



An Integrated Technical and Fundamental Stock Analysis Model
for the Saudi Stock Market (Tadawul)

Prepared By

Name: Malek M. Atarji

ID No.: 252000474

Submitted in fulfillment of the requirements for the degree Master of Science in Artificial
Intelligence

At

College of Informatics

Midocean University

UAE

Supervisor Name: Dr Amjad Qatish

College of Informatics, Midocean University, Comoros

Email: 252000474@st.midocean.edu.km

Acknowledgements

All praise and gratitude are due to Allah, whose guidance, mercy, and blessings have enabled me to complete this research.

I would like to express my sincere appreciation to Midocean University for providing an academic environment that encouraged learning, scientific research, and professional development throughout my master's studies.

My deepest gratitude is extended to my supervisor, Dr. Amjad Qtaish, whose continuous guidance, valuable feedback, academic expertise, and constructive discussions have greatly contributed to the successful completion of this thesis. His patience, encouragement, and commitment to academic excellence have been invaluable throughout every stage of this research.

I would also like to thank the faculty members of the College of Informatics for their support and for providing the knowledge and academic foundation that contributed to the completion of this work.

Finally, I extend my heartfelt appreciation to my family for their unwavering encouragement, patience, and continuous support throughout my academic journey. Their belief in me has been a constant source of motivation during the completion of this research.

Abstract

The rapid growth of financial markets and the increasing availability of digital financial data have created significant opportunities for the application of Artificial Intelligence (AI) in investment decision-making. Modern stock exchanges continuously generate large volumes of heterogeneous information, including historical market prices, trading activity, and corporate financial statements. Transforming these diverse data sources into reliable trading decisions remains one of the most challenging problems in financial forecasting due to the dynamic, nonlinear, and uncertain nature of financial markets.

Stock market prediction has traditionally been approached from two different perspectives. The first relies on **technical analysis**, which attempts to identify future market movements using historical price and trading information. The second relies on **fundamental analysis**, which evaluates the financial condition of companies using accounting information and financial ratios derived from corporate financial statements. While both approaches have demonstrated predictive value, each captures only part of the information available to investors. Technical indicators primarily reflect short-term market behavior, whereas financial variables describe the underlying economic condition of listed companies. Consequently, relying on either information source independently may limit prediction performance.

Recent advances in Artificial Intelligence, Machine Learning, and Deep Learning have enabled researchers to develop increasingly sophisticated prediction models capable of learning complex nonlinear relationships from financial data. Nevertheless, many existing studies continue to focus on improving individual prediction algorithms without giving sufficient attention to data engineering, temporal consistency, information leakage prevention, and practical investment evaluation. In addition, relatively few studies have investigated hybrid prediction frameworks specifically designed for the Saudi Stock Exchange while integrating both technical and financial information within a unified learning architecture.

To address these limitations, this thesis proposes a comprehensive hybrid artificial intelligence framework for short-term stock market prediction in the Saudi Stock Exchange (Tadawul). The proposed framework combines two specialized prediction models: a Bidirectional Long Short-Term Memory (BiLSTM) network designed to learn temporal patterns from sequential technical indicators, and an eXtreme Gradient Boosting (XGBoost) model developed to analyze structured financial variables extracted from quarterly corporate financial statements. Rather than selecting a single prediction model, the probabilistic outputs generated by both learning algorithms are integrated through a probability-level hybrid fusion strategy to produce unified prediction probabilities.

Unlike many previous studies, the contribution of this research extends beyond the development of prediction models. Particular emphasis is placed on constructing a complete engineering pipeline beginning with data collection and quality validation, followed by chronological financial integration, feature engineering, binary target generation, specialized model development, probability fusion, and portfolio-oriented evaluation. Throughout the proposed methodology, strict chronological consistency is maintained to eliminate information leakage and preserve the realism of the prediction process.

The effectiveness of the proposed framework is evaluated using historical data collected from companies listed on the Saudi Stock Exchange. The experimental evaluation considers both conventional machine learning performance metrics and practical investment performance through a virtual portfolio simulation, providing a comprehensive assessment of the proposed methodology from both computational and financial perspectives.

Overall, this thesis demonstrates that combining complementary technical and financial information within a carefully engineered hybrid artificial intelligence framework can improve short-term trading signal prediction and provide more reliable decision support for investors operating in the Saudi Stock Exchange.

Table of Contents

Contents

Acknowledgements	2
Abstract	3
Table of Contents	5
List of Abbreviations	10
List of Figures	10
List of Tables	11
Chapter 1 Introduction	13
1.1 Research Background.....	13
1.2 Problem Statement	15
1.3 Research Motivation	16
1.4 Research Aim	17
1.5 Research Objectives	18
1.6 Research Questions	19
1.7 Research Contributions	19
CHAPTER 2 Background and Related Work.....	22
2.1 Introduction	22
2.2 Financial Market Prediction	23
2.3 Artificial Intelligence in Financial Forecasting.....	24
2.4 Artificial Information in Financial Markets	26
2.4.1 Technical Market Information.....	27
2.4.2 Fundamental Financial Information	28
2.4.3 Complementary Nature of Technical and Fundamental Information.....	28
2.5 Machine Learning and Deep Learning for Financial Prediction.....	29
2.5.1 Machine Learning for Structured Financial Data	30
2.5.2 Deep Learning for Sequential Financial Data	30
2.5.3 Comparison Between Machine Learning and Deep Learning.....	31
2.6 Hybrid Artificial Intelligence Frameworks	32
2.6.1 Evolution of Hybrid Financial Prediction Models	32
2.6.2 Types of Hybrid Artificial Intelligence	33
2.6.3 Advantages of Hybrid Artificial Intelligence	34

2.6.4 Theoretical Foundation of the Proposed Framework	35
2.6.5 Chapter Relevance	36
2.7 Evaluation of Artificial Intelligence Models in Financial Prediction	36
2.7.1 Machine Learning Evaluation Metrics	37
2.7.2 Investment-Oriented Performance Metrics	37
2.7.3 Integrating Computational and Financial Evaluation	38
2.7.4 Explainability as a Third Evaluation Dimension.....	38
2.7.5 Relevance to the Present Research	39
2.8 Related Work.....	40
2.8.1 Evolution of Financial Prediction Research	40
2.8.2 Traditional Machine Learning Studies	40
2.8.3 Deep Learning Studies.....	41
2.8.4 Hybrid Artificial Intelligence Studies.....	41
2.8.5 Studies on the Saudi Stock Exchange.....	42
2.8.6 Summary of Existing Literature	44
2.9 Critical Analysis of Existing Literature	44
2.10 Research Gap	45
2.11 Positioning of the Proposed Framework	47
2.12 Chapter Summary	48
CHAPTER 3: Proposed Methodology.....	49
3.1 Introduction	49
3.2 Phase 1 - Blind Data Audit.....	50
3.2.1 Dynamic Dataset Exploration.....	50
3.2.2 Data Integrity Assessment	51
3.2.3 Ticker Validation	52
3.2.4 Temporal Validation.....	52
3.2.5 Baseline Data Retention	53
3.3 Phase 2 - Anti-Leakage and Temporal Alignment.....	53
3.3.1 Financial Reporting Delay	54
3.3.2 Time-Aware Financial Integration	55
3.3.3 Chronological Verification	55
3.3.4 Data Filtering and Retention Analysis.....	56

3.3.5 Value of Anti-Leakage Control	57
3.4 Phase 3 - Feature Engineering.....	57
3.4.1 Technical Feature Engineering.....	58
3.4.2 Financial Feature Engineering.....	59
3.4.3 Feature Coverage Validation	60
3.4.4 Feature Selection Strategy	61
3.4.5 Phase Output.....	61
3.5 Phase 4 - Dataset Construction and Binary Target Engineering.....	61
3.5.1 Construction of the Master Dataset	62
3.5.2 Binary Target Engineering	63
3.5.3 Prediction Horizon.....	63
3.5.4 Distribution of Binary Targets.....	64
3.5.5 Interpretation of Target Distribution	64
3.5.6 Validation of the Dataset	64
3.5.7 Output of Phase 4	65
3.6 Development of the Individual Prediction Models	65
3.6.2 Sequential Data Preparation	66
3.6.3 BiLSTM Training Strategy.....	66
3.6.4 Output of the BiLSTM Model.....	66
3.6.5 BiLSTM Model Configuration and Training Strategy	67
3.6.5.1 Sequential Observation Window	67
3.6.5.2 Network Architecture	67
3.6.5.3 Training Strategy	68
3.6.5.4 Early Stopping Strategy.....	69
3.6.5.5 Model Optimization.....	69
3.6.5.6 Training Monitoring	70
3.6.5.7 Output of the BiLSTM Training Phase	70
3.6.6 Development of the XGBoost Model.....	70
3.6.6.1 Financial Feature Representation	70
3.6.6.2 Gradient Boosting Strategy	71
3.6.6.3 Model Configuration	71
3.6.6.4 Training Procedure	72

3.6.6.5 Feature Learning Characteristics	72
3.6.6.6 Output of the XGBoost Model	72
3.6.7 Comparative Characteristics of the Individual Models	73
3.7 Phase 6 - Two-Level Probabilistic Fusion (Stacking Ensemble).....	73
3.7.1 Rationale for Probability Fusion.....	74
3.7.2 Probability Integration Strategy.....	74
3.7.3 Meta-Learner Training Strategy	75
3.7.4 Learned Meta-Feature Contributions.....	75
3.7.5 Hybrid Probability Output	76
3.7.6 Advantages of the Proposed Stacking Strategy	76
3.7.7 Decision Threshold Optimization.....	76
3.7.8 Output of Phase 6	77
3.8 Phase 7 - Portfolio Evaluation Framework	77
3.8.1 From Prediction to Trading Decision	78
3.8.2 Virtual Portfolio.....	78
3.8.3 Trading Rules	78
3.8.4 Portfolio Performance Recording.....	79
3.8.5 Performance Measurement Strategy.....	79
3.8.6 Separation Between Learning and Evaluation.....	79
3.9 Chapter Summary.....	80
CHAPTER 4 Experimental Results and Discussion.....	81
4.1 Introduction	81
4.2 Experimental Environment	81
4.3 Dataset Validation Results	82
4.3.1 Final Dataset Characteristics	83
4.3.2 Chronological Validation Results.....	83
4.3.3 Binary Target Verification.....	84
4.3.4 Feature Validation Results.....	84
4.3.5 Discussion.....	84
4.4 Performance Evaluation of the BiLSTM Model.....	85
4.4.1 Classification Performance	85
4.4.2 Interpretation of the Results	85

4.4.3 Discussion.....	86
4.5 Performance Evaluation of the XGBoost Model	87
4.5.1 Classification Performance	88
4.5.2 Interpretation of the Results	88
4.5.3 Comparative Analysis with BiLSTM.....	88
4.5.4 Discussion.....	89
4.6 Performance Evaluation of the Proposed Hybrid Model	91
4.6.1 Hybrid Classification Performance	91
4.6.2 Comparative Analysis.....	92
4.6.3 Interpretation of the Hybrid Performance	93
4.6.4 Discussion.....	94
4.6.5 Generalization Analysis.....	96
4.6.6 Explainability Analysis.....	97
4.6.6 Discussion of Explainability.....	98
4.7 Portfolio Performance Evaluation	99
4.7.1 Portfolio Evaluation Methodology	100
4.7.2 Portfolio Performance Results.....	100
4.7.4 Portfolio Performance Results.....	101
4.7.4 Interpretation of the Portfolio Results	101
4.7.6 Discussion.....	102
4.7.7 Why Sortino Ratio Was Selected	103
4.8 General Discussion.....	103
4.8.1 Study Limitations	104
4.9 Chapter Summary.....	104
CHAPTER 5 Conclusion and Future Work.....	105
5.1 Introduction	105
5.2 Summary of the Research	105
5.3 Achievement of the Research Objectives.....	105
5.4 Main Research Contributions.....	106
5.4.1 Contribution 1: Development of an End-to-End Hybrid Artificial Intelligence Framework.....	106
5.4.2 Contribution 2: Chronological Financial Data Engineering Framework	106
5.4.3 Contribution 3: Specialized Learning Architecture.....	106

5.4.4 Contribution 4: Two-Level Probabilistic Stacking Ensemble	107
5.4.5 Contribution 5: Separation Between Prediction and Investment Evaluation	107
5.4.6 Contribution 6: Comprehensive Evaluation Framework.....	107
5.4.7 Contribution 7: Application to the Saudi Stock Exchange.....	107
5.5 Practical Implications.....	108
5.6 Limitations of the Research.....	108
5.7 Future Research.....	109
5.8 Final Conclusion	110
References	111

List of Abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
BiLSTM	Bidirectional Long Short-Term Memory
XGBoost	Extreme Gradient Boosting
SHAP	SHapley Additive exPlanations
ROC	Receiver Operating Characteristic
ROC-AUC	Area Under the Receiver Operating Characteristic Curve
TP	Take Profit
SL	Stop Loss
TPR	True Positive Rate
FPR	False Positive Rate
EMA	Exponential Moving Average
RSI	Relative Strength Index
MACD	Moving Average Convergence Divergence
ATR	Average True Range
OBV	On-Balance Volume
CMF	Chaikin Money Flow

List of Figures

Figure 1 General Artificial Intelligence Pipeline for Financial Forecasting Based on the Proposed Hybrid Stacking Framework.....	26
Figure 2 Complementary Financial Information within AI Prediction Systems	29
Figure 3 Evolution of Financial Prediction Approaches.....	33
Figure 4 presents a conceptual framework for the trading system that developed within this research.	36
Figure 5 Three-Dimensional Evaluation Framework	39
Figure 6 Positioning of the Proposed Framework	47
Figure 7 Represents the architecture for the proposed framework for artificial intelligence models for prediction of the Saudi Stock Exchange.	50
Figure 8 Financial Reporting Timeline.....	55
Figure 9 Anti-Leakage Validation Pipeline	57
Figure 10 Shows the overall feature engineering workflow used in this research.	58
Figure 11 Financial Feature Engineering.....	60
Figure 12 Dataset Construction Workflow	62
Figure 13 Binary Target Engineering	63
Figure 14 depicts the development of these individual AI models.....	66
Figure 15 illustrates the architecture of the BiLSTM network.	68
Figure 16 Financial Learning Pipeline.....	71
Figure 17 Two-Level Probabilistic Fusion (Stacking Ensemble).....	74
Figure 18 Portfolio Evaluation Framework	77
Figure 19 illustrates this separation between the learning and evaluation of the trading portfolio.	80
Figure 20 Confusion Matrix of the BiLSTM Model	86
Figure 21 ROC Curve of the BiLSTM Model.....	87
Figure 22 Training History of the BiLSTM Model	87
Figure 23 Confusion Matrix of the XGBoost Model.....	89
Figure 24 ROC Curve of the XGBoost Model	90
Figure 25 Feature Importance	90
Figure 26 Performance Comparison Hybrid Model	95
Figure 27 Confusion Matrix of the Hybrid Model.....	95
Figure 28 ROC Curve of the Hybrid Model	96
Figure 29 SHAP Summary Plot.....	98
Figure 30 Portfolio Performance Metrics	100
Figure 31 Equity Curve.....	101
Figure 32 Drawdown Curve.....	101
Figure 33 Portfolio Comparison	102

List of Tables

Table 1 Comparison Between Machine Learning and Deep Learning in Financial Prediction ...	32
Table 2 Hybrid AI Fusion Strategies	34
Table 3 Common Machine Learning Evaluation Metrics.....	37
Table 4 Common Investment Performance Metrics	38
Table 5 Summary of Existing Literature	44
Table 6 Research Gap Analysis	46
Table 7 Data Sources Used in This Research	51
Table 8 Data Retention During Phase 2.....	56
Table 9 Technical Feature Categories.....	59
Table 10 Financial Feature Engineering	59
Table 11 Example Feature Coverage.....	61
Table 12 Summarizes the distribution of the binary target values.....	64
Table 13 BiLSTM Input Configuration	67
Table 14 BiLSTM Configuration.....	69
Table 15 XGBoost Hyperparameter Configuration.....	72
Table 16 Presents a comparison between the two individual models within the system.	73
Table 17 Two-Level Probabilistic Stacking Workflow	73
Table 18 Meta-Features Used by the Stacking Ensemble	75
Table 19 Learned Meta-Feature Contributions.....	75
Table 20 Advantages of Probability Fusion.....	76
Table 21 Portfolio Evaluation Rules.....	78
Table 22 Experimental Environment	82
Table 23 Final Dataset Characteristics	83
Table 24 Final Binary Target Distribution.....	84
Table 25 Performance of the BiLSTM Model.....	85
Table 26 Performance of the XGBoost Model	88
Table 27 Comparison Between Standalone Models	88
Table 28 Performance of the Hybrid Model	91
Table 29 Comparison of Prediction Models	92
Table 30 Portfolio Performance.....	100
Table 31 Summary of Research Contributions	107
Table 32 Future Research Opportunities	110

Chapter 1 Introduction

1.1 Research Background

Artificial Intelligence (AI) technology has gradually become one of the most influential technologies in the area of financial decision support systems. AI technology is able to analyze massive amounts of financial data to recognize and derive complex trading patterns from these data sets that would be difficult to recognize with conventional methods of financial data analysis. With the development of new advances in the field of machine and deep learning, AI technologies have been able to find significant improvements in their capacity to recognize the nonlinear relationships within the data, enabling their application to a variety of financial processes.

Among the various applications of AI technology to financial markets, one of the most challenging areas of research is the development of AI systems that are able to accurately predict the movements of the stock market. Due to the dynamic nature of financial markets, and the various economic and market conditions that impact those markets, the data from stocks exhibit non-linear relationships that pose challenges to AI systems that are not specifically designed to accommodate for such complexity. AI systems that can learn the nonlinear patterns within the stock market data without making assumptions regarding the distribution of the stock data can be applied to the resolution of this problem. Such systems include support vector machines, random forests, neural networks, and others. In particular, deep learning technologies such as LSTM and BiLSTM networks have been applied to financial markets in the attempt to recognize complex long-term temporal patterns within historical stock market data.

In addition to the difficulty of accurately predicting the movements of stocks, the development of systems that can recognize profitable trading opportunities within the stock market is an even more challenging problem to solve. Most existing approaches to stock market prediction attempt to model the problem as a classification problem in which the models must determine whether the closing price of a stock rise or fall the following trading day. In addition, most existing models have provided improvements in their classification results. However, such improvements in classification accuracy do not necessarily translate into improvements in investment performance. As an investor, one is interested in finding trading opportunities that exhibit certain profit and risk level objectives over a trading period of interest. Thus, methods that do not incorporate such objectives are lacking in their utility for investors.

Another issue with existing approaches to stock market prediction is their reliance upon specific classes of information from the market to train and operate their models. For example, many approaches to stock market prediction employ only technical indicators to inform the models of the stock's movements. Alternatively, other methods use only financial statements to determine the movements of stock prices. Technical indicators effectively provide information regarding the strength of stocks during short periods. However, they do not provide information regarding the financial strength of those companies. Financial statements provide information about the financial performance of those companies, but do not reflect the short-term dynamics of their stocks. Thus, AI systems that only use one of these sources of information are unlikely to develop models that can effectively learn of the movements of stocks.

To overcome these disadvantages of existing stock market AI prediction models, it is beneficial to develop model frameworks that can incorporate information from both technical indicators and financial statements for stocks. However, the development of such models introduces additional challenges in the AI systems engineering process. The integration of data from both sources of information, the preservation of the chronological nature of both data sets, and the avoidance of information leakage between data sets for training and evaluation of the models are all challenges to be overcome in the development of such systems.

These challenges are additionally compounded in the attempt to develop stock market prediction models for regions of the world whose characteristics are different from those of major international markets. The Saudi Stock Exchange, or Tadawul, is the largest stock exchange in the Middle East and North Africa region. The stock exchange has undergone extensive growth in the last few years as a result of the economic reforms that have been enacted within the Kingdom of Saudi Arabia. The growing interest of the Kingdom in developing as an industrial and trading power has enabled many new companies to be listed on the exchange, as well as has encouraged both domestic and foreign investors to trade on the exchange. As a result, the market has become more dynamic and information-rich. The complexity of the Tadawul market, however, presents challenges for the development of AI-based decision support systems that are able to account for both the market and financial data for stocks listed within the exchange.

Despite the proliferation of AI models for financial market prediction, significant challenges remain unaddressed: (1) most models rely on either technical or fundamental data in isolation, (2) chronological integration of financial statements with market data is often overlooked, leading to look-ahead bias, (3) evaluation focuses on classification accuracy rather than investment profitability, (4) probability outputs are used suboptimally without calibration or adaptive fusion, (5) model interpretability is neglected, and (6) the Saudi Stock Exchange (Tadawul) lacks a comprehensive AI framework that addresses these challenges simultaneously. This research aims to fill these gaps by proposing an integrated two-level probabilistic stacking framework specifically designed for the Tadawul market.

In response to the shortcomings of existing approaches to stock market prediction models, this research intends to propose an integrated **two-level probabilistic stacking framework** for the Saudi Stock Exchange. This framework simultaneously utilizes deep learning approaches based upon BiLSTM networks and machine learning approaches based upon XGBoost methods. Furthermore, in addition to providing the model with general classification accuracy rates, it will additionally aim to provide trading signals that can be evaluated for their effectiveness in the development of trading strategies. Additionally, the financial information that is employed by the framework is integrated in a time-aware manner to avoid distorting the chronological nature of the information, and to avoid information leakage between the training and testing data sets for the AI models. Finally, the approach will additionally utilize explainable AI techniques in the creation of its model to improve the interpretability of the framework.

Thus, overall, the aim of this research is to develop an AI framework that can perform stock market prediction in the Saudi Stock Exchange that incorporates data engineering, machine learning, deep learning, probability calibration, a two-level probabilistic stacking ensemble with

a Logistic Regression meta-learner, explainable AI techniques, and portfolio evaluation within a unified decision-support framework.

1.2 Problem Statement

Artificial intelligence has rapidly expanded the use of machine learning and deep learning models to perform the task of financial market price prediction. A plethora of prediction models exist in the literature for the task of forecasting the price movements of stocks. However, finding accurate and profitable investment opportunities continues to be a great challenge in the world of financial markets.

One limitation regarding existing stock prediction models is the use of only one source of financial information for prediction. Technical indicators based solely on the price movements of stocks have been used for stock prediction. These indicators provide valuable insight into the price movement of stocks, but they do not provide information regarding the financial state of the companies represented by those stocks. As such, using these indicators alone to inform trading and investment decisions may lead to inaccurate investment outcomes.

Another alternative to stock prediction using only one information source is the use of fundamental financial analysis. For this approach, various financial indicators based on financial statements of the companies are utilized in the stock market prediction process. These indicators provide valuable insight into the financial state of the represented companies. However, the information is published only periodically and does not represent the rapid price movement of stocks every trading day. Hence, using only financial indicators to represent stock price movements may not lead to the best trading and investment outcomes either.

Another challenge to the existing literature on stock price movement prediction models is the methods employed to evaluate the models' performances. While metrics like the F1-score are more appropriate for evaluating models with imbalanced datasets than accuracy, these machine learning metrics do not provide an insight into the financial performance that is required of a stock market prediction model. Hence, the models need to be evaluated based on financial indicators as well.

Moreover, another issue with the current literature on stock price movement prediction is the merging of financial statements with stock market price data without considering the delay between the release of financial statements and the availability of that information to the public. Using data that includes such future information may lead to the creation of bias in the machine learning models, which is referred to as Look-Ahead Bias. This bias can significantly impact the evaluation of models and their validity for use in stock market predictions.

In addition to these limitations in the existing literature on machine learning models for stock market price movement prediction, many of the current models are based on the prediction of the price of stocks for the following trading day. While it is true that stock price movements may trend similarly from trading day to trading day, investors are generally interested in knowing if the stock can reach a certain investment objective within a certain period. Hence, it would be

beneficial to create a machine learning model based on the achievement of these investment objectives rather than the price of the stock within a trading day.

Another critical problem for stock market price movement prediction that relates to the existing literature on the topic is the lack of investigation of these issues for the Saudi Stock Exchange (Tadawul). As the largest financial exchange in the Middle East and North Africa region, the Saudi Stock Exchange is underrepresented in the literature on artificial intelligence. Most of the existing studies regarding the exchange investigate limited datasets and use models created for other international markets. In addition, these studies do not evaluate the investment performance of the models proposed for the Tadawul Exchange.

Considering these limitations within the existing literature on stock price movement prediction using artificial intelligence for the Saudi Stock Exchange, there is a need for an integrated artificial intelligence framework that can address these challenges simultaneously. Therefore, a machine learning model for the prediction of stock market price movements for the Tadawul Exchange could incorporate financial information, ensure chronological integrity for the data, incorporate investment objectives for the stocks for investors, and evaluate the models according to both machine learning and financial performance indicators.

In response to these research needs and challenges, a framework for stock price movement prediction for the Saudi Stock Exchange proposed in this dissertation. The proposed model employs deep learning and machine learning models that are integrated into a single engineering framework for the achievement of the above goals. This framework employ Bidirectional Long Short-Term Memory (BiLSTM) and eXtreme Gradient Boosting (XGBoost) models to predict stock price movements while providing calibrated and interpretable probabilities and evaluating the models according to both machine learning and financial indicators to provide an accurate and valuable stock price movement prediction model.

1.3 Research Motivation

Currently, the increasing use of AI in financial markets presents opportunities for the development of intelligent systems to assist investors in understanding and analyzing market behaviour. Despite the availability of various AI models for market prediction, however, the performance of these models often fails to translate into improved financial market performance by investors. Thus, one of the main motivations for conducting this research is to address this discrepancy between prediction and financial performance.

Furthermore, market prediction problems differ from many traditional AI problems in that markets exhibit temporal information, nonlinear behaviours, and continuously changing market conditions. Thus, additional motivation for this research emerges in the development of an appropriate engineering framework for AI model development that accounts for these aspects of market data.

The technical and fundamental information available in markets is inherently complementary of each other. While market prices provide information regarding the short-term trading behaviour of individual listed companies, their financial statements provide information regarding the

financial and operational status of those firms. Thus, additional motivation for integrating both of these types of information into a single AI framework emerges in the potential ability of such a framework to provide more comprehensive representations of the financial markets and their behaviours.

Beyond prediction of market movements, investors are also often interested in evaluating the profitability of using those models for trading activities. So, additional motivation for the incorporation of measures of profitability and risk into the evaluation of the AI models for trading is provided.

Another motivation for the research was the need to increase the methodological rigor of AI models used in financial markets. Ensuring that temporal information is not inadvertently “leaked” into the training data, that data is chronologically whole, and that trading models utilize only information available during the indicated trading periods are all crucial aspects of ensuring that the developed models exhibit similar performance in live markets as that which is reported in testing environments.

