UNIVERSIDADE FEDERAL DE SERGIPE

CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Improving Stock Market Prediction with Feature Expansion and Explainable AI

Dissertação de Mestrado

Joao Paulo Euko

São Cristóvão – Sergipe

2024

UNIVERSIDADE FEDERAL DE SERGIPE

CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Joao Paulo Euko

Improving Stock Market Prediction with Feature Expansion and Explainable AI

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Sergipe como requisito parcial para a obtenção do título de mestre em Ciência da Computação.

Orientador(a): Leonardo Nogueira Matos

São Cristóvão – Sergipe

2024

Resumo

O mercado financeiro é um ambiente complexo, dinâmico e rápido, onde informações de qualidade, são altamente valorizadas. Neste contexto, a capacidade de acessar e interpretar dados financeiros de maneira eficiente pode significar a diferença entre o sucesso e o fracasso nas operações de mercado. As oscilações de preços e a volatilidade inerentes ao mercado de ações exigem análises contínuas e atualizações constantes.

O objetivo desta pesquisa é treinar modelos de aprendizado profundo e utilizar técnicas de explicabilidade para filtrar indicadores tanto para humanos quanto para máquinas, buscamos entender os fatores subjacentes que influenciam essas previsões. Com isso, pretendemos fornecer *insights* valiosos que possam ser utilizados para melhorar as estratégias de investimento.

Buscamos avaliar o desempenho dos modelos GRU e Conv2D em termos de tempo de execução e acerto de previsão. Para tanto, serão realizados diversos experimentos comparativos, onde diferentes arquiteturas de modelos serão testadas e ajustadas para otimizar seu desempenho. Além disso, a explicabilidade dos modelos será um foco central, visando garantir que os usuários possam confiar nas previsões geradas e compreender as razões por trás de cada decisão tomada pelos algoritmos.

Este estudo pretende contribuir para o campo de finanças computacionais, oferecendo soluções práticas e teóricas que possam ser aplicadas tanto por *traders* individuais quanto por grandes instituições financeiras. Através da implementação de técnicas de aprendizado profundo e da análise detalhada de seu desempenho, esperamos identificar as melhores práticas e propor melhorias que possam ser adotadas em futuras pesquisas e aplicações práticas.

Palavras-chave: xAI. Aprendizado Profundo. Mercado Financeiro.

Abstract

The financial market is a complex, dynamic, and fast-paced environment where high-quality, precise, and timely information is highly valued. In this context, the ability to access and interpret financial data efficiently can mean the difference between success and failure in market operations. Price fluctuations and the inherent volatility of the stock market demand continuous analyses and constant updates.

The objective of this research is to train deep learning models and utilize explainability techniques to filter indicators for both humans and machines. Through the use of advanced technologies, we aim not only to predict market movements but also to understand the underlying factors that influence these predictions. By doing so, we intend to provide valuable insights that can be used to improve investment strategies.

We seek to evaluate the performance of these models in terms of execution time and prediction accuracy. To achieve this, various comparative experiments will be conducted, where different model architectures will be tested and adjusted to optimize their performance. Additionally, the explainability of the models will be a central focus, aiming to ensure that users can trust the generated predictions and understand the reasons behind each decision made by the algorithms.

This study intends to contribute significantly to the field of computational finance, offering practical and theoretical solutions that can be applied by both individual traders and large financial institutions. Through the implementation of deep learning techniques and detailed analysis of their performance, we hope to identify best practices and propose improvements that can be adopted in future research and practical applications.

Keywords: xAI. Deep Learning. Stock Market.

List of Figures

List of Tables

List of abbreviations and acronyms

AI Artificial Intelligence

xAI Explainable Artificial Intelligence

POCID Prediction of Change in Direction

SHAP SHapley Additive exPlanations

VAE VAE (VAE)

OHLC Open, High, Low, Close

LLM Large Language Model

Contents

1

Introduction

One area that remains particularly challenging is understanding the indicators and features that influence price formation in stock market time series. This research originates from the pressing need to develop more effective methods for predicting stock market movements by leveraging advanced Artificial Intelligence (AI) techniques. Traditional methods often rely on technical indicators selected by traders, which can introduce human biases and may not capture the full complexity of market dynamics.

This work aims to address this gap by exploring the use of VAEs (VAEs) for feature expansion and Explainable Artificial Intelligence (xAI) techniques for feature selection. Notably, while autoencoders are typically used for dimensionality reduction, this study proposes using VAEs for dimensionality expansion to generate new features from raw closing price data, a novel approach not previously utilized in this context.

By focusing on the Prediction of Change in Direction (POCID) of stock prices, this research aims to develop a model that can better capture underlying patterns in the data. Additionally, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) analyses are incorporated to better understand temporal dependencies within the time series data.

A time window of 10 days is utilized for the analysis, a choice informed by ACF and PACF analyses that indicate significant autocorrelations within this period. The VAE is expected to improve the model by expanding the feature space, allowing it to learn more complex patterns that may not be evident from raw closing price data alone.

In this challenging environment, achieving even a modest improvement over baseline models like the Random Walk is significant. By comparing the proposed model against a Random Walk baseline, the aim is to demonstrate the merit of this approach and its potential contributions to financial market prediction.

It is common knowledge that significant advancements and breakthroughs have occurred in the field of Artificial Intelligence applied to various domains. These developments have led to remarkable tools, like "ChatGPT" (OPENAI, 2023), which have greatly enhanced productivity and expanded possibilities, even for those without coding or artificial intelligence knowledge. However, it is important to note that not all aspects of our lives have been equally impacted by these advancements.

Time series data can be analyzed and investigated to identify patterns and gain insights that help interpret the information it contains.

In the realm of financial market prediction, achieving high accuracy is particularly difficult. Various studies have highlighted the challenges and limitations in this field. For instance, algorithms in this domain often struggle to attain a desirable level of applicability, as noted by (LV; HOU; ZHOU, 2019). Despite these difficulties, even incremental improvements can be highly significant. Consider the work of (LENG, 2022a), which achieved around 57% accuracy; (NEVASALMI,), which reached 55% accuracy; and (JAQUART; DANN; WEINHARDT, 2021), which obtained 56% accuracy in predicting Bitcoin prices. These results underscore the inherent difficulty of the task and set a benchmark for evaluating new approaches. Given this challenging environment, achieving a result close to 60% is not just adequate but exceptional. It demonstrates significant progress, especially when considering the incremental nature of advancements in this field. The pursuit of higher accuracy, even by a small margin, represents substantial gains in practical applications, such as loss prevention and profit maximization. Therefore, a result close to 60% should be regarded as a noteworthy achievement, providing a solid foundation for further research and development in financial market prediction.

Time series data is present in various aspects of our lives, and the ability to predict events based on this data allows humanity to better manage resources. For example, by monitoring temperature readings, sunlight intensity, and average rainfall, humans have been able to understand the seasons of the year and determine the optimal timing for planting and harvesting crops. The use of sophisticated computers and Artificial Intelligence has further enhanced this knowledge. However, it's important to note that such advanced technology was not available when humans initially discovered these patterns.

While it is desirable for all problems to be as straightforward and simple as predicting the seasons of the year, it must be acknowledged that this particular problem has been studied for centuries. Extensive research and analysis have contributed to a better understanding of when environmental events will occur with greater precision. Over time, researchers have been able to filter out indicators that were initially thought to be useful in predicting seasons and have identified the most relevant ones.

With this perspective in mind, the focus is on forecasting the stock market and aiming to minimize errors when predicting the next value of an asset. This problem requires considering various indicators, such as price moving averages, economic indicators, political indicators, and any other indicators that traders or managers find relevant to their specific problem or asset.

As much as it is desirable for the problem to be straightforward, it is not that simple. If it were, everyone would be able to predict share prices before they happen, and that wouldn't be very interesting. Throughout the centuries, investors have been trying to forecast the next price movement in the stock market by utilizing the indicators mentioned earlier. Various strategies have been developed to aid in price prediction.

Due to the rapid and complex nature of price movements, it is essential to have a tool that is capable of processing information faster than the human eye and can capture patterns that may be difficult for humans to discern or require extensive training to identify. This is where deep learning comes into play. In the upcoming chapters, the discussion will focus on how deep learning, an artificial intelligence technique, has been proven to effectively predict price time series. However, it's important to note that the discussion will focus on historical data, and the real-time aspect will not be considered at this moment.

Finally, the goal is to develop a method that can filter through the multitude of indicators that investors, traders, and managers often get overwhelmed by. The idea is to identify and focus only on the most crucial indicators that accurately describe price formation. By understanding what is relevant for predicting a specific problem, performance can be significantly improved.

Given that the stock market operates at an incredibly fast pace, where microseconds can make a significant difference, it is essential to provide a solution that is both highly accurate in terms of prediction and efficient in terms of speed.

Ultimately, the goal is to reduce the dimensionality of the problem, making it easier for both humans and machines to comprehend and analyze.

1.1 Research Hypothesis

It is hypothesized that VAEs (VAEs) can effectively learn and represent the underlying structure of historical stock market data by capturing complex, non-linear relationships and expanding essential features. This feature expansion is anticipated to improve the predictive performance of machine learning models by providing a richer and more informative input space, thereby enhancing the models' ability to forecast market trends and price movements accurately.

Additionally, it is proposed that Explainable AI techniques, specifically "SHAP" (LUNDBERG; LEE, 2017), can be instrumental in identifying the most influential features contributing to model predictions. By enhancing model interpretability, SHAP is expected to provide a clear understanding of the factors driving predictions, offering valuable insights for financial decision-makers. This enhanced interpretability will enable more informed strategic decisions and facilitate the reduction of dimensionality by focusing on the most relevant features, thereby optimizing the computational efficiency of predictive models.

The integration of VAEs for feature expansion and SHAP for interpretability and feature selection is hypothesized to create a synergistic effect that improves both the accuracy and transparency of predictive models in financial forecasting. This combined approach aims to empower traders and investment professionals by delivering models that not only achieve higher prediction accuracy but also provide meaningful insights into the data's behavior, supporting more effective risk management and strategic planning in the financial markets.

1.2 Objectives of the Research

The main goal of this work is to improve model accuracy and interpretability in stock market prediction by creating indicators with VAEs (VAEs) and selecting relevant features with xAI techniques.

Currently, there are two approaches commonly used to filter information and indicators in stock market time series. One approach is the use of VAEs, a deep learning technique that aims to compress information, reduce dimensionality, and then reconstruct the compressed information to measure the loss of information. The ideal scenario is to have the output of the reconstruction match the input through an encoding layer.

VAE methods have proven to be effective in various problems, including time series analysis, and they serve as a necessary step in preprocessing the data before training the final model (LV; HOU; ZHOU, 2019). However, the encoding layer, which represents the compressed data, is not easily interpretable by humans. Nevertheless, it works well for reducing dimensionality in machine-based analysis.

Another solution that effectively reduces dimensionality is a technique called Explainable Artificial Intelligence (xAI). xAI attempts to determine the importance of features (in our case, stock market indicators) in the prediction. This technique determines the relevance of each input feature, allowing for manual or automated filtering based on a hyperparameter threshold.

Explainable AI appears to fulfill both the requirement of comprehensibility for humans and machines, as it provides insights into the importance of features. It allows for filtering and focusing on the most relevant information. However, as mentioned earlier, the speed aspect needs further investigation.

The proposal is to investigate whether Explainable AI can effectively reduce dimensionality and, if so, to what extent it can improve prediction performance. As this proposal is written, there are already several notable works in the field that provide explainability, such as SHAP and LIME. Further research will be conducted to explore the best alternative in terms of speed and prediction performance.

It is hypothesized that VAEs (VAEs) can effectively learn and represent the underlying structure of historical stock market data by capturing complex, non-linear relationships and expanding essential features from the closing price data. This feature expansion is anticipated to improve the predictive performance of machine learning models in predicting the change in direction (POCID) by providing a richer and more informative input space, thereby enhancing the models' ability to forecast market trends and price movements accurately.

Additionally, it is proposed that Explainable AI techniques, specifically SHAP, can be instrumental in identifying the most influential features contributing to model predictions. By enhancing model interpretability, SHAP is expected to provide a clear understanding of the factors driving predictions, offering valuable insights for financial decision-makers. This enhanced interpretability will enable more informed strategic decisions and facilitate the reduction of dimensionality by focusing on the most relevant features, thereby optimizing the computational efficiency of predictive models.

1.2.1 Specific Objectives

To ensure that the research stays on track with the objectives, specific goals have been set to aid progression towards the ultimate goal.

- Conduct a systematic review of VAEs and Explainable AI, specifically focusing on their application in the stock market or related domains. If necessary, consider studies utilizing similar data or structures.
- Replicate time series forecasting experiments to establish a baseline for dimensionality reduction and interpretation.
- Apply various methods of dimensionality reduction and measure their impact on prediction performance and processing time.
- Analyze the results and evaluate the ability of xAI techniques to provide human-interpretable insights for time series forecasting.
- Assess the effect of excluding specific features initially considered irrelevant on prediction performance to enhance both machine and human performance.
- Assess the effect of including specific features considered irrelevant to have a sanity check and determine if they indeed worsen the model.
- Deliver production-ready code for the implemented methods and techniques.

The following chapters will discuss the Theoretical Foundation, Related Works, Initial Results, Next Steps, Methodology, Proposal, and Project Phases.

2

Theoretical Foundation

In this chapter, a foundational background is provided for the topics discussed in the following chapters. It is assumed that the reader has prior knowledge in artificial intelligence, deep learning, and time series analysis applied to the stock market.

2.1 Theoretical Foundation

2.1.1 Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

The Autocorrelation Function (ACF) measures the correlation between a time series and its own lagged values over successive time intervals, helping in identifying patterns such as trends or seasonality in the data. The Partial Autocorrelation Function (PACF) measures the correlation between the time series and its lagged values while controlling for the correlations at shorter lags.

Analyzing ACF and PACF plots is crucial for determining the appropriate window size for time series models. In this study, these analyses indicated that significant autocorrelations exist within a 10-day window for the stock market closing prices. This informed the choice of using a 10-day window to capture the essential temporal dependencies in the data.

The ACF and PACF plots generated from the dataset are shown in Figure 1. The significant spikes within the first 10 lags in both plots justify the use of a 10-day window for the models.

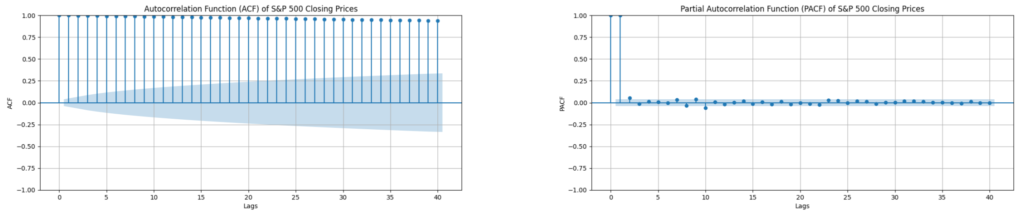


Figure 1 – (a) Autocorrelation Function (ACF) plot of S&P 500 closing prices. (b) Partial Autocorrelation Function (PACF) plot of S&P 500 closing prices.

Interpretation of ACF and PACF Plots:

ACF Plot: The ACF plot shows significant autocorrelation at lag 1, which gradually decreases but remains above the significance threshold up to lag 10.

PACF Plot: The PACF plot exhibits a significant spike at lag 1 and minor spikes up to lag 10, indicating that the first 10 lags contribute meaningful information for predicting future values.

$$\phi_{kk} = \frac{\mathrm{Cov}(X_t, X_{t-k} \mid X_{t-1}, X_{t-2}, \ldots, X_{t-(k-1)})}{\sqrt{\mathrm{Var}(X_t \mid X_{t-1}, TimeWindow_{k-1})} \cdot \mathrm{Var}(X_{t-k} \mid X_{t-1}, TimeWindow_{k-1})}$$

In both images the blue zone represents the 95% confidence interval, which is used to determine the significance of the autocorrelation values.

If the blue bar falls outside the confidence interval, the correlation is considered significant. In this case, the lag 10 are significant, indicating that the data has a strong temporal dependency within this window.

2.1.2 Variational Autoencoders (VAEs) for Dimensionality Expansion

VAEs (VAEs) are generative models that learn a probabilistic mapping from a high-dimensional input space to a lower-dimensional latent space and vice versa (KINGMA; WELLING, 2014). While VAEs are commonly used for dimensionality reduction, in this research, they are uniquely employed for dimensionality expansion. By encoding the closing price data into a higher-dimensional latent space, VAEs generate new features that capture complex, non-linear relationships inherent in the data.

How VAEs Improve the Model:

- **Capturing Non-Linear Patterns:** VAEs can model complex distributions and capture non-linear dependencies that traditional models might miss.
- **Feature Generation:** The latent variables generated by the VAE serve as new features that expand the input space for the predictive model.
- **Regularization:** The VAE framework includes a regularization term in the loss function, which encourages learning smooth latent representations and prevents overfitting.

2.1.3 Explainable Artificial Intelligence (xAI)

Explainable Artificial Intelligence (xAI) is a technique designed to identify and highlight the features that have the most significant impact on model predictions. According to (ALI et al., 2023), *"feature attribution, which involves evaluating the relative significance of input features in a particular model's decision, is a common kind of post-hoc explanation."* This technique provides valuable insights into which features have a negative impact or do not contribute significantly to training a model. One of the key advantages of Explainable AI is its human interpretability, allowing researchers and practitioners to understand and interpret the results more easily. By offering transparency, xAI helps build trust in AI systems, ensuring that stakeholders can comprehend how decisions are made and which variables are driving these decisions.

2.1.4 Investigating Explainability Methods in Recurrent Neural Network Architectures for Financial Time Series Data

The first step was to explore the existing work on xAI in the context of time series forecasting, and the paper (FREEBOROUGH; ZYL, 2022) proved to be a valuable starting point. This paper investigates various models, including GRU (Gated Recurrent Unit) and LSTM (Long Short-Term Memory), which are also utilized in the experimentation. The results obtained by (FREEBOROUGH; ZYL, 2022) were promising, as they stated, *"The results show that these methods are transferable to the financial forecasting sector, but further confirmation is needed for a more sophisticated hybrid prediction system"* (FREEBOROUGH; ZYL, 2022). This work demonstrated that it is possible to use explainability methods in time series forecasting, particularly in the stock market.

2.1.5 Explainable AI and Adoption of Financial Algorithmic Advisors: An Experimental Study

Motivated by the previous paper that demonstrated the potential of xAI in financial time series forecasting, an exploration was conducted into whether people are willing to trust explanations provided by artificial intelligence. A study by (DAVID; RESHEFF; TRON, 2021) developed a game where players had to decide how many lemons to buy for producing lemonade to be sold. As the game progressed, several reports were provided, including one generated by an artificial intelligence system. The results obtained by (DAVID; RESHEFF; TRON, 2021) revealed that people were willing to trust an artificial intelligence report when it could provide clear explanations for the decisions made. This experiment highlights the importance of transparency and the ability of xAI to enhance user trust in AI-driven systems.

2.1.6 SHAP

Since the previous subsection mentioned that SHAP will be used, it is important to understand what exactly SHAP is. The SHAP (SHapley Additive exPlanations) framework operates by taking a trained model and iterating through all its features. It evaluates the importance of each feature by considering their individual contributions and averaging the differences between the actual and expected values. SHAP incorporates properties such as *"local accuracy, missingness, and consistency"* (LUNDBERG; LEE, 2017).

The contribution of each feature to a prediction is computed using the following formula:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \left[f(S \cup \{i\}) - f(S) \right]$$

Where:

- ϕ_i represents the SHAP value for feature i,
- N is the set of all features,
- S is any subset of N excluding i,
- $f(S)$ is the model's prediction based on the features in subset S.

This formula calculates the contribution of each feature to the model's prediction, providing insights into the importance of individual features. SHAP has been widely adopted in various domains due to its ability to explain complex models and provide interpretable results.

This research explores different variants of SHAP and compares them with LIME and linear layer models to explain feature importance. Considering the availability of information and its suitability for explainable AI, SHAP has been chosen as the framework for this work. By leveraging SHAP, deeper insights into model behavior and feature significance can be gained, enhancing the overall interpretability of the AI systems.

2.1.7 The Best Way to Select Features? Comparing MDA, LIME, and SHAP

With the awareness of several free and open-source xAI frameworks available in the market, it was necessary to decide which one to use. After reading the (MAN; CHAN, 2021) paper, SHAP was chosen for several reasons. According to (MAN; CHAN, 2021), SHAP demonstrated stability and consistently provided reliable results. Additionally, the SHAP framework is compatible with various deep learning frameworks, such as "TensorFlow" (AUTHORS, 2023). Although several other options like "LIME" (RIBEIRO; SINGH; GUESTRIN, 2016) were considered, in the end, the advantages of SHAP outweighed the alternatives. The choice of SHAP is based on its ability to provide consistent and meaningful explanations, making it a powerful tool for feature selection and model interpretability.

2.2 Losses

2.2.1 Loss Function Mean Squared Error

The loss function measures the deviation between the predicted values and the actual values. There are several ways to calculate it, and different loss functions work better for different types of problems. Sometimes, more than one loss function can be suitable for a particular problem. For this problem, the Mean Squared Error (MSE) loss function will be used. The MSE squares the errors, which has been proven effective in reducing errors when predicting time series data. The mathematical representation of MSE is as follows:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2.1}$$

Where:

- n: Size
- y_i: Actual value of the target variable for the i-th instant
- $\hat{y}_i$: Predicted value of the target variable for the i-th sample

The MSE is widely used in regression problems due to its simplicity and effectiveness in penalizing larger errors more heavily than smaller ones.

2.2.2 R2 Score - Coefficient of Determination

The coefficient of determination, R-squared, aims to quantify the *"influence of one condition on another"* (WRIGHT, 1921). It serves as a valuable metric to assess the accuracy of predictions, especially in the context of time series analysis. R-squared allows understanding how well the predicted values match the original data, providing a measure of goodness-of-fit. The mathematical formula for R-squared is:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum i = 1^n (y_i - \bar{y})^2} \tag{2.2}$$

Where:

- R^2: Coefficient of determination.

- n: Number of samples in the dataset.
- y_i: Actual value of the target variable for the i-th sample.
- $\hat{y}_i$: Predicted value of the target variable for the i-th sample.
- $\bar{y}$: Mean of the actual target values.

The R-squared value ranges from 0 to 1, where a value closer to 1 indicates a better fit of the model to the data.

2.2.3 MAPE

The mean absolute percentage error (MAPE) metric is chosen because it demonstrates *"the consistency of the Empirical Risk Minimization"* (MYTTENAERE et al., 2016), which aligns with the problem aimed to be addressed: minimizing risk and increasing the expandability of the indicator. MAPE is particularly useful for understanding the accuracy of predictions in a relative sense. The formula for MAPE is:

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \tag{2.3}$$

Where:

- MAPE: Mean Absolute Percentage Error
- n: Number of samples
- Y_i: Actual value of the target variable for the i-th sample
- $\hat{Y}_i$: Predicted value of the target variable for the i-th sample

For this metric, the optimal value is zero, which indicates perfect accuracy. However, MAPE values can approach infinity, reflecting large errors in predictions. The further the result is from zero, the worse the prediction accuracy, but the result cannot be negative.

2.3 Related Work

The study cited as (KUMAR et al., 2020) explores an innovative approach utilizing SHAP for dimensionality reduction while simultaneously providing explanations for the most relevant features. The authors conclude that *"although the SHAP library is actually built for feature interpretation, it can also be effectively used for dimensionality reduction"* (KUMAR et al., 2020). However, the databases used in that study differ from those this proposal aims to investigate, which may affect the direct applicability of the findings to the research context.

Another relevant study applies SHAP to reduce dimensionality in a psychology dataset. Although the dataset used in their study is not similar to the one in focus here, the methodology employed is akin to this approach. The results demonstrate that *"SHAP values appear to yield the highest model accuracy"* (WILKERSON; LEAKE; CRANDALL, 2022). The authors also discuss the potential benefits of removing many features, which could facilitate explaining similarities between cases and improving model interpretability (WILKERSON; LEAKE; CRANDALL, 2022).

It is valid to mention (SELVARAJU et al., 2017) GradCam as a method to visualize the importance of features in Convolutional Neural Networks (CNNs). This method is used to generate heatmaps that highlight the regions of an image that are most relevant for the model's prediction.

In the financial domain, (LV; HOU; ZHOU, 2019) employs a Stacked Denoising VAE to predict the direction of price movements. While their main goal is to utilize autoencoders for trading applications, this proposal aims to use autoencoders to expand features using raw data, adding a layer of feature engineering intended to enhance prediction accuracy.

An innovative approach is seen in a study exploring the utility of ChatGPT-4 in managing tasks and predicting outcomes based on news prompts. The findings indicate that *"the average earnings announcement and attractiveness ratings are both positive,"* suggesting the potential of using advanced language models in financial decision-making processes (PELSTER; VAL, 2024). The methodology and prompt details are shared, providing a foundation for replication and further exploration.

A study integrating stock market data, including prices and textual information from platforms like Twitter/X and same-day news articles, employs various models, including VAE, Gated Orthogonal Recurrent Units (GORU), GRU, and LSTM, providing a comprehensive comparison of their effectiveness. The results indicated that the model achieved *"comparative experimental values on the Stocknet dataset with an accuracy of 57.081%"* (LENG, 2022b), highlighting the challenges and potential of integrating textual and numerical data for stock prediction.

Another paper proposes a hybrid approach combining a small set of technical indicators with machine learning models to predict market directions. The authors conclude that *"it is more profitable to make trading decisions using a combination of technical indicators with computational intelligence tools"* (DASH; DASH, 2016). This approach underscores the synergy between traditional trading strategies and modern AI techniques, potentially offering a more robust decision-making tool in volatile markets.

In the study (PFEIFER; MAROHL, 2023), the authors fine-tune RoBERTa, a large language model, to analyze and predict the sentiment of FED speeches. They manually labeled over 6,000 speeches across different sectors, assigning a sentiment score of 1 for positive and 0 for negative speeches. Their findings indicate that *"all LLMs outperform all non-LLM models"* (PFEIFER; MAROHL, 2023), showcasing the superiority of fine-tuned language models in processing complex linguistic data.

A paper discussing the application of various machine learning models to predict the S&P 500 index uses a dataset similar to the one used here but over a different time range. The results for the five models tested showed accuracies ranging from 52% to 61% on the training set, stabilizing around 55% on the test set (NEVASALMI, 2020). These findings highlight the variability and challenges inherent in financial market prediction.

Lastly, a study proposes using machine learning methods to predict short-term movements in the Bitcoin, gold, and oil markets. Several models, including GRU and LSTM, were used to attempt predictions over timeframes ranging from 1 minute to 60 minutes. Their best performance was 56% accuracy in the 60-minute timeframe, providing insights into the feasibility and limitations of short-term market predictions (JAQUART; DANN; WEINHARDT, 2021).

3

Dataset and Accuracy Context

3.1 Introduction

In this chapter, we discuss the dataset used for our experiments and provide context around the importance of accuracy in stock market predictions. Understanding both the data and the implications of predictive accuracy is crucial for appreciating the value of our models, particularly in the volatile and complex environment of the stock market.

3.2 Dataset

For both the image recognition and time series challenges, we utilized daily Open, High, Low, Close (OHLC) data from the S&P 500 on NASDAQ, representing the most traded companies on the exchange. The dataset spans from August 30, 2013, to August 29, 2023, with each row representing the price movement for a single day.

3.2.1 Data Preprocessing for Image Recognition

For image recognition, the data undergoes preprocessing, where it is transformed into candlesticks, a visual representation of price movements on a chart. This preprocessing is accomplished by converting the CSV containing historical prices into a candlestick chart, cropping the image, and saving it for later use. Following the methodology outlined in (FRANKEN, 2023), the data is grouped into blocks of 10 candles. This grouping allows for the observation of price movements, and each group is labeled based on whether the closing price of the sixth candle is above or below the closing price of the tenth candle.

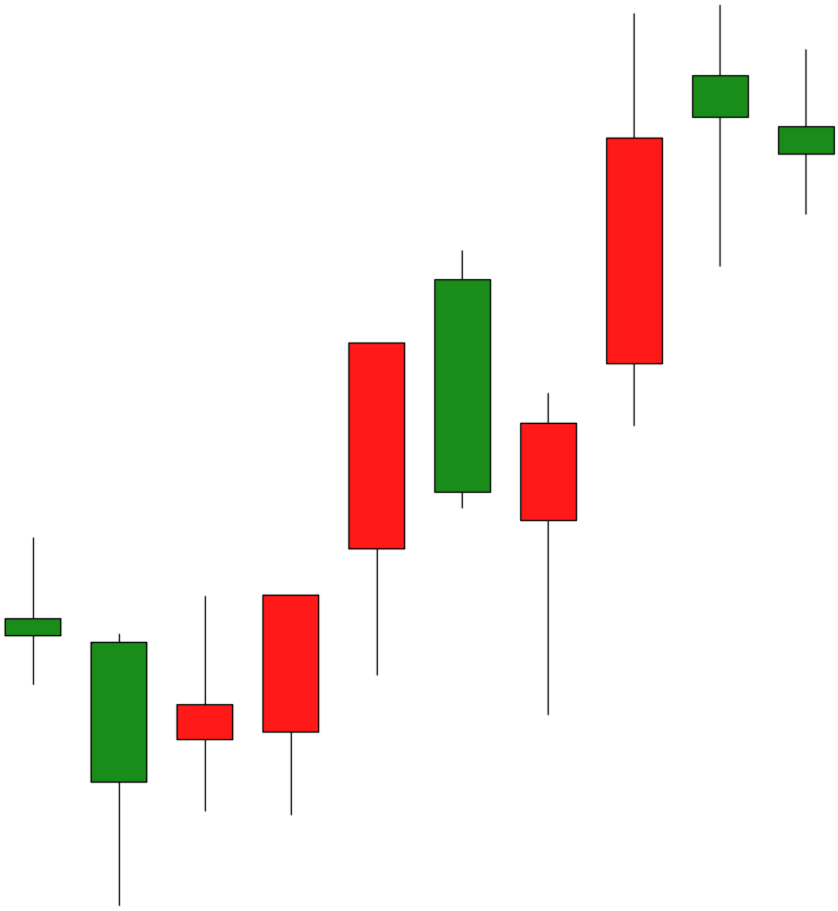


Figure 2 – Visualization of Financial Data: A 10-candle image representing grouped price movements in the S&P 500 on NASDAQ, where each block of candles captures the market dynamics over a specific time period.

Figure 7 presents an example of the obtained image. Each bar, referred to as a candlestick, conveys information about the open, close, high, and low prices. To enhance comprehension, the candlesticks are color-coded: red indicates that the price closed below its open, while green indicates closure above the open. The thin line atop the larger body indicates the maximum price at that moment, and the thin line at the bottom denotes the minimum price at that moment.

3.2.2 Data Preprocessing for Time Series Analysis

For the time series challenge, the data also underwent preprocessing. In some tests, only the closing price was used, which was the case for all Large Language Model (LLM) experiments. In other tests, moving averages of different timeframes were calculated and concatenated with the raw data to provide additional context. Part of the experiment involved using the closing price for feature expansion, while in other cases, the open, high, low, and close prices were used for this purpose.

3.3 Understanding Accuracy in Stock Market Predictions

Accuracy is a critical performance metric in predictive models, especially for binary outcomes such as stock price movements. However, it's important to provide context around what constitutes a meaningful accuracy rate, particularly when compared to random guessing or other baseline models.

3.3.1 Coin-Flip Analogy

To illustrate the importance of even small improvements in accuracy, consider a coin-flipping game where there is a 50% chance of guessing the correct outcome. Achieving an accuracy of 50% is equivalent to random guessing. However, if we can improve this probability—even slightly—through a predictive tool, it can significantly impact outcomes over time.

This analogy applies to stock market prediction, where the outcomes are binary (e.g., predicting whether the stock price will go up or down). By running multiple rounds of predictions with varying probabilities of success, even a small increase in predictive accuracy can have a substantial financial impact. Winning 60% of the time in such a scenario could lead to significant long-term gains.

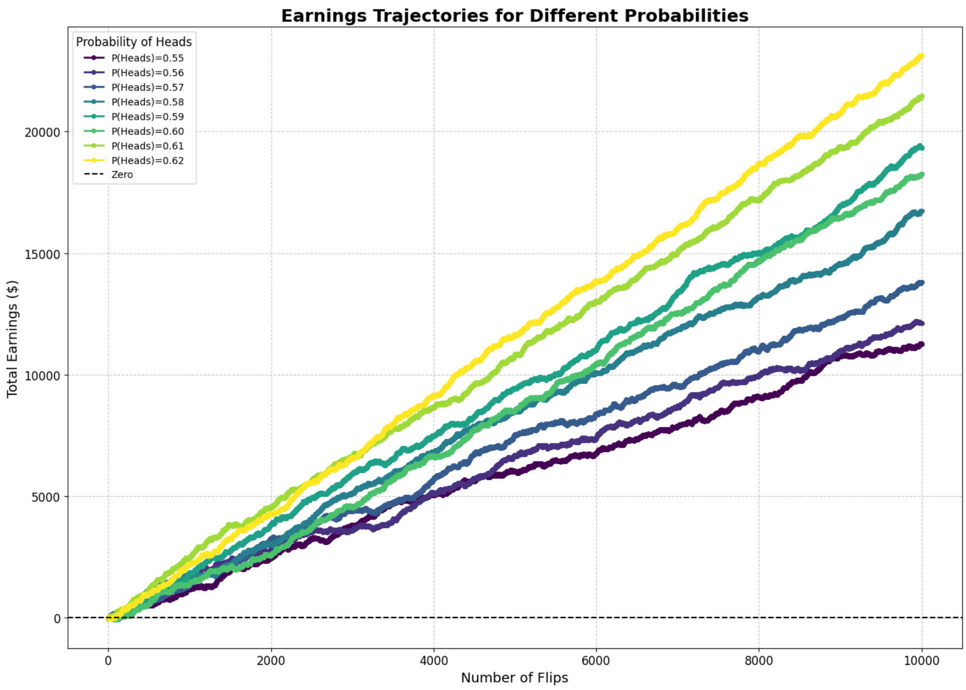


Figure 3 – Earnings trajectories for different probabilities of success in a coin-flipping analogy.

3.3.2 Significance in Financial Context

Achieving results around 60% accuracy is impressive in financial predictions. Although this may seem modest, in complex systems like the stock market, a consistent accuracy rate above random guessing can be leveraged for significant profit over time. Small improvements in prediction accuracy can have a large financial impact due to the compounding effects over numerous trades.

3.4 Stock Market Context

The stock market is an inherently volatile and complex environment. Traders are constantly seeking tools and techniques that provide them with an edge in predicting stock price movements. Machine learning models can offer this edge when developed and applied correctly.

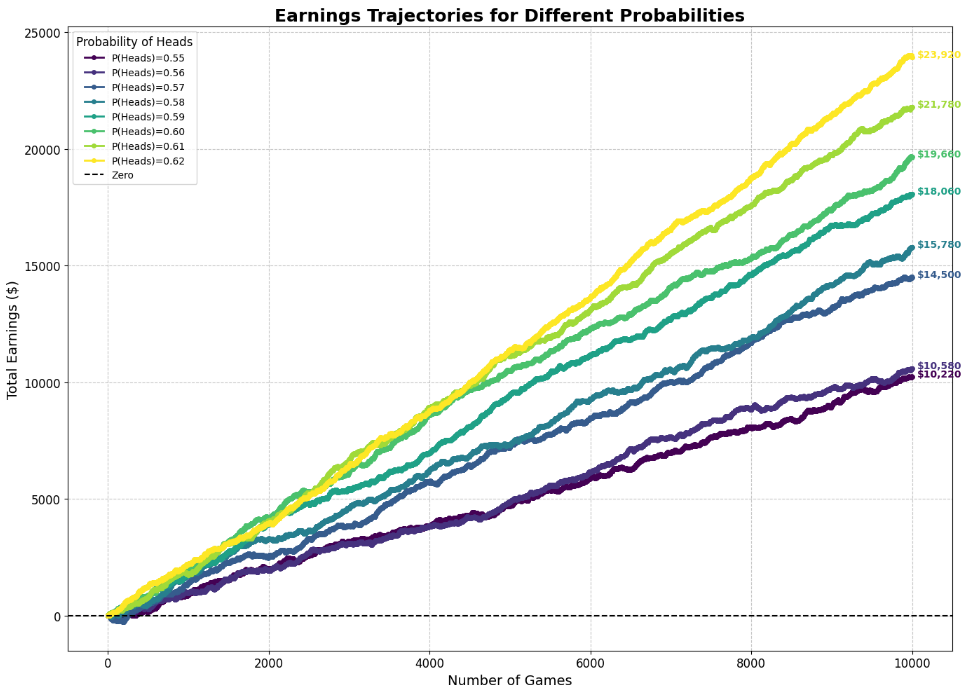


Figure 4 – Stock market prediction context illustrating earnings potential with small accuracy improvements.

Explanation of the Figure: Figure 4 further illustrates the cumulative earnings for varying probabilities in a stock market prediction context. Over 10,000 prediction instances, the smallest increase in accuracy (from 55% to 62%) results in progressively larger cumulative earnings. The highest accuracy of 62% results in earnings close to $26,000, reinforcing the notion that even a modest increase in predictive accuracy can result in large financial gains in the stock market.

3.4.1 Implications for Predictive Modeling

Predictive models that achieve a higher-than-random accuracy rate (above 50%) can generate consistent profits in financial markets. When compounded over time, these gains become highly valuable. Machine learning models, such as those developed in this work, seek to exploit patterns within historical data to offer better-than-random predictions, assisting traders in making more informed decisions.

3.5 Conclusion

Understanding the dataset and the context of accuracy is crucial for appreciating the value of machine learning models in stock market prediction. Even modest improvements in accuracy can have significant financial implications due to the cumulative nature of trading gains. Techniques such as xAI further enhance the interpretability of these models, making them more practical and trustworthy for decision-making in volatile markets.

3.6 Note for the Reader

The following chapters will delve deeper into two significant works. The first, *"Stock Market Prediction: Integrating Explainable AI with Conv2D Models for Candlestick Image Analysis,"* was published at the WorldCist 2024 Conference. The second, *"Stock Market Prediction: Integrating Feature Expansion, Explainable AI, and LLM Forecasts,"* has been submitted to a journal and is currently under review. These works explore advanced methodologies in stock market prediction, utilizing the dataset and accuracy context discussed in this chapter.

4

Integrating Explainable AI with Conv2D Models for Candlestick Image Analysis

In this chapter, the integration of Explainable AI with Conv2D models for analyzing candlestick images to predict stock market movements is explored. This chapter was published at the WorldCist 2024 Conference (EUKO; SANTOS; NOVAIS,). The importance of understanding stock market movements and the role of candlestick charts in this process are discussed, highlighting the complexities involved in predicting market trends. The chapter details the collection of daily OHLC data from the S&P 500, its transformation into candlestick images, and the grouping of this data into blocks of 10 candles. The training of Conv2D models on these images is described, focusing on predicting price movements based on the closing price of the 10th candle relative to the 11th candle.

Further, the application of Explainable AI techniques such as SHAP and GRAD-CAM is investigated to provide interpretability of these predictions. The evaluation of model performance is based primarily on accuracy, comparing predicted labels with the ground truth. Through this approach, the effectiveness of integrating explainability techniques with Conv2D models in providing insights into model behavior and improving predictive accuracy is demonstrated. The chapter concludes with a discussion on the insights gained from the experiments, the challenges faced, and the potential future directions for research in this domain. This research leverages explainability techniques to improve the interpretability and accuracy of Conv2D models for candlestick prediction in stock market images.

4.1 Introduction

In the dynamic, fast-paced financial market, where every second counts, having information and understanding why the market is moving in a certain direction can be immensely valuable. Information is the key to success, and decoding the nuances of the stock market presents a formidable challenge. The complexity of stock market time series, influenced by emotions, economic indices, politics, wars, and various other subtle factors that are difficult to discern, demands a need for understanding.

In this intricate dynamic world, where every second has a price, the significance of time series emerges. This form of data aggregates the temporal evolution of a stock market, taking into consideration price movement, emotions, indicators, and outsourced information. Being able to understand and predict these sequential patterns is key for assertive decision-making in the landscape of the financial market. When delving further into this challenge, the journey begins by exploring deep learning applied to image time series forecasting, an incredible tool to uncover patterns that humans struggle to identify or that take more time than should be allocated.

With this understanding in mind, this work aims to make time series prediction for the financial market clearer and more understandable. It explores deep learning applied to time series forecasting in an effort to predict the patterns generated by such information and explain these patterns.

Time series data is present in various aspects of our lives, and the ability to predict events based on this data allows humanity to better manage resources. For example, by monitoring temperature readings, sunlight intensity, and average rainfall, humans have been able to understand the seasons of the year and determine the optimal timing for planting and harvesting crops. The use of sophisticated computers and Artificial Intelligence has further enhanced this knowledge. However, it is important to note that such advanced technology was not available when humans initially discovered these patterns.

While it is desirable for all problems to be as straightforward and simple as predicting the seasons of the year, it is understood that this particular problem has been studied for centuries. Extensive research and analysis have contributed to a better understanding of when environmental events will occur with greater precision. Indicators that were initially thought to be useful in predicting seasons have been filtered out, and the most relevant ones have been identified.

