International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 03 | March - 2021 ISSN: 2582-3930

LIVE STREAMING DATA PREDICTION USING MACHINE LEARNING

S. Amudha*1, D. Dhanalakshmi*2, J. Jayashankari*3, Dr. P.Veeralakshmi*4

(1,2)Student- Prince Shri Venkateshwara Padmavathy Engineering College.

(3,4)Faculty- Prince Shri Venkateshwara Padmavathy Engineering College.

Abstract: In the past decades, there is an increasing interest in predicting markets among economists, policymakers, academics and market makers. The objective of the proposed work is to study and improve the supervised learning algorithms to predict the stock price. Stock Market Analysis of stocks using data mining will be useful for new investors to invest in stock market based on the various factors considered by the software. Stock market includes daily activities like Senses calculation, exchange of shares. The exchange provides an efficient and transparent market for trading in equity, debt instruments and derivatives. Our aim is to create software that analyses previous stock data of certain companies, with help of certain parameters that affect stock value. We are going to implement these values in data mining algorithms and we will be able to decide which algorithm gives the best result. This will also help us to determine the values that particular stock will have in near future. We will determine the patterns in data with help of machine learning algorithms.

Introduction:

Stock market trend prediction's is raising day-to-day in our society. As of now stock market is very uncertain and unstable and is difficult to predict exactly. There can be high risk in predicting the approximate values of the stock market. Most of the investors aware of the raise and fall of the stock price values. Based on the prediction ability, investors can able to make profit over the product. For these prediction many technologies like machine learning, Data Mining, Feature learning are mainly used by researchers. Machine learning is most tgcommonly used technique for predictions.

MACHINE LEARNING – OVERVIEW

Machine learning is a very hot topic for many key reasons, and because it provides the ability to automatically obtain deep insights, recognize unknown patterns, and create high performing predictive models from data, all without requiring explicit programming instructions. This high level understanding is critical if ever involved in a decision-making process surrounding the usage of machine learning, how it can help achieve business and project goals, which machine learning techniques to use, potential pitfalls, and how to interpret the results.

Machine learning is the application of artificial intelligence and based on the idea of the system that will learn data with less human intervention.

Supervised Learning

Supervised algorithms are trained using labeled examples, in different scenarios, as an input where the desired outcome is already known. An equipment, for instance, could have data points such as "F" and "R" where "F" represents "failed" and "R" represents "runs". A learning algorithm will receive a set of input instructions along with the corresponding accurate outcomes. The learning algorithm will then compare the actual outcome with the accurate outcome and flag an error, if there is any discrepancy. Using different methods, such as regression, classification, gradient boosting, and prediction, supervised learning uses different patterns to proactively predict the values of a label on extra unlabeled data. This method is commonly used in areas where historical data is used to predict events that are likely to occur in the future. For instance, anticipate when a credit card transaction is likely to be fraudulent or predict which insurance customers are likely to file their claims.

Unsupervised Learning

Unsupervised Learning in Machine Learning finds its application in areas were data has no historical labels. Here, the system will not be provided with the "right answer" and the algorithm should identify what is being shown. The main aim here is to analyze the data and identify a pattern and structure within the available data set. Transactional data serves as a good source of data set for unsupervised learning. For instance, this type of learning identifies customer segments with similar attributes and then lets the business to treat them similarly in marketing campaigns. Similarly, it can also identify attributes that differentiate customer segments from one another. Either ways, it is about identifying a similar structure in the available data set.

Reinforcement Learning

Reinforcement Learning used in navigation, robotics and gaming. Actions that yield the best rewards are identified by algorithms that use trial and error methods. There are three major components in reinforcement learning, namely, the agent, the actions and the environment. The agent in this case is the decision maker, the actions are what an agent does, and the environment is anything that an agent interacts with. The main aim in this kind of learning is to select the actions that maximize the reward, within a specified time.

MACHINE LEARNING APPLICATIONS:

STATISTICAL ARBITRAGE

In finance, statistical arbitrage refers to automated trading strategies that are typical of a short-term and involve a large number of securities. In such strategies, the user tries to implement a trading algorithm for a set of securities on the basis of quantities such as historical correlations and general economic variables. These measurements can be cast as a classification or estimation problem. The basic assumption is that prices will move towards a historical average.

MEDICAL DIAGNOSIS

ML provides methods, techniques, and tools that can help in solving diagnostic and prognostic problems in a variety of medical domains. It is being used for the analysis of the importance of clinical parameters and of their combinations for prognosis, e.g. prediction of disease progression, for the extraction of medical knowledge for outcomes research, for therapy planning and support, and for overall patient management.

