**Abstract**

**INTRODUCTION**

Generally, there are two prevailing methods for addressing various phenomena in financial markets: the Efficient Market Hypothesis (EMH) and the Adaptive Market Hypothesis (AMH) **(Fama, 1991)**. According to the first method, market prices react to various factors such as economic and social news. Since these news items stem from diverse phenomena that are unpredictable, it suggests that financial markets cannot be forecasted using any traditional method. The second method **(Lo, 2004)** seeks to identify potential causal relationships between different social phenomena and market developments. In this paper, our aim is to establish an effective relationship for the probabilistic prediction of the market. Furthermore, our emphasis will be on technical analysis as a key tool in this endeavor.

Creating reliable models of the stock market enables investors to make more informed choices, potentially reducing the risks associated with investments. A trading model may lower the risks linked to investing and allow traders to select companies that offer the highest dividend returns. However, due to a high degree of correlation among stock prices, analyzing the stock market with batch processing approaches becomes more challenging. The prediction of stock market values has entered a new era of advanced technology with the advent of global digitalization. Consequently, artificial intelligence models have gained significant importance due to the continuous increase in market capitalization. The novelty of the proposed study lies in the development of a robust time-series model based on deep learning for predicting future market values. As the industry and technology grow and more leading companies emerge in this field, financial markets have expanded significantly. Thus, it is essential to manage these new financial resources with an influx of capital from investors as well as the public. Prominent economists have always endeavored to predict financial markets in various ways to maximize profits. For this purpose, they have utilized various methods such as statistical, economic, and mathematical models, and many techniques have been invented to predict market trends. These methods include:

1. Technical Analysis
2. Fundamental Analysis

In this paper, our aim is to establish an effective relationship for the probabilistic prediction of the market. Furthermore, our emphasis will be on technical analysis as a key tool in this endeavor.

In Fundamental analysis, we aim to determine the intrinsic value of a symbol/company using various factors such as the company’s financial report, microeconomic indicators, and customer behavior psychology. It is evident that if the estimated value is higher than the current value, the symbol is suitable for entry, and vice versa, it is suitable for exit. An example of this method could be analyzing investor sentiment in the market. To achieve this, methods such as neural networks and machine learning can be employed **(Zheng et al., 2024)**.

In technical analysis, we focus solely on examining the past price movements of a symbol/company on the chart. As we know, specific patterns may appear on the time chart of different symbols, which can be used to derive conclusions for buying/selling the symbol. These patterns can be represented in various forms. Among the most popular chart types for market analysts are the line chart and the candlestick chart. Therefore, pattern recognition in different charts is of great importance for this section.

To diagnose patterns and predict upward and downward movements, artificial intelligence-based methods can generally be used. These methods include machine learning, neural networks, deep learning, etc. For instance, various well-known patterns in technical analysis can be classified using different machine learning methods and neural networks.

The stock market is a vital component of the financial market, playing a crucial role in wealth accumulation for investors, providing financing for listed companies, and fostering sustainable macroeconomic development. Significant fluctuations in the stock market can harm stock investors' profits and cause imbalances in the industrial structure, potentially disrupting the overall development of the national economy. Predicting stock price trends is a common research topic in academia. Predicting three trends—upward, sideways, and downward—can assist investors in making informed decisions regarding the buying, holding, or selling of stocks. Developing an effective prediction model for these trends is of substantial practical importance. Therefore, this research will present a new and efficient method for classifying stock market patterns.

The central part of a candlestick is known as the body. This section indicates whether the closing price of the stock was higher or lower than the opening price. Many stock market analysis software programs use colors to indicate price movements, with green and red commonly used to denote increases and decreases in price, respectively. Specifically, green signifies price growth and money entering the stock, while red indicates money leaving the stock and a decrease in price. In other words, if the closing price is higher than the opening price, the body of the candlestick is green. If the closing price is lower than the opening price, the body is red.

**Literature review**

For candlestick pattern detection, Convolutional Neural Networks (CNNs) can effectively identify complex candlestick patterns. The results showed that CNNs improved the prediction accuracy compared to traditional models **(Kim & Lee, 2021)**.

To understand sequential dependencies in candlestick data, an LSTM-based model was used **(Chen et al., 2021)**. This model was effective in predicting stock price trends and demonstrated good performance in distinguishing between different candlestick patterns.

**Methodology**

The most popular charts used in financial markets, such as the stock market, include: candlestick charts, bar charts, and line charts **(Pal and Kar, 2019)**.

Candlestick charts display the same information as bar charts, but in a more visually appealing and graphical format. Candlestick charts also show the high to low range with a vertical line. In each candlestick, the large block in the center represents the range between the opening and closing prices.

The patterns formed by candlesticks are divided into three categories based on the number of candles involved. These include single candlestick patterns, two candlestick patterns, and three candlestick patterns.

