1-INTRODUCTION

1.1- Understanding Stocks

A stock, also known as a share or equity, represents a unit of ownership in a company and entitles the holder to a portion of the company's assets and earnings. When investors acquire more stock, their ownership stake in the company increases, granting them greater influence and potential rewards. Stocks are traded on stock markets such as the New York Stock Exchange (NYSE) and Nasdaq, which provide a regulated environment for buying and selling shares. These markets ensure transparency, liquidity, and fair pricing, facilitating transactions that form the basis of many individual investment accounts. Companies use stock markets to raise capital by issuing shares, enabling them to fund operations, expand, and innovate. The initial public offering (IPO) process allows companies to sell shares to the public for the first time, providing a significant influx of capital that can be used for various purposes, including research and development, marketing, debt reduction, and acquisitions. For investors, owning stocks offers potential for capital appreciation and income through dividends, making them a crucial part of personal investment strategies. Capital appreciation occurs when the value of the stock increases over time, allowing investors to sell their shares at a higher price than they paid. Dividends are periodic payments made by companies to shareholders, typically from profits, providing a steady income stream. Stocks also offer the benefit of diversification, as investors can spread their investments across different companies, industries, and geographic regions to reduce risk. Moreover, stock ownership comes with voting rights, enabling shareholders to influence corporate governance by voting on key issues such as the election of directors and significant corporate actions. This democratic aspect of stock ownership empowers investors to have a say in the strategic direction of the companies they invest in. Overall, stocks play a vital role in the economy, driving capital formation, fostering innovation, and providing individuals with opportunities for wealth creation and financial security.

1.2- The Importance of Predicting Stock Prices

Predicting stock prices has long captivated the interest of investors, financial analysts, and academics alike. The ability to forecast stock movements accurately can yield substantial financial gains, underscoring its critical importance in financial analysis and decision-making. Traditionally, this endeavor relied on statistical methods and fundamental analysis, demanding a deep understanding of the financial market and proficiency in interpreting intricate data patterns. Analysts would meticulously examine a company's financial statements, market conditions, economic indicators, and historical price data to make informed predictions. This process involves assessing various factors that can influence stock prices, such as earnings reports, economic trends, interest rates, and geopolitical events. Accurate predictions enable investors to make strategic decisions, optimize their portfolios, and manage risks effectively.

The importance of predicting stock prices extends beyond individual investors to the broader financial market. For institutional investors, such as mutual funds, pension funds, and hedge funds, accurate stock price predictions are crucial for asset allocation, risk management, and achieving superior returns for their clients. By anticipating market movements, these institutions can adjust their investment strategies to capitalize on opportunities and mitigate potential losses. This ability to forecast stock movements also enhances market efficiency by ensuring that prices reflect all available information, thereby reducing the likelihood of mispricing and speculative bubbles.

Furthermore, accurate stock price predictions contribute to economic stability. They help companies in planning and financing their operations by providing insights into their stock performance. For instance, a company considering an expansion or a new project can use stock price forecasts to assess investor confidence and determine the feasibility of raising capital through equity issuance. Predictive insights can also guide corporate actions such as mergers, acquisitions, and share buybacks, aligning them with shareholder interests and market conditions.

In addition to aiding investment decisions and corporate planning, stock price predictions are vital for regulatory bodies and policymakers. Regulatory authorities use these forecasts to

monitor and prevent market manipulation, ensuring a fair and transparent trading environment. Policymakers, on the other hand, can gauge the health of the economy and the financial markets through stock price trends, informing their decisions on monetary and fiscal policies. For example, rising stock prices might indicate economic growth, prompting policies that support further expansion, while declining prices could signal economic distress, leading to measures aimed at stabilization.

Overall, the importance of predicting stock prices lies in its ability to provide actionable insights for a wide range of stakeholders. From individual investors seeking to grow their wealth to large institutions managing substantial portfolios, from companies strategizing their growth to regulators safeguarding market integrity, accurate stock price predictions are a cornerstone of financial planning and economic management. This multifaceted significance underscores why the pursuit of better forecasting methods remains a central focus in the field of finance.

1.3- Machine Learning in Stock Price Prediction

Machine learning techniques have brought about a paradigm shift in the realm of stock price prediction, fundamentally transforming the way analysts approach this complex task. Machine learning, a subset of artificial intelligence, utilizes intricate algorithms to autonomously learn from data and generate predictions without the need for explicit programming instructions. This self-learning capability is achieved through various models such as decision trees, neural networks, and support vector machines, which can analyze vast amounts of data far more quickly and accurately than traditional statistical methods.

By harnessing historical data, machine learning algorithms can not only forecast stock prices but also uncover latent patterns and trends that may elude manual analysis. These algorithms excel in detecting nonlinear relationships and interactions among variables, which are often too complex for human analysts to discern. This capacity to discern subtle trends and anomalies not only enhances the accuracy of stock price predictions but also facilitates more informed and efficient investment decision-making. For instance, algorithms can identify cyclical patterns or

correlations between different stocks and external factors, such as economic indicators or geopolitical events, providing a more comprehensive analysis of potential market movements.

Additionally, machine learning algorithms are capable of evolving over time when exposed to new data, continually refining their predictive capabilities. This adaptability is a significant advantage in the dynamic environment of financial markets, where conditions can change rapidly. For example, an algorithm might initially be trained on data from a stable economic period but will adjust its predictions as it encounters data from more volatile periods. This continuous learning process ensures that the model remains relevant and accurate, even as market conditions evolve.

Another key benefit of machine learning in stock price prediction is its being capable to analyze and deal with unstructured data, such as news articles, social media posts, and financial reports. Natural language processing (NLP) techniques enable algorithms to extract sentiment and other valuable information from text data, which can be incorporated into predictive models. This capability allows investors and analysts to consider a broader range of factors when making decisions, enhancing the robustness of their strategies.

Moreover, machine learning models can be backtested against historical data to evaluate their performance before being used in live trading environments. This rigorous testing process helps to identify potential weaknesses and biases in the models, enabling improvements to be made. Once validated, these models can be integrated into automated trading systems, executing trades based on predefined criteria and real-time data inputs. This automation reduces the risk of human error and ensures that trading decisions are made consistently and objectively.

As a result, machine learning has become an indispensable tool in the arsenal of investors and financial analysts, enabling them to navigate the complexities of the financial markets with greater precision and confidence. The continuous advancements in machine learning technology promise to further enhance the capabilities of these models, making them even more effective at predicting stock prices and identifying investment opportunities. Consequently, the adoption of machine learning in finance is likely to continue growing, driven by its proven ability to deliver superior predictive performance and its potential to unlock new insights from ever-expanding data sources.

1.4-Limitations in Stock Price Prediction

Despite the transformative potential of machine learning (ML) in stock price prediction, several limitations and challenges persist. One significant limitation is the quality and quantity of the data. Accurate predictions rely heavily on vast amounts of high-quality, relevant data. However, financial data can be noisy, incomplete, or inconsistent, which can negatively impact the performance of ML models. Furthermore, historical data may not always reflect future market conditions, especially during unprecedented events such as financial crises or global pandemics, leading to inaccurate predictions.

Another challenge is the black-box nature of many ML algorithms, particularly deep learning models. These models can produce highly accurate predictions but often lack transparency, making it difficult for analysts to understand how specific predictions are made. This opacity can be problematic for gaining the trust of stakeholders who require clear explanations for investment decisions.

Overfitting is also a common issue in ML models. When a model is too complex, it may perform exceptionally well on training data but fail to predict to new, unseen data. This can lead to poor performance in real-world scenarios, where the model encounters data that slightly deviates from the patterns it was trained on.

Additionally, ML models assume that historical patterns and relationships will continue into the future, which is not always the case in the dynamic and often unpredictable financial markets. Sudden shifts due to economic policy changes, geopolitical events, or technological disruptions can render models less effective.

The computational cost of developing and maintaining sophisticated ML models is another limitation. Training deep learning models, in particular, requires significant computational resources and expertise, which can be a barrier for smaller firms or individual investors.

Finally, the integration of unstructured data, such as news articles and social media sentiment, into predictive models introduces further complexity. While this data can enhance predictions, it

also requires advanced natural language processing techniques and can be difficult to interpret and validate.

In summary, while ML offers powerful tools for stock price prediction, challenges related to data quality, model transparency, overfitting, market unpredictability, computational costs, and the integration of unstructured data limit its effectiveness and widespread adoption.

1.5- Outline of the Paper

This paper will discuss the following:

1. Problem Statement: This section will describe the challenges associated with stock price prediction, emphasizing the volatility and complexity of financial markets.

2. Literature Review: A comprehensive analysis of existing research on the use of machine learning for forecasting stock prices will be conducted. This review will highlight the methodologies, techniques, and findings of previous studies, providing a foundation for the current research.

3. Objectives: The primary goals and objectives of this research will be outlined, focusing on enhancing the accuracy and the reliability of the stock price prediction models.

4. Hardware Specification: The specifications of the hardware used in this study will be detailed, ensuring transparency and reproducibility of the results.

5. Methodology: This section will explain the approach and methods used in the research. It will include details on collecting data, preprocessing, model selection, training, and evaluation.

6. Detailed Tasks: A breakdown of the tasks associated with the research will be provided, outlining the specific responsibilities and activities involved in each stage of the project.

7. Gantt Chart: A task schedule will be presented, providing a visual representation of the project's chronology and task dependencies.

8. Semester 1 Achievements: A summary of progress made during the first semester of the research will be provided, highlighting key milestones and outcomes.

9. Anticipated Problems and Solutions: Potential challenges that may arise during the research process will be discussed, along with proposed strategies to address them.

10. Result: The results of the research, including the performance of the developed stock price prediction model, will be presented and analyzed. This section will also include a comparison between the results obtained in this study and those reported in other relevant papers, providing insights into the effectiveness and novelty of the proposed approach.

11. Conclusion: The findings of the research will be summarized, highlighting the contributions to the field of stock price prediction and suggesting areas for future research.

2-Problem statement

Predicting stock prices using machine learning is a multifaceted challenge that necessitates the creation of a model capable of accurately forecasting stock price movements. This endeavor consists of several key steps, beginning with the collection and preprocessing of stock price data to ensure its accuracy and relevance.

2.1- Collecting Data and Preprocessing

The first step in predicting stock prices with machine learning is to gather relevant data. This data includes historical stock prices, trading volumes, and other financial metrics gathered from a variety of sources, including financial news websites, stock exchanges, and corporate financial statements. The raw data frequently contains noise and inconsistencies that must be addressed during preprocessing. Preprocessing includes handling missing values, normalizing data, and removing outliers. Data smoothing and feature scaling are common techniques used to prepare data for analysis. Furthermore, data augmentation techniques can be used to increase the dataset's size and diversity, thereby improving the model's robustness.

2.2- Feature Selection and Engineering

Once the data has been preprocessed, the next critical step is feature selection and engineering. Identifying relevant characteristics and variables that could affect stock prices necessitates a thorough understanding of financial markets and economic indicators. Historical price data, trading volumes, market indices, interest rates, and macroeconomic indicators such as GDP growth rates and inflation are all common features, thereby reducing dimensionality and improving model performance. Feature engineering is the process of creating new variables from existing data to better capture underlying patterns and trends. Moving averages, momentum indicators, and volatility measures are common features of stock price prediction models.

2.3- Selecting Model

Selecting the appropriate machine learning technique is critical for efficiently analyzing data and identifying meaningful patterns. Different models can be used, including linear regression, decision trees, random forests, support vector machines, and neural networks. The type of model used is determined by the nature of the data and the prediction task's specific requirements. Linear models, such as regression, are widely used because they are simple and easy to understand, whereas more complex models, such as neural networks, can capture intricate nonlinear relationships in data. Ensemble methods, which combine multiple models to improve prediction accuracy, are also widely used in stock price prediction.

2.4- Model Training

Training the model entails using previous data to teach it to make accurate predictions. This process usually involves dividing the data into training and testing sets. The training set is used to fit the model, and the testing set assesses its performance. During training, the model learns from historical stock prices and other relevant data by adjusting its parameters to reduce prediction errors. Cross-validation is a technique used to ensure model robustness and prevent overfitting. Hyperparameter tuning, or optimizing the model's configuration settings, is also an important part of the training process.

2.5- Model Evaluation

Once trained, the model's performance is assessed using a variety of metrics to determine its accuracy and reliability. MSE, MAE, and RMSE are some of the most commonly used evaluation metrics. More advanced metrics, such as the coefficient of determination (R^2) and Sharpe ratio, can offer insights into the model's predictive power and risk-adjusted returns. Backtesting, or testing the model on historical data to simulate its performance in real trading scenarios, is an important step in model evaluation. This aids in identifying any potential flaws and refining the model further.

2.6- Implementation and Continuous Improvement

Following a successful evaluation, the model is deployed in a real-world trading environment. Due to the dynamic nature of financial markets, the model must be continuously monitored and improved. Regular updates with new data, re-training, and recalibration are required to keep the model's accuracy over time. Additionally, incorporating feedback from actual trading results can help refine the model even further.

3- LITERATURE REVIEW

3.1- Survey on this field

Since stock market prediction using machine learning targets investors, companies, and anyone interested in the flow of stocks, it has garnered significant attention and led to several studies focusing on various approaches. Notably, Rouf et al. (2021) reviewed machine learning-based

approaches for stock prediction from 2011 to 2021, explaining that these approaches typically involve four phases in a generic framework: data collection, data preprocessing, feature extraction and reduction, and modeling. They discussed the importance of each phase, stating that data collection involves locating historical stock prices, trading volumes, and other relevant financial metrics from various sources, such as financial news websites and stock exchanges. Data preprocessing addresses inconsistencies and noise in the raw data through techniques like handling missing values, normalizing data, and removing outliers. Feature extraction and reduction are crucial for identifying and selecting relevant characteristics that influence stock prices, employing methods to enhance model performance. The modeling phase involves choosing appropriate machine learning techniques, such as Linear Regression, Decision Tree, Random Forests, Support Vector Machines, and Neural Networks, depending on the data's nature and the prediction task's requirements. Rouf et al. also suggested future developments and directions, including web scraping from social media, leveraging deep learning, and employing hybrid models that combine different machine learning approaches to improve prediction accuracy. They acknowledged that while machine learning in stock prediction shows promising results, it faces limitations and challenges, such as data quality issues, model overfitting, and the dynamic nature of financial markets.

Similarly, Strader et al. (2020) conducted a comprehensive review mirroring the phases and suggestions discussed by Rouf et al., reaffirming that data collection, preprocessing, feature extraction, and modeling are fundamental steps in machine learning-based stock prediction. They highlighted that the selection of relevant features and the appropriate model is critical for achieving accurate predictions. Strader et al. also echoed the future directions proposed by Rouf et al., stressing the potential of deep learning and hybrid models in enhancing stock prediction capabilities. They reiterated the challenges in this field, emphasizing the importance of continuous model improvement and adaptation to the ever-changing market conditions. Both reviews underscored that machine learning approaches provide a promising avenue for stock market prediction, but researchers and practitioners must remain vigilant about the inherent limitations and continuously strive for advancements in methodologies and techniques.

In addition to stock market prediction, there have been notable studies on using machine learning for predicting the forex market, which is the foreign exchange market where currencies are

traded globally. The forex market aims to profit from currency exchanges similarly to stock trading. Ayitey Junior et al. (2023) conducted a systematic literature review on machine learning applications for forex market forecasting, examining 60 papers from 2010 to 2021. They proposed a framework similar to that used in stock market prediction, encompassing data collection, preprocessing, feature extraction, and modeling. Their review found that Long Short-Term Memory (LSTM) networks and Artificial Neural Networks (ANN) are the most commonly used models for forex prediction due to their ability to capture complex temporal dependencies and nonlinear relationships in the data. Ayitey Junior et al. also discussed the challenges and future directions in forex market forecasting, highlighting the need for better data preprocessing techniques to handle the vast amount of unstructured data from various sources and the potential of combining different machine learning models to improve prediction accuracy and robustness.

Hu et al. (2021) conducted a similar survey focusing on the use of deep learning in forex market forecasting. They explored how deep learning models could be applied to the four phases of the prediction framework and discussed the advantages of these models in capturing intricate patterns and trends in the forex market. Hu et al. While deep learning models demonstrate significant promise, they also have challenges such as the need for large datasets, high computational resources, and the risk of overfitting. They proposed future research directions, including the development of more efficient deep learning algorithms, better feature extraction methods, and the incorporation of specialist expertise into machine learning models to improve interpretability and dependability.

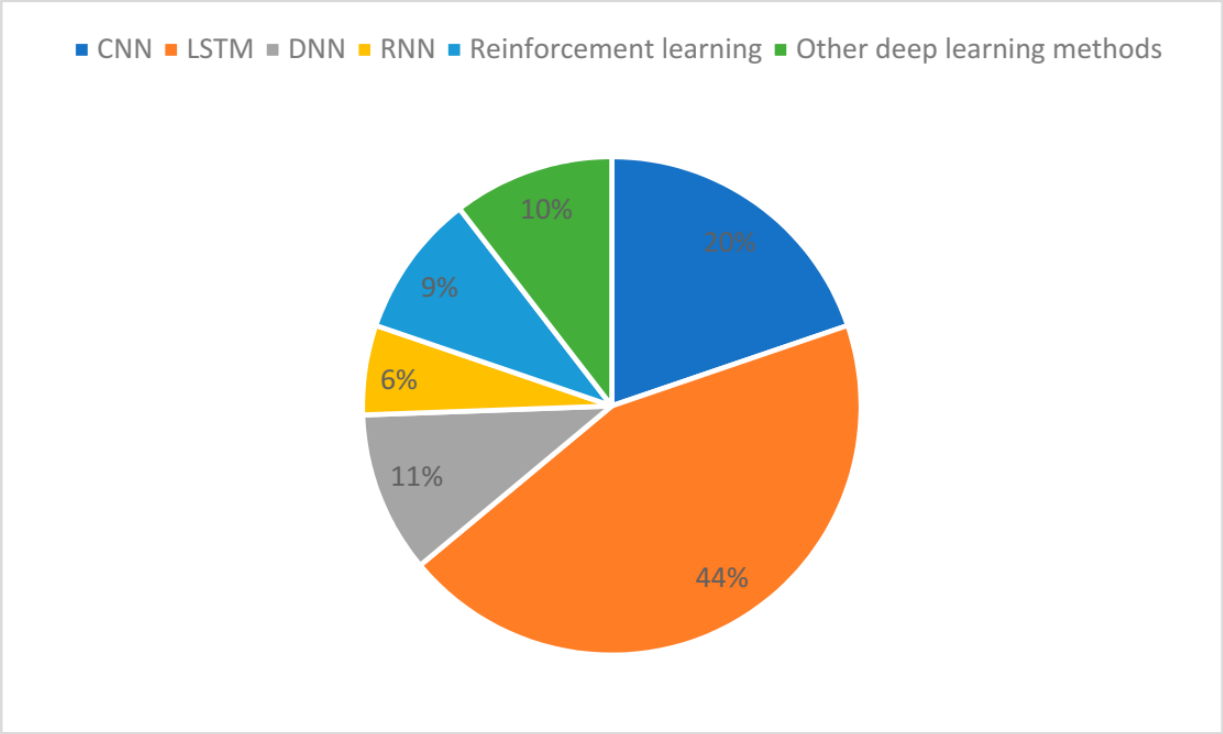


Figure 1. the percentage of the models used in Hu (2021) survey.

3.2-Race to find the best model

Predicting the stock market is not an easy task. It is not a default code that can be applied to the stocks and get accurate data. Some authors participated in the journey to find the best model, knowing that the best model between these two. Among these are also authors that will be mentioned, not most of them are the same. Starting with Guo (2023), which is a test of several machine learning models, linear regression, decision tree (DT), LSTM, and neural network (NN) on Netflix stock data from 2002 to 2021. The metric used in this paper is the mean square error (MSE). Notably, the NN outperformed the other models, achieving 10 times better results. On the other hand, Mukhallad M et al. (2019) tested different models with different datasets. The models used were deep neural networks, recurrent neural network, support vector regression (SVR), and support vector machine (SVM), using data from stocks such as Apple, Amazon, Google, and Facebook. The result favored SVM with the highest accuracy, achieving 82.91% using the directional accuracy metric. Some models have been shown in different comparisons. For instance, Chen (2020) used data from Apple, MasterCard, Ford, and ExxonMobil to see which model among SVR, LSTM, and convolutional neural network (CNN) was the best for

prediction using metrics such as mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). The result favored LSTM in this paper. On the other hand, Hiransha et al. (2018) conducted a comparison between only deep learning models. Their models included autoencoders (AE), multilayer perceptron (MLP), RNN, CNN, and LSTM, using data from the National Stock Exchange of India and the New York Stock Exchange. The metrics used were MAPE. The result of this comparison showed that CNN outperformed the other three models. In the next studies, LSTM was the model sitting on the throne as the best model for stock prediction. In the study by Nabipour et al. (2020), they tested nine different machine learning models, which included decision tree, bagging, random forest (RF), adaptive boosting, gradient boosting, extreme gradient boosting, ANN, and LSTM. The metrics used in this paper were MAPE, MAE, relative root mean square error (RRMSE), and MSE. Among all these models, LSTM was the best model. Also, in Ho et al. (2021), LSTM was compared with auto-regressive integrated moving average (ARIMA) and NN. The metrics used were MAE, MSE, and RMSE. The results indicated a close competition between LSTM and ARIMA, yet LSTM was the better model. Another study was conducted using data from the National Stock Exchange of India. Mehtab & Sen (n.d.) made a comparison with five different models. Two of them were regression models based on CNN, and three were predictive models based on LSTM, using RMSE as the metric. The most predictive model was the univariate encoder-decoder convolutional LSTM. Additionally, Houssein et al. (2021) presented a comprehensive study on the application of deep learning models, specifically nonlinear autoregressive artificial neural networks (NARX), trained using three different algorithms: Bayesian regularization (BR), Levenberg–Marquardt (LM), and scaled conjugate gradient (SCG), for predicting the closing prices of various indices of the Egyptian Stock Exchange. The study evaluated the performance of these models over short-term (1, 3, 5 days) and long-term (7, 15, 30 days) predictions, using mean squared error (MSE) and correlation coefficient as metrics. The results indicated that BR was more effective for short-term predictions, particularly for 3 days ahead, while LM showed better accuracy for long-term predictions, especially for 7-day forecasts. The research contributed to the field by demonstrating the potential of NARX-based models in stock price prediction and the importance of selecting appropriate training algorithms for different prediction horizons.

The literature on stock market prediction using machine learning is vast and diverse, with numerous studies exploring various models and techniques. The consensus among these studies is that no single model is universally the best for all types of stock prediction tasks. Each model has its strengths and weaknesses, and their performance can vary significantly depending on the dataset and the specific characteristics of the stocks being predicted. For example, in the study by Guo (2023), the neural network (NN) model outperformed other models such as linear regression, decision tree, and LSTM when tested on Netflix stock data from 2002 to 2021, achieving a significantly lower mean square error (MSE). This suggests that NN may be particularly well-suited for capturing the complex patterns in Netflix stock data over this period.

In contrast, Mukhallad M et al. (2019) found that the support vector machine (SVM) model performed best when tested on stocks such as Apple, Amazon, Google, and Facebook, achieving the highest accuracy of 82.91% using the directional accuracy metric. This indicates that SVM may be more effective for predicting the stock prices of technology companies. Similarly, Chen (2020) found that the LSTM model outperformed SVR and CNN when tested on stocks such as Apple, MasterCard, Ford, and ExxonMobil, using metrics such as MAE, RMSE, and MAPE. This suggests that LSTM may be particularly effective for capturing the temporal dependencies in the stock prices of these companies.

Furthermore, Hiransha et al. (2018) conducted a comparison between different deep learning models, including AE, MLP, RNN, CNN, and LSTM, using data from the National Stock Exchange of India and the New York Stock Exchange. The results indicated that CNN outperformed the other models, suggesting that CNN may be particularly well-suited for capturing the spatial patterns in the stock prices of these exchanges. On the other hand, Nabipour et al. (2020) found that LSTM was the best model among nine different machine learning models, including decision tree, bagging, random forest, adaptive boosting, gradient boosting, extreme gradient boosting, ANN, and LSTM, when tested on various stocks using metrics such as MAPE, MAE, RRMSE, and MSE. This indicates that LSTM may be particularly effective for capturing the complex temporal dependencies in stock prices across different stocks and exchanges.

Moreover, Ho et al. (2021) compared LSTM with ARIMA and NN, using metrics such as MAE, MSE, and RMSE. The results indicated a close competition between LSTM and ARIMA, yet LSTM was the better model, suggesting that LSTM may be more effective for capturing the long-term temporal dependencies in stock prices. Similarly, Mehtab & Sen (n.d.) found that the univariate encoder-decoder convolutional LSTM was the most predictive model when tested on data from the National Stock Exchange of India, using RMSE as the metric. This suggests that this specific type of LSTM model may be particularly effective for capturing the temporal dependencies in stock prices from this exchange.

Additionally, Houssein et al. (2021) conducted a comprehensive study on the application of deep learning models, specifically NARX, trained using three different algorithms: BR, LM, and SCG, for predicting the closing prices of various indices of the Egyptian Stock Exchange. The results indicated that BR was more effective for short-term predictions, particularly for 3 days ahead, while LM showed better accuracy for long-term predictions, especially for 7-day forecasts. This suggests that the choice of training algorithm can significantly impact the performance of NARX models for different prediction horizons. This research contributes to the field by demonstrating the potential of NARX-based models in stock price prediction and the importance of selecting appropriate training algorithms for different prediction horizons.

Author	Models Used	Metrics Used	Result
(Guo, 2023)	Linear Regression, Decision Tree (DT), LSTM, NN	MSE	NN
(Mukhallad M et al., 2019)	DNN, RNN, SVR, SVM	directional accuracy	SVM
(Chen, 2020)	SVR, CNN, LSTM	MAE, RMSE, MAPE	LSTM
(Hiransha et al., 2018)	MLP, LSTM, CNN, RNN	MAPE	CNN
(Nabipour et al., 2020)	DT, Bagging , RF , Adaptive Boosting (Adaboost), Gradient Boosting , eXtreme Gradient Boosting, (XGBoost) , (ANN) , (RNN) , (LSTM),	MAPE, MAE, RMSE, MSE	LSTM
(Hu et al., 2021)	LSTM, ARIMA, NN	MAE, MSE, RMSE	LSTM
(Mehtab & Sen, n.d.)	(CNNs): Two regression models built on CNNs. (LSTM) Networks: Three predictive models based on LSTM networks.	RMSE	Univariate encoder-decoder convolutional LSTM
(Houssein et al., 2021).	(BR), (LM), (SCG).	(MSE) and correlation	BR is more effective for short-term predictions, ,

		coefficient	while LM shows better accuracy for long-term

Table 1. results of the previous papers

3.3-Reliability of LSTM

LSTM (Long Short-Term Memory) has emerged as a pivotal tool in stock market prediction, credited for its robustness and adaptability in handling the complexities of financial data. Researchers, including Yu and Yan (2020), have extensively explored LSTM's potential by integrating it with diverse algorithms to amplify its predictive capabilities. Their research underscores LSTM's adeptness in capturing intricate dependencies and nonlinear patterns inherent in time series data. To this end, they proposed a hybrid model that merges Phase Space Reconstruction (PSR) with LSTM, benchmarking it against conventional models like ARIMA, Support Vector Regression (SVR), Multilayer Perceptron (MLP), and a standalone LSTM model. Through extensive experimentation across four prominent stock datasets and the evaluation of metrics such as Directional Accuracy (DA), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE), their findings consistently favored the proposed model's superior performance. This hybrid approach harnesses the strengths of PSR in reconstructing the state space of time series data, thereby providing LSTM with a more comprehensive dataset that enhances its ability to learn and predict future stock prices accurately.

In a similar vein, Khaled A. Al-Thelaya et al. (2019) concentrated on forecasting the future values of the Bahrain Bourse All Share Index (BAX) by employing advanced LSTM techniques. They included an LSTM Autoencoder for denoising and smoothing the input data, which helps in reducing noise and enhancing the quality of the input signals, thus improving the model’s performance. Additionally, they used stacked LSTM for forecasting, which involves multiple

LSTM layers stacked on top of each other to capture more complex features and dependencies in the data. Their comparative analysis against a basic LSTM model and a shallow Multilayer Perceptron (MLP) demonstrated the superior predictive power of their proposed models. The LSTM Autoencoder effectively preprocessed the data by removing noise, while the stacked LSTM captured long-term dependencies more efficiently, leading to more accurate predictions of the BAX index.

Conversely, Shastry Pm et al. (n.d.) opted for a more simplistic approach, utilizing LSTM in its standard form without modifications to predict India's national stock data. Their study affirmed LSTM's reliability and efficacy in stock market forecasting, highlighting that even without complex enhancements or hybrid models, LSTM alone is a powerful tool for capturing the temporal dependencies in stock market data. This straightforward application of LSTM provided significant insights into its baseline performance and established a benchmark for comparing more advanced models. The results from their study demonstrated that LSTM could effectively model the trends and patterns in stock data, providing reliable forecasts that are crucial for investors and financial analysts.

These studies collectively underscore LSTM's versatility and effectiveness in stock market prediction, showcasing its potential when combined with a diverse array of machine learning techniques. The growing body of research in this field continues to demonstrate LSTM's pivotal role in enhancing stock market prediction accuracy. For instance, hybrid models that integrate LSTM with other techniques, such as PSR or autoencoders, consistently show improved performance over traditional models and standalone LSTM applications. This indicates that LSTM's architecture, which includes memory cells and gates designed to handle long-term dependencies, is particularly well-suited for the dynamic and often volatile nature of stock market data.

Moreover, the use of LSTM in different market contexts, such as the Bahrain Bourse and India's national stock data, highlights its adaptability to various types of financial data and market conditions. Researchers have demonstrated that LSTM can be tailored to specific datasets and market environments, making it a flexible tool for global financial forecasting. This adaptability is crucial, as it allows for the development of more accurate and customized prediction models that can cater to the unique characteristics of different stock markets.

The continuous advancements in LSTM applications also point towards future research directions and potential improvements. For example, integrating LSTM with more sophisticated preprocessing techniques, like denoising autoencoders, or combining it with other deep learning models, such as CNNs, could further enhance its predictive capabilities. Additionally, exploring the use of LSTM in conjunction with real-time data feeds and high-frequency trading data could open new avenues for its application in short-term stock market predictions.

LSTM has firmly established itself as a cornerstone in the field of stock market prediction. Its ability to model complex temporal dependencies and nonlinear patterns makes it an invaluable tool for financial forecasting. The studies by Yu and Yan (2020), Khaled A. Al-Thelaya et al. (2019), and Shastry Pm et al. (n.d.) collectively highlight the diverse applications and superior performance of LSTM models in various stock market contexts. As research in this area continues to evolve, LSTM's role in enhancing the accuracy and reliability of stock market predictions will undoubtedly expand, offering significant benefits to investors, financial analysts, and the broader economic landscape.

3.4- Performance Evaluation Metrics

Machine learning models frequently evaluate their performance using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). MAE computes the average of the absolute differences between predicted and actual values, resulting in a simple measure of prediction error. MSE computes the average of the squared differences between predicted and actual values, giving greater weight to large errors. The square root of MSE, RMSE, provides a comprehensible measure of the average prediction error using the same unit as the target variable. These metrics are frequently used in regression tasks to evaluate the accuracy of the predictions made by the model and to compare different models or adjust the parameters.

Here is a briefly explanation of these metrics:

Mean Absolute Error (MAE): it is a metric used to measure the magnitude of difference between the predicted data and the actual data.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Mean Squared Error (MSE): It is a metric that measures the average square difference between the predicted value and actual target values of the used data.

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

The Root Mean Squared Error (RMSE): It is a metric is well known for using it for evaluating the prediction, Its function like (MSE).

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

Accuracy: classification refers to the ratio of correct predictions to total input samples. It works best with an equal number of samples from each class.

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

3.5- Limitations

3.5.1- Limitations:

Data Quality and Size: The quality and amount of training and testing
machine learning models can significantly impact their performance. Noise, outliers, and non-stationary data can pose challenges.

Black-Box Nature of Models: The black-box nature of deep learning models makes it difficult to analyze their judgments and grasp the reasons behind certain forecasts.

Computational Complexity: Deep learning model training may be computationally costly and need a significant amount of computer resources, which might be a constraint for some applications.