RELATED WORK:

Efficacy of News Sentiment for Stock Market Prediction- Stock Market trend prediction will always remain a challenging task due to stochastic nature. The enormous amount of data generated by the news, blogs, reviews, financial reports and social media are considered a treasure of knowledge for researchers and investors [1]. Stock Market Prediction based on Social Sentiments using Machine Learning- Machine learning and artificial intelligence techniques are being used in conjunction with data mining to solve a plethora of real world problems. These techniques have proven to be highly effective, yielding maximum accuracy with minimal monetary investment and also saving huge amounts of time. To add to their annual income, nowadays, people have started looking at stock investments as a lucrative option. With expert guidance and intelligent planning, we can almost double our annual revenue through stock returns [2]. Predicting the Effects of News Sentiments on the Stock Market- Stock market forecasting is very important in the planning of business activities. Stock price prediction has attracted many researchers in multiple disciplines including computer science, statistics, economics, finance, and operations research. Recent studies have shown that the vast amount of online information in the public domain such as Wikipedia usage pattern, news stories from the mainstream media, and social media discussions can have an observable effect on investors' opinions towards financial markets [3]. To predict the direction of US stock prices by integrating time-varying effective transfer entropy (ETE) and various machine learning algorithms. At first, we explore that the ETE based on 3 and 6 months moving windows can be regarded as the market explanatory variable by analyzing the association between the financial crises and Granger-causal relationships among the stocks. Then, we discover that the prediction performance on the stock price direction can be improved when the ETE driven variable is integrated as a new feature in the logistic regression, multilayer perceptron, random forest, XGBoost, and long short-term memory network. Meanwhile, we suggest utilizing the adjusted accuracy derived from the risk-adjusted return in finance as a prediction performance measure. Lastly, we confirm that the multilayer perceptron and long short-term memory network are more suitable for stock price prediction [5].In this paper, empirically investigating the predictability of stock market movement direction using an enhanced method of sentiments analysis. Precisely, we experiment on stock prices history, sentiments polarity, subjectivity, N-grams, customized text-based features in addition to features lags that are used for a finer-grained analysis. Five research questions have been investigated towards answering issues associated with stock market movement prediction using sentiment analysis. We have collected and studied the stocks of ten influential companies belonging to different stock domains in NASDAQ. Our analysis approach is complemented by a sophisticated causality analysis, an algorithmic feature selection and a variety of machine learning techniques including regularized models stacking. A comparison of our

approach with other sentiment-based stock market prediction approaches including Deep learning, establishes that our proposed model is performing adequately and predicting stock movements with a higher accuracy of 60% [4].

PROBLEM DESCRIPTION:

In this proposed system, the stock price trend prediction is implemented with three techniques of machine learning namely Support Vector Machine(SVM), Linear Regression(LR) and Random Forest(RF). With the help of these algorithms the accuracy of the datasets are identified and the final result is evaluated. The architecture of the proposed system is shown in figure:1.1.

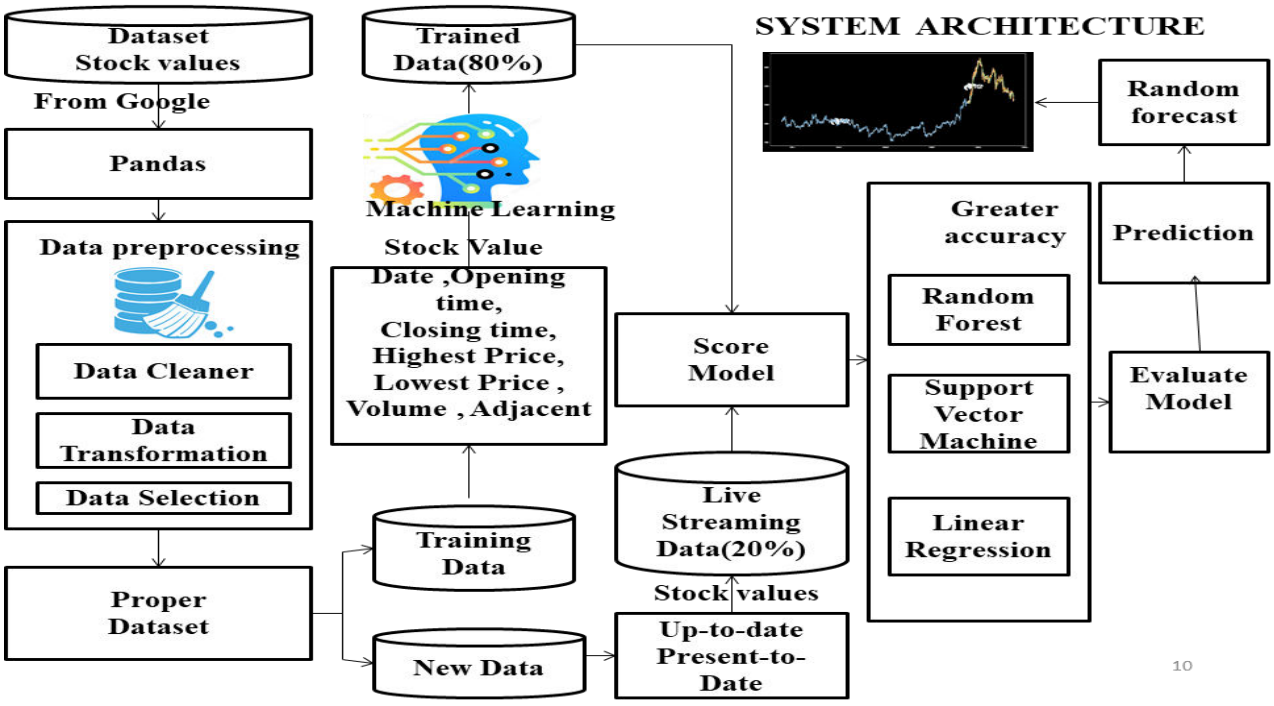

FIGURE: 1.1

In this architecture the datasets of any company is collected with the help of the unique key from the internet and those data's are read by using pandas (data reader). Then preprocessing process is done and the unwanted data, repetitions are removed. Accuracy of the algorithms are identified with the datasets. Each algorithm gives the different accuracy result. Based n that results the prediction of thee sock price is evaluated. And the random forecast is used to display the graph for showing the result after 30 days. At last the final graph is shown in the graph.

Methodology:

1. **Linear Regression (LR)**

Linear Regression algorithm in machine learning is a supervised learning algorithm. LR performs a regression task. Regression is widely used to find out the relation between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used.

Prediction in Linear regression takes place to predict the dependent variable (y), based on the independent variable(x). In this method regression technique finds out a linear relationship between independent variable(x), and dependent variable(y). The hypothesis function of the Linear Regression(LR) is shown .

$$y = \theta_1 + \theta_2.x$$

x: input
y: labels to data
θ_1: intercept
θ_2: coefficient of x

Cost Function (J):

The error difference between the predicted value and the true value is minimum, to predict y value, with this best fit regression line gets achieved. So, it is very important to update the θ_1 and θ_2 values, to reach the best value that minimize the error between predicted y value and true y value (y).

$$minimize \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

Cost function (J) of Linear Regression is the Root Mean Squared Error (RMSE) between predicted y value (pred) and true y value (y).

$$J = \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

2. Random Forest (RF)

Random Forest is one of the popular machine learning algorithm which belongs to the supervised learning technique. In ML, RF can be used for both Classification and Regression problems. A process of combining multiple classifiers to solve a complex problem and to improve the performance of the model is known as concept of ensemble learning. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." The random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output, instead of relying on one decision tree. The greater number of trees in the forest leads to higher accuracy and prevents the problem of over-fitting. Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase. The Working process can be explained in the below steps

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

3. Support Vector Machine (SVM)

The objective of the support vector machine algorithm is to find a hyper-plane in an N-dimensional space(N — the number of features) that distinctly classifies the data points. In machine learning the gradient descent is used to minimize said function, whereas in support vector machine Lagrangian is used. The Lagrange tells us to subtract the cost function by the summation over all the constraints where each of those constraints will be multiplied by some constant alpha (normally written as lambda for the Lagrangian).

$$L_p = \|w\|_2 2 + C(\sum_{i=1}^{N} \xi)_k - \sum_{i=1}^{N} \lambda_i (y_i (w \cdot x_i + b) - 1 + \xi_i) - \sum_{i=1}^{N} \mu_i \xi$$

Cost Function and Gradient Updates:

In the SVM algorithm, we are looking to maximize the margin between the data points and the hyper-plane. The loss function that helps maximize the margin is hinge loss.

$$c(x, y, f(x)) = \begin{cases} 0, & \text{if } y * f(x) \geq 1 \\ 1 - y * f(x), & \text{else} \end{cases}$$

Result:

Based on the accuracy of these algorithms the stock value prediction is evaluated in one algorithm which gives the greater accuracy. The accuracy of these algorithms are displayed and the graph is generated fr the prediction result of the collected datasets. The final graph is shown in figure: 1.2.

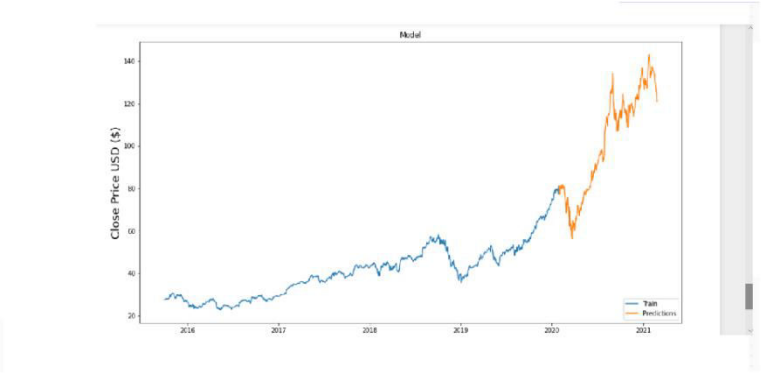

FIGURE: 1.2

CONCLUSION

By measuring the accuracy of the different algorithms, we found the most suitable algorithm for predicting the market price of a stock based on various data points from the historical data. The algorithm will be great asset for
brokers and investors for investing money in the stock market since it is trained on a huge collection of historical data and has been chosen after being tested on a sample data . The project demonstrates the machine learning model to predict the stock value with more accuracy as compared to previously implemented machine learning models.

FUTURE ENHANCEMENT

Future scope of this project will involve adding more parameters and factors like the financial ratios, multiple instances, etc. the more parameters are taken into account more will be the accuracy. The algorithms can also be applied for analyzing the contents of public comments and thus determine patterns or relationship between the customer and corporate employee. The use of traditional algorithms and determining techniques can also help predict the corporation's performance structure as a whole.

REFERENCES:

[1] Kalra, S., & Prasad, J. S. (2019). Efficacy of News Sentiment for Stock Market Prediction. 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). doi:10.1109/comitcon.2019.8862265

[2] Mankar, T., Hotchandani, T., Madhwani, M., Chidrawar, A., & Lifna, C. . (2018). Stock Market Prediction based on Social Sentiments using Machine Learning. 2018 International Conference on Smart City and Emerging Technology (ICSCET). doi:10.1109/icscet.2018.8537242

[3] Shah, D., Isah, H., & Zulkernine, F. (2018). Predicting the Effects of News Sentiments on the Stock Market. 2018 IEEE International Conference on Big Data (Big Data). doi:10.1109/bigdata.2018.8621884

[4] Bouktif, S., Fiaz, A., & Awad, M. (2020). Augmented Textual Features Based Stock Market Prediction. IEEE Access, 1–1. doi:10.1109/access.2020.2976725

[5] Kim, S., Ku, S., Chang, W., & Song, J. W. (2020). Predicting the direction of US stock prices using effective transfer entropy and machine learning techniques. IEEE Access, 1–1. doi:10.1109/access.2020.3002174

[6] Pesaran MH, Timmermann A. "Predictability of stock returns: Robustness and economic significance," The Journal of Finance, vol.50, no.4, pp.1201-1228, Sep. 1995.

[7] Ang A, Bekaert G. "Stock return predictability: Is it there?" The Review of Financial Studies, vol.20, no.3, pp.651-707, Jul. 2006.

[8] Rapach DE, Strauss JK, Zhou G. "International stock return predictability: what is the role of the United States?" The Journal of Finance, vol.68, no.4, pp.1633-1662, Aug. 2013.

[9] Qi M. "Nonlinear predictability of stock returns using financial and economic variables," Journal of Business & Economic Statistics, vol.17, no.4, pp.419-429, Oct. 1999.

[10] Aiolfi M, Favero CA. "Model uncertainty, thick modelling and the predictability of stock returns," Journal of Forecasting, vol.24, no.4, pp.233-254, Jul. 2005.

[11] Huang W, Nakamori Y, Wang SY. "Forecasting stock market movement direction with support vector machine," Computers & Operations Research, vol.32, no.10, pp.2513-2522, Oct. 2005.

[12] Atsalakis GS, Valavanis KP. "Forecasting stock market short-term trends using a neuro-fuzzy based methodology," Expert Systems with Applications, vol.36, no.7, pp.10696-10707, Sep. 2009.

[13] Schumaker R P, Chen H. "Textual analysis of stock market prediction using breaking financial news: The AZFin text system," ACM Transactions on Information Systems (TOIS), vol.27, no.2, pp.12, Feb. 2009.

[14] Adebiyi AA, Adewumi AO, Ayo CK. "Comparison of ARIMA and artificial neural networks models for stock price prediction," Journal of Applied Mathematics, 2014.

[15] Hinich MJ, Patterson DM. "Evidence of nonlinearity in daily stock returns," Journal of Business & Economic Statistics, vol.3, no.1, pp.69- 77, Jan. 1985.