In addition to the aforementioned motivations, the markets to be evaluated in this research are the companies listed on the Saudi Stock Exchange (Tadawul). The Tadawul is one of the largest stock exchanges in the Middle East and North Africa region, and has a variety of listed companies from different economic sectors. Despite the relatively high level of activity in the Tadawul, however, there is limited research regarding the use of AI frameworks for financial trading in the Saudi market.

Finally, motivation for this research is provided in the desire to develop an AI model that inherently accounts for the goals of investors. Typical AI models are developed with the intention of maximizing classification performance. However, investors are often interested in the profitability of utilizing AI models to trade stocks in the Tadawul market. Thus, motivation for the development an AI model that accounts for these factors exists.

Overall, then, there is motivation for the development of an AI framework that can successfully incorporate both technical and financial data from listed companies, that accounts for the goals of investors, and that can provide an evaluation of the performance of trading using those AI models. Such framework would contribute to both the field of AI and financial markets, and to the development of intelligent decision-support systems for the Saudi Stock Exchange market.

1.4 Research Aim

The aim of this research is to develop a novel **two-level probabilistic stacking framework** that improves BiLSTM and XGBoost models for short-term trading signal prediction that can be used to identify short-term trading opportunities within the Tadawul stock exchange. More specifically, the framework will simultaneously utilize both technical indicators and financial information to determine trading signals that can be used by investors to take advantage of those identified opportunities.

Unlike existing AI models that utilize only technical indicators to make trading decisions, the framework that developed will account for the financial status of the companies whose stocks are traded in the Tadawul. Furthermore, the framework optimized for investment performance rather than classification model performance, it will utilize explainable AI methods to improve the interpretability of its suggestions for trading decisions, and will include methods for evaluating the performance of those trading suggestions.

As a result, the aim of this research is to develop a novel, hybrid AI framework that improves the ability of investors in the Tadawul to achieve improved financial returns from their investments.

1.5 Research Objectives

In order to achieve the aim of this research, the following objectives pursued:

To design and implement a financial data engineering pipeline that incorporates both daily and quarterly financial data in a way that avoids look-ahead bias.

To engineer the various technical indicators and the binary return variable that represents the short-term return on investment of each company's stock.

To develop and train a Bidirectional Long Short-Term Memory (BiLSTM) deep learning model that learns from the technical indicators.

To develop and optimize an eXtreme Gradient Boosting (XGBoost) deep learning model that learns from the financial data variables.

To create a two-level ensemble model that combines the predictions of the BiLSTM and XGBoost models using Isotonic and Logistic regression models.

To optimize the threshold for the prediction probabilities in order to improve the performance of the trading portfolio.

To enhance the interpretability of the models using Explainable Artificial Intelligence (XAI) techniques based on the SHAP analysis.

To evaluate the proposed framework using various performance metrics for machine learning models as well as models that assess the performance of trading portfolios.

These objectives will enable the development of a novel and comprehensive framework for predicting short-term trading signals in the Tadawul stock exchange using artificial intelligence techniques.

1.6 Research Questions

In order to accomplish the stated objectives of this research and to fill the research gap identified in this research proposal, the following research questions utilized in the performance of the research:

RQ1: Does the two-level probabilistic stacking ensemble algorithm achieve a significantly higher F1-Score and Sortino Ratio compared to the BiLSTM and XGBoost models when used to predict trading signals on the Tadawul?

RQ2: Does the integration of technical indicators and financial statement ratios lead to improvements in the Recall and F1-Score of the trading signal prediction model compared to the models that use either type of data for trading signal predictions alone?

RQ3: More specifically, can the integration of financial data and chronological market data eliminate information leakage caused by the chronological market data?

RQ4: Does the use of isotonic regression to calibrate the probabilities of the models followed by the use of logistic regression to create a stacked ensemble model lead to an increase in the Sortino Ratio and the return of the investment portfolio compared to the probabilities of the models without calibration?

RQ5: Furthermore, will the use of thresholds other than those created using the conventional probabilities of the models increase the performance of the investment portfolio created with these models?

RQ6: Which of the financial variables contribute to the predictions of the XGBoost model, and how does this enhance the confidence that investors have in the model?

These research questions will help to determine the main direction of this research. Each of these research questions be answered through the development and implementation of the framework for the artificial intelligence models that will be used to attempt to predict the trading signals of the Saudi Stock Exchange.

1.7 Research Contributions

The research that performed in this dissertation contributes to the existing literature on artificial intelligence models for predicting financial markets through the development, implementation, and evaluation of the proposed framework for a hybrid artificial intelligence system for the prediction of trading signals on the Saudi Stock Exchange.

Many existing studies on artificial intelligence for financial markets have focused upon evaluating the accuracy of the predictions of individual machine learning models. However, the proposed framework addresses various aspects of the artificial intelligence system for financial markets prediction - from the engineering of the financial data itself to the evaluation of the performance of the resulting investment portfolio.

The main contributions of the research that be performed are as follows.

The major novelty of this research is the integration of the various proposed components into a single engineering framework for artificial intelligence models that is specifically developed for use within the Saudi Stock Exchange. Unlike many existing research studies that evaluated either the technical or financial aspects of the stocks listed on that exchange alone, this framework integrates chronological data engineering, specialized BiLSTM and XGBoost models, probability calibration, a two-level probabilistic stacking ensemble, explainable artificial intelligence, and portfolio-oriented evaluation into a unified engineering framework. Furthermore, like existing financial models, the framework evaluates the performance of the proposed artificial intelligence models as an indicator of the potential performance of investors who utilize the trading signals that are provided.

Contribution 1: Integrated Hybrid AI Framework for Short-Term Trading Prediction

The framework for artificial intelligence models for trading signal prediction will include the integration of deep learning models, specifically Bidirectional Long Short-Term Memory (BiLSTM) models, with machine learning models, specifically eXtreme Gradient Boosting (XGBoost) models.

Contribution 2: Time-Aware Financial Integration Strategy using a 45-day reporting buffer

An integration strategy for financial data from the companies listed on the exchange developed that incorporates the chronological nature of both the financial reports and trading signals to eliminate information leakage caused by the temporal information.

Contribution 3: Investment-Oriented Target Engineering

Rather than utilizing conventional strategies for engineering the target value of the artificial intelligence models to recognize whether the trading price of a stock will rise or fall the following trading day, the target of the models will instead be engineered to indicate whether the stock can produce an investment opportunity over a seven-day period.

Contribution 4: Two-Level Probabilistic Stacking Ensemble

A two-level probabilistic stacking strategy is proposed in which probability estimates generated by the BiLSTM and XGBoost models are first calibrated using Isotonic Regression. The calibrated probabilities are subsequently transformed into informative meta-features and supplied to a Logistic Regression meta-learner that learns the optimal integration strategy for producing the final hybrid trading probability.

Contribution 5: Investment-Oriented Evaluation Framework

In evaluating the performance of the developed framework, conventional metrics for evaluating machine learning models determined for the artificial intelligence system. Furthermore,

additional indicators of the performance of the portfolio that results from utilizing the trading signals from the AI models evaluated.

Contribution 6: Explainable Artificial Intelligence Integration

SHAP-based explainability is employed to interpret the contribution of financial variables within the XGBoost component before their calibrated probabilities are incorporated into the stacking framework.

These contributions to artificial intelligence models for financial markets go beyond the development of a single model for predicting trading signals on the Saudi Stock Exchange. Instead, the developments of each of these components result in the establishment of a complete engineering framework for AI models for trading signal prediction that can be implemented into the exchange itself.

CHAPTER 2 Background and Related Work

2.1 Introduction

Artificial intelligence is a multidisciplinary area that has radically transformed the problem of financial market prediction. The growth of computational power and data availability has enabled the development of artificial intelligence systems with the computational power to learn non-linear relationships within the financial markets. Unlike many other AI problems, financial market prediction involves the simultaneous influencing of numerous factors that determine stock prices. These factors can include financial results of the companies with stocks listed on the market, economic indicators, and various external events. As a result, researchers have turned to artificial intelligence methods that can recognize complex non-linear relationships within the financial markets. AI methods such as machine learning and deep learning have allowed researchers to improve their models for financial market prediction. Furthermore, various frameworks that utilize hybrid forms of artificial intelligence methods have allowed for the creation of even more sophisticated financial market prediction models.

Despite the variety of methods that have been developed for financial market prediction, their usefulness in practice is generally limited. For instance, most financial market prediction studies treat the problem as a classification problem of determining closing prices for stocks and calculating overall classification metrics for those models. However, few of these studies provide any evaluation of the potential performance of applying these models to the stock market as an investment opportunity. Furthermore, many of the classification models that have been created with high accuracy for stocks have not led to any profitable investing strategies. Thus, one challenge for financial market prediction is evaluating financial AI models according to their potential to achieve profitable investment returns. In addition to this problem, the information that is used to train financial market prediction models also presents additional challenges. Most stock market prediction models utilize information from markets prices and financial statements for the companies that are listed on those markets. These two types of information have different characteristics, and utilizing only one of these potential data sources could potentially limit the capability of financial market prediction models to accurately determine the prices of stocks. Consequently, financial market prediction methods that incorporate both types of information are of growing interest to researchers.

Furthermore, additional developments in artificial intelligence methods have also created new opportunities for developing improved models for financial market prediction. For example, methods in artificial intelligence that enable model explanations can help to enhance the confidence of the models and in their applications to financial markets. Consequently, incorporating methods in AI that allow for improved interpretation of models for financial market prediction can lead to the development of models that are trusted in financial environments.

One market of interest is that of the Saudi Stock Exchange (Tadawul). The Saudi Stock Exchange is the largest stock exchange in the Middle East and North Africa. It includes companies that span a variety of different economic sectors. As such, the exchange provides various financial information regarding each of the companies that are listed on the exchange.

Despite the size and growth of Tadawul, however, relatively little research has been performed on artificial intelligence methods for the Saudi market, especially in regard to addressing each of the challenges of financial market prediction that are discussed above.

In order to address these research opportunities, it is first important to provide a review of the relevant theories, developments, and research that has been performed on each of these topics. In general, then, this chapter will provide a review of the background of financial market prediction, the role of artificial intelligence in financial markets, the integration of technical and fundamental data for financial market prediction, various machine learning and deep learning models for financial markets, various frameworks for creating hybrid artificial intelligence models for financial market prediction, and finally, a review of the research that has been performed on financial markets within both the international and Saudi markets. Following this review, the gaps in the existing research on these topics stated and the research that be performed in this dissertation be positioned within the existing literature on the topic of artificial intelligence for financial markets prediction.

2.2 Financial Market Prediction

Financial market prediction is one of the most challenging problems in finance and artificial intelligence. Financial market prediction aims to estimate the prices of stocks in the future by utilizing information regarding the stocks in the past. However, as with many other financial problems, there are a variety of factors that influence the prices of those stocks. For instance, in addition to the financial results of the companies whose stocks trade in the markets, factors such as economic indicators and external events can also impact the prices of those stocks. Consequently, financial markets tend to be dynamic and non-linear in their behaviour.

One of the main characteristics of financial markets is their evolution over time. Financial markets are non-stationary in that they continuously change, and the relationship between various factors in the markets changes over time. As such, financial market prediction algorithms must be able to adapt to these changes in the stocks over time. Furthermore, the information available for financial markets is diverse. For instance, investors consider information regarding the price of stocks over time as well as financial statements of the companies whose stocks trade in those markets. The need to integrate these different types of information for stock market prediction models presents a major challenge to the development of financial market AI models.

Financial market prediction has traditionally used statistical models to predict prices for stocks. Some of the most common models have included models like autoregressive models, moving average models, and GARCH models. While these models have some use in the determination of prices of stocks, their assumptions regarding the stability of relationships between market factors limits their ability to accurately model the complex and non-linear relationships between those factors in the market.

Machine learning models have been developed that can overcome some of the limitations of statistical models. For instance, machine learning algorithms are able to learn the complex and non-linear relationships between the various factors in financial markets without having to make

assumptions regarding the nature of those relationships. Thus, the development of machine learning models has increased the performance of financial market prediction algorithms.

Deep learning models have also been applied to the problem of financial market prediction. Deep learning methods such as LSTM models and BiLSTM models are able to learn the temporal relationships between the factors that influence the prices of stocks over time. Furthermore, other models such as XGBoost have also become increasingly popular in financial markets as they are able to learn the non-linear relationships between the factors while maintaining high computational speeds in comparison to other deep learning algorithms.

Beyond forecasting methods, however, evaluations of financial market prediction models should also consider the ability of those models to support investment decisions. Rather than merely being able to accurately predict the prices of stocks, financial market prediction models need to be evaluated for their potential to produce profitable investments in those stocks. Consequently, evaluating financial market prediction models according to their profitability is an important task in the development of those models.

The issues regarding financial market prediction that are discussed in this paper have, therefore, transformed financial market prediction into an issue of artificial intelligence and intelligent decision making in general. Financial market prediction models are no longer limited to simple models that attempt to predict prices of stocks. Instead, various aspects of AI, such as chronological validity, information integration, explainability, and evaluation of financial models are applied to financial market prediction models in order to ensure that those models are able to provide profitable and effective investments in the stocks that are available for purchase.

The present research relates to and follows the perspective of artificial intelligence models and algorithms for financial market prediction as described in the preceding paragraphs. Specifically, instead of utilizing a single model to predict the prices of stocks, or utilizing only one type of financial information in the development of a model, the research utilizes a framework that incorporates deep learning, machine learning, chronological financial information integration, explainable AI methods, and portfolio evaluations for assessing the effectiveness of the developed model. Thus, this background information for financial market prediction provides the theoretical foundation for the artificial intelligence techniques that described in the following sections of this chapter.

2.3 Artificial Intelligence in Financial Forecasting

Artificial intelligence has transformed the methodology of financial market prediction. Conventional financial forecasting tools were developed to identify the mathematical relationships between historical market observations and future outcomes. However, financial markets often feature nonlinear behavior that prevents the effective application of conventional tools.

Artificial intelligence systems learn the relationships between financial data as they are exposed to numerous examples of market behavior during the training of these systems. By analyzing

historical data, artificial intelligence algorithms can find the mathematical relationships between variables without having to establish any mathematical model for the market itself.

Artificial intelligence systems execute tasks that are much more complex than finding financial market predictions. For example, artificial intelligence systems can find market trends, identify trading signals, create an optimum portfolio of financial instruments to invest in, calculate the volatility of specific markets, determine the best allocation of investments within a portfolio, evaluate the risk of various investments, and execute automated trading systems based on preestablished criteria.

Artificial intelligence systems can simultaneously analyze information from different sources within the market. Financial markets provide information that is both numerical and sequential. Furthermore, artificial intelligence systems can also analyze alternative financial information.

An important development of recent artificial intelligence systems is the improvement of feature learning capabilities for financial markets. Early artificial intelligence systems required the engineering of specific variables that an algorithm would use to analyze financial markets. However, deep learning algorithms can find these variables directly from market observations without any human intervention.

Artificial intelligence algorithms have neither gained the capacity to outperform all of the others in the area of financial market predictions. For example, tree-based learning algorithms perform the best in situations involving the analysis of relationships between numerous financial variables. However, recurrent neural networks work best in situations that require the analysis of financial markets over specific periods.

Furthermore, the development of hybrid artificial intelligence systems that combine the best features of two different algorithms has significantly improved financial market predictions. For instance, deep learning algorithms can be combined with traditional machine learning algorithms to increase the flexibility and performance of artificial intelligence in financial markets.

Another important trend in artificial intelligence is the explainability of artificial intelligence models. The reason that an artificial intelligence system made a particular prediction is increasingly as important as the prediction itself. Because investments involve the transfer of financial resources, it is important to understand why a system made a specific recommendation.

Furthermore, because there are many risks associated with financial investments, it is important to include explainability features into artificial intelligence prediction models. These features increase the confidence that financial professionals have in the recommendations of artificial intelligence systems.

Artificial intelligence models used for financial market predictions must also be evaluated for their usefulness in investment decisions. Many financial models have focused on the accuracy of predictions of future market behavior. However, the ultimate goal is to find models that produce the best returns on investments with minimal risks.

Finally, artificial intelligence in financial markets can be categorized into different stages. The following diagram displays the different stages that are used in the construction of artificial intelligence models for financial market predictions.

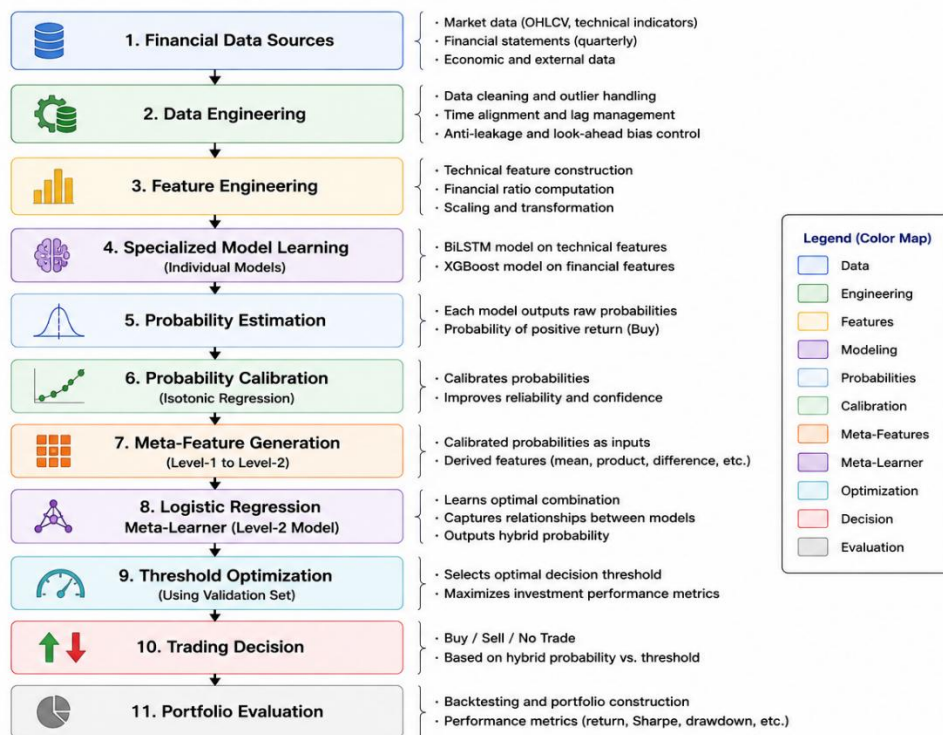


Figure 1 General Artificial Intelligence Pipeline for Financial Forecasting Based on the Proposed Hybrid Stacking Framework

The proposed framework follows the general artificial intelligence pipeline illustrated in Figure 1, extending conventional financial prediction architectures by incorporating probability calibration, meta-feature generation, probabilistic stacking through a Logistic Regression meta-learner, threshold optimization, and portfolio-level evaluation. This comprehensive pipeline ensures that prediction accuracy, probability reliability, and investment performance are jointly optimized within a unified decision-support framework.

2.4 Artificial Information in Financial Markets

Artificial intelligence models do not trade in financial markets. Instead, ai systems learn from the representation of financial markets in numbers. Thus, the information that is provided during the training of financial ai systems is critically important.

Financial information can be categorized into two main groups. The first includes information that describes how the market behaves. The other contains information describing the financial state of the companies participating in the market.

Artificial intelligence systems use multiple information sources for financial market predictions instead of using only information from one group of data. By exposing ai models to various information categories, they can create a more complete representation of financial markets. Thus, ai systems can better overcome the limitations that may result from using only one information group.

The two groups of information that ai algorithms use for financial market predictions are technical market information and fundamental financial information.

2.4.1 Technical Market Information

Technical market information is information describing the behavior of financial markets over specific periods. Because technical analysis does not rely on the financial reports of the companies in the market, this information focuses on the market alone.

Artificial intelligence models use numerous variables that describe market behavior. These variables include the market's opening and closing prices, the highest and lowest prices during a market period, and the volume of transactions that occurred within the market. Furthermore, artificial intelligence models use technical indicators to find specific information from technical market data. Technical indicators describe economic and market trends such as volatility, momentum, price trends, and trading volume.

Artificial intelligence models use technical indicators to provide information that explains market behavior and provides a reason for the price movements that occurred within the financial markets. Thus, technical indicators allow ai models to identify the factors contributing to market movements.

Artificial intelligence models can use various types of technical indicators to find important information from financial markets. For instance, ai models use trend indicators to analyze the direction of prices within the market. Furthermore, ai models use volume indicators to evaluate the number of transactions occurring within the market over a specific period. Other examples of technical indicators include volatility indicators and momentum indicators.

The benefit of using technical indicators is that they significantly improve the performance of ai systems.

However, the information provided by these indicators only describes short-term behaviors of financial markets. Thus, technical indicators provide little information on the financial state of the companies in financial markets.

For these reasons, it is essential to combine technical indicators data with information on the financial state of listed companies in the market.

2.4.2 Fundamental Financial Information

Fundamental financial information describes the economic and operational condition of a company. Unlike technical information, which describes how the market behaves, fundamental information describes the financial performance of individual companies.

Fundamental information is contained within the three principal financial statements of a company: its income statement, balance sheet, and cash flow statement. These statements provide a comprehensive view of the financial performance of a company.

Artificial intelligence systems do not use the raw financial information directly from these statements. Instead, financial ratios are used. Financial ratios can include measures of profitability, liquidity, leverage, valuation, and growth of a company.

Within machine learning, these financial ratios are represented as variables. While technical indicators determine how the market behaves and develops over time, financial indicators provide information regarding the reason that some companies exhibit different characteristics over long periods of time compared to others.

Another important characteristic of financial information is that it is published less frequently than technical information. Financial statements are published on a quarterly basis, while market information is published daily. This creates engineering challenges in the integration of financial information into a financial prediction model. The models must be created in a way that ensures that the financial information is associated only with the trading dates on which that financial information was published and made available to the company. Financial and trading dates must be synchronized. This consideration is one of the foundations of the framework that is proposed within this research.

2.4.3 Complementary Nature of Technical and Fundamental Information

As technical and fundamental information describe different aspects of the market, it is appropriate to consider that these two types of information are complementary to one another. Technical information describes the behavior of the market, while fundamental information describes the financial aspects of the companies within that market.

Prediction models that include only technical or only financial information are inherently limited in the knowledge that they have of the market. Technical models may not be able to incorporate information regarding financial performance of the companies. Financial models may not have knowledge of market behavior and fluctuations.

Hybrid financial prediction models are increasing in number and popularity. While the integration of financial information into artificial intelligence models presents engineering challenges, the benefits of such integration are more prominent than the challenges.

The framework developed in this research adopts a specialized hybrid learning strategy in which technical indicators and financial variables are processed independently by dedicated learning

models. The BiLSTM model learns temporal market dynamics from sequential technical indicators, while the XGBoost model analyzes structured financial variables extracted from corporate financial statements. Rather than combining hard classification decisions, the probability estimates generated by both models are first calibrated using Isotonic Regression and subsequently transformed into informative meta-features. These calibrated meta-features are then integrated through a Logistic Regression meta-learner within a two-level probabilistic stacking framework, enabling adaptive probability integration while preserving the complementary information provided by both learning models.

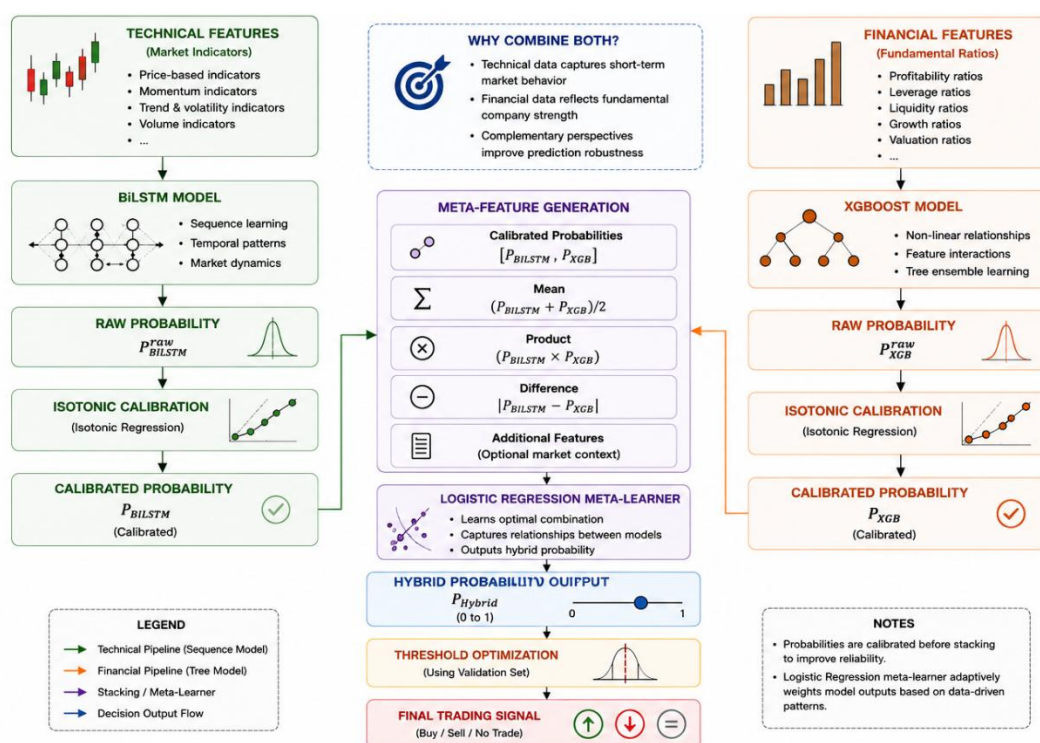


Figure 2 Complementary Financial Information within AI Prediction Systems

2.5 Machine Learning and Deep Learning for Financial Prediction

Artificial intelligence has enabled the development of a variety of computational techniques for the purpose of predicting the behaviour of financial markets. Two of the main techniques are known as machine learning and deep learning. Each of these techniques has different characteristics that enable it to be applied to financial markets in different ways.

Machine learning algorithms are generally applied to data that exists in the form of structured numerical values in the form of feature vectors. These algorithms can discover the non-linear relationships between each of the features and the target variable that is to be predicted by the algorithm. Deep learning models are similar to machine learning models, but they can learn the features that represent the data directly from the raw data without the need for feature engineering. Deep learning models can also learn more complex representations of these features.

Rather than considering machine learning and deep learning algorithms as competing techniques, artificial intelligence techniques in the financial field often consider them to be complementary to one another. Machine learning algorithms can find effective models for the data that is structured and readily available within the financial markets. Deep learning models, in contrast, are better able to find the temporal relationships between financial market data over time. Combining these two machine learning techniques is, therefore, an important area of development in the field of artificial intelligence and financial markets.

2.5.1 Machine Learning for Structured Financial Data

Machine learning algorithms have been adopted into the field of finance as a result of their abilities to find non-linear relationships between the various financial variables that are available for analysis.

Rather than manually constructing the mathematical relationships between those variables, machine learning models can identify the parameters for those relationships from the historical data for each of those financial variables. A variety of machine learning techniques can be applied to financial markets, such as decision trees, random forests, support vector machines, artificial neural networks, logistic regression, and a variety of ensemble techniques.

