over a 14-day time series for predicting Funding Ratio. 93
A.47 Performance of XGBoost on equity campaigns over a 14-day time series for predicting Funding Ratio. 93
A.48 Performance of Histogram Gradient Boosting on equity campaigns over a 14-day time series for predicting Funding Ratio. 94
A.49 Performance of MLP on equity campaigns over a 14-day time series for predicting Funding Ratio. 94
A.50 Performance of LSTM on equity campaigns over a 14-day time series for predicting Funding Ratio. 94
A.51 Performance of Decision Tree on equity campaigns over a 14-day time series for predicting Overfunding. 95
A.52 Performance of k-NN on equity campaigns over a 14-day time series for predicting Overfunding. 95
A.53 Performance of Ridge Regression on equity campaigns over a 14-day time series for predicting Overfunding. 95
A.54 Performance of Bagging on equity campaigns over a 14-day time series for predicting Overfunding. 96
A.55 Performance of Random Forest on equity campaigns over a 14-day time series for predicting Overfunding. 96
A.56 Performance of AdaBoost on equity campaigns over a 14-day time series for predicting Overfunding. 96
A.57 Performance of XGBoost on equity campaigns over a 14-day time series for predicting Overfunding. 96
A.58 Performance of Histogram Gradient Boosting on equity campaigns over a 14-day time series for predicting Overfunding. 97
A.59 Performance of MLP on equity campaigns over a 14-day time series for predicting Overfunding. 97
A.60 Performance of LSTM on equity campaigns over a 14-day time series for predicting Overfunding. 97

B.1 Performance of Decision Tree on lending campaigns over a 10-day time series for predicting Success. 99
B.2 Performance of k-NN on lending campaigns over a 10-day time series for predicting Success. 99

B.3 Performance of Logistic Regression on lending campaigns over a 10-day time series for predicting Success. 100
B.4 Performance of Bagging on lending campaigns over a 10-day time series for predicting Success. 100
B.5 Performance of Random Forest on lending campaigns over a 10-day time series for predicting Success. 100
B.6 Performance of AdaBoost on lending campaigns over a 10-day time series for predicting Success. 101
B.7 Performance of XGBoost on lending campaigns over a 10-day time series for predicting Success. 101
B.8 Performance of Histogram Gradient Boosting on lending campaigns over a 10-day time series for predicting Success. 101
B.9 Performance of MLP on lending campaigns over a 10-day time series for predicting Success. 102
B.10 Performance of LSTM on lending campaigns over a 10-day time series for predicting Success. 102
B.11 Performance of Decision Tree on lending campaigns over a 10-day time series for predicting Funding Ratio. 103
B.12 Performance of k-NN on lending campaigns over a 10-day time series for predicting Funding Ratio. 103
B.13 Performance of Ridge Regression on lending campaigns over a 10-day time series for predicting Funding Ratio. 103
B.14 Performance of Bagging on lending campaigns over a 10-day time series for predicting Funding Ratio. 104
B.15 Performance of Random Forest on lending campaigns over a 10-day time series for predicting Funding Ratio. 104
B.16 Performance of AdaBoost on lending campaigns over a 10-day time series for predicting Funding Ratio. 104
B.17 Performance of XGBoost on lending campaigns over a 10-day time series for predicting Funding Ratio. 105
B.18 Performance of Histogram Gradient Boosting on lending campaigns over a 10-day time series for predicting Funding Ratio. 105
B.19 Performance of MLP on lending campaigns over a 10-day time series for predicting Funding Ratio. 105
B.20 Performance of LSTM on lending campaigns over a 10-day time series for predicting Funding Ratio. 106
B.21 Performance of Decision Tree on lending campaigns over a 10-day time series for predicting Overfunding. 107
B.22 Performance of k-NN on lending campaigns over a 10-day time series for predicting Overfunding. 107
B.23 Performance of Ridge Regression on lending campaigns over a 10-day time series for predicting Overfunding. 107

B.24 Performance of Bagging on lending campaigns over a 10-day time series for predicting Overfunding. 108
B.25 Performance of Random Forest on lending campaigns over a 10-day time series for predicting Overfunding. 108
B.26 Performance of AdaBoost on lending campaigns over a 10-day time series for predicting Overfunding. 108
B.27 Performance of XGBoost on lending campaigns over a 10-day time series for predicting Overfunding. 109
B.28 Performance of Histogram Gradient Boosting on lending campaigns over a 10-day time series for predicting Overfunding. 109
B.29 Performance of MLP on lending campaigns over a 10-day time series for predicting Overfunding. 109
B.30 Performance of LSTM on lending campaigns over a 10-day time series for predicting Overfunding. 110
B.31 Performance of Decision Tree on lending campaigns over a 14-day time series for predicting Success. 111
B.32 Performance of k-NN on lending campaigns over a 14-day time series for predicting Success. 111
B.33 Performance of Logistic Regression Model on lending campaigns over a 14-day time series for predicting Success. 111
B.34 Performance of Bagging on lending campaigns over a 14-day time series for predicting Success. 112
B.35 Performance of Random Forest on lending campaigns over a 14-day time series for predicting Success. 112
B.36 Performance of AdaBoost on lending campaigns over a 14-day time series for predicting Success. 112
B.37 Performance of XGBoost on lending campaigns over a 14-day time series for predicting Success. 112
B.38 Performance of Histogram Gradient Boosting on lending campaigns over a 14-day time series for predicting Success. 113
B.39 Performance of MLP on lending campaigns over a 14-day time series for predicting Success. 113
B.40 Performance of LSTM on lending campaigns over a 14-day time series for predicting Success. L14 113
B.41 Performance of Decision Tree on lending campaigns over a 14-day time series for predicting Funding Ratio. 114
B.42 Performance of k-NN on lending campaigns over a 14-day time series for predicting Funding Ratio. 114
B.43 Performance of Ridge Regression on lending campaigns over a 14-day time series for predicting Funding Ratio. 115
B.44 Performance of Bagging on lending campaigns over a 14-day time series for predicting Funding Ratio. 115

B.45 Performance of Random Forest on lending campaigns over a 14-day time series for predicting Funding Ratio. 115

B.46 Performance of AdaBoost on lending campaigns over a 14-day time series for predicting Funding Ratio. 115

B.47 Performance of XGBoost on lending campaigns over a 14-day time series for predicting Funding Ratio. 116

B.48 Performance of Histogram Gradient Boosting on lending campaigns over a 14-day time series for predicting Funding Ratio. 116

B.49 Performance of MLP on lending campaigns over a 14-day time series for predicting Funding Ratio. 116

B.50 Performance of LSTM on lending campaigns over a 14-day time series for predicting Funding Ratio. 116

B.51 Performance of Decision Tree on lending campaigns over a 14-day time series for predicting Overfunding. 117

B.52 Performance of k-NN on lending campaigns over a 14-day time series for predicting Overfunding. 117

B.53 Performance of Ridge Regression on lending campaigns over a 14-day time series for predicting Overfunding. 118

B.54 Performance of Bagging on lending campaigns over a 14-day time series for predicting Overfunding. 118

B.55 Performance of Random Forest on lending campaigns over a 14-day time series for predicting Overfunding. 118

B.56 Performance of AdaBoost on lending campaigns over a 14-day time series for predicting Overfunding. 118

B.57 Performance of XGBoost on lending campaigns over a 14-day time series for predicting Overfunding. 119

B.58 Performance of Histogram Gradient Boosting on lending campaigns over a 14-day time series for predicting Overfunding. 119

B.59 Performance of MLP on lending campaigns over a 14-day time series for predicting Overfunding. 119

B.60 Performance of LSTM on lending campaigns over a 14-day time series for predicting Overfunding. 119

List of Acronyms

Acronym	Definition
AI	Artificial Intelligence
API	Application Programming Interface
FAIR	Findable-Accessible-Interoperable-Reusable
FN	False Negative
FP	False Positive
GDPR	General Data Protection Regulation
IPO	Initial Public Offering
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MLP	Multi-Layer Perceptron
MSE	Mean Squared Error
NLP	Natural Language Processing
RNN	Recurrent Neural Network
ROE	Return on Equity
R^2	Coefficient of Determination
SMOGN	Synthetic Minority Over-sampling Technique for Regression with Gaussian Noise
SMOTE	Synthetic Minority Over-sampling Technique
TN	True Negative
TP	True Positive
VC	Venture Capital
XGBoost	Extreme Gradient Boosting

Chapter 1

Introduction

Crowdfunding has emerged as a transformative financing mechanism, enabling collective contributions from diverse investors to support a wide range of initiatives. This funding model has gained significant global traction and has become a focal point of extensive research, underscoring its efficacy as an alternative to conventional financing options such as bank loans, venture capital, and private equity [1, 2, 3].

Through digital platforms, entrepreneurs can present detailed project proposals, allowing potential investors to evaluate opportunities and decide whether and how to contribute [4]. Depending on the campaign type, investors may acquire equity stakes in businesses or act as creditors in lending-based models, offering loans to project creators.

Crowdfunding is accessible to almost anyone with a viable idea. Yet, many platforms impose specific constraints, such as geographic limitations or restrictions based on the type of project, particularly on niche platforms tailored to specific industries. This inclusivity, coupled with certain platform guidelines, has contributed to crowdfunding's appeal as a flexible and innovative funding option.

In the early stages of a campaign, initial backers are often individuals within the creator's network. These supporters frequently base their decisions not only on the project's potential but also on personal trust and the entrepreneur's reputation. As the campaign progresses, social media plays a pivotal role in expanding its reach, helping the project move beyond immediate connections to attract interest from a broader audience, including acquaintances and strangers. Ultimately, the success of a crowdfunding campaign hinges on various factors, including the entrepreneur's social capital, the quality of communication, the sense of community identity fostered by the campaign, and, critically, the trust investors place in the project creator [5].

1.1 Objectives and Purposes

This research involves crowdfunding platforms using models, including equity crowdfunding and lending crowdfunding:

- Analysing Italian portals to provide a broader and more detailed perspective on crowdfunding dynamics in this country.

- Studying multiple websites simultaneously enables a more comprehensive and accurate analysis of crowdfunding campaigns.
- Using advanced machine learning methods, such as XGBoost and LSTM networks, to analyse crowdfunding data proves to be more effective than previous approaches.
- Focusing on time series data associated with the crowdfunding phenomenon addresses a commonly overlooked aspect due to the challenges in obtaining time-related data.
- Introducing new success indices that are more detailed than a simple Boolean index, facilitating a more thorough comparison between different crowdfunding campaigns.

This research proposes an innovative framework for interpreting crowdfunding dynamics by integrating investment time series, investor behaviour, and financial transactions from both equity and lending campaigns in Italy. While academic interest in crowdfunding is growing, a limited understanding of its mechanisms within specific national contexts underscores the need for a more refined, data-driven approach.

To address this, the study introduces two original datasets that capture temporal investment dynamics and investor activity, enabling a more granular analysis of Italian crowdfunding platforms. These datasets uncover distinctive behavioural patterns and structural elements that influence campaign performance, offering targeted insights into the Italian market.

Building on this empirical foundation, machine learning techniques are employed to predict campaign outcomes by leveraging time series data. These models detect latent patterns in investor behaviour over time, contributing to a deeper understanding of success drivers in both equity and lending contexts.

Success in crowdfunding is traditionally measured by reaching fundraising targets, but this study adopts a broader definition that also accounts for project visibility, supporter engagement, and reputational growth. This comprehensive view moves beyond binary outcomes, capturing the multi-dimensional nature of campaign effectiveness.

Such a perspective enhances predictive accuracy and supports practical decision-making: improved allocation of platform resources, optimised campaign design, and more effective strategies for engaging potential investors. By incorporating both financial and non-financial indicators, the approach supports sustainable growth for project creators and platforms alike.

Moreover, aligning predictive insights with broader business strategies enables companies to treat crowdfunding not merely as a funding mechanism but as a tool for market testing, product development, and customer engagement. This strategic integration is essential as the ecosystem continues to evolve.

The study further explores underexamined facets of crowdfunding, especially within alternative financing models. By examining multiple Italian platforms, it offers a richer comparative perspective and highlights how platform-specific features shape outcomes.

Advanced machine learning models—more adaptive than traditional analytical techniques—are applied to reveal complex correlations in campaign behaviour. The focus on time series data, often excluded due to access difficulties, enables a dynamic view of campaign evolution. New success indicators are proposed to allow more nuanced cross-campaign comparison, replacing simplistic binary classifications.

Despite growing academic and professional interest, crowdfunding's temporal and behavioural dimensions remain insufficiently explored. By focusing on investment trajectories and decision-making patterns, and analysing two campaigns led by Italian entrepreneurs, the study contributes actionable knowledge on how to better structure and manage crowdfunding initiatives.

In contrast to prior work based on static or aggregate data, this approach captures temporal investor behaviour through dynamic modelling. The result is a more refined understanding of how investor actions unfold over time and interact with campaign features, helping stakeholders attract higher investments and meet strategic objectives more effectively.

1.2 Overview of the Thesis

The remainder of this thesis is structured as follows:

- **Chapter 2 - State of the Art**: This chapter reviews the existing literature on crowdfunding, with a particular focus on equity and lending models. It examines previous studies on campaign success prediction and identifies key limitations, such as the scarce use of temporal data and oversimplified success metrics. The chapter underscores the need for more advanced predictive approaches that account for dynamic investor behaviour and multifaceted indicators of success.

- **Chapter 3 - Modelling for Crowdfunding Campaigns**: This chapter addresses the methodological challenges related to processing crowdfunding data, concentrating on key tasks such as handling missing values, standardising time series, and encoding categorical variables. The objective is to establish robust procedures for data cleaning and transformation, ensuring consistency and reliability for subsequent predictive modelling. Various types of missing data are discussed, along with preprocessing strategies designed to preserve data integrity and improve informational quality for machine learning applications.

  The chapter also explores the dynamic nature of crowdfunding campaigns, emphasising the distinction between static features available at launch and evolving variables that change over time. This distinction enables machine learning models to better capture behavioural patterns and campaign developments by leveraging temporal structures. The operations carried out in this stage lay the methodological foundation for the predictive modelling presented in the next chapter. Key techniques such as data imputation, feature construction, and the management of class imbalance are introduced as essential tools for enhancing the accuracy and robustness of predictive models. These techniques are particularly suited for preparing inputs for deep learning and advanced machine learning algorithms.

- **Chapter 4 - Experimental Framework**: This chapter represents the natural continuation of the methodology presented in Chapter 3, where the theoretical design is translated into a structured experimental workflow. The primary objective is to assess the forecasting capabilities of selected machine learning algorithms by applying them to purpose-built datasets covering both equity and lending crowdfunding domains, enriched through sophisticated feature engineering.

  The empirical strategy is articulated through a sequence of interdependent phases—data preparation, modelling, and evaluation—that collectively define the analytical pipeline. Each step is illustrated in detail to highlight the interplay between investment time series, behavioural investor features, and predictive models.

  The chapter begins with an in-depth analysis of pre- and post-launch variables, followed by a detailed presentation of the datasets, including their temporal organisation, data collection protocols, and encoding procedures. It then introduces derived features aimed at uncovering latent investment behaviours and campaign dynamics.

  Concerning outcome definition, the chapter proposes an expanded set of success metrics that move beyond binary classification. These include measures such as the funding ratio and overfunding level, allowing for a more nuanced assessment of campaign performance.

  The core of the chapter is dedicated to the benchmarking of predictive models. It presents the results of hyperparameter tuning, performance evaluation, and model interpretability analysis, using SHAP values to identify the most influential factors driving campaign success.

  Finally, the chapter offers a critical discussion of the experimental results, including a cross-country comparison between Italian and U.S. campaigns, and considerations regarding the implementation and scalability of the proposed models in real-world crowdfunding environments.

- **Chapter 5 - Conclusions**: The final chapter summarises the main findings of the research and discusses its theoretical and practical implications. It reflects on the contributions made to the study of crowdfunding dynamics and predictive modelling, outlining potential future developments. Particular attention is given to the integration of real-time behavioural indicators and the applicability of the proposed approach to broader contexts within alternative finance and digital entrepreneurship.

The following section outlines the structure of the thesis, guiding the reader through the analytical path adopted.

Chapter 2

State of the Art

The application of advanced data mining techniques, machine learning algorithms, and big data analytics has revolutionised our ability to extract meaningful insights from crowdfunding platforms [6]. This integrated approach significantly enhances the accuracy of success predictions. The utilisation of artificial intelligence and data-driven decision-making systems has become increasingly crucial in understanding and predicting crowdfunding dynamics [7]. As noted by Ziegler, 38% of European users face limited access to traditional financial services. Fintech platforms leverage digital systems and technological innovations to offer efficient and personalised financial solutions, enhancing financial inclusion for unbanked or underbanked individuals, especially in regions like Sub-Saharan Africa and the Asia-Pacific. The rise in annual transaction values in the lending and equity sectors indicates a paradigm shift in finance.

By employing machine learning models and time series analysis, this study delves into how temporal variations in investor patterns and the evolution of financial data influence the outcomes of crowdfunding initiatives. The application of deep learning architectures, such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, has shown promising results in capturing the temporal dynamics of crowdfunding campaigns [8].

These insights are crucial for optimising economic strategies and enhancing the management and success rates. In addition to contributing to the existing academic literature, this comprehensive approach equips industry practitioners with more effective tools for evaluating and managing these financial endeavours. Integrating natural language processing (NLP) techniques has enabled researchers to analyse unstructured data from project descriptions and social media interactions, providing a more holistic view of campaign success factors [9].

2.1 Crowdfunding and Digital Transformation

Crowdfunding enables individuals and businesses to raise funds through online platforms by collecting small contributions from a large number of backers. This mechanism allows project creators to present and finance a wide range of initiatives, while supporters may choose to contribute without expecting financial returns or in exchange for rewards, equity, or financial rights [10, 11].

These platforms act as digital intermediaries, connecting fundraisers with contributors through recommendation systems often powered by collaborative filtering algorithms tailored to user preferences and behaviours [8]. Their technological infrastructure includes scalable cloud environments, secure database management systems, and protocols that ensure transaction integrity at scale [12]. Real-time campaign monitoring is typically facilitated by interactive dashboards and analytics tools [6].

From a campaign design perspective, key elements include a financial goal, a detailed project description, and a clear deadline. Structural features such as openness (anyone can propose a project), reputation, and participatory engagement also play critical roles. Transparency is encouraged through the disclosure of contributions and funder identities to enhance credibility and trust [10].

Technological advancements further shape the crowdfunding landscape. Blockchain offers immutable transaction records that support transparency and accountability [13]. Smart contracts and tokenization facilitate efficient ownership transfer, especially in equity crowdfunding. In parallel, fraud and anomaly detection algorithms have become vital to ensure platform integrity [9].

Social media integration has significantly amplified campaign visibility. Platforms leverage APIs and sentiment analysis tools to track public perception and engagement in real time, offering dynamic feedback for campaign optimisation [7].

Crowdfunding also intersects with fields such as entrepreneurship, analytics, and behavioural science. Piva et al. [14] examined how signals of human capital affect campaign success in equity models. Natural language processing enhances these analyses by extracting semantic features from project narratives [9], while real-time analytics pipelines support data-driven decisions throughout campaign execution [6].

Equity crowdfunding, in particular, allows investors to acquire ownership stakes in startups and SMEs. It promotes capital access and rapid market validation outside traditional financial channels [15, 16, 17]. Investors benefit from algorithmic tools for risk assessment and portfolio optimisation, supporting informed decision-making in high-risk environments [12].

The predictive aspect of crowdfunding has attracted growing scholarly attention. Ralcheva and Roosenboom [18] developed a model incorporating variables such as equity offered, founder experience, and company age. Kuramoto et al. [19] studied the effect of venture capital and fiscal incentives like Japan's Angel Tax System. Edward et al. [20] analysed trust and equity values in Sharia-compliant peer-to-peer lending.

Crowdfunding can be categorised into four main models: donation-based, reward-based, equity-based, and lending-based [11]. Each model appeals to different motivations and involves specific structural and regulatory dynamics. Platforms increasingly deploy tailored algorithms to optimise campaign performance across these models [7].

A more nuanced platform classification—based on sector focus, geographical scope, and the adoption of technologies such as blockchain or AI—offers insights into market segmentation and strategic behaviour [13, 21]. Yet, challenges remain, including data scarcity, platform heterogeneity, and inconsistent reporting standards [22]. Addressing these issues requires methods such as web scraping, API integration, and stronger academic-industry collaboration. Advanced machine learning may also help infer missing platform attributes from partial data [23, 24], enhancing the quality of comparative studies and predictive modelling [25].

Recent work has expanded the methodological toolkit for crowdfunding analysis. Collaborative filtering techniques have been used to predict campaign outcomes [26], while multimodal data approaches—including text, image, and audio mining—improve platform strategy and user engagement [6, 27].

Finally, cultural factors play a key role in shaping campaign design and investor response. Research by Cicchiello et al. [28] and Gao et al. [29] shows how national culture influences communication style and emotional tone, with variations between Western and Eastern contexts. Sentiment and emotion analysis techniques further quantify these cultural dimensions [9]. Country-specific case studies—such as Kenya [30], Germany [31], Algeria [32], and Japan [33]—reveal how local economic, regulatory, and social environments affect crowdfunding outcomes.

2.2 Regulatory Framework

To fully understand crowdfunding and its regulatory implications, it is essential to explore the innovative approach taken by Italian authorities. Regulatory measures play a fundamental role in shaping crowdfunding platforms' operations, affecting investor confidence and participation. This structured framework has been instrumental in establishing crowdfunding as a credible financial tool, promoting transparency, and expanding fundraising opportunities.

Italy was the first European country to enact specific legislation for equity crowdfunding, creating a dual regulatory system that governs both platform managers and companies. This framework operates at two levels: primary legislation through amendments to the Consolidated Law on Finance (TUF), and secondary regulations established by CONSOB. Together, these laws aim to reduce information asymmetries and provide a secure environment for investors.

The regulatory journey began with Decree Law 179/2012, followed by CONSOB's 2013 regulation, which introduced a flexible framework focused on investor protection and transparency. Amendments to the TUF granted CONSOB supervisory authority, defined the responsibilities of platform operators, and set rules for public offerings via web portals (Art. 50-quinquies and Art. 100-ter). CONSOB's "Regulation on Raising Risk Capital" further clarified registration requirements, conduct standards, and investor protection measures, including a 5% minimum investment requirement from professional investors to build credibility and trust.

Over time, additional regulations broadened the scope of crowdfunding. The Growth Decree 3.0 (2015) extended access to innovative SMEs, while the 2017 Budget Law opened equity crowdfunding to all SMEs and introduced tax benefits for investors. Legislative updates also allowed limited liability companies (S.r.l.) to issue financial instruments through crowdfunding, aligning their fundraising capabilities with those of publicly held corporations (S.p.a.).

In 2019, CONSOB revised its regulations, allowing SMEs to issue equity shares and debt instruments, such as bonds, through crowdfunding. This revision also introduced a secondary market for trading SME shares and bonds, providing investors with greater liquidity and making crowdfunding a more accessible and appealing investment option. These changes foster a more dynamic investment environment, enabling companies to diversify their capital-raising methods and investors to broaden their portfolios.

Further incentives were introduced with the 2020 Relaunch Decree, which provided a 50% tax deduction for individuals investing in innovative startups and SMEs. In line with European standards, Directive EU 2020/1504 clarified crowdfunding's scope within the MiFID II framework, ensuring regulatory consistency for crowdfunding providers across Europe. This evolving legislative framework highlights the government's commitment to a secure, transparent, and adaptable crowdfunding environment. By enhancing investor protection, reducing information gaps, and creating new avenues for SME financing, these regulations have positioned crowdfunding as a valuable tool for companies and investors. To fully understand crowdfunding and its regulatory implications, it is essential to explore the innovative approach taken by Italian authorities. Regulatory measures play a crucial role in defining the operational frameworks of crowdfunding platforms, influencing investor confidence and participation. This structured framework has significantly contributed to positioning crowdfunding as a credible financial tool, promoting transparency, and expanding fundraising opportunities. Beyond the basic analysis of Italian regulations presented in this thesis, it is important to explore how the evolution of legislative frameworks may affect both the predictive performance of models and the availability of data. In particular, regulatory changes such as updates to transparency and disclosure requirements, modifications to the authorisation criteria for crowdfunding platforms, or the introduction of new investor protection rules can significantly impact the quality and granularity of collected data. Recent international studies highlight how stricter regulations on information disclosure and platform management enhance market transparency and investor protection, contributing to the reduction of fraudulent activities and increasing the overall reliability of collected data [34]. At the same time, overly restrictive regulation may limit the operational flexibility of platforms, thereby reducing the variety and timeliness of available data, which negatively affects the granularity of information used in predictive models. As also pointed out by Pohulak-Żołędowska and Wójcik-Czerniawska [35], regulation must carefully balance the need to protect investors with the necessity to stimulate market innovation and growth. The integration of specific regulatory indicators, such as variables reflecting the direct impact of legislative changes, could serve as an additional tool to adapt models to the dynamics of a continuously evolving regulatory context. In this way, a critical reflection on regulatory impact not only enriches the theoretical framework of the research but also provides concrete insights for future improvement strategies and the development of more robust and contextually sensitive models. Moreover, new technologies, particularly the tokenisation of corporate shares and the issuance of Security Token Offerings (STOs), represent a high-potential area of experimentation that could redefine the boundaries of crowdfunding and capital raising [36]. In a coherent and flexible regulatory environment, the integration of solutions based on blockchain and smart contracts would ensure, on the one hand, greater transparency and traceability of operations, thus improving both the quality and timeliness of collected data and on the other hand, broaden the base of potential investors with advantages in terms of liquidity and diversification. However, the absence of shared guidelines or regulatory standards can hinder the adoption of these innovations, generating uncertainty for platforms and investors, as well as impeding the formation of homogeneous and comparable datasets over time. In the medium to long term, a balanced legislative intervention—including experimental sandboxes, more granular disclosure requirements, and a clear framework of responsibilities—would help mitigate the risks of overregulation without stifling the growth of platforms, while also promoting a competitive advantage for the entire financial system and empirical research. Regarding security and privacy, the use of sensitive data in the context of crowdfunding and token issuance necessitates advanced encryption techniques, multifactor authentication, and ongoing compliance with personal data protection regulations, such as the GDPR in Europe. Indeed, solutions like blockchain and AI offer significant innovation potential, but demand delineated platform competencies and robust cybersecurity strategies to avoid uncertainties and vulnerabilities in digital markets. In summary, regulation has a dual impact: on one hand, it enables the development of more robust predictive models thanks to transparency and disclosure requirements; on the other hand, it can limit their generalizability, especially when data collection is constrained by privacy regulations or by highly heterogeneous legal frameworks across countries. Looking ahead, incorporating regulatory features into predictive models (e.g., capitalisation requirements, the presence of investor protection clauses) could enhance both model accuracy and their potential use by regulatory authorities as tools for pre-screening or ex-post analysis. Furthermore, regulatory developments related to tokenisation and smart contracts—if properly framed—could lead to more granular and timely data, further improving the effectiveness of predictive models in the crowdfunding domain.

2.3 Equity Crowdfunding

Equity crowdfunding consists of selling a stake in a business to multiple investors in return for capital. This model modernises traditional forms of corporate financing—such as private equity, venture capital, and angel investing—by leveraging digital platforms to enable broader participation, including from non-professional investors and consumers [37]. The process is supported by secure infrastructures, often based on blockchain for transaction integrity and smart contracts for equity management [13]. These platforms eliminate the need for direct, in-person negotiations by facilitating access to a global and diversified investor base. They employ advanced matching algorithms and machine learning models that assess investor preferences, risk tolerance, and investment histories to recommend suitable opportunities [38, 39, 40]. While these elements will be explored in detail in Paragraph **??**, here we focus on the key benefits these technologies provide to each actor within the equity crowdfunding ecosystem. Crowdfunding models are typically divided into four categories: donation, reward-based, lending, and equity [11]. Equity-based models distinguish themselves through the adoption of more advanced fintech tools, particularly for valuation, compliance, and risk assessment [41, 42].

Equity crowdfunding platforms offer strategic advantages enhanced by technological innovation:

1. **Investor Attraction:** AI-powered recommendation systems automate investor-project matching, improving efficiency and reducing resource needs [8].

2. **Ownership Protection:** Blockchain ensures secure, transparent records of share distribution [13].

3. **Diverse Investor Base:** Big data analytics support targeting and engagement strategies, broadening investor participation [43].

4. **Early Feedback:** Natural Language Processing (NLP) tools extract actionable insights from investor comments and market sentiment [44].

Compared to Initial Public Offerings (IPOs), equity crowdfunding is more accessible and cost-effective due to its digital nature. APIs allow seamless data exchange with regulatory entities and financial institutions, ensuring compliance and transparency.

