International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 4 ǁ 2022 ǁ PP. 21-26

Predicting Stock Markettrends using Machine Learning and Deep Learning

Miss. Sanjana pisal[1], Mr. Shaunak Deokute[2],Mr.YashLaddha[3],Miss.BushraNazSayyed[4]
Prof.J.Y.Kapadnis[5](ProjectGuideofCOMPDept.)
UGscholar[1,2,3,4],ProjectGuide[5],Dept.ofComputerEngineering,punevidhyarthiGriha'scollegeofEngineeringNashik,MaharashtaIndia[1,2,3,4]

Abstract: *Equity market forecasting has long been an area of interest to investors and researchers due tovolatility, complexity and the ever-changing nature of, making it difficult to reliably predict.Machinelearning (ML) and data mining are based on the theory that historical data contains important storage forpredicting the future direction. This technology is designed to help investors discover hidden patterns of historical data that are likely to be predictable when making investment decisions.Forecasting stockmarketsisconsideredadifficulttaskinfinancialtimeseriesforecastingusingARIMAmodel.Dataanalysisis a way to predict whether future stock prices will rise or fall.We also looked at various global events andforecasts for the stock market. The stock market can be seen as a special data mining topic. Text miningapproaches are also used to measure the impactof real-time news on stocks. Predict the ups and downs ofthestockmarketusingavarietyoftechniquesandstrategies.*

Keywords*giventothevoicerecognition featuresareasfollows:-*
Forecasting,Dataanalysis,Prediction,Textmining

Date of Submission: 06-04-2022 Date of acceptance: 22-04-2022

I. INRODUCTION

Rapid advances in digital data capture are rapidly increasing the amount of data stored in databases, datawarehouses, and other types of data storage. Data can hide valuable information, but the vast amount of data makes it difficult for humans to extract without powerful tools. Until the beginning of the last decade, news wasnot available easily and quickly. In this information age, content providers and content locators, such as online news services, have emerged on the World Wide Web, making news easily accessible. The constant availability of more news articles in digital format, the latest developments in natural language processing (NLP), and the availability of faster computers, how more information can be extracted from news articles. It leads to the problem of. Financial analysts investing in the stock market are usually unaware of stock market behavior. They face stock trading problems because they don't know which stocks to buyand which stocksto sell in order to make more profits. All these users know that the development of the stock market depends heavily on the relevant news and needs to process vast amounts of information on a daily basis. You need to analyze all the news that appears in newspapers, magazines, and other text sources. However, analyzing such an amount of financial news and articles to extract useful knowledge is beyond human ability. Text mining technology helps to automatically extract useful knowledge from text resources. We will use text mining technology to develop a system that can model and predict the reaction of the stock market to news articles. This allows investors to anticipate future behavior of the stock and take immediateaction when relevant news is released. Use real-time news articles and daytime stock prices from several companies on the Bombay Stock Exchange as input.

The overall purpose of the survey can be summarized in the following survey questions:
• How can you use textual financial news to predict stock price reactions?
• How can data and text mining techniques help generate this predictive model?
To investigate the impact of news on the movement of stock trends, we need to build a forecast model.
Predicting the development of the stock market is one of the most difficult tasks. There are numerous factors involved in forecasting. Physical factors and psychological, rational and irrational behavior. All of these factors can cause stock prices to fluctuate, making it extremely difficult to predict accurately.

II. MOTIVATION

Predicting how the inventory marketplace works is one of the maximum hard tasks. There are severa elements concerned in prediction- bodily . All those elements integrate to make percentage fees risky and really difficultto expect with a excessive diploma of accuracy. Predicting inventory marketplace moves is a famous and thrilling topic. Today, social media is a whole expression of public opinion and critiques approximately modern events. Twitter mainly attracted the eye of researchers reading public opinion.Opinion-based market forecastingexpressed on Twitter is a fascinating areaof research.A previous study concludedthatthegeneralpublicopinioncollectedby Twitter correlates well with the DowJones Industrial Average (DJIA). Thepurpose of this paper is to observe how thevolatility, riseandfallofacompany'sstockpricecorrelateswiththe4,444public opinion expressed in a tweet aboutthat company. Understanding the author'sopinion from the text is the goal ofsentimentanalysis.

III. OBJECTIVE

To predict the future value of thefinancialstocks ofacompany.Toimplement prediction technologies withuse of machine learning which makesprediction based on the values of currentstock market. To use twitter data forprediction via sentimental analysis toincrease the efficiency of system. Toachievemaximumaccuracy.

IV. LIMITATION

Continuesrequirement ofInternet connection. Accurate Input DataRequired.Forsentimentalanalysispropersegregation of the acquired data isrequired.

V. LITERATURE SURVEY

This section describes related work. We have identified related work in two different areas: technology and finance.Kim and Han constructed a discretized model of the function for predicting stock indexes asa combinationof artificial neural networks (ANN) and genetic algorithms (GA). The data used intheir study include technical indicators and the direction of daily changes in the Korea Stock Index(KOSPI). They used data containing a sample of 2928 trading days from January 1989 to December1998 to report selected characteristics and formulas. They also applied feature discretization optimization as a technique similar to dimensionality reduction. The strength of their work is the implementation of GA to optimize ANN. First, the number of hidden layer input features and processing elements is 12, which cannot be adjusted. Another limitation lies in the learning process ofANN, where the author focuses only on two elements of optimization. Although they still believed that GA had great potential for optimizing the discretization of functions. The initialized feature pool is related to the selectedfeature. In this work, the author uses a genetic algorithm with a modelbased on an artificial neural network, which he calls a hybrid GAANN model.Piramuthu has conducted a thorough evaluation of various feature selection methods for data mining applications. He used datasets that are loan approval data, loan default data, web traffic data, tom and kian data, and compared how various feature selection methods optimize the performance of decision trees. Among the methods he compared with the features he chose were probabilistic distance measures classes: Minkovsky distance measures, city block distance measures, Euclidean distance measures, Chebisev distance measures, non-linear (Parsen and Hypersphere kernel) distance measures. Hassan and Nath applied Hidden Markov Model (HMM) to stock market forecasts Stock prices of four different airlines.Reduces the state of the model to four states: open price, close price, highest price, and lowest price. The strength of this white paper is that this approach does not require expertise to build a predictive model. This task is limited within the aviation industry and is interpreted on very small datasets, but it may not be a predictive model with general validity. One of the working approachesto stock market forecasting can be used for comparison work. The author selected up to 2 years as the date range for the training and testing dataset. This provided a reference to the date range of the review. Lei used Wavelet Neural Network (WNN) to predict stock price trends. For optimization, the author also uses an approximate set (RS) to reduce the attribute. Approximate sets are used to reduce the dimension of stock price trends. It has also been used to determine the structure of wavelet neural networks. The dataset for this work includes five well-known stock indices (1) SSEComposite Index (China), (2) CSI 300 Index (China), (3) All Common Indexes (Australia), (4). It is composed of Nikkei. 225 Index (Japan) and (5) Dow Jones Index (USA). The model's valuation is based on various stock market indices, and the results were convincing becauseof its general validity. Using rough sets to optimize feature dimensions before processing reduces computational complexity. However, in the discussion part, the author emphasized only parameter adjustments and did not identify weaknesses in the model itself. In the meantime, I also noticed that the index was beinge valuated. If youapply the same model to a particular stock, you may not get the same performance. Sirignano and Cont were used in a deep learning solution trained on the universal functions of financial markets. The dataset used consisted of trading records for all transactions and cancellation of orders for approximately 1000 NASDAQ shares via stock exchange purchase orders.NN consists of three layers with LSTM units and finally a lookahead layer with rectified linear unit (ReLU), using a stochastic gradient descent (SGD) algorithm for optimization.

Their universal model can generalize and cover constraints other than those of the training data. They mentioned the benefits of the universal model, but the cost of training was still high. In the meantime, it's unclear if the mysterious programming of deep learning algorithms contaminates useless functions when data is fed into the model. The authors found that it was better to perform the feature selection part before training the model, and found that this was an effective way to reduce computational complexity.

