**Abstract**

**Stock Closing Price Prediction**

**Using Machine Learning Algorithms :**

**A Critical Review Analysis of Techniques**

Anjaneyulu Gudla A Dr P. ChandraShekhar Reddy B

***A) Anjaneyulu Gudla,Phd Scholar from SR Unversity ,Warangal and working as Assistant Professor at Trinity College of Engineering & Technology, Peddapally, state: Telangana, country: India ,******Email:*** [***anjugudla@gmail.com***](mailto:anjugudla@gmail.com)

***B) Dr P. ChandraShekhar Reddy*** *Dr P.Chandrashekhar Reddy sir is working as a Professor at SR University Warangal in the Department of CS & AI, state:Telangana, country: India Email:* [*p.chandrashakerreddy@sru.edu.in*](mailto:p.chandrashakerreddy@sru.edu.in)

**Stock market forecasting is a interesting but challenging problem that has attracted the attention of economics & computer scientists. Generally today’s financial world stock exchange has become one of the most significant events. The world’s economy today is totally dependent on the stock market price. Individuals make their stock market investment decisions based on predictions or past research knowledge. Due to eccentricity and non-linear nature of economic inventory markets, as is should be predicting stock market returns is extraordinarily tough why because, the creation of machine studying and stepped forward processing functionality, programmable prediction techniques have established to be more available in predicting stock values. In terms of prediction people often look for tools or methods that will minimize their risks and maximize profits so the stock price prediction takes on an influential role in the stock-market-business. Research provides subject represents non AI machine learning techniques. Adopting traditional methodologies such as fundamental &technical analysis doesn’t seem to ensure the correct final value and accuracy in the prediction’s result the Machine Learning techniques have become the** **recent trend in**

**the stock market prediction whose prediction price based on the eventually as on outcome of training on their last values. This paper focus on forecast techniques in predicting the stock closing price, with supervised algorithms Random Forest and KNN with future work of RNN (Recurrent Neural Networks) and LSTM(Long Short Term Memory)**

***keywords:***

***Machine learning algorithms, supervised Random forest, KNN , deep learning algorithms LSTM,RNN***

**I. Introduction:**

The stock marketplace is dynamic, unpredictable, and nonlinear in nature. Predicting inventory values is a tough assignment why because they are affected by a spread of factors, which include financial system, an enterprise’s economic-reviews overall performance. Buying &selling financial assets or crypto currencies in various markets to make money by financial assets or trading involves high risks, such as the risk of losing the amount invested, and may not be suitable for all traders. Crypto currency prices are affected by external many factors such as economic, political or regulatory events and are highly volatile. Before deciding to invest in financial instruments or crypto currencies, one should understand the risks and costs involved in financial markets, consider invest objectives, trading experience, risk tolerance, &the seek professional advice if necessary. Predicting the price changes of crypto currencies helps us to predict and control out daily purchasing activities and manage our experience.

This enables us to create strategic and get ready for unforeseen events. Managing the budget is as important as having a global vision in order to plan projects, anticipate change, and stay out of the way of several limit. With the advent of automation technology, it is now feasible to forecast balances in real time and with a level of relevance that is higher than usual. As can be seen from the above, managing every spending helps with financial tracking and figuring out future course. In the global financial markets, professional traders have typically relied on computers to help them with their increasingly complicated large-scale arrangements. It is estimated that the majority of transactions in the world today are made through algorithm-based trading. While popular trading tools are made to acquire and sell financial instruments as quickly as feasible. AI-based model search for long term optimal periods to buy and sell. With so much data at its disposal, artificial intelligence (AI) can be a priceless trading aid. Algorithms in the financial markets are already more efficient than people, and technology is always developing.

Data, economic activity, climate indicators, and temporal behaviours are just a few of the well-known phenomena that are necessary for precise predicting of future events. In addition, one of the most important activities that business owners and individual depositors do when choosing how to divide their funds is stock market forecasting. Although the stock market s a dynamic, non-linear system, it can be challenging to predict the future of the stock market with any degree of accuracy [7].

The main objective of stock market prediction is to accretion a possibly profitable rate at which company shares may be traded on exchange. Furthermore, accurate forecast of future stock values could result in significant profits. Forecast movement generally falls into three categories: short, medium, and long. The following are these categories: short-term forecasting is the act of predicting the time in a span of minutes, hours, or days within a week; medium-term forecasting is the act of predicting the time in a span of weeks to months. Similarly, a prognosis that is one year or greater in duration denotes long term [8].

**Prediction Analysis Methods:**

Major Prediction analysis methods are fundamental analysis, technical analysis and machine learning method.

**Fundamental Analysis:**

Fundamental Analysis considers economic factors as fundamentals. Fundamental analysis is the physical study of a company with respect to product sales, workers, infrastructure, and quality. This analysis is mostly suitable for long terms prediction as it depends on statistical data of the company.