Assumption of Repeatable Patterns: Machine learning models assume that previous patterns will continue to occur in the future, which may not always be true in the volatile and unpredictable stock market environment.

Overfitting: Overfitting is a danger while training machine learning models, particularly with complex datasets, which can result in poor generalization performance on previously unknown data.

3.5.2- Impact of Unexpected Real-World Events:

In addition to the key findings and limitations mentioned, it is important to note that unexpected real-world events, such as financial crises or geopolitical events, can significantly impact stock prices and pose challenges for machine learning models.

Event-driven Volatility: Events like the financial crisis of 2008 can lead to sudden and drastic changes in stock prices, making them difficult to predict using historical data alone.

Model Robustness: Machine learning models trained on historical data may not be robust enough to handle extreme events or outliers that deviate significantly from past patterns.

Need for Adaptability: To address these challenges, machine learning models need to be adaptable and capable of learning from new data in real-time to adjust their predictions accordingly.

Risk Management: Understanding the limitations of machine learning models in predicting unexpected events is crucial for effective risk management and decision-making in volatile markets.

3.6- SVM VS LSTM

Support Vector Machines (SVM) and Long Short-Term Memory (LSTM) networks are two different types of machine learning models used for different types of tasks. Here are the key differences between SVM and LSTM:

Feature	SVM	LSTM
Model Type	Supervised learning, shallow learning	Supervised learning, deep learning
Underlying Algorithm	Statistical learning, optimizes a margin between hyperplanes	Recurrent neural network, learns long-term dependencies
Strengths	Efficient training, good for smaller datasets, effective for classification problems	Captures complex temporal relationships, handles sequential data well
Weaknesses	Limited feature engineering capabilities, not ideal for highly non-linear data	Complex architecture, requires more data for training, prone to overfitting
Suitability for Stock Price Prediction	May be suitable for short-term predictions or classification tasks like predicting price direction	More suitable for long-term predictions and capturing non-linear patterns in historical data
Performance	Less accurate than LSTMs for stock	Often outperforms SVMs in terms of accuracy, especially for complex

	price prediction	datasets
Interpretability	Easier to interpret model behavior	Difficult to interpret internal workings of the model
Computational Cost	Lower computational cost	Higher computational cost because of complex architecture

Table 2. SVM VS LSTM

3.7- Datasets of the previous studies

As shown in the previous studies, the datasets used for stock market prediction have predominantly focused on well-known companies such as Apple, Facebook, Amazon, Google, and Netflix. There are several compelling reasons for this choice. One primary reason is the availability of data. The stock data for these companies is publicly accessible on popular financial data platforms such as Yahoo Finance, Google Finance, and Kaggle. This easy access to comprehensive datasets simplifies the data collection process for researchers, as it eliminates the need for complex data acquisition steps.

Another reason for using these companies is the clarity and quality of their data. These datasets typically have less noise and fewer anomalies, making them ideal for the training process in machine learning models. Cleaner data leads to more accurate predictions, as the models can learn from more consistent and reliable patterns without being misled by extraneous noise. This clarity is crucial for building robust predictive models that can effectively forecast stock prices.

Furthermore, these companies have extensive historical data available. For example, researchers can obtain historical stock prices and related financial metrics from the day these companies went public. This rich historical data provides a solid foundation for training machine learning models, allowing them to capture long-term trends and patterns that are essential for accurate stock market prediction. The availability of long-term data helps in building models that are not only accurate but also resilient to market fluctuations and anomalies.

On the other hand, predicting stock prices for companies with less accessible data presents significant challenges. Datasets that are not readily available or do not contain all the necessary information can hinder the accuracy of predictions. Incomplete data sets often lack the critical historical context needed for the models to learn effectively. Additionally, these datasets might be plagued with higher levels of noise and inconsistencies, further complicating the training process and leading to less reliable predictions.

Even if such datasets are used, the resulting models often suffer from lower accuracy. The absence of comprehensive and clean data makes it difficult for machine learning algorithms to discern meaningful patterns, resulting in predictions that are less reliable and robust. Consequently, researchers tend to focus on well-documented and widely studied companies, as the quality and availability of their data significantly enhance the potential for accurate and effective stock price predictions.

In summary, the preference for using stock data from prominent companies like Apple, Facebook, Amazon, Google, and Netflix in predictive modeling studies is driven by the easy availability, high clarity, and extensive historical records of these datasets. These factors all help to improve the accuracy and reliability of stock market prediction models.

4- Objectives

The main objective of this research paper is to conduct an in-depth investigation into the predictive capabilities of Long Short-Term Memory (LSTM) models and Support Vector Machines (SVM) for forecasting Egyptian stock prices. The study aims to evaluate the effectiveness of LSTM models in comparison to SVM models. This evaluation will be based on a range of performance metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), to provide a comprehensive analysis of each model's predictive accuracy. The results of these models will be compared with findings from previous studies to determine relative performance improvements or deficiencies. Additionally, the research will explore the impact of different hyperparameters and model configurations on the performance of both LSTM and SVM models, with a focus on identifying the most effective approach for stock price prediction. Practical

testing will be conducted on hardware devices to assess the real-world applicability of these models and their ability to comprehend unseen data. Furthermore, the study will investigate potential challenges faced by LSTM and SVM models in stock price prediction, such as overfitting and data preprocessing issues, and propose strategies to mitigate these challenges. The findings of this research are expected to contribute valuable insights into the use of LSTM and SVM models for stock price prediction and provide guidance for future research in this field.

Initially, the study will involve a comprehensive review of existing literature on stock price prediction using LSTM models and SVM models. This review will help to identify the strengths and weaknesses of each approach and provide a foundation for developing effective LSTM and SVM models for the Egyptian stock market. The review will cover various aspects of stock price prediction, including the types of data used, preprocessing techniques, model architectures, and performance metrics. By analyzing the existing literature, the study will identify gaps in the current research and propose new methodologies to improve the accuracy and reliability of stock price predictions.

Following the literature review, the study will focus on the collection and preprocessing of data from the Egyptian stock market. Historical stock prices, and other relevant financial indicators will be gathered from reliable sources such as the Egyptian Exchange (EGX) and financial news websites. The data will be cleaned and processed to remove any noise and inconsistencies, ensuring that the dataset is suitable for training and testing the LSTM and SVM models. Data preprocessing methods such as normalization, feature scaling, and data augmentation will be used to enhance the quality and robustness of the dataset.

Once the data is prepared, the study will involve the development and training of LSTM and SVM models with various hyperparameters and configurations. The hyperparameters to be optimized include the number of layers, the number of neurons in each layer, the learning rate, the batch size, and the number of epochs for the LSTM models. For the SVM models, key parameters such as the kernel type, regularization parameter (C), and gamma will be optimized.

The study will employ techniques such as grid search and random search to identify the optimal hyperparameters that result in the best predictive performance for both models. Additionally, different LSTM architectures such as stacked LSTM, bidirectional LSTM, and LSTM with attention mechanisms will be evaluated to determine the most effective model for stock price prediction. The training process will involve splitting the dataset into training, validation, and test sets to evaluate the models' performance and generalization capability.

To assess the performance of the LSTM and SVM models, the study will compare their predictive accuracy with findings from previous studies. The models will be evaluated based on performance metrics such as MSE, RMSE, MAE, and MAPE, which measure different aspects of prediction errors. MSE measures the average squared difference between predicted and actual values, RMSE is the square root of MSE, MAE measures the average absolute difference, and MAPE measures the average percentage error. These metrics provide a comprehensive assessment of each model's accuracy and reliability in predicting stock prices. Additionally, the study will analyze the directional accuracy of the models, which measures the ability of the model to correctly predict the direction of stock price movements. This is particularly important for investors who rely on accurate predictions of price trends to make informed trading decisions.

Furthermore, the study will investigate the impact of different hyperparameters and model configurations on the performance of LSTM and SVM models. By conducting experiments with various configurations, the study aims to identify the most effective approach for stock price prediction. This includes exploring the use of different activation functions, optimization algorithms, and regularization techniques to improve the models' accuracy and prevent overfitting. Overfitting is a common challenge in machine learning, where the model performs well on the training data but fails to generalize to unseen data. The study will employ strategies such as dropout, early stopping, and cross-validation to mitigate the risk of overfitting and ensure that the LSTM and SVM models are robust and reliable.

In addition to evaluating the predictive performance of the models, the study will also assess their real-world applicability by conducting practical testing on hardware devices. This involves deploying the trained LSTM and SVM models on different hardware platforms such as CPUs,

GPUs, and edge devices to evaluate their computational efficiency and scalability. The study will analyze the time taken for training and inference, the memory usage, and the overall computational cost of deploying the models in a real-world trading environment. This is crucial for determining the feasibility of using LSTM and SVM models for stock price prediction in practical applications, where computational resources and efficiency are important considerations.

Moreover, the study will investigate potential challenges faced by LSTM and SVM models in stock price prediction and propose strategies to address these challenges. This includes issues related to data preprocessing, model selection, and hyperparameter tuning. For example, the quality of the input data is critical for the accuracy of the predictions, and any noise or inconsistencies in the data can significantly affect the models' performance. The study will explore advanced data preprocessing techniques to ensure that the dataset is clean and reliable. Additionally, the selection of appropriate hyperparameters and model configurations is crucial for achieving the best performance, and the study will employ systematic optimization techniques to identify the optimal settings.

Finally, the study will propose strategies to enhance the predictive capabilities of LSTM and SVM models and provide guidance for future research in this area. This includes exploring the integration of LSTM and SVM models with other machine learning techniques such as reinforcement learning, ensemble methods, and hybrid models to improve accuracy and robustness. The study will also investigate the potential of using external data sources such as financial news, social media sentiment, and macroeconomic indicators to enhance the predictive power of LSTM and SVM models. By incorporating additional information, the models can capture a broader range of factors influencing stock prices and provide more accurate predictions.

The findings of this research are expected to contribute valuable insights into the use of LSTM and SVM models for stock price prediction and provide practical guidance for researchers and practitioners in the field. The study aims to demonstrate the effectiveness of LSTM and SVM models in capturing the complex dependencies and nonlinear patterns in stock market data and their potential to outperform findings from previous studies. Additionally, the research will highlight the importance of data quality, hyperparameter tuning, and model selection in

achieving accurate and reliable predictions. The insights gained from this study will contribute to the advancement of machine learning techniques for financial forecasting and support the development of more effective tools for investors and financial analysts.

5- Device Hardware

with Python 3.8, the hardware specifications include an Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz processor, 32.0 GB of RAM, a 1TB SSD, and an NVIDIA GeForce GTX 1660Ti graphics card. Choosing Python 3.8 for its stability ensures that the Python code runs smoothly and reliably, especially when working with machine learning algorithms and handling large datasets. This setup is well-equipped to handle the computational demands of machine learning algorithms, providing efficient processing, fast data access, and faster training times for models like LSTM neural networks.

The Intel Core i7-10750H CPU offers robust performance with its six cores and twelve threads, making it capable of handling complex calculations and parallel processing tasks required in machine learning. The 32 GB of RAM ensures that large datasets can be loaded into memory without causing bottlenecks, while the 1TB SSD provides quick data access and ample storage for datasets and model outputs. Additionally, using a 64-bit Windows 11 operating system ensures compatibility and support for running Python 3.8 and its extensive libraries effectively, making this setup ideal for developing and deploying advanced stock price prediction models.

6- Detailed tasks

The academic year is a journey filled with tasks that contribute to the completion of a research project. This essay outlines these tasks, focusing on a research project that aims to predict stock prices using deep learning.

Task 1: Gathering Background Information
The first step in any research project is to gather background information. In this case, the focus is on understanding the need for stock price prediction and why deep learning is a suitable method for this task. Stock price prediction is crucial for investors and financial analysts as it

helps them make informed decisions. Deep learning, with its ability to learn from large datasets and capture complex patterns, has shown promising results in this field.

Task 2: Initiating the Report

Once the background information is collected, the next step is to start writing the report. This involves documenting the research objectives, methodology, and initial findings. It serves as a blueprint for the entire research project.

Task 3: Model Selection

The third task is to decide on the model to be used for the task. In this project, the (LSTM) model, and (SVM), is selected

Task 4: Literature Review

The fourth task involves a detailed study of related work, particularly focusing on mixed-code sentiment analysis. This helps to understand the current state of the field and identify gaps that the research can fill.

Task 5: Interim Report

After the literature review, an interim report is written. This report provides an update on the progress of the research and outlines the next steps.

Task 6: Dataset Examination

The sixth task is to examine the availability of datasets related to the stock price of EGX. The quality and relevance of the dataset significantly impacts the accuracy of the prediction model.

Task 7: Data Pre-processing

Once the dataset is obtained, it needs to be pre-processed. This involves cleaning the data, handling missing values, and normalizing the data to make it suitable for the LSTM model, and for SVM.

Task 8: Model Implementation

The pre-processed data is then applied to the LSTM ,and SVM. This involves training the model on the dataset and adjusting its parameters to improve its performance.

Task 9: Accuracy Calculation

The ninth task is to calculate the accuracy of the model using appropriate metrics. This helps to evaluate the performance of the model and identify areas for improvement.

Task 10: Documentation

All the implementation stages are documented in the report. This includes the methods used, the challenges faced, and how they were overcome.

Task 11: Model Tuning

The model parameters are iteratively tuned to improve its performance. This involves adjusting the learning rate, the number of layers in the model, and other parameters.

Task 12: Result Documentation

All the results, including the accuracy of the model and the insights gained from the data, are documented in the report.

Task 13: Finalizing the Research Writing

The research writing is finalized by revising the report, ensuring that all the information is accurate and clearly presented.

Task 14: Executive Summary

The final task is to write an executive summary. This provides a brief overview of the research, its findings, and its implications.

7- Gantt chart (Timetable of tasks)

This Gant chart table will explain the tasks that are Done Will done in the full academic year and the duration in weeks and days.

Task	Duration (Weeks)	Duration (Days)
Collect Background Information	4	20
Decide on Model (LSTM),(SVM)	4	20
Related Work & Literature Review	6	30
Interim Report	13	65
Examine Datasets (EGX)	4	20
Pre-process Dataset	7	35
Apply Data to LSTM,SVM	1	5
Calculate Accuracy Metrics	1	5
Document Implementation	24	120
Model Tuning	14	70
Document Tuning Results	14	70
Finalize Research Writing	19	95

Table 3.table of tsaks

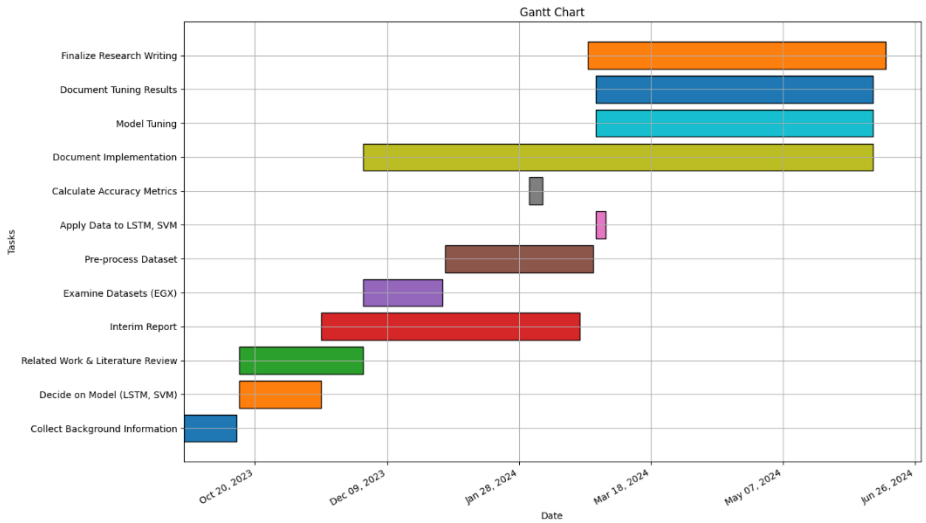


Figure 2. Gantt chart

8- Achievements of semester 1

Having successfully completed a comprehensive course in Machine Learning, I have developed a strong fundamental understanding of the principles and techniques associated with this field. This understanding is indispensable for the effective deployment of cutting-edge models and will serve as a solid foundation for my research endeavors.

In the process of conducting a comprehensive literature review, I meticulously analyzed a variety of research papers on stock price prediction. This thorough examination yielded invaluable insights into the methodologies used, the challenges faced, and emerging trends in the field. Armed with this knowledge, I am better equipped to navigate the complexities of stock market prediction and contribute to the advancement of this area of study.

After careful evaluation, I have chosen the LSTM model and SVM for my research project. LSTM networks are particularly adept at analyzing time-series data, making them exceptionally suitable for tasks related to stock price prediction. The capacity of the LSTM architecture to maintain long-term dependencies in sequential data is beneficial for identifying the intricate

patterns inherent in stock market data, enhancing the accuracy and reliability of predictions. Similarly, SVMs are known for their robust performance in classification and regression tasks, making them a strong candidate for stock price prediction as well. By utilizing the strengths of both the LSTM and SVM models, this research aims to provide a comprehensive evaluation of their predictive capabilities and determine the most effective approach for forecasting Egyptian stock prices.

To obtain the necessary data for the project, I have secured the EGX30, EGX50, EGX70, and EGX100 datasets from (Houssein et al., 2021). This dataset includes historical stock market data specifically designed for the Egyptian Stock Exchange Index, which will form the basis for my analysis and model training. The dataset comprises a variety of features such as opening, closing, high, low, and trading volume for each trading day, providing a comprehensive and rich source of information for my research.

With the dataset in hand, the next steps involve data preprocessing, feature engineering, model implementation utilizing both the LSTM architecture and SVM, and model evaluation. I plan to employ techniques such as normalization and scaling to preprocess the data and extract significant features for input into both the LSTM and SVM models. Additionally, the dataset will be divided into validation and testing sets to assess the models' performance and ensure their effectiveness in real-world scenarios. This rigorous approach will enable me to develop robust and reliable models for stock price prediction, contributing to the advancement of the field.

Furthermore, the research will delve into the analysis of the LSTM and SVM models' performance on various time frames, such as daily, weekly, and monthly predictions. This analysis will provide insights into each model's ability to capture different patterns and trends in the stock market, enhancing their versatility and applicability in different trading scenarios. Additionally, the research will explore the use of ensemble methods and hybrid models to further improve prediction accuracy and robustness. These methods will be evaluated against the baseline LSTM and SVM models to assess their effectiveness in enhancing prediction performance. Overall, the research aims to provide a comprehensive and insightful analysis of the capabilities of LSTM and SVM for stock price prediction, with the ultimate goal of contributing to the development of more accurate and reliable forecasting models in the field.

9 - Anticipated problems and suggested ways to solve them

One of the primary challenges encountered in this study pertains to the research literature itself, where a notable issue is the disparity among research papers in terms of their data, methodologies, and recommendations. The absence of standardized models tailored to this field is attributed to the diverse nature of the datasets used for prediction. To address this, a narrower focus has been adopted, limiting the scope of consideration to research papers published from 2018 to the present. This meticulous approach has facilitated the identification of recent models applied in this domain, thereby aiding in the selection of a suitable model (specifically, LSTM in this instance).

Another significant challenge faced in this research pertains to the availability and quality of the dataset. Initially, the dataset concerning the Egyptian stock market was not readily accessible, requiring alternative measures to be taken. Consequently, the data had to be sourced from a previous scholarly work, which not only introduced potential biases but also added complexity to the data processing and analysis stages. Despite these challenges, efforts were made to ensure the integrity and reliability of the dataset through thorough validation and verification processes.

Furthermore, a notable challenge encountered was the limited availability of academic resources focusing specifically on Egyptian stocks. Despite conducting a thorough search for articles and papers related to this topic, only a handful of relevant pieces discussing Egyptian stocks were identified. However, this limitation is mitigated by the primary aim of this research, which is to empirically evaluate the effectiveness of machine learning in predicting Egyptian stocks. This emphasis on empirical analysis underscores the innovative approach of this study in addressing the dearth of academic resources in this field, contributing to the advancement of knowledge and methodologies in stock market prediction research. The findings of this study are expected to not only enhance the understanding of stock market dynamics in the Egyptian context but also provide valuable insights for future research in this area.

10- Achievements of semester 2

The tasks outlined for the second semester of the academic year are strategically designed to encompass a series of intricate activities aimed at not only enhancing the research outcomes but also meticulously documenting the findings to contribute significantly to the existing body of knowledge.

The first task involves a continuation of experiments focused on model tuning, aimed at improving the accuracy of the predictions. This process involves a systematic exploration of various parameters and configurations for both LSTM and SVM models, with detailed documentation of the effects of each change. This meticulous approach is crucial for gaining a deep understanding of each model's behavior and optimizing their performance effectively.

Subsequently, the application of both Long Short-Term Memory (LSTM) and Support Vector Machine (SVM) models to the acquired dataset is planned, followed by a comprehensive analysis of the results. LSTM, a type of recurrent neural network (RNN), is particularly well-suited for sequential data such as stock prices, making it a valuable tool for this research. SVM, known for its robust performance in classification and regression tasks, will also be employed to predict stock prices. Analyzing the results of both LSTM and SVM applications provides insights into their efficacy in predicting stock prices based on the dataset under study.

A critical aspect of the research involves comparing the results with existing works in the field. This comparative analysis not only contextualizes the current findings but also highlights the advancements and contributions made by the present study. Such an analysis is essential for understanding the state-of-the-art in stock price prediction using both deep learning and traditional machine learning techniques.

Moreover, the meticulous writing of the research paper is planned, which entails expanding on various aspects such as the stock market dynamics, the underlying principles of deep learning and machine learning, the architecture of the LSTM model, and the characteristics of the dataset used. Additionally, details regarding the research environment, programming languages, and

libraries utilized in the implementation of the code for both LSTM and SVM models will be elaborated upon. Furthermore, a comprehensive discussion of all results, their implications, and suggestions for future research directions will be included, adding depth and rigor to the research paper.

The finalization of the research report is also scheduled, ensuring that all sections are coherent, well-structured, and aligned with academic standards. This phase involves reviewing and revising the entire document to ensure clarity, accuracy, and consistency in presenting the research findings.

Lastly, the completion of the Executive Summary of the research is planned, providing a concise yet informative overview of the research objectives, methodology, key findings, and implications. This summary serves as a snapshot of the research for a broader audience, highlighting its significance and potential impact in the field of stock price prediction using both deep learning (LSTM) and machine learning (SVM) techniques.

11- Methodology

11.1- LSTM

One of the key advantages of using LSTM (Long Short-Term Memory) networks over traditional RNNs (Recurrent Neural Networks) is their ability to effectively capture and learn long-term dependencies in sequential data. This is accomplished through the use of a memory cell and gating mechanisms, which enable LSTM networks to keep and update information over time steps, thereby avoiding the issue of vanishing gradients that can occur in traditional RNNs.

Another advantage of LSTM networks is their ability to handle sequences of varying lengths. Traditional RNNs process input sequences one step at a time, making them less effective when dealing with sequences of different lengths. LSTM networks, on the other hand, can process sequences of varying lengths due to their ability to selectively read, write, and reset information in the memory cell, making them more versatile for tasks where the length of the input sequence may vary.

Additionally, LSTM networks are better at capturing and remembering long-term dependencies while avoiding the issues of exploding or vanishing gradients. This is crucial for tasks such as natural language processing (NLP) and speech recognition, where understanding the context of a word or phrase requires remembering information from earlier in the sequence.

Overall, the key advantages of using LSTM networks over traditional RNNs include their ability to capture long-term dependencies, handle sequences of varying lengths, and mitigate issues related to gradient vanishing or exploding. These factors make LSTM networks a powerful tool for tasks involving sequential data analysis and prediction.

11.2- SVM

Support Vector Machines (SVMs) are supervised machine learning algorithms used for regression and classification. SVMs work by determining which hyperplane best divides the data into classes. For regression tasks, the goal is to locate a hyperplane that can predict values that are continuous. SVMs are well-known for their ability to handle high-dimensional data and their effectiveness in situations where the number of dimensions outnumbers the number of samples.

In the context of predicting the close price of stocks, SVM can be a powerful tool due to its robustness in handling non-linear relationships through the use of kernel functions. Kernels transform the input data into higher dimensions where a linear separation is possible. This ability to capture complex patterns makes SVM suitable for stock price prediction, where the relationships between variables can be highly non-linear and intricate.

For predicting stock prices, SVM can be trained using historical data that includes various financial indicators, technical analysis metrics, and other relevant features. The model learns the relationship between these inputs and the closing prices, allowing it to make future predictions. One of the advantages of SVM is its ability to avoid overfitting, especially in high-dimensional spaces, by maximizing the margin between data points and the hyperplane.

11.3- Dataset used in the research

11.3.1- Dataset requirements

To prepare a dataset for stock price prediction using both LSTM, a recurrent neural network, and SVM, a supervised machine learning algorithm, the following are typically required:

11.3.1.1- Data Structure

- **Time-Series Format:** Data should be in a time-series format, with each row representing a date and time and each column representing a stock-related feature or variable.

11.3.1.2- Features

- **Stock-Related Features:** The dataset should include features that have the potential to influence stock prices. Examples include opening, closing, highest, and lowest prices, as well as trading volume.
- **Additional Data:** Incorporate macroeconomic indicators, company-specific information, and other relevant financial metrics to enhance prediction accuracy.

11.3.1.3- Consistency

- **Time Span and Frequency:** The data in all files should be consistent in terms of the time span and frequency of data points (e.g., daily, weekly, monthly). Ensuring consistency is crucial for both LSTM and SVM models.

11.3.1.4- Cleanliness

- **Data Quality:** The data should be free of missing values, outliers, and errors. Any such concerns should be addressed prior to modeling through imputation or removal of anomalies.

11.3.1.5- Normalization

- **Scaling:** Both LSTM and SVM models are sensitive to the scale of data. Typically, data is normalized or standardized before being fed into the models to ensure all features contribute equally.

11.3.1.6- Sequence Length (for LSTM)

- **Input Sequences:** LSTM requires a sequence length, which is the number of previous time steps used as input variables to predict the next period. This sequence length needs to be defined and the data reshaped accordingly.

11.3.1.7- Training and Validation

- **Data Split:** Divide the dataset into training, validation, and testing sets to evaluate the models' performance and ensure their effectiveness in real-world scenarios.

By following these steps, the dataset will be well-prepared for developing robust LSTM and SVM models for stock price prediction, facilitating a comprehensive comparison of their predictive capabilities.

11.3.2- Dataset used

The datasets utilized in this research, which include the Egyptian stock market indices EGX30, EGX50, EGX70, and EGX100, were meticulously selected from the comprehensive study conducted by Houssein et al. (2021). These datasets play a pivotal role in providing a thorough understanding of the dynamics and trends observed in the Egyptian stock market over an extended period.

The EGX30 dataset, covering the period from November 27, 2008, to August 27, 2019, offers a substantial historical perspective on the performance of the Egyptian stock market. Similarly, the EGX50 dataset, spanning from August 2, 2015, to August 27, 2019, provides a more recent dataset for analysis. The EGX70 dataset, with data points from March 1, 2009, to August 27, 2019, and the EGX100 dataset, covering February 8, 2009, to August 27, 2019, complement the

EGX30 and EGX50 datasets, collectively offering a comprehensive view of the Egyptian stock market's historical behavior and trends.

The decision to utilize these specific datasets was primarily motivated by the lack of complete and up-to-date Egyptian stock market data available on mainstream financial platforms such as Yahoo Finance, Kaggle, and Google Finance. Given the importance of having a reliable and complete dataset for accurate analysis, the Houssein et al. dataset emerged as the most suitable option for this research endeavor.

Furthermore, the daily frequency of the data points in these datasets allows for a detailed and granular analysis of stock market trends and patterns. This level of granularity is instrumental in developing robust predictive models and gaining deeper insights into the behavior of the Egyptian stock market. In essence, these datasets serve as the cornerstone of this research, providing a solid foundation for analyzing and predicting stock prices in the Egyptian stock market using machine learning techniques.

11.4- Code apply

The LSTM code processes and analyzes financial information from the EGX30 index using a Long Short-Term Memory (LSTM) model to forecast future stock prices. It starts by loading data from a CSV file, focusing on the 'INDEXCLOSE' column, and cleaning the numerical data by removing commas and converting it to float. The index is converted to datetime format for accurate time series analysis. The data is then normalized using MinMaxScaler. The script prepares the dataset by creating sequences of 30 past days' prices as features (X) and the next day, 3 days, and 5 days' prices as targets (y). These sequences are reshaped to fit the input shape required by LSTM models. The data is divided into training and test sets (80% training, 20% testing). An LSTM model is created and trained using 150 units in the LSTM layer, followed by a Dense layer with a linear activation function, the 'adam' optimizer, and mean squared error loss. Following training, the model makes predictions for the test set. To evaluate, the predicted and actual prices are inversely transformed to their original scale. The script computes and outputs evaluation metrics (MSE, RMSE, MAE, and MAPE) for the 5-day prediction horizon. It also generates and

displays plots comparing actual and predicted prices for visual analysis, in addition to the original data plot.

The SVM code processes and analyzes stock market data from four indices (EGX30, EGX50, EGX70, and EGX100) using Support Vector Regression (SVR). It begins by loading and preprocessing the data, converting date columns, handling missing values, and cleaning numerical data. The script then creates features and target variables for 1-day, 3-day, and 5-day price predictions. The features used include 'INDEXOPEN', 'INDEXHIGH', 'INDEXLOW', 'INDEXCLOSE', 'TRADE_VOLUME', 'TRADE_VALUE', the difference between 'INDEXOPEN' and 'INDEXCLOSE' ('OPEN-CLOSE'), the difference between 'INDEXHIGH' and 'INDEXLOW' ('HIGH-LOW'), a 5-day moving average of 'INDEXCLOSE' ('MOVING_AVG'), and volatility ('VOLATILITY'). It splits the data into training and test sets, standardizes the features, and uses GridSearchCV for hyperparameter tuning to find the optimal SVR model parameters. The models are trained and used to make predictions, which are evaluated using metrics such as MSE, RMSE, MAE, MAPE. The script also generates and saves plots comparing actual vs. predicted values. The results, including evaluation metrics and visual plots, are stored and displayed for each index to assess the model's performance.

12- Results

In this section first the results of this research will be shown first and then we will compare it with the papers that have been featured in this paper, the result will be about predicting 1,3,5 days of prediction

METRIC	ALGORITHM	EGX30	EGX50	EGX70	EGX100
MSE	LSTM	24123.82	1601.85	40.66	289.67
	SVM	17676.45	746.97	72.64	219.89
RMSE	LSTM	155.32	40.02	6.38	17.02
	SVM	132.95	27.33	8.52	14.83
MAE	LSTM	119.47	32.09	4.67	12.99
	SVM	92.14	19.93	5.82	10.09
MAPE	LSTM	0.01	0.01	0.01	0.01

	SVM	1.21	1.03	1.02	0.95

Table 4. result of next day prediction

METRIC	ALGORITHM	EGX30	EGX50	EGX70	EGX100
MSE	LSTM	80047.30	4309.77	336.26	938.26
	SVM	54883.59	2605.02	280.89	562.61
RMSE	LSTM	282.93	65.65	18.34	30.63
	SVM	234.27	51.04	16.76	23.72
MAE	LSTM	222.26	54.49	14.58	23.53
	SVM	165.60	39.11	11.92	17.94
MAPE	LSTM	0.02	0.03	0.02	0.01
	SVM	2.23	2.01	2.11	1.72

Table 5.results of 3 day prediction

METRIC	ALGORITHM	EGX30	EGX50	EGX70	EGX100
MSE	LSTM	193863.12	5874.77	714.78	2621.38
	SVM	108332.13	4505	469.35	1314.99
RMSE	LSTM	440.30	76.65	26.74	51.20

	SVM	329.14	67.12	21.66	36.26
MAE	LSTM	354.00	62.26	21.60	40.01
	SVM	235.71	51.44	15.80	26.24
MAPE	LSTM	0.02	0.03	0.03	0.02
	SVM	3.1	2.65	2.79	2.53

Table 6.result of 5 day prediction

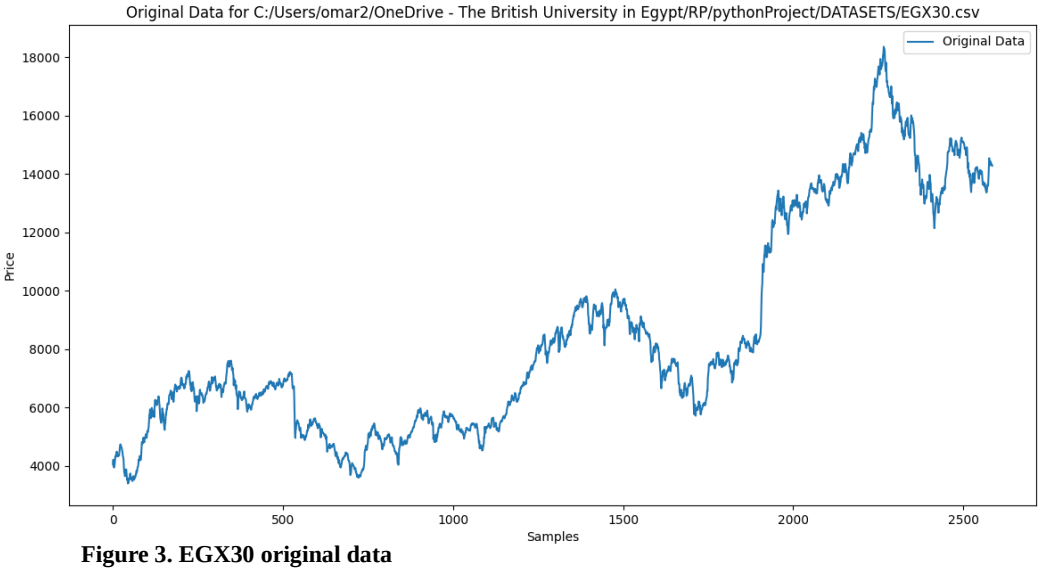


Figure 3. EGX30 original data

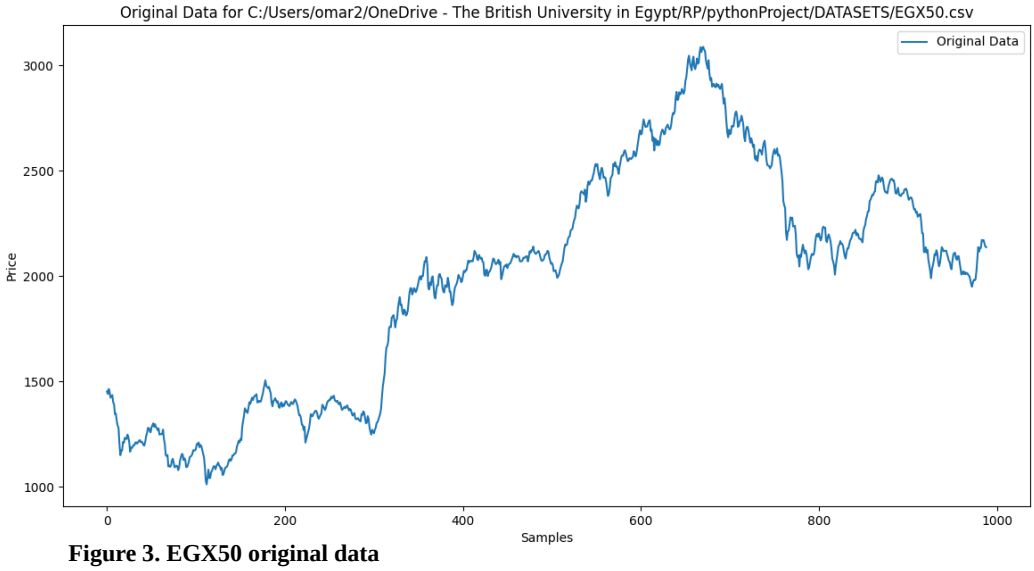


Figure 3. EGX50 original data

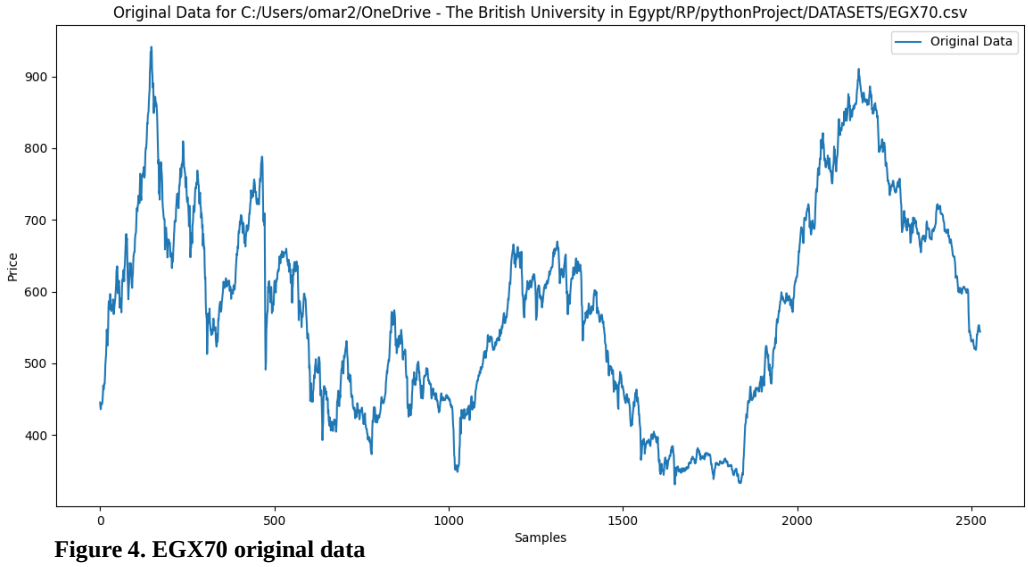


Figure 4. EGX70 original data

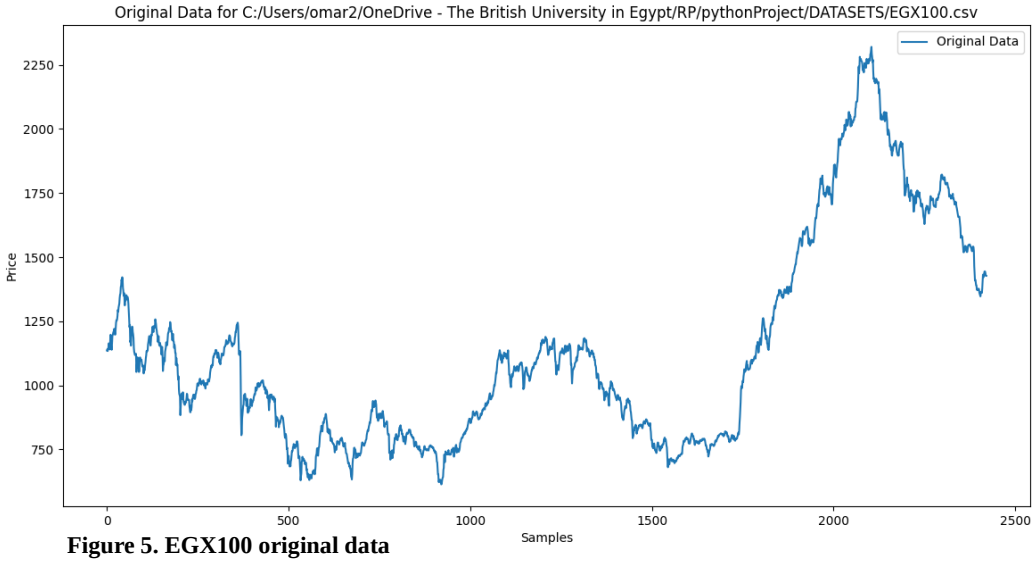


Figure 5. EGX100 original data

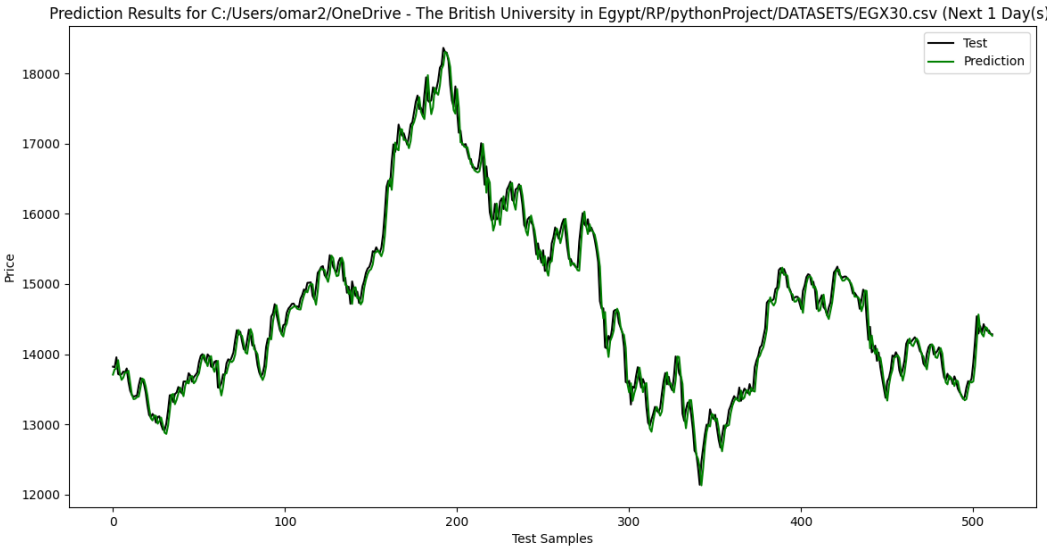


Figure 6. LSTM EGX30 1 day prediction

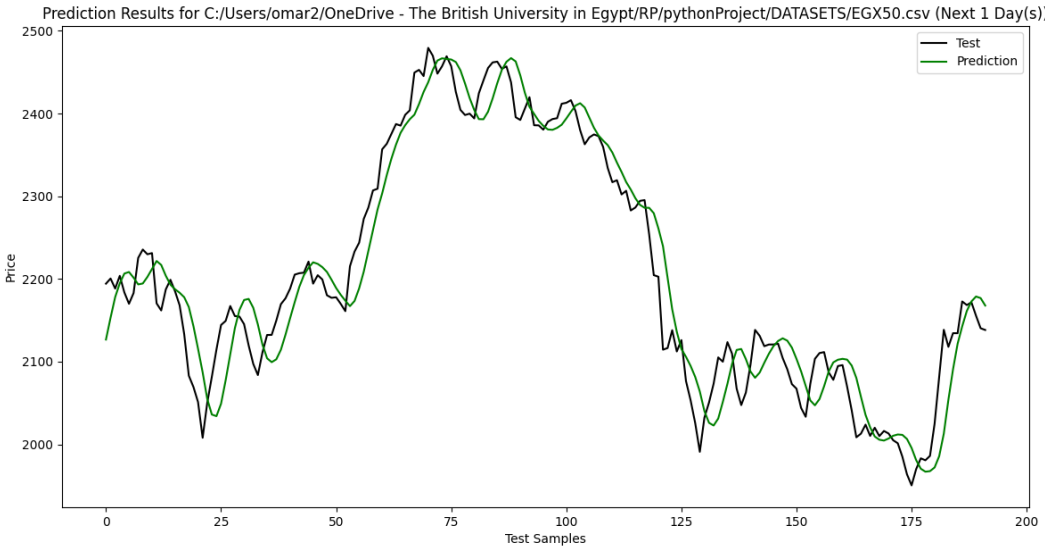


Figure 7. LSTM EGX50 1 day prediction

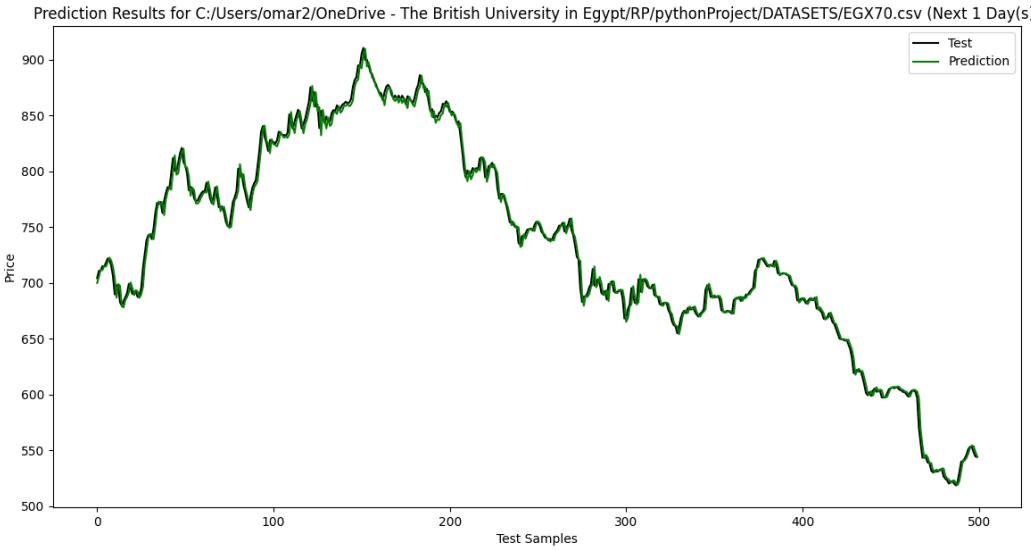


Figure 8. LSTM EGX70 1 day prediction

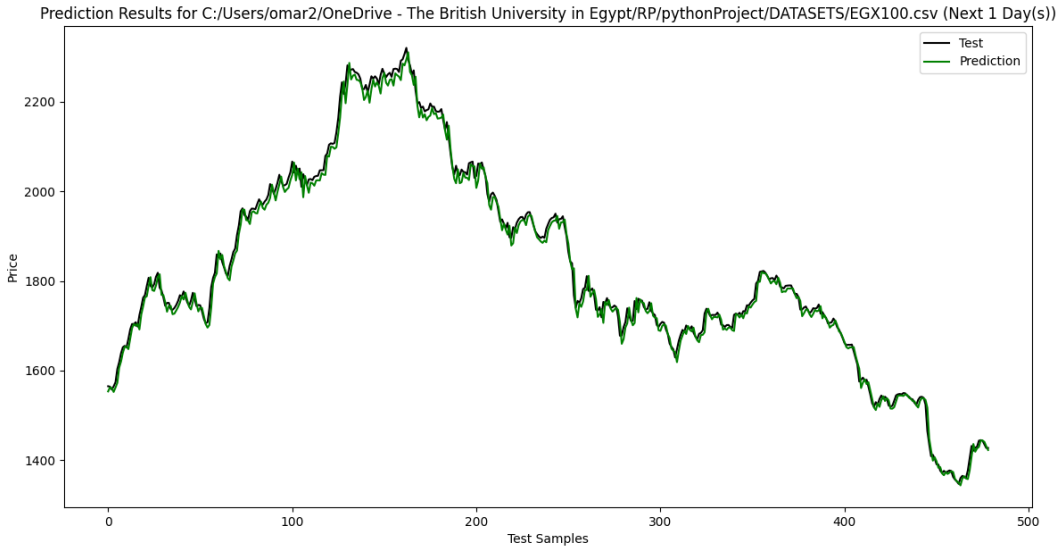


Figure 9. LSTM EGX100 1 day prediction

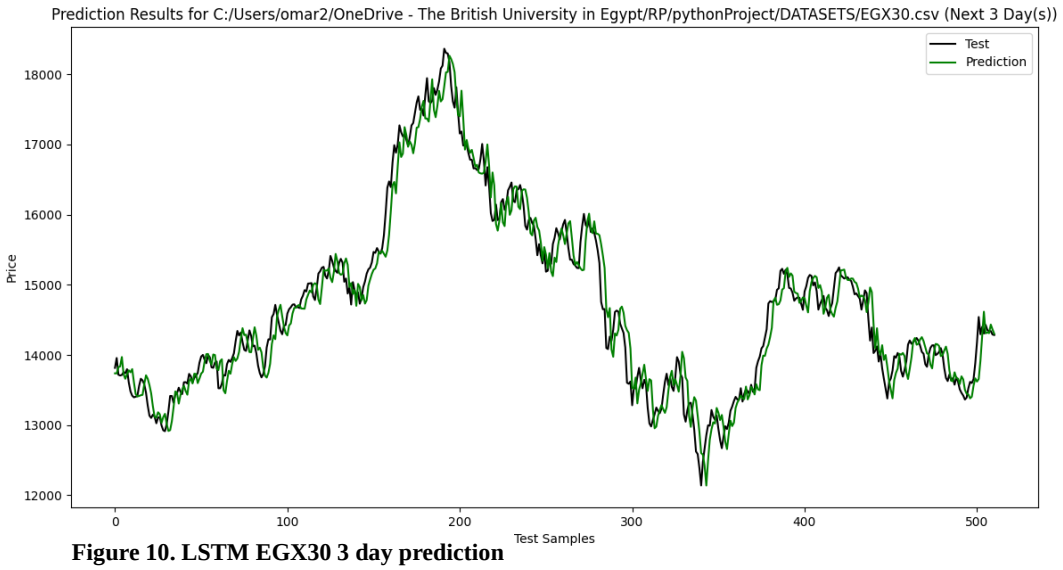


Figure 10. LSTM EGX30 3 day prediction

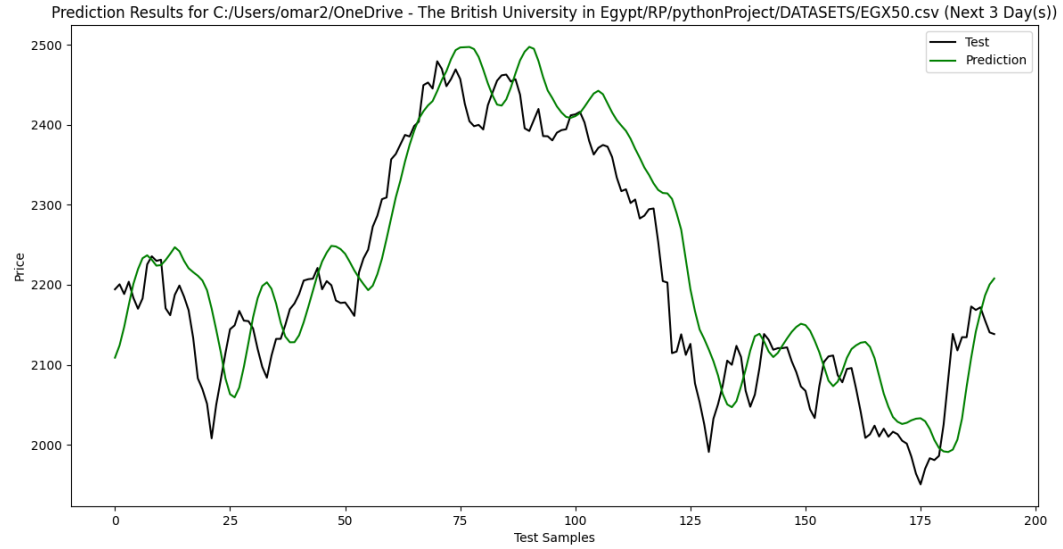


Figure 11. LSTM EGX50 3 day prediction

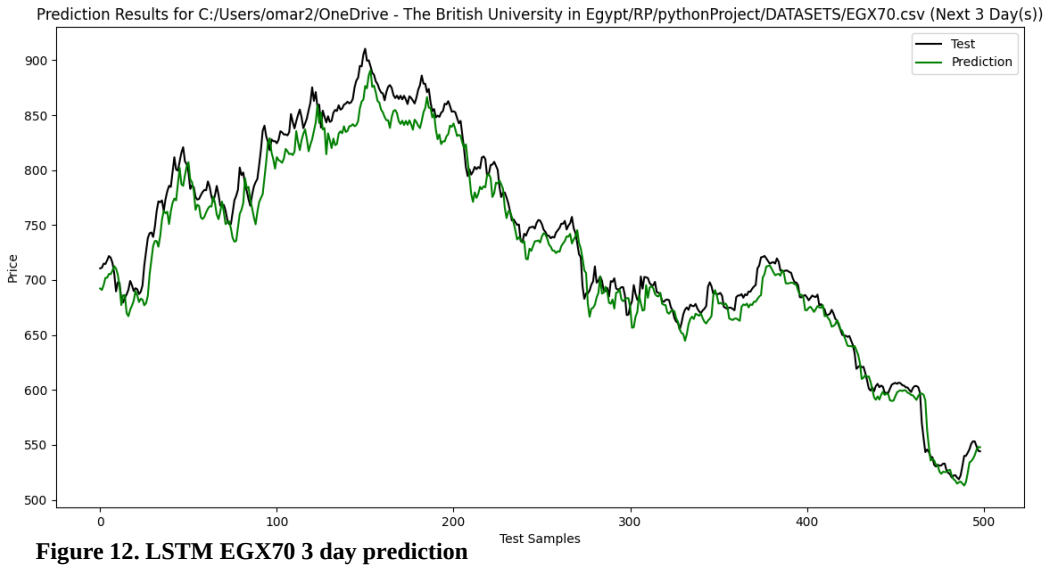


Figure 12. LSTM EGX70 3 day prediction

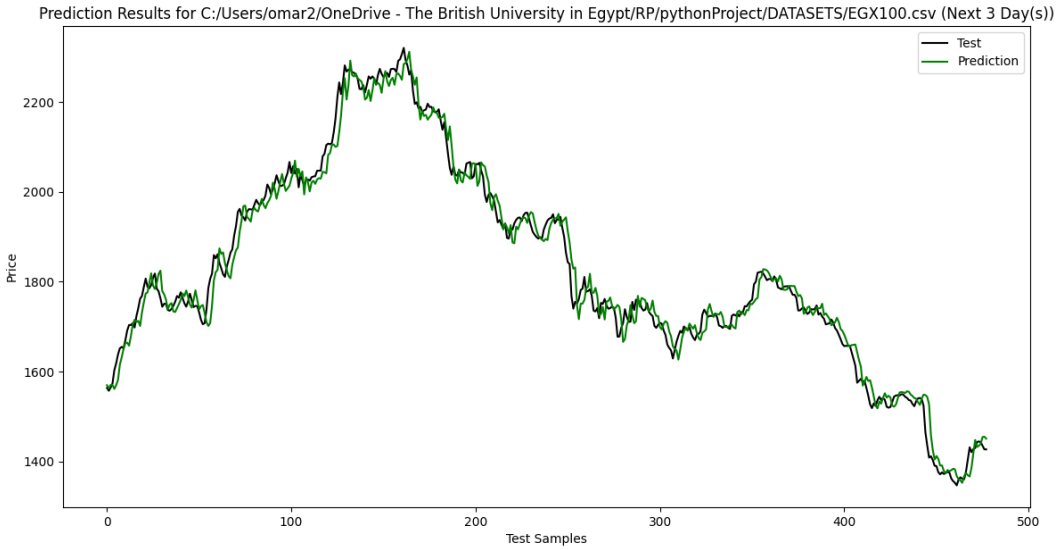


Figure 13. LSTM EGX100 3 day prediction

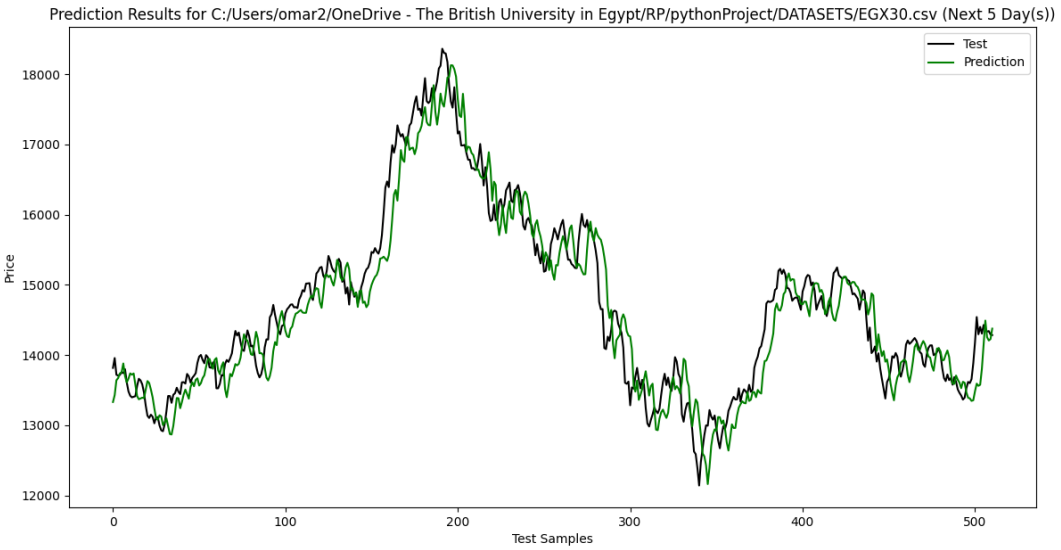


Figure 14. LSTM EGX30 5 day prediction

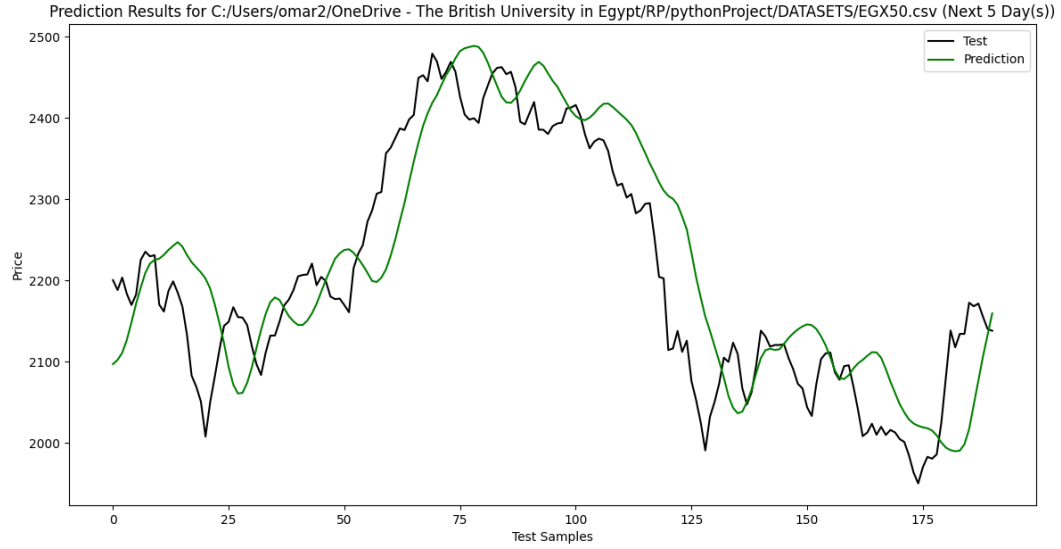


Figure 15. LSTM EGX50 5 day prediction

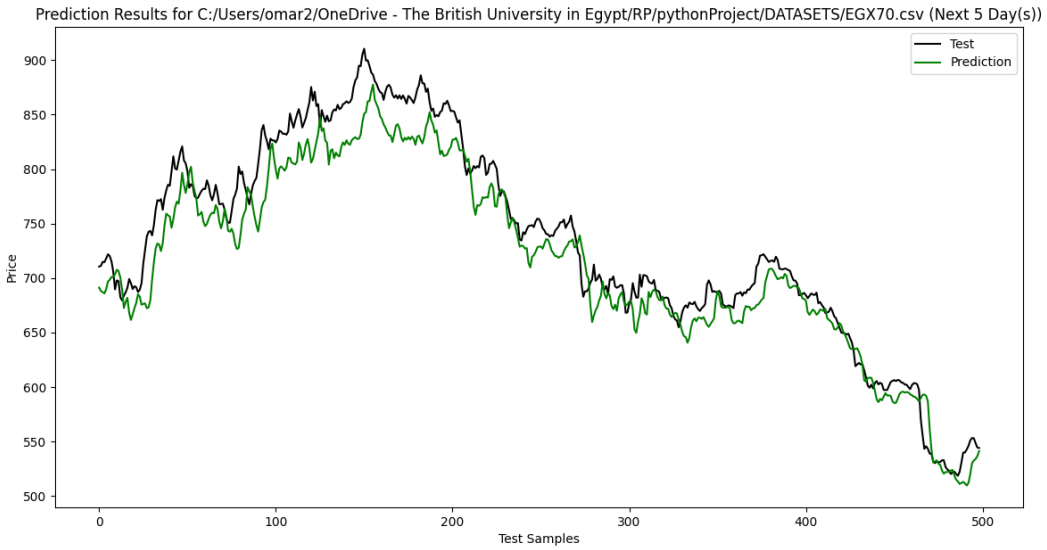


Figure 16. LSTM EGX70 5 day prediction

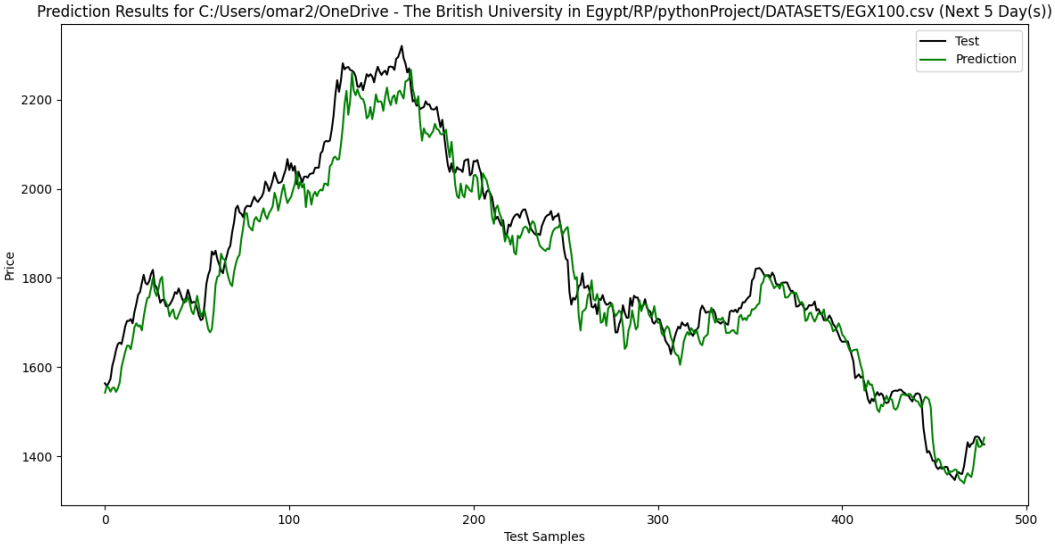


Figure 17. LSTM EGX100 5 day prediction

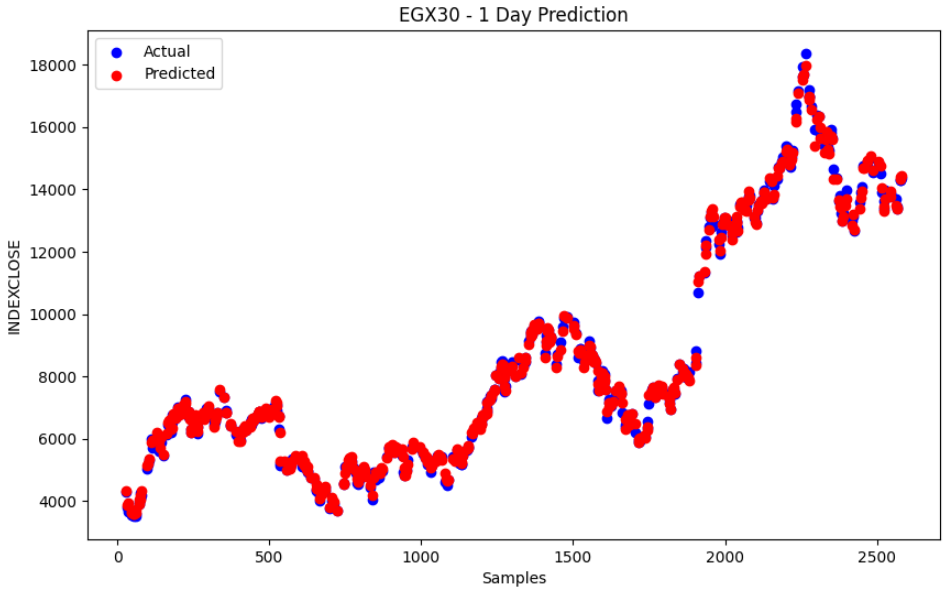


Figure 18. SVM EGX30 predicting next day

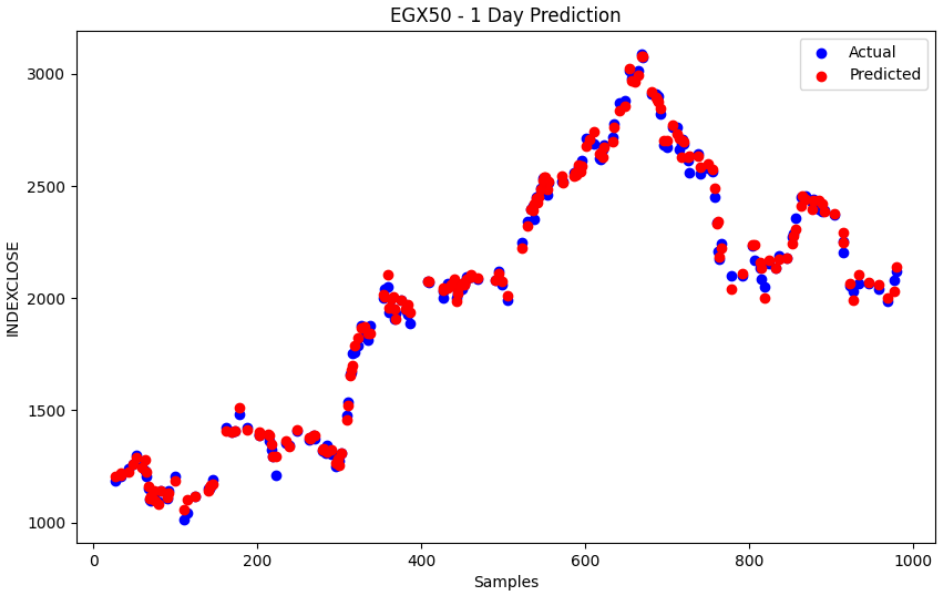


Figure 19 . SVM EGX50 predicting next day

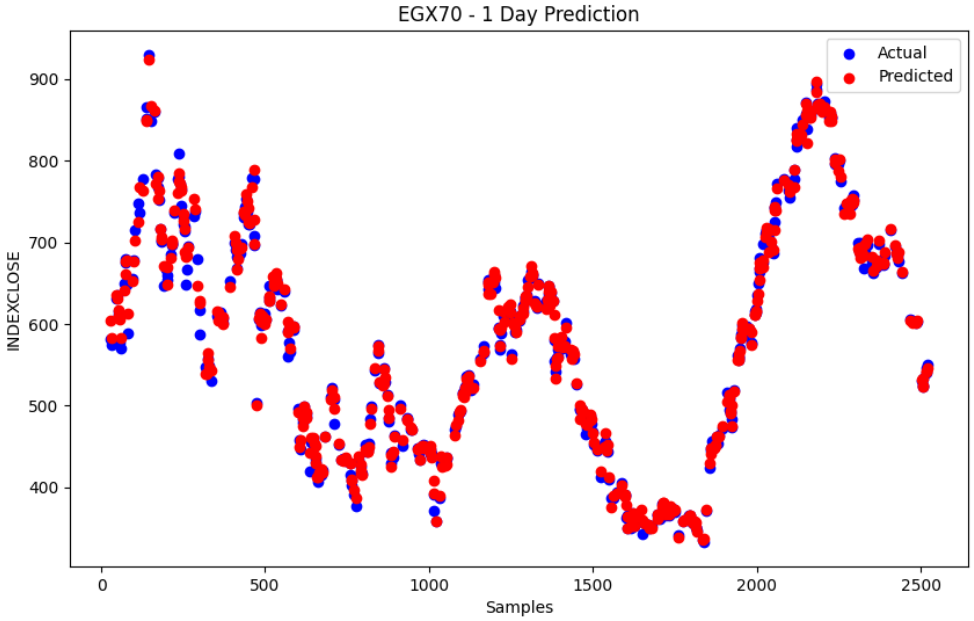


Figure 20. SVM EGX70 predicting next day

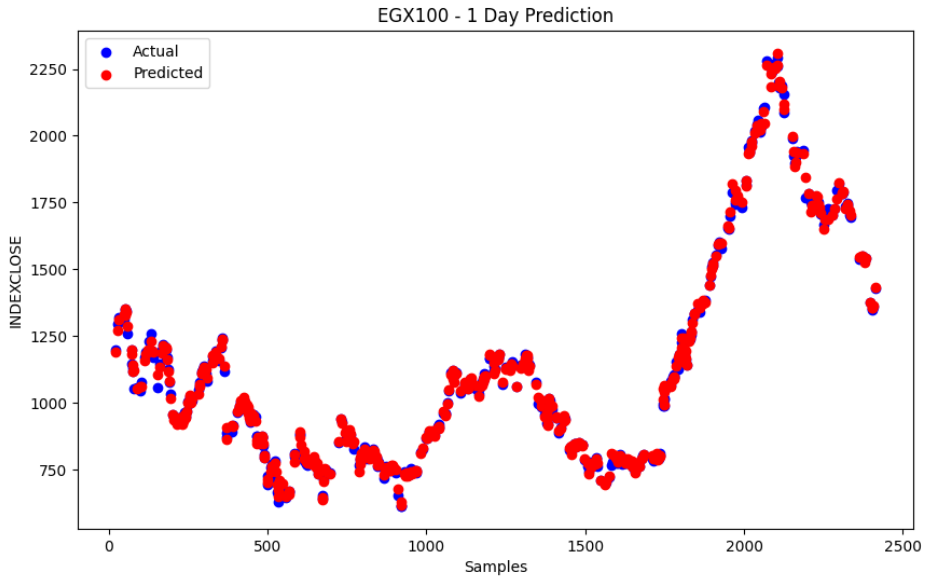


Figure 21. SVM EGX100 predicting next day

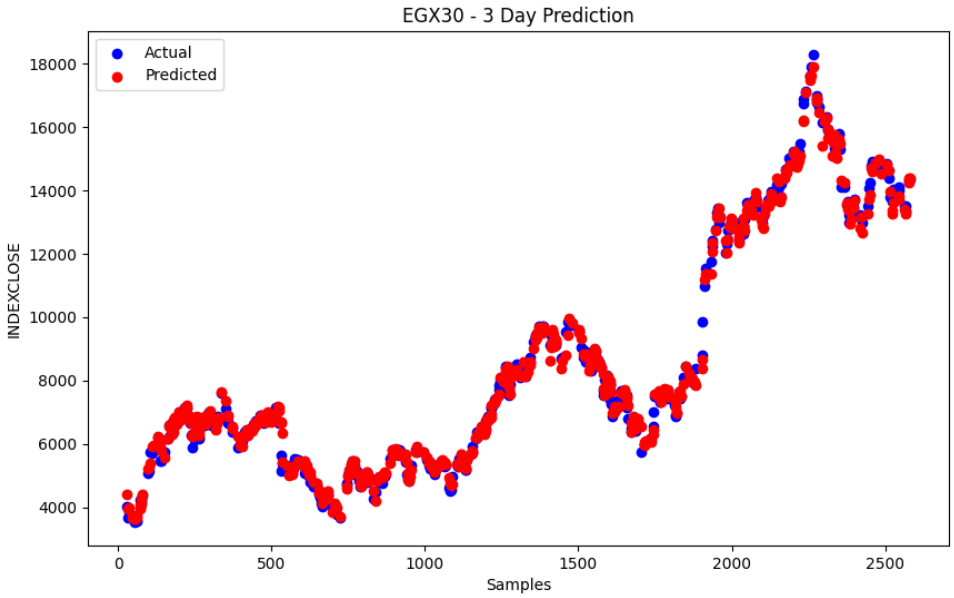
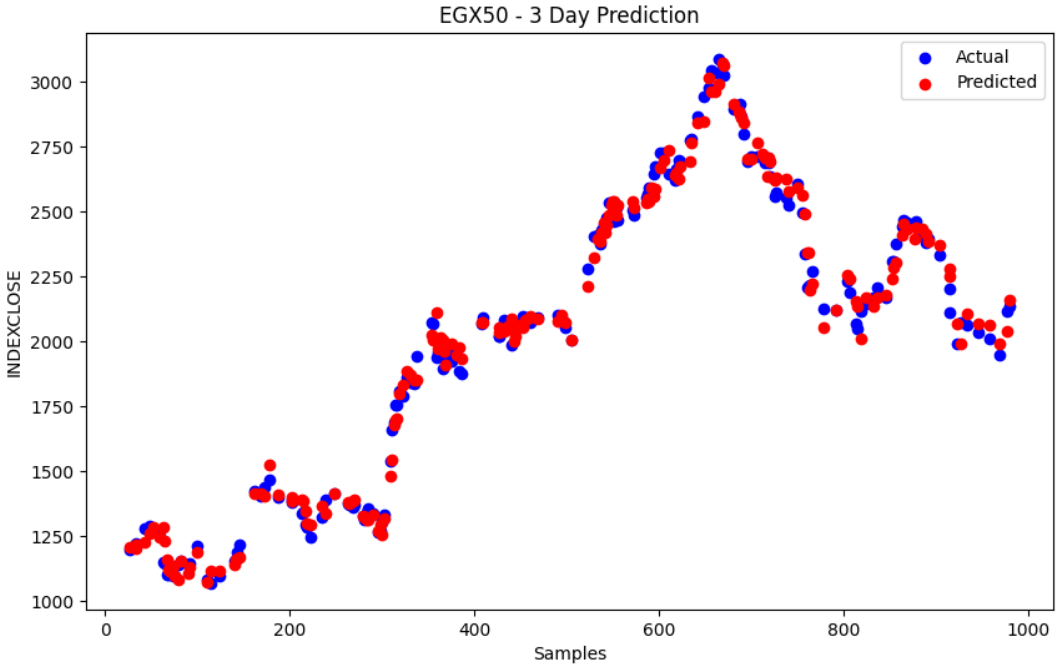


Figure 22. SVM EGX30 predicting next 3 days



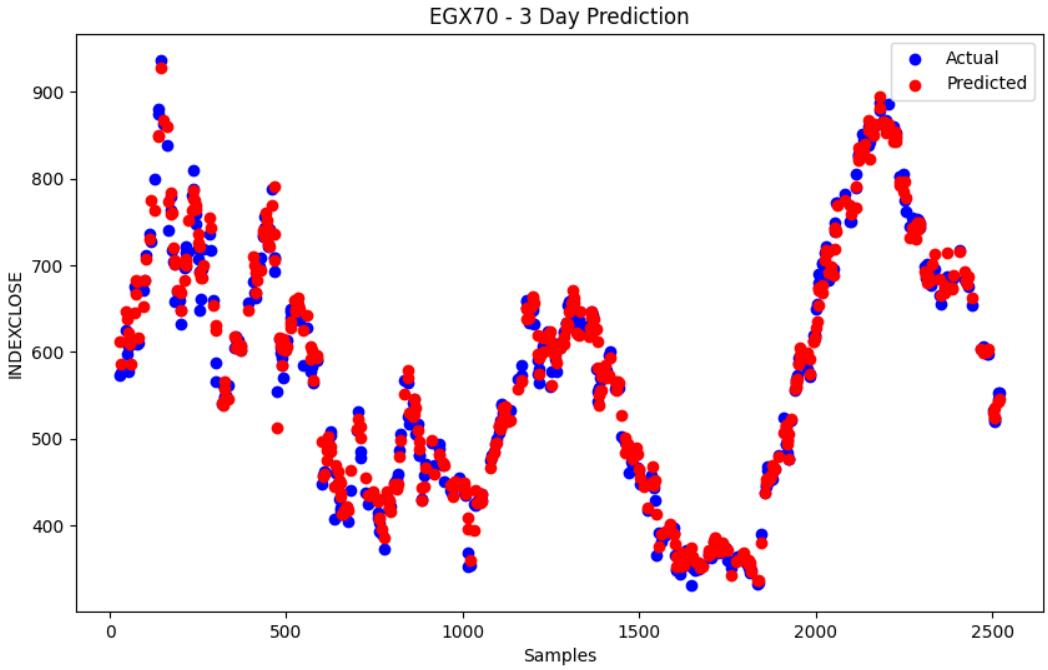
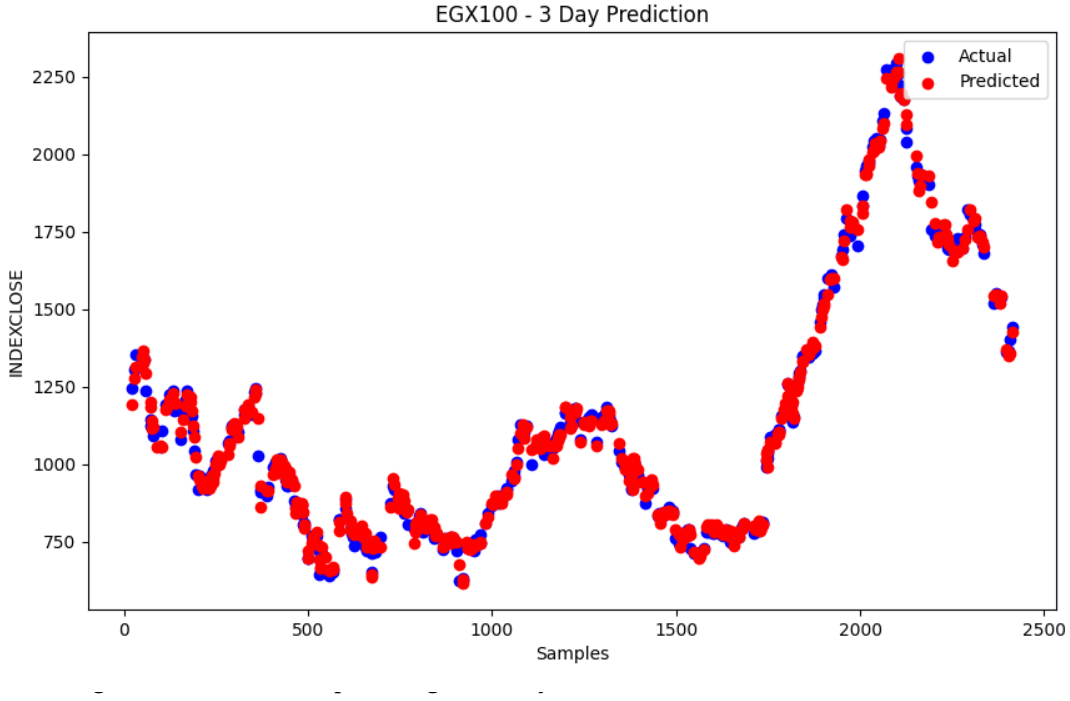
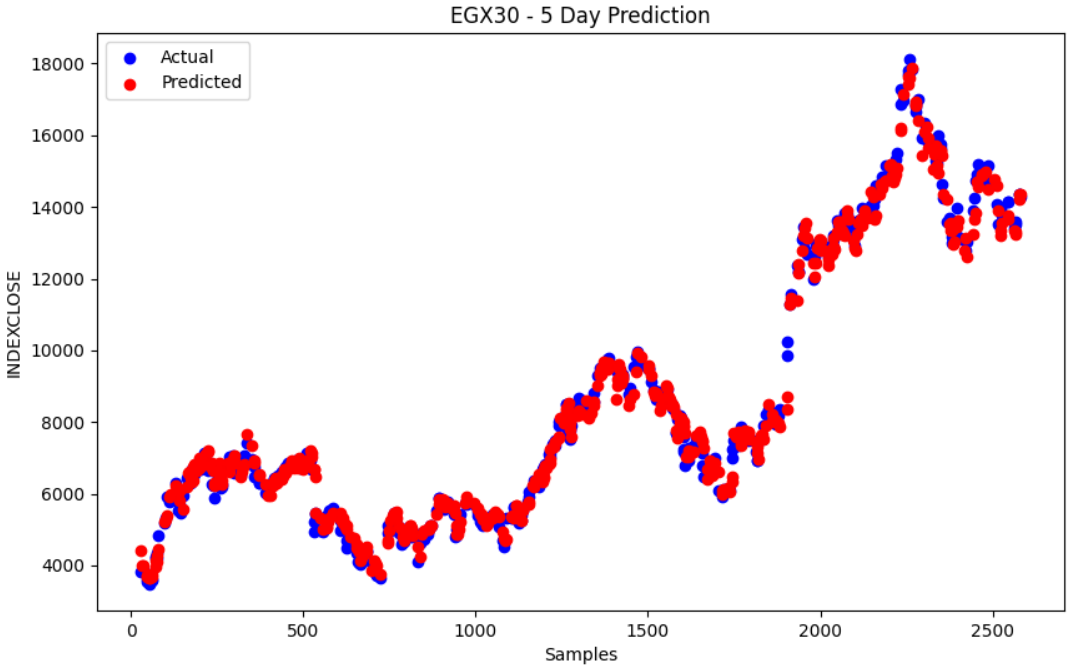
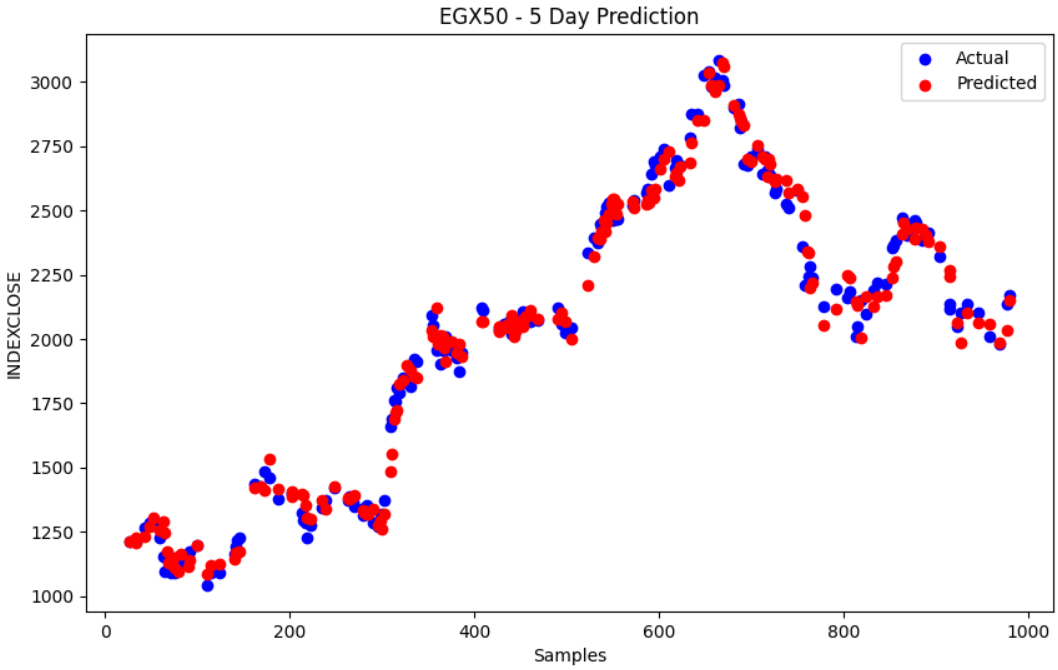
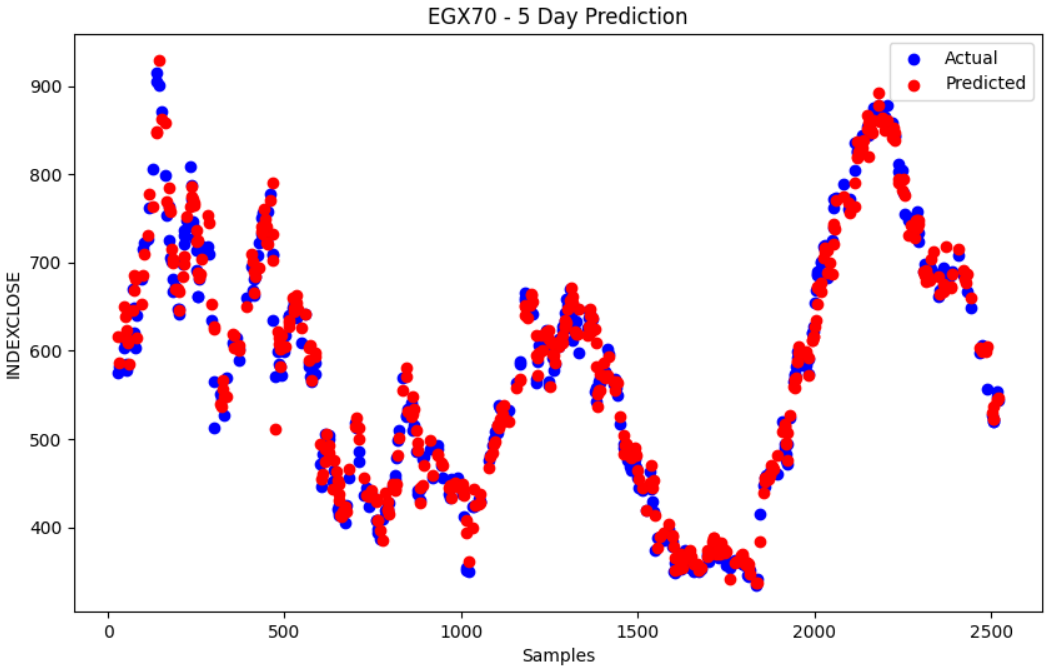


