

Stock Price Prediction System

Prof. Sulochana Sonkamble¹, Vaibhav Vyas², Prathamesh Shimpi³, Aniket Mule³, Mihir Sonawane⁴

¹Prof. Sulochana Sonkamble, Department of Computer Engineering, JSPM's Rajarshi Shahu School of Engineering and Research, Pune, Maharashtra, India

^{2,3,4,5} BE Students, Department of Computer Engineering, JSPM's Rajarshi Shahu School of Engineering and Research, Pune, Maharashtra, India

ARTICLE INFO

Article History:

Accepted: 01 April 2023

Published: 09 April 2023

Publication Issue

Volume 10, Issue 2

March-April-2023

Page Number

273-277

ABSTRACT

To determine the future stock value of a company is the main purpose of stock price prediction there is a continuous change in the price of stocks which is affected by different industries and market conditions. The high dimensionality of data is a challenge for machine learning models because highly correlated dimensions/attributes may exert influence on precision of the model. PCA is used to reduce dimensionality to fit linear regression algorithm for future stock price prediction. Linear regression algorithm is used prior to and subsequent to implementation of Principal Component Analysis on the Tesla stock price data. Results manifest that production of machine learning models can be boosted by PCA, reducing the correlation and appropriate selection of principal components for high redundancy of data. Root mean square value and R-square value is used for assessment. Keywords: Principal component analysis, Linear regression, Root mean square error, r square value.

Keywords: Principal component analysis, Linear regression, Root mean square error, r square value.

I. INTRODUCTION

The stock price prediction is one of the popular topics in financial field. Economic behaviors, social economy gets affected directly due to trends of the stocks. To erect a forecast model traditional stock method is used like ARIMA model to construct an autoregressive model to forecast stock price. Nonlinear forecasting model requires huge financial data based on rough sets, suggest a decision tree system. The advantages of rough sets and decision trees are mixed in this method, but overfitting occurs.

The trend will be distressed because of huge amount of noise in the datasets. To overcome this, an Artificial neural network is preferred as it deals with nonlinear data models, so it is widely used in time series prediction. In practical application ANN is used to the local optimal problem and SVM is used to reduce it. SVM improves the generalizability of the model. Random forest gives better results than SVM after optimization of parameters. Layered feature representation is mainly analyzed by DNN with complex, deep non-linear relationships. Aim is to predict trend of stock price with the help of liner

regression as a classification model. PCA is used to optimize the results with accurate predicted values by reducing dimensionality and redundancy.

II. LITERATURE SURVEY

Shah D. et al. [1] examined the results of news. Specifically, they advanced a dictionary-based sentiment analysis model and examined the results of news sentiments on stocks for the pharmaceutical market. They achieved a directional correctness of 70.59%. Mehtab S [2] In this paper, regression models consist of NIFTY 50 of certain periods. Afterward, they used deep learning-based regression models using LSTM for prediction. Ne- koeiQachkanloo H [3] A trading system that serves as an artificial counselor for us. Support vector regression helps to predicts future prices. For the recommendation part, two methods were used, first Markowitz's portfolio theory, second a fuzzy investment counselor. Mehtab S, and Sen J [4] developed eight regression and classification methods. For augmenting a predictive framework, they bring public sentiments, and as the two inputs, a fuzzy neural network-based SOFNN algorithm is used. Focus is the ability of various neural networks. Solving the accurate problem in the model by using Genetic Algorithm by Chen [5]. Used two-stage price prediction model GA-LSTM with the help of feature selection and GA-LSTM Algorithm reaches a minimum and 0.0039 MSE score. Using crossover and mutation rate, and several factor combined improves the performnace. Reinforcement learning, agent-based learning apart from supervised and unsupervised learning. Q-learning is a model-free learning algorithm were used R. Sathya [6] Buying and Selling actions can be done with layered neural network. There is method called expReplay when memory gets full. Identifying the trends by data mining and developing time-series data to forecast periodical changes by D. Sorna Shanthi [7] using regression analysis model. The Gaussian process recognizes the trend in stock data. The algorithm will

be learning geometrical shapes. Generative adversarial network, predicts the price and CNN as discriminator used with LSTM and RNN. The patterns are recognized this model in historical market. Chacón HD et al. [8] Empirical mode decomposition, improved the correctness of financial time series. EMD and sample entropy to reduce complexity of forecasting accuracy. LSTM model is the best suited model for forecasting. Gaussian distribution with parameters ($\mu = \sigma$), used for noise removal. First, two IMF's will have lower prediction accuracy than other cases. Stock price trend prediction (classification) and stock price forecast (regression) are forecasting types. To predict yesterday's or 5-day closing price, by using short-term. SVM (Support Vector Machine), used to maximize the minimum interval, the algorithm altered into an optimization problem. Pre-processing of missing data is essential, can be obtained by the fill method implemented by Yuan X [9]. The best suited for the stock price trend is the RF model. Ketter W et al. [10] used time series, interval forecasting, financial services industry, least-square support vector regression, sliding window method. Maio et al developed its forecasting uses ANN. Later, Xiong et al introduced multiple output support vector regression methods used for predicting stock price value. This system provides guessing performance in the financial field. The prediction mainly uses tweet mining, machine learning, sentiment analysis, model stacking, stock movement direction prediction, and textual features extraction. To utilize Twitter data as information of six well-known NASDAQ companies. To predict stock market prediction Chou JS, Truong DN, Le TL [14] need a sophisticated way of sentiment analysis. Li X. [11] proposed a deep neural network for prediction, incorporates the news articles as hidden. For the prediction, Cho P. et al. [12] used finance, equality research report, natural language processing, stock market, investment strategy, and binary classification. Using binary classification, dependent on NLP-elements, framework was explored. Therefore, they

predict positive stock prices. Shen J [15] By using Recursive feature elimination for feature engineering and LSTM model for forecasting the price of the trend of the market included financial aspects as input data. The binary accuracy is 93.25%. The limitation is REE is not suitable for the long term. Scanning the significant acutance between markets and experts. Calculated the Kendall-Tau rankorder correlation and Mean Absolute Percent Error to Analyze the market behavior and elasticity on trading performance. Blohm I. et al.[13] conclude that Price elasticity affects the market design on trading performance thus, the difference in trading performance will be large for moderate elasticity settings and smaller for low and high elasticity settings

III. TECHNIQUES

1 SUPPORT VECTOR MACHINE: SVM is the flexible algorithm. kernel function with VM solves the linear indivisibility problem. The chief purpose is to find a hyperplane. Small design, non-linear and pattern recognition for that solving SVM is involved. It gives better prediction than neural networks.

2 RANDOM FOREST ALGORITHM (RF): It is based on a decision tree algorithm in which record branched into small features and specially used for classification and regression analysis. In this, the random data of the stocks are taken and then train the models to achieve better accuracy.

3 ARTIFICIAL NEURAL NETWORKS (ANN): For non-linear models, ANN mostly requires. This algorithm is similar to the human brain. Input, output the hidden layer are used. The input to the model is the features data and the output is obtained by calculation of hidden layers in between. This algorithm contains nodes and is calculated by the error backpropagation algorithm which contains various steps like Parameter Initialization, Forward propagation, calculate the total error, and error backpropagation.

4 EMPIRICAL MODE DECOMPOSITION (EMD): Development in the accuracy and reduced complexity EMD is used. This method depends on Fourier transformation that decomposes signal based on time scale and finding the regular pattern in time series ARIMA used.

5 LONG SHORT-TERM MEMORY (LSTM): It is the most useful algorithm and is used in many projects of the stock market [10][11]. In this, certain operations are performed and then we train the LSTM model SVM is involved in solving small design, non-linear and pattern recognition. It gives better prediction than neural network.

IV. METHODOLOGY

Tesla's closing price is predicted for the system. Using various Machine Learning Algorithms to prediction the future price of stocks we train the machine from the numerous data points of historical values to make a future price prediction.

1. DESCRIPTION OF DATA: The dataset includes 4 years data from 2010-06-29 to 2020-02-03 of Tesla. High, Low, Open, Close, Adjacent close and Volume are the attributes of data. The extracted prices are of daywise closing.

2. DATA PRE-PROCESSING: It is a very important step that our model will outcome important details from data. Every time we don't get the clean and formulated data Therefore before assign the data in our model, we pre-process and removing outliers, noises, missing values. Data pre-processed by normalization. Normalization changes the numerical values without making differences in the ranges of values.

3. PRINCIPAL COMPONENT ANALYSIS: Dimensions are reduced of various variables from given data set. The principal component is to outcome uncorrelated variables from correlated variables by using the mathematical procedure. Reduce the extension of convert data, only the first few element well-advised as the first principal component have

the largest conflict in data. Less important variables are still present in the most valuable parts of all the variables so it is important as new variables after PCA are independent of one another this is helpful for the linear model requires independent variables should be independent of one another.

5. LINEAR REGRESSION : Supervised technique to fit best line depends on slope and intercept by the Independent variable X and another scalar on dependent the variable Y on X . Insert the input in the form of (X_i, Y_i) , We have to predict Y_{n+1} where $i = 1$ to n , and for a new point X_{n+1} . Equation is: i. $y^{\wedge} = a + b_0$ (I) ii. $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_n x_{1n}$ (II).