VI. ARCHITECHTURE DAIGRAM

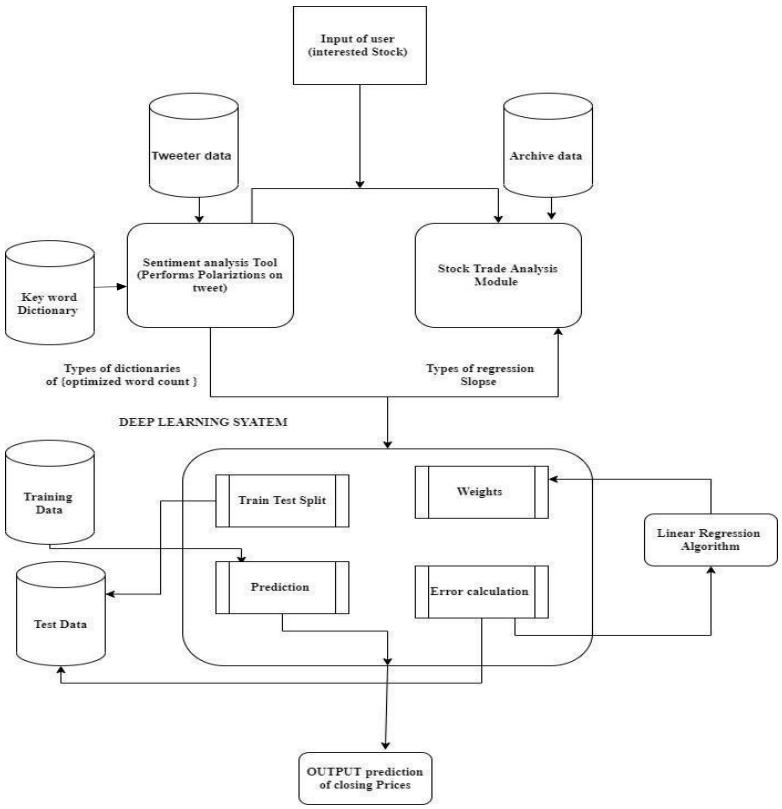


Figure1:SystemArchitecture

Accuratelypredictingstockscanbringenormousprofits to both sellersand brokers. Itis often pointed out that predictions are morechaotic than random. This means that you canmakepredictionsbycarefullyanalyzingthehistory of the stock market in question. Machinelearningisanefficientwaytomapsuchprocesses.Predictmarketvalueclosetomaterialvalue,therebyincreasingaccuracy.The advent of machine learning in the field ofinventoryforecastinghasfascinatedmanyresearchersforitsefficientandaccuratemeasurementas small changes inthedatacan persist largechangesintheresults.Inthissystem,supervisedmachinelearningisappliedtodatasetssuppliedbyTweeterFinance. Thisdatasetcontainsfivevariables: Open, Close, Low, High, and Volume.Open, close, low, and high prices are differentbid prices for stocks at different points in timethat have almost direct names. Volume is thenumberof sharespassed from one ownertoanother during the period. The model isthentested againstthetestdata.

VII. **PROJECTIMPLEMENTATION:Module1**(LoginModule)

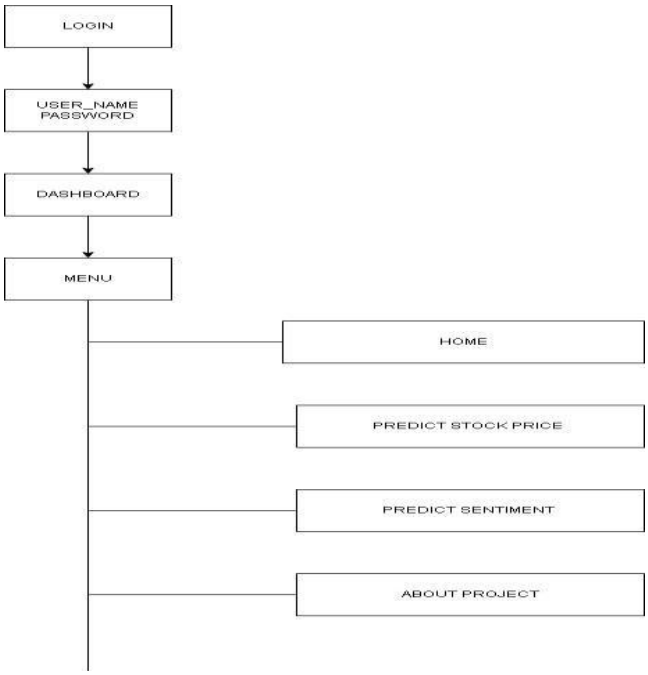
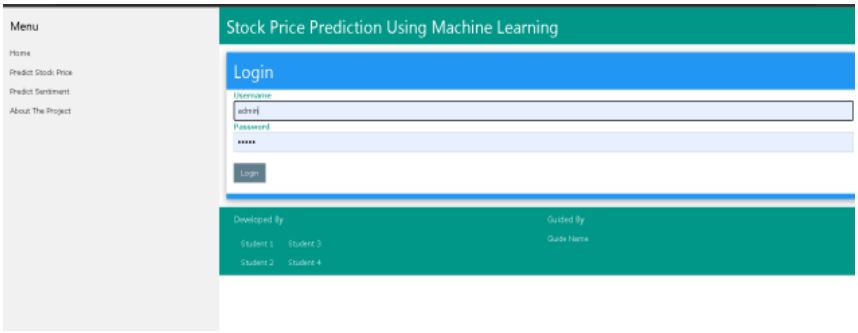


Figure 2: Login Module

The Login Module Consists of Dashboard Wherethe user can login into the account and will bedirectedtothehomepage.
TheMenuoptionconsistsofHome,Predictstockprice,PredictSentimentandAbouttheproject.
Home: The home Page Consists of the introductionpart of the topic opted for project i.e Informationregardingthestockmarket.
Predict Stock Price: This page consists of theprediction part . The price of the stocks arepredictedherewiththeentryofstockname,startdateandenddate.
Predicted sentiment : On this page the sentimentsanalysis of the stock is done and the increase ordecreaseinthepriceispredicted.
About the Project: Here the details about theproject is Displayed along with ArchitectureDiagram.



VIII. Module 2(ARIMA)

1.AutoRegressiveIntegratedMovingAverage(ARIMA)
- A time series forecasting model is a model capable of predicting future values based on previously observed values.
- Time series forecasting is widely used for non-stationary data.
- Non-stationary data is called data whose statistical properties, such as mean and standard deviation, are not fixed over time, but instead, these measurements change over time.

Breakingdown thesetermsweget:
1) AR: means that the model uses a dependency between the observations and a predefined number of delay observations. (Also known as "time lag" or "delay").
2) I: means that the model uses the derivative of the raw observations of

3) MA: means that the model utilizes the relationship between residual error, residual error and observed values.Stepsforimplementationareasfollows:

1. Visualizethetimeseries data.
2. Makethetimeseriesdatastationary.
3. Plotthecorrelationandauto-correlationchart.
4. ConstructtheARIMAmodelbasedonthedata.
5. Usethemodeltomakepredictions.

Mathematical Equation For ARIMA is asfollows :

$z_t = a_{t+1} - a_t$

$z_t = \phi_1 z_{t-1}{}^{+} \theta_1 \varepsilon_{t-1}{}^{+} \varepsilon_t$

$Recover ak$

$ak = zk-1 + ak-1 = zk-1 + zk-2 + ak-2$

$= \cdots$

$k-l$

$\sum z_{k-i+al}$

$i=1$

IX. Module2(SentimentalAnalysis)

Sentiment analysis or emotionclassification is broadly divided into awide range of text classification tasks. Itdisplays a sentence or a list of sentences,letting the classifier know if the emotionbehinditispositive, negative,orneutral. To maintain the problem of binaryclassification, the third attribute may notberetrieved.Forcurrent tasks,moodssuchas "quitepositive"and"quitenegative"arealsotakenintoaccount.

Step 1: Read the Dataframe. importpandasaspd. ...
Step2:DataAnalysis.Next, let's look at the variable "score" to see if most customer reviews are positive or negative.
Step3:ClassifyingTweets....
Step4:MoreDataAnalysis....
Step5:BuildingtheModel
Step6:Testing

NaiveBayesalgorithm

The Naive Bayes classifier calculates the probability of an event by doing the following:

• Step 1: Calculate prior probabilities for a particular class label
• Step 2: Find the probability probability for each attribute of each class
• Step 3: Substitute these values into Bayesian equations to calculate posterior probabilities.
• Step 4: If the input belongs to a class with a higher probability, find out which class has the higher probability.

MathematicalequationforNaïveBayesAlgorithm:

$$P(h\backslash D) = \frac{P(D\backslash h)P(h)}{P(D)}$$

- P(h): The probability of hypothesis hbeing true (regardless of the data).This isknownaspriorprobabilityofh.
- P(D): The probability of thedata(regardless of the hypothesis).This isknownas priorprobability.
- P(h/D): The probability of hypothesis h isgiven the data D. This is known asposteriorprobability.
- P(D/h):Theprobabilityofdatadgiventhat the hypothesis h was true .This isknown as posterior probability.

X. NLP(NaturalLanguageProcessing)

Natural Language Processing (NLP) refers to the field of computer science, especially artificial intelligence or AI. This allows computer systems to understand text and spoken language as humans do.NLP combines complete rule-based modeling of human language in computational linguistics with statistics, device-related, and deep learning modes. The combination of these technologies allows computer systems to systematize human audio in the form of textual or audio data and "understand" the whole, the purpose and

emotion of the speaker or creator. NLP drives a laptop package that translates text content from one language to another, responds to voice commands, and quickly summarizes large amounts of text content in real time. It is very likely that you have interacted with NLP in the form of voice-managed GPS structures, virtual assistants, voice reputation dictation software, customer support chatbots, and other customer conveniences. However, NLP also plays a developmental role in the organization's response to streamline business operations, increase employee productivity, and simplify critical business processes. Laptop algorithms, combined with modal learning and deep learning devices, mechanically extract, classify, and label

factors from textual content and audio data, statistically all possible means of these factors. To capture. Enter the statistic NLP that provides the gender. The NLP structure is very raw, unstructured, untagged, and in-situ. I have. You can "learn" with you. You can extract more detailed information from the text content. And voice recording.

XI. FUTURESCOPE

- Livechartofaparticularstock.
- Displayingthe tweets of a particularstock.
- Show green signal if stock is bullish andred signal if the stock is bearish on theanalysis chart.
- Showthestockpriceofaparticularstock24hrbeforeprediction.

XII. CONCLUSION

The purpose of this study is to use a combination of sentiment analysis, stock trend analysis, and deep learning to improve the accuracy of stock forecasting. The ultimate goal was to increase the rate of return on investment.The accuracy was found to lean towardsprecisionwithincreaseddata.Inconclusion we can say that if we train oursystem withmore input data set, modifythe deep learning module to perform moretrial and errors tests, it has the potential togeneratemoreaccurate andsignificantlymoreconsistentminimalerrorpredictions.

REFERENCES

[1]. S. Selvin , R. Vinayakumar , E. A. Gopalakrishnan , V. K. Menon and K. P. Soman.(2017) ”Stock price prediction using LSTM, RNN and CNN-sliding window model.” International Conference on Advances in Computing, Communications and Informatics: 1643-1647.

[2]. Rather A. M., Agarwal A., and Sastry V. N. (2015). ”Recurrent neural network and a hybrid model for prediction of stock returns.” Expert Systems with Applications 42 (6): 3234-3241.

[3]. Zhang G., Patuwo B.E., and Hu M. Y. (1998). ”Forecasting with artificial neural networks: Thestate of the art.” International journal of forecasting 14 (1): 35-62.

[4]. Heaton J. B., Polson N. G., and Witte J. H. (2017). ”Deep learning for finance: deep portfolios.” Applied Stochastic Models in Business and Industry 33 (1): 3-12.

[5]. Jabin S. (2014). ”Stock market prediction using feed-forward artificial neural network”. growth 99 (9).

[6]. Hamzaebi C., Akay D. and Kutay F. (2009). ”Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting.” Expert Systems with Applications 36 (2): 3839-3844.

[7]. Rout A. K., Dash P. K., Dash R., and BisoiR. (2015). ”Forecasting financial time series using alow complexity recurrent neural network and evolutionary learning approach.” Journal of King SaudUniversity-Computer and Information Sciences 29 (4):536-552.

[8]. Moghaddam A. H., Moghaddam M. H., and Esfandyari M. (2016). ”Stock market index prediction using artificial neural network.” Journal of Economics, Finance and Administrative Science 21 (41): 89-93.

[9]. Zhang G. P. (2003). ”Time series forecasting using a hybrid ARIMA and neural network model.”Neurocomputing50:159-175.

[10]. Menon V. K., Vasireddy N. C., Jami S. A., Pedamallu V. T. N., Sureshkumar V., and Soman K. P. (2016, June). ”Bulk Price Forecasting Using Spark over NSE Data Set.” In International Conference on Data Mining and Big Data : 137-146.