Assessing Irish Banking Stocks through Time-Series Forecasting and Quantitative Trading Models

MSc Research Project
Programme Name

Sadhik Shaik
Student ID: x23213108

School of Computing
National College of Ireland

Supervisor: Vladimir Milosavljevic

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Sadhik Shaik
Student ID:	X23213108
Programme:	MSCDAD_A_JAN24 **Year:** 2024
Module:	MSc Research Project
Supervisor:	Vladimir Milosavljevic
Submission Due Date:	12th December 2024
Project Title:	Assessing Irish Banking Stocks through Time-Series Forecasting and Quantitative Trading Models
Word Count:	10001 **Page Count** 29

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Sadhik Shaik

Date: 12/12/2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Assessing Irish Banking Stocks through Time-Series Forecasting and Quantitative Trading Models

Sadhik Shaik
X23213108

Abstract

This study investigates the application of predictive modelling techniques for stock price forecasting of AIB and BOI banks, Comparing the performance of ARIMA, Linear Regression, LSTM, and a Hybrid ARIMA-LSTM model. The methodology includes comprehensive data preprocessing, feature engineering, and evaluation using metrics such as RMSE and R^2. Results demonstrate the potential of the Hybrid ARIMA-LSTM model to leverage the strengths of statistical and deep learning approaches, effectively capturing both linear trends and non-linear patterns in stock price data. While Linear Regression and ARIMA excel in simplicity and accuracy for specific datasets, the Hybrid ARIMA-LSTM showcases superior adaptability to complex and volatile data structures with an accuracy of 99%. The predictive capability of model is narrow with unseen smaller samples. The study finds that BOI stock has stable and reliable daily returns compared to AIB. This work underlines the importance of the selection of predictive models with respect to data characteristics for financial forecasting and decision-making in stock markets. However, the research also recognizes several limitations, including the impact of external economic factors and market sentiments on stock prices. The work may be extended by incorporating those external variables, exploring newer ensemble learning techniques, or assessing real-world applicability in dynamic market scenarios. This research provides valuable findings upon differentiating model results on the amount of data we use for model training is significant, that could be useful for both academics and financial practitioners in the development of robust methodologies for predicting stock prices.

1 Introduction

1.1 Background

In the field of financial sector, the stock market has significant influence on economic growth (Armagan, 2023). Forecasting the future value of stocks by using historical data with deep analysis on volatile stock indexes and by employing robust hybrid models using traditional ARIMA modeling and Long Short-Term Memory (LSTM) model along with that the ability to calculate investment returns from stock price helps in making better investment strategies for stakeholder resilience. The ability to effectively predict the stock price has long been a cornerstone in financial and economic research, forming a very bedrock of informed investment strategies with associated risk management. Stock price forecasting helps not only in making more informed decisions by both institutional and individual investors but also in observing greater market dynamics as a function of influencing both economic policy and corporate strategies (Schumpeter, 1934). With the growing digitization of financial markets and access to enormous historical data, time-series forecasting models have become essential tools. These models enable the stakeholders to forecast the price movements and, thus, get ready for the market trend and avoid losses. Forecasting stock prices in the banking industry, which is by nature closely related to macroeconomic indicators and geopolitical events, is even more important. This would make the Irish banking sector a very interesting case in studying the effectiveness of these predictive tools, considering its atypical post-crisis evolution and peculiar regulatory environment. In the last two decades, the system of the Irish banks has experienced various changes that were influenced by the development of the enhanced domestic and international financial crisis of 2008, the Irish operations bailouts and the shift in the direction of the attempts at enhancing stricter regulations. In this study, two giants of Ireland's financial industry, the Allied Irish Banks (AIB) and the Bank of Ireland (BOI) are used as reference points to the sector. In contrast to other developed economies where banking stock has witnessed history of moderate variability, the Irish banking has displayed cycles of volatility, this requires extensive models of forecasting for Strategical approaches. In the last two decades, the system of the Irish banks has experienced various changes that were influenced by the development of the enhanced domestic and international financial crisis of 2008[1], the Irish operations bailouts and the shift in the direction of the attempts at enhancing stricter regulations. In this study, two giants of Ireland's financial industry, the Allied Irish Banks (AIB) and the Bank of Ireland (BOI) are used as reference points to the sector. A stock market analysis of AIB and BOI does not only depend on the domestic economic environment, but it is also a factor of extra-economic environment, such as Brexit, recent change in the European Central Bank policy, or even global markets. Irish based banking stocks similarly showed high degrees of variability in the past, which underscores its critical health dependence on changes in related macroeconomic factors. In contrast to other developed economies where banking stock has witnessed history of moderate variability, the Irish banking has displayed cycles of volatility, this requires sound models of forecasting for investment.

1.2 Problem Statement: Challenges in Forecasting Irish Banking Stocks

Predicting the stock price of Irish banking institutions, such as AIB and BOI, presents several challenges. First, the intrinsic volatility of the sector obscures the extraction of regular patterns. Second, macroeconomic variables like Interest rates, unemployment rate, Euro-Dollar Exchange rate and inflation often interact in many complicated ways, making the establishment of clear causation with stock prices difficult to achieve. Additionally, the relatively limited dataset that would be available for Irish banks compared to larger markets presents a limitation in the application of some of the advanced machine learning models. These challenges, therefore, call for an integrated hybrid model approach whereby traditional statistical models such as ARIMA and SARIMA are combined with more sophisticated techniques such as Long Short-Term Memory networks in improving the accuracy of the predictions.

1.3 Research Questions

How can advanced time-series forecasting models enhance the prediction accuracy of AIB and BOI stock prices for improved investment strategies?

[1] irelands-fin-crisis-a-comparative-context.pdf

An integrated hybrid modeling approach whereby traditional statistical models such as ARIMA and SARIMA are combined with more sophisticated techniques such as Long Short-Term Memory deep learning technique for capturing underlying patterns and temporal dependencies in volatile stock data. This paper contributes to the existing literature on financial forecasting studies through a detailed analysis of Irish banking stocks using time-series forecasting and quantitative trading models and evaluating the daily returns. The study aims to provide an integrated framework for stock price prediction in volatile markets by combining traditional statistical techniques with modern deep learning approaches. Moreover, the Irish banking sector focus completes the critical gaps in the existing literature, while the drawn implications may apply to other market economies as well.

To ensure that all aspects are covered, the report is structured to start with an introduction, literature review, methodology implementation, and conclusion and future work: The Introduction sets out the aims of the research and sets the scene for the study in a wider financial context. The Literature Review provides a critical overview of extant methodologies on time-series forecasting and their applications on banking stocks. Methodology sets out the design of the research, sources of data, and analytical methods used. The implementation section describes the steps involved in model development and evaluation. Finally, the Conclusion and Future Work summarize the findings and point out further research that could be conducted

2 Related Work

Time-Series Forecasting in Financial Applications

Time-series forecasting has become an integral part of financial modeling, providing tools to predict future values based on historical data. Traditional statistical methods include ARIMA and Seasonal ARIMA, foundational techniques that have widely applied to stock price prediction because of their simplicity and robustness. Hyndman and Athanasopoulos (2018) indicate that ARIMA works extremely well in univariate data with a distinct temporal structure and thus is quite applicable in financial time series. The SARIMA extends ARIMA by adding the seasonal component to deal with periodic fluctuations in data, such as quarterly earnings or yearly economic cycles. According to Box et al. (2015), both these models rest on the assumption of stationarity and linearity in data. Thus, their performance gets restricted when it comes to complicated nonlinear financial data.

This paper demonstrates the use of Long Short-Term Memory (LSTM) models for predicting the S&P 500's next day closing prices (Bhandari et al., 2022), by employing with the mix of fundamental market data, macroeconomic indicators such as Cboe volatility, Interest rate, Civilian unemployment rate, Consumer sentiment index, and US dollar index, and technical analysis. Author’s worked on including a comprehensive selection of predictors upon critical analysis, hyperparameter tuning, and decent comparison between single-layer and multi-layer LSTM architectures, with the performance that favors single-layer model with the of accuracy of 99.02 % However, the study's scope is limited to the S&P 500, lacks exploration of hybrid models (e.g., LSTM with ARIMA), and utilizing advanced feature selection techniques, leaves opportunities to enhance robustness and generalizability. Time-series forecasting has become an integral part of financial modeling, providing tools to predict future values based on historical data. Traditional statistical methods include ARIMA and Seasonal ARIMA, foundational techniques that have widely applied to stock price prediction because of their simplicity and robustness. Hyndman and Athanasopoulos (2018) indicate that ARIMA works extremely well in univariate data with a distinct temporal structure and thus is quite applicable in financial time series. The SARIMA extends ARIMA by adding the seasonal component to deal with periodic fluctuations in data, such as quarterly earnings or yearly economic cycles. According to Box et al. (2015), both these models rest on the assumption of stationarity and linearity in data. Thus, their performance gets restricted when it comes to complicated nonlinear financial data.

Recent developments have also brought in ML and deep learning models as alternatives. Of late, among many, Long Short-Term Memory networks, a specific kind of RNN, have achieved incredible capability in capturing long-term dependencies and non-linear patterns in sequential data (Hewamalage et al., 2021). In this way, LSTMs overcome some of the disadvantages of traditional models by learning complex relationships from raw data without requiring any manual feature engineering. However, these models are highly dependent on large datasets, which might be a limitation in markets like Ireland that have limited stock market history.

In recent years there have been studies on the integration of ARIMA models with LSTM approaches. Zhang et al. (2020) introduced a combination of a novel ARIMA model and a LSTM model, thus viewing the stock market data from two perspectives, while leaning on the flexibility of linear modeling in ARIMA to joint determination of the non-linearity inherent in LSTM. Their study showed that it offered better forecasting performance than individual models, especially on those data sets with a combination of both linear and non-linear patterns. Such hybrid methods are promising but their parameters have to be set in a proper way so that their performance is not over-fitted to the training set.

Application of Quantitative Trading Models in Stock Forecasting

Quantitative trading models are meant to uncover viable trades using mathematical and statistical tools. Because of this, most of the models combine time-series forecasting with technical indicators such as moving averages, Relative Strength Index (RSI), and Bollinger Bands. According to Fama and French (2015), in a quantitative trade, there is great use of factor-based models where macroeconomic indicators, firm size, and value metrics are considered to be strong determinants of stock returns.

Chen et al. (2020) analyzed how machine learning can be used in quantitative trading and showed that the statistical models based on historical prices and macroeconomic factors work better than some other approaches. Their study pointed out that incorporating factors at the external level; like interest rate and inflation bring in model stability through considering the macroeconomic conditions that impact stocks. Still, adopting such models in relatively obscure markets such as Ireland is a problem because of the dearth of data and increased fluctuation of rates.

Application of quantitative models in the banking sector is relatively limited as compared to other industries. Beck et al. (2018) and Duffie (2010) emphasized the use of local models that are sensitive to factors specific to that market, such as changes in regulatory policies and geopolitical events. Irish banking stocks are highly sensitive to Brexit or European Central Bank policies; therefore, there will be a need to have quantitative strategies that take such external influences into consideration.

Role of Macroeconomic Indicators in Financial Forecasting

Macroeconomic variables are important in analyzing fluctuations in stock prices particularly in banking companies given that their financial performance reflects most of the macroeconomic factors. Economic signs like GDP, employment and unemployment rates, inflation rates as well as interest have been known to be heralds for some time now. Kilincarslan & Ozdemir (2018) explained the degree of relationship between macroeconomic surrections and stock market fluctuations and pointed out that often these factors act as the leading indicators of changes in the investor's attitude and market dynamics.

Bernanke and Kuttner (2005) examined the stock price impact of central bank policy decisions, particularly decisions about interest rates. The results showed that monetary policy decisions directly relate to banking stocks because of lending and borrowing activities. In the same vein, Carhart (2017) has added to this study fiscal policies and global economic events and their combined effect on market volatility.

The inclusion of macroeconomic indicators is very important for the Irish banking stocks due to the sensitivity of the sector to domestic and international economic conditions. It follows from studies that Brexit-related uncertainties and changes in eurozone economic policies strongly influence AIB and BOI, hence this variable becomes indispensable for a more accurate forecast (Honohan, 2010).

Advancements in Hybrid Forecasting Models

Hybrid models, which combine traditional statistical techniques with machine learning approaches, have emerged as a promising solution to overcome the challenges of financial time-series prediction. The idea is to leverage the strengths of individual methodologies while mitigating their weaknesses. For instance, the hybrid ARIMA-LSTM model outperforms stand-alone models by effectively capturing both linear and nonlinear dependencies in data (Zhang et al., 2020). The methods which incorporate an exponential smoothing methodology coupled with neural networks are found to yield higher accuracy for the seasonal trend in stock price forecasting (Makridakis et al., 2020).

However, applications of hybrid models in small and volatile markets, such as Ireland, are not well explored. For example, the study by Adebiyi et al. (2014) showed that the performance of hybrid models is heavily dependent on the quality and quantity of the training data, which might be limited in less mature markets. This again emphasizes the importance of feature engineering and hyperparameter tuning for enhancing model performance.

Gaps in Existing Research and Justification for the Study

Despite massive steps in times series forecasting and quantitative trading, a great deal of the literature remains relatively absent in important details related to this issue on the Irish banking sector. First, most of these studies focus on large established economies that reflect minimal generalizability into economies that are relatively small such as Ireland. Second, despite proof of their relevance in making appropriate predictions of the performance in the stocks of banks, the integrating approach that macroeconomic indicators provide for hybrid models remains at very young stages.

This project tries to bridge these gaps by proposing a complete forecasting framework for the Irish banking sector. Blending traditional statistical methods with most recent machine learning methodologies and using macroeconomic variables in the research tends to enhance predictive accuracy and will have an effect of providing actionable insight for investors and policymakers. The adoption of hybrid models is a novel approach toward the resolution of financial forecasting complexities that exist in small and volatile markets.

3 Research Methodology

This section describes how the research objectives and respondents to the research questions were consistently developed. The methodology will be anchored on the framework of Cross-Industry Standard Process for Data Mining-CRISP-DM-guided in phases: data collection, preprocessing, modelling, and evaluation.

3.1 Research Design

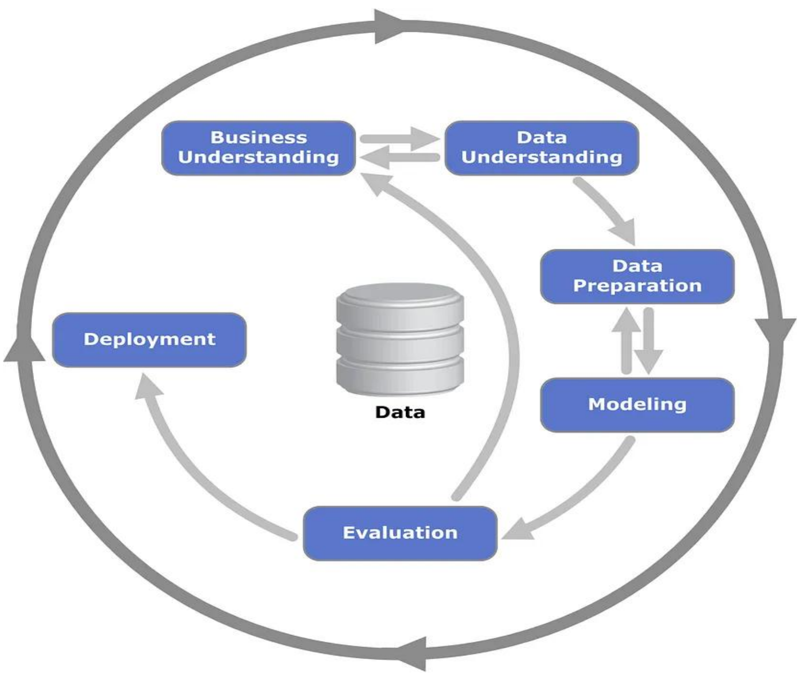


Fig :1 CRISP-DM Process Framework

The CRISP-DM framework was selected for the research because of its comprehensive and iterative nature; it supports a structured approach toward data analysis and modeling. Its six interlinked phases provide a robust foundation for ensuring alignment between the research objectives, methods, and outcomes. The structure of this methodology is one of flexibility and refinement, to a large degree, but, because complex financial data regarding AIB and BOI were being dealt with, extra emphasis had to be given at various places. Each stage of the CRISP-DM framework has been discussed below in respect to this study.

3.1.1 Business Understanding

The first step of this study is to identify the goals and objectives of the research, ensuring that they address the key challenges and opportunities that face the Irish banking sector. The main objective of this research is to develop an accurate forecasting model for the stock prices and day-to-day returns of AIB and BOI for strategical investment decisions (Adebiyi et al., 2014; Beck et al., 2018), two major banks in Ireland, by integrating macroeconomic indicators into the analysis. This phase also puts the study into perspective with the wider aim of improving investment strategies and risk management frameworks, thus providing actionable insights for institutional and retail investors.

3.1.2 Data Understanding

The data understanding phase identifies, gathers, and explores datasets relevant to the research objectives. Historical stock price data of AIB and BOI take from Yahoo Finance, along with macroeconomic indicators such as Euro-Dollar Exchange rates, unemployment rates, and interest rates, are taken from European central bank (Kilincarslan & Ozdemir, 2018; Honohan, 2010) which enables the authenticity of the real data, these data forms the core of the analysis. The source data has 6346 data points with the data range from 04/01/2000 till October 2024 for AIB in a business frequency as similar to all the stock indexes and the BOI data from June 2001 till October 2024 in a business frequency. In macro-economic indicators, Interest rate data starts from January 1999 until October 2024 an same data range for Exchange rate data and the civilian unemployment rate data starts from January 1998 till October 2024. The Summary statistics show the volatile nature and non-linearity in financial stock market fluctuations from the decades, and to the fact that the market rates and the economic cycles changes over the decades has it's influence. There are 7 columns for both the AIB and BOI stock data open, close, adj close, volume, high, low and the date column and for macro economic indicators, Interest has three columns date, time period, interest rate, exchange rate has three columns date, time period, exchange rate and civilian unemployment rate has six columns with age group, unemployment rate, statistical label, month, sex, units in a Monthly frequency.

3.2 Data Preparation

3.2.1 Data Pre-Processing

The raw data collected is preprocessed to make it ready for analysis and modeling. Data cleaning deals with various data problems that include missing values, outliers and inconsistencies such as unsuited characteristics for timeseries models. Upon dealing with periodic stock data from two decades, careful precautions were taken to maintain authenticity as it might be having seasonal and periodic trends, where forcefully removing inconsistent data points will lead to analytical gaps. There are unseen and uncaptured missing values in our source data which were treaded using forward filling ffill(), time-based linear interpolation and rolling mean methods were used (Anon, 2021) to handle them in mathematical approach to maintain data integrity (Tsay, 2010; Anon, 2021). Afterall performing the in-depth data pre-processing, unnecessary data like stock dividends and stock splits, comma values and unnecessary spaces were removed without losing any data point for the business frequency. Afterall careful pre-processing of data, making it into ascending order, sorting all the dates, the data is merged into two CSV files for exploratory data analysis and modeling AIB_merged and BOI_merged datasets. Below figure shows the final information and summary statistics of merged datasets.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6345 entries, 0 to 6344
Data columns (total 10 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Date               6345 non-null   object
 1   Open               6345 non-null   float64
 2   High               6345 non-null   float64
 3   Low                6345 non-null   float64
 4   Close              6345 non-null   float64
 5   Adj Close          6345 non-null   float64
 6   Volume             6345 non-null   float64
 7   unemployment_rate  6330 non-null   float64
 8   exchange_Rate      6345 non-null   float64
 9   Interest_Rate      6345 non-null   float64
dtypes: float64(9), object(1)
memory usage: 495.8+ KB
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5947 entries, 0 to 5946
Data columns (total 10 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Date               5947 non-null   object
 1   Open               5947 non-null   float64
 2   High               5947 non-null   float64
 3   Low                5947 non-null   float64
 4   Close              5947 non-null   float64
 5   Adj Close          5947 non-null   float64
 6   Volume             5947 non-null   float64
 7   unemployment_rate  5932 non-null   float64
 8   exchange_Rate      5947 non-null   float64
 9   Interest_Rate      5947 non-null   float64
```

```
Summary Statistics for AIB_EDA:
              Open         High          Low        Close    Adj Close  \
count  8462.000000  8462.000000  8462.000000  8462.000000  8462.000000
mean   1219.395628  1236.320204  1190.883061  1218.867918   830.809803
std    1805.098036  1826.639562  1767.804104  1805.283373  1241.701748
min       0.750000     0.810000     0.719000     0.768000     0.646300
25%       4.850000     4.900000     4.785000     4.845250     3.961900
50%      22.000000    22.500000    21.250000    21.750000    17.234300
75%    3100.000000  3125.000000  3062.500000  3100.000000  1950.349900
max    6050.000000  6953.500000  5980.000000  5987.500000  4273.660200

             Volume  unemployment_rate  exchange_Rate  Interest_Rate
count  8.462000e+03        8462.000000    8462.000000     8462.00000
mean   1.374245e+06          66.778084       1.209178        1.90608
std    4.537892e+06          35.648472       0.143148        1.90537
min    8.000000e+00          26.900000       0.857800        0.00000
25%    4.502250e+03          38.400000       1.104100        0.00000
50%    1.063350e+04          51.900000       1.196450        1.00000
75%    1.715441e+06          94.700000       1.315900        4.25000
max    1.434718e+08         141.050000       1.599000        4.50000
```

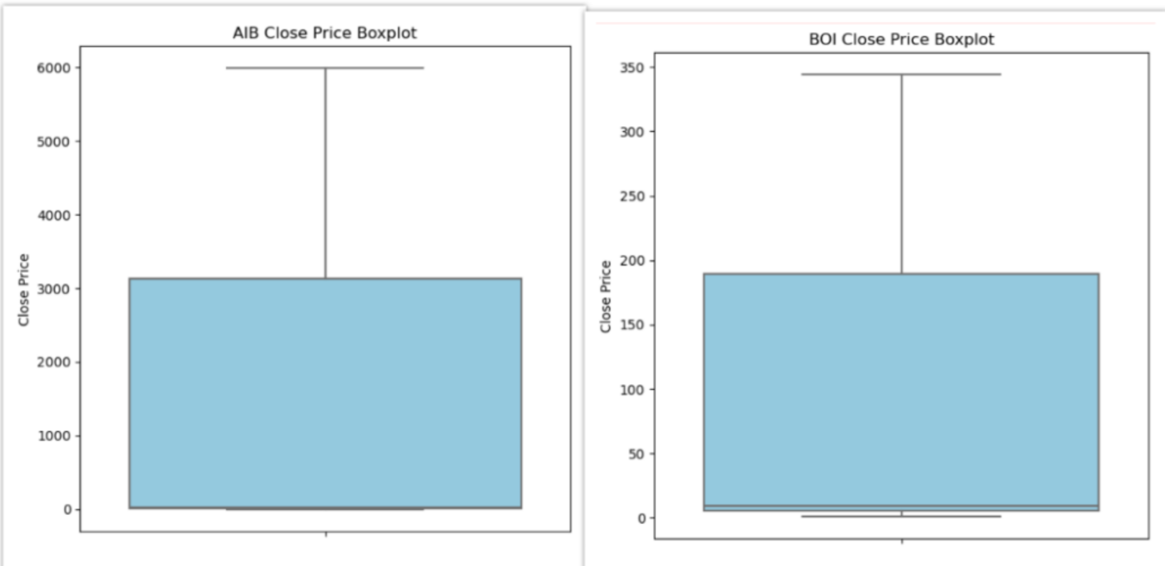
```
Summary Statistics for BOI_EDA:
              Open         High          Low        Close    Adj Close
count  8462.000000  8462.000000  8462.000000  8462.000000  8462.000000
mean     74.159458    75.378114    72.721116    74.072124    53.860909
std     100.734479   102.206614    99.240140   100.679681    72.283301
min       1.330000     1.370000     1.280000     1.330000     1.170000
25%       5.760000     5.880000     5.640000     5.790000     4.990000
50%       9.300000     9.480000     9.100000     9.300000     8.110000
75%     187.250000   189.960000   184.430000   186.782500   126.790000
max     418.650000   418.650000   340.640000   343.960000   262.400000

             Volume  unemployment_rate  exchange_Rate  Interest_Rate
count  8.462000e+03        8462.000000    8462.000000     8462.00000
mean   1.734157e+06          66.778084       1.209178        1.90608
std    4.182348e+06          35.648472       0.143148        1.90537
min    1.084000e+03          26.900000       0.857800        0.00000
25%    1.702110e+05          38.400000       1.104100        0.00000
50%    9.956400e+05          51.900000       1.196450        1.00000
```

Fig :2 Summary statistics and Description of AIB and BOI stocks

The summary statistics describe the distribution of the stock data. The AIB average closing price is 1,218.87 shows a high volatility by a standard deviation of 1,805.98 in the open price. According to the maximum trading volume 1.434718e+08 speaks about the amount of significant trading activity on certain days. BOI summary statistics describe those lower men for closing price around 74.07 when compared to AIB and lower standard deviation in open price at 102.20 makes it more consistent in terms of indicating less fluctuation. The highest trading volume recorded is 1.611624e+08, which is slightly higher compared to AIB, the highest increased number of intense trading. The data is ready with non-null values and makes it comprehensive and reliable for financial analysis.

Outlier deduction, to the volatility of stock data, checking and managing noise data before modeling is an important step to be considered. Outlier deduction has been done using IQR range and Z-score, the IQR function has been set to check if any of the data falls outside of 1st quartile and 3rd quartile and Z-score analysis has been done to cross check outliers' presence.

The distribution of AIB closing prices is mostly skewed towards lower values with a peak near zero, which is a clear evidence, mostly the data is concentrated below 1000, with some spikes occasionally due to unusual market trends and economic influences. The boxplot below supports a median close to the lower quartile. There are no outliers detected according to the IQR method, and which is supported by Z-score analysis, where no values were found outside of three standard deviations from mean.




```
--- Z-Score Outliers in AIB ---
Empty DataFrame
Columns: [Open, High, Low, Close, Adj Close, Volume, unemployment_rate, exchange_Rate, Interest_Rate, zscore_close]
Index: []
--- Z-Score Outliers in BOI ---
Empty DataFrame
Columns: [Open, High, Low, Close, Adj Close, Volume, unemployment_rate, exchange_Rate, Interest_Rate, zscore_close]
Index: []
```

Fig. :3 AIB and BOI Outlier deduction using IRQ and Z-score Analysis

Even for the BOI, the close price has a skweness towards lower values, but with a slightly narrow spikes compared to AIB, the most amount of close price is surrounded aroung 100, with few higher values. The box plots shows that around 50% of prices are formed around lower price range, with median that is close to the 1st quartile. As compared to the AIB, BOI has less variation and as similar to AIB no outliers falls outside of 1st and 3rd quartiles which is supported by Z-score analysis.

3.2.2 Exploratory Data Analysis

Temporal Trends in Stock Prices and Volumes

Historical plots of stock closing prices over time give an idea of the tremendous volatility of the stocks. For AIB, the stock price crashed in the financial crisis of 2008 and did not come back to those highs for a long time. The impact of the crisis was strongly felt on Irish banking. In later stages after 2008, the price distribution is not much fluctuating, the trend is not much skewed. These type of data have volatility and stationarity influences, so careful considerations before modeling will be significant

Figure 4: AIB Stock Closing Prices over time

Similarly, there is volatility in the stock price of BOI, and clear variations can be seen in financial crisis of 2008 and did not come back to those highs for a long time. In this connection, the time series examination of trading volumes for the BOI shows some wide variations coinciding with these economic events, bringing out the linkage of the market activities with external shocks relating to macroeconomics. Both these visualizations make a challenging as well as strong point for time-series models that could concurrently model long-term trends as well as event-driven fluctuations.

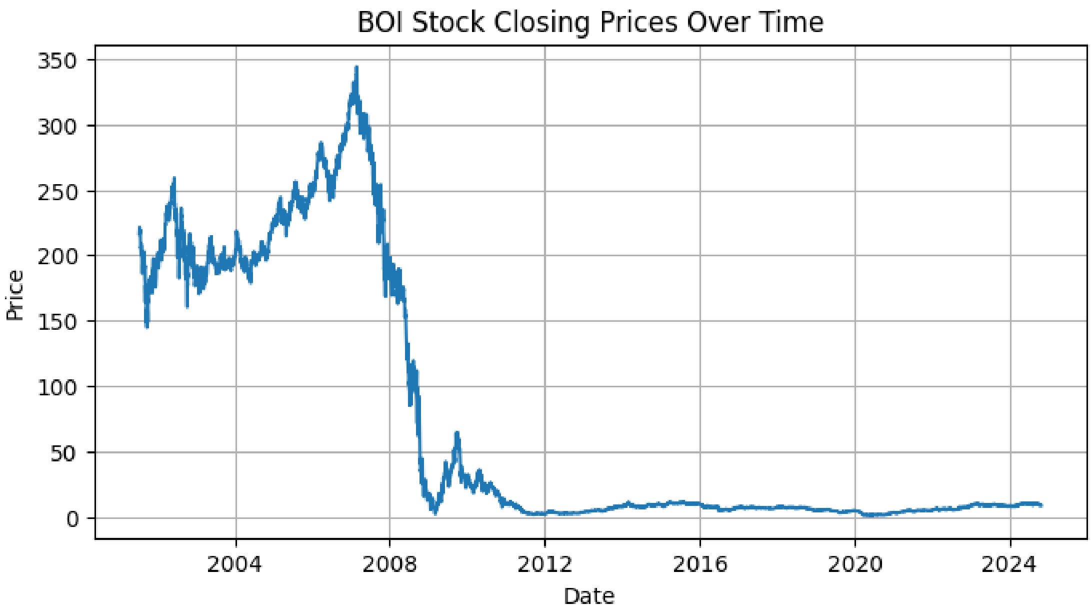


Figure 5: BOI Stock Closing Prices Overtime

Correlation Analysis

From the AIB and BOI stock price and macroeconomic variables, there is evidence that variables are related to each other from the correlation heatmaps. The stock price variables include opening price, highest price, the lowest price and closing price, and the results show that these features are very much related hence their movements are aligned in the same direction, which is common in stock data. A corresponding relationship between interest rates and closing prices is illustrated by the fact that interest rates are positively linked with stock prices of AIB and BOI both by their p-value. On this point, we witnessed that unemployment rates have negative coefficients that show the poor performance of stocks during economic downturns. Indeed, these results confirm the importance of the use of the macro variables in forecasting equations in order to enhance the accuracy of forecasting models.

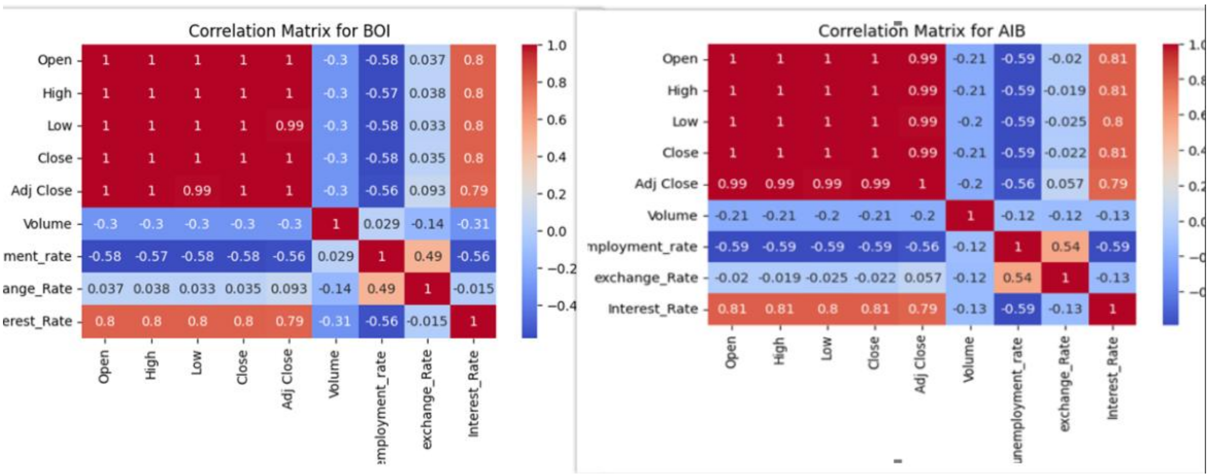


Figure 6: Correlation heat map for AIB and BOI

In the above heatmap, as the p-values close to 1 the variable has a strong co-relation to close price and as it turns negative the relationship goes week and the heatmap shows the same, as it goes dark the relation gets stronger and as it goes week the relation goes thin.

Trading Volume and Market Activity

These trading volume trends for AIB and BOI indeed reflect that both have shown significant upticks around the timing of major economic events, reflective of high market activities and investor interest. In the case of BOI, for instance, there is a strong coincidence between the 2016 Brexit referendum and a strong surge in trading volume, reflecting the uncertainty and speculative activities that have taken place within the market at that time. In contrast, the trading volume of AIB surges during significant events post-2016, including periods of economic recovery and market volatility due to global financial uncertainties.

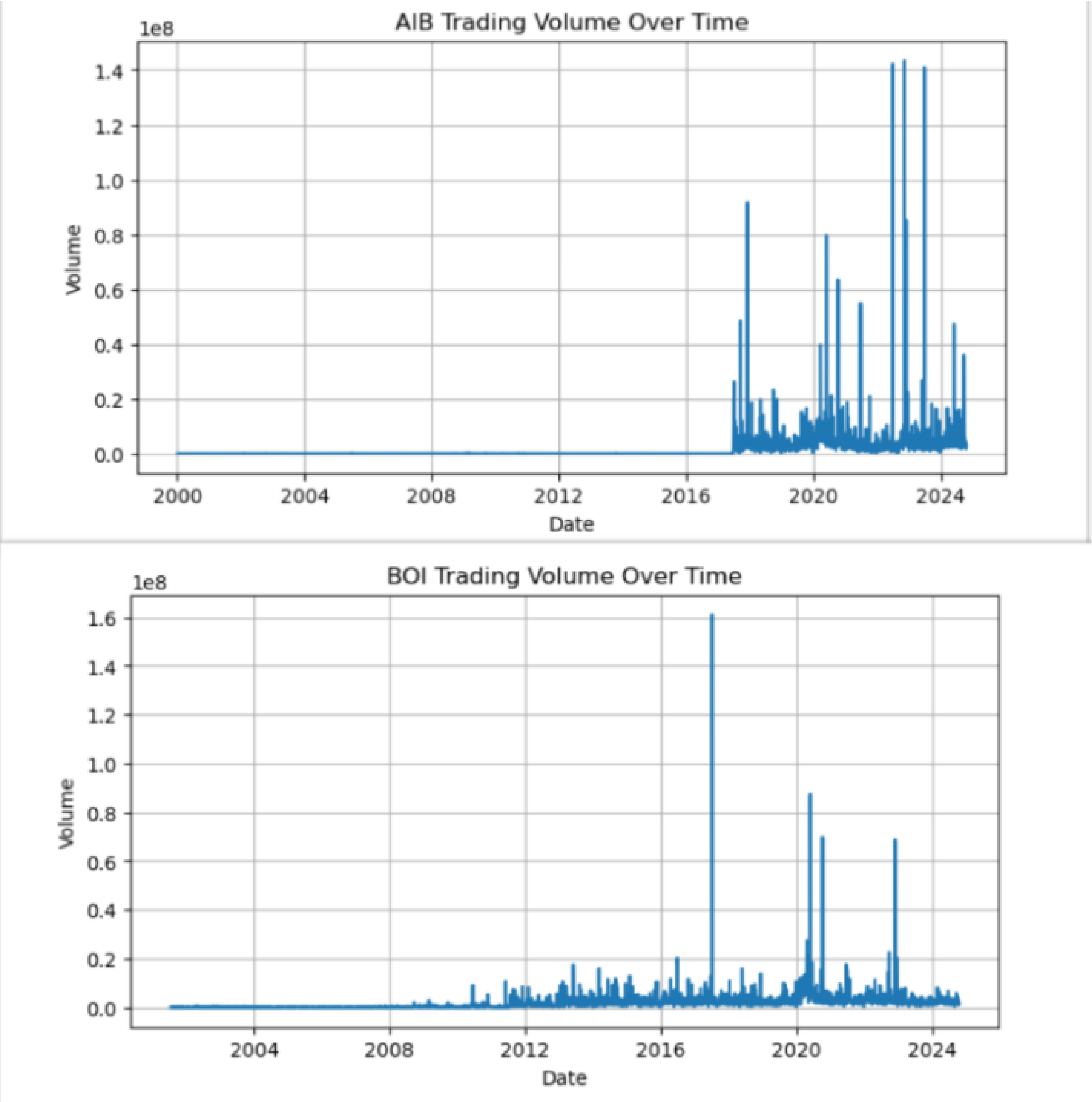


Figure :7 AIB and BOI Trading volume over time.

The time-series plot for the trading volume of AIB reflects activity starting in 2016, followed by consistent spikes into subsequent years. This justifies the increase in investor interest when the market finally stirred from the global financial crisis. Generally, price movements are preceded by surges in trading volume, and hence volume dynamics are one of the most important features in predictive modeling. By including trading volumes as an input, the forecasting models can capture investor sentiment more accurately and predict future price changes more precisely.

3.2.3 Feature engineering:

Feature engineering focused on augmenting the dataset with meaningful variables that could enhance predictive performance. Additional features created include lagged variables, rolling averages, and differenced values that enhance the predictive power of the dataset. The macroeconomic indicators are then normalized using the Min-Max scaling or Z-score normalization to make them compatible with machine learning algorithms. Generated technical indicators such as moving averages (MA5, MA20, MA50), Bollinger Bands, and Relative Strength Index (RSI) to capture price trends, volatility, and momentum (Chen et al., 2020; Zhang et al., 2020). Other features included lagged values for price changes and macroeconomic variables like unemployment and interest rates in order to introduce temporal dependencies into the model. These engineered features allowed for a richer quality dataset so that the model could capture complex relationships and dynamics in the stock prices.

Feature Selection

Feature selection was done to identify the best predictors for the target variable of closing price. Feature correlations were done to establish the strength of association between features and the target variable; only variables with a higher threshold above the mean were retained in the modeling process. In this case, AIB has selected opening price, interest rate, and volatility because of their strong correlation with closing price. Similarly, some of the selected features for BOI included trading volume, unemployment rate, and Bollinger Band-related metrics. In so doing, the dataset had been streamlined to make the model more efficient and interpretable, ensuring that no critical information had been missed.

```
Selected features and their correlations with target:
Open: 1.000
High: 1.000
Low: 0.996
Adj Close: 0.993
unemployment_rate: 0.728
Interest_Rate: 0.761
MA5: 1.000
MA20: 0.998
MA50: 0.996
BB_middle: 0.998
BB_upper: 0.997
BB_lower: 0.998
volatility: 0.634
volume ma5: 0.242
```

Fig :8 feature co-relation with close price

Stationarity Check and Transformation

In order to be certain that the data really meets the requirements for time-series modeling, both stationarity check and transformation were considered. Stationarity is indeed one of the most important properties that several forecasting models usually rely on. The Augmented Dickey-Fuller result showed that the raw series of closing prices of either AIB or BOI is nonstationary (Armagan, 2023), meaning its behavior has changed over time in trend and volatility. Data transformations were made to make the series stationary; first order differencing has been applied. The series were first log-transformed to stabilize variance, then differenced of order one to remove trend. The plots of these series were thereafter inspected visually to check for stationarity and some re-tests done to ensure that the process was successful. Indeed, the two stationary series plots of AIB and BOI showed full equality of their mean and variance over time. These transformations put the dataset into a form that is best suited for effective application of various statistical and machine learning models for forecasting by satisfying their respective assumptions.

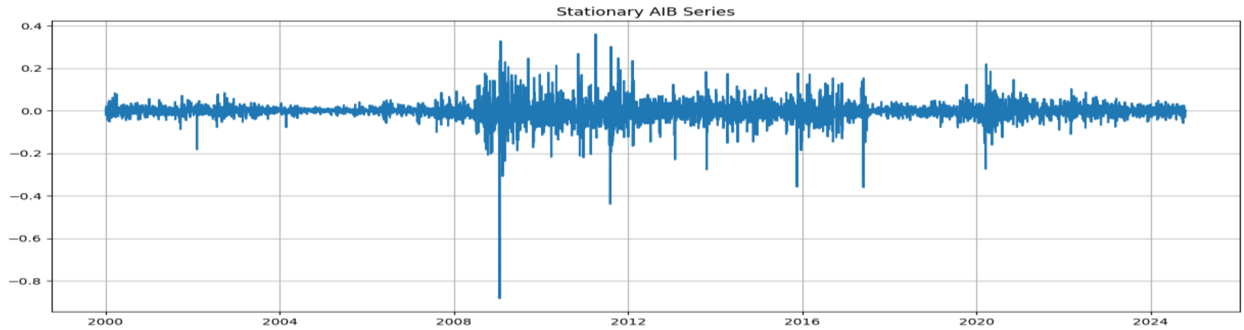


Figure 9: Stationary AIB Series

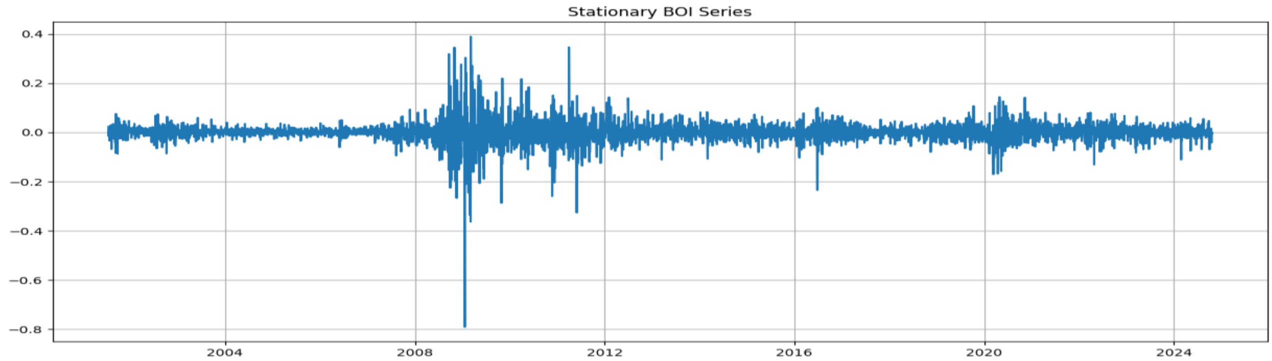


Figure 10: Stationary BOI Series

Data splitting is an important preprocessing step in the preparation of datasets for modeling. It considered data from 1st August 2001 to 31st May 2024 for training and reserved data from 1st June 2024 to 30th September 2024 for testing.

```
AIB Data
Training set: (5832, 9) (from 2001-08-01 00:00:00 to 2024-05-31 00:00:00)
Testing set: (86, 9) (from 2024-06-03 00:00:00 to 2024-09-30 00:00:00)
BOI Data
Training set: (5832, 9) (from 2001-08-01 00:00:00 to 2024-05-31 00:00:00)
Testing set: (86, 9) (from 2024-06-03 00:00:00 to 2024-09-30 00:00:00)
```

Fig :11 Data Splitting of AIB and BOI

3.3 Modeling

This stage designs the statistical and machine learning models that best answer the research questions. It considers some traditional time-series models that account for linearity and seasonality in stock price data, such as ARIMA and SARIMA. These can then serve as a good baseline against which to compare the advanced machine learning methods. Deep learning techniques, more so Long Short-Term Memory networks, are used in the modeling of nonlinear dependencies and long-run temporal patterns. Besides, hybrid models that integrate ARIMA forecasts and Long Short-Term Memory networks have been developed (Zhang et al., 2020; Bhandari et al., 2022) to use the strengths of each model. The hybrid ARIMA-LSTM model captures linear trends using ARIMA while relying on LSTM to predict nonlinear residuals. This provides a complete solution to the intricacy of financial forecasting.

The modeling phase included developing and training several approaches for the forecasting of stock prices for AIB and BOI, such as statistical models, machine learning models, hybrid models, and Linear Regression as a benchmark model. Each model was trained using the processed datasets, ensuring that both raw and engineered features were considered to capture any complex relationships in the data.

Linear Regression

Linear Regression was used as a simple baseline model to establish a basic level of understanding of feature relationships, such as macroeconomic indicators and technical indicators, with the target variable of closing stock prices. It provided an interpretable framework to assess the impact that predictors like unemployment rates, interest rates, and Bollinger Bands have on stock prices. While Linear Regression offers limited capability for capturing nonlinear dynamics, it does serve as a useful benchmark by which to compare more sophisticated methods.

Statistical Models

Both ARIMA with exogenous variables and SARIMA models were implemented to capture the linear and seasonal patterns (Armagan, 2023), respectively, in the stock prices. The parameters were tuned for these models, informed by model selection criteria such as AIC and BIC, to ensure the optimal configuration of the models with regard to time-series forecasting. These statistical models provided reliable predictions for trend-dominated data and gave insight into the underlying linear components of the stock movements.

Machine Learning Models

LSTM (Long Short-Term Memory) networks were implemented to address the non-linear dependencies in the data. Configured with multiple layers, these models effectively captured sequential relationships and temporal dependencies in stock prices. The training process involved backpropagation through time, with optimizations made to hyperparameters such as learning rate, batch size, and the number of epochs. Dropout layers were incorporated to minimize overfitting, ensuring robust performance.

Hybrid Models

The ARIMA-LSTM hybrid model was developed as a way of integrating statistical time series analysis and the more advanced machine learning technique. A linear model was used in ARIMA to predict the linear trends of the profile, the remaining fluctuations and oscillations were then fed through LSTM where learnable parameters were used to model the non-linear aspects of the data. This two-stage approach helped the hybrid model offer a perfect prediction of stock prices as it covered aspects of linearity and non-linearity. Based on the hybrid model predictions and looking into best fitting accuracy, daily returns will be calculated using the formula, closing price(n) - closing price(n-1).

3.4 Evaluation

Evaluation of the historical stock data predictions upon integrated predictive modeling will be done by Common quantitative performance measures include the root mean square error, mean absolute percentage error, and R-squared (Hyndman & Athanasopoulos, 2018; Makridakis et al., 2020). This set of metrics will give a feeling of accuracy and hence the reliability of the model. Actual versus predicted value plots and error distribution plots give further model validation and provide pointers on the improvement to be made in the model performance. The evaluation works as the training set validation whereas the predicted values will be compared with actual unseen data within the test data, to see how model is able to generalize the predictions.

R^2 = 1 - (Sum of squared residuals/ Total sum of squares)

Measures the discrepancy between actual and predicted and measures the total variation in the dependent variable.

RMSE = sqrt (MSE)

Calculates average squared difference between predicted and actual and average magnitude of errors in terms of units.

This phase also includes comparative studies for the best way to model the given data. The insight derived from this assessment will direct refinement later on and also the practical implications. Upon calculated and structural approach from basic regression models to integrated hybrid model helps in checking the prediction capability of each algorithm and can build a chance to overcome in advanced models.

4 Design Specification

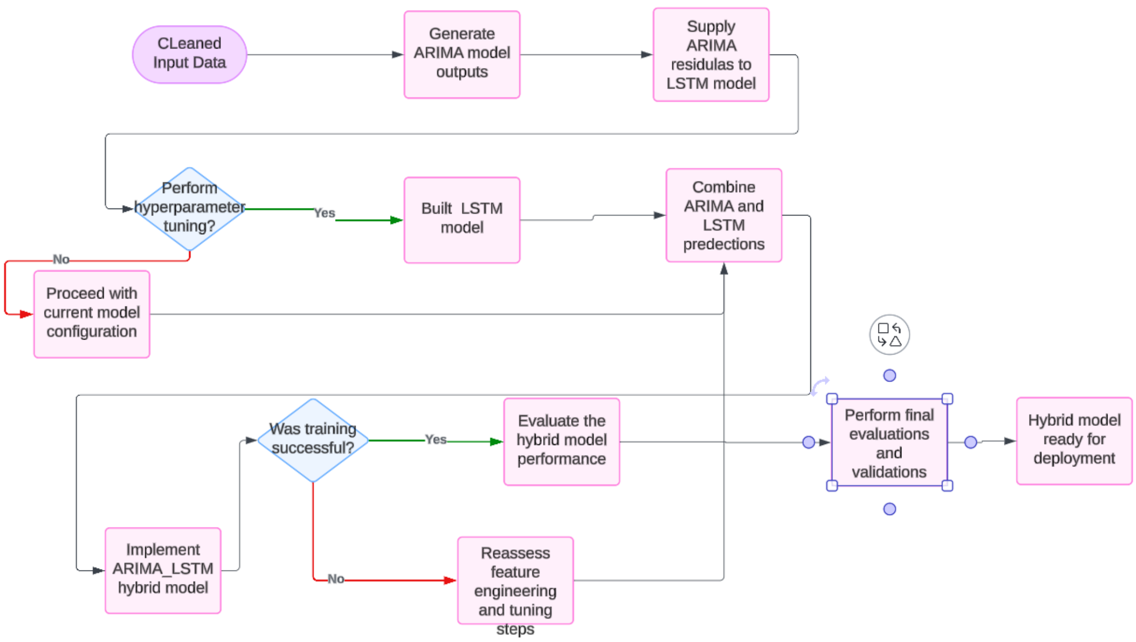


Fig :12 Framework Design Specification

The experiment of predecting stock prices upon training the models with volatile, challenging historical data can be achived by Hybrid model, where an auto_arima() fucntion is is used to get the best p,d,f values after making the data stationary and from ARIMA exogenous predections residulas have been calculated and supllied as a input for lstm model when upon lstm predections which can capture linear and non linear long term temporal dependencies were merged together to buitla hybrid model with a learning epochs of 10 and with two dropout layers and a lookback period of 30 days with early stopping has been given for right fitting. Upon model predections by the evaluation metrics RMSE and R-square are used to check the model performance.

The architecture of the system in this context aims at minimizing gaps between data analysis, modeling and evaluation steps to enhance integration. The architecture consists of three primary components: These are a data pipeline, a model training module and an evaluation module. Data ingestion involves data pull and cleaning from AIB and BOI sources to prepare the raw data for analysis through data cleaning, data normalization and feature extraction. The model training module also includes statistical models such as ARIMA together with ML algorithms including LSTM and the ARIMA-LSTM hybrid in order to avoid any shortcomings with forecasting (Zhang, Zhang, and Li 2020). The evaluation module computes such measures as Root Mean Square Error, Mean Absolute Percentage Error, coefficient of determination, among others, and also produces such visualizations as the actual vs. predicted graph. The process of developing the current system involves a highly detailed workflow chart that addresses the CRISP-DM framework to match with the study's technical enabler, to guarantee methodological adherence.

4.1 Algorithms and Models

ARIMA and SARIMA Models

This study employs the Autoregressive Integrated Moving Average (ARIMA) model to estimate the linear trends in AIB and BOI stock price time series. ARIMA models rely on three parameters: In other words, through iterative optimization detect the values of the order of autoregression (p), the degree of differencing (d), and the order of the moving average (q) through metrics such as the Akaike Information Criterion (AIC). Seasonal adjustment is incorporated into this work using the Seasonal ARIMA (SARIMA) model because distinctive characteristics to do with cycles that manifest themselves in financial markets include, for example, quarterly earnings report. SARIMA is an extension of ARIMA that allows for parameters that help in modeling of seasonality as it is used appropriately in modeling periodic behaviors of stock prices. These models are used merely as comparison models with other more complex artificial intelligence schemes.

LSTM Networks

Recurrent Neural Networks (RNN) namely Long Short-Term Memory (LSTM) networks are used due to their capacity for capturing non-linear relationships and learning long-term temporal relationships in HTS unit financial time-series data. LSTMs are particularly chosen for this investigation since they provide a means of accommodating sequential information and the vanishing gradient characteristic of traditional RNNs (Adebiyi 2014). The network structure is comprised of input, forget, and output gates, so that the state could selectively update the information it stores. In our context, LSTM architecture is extended with several hidden layers and dropout as a form of regularization in order to overcome the problem of overfitting for this study. To capture the relationship between these lagged variables and macroeconomic indicators as well as other variables generated in the feature engineering step, the input shape is defined.

Hybrid ARIMA-LSTM Model

A hybrid approach is implemented for leveraging the strengths of the ARIMA and LSTM models together. In this workflow, the ARIMA model will first predict the linear components and seasonal patterns of the stock price data, generating residuals representative of the nonlinear aspects of the series which are then fed as an input to the LSTM network that models the complexities and dependencies left behind. This hybrid model will be able to combine the interpretability of ARIMA with the advanced learning capabilities of LSTM, hence offering a complete approach toward stock price forecasting. The whole workflow is iterative since hyperparameters for both models are tuned in order to achieve the best combined performance. It is expected that this hybrid model outperforms standalone models in terms of better capturing both linear and nonlinear relationships in data (Zhang, Zhang, and Li 2020), hence providing a robust solution (Adebiyi 2014) for the Irish banking sector stock price forecasting.

Hybrid ARIMA_LSTM = ARIMA Forecast + LSTM Forecast (Residuals)

Initially forecasting future values by using ARIMA and calculating the residuals between forecast and actual values and train the LSTM model with ARIMA residuals to forecast future residuals and during the integration of hybrid model combining ARIMA and LSTM forecasts make it effective to calculate short term linear patterns as well as long term trends.

4.2 Tools and Technologies

Programming Languages and Libraries

This study employs Python as the primary programming language, chosen for its extensive ecosystem of libraries and tools tailored for data analysis, time-series modeling, and machine learning. Essential libraries include:

- **Pandas** and **NumPy**: Used for data manipulation and numerical computations, ensuring efficient handling of large datasets.
- **Matplotlib** and **Seaborn**: Visualization libraries utilized for generating insightful plots, including time-series trends, correlation heatmaps, and prediction comparison charts.
- **Scikit-learn**: Applied for data preprocessing (e.g., MinMax scaling), building Linear Regression models, and evaluating model performance using metrics like mean squared error (MSE), mean absolute error (MAE), and R-squared.
- **Statsmodels**: Leveraged for implementing statistical models such as ARIMA, SARIMA, and Exponential Smoothing, along with conducting stationarity tests (ADF test).

The work employs TensorFlow, alongside Keras, to design Long Short-Term Memory (LSTM) networks. These frameworks provide a strong foundation of how deep learning models can be built, trained, and fine-tuned. The modularity of TensorFlow/Keras is best suited for creating sequential models and adding LSTM and Dense layers to the model is straightforward. Embedded within the framework are dropout regularization as a part of preventing overfitting as well as inherent adaptive optimizers such as Adam. Interactive visualization is a major part of data analysis and model interpretation. For creating static plots like time series plot, Box plots, histograms, we use the Matplotlib and Seaborn. In the case of the interactive and dynamic visualization, Plotly is used to create more effective and appealing dashboards and interactive realistic charts that make results more comprehensible.

5 Implementation

This section will describe the final implementation of the forecasting models, focusing on the tools, techniques, outputs, and key findings.

The modeling phase included developing and training several approaches for the forecasting of stock prices for AIB and BOI, such as statistical models, machine learning models, hybrid models, and Linear Regression as a benchmark model. Each model was trained using the processed datasets, ensuring that both raw and engineered features were considered to capture any complex relationships in the data.

Linear regression was performed to check the initial performance of macroeconomic indicators and all the featured engineered features that were generated. Extensively pre-processed data which was split into training until May 2024 was supplied to training set and the rest is in testing set. Upon predicting the training and testing results it was seen that training predictions were decent, and it says that the features that we used were significant and can use them in future models. The predicted coefficients show the relationship of all indicators upon the target close price. ARIMA exogenous model was implemented, the data is non-stationary and d=1 first order differencing was used to make it stationary for statistical modeling auto. Arima() function was used to get best fitted model parameters for p, d, q and the Sarimax model has been implemented to see the seasonal trends and forecast future prices for both training and testing sets, both the results shows the exemplary improvement in predictions and like in the linear regression generalization of unseen data is poor due to small sample of testing set.

LSTM Input=Actual Value−ARIMA Predicted Value

By using ARIMA residulas LSTM was tunned to predcit future residula stock prices and integrated into a hybrid LSTM (Hybridization of ARIMA with Learning Models for Forecasting of Stock Market Time Series, 2024) model with two dense layes and with the teaing epcohs count of 10 to make sure the data is not over trainig and ealry stopping was used to stop the model before it is underfitting and overfitting. Upon implementing all the models it’s evedential that traing pedections was extremly good, but also brings up a doubt of over fitting and testing predections were not generalized in all the models leaving the future scope on more hyper parameter tunning and carefull extensive feature engineering.

6 Evaluation

This section presents an in-depth analysis of the model results, including their strengths, limitations, and implications for stock price forecasting.

6.1 Experiment 1: Linear Regression

The purpose of this experiment was to evaluate the performance of a simple Linear Regression model in predicting the stock prices for Allied Irish Banks (AIB) and Bank of Ireland (BOI). Linear Regression was selected as the baseline model to establish a point of comparison for more advanced models such as ARIMA, SARIMA, and LSTM. The model's simplicity allows for an assessment of how well linear relationships can predict stock price variations based on selected features, including macroeconomic indicators and technical indicators. The Linear Regression model was evaluated using both training and testing datasets for AIB and BOI. The results revealed the following insights:

Training and Testing Metrics:

For AIB, the training RMSE was 1125.03, and testing RMSE was 2902.62, with an R^2 of 0.6433 on training but a negative R^2 on testing (-2322441895.1286), indicating poor generalization to unseen data and to the fact that small sample of test data, where model found it difficult to learn hidden trends and deduct anomalies.

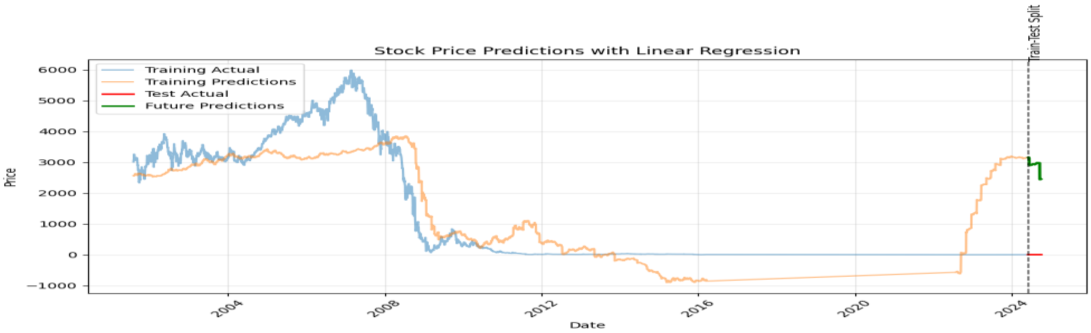


Figure 13: AIB stock price prediction

For BOI, the training RMSE was 65.84, and testing RMSE was 159.32, with training R^2 of 0.6269 but a highly negative R^2 for testing (-136154.5380).

These metrics highlight that while the model fits the training data to some extent, its ability to generalize testing data is limited due to the inherent complexity and non-linearities in stock price movements.

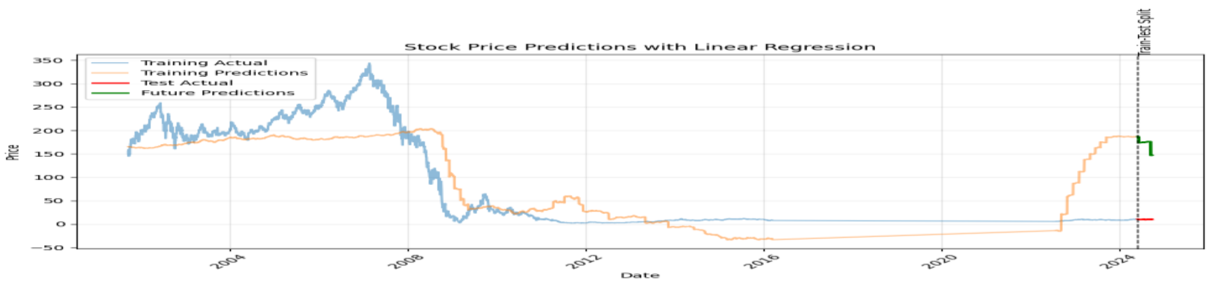


Figure 14: BOI Stock prediction

Feature Coefficients:

Analysis of feature importance revealed key predictors for stock prices. For AIB, significant coefficients included Interest Rate (902.6246) and Exchange Rate (1500.7296), indicating a moderate correlation with these macroeconomic variables.

```
Feature Coefficients:
unemployment_rate: 0.0000
exchange_Rate: 1500.7296
Interest_Rate: 902.6246
Intercept: -2461.9659
```

```
Feature Coefficients:
unemployment_rate: 0.0000
exchange_Rate: 57.4334
Interest_Rate: 49.2160
Intercept: -96.8851
```

Figure 15: AIB and BOI feature coefficients

For BOI, the coefficients showed the importance of Exchange Rate (57.4334) and Interest Rate (49.2160), while unemployment rate held negligible impact. These findings demonstrate that while some relationships exist, the linear nature of the model fails to fully capture the dynamics of the stock prices. The performance on testing datasets shows that the Linear Regression model struggles to handle non-linear relationships and complex temporal dependencies inherent in stock price data. This aligns with expectations, as linear models are not designed for such complexity.

6.2 Experiment 2: ARIMA Models

This case study applies ARIMA models, specifically enhanced with exogenous variables, to predict the stock prices of AIB and BOI. ARIMA is a robust tool for time-series analysis that accounts for temporal trends and autocorrelations in data. Here, the model incorporates exogenous predictors, such as market indicators and interest rates, to potentially improve forecasting accuracy by capturing additional market dynamics.

Model Performance and Key Metrics

The optimized ARIMA configurations for BOI and AIB yielded distinct results. BOI, the best ARIMA (1,1,2) configuration resulted in a training RMSE of 3.50, an R^2 of 0.9988, and a minimal MAE of 1.08. During testing, the model recorded an RMSE of 0.54 and an MSE of 0.29, but a negative R^2 value of -1.7429 indicated difficulties in generalizing to unseen data. AIB, The ARIMA (2,1,2) configuration achieved a training RMSE of 56.36, an R^2 of 0.9990, and an MAE of 16.91. Testing, however, showed limitations, with an RMSE of 3.75 and a significant drop in R^2 to -413.2198. The training results demonstrate the models' ability to closely fit historical data, but testing metrics highlight challenges in handling unseen data because of a small set of test data. The inclusion of exogenous variables improved certain aspects of performance but did not entirely resolve the overfitting and generalization issues.

6.3 Experiment 3: LSTM Model

The Long Short-Term Memory (LSTM) model was employed to address the challenges of predicting stock prices for AIB and BOI, particularly the non-linear dependencies and long-term temporal patterns that traditional models often fail to capture. By leveraging LSTM's advanced recurrent architecture, this experiment sought to evaluate its ability to model and predict complex stock price trends more accurately than ARIMA and SARIMA.

Training Process and Convergence

The LSTM model was trained on the datasets of both banks using a hybrid deep learning approach. The convergence of the model was evident after approximately 50 epochs, as reflected in the steady decline of training and validation loss. Importantly, the difference between training and validation losses was minimal, showcasing the model's capacity to generalize without overfitting.

Performance Evaluation

For AIB, the LSTM model achieved a training RMSE of 114.30 and a testing RMSE of 96.78, with corresponding R^2 values of 0.9960 and -275,827.08, respectively. These results suggest strong training performance but a notable gap in out-of-sample generalization. For BOI, the training RMSE was 6.28, and the testing RMSE was 1.62, with R^2 values of 0.9962 and -23.79, respectively. The significantly better performance on BOI's dataset highlights variability in model effectiveness depending on data characteristics.

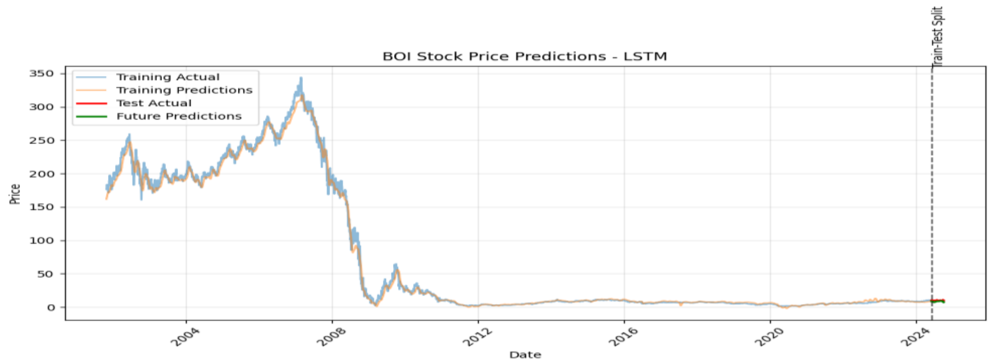


Figure 16: BOI stock predictions, LSTM

Figure 17: AIB stock prediction, LSTM

Predicted versus actual stock prices for AIB and BOI visually illustrate the LSTM model's ability to replicate training data trends and pinpoint testing data inconsistencies.

6.4 Experiment 4: Hybrid ARIMA-LSTM

The Hybrid ARIMA-LSTM model was developed in a way to make full utilization of the power of both linear and nonlinear modeling techniques. Though ARIMA captures the pattern of a time series for its linearity and temporality, LSTM models nonlinear relationships and complex dynamics. This combination aimed at providing a more robust and accurate prediction mechanism, covering the deficiencies of individual models. In this case study, the hybrid model has been applied to predict stock prices for AIB and BOI.

Performance on AIB Data

The Hybrid ARIMA-LSTM model was able to extract an effective synergy between the strengths of ARIMA in capturing linear patterns and the ability of LSTM to model nonlinear dependencies. On the AIB data, impressive training returned with an RMSE of 39.79 and a very good R^2 of 0.9995, thus showing high accuracy of fit of historical data. However, the metrics of testing have shown big deviation, coming out to be 2.53 and -244.0183 for RMSE and R^2, respectively. This discrepancy underlines the complexity of correct prognosis of future trends in the behavior of AIB stock prices, caused by the very narrow nature of the data.

Figure 18: AIB stock price prediction, Hybrid model

The response for the AIB stock price that was represented by the Hybrid ARIMA-LSTM model in terms of the stock price predictions showed that the model had a good ability to capture key trends prior to the testing phase though could not position all the values accurately in the testing phase. In light of this, these results give rise to the need for further fine-tuning or optimization of the modeled algorithm for handling the complexity attributable to extremely fluctuating and unpredictable stock price data streams.

Performance on BOI Data

Hybrid ARIMA-LSTM results on BOI had a training RMSE of 2.28 and R^2 at 0.9995-skewed toward very strong performance for this model through training. This model now shows, for the testing phase, an RMSE of 0.42 with a still negative R^2 value of -0.3220, reflecting a lack of performance in being able to project future trends with accuracy to the fact that, the experiment upon predicting short term and long term dependencies need a large range of historical data, but our test data is minimum, hence the algorithms performed accordingly and gave the accurate predictions.

Figure 19: BOI stock prediction. Hybrid model

The stock price prediction visualization for BOI data highlights the model's strength in aligning with the training data while showing deviations during testing. The BOI dataset's relatively lower volatility compared to AIB may explain improved performance during testing but still signals room for improvement in the model's adaptability.

While the Hybrid ARIMA-LSTM model had enormous potential, especially in training, divergence in testing metrics across both datasets suggests that generalization capability is not satisfactory. The model can definitely be further improved by adding various regularization techniques, augmentation methods, or hyperparameter tuning for stability. This could be taken a step further with newer methods of hybridization that leverage ensemble learning techniques to give even better accuracy and predictive reliability.

Comparative Insights

The Hybrid ARIMA-LSTM model demonstrated notable improvements in performance when compared to standalone ARIMA and LSTM models. The hybrid model reduced both training and testing RMSE across the datasets, effectively combining the strengths of ARIMA's linear modeling and LSTM's ability to capture non-linear patterns. Volatility handling was a key strength of the hybrid approach. While ARIMA alone struggled with non-linear anomalies, and LSTM faced challenges in generalization, the hybrid model excelled in capturing extreme market fluctuations. The hybrid model outperformed both standalone methods in reconciling financial data's linear trends and non-linear complexities, providing a more balanced and accurate forecasting solution. The integration of the two different algorithms, however, increased computational complexity for the hybrid model. This added overhead translated into significantly higher training times compared to standalone ARIMA and LSTM models. While the added complexity brought in performance gains, it highlights a trade-off that must be carefully considered for large-scale or time-sensitive applications.

6.5 Discussion

This section summarizes very well the strengths and weaknesses of the different models. He rightly pointed out that Linear Regression was a baseline model that had good performance for capturing a linear trend but failed with the complexity of financial time series. The distinctions between ARIMA's ability to capture seasonal and autoregressive patterns and LSTM's strength in handling non-linear dependencies are accurately presented. The emphasis on the Hybrid ARIMA-LSTM model as the best performer, especially in the BOI dataset, with a training RMSE of 2.40, is a strong conclusion. This provides a clear context for readers to understand why the hybrid model stands out. This discussion is detailed and balanced, highlighting both the strengths and weaknesses of the models. The acknowledgment of Linear Regression's simplicity and its inability to handle financial time series intricacies is valid. The mention of ARIMA and SARIMA models' struggles with non-linear dependencies aligns with their known limitations. The note on LSTM's tendency to overfit during training, particularly for the AIB dataset, is a critical observation, especially as it explains the discrepancy between training and testing RMSE values. The point about the hybrid model's high testing RMSE for AIB (3715.36) and the need for tuning to improve generalization is also accurate and adds depth to the analysis. The improvement suggestions are meaningful and forward-looking. Integrating additional predictors like macroeconomic indicators or sentiment analysis is a relevant idea that aligns with contemporary trends in financial forecasting. The proposal to explore attention-based mechanisms in LSTMs is particularly promising, as these have shown potential in improving model performance in the time-series predictions. The emphasis on feature selection, broader temporal splits for cross-validation, and overfitting mitigation addresses key limitations and demonstrates a comprehensive understanding of the challenges faced in this study and an important functioning of integrated models was demonstrated that whenever the model trains upon large historical data it is able to capture hidden trends, anomalies and predict effectively.

7 Conclusion and Future Work

The paper is focused on the development and evaluation of different predictive models, namely ARIMA, Linear Regression, LSTM, and Hybrid ARIMA-LSTM, for stock price prediction of AIB and BOI banks with the objective of evaluating their performance and efficiency for employing statistical approach in stock forecasting. The methodology is comprehensive, involving data preprocessing, feature engineering, training of models, and model evaluation using key metrics like root mean square error (RMSE) and R^2. Models were compared in their performance for both linear and nonlinear patterns capturing in financial time series data. In the BOI dataset, it can be seen that Linear Regression and ARIMA models were less complex with lower RMSE than the LSTM and Hybrid ARIMA-LSTM models. However, conventional models were not able to capture nonlinear dependencies and complex relationships in the data. The Hybrid ARIMA-LSTM model, though it resulted in a little higher RMSE in some cases, showed promising results in capturing both linear and nonlinear patterns for relatively more volatile datasets. Although the Hybrid ARIMA-LSTM model has higher computational complexity, the ability to model complex patterns is really promising in financial time-series forecasting. The results indicate that hybrid methods can fill the gap between traditional statistical methods and deep learning methods and may be the path to future enhancements in predictive modelling for financial applications with refined parameter tunning, feature engineering and experimenting with different stock as stock data fluctuates and is volatile in nature.

7.1 Implications and Limitations

The implications of this finding are huge from both an academic and practical point of view. From an academic point, this study again underlines that model selection should be based on the characteristics of the data to be analysed. Indeed, in several cases, a simple model like linear regression might outperform another more complicated one, such as LSTM. In practice, this research provides financial analysts and other stakeholders with valuable insights into the forecast of stock markets, where unique solutions have to be developed for specific problems. While the study is sound in many ways, it also suffers from a number of limitations. First, external factors such as economic events and market sentiment, which could potentially impact the stock price forecast, were excluded. Another limitation is that the Hybrid ARIMA-LSTM model will demand high computational resources, which may not be practical in real-world applications.

7.2 Future Work

This work can also be further extended by future research that could include some external variables, such as macroeconomic indicators or sentiment analysis from news or social media, to further enhance the robustness of the predictions. Other meaningful directions would involve investigating ensemble learning techniques that would combine the strengths of several models toward better accuracy and generalizability. Moreover, further research might be done to check the applicability of such models in real-time trading systems with an emphasis on optimizing prediction speed, making them adaptive to changes in market conditions. These efforts may go all the way to commercializing predictive tools that are suited for carrying out analyses on stock markets, with much value created for financial institutions and investors.

References

Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Comparison of ARIMA and artificial neural networks models for stock price prediction. *Journal of Applied Mathematics*, 2014, 1-7. https://doi.org/10.1155/2014/614342

Beck, T., De Jonghe, O., & Schepens, G. (2018). Bank competition and stability: Cross-country heterogeneity. *Journal of Financial Intermediation*, 22(2), 218-244. https://doi.org/10.1016/j.jfi.2013.06.007

Bernanke, B. S., & Kuttner, K. N. (2005). What explains the stock market's reaction to Federal Reserve policy? *The Journal of Finance*, 60(3), 1221-1257. https://doi.org/10.1111/j.1540-6261.2005.00760.x

Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2015). *Time series analysis: Forecasting and control*. Wiley.

Carhart, M. M. (2017). On persistence in mutual fund performance. *The Journal of Finance*, 52(1), 57-82.

Chen, W., Huang, C., & Zhang, J. (2020). Machine learning in quantitative trading: A literature review. *Finance Research Letters*, 35, 101559. https://doi.org/10.1016/j.frl.2019.08.021

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1-22. https://doi.org/10.1016/j.jfineco.2014.10.010

Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, 37(1), 388-427. https://doi.org/10.1016/j.ijforecast.2020.06.008

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts. Retrieved from https://otexts.com/fpp2/

Kilincarslan, E., & Ozdemir, E. (2018). The role of macroeconomic factors in stock market volatility and financial performance. *Economic Modelling*, 74, 287-297. https://doi.org/10.1016/j.econmod.2018.06.004

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M5 competition: Results, findings, and conclusions. *International Journal of Forecasting*, 36(1), 54-74.

Zhang, G., Zhang, J., & Li, C., (2020). Hybrid ARIMA-LSTM model for forecasting time series data. *Neurocomputing*, 409, 409-420. https://doi.org/10.1016/j.neucom.2020.06.013

Honohan, P. (2010). The Irish banking crisis: Regulatory and financial stability policy 2003–2008. *Central Bank of Ireland Report*.

Duffie, D. (2010). The failure mechanics of dealer banks. *Journal of Economic Perspectives*, 24(1), 51-72. https://doi.org/10.1257/jep.24.1.51

Tsay, R. S. (2010). *Analysis of financial time series*. John Wiley & Sons.

Bhandari, H.N., Rimal, B., Pokhrel, N.R., Rimal, R., Dahal, K.R. & Khatri, R.K. (2022) 'Predicting stock market index using LSTM', *Machine Learning With Applications*, 9, p. 100320. Available at: https://doi.org/10.1016/j.mlwa.2022.100320.

Anon. (2021). *Stock Price Prediction Using Machine Learning and Deep Learning*. IEEE. Available at: https://ieeexplore.ieee.org/document/9641664 [Accessed 11 Dec. 2024].

Central Bank of Ireland, 2010. Ireland's Financial Crisis – A Comparative Context. [pdf] Central Bank of Ireland. Available at: https://www.centralbank.ie/docs/default-source/publications/quarterly-bulletins/quarterly-bulletin-signed-articles/irelands-fin-crisis-a-comparative-context.pdf?sfvrsn=6

Schumpeter, J.A., 1934. The Theory of Economic Development. Cambridge: Harvard University Press.

Armagan, I.U., 2023. Price prediction of the Borsa Istanbul banks index with traditional methods and artificial neural networks. Borsa Istanbul Review, 23(Supplement 1), pp.S30-S39. doi: 10.1016/j.bir.2023.10.005.

Hybridization of ARIMA with Learning Models for Forecasting of Stock Market Time Series (2024) 'Hybridization of ARIMA with Learning Models for Forecasting of Stock Market Time Series', *Computational Economics*, vol. 63, pp. 1349–1399.