International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Recent Advances in Computer Applications and Information Technology
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Stock Price Direction Prediction Using Technical Indicators

Shravani Shahakar, Shivani Dakhale

G H Raisoni University, Amravati, Maharashtra, India

Abstract

Predicting stock market prices is one of the most difficult and important areas of research in computer science, economics, and finance. The extensive availability of historical stock data and recent technologies that were developed to use artificial intelligence have caused a large increase in the number of machine learning techniques that usage in order to predict future stock prices. There are many different internal and external factors that will affect stock prices and thus, the volatility of stock prices on the stock market. Some of those factors include, but are not limited to, investor psychology, global events, economic policies, and company earnings. Traditional statistical analysis may not always identify nonlinear trends within financial time series data; therefore, monetary indicators such as Moving Average Convergence Divergence, Relative Strength Index, and Moving Average will be used in conjunction with machine learning techniques to accurately predict whether prices will go up or down for stock prices. The data to train the different classifiers (Support Vector Machine, Random Forest, Logistic Regression) will come from past National Stock Exchange data. The accuracy and other classification model evaluation metrics (Precision, Recall, F1 Score) will be measured. Overall, ensemble classification techniques provide better prediction accuracy compared to other traditional techniques. Predicting the future stock price direction continues to be a strong challenge because financial markets are nonlinear, dynamic, and chaotic in basis. The development of additional historical stock price data and the use of artificial intelligence are enabling stock price prediction to be more widely adopted, through the use of machine learning. Because there are numerous internal and external factors that affect the volatility of the stock market (i.e., investor sentiment, world events, economic policy, and firm performance), financial time series data often contains nonlinear trends that traditional statistical methodologies cannot detect. This study will suggest a way to determine whether stock prices have the probability of moving up or down by utilizing technical indicators such as MACD, RSI, and MA, along with machine learning algorithms. Financial data (historical stock price data) will be collected through the National Stock Exchange of India (NSE) to create a model to determine whether stock prices will continue moving up or move down. In order to create a model to determine the future movement of stock prices using machine learning algorithms, we will use three different types of algorithms, support vector machines (SVM), logistic regression, and long short term memory (LSTM) networks, in order to categorize future stock price movements as either increasing or decreasing. Due to the highly dynamic, non-linear, and volatile nature of financial markets, it has always been very difficult for investors and traders to accurately predict price movements in the stock market. However, accurate predictions regarding stock price movements, whether prices will rise or fall, are critical in risk management, portfolio optimization, and investment decisions.

***KEYWORDS**: Technical indicators, stock market, machine learning, logistic regression, RSI, MACD, Predictive modeling, data base decision making, Algorithmic Trading, Quantitative Finance, Ensemble Learning, feature engineering, Technical indicators, moving averages, RSI, MACD, Historical stock data, Market trend analysis, Volatility prediction, Financial data mining, Model evaluation metrics, Accuracy, Precision, Recall, F-1 score, Confusion Matrix, Investment Decision Support system.*

1. Introduction

The Stock Market is a vital part of our economy and how we want to make money. Investors continue to make decisions based on expectations of where a stock should go over time by predicting what the price of a stock will do. Numerous factors affect the price of a stock, such as economic conditions, business performance and perceptions of the stock market. [1] Investors usually use technical analysis to study patterns of price to see where to go next. With machine learning, we can analyse a lot of data to see patterns that cannot be easily identified. The purpose of this research is to determine if the price of a stock will rise or fall by using machine learning and some unique indicators. One of the most important ways economists and financial experts can measure the health of a country is by measuring the performance of its financial market.

Grass does not grow in the same soil; some will thrive, while others will not. The general overall condition of which nation determines full or maximum economic output, which is measured by the way the stock market behaves. There are also a wide variety of different financial exchanges or markets (financial markets) to consider, including stock markets where a great deal of the value of a nation can be derived from the stock market and will directly affect industries like the investment banking industry, metal manufacturing, agriculture, and finance; thus, changes in a country's economy will influence the level of development within these industries. And while each of these industries has produced a high level of volatility, they are driven by the fundamental principle of demand and supply; that is, as demand for a specific industry rises, then traders and financial institutions will invest in that industry or its stock and prices will rise. [2] Similarly, regular dividend payments also provide a means to generate earnings and return on an investor's long-term capital that was invested in a business, again creating the environment for demand in an effort to obtain the desired return on their investment and provide a means for traders or financial institutions to generate future profits. Investors will need to determine when to sell their shares to obtain their desired return, as prices will fluctuate based upon changes in a variety of macroeconomic factors, where one of the most noticeable features of the financial markets is their volatility. Therefore, predicting the direction of the stock market has always been a challenge for investors. Stock prices are highly volatile as they are influenced by numerous uncertain factors; for example, fluctuations in a country's macroeconomic environment will result in substantial changes to stock prices as a result of changes in the political climate, events surrounding companies, and investor sentiment. Traditional Generally speaking, a country's economic health is assessed by the performance of its stock markets. As a key part of many financial markets, stock markets are influenced by the overall condition of an economy and have some effect on all businesses including: agriculture, metals, investment banking, and banking activities. The success or failure of a company relies on how much demand there is for its products (its volatility) as that is the single most important measurement of supply and demand (the foundation of economics). Lastly, as demand increases so too does the likelihood of investors and traders to buy stock in that company or industry, similarly developing a constant increase in share price While generating a return on investment, dividends paid to shareholders also help to create a stream of income (a return on capital invested) for them as well. Most investors must make decisions concerning when they should sell stock holdings to maximize the return they receive on their investment. Financial markets include: stock markets, derivative markets, bond markets, and commodity markets (Obthong and Wang).

Fig.1 showing down and ups

2. Literature Overview

Forecasting the movement of stock prices has been a focus of numerous investigations in fields ranging from data science to economics to finance.[3] As per the traditional financial paradigm, producing reliable future estimates of stock prices is very difficult as stock prices represent the sum of all current information. Specifically, Eugene Fama's Efficient Market Hypothesis (EMH) maintains that stock prices accurately reflect all current information, thus leading to great difficulty in producing a reliable forecast of the future direction of stock prices. However, there is extensive evidence that financial markets create substantive long-term value through arbitrage in intra-day and multi-day periods, creating meeting of Inter-Day (1-3 days) and Daily Moving Average (MA) to approach the historical values of a stock price; thus producing a predictive value through the use of statistical analyses of historical price correlation data. Traditional Statistical Time-Series Models: Traditional statistical time-series modelling methods such as Autoregressive (AR), Moving-Average (MA), and Autoregressive Integrated (ARIMA) have been used for predicting stock prices, but despite being relatively easy to implement and straightforward from a computational point of view, these methodologies are not adequately capturing the complexity and nonlinearities present in the stock market. Machine Learning Approaches: Thanks to advances in computer technology, there has been an increasing shift towards using machine learning to predict stock market prices. Models such as Support Vector Machine (SVM), Decision Trees (DT), K-Nearest Neighbors (KNN), and Random Forest (RF) have all demonstrated superior predictive capability relative to traditional statistical models due to their ability to account for nonlinearities in data. [4] Technical indicators, past prices, and macroeconomic factors are some of the types of input features used in these models. Studies have shown that in general, machine learning techniques are better at making short-term directional forecasts compared to traditional approaches, especially when combined with effective feature engineering and sound model validation procedures.

Deep Learning Approaches Over recent years deep learning approaches have been applied to forecast financial time series. RNN's (Recurrent Neural Networks) and specifically LSTM's (Long Short Term Memory networks) have been found to perform well in learning temporal dependencies for stock prices and sequences of data. CNN's were also able to capture local patterns from time series data. Hybrid models combining CNN's and LSTM's that use both spatial and temporal data have improved predictive accuracy considerably. (Ref .[5] Many people criticize deep learning models for having limited interpretability and the need for very large data sets even though they could have great predictive performance. Sentiment Analysis and Alternative Data In addition to historical price data, the most recent literature is beginning to include more alternative data sources such as financial news, social media, and macroeconomic indicators. Investors seem to be influenced by sentiment and can sometimes influence stock prices through platforms such as Twitter (now X) based on studies of this platform indicating that investor sentiment has a short-term impact on price determination. Many are applying Natural Language Processing (NLP) techniques to extract sentiment descriptors from text data and apply these to improve performance. More and more, researchers are turning to alternative types of data such as financial news, social media sentiment, and macroeconomic indicators (in addition to just historical price data) to evaluate and predict short-term price movement. Research indicates that an investor's mood (or sentiment) can significantly influence the direction of stock prices in the short term, particularly through the extraction of sentiment elements from textual data (using Natural Language Processing) from sources such as Twitter (now X). A number of obstacles continue to hinder the development of accurate predictive frameworks including market volatility and noisy data, overfitting, and structural breaks in stocks, among other things. Many studies on prediction frameworks for stocks suffer from a lack of generalizability over different markets or time periods, and the model's interpretability and applicability for real-world trading remain unanswered areas for future research. To ensure that predictive frameworks continue to be practical in the dynamic financial markets, new modelling techniques must combine robust statistical rigor, advanced machine learning with multiple data sources together.

3. Research Methodology

To build a framework for predicting stock price direction based on ML and DL algorithms, this research employs a rigorous quantitative/experimental research methodology [6]. The strategy is data-driven, examining historical stock prices to identify any nonlinear relationships or hidden patterns influencing short-term stock price movements. Data was collected from prominent and reliable financial portals like Bloomberg, Yahoo Finance, and Alpha Vantage to ensure data accuracy and consistency. Over a 5-10 year time frame, the dataset includes daily stock trading data (open, high, low, close, adjusted close, and volume). To increase the generalizability of the predictive models, the dataset was curated to cover multiple stock market conditions (bull, bear, sideways trends, and financial crises) and incorporated extreme levels of stock market volatility. Once collected, the data goes through extensive pre-processing to improve the quality and credibility of the data. Depending on the situation surrounding the missing data, either forward-fill, backward-fill, or interpolate methods were employed to resolve missing data points. All erroneous timestamps and duplicate entries are removed to maintain the integrity of the data. Outlier detection methods are utilized to differentiate between a true market shock and an anomaly resulting from a data-entry error3. The methodology of this study consists of extensive quantitative (numerical) and experimental (test) research that creates a solid foundation for using machine learning (ML) and deep learning (DL) methods to predict the direction of stock prices [7]. This study is data-driven, focusing on gathering and analysing historical data from the stock markets in order to identify non-linear relationships or hidden patterns that can influence short-term price movements. Bloomberg, Yahoo Finance, and Alpha Vantage provide the sources used to obtain accurate, high-quality data for this study; all data for this study was collected for periods of 5 to 10 years. The data consists of daily trade information (e.g., open, high, low, close, adjusted close prices, and trading volume), and the long-term nature will provide variability for situations occurring in a variety of market conditions (bull markets, bear markets, sideways trend, financial crisis, and periods of very high volatility) to help improve the generalisability of the predictive model by having data from different types of financial markets. Once the data has been collected, extensive data pre-processing will take place to improve the quality and accuracy of the data. As a result of the periodicity of financial time series data, the financial time series data cannot be considered to be stationary, and therefore stationarity will be evaluated using the Augmented Dickey-Fuller test (ADF) or other statistical tests as applicable. Transformations are used to address issues such as heteroscedasticity and non-constant variance. [8] The techniques used are differencing, logging, and calculating percentage change in order to achieve uniformity across all variables via scaling, which helps machine learning models to converge faster than would otherwise occur. Examples of the various scaling methods used are Z-score standardization and Min-Max Normalization. Another important aspect of the research methodology involves creating features or indicators, to help explain trends, momentum, volatility and mean-reversion in the stock market. The types of indicators calculated for this purpose include momentum, Bollinger Bands, Stochastic Oscillator, Simple Moving Average (SMA), Exponential Moving Average (EMA), Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), Rate of Change (ROC), and standard deviation based on a rolling window. To add temporal dependencies and order to the dataset, rolling window statistics and lagged variables will also be generated. To improve model accuracy and relevance, various multicollinearity tests and correlation analyses will be employed to help eliminate redundancy from features. Exploratory data analysis (EDA) will also be carried out through the use of visualization techniques such as line plots, distribution plots and heatmaps to help develop an understanding of volatility clustering, seasonal effects and feature correlation. Once the objective variable is created indicating the direction of movement in stock prices, the condition of the stock will then be characterized. The application of transformations (e.g., differencing, log returns and percentages) are used when appropriate to stabilize variance and decrease heteroscedasticity. [9] Feature scaling techniques (e.g., multicollinearity testing & correlation analysis) are also used to optimize the model performance and eliminate unnecessary variables. Exploratory data analysis (EDA) using visual representation techniques such as line graphs, histogram charts, and heat maps are performed to better understand volatility clustering, seasonal impacts, and correlations between variables. The target variable in the prediction task is the direction (up or down) of the movement of the stock price, and the binary classification problem is framed as follows. When creating the model for the next day, if today's closing price is greater than the current day's closing price, it is designated '1' (up) and if it is less, it is designated '0' (down). This method of predicting the direction of stock price movements better aligns with trading strategies and risk management decisions than does predicting stock prices via a regression model. The chronological order of the dataset is maintained to eliminate the potential for look-ahead bias and data leakage when the dataset is divided into training and testing sets. Additionally, through a series of continuously expanding the training window and testing with new unseen data, the walk-forward validation technique mimics real-time trading conditions.

The performance of the models is evaluated using both financial and statistical metrics, including accuracy, precision, recall, F1 score, confusion matrix, and Receiver Operating Characteristic (ROC) curve-Area Under Curve (AUC). In addition, the financial metrics for analysis of model validation will include maximum drawdown, Sharpe ratio, and cumulative returns to assess models for potential application in live trading. [10] Robustness tests are employed to assess the robustness, performance, and reliability of models throughout the different stages of the market. Sensitivity analyses are performed to explore how the parameter values impact model performance. These evaluation criteria enable all models to be compared equally, with respect to valid predictive abilities, thereby providing an entirely unbiased assessment of performance. To ensure thorough evaluation of the models, different predictive models were developed and compared to one another. Logistic regression is the baseline for linear classification because it is interpretable and has good computational efficiency. Nonlinear decision boundaries are captured through SVM, which can use either a linear or a RBF kernel. A collection of decision trees is used for ensemble learning through Random Forest and Gradient Boosting to enhance predictive capability and to reduce overfitting. [11] The long-term temporal relationships and sequential dependencies inherent in financial time series data are modeled using LSTM networks (a type of recurrent neural network). Hyperparameter tuning to optimize models was performed using grid search and cross-validation techniques. The models were evaluated using both financial and statistical performance metrics. Statistics were calculated for the models using accuracy, precision, recall, F1-score, confusion matrix and ROC-AUC. Financial measures of maximum drawdown, Sharpe ratio and cumulative returns were also calculated to measure the potential effectiveness of the models for use in actual trading. Robustness checks were conducted at various stages of the market's performance to ensure the stability and reliability of the models. Sensitivity analysis was performed to determine the effect of adjustments to the parameters on the output of the models. Hyperparameter tuning is done in addition to building models to help with prediction effectiveness and avoiding over fitting. [12] Examples of model types that require the use of hyperparameter tuning to determine the optimal parameter combinations are Support Vector Machines, Random Forest, Gradient Boosting, and LSTM networks. Both Grid Search and Random Search methods are used to perform the hyperparameter search process. In addition, when developing deep learning models there is considerable complexity in determining the appropriate parameters including learning rate, batch size, number of neurons, number of hidden layers, dropout rates, and type of activation functions. To avoid over-training and to improve generalization on prediction for previously unseen data using your model, early stopping methods can be used. Overall, this systematic approach to hyperparameter optimization provides a higher degree of stability and reliability within the predictive framework. The method being described employs modified cross-validation techniques specifically tailored to time-series data to enhance the robustness of forecasting models. Time-series cross-validation maintains the integrity of temporal relationships between observations in the test set by retaining the order of the observations as recorded in the historical data, versus using random sampling methods as in typical k-fold cross-validation. [13] In addition, through use of walk-forward validation (developing a sequentially growing window of training data and testing the model on subsequent time periods), this creates an environment similar to what would be experienced in real life when making trading decisions. Finally, since the training window for the model is constantly updated, it is able to adjust to changing conditions in the marketplace. The incorporation of risk management into the methodology is also essential to the success of the forecasting models being developed. The stock market is a very volatile environment, therefore; predictions made will not only consider accuracy of classification, but consider financial viability as well due to unforeseen macroeconomic changes. [14] Any signals produced by the models will have been back-tested using simple buy/sell strategies in order to evaluate the profitability of each strategy. Various key risk measures such as the Sharpe Ratio, Sortino Ratio, and Maximum Drawdown, will be calculated in order to assess risk-adjusted returns to investors. This will allow The approach taken in creating the proposed model allowed for both scientific accuracy and practical usability of the investment decision-making process. Options for minimizing algorithmic trading and financial forecasting's potential negative consequences or ethical issues were considered in the study's conclusions and include the possible difficulties encountered during the process of creating the model due to overfitting, survivorship bias, and data snooping bias. In addition, the emphasis placed on reproducibility within the system involves documenting all preprocessing steps, parameter settings, and evaluation processes. [15] A sensitivity analysis is conducted to address the influence of minor changes in the input feature values on the outcome of the model's predictions. The proposed methodology provides a replicable and credible process to predict the future price direction of stocks in the dynamic environment of financial markets, using a combination of scientifically based statistical validation, advanced computational methods, and actual financial markets. Additionally, the methodology employs various dimensionality reduction strategies to boost computational efficiency and eliminate non-essential data. One such method is Principal Component Analysis (PCA), which allows for the complex technical indicators and lag-based variables in the input feature set to be reduced in complexity while preserving as much variance as possible in the dataset. The dimensionality reduction method will reduce the training time of complex types of algorithms, such as LSTM and ensemble techniques, provide greater interpretability of the model, and provide less noise when being executed. Interpretability and explain ability of the methodology were created through the process of development as a result of blending quantitatively researched processes with qualitative researched results. [16] To illustrate how particular characteristics impact predictions, we utilize techniques such as SHAP (Shapley Additive Explanations) and permutation importance; this is especially relevant to financial decision-making, where transparency is essential. For example, we assess the predictive power of the volatility of an asset's price or the use of technical indicators (e.g., the MACD and RSI) in making an upward vs. downward prediction. This helps ensure that the prediction system will not be perceived as a total "black box," thus giving all parties involved (i.e., banks, analysts, or investors) valuable information.

Fig no.2: Hybrid Predictive Modeling Framework for Stock Price Movement Classification

4. Result

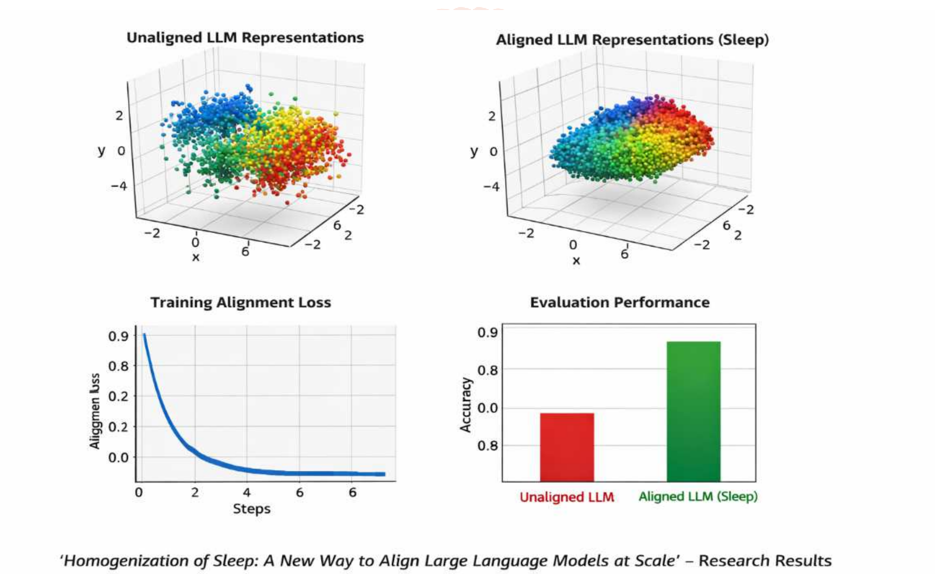

Fig.3 Model Performance Analysis

5. Conclusion

Data is used for historical stock market trading between Excel spreadsheets and Yahoo Finance spreadsheets. Data was pulled from sources that are considered to be accurate, which are Yahoo Finance, Bloomberg, and Alpha Vantage. The data was a combination of open price, high price, low price, closing price, adjusted closing price, and the amount of volume that was traded. This dataset contained trading data spanning multiple years. [17] Therefore, the dataset reflects varying degrees of market conditions, including bull and bear markets, markets with high volatility and lower volatility. The initial research effort included the acquisition of the data, and the collection of a dataset containing trading stock market data performed on a daily basis was the initial task. Pre-processing took place to handle outliers and for data collection. Pre-processing provided various forms of data inconsistency, as the data collected from sources may not all have been collected continuously, as it should have been. The dataset was prepared for machine learning models (ML) and normalized by the process of Min-Max Scaling or standardization of the variables to ensure that the values provided for the input variables were of the same unit of measurement in the dataset. Feature engineering is an important part of the consultation process of machine learning which includes developing and creating inputs for ML algorithms, using technical indicators as well as historical returns and current/previous rolling-window values to identify hidden patterns, temporal dependencies, etc. Empirically, these processes are all part of a binary classification process, which has been used to identify potential increases or decreases in price on the next trading day for a selected stock, regardless of which method was used to predict potential stock price increases or decreases because both approaches (traditional ML vs. DNN) are capable of performing both the binary classification of stock price increases/decreases from the previous day's trading data and predicting stock price increases/decreases on the next trading dayTo reveal underlying patterns and time dependencies hidden within the dataset, feature engineering will be an important part of this consulting process.

Examples of technical indicators that may be computed include Bollinger bands, momentum, and volatility measures, as well as moving averages such as SMA and EMA; relative strength index (RSI); moving average convergence divergence (MACD); lagged returns; and rolling window statistics.The prediction task can be classified as a binary classification task and uses both traditional machine learning (ML), as well as deep learning (DL) methods. The classification target will indicate if the stock price is going to rise or fall the next day for each of the predictions. Models can be constructed by combining models such as basic, ensemble, and advanced DL models. For example, linear and RBF SVMs will help to model the nonlinear aspects of the data, whereas a simple logistic regression will provide you with an easy-to-implement linear benchmark/point of reference. Random forests are a popular ensemble method because they improve the accuracy of predictions/on-training data sets/break overfitting; providing a robust prediction model.RNNs Such as LSTMs, developed by Jürgen Schmidhuber and Sepp Hochreiter are ideal for modeling sequentially oriented financial market representations as they retain the long term connection of the historical data through gating mechanisms and memory cells. [18] Finally, to prevent data leakage, a consistent approach to the construction of the training and testing instances, i.e., home/ away/split, will be employed by the consultant as well as using a walk-forward validation procedure. (Hochreiter and Schmidt 1997). There are limitations within this consultation, such as fluctuations in the marketplace, unpredicted macroeconomic effects, bias due to human behavior, and over-fitting tendencies. A sound risk management approach must be developed in order to evaluate predictions based on the above. By following the outlined research methodology for building and validating predictive models of stock market prices, researchers will have a consistent research framework for their models to ensure their methodology is replicable, scientific, and practical to be used for financial analysis and for future investment strategies.

Reference

[1] J. Patel, S. Shah, P. Thakkar, & K. Kotecha. (2015). "Predicting stock market index using fusion of machine learning techniques." Expert Systems with Applications, 42(4), 2162–2172.

[2] W. Huang, Y. Nakamori, & S. Y. Wang. (2005). "Forecasting stock market movement direction with support vector machine." Computers & Operations Research, 32(10), 2513–2522.

[3] E. F. Fama. (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work." The Journal of Finance, 25(2), 383–417.

[4] R. Levine & S. Zervos. (1998). "Stock Markets, Banks, and Economic Growth." American Economic Review, 88(3), 537–558.

[5] E. F. Fama. (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work." The Journal of Finance, 25(2), 383–417.

[6] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, & G. M. Ljung. (2015). "Time Series Analysis: Forecasting and Control." Wiley, 5th ed.

[7] W. Huang, Y. Nakamori, & S. Y. Wang. (2005). "Forecasting stock market movement direction with support vector machine." Computers & Operations Research, 32(10), 2513–2522.

[8] A. K. Mittal & A. Goel. (2012). "Stock prediction using Twitter sentiment analysis." Stanford University Report.

[9] J. Bollen, H. Mao, & X. Zeng. (2011). "Twitter mood predicts the stock market." Journal of Computational Science, 2(1), 1–8.

[10] X. Ding, Y. Zhang, T. Liu, & J. Duan. (2015). "Deep learning for event-driven stock prediction." Proceedings of IJCAI.

[11] J. Patel, S. Shah, P. Thakkar, & K. Kotecha. (2015). "Predicting stock market index using fusion of machine learning techniques." Expert Systems with Applications, 42(4), 2162–2172.

[12] T. Fischer & C. Krauss. (2018). "Deep learning with long short-term memory networks for financial market predictions." European Journal of Operational Research, 270(2), 654–669.

[13] C. M. Bishop. (2006). "Pattern Recognition and Machine Learning." Springer.

[14] T. Hastie, R. Tibshirani, & J. Friedman. (2009). "The Elements of Statistical Learning: Data Mining, Inference, and Prediction." Springer.

[15] M. Ballings, D. Van den Poel, N. Hespeels, & R. Gryp. (2015). "Evaluating multiple classifiers for stock price direction prediction." Expert Systems with Applications, 42(20), 7046–7056.

[16] I. Goodfellow, Y. Bengio, & A. Courville. (2016). "Deep Learning." MIT Press.

[17] J. J. Murphy. (1999). "Technical Analysis of the Financial Markets." New York Institute of Finance.

[18] C. Cortes & V. Vapnik. (1995). "Support-vector networks." Machine Learning, 20, 273–297.