Within this ecosystem, each stakeholder benefits from dedicated technologies:

- **Fundraisers:** Leverage data visualisation and predictive analytics to craft compelling campaigns and estimate funding outcomes. Historical data guides optimisation strategies [45].

- **Backers:** Rely on AI-based recommendations and blockchain smart contracts for secure investment processes. Social network analysis identifies behavioural patterns among investor groups [46].

- **Platforms:** Operate as intermediaries using robust cybersecurity protocols, cloud scalability, and fraud detection algorithms based on machine learning [47].

Recent innovations have further refined platform capabilities:

1. **Visual Signaling:** Computer vision algorithms assess multimedia content to forecast campaign success [46].

2. **Social Media Integration:** NLP and sentiment analysis gauge campaign reception and forecast impact [48].

3. **Predictive Modelling :** Ensemble and deep learning methods estimate success probabilities and optimal funding targets [49].

4. **Blockchain Integration:** Smart contracts automate processes such as dividend distribution [13].

5. **Big Data Analytics:** Extract patterns and success factors from large volumes of historical data [50].

These technological solutions not only facilitate the funding process but also contribute to innovation in business models and product development. By leveraging AI and data analytics, equity crowdfunding platforms create collaborative innovation environments, where funders act as knowledge generators through continuous data-driven exchanges with entrepreneurs [49].

Return on Equity (ROE) is widely regarded as a primary profitability indicator, reflecting how effectively a company uses shareholders' capital to generate profit. A higher ROE implies better management of equity capital, which may signal a higher probability of long-term success [51]. My study emphasises ROE as it offers a clear measure of a company's financial health, particularly valuable in crowdfunding contexts where detailed financials are often limited [52]. Modern crowdfunding platforms increasingly use data analytics and machine learning to provide real-time ROE calculations and comparisons, equipping investors with valuable insights. Key functionalities of these tools include:

- Automated ROE Calculations: Platforms retrieve financial data via APIs, allowing for up-to-date ROE metrics [52].

- Comparative Analysis: Machine learning algorithms evaluate a company's ROE relative to industry standards, offering contextual insights [51].

- Predictive Modeling: Advanced analytics forecast potential ROE trends based on historical and market data, supporting investors in assessing risk [52].

For example, a company with a 15% ROE indicates it generates 15 cents in profit for every euro invested, outperforming a similar company with a 5% ROE. Such insights enable investors in equity crowdfunding to identify more promising opportunities [52].

Additionally, ROE can play a role in mitigating the risks associated with equity crowdfunding, a domain where startups face higher chances of failure than established companies. As a metric of profitability, ROE allows investors to make informed decisions based on measurable data [52, 51]. To further reduce investment risks, crowdfunding platforms integrate:

- AI-Driven Risk Scoring: Machine learning models assess financial indicators, including ROE, to develop comprehensive risk profiles [51].
- Blockchain Transparency: Blockchain technology ensures transparency and immutability of financial data [13].
- NLP for Data Validation: NLP algorithms analyse company reports and news to corroborate ROE data with qualitative information [9].

Overall, the integration of Return on Equity as a key profitability metric in equity crowdfunding aids investors in making more informed choices and strengthens the platforms' capacity to provide transparent, data-driven insights. This approach aligns with the broader goals of equity crowdfunding, enhancing decision-making processes and supporting a more sustainable investment ecosystem.

2.4 Lending Crowdfunding

Lending crowdfunding, also known as peer-to-peer (P2P) lending or marketplace lending, is a form of crowdfunding where a large number of individual investors collectively finance loans to individuals or businesses through an online platform. This innovative financial model, powered by advanced digital technologies, has significantly opened up new opportunities for both entrepreneurs and private individuals by simplifying access to financing [53]. The concept of lending crowdfunding was first actualised in the United Kingdom in 2005, with the emergence of the first P2P lending platforms. These platforms aimed to provide personal loans, representing a tech-driven alternative to traditional banking services. Early adopters of P2P lending recognised the potential to streamline the borrowing process through digital means, reduce overhead costs associated with traditional banking, and provide more attractive returns to lenders.

The operational framework of these platforms involves a comprehensive evaluation process, where the platform uses advanced algorithms and data analytics to assess the creditworthiness of borrowers and categorise projects into various risk levels [7]. This data-driven risk assessment helps mitigate potential losses for lenders by offering detailed information about each borrowing entity. Lenders then indicate the interest rates at which they are willing to lend, and borrowers select the most encouraging offers through an automated matching system. This reverse auction mechanism, facilitated by intelligent algorithms, benefits borrowers by potentially lowering the interest rates they must pay.

A key feature of lending crowdfunding is the centralised repayment management system, built on a robust financial technology infrastructure. Borrowers repay their loans through direct electronic payments to the platform, which then automatically distributes the repayments to the individual lenders. This ensures a transparent and efficient process for all parties involved, enabled by secure transaction processing systems. Additionally, these platforms often provide real-time monitoring and reporting tools, allowing lenders to track their investments and returns through user-friendly digital interfaces.

Lending crowdfunding offers several strategic benefits, making it an attractive option for both borrowers and lenders. One of the most significant advantages is the ease of access to funds, facilitated by streamlined digital processes. Entrepreneurs and individuals can secure financing much faster and more easily compared to traditional banking channels, which are often slower and involve more rigid procedures. This tech-enabled model is particularly advantageous in sectors like real estate, where developers can quickly raise the necessary funds for construction or renovation projects without the need to purchase property outright.

Inclusivity is another positive aspect of lending crowdfunding, enhanced by the accessibility of online platforms. The low minimum loan threshold encourages a wide range of investors to participate, making the fundraising process more inclusive and democratised. This model allows small investors to engage in funding opportunities with modest investments through user-friendly web and mobile interfaces, broadening the pool of potential financiers and diversifying the risk.

Efficiency and transparency are further strengths of lending crowdfunding, underpinned by advanced digital technologies. Online platforms facilitate efficient connections between lenders and borrowers, streamlining the process and making it more transparent through real-time data sharing and automated transactions. The reverse auction mechanism, powered by sophisticated matching algorithms, often results in competitive interest rates for borrowers, making financing more affordable.

However, lending crowdfunding is not without its drawbacks. One significant risk for lenders is liquidity limitation, as the digital nature of these investments doesn't yet allow for easy secondary market trading. Once funds are committed to a loan, lenders cannot request their return before the loan matures, which can pose a liquidity risk. Despite the rigorous evaluation process supported by AI and machine learning, there is always a risk of borrower default, which could result in lenders losing their investments. Market fluctuations can also impact interest rates and loan demand, affecting both lenders and borrowers and requiring platforms to constantly update their algorithms and risk models.

Regulatory risks are another consideration, as the legal environment for lending crowdfunding is still evolving, particularly in response to rapid technological advancements. Changes in regulations can affect the operation and viability of platforms, requiring them to adapt their technological infrastructure accordingly. Investors also have limited control over how their money is used once the loan is given, relying heavily on the platform's digital due diligence and monitoring systems.

Despite these risks, the strategic benefits of lending crowdfunding, such as technology-enabled ease of access to funds, digital inclusivity, algorithmic efficiency, and competitive interest rates, make it a valuable tool in the modern financial landscape. This model continues to evolve, offering new opportunities and challenges as it grows, driven by ongoing advancements in financial technology and data science [24].

2.5 Research Opportunities

Equity-based and lending crowdfunding platforms remain significantly underexplored, particularly in data-driven analyses.

This gap highlights the need for advanced machine learning applications to better understand the unique dynamics of these platforms [22].

The success of certain reward-based platforms has generated detailed studies leveraging project data to examine success factors, utilising natural language processing and computer vision to analyse project descriptions and visual content [54, 55, 56, 6, 8, 7, 9].

However, studies on lending and equity platforms have primarily focused on economic and legal perspectives, with limited application of time series analysis or sophisticated machine learning models beyond standard linear and logistic regression [55, 8].

Transfer learning and multi-task learning represent promising approaches to extend insights across platforms and different crowdfunding models [7].

Another noteworthy finding from the study of crowdfunding phenomena is the varied interpretation of campaign success.

Many studies measure success using a Boolean approach, assessing only whether a campaign reached its funding goal [54, 6, 56, 7, 9].

However, some research has proposed alternative indicators, such as the final amount raised or the ratio of funds raised to the minimum goal [22, 57]. These approaches suggest the potential to develop more refined success metrics that better capture the complexity of crowdfunding outcomes.

This topic will be discussed in greater detail in the following chapters, where different approaches to measuring crowdfunding success will be evaluated.

Additionally, recent research emphasises the influence of national culture on crowdfunding practices. Studies by Cicchiello, Bernardino, and Xiang have explored how cultural factors affect various aspects of crowdfunding, from adoption and perception to active engagement.

Methods like cross-cultural text analysis and multilingual natural language processing have proven effective in quantifying these cultural impacts across diverse markets [29].

Table 2.2 provides a comparative overview of key articles, illustrating various methods to define and measure crowdfunding success. In contrast, research such as Yuji Honjo Kurihara's study on Japanese equity crowdfunding applies binary response models (logit and probit) and survival analysis to explore campaign success factors.

Combining traditional statistical techniques with machine learning, this methodological approach demonstrates the potential for richer insights into crowdfunding dynamics [33].

To conclude, Table 2.1 offers a comprehensive summary of the main characteristics of various articles on crowdfunding. This table is designed to streamline the comparison of critical insights, methodologies, and outcomes across existing literature, focusing on the data science and machine learning techniques employed in each study.

Table 2.1: *State-of-the-art on predicting the success of crowdfunding campaigns.*

Ref.	Year	Typology	Methodology	Success measure	Platform(s)	Number of dataset instances	Nation(s)	Time series
(*)	-	Equity/Lending	ML/DL	Boolean(98%) + New indexes	40 platforms	907+1084	Italy	Yes
[54]	2014	Reward	Logistic Reg	Boolean	Kickstarter	48,500	-	No
[58]	2016	Equity	Logit/Probit Reg	Boolean	Seedrs	636	UK	No
[59]	2016	Reward	Linear Reg	Funded/Min amount†	Kickstarter	15,824	-	No
[55]	2018	Reward	Gaussian Process	No (only forecast)	Indiegogo	14,143	-	Yes
[14]	2018	Equity	Probit Reg	Boolean	SiamoSoci	129	Italy	No
[6]	2019	Reward	ML/DL	Boolean	Kickstarter	N/A (webrobots.io)	-	No
[56]	2019	Reward	Logistic Reg	Boolean	Kickstarter	9,962 (webrobots.io)	-	No
[57]	2020	Donation	ML/DL	Funded/Min amount	GoFundMe	9,948	-	No
[8]	2020	Reward	ML/DL	No (Only forecast)	Indiegogo	14,143	-	Yes
[26]	2020	Reward	ML	Boolean	Kickstarter	N/A	-	Yes
[18]	2020	Equity	Logistic Reg	Boolean	Crowdcube/Seedrs	2,171 (TAB)	UK	-
[60]	2020	Equity	Linear Reg	Funded/Min amount†	Crowdcube	299	UK, Spain	No
[61]	2020	Equity	Linear Reg	Funded/Min amount†	-	192	UK	No
[19]	2022	Equity	Probit Reg	Boolean	FUNDINNO	242	Japan	No
[32]	2022	Equity	Linear Reg	Investors	FUNDINNO	63	Japan	No
[62]	2022	Equity	Logit/Tobit Reg	Boolean + Funded/Min amount	35 platforms	580	USA	No
[7]	2022	Reward	ML/DL	Boolean	Kickstarter	126,593	-	No
[9]	2022	Reward	ML/DL	Boolean	JD Finance	15,384	China	No
[63]	2022	Reward	Logistic Reg	Boolean	Kickstarter	563 (drones)	-	No
[64]	2023	Reward	ML/DL	Boolean	Kickstarter	16,439 (webrobots.io+)	-	No
[65]	2023	Reward	ML	Boolean	Kickstarter	250	-	No
[66]	2023	Reward	ML	Boolean	Kickstarter	51,513 (webrobots.io)	-	No
[67]	2023	Equity	Linear Reg	Funded/Min amount	Crowdcube	954	UK	No
[68]	2023	Equity	Linear Reg	Money and Investors	4 platforms	190	France	No
[20]	2023	P2P Lending	Linear Reg	Money	Ammana	1,153	Indonesia	No

(*) is this thesis.
"†": only for overfunding.
"-": lacks of information and/or uncertainty
Notations:
ML: Machine Learning; DL: Deep Learning; Reg: Regression.

Chapter 3

Modelling for Crowdfunding Campaigns

This chapter addresses the challenges related to processing crowdfunding campaign data, focusing on key aspects such as handling missing values, standardising time series, and encoding categorical variables. The goal is to define effective methods for data cleaning and transformation, ensuring consistency and reliability for predictive modelling.

In this context, different types of missing data are analysed, and preprocessing techniques aimed at preserving data integrity and enhancing informational quality for machine learning applications are illustrated.

Moreover, the chapter examines the dynamic characteristics of crowdfunding campaigns, highlighting the fundamental distinction between static information available at the campaign's launch and dynamic data, which evolves. This distinction allows artificial intelligence models to more effectively recognise investor patterns and campaign trends by leveraging temporal dependencies and pattern recognition.

The work carried out in this phase lays the methodological groundwork necessary for the predictive modelling experiments discussed in later chapters. Techniques such as data imputation, feature engineering, and managing imbalanced datasets are presented as fundamental for optimising model accuracy and robustness. These methodologies prove particularly useful for structuring inputs intended for deep learning models and advanced machine learning algorithms, which will be explored in the next chapter.

3.1 Data Preparation for predictive modelling

To ensure consistency and analytical comparability across crowdfunding campaigns, a series of preprocessing steps was implemented to transform raw inputs into a structured format aligned with the forecasting goals. These operations are fundamental to enabling reliable model training and reproducibility. In line with the study's predictive strategy, a modular framework was adopted, articulated into four main phases:

data preparation, input structuring, model development, and performance evaluation (Figure 3.1). This high-level overview introduces the operational procedures detailed in the following sections.

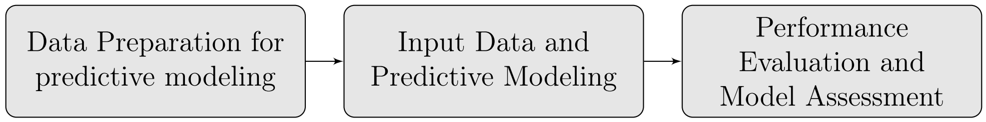

Figure 3.1: Modular structure of the modeling framework: from data preparation to model evaluation

Following the feature engineering process, the resulting variables were systematically organised into three main categories—structural, temporal, and textual—each representing a distinct informational dimension relevant to campaign performance.

- **Structural variables** encompass contextual attributes such as platform type, funding goals, and sector classification, offering a foundational understanding of the campaign environment.
- **Temporal variables** reflect the dynamic evolution of campaign activity, capturing metrics such as daily funding inflows and investor engagement over time.
- **Textual variables**, extracted from campaign titles and descriptions, encapsulate communicative strategies that may influence investor behaviour and decision-making.

Each category was processed through a standardised transformation pipeline, which included cleaning, encoding, and formatting procedures. This workflow was designed to ensure that all features—regardless of type or source—could be seamlessly incorporated into a variety of machine learning models, from traditional ensemble methods to advanced deep learning architectures. The specific preprocessing techniques—including encoding schemes for categorical and textual data, as well as resampling strategies to address class imbalance—are detailed in the following section. There, we clarify how each input category contributes to the predictive modelling framework and how these transformations enhance both model performance and generalisability.

3.2 Input Data and Predictive Modeling in Crowdfunding Dynamics

After addressing the critical aspects of data quality and preprocessing, the next phase involves a comprehensive analysis of both the structural and temporal characteristics of the dataset. Within the crowdfunding environment, input variables typically fall into two distinct categories: static and dynamic. Static variables—such as the economic sector, initial funding target, or selected platform—remain constant throughout the campaign and provide a baseline for assessing project potential. In contrast, dynamic variables evolve over time, capturing real-time investor interest, funding velocity, and behavioural responses to campaign updates.

Effectively capturing these dynamic features is essential for predictive modelling, as they encode the temporal fundraising trajectory that underpins short-term forecasting. Time series analysis enables the identification of recurring patterns, investment plateaus, and behavioural shifts triggered by strategic interventions or promotional bursts. Modelling these temporal dynamics accurately enhances both the predictive performance and interpretability of the models, reducing the influence of random noise and temporal misalignments.

A complementary dimension concerns the textual content embedded in campaign titles and descriptions. These elements serve as the first interface between the project and potential backers, playing a critical role in shaping perceptions and influencing engagement. Linguistic attributes—including title length, capitalisation, and the presence of numerical elements—are systematically extracted and encoded to retain semantic relevance while ensuring compatibility with supervised learning models.

Categorical variables, such as platform choice and sector classification, also exert significant influence on campaign outcomes. Appropriate encoding techniques have been employed to preserve their predictive contribution and minimise the risk of introducing modelling bias. When integrated with temporal and textual features, these categorical attributes contribute to a robust and multidimensional input space.

Beyond binary success indicators, this study also incorporates continuous performance metrics, including funding ratios and overfunding levels. These refined targets enable a more granular interpretation of campaign dynamics and facilitate the development of context-sensitive predictive models. Class imbalance, frequently observed in crowdfunding datasets, has also been addressed through the use of resampling techniques. Specifically, Synthetic Minority Over-sampling Technique (SMOTE) for classification tasks and SMOGN for regression settings were employed to mitigate skewness and ensure balanced learning. These strategies enhance the reliability of model training by improving representation across outcome classes. The following subsection presents the predictive modelling framework adopted in this study, outlining the rationale for algorithm selection and illustrating how static, dynamic, and semantic features are jointly leveraged to forecast campaign performance. A comparative evaluation is then conducted to assess the models' capacity to capture the multifaceted nature of crowdfunding dynamics across both classification and regression scenarios.

3.2.1 Resampling Techniques for Crowdfunding Data

Crowdfunding datasets—particularly in equity and lending contexts—often suffer from imbalanced outcome distributions, where highly successful campaigns are significantly outnumbered by less successful or failed ones. To mitigate this issue, dedicated resampling techniques were employed to enhance model training and generalisation.

For classification tasks, such as predicting campaign success, the Synthetic Minority Over-sampling Technique (SMOTE) is used. This method generates synthetic examples of the minority class to balance the dataset and improve the model's sensitivity to underrepresented outcomes.

For regression tasks—such as forecasting the funding ratio or overfunding amount (see Section 4.2 for definitions)—we applied SMOGN (Synthetic Minority Over-sampling Technique for Regression with Gaussian Noise). This technique is particularly well-suited to crowdfunding scenarios, as it emphasises the tails of the distribution, where exceptional or extreme performances are often observed but sparsely represented.

These resampling strategies proved especially relevant for equity crowdfunding, where campaigns may experience highly skewed distributions due to large investments by a small number of backers, and for lending platforms, where repayment dynamics and loan sizes vary substantially. By addressing class imbalance in both domains, the models were able to capture a more complete range of outcomes, ultimately improving both predictive accuracy and robustness.

3.2.2 Machine Learning Models

The selection of machine learning models plays a central role in predicting crowdfunding campaign outcomes. This study adopts a dual strategy that combines traditional algorithms with deep learning approaches, aiming to balance interpretability with the capacity to model complex behavioural dynamics. Classical models—such as decision trees, logistic regression, and ensemble methods like Random Forest and XGBoost—were applied to both static features and hybrid representations that incorporate early-stage temporal signals (e.g., first-day investment activity). These models established strong baseline performances and highlighted significant correlations between initial campaign momentum and final outcomes. To thoroughly investigate the various aspects of the prediction task, a diverse set of traditional supervised learning models was employed:

- **Decision Tree**: A non-parametric, supervised method for classification and regression that relies on decision rules derived from input features.

- **k-Nearest Neighbours (k-NN)**: Computes the distance between a new input and existing data points, returning the output based on the values of its k nearest neighbours.

- **Logistic Regression**: Used to predict campaign success as a binary outcome.

In this study, logistic models incorporated both static attributes and partial time-based variables such as day-one investment.

- **Ridge Regression**: Applied to continuous targets such as funding ratios and overfunding. This regularised linear model mitigates multicollinearity and enhances prediction stability.

- **Bootstrap Aggregating (Bagging)**: An ensemble approach that improves robustness by combining multiple weak learners, especially useful in the presence of outlier campaigns.

- **Random Forest**: An ensemble of decision trees used for both classification and regression. It demonstrated strong performance while retaining a degree of interpretability, which is essential for understanding static predictors.

- **Adaptive Boosting (AdaBoost)**: A boosting technique that iteratively focuses on misclassified instances, enhancing classification accuracy in campaign success prediction.

- **Extreme Gradient Boosting (XGBoost)**: A highly optimised implementation of gradient boosting that outperformed other models in capturing non-linear patterns within static campaign attributes.

- **Histogram Gradient Boosting**: A scalable variant of gradient boosting, effective for processing high-dimensional static data, improving training speed without compromising accuracy.

While these traditional models excelled at leveraging structural campaign features, they are inherently limited in capturing the evolving nature of crowdfunding campaigns particularrly the sequential dynamics that unfold over time.

In contrast, LSTM networks are specifically designed for sequential data analysis, making them particularly suitable for predicting crowdfunding outcomes where key variables evolve over time. Daily changes in investment amounts, variations in investor counts, and cumulative returns can be integrated into LSTM architectures to uncover dynamic patterns that static models cannot capture.

- **Multi-Layer Perceptron (MLP)**: A fully connected feedforward neural network used as a baseline deep model. The MLP was applied to static features, serving as a reference point for evaluating more advanced architectures.

- **Long Short-Term Memory (LSTM)**: A recurrent neural network designed to capture long-term temporal dependencies. In this study, LSTM models were trained on daily investment time series to trace the temporal evolution of campaign performance. Their ability to model sequences enabled the identification of shifts in momentum and investor engagement that traditional models could not detect.

Although LSTM networks require greater computational resources and are less interpretable than classical methods, their strength lies in capturing dynamic fundraising patterns over time. This makes them particularly suited for forecasting daily investment trends and detecting behavioural shifts throughout a campaign's duration. The selection of predictive models was carefully aligned with the analytical objectives of the study. Models such as XGBoost and Random Forest proved effective in extracting information from structural features, due to their interpretability and efficiency. Conversely, LSTM networks excelled in modelling sequential investment behaviour. In several cases, a hybrid approach—integrating static campaign data with partial time-series inputs—offered a balanced perspective that enhanced both accuracy and interpretability. This combined strategy allowed the models to account for early momentum signals while retaining the explanatory clarity of structural variables.

Having outlined the core machine learning algorithms and their respective roles in analysing crowdfunding dynamics, the next section presents a comparative evaluation of these models. The benchmarking focuses on their performance across key prediction tasks and discusses the practical implications for forecasting success in equity-based campaigns.

3.2.3 Temporal Feature Integration and Modelling Strategy

This section compares machine learning models, evaluating their effectiveness in predicting campaign success and understanding investor behaviour over time. Both traditional and advanced machine learning models were assessed to predict crowdfunding campaign outcomes. Decision tree-based algorithms—such as XGBoost and Random Forest—were employed to analyse static campaign attributes (e.g., economic sector, platform type, and funding goal), whereas recurrent neural networks (particularly LSTM) were utilised to capture the temporal evolution of fundraising patterns. This methodological choice balances model interpretability with the ability to effectively capture dynamic temporal behaviors, which proves crucial for understanding how investor activity evolves over time.

Throughout each iteration, a consistent core set of additional data was maintained to ensure that emerging temporal information was evaluated within a stable contextual framework. In the iterative training algorithm, `m` represents the time series of money raised, `i` is the number of investors, $\mathtt{r_m}$ denotes returns on money raised, $\mathtt{r_i}$ refers to returns on the number of investors, and `d` captures daily changes in the amount raised. Additionally, `additionalData` integrates other relevant contextual information such as campaign duration, categorisation (e.g., equity or lending), and platform-specific features.

In addition to the static attributes of each campaign, dynamic variables play a crucial role in capturing the evolving nature of crowdfunding activities. These variables are derived from daily transactional data and reflect the temporal progression of fundraising dynamics. A formal summary of all variables used in the subsequent analysis is provided in Table 4.2, which outlines their definitions and analytical roles within the study.

Specifically, the following time series were extracted:

- $\mathtt{m}$: cumulative amount of money raised over time,
- $\mathtt{i}$: number of investors active on each day,
- $\mathtt{r_m}$: return on the invested amount,
- $\mathtt{r_i}$: return related to investor participation,
- $\mathtt{d}$: daily change in the amount raised.

These dynamic variables were derived from raw platform data through a series of preprocessing steps, including normalisation, noise reduction, and temporal smoothing, in order to mitigate the effects of outliers and short-term anomalies. By structuring this information into sequential time series, it was possible to integrate them into predictive models based on recurrent architectures, particularly Long Short-Term Memory (LSTM) networks. This integration is essential for modelling both short-term fluctuations and longer-term trends, ultimately improving forecasting accuracy.

In addition to providing a detailed view of daily fundraising patterns, dynamic data enhance the overall model by supplying a continuous stream of real-time signals that enrich and complement the campaign's static characteristics. Finally, a thorough feature engineering process was applied to transform these raw time-based indicators into meaningful variables compatible with the model requirements. The iterative training strategy enabled the progressive incorporation of new data as the campaign advanced, refining predictions based on the most recent information. This systematic treatment of dynamic variables constitutes a key component of the predictive framework, ensuring a robust and realistic representation of campaign evolution in crowdfunding environments.

The comparative strengths and weaknesses of these models will be explored through targeted empirical evaluations, highlighting their ability to capture the complexity of crowdfunding dynamics in both classification and regression settings.

3.3 Performance Evaluation and Model Assessment

The evaluation of machine learning models in this study relies on performance metrics specifically tailored to the nature of each prediction task. As the analysis includes both classification and regression problems, distinct sets of indicators were used to provide meaningful insights into model effectiveness and reliability.

For classification models—designed to estimate whether a campaign will reach its funding goal—the following metrics were employed:

- **Accuracy**: The proportion of correctly classified campaigns out of the total number of cases.

- **Precision**: The proportion of predicted successful campaigns that are truly successful, particularly important for minimizing false positives.
- **Recall**: The proportion of actual successful campaigns that were correctly identified by the model, capturing sensitivity.
- **F1-score**: The harmonic mean of Precision and Recall is a crucial metric, particularly effective in situations with class imbalance..

These metrics play a central role in assessing model performance within the iterative predictive framework adopted in this study. As models are progressively updated with new data throughout each campaign, monitoring changes in Precision and Recall becomes especially important to mitigate the risk of biased or misleading classifications.

For regression tasks—focused on estimating continuous outcomes such as the funding ratio or the overfunding amount—the following metrics were employed:

- **R-squared ($\mathbf{R^2}$)**: Indicates the proportion of variance in the target variable that is explained by the model. Higher values suggest a better fit.
- **Adjusted R-squared ($\mathbf{R^2_{adj}}$)**: A refined version of R^2 that accounts for the number of predictors, mitigating the risk of overfitting.
- **Mean Squared Error (MSE)**: Represents the average squared difference between actual and predicted values, assigning greater weight to larger errors.
- **Mean Absolute Error (MAE)**: The average of the absolute differences between predictions and actual values, offering a more interpretable view of error magnitude.

The selected performance metrics ensure consistent evaluation of predictive accuracy across both classification and regression tasks. This evaluation framework serves as the basis for the comparative analysis of modelling approaches discussed in the following chapter.

Chapter 4

Experimental framework

This chapter represents the natural continuation of the methodological framework outlined in Chapter 3, where the theoretical models and design principles are translated into a structured experimental workflow. The main objective is to evaluate the predictive capacity of the proposed machine learning models by applying them to purpose-built datasets related to both equity and lending crowdfunding, enriched through advanced feature engineering techniques.

The high-level modular structure introduced in Figure 3.1—comprising the three macro-phases of data preparation, modeling, and evaluation—provides the conceptual backbone for the empirical strategy. This framework is further detailed in Figure 4.1, which operationalises each phase through a sequence of interdependent modules. The refined diagram captures the dynamic interaction between the temporal features of the investment, investor behavioural indicators, and predictive models, offering a more granular and actionable representation of the entire pipeline.

The analysis begins with a thorough examination of pre- and post-launch features, followed by a detailed presentation of the datasets employed in the study, with attention to their temporal organisation, data acquisition protocols, and encoding strategies. The subsequent section outlines the creation of derived variables aimed at capturing latent investment behaviours and campaign dynamics.

In terms of outcome definition, the chapter introduces a set of success indicators that go beyond the traditional binary outcome, including the funding ratio and overfunding level, allowing for a more articulated and realistic assessment of campaign performance.

The core part of the chapter is dedicated to the comparative evaluation of predictive models, reporting results from hyperparameter tuning, performance metrics, and model interpretability analyses using SHAP values to identify the key drivers of success.

The chapter concludes with a critical discussion of the experimental findings, including a cross-country comparison between Italian and U.S. campaigns, and considerations on the implementation and scalability of the proposed models in real-world crowdfunding platforms.

4.1 Data preparation for predictive modeling

4.1.1 Pre- and Post-Launch Features

In crowdfunding dynamics, campaign performance is shaped by a combination of static and dynamic factors that can be temporally categorised into two main groups: pre-launch features and post-launch features [69]. This distinction reflects the two phases of campaign evolution—before and after public launch—and is the foundation for the predictive modelling framework introduced in Chapter 3.2. Understanding the nature and timing of each feature is fundamental for building robust and flexible forecasting models:

- **Pre-launch features:** variables defined before the campaign's start remain fixed throughout its duration.
- **Post-launch features:** time-dependent variables that evolve dynamically as the campaign progresses.

Each group contributes distinct predictive value: pre-launch features offer early insights into the campaign's structure, scope, and strategic setup, while post-launch features capture real-time investor behaviour and funding dynamics. Table 4.2 titled *Summary table of variable definitions used in the subsequent analysis* reports the full list of variables used in this study and their definitions and encoding types. These variables have been extracted from two distinct datasets—one for equity-based campaigns and one for lending-based campaigns—reflecting the methodological separation applied in the modelling process. This distinction is conceptual and central to the design of the predictive architecture adopted in this work. (See figure 4.1 module Pre and Post Launch Features)

Pre-launch features represent the initial conditions of each campaign and serve as early indicators of its underlying potential. These include key financial parameters such as `Min_amount` (minimum funding target), `Max_amount_Imputed` (imputed upper funding limit), and `Minimum_investment` (minimum individual contribution). In addition, several categorical variables—`Platform`, `Region`, `EP Categories`, and `Sector`—capture platform-specific, geographic, and sectoral characteristics, offering contextual information that may influence investor behaviour.

Textual features extracted from campaign titles, including `lenName` (number of characters), `NumWordName` (number of words), `UpperCaseinName` (uppercase letters), and `NuminName` (numerical characters), provide linguistic cues potentially linked to engagement levels. Temporal variables such as `year_Startdate`, `month_Startdate`, and `dayofweek_Startdate` situate each campaign within a specific calendar context, capturing seasonality and timing effects. For equity campaigns, the variable `Pre-money valuation` was also included to reflect the company's financial expectations, while lending campaigns integrated variables like `Holding time` and `Annual ROI` to model investment duration and profitability.

Post-launch features reflect the campaign's unfolding dynamics once it goes live. These include daily and cumulative fundraising data (m), investor counts (i), and derived returns—r^m for monetary returns and r^i for returns on investor engagement. To capture behavioural shifts and funding momentum, the variable d was introduced to represent the daily change in capital raised. These variables were constructed from raw platform data through a series of preprocessing steps involving normalisation, smoothing, and outlier reduction, which ensured consistency across campaigns with differing time horizons and update frequencies.

Thanks to their sequential nature, post-launch variables were integrated into recurrent neural networks, particularly Long Short-Term Memory (LSTM) architectures, as described in Section: Input Data and Predictive Modeling in Crowdfunding Dynamics. Their inclusion enabled the models to identify non-linear trajectories and time-dependent patterns, offering a richer representation of campaign performance over time.

The combined use of pre- and post-launch features aligns with existing research highlighting the benefits of hybrid predictive approaches [24, 26]. In this study, the modelling approach was designed to progressively incorporate both static and temporal data, allowing early predictions to be refined as new temporal inputs become available. The integration of features occurs within a unified framework rather than strictly separated phases. In practice, this structure allowed the models to leverage the stability of pre-launch variables and the responsiveness of post-launch dynamics, improving predictive performance at every stage.

Finally, to move beyond binary classification, the analysis also incorporates advanced indicators such as the *Funding Ratio* and *Overfunding*, discussed in the following sections.

These continuous success metrics enable a more refined understanding of campaign performance and facilitate meaningful comparisons across different initiatives [60].

4.1.2 Dataset Description

The structure and characteristics of each dataset are detailed in Figure 4.1, under the section "Dataset Description" within the broader framework of *Data Preparation for Predictive Modelling.* This chapter presents a comprehensive overview of the crowdfunding campaigns analysed in Italy, with particular emphasis on the key features that influence success.

The data were independently collected during this research from a diverse set of Italian platforms. To enhance robustness and enable cross-national comparisons, selected reference features from European platforms were also incorporated.

As previously noted, the dataset includes information from approximately 40 Italian platforms active in 2022, accounting for about half of those operating nationally. This extensive collection offers a representative snapshot of the crowdfunding landscape, encompassing various variables relevant to campaign performance and success.

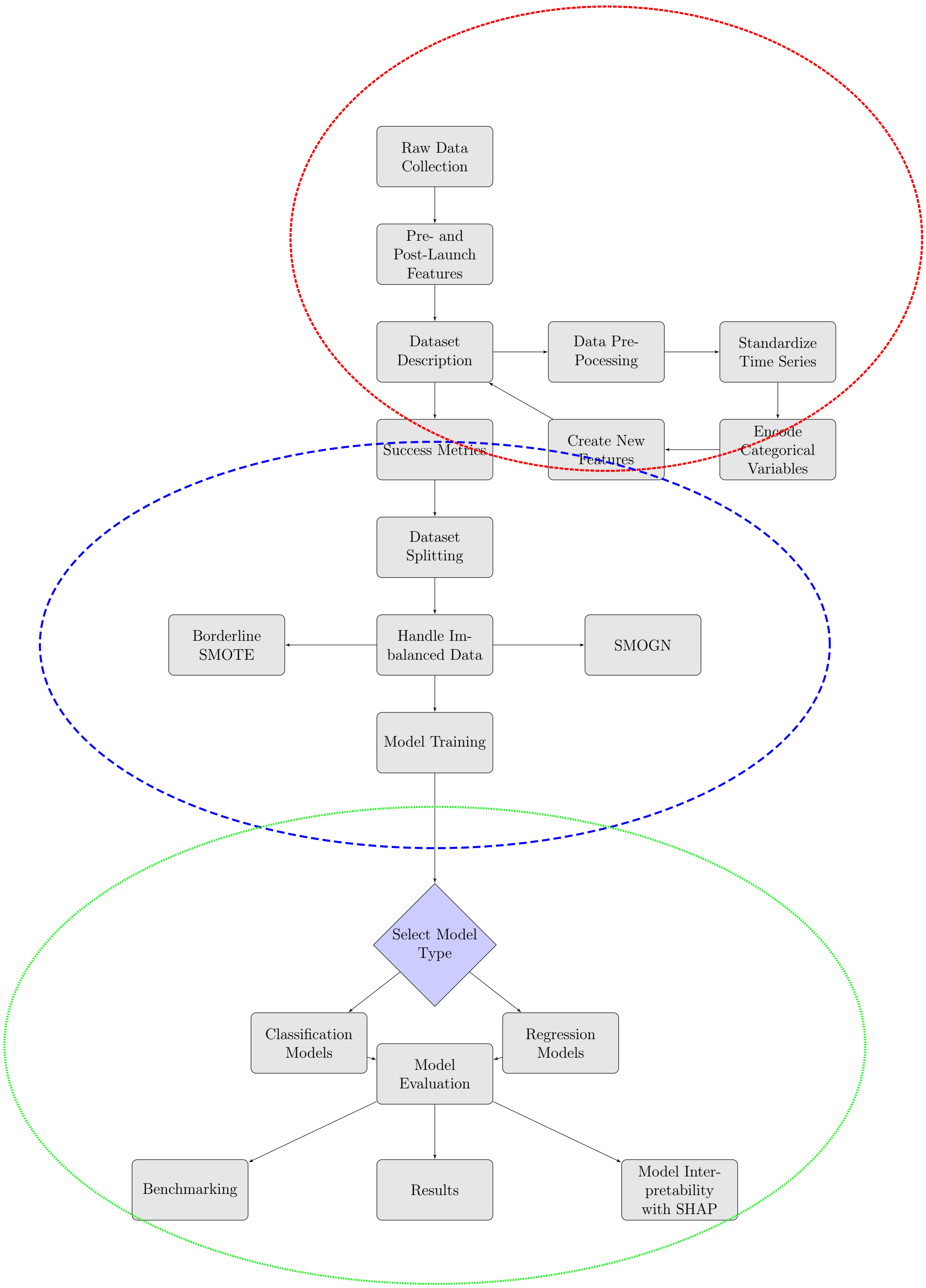

Figure 4.1: Framework of Crowdfunding Success Prediction Methodology

To structure the experimental workflow, three distinct analytical phases were defined and visually distinguished by colour Figure 4.1:

- **Red ellipse:** *Data Preparation for Predictive Modeling*, including raw data collection, feature engineering, data cleaning, and preprocessing operations such as time series standardisation and categorical encoding.

- **Blue ellipse:** *Input Data and Predictive Modeling*, covering dataset partitioning, class imbalance correction (e.g., Borderline SMOTE and SMOGN), and preparation for model training.

- **Green ellipse:** *Performance Evaluation and Model Assessment*, comprising model selection, metric evaluation, algorithm benchmarking, and interpretability techniques such as SHAP.

Together, these phases constitute a rigorous and systematic framework for developing and validating predictive models to forecast crowdfunding outcomes.

The dataset supports this experimental workflow and informs the subsequent analytical steps. Following an overview of its structure, the next sections explore the variables most relevant to campaign performance.

Campaigns are categorised by project type, distinguishing between equity and lending models. This classification allows for a clearer analysis of temporal trends and category-specific dynamics. Approximately 60% of campaigns are equity-based, while 40% fall under lending. This distribution indicates a strong market preference for equity crowdfunding in Italy, aligning with trends observed in other European contexts. Such a pattern highlights the importance of examining the operational mechanisms and enabling conditions of participatory finance, with a particular focus on equity and peer-to-peer lending initiatives.

Beyond general attributes, the dataset includes key features that deepen our understanding of crowdfunding processes:

- **Campaign Type:** Indicates whether a project is lending- or equity-based. In total, 1,084 projects fall under lending, and 907 under equity. See Fig. 4.2 for cumulative annual trends, in line with data from [70].

- **Campaign Name:** Serves as a unique identifier and promotional element, often influencing initial investor interest.

- **Project Region:** Specifies the geographical base. Northern regions, especially Lombardy and Veneto, account for over 50% of campaigns—possibly due to stronger economic ecosystems.

- **Fundraising Goals:** Defines minimum and maximum capital targets. Campaigns exceeding €500,000 tend to exhibit lower success rates, likely due to heightened perceived risk.

- **Campaign Duration:** Averages around 60 days, with most falling between 45 and 75 days. Duration plays a key role in structuring engagement strategies.
- **Time Series Data:** Captures daily investor counts and funds raised, offering insights into campaign progression.
- **Pre-Commitment:** Indicates the volume of funds and backers secured prior to launch—often a signal of early momentum.
- **Economic Categories:** Assigns campaigns to one or more economic sectors.
- **Pre-Money Valuation:** Reflects the enterprise's estimated value before the campaign; 30% exceed €1 million, shaping investor expectations.
- **Investor Engagement:** Time series of investor behaviour, typically peaking at launch and near closing dates.
- **Minimum Investment Required:** Ranges from €50 to €10,000. Lower thresholds attract a broader base; higher ones appeal to institutional investors.
- **Investment Duration:** Generally spans 12 to 36 months, influencing perceived risk and expected return.
- **Annual ROI:** Indicates average annual return—around 7% in this dataset. Real estate projects typically yield slightly higher returns than corporate campaigns.
- **Campaign Start and End Dates:** Mark the official fundraising window, crucial for tracking performance over time.

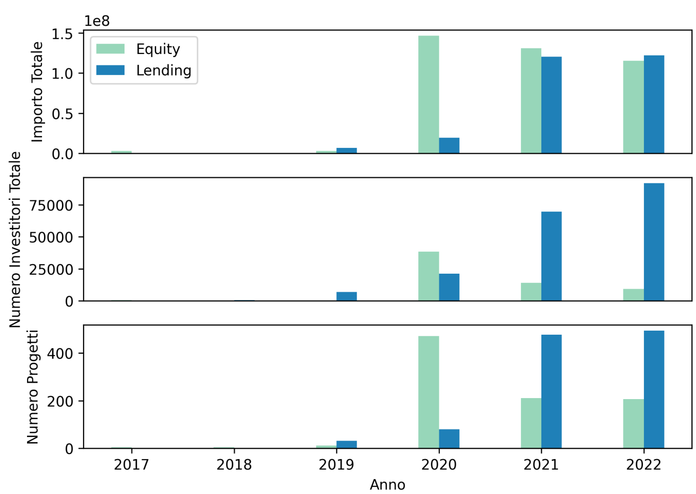

Figure 4.2: Cumulative amount of money, investors, and projects by year and type of crowdfunding in the analysed sample.

Several variables were excluded from the analysis due to limited availability or low predictive value:

- **Start/End of "Coming Soon" Phase:** Often missing or inaccessible post-campaign.
- **Cities and Provinces:** Incomplete and highly correlated with regional data already used.
- **Company Type:** Mostly identified as "innovative SME" or "startup," offering limited variability.
- **Campaign URL:** Used only for validation and excluded from modelling.
- **Country of Project Launch:** Omitted due to the negligible number of non-Italian campaigns (e.g., 2 equity campaigns from Estonia).

Although the exclusion criteria helped simplify the structure of the dataset, several technical challenges still arose. A significant portion of the time series contained missing values, primarily due to scraping errors or discontinuities in the original platform data. Moreover, the varying durations of the campaigns led to substantial heterogeneity in the lengths of the time series. These issues required a rigorous preprocessing phase, including interpolation, normalization, and resampling operations, to ensure both internal consistency within the dataset and comparability across observations.

The datasets analysed were obtained from Startwallet, a specialised information source that monitors the Italian crowdfunding market. This platform provided an extensive and detailed collection of data documenting the evolution of campaigns and platforms over time. The information gathered includes fundraising volumes, investment flows, project characteristics, and platform behaviours, offering a comprehensive overview of both equity and lending crowdfunding dynamics.

Despite its depth and granularity, the empirical foundation presented notable technical complexities. Incomplete records, inconsistent values, and heterogeneous temporal structures required a thorough phase of data cleaning, imputation, and transformation. The strategies outlined in the following section reflect a rigorous and technically demanding workflow that guarantees methodological soundness and analytical reliability. Finally, to improve model accuracy and reflect the structural differences between funding models, the dataset was partitioned into two analytical subsets: equity and lending. This distinction, consistent with the typology introduced earlier, enables separate evaluations of campaign dynamics and forecasting performance in the subsequent chapters.

4.1.3 Data Preprocessing

The datasets analysed in this study presented a significant number of missing values, especially within the two core time series: *money* ($\mathbf{m}$) and *investors* ($\mathbf{i}$). These variables are essential for understanding the evolution of crowdfunding campaigns, as they capture both the funds raised and the level of investor participation over time. Addressing these gaps was crucial to preserving the accuracy and reliability of the analysis. Standardisation and encoding procedures ensure that the time series inputs are model-ready (see Figure 4.1, modules "Standardize Time Series" and "Encode Categorical Variables"; Data Preparation for Predictive Modeling).

Most missing values stemmed from scraping issues—failures in the automated extraction of complete or up-to-date data from crowdfunding platforms. In other instances, data remained unchanged across multiple time points due to inactivity, making it unclear how to handle early values that were never recorded. Simply carrying forward the first available entry would have risked distorting a campaign's initial dynamics.

To formalize the structure, let (m_j, i_j) denote the observed values in the series $\mathbf{m} = [m_1, \ldots, m_n]$ and $\mathbf{i} = [i_1, \ldots, i_n]$, where each pair refers to a specific time point j. At time j, m_j corresponds to the cumulative amount raised, while i_j indicates the number of investors up to that point.

During the analysis, several recurring patterns of missing data emerged:

1. **Simultaneous absence in both series**: m_j and i_j missing at the same time point, typically caused by scraping failures or complete inactivity.

2. **Missing initial values**: Absence of one or both starting values (m_1, i_1), making it difficult to assess early campaign momentum.

3. **Complete absence of one series**: No valid entries in either $\mathbf{m}$ or $\mathbf{i}$ across all time points, severely limiting analysis.

4. **Irregular and combined missing patterns**: Partial, inconsistent gaps requiring flexible and context-aware imputation strategies.

To mitigate the impact of these issues, a multi-step preprocessing pipeline was developed.

In case (1), when both values were missing for the same date, the data pair (m_j, i_j) was removed to maintain alignment and avoid inconsistencies.

In case (2), where a time series had only one data point and initial values were missing, zeros were imputed for both series to define a clear baseline.

In case (3), campaigns missing the full *money* series were excluded, as reconstruction was not feasible. However, when only the *investors* series was missing, the campaign was retained and flagged, since financial data still offered useful insights into campaign progress.

For irregular gaps (case 4), a dual linear regression approach was adopted to estimate missing values in one series using the available values in the other, leveraging their correlation while preserving temporal consistency.

Once these gaps were addressed and the series completed, the next step was to standardise their length. This was essential for enabling consistent comparisons across campaigns and preparing the dataset for machine learning applications and temporal pattern analysis. To ensure the compatibility of time series with machine learning algorithms, their length was standardised to two specific values: 10 and 14 days. The 10-day window was selected as it reflects the median duration of the campaigns in the dataset, while the 14-day window was introduced to represent a biweekly interval, which is often relevant in behavioural time-based analyses. These two lengths will be referred to as L10 and L14 throughout the remainder of this thesis. This decision strikes a balance between accurately capturing campaign dynamics and maintaining computational stability and efficiency during model training. Shorter time series were interpolated to reach the desired length, whereas longer series were uniformly sampled while preserving both the first and the last data points. The distinction between L10 and L14 will also be maintained in the benchmarking phase, allowing for a consistent comparative analysis of model performance.

4.1.4 Standardize Time Series

To overcome the limitations of the original cumulative time series, a differential version, hereafter referred to as d, was introduced and used throughout the analysis. This transformation was essential to obtain a clearer and more informative representation of campaign dynamics. Indeed, cumulative series, such as the amount raised or the number of investors, often obscure short-term fluctuations and are influenced by irregular update frequencies among campaigns.

By converting cumulative data into daily variations, the d series allowed for a more standardized comparison between campaigns. In particular, d ensures temporal uniformity, as the variation is distributed over regular daily intervals, eliminating distortions caused by inconsistent data recording. This enables the identification of critical events, such as participation spikes or consistent contributions, which might otherwise be masked in a cumulative visualization.

To ensure comparability across campaigns, all time series were standardized to a fixed length of $K = 5$ days. This value was selected as a shared benchmark for both equity and lending crowdfunding, as it represents a balanced trade-off between temporal detail and data coverage. Standardizing the time window ensured a consistent input format for the predictive models, avoided biases due to heterogeneous campaign durations, and allowed each campaign to contribute an equal amount of temporal information to the analysis.

For example, suppose a campaign raises €100 on the first day, €150 on the second day, and €200 on the fourth day. The d series calculates a variation of $+50$€ between the first and second day, while the variation between the second and fourth day, distributed over two days, results in an average of $+25$€ per day. This process provides a more realistic view of the dynamics at play, especially when aggregated across a large number of campaigns.

After standardizing the length of the time series for both money and investors, two additional variables were calculated: the daily return series R_M for money and R_{io} for investors. These were derived using the following formulas:

$$R_M^J = \frac{M_J - M_{J-1}}{M_{J-1}} \times 100, \quad R_{io}^J = \frac{io_J - io_{J-1}}{io_{J-1}} \times 100$$

where J indexes the days in the differential series d, and by definition $R_M^1 = 0$. The resulting return series capture the relative changes between consecutive time points, offering an additional lens to interpret campaign performance.

These models are particularly effective for analyzing temporal trends, as they capture percentage variations rather than absolute values, thus revealing phases of acceleration or deceleration in campaign performance. The length of the return series corresponds to that of the original time series minus one, since each return R^J is calculated as the relative change between two consecutive observations.

To illustrate this, consider a crowdfunding campaign with daily contributions X_J recorded over a period of 10 days (in monetary units):

$$X = [100, 120, 140, 200, 220, 240, 300, 320, 340, 400]$$

The following table presents the return series R_J calculated:

Day j	Contribution x_j	Return r_j (%)
2	120	20.00
3	140	16.67
4	200	42.86
5	220	10.00
6	240	9.09
7	300	25.00
8	320	6.67
9	340	6.25
10	400	17.65

Table 4.1: *Calculation of daily returns r_j for a 10-day crowdfunding campaign*

As shown in Table 4.1, the return series offers valuable insights into the momentum of the campaign. Notable trends include:

- Day 4: A significant increase of 42.86- Days 7–9: Returns stabilize below 10- Day 10: A late surge of 17.65

The return-based approach enables a more precise identification of key phases in campaign development. While cumulative or absolute values may highlight major milestones, the return series accentuates relative variations, offering a dynamic lens through which to assess performance. For instance, the peak variation on day 4 signals a critical inflection point, followed by a period of slowdown and then renewed momentum toward the campaign's conclusion.

Overall, the d and R series serve as key methodological enhancements, enabling a more detailed understanding of daily trends. They help highlight the effectiveness of specific strategic actions—such as marketing boosts or media coverage—and provide valuable insights for optimizing future strategies, enhancing overall management, and improving the outcomes of subsequent campaigns.

Finally, to further refine the analysis and capture the diverse characteristics of the crowdfunding campaigns, one-hot encoding was applied to categorical variables such as *campaign categories*, *platform types*, and *geographical regions*.

One-hot encoding transforms each category into a binary vector, where a value of 1 indicates the presence of the category and 0 represents its absence. This encoding scheme enables machine learning algorithms to process and learn from categorical data effectively, allowing us to incorporate relevant contextual dimensions into the predictive framework. By explicitly modeling these categorical features, the analysis can capture potential differences or similarities in campaign behavior and outcomes across sectors, platforms, and regional settings. For example, platform-specific dynamics or regional investor behavior may significantly impact campaign performance and are therefore valuable to consider in a structured way.

Additionally, multi-hot encoding was applied to multi-class categorical variables, such as the economic-productive categories (*Categorie EP*), to allow the models to account for the multifaceted nature of certain campaigns, which may span across multiple sectors simultaneously.

The transformation and standardization of both numerical and categorical variables ensured that all features were in a suitable format for subsequent modeling, facilitating the learning process and improving interpretability of the results.

The following provides a comprehensive description of the features included in the dataset, Table 4.2:

4.1.5 Encode Categorical Variables

A critical component of data preprocessing for machine learning applications is the encoding of categorical variables. These variables—such as the crowdfunding platform, geographical region, and economic sector—are typically represented as non-numeric labels, which must be transformed into numerical formats to be processed by predictive models.

One of the most widely adopted techniques for this transformation is *One-hot encoding*, which creates a binary feature for each category within a given variable. In this representation, a "1" denotes the presence of a specific category, while a "0" indicates its absence. For example, if the variable *Platform* includes categories like "Kickstarter," "Indiegogo," and "GoFundMe," One-hot encoding generates three binary columns, each corresponding to one platform.

Table 4.2: *Summary table of variable definitions used in the subsequent analysis*

Variable Name	Variable Description	Type
Tipologia	Type (equity/lending) of the crowdfunding campaign	String
Min_amount	Minimum target of the campaign	Float
Max_amount_Imputed	Maximum target of the campaign, imputed from Max_amount	Float
Minimum_investment	Minimum subscription that can be invested in the project (if absent, imputed with the median)	Float
lenName	Number of characters in the campaign name	Int
NumWordName	Number of words in the campaign name	Int
UpperCaseinName	Number of uppercase characters in the campaign name	Int
NuminName	Number of numeric characters in the campaign name	Int
Duration	Duration of the campaign	Int
year_Startdate	Start year of the campaign	Int
year_endDate	End year of the campaign	Int
month_Startdate	Start month of the campaign	12th roots
month_endDate	End month of the campaign	12th roots
quarter_Startdate	Start quarter of the campaign	4th roots
quarter_endDate	End quarter of the campaign	4th roots
dayofweek_Startdate	Start day of the week of the campaign	7th roots
dayofweek_endDate	End day of the week of the campaign	7th roots
Piattaforma	Name of the crowdfunding platform/portal/website	One Hot Encoding
Regione	Italian region where the project/company is based	One Hot Encoding
Categorie EP	Economic-productive categories	Multi Hot Encoding
Settore	Sector (Real Estate/Company)	Boolean
con_srl	Label indicating if the name contains the word "srl"/"S.R.L."	Boolean
con_milano	Label indicating if the name contains the word "Milano"	Boolean
con_via	Label indicating if the name contains the word "via" [lending only]	Boolean
con_a	Label indicating if the name contains the word "a" [lending only]	Boolean
con_appartamento	Label indicating if the name contains the word "appartamento" [lending only]	Boolean
Holding time	Duration of the investment [lending only]	Float
Annual ROI	Annual ROI [lending only]	Float
Pre-money valuation	Pre-money valuation of the company [equity only]	Float

This method preserves the informational content of categorical variables while avoiding the introduction of arbitrary ordinal relationships. It integrates well with a wide range of models, including neural networks, and contributes to predictive accuracy by providing a structured representation of nominal data.

In the context of crowdfunding, One-hot encoding is particularly useful for variables such as platform, geographical location, and project category. Once encoded, these features can be used directly in model training, enhancing the interpretability and predictive power of the model.

However, when categorical variables exhibit high cardinality, as is often the case with crowdfunding platforms or sector classifications, One-hot encoding may result in an excessively large number of features. This can increase data sparsity, reduce model efficiency, and raise the risk of overfitting.

To address these issues, *embedding techniques* offer a more efficient alternative. Embedding maps each category into a low-dimensional, dense numerical vector, capturing latent relationships between categories. This method significantly reduces feature dimensionality while preserving semantic structure, making it particularly effective for deep learning models.

In crowdfunding datasets, where variables such as platform or region may include dozens of categories, embeddings help identify meaningful patterns and similarities. Unlike One-hot encoding, embeddings enable the model to generalize better across categories and reduce the computational complexity associated with high-dimensional input spaces.

Moreover, embeddings enhance model scalability by decreasing the number of trainable parameters—an advantage in scenarios with high variability, such as those observed in this analysis.

The following sections of this thesis explore these techniques in greater detail, outlining their implementation and evaluating their impact on model performance in the context of crowdfunding campaign prediction. Specific examples demonstrate how encoding strategies affect both the accuracy and computational efficiency of the predictive framework.

This encoding step integrates seamlessly with previous transformations—such as the time series standardisation to $K = 5$ and the use of first-order differentials—ensuring a reliable and harmonised dataset, particularly when time series segments are incomplete or missing early values for monetary and investor variables. Additionally, transforming raw observations into differential series has proven effective in addressing data irregularities and temporal misalignment. The *diff* format not only harmonises disparate data intervals but also enhances the capacity of models, especially LSTM architectures, to capture fine-grained campaign dynamics and long-term dependencies in financial time series.

The analysis also introduces refined success indicators, such as the *Funding Ratio* and *Overfunding*, which provide a more nuanced evaluation of performance beyond binary outcomes. These indicators incorporate thresholds and statistical benchmarks to measure outcomes relative to campaign goals and investors.

4.1.6 Create new feature

The feature engineering phase plays a pivotal role in the data preprocessing pipeline, substantially enhancing the accuracy and robustness of the predictive models developed for crowdfunding campaigns. Building upon the classification of pre-launch and post-launch variables detailed in Section 4.1.1, this section focuses on the transformation, expansion, and enrichment of those features to extract deeper behavioural and structural signals from the raw data.

In particular, quantitative indicators such as the Funding Ratio and Overfunding Index were derived to complement traditional binary success metrics. These continuous measures support the implementation of regression models and recurrent neural networks (RNNs), enabling the identification of latent temporal dynamics and campaign momentum (see Figure 4.1, module "Create New Features", within the broader workflow in Figure 3.1).

Beyond numerical transformations, linguistic elements extracted from campaign titles were also incorporated. These textual features serve not only as identifiers but also as strategic communication tools capable of influencing investor perception and engagement. Prior research (e.g., Yang So et al., 2019; Ethan Mo, 2014; Xiaoying Re et al., 2020) has demonstrated the predictive relevance of language-related attributes in crowdfunding.

A dedicated linguistic analysis was conducted to identify variables likely to shape investor response. For instance, title length was examined both in terms of character and word count to detect patterns of clarity or verbosity. Capitalisation usage was also considered, distinguishing between neutral and emphatic textual styles. Likewise, the presence of numbers in titles was analysed to capture project-specific signals related to timing, financial targets, or quantifiable objectives. These elements were encoded through standard techniques such as one-hot encoding and vectorisation, with potential for future expansion via embeddings or tokenisation.

These insights were integrated into machine learning workflows through classifiers including decision trees, support vector machines (SVM), and ensemble models like random forests.

As described in Section 4.1.5, categorical variables—such as platform, geographical region, and sector—were encoded to ensure interpretability and compatibility with standard model architectures. Additional boolean indicators were constructed to distinguish between equity and lending campaigns based on keywords in the campaign titles, contributing to the model's ability to discern tone and perceived campaign typology.

By combining these engineered features with the structural and temporal inputs previously introduced, the model attains a more comprehensive and nuanced view of crowdfunding dynamics. This integrated feature set, especially when paired with advanced neural network architectures such as LSTM or Transformer models, enhances the system's ability to uncover non-linear trajectories and support more informed, data-driven strategic decisions.

The next section builds upon these transformations to introduce an extended definition of campaign Success, moving beyond binary classifications and enabling a richer, more realistic predictive framework.

4.2 Output data and predictive modeling

To ensure robust and multidimensional model comparison, this study adopts multiple evaluation metrics across both classification and regression tasks. This methodological choice enhances the interpretability and reliability of model benchmarking

4.2.1 Handle imbalance data: SMOTE AND SMOGN

Class imbalance represents a relevant methodological concern in predictive modelling applied to crowdfunding. The observed data distribution tends to be concentrated around recurring patterns, such as campaigns that reach funding levels close to the average. At the same time, cases characterised by extreme targets or atypical behaviours are under-represented. This imbalance may adversely affect the performance of both classification and regression models, limiting their ability to learn from infrequent cases effectively. The issue has been discussed in the literature, where unbalanced datasets are shown to lead to biased models that systematically favour majority classes [71, 72].

In this study, the class imbalance problem is addressed within the macro-phase Input Data and Predictive Modeling, and more specifically in the Data Imbalance Handling module of the proposed analytical framework (see Figure 4.1, module "Data Imbalance Handling"; Input Data and Predictive Modeling phase). To mitigate this issue, advanced resampling techniques were adopted and tailored to the specific nature of the predictive task to enhance the model's ability to learn from both common and rare patterns.

For classification tasks, the Synthetic Minority Over-sampling Technique (SMOTE) was employed [73]. SMOTE generates synthetic instances for the minority class by interpolating between neighbouring real observations. This strategy prevents the overfitting typically associated with naive oversampling and contributes to a more balanced training set, improving the model's classification performance. In this context, SMOTE was applied to balance observations related to campaigns classified as "successful" or "unsuccessful", based on the 98% threshold concerning the minimum funding target.

For regression tasks, the SMOGN technique (SMOTE for Regression with Gaussian Noise) was applied [74]. SMOGN extends the SMOTE methodology to continuous target variables by adding controlled Gaussian noise to the generated synthetic samples. This approach increases the density of instances in regions corresponding to rare or extreme values—such as very high or near-zero overfunding—and improves the model's ability to generalise across sparsely populated segments of the target space.

Integration of SMOTE and SMOGN into the modelling pipeline contributed to the development of more robust, fair, and generalisable models, capable of adequately capturing the heterogeneity of real crowdfunding dynamics. These techniques enabled the framework to overcome a structural limitation that remains insufficiently addressed in the current literature, ultimately enhancing predictive reliability, particularly in contexts affected by highly imbalanced data distributions.

4.2.2 Model training

The methodological approach in this study follows a structured and comprehensive process encompassing all critical phases of predictive model development and evaluation: data collection, preprocessing, feature engineering, model training, and performance assessment. Each step ensures the accuracy and robustness of models, addressing key challenges in crowdfunding data analysis, such as handling imbalanced datasets and missing values. A fundamental aspect of this methodology is the iterative training process, which continuously refines predictions and enhances model performance through supervised and unsupervised learning techniques.

Managing imbalanced data and missing values is particularly crucial in the analysis phase. Class imbalance, a common issue in machine learning, can skew model performance by favouring the majority class, reducing the ability to predict underrepresented cases correctly. This phenomenon, known as class imbalance bias, leads to lower predictive accuracy and a higher rate of false negatives or false positives.

To mitigate this, techniques such as SMOTE (Synthetic Minority Over-sampling Technique) and undersampling are applied, ensuring a balanced dataset and improving classification metrics such as precision, recall, and F1-score. Models were trained on balanced datasets using SMOTE and SMOGN (see Figure 4.1, modules "Handle Imbalanced Data" and "Model Training"; Input Data and Predictive Modeling). Similarly, missing values are managed using advanced imputation techniques, including linear regression and K-Nearest Neighbors (k-NN), to maintain dataset integrity without compromising statistical validity.

The iterative training process integrates new data at each cycle, allowing the model to adapt dynamically to evolving trends in crowdfunding campaigns. Initially, the model is trained on historical data; as new data becomes available, it is incorporated into the learning process, enhancing predictive accuracy and enabling the model to capture temporal patterns more effectively. To achieve this, recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, are employed to model long-term dependencies in temporal data.

This approach ensures that predictions remain adaptive to the shifting context of a campaign, improving the model's ability to anticipate outcomes with greater accuracy. Given that investor behaviour and funding dynamics fluctuate throughout a campaign, continuously integrating new information optimises strategic decision-making and facilitates targeted interventions. This progressive refinement process strengthens the model over time, making it increasingly precise and capable of supporting real-time decision-making.

Additionally, continuous monitoring of model performance allows for regular evaluation and adjustments. Each training iteration refines predictive capabilities by incorporating insights from active campaigns, making the model more responsive to shifts in investor behaviour and enabling real-time strategic adjustments. The use of deep learning and advanced machine learning techniques ensures the detection of complex temporal relationships, enhancing prediction accuracy as data evolves.

The models demonstrating the best performance in the test set include:

- **Success Prediction:**
  - **Equity:** AdaBoost and Histogram Gradient Boosting performed best for L10 (Table 4.4) and L14 (Table 4.10).
  - **Lending:** Random Forest and Histogram Gradient Boosting outperformed other models for L10 (Table 4.7) and L14 (Table 4.13).
- **Funding Ratio Prediction:**
  - **Equity:** Histogram Gradient Boosting yielded the best results for both L10 (Table 4.5) and L14 (Table 4.11).
  - **Lending:** LSTM exhibited superior performance for L10 (Table 4.8) and L14 (Table 4.14).
- **Overfunding Prediction:**
  - **Equity:** XGBoost and LSTM were the top performers for L10 (Table 4.6) and L14 (Table 4.12), respectively.
  - **Lending:** Bagging outperformed other models for both L10 (Table 4.9) and L14 (Table 4.15).

4.2.3 Hyperparameter Optimization

A dedicated hyperparameter optimisation phase was conducted to ensure that each predictive model reached its highest potential in terms of performance. The process involved systematically exploring the most relevant configurations for each model using well-established tuning techniques, namely `GridSearchCV`, `RandomizedSearchCV`, and manual tuning procedures inspired by `KerasTuner` for deep learning architectures.

The choice of optimisation strategy depended on the model's complexity and the dimensionality of the parameter space. For instance, `RandomizedSearchCV` was preferred for tree-based models such as XGBoost and Random Forest, to reduce computational burden while preserving exploration efficiency. For the LSTM model, due to its neural nature and longer training time, a more targeted manual tuning was adopted based on iterative validation.

Each configuration was evaluated using appropriate metrics—*F1 score* for classification, R^2 *score* for regression, and *validation loss* for sequence models—with cross-validation where applicable.

Table 4.3 summarizes the hyperparameter tuning strategies adopted for the main predictive models used in this study, covering classification, regression, and sequence modeling tasks. The table reports the optimization technique, the selected evaluation metric, the best hyperparameter configuration, and the corresponding performance score.

Table 4.3: *Summary of Hyperparameter Optimization for Key Models*

Model	Objective	Tuning	Metric	Best Params	Score
XGBoost	Success (Classification)	Randomized	F1 Score	n_estimators=300 max_depth=6 learning_rate=0.1 subsample=0.8	0.84 (CV)
Random Forest	Funding Ratio (Regression)	Randomized	R^2 Score	n_estimators=200 max_depth=20 min_samples_split=5	0.90 (Test)
LSTM	Temporal Modeling	Manual	Val Loss	units=128 dropout=0.3 learning_rate=0.0005	0.31 (Val)

While all models underwent systematic tuning to identify optimal hyperparameter configurations, the tuning process of the LSTM is particularly illustrative due to its sensitivity to architectural choices. Figure 4.3 shows the relationship between the number of LSTM units and the validation loss. The best configuration was achieved using 128 units, a dropout rate of 0.3, and a learning rate of 0.0005.

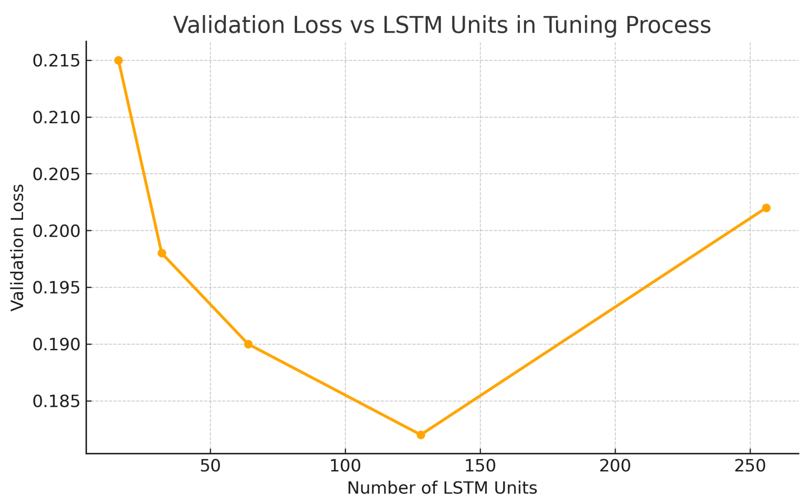

Figure 4.3: Validation Loss vs LSTM Units in Tuning Process

To conclude, the hyperparameter optimisation phase ensured that each predictive model operated under optimal conditions, thereby supporting a rigorous and equitable evaluation framework. The following section provides a comparative benchmarking of these models across multiple predictive tasks and datasets. This hyperparameter optimisation step is a fundamental component of the model training process (see Figure

4.1, module "Model Training"; part of the "Input Data and Predictive Modelling" phase in Figure 3.1).

4.2.4 Results

These selections highlight the most effective models across different prediction tasks, reinforcing the comprehensive approach taken in this study to identify robust solutions.

Table 4.4: *Performance of AdaBoost on equity campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.94	0.945	0.934	0.956	0.951	0.973	0.973	0.967	0.984	0.978
Precision	0.973	0.966	0.966	0.98	0.973	0.986	0.98	0.98	0.987	0.98
Recall	0.953	0.966	0.953	0.966	0.966	0.98	0.987	0.98	0.993	0.993
F1	0.963	0.966	0.959	0.973	0.97	0.983	0.983	0.98	0.99	0.987

Table 4.5: *Performance of Histogram Gradient Boosting on equity campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.778	0.854	0.89	0.906	0.91	0.954	0.959	0.954	0.952	0.955
R^2_{adj}	0.645	0.759	0.812	0.833	0.834	0.912	0.918	0.904	0.895	0.897
MSE	0.038	0.025	0.019	0.016	0.016	0.008	0.007	0.008	0.008	0.008
MAE	0.122	0.092	0.08	0.072	0.069	0.058	0.054	0.055	0.054	0.052

The models' performance systematically improved as more time-series data became available, validating the expected positive correlation between dataset size and predictive accuracy. Additional data points enabled models to better identify patterns, leading to more refined predictions.

Notably, this improvement was evident even in early-stage predictions. Despite limited initial datasets, the most effective models consistently delivered reliable forecasts of campaign success. This capability is crucial for campaign managers, allowing them to make informed decisions early in the campaign lifecycle. The models' consistency across both classification and regression tasks highlights their robustness and adaptability for various predictive applications.

Table 4.6: *Performance of XGBoost on equity campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.563	0.644	0.68	0.734	0.766	0.8	0.819	0.837	0.862	0.889
R^2_{adj}	0.302	0.412	0.452	0.528	0.568	0.616	0.638	0.66	0.699	0.746
MSE	0.067	0.055	0.049	0.041	0.036	0.031	0.028	0.025	0.021	0.017
MAE	0.179	0.152	0.134	0.118	0.109	0.101	0.096	0.091	0.078	0.064

Table 4.7: *Performance of Random Forest on lending campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.968	0.972	0.977	0.977	0.977	0.995	0.982	0.991	0.977	0.986
Precision	0.977	0.977	0.981	0.981	0.981	0.995	0.986	0.991	0.986	0.991
Recall	0.99	0.995	0.995	0.995	0.995	1.0	0.995	1.0	0.99	0.995
F1	0.983	0.986	0.988	0.988	0.988	0.998	0.991	0.995	0.988	0.993

Table 4.8: *Performance of LSTM on lending campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9,	t_{10}
R^2	0.942	0.945	0.951	0.947	0.952	0.948	0.944	0.955	0.96	0.95
R^2_{adj}	0.917	0.919	0.926	0.918	0.924	0.916	0.907	0.923	0.93	0.91
MSE	0.008	0.008	0.007	0.007	0.007	0.007	0.008	0.006	0.006	0.007
MAE	0.033	0.036	0.037	0.039	0.035	0.04	0.04	0.037	0.036	0.037

Our analysis revealed that predictive models for lending consistently outperformed those for equity, achieving classification scores above 0.9. This difference suggests that lending campaigns exhibit more predictable patterns or that the features used are more indicative of success in lending contexts. The strong performance of lending models is probably due to structured financial metrics and standardised lending agreements, which may be easier to model in comparison to more variable equity campaigns.

Interestingly, the time-series length (10 vs. 14 days) did not significantly impact

Table 4.9: *Performance of Bagging on lending campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.889	0.89	0.89	0.884	0.886	0.89	0.887	0.89	0.893	0.895
R^2_{adj}	0.841	0.839	0.835	0.821	0.82	0.822	0.812	0.813	0.813	0.811
MSE	0.024	0.024	0.024	0.025	0.025	0.024	0.025	0.024	0.024	0.023
MAE	0.062	0.063	0.062	0.066	0.065	0.064	0.063	0.064	0.061	0.061

Table 4.10: *Performance of Histogram Gradient Boosting on equity campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.918	0.934	0.951	0.945	0.951	0.934	0.945	0.956	0.962	0.967	0.973	0.978	0.978	0.984
Precision	0.953	0.948	0.961	0.966	0.973	0.966	0.966	0.973	0.973	0.98	0.98	0.987	0.98	0.987
Recall	0.946	0.973	0.98	0.966	0.966	0.953	0.966	0.973	0.98	0.98	0.987	0.987	0.993	0.993
F1	0.949	0.96	0.97	0.966	0.97	0.959	0.966	0.973	0.977	0.98	0.983	0.987	0.987	0.99

Table 4.11: *Performance of Histogram Gradient Boosting on equity campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.791	0.86	0.866	0.894	0.905	0.911	0.912	0.934	0.942	0.938	0.939	0.941	0.949	0.964
R^2_{adj}	0.666	0.769	0.771	0.812	0.825	0.829	0.824	0.862	0.873	0.858	0.853	0.851	0.864	0.898
MSE	0.035	0.023	0.022	0.017	0.016	0.015	0.014	0.011	0.01	0.01	0.01	0.01	0.008	0.006
MAE	0.117	0.09	0.082	0.069	0.066	0.06	0.057	0.054	0.05	0.052	0.049	0.049	0.047	0.043

regression models' performance, indicating that even shorter time frames provide sufficient information for accurate predictions. This is particularly beneficial for real-time scenarios where data availability is limited. However, deep learning models, particularly LSTM, excelled in predicting the Funding Ratio for lending campaigns, demonstrating their ability to capture sequential dependencies. In other cases, dataset size constraints may have limited deep learning models' effectiveness, resulting in suboptimal outcomes. Nonetheless, LSTM models showed strong capabilities in forecasting Overfunding within equity campaigns, highlighting their potential to detect complex

Table 4.12: *Performance of LSTM on equity campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.38	0.484	0.481	0.579	0.539	0.548	0.577	0.634	0.664	0.672	0.705	0.692	0.706	0.74
R^2_{adj}	0.01	0.147	0.111	0.252	0.149	0.132	0.154	0.236	0.266	0.249	0.291	0.22	0.214	0.263
MSE	0.095	0.079	0.08	0.065	0.071	0.07	0.065	0.056	0.052	0.05	0.045	0.047	0.045	0.04
MAE	0.191	0.173	0.168	0.151	0.156	0.156	0.154	0.14	0.134	0.129	0.124	0.132	0.118	0.111

Table 4.13: *Performance of Histogram Gradient Boosting on lending campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.972	0.982	0.986	0.982	0.986	0.986	0.991	0.986	0.991	0.982	0.986	0.986	0.991	0.995
Precision	0.977	0.981	0.986	0.981	0.986	0.986	0.991	0.986	0.991	0.99	0.991	0.986	0.991	0.995
Recall	0.995	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.99	0.995	1.0	1.0	1.0
F1	0.986	0.991	0.993	0.991	0.993	0.993	0.995	0.993	0.995	0.99	0.993	0.993	0.995	0.998

Table 4.14: *Performance of LSTM on lending campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.937	0.959	0.956	0.957	0.955	0.956	0.957	0.955	0.958	0.959	0.958	0.96	0.96	0.961
R^2_{adj}	0.91	0.94	0.934	0.934	0.929	0.929	0.929	0.923	0.926	0.926	0.922	0.924	0.921	0.921
MSE	0.009	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.005	0.005	0.005
MAE	0.035	0.033	0.033	0.035	0.034	0.032	0.035	0.034	0.035	0.036	0.036	0.034	0.032	0.031

patterns that traditional models may overlook.

Ensemble methods, particularly Gradient Boosting variations (eXtreme and Histogram) and Bagging, consistently achieved high performance across classification and regression tasks. Their adaptability and robustness make them effective for early success prediction, as combining multiple weak learners enhances data representation, leading to more reliable forecasts.

Key features significantly influencing crowdfunding success were identified through feature importance analysis within Random Forest models. The total amount of money raised emerged as one of the most critical predictors, except in lending overfunding sce-

Table 4.15: *Performance of Bagging on lending campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.891	0.892	0.887	0.885	0.871	0.881	0.883	0.878	0.882	0.878	0.88	0.884	0.886	0.888
R^2_{adj}	0.844	0.842	0.83	0.823	0.797	0.808	0.806	0.792	0.793	0.78	0.778	0.778	0.776	0.772
MSE	0.024	0.024	0.025	0.025	0.028	0.026	0.026	0.027	0.026	0.027	0.026	0.026	0.025	0.025
MAE	0.061	0.061	0.063	0.065	0.07	0.066	0.066	0.069	0.067	0.068	0.066	0.066	0.065	0.064

Table 4.16: *TN, FP, FN and TP values for confusion matrices on a set of 10 success lending observations L10 for random forest*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
TN	2	2	3	3	3	6	4	5	4	5
FP	5	5	4	4	4	1	3	2	3	2
FN	2	1	1	1	1	0	1	0	2	1
TP	208	209	209	209	209	210	209	210	208	209

narios, where other dynamics prevail. Early financial momentum plays a fundamental role, as substantial initial contributions can signal credibility and attract further investors. In lending campaigns, a lower minimum investment threshold was particularly impactful, as it reduced entry barriers for a broader investor base. Conversely, in equity campaigns, a lower minimum funding goal had a greater influence, presumably due to the higher risk-reward dynamics that create urgency and make targets more attainable.

An important insight was that shorter loan durations were a stronger predictor of success than higher interest rates, suggesting that lenders perceive shorter commitments as less risky. Additionally, in lending campaigns, overfunding appeared to be driven more by platform-specific features than general loan terms, whereas equity campaigns followed different influencing factors. Furthermore, the inclusion of geographical descriptors in project titles significantly influenced success in lending, potentially due to localized trust dynamics. In contrast, pre-money valuation had minimal impact outside of overfunding scenarios, indicating that other factors play a more substantial role in initial investment decisions.

To validate predictions in lending crowdfunding and assess model robustness, confusion matrices were analysed across 10 consecutive observations. The results consistently showed high True Positive (TP) values, ranging from 208 to 210, confirming the Random Forest model's accuracy in predicting successful lending campaigns. This reinforces its utility as a key predictive tool.

Equally important, False Positive (FP) and False Negative (FN) values remained low, indicating minimal misclassification errors. This high precision is particularly critical in lending crowdfunding, where inaccurate predictions can lead to financial risks for investors, potentially undermining confidence in the platform. The model's ability to minimise errors enhances its reliability in making strategic investment recommendations.

The stability of True Negative (TN) and FP values across observations further underscores the model's robustness. The relatively small variations suggest a consistent ability to classify unsuccessful campaigns accurately. This consistency is essential for understanding crowdfunding dynamics, enabling more precise long-term predictions.

Integrating time-series analysis further improved model adaptability, allowing it to adjust dynamically to shifting data patterns. By linking time-based predictions with performance metrics, the model effectively identified evolving trends in campaign success and investor behaviour. The ability to capture fluctuations in investment dynamics significantly enhanced forecasting accuracy and decision-making.

Despite achieving high True Positive rates, the Random Forest model exhibited misclassification errors in borderline cases, particularly campaigns with marginal financial signals near the success-failure threshold. These errors probably result from an overreliance on aggregated features, such as cumulative amounts and daily averages, which obscure transient dynamics. Incorporating volatility indicators and risk-specific measures could help detect fine-grained daily variations and critical inflexion points.

Moreover, the absence of contextual features—such as sentiment-based variables from campaign titles and descriptions—limits the model's ability to interpret investment flows accurately, particularly in ambiguous cases. Including qualitative indicators related to marketing strategy and promotional events could reduce misclassification in borderline scenarios and improve predictive accuracy.

Detailed error analysis revealed that 25.6% of campaigns exhibited significant volatility in fundraising and investor numbers, complicating trend detection. For example, Comehome! on Crowdfundme saw a 205% increase in funds raised between day 7 and day 10, while Hinelsen on Mamacrowd experienced a 140% surge in the final three days. Such fluctuations make predictions particularly challenging.

Additionally, eight campaigns were misclassified due to unexpected funding accelerations. Aulab on Crowdfundme, for instance, was predicted to fail but ultimately reached its target due to a last-minute funding surge. Similarly, Blue Taste on Mamacrowd saw a 185% increase in the final two days, leading to incorrect predictions.

These observations highlight the need for enhanced modeling techniques to account for rapid funding shifts and nonlinear investment patterns. Improving predictions may require incorporating temporal variables and economic strategy data. A future study could explore the use of moving averages to smooth extreme variations and introduce fundraising acceleration indicators (e.g., funds raised in the first three days) to identify strong initial momentum campaigns.

Segmenting predictions by platform may also be beneficial, as Mamacrowd and Crowdfundme exhibit distinct investment behaviours. The analysis shows that high-volatility campaigns are the most difficult to predict, with errors concentrated in campaigns experiencing rapid late-stage growth or strong pre-commitments. Introducing additional features and refining temporal modeling could improve prediction accuracy.

In conclusion, integrating machine learning with time-series analysis significantly enhances the prediction of crowdfunding campaign outcomes. This dual approach not only improves predictive accuracy but also provides deeper insight into the temporal dynamics of campaign performance. The analysis of key misclassification patterns—such as irregular funding trends, abrupt investor inflows, and platform-specific behaviours—highlights the importance of adaptive modelling strategies to capture the evolving complexity of crowdfunding ecosystems.

These results directly address the research goals outlined in Section 1.2, particularly concerning predictive reliability, model interpretability, and applicability to both equity and lending contexts. Among the evaluated models, Random Forest demonstrated the best performance in terms of MAE, while LSTM models—despite their temporal sophistication—showed comparatively lower predictive accuracy. These findings are further examined in the benchmarking section that follows, which offers a structured comparison and practical implications for model selection and deployment.

4.2.5 Benchmarking of Predictive Models in Equity Crowdfunding

ARIMA, Logistic Regression, Random Forest, XGBoost, and LSTM were tested to evaluate their predictive capabilities on crowdfunding data. The results highlighted the limitations of traditional time series models—such as ARIMA—in capturing the nonlinear dynamics and complex interactions typical of campaign behaviour. In contrast, machine learning approaches, particularly ensemble methods like Random Forest and XGBoost, demonstrated superior predictive accuracy, as reflected by consistently lower Mean Absolute Error (MAE) scores.

A comparative model evaluation was carried out across different tasks and datasets, as shown in Figure 4.1 (module "Model Evaluation" within the "Performance Evaluation and Model Assessment" phase). To further explore variability and robustness, boxplots and error distribution graphs were introduced, offering a visual assessment of each model's stability. These representations reinforce the advantages of ensemble-based methods over traditional forecasting techniques.

A summary of the obtained results is provided in Table 4.17, detailing model performance across evaluation metrics and predictive scenarios.

The analysis began with ARIMA, one of the most recognised and commonly used time series analysis methods. Despite being well-established, ARIMA struggled to capture the peculiar and nonlinear dynamics of crowdfunding campaigns, as reflected in its MAE of 39,275.33.

Logistic regression, previously identified in the thesis as an effective tool for assessing the probability of campaign success, produced an accuracy of 95.05%. This finding

Model	Metric	Value
ARIMA	MAE	39,275.33
Logistic Regression	Accuracy	95.05%
Random Forest	MAE	26,891.11
XGBoost	MAE	49,863.77
LSTM	MAE	578,717.81

Table 4.17: *Summary of model benchmarking results*

reinforces the idea that qualitative variables and early campaign performance reliably predict the outcome.

Among the advanced machine learning models, Random Forest demonstrated the best performance, recording the lowest MAE of 26,891.11. This confirms the thesis's intuition that ensemble algorithms based on decision trees are highly effective in capturing the complexity of crowdfunding data.

XGBoost, while known for its capability in handling structured and temporal data, showed a higher MAE of 49,863.77, suggesting that further hyperparameter tuning could enhance its predictive accuracy.

LSTM, a recurrent neural network designed to capture complex temporal patterns, exhibited the highest MAE of 578,717.81, indicating its difficulty in interpreting the structure of the available data. This aligns with previous discussions on the necessity of extensive and well-structured datasets for deep learning models.

This visualisation also confirms that LSTM performed significantly worse, reinforcing the need for further analysis and optimisation.

Advanced models such as Random Forest prove more accurate and reliable than traditional methods such as ARIMA. While logistic regression is useful for rapidly classifying campaign success, decision-tree-based models offer superior predictive capabilities. Future optimisation of XGBoost and LSTM could enhance this framework, improving the understanding and forecasting of crowdfunding dynamics. Furthermore, integrating real-time market signals with historical data may provide more refined and adaptable predictive frameworks. SHAP-based interpretability supports a more transparent understanding of model decisions (see Figure 4.1, module "Model Evaluation"; Performance Evaluation and Model Assessment phase).

4.2.6 Model Interpretability with SHAP

In predictive modelling for equity crowdfunding campaigns, a crucial aspect lies in selecting a metric capable of effectively capturing the degree to which predefined objectives are met. Among the various metrics explored and thoroughly discussed in

Chapter 4 (Section 4.5.2), the *Funding Ratio*—defined as the ratio between the amount raised and the campaign's minimum funding target—was selected as the main analytical variable. This choice is motivated by the fact that, unlike binary variables (such as *Success*) or more complex measures (such as *Overfunding*, which—although continuous—only captures campaigns that exceeded their funding goal), the Funding Ratio provides a continuous, intuitive, and directly interpretable measure. It is thus more suitable for early-stage monitoring and assessment of campaign dynamics. These characteristics align well with the interpretability framework provided by SHAP, which relies on the ability to transparently explain the contribution of each input variable to the model's output.

To enhance the transparency and comprehensibility of the predictions related to the *Funding Ratio* on the L10 dataset, the *SHapley Additive exPlanations* (SHAP) framework introduced by [75] was employed. SHAP builds upon cooperative game theory to assign each feature a *Shapley value*, which quantifies the average marginal contribution of that feature across all possible feature coalitions. In this analysis, the `TreeExplainer` implementation was adopted, as it is optimised for ensemble-based models such as XGBoost. This explainer leverages the internal structure of tree-based models to efficiently compute exact Shapley values, ensuring consistency and local accuracy.

The predictive model was trained using the XGBoost algorithm, with a squared error loss function (reg:squarederror) suitable for continuous regression tasks. The model was trained on a subset comprising 517 observations and 12 numerical variables. Hyperparameters were selected through a grid search with 5-fold cross-validation to mitigate overfitting and ensure generalizability. To address scale heterogeneity and skewed distributions, a logarithmic transformation (natural logarithm) was applied to monetary variables—specifically `Pre money evaluation`, `Min amount`, and `Max amount`. This transformation preserved the monotonicity of the data while stabilising variance, reducing the influence of outliers, and improving model convergence. An illustrative example of this transformation is provided in Table 4.18, where both the original and log-transformed values are shown.

Table 4.18: *Example of Log-Transformed Variables*

Variable	**Original Value**	**Log-Transformed Value**
Pre money evaluation	1,000,000	13.82
Min amount	100,000	11.51
Max amount	500,000	13.12

This preprocessing step helps enforce a more linear relationship between predictors and the response variable.

The model performance, evaluated on the test set, is summarised in Table 4.19, which reports

Table 4.19: *Model Performance Metrics*

Metric	**Value**
Mean Squared Error (MSE)	0.084
Coefficient of Determination (R^2)	0.046

While predictive accuracy remains moderate, the model retains sufficient internal consistency to support an informative interpretability analysis.

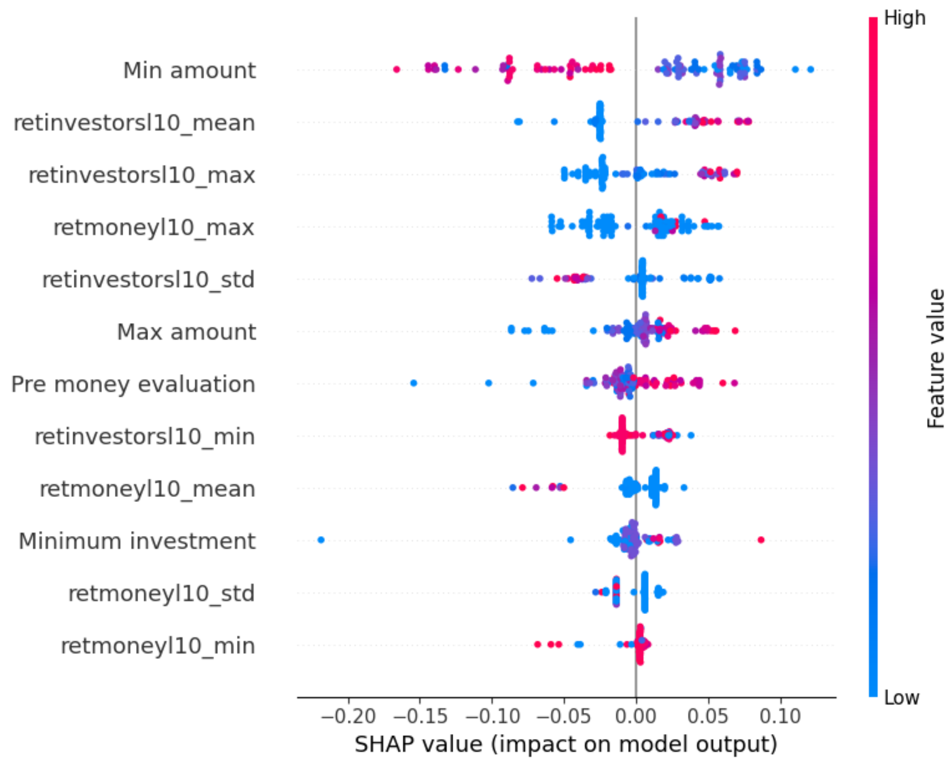

Figure 4.4: SHAP summary plot for the XGBoost model predicting the Funding Ratio (Equity L10 dataset).

The SHAP analysis reveals that aggregate features derived from time series variables—particularly `RetMoneyl10_mean` (mean daily funds raised in the first 10 days) and `RetInvestorsl10_max` (maximum number of daily investors)—exert a strong influence on the model's predictions. This suggests that early performance dynamics, such as initial fundraising velocity and peak investor engagement, are key determinants of campaign outcomes. Moreover, high values of `RetMoney110_mean` are associated with positive SHAP values, indicating a strong positive contribution to the Funding Ratio prediction.

Other influential variables include the log-transformed `Min amount` and `Pre money evaluation`, which also exhibit clear directional effects: higher minimum targets (log-transformed) tend to negatively influence predictions, potentially reflecting increased funding thresholds perceived as more challenging to achieve by potential investors.

In conclusion, adopting interpretability techniques such as SHAP allows for a granular and transparent understanding of the internal decision logic of machine learning models. This approach fits within the broader paradigm of *Explainable Artificial Intelligence* (XAI), which is particularly valuable in high-impact financial and strategic decision-making settings [76]. SHAP enhances model transparency and supports actionable insights, guiding platform managers and campaign designers in identifying the most influential features contributing to early-stage success. This section summarises the main experimental insights and performance outcomes (see Figure 4.1, module "Benchmarking ...(see Figure 4.1, module "Benchmarking & Results"; part of the "Performance Evaluation and Model Assessment" phase in Figure 3.1).

4.3 Key Findings and Advancements Over Previous Research

My research represents a substantial advancement in the field of predicting the success of crowdfunding campaigns through the application of sophisticated machine learning techniques, such as XGBoost and Long Short-Term Memory (LSTM) networks. These advanced methods allowed us to incorporate temporal data into our models—a crucial aspect that has often been neglected in earlier studies. By doing so, we achieved a significant improvement in predictive accuracy. Unlike conventional approaches that rely solely on static or aggregated data, LSTM networks excel in capturing the temporal evolution of investor behaviour. This capability offers a deeper and more dynamic understanding of how investors engage with campaigns over time, yielding predictions that are not only highly accurate but also contextually nuanced and adaptable. A key contribution of this research lies in the development of two innovative datasets specifically tailored for analysing crowdfunding campaigns and lending activities. These datasets, meticulously gathered through advanced web scraping methods, serve as a unique and comprehensive resource tailored to the specific characteristics of the Italian crowdfunding ecosystem. This facilitates a more detailed examination of campaign dynamics. Unlike many previous studies, which have been constrained by a narrow focus on a single type of platform or limited datasets, our research leverages the inclusion of temporal data. This enables the real-time tracking and analysis of campaign progress, allowing us to capture the evolving nature of these initiatives with unparalleled precision.

In addressing the limitations of earlier predictive models, our study provides notable improvements. For instance, prior work, such as that by Ralcheva and Roosenboom (2020), predominantly relied on regression models to predict equity crowdfunding campaign success, focusing on static variables like offered equity and company age [18]. By contrast, our approach integrates temporal variables and utilises more sophisticated machine learning models, such as XGBoost and LSTM, which are designed to handle the intricate complexities of dynamic datasets effectively [77, 78]. This innovative combination of methods enabled us to deliver predictions that are not only more accurate but also actionable, particularly in the lending sector, where the behaviour of investors tends to be highly variable over time.

To further enhance the robustness and effectiveness of our predictive models, we implemented advanced techniques such as BorderlineSMOTE to address issues associated with imbalanced datasets.

This method proved particularly beneficial in lending campaigns, where the occurrence of successful outcomes is far more common than in equity campaigns. The ability to effectively manage such imbalances represents a substantial leap forward compared to traditional methods, which have often overlooked this critical challenge. By doing so, our approach ensured a more balanced and fair analysis, leading to improved model performance and greater predictive reliability.

The practical implications of these findings are far-reaching, particularly for crowdfunding platform managers and entrepreneurs seeking funding for their projects. The ability to accurately predict the success of a campaign during its early stages provides a valuable tool for optimising both project selection and marketing strategies. Platforms can utilise these insights to allocate resources more efficiently, minimise the likelihood of failed campaigns, and enhance overall operational effectiveness. However, one notable challenge encountered during this study was the limited collaboration with platform managers. In some cases, managers were hesitant to fully engage due to concerns about the transparency of the algorithm, which could reveal potential weaknesses or failures in past campaigns. This reluctance posed constraints on data availability, and the closure of Startwallet further limited the scope of our dataset.

My study also introduced more sophisticated metrics for assessing campaign success, such as the Funding Ratio and Overfunding. These metrics move beyond the simple binary categorisations that have dominated earlier research, offering a more detailed and refined evaluation of success. This enables platform managers to make more informed decisions and better identify and support campaigns with high potential. Adopting these advanced metrics represents a significant step forward in providing actionable insights for stakeholders in the crowdfunding ecosystem.

Despite the valuable insights gained from our current dataset, the development of more comprehensive data remains essential for guiding fundraisers in selecting optimal platforms. Recent advancements in generative AI, especially in the domain of financial applications [79], suggest that techniques such as diffusion models could play a pivotal role in overcoming data limitations. Building on existing approaches to synthetic data generation [80], future research could explore the use of generative models to create realistic and highly representative campaign data. Such data would preserve the intricate relationships observed in successful campaigns, enabling a deeper and more robust analysis.

Generative models offer capabilities that go beyond mere data augmentation. They can simulate the performance of campaigns across different platforms and market conditions, providing insights that are otherwise difficult to obtain through conventional data collection methods. By integrating synthetic data generation with traditional predictive modeling techniques, researchers can establish a more comprehensive framework for understanding and analysing crowdfunding dynamics. Future studies should focus on refining these methods to ensure that the generated data accurately mirrors the complexities of real-world scenarios. Another promising avenue involves incorporating qualitative data, such as investor engagement on social media platforms. Recent findings suggest that online interactions play a significant role in determining crowdfunding success, and their inclusion could further improve the accuracy of predictions and offer a holistic perspective on investor behaviour [81].

Finally, the predictive algorithm developed in this thesis is not restricted to national applications but is adaptable for use across various European contexts. This adaptability is rooted in the reliance on standardised features, such as the number of investors, the amount raised, and the duration of campaigns—variables commonly utilised by numerous European platforms and aggregators. Leading platforms like Nebula, NextFin, Findcrowdfunding, Hellocrowdfunding, and Crowdfunding.de base much of their predictive analyses on similar parameters [52]. The widespread standardisation of these metrics ensures that our model can be seamlessly applied across borders, offering a valuable tool for comparative analyses. These comparative insights confirm the proposed modelling approach's adaptability and robustness (see Figure 4.1, module "Benchmarking & Results"; it is part of the "Performance Evaluation and Model Assessment" phase in Figure 3.1).

Incorporating these standardised metrics, our findings align with broader European trends in crowdfunding research. Scholars such as Blaseg et al. (2021) have highlighted the importance of integrating similar success metrics across European platforms, demonstrating the cross-context relevance of our approach [69]. Furthermore, platforms like EvoEstate and BrickCrowdfunding, which track investor interactions and funds raised over time, validate the effectiveness of employing standardised parameters in predictive modeling [52]. Research on overfunding, such as that by Martínez-Gómez et al. (2020) and Koch (2016), further supports the inclusion of dynamic metrics to adapt predictive models to diverse national contexts [60, 59].

By building on these insights, our model proves its applicability to leading platforms, including Crowd Circus and Crowdfunding Finder, which adopt similar methodologies to analyse investor engagement and campaign performance. This alignment ensures that the algorithm remains attuned to contemporary trends and demands in the European crowdfunding landscape.

4.4 Key Drivers of Crowdfunding Success

The achievement of success in crowdfunding campaigns is contingent upon a variety of factors that can be categorised into pre-launch and post-launch elements. As outlined by [59], pre-launch factors such as pre-committed funds and the initial number of investors play a pivotal role in the development of a campaign. These elements serve as indicators of the campaign's credibility, thereby attracting early investments. The presence of pre-committed funds and a substantial initial investor base are critical predictor of the campaign's potential for future success, as they establish a foundation of trust and momentum that can drive further investor engagement.

In contrast, the dynamics that unfold after the campaign has launched focus primarily on the rate at which new investors engage with the campaign and the daily inflows of capital as the campaign progresses. Research, including that conducted by [60], highlights the importance of post-launch metrics such as overfunding, which serve as significant indicators of a campaign's ongoing performance and investor confidence. The introduction of advanced metrics like the Funding Ratio and Overfunding in our study facilitates a more granular and comprehensive analysis of campaign performance, moving beyond traditional measures to capture the nuanced aspects of campaign dynamics.

Building upon the existing body of literature, our predictive model integrates both pre-launch and post-launch data, thereby enhancing the accuracy of success predictions throughout the entire lifecycle of a crowdfunding campaign.

By adopting a two-phase approach, our methodology enables the formulation of early predictions based on pre-launch factors, such as the number of initial investors and the amount of pre-committed funds, and subsequently refines these predictions with post-launch data, including daily contributions and ongoing investor engagement. This iterative process ensures that our predictions remain aligned with the continuously evolving behaviour of investors, allowing for real-time adjustments and more precise forecasting.

The amalgamation of these diverse metrics allows us to transcend the simplistic binary classification of success versus failure that has been prevalent in earlier studies. By quantifying success using continuous variables like the Funding Ratio, our approach provides a more sophisticated and nuanced understanding of campaign performance. This level of detail is crucial for crowdfunding platforms that aim to optimise campaign outcomes and enhance investor engagement, as it allows for more informed decision-making and strategic planning based on a comprehensive set of performance indicators.

My empirical findings concur with the work of researchers such as [69], who advocate for integrating static and dynamic factors to achieve a more thorough and comprehensive analysis.

Emphasising overfunding as a dynamic measure of investor confidence aligns with broader European trends in crowdfunding campaigns. This focus has practical implications for campaign managers and platform developers, underscoring the importance of maintaining campaign momentum and sustaining investor interest over time. By highlighting these dynamic measures, our study provides actionable insights that can inform strategies to foster continuous investor engagement and ensure the long-term success of crowdfunding initiatives.

Our dataset includes features derived from an exhaustive study of various European crowdfunding platforms, including Crowdcube, NextFin, Findcrowdfunding, Hellocrowdfunding, and Crowdfunding.de. This foundational research underscores the adaptability of our algorithmic approach across different international contexts by leveraging comparable features and structural elements observed in these diverse platforms. The consistency in key variables across platforms facilitates the application of our predictive model in various European settings, enhancing its robustness and generalizability.

Integrating pre-launch and post-launch factors into a unified predictive model represents a significant advancement in the analysis of crowdfunding campaigns. It provides a robust framework for understanding and forecasting campaign success.

As new data becomes available, the continuous refinement of predictions further strengthens the model's accuracy and reliability. This comprehensive approach ensures that the model remains responsive to the dynamic nature of crowdfunding campaigns, allowing for ongoing improvements and adjustments based on the latest data and trends.

While the methodologies employed in our study exhibit broad applicability, testing them across diverse geographical settings is imperative to ensure their robustness and reliability. Future analyses should aim to incorporate a wider range of datasets, encompassing different cultural and regional contexts and additional qualitative and cultural variables. Such enhancements would significantly deepen the model's analytical capabilities and bolster its resilience against varying market conditions and investor behaviours.

Incorporating factors like social media engagement, which recent studies have identified as a critical influence on investor behaviour, would further elevate the model's predictive accuracy and provide a more holistic view of the mechanisms driving crowdfunding success [81].

4.4.1 Applicability of the Model to European Platforms

The analysis conducted in this research demonstrates that the predictive model developed is not only applicable to the Italian market but can also be extended to European crowdfunding platforms. This possibility is largely due to the transparency and comprehensive data provided by these platforms, which report key indicators such as total funds raised, number of investors, success rate, average investment value, fundraising speed, and conversion rate—features that closely align with those used in our model.

A distinctive aspect of our model is its incorporation of the pre-launch phase, a component that is not explicitly captured in the datasets of many international platforms. However, alternative proxies can be identified. For example, platforms like Kickstarter, Crowdcube, and Seedrs provide metrics on the number of followers prior to launch, social media interactions, and website traffic to campaign pages, which can serve as substitutes to reconstruct pre-launch dynamics similar to those analysed in this study.

Moreover, the European context is characterised by stricter regulatory standards, which mandate greater transparency and systematic reporting of fundraising data. Far from being an obstacle, this regulatory environment enhances data quality and granularity, potentially improving the model's predictive accuracy. The availability of structured data on institutional investments, overfunding levels, and investor engagement further enriches the model, allowing it to be adapted to the specificities of the European market.

In essence, applying the proposed method to non-Italian platforms relies on our capacity to pinpoint suitable indicators for the pre-launch phase and recalibrate key variables per the specific data structures available.

Although differences in market success criteria might necessitate some customisation, the scalability and adaptability of the method render it a valuable tool for European platforms and investors aiming to enhance risk assessment and forecast campaign success.

In summary, the predictive model developed in this research is transferable to European crowdfunding platforms with only minor adjustments required to optimise its calibration for each market's specific characteristics. By adopting suitable feature engineering strategies, the model's capability to predict campaign success in a regulated international environment can be further enhanced, thereby broadening its scope of application.

This perspective opens promising avenues for future research, such as integrating indicators related to country-specific regulations, tailoring the model to markets with varying levels of maturity, and leveraging advanced machine learning techniques to refine predictions based on the unique data availability across different crowdfunding ecosystems.

4.5 U.S. dataset analysis and international comparison

An experimental extension of the computational framework was carried out using data from the U.S. Securities and Exchange Commission (SEC) Reg CF filings [82]. This dataset provides detailed information on each campaign, including the amount raised, the type of security offered, and the platform used. A total of **5,369 campaigns** were analysed, reflecting the broader scale and greater maturity of the U.S. market compared to the Italian sample.

The availability of such a comprehensive dataset enabled the full application of the methodological framework, encompassing time-series normalisation, feature engineering, and the implementation of supervised machine learning models (XGBoost, LSTM) — to explore the variability and distribution of campaign outcomes.

Distribution of Funds Raised by Platform (Top 10 - USA) fig. 4.5 presents a boxplot showing how funds are distributed across the ten most active platforms.

A few major players, such as Republic, SeedInvest, StartEngine, and Wefunder display higher median amounts and broader distributions, indicating their ability to host large-scale campaigns and attract significant funding volumes.

Conversely, platforms like HoneyComb and MainVest show lower medians and a more concentrated distribution, indicating a uniform campaign size typically associated with smaller funding targets. However, several outliers demonstrate that some campaigns have raised exceptionally high amounts, underscoring the significant impact of factors such as platform specialisation and investor base on fundraising outcomes.

Overall, the analysis reveals a market in which capital is concentrated among a few leading platforms, while a broad array of smaller operators manage more stable campaigns with lower funding levels.

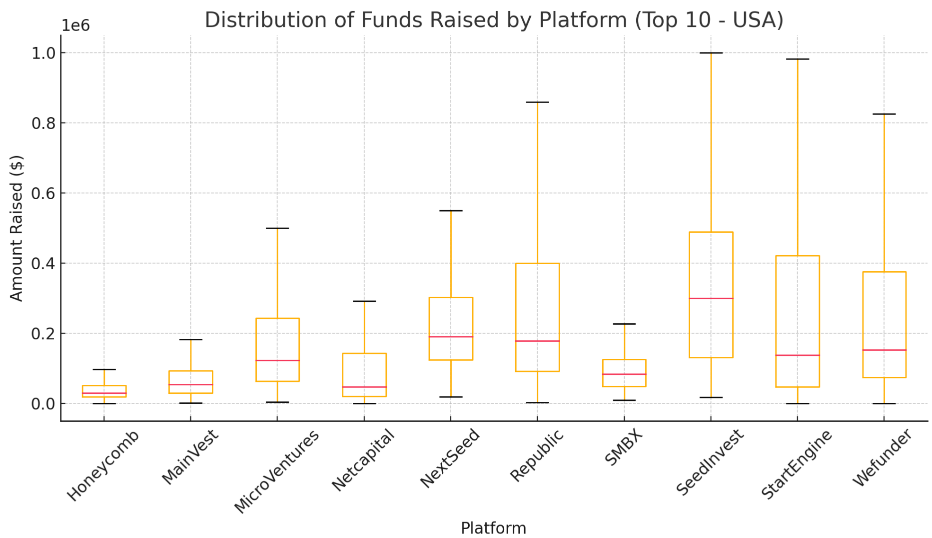

Figure 4.5: Distribution of Funds Raised by Platform (Top 10 - USA)

To better contextualise the dynamics of the U.S. market, we analysed several descriptive statistics derived from the dataset. These figures not only offer a clear view of the overall scale of activity but also serve as a valuable benchmark for comparison with other countries.

The results further confirm the distinctive characteristics of the U.S. market:

- Total number of campaigns: **5,369**
- Mean amount raised per campaign: **$325,261**
- Median amount raised: **$114,850**
- Maximum amount raised: **$5,000,000**

Building on this analysis, the following table provides a comparative overview of the Italian and U.S. equity crowdfunding markets, highlighting key structural and performance differences between the two contexts:

Variable	**Italy**	**USA (Reg CF)**
Number of campaigns	200	5,369
Estimated success rate	65%	60% (estimated)
Average funding target	€150,000	$180,000
Maximum amount raised	€2,500,000	$5,000,000
Average number of employees	8	10
Average revenue	€400,000	$500,000

Table 4.20: *Comparison of equity crowdfunding campaigns in Italy and the USA*

To further assess the predictive potential of the available features, we conducted a linear regression analysis on the U.S. dataset. The model aimed to estimate the *Mean Absolute Error (MAE)* and the *Coefficient of Determination (R^2)*, providing a first empirical insight into the framework's performance when applied to this context.

Given the limited availability of dynamic variables, such as specific funding targets and detailed daily investment flows, the regression was carried out using only the *Platform* variable as a predictor. This choice reflects the constraints of the dataset but also offers an opportunity to explore the explanatory power of platform-related features in isolation.

Beyond the numerical results, this exercise highlights the crucial importance of incorporating granular and time-dependent variables to improve model accuracy. It also provides a practical perspective on the challenges of applying predictive models across different datasets and market environments.

The regression results are reported below:

- **MAE**: $305,915
- **R^2**: 0.118

While these values are relatively modest, they clearly underscore the central role of high-quality, context-specific features in predictive modeling. The limited performance of the model is largely attributable to the absence of dynamic variables such as the *Funding Ratio* and *Overfunding* that represent key components of the framework developed in this thesis.

Nevertheless, this result reinforces the framework's adaptability when applied to different market contexts. Despite structural differences—such as higher market concentration and the use of more complex financial instruments—the methodology proves robust and capable of capturing the underlying dynamics of equity crowdfunding performance.

The comparative analysis between Italian and U.S. crowdfunding campaigns revealed structural similarities as well as notable differences, particularly in terms of investment volumes, campaign duration, and early-stage investor response. These findings reinforce the validity of the predictive models developed in this work, which demonstrate strong adaptability even when applied to heterogeneous datasets and suggest their potential transferability to broader and more diverse operational contexts.

4.6 Deployment, Scalability and Integration

In light of the obtained results and the demonstrated generalizability of the models on international datasets, it becomes crucial to investigate the operational implications of deploying these predictive models in real-world environments. This section aims to provide a technical discussion on the infrastructural, scalability, and deployment aspects that are key to integrating the predictive models into existing crowdfunding platforms, thereby enhancing the applied scientific contribution of this doctoral research.

As detailed in previous chapters, the predictive models—specifically XGBoost and LSTM—were rigorously trained using real-world data from equity and lending crowdfunding campaigns gathered from multiple Italian platforms. These models were engineered to analyze investment time series, enabling the estimation of advanced performance indicators, such as the Funding Ratio and the Overfunding level, beyond the conventional binary outcomes. Ensuring the applicability of these sophisticated predictive tools in production requires addressing several challenges concerning scalability and model availability.

A cloud-based infrastructure emerges as a technically viable solution due to its inherent ability to manage computational loads elastically, dynamically scaling resources in response to the fluctuating demands typical of active crowdfunding campaigns. Such infrastructures provide access to specialized hardware resources—for instance, GPUs that are critical for efficient sequential model inference—and facilitate integration with advanced services for security, real-time monitoring, and version control.

Containerization, particularly through Docker, plays a pivotal role in this context by isolating the model execution environment and ensuring portability across diverse operating systems. This strategy is essential for managing library dependencies—such as xgboost, tensorflow, scikit-learn, pandas, and numpy—in a consistent manner, thereby guaranteeing the reproducibility of experimental results. Containers mitigate potential conflicts from version incompatibilities or environmental discrepancies, which is a critical consideration for large-scale machine learning architectures.

Moreover, the integration with Kubernetes, a well-established container orchestration system in cloud-native environments, provides granular control over the distributed infrastructure. Key technical benefits include:

Horizontal Auto-Scaling: This feature facilitates the automatic instantiation of additional inference service instances in response to increased demand. Nevertheless, the effectiveness of auto-scaling is contingent on the instance startup latency and the speed at which models are loaded into memory, particularly for complex architectures like LSTM. Implementing warm-up strategies or maintaining pre-warmed instances is therefore advisable.

Load Balancing: By distributing inference requests evenly across multiple model replicas, load balancing minimises bottlenecks and enhances real-time service quality, as observed in campaign monitoring dashboards.

Fault Tolerance: Mechanisms that automatically restart containers in the event of a failure bolster the system's resilience. This feature is indispensable for maintaining continuous service in production environments characterised by high loads or intermittent instability.

Rolling Updates: The ability to perform rolling updates permits the continuous deployment of new model versions without disrupting the service—a critical attribute for platforms that require frequent model updates driven by the influx of new campaign data.

To ensure long-term system reliability, implementing an end-to-end MLOps pipeline is essential. Such a pipeline should encompass systematic monitoring of predictive performance, prompt detection of concept and data drift phenomena, and automation of the model retraining process. In the dynamic crowdfunding environment, where investment behaviours and user interactions can evolve rapidly, the absence of an adaptive mechanism could undermine model validity. Continuous monitoring systems capable of identifying decreases in accuracy metrics (e.g., MAE, R^2) and shifts in feature distributions are therefore imperative to sustain the effectiveness and reliability of the predictive system over time.

This technical framework bolsters the proposed models' operational viability and reinforces the thesis's applied contributions by bridging the gap between theoretical advancements and practical deployment in complex, real-world financial ecosystems. The findings presented in this chapter validate the adoption of a multilevel predictive framework for modelling crowdfunding campaign performance. By leveraging temporal fundraising patterns, success indicators, and a comparative evaluation of machine learning algorithms, the analysis establishes a structured approach to forecasting campaign outcomes. The concluding section builds on these results by reflecting on their methodological implications and outlining practical and research-oriented perspectives for future development.

Conclusion

This thesis has explored the application of machine learning techniques to predictive modelling in equity and lending crowdfunding. Leveraging a modular and operational framework, the research has developed a data-driven pipeline capable of addressing the temporal, categorical, and behavioural complexity of campaign dynamics. Key methodological contributions include the construction of pre- and post-launch features, the treatment of missing data, and the design of success metrics extending beyond binary classification. By integrating static campaign characteristics with dynamic investor behaviours, the forecasting system gains temporal awareness and contextual relevance.

Empirically, the results confirm the effectiveness of tree-based models—particularly Random Forests—in predicting early campaign outcomes, outperforming both linear baselines and deep learning alternatives in terms of mean absolute error. The use of SHAP values has enhanced model interpretability by identifying the most influential features and behavioural signals associated with successful campaigns. Furthermore, the comparative analysis between Italian and U.S. platforms illustrates how predictive frameworks can be adapted across regulatory environments, provided a rigorous preprocessing strategy is adopted.

From a theoretical standpoint, the study contributes to the literature on alternative finance by presenting a methodological approach that combines time series modelling, feature engineering, and explainable AI. This foundation supports future applications aimed at leveraging real-time behavioural data to optimise campaign strategies and inform platform-level decision-making. The findings open avenues for integrating behavioural finance, temporal analytics, and algorithmic forecasting within the broader context of digital entrepreneurship.

During the course of this research, several critical issues emerged that offer valuable insights for future investigation.

First, the limited availability of open, structured, and FAIR-compliant datasets constrained the potential for large-scale replication and cross-platform validation.

Second, structural heterogeneity among crowdfunding platforms—including differences in reporting standards, goal-setting logic, and the availability of pre-launch metrics—complicates the immediate transferability of the proposed models to new contexts. These differences call for tailored adaptation and calibration procedures.

Finally, although a comparative analysis with the U.S. SEC dataset was conducted to assess international applicability, this represents a preliminary validation. Extending the framework to other platforms such as Seedrs or Indiegogo will require further alignment in terms of dataset structure, feature granularity, and campaign design.

Addressing these issues will be essential for advancing the generalisability and practical deployment of predictive models in the crowdfunding landscape. Future research could prioritise the integration of real-time behavioural indicators, the promotion of data standardisation protocols, and the testing of model robustness across diverse regulatory and geographical contexts.

Bibliography

[1] Joshua Madsen and Jeff L McMullin. "Economic consequences of risk and ability disclosures: evidence from crowdfunding". In: *Available at SSRN 3202453* (2018).

[2] Carla Martnez-Climent, Ana Zorio-Grima, and Domingo Ribeiro-Soriano. "Financial return crowdfunding: literature review and bibliometric analysis". In: *International Entrepreneurship and Management Journal* 14 (2018), pp. 527–553.

[3] Xue Yang, Huiling Wang, and Fangyue Li. "The impact of past fundraising experiences on the fundraising performance of equity crowdfunding projects". In: *Electronic Commerce Research* (2024), pp. 1–21.

[4] Alexandra Mora, José Saura, and Pedro Palos-Sanchez. "Social Media and User-Generated Content as a Teaching Innovation Tool in Universities". In: IGI Global, June 2022, pp. 52–67. ISBN: 9781668444412. DOI: `10.4018/978-1-6684-4441-2.ch004`.

[5] Prince Baah-Peprah, Rotem Shneor, and Ziaul Haque Munim. ""In this together": on the antecedents and implications of crowdfunding community identification and trust". In: *Venture Capital* (2024), pp. 1–31.

[6] Chaoran Cheng et al. "Success Prediction on Crowdfunding with Multimodal Deep Learning". In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*. 2019, pp. 2158–2164.

[7] Yenchun Jim Wu Wei Wang a Lihuan Guo b. "The merits of a sentiment analysis of antecedent comments for the prediction of online fundraising outcomes". In: *Technological Forecasting and Social Change* (Jan. 2022). DOI: `https://doi.org/10.1016/j.techfore.2021.121070`.

[8] Jun Wang et al. "Crowdfunding Dynamics Tracking: A Reinforcement Learning Approach". In: *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 2020, pp. 6210–6218.

[9] Chao Zhong, Wei Xu, and Wei Du. “Success prediction of crowdfunding campaigns with project network: A machine learning approach”. In: *Journal of Electronic Commerce Research* 23.2 (2022), pp. 99–114.

[10] Ivana Pais, Paola Peretti, and Chiara Spinelli. *Crowdfunding: la via collaborativa all’imprenditorialità*. EGEA spa, 2018.

[11] Khalid Usman Sheng Bi Zhiying Liu. “The influence of online information on investing decisions of reward-based crowdfunding”. In: *Journal of Business Research* 71 (2017), pp. 10–18. ISSN: 0148-2963. DOI: https://doi.org/10.1016/j.jbusres.2016.10.001. URL: https://www.sciencedirect.com/science/article/pii/S0148296316305707.

[12] Manuchehr Shahrokhi and AM Parhizgari. “Crowdfunding in real estate: evolutionary and disruptive”. In: *Managerial Finance* 46.6 (2020), pp. 785–801.

[13] Tobias Guggenberger et al. “Kickstarting Blockchain: Designing Blockchain-based Tokens for Equity Crowdfunding”. In: *Electronic Commerce Research* 24 (Feb. 2023). DOI: 10.1007/s10660-022-09634-9.

[14] Evila Piva and Cristina Rossi-Lamastra. “Human Capital Signals and Entrepreneurs’ Success in Equity Crowdfunding”. In: *Small Business Economics* 51 (2018), pp. 667–686.

[15] Paul Belleflamme, Nessrine Omrani, and Martin Peitz. “The economics of crowdfunding platforms”. In: *Information Economics and Policy* 33 (2015), pp. 11–28. ISSN: 0167-6245. DOI: https://doi.org/10.1016/j.infoecopol.2015.08.003. URL: https://www.sciencedirect.com/science/article/pii/S0167624515000463.

[16] Paolo Roma, Maria Vasi, and Christos Kolympiris. “On the signaling effect of reward-based crowdfunding:(When) do later stage venture capitalists rely more on the crowd than their peers?” In: *Research Policy* 50.6 (2021), p. 104267.

[17] Daniel Blaseg, Douglas Cumming, and Michael Koetter. “Equity Crowdfunding: High-Quality or Low-Quality Entrepreneurs?” In: *Entrepreneurship Theory and Practice* 45.3 (2021), pp. 505–530. DOI: 10.1177/1042258719899427. URL: https://doi.org/10.1177/1042258719899427.

[18] Aleksandrina Ralcheva and Peter Roosenboom. “Forecasting Success in Equity Crowdfunding”. In: *Small Business Economics* 55 (2020), pp. 39–56.

[19] Koki Kurihara and Yuji Honjo. *Determinants of Campaign Success: Empirical Evidence from Equity Crowdfunding in Japan*. Tech. rep. Research Institute of Economy, Trade and Industry (RIETI), 2022.

[20] Mohammad Yunies Edward et al. “Success Factors for Peer-to-peer Lending for SMES: Evidence from Indonesia”. In: *Innovations* 20.2 (2023), pp. 16–25.

[21] Anoosheh Rostamkalaei and Mark Freel. “Some initial observations on the geography of the supply of equity crowdfunding”. In: *Venture Capital* 25.1 (2023), pp. 65–90.

[22] Benjamin Le Pendeven and Armin Schwienbacher. “Equity crowdfunding: the influence of perceived innovativeness on campaign success”. In: *British Journal of Management* 34.1 (2023), pp. 280–298.

[23] Annabeth Aagaard and Filip Rezac. “Governing the interplay of inter-organizational relationship mechanisms in open innovation projects across ecosystems”. In: *Industrial Marketing Management* 105.December 2020 (2022), pp. 131–146. DOI: 10.1016/j.indmarman.2022.06.003.

[24] Zihui Cai et al. “Multimodal dynamic graph convolutional network for crowdfunding success prediction”. In: *Applied Soft Computing* 154 (2024), p. 111313. DOI: 10.1016/J.ASOC.2024.111313.

[25] Y. Xia, Y. Sun, and F. Li. “Social Media and Its Impact on Crowdfunding Campaigns: A Sentiment Analysis Approach”. In: *Journal of Business Research* 145 (2024), pp. 89–102. DOI: 10.1016/j.jbusres.2023.01.010.

[26] Lafaiet Silva, Nádia Félix Silva, and Thierson Rosa. “Success prediction of crowdfunding campaigns: A two-phase modeling”. In: *International Journal of Web Information Systems* 16.4 (2020), pp. 387–412. DOI: 10.1108/IJWIS-05-2020-0026.

[27] Michael Safo Oduro, Han Yu, and Hong Huang. “Predicting the Entrepreneurial Success of Crowdfunding Campaigns Using Model-Based Machine Learning Methods”. In: *International Journal of Crowd Science* 6 (Mar. 2022), p. 7. DOI: 10.26599/IJCS.2022.9100003.

[28] Susana Bernardino and José De Freitas Santos. “Cultural differences and crowdfunding: a four-country study”. In: *E-Revista de Estudos Interculturais* (2022).

[29] Xiang Gao, Weige Huang, and Sunghan Ryu. “Effects of sentiment and emotion of campaign pitch on crowdfunding performance: A cross-cultural comparison”. In: *Journal of Research in Emerging Markets* 3 (Sept. 2021). DOI: 10.30585/jrems.v3i4.711.

[30] Esther Wanjiru Wachira and Virginia Kirigo Wachira. “Crowdfunding in Kenya: Factors for Successful Campaign — The Case of Kickstarter Crowdfunding Platform”. In: *Public Finance Quarterly* 66.3 (2021), pp. 413–428. URL: https://EconPapers.repec.org/RePEc:pfq:journl:v:66:y:2021:i:3:p:413-428.

[31] Martin Angerer et al. “Start-up Funding via Equity Crowdfunding in Germany – A Qualitative Analysis of Success Factors”. In: *The Journal of Entrepreneurial Finance* 19 (Jan. 2017). DOI: 10.57229/2373-1761.1290.

[32] Guerram Adel, Pr Maeda Yuji, and Pr MEZAACHE Ali. “Success factors in equity crowdfunding: The case of Japan”. In: *Journal of Financial, Accounting and Managerial Studies* (2022). ISSN: 2352-9962.

[33] Yuji Honjo and Koki Kurihara. "Target for campaign success: an empirical analysis of equity crowdfunding in Japan". In: *The Journal of Technology Transfer* (May 2023), pp. 1–33. DOI: 10.1007/s10961-023-10010-4.

[34] Muna M. Alhammad, Rawan AlOthman, and Chekfoung Tan. "Review of Crowdfunding Regulations across Countries: A Systematic Review Study". In: *Journal of Information Systems Engineering and Management* 6.4 (2021), em0145. DOI: 10.21601/jisem/11395. URL: https://doi.org/10.21601/jisem/11395.

[35] Elżbieta Pohulak-Żołędowska and Agnieszka Wójcik-Czerniawska. "Impact of regulation on investment crowdfunding". In: *Nierówności Społeczne a Wzrost Gospodarczy* 3.79 (2024).

[36] Umair Baig et al. "FinTech Past and Future: Ecosystem, Business Model and its Proximate Challenges". In: *Pakistan Business Review* 24.1 (2022), pp. 40–61.

[37] Internal Market, Industry, Entrepreneurship and SMEs. *Equity Crowdfunding*. Accessed: 2023-06-09. 2023.

[38] Sandra Correia, Miguel Sousa, and Elísio Brandão. "What do we know about the choices of entrepreneurs before the equity crowdfunding campaign?" In: *Small Business Economics* (2024), pp. 1–31.

[39] Giuseppe Modaffari. "Equity Crowdfunding: Literature Analysis". In: *Equity Crowdfunding for SMEs: An Alternative Financing Tool* (2024), pp. 75–112.

[40] Pau Sendra-Pons, Alicia Mas-Tur, and Dolores Garzon. "Anchor investors and equity crowdfunding for entrepreneurs". In: *European Journal of Management and Business Economics* 33.1 (2024). Universitat de València, pp. 20–36.

[41] Mónika Kuti and Gábor Madarász. "Crowdfunding". In: *Public Finance Quarterly* 59.3 (2014), pp. 355–366.

[42] Pema Wangchuk. "Common types of Crowdfunding Models, Related Concepts and Its Impact on Business: A Brief Literature Review". In: *Asian Journal of Economics, Business and Accounting* (Oct. 2021), pp. 56–63. DOI: 10.9734/ajeba/2021/v21i1430471.

[43] Susana Bernardino and J Freitas Santos. "Crowdfunding: an exploratory study on knowledge, benefits and barriers perceived by young potential entrepreneurs". In: *Journal of Risk and Financial Management* 13.4 (2020), p. 81.

[44] Sıdıka Tunç Candoğan et al. "Product Development in Crowdfunding: Theoretical and Empirical Analysis". In: *Manufacturing & Service Operations Management* 26.2 (2024). Published online: 17 January 2024. DOI: 10.1287/msom.2022.0344. URL: https://doi.org/10.1287/msom.2022.0344.

[45] Xue Yang, Hao Wang, and Feng Li. "The impact of past fundraising experiences on the fundraising performance of equity crowdfunding projects". In: *Electronic Commerce Research* (2024). DOI: 10.1007/s10660-024-09815-8. URL: https://doi.org/10.1007/s10660-024-09815-8.

[46] Pau Sendra-Pons, Dolores Garzón, and María-Ángeles Revilla-Camacho. "Catalyzing success in equity crowdfunding: trust-building strategies through signaling". In: *Review of Managerial Science* (2024). DOI: 10.1007/s11846-024-00734-4. URL: https://doi.org/10.1007/s11846-024-00734-4.

[47] Nohyoon Seong et al. *Equity Crowdfunding and Startup Performance.* Working Paper SSRN 4559496. 66 Pages, Posted: 20 Sep 2023, Last revised: 28 Nov 2023. College of Business, Korea Advanced Institute of Science and Technology (KAIST); Peking University - HSBC School of Business; The Chinese University of Hong Kong, Sept. 2023. DOI: 10.2139/ssrn.4559496. URL: https://ssrn.com/abstract=4559496.

[48] Ali Alalwan et al. "Entrepreneurial e-equity crowdfunding platforms: antecedents of knowledge acquisition and innovation performance". In: *European Journal of Innovation Management* 27 (Aug. 2022). DOI: 10.1108/EJIM-03-2022-0167.

[49] Michael A. Stanko and David H. Henard. "Toward a better understanding of crowdfunding, openness and the consequences for innovation". In: *Research Policy* 46.4 (2017), pp. 784–798. ISSN: 0048-7333. DOI: https://doi.org/10.1016/j.respol.2017.02.003. URL: https://www.sciencedirect.com/science/article/pii/S0048733317300379.

[50] Ramya Raghavan. *Wellbeing and Prosocial Crowdfunding for Achieving Sustainable Development Goals.* 2024. DOI: 10.31235/osf.io/b76u9.

[51] G. Paolucci. *Analisi di bilancio. Logica, finalità e modalità applicative. Aggiornata con le novità bilancio 2016.* Milano: Il Sole 24 Ore, 2016.

[52] Money.it. *Cos'è il ROE e come si usa per valutare la redditività aziendale.* https://www.money.it/Cos-e-il-ROE-e-come-si-usa-per-valutare. Accessed: 2024-09-10. 2016.

[53] Wen-Shan Lin, Hong-Ren Chen, and Meng-Li Yang. "How do we learn to get success together through crowdfunding platform? From the perspectives of system learning and multi-motivations". In: *Telematics and Informatics* 52 (May 2020), p. 101428. DOI: 10.1016/j.tele.2020.101428.

[54] Ethan Mollick. "The dynamics of crowdfunding: An exploratory study". In: *Journal of Business Venturing* 29.1 (2014), pp. 1–16. ISSN: 0883-9026. DOI: https://doi.org/10.1016/j.jbusvent.2013.06.005.

[55] Xiaoying Ren et al. "Tracking and Forecasting Dynamics in Crowdfunding: A Basis-Synthesis Approach". In: *2018 IEEE International Conference on Data Mining (ICDM).* 2018, pp. 1212–1217.

[56] Yang Song et al. "Mining and investigating the factors influencing crowdfunding success". In: *Technological Forecasting and Social Change Volume 148* 148 (2019), pp. 43–58. DOI: org/10.1016/j.techfore.2019.119723..

[57] Massara Alazazi, Bin Wang, and Tareq Allan. "Success Factors of Donation-Based Crowdfunding Campaigns: A Machine Learning Approach". In: *Hawaii International Conference on System Sciences*. Jan. 2020, pp. 1–10.

[58] Nir Vulkan, Thomas Åstebro, and Manuel Fernandez Sierra. "Equity Crowdfunding: A New Phenomena". In: *Journal of Business Venturing Insights* 5 (2016), pp. 37–49.

[59] Jascha-Alexander Koch. "The Phenomenon of Project Overfunding on Online Crowdfunding Platforms -– Analyzing the Drivers of Overfunding". In: *Banking & Insurance eJournal* (2016). URL: `https://api.semanticscholar.org/CorpusID:29781112`.

[60] Cristina Martínez-Gómez, Francisca Jiménez-Jiménez, and M. Virtudes Alba-Fernández. "Determinants of Overfunding in Equity Crowdfunding: An Empirical Study in the UK and Spain". In: *Sustainability* 12.23 (2020). ISSN: 2071-1050.

[61] Yijing Li et al. "Exploring the Impact of Initial herd on Overfunding in Equity Crowdfunding". In: *Information & Management* 59.3 (2022), p. 103269. ISSN: 0378-7206.

[62] Sarah Borchers and Lee M Dunham. "Tapping of the Crowd: The Effect of Entrepreneur Engagement on Equity Crowdfunding Success". In: *Journal of Economics and Finance* 46.2 (2022), pp. 324–346.

[63] Xun Zhu. "Proximal Language predicts Crowdfunding Success: Behavioral and Experimental Evidence". In: *Computers in Human Behavior* 131 (2022), p. 107213. ISSN: 0747-5632.

[64] Simon J. Blanchard et al. "Extraction of Visual Information to predict Crowdfunding Success". In: *Production and Operations Management* 32.12 (2023), pp. 4172–4189. DOI: `https://doi.org/10.1111/poms.14083`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/poms.14083`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/poms.14083`.

[65] Vahid Godarzi, Mohammad Mashhadizadeh, and Sayyed Mohammad Reza Davoodi. "Providing a Dynamic Investment Model for financing Knowledge-based Companies with a Data Mining Approach". In: *International Journal of Nonlinear Analysis and Applications* (2023).

[66] Sarifah Putri Raflesia et al. "Using machine learning approach towards successful crowdfunding prediction". In: *Bulletin of Electrical Engineering and Informatics* 12.4 (2023), pp. 2438–2445.

[67] Anh Nguyet Vu and Jose Christian. "UK Equity Crowdfunding Success: The Impact of Competition, Brexit and Covid-19". In: *British Journal of Management* (2023).

[68] Benjamin Le Pendeven and Armin Schwienbacher. "Equity Crowdfunding: the Influence of Perceived Innovativeness on Campaign Success". In: *British Journal of Management* 34.1 (2023), pp. 280–298.

[69] Daniel Blaseg, Douglas Cumming, and Michael Koetter. "Equity crowdfunding: High-quality or low-quality entrepreneurs?" In: *Entrepreneurship Theory and Practice* 45.3 (2021), pp. 505–530. DOI: 10.1177/1042258719899427.

[70] Osservatori Entrepreneurship Finance & Innovation. *7° Report italiano sul CrowdInvesting*. Tech. rep. Politecnico di Milano, July 2022. URL: https://www.osservatoriefi.it/efi/wp-content/uploads/2022/07/reportcrowd2022.pdf.

[71] Haibo He and Edwardo A Garcia. "Learning from imbalanced data". In: *IEEE Transactions on Knowledge and Data Engineering* 21.9 (2009), pp. 1263–1284.

[72] Gustavo EAPA Batista, Ronaldo C Prati, and Maria Carolina Monard. "A study of the behavior of several methods for balancing machine learning training data". In: *ACM SIGKDD Explorations Newsletter* 6.1 (2004), pp. 20–29.

[73] N. V. Chawla et al. "SMOTE: Synthetic Minority Over-sampling Technique". In: *Journal of Artificial Intelligence Research* 16 (June 2002), pp. 321–357. DOI: 10.1613/jair.953. URL: https://doi.org/10.1613%5C%2Fjair.953.

[74] Paula Branco, Luís Torgo, and Rita P. Ribeiro. "SMOGN: A Pre-processing Approach for Imbalanced Regression". In: *First International Workshop on Learning with Imbalanced Domains: Theory and Applications (LIDTA 2017)*. 2017, pp. 36–50.

[75] Scott M Lundberg and Su-In Lee. "A unified approach to interpreting model predictions". In: *Advances in neural information processing systems* 30 (2017).

[76] Alejandro Barredo Arrieta et al. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI". In: *Information Fusion* 58 (2020), pp. 82–115.

[77] Tianqi Chen and Carlos Guestrin. "XGBoost: A scalable tree boosting system". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016.

[78] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-term Memory". In: *Neural computation* 9 (Dec. 1997), pp. 1735–80. DOI: 10.1162/neco.1997.9.8.1735.

[79] Haohan Xie et al. *A Survey on Synthetic Data Generation with Deep Generative Models: Applications, Methods, and Challenges*. 2024. arXiv: 2401.00081 [cs.LG].

[80] Wenglei Wu, Nicholas Kunz, and Paula Branco. "ImbalancedLearningRegression-A Python Package to Tackle the Imbalanced Regression Problem". In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer. 2022, pp. 645–648.

[81] Yu Cao Jing Wu Ling Liu. "Understanding Investor Co-investment in a Syndicate on Equity Crowdfunding Platforms". In: *Industrial Management & Data Systems* 123.5 (2023), pp. 1599–1623. ISSN: 0263-5577.

[82] U.S. Securities and Exchange Commission. *Crowdfunding Offerings Archive.* https://www.sec.gov/about/crowdfunding-offerings-archive. Consultato il 19 marzo 2025. 2025.

Appendices

Appendix A

Equity Results

A.1 Equity L10 results: Success

Table A.1: *Performance of Decision Tree on equity campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.846	0.918	0.879	0.94	0.94	0.962	0.918	0.929	0.945	0.94
Precision	0.923	0.972	0.932	0.966	0.973	0.967	0.965	0.966	0.973	0.973
Recall	0.886	0.926	0.919	0.96	0.953	0.987	0.933	0.946	0.96	0.953
F1	0.904	0.948	0.926	0.963	0.963	0.977	0.949	0.956	0.966	0.963

Table A.2: *Performance of k-NN on equity campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.632	0.692	0.703	0.698	0.736	0.764	0.758	0.764	0.758	0.769
Precision	0.866	0.927	0.936	0.943	0.955	0.949	0.957	0.965	0.957	0.965
Recall	0.651	0.678	0.685	0.671	0.711	0.752	0.738	0.738	0.738	0.745
F1	0.743	0.783	0.791	0.784	0.815	0.839	0.833	0.837	0.833	0.841

Table A.3: *Performance of Logistic Regression on equity campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.846	0.879	0.89	0.896	0.885	0.89	0.912	0.896	0.907	0.907
Precision	0.969	0.97	0.971	0.971	0.951	0.957	0.959	0.958	0.958	0.958
Recall	0.839	0.879	0.893	0.899	0.906	0.906	0.933	0.913	0.926	0.926
F1	0.899	0.923	0.93	0.934	0.928	0.931	0.946	0.935	0.942	0.942

Table A.4: *Performance of Bagging on equity campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.907	0.923	0.929	0.951	0.962	0.967	0.962	0.973	0.973	0.962
Precision	0.958	0.966	0.966	0.979	0.986	0.98	0.973	0.986	0.986	0.973
Recall	0.926	0.94	0.946	0.96	0.966	0.98	0.98	0.98	0.98	0.98
F1	0.942	0.952	0.956	0.969	0.976	0.98	0.977	0.983	0.983	0.977

Table A.5: *Performance of Random Forest on equity campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.89	0.907	0.885	0.912	0.923	0.929	0.934	0.94	0.94	0.934
Precision	0.939	0.978	0.964	0.972	0.979	0.979	0.979	0.979	0.979	0.979
Recall	0.926	0.906	0.893	0.919	0.926	0.933	0.94	0.946	0.946	0.94
F1	0.932	0.941	0.927	0.945	0.952	0.955	0.959	0.962	0.962	0.959

Table A.6: *Performance of AdaBoost on equity campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.94	0.945	0.934	0.956	0.951	0.973	0.973	0.967	0.984	0.978
Precision	0.973	0.966	0.966	0.98	0.973	0.986	0.98	0.98	0.987	0.98
Recall	0.953	0.966	0.953	0.966	0.966	0.98	0.987	0.98	0.993	0.993
F1	0.963	0.966	0.959	0.973	0.97	0.983	0.983	0.98	0.99	0.987

Table A.7: *Performance of XGBoost on equity campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.912	0.923	0.923	0.951	0.956	0.973	0.967	0.967	0.973	0.978
Precision	0.94	0.959	0.959	0.979	0.98	0.98	0.98	0.98	0.98	0.98
Recall	0.953	0.946	0.946	0.96	0.966	0.987	0.98	0.98	0.987	0.993
F1	0.947	0.953	0.953	0.969	0.973	0.983	0.98	0.98	0.983	0.987

Table A.8: *Performance of Histogram Gradient Boosting on equity campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.923	0.934	0.929	0.94	0.951	0.973	0.973	0.984	0.978	0.989
Precision	0.953	0.954	0.966	0.966	0.967	0.98	0.986	0.987	0.98	0.987
Recall	0.953	0.966	0.946	0.96	0.973	0.987	0.98	0.993	0.993	1.0
F1	0.953	0.96	0.956	0.963	0.97	0.983	0.983	0.99	0.987	0.993

Table A.9: *Performance of MLP on equity campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.769	0.962	0.945	0.94	0.94	0.956	0.956	0.918	0.94	0.951
Precision	0.845	0.98	0.966	0.96	0.966	0.986	0.986	0.947	0.973	0.961
Recall	0.879	0.973	0.966	0.966	0.96	0.96	0.96	0.953	0.953	0.98
F1	0.862	0.976	0.966	0.963	0.963	0.973	0.973	0.95	0.963	0.97

Table A.10: *Performance of LSTM on equity campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.769	0.929	0.929	0.929	0.929	0.945	0.923	0.918	0.912	0.918
Precision	0.845	0.959	0.953	0.959	0.942	0.96	0.947	0.947	0.94	0.935
Recall	0.879	0.953	0.96	0.953	0.973	0.973	0.96	0.953	0.953	0.966
F1	0.862	0.956	0.957	0.956	0.957	0.967	0.953	0.95	0.947	0.95

A.2 Equity L10 results: Funding Ratio

Table A.11: *Performance of Decision Tree on equity campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.47	0.755	0.819	0.791	0.874	0.893	0.927	0.884	0.931	0.958
R^2_{adj}	0.153	0.595	0.69	0.629	0.768	0.795	0.854	0.758	0.849	0.904
MSE	0.092	0.042	0.031	0.036	0.022	0.018	0.013	0.02	0.012	0.007
MAE	0.162	0.094	0.075	0.079	0.058	0.059	0.05	0.064	0.048	0.037

Table A.12: *Performance of k-NN on equity campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.365	0.378	0.404	0.443	0.467	0.452	0.446	0.453	0.446	0.449
R^2_{adj}	-0.014	-0.028	-0.02	0.011	0.017	-0.052	-0.108	-0.142	-0.21	-0.261
MSE	0.11	0.107	0.103	0.096	0.092	0.095	0.096	0.094	0.096	0.095
MAE	0.234	0.232	0.224	0.215	0.21	0.211	0.211	0.208	0.21	0.208

Table A.13: *Performance of Ridge Regression on equity campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.39	0.416	0.427	0.43	0.426	0.415	0.409	0.396	0.397	0.389
R^2_{adj}	0.026	0.035	0.019	-0.012	-0.059	-0.123	-0.182	-0.261	-0.317	-0.398
MSE	0.105	0.101	0.099	0.098	0.099	0.101	0.102	0.104	0.104	0.106
MAE	0.255	0.252	0.246	0.245	0.245	0.245	0.246	0.248	0.248	0.249

Table A.14: *Performance of Bagging on equity campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.739	0.818	0.879	0.889	0.891	0.923	0.942	0.941	0.945	0.954
R^2_{adj}	0.583	0.699	0.793	0.803	0.799	0.852	0.884	0.877	0.88	0.895
MSE	0.045	0.031	0.021	0.019	0.019	0.013	0.01	0.01	0.01	0.008
MAE	0.117	0.092	0.071	0.063	0.061	0.055	0.048	0.047	0.045	0.042

Table A.15: *Performance of Random Forest on equity campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.716	0.807	0.861	0.884	0.882	0.923	0.939	0.933	0.939	0.95
R^2_{adj}	0.546	0.681	0.762	0.794	0.782	0.852	0.878	0.86	0.867	0.886
MSE	0.049	0.033	0.024	0.02	0.02	0.013	0.011	0.012	0.01	0.009
MAE	0.139	0.103	0.081	0.068	0.068	0.061	0.054	0.055	0.051	0.048

Table A.16: *Performance of AdaBoost on equity campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.588	0.663	0.724	0.77	0.812	0.833	0.843	0.862	0.866	0.882
R^2_{adj}	0.342	0.443	0.527	0.591	0.653	0.679	0.686	0.712	0.707	0.73
MSE	0.071	0.058	0.048	0.04	0.032	0.029	0.027	0.024	0.023	0.02
MAE	0.218	0.219	0.193	0.171	0.154	0.151	0.146	0.136	0.136	0.129

Table A.17: *Performance of XGBoost on equity campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.689	0.821	0.897	0.892	0.91	0.917	0.952	0.938	0.95	0.956
R^2_{adj}	0.503	0.704	0.824	0.808	0.834	0.841	0.904	0.871	0.891	0.899
MSE	0.054	0.031	0.018	0.019	0.016	0.014	0.008	0.011	0.009	0.008
MAE	0.123	0.087	0.067	0.062	0.054	0.052	0.042	0.046	0.039	0.04

Table A.18: *Performance of Histogram Gradient Boosting on equity campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.778	0.854	0.89	0.906	0.91	0.954	0.959	0.954	0.952	0.955
R^2_{adj}	0.645	0.759	0.812	0.833	0.834	0.912	0.918	0.904	0.895	0.897
MSE	0.038	0.025	0.019	0.016	0.016	0.008	0.007	0.008	0.008	0.008
MAE	0.122	0.092	0.08	0.072	0.069	0.058	0.054	0.055	0.054	0.052

Table A.19: *Performance of MLP on equity campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.774	0.822	0.802	0.820	0.832	0.828	0.845	0.864	0.873	0.856
R^2_{adj}	0.639	0.706	0.661	0.68	0.69	0.67	0.69	0.716	0.723	0.67
MSE	0.039	0.031	0.034	0.031	0.029	0.030	0.027	0.023	0.022	0.025
MAE	0.100	0.08	0.09	0.086	0.087	0.086	0.079	0.076	0.069	0.072

Table A.20: *Performance of LSTM on equity campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.774	0.843	0.839	0.837	0.856	0.877	0.896	0.892	0.874	0.902
R^2_{adj}	0.639	0.741	0.724	0.71	0.734	0.764	0.792	0.775	0.725	0.776
MSE	0.039	0.027	0.028	0.028	0.025	0.021	0.018	0.019	0.022	0.017
MAE	0.100	0.076	0.078	0.079	0.078	0.071	0.065	0.068	0.072	0.062

A.3 Equity L10 results: Overfunding

Table A.21: *Performance of Decision Tree on equity campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.351	0.394	0.436	0.426	0.442	0.439	0.509	0.498	0.602	0.6
R^2_{adj}	-0.037	-0.002	0.034	-0.019	-0.029	-0.077	0.018	-0.048	0.131	0.085
MSE	0.1	0.093	0.087	0.088	0.086	0.086	0.076	0.077	0.061	0.062
MAE	0.214	0.192	0.187	0.192	0.178	0.181	0.163	0.163	0.145	0.142

Table A.22: *Performance of k-NN on equity campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.221	0.245	0.282	0.294	0.302	0.335	0.358	0.354	0.348	0.347
R^2_{adj}	-0.244	-0.248	-0.229	-0.254	-0.288	-0.276	-0.284	-0.349	-0.424	-0.494
MSE	0.12	0.116	0.111	0.109	0.107	0.102	0.099	0.099	0.1	0.101
MAE	0.235	0.228	0.221	0.219	0.217	0.209	0.203	0.201	0.2	0.2

Table A.23: *Performance of Ridge Regression on equity campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.233	0.283	0.291	0.299	0.306	0.29	0.292	0.291	0.281	0.238
R^2_{adj}	-0.225	-0.185	-0.214	-0.245	-0.28	-0.363	-0.416	-0.48	-0.57	-0.744
MSE	0.118	0.11	0.109	0.108	0.107	0.109	0.109	0.109	0.111	0.117
MAE	0.268	0.256	0.251	0.247	0.244	0.245	0.24	0.24	0.238	0.242

Table A.24: *Performance of Bagging on equity campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.576	0.645	0.692	0.728	0.722	0.755	0.772	0.777	0.793	0.82
R^2_{adj}	0.323	0.413	0.473	0.517	0.487	0.53	0.544	0.534	0.548	0.588
MSE	0.065	0.055	0.047	0.042	0.043	0.038	0.035	0.034	0.032	0.028
MAE	0.178	0.15	0.135	0.126	0.123	0.116	0.109	0.107	0.102	0.089

Table A.25: *Performance of Random Forest on equity campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.536	0.575	0.626	0.652	0.637	0.658	0.687	0.684	0.707	0.729
R^2_{adj}	0.259	0.298	0.36	0.382	0.33	0.344	0.374	0.34	0.36	0.38
MSE	0.071	0.065	0.058	0.054	0.056	0.053	0.048	0.049	0.045	0.042
MAE	0.202	0.179	0.166	0.156	0.157	0.148	0.14	0.14	0.133	0.124

Table A.26: *Performance of AdaBoost on equity campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.369	0.454	0.447	0.559	0.413	0.568	0.632	0.476	0.603	0.64
R^2_{adj}	-0.008	0.098	0.053	0.217	-0.083	0.171	0.264	-0.094	0.133	0.176
MSE	0.097	0.084	0.085	0.068	0.09	0.067	0.057	0.081	0.061	0.055
MAE	0.271	0.237	0.256	0.201	0.275	0.182	0.177	0.256	0.176	0.177

Table A.27: *Performance of XGBoost on equity campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.563	0.644	0.68	0.734	0.766	0.8	0.819	0.837	0.862	0.889
R^2_{adj}	0.302	0.412	0.452	0.528	0.568	0.616	0.638	0.66	0.699	0.746
MSE	0.067	0.055	0.049	0.041	0.036	0.031	0.028	0.025	0.021	0.017
MAE	0.179	0.152	0.134	0.118	0.109	0.101	0.096	0.091	0.078	0.064

Table A.28: *Performance of Histogram Gradient Boosting on equity campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.552	0.625	0.677	0.691	0.714	0.745	0.766	0.777	0.802	0.848
R^2_{adj}	0.284	0.38	0.447	0.451	0.472	0.511	0.532	0.534	0.568	0.652
MSE	0.069	0.058	0.05	0.048	0.044	0.039	0.036	0.034	0.03	0.023
MAE	0.191	0.169	0.157	0.15	0.142	0.13	0.124	0.12	0.109	0.09

Table A.29: *Performance of MLP on equity campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.436	0.477	0.439	0.498	0.584	0.674	0.537	0.697	0.656	0.694
R^2_{adj}	0.099	0.136	0.039	0.108	0.232	0.374	0.074	0.367	0.249	0.3
MSE	0.087	0.081	0.086	0.077	0.064	0.05	0.071	0.047	0.053	0.047
MAE	0.19	0.168	0.174	0.161	0.146	0.132	0.159	0.127	0.136	0.124

Table A.30: *Performance of LSTM on equity campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.436	0.458	0.686	0.554	0.583	0.586	0.664	0.672	0.716	0.748
R^2_{adj}	0.099	0.104	0.462	0.208	0.231	0.205	0.328	0.315	0.38	0.423
MSE	0.087	0.083	0.054	0.069	0.064	0.064	0.052	0.051	0.044	0.039
MAE	0.19	0.176	0.136	0.154	0.147	0.147	0.132	0.128	0.118	0.111

A.4 Equity L14 results: Success

Table A.31: *Performance of Decision Tree on equity campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.835	0.901	0.918	0.901	0.918	0.912	0.912	0.923	0.962	0.945	0.951	0.934	0.951	0.945
Precision	0.941	0.965	0.965	0.952	0.965	0.959	0.965	0.972	0.98	0.973	0.973	0.96	0.961	0.96
Recall	0.852	0.913	0.933	0.926	0.933	0.933	0.926	0.933	0.973	0.96	0.966	0.96	0.98	0.973
F1	0.894	0.938	0.949	0.939	0.949	0.946	0.945	0.952	0.976	0.966	0.97	0.96	0.97	0.967

Table A.32: *Performance of k-NN on equity campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.654	0.692	0.703	0.736	0.753	0.764	0.78	0.786	0.797	0.808	0.758	0.769	0.808	0.83
Precision	0.891	0.919	0.928	0.947	0.948	0.965	0.966	0.966	0.967	0.967	0.957	0.973	0.983	0.992
Recall	0.658	0.685	0.691	0.718	0.738	0.738	0.758	0.765	0.779	0.792	0.738	0.738	0.779	0.799
F1	0.757	0.785	0.792	0.817	0.83	0.837	0.85	0.854	0.862	0.871	0.833	0.84	0.869	0.885

Table A.33: *Performance of Logistic Regression on equity campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.846	0.863	0.879	0.912	0.89	0.89	0.885	0.89	0.907	0.901	0.89	0.885	0.885	0.896
Precision	0.969	0.963	0.964	0.972	0.971	0.957	0.951	0.951	0.978	0.978	0.971	0.957	0.957	0.958
Recall	0.839	0.866	0.886	0.919	0.893	0.906	0.906	0.913	0.906	0.899	0.893	0.899	0.899	0.913
F1	0.899	0.912	0.923	0.945	0.93	0.931	0.928	0.932	0.941	0.937	0.93	0.927	0.927	0.935

Table A.34: *Performance of Bagging on equity campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.918	0.934	0.945	0.94	0.94	0.945	0.945	0.951	0.967	0.962	0.956	0.967	0.973	0.973
Precision	0.953	0.954	0.96	0.966	0.973	0.979	0.973	0.973	0.98	0.98	0.98	0.98	0.98	0.98
Recall	0.946	0.966	0.973	0.96	0.953	0.953	0.96	0.966	0.98	0.973	0.966	0.98	0.987	0.987
F1	0.949	0.96	0.967	0.963	0.963	0.966	0.966	0.97	0.98	0.976	0.973	0.98	0.983	0.983

Table A.35: *Performance of Random Forest on equity campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.907	0.896	0.912	0.929	0.923	0.929	0.923	0.94	0.929	0.945	0.956	0.951	0.951	0.945
Precision	0.958	0.978	0.972	0.972	0.972	0.979	0.979	0.973	0.972	0.979	0.973	0.973	0.967	0.973
Recall	0.926	0.893	0.919	0.94	0.933	0.933	0.926	0.953	0.94	0.953	0.973	0.966	0.973	0.96
F1	0.942	0.933	0.945	0.956	0.952	0.955	0.952	0.963	0.956	0.966	0.973	0.97	0.97	0.966

Table A.36: *Performance of AdaBoost on equity campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.923	0.945	0.934	0.94	0.918	0.945	0.945	0.945	0.956	0.973	0.956	0.973	0.967	0.973
Precision	0.959	0.96	0.954	0.96	0.959	0.966	0.966	0.966	0.973	0.98	0.973	0.98	0.974	0.974
Recall	0.946	0.973	0.966	0.966	0.94	0.966	0.966	0.966	0.973	0.987	0.973	0.987	0.987	0.993
F1	0.953	0.967	0.96	0.963	0.949	0.966	0.966	0.966	0.973	0.983	0.973	0.983	0.98	0.983

Table A.37: *Performance of XGBoost on equity campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.918	0.929	0.94	0.962	0.945	0.94	0.94	0.956	0.956	0.956	0.962	0.962	0.973	0.967
Precision	0.953	0.942	0.954	0.98	0.979	0.966	0.973	0.973	0.973	0.973	0.973	0.973	0.98	0.98
Recall	0.946	0.973	0.973	0.973	0.953	0.96	0.953	0.973	0.973	0.973	0.98	0.98	0.987	0.98
F1	0.949	0.957	0.963	0.976	0.966	0.963	0.963	0.973	0.973	0.973	0.977	0.977	0.983	0.98

Table A.38: *Performance of Histogram Gradient Boosting on equity campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.918	0.934	0.951	0.945	0.951	0.934	0.945	0.956	0.962	0.967	0.973	0.978	0.978	0.984
Precision	0.953	0.948	0.961	0.966	0.973	0.966	0.966	0.973	0.973	0.98	0.98	0.987	0.98	0.987
Recall	0.946	0.973	0.98	0.966	0.966	0.953	0.966	0.973	0.98	0.98	0.987	0.987	0.993	0.993
F1	0.949	0.96	0.97	0.966	0.97	0.959	0.966	0.973	0.977	0.98	0.983	0.987	0.987	0.99

Table A.39: *Performance of MLP on equity campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.775	0.951	0.951	0.978	0.951	0.967	0.929	0.934	0.956	0.945	0.967	0.951	0.951	0.934
Precision	0.842	0.973	0.961	0.987	0.967	0.98	0.953	0.954	0.967	0.966	0.98	0.961	0.961	0.954
Recall	0.893	0.966	0.98	0.987	0.973	0.98	0.96	0.966	0.98	0.966	0.98	0.98	0.98	0.966
F1	0.866	0.97	0.97	0.987	0.97	0.98	0.957	0.96	0.973	0.966	0.98	0.97	0.97	0.96

Table A.40: *Performance of LSTM on equity campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.775	0.934	0.929	0.918	0.94	0.945	0.923	0.951	0.951	0.945	0.945	0.923	0.923	0.934
Precision	0.842	0.96	0.953	0.947	0.954	0.96	0.941	0.961	0.955	0.96	0.954	0.941	0.941	0.96
Recall	0.893	0.96	0.96	0.953	0.973	0.973	0.966	0.98	0.987	0.973	0.98	0.966	0.966	0.96
F1	0.866	0.96	0.957	0.95	0.963	0.967	0.954	0.97	0.97	0.967	0.967	0.954	0.954	0.96

A.5 Equity L14 results: Funding Ratio

Table A.41: *Performance of Decision Tree on equity campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.584	0.721	0.766	0.828	0.798	0.838	0.898	0.88	0.882	0.875	0.87	0.898	0.927	0.939
R^2_{adj}	0.336	0.539	0.599	0.695	0.627	0.689	0.796	0.749	0.742	0.714	0.687	0.742	0.805	0.827
MSE	0.069	0.046	0.039	0.028	0.033	0.027	0.017	0.02	0.02	0.021	0.022	0.017	0.012	0.01
MAE	0.137	0.098	0.083	0.073	0.079	0.067	0.057	0.058	0.054	0.063	0.063	0.053	0.046	0.042

Table A.42: *Performance of k-NN on equity campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.403	0.417	0.428	0.422	0.427	0.446	0.45	0.442	0.43	0.454	0.446	0.439	0.445	0.437
R^2_{adj}	0.046	0.036	0.021	-0.027	-0.057	-0.063	-0.1	-0.165	-0.245	-0.25	-0.332	-0.42	-0.484	-0.595
MSE	0.099	0.097	0.095	0.096	0.095	0.092	0.091	0.092	0.094	0.09	0.092	0.093	0.092	0.093
MAE	0.223	0.221	0.22	0.221	0.221	0.215	0.212	0.213	0.212	0.206	0.207	0.209	0.208	0.207

Table A.43: *Performance of Ridge Regression on equity campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.372	0.383	0.373	0.388	0.382	0.352	0.335	0.344	0.319	0.324	0.309	0.298	0.309	0.317
R^2_{adj}	-0.003	-0.02	-0.074	-0.087	-0.14	-0.244	-0.33	-0.37	-0.487	-0.547	-0.661	-0.777	-0.848	-0.935
MSE	0.104	0.102	0.104	0.101	0.102	0.107	0.11	0.109	0.113	0.112	0.114	0.116	0.114	0.113
MAE	0.252	0.251	0.252	0.247	0.249	0.251	0.253	0.251	0.253	0.253	0.253	0.255	0.254	0.253

Table A.44: *Performance of Bagging on equity campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.767	0.846	0.857	0.886	0.892	0.916	0.92	0.923	0.935	0.941	0.94	0.944	0.954	0.966
R^2_{adj}	0.628	0.745	0.755	0.798	0.801	0.839	0.84	0.839	0.858	0.865	0.856	0.858	0.877	0.904
MSE	0.039	0.026	0.024	0.019	0.018	0.014	0.013	0.013	0.011	0.01	0.01	0.009	0.008	0.006
MAE	0.106	0.083	0.076	0.063	0.06	0.053	0.054	0.052	0.046	0.044	0.044	0.041	0.04	0.036

Table A.45: *Performance of Random Forest on equity campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.723	0.827	0.84	0.882	0.883	0.907	0.913	0.921	0.93	0.934	0.933	0.932	0.943	0.956
R^2_{adj}	0.558	0.714	0.726	0.79	0.784	0.822	0.826	0.835	0.847	0.849	0.839	0.828	0.848	0.875
MSE	0.046	0.029	0.026	0.02	0.019	0.015	0.014	0.013	0.011	0.011	0.011	0.011	0.009	0.007
MAE	0.131	0.094	0.086	0.071	0.069	0.063	0.058	0.055	0.051	0.049	0.049	0.05	0.046	0.041

Table A.46: *Performance of AdaBoost on equity campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.636	0.635	0.705	0.778	0.791	0.825	0.81	0.842	0.831	0.852	0.849	0.854	0.864	0.907
R^2_{adj}	0.419	0.397	0.495	0.606	0.614	0.664	0.62	0.67	0.631	0.661	0.637	0.63	0.636	0.737
MSE	0.06	0.06	0.049	0.037	0.035	0.029	0.031	0.026	0.028	0.025	0.025	0.024	0.022	0.015
MAE	0.162	0.221	0.2	0.168	0.163	0.148	0.159	0.143	0.15	0.138	0.139	0.135	0.131	0.109

Table A.47: *Performance of XGBoost on equity campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.759	0.834	0.854	0.9	0.891	0.901	0.929	0.928	0.95	0.939	0.945	0.928	0.961	0.963
R^2_{adj}	0.615	0.726	0.75	0.822	0.799	0.81	0.858	0.85	0.891	0.86	0.868	0.818	0.896	0.895
MSE	0.04	0.028	0.024	0.016	0.018	0.016	0.012	0.012	0.008	0.01	0.009	0.012	0.007	0.006
MAE	0.104	0.078	0.075	0.055	0.055	0.052	0.044	0.042	0.036	0.042	0.037	0.041	0.033	0.031

Table A.48: *Performance of Histogram Gradient Boosting on equity campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.791	0.86	0.866	0.894	0.905	0.911	0.912	0.934	0.942	0.938	0.939	0.941	0.949	0.964
R^2_{adj}	0.666	0.769	0.771	0.812	0.825	0.829	0.824	0.862	0.873	0.858	0.853	0.851	0.864	0.898
MSE	0.035	0.023	0.022	0.017	0.016	0.015	0.014	0.011	0.01	0.01	0.01	0.01	0.008	0.006
MAE	0.117	0.09	0.082	0.069	0.066	0.06	0.057	0.054	0.05	0.052	0.049	0.049	0.047	0.043

Table A.49: *Performance of MLP on equity campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.736	0.781	0.784	0.809	0.786	0.851	0.857	0.849	0.893	0.873	0.865	0.855	0.87	0.864
R^2_{adj}	0.578	0.638	0.63	0.661	0.605	0.714	0.714	0.685	0.766	0.709	0.675	0.633	0.652	0.615
MSE	0.044	0.036	0.036	0.032	0.035	0.025	0.024	0.025	0.018	0.021	0.022	0.024	0.021	0.023
MAE	0.103	0.097	0.087	0.091	0.091	0.076	0.075	0.074	0.067	0.067	0.077	0.076	0.068	0.081

Table A.50: *Performance of LSTM on equity campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.736	0.716	0.823	0.839	0.846	0.877	0.88	0.848	0.869	0.886	0.871	0.89	0.912	0.900
R^2_{adj}	0.578	0.531	0.697	0.714	0.716	0.764	0.76	0.683	0.714	0.739	0.69	0.721	0.765	0.717
MSE	0.044	0.044	0.029	0.027	0.025	0.02	0.02	0.025	0.022	0.019	0.021	0.018	0.015	0.017
MAE	0.103	0.118	0.084	0.082	0.08	0.071	0.068	0.074	0.071	0.068	0.07	0.066	0.062	0.065

A.6 Equity L14 results: Overfunding

Table A.51: *Performance of Decision Tree on equity campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.325	0.309	0.327	0.422	0.503	0.483	0.606	0.486	0.534	0.488	0.513	0.527	0.471	0.559
R^2_{adj}	-0.078	-0.142	-0.152	-0.027	0.083	0.008	0.212	-0.073	-0.018	-0.172	-0.171	-0.198	-0.415	-0.249
MSE	0.104	0.106	0.104	0.089	0.076	0.08	0.061	0.079	0.072	0.079	0.075	0.073	0.081	0.068
MAE	0.214	0.212	0.206	0.186	0.173	0.169	0.146	0.173	0.162	0.171	0.163	0.165	0.172	0.152

Table A.52: *Performance of k-NN on equity campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.275	0.285	0.292	0.303	0.285	0.308	0.324	0.323	0.334	0.349	0.343	0.345	0.363	0.363
R^2_{adj}	-0.158	-0.182	-0.212	-0.238	-0.319	-0.328	-0.352	-0.413	-0.454	-0.49	-0.579	-0.658	-0.703	-0.805
MSE	0.112	0.11	0.109	0.107	0.11	0.107	0.104	0.104	0.103	0.1	0.101	0.101	0.098	0.098
MAE	0.227	0.223	0.217	0.217	0.216	0.211	0.208	0.205	0.201	0.197	0.197	0.198	0.194	0.195

Table A.53: *Performance of Ridge Regression on equity campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.232	0.27	0.265	0.282	0.274	0.268	0.26	0.256	0.243	0.245	0.238	0.245	0.246	0.23
R^2_{adj}	-0.227	-0.207	-0.258	-0.275	-0.339	-0.405	-0.48	-0.553	-0.653	-0.728	-0.832	-0.912	-1.016	-1.182
MSE	0.118	0.112	0.113	0.11	0.112	0.113	0.114	0.114	0.116	0.116	0.117	0.116	0.116	0.119
MAE	0.267	0.257	0.254	0.251	0.252	0.248	0.245	0.243	0.242	0.241	0.239	0.237	0.236	0.24

Table A.54: *Performance of Bagging on equity campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.578	0.669	0.641	0.668	0.699	0.715	0.728	0.754	0.759	0.775	0.782	0.801	0.81	0.825
R^2_{adj}	0.326	0.453	0.385	0.41	0.445	0.453	0.456	0.486	0.474	0.485	0.476	0.496	0.492	0.504
MSE	0.065	0.051	0.055	0.051	0.046	0.044	0.042	0.038	0.037	0.035	0.034	0.031	0.029	0.027
MAE	0.177	0.148	0.148	0.142	0.134	0.13	0.124	0.119	0.115	0.11	0.105	0.1	0.094	0.089

Table A.55: *Performance of Random Forest on equity campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.538	0.622	0.594	0.616	0.637	0.652	0.647	0.663	0.667	0.679	0.694	0.695	0.706	0.726
R^2_{adj}	0.262	0.375	0.305	0.318	0.33	0.332	0.294	0.296	0.273	0.265	0.264	0.228	0.214	0.224
MSE	0.071	0.058	0.063	0.059	0.056	0.054	0.054	0.052	0.051	0.049	0.047	0.047	0.045	0.042
MAE	0.202	0.174	0.174	0.166	0.16	0.155	0.155	0.15	0.146	0.143	0.137	0.137	0.132	0.126

Table A.56: *Performance of AdaBoost on equity campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.348	0.443	0.453	0.51	0.486	0.517	0.497	0.535	0.486	0.47	0.54	0.587	0.585	0.609
R^2_{adj}	-0.041	0.079	0.063	0.13	0.052	0.073	-0.006	0.029	-0.122	-0.213	-0.106	-0.046	-0.11	-0.108
MSE	0.1	0.086	0.084	0.075	0.079	0.074	0.077	0.072	0.079	0.082	0.071	0.064	0.064	0.06
MAE	0.28	0.243	0.244	0.227	0.244	0.231	0.243	0.219	0.251	0.258	0.181	0.173	0.17	0.156

Table A.57: *Performance of XGBoost on equity campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.593	0.688	0.687	0.717	0.723	0.781	0.766	0.773	0.777	0.813	0.833	0.835	0.855	0.906
R^2_{adj}	0.35	0.484	0.464	0.497	0.489	0.58	0.532	0.526	0.513	0.572	0.599	0.582	0.612	0.734
MSE	0.063	0.048	0.048	0.044	0.043	0.034	0.036	0.035	0.034	0.029	0.026	0.025	0.022	0.015
MAE	0.173	0.142	0.134	0.127	0.127	0.107	0.112	0.11	0.106	0.095	0.087	0.087	0.075	0.064

Table A.58: *Performance of Histogram Gradient Boosting on equity campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.557	0.666	0.656	0.673	0.697	0.701	0.701	0.723	0.746	0.771	0.779	0.793	0.809	0.841
R^2_{adj}	0.292	0.448	0.411	0.419	0.441	0.426	0.402	0.422	0.445	0.476	0.469	0.476	0.489	0.549
MSE	0.068	0.051	0.053	0.05	0.047	0.046	0.046	0.043	0.039	0.035	0.034	0.032	0.029	0.024
MAE	0.189	0.163	0.161	0.158	0.151	0.147	0.144	0.139	0.132	0.124	0.119	0.114	0.103	0.094

Table A.59: *Performance of MLP on equity campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.38	0.462	0.438	0.408	0.541	0.561	0.611	0.623	0.665	0.7	0.731	0.717	0.751	0.751
R^2_{adj}	0.01	0.111	0.038	-0.051	0.153	0.157	0.222	0.213	0.269	0.313	0.353	0.283	0.334	0.294
MSE	0.095	0.083	0.087	0.091	0.071	0.068	0.06	0.058	0.052	0.046	0.041	0.043	0.038	0.038
MAE	0.191	0.174	0.174	0.182	0.163	0.154	0.146	0.143	0.135	0.127	0.118	0.123	0.121	0.114

Table A.60: *Performance of LSTM on equity campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.38	0.484	0.481	0.579	0.539	0.548	0.577	0.634	0.664	0.672	0.705	0.692	0.706	0.74
R^2_{adj}	0.01	0.147	0.111	0.252	0.149	0.132	0.154	0.236	0.266	0.249	0.291	0.22	0.214	0.263
MSE	0.095	0.079	0.08	0.065	0.071	0.07	0.065	0.056	0.052	0.05	0.045	0.047	0.045	0.04
MAE	0.191	0.173	0.168	0.151	0.156	0.156	0.154	0.14	0.134	0.129	0.124	0.132	0.118	0.111

Appendix B

Lending Results

B.1 Lending L10 results: Success

Table B.1: *Performance of Decision Tree on lending campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.945	0.908	0.926	0.954	0.945	0.959	0.954	0.968	0.959	0.949
Precision	0.976	0.97	0.98	0.976	0.976	0.995	0.985	0.986	0.99	0.981
Recall	0.967	0.933	0.943	0.976	0.967	0.962	0.967	0.981	0.967	0.967
F1	0.971	0.951	0.961	0.976	0.971	0.978	0.976	0.983	0.978	0.974

Table B.2: *Performance of k-NN on lending campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.963	0.945	0.935	0.94	0.931	0.926	0.922	0.94	0.954	0.935
Precision	0.976	0.981	0.976	0.98	0.98	0.975	0.98	0.98	0.981	0.98
Recall	0.986	0.962	0.957	0.957	0.948	0.948	0.938	0.957	0.971	0.952
F1	0.981	0.971	0.966	0.969	0.964	0.961	0.959	0.969	0.976	0.966

Table B.3: *Performance of Logistic Regression on lending campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.945	0.949	0.935	0.954	0.949	0.945	0.935	0.94	0.945	0.926
Precision	0.981	0.981	0.98	0.981	0.981	0.981	0.98	0.98	0.981	0.98
Recall	0.962	0.967	0.952	0.971	0.967	0.962	0.952	0.957	0.962	0.943
F1	0.971	0.974	0.966	0.976	0.974	0.971	0.966	0.969	0.971	0.961

Table B.4: *Performance of Bagging on lending campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.972	0.982	0.972	0.972	0.977	0.986	0.982	0.982	0.972	0.986
Precision	0.977	0.986	0.986	0.981	0.986	0.991	0.986	0.986	0.99	0.986
Recall	0.995	0.995	0.986	0.99	0.99	0.995	0.995	0.995	0.981	1.0
F1	0.986	0.991	0.986	0.986	0.988	0.993	0.991	0.991	0.986	0.993

Table B.5: *Performance of Random Forest on lending campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.968	0.972	0.977	0.977	0.977	0.995	0.982	0.991	0.977	0.986
Precision	0.977	0.977	0.981	0.981	0.981	0.995	0.986	0.991	0.986	0.991
Recall	0.99	0.995	0.995	0.995	0.995	1.0	0.995	1.0	0.99	0.995
F1	0.983	0.986	0.988	0.988	0.988	0.998	0.991	0.995	0.988	0.993

Table B.6: *Performance of AdaBoost on lending campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.963	0.963	0.968	0.968	0.972	0.972	0.968	0.977	0.986	0.982
Precision	0.976	0.976	0.977	0.981	0.986	0.986	0.986	0.99	0.986	0.986
Recall	0.986	0.986	0.99	0.986	0.986	0.986	0.981	0.986	1.0	0.995
F1	0.981	0.981	0.983	0.983	0.986	0.986	0.983	0.988	0.993	0.991

Table B.7: *Performance of XGBoost on lending campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.972	0.977	0.972	0.972	0.977	0.986	0.977	0.977	0.982	0.991
Precision	0.977	0.986	0.981	0.986	0.986	0.995	0.986	0.986	0.99	0.991
Recall	0.995	0.99	0.99	0.986	0.99	0.99	0.99	0.99	0.99	1.0
F1	0.986	0.988	0.986	0.986	0.988	0.993	0.988	0.988	0.99	0.995

Table B.8: *Performance of Histogram Gradient Boosting on lending campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.972	0.963	0.977	0.977	0.972	0.986	0.982	0.977	0.977	0.991
Precision	0.977	0.972	0.981	0.986	0.986	0.986	0.986	0.986	0.986	0.991
Recall	0.995	0.99	0.995	0.99	0.986	1.0	0.995	0.99	0.99	1.0
F1	0.986	0.981	0.988	0.988	0.986	0.993	0.991	0.988	0.988	0.995

Table B.9: *Performance of MLP on lending campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.968	0.977	0.982	0.968	0.972	0.977	0.968	0.982	0.977	0.972
Precision	0.972	0.981	0.981	0.977	0.977	0.981	0.972	0.981	0.977	0.981
Recall	0.995	0.995	1.0	0.99	0.995	0.995	0.995	1.0	1.0	0.99
F1	0.984	0.988	0.991	0.983	0.986	0.988	0.984	0.991	0.988	0.986

Table B.10: *Performance of LSTM on lending campaigns over a 10-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Accuracy	0.968	0.972	0.972	0.977	0.968	0.972	0.977	0.972	0.977	0.977
Precision	0.972	0.977	0.972	0.977	0.972	0.972	0.977	0.972	0.977	0.977
Recall	0.995	0.995	1.0	1.0	0.995	1.0	1.0	1.0	1.0	1.0
F1	0.984	0.986	0.986	0.988	0.984	0.986	0.988	0.986	0.988	0.988