To familiarize oneself with candlestick patterns in technical analysis, 18 important examples from various candlestick charts have been examined. To define these patterns in a rule-based manner, two sources have been utilized. One of them was published in 1991 **(Nison, 1991)** and the other in 2008 **(Bulkowski, 2008)**.

Some of these patterns include the engulfing, doji, morning star, hammer, shooting star, three white soldiers, and etc. Additionally, patterns that do not follow these identified types are labeled as 'unknown'.

**Random Forest** is a powerful ensemble learning method that uses multiple decision trees to enhance prediction accuracy and robustness. During training, each tree is constructed using a random subset of data (both features and samples), and these trees are built through a non-overlapping random sampling process. This helps to prevent overfitting, which is common in individual decision trees. During prediction, the model aggregates the outputs of the individual trees using either averaging or majority voting to determine the final output. By combining predictions from different trees, Random Forest effectively reduces variance and minimizes overfitting, making it capable of handling high-dimensional datasets.

In Random Forest, the Bootstrap Aggregation (Bagging) method is used, where multiple decision trees are trained in parallel using random samples of data. Each tree is trained independently. After training all the trees, the model uses averaging for regression tasks and majority voting for classification tasks to make the final prediction. This method is suitable for both classification and regression tasks, and it performs well even when dealing with missing data or datasets with a large number of features **(Breiman, 2001)**.



**Fig.2. Bagging architecture**

**XGBoost** is a very powerful gradient boosting algorithm suitable for structured and tabular data. The algorithm uses gradient boosting to create an ensemble of weak learners, typically decision trees, to make accurate predictions. The key idea behind XGBoost is to sequentially add weak models and correct the errors made by previous models, a process known as "boosting."

This algorithm starts by calculating the residuals (the difference between actual values and predicted values). Then, a new weak model (another decision tree) is trained to predict these residuals. The goal of this new model is to reduce the errors of the initial predictions. By adding this new model to the original model, the predictions gradually become closer to the actual values.

XGBoost optimizes predictions using an objective function that consists of two parts: one is the error function for measuring the alignment of predictions with actual values, and the other includes regularization terms to control model complexity and prevent overfitting. This algorithm also uses techniques like tree pruning and depth control to prevent overfitting. Additionally, XGBoost supports parallel processing, which helps increase the speed of model training **(Chen & Guestrin, 2016)**.

**LightGBM** works based on a gradient boosting framework, similar to XGBoost, and uses decision trees as weak learners. However, it is designed to be very fast and efficient for large datasets. One of the key features of LightGBM is the use of a histogram-based approach for building trees. Instead of examining all possible split points, the data is discretized into bins, and a histogram is created for each feature. This method significantly reduces computation and memory usage, making LightGBM highly efficient for large datasets.

LightGBM uses a leaf-wise growth strategy for building decision trees. Unlike traditional methods that expand trees level by level, it expands the leaves that result in the greatest reduction in the error function. This approach makes the trees deeper and more focused, which can improve accuracy. However, if not controlled properly, it may lead to overfitting.

LightGBM also employs methods like gradient-based one-side sampling (GOSS), which focuses more on samples with high gradients instead of using the entire dataset. This reduces training time without compromising model quality. Additionally, it uses feature bundling techniques to reduce the number of features, thereby enhancing efficiency.

LightGBM has specific capabilities to handle missing data and automatically learns an appropriate strategy for splits that involve missing values. This makes it particularly effective for large datasets and high-dimensional data, allowing it to operate more efficiently and quickly compared to XGBoost (**Meng and et al, 2017**).

**Results**

**Future Works**

**Conclusion**

**References**

1. Fama, E. F. (1991). "Efficient capital markets: II." The journal of finance **46**(5): 1575-1617.
2. Lo, A. W. (2004). "The adaptive markets hypothesis: Market efficiency from an evolutionary perspective." Journal of Portfolio Management, Forthcoming.
3. Zheng, J., D. Xin, Q. Cheng, M. Tian and L. Yang (2024). "The Random Forest Model for Analyzing and Forecasting the US Stock Market in the Context of Smart Finance." arXiv preprint arXiv:2402.17194.
4. Kim, H., & Lee, S. (2021). Candlestick Pattern Recognition Using Convolutional Neural Networks for Stock Trading Strategies. IEEE Access, 9, 126798-126808.
5. Chen, Z., Huang, Q., & Zhang, Y. (2021). Stock Price Prediction Using Candlestick Chart Patterns and LSTM Neural Networks. Journal of Computational Finance, 25(2), 123-138.
6. Pal, S. S. and S. Kar (2019). "Time series forecasting for stock market prediction through data discretization by fuzzistics and rule generation by rough set theory." Mathematics and Computers in Simulation 162: 18-30.
7. Nison, S. (1991). Japanese Candlestick Charting Techniques. New York Institute of Finance.
8. Bulkowski, T. (2008). Encyclopedia of Candlestick Charts. Wiley Trading.
9. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32
10. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
11. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017), LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS).