





**October University for Modern Sciences and Art**

**Faculty of Computer Science**

**Graduation Project**

**AI-based system for stock market predictions**

**Supervisor: Dr. Hatem Elattar**

**Name: Omar Tamer, Karim Akmal**

**ID: 220557, 223309**

**Abstract**

After evaluating and comparing multiple models to enhance prediction accuracy, the final outcome of the project includes the development of a comprehensive financial dashboard. This dashboard displays historical and current stock prices, and it enables users to forecast stock price trends up to 15 days into the future. Moreover, an AI-powered chatbot assistant has been integrated to serve as a virtual financial advisor, helping users interpret predictions, understand market movements, and receive tailored insights based on model outputs. This expansion bridges the gap between predictive analytics and practical investor decision support.

The stock market forecasting is a paramount and challenging problem because the stock market fluctuates and is influenced by several external factors such as company performance, investor sentiment, and global economic events. The developed traditional financial model can hardly capture the non-linear complex relationships based on the stock market data, and thus, artificial intelligence is introduced as a method to improve the forecasting model. The task of the project is to explore the different artificial intelligence methods to improve the stock price forecasting. This is an essential point in the advancement of the project thus will be studied in detail in the following sections.

The research encompasses a variety of models such as Long-Short Term Memory (LSTM) networks, AutoRegressive Integrated Moving Average (ARIMA), Random Forest, Transformers, and Graph Neural Networks (GNNs), which are evaluated based on their capability to analyze historical stock data, identify trends, and make reliable forecasts. The feature engineering techniques employed include Simple Moving Averages (SMA), Exponential Moving Averages (EMA), and the calculation of volatility, which are used to improve model performance. In addition, the use of hybrid models combining multiple models is also explored in order to optimise the level of accuracy while maintaining model efficiency, which is a critical aspect of project development that will be further explored in the subsequent sections.

For the purpose of real-time predictive capabilities, we make use of datasets from Yahoo Finance. Key indicators for financial data includes opening and closing prices, trading volumes, and adjusted close prices. The evaluation process involves training models on a large amount of historical data and then testing the performance on unseen data. The comparison of results is based on RMSE and accuracy metrics. LSTM's and Random Forests show the highest predictive performance; however, the use of advanced architectures, such as Transformers and GNN's, can enhance adaptability and robustness in volatile market conditions. This is a crucial point of the project and will be thoroughly discussed in the next sections.

**Table of Contents**

1. **Introduction**  
   1.1 Introduction  
   1.2 Problem Statement  
   1.3 Objective  
   1.4 Motivation  
   1.5 Thesis Layout
2. **Background and Literature Review**  
   2.1 Background  
   2.1.1 Machine Learning  
   2.1.2 Deep Learning  
   2.1.3 Transfer Learning  
   2.2 Previous Work
3. **Materials and Methods**  
   3.1 Materials  
   3.2 Methods  
   3.3 System Architecture Overview

**Chapter 1: Introduction**  
Dashboard and Visualization Layer: A user-friendly web interface that displays current and historical stock price data, performance metrics, and model-driven forecasts up to 15 days into the future. Users can interact with different visualizations to explore stock performance.  
  
AI Chatbot Assistant: An intelligent conversational interface integrated into the dashboard, designed to answer user queries, explain stock prediction results, and provide context-aware financial advice based on model outputs and market data.

**1.1 Introduction**

Stock market prediction is a substantial though intricate chapter of the financial studies. Global economic trends, market sentiment, and real-time events have an increasing impact on the market, and traditional methods of analysis are often found lacking in efficacy. The advancement of artificial intelligence, specifically machine learning and deep learning, provided a new method of analyzing stock data. AI-based models, using historical data and real-time market inputs, are able to capture subtle patterns and help investors make the right decisions. This point has great importance for the project and will be analyzed in a thorough manner throughout the following sections.

The paper investigates the processes and methods of using artificial intelligence, integrating machine learning and deep learning algorithms to increase the accuracy of predicting stock prices of companies and optimizing computational efficiency. The developed system evaluates the most popular models and methods of machine learning and deep learning, such as the LSTM method, the ARIMA method, Random Forest, the Transformer model, the Graph Neural Network (GNN) method, and other advanced methods. This point is crucial for the development of the project and will be discussed in detail in the next section.

**1.2 Problem Statement**

SThe stock market forecast is one of the significant problems due to the dynamics and volatility of financial markets themselves. Classic methods, such as linear regression or primary statistical models, do not catch on to the intricate and nonlinear dependencies within the movements of stock prices, therefore, the predictions made by such conventional methods are not always correct. This point plays an important role in the project and will be explored in detail in the following sections.

Machine learning and deep learning models provide more accurate predictions. They are more reliable. However, they have a downside. They are computationally expensive and a real-time process is a tough nut to crack. Also, many AI models fail to incorporate other market factors such as macroeconomic factors, investor sentiments or real-time events. The central issue this research addresses is the need for a model that can have high accuracy and high computational efficiency. This will enable it to adapt to real-time market changes. This is the pivotal point in the project and will be elaborated in the upcoming sections.

**1.3 Objective**

We aim to research a method for enhancing stock market prediction using artificial intelligence. The project is based on a concept of using advanced machine learning and deep learning methods. The research will evaluate and implement a variety of models, such as LSTM, ARIMA, Random Forests, XGBoost, LightGBM, and other emerging models to improve prediction accuracy. The project will also involve incorporating external financial indicators to improve predictability. It will also involve refining data filtering techniques and using hyperparameter estimation with the goal of creating a scalable model that works in real time. Looking ahead, the goal of the project is to develop a decision support system for investors. This will improve the outcome of investment activities and reduce the uncertainty of investment decisions. This issue is of significant importance in investment decisions, and it will be exposed in greater detail in the subsequent sections.

**1.4 Motivation**

Stock Market Prediction is an important field of study as it has a huge effect on financial decisions, investment strategies and the economic stability of a country. The rapid development of artificial intelligence and the wide availability of large-scale financial data have allowed more advanced predictive models to be developed. Unfortunately, the financial markets are too complex to be determined by any single method and are based on a number of historical patterns, investor sentiment and macroeconomic trends. All of these factors are not considered by traditional models, which is why they give inaccurate forecasts. This is one of the most important aspects when developing the project and it will be described in detail later in the paper.

The fusion of deep learning and machine learning techniques opens the door to better predictive accuracy and computational efficiency. For example, LSTM networks perform best in capturing long-term dependencies within time series data and Random Forest delivers a robust feature selection mechanism. In the recent past, transformer-based models and graph neural networks have shown promising results in pattern recognition and relational learning, respectively. The motivation for this research is to bridge the gap between the latest artificial intelligence research and realistic financial forecasting applications. This research is supposed to assist in the development of an innovative model that will be able to contribute to the intelligent stock market analysis, informing investors and financial institutions. This is a critical assumption in the development of the research and will be explored further in the upcoming sections.

**1.5 Thesis Layout**

This thesis is structured as follows:

* Chapter 2 provides a comprehensive review of existing stock market prediction techniques and the role of AI in financial forecasting. It explores various literature sources and discusses their contributions and limitations. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* Chapter 3 details the materials and methods used in the research, including data sources, preprocessing techniques, model selection, and performance evaluation metrics. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.

**Chapter 2: Background and Literature Review**

**2.1 Background**

The topic of stock market prediction is central to the research not least because the results can influence financial decisions. Traditional methods of prediction were based on fundamental and technical analysis, but with the development of artificial intelligence, new ways of predicting have appeared. Artificial intelligence is successfully applied in the fields of natural language processing, machine learning and deep learning, and the prediction of stock prices is not an exception. The given point is crucial for the project development, so it will be analyzed in the subsequent sections.

Quick growth of fintech allows to adopt AI-driven predictive models that can combine a wide range of data sources such as historical stock prices, economic indicators, news sentiment, and trading volume trends. However, despite the significant progress, there are still challenges in a task of getting the most accurate, computationally efficient, and most adaptable to the market’s changes model. This aspect plays a crucial role in the development of the project and will be considered in further details in the following sections.

**2.1.1 Machine Learning**

Stock market prediction has been revolutionized by machine learning. Patterns and relationships within financial data are identified by the machine learning algorithm. These patterns are then used to predict future stock prices. Decision Trees, Support Vector Machines, and ensemble methods like Random Forest have been extensively used to classify stock trends and predict future stock prices. These models learn from stock data and adapt to new market conditions, improving prediction accuracy. This is a critical point in the development of the project and will be explored in detail in the upcoming sections.

**2.1.2 Deep Learning**

Deep learning techniques like LSTM networks and CNNs have significantly improved predictive potential. These models are good at capturing sequential dependencies and extracting high-level features from stock data. Also, transformer-based architectures such as BERT and GPT have been increasingly explored for financial forecasting, providing more context-aware predictions. This point plays the key role in the development of the project and will be explored in detail throughout the upcoming sections.

**2.1.3 Transfer Learning**

Transfer learning allows the usage of existing models for financial data. This method eliminates the need to train a model from scratch. One can fine-tune pre-trained models on stock market datasets, using the previous knowledge to improve the accuracy of the system. This approach results in faster model acquisition and improved performance in various market conditions. This point is essential for the project, and it will be explained in detail using the upcoming sections

.**2.2 Previous Work**

**2.2.1 Stock Price Analysis and Prediction (LSTM, RF)**

A study by Vora et al. (2021) explored stock price prediction using LSTM and Random Forest models. Key findings included:

* Data: Historical stock prices from Yahoo Finance. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* Methods: Linear regression, decision trees, random forest, and LSTM. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* Results:
  + Random Forest achieved 68% accuracy in regression tasks. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
  + LSTM outperformed traditional methods with over 75% accuracy for short-term predictions. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* Limitations: The study did not consider external factors such as economic indicators and news sentiment, limiting its applicability across different market conditions. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.

**2.2.2 Prediction and Analysis of Stock Markets Using ARIMA Models**

Deepan et al. (2024) proposed a hybrid approach combining AutoRegressive Integrated Moving Average (ARIMA) with LSTM to enhance stock price predictions. Key aspects included:

* Data: Historical stock prices and external factors such as trading volumes and news sentiment. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* Methods:
  + ARIMA for capturing linear dependencies. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
  + LSTM for modeling non-linear relationships. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* Results:
  + ARIMA Accuracy: Amazon: 94.86%, Apple: 97.86%, Microsoft: 94.56%. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
  + LSTM Accuracy: Amazon: 93.40%, Apple: 97.81%, Microsoft: 93.76%. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* Limitations:
  + ARIMA's reliance on linear trends reduces its effectiveness in volatile markets. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
  + LSTM models risk overfitting without proper optimization. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
  + The hybrid approach was not tested in real-time scenarios. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.

**Chapter 3: Materials and Methods**

**3.1 Materials**

This study utilizes stock market data sourced from Yahoo Finance API, which includes various financial indicators. The dataset consists of historical stock prices from multiple companies, covering a broad range of industries to ensure a diverse and generalizable prediction model. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.

**Dataset Features:**

* **Date:** The trading date for each stock entry. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* **Opening Price:** The initial stock price at the beginning of the trading day. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* **Closing Price:** The final stock price at the end of the trading day. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* **High and Low Prices:** The highest and lowest prices reached within a trading session. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* **Adjusted Closing Price:** Adjusted for stock splits, dividends, and other corporate actions. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* **Trading Volume:** The total number of shares exchanged during a specific trading period. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.

**Tools and Techniques Used:**

* Programming Language: Python
* Data Processing: Pandas, NumPy
* Visualization: Matplotlib, Seaborn
* Machine Learning Libraries: Scikit-learn, TensorFlow, PyTorch
* Deep Learning Architectures: LSTM
* Machine Learning: Random Forest, XGBoost, LightGBM
* Time Series Analysis: ARIMA, Exponential Smoothing
* Feature Engineering: Moving averages, sentiment analysis, volatility calculations

**3.2 Methods**

This research follows a structured methodology for data preprocessing, model selection, training, and evaluation. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.

**3.2.1 Data Preprocessing**

The raw stock market data undergoes various preprocessing steps to ensure it is clean and suitable for analysis. These steps include:

1. Handling Missing Values: Removing or imputing missing entries in the dataset. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
2. Normalization and Scaling: Applying Min-Max Scaling or Standard Scaling to standardize input features. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
3. Feature Engineering: Introducing additional relevant features such as moving averages (SMA, EMA), trading volume trends, and sentiment analysis from financial news. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.

**3.2.2 Model Selection and Training**

Multiple AI models will be implemented and evaluated, including:

* Traditional Machine Learning Models: Decision Trees, Random Forest, Support Vector Machines (SVM). This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* Deep Learning Models: LSTM (Long Short-Term Memory), GRU (Gated Recurrent Units), Transformer-based architectures. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* Hybrid Models: Combining statistical approaches like ARIMA with deep learning to leverage the strengths of both. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* Other Advanced Models: Exploring Graph Neural Networks (GNNs) and Reinforcement Learning models for optimizing trading strategies. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.

Each model will be trained using 80% of the dataset, while the remaining 20% will be reserved for testing. The training will involve hyperparameter tuning using techniques such as Grid Search and Bayesian Optimization to optimize performance. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.

**3.2.3 Performance Evaluation Metrics**

The models will be assessed based on various metrics, including:

* Root Mean Square Error (RMSE): Measures prediction accuracy by quantifying errors in continuous values. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* Mean Absolute Percentage Error (MAPE): Evaluates the percentage deviation between predicted and actual stock prices. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* R-Squared Score: Determines the proportion of variance explained by the model. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
* Computational Efficiency: Assesses model execution time and memory consumption. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.

**3.3 System Architecture Overview**

The system is designed to ensure a seamless pipeline from data collection to real-time stock prediction. It consists of the following components:

1. Data Ingestion Module: Fetches stock market data from Yahoo Finance API and processes it for analysis. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
2. Preprocessing Engine: Performs data cleaning, normalization, and feature extraction. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
3. Model Training Pipeline: Trains and fine-tunes AI models for optimal predictive performance. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
4. Prediction Service: Generates real-time stock price predictions and trend forecasts. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.
5. User Interface: Provides an interactive dashboard to visualize stock trends and model performance. This point plays a critical role in the development of the project and will be explored in detail throughout the upcoming sections.