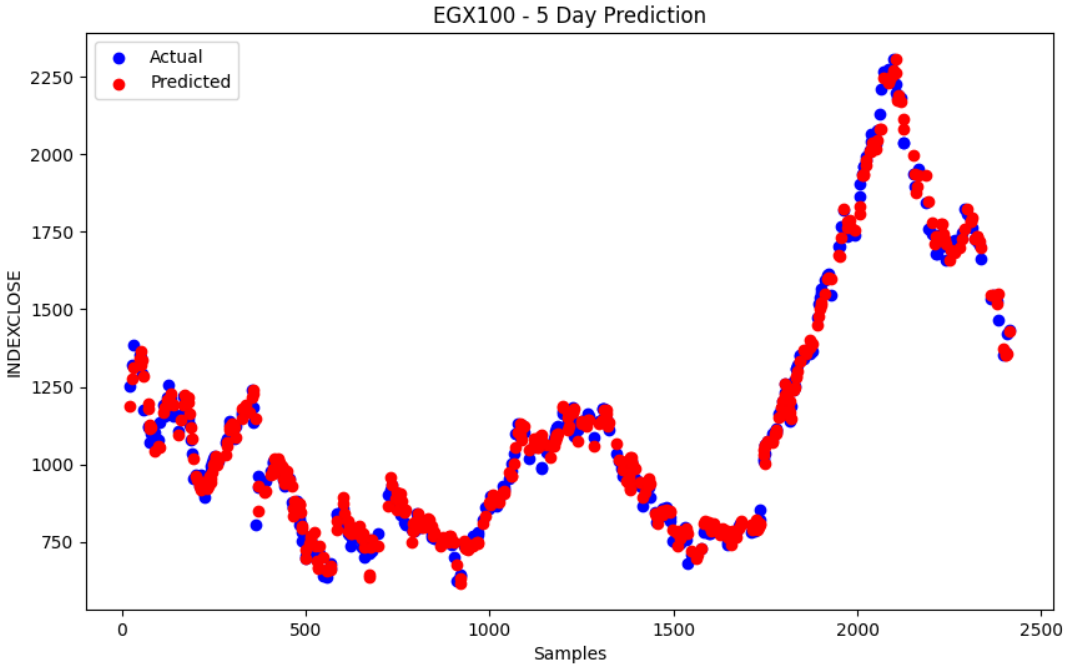
Figure 24. SVM EGX70 predicting next 3 days




EGX30 - 5 Day Prediction
Actual
Predicted
INDEXCLOSE
Samples
18000
16000
14000
12000
10000
8000
6000
4000
0
500
1000
1500
2000
2500


EGX50 - 5 Day Prediction
Actual
Predicted
INDEXCLOSE
Samples
3000
2750
2500
2250
2000
1750
1500
1250
1000
0
200
400
600
800
1000


EGX70 - 5 Day Prediction
Actual
Predicted
INDEXCLOSE
900
800
700
600
500
400
0
500
1000
1500
2000
2500
Samples


EGX100 - 5 Day Prediction
Actual
Predicted
INDEXCLOSE
2250
2000
1750
1500
1250
1000
750
0
500
1000
1500
2000
2500
Samples

As the results shows SVM outperforms LSTM in MSE,RMSE,MAE metrics but in MAPE LSTM has the lowest MAPE .

Number of train samples and test samples for each file:

EGX30: Number of train samples: 2044 , Number of test samples: 512

EGX50: Number of train samples: 767 , Number of test samples: 192

EGX70: Number of train samples: 1996 , Number of test samples: 500

EGX100: Number of train samples: 1912 , Number of test samples: 479

12.1- Compare it with previous studies

in this section the results provided in previous section will be compared with previous studies that have been used in this paper. And it will be started with the paper that have used the same dataset for predicting EGX stock

(Houssein et al., 2021). Stated BR is the best algorithm for predicting stock and used metric MSE X 10^-4 , it will be compared with this paper LSTM algorithm.

Prediction period	ALGORITHM	EGX30	EGX50	EGX70	EGX100
1 day	BR	0.54842	1.11751	1.80849	0.56986
	LSTM	2.4123	0.1601	0.004	0.0289
	SVM	1.7676	0.0746	0.0072	0.0219
3 days	BR	0.60867	1.09449	1.49053	0.46220
	LSTM	8.0047	0.4309	0.0336	0.0938
	SVM	5.4883	0.2605	0.0280	0.0562
5 days	BR	0.52088	1.32792	2.01257	0.53589
	LSTM	19.3863	0.5874	0.0714	0.2612
	SVM	10.8332	0.4504	0.0469	0.1314

Table 7. compare the results with (Houssein et al.,2021)

As the result shows SVM have better results in EGX50,EGX70,EGX100. Also LSTM too compared to results of (Houssein et al., 2021).

For comparing to the other papers, the algorithms will be tested on another dataset most of it will be from yahoo finance.

Comparing with (Chen, 2020)		
company	algorithm	MAPE
AAPL	CNN	2.18
	LSTM	0.03
MAST	CNN	4.17
	LSTM	0.03
FORD	CNN	0.55
	LSTM	0.07
EXON	CNN	0.77
	LSTM	0.02

Table 8. comparing results with (Chen, 2020)

Comparing with (Ho et al., 2021)		
Metric	(Ho et al., 2021)	This paper
RMSE	16.8410	0.2746524635810633
MAPE	0.8184	0.3735519213544084

Table 9. comparing results with (Ho et al., 2021)

As it shows in table 8 and 9 the results is better in these studies taken the exact data and exact historical range taken from these papers and with these results it can be said that LSTM is

reliable for predicting the stocks, the results of SVM is not applied because the SVM model created doesn’t handle the long historical data well.

13- Problems faced on this study

13.1- Faced in SVM and LSTM

Problem: The data preparation was based on reading local CSV files for different indices, which may not have been consistent in formatting.

Solution: Ensured all columns were converted to the correct numeric types and missing values were handled.

Problem: Data contained commas and needed to be converted to numeric types.

Solution: Used string replacement and type conversion to clean the data columns.

Problem: Limited features were used initially, potentially missing important information for the SVM model.

Solution: Added additional features such as OPEN-CLOSE, HIGH-LOW, MOVING_AVG, and VOLATILITY to improve model performance.

Problem: Initial code lacked hyperparameter tuning, potentially leading to suboptimal model performance in SVM.

Solution: Introduced GridSearchCV for hyperparameter tuning to find the best parameters for the SVM model.

Problem: Initial scaling used StandardScaler, which might not handle outliers well in SVM.

Solution: Switched to RobustScaler for better handling of outliers in the data.

Problem: Initial plots were cluttered and hard to interpret in SVM.

Solution: Improved plotting by using scatter plots with markers and dashed lines to make them cleaner and more informative.

13.2- Challenges in Stock Price Prediction with Limited Datasets

The pursuit of accurate stock price prediction using Long Short-Term Memory (LSTM) models is often hindered by the availability and quality of datasets. In this study, the main dataset's historical range spans between 4 to 9 years. However, these ranges are insufficient for highly accurate predictions. Stock market data spanning several decades often provides the rich, varied context necessary for robust modeling. The limited historical data available in this study underscores the difficulties encountered in achieving high prediction accuracy.

13.3- Dataset Acquisition Difficulties

The datasets used in this research were acquired from authors who published their papers, highlighting another significant challenge: the lack of publicly accessible, comprehensive datasets. Ideally, stock data should be freely available from public sources like Yahoo Finance, Google Finance, or other well-known stock data providers. However, the datasets obtained for this study were incomplete, often not containing even a full year of historical data. This limitation not only affects the reliability of the models but also restricts the scope of analysis.

13.4- Contrasts with Previous Research

Previous studies in this field have predominantly used datasets from well-known global trademarks. These brands' stock historical data is readily available from the inception of the stock to the present day, providing a rich dataset for training models. This availability allows researchers to build models that can learn from extensive historical trends and market behaviors, thus enhancing prediction accuracy. The contrast in dataset quality and availability significantly impacts the outcomes of this study compared to those utilizing more comprehensive data.

13.5- Constraints of the Study

The constraints of this study are further amplified by the academic timeline within which it was conducted. This research was undertaken by a single student within the confines of an academic year, limiting the time available for extensive data collection and analysis. Consequently, the information provided and the models developed may not fully capture the complexities of the stock market, and the results should be interpreted with this context in mind.

14- Future plans for the research

The use of Support Vector Machines (SVM) and Long Short-Term Memory (LSTM) networks in stock price prediction has shown promising results, but future improvements can enhance their performance significantly. For SVM, feature engineering can be advanced by incorporating alternative data sources like social media sentiment, news articles, and economic indicators, and implementing dimensionality reduction techniques to reduce noise and enhance generalization. Kernel optimization can be improved through the development of custom kernel functions tailored to specific stock price patterns and enhancing parameter optimization methods using techniques like Bayesian optimization. Ensemble methods, such as hybrid models combining SVM with Random Forests or Gradient Boosting and using bagging and boosting techniques, can increase robustness and accuracy. Time-series specific approaches, including incorporating lagged features and moving averages, and employing various windowing techniques for data segmentation, can also enhance SVM performance.

For LSTM networks, model architecture can be enhanced by integrating attention mechanisms to focus on relevant input sequences and using bidirectional LSTMs to capture dependencies in both forward and backward directions. Data augmentation strategies, such as using Generative Adversarial Networks (GANs) for synthetic data generation and injecting noise into training data, can improve model robustness against market volatility. Regularization and optimization techniques, like implementing dropout layers and batch normalization, and experimenting with advanced optimizers such as AdamW, Ranger, or Lookahead, can prevent overfitting and enhance convergence speed and stability. Hybrid models that combine LSTM with statistical models like ARIMA or GARCH, and multi-model approaches using LSTM with other deep learning models like Convolutional Neural Networks (CNNs), can capture both linear and non-linear patterns in stock prices.

15- Conclusion

In conclusion, this research has explored the application of machine learning models, specifically Long Short-Term Memory (LSTM) networks and Support Vector Machines (SVM), for predicting stock prices within the Egyptian stock market. By meticulously preprocessing data, selecting relevant features, and optimizing model parameters, the study has demonstrated the potential of these advanced models in capturing the complex and nonlinear patterns inherent in financial data.

The findings reveal that while both LSTM and SVM models exhibit robust predictive capabilities, each model has its unique strengths. The LSTM model excels in capturing long-term dependencies and trends in the data, making it highly effective for sequential data analysis. On the other hand, the SVM model, with its proficiency in handling high-dimensional data and non-linear relationships, also performs admirably, particularly in shorter prediction horizons.

Comparative analysis with previous studies further underscores the reliability of these models. The LSTM model's performance in various metrics, such as Mean Absolute Percentage Error (MAPE), indicates its superiority in certain contexts, while the SVM model's lower Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) in specific datasets highlight its efficacy.

Despite the promising results, the study acknowledges several limitations, including data quality issues and the challenges posed by limited historical data. The availability of comprehensive datasets and the inclusion of alternative data sources such as social media sentiment and news articles could significantly enhance model accuracy and robustness.

Future research should focus on integrating hybrid models and ensemble techniques to leverage the strengths of both LSTM and SVM models. Additionally, incorporating real-time data feeds and exploring more sophisticated preprocessing techniques could further improve predictive performance. By addressing these areas, the predictive capabilities of machine learning models in stock price forecasting can be significantly enhanced, providing valuable tools for investors and financial analysts.

Overall, This study adds to the expanding body of research in the field of financial prediction, demonstrating the potential of machine learning models in predicting stock prices and offering insights for further advancements in this domain.