With this perspective in mind, the focus is on forecasting the stock market with the aim of minimizing errors when predicting if an asset is going up or down. This problem requires considering various indicators, such as price moving averages, economic indicators, political indicators, and any other indicators that traders or managers find relevant to their specific problem or asset. However, in this approach, the concentration is solely on OHLC prices. As much as it is desired for the problem to be straightforward, it is not that simple. If it were, everyone would be able to predict share prices before they happen, and that would not be very interesting. Throughout the centuries, investors have been trying to forecast the next price movement in the stock market by utilizing the indicators mentioned earlier. Various strategies have been developed to aid in price prediction.

Due to the rapid and complex nature of price movements, it is believed that it is essential to have a tool that is capable of processing information faster than the human eye and can capture patterns that may be difficult for humans to discern or require extensive training to identify. This is where deep learning comes into play. In the upcoming sections, how deep learning, an artificial intelligence technique, has been proven to effectively predict price time series will be discussed. However, it is important to note that the discussion will focus on historical data, and the real-time aspect will not be considered at this moment.

Finally, the goal is to develop a method that can explain the decisions made during the forecasting process, exclusively utilizing OHLC data. The idea is to identify and focus only on the most crucial indicators that accurately describe price formation. By understanding what is relevant for predicting a specific problem, performance can be significantly improved.

4.1.1 Objectives of the Research

Explainable Artificial Intelligence (xAI) represents a sophisticated subfield devoted to enhancing model interpretability. Specifically applied to stock market patterns in this context, xAI aims to determine the significance of features in the prediction process. This subfield involves assessing the relevance of each input feature, facilitating manual or automated filtering or validation based on a hyperparameter threshold. Notably, approaches such as GRAD-CAM and SHAP show promise in meeting the demand for comprehensibility among both humans and machines. By offering insights into the importance of features, they enable strategic filtering and focusing on the most pertinent information for improved interpretability.

The proposal involves investigating whether xAI can effectively enhance the interpretability of Conv2D models, improve prediction performance, and clarify which features are being more considered. As of now, there are already notable works in the field that offer explainability, such as SHAP and LIME, and further research aims to explore the best alternatives, considering both speed and prediction performance.

This study explores the integration of Conv2D models and explainability techniques, specifically SHAP and GRAD-CAM, to enhance the interpretability of these models in predicting candlestick patterns. The primary focus is on delivering more answers than questions.

4.2 Method

The analysis method involves collecting daily OHLC data from the S&P 500, transforming it into candlestick images, and grouping the data into blocks of 10 candles. These grouped images are labeled based on the closing price of the 10th candle relative to the 11th candle. The primary focus is on Conv2D models for predicting price movements. Explainability is enhanced using SHAP and GRAD-CAM techniques. The main evaluation metric is accuracy, comparing predicted labels with ground truth. The experimental design includes training Conv2D models and applying explainability techniques.

4.3 Model Architecture

For the image challenge, the primary focus is on Conv2D models. These models undergo training and comparative analysis to determine their performance in predicting price movements of stock market candlestick images.

The architectural details of the Conv2D models utilized in the experiments are presented below.

$$\boxed{\text{Input}} \longrightarrow \boxed{\text{Conv2D}} \longrightarrow \boxed{\text{Flatten}} \longrightarrow \boxed{\text{Dense}} \longrightarrow \boxed{\text{Output}}$$

Figure 5 – Model Architecture Diagram

4.4 Evaluation Losses

The most important metric and the principal one is accuracy, as the objective is to predict the next candlestick closing price. The accuracy is calculated by comparing the predicted label with ground truth.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

The ground truth is the label of the next candlestick closing price. The accuracy is calculated by dividing the number of correct predictions by the total number of predictions.

4.5 Experimental Design

The experiment consists of gathering historical data, converting it to images containing five candles each, and labeling it with the next candlestick closing price. After that, the data is split into train and validation sets. The train set is used to train the Conv2D model, and the test set is used to evaluate the model. The accuracy is calculated by comparing the predicted label with ground truth, representing the label of the next candlestick closing price. Afterward, SHAP and GRAD-CAM are applied as explainability techniques to the Conv2D model, and the model is retrained. The accuracy is calculated again and compared with the accuracy of the original Conv2D model.

Both images from xAI from non-filtered and filtered, as well as the accuracy level, are compared.

4.6 Model Training and Accuracy

The Conv2D models were trained using images generated from S&P 500 OHLC data. To facilitate this, the tabular data underwent a preprocessing step where it was transformed into images, as illustrated in Figure 7. Each image in the dataset represents a group of price movements within a block of 10 candles.

The training objective was to predict whether the closing price of the 11th candle would be higher or lower than the closing price of the 10th candle. The models were evaluated based on their accuracy in predicting this directional movement.

Figure 6 – 10-candle image for price movement prediction in the S&P 500 on NASDAQ.

The models achieved an accuracy close to 100% (see Figure 7) and had an error close to zero during training. However, in the validation data, the accuracy remained around 55%.

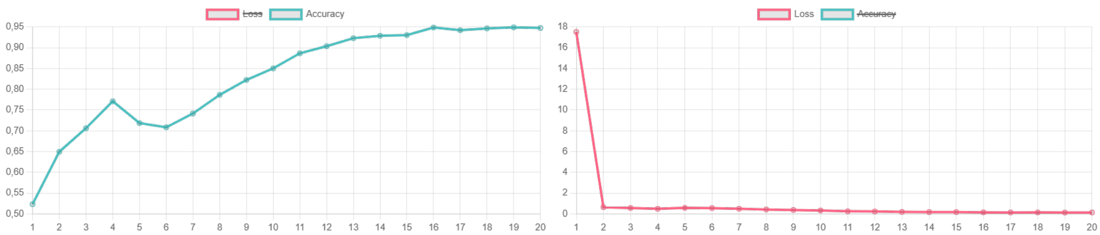


Figure 7 – Accuracy (a) and Loss (b) charts during training, with the x-axis representing the number of epochs and the y-axis indicating the corresponding metric.

4.7 Discussion

4.7.1 Explainability

In the previous section, Figure 7 was utilized for predicting the next candle, yielding a prediction accuracy of 99.09% for a green candle. Remarkably, the prediction proved accurate, as the subsequent candle indeed turned out to be green.

Upon closer inspection, the price trend displayed a succession of red candles, suggesting a potential market decline. To delve deeper into the market dynamics during this period, Figure 7 provides insights.

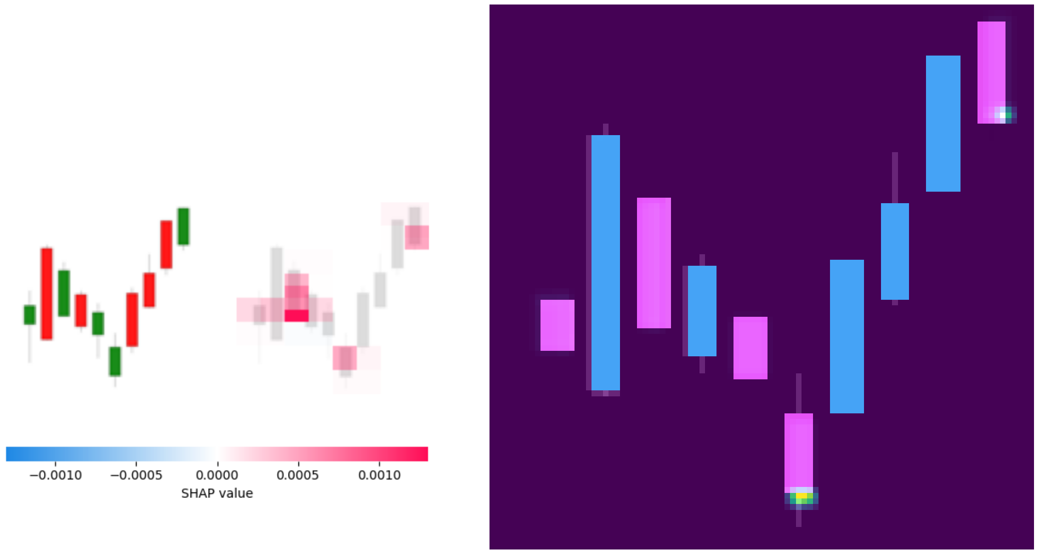


Figure 8 – Explainability: The leftmost image is the Original, the middle image is SHAP, and the rightmost image is Grad-CAM.

It is possible to see in Figure 7 three images. The leftmost one is the original image, sharing information with the SHAP visualization, and on the rightmost is the GradCam visualization. It is interesting to note that in both methods, the most important region is almost the same. The green candle in the middle of the image, representing the price reversion for that moment, had its open price checked by the models, and the open price on the last candle was also important for the correct price prediction. SHAP also had a good amount of importance on the green candle before the one in the middle.

4.7.2 Insights Gained

The results obtained from the experiments revealed that the Conv2D models were able to predict the next candlestick closing direction with high accuracy in training. However, the accuracy of the models in predicting the direction of the next candlestick closing price was significantly lower in validation.

This is understandable since it is a very complex task to predict the next candlestick closing price direction. The models were able to learn the patterns in the training data, but they were not able to generalize well to the validation data.

However, the ability to explain the market was found to be satisfactory and indicated that the model is looking at the right places to make the prediction.

4.8 Conclusion

This research delved into the domain of the stock market, exploring the utilization of Conv2D models for predicting the next candlestick closing price direction. Additionally, the application of Explainable Artificial Intelligence (xAI) techniques was investigated to provide insights into the predictions made by these Conv2D models.

The primary goal was to elucidate the predictions of the Conv2D models and offer a deeper understanding of why specific predictions were generated.

The results obtained (1) from the experiments indicated that the Conv2D models exhibited high accuracy in predicting the next candlestick closing direction during training. However, their accuracy significantly decreased when predicting the direction of the next candlestick closing price in validation.

Dataset	**Accuracy (%)**
Train Set	99
Test Set	55

Table 1 – Accuracy of the model on the Train and Test sets

This decline in accuracy is comprehensible, given the intricate nature of predicting the next candlestick closing price direction. It also emphasizes the necessity for further investigation into potential overfitting and the exploration of more robust models.

The use of xAI produced significant results, providing satisfactory outcomes that indicated the model's ability to focus on relevant information for making predictions.

Despite the potential decrease in model performance during validation, the impressive explainability intuition underscores the significance of acknowledging the limitations faced during experimental development, particularly concerning computational power.

In conclusion, this research showcases the viability of employing xAI to explicate Conv2D models for image classification in stock market prediction. While the results warrant further investigation, they establish a solid foundational point.

Future efforts will involve replicating the experiment with an improved computational infrastructure and an expanded dataset to assess whether the model can generalize more effectively. This endeavor aims to address the encountered limitations in this research.

5

Integrating Feature Expansion, Explainable AI, and LLM Forecasts.

In this chapter, the potential of improving stock market prediction capabilities by integrating feature expansion techniques, Explainable AI, and Large Language Model (LLM) forecasts is explored. Initially, VAEs (VAEs) are leveraged to augment the dimensionality of traditional stock market data, represented by open, low, high, and close prices (OHLC). Following this, Explainable AI methodologies are applied to conduct feature selection on the expanded dataset, aiming to encapsulate pertinent information for model training. Finally, LLM-based approaches are employed to forecast market direction and provide interpretability alongside predictions. The findings indicate that while feature expansion boosts predictive performance, Explainable AI methodologies yield comparable results. Nevertheless, there is significant potential for improvement in the performance of LLM models for prediction and explanation. This chapter details the methods, experiments, and results of the investigation, providing insights into the effectiveness of these advanced techniques in stock market prediction.

5.1 Introduction

Technical analysis in the stock market involves studying historical price and volume data to predict future price movements. It is a widely used method, utilizing open, low, high, and close prices (OHLC), as well as volume, to calculate indicators used by traders and automated systems for prediction.

These technical indicators provide valuable insights into market dynamics, representing market behavior and serving as foundations for traders to explain their decisions and formulate strategies. When applied to historical data, they enable traders to identify patterns and understand market behavior over specific time periods. However, reliance on technical indicators presents several challenges, particularly as the number of indicators increases.

One challenge is their reliance on historical data, which may not accurately reflect current market conditions and can result in a time lag between data updates and the present moment. Additionally, as traders use visual inspection of charts to identify patterns, an increase in the number of charts can overwhelm traders and make pattern identification difficult, slowing down the analysis process.

Furthermore, some indicators are hard to understand and interpret, and correlations between indicators can lead to redundancy or contradictory signals. Additionally, human limitations make it impossible to analyze all data and indicators comprehensively or identify all patterns effectively.

Despite these challenges, technical indicators remain widely used to help traders navigate the complexities of the stock market.

Recognizing the importance of technical indicators and the potential biases that may arise from their selection by traders, this study aims to address this issue by allowing machines to create their own "indicators." This is achieved by expanding the OHLC data using a VAE (VAE) and then using Explainable AI to select the most relevant features to train a new model.

Furthermore, Large Language Model-based (LLM) approaches are leveraged via API prompts to predict market direction and provide explanations. This investigation seeks to determine if LLM predictions can offer valuable insights to traders while maintaining high accuracy levels.

5.1.1 Objectives of the Research

The motivation for this research is to explore the potential of using VAEs to expand OHLC data and improve stock market prediction, to understand if it is possible to create new "indicators" for models to use. Additionally, the potential of using Explainable AI to select the most relevant features and train a new model to improve the model's performance is explored. Furthermore, the study aims to integrate Language Model-based (LLM) forecasts to offer actionable insights for traders and decision-makers within the stock market domain.

Moreover, the study seeks to leverage the potential of VAEs to expand OHLC data and improve stock market prediction, as well as to explore the potential of using Explainable AI to select the most relevant features and train a new model to improve the model's performance.

Since historical data in the stock market carries valuable information about psychology and market behavior, it is believed that VAEs can capture this information and improve the model's performance. It is also believed that Explainable AI can help understand the most relevant features and improve the model's performance, as VAE data is not human-readable.

By allowing the machine to capture such information using the same data that traders use to create their indicators, a new perspective on the stock market can be provided, potentially improving stock market prediction. This involves generating a new set of data by the machine and utilizing it, not by humans, in an attempt to reduce human bias.

In conclusion, the objective is to examine whether a machine learning model can capture stock market information akin to how human traders endeavor to do so with their indicators. While it is acknowledged that machines may not replicate the entirety of human insight, it is posited that they can discern information beyond human capacity. This potential is explored alongside the utilization of LLMs.

5.2 Theoretical Foundation

5.2.1 Context

The stock market has always presented a challenge for traders and investors, who continually seek to decipher the puzzle of predicting its movements. In their quest to predict the stock market, traders and investors explore new strategies and methods to understand its behavior, often by introducing indicators through historical data and developing new trading approaches.

However, technology is evolving and gaining autonomy in decision-making, even trading independently. With this in mind, it becomes intriguing to explore whether machines can create their own indicators and select data autonomously from raw datasets, as well as forecast direction and provide explanations for their predictions.

5.2.2 Problem Statement

Given the complexities and challenges associated with predicting the stock market, along with the biases that may arise from humans introducing technical indicators to historical data, this study aims to address these issues by allowing machines to create their own "indicators." By enabling machines to generate their own indicators, the study facilitates data expansion without human biases, providing a new perspective on the stock market.

5.2.3 Research Questions

This study aims to answer the following research questions:

Can VAEs (VAEs) effectively capture valuable information about market behavior from historical data in the stock market to improve the performance of predictive models?

Can xAI effectively select the most relevant features from the expanded data to train a new model and improve the model's performance?

Can Large Language Model-based approaches (LLMs) accurately predict market direction while providing explanations for their predictions?

5.3 Method

The research involves collecting daily OHLC data from the S&P 500, grouping the data into blocks of 10 rows. These grouped data are labeled based on the closing price of the 10th day relative to the 11th day.

With the data prepared, the next step is to predict the market direction using LLMs and define the base case, where a GRU model is trained using only the closing price. Subsequently, the model is evaluated based on its accuracy in predicting the direction of the next closing price.

After that, VAE models are trained with different architectures and hyperparameters. The VAE model is trained using two differentiations of the raw data: one considering only the closing price for feature expansion, and a second strategy using the four prices (Open, High, Low, and Close) to expand the data.

With the VAE models trained, the data can be expanded and saved for later use.

The subsequent step involves utilizing the expanded data to train another GRU model. This model is trained using the expanded data and evaluated based on its accuracy in predicting the direction of the next closing price.

The final step is to use SHAP to select the most relevant features from the expanded data and train a new GRU model. This model is trained using the selected features and evaluated based on its accuracy in predicting the direction of the next closing price.

The models are evaluated based on their accuracy in predicting the direction of the next closing price. The ground truth is the label of the next closing price. The accuracy is calculated by dividing the number of correct predictions by the total number of predictions.

Each part of the experiment is executed 32 times, with the exception of the LLMs, and the average results are recorded for comparison.

5.4 Model Architecture

For this challenge, the primary focus will be on Figure 9 VAE and Figure 10 GRU models. These models will undergo training and comparative analysis to determine their performance in predicting price direction of stock market and if it can do feature expansion.

The architectural details of the VAE model utilized in the experiments are presented below.

Figure 9 – VAE Model Architecture Diagram

The hyperparameter used can be found on table 2 in hidden dim.

Input → GRU Encoder → Latent Space → GRU Decoder → Output

Figure 10 – GRU Model Architecture Diagram

5.5 Evaluation Metrics

Accuracy will be the main metric for the experiment, since the objective is to determine if the model is capable of predicting the direction of the next closing price. The accuracy is calculated by dividing the number of correct predictions by the total number of predictions.

5.6 VAE Results

Below we can see on Table 2 and Table 3 the validation accuracy of several models trained with different hyperparameters and architectures.

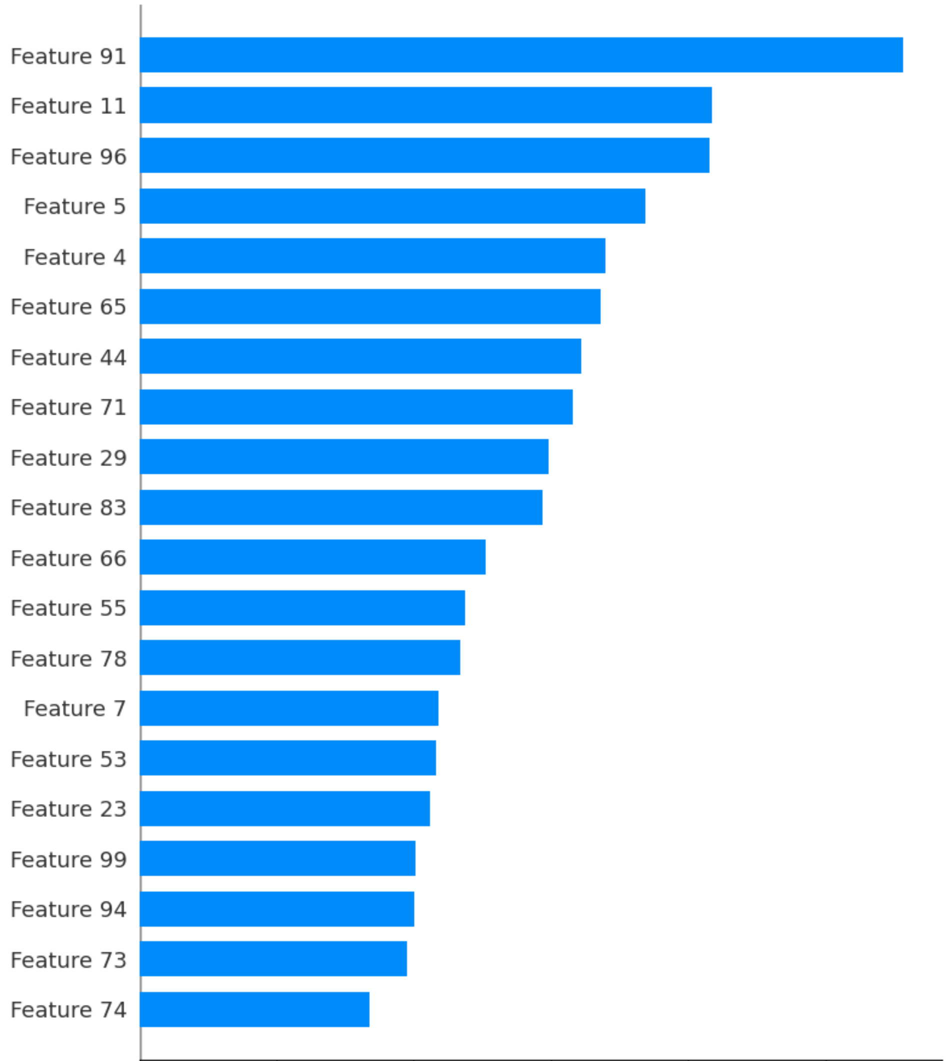
Base data	**Hidden dim**	**Features Expanded**
Closing Price	512	100
Closing Price Shap	512	20
OHLC	512	100
OHLC Shap	512	1000
OHLC	8	1000
OHLC Shap	8	20
Closing Price	8	1000
Closing Price Shap	8	20
Closing Price	16	1000
OHLC	16	1000
OHLC Shap	16	20
Closing Price Shap	16	20
Closing Price	512	1000
Closing Price + 100 Moving Average	-	64
Closing Price Shap	512	20
Closing Price	1024	100
Closing Price Shap	1024	20
OHLC	128	100
OHLC Shap	128	1000
Random Choice	-	-

Table 2 – Validation Accuracy (Part 1)

And here we can see the SHAP values in Figure 11 for the most important features for the model trained with the OHLC Expanded (100 Features) Shap Filtered 20 Features architecture.

Base data	**Filtered Size**	**Validation Accuracy**
Closing Price	**-**	**0.5935**
Closing Price Shap	20	0.5935
OHLC	-	0.5909
OHLC Shap	5	0.5909
OHLC	-	0.5851
OHLC Shap	20	0.5851
Closing Price	-	0.5846
Closing Price Shap	20	0.5846
Closing Price	-	0.5842
OHLC	-	0.5836
OHLC Shap	20	0.5836
Closing Price Shap	20	0.5825
Closing Price	-	0.5814
Closing Price + 100 Moving Average	-	0.581
Closing Price Shap	20	0.5796
Closing Price	-	0.5771
Closing Price Shap	20	0.5771
OHLC	-	0.5769
OHLC Shap	5	0.5769
Random Choice	-	0.4744

Table 3 – Validation Accuracy (Part 2)



5.7 LLM Results

In this section, the results of the LLM experiments are discussed.

The experiment utilized locally hosted LLMs to predict market directions and provide explanations for these predictions. Hosting models locally allows for greater control over the experimental environment, reducing dependencies on external services and minimizing latency issues. The models were run on an RTX 2060 6GB GPU, which influenced the processing capabilities and time required for inference.

To interact with the LLMs, the Ollama API was used, facilitating the sending of prompts and receiving predictions and explanations in a structured format. This API allowed for a streamlined process to query the models using consistent and well-defined prompts, ensuring uniformity across different model runs.

Despite the advantages of local hosting and streamlined API interactions, several limitations were identified. Running LLMs locally on an RTX 2060 6GB GPU limited the size and complexity of the models that could be used. Larger models, which might provide better performance, require more computational power and memory. The time required for inference varied significantly between models, with some models taking much longer than others. This variation impacts the feasibility of using certain models in time-sensitive applications.

5.7.1 LLM Prompt

For all the LLM models, the same prompt was used:

"As a time series analyst, your goal is to predict whether the next price in the series will be up or down. You will be provided with a time series of prices. Analyze the series and provide a prediction for the next price direction along with an explanation for your prediction. Your responses should be in a JSON format, using this template: 'direction': 'up/down', 'explanation': 'Provide your explanation here'. Do not give any information outside the JSON. Here is the time series" + str(data)

Python language was used to call the Ollama API running locally and get the predictions.

5.7.2 Results

These models are available via the Ollama API, which offers several other models beyond the ones used in this experiment. The models listed are among the most popular multipurpose open-source models that are compatible with the current computational capabilities available in this setup. Each model has its strengths and weaknesses, influencing their performance in terms of accuracy and data loss. The results are shown in Table 4 and Table 5.

The same data used in the VAE experiments was also used for prediction, and the results are shown below. The inference set consists of 2020 time windows, the same as used in the VAE.

Model	Time	Average Inference Time (seconds)
mistral:7b	2h 57m	5.2623
llama2-uncensored:7b	4h 43m	8.4139
llama2:7b	7h 22m	13.1255
gemma:7b	13h 53m	24.7538

Table 4 – LLM Inference time for 2020 time windows

On Table 5, we can observe the accuracy, data loss, and usable data length of the models.

A total of 2020 time windows were utilized for inference, and the data loss percentage is calculated by dividing the number of unusable data by the total number of data. Data loss signifies instances where the model couldn't provide answers as expected. Although all models have responses for all 2020 time windows, some of the answers were not usable for automating actions; they were considered data loss and removed from the accuracy calculation.

All the models have the temperature set to zero, which means they will prioritize precision over creativity.

Model	Usable data	Accuracy	Data Loss Percentage
gemma:7b	2019	0.41506	0.00050
mistral:7b	1814	0.46196	0.10198
llama2:7b	1760	0.47614	0.12871
llama2-uncensored:7b	1426	0.58766	0.29406

Table 5 – LLM Results - Accuracy, Data Loss, and Length

Below, sample responses for each model are shared, separated by a line.

5.7.3 Gemma Return Sample:

````
"```json
{'direction': 'down', 'explanation': 'The time series is showing a general downward trend, with the most recent price being lower than the previous peak. The trend is likely to continue, and the next price is likely to be down.'}
```

**Explanation:**

* The time series is showing a general downward trend, with the most recent price being lower than the previous peak.
````
````

```
* The trend is likely to continue, as the moving average is
  decreasing and the exponential smoothing average is also
  decreasing.
* The next price is likely to be down, as there is a high
  probability of the trend continuing."
```

```
"‘‘‘json
{’direction’: ’up’, ’explanation’: ’The time series is showing an
  upward trend, with the most recent price being at a higher level
  than the previous price. This trend is likely to continue,
  therefore the next price direction is predicted to be up.’}
‘‘‘

**Explanation:**

* The time series exhibits an increasing trend, with each
  consecutive price point being higher than the previous one.
* The most recent price is at a higher level than the previous
  price, indicating an upward trend continuation.
* Therefore, the next price direction is predicted to be up."
```

5.7.4 Mistral Return Sample:

```
{"direction": "up", "explanation": "The time series shows a general
  uptrend with some short-term fluctuations. The last two values in
  the series are higher than the previous ones, indicating a
  potential continuation of the uptrend."}
```

```
{"direction": "down", "explanation": "The time series shows a
  significant downward trend in the last few data points, with
  prices decreasing from 2090.11 to 2050.44. This suggests that the
  next price is likely to continue this trend and decrease
  further."}
```

5.7.5 Llama2 Return Sample:

```
"
Based on the provided time series, we predict that the next price will be up. Here's my explanation:

{ 'direction': 'up', 'explanation': 'The time series shows a clear trend of increasing prices over time. The most recent data point (2984.42) is higher than the previous data point (2976.61), which suggests that the upward trend is continuing. Additionally, the moving average of the time series is also trending upwards, which supports my prediction. Based on these observations, we believe that the next price in the series will be higher than the current price.' }"
```

```
"
Based on the provided time series, we predict that the next price will be down. Here's my explanation:

{ 'direction': 'down', 'explanation': 'The time series shows a clear trend of decreasing prices over time. The most recent price (1837.88) is lower than the previous price (1845.89), which suggests that the downward trend is likely to continue. Additionally, the slope of the line is negative, indicating a stronger downward momentum. Based on these factors, we predict that the next price will be lower than the current price.' }"
```

5.7.6 Llama2 Uncensored Return Sample:

```
"{'direction': 'down', 'explanation': 'The next price in the series is predicted to be down based on the following reasons:
- The current price is lower than the previous prices, indicating a downward trend.
- There are no significant changes or fluctuations in the series that could indicate an upward trend.
```

```
- The overall trend of the series is downward, which increases the
  likelihood of a further decrease in the next price.'}"
```

```
{'direction': 'down', 'explanation': 'The time series shows a
  downward trend, with the price decreasing over time. The most
  recent data point is 1926.82, which is lower than the previous
  data points. This indicates that the next price in the series
  will be down.'}
```

5.8 Discussion

In this section, the insights gained from the experiments and the limitations and future work are discussed.

5.8.1 VAE Results

After several experiments, it is evident that the model trained using only the closing price with 100 expanded features achieved the highest validation accuracy of 59.35

The model achieved a 1.58% improvement over the base model.

At first impression, it seems that there was no improvement or not enough relevance in any of the experiments.

However, most of the experiments yielded better results than the base case, even though the improvements in performance were minor.

5.8.2 LLM Results

Overall, 3 out of 4 models performed poorly, with only "llama2-uncensored:7b" achieving results close to the VAE experiment.

However, "llama2-uncensored:7b" had the most diverse responses, resulting in an almost 30% data loss. The algorithm failed to find any resemblance to the JSON format intended to be returned.

On the other hand, "gemma:7b" had the worst result in accuracy but was the best model in terms of automation, with only one response unable to be automated and returned.

5.9 Limitations and Future Work

It is understood that the stock market is a very complex environment and predicting its movements is a very hard task. The dataset used in this experiment includes only the S&P 500 index, and the model used is relatively simple. It is believed that with a larger dataset and a more complex model, better results could be achieved.

Due to hardware limitations, it was not possible to train more complex models and use larger datasets. With more computational power, better results could be achieved with more complex models, including larger LLMs such as ":70b" models.

In future work, more complex and elaborate VAE models will be employed to expand the data. Although the results shown are not the most impressive, it is believed that by expanding the data and improving the model's performance, significant enhancements are possible.

5.10 Conclusion

In conclusion, despite the observation that "most algorithms have not yet attained a desirable level of applicability" (LV; HOU; ZHOU, 2019), there is still considerable potential for improvement. A 1.58% enhancement in such a challenging environment is a noteworthy achievement, especially when compared with similar papers like (LENG, 2022b) that achieved results around 57% and (NEVASALMI, 2020) that achieved 55% on the test set, as well as (JAQUART; DANN; WEINHARDT, 2021) which obtained 56% when attempting to predict Bitcoin. In a complex environment, even a modest 1.58% improvement can translate into significant gains, whether it's in terms of loss prevention or increased profit. The results of this experiment are promising. With access to greater computational power and the implementation of more sophisticated models, even better outcomes are anticipated.

Additionally, the performance of the LLMs did not meet expectations. Perhaps larger models or non-open models could yield better results.

6
Discussion

In this chapter, we discuss the insights gained from the experiments, address the limitations encountered, and outline directions for future work.

6.1 Insights Gained

The experiments demonstrated a meaningful improvement over the Random Walk model, with an accuracy increase of approximately 12%. While this may seem modest, in financial contexts, such an improvement is significant due to the cumulative impact over numerous trades.

How the VAE Improves the Model:

- **Complex Pattern Capture:** The VAE captures non-linear relationships and complex patterns in the data that are not apparent in the raw closing prices.
- **Feature Enrichment:** By expanding the feature space, the model has access to a richer set of information, enhancing predictive capabilities.

Limitations of Baseline Models:

- **Random Choice Model:** Assumes stock prices follow a purely random process, which does not account for underlying market dynamics.

6.1.1 VAE Results

Through several experiments, the model trained using only the closing price with 100 expanded features achieved the highest validation accuracy of 59.35%. This model was trained with a hidden dimension of 512.

This represents a 1.58% improvement over the base model and a 12% improvement over the Random Walk model. While these gains may appear modest, they are significant given the complexity of financial forecasting. The improvement underscores the potential of using expanded feature sets to capture more nuanced patterns in the data.

Initially, it might seem that the improvements are not substantial. However, upon closer analysis, most experiments yielded better results than the base case, even if the performance gains were incremental. In the financial domain, small percentage improvements can translate into substantial financial impacts over time due to the cumulative effect of numerous transactions.

Further exploration into the specific features that contributed most to the accuracy could provide valuable insights into market dynamics, aiding in future model refinements.

6.1.2 LLM Results

Large Language Models (LLMs) have emerged as a significant trend in Natural Language Processing (NLP), with applications in text generation, translation, and summarization. In this study, LLMs were utilized to predict the next candlestick in stock market data, with all models operated locally.

The results, however, did not meet expectations. The models generally underperformed, with only `llama2-uncensored:7b` achieving results comparable to the VAE experiment. Despite producing diverse responses, this model experienced nearly 30% data loss, failing to consistently output the intended JSON format. Nevertheless, it performed best in terms of result returns.

Conversely, `gemma:7b` had the lowest accuracy but excelled in terms of automation, failing to process only one response. This highlights the trade-offs between model complexity and operational efficiency.

The performance of `llama2-uncensored:7b` suggests it is closest to achieving the prediction goals. However, even with the temperature set to prioritize precision over creativity, significant data loss occurred. This indicates a possible misalignment between the model's settings and the structured requirements of financial data outputs. In contrast, `gemma:7b` was highly consistent in format but lacked accuracy, which may be related to its design focus. Refining the prompts or employing more specialized models could potentially yield better results, indicating an area for future research.

6.1.3 Comparative Analysis

Comparing the VAE and LLM models provides deeper insights into their respective strengths and weaknesses. The VAE models showed modest but consistent improvements over baseline models, suggesting they are more suited for capturing complex numerical patterns in financial data. LLMs, while powerful in language tasks, struggled with the structured prediction required in this context.

This comparative analysis suggests that for time series prediction in financial markets, models specifically designed for numerical data, like VAEs combined with GRUs, may be more effective than general-purpose LLMs. This understanding can guide future efforts in selecting appropriate model architectures and feature engineering strategies to enhance predictive performance.

6.1.4 Future Directions

The insights from these experiments lay the groundwork for several future research directions:

- **Enhancing LLM Performance:** Exploring different configurations, larger models, or more domain-specific LLMs may improve performance. Adjusting model parameters, refining prompts, or incorporating additional training data could enhance accuracy and reliability.
- **Refining VAE Models:** Further expanding the feature set and optimizing the architecture of the VAE models could lead to better predictive accuracy. Investigating different latent dimensions, activation functions, and training strategies may yield improvements.
- **Hybrid Approaches:** Combining the strengths of VAEs and LLMs, perhaps through ensemble methods or multi-modal architectures, could leverage the advantages of both model types.
- **Feature Importance Analysis:** Conducting a detailed analysis of the features that contribute most to prediction accuracy can inform feature selection and engineering, potentially leading to more efficient models.

Overall, while the experiments presented challenges, they also highlight opportunities for significant advancements in predictive accuracy and operational efficiency with continued research and development.

7 Conclusion

The stock market is a highly complex environment, and predicting its movements is an inherently challenging task. This study utilized a dataset limited to the S&P 500 index and employed a relatively simplistic model. It is hypothesized that utilizing a more extensive dataset encompassing a broader range of indices and implementing a more complex model structure might yield improved results.

A primary constraint encountered during this experiment was related to hardware capabilities. The limited computational resources restricted the ability to train more sophisticated models and to process larger datasets effectively. In particular, it was not possible to explore the potential of advanced Large Language Models (LLMs), such as the 70-billion-parameter models known for their ability to handle and analyze vast amounts of data with higher accuracy.

7.1 Future Works

Future work aims to overcome these challenges by securing access to more powerful computational resources. This enhancement will allow experimentation with more intricate models, including advanced VAEs (VAEs) and other deep learning architectures, which could significantly improve predictive capabilities.

Additionally, future experiments plan to diversify the data sources. By integrating data from various global indices and incorporating different market indicators, it is intended to develop a more robust model that can generalize across different market conditions and environments. This approach will not only test the scalability of the models but also enhance their applicability to real-world scenarios.

Moreover, subsequent studies intend to employ more complex and elaborate VAE models to expand the data. Although the results presented in this study may not be the most impressive, it is believed that by expanding the dataset and enhancing the model's performance through fine-tuning and advanced feature engineering, significant improvements can be achieved. This process will involve a detailed analysis of model behavior under various market conditions, providing a deeper understanding of the factors that influence stock market trends.

An essential component of future research will be to implement and evaluate the impact of xAI techniques. This will not only aid in understanding model decisions and predictions but also in refining model parameters to better capture the nuances of market dynamics. Through these methods, the goal is to build models that are not only performant but also transparent and interpretable, enhancing their utility for decision-makers in the financial sector.

In summary, while this study provides valuable insights into stock market prediction using machine learning models, it underscores the necessity for more advanced computational resources and methodologies. By addressing the identified limitations and pursuing the outlined future directions, there is potential for significant advancements in predictive accuracy and model interpretability, ultimately benefiting decision-makers in the financial sector.

Bibliography

ALI, S. et al. Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. v. 99, p. 101805, 2023. ISSN 15662535. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1566253523001148>. Citado na página 16.

AUTHORS, T. *TensorFlow: An end-to-end open source machine learning platform*. Google, 2023. Accessed: 2024-10-21. Disponível em: <https://www.tensorflow.org>. Citado na página 17.

DASH, R.; DASH, P. K. A hybrid stock trading framework integrating technical analysis with machine learning techniques. v. 2, n. 1, p. 42–57, 2016. ISSN 24059188. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2405918815300179>. Citado na página 20.

DAVID, D. B.; RESHEFF, Y. S.; TRON, T. Explainable AI and Adoption of Financial Algorithmic Advisors: an Experimental Study. arXiv, jun. 2021. ArXiv:2101.02555 [cs]. Citado na página 16.

EUKO, J. P.; SANTOS, F.; NOVAIS, P. Stock market prediction: Integrating explainable ai with conv2d models for candlestick image analysis. In: . [s.n.]. Disponível em: <https://doi.org/10.1007/978-3-031-60218-4_2>. Citado na página 27.

FRANKEN, G. Applications of content-based image retrieval in financial trading. 2023. Citado na página 22.

FREEBOROUGH, W.; ZYL, T. van. Investigating Explainability Methods in Recurrent Neural Network Architectures for Financial Time Series Data. *Applied Sciences*, v. 12, n. 3, p. 1427, jan. 2022. ISSN 2076-3417. Disponível em: <https://www.mdpi.com/2076-3417/12/3/1427>. Citado na página 16.

JAQUART, P.; DANN, D.; WEINHARDT, C. Short-term bitcoin market prediction via machine learning. v. 7, p. 45–66, 2021. ISSN 24059188. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2405918821000027>. Citado 3 vezes nas páginas 10, 21, and 45.

KINGMA, D. P.; WELLING, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2014. Citado na página 15.

KUMAR, C. S. et al. Dimensionality Reduction based on **shap!** (**shap!**) Analysis: A Simple and Trustworthy Approach. In: *2020 International Conference on Communication and Signal Processing (ICCSP)*. Chennai, India: IEEE, 2020. p. 558–560. ISBN 978-1-72814-988-2. Disponível em: <https://ieeexplore.ieee.org/document/9182109/>. Citado na página 19.

LENG, J. Stock movement prediction model based on gated orthogonal recurrent units. 2022. Citado na página 10.

LENG, J. Stock movement prediction model based on gated orthogonal recurrent units. 2022. Citado 2 vezes nas páginas 20 and 45.

LUNDBERG, S.; LEE, S.-I. *A Unified Approach to Interpreting Model Predictions*. arXiv, 2017. ArXiv:1705.07874 [cs, stat]. Disponível em: <http://arxiv.org/abs/1705.07874>. Citado 2 vezes nas páginas 11 and 17.

LV, S.; HOU, Y.; ZHOU, H. Financial market directional forecasting with stacked denoising VAE. arXiv, n. arXiv:1912.00712, 2019. Disponível em: <http://arxiv.org/abs/1912.00712>. Citado 4 vezes nas páginas 10, 12, 20, and 45.

MAN, X.; CHAN, E. P. The Best Way to Select Features? Comparing MDA, LIME, and **shap!**. *The Journal of Financial Data Science*, v. 3, n. 1, p. 127–139, jan. 2021. ISSN 2640-3943. Disponível em: <http://jfds.pm-research.com/lookup/doi/10.3905/jfds.2020.1.047>. Citado na página 17.

MYTTENAERE, A. D. et al. Mean Absolute Percentage Error for regression models. *Neurocomputing*, v. 192, p. 38–48, jun. 2016. ISSN 09252312. ArXiv:1605.02541 [stat]. Citado na página 19.

NEVASALMI, L. Forecasting multinomial stock returns using machine learning methods. v. 6, p. 86–106. ISSN 24059188. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2405918820300143>. Citado na página 10.

NEVASALMI, L. Forecasting multinomial stock returns using machine learning methods. v. 6, p. 86–106, 2020. ISSN 24059188. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2405918820300143>. Citado 2 vezes nas páginas 21 and 45.

OPENAI. *ChatGPT: Language model for conversational AI*. 2023. Accessed: 2024-10-21. Disponível em: <https://openai.com/chatgpt>. Citado na página 10.

PELSTER, M.; VAL, J. Can ChatGPT assist in picking stocks? v. 59, p. 104786, 2024. ISSN 15446123. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1544612323011583>. Citado na página 20.

PFEIFER, M.; MAROHL, V. P. CentralBankRoBERTa: A fine-tuned large language model for central bank communications. v. 9, p. 100114, 2023. ISSN 24059188. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2405918823000302>. Citado 2 vezes nas páginas 20 and 21.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. “why should i trust you?” explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016. p. 1135–1144. Disponível em: <https://arxiv.org/abs/1602.04938>. Citado na página 17.

SELVARAJU, R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2017. p. 618–626. Citado na página 20.

WILKERSON, Z.; LEAKE, D.; CRANDALL, D. Leveraging **shap!** and CBR for Dimensionality Reduction on the Psychology Prediction Dataset. 2022. Citado na página 20.

WRIGHT, S. *Correlation and causation. Journal of Agricultural Research*. 1921. 557–585 p. Citado na página 18.