B.2 Lending L10 results: Funding Ratio

Table B.11: *Performance of Decision Tree on lending campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.846	0.89	0.852	0.87	0.848	0.858	0.867	0.894	0.886	0.922
R^2_{adj}	0.779	0.839	0.778	0.8	0.76	0.77	0.779	0.819	0.8	0.86
MSE	0.022	0.016	0.021	0.018	0.021	0.02	0.019	0.015	0.016	0.011
MAE	0.057	0.047	0.052	0.047	0.049	0.048	0.049	0.048	0.046	0.039

Table B.12: *Performance of k-NN on lending campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.937	0.932	0.929	0.93	0.929	0.927	0.926	0.929	0.928	0.932
R^2_{adj}	0.91	0.9	0.893	0.892	0.888	0.882	0.877	0.879	0.874	0.878
MSE	0.009	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
MAE	0.036	0.038	0.038	0.038	0.04	0.04	0.039	0.039	0.039	0.038

Table B.13: *Performance of Ridge Regression on lending campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.857	0.861	0.864	0.865	0.864	0.867	0.864	0.864	0.867	0.863
R^2_{adj}	0.795	0.796	0.796	0.792	0.786	0.785	0.774	0.768	0.767	0.753
MSE	0.02	0.02	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019
MAE	0.096	0.094	0.093	0.093	0.093	0.092	0.094	0.095	0.094	0.096

Table B.14: *Performance of Bagging on lending campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.913	0.925	0.932	0.928	0.93	0.928	0.935	0.938	0.942	0.945
R^2_{adj}	0.875	0.89	0.898	0.889	0.89	0.884	0.892	0.894	0.898	0.901
MSE	0.012	0.011	0.01	0.01	0.01	0.01	0.009	0.009	0.008	0.008
MAE	0.04	0.038	0.038	0.039	0.04	0.039	0.038	0.036	0.033	0.031

Table B.15: *Performance of Random Forest on lending campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.883	0.915	0.915	0.912	0.909	0.912	0.912	0.917	0.924	0.926
R^2_{adj}	0.832	0.875	0.872	0.865	0.856	0.858	0.854	0.859	0.867	0.867
MSE	0.016	0.012	0.012	0.012	0.013	0.012	0.012	0.012	0.011	0.01
MAE	0.054	0.046	0.047	0.048	0.049	0.048	0.049	0.045	0.042	0.04

Table B.16: *Performance of AdaBoost on lending campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.839	0.84	0.84	0.819	0.811	0.818	0.835	0.821	0.841	0.851
R^2_{adj}	0.769	0.765	0.76	0.721	0.702	0.706	0.726	0.695	0.722	0.732
MSE	0.023	0.023	0.023	0.026	0.027	0.026	0.023	0.025	0.022	0.021
MAE	0.088	0.115	0.122	0.141	0.144	0.139	0.129	0.136	0.132	0.122

Table B.17: *Performance of XGBoost on lending campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.92	0.93	0.939	0.938	0.931	0.934	0.931	0.937	0.951	0.947
R^2_{adj}	0.885	0.897	0.908	0.905	0.891	0.893	0.885	0.893	0.914	0.905
MSE	0.011	0.01	0.009	0.009	0.01	0.009	0.01	0.009	0.007	0.007
MAE	0.043	0.039	0.034	0.036	0.036	0.038	0.039	0.038	0.031	0.03

Table B.18: *Performance of Histogram Gradient Boosting on lending campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.885	0.91	0.918	0.917	0.915	0.912	0.911	0.912	0.922	0.931
R^2_{adj}	0.835	0.868	0.877	0.872	0.866	0.858	0.852	0.85	0.863	0.876
MSE	0.016	0.013	0.012	0.012	0.012	0.012	0.012	0.012	0.011	0.01
MAE	0.059	0.054	0.05	0.051	0.051	0.051	0.051	0.048	0.046	0.04

Table B.19: *Performance of MLP on lending campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.942	0.947	0.943	0.953	0.934	0.942	0.924	0.922	0.931	0.940
R^2_{adj}	0.917	0.867	0.914	0.928	0.896	0.906	0.874	0.922	0.879	0.892
MSE	0.008	0.007	0.008	0.007	0.009	0.008	0.011	0.011	0.01	0.008
MAE	0.033	0.036	0.037	0.034	0.039	0.037	0.049	0.042	0.038	0.038

Table B.20: *Performance of LSTM on lending campaigns over a 10-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9,	t_{10}
R^2	0.942	0.945	0.951	0.947	0.952	0.948	0.944	0.955	0.96	0.95
R^2_{adj}	0.917	0.919	0.926	0.918	0.924	0.916	0.907	0.923	0.93	0.91
MSE	0.008	0.008	0.007	0.007	0.007	0.007	0.008	0.006	0.006	0.007
MAE	0.033	0.036	0.037	0.039	0.035	0.04	0.04	0.037	0.036	0.037

B.3 Lending L10 results: Overfunding

Table B.21: *Performance of Decision Tree on lending campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.717	0.702	0.702	0.689	0.664	0.718	0.718	0.718	0.718	0.718
R^2_{adj}	0.594	0.563	0.552	0.521	0.47	0.544	0.532	0.519	0.506	0.492
MSE	0.062	0.065	0.065	0.068	0.074	0.062	0.062	0.062	0.062	0.062
MAE	0.114	0.117	0.117	0.121	0.126	0.115	0.115	0.115	0.115	0.115

Table B.22: *Performance of k-NN on lending campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.819	0.846	0.847	0.855	0.851	0.85	0.833	0.836	0.833	0.84
R^2_{adj}	0.74	0.774	0.77	0.777	0.765	0.757	0.723	0.72	0.708	0.712
MSE	0.04	0.034	0.033	0.032	0.033	0.033	0.037	0.036	0.037	0.035
MAE	0.08	0.077	0.077	0.074	0.077	0.076	0.081	0.079	0.079	0.077

Table B.23: *Performance of Ridge Regression on lending campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.854	0.853	0.855	0.858	0.86	0.853	0.856	0.858	0.855	0.856
R^2_{adj}	0.791	0.784	0.782	0.781	0.779	0.762	0.761	0.758	0.746	0.741
MSE	0.032	0.032	0.032	0.031	0.031	0.032	0.032	0.031	0.032	0.032
MAE	0.093	0.093	0.093	0.093	0.094	0.096	0.095	0.095	0.096	0.096

Table B.24: *Performance of Bagging on lending campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.889	0.89	0.89	0.884	0.886	0.89	0.887	0.89	0.893	0.895
R^2_{adj}	0.841	0.839	0.835	0.821	0.82	0.822	0.812	0.813	0.813	0.811
MSE	0.024	0.024	0.024	0.025	0.025	0.024	0.025	0.024	0.024	0.023
MAE	0.062	0.063	0.062	0.066	0.065	0.064	0.063	0.064	0.061	0.061

Table B.25: *Performance of Random Forest on lending campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.806	0.804	0.803	0.808	0.809	0.804	0.809	0.81	0.818	0.817
R^2_{adj}	0.722	0.712	0.704	0.705	0.699	0.683	0.683	0.676	0.681	0.67
MSE	0.043	0.043	0.043	0.042	0.042	0.043	0.042	0.042	0.04	0.04
MAE	0.122	0.121	0.125	0.124	0.121	0.123	0.122	0.12	0.12	0.121

Table B.26: *Performance of AdaBoost on lending campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.679	0.683	0.691	0.67	0.717	0.683	0.673	0.685	0.69	0.649
R^2_{adj}	0.54	0.535	0.536	0.492	0.554	0.487	0.457	0.463	0.457	0.368
MSE	0.071	0.07	0.068	0.072	0.062	0.07	0.072	0.069	0.068	0.077
MAE	0.198	0.188	0.185	0.186	0.181	0.194	0.195	0.185	0.188	0.228

Table B.27: *Performance of XGBoost on lending campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.877	0.859	0.864	0.877	0.877	0.878	0.873	0.896	0.884	0.906
R^2_{adj}	0.824	0.793	0.796	0.811	0.806	0.803	0.789	0.823	0.797	0.831
MSE	0.027	0.031	0.03	0.027	0.027	0.027	0.028	0.023	0.026	0.021
MAE	0.067	0.076	0.07	0.072	0.077	0.072	0.075	0.065	0.071	0.066

Table B.28: *Performance of Histogram Gradient Boosting on lending campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.848	0.848	0.851	0.857	0.859	0.858	0.854	0.86	0.863	0.867
R^2_{adj}	0.782	0.777	0.776	0.78	0.778	0.77	0.758	0.761	0.76	0.76
MSE	0.033	0.033	0.033	0.031	0.031	0.031	0.032	0.031	0.03	0.029
MAE	0.102	0.102	0.102	0.1	0.101	0.101	0.102	0.101	0.099	0.101

Table B.29: *Performance of MLP on lending campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.848	0.837	0.877	0.854	0.854	0.893	0.858	0.848	0.875	0.868
R^2_{adj}	0.782	0.761	0.815	0.775	0.77	0.827	0.764	0.741	0.781	0.762
MSE	0.033	0.036	0.027	0.032	0.032	0.023	0.031	0.033	0.027	0.029
MAE	0.053	0.058	0.046	0.049	0.05	0.042	0.05	0.054	0.048	0.048

Table B.30: *Performance of LSTM on lending campaigns over a 10-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
R^2	0.848	0.872	0.833	0.86	0.88	0.845	0.896	0.873	0.863	0.861
R^2_{adj}	0.782	0.812	0.749	0.785	0.811	0.749	0.827	0.784	0.76	0.75
MSE	0.033	0.028	0.037	0.031	0.026	0.034	0.023	0.028	0.03	0.031
MAE	0.053	0.048	0.056	0.052	0.048	0.055	0.041	0.046	0.049	0.049

B.4 Lending L14 results: Success

Table B.31: *Performance of Decision Tree on lending campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.945	0.903	0.963	0.972	0.959	0.977	0.963	0.954	0.954	0.959	0.949	0.931	0.959	0.959
Precision	0.976	0.97	0.986	0.99	0.99	0.986	0.99	0.99	0.985	0.99	0.985	0.985	0.981	0.99
Recall	0.967	0.929	0.976	0.981	0.967	0.99	0.971	0.962	0.967	0.967	0.962	0.943	0.976	0.967
F1	0.971	0.949	0.981	0.986	0.978	0.988	0.981	0.976	0.976	0.978	0.973	0.964	0.979	0.978

Table B.32: *Performance of k-NN on lending campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.963	0.945	0.931	0.931	0.926	0.931	0.935	0.94	0.935	0.949	0.949	0.954	0.949	0.959
Precision	0.976	0.976	0.976	0.976	0.975	0.976	0.98	0.98	0.98	0.981	0.976	0.981	0.981	0.986
Recall	0.986	0.967	0.952	0.952	0.948	0.952	0.952	0.957	0.952	0.967	0.971	0.971	0.967	0.971
F1	0.981	0.971	0.964	0.964	0.961	0.964	0.966	0.969	0.966	0.974	0.974	0.976	0.974	0.978

Table B.33: *Performance of Logistic Regression Model on lending campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.945	0.945	0.945	0.931	0.931	0.935	0.926	0.926	0.926	0.922	0.926	0.94	0.912	0.926
Precision	0.981	0.981	0.981	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
Recall	0.962	0.962	0.962	0.948	0.948	0.952	0.943	0.943	0.943	0.938	0.943	0.957	0.929	0.943
F1	0.971	0.971	0.971	0.964	0.964	0.966	0.961	0.961	0.961	0.959	0.961	0.969	0.954	0.961

Table B.34: *Performance of Bagging on lending campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.972	0.972	0.986	0.986	0.982	0.986	0.986	0.986	0.968	0.977	0.982	0.986	0.968	0.982
Precision	0.977	0.981	0.991	0.986	0.986	0.991	0.991	0.991	0.986	0.99	0.99	0.991	0.99	0.99
Recall	0.995	0.99	0.995	1.0	0.995	0.995	0.995	0.995	0.981	0.986	0.99	0.995	0.976	0.99
F1	0.986	0.986	0.993	0.993	0.991	0.993	0.993	0.993	0.983	0.988	0.99	0.993	0.983	0.99

Table B.35: *Performance of Random Forest on lending campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.968	0.972	0.977	0.982	0.977	0.986	0.982	0.977	0.982	0.977	0.986	0.991	0.982	0.977
Precision	0.977	0.977	0.981	0.981	0.981	0.991	0.986	0.99	0.981	0.986	0.986	0.991	0.986	0.986
Recall	0.99	0.995	0.995	1.0	0.995	0.995	0.995	0.986	1.0	0.99	1.0	1.0	0.995	0.99
F1	0.983	0.986	0.988	0.991	0.988	0.993	0.991	0.988	0.991	0.988	0.993	0.995	0.991	0.988

Table B.36: *Performance of AdaBoost on lending campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.963	0.977	0.968	0.963	0.972	0.972	0.977	0.982	0.977	0.982	0.982	0.977	0.977	0.982
Precision	0.976	0.981	0.986	0.981	0.986	0.986	0.99	0.986	0.981	0.99	0.986	0.986	0.986	0.99
Recall	0.986	0.995	0.981	0.981	0.986	0.986	0.986	0.995	0.995	0.99	0.995	0.99	0.99	0.99
F1	0.981	0.988	0.983	0.981	0.986	0.986	0.988	0.991	0.988	0.99	0.991	0.988	0.988	0.99

Table B.37: *Performance of XGBoost on lending campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.977	0.977	0.977	0.982	0.968	0.977	0.977	0.972	0.963	0.968	0.972	0.977	0.982	0.991
Precision	0.977	0.981	0.986	0.981	0.981	0.986	0.986	0.986	0.981	0.986	0.986	0.986	0.99	0.995
Recall	1.0	0.995	0.99	1.0	0.986	0.99	0.99	0.986	0.981	0.981	0.986	0.99	0.99	0.995
F1	0.988	0.988	0.988	0.991	0.983	0.988	0.988	0.986	0.981	0.983	0.986	0.988	0.99	0.995

Table B.38: *Performance of Histogram Gradient Boosting on lending campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.972	0.982	0.986	0.982	0.986	0.986	0.991	0.986	0.991	0.982	0.986	0.986	0.991	0.995
Precision	0.977	0.981	0.986	0.981	0.986	0.986	0.991	0.986	0.991	0.99	0.991	0.986	0.991	0.995
Recall	0.995	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.99	0.995	1.0	1.0	1.0
F1	0.986	0.991	0.993	0.991	0.993	0.993	0.995	0.993	0.995	0.99	0.993	0.993	0.995	0.998

Table B.39: *Performance of MLP on lending campaigns over a 14-day time series for predicting Success.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.968	0.972	0.972	0.968	0.977	0.972	0.972	0.977	0.972	0.959	0.968	0.963	0.968	0.972
Precision	0.977	0.977	0.981	0.977	0.981	0.977	0.977	0.977	0.977	0.972	0.977	0.976	0.972	0.977
Recall	0.99	0.995	0.99	0.99	0.995	0.995	0.995	1.0	0.995	0.986	0.99	0.986	0.995	0.995
F1	0.983	0.986	0.986	0.983	0.988	0.986	0.986	0.988	0.986	0.979	0.983	0.981	0.984	0.986

Table B.40: *Performance of LSTM on lending campaigns over a 14-day time series for predicting Success. L14*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
Accuracy	0.968	0.972	0.968	0.972	0.972	0.977	0.972	0.977	0.963	0.977	0.963	0.972	0.968	0.972
Precision	0.977	0.977	0.977	0.977	0.977	0.977	0.977	0.977	0.976	0.977	0.976	0.977	0.977	0.977
Recall	0.99	0.995	0.99	0.995	0.995	1.0	0.995	1.0	0.986	1.0	0.986	0.995	0.99	0.995
F1	0.983	0.986	0.983	0.986	0.986	0.988	0.986	0.988	0.981	0.988	0.981	0.986	0.983	0.986

B.5 Lending L14 results: Funding Ratio

Table B.41: *Performance of Decision Tree on lending campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.88	0.907	0.909	0.906	0.897	0.891	0.899	0.86	0.9	0.907	0.863	0.913	0.86	0.945
R^2_{adj}	0.828	0.7614	0.746	0.855	0.838	0.824	0.832	0.724	0.825	0.832	0.863	0.834	0.86	0.888
MSE	0.016	0.013	0.012	0.013	0.014	0.015	0.014	0.019	0.014	0.013	0.019	0.012	0.019	0.007
MAE	0.047	0.042	0.042	0.041	0.041	0.045	0.043	0.053	0.046	0.044	0.052	0.04	0.046	0.032

Table B.42: *Performance of k-NN on lending campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.935	0.935	0.924	0.925	0.922	0.923	0.919	0.918	0.917	0.909	0.922	0.924	0.926	0.925
R^2_{adj}	0.907	0.905	0.886	0.885	0.877	0.875	0.866	0.86	0.855	0.836	0.855	0.855	0.854	0.848
MSE	0.009	0.009	0.01	0.01	0.011	0.01	0.011	0.011	0.011	0.012	0.011	0.01	0.01	0.01
MAE	0.036	0.036	0.041	0.041	0.042	0.042	0.043	0.043	0.043	0.045	0.042	0.042	0.042	0.042

Table B.43: *Performance of Ridge Regression on lending campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.845	0.848	0.852	0.854	0.475	0.622	0.587	0.567	0.582	0.612	0.62	0.692	0.717	0.755
R^2_{adj}	0.778	0.777	0.778	0.775	0.172	0.389	0.315	0.262	0.268	0.301	0.296	0.412	0.443	0.502
MSE	0.021	0.021	0.02	0.02	0.072	0.052	0.056	0.059	0.057	0.053	0.052	0.042	0.039	0.033
MAE	0.096	0.095	0.095	0.094	0.111	0.107	0.109	0.108	0.109	0.109	0.111	0.108	0.105	0.103

Table B.44: *Performance of Bagging on lending campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.931	0.931	0.938	0.926	0.937	0.944	0.941	0.945	0.952	0.947	0.947	0.951	0.957	0.957
R^2_{adj}	0.901	0.899	0.907	0.886	0.901	0.909	0.902	0.906	0.916	0.905	0.902	0.906	0.915	0.913
MSE	0.009	0.009	0.008	0.01	0.009	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.006	0.006
MAE	0.036	0.037	0.037	0.037	0.035	0.035	0.035	0.034	0.031	0.032	0.032	0.03	0.028	0.027

Table B.45: *Performance of Random Forest on lending campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.904	0.922	0.923	0.915	0.923	0.923	0.927	0.929	0.934	0.936	0.933	0.938	0.943	0.947
R^2_{adj}	0.862	0.886	0.884	0.869	0.879	0.875	0.879	0.879	0.884	0.885	0.876	0.882	0.888	0.892
MSE	0.013	0.011	0.01	0.012	0.01	0.01	0.01	0.01	0.009	0.009	0.009	0.008	0.008	0.007
MAE	0.047	0.046	0.046	0.047	0.045	0.046	0.044	0.044	0.043	0.041	0.041	0.04	0.037	0.035

Table B.46: *Performance of AdaBoost on lending campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.805	0.824	0.819	0.826	0.821	0.823	0.821	0.832	0.854	0.845	0.844	0.852	0.828	0.817
R^2_{adj}	0.72	0.742	0.728	0.732	0.718	0.714	0.703	0.714	0.744	0.721	0.711	0.717	0.661	0.628
MSE	0.027	0.024	0.025	0.024	0.024	0.024	0.024	0.023	0.02	0.021	0.021	0.02	0.023	0.025
MAE	0.116	0.121	0.134	0.132	0.137	0.137	0.136	0.13	0.118	0.125	0.124	0.12	0.134	0.139

Table B.47: *Performance of XGBoost on lending campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.929	0.943	0.931	0.932	0.948	0.942	0.945	0.95	0.949	0.946	0.938	0.939	0.957	0.959
R^2_{adj}	0.898	0.916	0.896	0.895	0.918	0.906	0.909	0.915	0.911	0.903	0.885	0.884	0.915	0.917
MSE	0.01	0.008	0.009	0.009	0.007	0.008	0.007	0.007	0.007	0.007	0.008	0.008	0.006	0.006
MAE	0.038	0.035	0.034	0.036	0.03	0.03	0.031	0.028	0.031	0.03	0.034	0.034	0.029	0.027

Table B.48: *Performance of Histogram Gradient Boosting on lending campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.903	0.913	0.928	0.92	0.924	0.924	0.926	0.93	0.937	0.935	0.935	0.937	0.947	0.951
R^2_{adj}	0.861	0.872	0.892	0.877	0.88	0.877	0.877	0.881	0.89	0.883	0.88	0.88	0.896	0.9
MSE	0.013	0.012	0.01	0.011	0.01	0.01	0.01	0.01	0.009	0.009	0.009	0.009	0.007	0.007
MAE	0.056	0.055	0.048	0.05	0.048	0.048	0.047	0.045	0.042	0.042	0.042	0.041	0.037	0.034

Table B.49: *Performance of MLP on lending campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.937	0.954	0.952	0.963	0.962	0.954	0.955	0.934	0.96	0.96	0.957	0.958	0.962	0.954
R^2_{adj}	0.91	0.932	0.928	0.943	0.94	0.926	0.925	0.888	0.93	0.928	0.92	0.92	0.925	0.907
MSE	0.009	0.006	0.007	0.005	0.005	0.006	0.006	0.009	0.005	0.005	0.006	0.006	0.005	0.006
MAE	0.035	0.034	0.036	0.028	0.028	0.033	0.033	0.04	0.032	0.027	0.032	0.032	0.033	0.035

Table B.50: *Performance of LSTM on lending campaigns over a 14-day time series for predicting Funding Ratio.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.937	0.959	0.956	0.957	0.955	0.956	0.957	0.955	0.958	0.959	0.958	0.96	0.96	0.961
R^2_{adj}	0.91	0.94	0.934	0.934	0.929	0.929	0.929	0.923	0.926	0.926	0.922	0.924	0.921	0.921
MSE	0.009	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.005	0.005	0.005
MAE	0.035	0.033	0.033	0.035	0.034	0.032	0.035	0.034	0.035	0.036	0.036	0.034	0.032	0.031

B.6 Lending L14 results: Overfunding

Table B.51: *Performance of Decision Tree on lending campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.717	0.699	0.707	0.713	0.661	0.675	0.689	0.699	0.699	0.713	0.699	0.699	0.688	0.673
R^2_{adj}	0.594	0.558	0.56	0.558	0.465	0.474	0.484	0.487	0.473	0.483	0.442	0.425	0.386	0.336
MSE	0.062	0.066	0.064	0.063	0.074	0.071	0.068	0.066	0.066	0.063	0.066	0.066	0.069	0.072
MAE	0.114	0.119	0.118	0.116	0.127	0.125	0.121	0.119	0.119	0.116	0.119	0.119	0.121	0.125

Table B.52: *Performance of k-NN on lending campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.82	0.842	0.85	0.856	0.85	0.851	0.847	0.844	0.842	0.839	0.846	0.849	0.852	0.844
R^2_{adj}	0.742	0.768	0.775	0.778	0.763	0.759	0.746	0.734	0.723	0.71	0.715	0.712	0.709	0.683
MSE	0.04	0.035	0.033	0.032	0.033	0.033	0.034	0.034	0.035	0.035	0.034	0.033	0.032	0.034
MAE	0.08	0.077	0.075	0.073	0.075	0.075	0.077	0.078	0.078	0.079	0.077	0.076	0.074	0.077

Table B.53: *Performance of Ridge Regression on lending campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.855	0.852	0.851	0.854	0.751	0.814	0.81	0.783	0.736	0.822	0.806	0.809	0.696	0.727
R^2_{adj}	0.792	0.63	0.776	0.775	0.607	0.699	0.685	0.783	0.538	0.679	0.64	0.635	0.401	0.445
MSE	0.032	0.032	0.033	0.032	0.055	0.041	0.042	0.048	0.058	0.039	0.043	0.042	0.067	0.06
MAE	0.092	0.093	0.093	0.093	0.104	0.1	0.101	0.103	0.105	0.098	0.101	0.1	0.105	0.106

Table B.54: *Performance of Bagging on lending campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.891	0.892	0.887	0.885	0.871	0.881	0.883	0.878	0.882	0.878	0.88	0.884	0.886	0.888
R^2_{adj}	0.844	0.842	0.83	0.823	0.797	0.808	0.806	0.792	0.793	0.78	0.778	0.778	0.776	0.772
MSE	0.024	0.024	0.025	0.025	0.028	0.026	0.026	0.027	0.026	0.027	0.026	0.026	0.025	0.025
MAE	0.061	0.061	0.063	0.065	0.07	0.066	0.066	0.069	0.067	0.068	0.066	0.066	0.065	0.064

Table B.55: *Performance of Random Forest on lending campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.81	0.805	0.801	0.81	0.806	0.806	0.805	0.808	0.807	0.812	0.808	0.814	0.816	0.816
R^2_{adj}	0.727	0.714	0.701	0.708	0.694	0.686	0.676	0.673	0.662	0.661	0.644	0.645	0.638	0.626
MSE	0.042	0.043	0.044	0.042	0.043	0.043	0.043	0.042	0.042	0.041	0.042	0.041	0.04	0.04
MAE	0.123	0.122	0.125	0.123	0.121	0.124	0.123	0.122	0.123	0.123	0.123	0.122	0.121	0.122

Table B.56: *Performance of AdaBoost on lending campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.666	0.682	0.682	0.682	0.682	0.682	0.682	0.7	0.685	0.663	0.669	0.692	0.707	0.669
R^2_{adj}	0.521	0.533	0.522	0.511	0.498	0.486	0.472	0.489	0.448	0.393	0.386	0.412	0.423	0.327
MSE	0.073	0.07	0.07	0.07	0.07	0.07	0.07	0.066	0.069	0.074	0.073	0.067	0.064	0.073
MAE	0.195	0.188	0.188	0.188	0.188	0.188	0.188	0.188	0.193	0.183	0.185	0.192	0.178	0.195

Table B.57: *Performance of XGBoost on lending campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.857	0.884	0.887	0.868	0.858	0.872	0.848	0.857	0.87	0.866	0.88	0.884	0.869	0.898
R^2_{adj}	0.795	0.83	0.83	0.797	0.776	0.793	0.748	0.756	0.772	0.759	0.778	0.778	0.742	0.793
MSE	0.031	0.025	0.025	0.029	0.031	0.028	0.033	0.031	0.028	0.029	0.026	0.025	0.029	0.022
MAE	0.074	0.067	0.064	0.074	0.077	0.074	0.077	0.077	0.077	0.078	0.071	0.07	0.073	0.067

Table B.58: *Performance of Histogram Gradient Boosting on lending campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.848	0.847	0.85	0.849	0.849	0.851	0.853	0.854	0.853	0.854	0.857	0.862	0.861	0.866
R^2_{adj}	0.782	0.775	0.775	0.768	0.762	0.759	0.756	0.751	0.743	0.737	0.735	0.736	0.726	0.728
MSE	0.033	0.034	0.033	0.033	0.033	0.033	0.032	0.032	0.032	0.032	0.031	0.03	0.03	0.029
MAE	0.102	0.101	0.101	0.102	0.102	0.101	0.101	0.101	0.101	0.102	0.101	0.1	0.1	0.101

Table B.59: *Performance of MLP on lending campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.832	0.874	0.866	0.857	0.858	0.852	0.879	0.863	0.882	0.857	0.873	0.883	0.875	0.868
R^2_{adj}	0.759	0.815	0.799	0.78	0.776	0.761	0.799	0.767	0.793	0.742	0.765	0.777	0.754	0.732
MSE	0.037	0.028	0.029	0.031	0.031	0.032	0.027	0.03	0.026	0.031	0.028	0.026	0.027	0.028
MAE	0.056	0.048	0.049	0.052	0.05	0.054	0.047	0.047	0.043	0.051	0.046	0.044	0.047	0.045

Table B.60: *Performance of LSTM on lending campaigns over a 14-day time series for predicting Overfunding.*

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
R^2	0.832	0.881	0.868	0.861	0.889	0.852	0.874	0.843	0.875	0.87	0.871	0.869	0.86	0.855
R^2_{adj}	0.759	0.825	0.802	0.786	0.825	0.761	0.791	0.732	0.781	0.766	0.761	0.75	0.724	0.705
MSE	0.037	0.026	0.029	0.031	0.024	0.033	0.028	0.035	0.027	0.029	0.028	0.029	0.031	0.032
MAE	0.056	0.046	0.048	0.051	0.044	0.053	0.049	0.055	0.048	0.049	0.049	0.051	0.051	0.052