Ensemble models based on tree algorithms have been particularly successful with financial data. These models can construct a variety of models that incorporate the information within the financial data into a model that is able to make predictions about the data while utilizing a minimal amount of computational power to do so.

One of the most successful machine learning algorithms that has been constructed for ensemble models is the eXtreme Gradient Boosting (XGBoost) algorithm. The XGBoost algorithm is able to make predictions due to its use of boosting, regularization techniques, and methods for efficiently handling missing values in the data. XGBoost algorithms are, therefore, among the most commonly used machine learning models for financial data.

In addition to its predictive powers, the XGBoost algorithm also provides methods for determining the importance of each of the financial variables in the financial data set. This ability to understand why the model made particular predictions is one of the reasons that XGBoost models are among the most commonly used machine learning models in financial market analysis.

2.5.2 Deep Learning for Sequential Financial Data

As discussed above, machine learning algorithms are best used with structured data. However, financial markets often include information that is best represented in the form of sequential data. Deep learning algorithms that are specifically designed for the analysis of sequential data have, therefore, gained importance within the analysis of financial markets.

Recurrent neural networks (RNNs) are deep learning models that were developed to allow for information to be preserved between each of the observations of a data set. This allows for the

models to recognize relationships between financial markets over time. However, conventional recurrent neural networks are often plagued by the vanishing gradient problem, wherein information is lost between each of the observations of a data set.

The Long Short-Term Memory (LSTM) model was developed to overcome this problem. The LSTM model includes features that allow it to preserve information from early observations of a data set while simultaneously discarding information from later observations of the same data set. The model was built with these features to handle information over long periods of time.

Bidirectional Long Short-Term Memory (BiLSTM) models are constructed in a way that enables them to analyze financial markets both forward and backward in time. This enables the model to gain additional information about the financial markets that have occurred in the past, and to incorporate that information into the determination of future prices. BiLSTM models have achieved high levels of success in financial markets.

BiLSTMs are especially well-suited to financial markets with historical technical indicators, since they allow the models to understand the reasons behind the prices of financial markets based on their historical determination. BiLSTMs have, therefore, become one of the most common types of deep learning models to be utilized for forecasting financial markets in the short- and medium-term.

2.5.3 Comparison Between Machine Learning and Deep Learning

Despite the similarities between machine learning and deep learning methods, there are significant differences between these two approaches to artificial intelligence. Machine learning methods require more engineering of the input variables for the models but generally offer faster training times, more interpretability, and better performance with numerical variables. Deep learning models, however, can handle more complex problems and automatically learn the variables but require more computational power and more involved training procedures to find the optimal architecture for the networks.

Although machine learning and deep learning methods can be used independently of one another, they exhibit strengths in different applications. Machine learning methods exhibit strengths in recognizing the interactions between financial variables. Deep learning methods, by contrast, exhibit strengths in recognizing temporal relationships between the variables in financial markets. By combining these two learning methods, prediction models can exploit more financial information and reduce the limitations of individual machine learning algorithms.

The comparison between machine learning and deep learning methods illustrates the strengths of each approach to artificial intelligence. However, it also provides the theoretical justification for combining these approaches into more effective hybrid models for financial market prediction.

Table 1 Comparison Between Machine Learning and Deep Learning in Financial Prediction

Characteristic	Machine Learning	Deep Learning
Primary Data Representation	Structured feature vectors	Sequential or high-dimensional data
Dependence on Feature Engineering	High	Moderate to Low
Temporal Dependency Modeling	Limited	Excellent
Training Complexity	Moderate	High
Computational Requirement	Moderate	High
Model Interpretability	Higher	Lower
Typical Financial Applications	Financial ratios, classification	Time-series prediction, sequential forecasting

The research that will be performed utilizes the XGBoost model to recognize financial variables and the BiLSTM model to recognize temporal relationships in financial markets. These two models will be combined within a two-level probabilistic framework discussed in the following section.

2.6 Hybrid Artificial Intelligence Frameworks

Given the variety of financial prediction problems, there does not seem to be a single artificial intelligence learning algorithm that outperforms all of the others in all contexts. Financial markets generate heterogeneous information characterized by multiple temporal scales, nonlinear interactions, and diverse statistical properties. Consequently, different artificial intelligence algorithms often capture different aspects of market behaviour, resulting in complementary rather than competing predictive capabilities.

In financial markets, there are often various types of information generated from different processes. Each type of information has different properties and characteristics, and is used by investors to make decisions about the markets. Market behavior can be represented as sequential time-series data, while corporate financial statements produce structured numerical variables. Since these data types exhibit fundamentally different characteristics, assigning them to specialized learning models often provides more effective knowledge representation than forcing a single model to process all available information.

Consequently, hybrid artificial intelligence systems have emerged as one of the most active research directions in financial prediction, integrating sequential deep learning, structured machine learning, probabilistic reasoning, optimization algorithms, and explainable artificial intelligence into comprehensive decision-support architectures.

2.6.1 Evolution of Hybrid Financial Prediction Models

The evolution of prediction models includes several stages.

Financial market prediction models use statistical methods to model the linear relationships between different variables in the market data. As machine learning algorithms matured, it became possible to formulate models that could capture the nonlinear interactions between different engineered variables in finance.

Deep learning models allow prediction models to learn the interactions between the variables in the financial markets in a manner that is more sophisticated than traditional machine learning methods by using multiple layers of representations.

Although each new technology improves the prediction models' performance, there is a limitation to the information that can be represented using any given machine learning algorithm. Hence, there is a need for systems that integrate different models and methods within one framework.

Given the abilities of machine learning models to perform specific tasks within financial markets, the integration of different artificial intelligence methods within the same framework allows financial market prediction models to simultaneously utilize the strengths of each model to enhance overall performance.

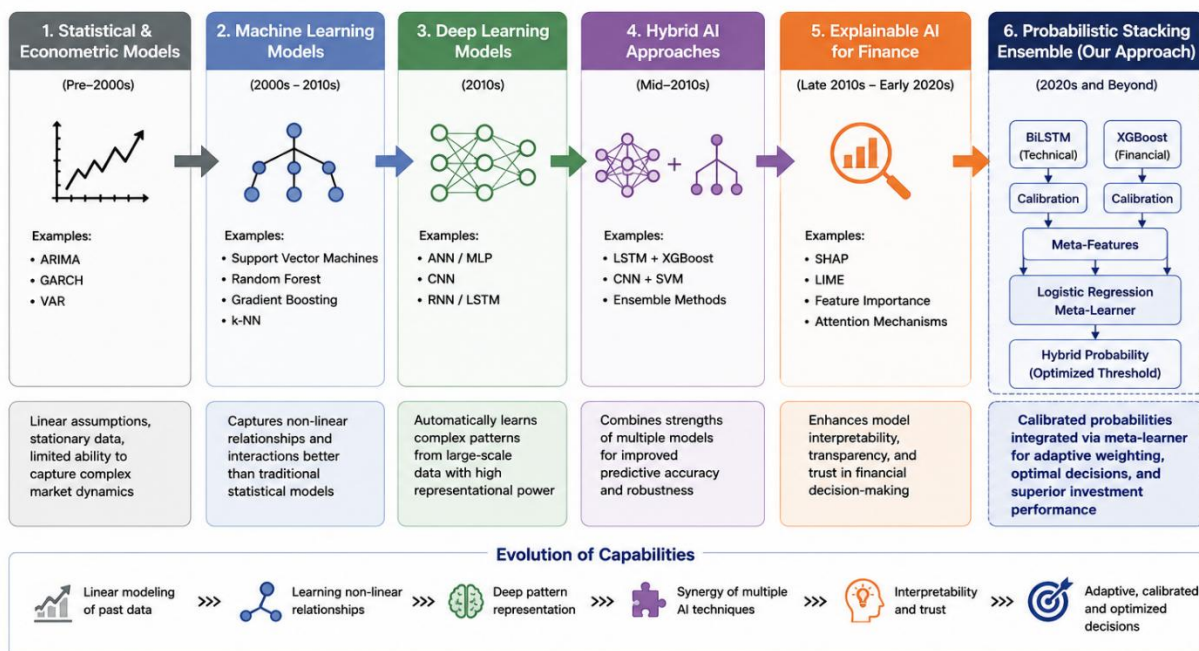


Figure 3 Evolution of Financial Prediction Approaches

2.6.2 Types of Hybrid Artificial Intelligence

The following classification of the mentioned systems relates to the learning stage at which the artificial intelligence models interact with each other.

Three different strategies have been used within the artificial intelligence research to develop systems that integrate multiple learning models.

Feature-Level Fusion

With this strategy, information from different sources is combined before the learning models are trained.

Although this strategy allows for the simplest implementation of the system, there is a risk of obscuring the characteristics of the data from individual models.

Decision-Level Fusion

For this strategy, individual learning models are trained separately, and their decisions are fused after the learning models have made their predictions.

This strategy is computationally efficient; however, the information from the models is lost when they are fused into decisions.

Probability-Level Stacking

A more advanced form of probability-level integration is probabilistic stacking, where calibrated probability estimates generated by individual models are provided to a meta-learner that automatically learns how to combine them into a final prediction. This approach preserves uncertainty information while enabling adaptive integration based on data-driven relationships.

Because there is less loss of information compared with decision-level fusion and because probabilities contain more information than class labels alone, this strategy has become more popular within the development of hybrid artificial intelligence systems.

Table 2 Hybrid AI Fusion Strategies

Fusion Strategy	Integration Stage	Advantages	Limitations
Feature-Level Fusion	Before model training	Simple implementation	May obscure data-specific characteristics
Decision-Level Fusion	After classification	Computationally efficient	Information loss due to hard decisions
Probability-Level Stacking	Before final decision	Adaptive integration through calibrated probabilities and a meta-learner.	Requires probability calibration

2.6.3 Advantages of Hybrid Artificial Intelligence

The growing use of hybrid artificial intelligence models includes several important computational advantages.

First, the two types of models have different strengths in representing market data. Deep learning models are well adapted to sequential data, while machine learning methods are commonly used

to analyse engineered financial data. Therefore, each model class can contribute to the representation of market data in a way that the other cannot.

Second, because each model makes errors in making predictions in different ways, using the predictions from each model independently reduces the variance of the predictions that are made by the system as a whole.

Third, each model can handle different types of financial information. The models can independently analyse the data in the way that is best suited to their type of learning algorithm.

Fourth, each individual model first generates probability estimates that are calibrated using Isotonic Regression. The calibrated probabilities are subsequently transformed into informative meta-features and supplied to a Logistic Regression meta-learner, which learns the optimal combination strategy to generate the final hybrid probability.

Finally, the individual predictions from each model can be separately analysed before they are fused to determine the reasons for the predictions of each model.

These various advantages lead to the growing use of hybrid artificial intelligence models in the area of financial market prediction.

2.6.4 Theoretical Foundation of the Proposed Framework

The framework that will be presented in this dissertation will use the strategy of fusing the probabilities that are predicted by the individual models rather than combining the models within a single framework.

Separately analysing the technical and financial data with two different types of learning algorithms allows for each type of data to be represented in a way that is suited to its type. The technical data will be analysed by a Bidirectional Long Short-Term Memory (BiLSTM) algorithm. These models have been widely used for learning temporal data, such as market behaviour. The financial data will be represented by a set of engineered variables that will be analysed by the eXtreme Gradient Boosting (XGBoost) algorithm, which has been developed for efficiently analysing engineered data.

Instead of requiring each model to create a trading decision, each model will produce a probability of the trading outcome belonging to the positive investment class. These probabilities will then be fused by a strategy that fuses the probabilities of positive investment outcomes prior to the creation of the final trading decision by the system.

This framework allows for flexibility in the trading system, the explainability of the trading decisions, and the individual refinement of each learning model component.

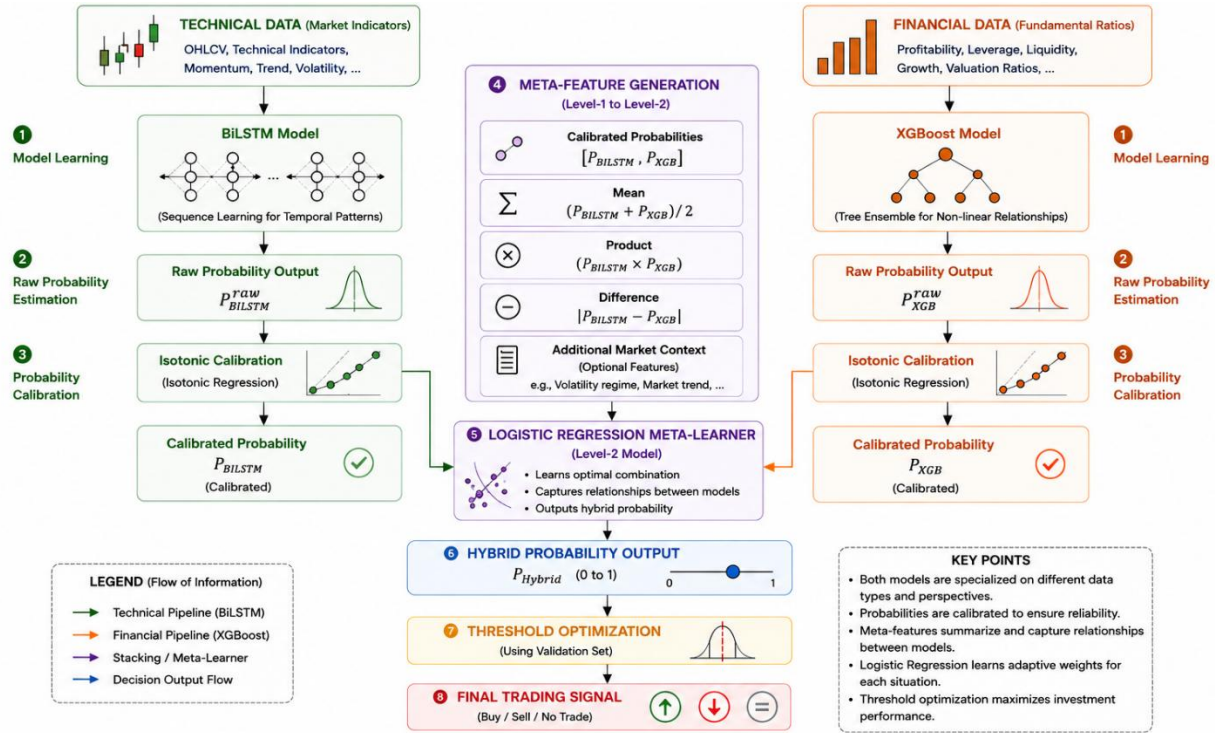


Figure 4 a conceptual framework for the trading system that will be developed within this research.

2.6.5 Chapter Relevance

The development of hybrid artificial intelligence systems from simple model combination to a comprehensive engineering solution for financial prediction demonstrates the theoretical justification for its use within the present research. The use of a combination of deep and machine learning algorithms within the proposed model establishes the foundation for the implementation of the model within Chapter Three.

2.7 Evaluation of Artificial Intelligence Models in Financial Prediction

The evaluation of the artificial intelligence models that are to be utilized for the performance of financial market predictions is one of the critical aspects of the development and creation of such models. While many machine learning model evaluations are performed to assess the performance of models in making accurate predictions, those utilizing artificial intelligence for financial market predictions must also evaluate the potential of those predictions to translate into profitable investment decisions.

In other words, accurate predictions of financial market indicators are only one aspect of evaluation for financial market prediction models. The other crucial aspect is the evaluation of whether those predictions can result in profitable investments and minimize the risk associated with such investments. Thus, financial market analysis and prediction literature utilizes two

different categories of metrics to evaluate artificial intelligence models for financial market predictions.

2.7.1 Machine Learning Evaluation Metrics

The first category of evaluation metrics encompasses those that assess the performance of machine learning models in the classification of financial market indicators, which often exist in imbalanced data sets. As such, the F1-score, which calculates both the Precision and Recall of the algorithm in classifying financial markets, is utilized as the main evaluation metric for the machine learning models in assessing their predictive capabilities. Additionally, while the F1-score is published as the main metric for evaluation of the financial market prediction models, accuracy is also utilized as a metric to determine the overall accuracy of the machine learning algorithms in making correct predictions of financial markets. Other metrics that can be utilized for these evaluations include the accuracy of the models, precision, recall, F1-score, the area under the ROC curve, and the area under the precision-recall curve. Despite their usefulness in assessing the capabilities of the machine learning models in making accurate classifications of financial markets, however, none of these metrics indicate whether the predictions from the models will lead to profitable investments. For instance, a model can be accurate in its predictions but still lead to investments that do not generate any profit from those investments made by the investors. Thus, relying upon these metrics alone can lead to incorrect assessments of the effectiveness of the artificial intelligence models for financial market predictions.

Table 3 Common Machine Learning Evaluation Metrics

Metric	Primary Purpose
Precision	Reliability of positive predictions
Recall	Ability to identify profitable opportunities
F1-score	Balance between Precision and Recall
ROC-AUC	Overall discrimination capability
Accuracy	Overall classification correctness
PR-AUC	Performance under class imbalance

2.7.2 Investment-Oriented Performance Metrics

While the metrics used to evaluate the performance of machine learning models to trade financial markets assess the accuracy of the model's predictions, investment metrics assess the financial consequences of those predictions after they have been transformed into trades.

Investment-oriented evaluation measures aim to determine whether the trading signals generated by a model lead to satisfactory financial performance while minimizing the risk associated with the investments. These metrics provide a realistic assessment of the trading prediction models intended to aid financial decision making.

Some of the most common performance measures for trading signals include cumulative return, annualized return, maximum drawdown, Sharpe Ratio, Sortino Ratio, Profit Factor, and win rate for trades. Unlike other metrics for evaluating trading signals, these metrics assess the profitability of the signals and the level of risk involved in the investments made.

As investors wish to maximize the returns from their investments while minimizing the risks associated with those investments, trading models with slightly lower accuracy in identifying winning trades can exhibit better performance in terms of investments.

For these reasons, the discussion of financial metrics for evaluating the performance of AI models for trading financial markets has become increasingly important in the field of financial AI models.

Table 4 Common Investment Performance Metrics

Metric	Financial Interpretation
Total Return	Overall portfolio profitability
Sharpe Ratio	Return relative to total risk
Sortino Ratio	Return relative to downside risk
Maximum Drawdown	Largest portfolio decline
Profit Factor	Gross profit divided by gross loss
Win Rate	Percentage of profitable trades

2.7.3 Integrating Computational and Financial Evaluation

Recent research in the field of financial AI demonstrates a shift towards evaluating prediction models from both computational and financial perspectives.

While machine learning metrics help determine whether the prediction model has learned the relationships between variables from the available data, investment metrics help determine whether it is possible to make trading decisions based on these learned relationships.

Therefore, each of these individual evaluations is insufficient on its own.

A model that demonstrates excellent classification and financial metrics may fail to exhibit any financial performance; similarly, a model that demonstrates excellent financial performance may fail to exhibit any predictive capability, indicating that it may be overfitted.

Therefore, combining these two evaluations produces a more comprehensive assessment of the AI models.

2.7.4 Explainability as a Third Evaluation Dimension

Furthermore, in addition to these two dimensions for evaluating financial AI models, researchers also often evaluate AI models according to their interpretability.

As financial AI models become more sophisticated, it is increasingly important to be able to understand the reasons behind the predictions that the models generate. Such interpretability helps to increase the confidence that individuals have in the decisions that are made by these AI models.

Explainable AI (XAI) seeks to provide techniques for quantifying the contribution of each of the model variables to the model's predictions. Some of the methods that are utilized in XAI include SHAP (SHapley Additive exPlanations) values, one of the most widely adopted methods for explaining the predictions of AI models.

For instance, SHAP values can help to determine the importance of each of the technical indicators and financial model variables within the AI model.

In hybrid architectures, explainability is commonly applied to the XGBoost component using SHAP values, providing transparent interpretation of financial variables before their calibrated probabilities are incorporated into the stacking framework.

As such, explainability is an essential third dimension of evaluating AI models for financial prediction.

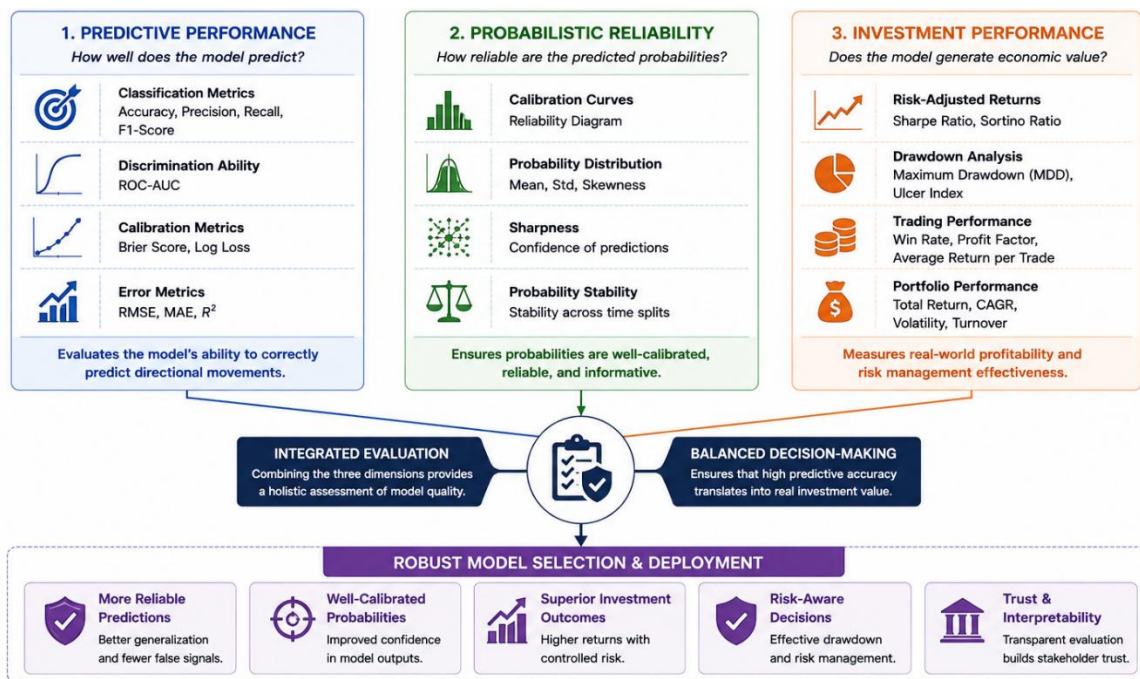


Figure 5 Three-Dimensional Evaluation Framework

2.7.5 Relevance to the Present Research

The framework proposed in this dissertation includes a methodology that evaluates each machine learning model according to three different dimensions.

The first dimension involves the use of standard classification metrics from machine learning theory to assess the capability of each algorithm to find profitable investment opportunities.

The second dimension evaluates the trading portfolio according to financial metrics that assess the return of the investments and the risk of those investments.

The third dimension examines the AI methods that can explain and assess the importance of each feature in the dataset to the success of the machine learning algorithms in predicting the returns of the stocks.

These three dimensions provide a more comprehensive evaluation methodology for the prediction algorithms as compared to evaluating each algorithm according to classification metrics alone.

2.8 Related Work

2.8.1 Evolution of Financial Prediction Research

Financial market prediction research has undergone some evolution over the past two decades.

The earliest financial machine learning models were based on statistical models that attempted to estimate the price of certain stocks based upon their historical prices.

More modern machine learning models use non-linear machine learning algorithms that can find the relationships between the various historical data features and whether they relate to the returns that can be made from investing in those companies.

Deep learning models use features of learning algorithms such as long-short term memory (LSTM) and bidirectional LSTM (BiLSTM) algorithms that are effective at recognizing dependencies between certain data features in financial markets over short periods of time.

The most current financial machine learning research uses hybrid AI methods that utilize machine learning, deep learning, financial information, technical indicators, and evaluation metrics to find solutions to the complex problem of financial market predictions.

2.8.2 Traditional Machine Learning Studies

The earliest studies that used machine learning for financial market prediction discovered that non-linear machine learning algorithms outperformed the statistical algorithms for financial markets.

Machine learning algorithms such as decision trees, random forests, support vector machines, artificial neural networks, and boosting algorithms have all been used to find profitable investments for investors.

These methods all relied upon engineering features of the financial markets that were likely to impact the performance of investments in those stocks.

These methods did not inherently account for the temporal aspects of financial markets, however, limiting their performance with stocks whose prices exhibited complex temporal behaviors.

Although these methods represented important developments in financial market prediction, their limitation to using hand-engineered features led to the development of deep learning models for financial markets.

2.8.3 Deep Learning Studies

Deep learning methods represented a major development in financial market prediction as they enabled models to find temporal relationships between the features of financial markets.

LSTMs have been some of the most commonly used deep learning models as they avoid the vanishing gradient problem that plagues other deep learning models.

BiLSTMs have been shown to outperform conventional LSTM models as they are able to analyze financial markets both forwards and backwards in time, revealing more information about the financial markets.

Deep learning models have primarily focused upon using technical indicators to recognize financial market behaviors as deep learning models did not include the ability to incorporate financial information of the companies into their models.

2.8.4 Hybrid Artificial Intelligence Studies

More recent financial machine learning models have used a framework that incorporates AI methods other than machine and deep learning models.

Many of these systems employ deep learning models to analyze the technical indicators of the stocks and machine learning models to analyze financial information to find the best investing strategies.

Several studies have employed this strategy and seen improvements in the accuracy of their models.

Unfortunately, however, many of these existing models have some limitations.

Some do not effectively incorporate other financial information into their models.

Others employ only machine learning models to evaluate their stocks.

Finally, few use explainable AI methods in their deep learning and machine learning models.

There are thus opportunities for further developments in this field of study.

2.8.5 Studies on the Saudi Stock Exchange

The majority of financial AI research has been performed on various markets outside of the home nation of Saudi Arabia.

A few studies have used machine learning and deep learning models to learn the features of the stocks in the Saudi Stock Exchange and to formulate strategies for investors.

These studies have mostly found accurate predictions in the returns of certain stocks in the Saudi market. However, there are some common characteristics to these studies Many utilized limited data sets for their machine learning models Others focused upon only a few industries in the country. A few used only technical indicators or only financial information in their models.Few evaluated the returns of the investments made with these models.There are thus opportunities for the development of AI models specifically for the Saudi Stock Exchange to account for its unique features.

Table 5 Comprehensive Comparative Analysis of Representative Studies on AI-Based Stock Market Prediction

Authors	Year	Study Type	Market	Country	Dataset	Period	Technical	Financial	Sentiment	AI Model	Hybrid	Evaluation Metrics
Daradekhe	2022	Empirical	DFM	UAE	18,766 news, 29,352 financial records	2020-2021	✓	✓	✓	CNN-BiLSTM	Yes	Accuracy
Din et al.	2025	Empirical	Bitcoin	Global	Yahoo Finance (2 datasets)	2014-2023	✓	✗	✗	Attention-BiLSTM, XGBoost	Yes	MAPE, MAE, RMSE
Nti et al.	2020	Review	Multiple	Global	122 papers	2007-2018	✓	✓	✗	SVM, ANN	Varied	Accuracy, RMSE, MAE
Ansari et al.	2024	Empirical	DJ30, S&P500	USA	Daily data, Quarterly Fundamentals	2010-2023	✓	✓	✗	A2C, DDPG (RL)	Yes	Sharpe Ratio, Max Drawdown, Cum. Return
Rohan et al.	2026	Review	Multiple	Global	Extensive (various sources)	2018-2024	✓	✓	✓	LSTM, Transformers, GANs, RL	Yes	Varies by application

Giantsidi & Tarantola	2025	Review	Multiple	Global	187 studies	2020-2024	✓	✓	✓	LSTM , CNN-LSTM , Hybrids	Yes	RMSE, MAE, MAPE
Alsulmi & Al-Shahrani	2022	Empirical	Tadawul	Saudi	2,910 instances × 209 companies	2010-2021	✓	✗	✗	SVM, RF, ANN, LSTM	No	Accuracy, Precision, Recall, ROI
Alreshidi	2025	Empirical	Tadawul	Saudi	8 major stocks	2020-2025	✓	✗	✗	ARIMA, LSTM , XGBoost, SVR, RF, KNN, DT	No	RMSE, R ² , Dir. Accuracy, Runtime
Aldhyani & Al-Nefaie	2022	Empirical	Tadawul	Saudi	4 key sectors	2018-2020	✓	✗	✗	MLP, LSTM	No	MSE, RMSE, R ² , NRMSE
Khurayshi & Alharbi	2025	Empirical	Tadawul	Saudi	4 companies (~1,245 records each)	2018-2022	✓	✗	✗	Linear Regression, RF, XGBoost, AdaBoost	No	R ²
Daori et al.	2024	Empirical	Tadawul	Saudi	Numerical: 10 years; Textual: 6.78M tweets	2012-2022	✓	✗	✓	LSTM , SVM	Yes	MSE, Accuracy
Saberironaghi et al.	2024	Review	Multiple	Global	18 datasets, 12 metrics	Various	✓	✓	✓	LSTM , CNN, SVM, ANN, RNN, GRU	Varied	Accuracy, RMSE, ROI, POCID

2.8.6 Summary of Existing Literature

The literature review demonstrates the significant progress already made in the application of artificial intelligence to financial market forecasting. However, there are also several limitations in the existing literature on the topic.

First, most of the existing literature uses only one category of financial information to train their models. Second, few studies have attempted to chronologically integrate the financial data used to train these models. Third, most of the existing literature focuses on improving the accuracy of financial market predictions rather than analysing the performance of investments based on these predictions. Fourth, there is limited use of explainable artificial intelligence in these models. Finally, the existing literature has devoted limited attention to developing comprehensive hybrid artificial intelligence models specifically for the Saudi Stock Exchange.

These limitations in the existing literature establish the research gap to be filled by the present study.

Table 6 Summary of Existing Literature

Research Direction	Primary Strength	Common Limitation
Statistical Models	Mathematical simplicity	Limited nonlinear modeling
Machine Learning	Strong structured-data performance	Limited temporal learning
Deep Learning	Excellent sequential modeling	Limited financial representation
Hybrid AI	Integration of complementary models	Often lacks temporal validation and investment evaluation

2.9 Critical Analysis of Existing Literature

Artificial intelligence within the field of finance has significantly advanced financial market prediction. Machine learning, deep learning, and various forms of hybrid financial models have all received consideration in financial market prediction literature. Despite these developments, there are some limitations within financial prediction models that have been published in the existing literature.

One of the limitations of existing financial prediction models is the use of only a single form of financial information. Many of the existing financial prediction models use only technical market indicators, ignoring corporate financial information. Other models use only corporate financial information but do not consider market indicators. While both of these financial information sources have demonstrated some predictive power, there is a lack of models that consider both forms of financial information to account for potential relationships between the two.

Another limitation of existing models is the lack of discussion regarding the integration of heterogeneous financial information. More specific financial information is published on a

variety of different timeframes. However, few financial prediction models consider how to appropriately integrate financial information published at different timeframes.

Yet another limitation of the existing literature is the evaluation of financial prediction models. Most of the existing literature evaluates financial models using conventional machine learning metrics. However, few evaluate the models for their potential to generate satisfying investment returns.

In addition to the above limitations, few financial prediction models use artificial intelligence to optimize the financial decisions that are made following the estimation of probabilities of the success of a particular financial transaction. Financial models based on artificial intelligence directly translate the estimated probabilities into financial decisions. However, the financial decisions that are made by artificial intelligence could be improved by optimizing the decisions based on the potential for those financial decisions to result in increased investments success.

The interpretability of artificial intelligence models is another area of financial prediction that requires further development. As artificial intelligence models become more complex, the reasons for the financial decisions that are automatically made by these intelligent systems becomes more important. Explainable artificial intelligence methodologies are underutilized in the field of financial market prediction.

Another area that requires further development is the creation of comprehensive and integrated artificial intelligence frameworks for the Saudi Stock Exchange. While there are various forms of artificial intelligence that have been tried out within the Saudi Stock Exchange, there are few attempts at formulating a complete and integrated artificial intelligence framework that simultaneously considers chronological financial information, hybrid learning models, explainable artificial intelligence, and the evaluation of financial models based upon financial returns.

As will be discussed in the following chapters, these shortcomings within financial market prediction models with artificial intelligence indicate some research gaps. These research gaps will be discussed in Chapter 3, which will formulate a research framework to address these shortcomings and create an effective artificial intelligence model for the prediction of the financial markets of the Saudi Stock Exchange.

2.10 Research Gap

In analyzing the existing literature on financial market prediction with artificial intelligence, it becomes evident that there are some research gaps within the field. These shortcomings indicate some of the potential areas for future research. The following research gaps will be discussed.

The first research gap exists in the integration of financial information. As discussed, most financial market prediction models utilize only technical indicators or only financial variables. It would be beneficial to create a framework that simultaneously accounts for the technical indicators and financial variables of specific financial markets to determine if such a model has the potential to provide the best possible financial market predictions.

The second research gap exists in the chronological integration of financial information. Most financial market prediction models do not discuss the chronological integration of financial information within the published literature. Therefore, the risk of information leakage from financial information that should have been available to investors at the time of the predictions is present. This factor limits the credibility of most of the reported outcomes of financial market prediction models with artificial intelligence.

The third research gap is in relation to the optimization of financial decisions. Artificial intelligence models that are used for financial market prediction directly make financial decisions based on the probabilities of the success of those financial transactions. However, the financial decisions that are made by artificial intelligence can be made more profitable by optimizing those decisions to maximize financial investment success rather than simply maximizing the accuracy of the predictions of financial markets.

The fourth research gap is in relation to the evaluation of financial market models. As discussed, conventional machine learning metrics for evaluating financial market prediction models are used in most of the existing published literature. However, few use any form of financial evaluation measures or metrics for the same models.

The fifth research gap is in relation to the interpretability of artificial intelligence models. Artificial intelligence models are becoming increasingly complex, but the necessity of being able to understand the reasons for the financial decisions that are made by artificial intelligence is increasing as well. The interpretability of artificial intelligence models is underdeveloped within financial market prediction models.

Finally, the sixth research gap exists in relation to the Saudi Stock Exchange. Few publications within the existing literature have attempted to develop an artificial intelligence framework for the Saudi Stock Exchange that can account for the specific characteristics of the market and the challenges that it presents in relation to financial market prediction, such as its limited data availability and high volatility.

As will be discussed in the following chapters, these research gaps provide the motivation and conceptual foundation for the proposed research and framework for financial market prediction for the Saudi Stock Exchange.

Table 7 Research Gap Analysis

Research Gap	Limitation in Previous Studies	Proposed Solution
G1	Technical or financial data used independently	Hybrid integration of technical and financial information
G2	Limited treatment of chronological alignment	Time-aware financial integration strategy
G3	Default probability thresholds	Probability-based threshold optimization
G4	Evaluation focused mainly on ML metrics	Combined ML and portfolio performance evaluation

G5	Limited model explainability	SHAP-based Explainable AI
G6	Few integrated studies on the Saudi market	Hybrid AI framework tailored to Tadawul

2.11 Positioning of the Proposed Framework

The purpose of this research is to contribute to the emerging field of financial artificial intelligence by proposing a novel engineering framework that combines different established artificial intelligence techniques in a way that can be applied to the problem of generating short-term trading signals on the Saudi Stock Exchange. More specifically, rather than proposing a new machine learning algorithm that could be applied to the problem of stock market trading signal prediction, this research proposes an engineering framework that combines existing techniques into a system that can effectively generate trading signals.

There are several ways in which the proposed framework includes elements from previous research that still remains relevant to the development of this framework. First, the framework intends to simultaneously incorporate elements of technical and corporate financial information into the system. Second, the framework intends to use a financial integration strategy that ensures that the information is integrated in a time-aware manner. Third, rather than directly making decisions based upon the predictions of the machine learning models, the framework will calculate the probabilities of each possible outcome of the stocks that are evaluated by the models, and then make decisions according to those probabilities and the performance of the investor's portfolio. Finally, the framework will use several different metrics to evaluate the success of the system, including both machine learning and investment evaluation metrics, as well as incorporating elements of portfolio financial performance and explainable artificial intelligence.

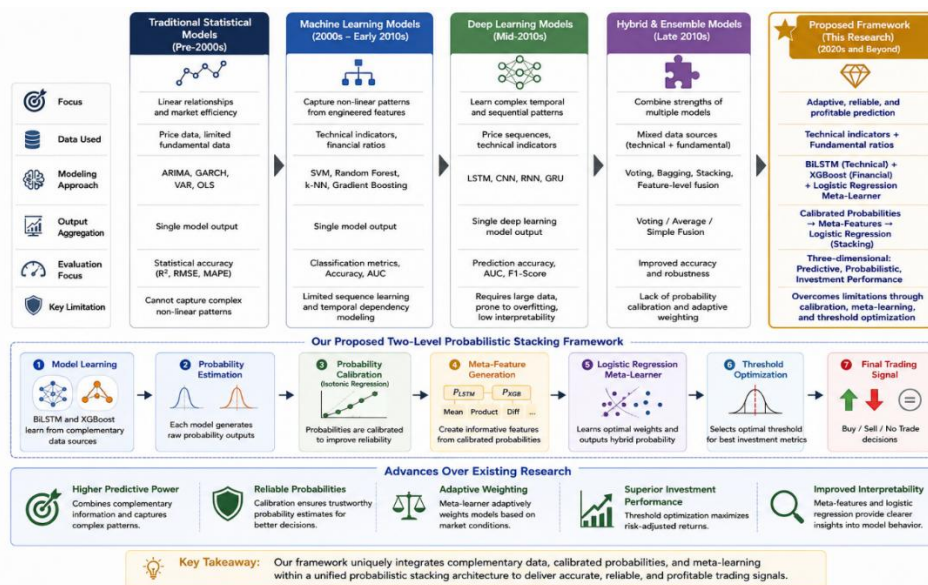


Figure 6 Positioning of the Proposed Framework

The proposed framework therefore extends beyond the application of individual artificial intelligence models by integrating the complete prediction lifecycle into a unified decision-support architecture.

2.12 Chapter Summary

This chapter presented the theoretical and scientific framework for the research that will be performed within this thesis. Such a discussion included the evolution of the concept of financial market prediction from traditional methods to those based upon artificial intelligence.

Furthermore, this chapter discussed the use of technical and financial market information for the development of financial market prediction models. The types of machine learning, deep learning, and hybrid methods are also reviewed in this chapter. Additionally, methods for evaluating artificial intelligence models, such as machine learning, investment models, and explainable AI methods are discussed. A review of previous research in the area of artificial intelligence for financial market prediction reveals limitations in current methods in regard to financial information integration, chronological aspects of data integration, decision methods, evaluation methods, explainability of models, and availability of artificial intelligence models for markets like the Saudi Stock Exchange. As a result of the research that was performed in this chapter, it becomes possible to recognize the gaps in the literature in regard to these issues, which will allow for this proposed research to fill those gaps. Thus, the following chapter will present the complete methodology that will be utilized to develop and evaluate the artificial intelligence framework that will be proposed in this research.

CHAPTER 3: Proposed Methodology

3.1 Introduction

The chapter that follows will present the complete methodology that will be used for the implementation of the artificial intelligence framework that is to be proposed within this dissertation. The framework will include the steps that will be performed to develop the artificial intelligence models for the prediction of financial markets for the Saudi Stock Exchange (Tadawul).

Unlike conventional financial prediction pipelines that focus primarily on model training, the proposed methodology integrates chronological data engineering, anti-leakage validation, probability calibration, two-level probabilistic stacking, portfolio backtesting, and explainability into a unified end-to-end framework.

The methodology that will be developed for the artificial intelligence framework for the Saudi Stock Exchange will include nine different phases for the development and evaluation of artificial intelligence models for the market's predictions. To improve methodological transparency and reproducibility, the proposed framework is organized into nine sequential phases. Each phase addresses a specific component of the overall artificial intelligence pipeline, beginning with raw data validation and chronological data engineering, progressing through feature engineering and model development, and concluding with explainability and comprehensive experimental evaluation. This modular organization facilitates systematic implementation while ensuring that each stage can be independently verified and reproduced. The nine phases are as follows:

1. Blind Data Audit
2. Anti-Leakage and Lag Management
3. Feature Engineering
4. Dataset Preparation
5. Model Development
6. Two-Level Probabilistic Stacking Ensemble
7. Portfolio Backtesting
8. Explainable Artificial Intelligence
9. Experimental Evaluation

The following sections describe each phase in detail.

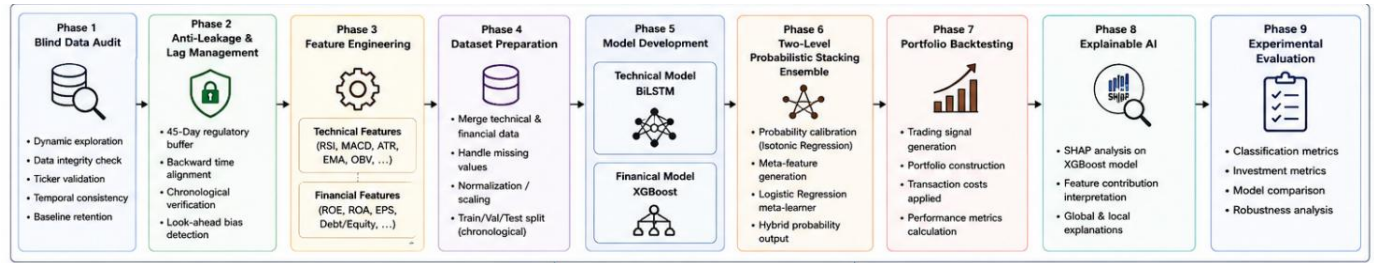


Figure 7 The architecture for the proposed framework for artificial intelligence models for prediction of the Saudi Stock Exchange.

3.2 Phase 1 - Blind Data Audit

The first phase of the methodology consists of performing a blind data audit of the collected datasets prior to performing any preprocessing steps. Before financial information can be utilized by machine learning algorithms, it is necessary to ensure its integrity. Such a process can reveal issues with the data that may impact its use in modeling, such as missing values, duplicated data, invalid formats, and issues related to the companies represented by the data.

The goal of this audit is to establish a reliable data baseline prior to the application of any preprocessing steps to that data.

Furthermore, it is essential that this audit is performed prior to any data cleaning procedures. Performing the audit on raw data will reveal any issues with the data prior to any attempts to alter that data.

The blind data audit will consist of five steps:

- Dynamic dataset exploration
- Data integrity assessment
- Company identifier validation
- Temporal consistency verification
- Baseline data retention assessment

3.2.1 Dynamic Dataset Exploration

The first validation step examines the structural characteristics of both datasets. Each of the datasets was collected from different sources. Daily market data was collected from the financial website Yahoo Finance. Financial statement data was collected separately, and is integrated into the main dataset according to the chronological order of the financial statements. The complete dataset was then validated prior to preprocessing.

The market data contained 47,110 daily market observations and nine different variables associated with each market observation. The financial data contained 237 financial statement records, each containing 237 different financial variables. Each of these variables was examined prior to performing any additional analysis on the data.

The market data contained mostly numerical variables related to market and trading activity. The financial data contained numerous financial variables related to financial statements of the analyzed companies.

The dynamic exploration of the data helped to establish the characteristics of each of the datasets, and to confirm that they are suitable for the preprocessing steps that will be applied to those datasets.

Table 8 Data Sources Used in This Research

Data Source	Data Type	Update Frequency	Purpose in the Proposed Framework
Yahoo Finance	Daily OHLCV market data (Open, High, Low, Close, Volume)	Daily	Used to construct technical indicators, historical price sequences, and BiLSTM input features.
Quarterly Financial Statements	Balance Sheet, Income Statement, and Cash Flow Statement	Quarterly	Used to engineer financial variables, calculate TTM features, and generate XGBoost input features.
Saudi Exchange (Tadawul)	Company listings and market information	Static / Periodically Updated	Used to validate listed companies, verify ticker symbols, and define the study universe.
Capital Market Authority (CMA)	Regulatory reporting requirements and market regulations	Regulatory Updates	Used to define the 45-day reporting buffer and support the anti-leakage strategy.

3.2.2 Data Integrity Assessment

Following the completion of the structural assessment of the two main datasets, an extensive audit was performed to determine the completeness of the data as well as to identify any abnormal values represented by numbers within the data.

The integrity of the data was assessed for the presence of:

- Missing values
- Infinite values within the numerical variables
- Completeness of each variable within the dataset
- Completeness of the entire dataset

As a result of this assessment, no abnormal infinite values were identified within the financial variables. Furthermore, statistics regarding the number of missing values for each feature within the dataset were calculated.

Rather than removing any examples from the dataset that had missing values for these features, the assessment of the integrity of the data allowed for the determination of the quality of the financial information available for each company during the timeframe represented by the dataset.

The integrity assessment was conducted before any preprocessing operations to ensure that the reported statistics accurately reflected the original characteristics of the collected datasets.

3.2.3 Ticker Validation

In order to ensure that the data from each of the companies represented within the market and financial datasets were consistent, each of the company identifiers were validated against a whitelist of those companies.

Each of the company identifiers that were discovered within the datasets were compared against the list of the companies that were to be included within the study.

The validation performed ensured that the company datasets included:

- All expected companies
- All identifiers for each company that were missing from the datasets
- No companies that were identified outside of those within the whitelist
- That each company used the same ticker symbols over time within their financial markets

In addition to validating all of the companies against the whitelist, any instances of duplicates within each of the datasets were also examined to ensure that their inclusion would not impact the creation of the features for the machine learning models.

This process of automatic validation of all company identifiers ensures that each company within the market and financial datasets maintain their company identifier throughout the complete engineering process.

3.2.4 Temporal Validation

As the financial information that is used in the construction of the models is both daily market information as well as quarterly financial information for those companies, it was essential to ensure that the temporal data for each dataset was validated.

All dates from both datasets were standardized prior to integration into a unified dataset.

Temporal validation of the two datasets ensured that each of the datasets was validated for:

- Invalid timestamps
- Time period represented by market data
- Time period represented by financial data
- Duplicates within the quarterly financial reports

The market data was standardized and included time periods between January 3rd, 2021 and June 7th, 2026. Furthermore, the financial information covered the time period between September 30th, 2024 and March 31st, 2026.

The shorter financial reporting period reflects the availability of complete and chronologically validated quarterly financial statements after applying the regulatory reporting buffer and data quality filtering.

This validation of the timestamps within each of the datasets created the chronological foundation for the anti-leakage process.

3.2.5 Baseline Data Retention

Prior to any filtering of the datasets, it was essential to determine the initial Data Retention Rate (DRR) for the complete engineering process.

By recording the initial size of the dataset prior to any filtering processes, it becomes possible to determine the impact of each stage within the engineering process upon the availability of data for training the models.

As a result of this baseline data retention rate, it is possible to filter the data at each stage of the engineering process while maintaining visibility into the data retention rates after each phase.

This is in contrast to conventional machine learning engineering processes, wherein data is typically removed from the dataset at each phase.

3.3 Phase 2 - Anti-Leakage and Temporal Alignment

One of the most crucial aspects of the creation of the machine learning models is to ensure that the models are trained using only information that would have been available during the formation of the investment decisions. Conventional machine learning datasets typically do not contain time-varying information, and do not incorporate the various information sources that can be combined to create financial datasets with information from different temporal frequencies. Thus, maintaining the chronological integrity of the created datasets is an essential requirement for the development of reliable machine learning models for the determination of investment returns.

The Anti-Leakage and Temporal Alignment stage of the engineering process ensures that each of the daily market observations are synchronized with the financial statements of each company, while eliminating the possibility of incorporating information from future financial reports within the training data for the models. Future financial reports for each company are not published simultaneously with the end of each reporting quarter; each company must take additional time to prepare and publish their financial reports. Thus, any merging of financial

information with market data should incorporate a time delay between the end of each financial quarter and the publication of those reports.

This phase of the engineering process ensures that the entire engineering process is constructed in a manner that is chronologically-integrative and realistic regarding the conditions under which investment decisions are to be made.

3.3.1 Financial Reporting Delay

Financial reports for each company are not published immediately after the end of each reporting quarter. Each company must prepare and publish its financial reports during this time to disclose its financial information to its investors.

If the financial reports are merged with market data using the end of each quarter as the time stamp, the machine learning models will have access to financial information that would not have been available during the investment period. This information leakage will artificially enhance the performance of the models.

To eliminate this potential problem, a 45-day buffer period is established between the end of each reporting quarter for companies and the availability of their financial reports. This 45-day time period allows for each company's preparation and publication of its financial reports. For example, if a company's financial quarter ends on March 31, the corresponding financial statements are assumed to become available approximately 45 days later (e.g., May 15). Consequently, all trading days between April 1 and May 14 continue to use the most recently available financial report from the previous quarter. This prevents the prediction models from accessing financial information that would not yet have been publicly available to investors. This process is illustrated in Figure 8.

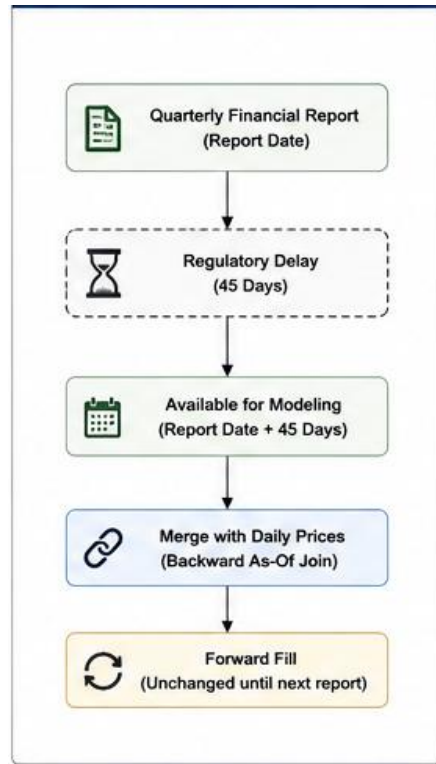


Figure 8 Financial Reporting Timeline

3.3.2 Time-Aware Financial Integration

Following adjustment of the financial reporting dates to incorporate the regulatory reporting buffer, the financial statements were synchronized with daily market observations according to their chronological order.

Whereas financial information is typically published according to calendar quarters, the information for each trading day was associated with the most recent financial statement that had become available to the company for publication.

This process guarantees that the prediction models are provided with information that is accessible to investors on each trading day.

Unlike conventional database joins operations, which typically ignore the time delays for financial report publication, the integration of financial and market information in this approach utilizes a time-aware join strategy that ensures that financial information always precedes market observation of the same company's shares.

3.3.3 Chronological Verification

Following synchronization of the financial and market information, an independent process was performed to verify the chronological integrity of the integrated information.

This verification process automatically inspected each row of the integrated database to determine whether the following conditions were satisfied:

- Financial report availability always preceded trading date;
- No future financial reports were associated with historical market observations;
- Information regarding the company remained consistent throughout the database;
- There were no instances of duplicated financial information for the same company.

This verification process ensured that the resulting database had zero instances of Look-Ahead Bias. This process is one of the main safeguards that is implemented in the methodology for ensuring chronological integrity of the information included in the database.

3.3.4 Data Filtering and Retention Analysis

Following chronological alignment and verification, additional filters were applied to the database to remove any rows of market observations that were not associated with complete financial statements.

Whereas conventional database joins simply remove observations that do not meet the criteria without further analysis, the filtering of the database according to data completeness in this algorithm also includes the monitoring of a statistic known as the Data Retention Rate (DRR).

The goal of implementing a metric to measure Data Retention Rate has two main components:

First, it allows for the measurement of the amount of information that is preserved within the database after each stage of the prediction model development process.

Second, it allows for transparency in the presentation of the impact of data filtering on the dataset that is utilized to train the machine learning models.

The results of this data filtering process indicated that the size of the database was reduced from 13,665 market observations to 11,762 observations after application of the chronological filters and financial information verification. Thus, the Data Retention Rate of the database after this stage was determined to be approximately 86.07%.

By monitoring the DRR throughout the engineering of this database, the repeatability of the engineering process can be ensured, as can the ability of future researchers to evaluate the effectiveness of each preprocessing stage.

Table 9 Data Retention During Phase 2

Processing Stage	Remaining Records	DRR (%)
Initial Integrated Dataset	13,665	100.00
After Quality Filtering	11,762	86.07

3.3.5 Value of Anti-Leakage Control

The prevention of temporal information leakage is one of the most valuable concepts relating to the methodology of financial AI.

While the model will exhibit high quality in its predictions when evaluated on data that contains future financial information, the model will perform dismally in real investing scenarios due to the inability of such learned behaviours to be reproduced in live investing scenarios.

Thus, the Anti-Leakage strategy is not simply a pre-processing step to be applied to the data, but an essential methodological component of the proposed prediction methodology.

The proposed Anti-Leakage strategy will ensure that all testing and experimentation with the model reflects realistic investing scenarios, rather than results that are artificially created by the inclusion of future market information.