V. CONCLUSION

1. Stock market investment is a fascinating and trending topic. However, there are many factors involved in this making decision a complex task. A clever prediction system will help investors make investments accurately and profitably with information of future prediction direction of stock prices.
2. The high-dimension data is an improvement in the ML performance in the classification of PCA. For PCA the impact of features aided, has been investigated.
3. This paper is helpful for researchers to acknowledge the effect of implementing PCA on highly correlated datasets and helpful for selecting various optimum principal components, used while scrutinize the automatic selection of parameters for techniques, which are k for k -NN, kernel parameters in SVM, and the number of PCs.
4. Sentiment analysis has a high impact on future prices. Thus, the investigation can produce a highly efficient prediction of the mixture. For stock price prediction is not able to predict dynamic and rapid changing patterns in stock

price movement a disadvantage of the existing propositions.

5. Many studies estimate their machine learning model by using

VI. REFERENCES

- [1]. Shah D, Isah H, Zulkernine F, "Predicting the effects of news sentiments on the stock market", In 2018 IEEE International Conference on Big Data (Big Data) 2018 Dec 10 (pp.4705-4708) IEEE.
- [2]. Mehtab S, Sen J, Dutta A, "Stock price prediction using machine learning and LSTM-based deep learning models. symposium on Machine Learning and Metaheuristics Algorithms, and Applications", Springer, Singapore. 2020 Oct 14 (pp. 88-106).
- [3]. Nekoei Qachkanloo H, Ghoghogh B, Pasand AS, Crowley M, "Artificial counselor system for stock investment", In Proceedings of the AAAI Conference on Artificial Intelligence 2019 Jul 17 (Vol. 33, No. 01, pp. 9558- 9564).
- [4]. Mehtab S, Sen J. A robust "Predictive model for stock price prediction using deep learning and natural language processing." Available at SSRN 3502624. 2019 Dec 12.
- [5]. Chen S, Zhou C, "Stock prediction based on genetic algorithm feature selection and long short-term memory neural network", IEEE Access. 2020 Dec 24;9:9066-72.
- [6]. R. Sathya, Prateek Kulkarni, Momin Nawaf Khalil, Shishir Chandra Nigam, "Stock Price Prediction using Reinforcement Learning and Feature Extraction", IJRTE Volume-8 Issue-6, March 2020
- [7]. "Pattern Recognition in Stock Market", D. Sorna Shanthi*1; T. Aarthi*2; A.K. Bhuvanesh*3; RA. Chooriya Prabha*4. 2020.
- [8]. Chacón HD, Kesici E, Najafirad P, "Improving financial time series prediction accuracy using ensemble empirical mode decomposition and

- recurrent neural networks” ,IEEE Access. 2020 May 25;8:117133-45
- [9]. Yuan X, Yuan J, Jiang T, Ain QU, “ Integrated long-term stock selection models based on feature selection and machine learning algorithms for China stock market”, IEEE Access. 2020 Jan 24;8:22672-85.
- [10]. Ketter W, Collins J, Gini M, Gupta A, Schrater P, “A Computational Approach to Predicting Economic Regimes in Automated Exchanges”, 2005 Jan 1 (pp. 147-152).
- [11]. Li X, Li Y, Yang H, Yang L, Liu XY, “DP-LSTM: Differential privacy-inspired LSTM for stock prediction using financial news.”, arXiv preprint arXiv:1912.10806. 2019 Dec 20.
- [12]. Cho P, Park JH, Song JW., “Equity Research Report-Driven Investment Strategy in Korea Using Binary Classification on Stock Price Direction.” ,IEEE Access. 2021 Mar 22;9:46364-73.
- [13]. Blohm I, Riedl C, Füller J, Köroglu O, Leimeister JM, Krcmar H, “The effects of prediction market design and price elasticity on the trading performance of users: an experimental analysis.” , arXiv preprint arXiv:1204.3457. 2012 Apr 16.
- [14]. Chou JS, Truong DN, Le TL, “Interval forecasting of financial time series by accelerated particle swarmoptimized multi-output machine learning system.” ,IEEE Access. 2020 Jan 10;8:14798-808.
- [15]. Shen J, Shafiq MO, “Short-term stock market price trend prediction using a comprehensive deep learning system.”, Journal of Big Data. 2020 Dec;7(1):1-33

Cite this article as :

Prof. Sulochana Sonkamble, Vaibhav Vyas, Prathamesh Shimpi, Aniket Mule, Mihir Sonawane, "Stock Price Prediction System", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 2, pp.273-277, March-April-2023. Available at doi : <https://doi.org/10.32628/CSEIT2390229>
Journal URL : <https://ijsrcseit.com/CSEIT2390229>