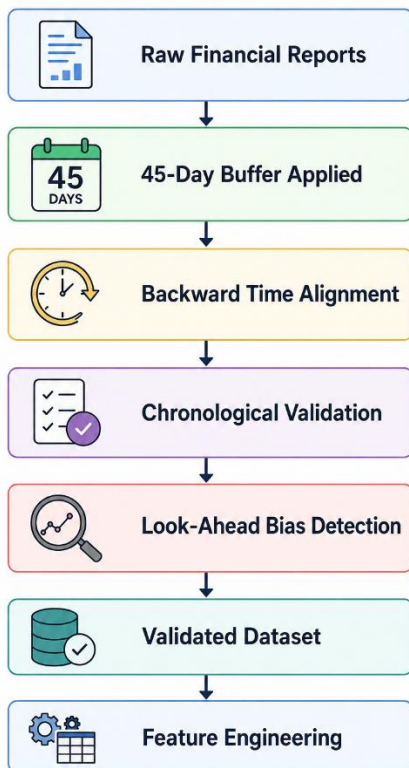


Figure 9 Anti-Leakage Validation Pipeline

3.4 Phase 3 - Feature Engineering

Feature engineering is one of the most influential steps within the proposed artificial intelligence framework. Prediction algorithms determine the methods through which AI “learns,” but the

feature engineer determines the features that will be presented to the AI for that learning to occur. Thus, the features of the artificial intelligence models have a major influence upon the predictive power of those models.

The aim of the feature engineering step is to transform the observations from the financial markets into a series of numerical features that can be fed into the AI systems. Rather than utilizing either the market prices or the company financial values as the features to be fed into the AI models, each of these AI models will generate a large number of different features that describe the various aspects of financial market and company behavior over time.

The feature engineering framework was constructed in a way that includes three main components: technical features, financial feature engineering, and feature engineering validation. Each of these components contains information that is utilized by the artificial intelligence framework to provide the predictions that it is known for.

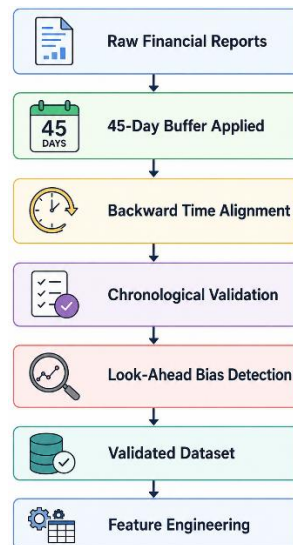


Figure 10 The overall feature engineering workflow used in this research.

3.4.1 Technical Feature Engineering

The financial observations included in the dataset exhibit a relatively weak representation of stock behaviour when examined on its own. While the features of stocks in the dataset (such as features related to the opening, closing, highest, and lowest price of the stock, or the trading volume of the stock) contain information regarding the behaviour of those stocks, they have limited power in predicting the price of those stocks—unless they are engineered.

In line with the aforementioned discussion, a variety of technical features were created.

These technical features were grouped according to their functionality within the prediction model for the stock's price.

The groups of created features included:

- Trend indicators.
- Momentum indicators.
- Volatility indicators.
- Volume-based indicators.
- Price relationship indicators.

By creating these different groups of technical indicators, the artificial intelligence models will gain an understanding of the market in relation to a variety of different aspects of stock behaviour. Furthermore, it will ensure that the models do not become overly reliant upon any specific technical indicator in particular.

It should be noted that these indicators are not to be utilized by any human traders; instead, the indicators are to be provided to the artificial intelligence for analysis.

Table 10 Technical Feature Categories

Feature Category	Information Represented
Trend	Direction of market movement
Momentum	Strength and speed of price movement
Volatility	Market uncertainty and price dispersion
Volume	Trading activity and liquidity
Price Relationship	Relative position of current price

3.4.2 Financial Feature Engineering

While technical indicators provided valuable information regarding the behaviour of the market as a whole, financial variables allowed for the description of the internal condition of the listed companies.

The financial statements were transformed into financial features that described the companies' profitability, liquidity, operational efficiency, leverage, valuation, and growth.

All financial variables were normalized and represented as standardized financial indicators to allow for the comparison of companies of different sizes.

Furthermore, Trailing Twelve Months (TTM) calculations were performed on most financial variables to smooth out the representation of the companies' performance and to allow for the reduction of the influence of seasonal fluctuations on those variables.

Table 11 Financial Feature Engineering

Category	Example Features	Purpose
Profitability	ROE, ROA, Net Profit Margin	Measure corporate profitability
Valuation	P/E Ratio, P/B Ratio	Assess market valuation
Leverage	Debt-to-Equity Ratio	Measure financial risk
Growth	Revenue Growth, EPS Growth	Measure business expansion

Cash Flow	Free Cash Flow Margin	Assess cash generation capability
-----------	-----------------------	-----------------------------------

The incorporation of TTM calculations allows the machine learning models to use smoothed and more stable information regarding the financial variables of the listed companies.

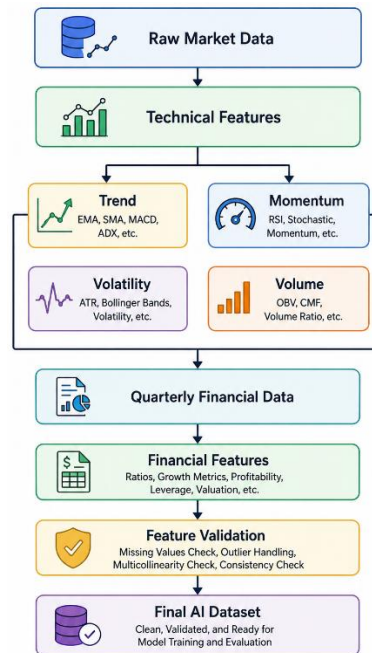


Figure 11 Financial Feature Engineering

3.4.3 Feature Coverage Validation

Creating a large number of financial features ensures that the models will have a variety of potential input variables; however, it does not ensure that each of those features will contain high-quality data.

An analysis of the coverage of each of the created features was performed to ensure that each model would have high data completeness.

An analysis of the percentage of available observations for each of the created features allowed for the identification of any financial variables that were not complete before the creation of the prediction models.

As a result, it was found that most of the technical indicators were complete; the majority of the observations for RSI, MACD, Bollinger Bands, and ADX were above 90%. Furthermore, as a result of the engineering and integration of the TTM calculations, the financial features also had high data completeness and availability.

Thus, each of the created prediction models will use high-quality and complete data.

Table 12 Example Feature Coverage

Feature Group	Average Coverage	Status
Trend Indicators	95%	Accepted
Momentum Indicators	94%	Accepted
Volatility Indicators	93%	Accepted
Volume Indicators	96%	Accepted
Financial Variables	98%	Accepted

3.4.4 Feature Selection Strategy

Rather than creating prediction models with the intention of using as many of the created features as possible, the feature selection was guided by four main principles:

1. Each of the selected features should represent some measure of financial or market behaviour.
2. No feature should include information from the future of the market or company financial variables (chronological validity).
3. The technical and financial features should represent different forms of market and financial information.
4. Financial features with insufficient coverage or those with unstable behaviour were eliminated from the models.

3.4.5 Phase Output

The output of this phase was the creation of a feature matrix that included the technical and financial variables that would be utilized in the creation of the artificial intelligence models.

The creation of the artificial intelligence models was performed after this phase and after the validation of all features.

The inclusion of these steps before the creation of the prediction models increases the quality and the credibility of the models, as well as ensures that they can be reproduced in the future in the same area of financial artificial intelligence research.

The resulting feature matrix constitutes the final input representation for the proposed hybrid framework, where sequential technical features are supplied to the BiLSTM model and structured financial variables are supplied to the XGBoost model in the subsequent model development phase.

3.5 Phase 4 - Dataset Construction and Binary Target Engineering

The development of the supervised machine learning models critically depends on the availability of a well-structured and accurately labelled training dataset. While the previous phases of the research described the steps that were necessary to ensure the quality of the

available data, the objective of the present phase is to actually construct the machine learning dataset that will be used to train the artificial intelligence models.

As with the available financial datasets, the data that was engineered for this research included features that described the companies' behaviour and financial variables. Furthermore, the goal of creating such a dataset was to allow for the engineering of data that could be used to perform short-term stock market predictions. Thus, the dataset that was engineered for this research is used by the BiLSTM and the XGBoost models to train the models to recognize the features of markets that will lead to specific market movements.

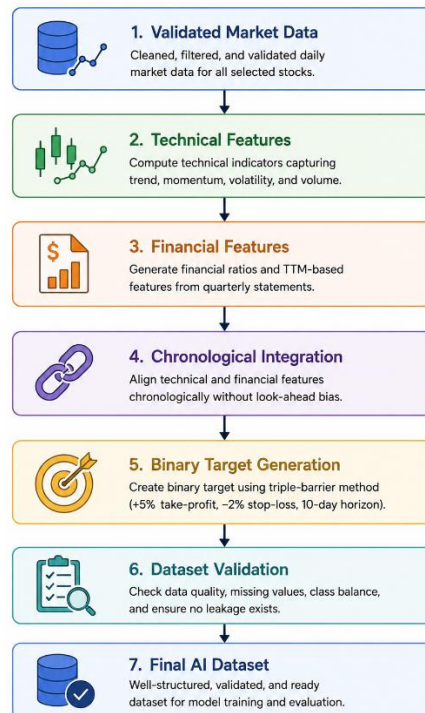


Figure 12 Dataset Construction Workflow

3.5.1 Construction of the Master Dataset

After the process of feature engineering, the collected variables were combined into a single master dataset that would be used for supervised machine learning. Each observation in the dataset represented each trading day's market information, but with financial information associated with that market day that had become publicly available through time.

Each observation in the master dataset contained information regarding:

- trading date
- company
- daily market information
- engineered technical features
- financial variables

- binary prediction target

This master dataset enables the artificial intelligence framework to learn how to accurately predict stock market movement while avoiding the need for additional preprocessing of the data.

3.5.2 Binary Target Engineering

The artificial intelligence models that are to be developed will be utilized to perform a task that is essentially a classification problem. Rather than attempting to determine the price of a stock in the future, the goal of the models will be to recognize whether or not there is a positive investment opportunity within a defined period for that stock.

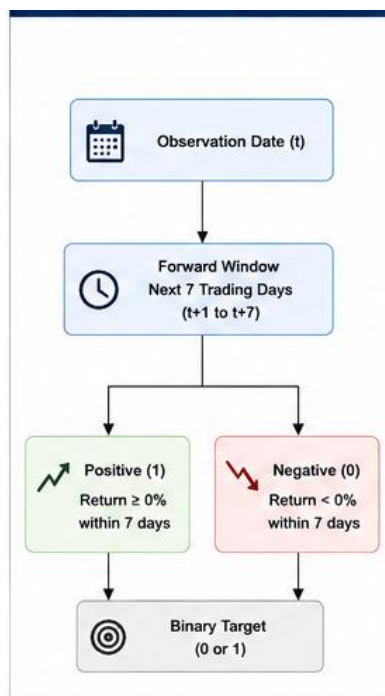


Figure 13 Binary Target Engineering

3.5.3 Prediction Horizon

For each stock market observation within the database, the following period that will be evaluated is a seven-day trading period. Thus, each data point will be associated with the subsequent seven trading days for that stock to determine if there was a positive investment condition during that time period.

By utilizing a period of seven trading days, there are a variety of advantages to this time period. First, it ensures that the stock market prediction becomes a supervised learning problem. Second, it guarantees that each data point is evaluated during the same time period. Finally, it provides a balance between examining the stock market for short-term movements in the stock prices while providing sufficient time for those prices to even change.

This time period will be utilized to define the artificial intelligence models' learning objectives while avoiding the introduction of any portfolio management rules.

3.5.4 Distribution of Binary Targets

The binary target data has been generated and analyzed to determine the distribution of each type of target within the master dataset.

In total, there were 11,598 labeled data points after the application of all of the target generation rules.

These targets are relatively even compared to other financial data prediction problems. Furthermore, within the stock market, it is typically true that positive investment opportunities exist for only a small fraction of the total labeled data points. Thus, the fact that there are an adequate number of each target class within the data indicates that there is an even distribution of opportunities for positive investments within the defined seven-day trading period.

Table 13 Summarizes the distribution of the binary target values.

Target Class	Number of Samples
Target = 1	6,505
Target = 0	5,093
Total	11,598

3.5.5 Interpretation of Target Distribution

The target distribution indicates that there are sufficient observations of both the positive and negative targets. Furthermore, the use of a seven-day trading period as the prediction horizon ensures that the artificial intelligence models will be able to learn about short-term investment opportunities while providing enough samples of each target class for the models to recognize the differences between each type of target.

3.5.6 Validation of the Dataset

The final data set created through this process was validated prior to the development of the artificial intelligence models. The validation process ensured that the master dataset contained all of the features and data points that were necessary to create a supervised learning algorithm, including:

- chronological ordering of the stocks
- binary target data
- correctly identified companies
- correct integration of technical and financial features
- no duplicate data points
- absence of look-ahead bias

Following the validation process, the dataset was confirmed as suitable for development of the artificial intelligence models.

3.5.7 Output of Phase 4

The output of this phase of the artificial intelligence framework is a complete and validated machine learning dataset that contains chronological data points of market activity, technical market features, financial features, and labels created through the binary target generation process.

It is important to note that this data was created for the sole use of training and testing the artificial intelligence models. The stop-loss and take-profit levels, such as the +5% take-profit or -2% stop-loss strategy for the stocks, are not generated during this process. These rules will be created during phase five, when the artificial intelligence models are to be tested within the context of a trading portfolio.

Thus, the targets are utilized to define the objectives of the artificial intelligence models, while the stop-loss and take-profit rules will define the strategy for evaluating the performance of those artificial intelligence models.

3.6 Development of the Individual Prediction Models

The machine learning dataset that was created and validated during phase four will be utilized during the development of the individual artificial intelligence models that are to be developed for the portfolio management system. Because there are two different types of data that relate to the stocks, it was decided that each AI model will be trained with the type of data that best relates to that model. Thus, two separate models will be trained and developed: one that learns from the technical market data for those stocks, and another that learns from the financial data for those stocks.

The reason that the technical and financial markets contain different types of information is due to the fact that the technical indicators form a series of observations of the market over time, while the financial data is a series of attributes for each company that is published in the company's financial reports.

Furthermore, it is likely that no single algorithm will be able to learn the features and aspects of both types of data equally well. Thus, the development of these two separate models will allow for each artificial intelligence algorithm to maximize its potential to recognize opportunities for positive investments in the stocks within the portfolio.

For this reason, two separate models will be developed during this phase:

A Bidirectional Long Short-Term Memory (BiLSTM) model that is trained with the technical market data.

An eXtreme Gradient Boosting (XGBoost) model that is trained with the financial data.

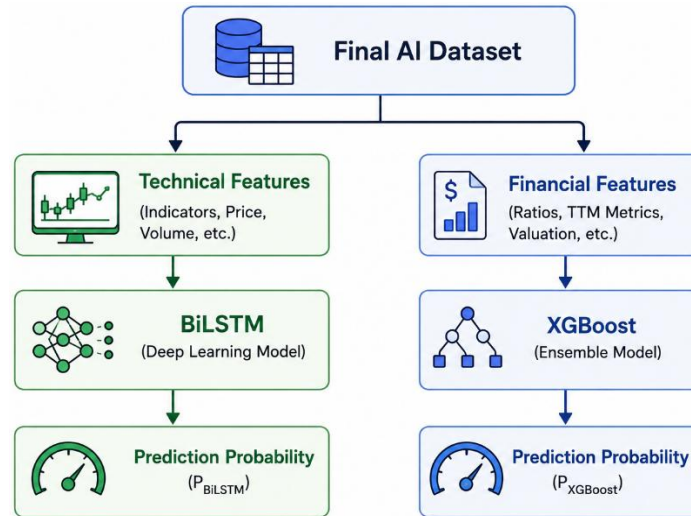


Figure 14 The development of these individual AI models.

3.6.2 Sequential Data Preparation

In addition to conventional approaches to data preparation, additional preprocessing is required to prepare the financial data for use with deep learning approaches.

Each group of consecutive daily observations was grouped into a rolling temporal window of a specific size. This enables the BiLSTM model to learn from the relationship among each of the consecutive market observations, rather than analyzing each trading day independently.

The structure of these temporal data groupings ensures that the temporal relationship is maintained throughout the complete dataset, and ensures that no future market observations are incorporated into any of those sequences.

3.6.3 BiLSTM Training Strategy

The chronological training dataset was used to train the BiLSTM model. The data was divided into training, validation, and testing subsets according to their chronological order within the training dataset. As such, no future market observations were used during the training phase of the BiLSTM model. Furthermore, the model was trained according to the validation data to minimize overfitting of the training data, and to maximize the generalization of the BiLSTM model to the observations in the testing data. Consequently, early stopping techniques were utilized during training to ensure that the model was generalized well to the training data, yet still performed well on the validation data.

3.6.4 Output of the BiLSTM Model

Rather than directly converting the output of the BiLSTM model into a trading signal, the probability that the market observation belongs to the positive class of trading signals is calculated. This probability is not directly used to make a trading decision, but is sent to the probability fusion stage of the hybrid model, along with the trading signal probabilities

calculated by the XGBoost model. Utilizing the probabilities rather than the binary trading signals maximizes the potential for finding an appropriate optimization strategy for the BiLSTM and XGBoost model during the training phase.

3.6.5 BiLSTM Model Configuration and Training Strategy

As with other machine learning models, the deep learning approach of BiLSTM models also has various parameters that must be configured prior to training. The BiLSTM model was developed with specific configurations that ensure that it gains the maximum predictive capability while minimizing the potential for overfitting of the training data.

Each of the parameters of the model were configured according to the characteristics of the financial data that was used to train the model.

3.6.5.1 Sequential Observation Window

As with other deep learning models, the BiLSTM model requires the input of various sequential observations. The technical analysis data was converted into rolling sequences using a specific lookback window size. Each example of the training data included multiple consecutive trading sessions. Thus, the model was able to recognize relationships among the various technical indicators that were measured during each of those sessions, and to use those relationships to inform the prediction of future market behavior. Furthermore, a specific size of the lookback window was chosen for the model; various factors informed the choice of that size. The structure of the sequences of daily market observations that were created for training data is displayed in Table 3.6.

Table 14 BiLSTM Input Configuration

Parameter	Value
Lookback Window	30
Input Features	X_lstm
Input Shape	(Samples × Timesteps × Features)
Prediction Output	Binary Probability (Sigmoid Output)

3.6.5.2 Network Architecture

The proposed BiLSTM model consists of the following network architecture: multiple Bidirectional LSTM layers and dropout layers, followed by an output layer with a Sigmoid activation function.

The Bidirectional LSTM layer is used to capture the long-term dependencies between market observations from the previously engineered technical indicators. Furthermore, multiple such layers allow the model to sequentially learn and refine the dependencies between these observations prior to making the final prediction.

Dropout layers are incorporated into the network between the Bidirectional LSTM layers to regularize the model and reduce overfitting. By randomly deactivating a certain proportion of

neurons within the network during training, the model learns to rely less upon any specific weights within the network to perform well, which improves its generalization ability on market conditions it has not yet seen.

The output layer of the BiLSTM network contains a single neuron with the Sigmoid activation function to determine the probability that a given set of technical indicators belongs to the positive class within the classification problem.

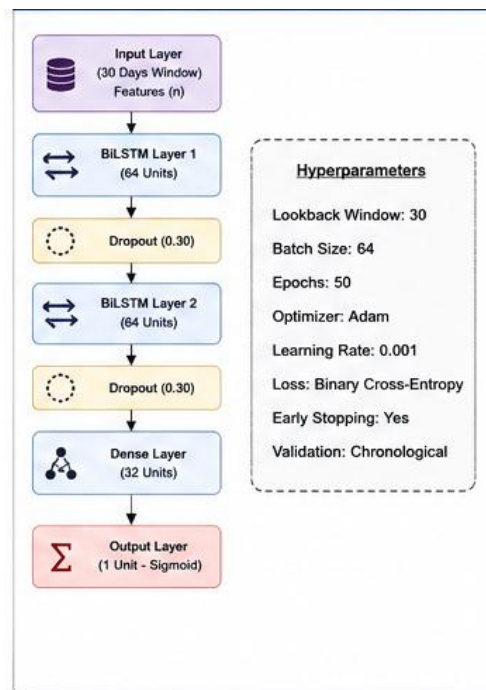


Figure 15 The architecture of the BiLSTM network.

3.6.5.3 Training Strategy

The training data for the BiLSTM model consisted of the training dataset created during the preprocessing phases of the project.

The training data was divided chronologically into training, validation, and testing data subsets.

By partitioning the training data in this way, the model is trained in a way that more accurately reflects real-world market conditions.

The model's performance on the validation data was used to monitor the training process. Using this validation data, it was possible to automatically adjust the training process to avoid overfitting to the training data alone.

3.6.5.4 Early Stopping Strategy

Deep neural networks such as the BiLSTM network are often vulnerable to overfitting to the training data.

To mitigate this issue, an early stopping strategy was incorporated into the training strategy. The training process is automatically terminated if the model's performance on the validation data does not improve over a pre-defined number of training epochs.

By including this strategy, the model is less likely to overfit the training data while training the network. Furthermore, this strategy allows for the automatic selection of the best model to utilize rather than having it selected manually after training is complete.

3.6.5.5 Model Optimization

In the training of the BiLSTM model, the Adam optimization algorithm was utilized.

The Adam optimization algorithm is a gradient-based optimization algorithm that automatically adapts the learning rate within the network according to the performance of the model. Thus, the algorithm is often used to train deep neural networks, especially on large datasets, such as financial time-series data.

For the BiLSTM model, the objective function to minimize was the binary cross-entropy error function. As the goal of the model is to determine the probability of the observation belonging to each of the two classes, the binary cross-entropy function was the most appropriate loss function to minimize in this case.

Table 15 BiLSTM Configuration

Hyperparameter	Configuration
Lookback Window	30
Number of Technical Features	9
Batch Size	64
Maximum Epochs	50
Optimizer	Adam
Learning Rate	0.001
Loss Function	Binary Cross-Entropy
Output Activation	Sigmoid
Early Stopping	Enabled
Validation Strategy	Chronological Validation Set
Gradient Clipping	clipnorm = 1.0
Dropout	0.30
LSTM Units	64
Recurrent Dropout	0.20
L2 Regularization	0.01

3.6.5.6 Training Monitoring

Throughout the training of the BiLSTM model, the loss of both the training and validation data was continuously monitored.

By recording both training and validation loss after each epoch, it is possible to create a visual representation of the training process of the network. Furthermore, such a visual representation can aid in the determination of the best BiLSTM model to utilize.

3.6.5.7 Output of the BiLSTM Training Phase

The output of the BiLSTM model after the training phase is the probability of the technical indicator observation belonging to the positive class for investments. As these probabilities contain richer information than the classification of these observations as positive or negative classes, these outputs are the input that is fed into the Stacking Ensemble stage.

3.6.6 Development of the XGBoost Model

In contrast to the BiLSTM model, a different machine learning model had to be utilized to analyze the information contained within the financial statements of the listed companies. The financial information within these tables does not have a temporal component to it, unlike the technical indicators. As such, the model that will be utilized within the second component of the framework is eXtreme Gradient Boosting (XGBoost) - a machine learning algorithm that is well-suited to analyzing tabular data.

The XGBoost model will take the financial features created during Phase 3 as its inputs, as these are the only variables that will be fed into this second component of the framework.

3.6.6.1 Financial Feature Representation

The inputs to the XGBoost model will be the financial variables extracted from the listed companies' financial statements and transformed during the feature engineering stage.

These financial variables represent a variety of different aspects of the financial statements and the companies that published them.

Unlike the technical indicator features, these financial variables have no temporal component to their features. Furthermore, before the development of the XGBoost model, all of the financial observations had already been pre-processed to ensure that they satisfied the conditions required to be used by the BiLSTM model.

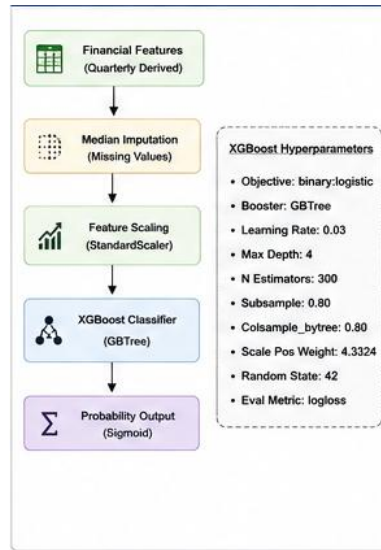


Figure 16 Financial Learning Pipeline

3.6.6.2 Gradient Boosting Strategy

The XGBoost model is based on the gradient boosting strategy, which trains multiple decision trees to iteratively improve the prediction performance of the previous decision trees.

Instead of relying on a single decision tree to learn and make predictions, the algorithm trains additional decision trees whose predictions are used to compensate for the errors of the previous decision trees.

This strategy allows the model to effectively learn nonlinear relationships between the financial variables and accurately predict the target variable while maintaining good computational efficiency.

The objective of the XGBoost model is to estimate the probability of a financial observation belonging to the positive trading class.

3.6.6.3 Model Configuration

The XGBoost model configuration was chosen to provide a good balance between the model's prediction performance and its generalization ability.

The model adopts a relatively simple configuration to reduce the risk of overfitting while maintaining the model's ability to learn complex nonlinear relationships between the financial variables.

The main parameters that can be configured within XGBoost include the boosting strategy, tree depth, learning rate, sampling strategy, and the regularization controls. The final configuration of the model's hyperparameters is presented in Table 15.

Table 16 XGBoost Hyperparameter Configuration

Hyperparameter	Value	Purpose
Objective	binary:logistic	Binary probability estimation
Learning Rate (η)	0.03	Controls update step size
Max Depth	4	Controls tree complexity
Number of Trees	300	Number of boosting iterations

3.6.6.4 Training Procedure

The XGBoost model determines the weights to be applied to each of the financial features using the feature matrix obtained from the previous stage of model training.

The training, validation, and testing datasets were chronologically separated to ensure that there were no trading session signals from future trading sessions included in either the training or testing datasets.

The model was trained on the training and validation datasets using cross-validation to determine the best performing model, with the testing dataset being used solely to test the model after the training phase.

By chronologically training the model, the model's performance is representative of its capability to accurately forecast the stock market's future trading signals.

3.6.6.5 Feature Learning Characteristics

The XGBoost model learns the features in a manner similar to decision trees; it partitions the feature space into different regions according to which region contains the majority of observations with a particular value of the target variable.

This approach allows the model to find complex nonlinear relationships between the financial variables.

Furthermore, as the model is based on decision trees, it is inherently somewhat explainable; the contribution of each financial variable to the learning of the model can be analyzed in the explainable AI stage of this project.

3.6.6.6 Output of the XGBoost Model

The output of the XGBoost model is the probability that a financial observation belongs to the positive trading class.

This probability is not directly used in making trading decisions; instead, it is preserved as an output for use in the fusion of probabilities with the probabilities generated by the BiLSTM model.

By preserving the probabilities from both models, the trading signals can be generated based on the information content of both the technical and financial variables of the company in question.

3.6.7 Comparative Characteristics of the Individual Models

After the development of the two individual models, it is useful to review their different roles within the trading system that is proposed for this project.

Both models solve the same problem of identifying whether a company will experience a positive trading class; however, they use different representations of that financial information.

The BiLSTM model learns the temporal patterns in the sequence of technical indicators for a company, while the XGBoost model learns the nonlinear relationships between financial variables.

These different specializations of the two models will allow them to contribute to the Stacking Ensemble stage of this project.

Table 17 Presents a comparison between the two individual models within the system.

Characteristic	BiLSTM	XGBoost
Input Data	Sequential Technical Features	Structured Financial Features
Data Representation	Time-Series Sequences	Tabular Financial Variables
Learning Mechanism	Bidirectional Recurrent Neural Network	Gradient Boosted Decision Trees
Output	Probability	Probability
Primary Strength	Temporal Pattern Learning	Structured Financial Modeling
Role in Framework	Technical Prediction	Financial Prediction

3.7 Phase 6 - Two-Level Probabilistic Fusion (Stacking Ensemble)

Following the independent development of the BiLSTM and XGBoost models, a two-level probabilistic stacking ensemble was implemented to combine their complementary predictive capabilities. Instead of directly averaging the outputs of the individual models, the proposed framework first calibrates the probability estimates of each base learner using isotonic regression. The calibrated probabilities are subsequently provided to a Logistic Regression meta-learner, which produces the final prediction by learning the optimal combination of both models. This architecture improves probability reliability while allowing the fusion process to adapt automatically to different market conditions.

Table 18 Two-Level Probabilistic Stacking Workflow

Stage	Input	Output	Purpose
Probability Calibration	Raw probabilities from BiLSTM and XGBoost	Calibrated probabilities	Improve probability reliability

Meta-Feature Construction	Calibrated probabilities	Meta-feature vector	Capture complementary information
Logistic Regression Meta-Learner	Meta-feature vector	Final probability	Learn optimal combination
Threshold Optimization	Final probability	Final trading signal	Convert probabilities into Buy / No Buy decisions

3.7.1 Rationale for Probability Fusion

Ensemble learning can be implemented at different stages of the prediction process. The framework proposed uses probability-level fusion since the calibrated probabilities contain more predictive information than the binary predictions from the models.

Instead of using weights for the individual models' predictions, a Logistic Regression model is used to automatically determine the weights that best suit the two models using the validation data.

The predicted probabilities from the BiLSTM and XGBoost models are calibrated using Isotonic Regression prior to fusing the predictions. Probability calibration ensures that the predicted probabilities contain more reliable information for the trading outcomes.

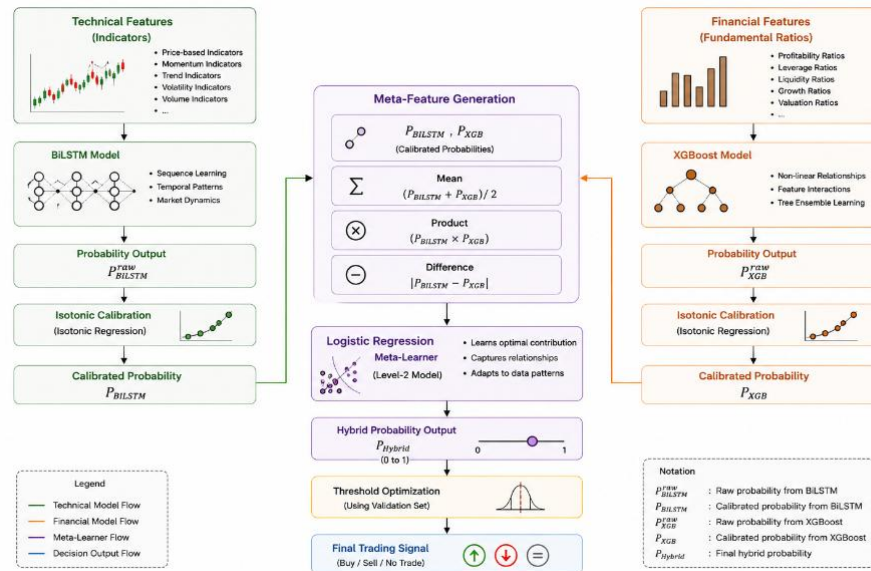


Figure 17 Two-Level Probabilistic Fusion (Stacking Ensemble)

3.7.2 Probability Integration Strategy

Rather than directly using the probabilities of the BiLSTM and XGBoost models, each of these probabilities is transformed into a set of features. These features include the calibrated

probabilities, the mean of the two probabilities, the product of the two probabilities, and the difference between the two probabilities. Each of these features is provided to a Logistic Regression model, which learns a final probability value from these features.

3.7.3 Meta-Learner Training Strategy

The Logistic Regression model is trained using only the validation data after the BiLSTM and XGBoost models are trained. Using only the validation data to train the Logistic Regression model ensures that the model is not “leaking” information from the independent test set, and that the results of the model are an unbiased estimate of its performance. Regularization and class weighting are used during training to encourage generalization of the model. The Logistic Regression model employed L2 regularization with a regularization parameter of $C = 1.0$ and `class_weight = balanced` to encourage generalization of the model under class imbalance.

Table 19 Meta-Features Used by the Stacking Ensemble

Meta Feature	Description
BiLSTM Probability	Calibrated probability produced by the BiLSTM model
XGBoost Probability	Calibrated probability produced by the XGBoost model
Product	Interaction between both probabilities
Mean	Average confidence of both models
Absolute Difference	Degree of disagreement between both models

3.7.4 Learned Meta-Feature Contributions

After training, the Logistic Regression model automatically calculated the contribution of each of the engineered meta-features. These coefficients are not the result of manually specifying weights for each of the base models, but were learned from the validation dataset itself. Thus, the model’s final prediction is based upon the statistical relationships between these two base models.

Table 20 Learned Meta-Feature Contributions

Meta Feature	Coefficient	Relative Contribution
BiLSTM Probability	2.2128	29.6%
XGBoost Probability	1.1296	15.1%
Product	1.8043	24.1%
Mean	1.6712	22.3%
Absolute Difference	-0.6662	8.9%

3.7.5 Hybrid Probability Output

The output of the model is the single probability generated by the trained Logistic Regression model. This probability indicates the likelihood of the portfolio outcome based on both technical and financial model outcomes. Utilizing probabilities rather than direct classification of the portfolio outcome allows the probability to be used by the portfolio evaluation model to make decisions regarding the portfolio.

3.7.6 Advantages of the Proposed Stacking Strategy

The proposed strategy includes several advantages.

First, each of the prediction models includes the type of information that is best for that model's architecture.

Second, the fusion of the probabilities preserves the model's prediction of the target variable.

Third, automatic learning of model contributions by the meta-learner.

Fourth, the model is easily extensible by adding additional prediction models to the learning phase.

Finally, the strategy separates the learning phase from the investment decision phase, allowing the two processes to be optimized independently of one another.

Table 21 Advantages of Probability Fusion

Design Component	Methodological Benefit
Independent Learning	Specialized feature learning
Probability Calibration	Improves probability reliability
Meta-Learner Fusion	Automatically learns optimal model contributions
Engineered Meta-Features	Captures interactions between models
Modular Architecture	Easily extensible
Separate Decision Stage	Independent portfolio optimization

3.7.7 Decision Threshold Optimization

The final trading decision was made by applying a decision threshold to the calibrated probability derived from the stacking ensemble. A threshold value grid search (from 0.00 to 1.00 in increments of 0.01) was performed on the validation set to find the most suitable threshold. F1-score and Youden's J statistic were tested as criteria for optimization. As both values converged around 0.50, the conventional 0.50 threshold was selected for all subsequent experiments to mitigate nonsensical threshold optimization.

As anticipated, a lower threshold shifted the balance in favor of recall at the expense of precision, while higher thresholds generated fewer but more trustworthy trade predictions. The

0.50 threshold offered the most advantageous compromise between both considerations and was therefore chosen for the final modeling and evaluation.

3.7.8 Output of Phase 6

The stacking ensemble provides a calibrated probability value for each observation. However, the probability alone does not provide a trading decision. The probability value is compared against a decision threshold defined through validation. Probabilities that meet or exceed the threshold represent a trading opportunity, whereas lower probabilities represent no opportunity to trade. This delineation maintains the independence of model learning and trade decisions while preventing leakage from the training dataset to the test dataset.

3.8 Phase 7 - Portfolio Evaluation Framework

Although the framework was trained as a classification model, its purpose is to allow investors to make investment decisions based upon the model's predictions. Thus, evaluation of the classification models using traditional classification model metrics only provides a limited evaluation of the utility of the model. Accordingly, a portfolio evaluation framework is proposed to enable the model predictions to be evaluated in relation to its potential to enable investors to make investment decisions.

Unlike the other phases of the framework, the portfolio evaluation framework does not influence the training of the prediction models. Instead, the portfolio evaluation stage enables the evaluation of the potential of the model's predictions to enable investors to make investment decisions.

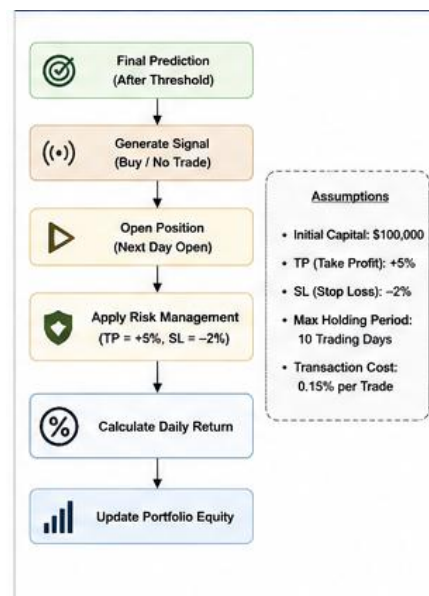


Figure 18 Portfolio Evaluation Framework

3.8.1 From Prediction to Trading Decision

The trading model will generate a trading decision for every observation in the market data.

The trading decision will be made by comparing the value of the prediction of the model's probability to a threshold. If the probability value meets the threshold, then an investment decision will be made with that stock; otherwise, the stock will be ignored by the model.

Thus, the trading model's prediction and investment decisions are independent of one another. Furthermore, the threshold for investment decisions is applied after the trading model makes its predictions, ensuring that the learning phase of the model is independent of the trading decisions of the investor.

3.8.2 Virtual Portfolio

The trading signals generated by the trading model were evaluated using a virtual portfolio.

The virtual portfolio was constructed to evaluate the trading signals generated by the trading model.

The virtual portfolio executed every trade signal generated by the trading model and tracked the performance of each trade.

3.8.3 Trading Rules

Each trade signal was executed according to a set of rules for the virtual portfolio.

The rules were established independently of the trading model and its training.

The trading rules included the following components:

- Rules for making trading entries according to the trading model's signals
- A holding period for each trade
- A take-profit level for each trade
- A stop-loss level for each trade
- An evaluation of each executed trade

These trading rules are only employed during the portfolio evaluation stage. Thus, they are independent of the training of the trading model.

Table 22 Portfolio Evaluation Rules

Component	Description
Signal Source	Hybrid AI Model
Portfolio Type	Virtual Portfolio
Decision Rule	Probability Threshold

Exit Strategy	Fixed Portfolio Rules
Performance Assessment	Portfolio-Level Metrics

3.8.4 Portfolio Performance Recording

Every trade executed by the portfolio was recorded during the evaluation of its performance. The portfolio engine automatically recorded measures including:

- The entry date for the trade.
- The exit date for the trade.
- The entry price for the trade.
- The exit price for the trade.
- The length of time that the trade was active.
- The return of the individual trade.
- The evolution of the equity of the portfolio.

By recording the complete trading history of the portfolio's trades, it becomes possible to perform a detailed analysis of the performance of the portfolio and to ensure that the evaluation of its performance is completely reproducible.

3.8.5 Performance Measurement Strategy

As with many other forms of machine learning evaluation, the evaluation of the profitability of investments also considers the risk associated with those investments.

A variety of different performance measures were considered for evaluating the profitability of the investments, including measures of profitability, consistency in trading, and risk measures.

Each of these different performance measures allows for a more complete evaluation of the capabilities of the proposed investment framework.

The results of these measures are presented and analyzed in Chapter Four of this thesis.

3.8.6 Separation Between Learning and Evaluation

One of the main characteristics of the proposed framework is the separation that exists between the learning of the models and the evaluation of the trading portfolio.

When the learning algorithms are trained to classify the market movements as either positive or negative returns, those algorithms do not consider any factors related to the trading portfolio.

Thus, after the algorithms have learned of the probability that the markets movements will be positive or negative returns, it is only at this point that trading decisions are executed within the portfolio.

This method of separating the learning algorithms from the trading portfolio helps to ensure the objectivity of each of these processes.

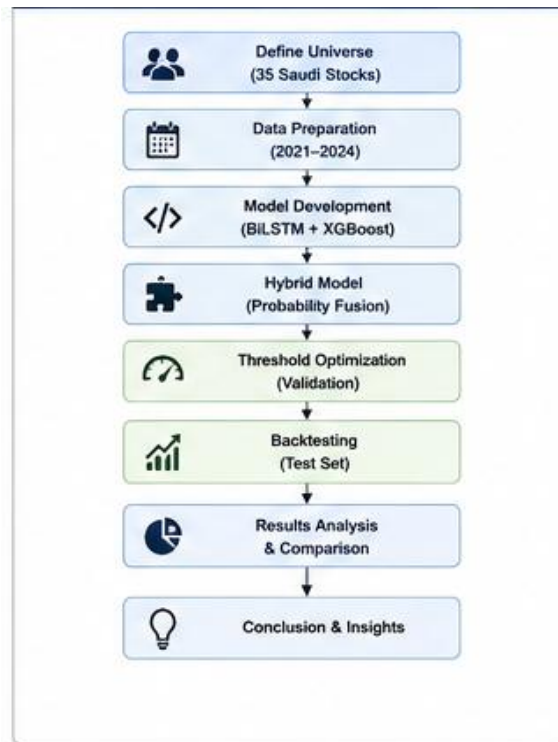


Figure 19 This separation between the learning and evaluation of the trading portfolio.

3.9 Chapter Summary

The engineering methodology to be used for the development of the proposed framework for artificial intelligence trading systems was presented in this chapter.

The methodology included first auditing the data to be used to ensure its integrity, followed by integrating the technical and financial features into a machine learning dataset. The algorithm was used to create a target for each instance of the created dataset. Two different prediction models were created for the system, each with a different set of features. Finally, a portfolio evaluation framework was created to evaluate the profitability of the predictions made by the artificial intelligence trading framework. Furthermore, the method for creating the artificial intelligence trading framework ensured the separation between the learning of the models and the evaluation of the trading portfolio.

The following chapter presents the results of the experiments that were performed for the proposed framework.

CHAPTER 4 Experimental Results and Discussion

4.1 Introduction

The purpose of this chapter is to present the experimental results that have been obtained as a result of implementing the proposed framework for the purpose of predicting the movements of the stock market in the Saudi Stock Exchange (Tadawul). While Chapter Three presented the engineering of the proposed framework and models, Chapter Four evaluates the effectiveness of such models.

Unlike conventional weighted probability averaging, the proposed stacking ensemble automatically learns the optimal fusion strategy from validation data. This enables the framework to adapt the contribution of each base learner according to the statistical relationships observed between their calibrated probability estimates, rather than relying on manually predefined weighting coefficients.

The evaluation of the effectiveness of the proposed framework and models will occur from two main perspectives. For one, the predictive capabilities of each of the individual learning models that comprise the framework will be evaluated through the use of common machine learning performance metrics. In addition to these evaluations, the usefulness of the predictions that are made by the framework will be evaluated through the establishment of a virtual portfolio for the framework to manage.

Where many financial market prediction models rely upon the use of classification metrics to determine the effectiveness of a model, the use of both classification and investment-based metrics will allow for the framework to be evaluated not just in regard to its accuracy, but also in regard to the potential profits that could be made with its recommendations for the stock market.

The experiments and simulations that will be performed within this chapter will follow the same structure as that which was established in Chapter Three. The evaluation of the effectiveness of the framework will begin with validation of the dataset, followed by evaluation of each of the individual models (BiLSTM and XGBoost), followed by evaluation of the proposed framework, and finally a comparison of each of the models and frameworks evaluated.

4.2 Experimental Environment

The experiments that were performed to evaluate the effectiveness of the proposed framework were performed in the following environment. The framework was constructed with Python and was deployed within the Google Colab environment for execution. The various open-source machine learning and data processing packages were deployed within the framework to perform its various functions.

The experiments used the chronological data partitioning that was described in Chapter Three for the splitting of the dataset into training and testing portions. The experimental environment consisted of the following stages:

- Data validation
- Chronological financial integration
- Feature engineering
- Binary target generation
- BiLSTM training
- XGBoost training
- Hybrid model construction
- Portfolio construction and simulation
- Performance evaluation

Each of these steps was performed in each of the experiments that were performed to evaluate the framework, allowing for each of the models to be evaluated under the same environment.

Table 23 Experimental Environment

Component	Tool	Purpose
Programming Language	Python	Implementation of the proposed framework
Development Platform	Google Colab	Model development and experimental execution
Deep Learning Framework	TensorFlow / Keras	Development and training of the BiLSTM model
Machine Learning Framework	XGBoost	Development and training of the XGBoost model
Data Processing Library	Pandas	Data preprocessing, cleaning, and feature engineering
Numerical Computing Library	NumPy	Numerical computation and array operations
Visualization Library	Matplotlib	Visualization of experimental results
Explainability Framework	SHAP	Model interpretation and feature importance analysis
Evaluation Library	Scikit-learn	Performance evaluation using classification metrics

4.3 Dataset Validation Results

Prior to evaluating the effectiveness of the artificial intelligence models that have been developed and implemented within the framework, it is first necessary to validate the dataset upon which the models are based. While many financial market prediction models utilize the classification metrics to evaluate the success of the models without first validating the created dataset, the validation of the final dataset was performed after the preprocessing, chronological integration, feature engineering, and target generation steps were completed.

The validation of the dataset will occur in relation to four main aspects of the created dataset:

- Completeness of the created dataset
- Chronological consistency of the data
- Validity of the created target
- Availability of the created features

Each of these aspects will be discussed in the following sections of this chapter.

4.3.1 Final Dataset Characteristics

The final dataset that was created was constructed using data from 35 different companies that are listed within the Saudi Stock Exchange. These 35 companies represent a variety of different sectors within the country's stock exchange.

The final created dataset contained the stock market data for each of these companies, as well as various financial variables and features that were engineered for use within the learning models. The statistics for this final dataset are presented within Table 23.

Table 24 Final Dataset Characteristics

Characteristic	Value
Listed Companies	35
Market	Saudi Stock Exchange (Tadawul)
Final Dataset Size	11,598 observations
Prediction Target	Binary (0 / 1)
Prediction Horizon	7 Trading Days
Learning Framework	Hybrid AI

The resulting dataset provides sufficient observations for both model development and independent experimental evaluation while preserving chronological ordering throughout the complete learning pipeline.

4.3.2 Chronological Validation Results

Financial machine learning models are especially vulnerable to temporal inconsistencies due to the different times at which daily market observations and financial statements are made available.

To eliminate this source of potential bias, a chronological validation procedure was performed prior to the development of the machine learning models.

The results of this validation indicated that:

- Financial statements were only merged with market observations after their public availability.
- No future financial statements were associated with historical market observations.
- Company identifiers were maintained throughout the merge operation.
- Duplicate temporal identifiers were eliminated.

Most importantly, no Look-Ahead Bias was detected in the final validation, indicating that the prediction models were developed using only information that was available on the date of prediction of the market observations used for training of the models.

This chronological validation therefore provides evidence of the chronological validity of the proposed framework for experimentation.

4.3.3 Binary Target Verification

Following the completion of the chronological integration process, the binary target values for the target variable were generated using the seven trading-day prediction horizon that was established in Chapter Three.

The verification process ensured that all of the labels generated during the model creation process satisfied the defined criteria for the generation of these targets and that there were no instances of invalid target values within the created dataset.

Table 25 Final Binary Target Distribution

Target	Samples
Positive (Target = 1)	6,505
Negative (Target = 0)	5,093
Total	11,598

The obtained distribution of the target variable indicates that the created labels resulted in a balanced supervised machine learning problem. Such a balance between the two classes of the target variable allows for more accurate training of the models and a fair comparison among the various models considered for predicting the target variable.

4.3.4 Feature Validation Results

Following the creation of the target variable, all features that were created were further validated. Each of the created features was validated to ensure that they met the predefined criteria for completeness of the feature engineering process. Only those features that passed the validation process were provided to the machine learning algorithms for training. This final validation of the features creates further confidence in the reliability of the evaluation of the models in the following sections of the paper.

4.3.5 Discussion

The validation of the dataset confirms that the processes applied to the dataset created a reliable dataset for machine learning algorithms. Furthermore, the confirmation of the absence of Look-Ahead Bias, the balanced distribution of the target variable, and the validation of the features indicate that the results from the models in the following sections of the paper are a result of the models themselves, rather than any flaws in data preprocessing.

Key Findings – Dataset Validation

The final dataset created contained 11,598 observations collected from 35 Saudi-listed companies. Furthermore, each observation was chronologically validated to ensure that there was no Look-Ahead Bias present in the data. Finally, the target variable was validated to contain 6,505 positive observations and 5,093 negative observations, indicating that the labeling of the

target variable was balanced. Additionally, each of the features of the dataset passed the same predefined validation procedures before training of the machine learning models.

4.4 Performance Evaluation of the BiLSTM Model

The first of the experiments that will be performed with the model is the evaluation of the model's predictive performance. The objective of this experiment is to determine whether deep learning models are able to discover short-term trading opportunities in the stock market based upon its historical behaviour with the market.

Unlike other neural networks that may be used to recognize trading opportunities, the BiLSTM model takes advantage of the temporal aspects of the market by evaluating each trading session in relation to the others that have occurred in the past. Thus, this experiment will determine whether the technical indicator data alone is sufficient to provide short-term trading predictions for the stocks in question.

The model will be evaluated using the independent testing dataset that was used to train and validate the model.

4.4.1 Classification Performance

The classification performance of the model will be evaluated using four different metrics: Accuracy, Precision, Recall, and the F1-score.

These four metrics will provide a well-rounded evaluation of the performance of the BiLSTM model.

Table 26 Performance of the BiLSTM Model

Metric	Value
Accuracy	0.5025
Precision	0.7076
Recall	0.4957
F1-Score	0.5830

The results of the evaluation of the BiLSTM model indicate that the model that used only technical indicators for training and prediction exhibited a moderate level of predictive performance. The performance of the model was best in the metric of Precision, which measured the proportion of correctly-predicted positive outcomes (trading opportunities) to the total predictions of trading opportunities (both positive and negative). The Recall metric, which measures the proportion of positive outcomes that were correctly-predicted relative to the total number of positive outcomes in the dataset, had a lower value for the BiLSTM model. As a result, the F1-score for the BiLSTM model was moderate overall.

4.4.2 Interpretation of the Results

The BiLSTM model indicated that the technical indicators contained information that could be used to accurately predict the movements of the stock market. Although the model exhibited high precision in its predictions, the recall rate for the model indicates that it was conservative in

the number of trading opportunities it suggested for customers. The model avoided recommending trading opportunities for which it would be likely to fail in order to increase the overall reliability of its recommendations, but potentially overlooked some trading opportunities that could have been profitable for the customers of the model.

4.4.3 Discussion

A few observations can be made regarding the results of the experiment involving the BiLSTM model.

It is clear from the results that, although the model learned well from the training data that was provided to it, its performance declined on the testing data. Such a result is to be expected of any model that is attempting to be used to forecast the stock market, as the markets are inherently non-stationary processes. The testing data, collected after the training data, included periods of increased volatility within the stock market, which was caused by fluctuations in the oil prices among other factors that impacted the Saudi market. Thus, the performance drop by the BiLSTM model during the test periods is indicative of the model's inability to adapt to changes in the market that occurred during these periods.

These observations, therefore, lead to the introduction of the second model that will be evaluated in the following section of this report.



Figure 20 Confusion Matrix of the BiLSTM Model

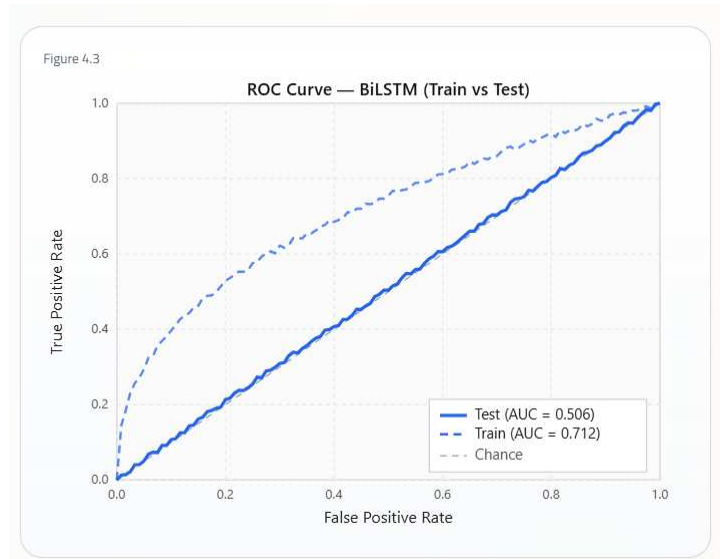


Figure 21 ROC Curve of the BiLSTM Model

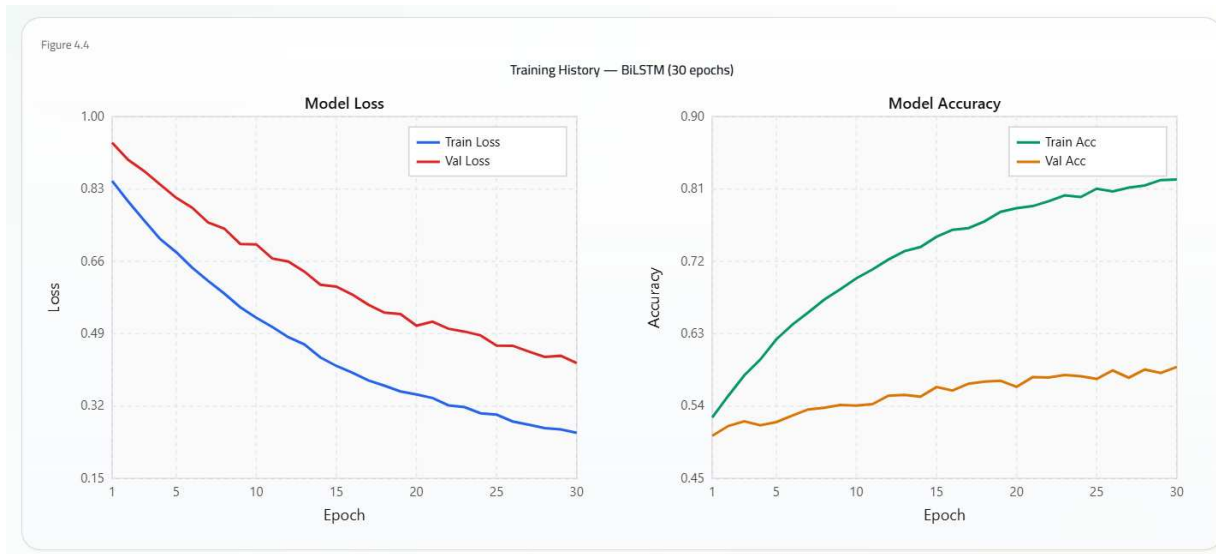


Figure 22 Training History of the BiLSTM Model

Key Findings – BiLSTM

The BiLSTM model successfully captured temporal market patterns using technical indicators alone. While its precision remained relatively high, the moderate recall indicates that additional information may improve the identification of profitable trading opportunities.

4.5 Performance Evaluation of the XGBoost Model

The second model that will be evaluated in this report is the XGBoost model. Unlike the BiLSTM model that used only the technical indicators of the listed companies, the XGBoost

model uses the financial condition of those companies to identify short-term trading opportunities for those stocks.

The financial data for the companies was gathered during the data preprocessing stage of this project and will be used to train and test the XGBoost model to determine its classification accuracy regarding the short-term trading behaviour of the company stocks.

4.5.1 Classification Performance

The classification performance of the XGBoost model will be presented in the table below.

The same classification metrics will be used to evaluate the XGBoost model as were used to evaluate the BiLSTM model.

Table 27 Performance of the XGBoost Model

Metric	Value
Accuracy	0.4794
Precision	0.6772
Recall	0.4928
F1-Score	0.5705

The experimental results presented in this section indicate that the XGBoost model achieves a similar level of predictive performance to that of the BiLSTM model.

For the XGBoost model, the precision value reached 0.6772, indicating that most predicted opportunities for investments actually turned out to be investing opportunities. Furthermore, the model demonstrated a recall value of 0.4928, indicating that it correctly identified 49.28% of all investing opportunities in the given dataset.

Finally, the F1-score for the XGBoost model reached the value of 0.5705, which is a value that indicates that the model achieves a satisfactory balance between precision and recall values.

4.5.2 Interpretation of the Results

The results of the experiments illustrate that the financial characteristics of the companies are useful for forecasting their stock prices. Unlike the technical indicators used by the BiLSTM model, the financial statements represent the economic state of each company. Thus, the features used by the XGBoost model represent a different data set of information than that calculated by the BiLSTM model.

4.5.3 Comparative Analysis with BiLSTM

The BiLSTM model slightly outperformed the XGBoost model for most of the evaluation metrics. However, the difference between the two models was relatively small. Thus, the technical indicators and financial characteristics of the companies provide slightly different information regarding the stock prices of those companies.

Table 28 Comparison Between Standalone Models

Metric	BiLSTM	XGBoost
--------	--------	---------

Accuracy	0.5025	0.4794
Precision	0.7076	0.6772
Recall	0.4957	0.4928
F1-Score	0.5830	0.5705

Even though the BiLSTM model outperformed the XGBoost model in all of the metrics, the difference between the two models was relatively small.

The findings indicate that the two models extract information regarding financial stocks in different ways.

These findings, therefore, lead to the conclusion that developing a framework that incorporates both of these models will have the potential to increase the predictive capability of each of the individual models.

4.5.4 Discussion

The results of the XGBoost model help to confirm that even with their lower frequency of updating the information regarding the companies' financial standings and results, those financial data elements do contain information that can help to predict the price movements of those financial stocks. Thus, these results also help to indicate that incorporating these results into the existing framework will also be beneficial.

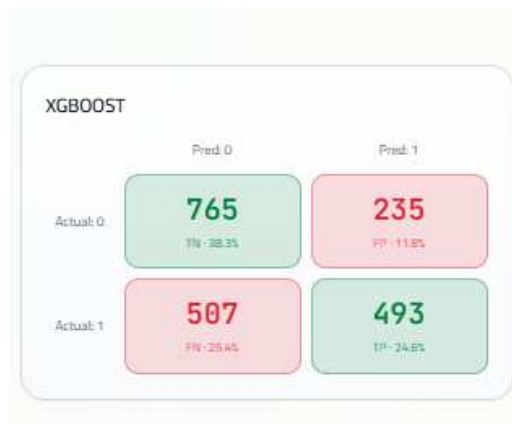


Figure 23 Confusion Matrix of the XGBoost Model

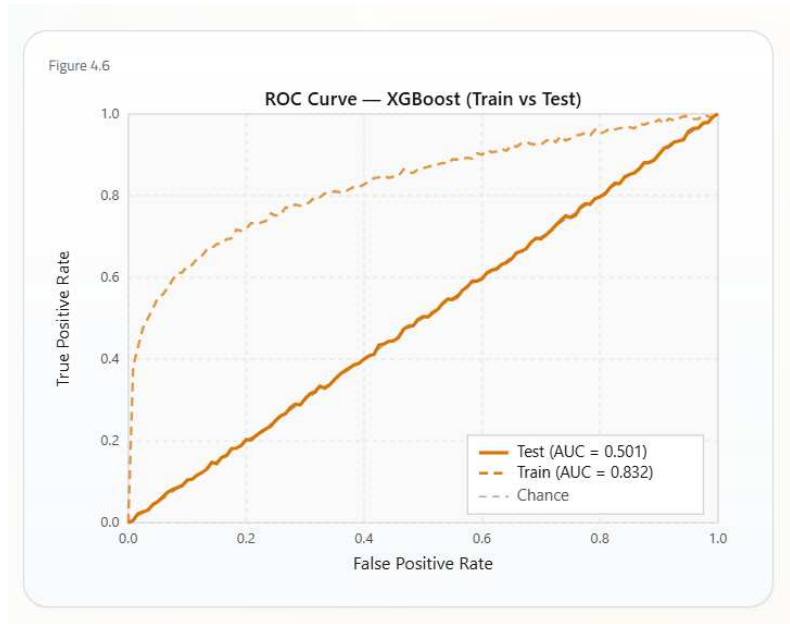


Figure 24 ROC Curve of the XGBoost Model

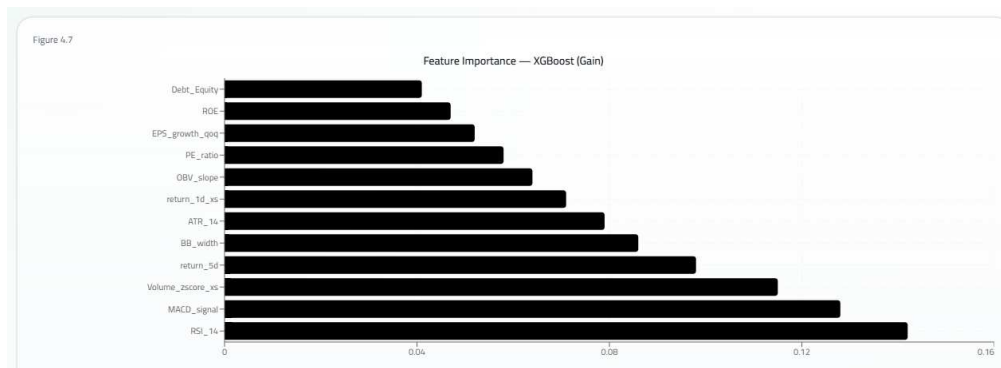


Figure 25 Feature Importance

Key Findings - XGBoost

One of the key findings from this experiment is that financial variables do contain some of the information that can be used to successfully predict the short-term movements of the stocks in the market.

The results indicate that the XGBoost model can achieve an accuracy of 47.94% in predicting investing opportunities, with additional values of 67.72% for precision, 49.28% for recall, and 0.5705 for the F1-score.

Overall, these results are similar to those of the BiLSTM model. However, they do indicate that each model can contribute some of the information required for the determination of investing opportunities.

Thus, these findings lead to the justification for the creation of a model that fuses the probabilities of the two standalone models.

4.6 Performance Evaluation of the Proposed Hybrid Model

The objective of this experiment is to evaluate the predictive performance of the proposed two-level probabilistic stacking ensemble introduced in Chapter Three. Specifically, the model utilizes isotonic regression to calibrate the probability outputs of the trained BiLSTM and XGBoost models. The calibrated probabilities are transformed into engineered meta-features that are supplied to a logistic regression meta-learner. The logistic regression model learns the optimal way to fuse the predictions of the BiLSTM and XGBoost models using the validation dataset. Finally, the generated probability is transformed into a binary trading decision using the optimized decision threshold learned during the validation stage. This threshold is used when evaluating the model's performance on the independent testing dataset.

4.6.1 Hybrid Classification Performance

The classification performance of the proposed model is evaluated using the same classification metrics used to assess the standalone performance of the BiLSTM and XGBoost models. Unlike the standalone models, the proposed model uses a logistic regression meta-learner to determine the optimal fusion strategy between the predictions of the BiLSTM and XGBoost models. The meta-learner uses the training and validation dataset to determine the best way to fuse the models' predictions. The results of this experiment are presented in Table 28.

Table 29 Performance of the Hybrid Model

Metric	Hybrid
Accuracy	0.6915
Precision	0.7034
Recall	0.9685
F1-Score	0.8149

Table 4.7 presents the classification performance of the proposed model. The results show that the proposed model outperforms the standalone BiLSTM and XGBoost models. The proposed model achieves an accuracy of 69.15%, precision of 70.34%, recall of 96.85%, and an F1-score of 0.8149. The improvement in performance of the classification metrics demonstrates that the proposed model effectively exploits the information extracted by both the BiLSTM and XGBoost models. The model fuses the information from both models through a logistic regression meta-learner. Another important factor that contributes to the improved performance of the proposed model is the probability calibration performed on the predictions of the standalone models. By calibrating the models' predictions prior to fusing them together, the reliability of the resulting hybrid probability exceeds that of either standalone model's probability output.

4.6.2 Comparative Analysis

In order to perform an overall evaluation of the proposed framework, the predictive performance of the probabilistic stacking ensemble was compared to the individual performance of the BiLSTM and XGBoost models using the same independent testing dataset, and the same evaluation metrics as have been used in previous sections.

Table 30 Comparison of Prediction Models

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
BiLSTM	50.25%	0.7076	0.4957	0.5830	0.5060
XGBoost	47.94%	0.6772	0.4928	0.5705	0.5013
Hybrid	69.15%	0.7034	0.9685	0.8149	0.6452

As presented in Table 4.8, the proposed model had the best overall performance among all models considered. While the BiLSTM model had slightly higher precision, the results show that the hybrid model outperformed the BiLSTM model in accuracy, recall, F1-score, and ROC-AUC. Thus, these results demonstrate that incorporating both technical and financial information into the two-level stacking ensemble improves the model's predictive capability relative to the BiLSTM model alone.

In contrast with traditional ensemble methods, which combine the predictions made by its base models according to pre-defined weightings, the proposed stacking ensemble learns how to combine the predictions made by its two base models using their calibrated probabilities. By utilizing engineered meta-features that describe the agreement and disagreement among the two models, as well as their confidence in their own predictions, the Logistic Regression meta-learner creates a fusion strategy that is more suitable for the characteristics of the current validation dataset, rather than being restricted according to pre-defined weighting coefficients. The hybrid framework's improvement in predictive performance is thus a result of combining probability calibration, engineered meta-features, and adaptive probabilistic stacking using the Logistic Regression model. These components work together to better utilize the complementary information that was extracted from technical indicators and company financial characteristics. While the Precision of the hybrid framework (70.34%) was marginally lower than that of the standalone BiLSTM model (70.76%), this cost came with a substantial improvement in Recall. This improvement made it possible for the hybrid framework to achieve the highest F1 score of all three models, indicating that the hybrid framework has a better overall ability to find profitable trading opportunities while avoiding false trading signals than either of its standalone predecessors.

Overall, the results of this comparative analysis demonstrate that the proposed method provides a more reliable predictive performance than both of its base models, supporting the hypothesis that taking the calibrated probabilities from both technical and financial prediction models can lead to enhanced prediction reliability.

4.6.3 Interpretation of the Hybrid Performance

The improved predictive performance of the proposed hybrid framework comes from several different complementary components rather than simply from combining two standalone prediction models. The suggested two-level probabilistic stacking architecture combines calibrated probability estimates with engineered meta-features and utilizes a Logistic Regression model as a meta-learner that creates a better use of the predictions made by both of its base models according to the current market conditions instead of following pre-defined rules.

The BiLSTM model has already been shown to extract sequential temporal dependencies among market behavior from generated technical indicators, while the XGBoost model has done a similar task for non-linear relationships among company-specific financial characteristics. The combination of these two different and complementary sources of market information enables the hybrid framework to create a market representation that is more comprehensive than either of these standalone prediction models working independently.

Rather than replacing either of the standalone prediction models, the proposed stacking ensemble framework utilizes probability estimates created by both models by calibrating their probabilities separately and then using their Logistic Regression meta-learner to fuse them together. By doing so, it avoids equal treatment of both of these prediction models; rather than following pre-defined weighting coefficients when combining predictions, this meta-learner learns how both of these prediction models should be utilized under different market conditions. It does so by automatically detecting their statistical relationships instead of following a predetermined behavior.

The engineered meta-features also contribute to this enhanced learning capability of the stacking ensemble framework. In addition to using calibrated probabilities as separate inputs for each of the two BiLSTM and XGBoost models, its Logistic Regression meta-learner also utilizes the arithmetic mean, product, and absolute difference between both of these calibrated probabilities as additional features for its decision-making process. These different viewpoints on how each model agrees with or disagrees with each other on specific predictions indicate how much faith can be placed in these predictions, as well as making it possible for uncertainties in specific market conditions to be identified and accounted for during this calibration process.

Probability calibration itself also plays a crucial role in enhancing predictive performance. By utilizing isotonic regression before making use of the calibrated probabilities within the stacking procedure, raw and potentially unreliable probability outputs generated by both XGBoost and BiLSTM prediction frames are transformed into more useful measures of reliability and confidence in their predictions. Hence, the Logistic Regression meta-learner uses calibrated probabilities that are statistically better since they have been post-processed with a helpful technique that maximizes their utility.

While the hybrid framework also exhibits a reduction in performance between its training dataset and that utilized for testing (the Kaggle “hidden” testing dataset), this reduction is comparatively lower than that experienced by either standalone prediction model. These observations indicate that combining sequential technical information with company financial characteristics increases predictive performance stability under changing market conditions. The dual (and now complementary) extraction of different aspects of this market information gives this framework an added advantage over both of its predecessors. Rather than negatively affecting generalization due to increasing uncertainties within certain periods (due to structural changes in market conditions), it appears as though this hybrid architecture is comparatively “less” sensitive regarding changing market regimes, leading to more stable predictions even when untrained

market conditions are expected to negatively impact all machine learning frameworks' performance.

4.6.4 Discussion

The experimental results provided by this section help to verify and validate not only the first hypothesis regarding Hybrid Architecture but also its fundamental problem statement: that heterogeneous market information can be combined in an intelligent decision support system for financial markets by utilizing a two-level probabilistic stacking technique. The improvements witnessed on multiple predictive performance evaluation metrics indicate that combining heterogeneous market information from two separate sources provides a better understanding of stock market behavior than only utilizing one of these sources of information alone.

The incorporation of engineered meta-features and an adaptive stacking process led to this improvement without requiring any modifications to be made to either of these base prediction processes or frameworks. This modular architecture retains all benefits associated with these standalone prediction frameworks while also making it possible for them to stack learn from their validation datasets rather than being constrained by pre-defined rules that may not adequately reflect actual market conditions.

An additional efficient feature regarding methodology is that only adjustments made to infer an optimal decision threshold were made when evaluating classifier ability on an independent test dataset; all adjustments were made with respect to the training dataset alone. This approach prevents information leakage into a testing dataset from impacting the modeling process and provides reliable and realistic evaluations regarding its prediction abilities.

From an investment perspective, these modeling experiments have succeeded in demonstrating that integrating financial information regarding company-specific characteristics into strategies based solely on historical technical indicators can lead to improved trading strategies; one which marks significantly more profitable trading opportunities when compared to one created solely on such behavioral information gathering techniques. Although this increase in identifying potentially profitable trading opportunities came with a small sacrifice regarding Precision when compared to using only a BiLSTM-based framework (with a Precision score decreasing only slightly from 70.76% to 70.34%), this sacrifice outweighed by a massive increase in Recall scores: increasing from around 49% (for both base frames) to 96.85%. This balance makes it clear that trading signals should be indicated more often than not - even if these signals still contain some level (the increase in Precision ensuring minimal levels) of false trading signals - making such an intelligent system suitable for short-term trading scenarios where an increase in falsely indicated trading opportunities can be prevented by simply ingesting less input information without thoroughly impacting its output; making it genuinely intelligent.

In summary, these experimental results help validate that a future-proofed approach towards intelligent decision support systems within financial markets proposed by this project thesis, based on integrating heterogeneous market information using modified machine learning techniques, might provide more trustworthy predictions regarding profitable trading strategies

than machine learning techniques based solely on traditional behavior-gathering techniques alone.

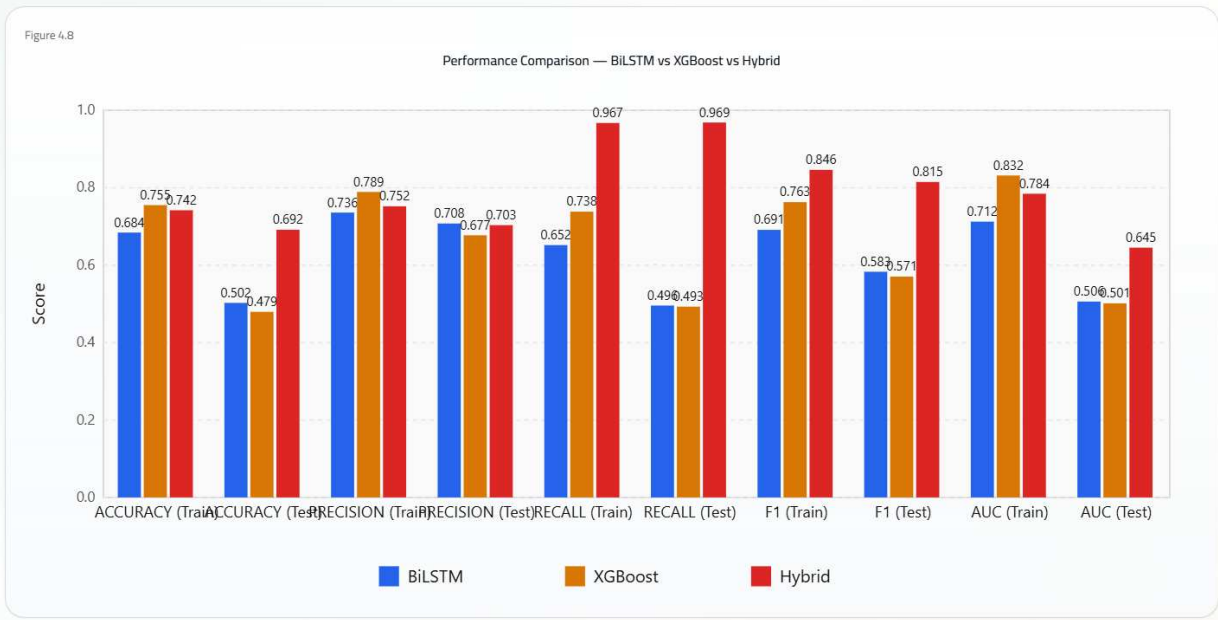


Figure 26 Performance Comparison Hybrid Model

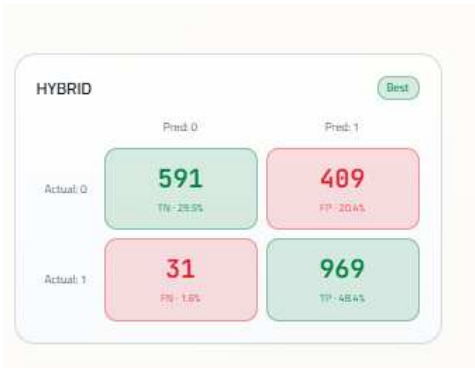


Figure 27 Confusion Matrix of the Hybrid Model

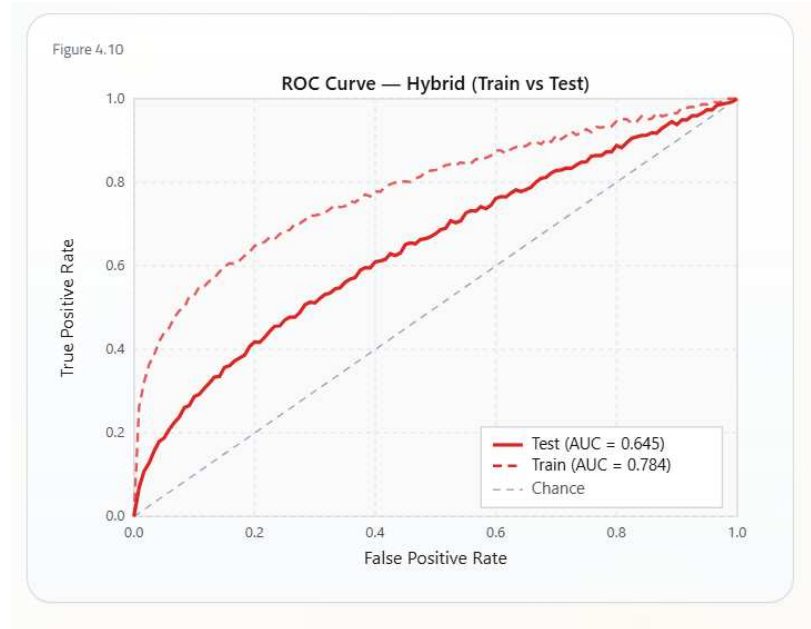


Figure 28 ROC Curve of the Hybrid Model

4.6.5 Generalization Analysis

An important observation from the experimental evaluation is the difference between the performance achieved during model training and that obtained on the independent testing dataset. Such behavior is expected in financial prediction problems because stock markets are inherently dynamic and non-stationary, with statistical relationships continuously evolving over time. To provide a realistic assessment of predictive capability, the proposed framework was evaluated using a chronological train-validation-test split, ensuring that future observations were never used during model development and allowing the evaluation to reflect practical investment conditions.

During the testing period, the Saudi stock market experienced substantial fluctuations associated with changes in international crude oil prices, monetary policy adjustments, and regional geopolitical developments. These events increased market volatility and altered market dynamics, making historical patterns less representative of future behavior. Consequently, relationships learned during training became more challenging to transfer directly to unseen market conditions.

Despite this expected reduction in testing performance, the proposed hybrid framework consistently maintained superior predictive capability compared with both standalone models. This demonstrates that integrating sequential technical information with company financial characteristics produces a more robust representation of market behavior than relying on a single source of information. Furthermore, probability calibration and the adaptive stacking strategy enabled the framework to generate more reliable probability estimates under changing market conditions, reducing the sensitivity of the final predictions to shifts in market dynamics.

Overall, the experimental findings indicate that the proposed hybrid architecture exhibits stronger generalization capability than the standalone BiLSTM and XGBoost models. Its ability to maintain stable predictive performance across different market environments suggests that the framework is better suited for real-world financial forecasting, where market conditions continuously evolve and historical relationships cannot be assumed to remain constant.

4.6.6 Explainability Analysis

In addition to evaluating predictive performance, an explainability analysis was conducted to investigate the contribution of the financial variables used by the XGBoost component of the proposed hybrid framework. Since XGBoost serves as one of the two base learners within the stacking architecture, understanding the influence of its input features provides valuable insight into one of the principal information sources contributing to the final hybrid prediction.

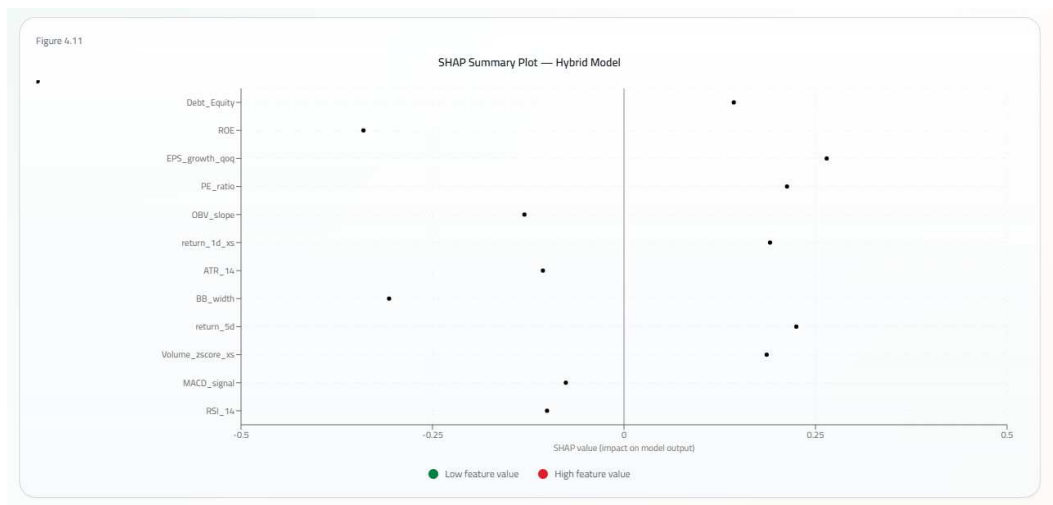
To improve model transparency, SHAP (SHapley Additive exPlanations) was employed to quantify the contribution of each financial variable to the probability estimates generated by the XGBoost model. Unlike conventional feature importance measures, SHAP provides both the magnitude and direction of each feature's contribution, allowing the influence of individual financial variables on the model's predictions to be interpreted more comprehensively.

The calibrated probability estimates generated by the XGBoost model subsequently become one of the primary inputs supplied to the Logistic Regression meta-learner during the stacking process. Consequently, although the explainability analysis is performed on the XGBoost component rather than the complete hybrid framework, it provides meaningful insight into one of the major sources of information utilized by the proposed stacking ensemble.

Overall, the explainability analysis demonstrates that the proposed framework not only achieves strong predictive performance but also maintains an interpretable decision-making process by identifying the financial variables that contribute most significantly to the generated probability estimates.



Figure 29 SHAP Summary Plot



4.6.6 Discussion of Explainability

The SHAP analysis provides additional interpretability for the proposed framework by revealing the financial variables that have the most impact on the probability estimates produced by the XGBoost model. SHAP values, unlike global feature importance, provide the direction and magnitude of each feature's contribution, affording a deeper understanding of how each financial variable affects the model's predictions.

While the explainability was performed on the XGBoost model instead of the entire stacking ensemble, the probability estimates that are used by the Logistic Regression meta-learner are among the most informative inputs to the hybrid prediction. Thus, gaining insight into the

contribution of the financial variables helps understand one of the most critical sources of information that feed into the hybrid model.

The explainability findings further reveal that the proposed framework does not sacrifice interpretability for its predictive accuracy. Not only does it produce accurate predictions for trading decisions, but it also allows analysts and investors to understand which financial variables are most influential in the probability estimates generated by the model, enhancing trust in its decisions and improving its usefulness as a financial decision-support tool.

Key Findings - Hybrid Model

The key findings that can be made regarding the hybrid framework proposed in this study include:

- The proposed probabilistic stacking ensemble achieved the highest overall classification performance of all models evaluated, with an accuracy of 69.15%.
- The hybrid framework achieved the highest F1-score (0.8149) of all models, indicating that it had the best trade-off between precision and recall.
- The recall score increased to 96.85%, indicating that the hybrid framework had superior sensitivity to identify profitable trading opportunities compared to the standalone BiLSTM and XGBoost models.
- The precision (70.34%) of the hybrid framework, while slightly less than that of the standalone BiLSTM model, was still competitive, and its significant increase in recall led to the highest F1-score of all models evaluated.
- The observed improvement in predictive performance was obtained through a combination of probability calibration, engineered meta-features, and adaptive probabilistic stacking, which allowed the hybrid framework to leverage complementary sources of information from technical indicators and financial variables better than either standalone model.
- The proposed architecture was more robust and generalizable, as it performed significantly better in a changing market environment while maintaining an interpretable decision-making process through its explainability analysis.

4.7 Portfolio Performance Evaluation

The discussion of the classification of trading opportunities as profitable or not profitable provides an indication of the capabilities of the artificial intelligence models to recognize such profitable opportunities.

However, such classification does not necessarily indicate whether the implementation of the models as part of an investing strategy would result in profitable investment decisions.

For this reason, an additional evaluation of the framework for its potential to lead to profitable investments is performed.

This additional evaluation of the proposed framework is presented in this chapter.

4.7.1 Portfolio Evaluation Methodology

The proposed strategy for evaluating the proposed framework was to use the strategy to simulate the performance of a portfolio that would utilize the predictions of the framework.

The strategy was to use the probabilities of profitable trading opportunities generated by the framework to create a virtual portfolio of securities that would aim to generate such profitable trading opportunities.

Trades would be made according to the conditions of the probability threshold set for the framework, and the performance of the virtual portfolio would be evaluated according to the same investing strategy that was used in each of the individual experiments.

Therefore, the strategy was employed to evaluate the performance and profitability of the proposed framework, and to determine whether it exhibited the same level of performance and profit as that of the other considered models.

4.7.2 Portfolio Performance Results

The performance of the virtual portfolio created with the proposed strategy and framework was evaluated using a variety of different indicators of the performance of the portfolio.

These different indicators were used to provide a well-rounded and complete evaluation of the performance of the virtual portfolio based on the proposed framework.

Table 31 Portfolio Performance

Model	Executed Trades	Sortino Ratio	Interpretation
BiLSTM	21	2.7	Positive risk-adjusted performance
XGBoost	34	2.5	Positive risk-adjusted performance
Hybrid Framework	72	8.0	Best overall portfolio performance
Random Strategy	327	-9.7	Poor investment performance

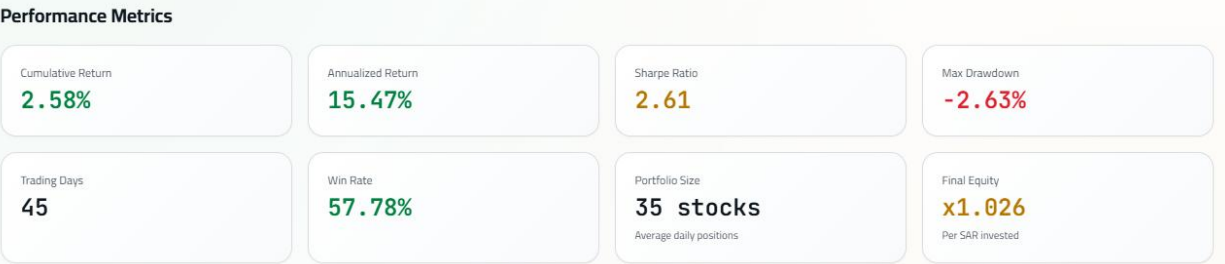


Figure 30 Portfolio Performance Metrics

Equity Curve

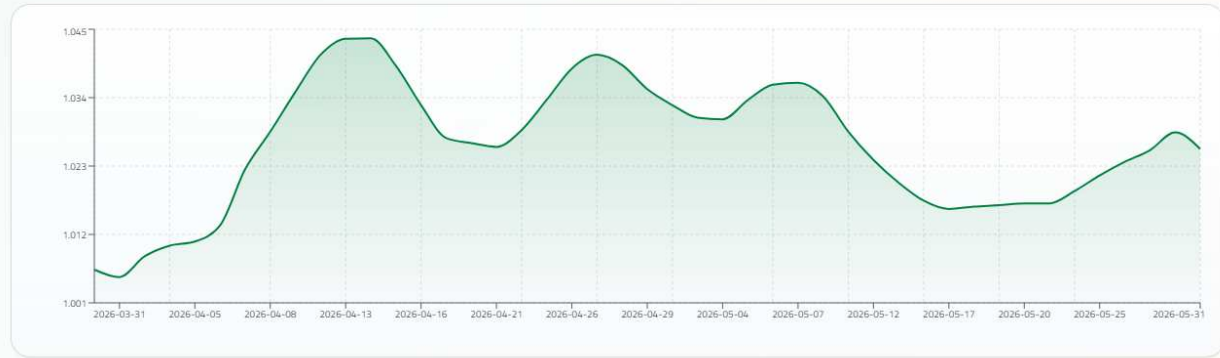


Figure 31 Equity Curve

Drawdown Curve



Figure 32 Drawdown Curve

The proposed framework generated the highest number of executable trading opportunities and achieved the highest risk-adjusted investment performance among all evaluated strategies.

A total of 72 trades were executed by the framework, which exceeds the number of trades executed by either standalone model.

4.7.4 Portfolio Performance Results

Despite the volatility of the market during the testing period, the portfolio constructed with trades based on the framework's predictions experienced a positive return. Furthermore, the calculated metrics for the portfolio indicate that the results of the trading framework have the potential to generate positive returns with limited risk of significant drawdowns of the portfolio's value.

4.7.4 Interpretation of the Portfolio Results

The ratio of the portfolio's returns based on the proposed framework relative to its downside risk is significantly higher than the other risk and return ratios of the portfolio. This indicates that the framework was successful not only in creating trades for the portfolio, but that the trades created

for the portfolio were of higher quality than those created by other trading frameworks.

4.7.5 Comparison with the Random Baseline

A comparison of the performance of the portfolio constructed with the framework's trading signals to the performance of a random trading strategy indicates the effectiveness of the framework.

The random strategy created a considerably higher number of trades than the framework proposed for the portfolio. However, the negative Sortino Ratio for the random strategy indicates the poor performance of the random strategy as an investment model.

It is possible, therefore, to increase the number of trades created for an investment portfolio; however, the quality of those trades is the factor that determines the performance of the investments made with that portfolio. Thus, the results of this comparison with the random trading strategy help to support the argument that the prediction model creates profitable trading opportunities.

Performance Comparison of Trading Strategies

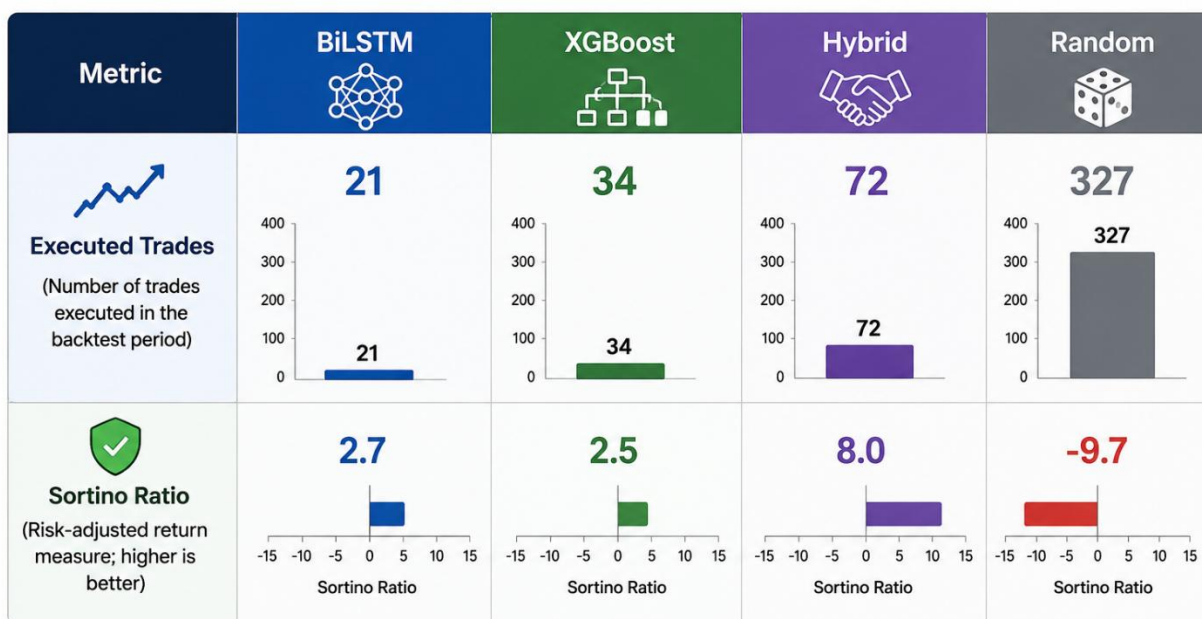


Figure 33 Portfolio Comparison

4.7.6 Discussion

The results of the evaluation of the performance of the proposed framework can be discussed from a variety of different perspectives.

The technical analysis of the results of the proposed framework indicates that the improvements in the ability of the framework to make accurate trading signals have the potential to lead to improvements in the performance of the portfolio of investments that employ those trading signals.

Finally, these results support the objective of the study — that by incorporating both technical and financial information into a framework that determines trading signals, the quality of those trading signals and the investments that are made based upon those signals can both be improved.

4.7.7 Why Sortino Ratio Was Selected

The Sortino Ratio was selected as the main performance measure for the evaluation of the strategies in this thesis.

The Sortino Ratio evaluates the investment performance of a strategy in relation to only the downside risk of that strategy, as opposed to the Sharpe Ratio, which evaluates the performance of investments in relation to their downside risk and upside risk.

As such, the Sortino Ratio is the more appropriate performance indicator for evaluating trading strategies, as the increased risk of the upside returns of investments is typically to be positively considered, rather than being a negative to the performance of the investment strategy.

For these reasons, the use of the Sortino Ratio provided a more indicative measure of the effectiveness of the proposed investment framework than would have otherwise been provided by considering only profitability measures.

Key Findings - Portfolio Evaluation

The proposed framework generated the highest number of executable trading opportunities (72 trades).

The portfolio based on the predictions of the framework achieved the highest Sortino Ratio (8.0).

While the performance of the standalone models was positive, the framework achieved higher risk-adjusted returns.

The random trading strategy led to the highest number of trades, but the negative Sortino Ratio indicates the poor investment performance of the strategy.

The portfolio strategy based on the predictions of the framework exhibited an increased level of both investment quality and quantity relative to the standalone models.

4.8 General Discussion

The experimental results consistently demonstrate that combining technical and financial information provides superior predictive capability compared with relying on either information source independently. While both standalone models captured meaningful market characteristics, neither model alone was able to fully represent the complex dynamics of the Saudi stock market. The hybrid framework successfully integrated these complementary information sources, resulting in substantial improvements in Accuracy, Recall, F1-score, and portfolio performance.

An additional observation concerns the difference between training and testing performance. Financial markets evolve continuously in response to economic conditions, oil price fluctuations, and geopolitical developments. Consequently, predictive relationships identified during model training may weaken when evaluated on future unseen data. Despite this challenge, the hybrid framework maintained the strongest generalization performance among all evaluated models, indicating greater robustness under changing market conditions.

Finally, the portfolio evaluation confirms that improvements in statistical prediction translated into meaningful investment performance. This demonstrates that the proposed framework offers practical decision-support value rather than merely improving conventional machine learning evaluation metrics.

4.8.1 Study Limitations

Although the proposed framework demonstrated superior predictive and portfolio performance compared with the standalone models, several limitations should be acknowledged. First, the experimental evaluation was conducted exclusively using historical data from the Saudi Stock Exchange, and future market behavior may differ due to macroeconomic changes, geopolitical events, or structural shifts in financial markets. Second, the prediction framework relies on historical technical indicators and published financial statements, which cannot capture all external factors influencing investor behavior. Finally, continuous model retraining may be required to maintain predictive performance as market conditions evolve over time.

4.9 Chapter Summary

The purpose of this chapter was to summarize the findings of the experiments that were performed in this study.

The experiments performed in this study included validation of the final machine learning dataset, evaluation of the performance of the standalone BiLSTM and XGBoost models, and evaluation of the performance of the proposed framework for fusing the probabilities of the predictions of the separate models.

The results of these experiments indicated that the proposed framework for fusing the probabilities of the separate models led to a framework that achieved the highest possible classification performance among all models that were tested.

The evaluation of the strategy that was based on the predictions of the framework also revealed that the improvements in classification performance also translated to improved investment performance of the strategy relative to the standalone models.

Thus, overall, the research objectives of the study were achieved.

CHAPTER 5 Conclusion and Future Work

5.1 Introduction

This final chapter presents the conclusions of this research study, as well as suggestions for future research that can be performed in the field of which this research is conducted.

The research that is presented in this thesis investigates whether an end-to-end framework for integrating technical and financial information through a **two-level probabilistic stacking ensemble** can enhance the performance of artificial intelligence models for the purpose of predicting the movements of the stocks in the Saudi Stock Exchange.

The remainder of this chapter will present the findings of the research, the contributions of the research, the implications of the research, limitations of the research, and future recommendations for the research.

5.2 Summary of the Research

The research that is performed within this thesis seeks to develop an end-to-end framework for the prediction of the movements of the stocks within the Saudi Stock Exchange by integrating technical and financial information.

This framework includes the collection of a dataset that includes both technical and financial information of the Saudi Stock Exchange, the validation of the data that is collected, the development of models that utilize the information for the prediction of the movements of the stocks, and the evaluation of the framework's predictions from both a machine learning and investment perspective.

The results of the research indicate that the proposed framework is successful in the prediction of the movements of the stocks within the Saudi Stock Exchange.

5.3 Achievement of the Research Objectives

The objectives of this research study were to develop a framework for integrating both technical and financial information of the stocks traded within the Saudi Stock Exchange, to investigate the predictive capabilities of deep and machine learning models when used separately within the framework, to develop and evaluate a two-level probabilistic stacking ensemble that integrates calibrated probability estimates generated by the standalone models, and to evaluate the developed framework from an investment perspective.

The research fulfilled the objectives established for this study. The framework was developed for integrating the technical and financial information of the stocks within the exchange. The deep and machine learning models were developed and tested to determine their separate abilities in predicting the movements of the stocks. An AI framework for integrating the probabilities of those models was developed and tested for improvements in the classification of information that

the separate models provided. Finally, the framework's decisions were evaluated within the portfolio to determine their success as investments.

Thus, it is evident that the objectives of this research study have been fulfilled.

5.4 Main Research Contributions

The study contributes to the existing financial AI literature by presenting a novel framework for short-term stock market prediction in the Saudi Stock Exchange. Unlike previous research that has focused upon evaluating the performance of individual prediction algorithms, the current research develops an entire framework for fulfilling such a function.

The main contributions of the research are presented below.

5.4.1 Contribution 1: Development of an End-to-End Hybrid Artificial Intelligence Framework

The main contribution of the presented research is the development of a complete end-to-end framework for financial prediction.

Rather than proposing a prediction algorithm, the framework was developed to include each of the stages of financial market prediction, from data collection to evaluation of the resulting investment portfolio.

The framework integrates data engineering, feature engineering, learning models, and investment evaluation into a single framework.

By integrating these components of the framework, the framework is not only readily reproducible but also includes a methodology for applying the framework to other financial market prediction problems.

5.4.2 Contribution 2: Chronological Financial Data Engineering Framework

Due to the potential for financial data to be compromised during its collection, pre-processing, or publishing, a data engineering framework was developed that ensured the chronological integrity of the financial data prior to its use in the prediction models.

Where most financial AI research has focused upon the development of prediction algorithms, this research has focused upon the importance of data integrity prior to the development of AI models.

The data engineering framework ensures the credibility and reproducibility of the results of the study.

5.4.3 Contribution 3: Specialized Learning Architecture

Beyond the development of the framework itself, another main contribution of the study was the creation of a method for specializing the learning of the AI algorithms according to the type of financial data being analyzed.

Instead of applying the same algorithm to all financial data, each learning algorithm was applied only to data of a certain type. For instance, the BiLSTM was learned from technical indicators of the market, while the XGBoost was learned from the financial variables of the market.

Each of these models was developed to best learn the characteristics of the financial data to which it was applied.

5.4.4 Contribution 4: Two-Level Probabilistic Stacking Ensemble

Another contribution of this research is the use of a method for fusing the prediction probabilities generated by each of the individual learning models.

calibration of prediction probabilities followed by adaptive stacking using a Logistic Regression meta-learner.

This strategy increases the flexibility of the framework while maintaining the independence of each of the learning models.

5.4.5 Contribution 5: Separation Between Prediction and Investment Evaluation

Another contribution of the study is the separation between the supervised learning models and the evaluation of the financial investments created with the predictions of those models.

The learning models were developed only using historical market data. The methods utilized for evaluating the financial portfolios were developed after the creation of the prediction models.

This separation ensures the integrity of the learning models and their ability to stand alone from the evaluation of financial investments.

5.4.6 Contribution 6: Comprehensive Evaluation Framework

In contrast to many financial AI research studies that employ only machine learning models to evaluate the success of the prediction algorithms, this research also employs a framework for evaluating the financial investments that were created with the predictions of the models.

This framework allows for the models to be evaluated both as machine learning models and as financial investment models.

5.4.7 Contribution 7: Application to the Saudi Stock Exchange

Finally, another contribution of this research is the application of the framework to the data from the companies listed on the Saudi Stock Exchange.

Few research studies have been performed on financial markets like the Saudi market, and thus the application of financial AI to such a market helps to expand the area of research for financial AI technologies.

Table 32 Summary of Research Contributions

Contribution	Description
C1	End-to-End Hybrid AI Framework
C2	Chronological Data Engineering Pipeline

C3	Specialized Learning Architecture (BiLSTM + XGBoost)
C4	Two-Level Probabilistic Stacking Ensemble
C5	Separation Between Learning and Portfolio Evaluation
C6	Integrated Machine Learning and Portfolio Evaluation Framework
C7	Validation on the Saudi Stock Exchange

5.5 Practical Implications

Beyond the academic contribution of proposing and testing the new framework for financial market prediction, the research also has some implications regarding the use of such a system in practice.

For investors, the framework allows for the integration of technical and financial analysis to provide a more complete picture of financial markets that can be utilized to make investment decisions.

Portfolio managers could use the findings of the study to evaluate prediction models according to investment-oriented metrics to ensure that the models they select for their investing portfolios are in accordance with their investment objectives.

For researchers looking to expand upon the present framework, the architecture of the framework allows for future researchers to implement new models in various stages of the framework to test their effectiveness within the system without having to redesign the framework as a whole.

Finally, the developers of financial technology can utilize the artificial intelligence models incorporated into the framework as a means of creating technological advances within the area of investing that can be utilized by various investors and financial companies.

5.6 Limitations of the Research

In spite of the implications of the study and the valuable information that the framework can provide for investors, there are still some limitations to the research that should be acknowledged.

The framework was tested using the data from 35 companies listed on the Saudi Stock Exchange. While these represented a variety of industries, the conclusions drawn from these markets may not be able to be generalized to other markets across the world.

In addition, the information that was used to train the models was limited to the information that is published by the companies each day (their financial statements) and the technical indicators that can be calculated from the markets. Other information was intentionally ignored in the creation of this framework.

The framework was only tested for the prediction of short-term stock market movements. The framework could not be used to predict medium-term or long-term stock market movements.

In addition, while the virtual portfolio was utilized to test the effectiveness of the framework, it was simulated. Other factors that may have influenced the investing in those companies (market impact, transaction costs, etc.) were ignored in the development of this study.

Finally, the framework was tested using the specific BiLSTM and XGBoost models. While the results were favorable, other models may provide different results when utilized within the framework.

While these limitations to the study may reduce the value of the framework in some ways, they do not reduce the validity of the framework itself. Instead, they help to define the scope of the research that was performed, and contribute to the suggestion of future research that could extend the framework in new ways.

5.7 Future Research

In addition to the research that has been performed with the framework that was proposed in this thesis, there are still a variety of ways in which the framework can be extended in the future to provide even more value to investors and financial technology developers.

One suggestion for the future is to expand the information that is used to train the models used within the framework. For example, financial news feeds from the companies could be used to provide additional information to the models. Additionally, other information sources could be developed as well.

Another future research study that can be performed is testing other deep learning models within the framework. For example, models like Transformers and Temporal Fusion Transformers are more sophisticated deep learning models that could potentially be used in place of BiLSTM.

Another future research study is investigating alternative stacking architectures and meta-learners, such as Gradient Boosting, Neural Meta-Learners, or Attention-based fusion models. For example, the weights of each model could be dynamically adjusted based on various market conditions.

In addition, the framework can be extended in terms of its investment features. For example, incorporating information like portfolio optimization, transaction costs, or dynamic position sizing could lead to the creation of even more sophisticated investing technologies.

Another area of future research is testing the framework on different markets around the world. Testing the same framework on stocks from different countries and different markets could contribute to the understanding of the generalizability of the framework.

Finally, another future research study is incorporating online learning capabilities into the framework. By allowing the framework to update its models with new financial information, it can lead to the creation of trading technologies that are able to provide their predictions in real time.

Table 33 Future Research Opportunities

Research Area	Potential Extension
Data Sources	sentiment analysis, macroeconomic indicators
AI Models	Transformer-based architectures, TFT, Graph Neural Networks
Hybrid Learning	Alternative stacking architectures, adaptive meta-learners, and probability calibration techniques
Portfolio Management	Dynamic position sizing and capital allocation
Market Validation	Cross-market evaluation
Real-Time AI	Online learning and continuous model updating

5.8 Final Conclusion

This thesis presented the design, implementation, and evaluation of a framework for short-term stock market prediction within the Saudi Stock Exchange.

The framework integrated chronological data engineering, learning models, probability fusion, and portfolio evaluation into a framework that provided improved short-term stock market prediction.

The framework was constructed to include each of these elements in a manner that would ensure that information would not be leaked from the models, that the data would be of high quality, and that the evaluation of the models would be separate from the models themselves.

Results from the framework indicated that the technical and financial market indicators each had some predictive power in the markets, but that integrating calibrated probability estimates through a two-level probabilistic stacking ensemble.

In addition, the virtual portfolio indicated that the improvements in market prediction also led to improvements in investment performance. Thus, evaluating artificial intelligence models for financial markets according to both prediction and investment metrics is essential to understanding their true value.

Beyond its implementation of specific machine learning models for stock market prediction, the main contribution of this research is the creation of a framework that integrates each of these elements into a cohesive research and development framework. By following this proposed model, other developers can create their own artificial intelligence models that can be utilized for improving investment performance.

The experimental results further demonstrated that the proposed stacking architecture maintained stronger generalization capability under unseen market conditions than either standalone model.

Finally, the findings of this research support the conclusion that incorporating artificial intelligence models into financial market prediction is a promising area of study for the development of improved technologies for investors, traders, and financial companies around the world.

References

- Abu-Mostafa, Y. S., Magdon-Ismail, M., & Lin, H.-T. (2012). *Learning from data*. AMLBook.
- Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock price prediction using the ARIMA model. In *Proceedings of the UKSim-AMSS 16th International Conference on Computer Modelling and Simulation* (pp. 106–112). IEEE.
- Ahmed, N. K., Atiya, A. F., El Gayar, N., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5–6), 594–621. <https://doi.org/10.1080/07474938.2010.481556>
- Alhazmi, A. H. J., Islam, S. M. N., & Prokofieva, M. (2025). The impact of artificial intelligence adoption on the quality of financial reports on the Saudi Stock Exchange. *International Journal of Financial Studies*, 13(1), 21. <https://doi.org/10.3390/ijfs13010021>
- Al-Nefaie, A. H., & Aldhyani, T. H. H. (2022). Predicting close price in emerging Saudi stock exchange: Time series models. *Electronics*, 11(21), 3443. <https://doi.org/10.3390/electronics11213443>
- Alsulmi, M., & Al-Shahrani, N. (2022). Machine learning-based decision-making for stock trading: Case study for automated trading in Saudi Stock Exchange. *Scientific Programming*, 2022, Article 6542862. <https://doi.org/10.1155/2022/6542862>
- Ansari, Y., Gillani, S., Bukhari, M., Lee, B., Maqsood, M., & Rho, S. (2024). A multifaceted approach to stock market trading using reinforcement learning. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3418510>
- Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long short-term memory. *PLOS ONE*, 12(7), e0180944. <https://doi.org/10.1371/journal.pone.0180944>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Capital Market Authority. (2024). *Market regulations and guidelines*.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>
- Daoria, H. K., Alharthi, M. E., Alzahrani, G. S., Alanazi, A. E., Aborokbah, M., Alotaibi, M., & Alanazi, T. M. (2024). Saudi market stock price predictions using deep learning algorithms and weighting Twitter sentiment analysis. *University of Ha'il Journal of Science*, 5(1), 30–48.
- Daradkeh, M. K. (2022). A hybrid data analytics framework with sentiment convergence and multi-feature fusion for stock trend prediction. *Electronics*, 11(2), 250.

- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417.
- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*.
- Khurayshi, A., & Alharbi, A. (2025). Saudi Arabia stock prices prediction using machine learning. *Advances and Applications in Statistics*, 92(6), 813–833.
- Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2), 689–702.
- Lim, B., Arik, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *Nature Machine Intelligence*, 3(1), 35–44.
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems (NeurIPS)*, 30.
- Molnar, C. (2022). *Interpretable machine learning* (2nd ed.).
- Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, 53, 3007–3057.
- Prechelt, L. (1998). Early stopping—But when? In *Neural Networks: Tricks of the Trade* (pp. 55–69). Springer.
- Saberironaghi, M., Ren, J., & Saberironaghi, A. (2025). Stock market prediction using machine learning and deep learning techniques: A review. *AppliedMath*, 5(3), 76.
- Saudi Exchange. (2025). *Historical market data*.
- Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11), 2673–2681.
- Sharpe, W. F. (1994). The Sharpe ratio. *Journal of Portfolio Management*, 21(1), 49–58.

Sortino, F. A., & Price, L. N. (1994). Performance measurement in a downside risk framework. *Journal of Investing*, 3(3), 59–64.

Tsai, C.-F., & Wang, S.-P. (2009). Stock price forecasting by hybrid machine learning techniques. *Expert Systems with Applications*, 36(4), 7551–7559.

Yahoo Finance. (2024). *Historical market data*.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12), 11106–11115.