**Technical Analysis:**

Technical analysis mainly considers indicators on stock charts that will decide the future movement. It normally uses technical data price, volume, highest and lowest prices to forecast price moments. This kind of analysis is normally suitable for short time span.

**Machine Learning:**

Machine learning method uses Artificial Intelligence (AI), for training the system and then trained system for forecasting future movements in stock.

The structure of this work is as follows: in section II, we outline the typical data types utilized in stock market forecasting and literature review, and in section III, we offer worldwide machine learning(ML) methodologies. The executed examinations of **RF, KNN** are then covered in section IV. Next, we go over the suggested research approach **LSTM** that was used in section V.

**II Data Types for Stock Market**

Most frequently use market data or historical data, which covers all trading activities that take place on exchanges, such as open/high/low/close prices, volume, etc.(historical prices in look back windows) and forecasts a target(such as the closing price for the next day) [1]. Textual data connected to texts produced by people, such news and social media, can also be employed. These data, as another kind, are more challenging to gather and analyze, but they can include information not available in previous data that could be helpful in making predictions. This textual data can be subjected to sentiment analysis techniques like Natural Language Processing (NLP) to provide sentiment variables that can then be utilized for forecasting. Because macro economic statistics, such as GDP reflect the economic conditions of a specific sector, area, or nation, they are also a particularly fascinating sort of data [2]. These data points can validate or refute the quality of the stock market rise and show the general health of the market, making them pertinent to the stock market. Other authors, including jiawei long et al [3], opted to employ knowledge graph data date the high correlation between various markets and companies; for instance, new can have as impact on stock movements within the same industry. With the aid of right away created graph neural networks, open source knowledge graph data is now available to enhance predictive performance.

The gross domestic product (GDP) can be influenced by the stock market, which is widely regarded as a sentiment indicator. The overall output of goods and services produced by an economy is gauged by its GDP. Economic mode fluctuates along with the stock market [4,5]. Conversely, the stock market can impact GDP in both good and negative ways.

A percentage increase form one period to the next is typically used to express GDP. For instance, if the growth rate from quarter to quarter is 2%, it indicates a 2% yearly growth in the economy for that particular quarter. It’s critical to keep in mind that consumption drives the majority of India’s economy. Almost three-fifths of Indian economy has been made up of citizen’s private consumer spending, or the money they spend on goods and services [6].

**Literature Review:**

Stock market prediction was simulated using a three-stage process [9]. Financial and economic factors with stronger associations were found using multiple regression analysis. Type-2 Fuzzy Clustering based on differential Evaluation was used to develop the prediction model. The final phase was the prediction using a fuzzy type-2 neural network. Although this method is quite computationally difficult, its prediction error is reduced.

Interrelated Time Series Data was used to predict the stock market[10], this methods extracted time series includes foreign exchanges, global stock market indexes. This prediction approach was composed of two stages: prediction and interrelation finding.

In order to predict the stock market, the authors of [1] examined the behaviour of the market and identified the best-fit model among a number of well-known machine learning techniques, including as Random Forest (RF), Support Vector Machine(SVM), Naive Bayes, K-Nearest Neighbour (KNN). The authors carried out a comparative analysis of these methods, using data acquired from various sources such as Yahoo and NSE India, and applying a number of technical indicators to the data. When the accuracy of each model was examined, it was found that for large datasets, RF produced the most gratifying results, while Naive Bayesian produced the highest accuracy for small datasets. Another finding was that the model’s accuracy dropped as the number of technical indications decreased. The supervised learning method of linear regression is used in the suggested research project essentially describes how to use a dataset to predict the closing value by analyzing the GOOGLE stock and extracting data spanning over 14 years.

According to the research [11] by Meghana Misra et al., after doing Principal Component Analysis (PCA) on the data to identify the most pertinent components, predictions made using the Linear Regression Model had a higher accuracy rate. When it comes to non-linear classification data, SVM shows excellent accuracy; for linear data, linear regression is preferred due to its high confidence value; for binary classification models, the Random Forest Approach yielded a high accuracy rate; and for making predictions, the Multilayer Perception (MLP) produced the least amount of error.

As the authors in the paper[12] conclude by studying the application of machine learning models to analyze financial trading and to design optimal strategies for the same, many of the aforementioned techniques are not just limited to stock price prediction but can be used broadly in the financial markets. Following a quantitative examination of several methods, the authours advice going deeper into behavioural finance to assess investor or market psychology &comprehend changes in the market &author suggest using machine learning and text mining techniques to track user behaviour on digital financial trading platforms.

**III. Machine Learning Methodologies:**

**1) Support Vector Machine:** A machine learning approach called Support Vector Machine (SVM) is used to solve regression and classification issues. It uses a kernel trick technique to alter the input, and then it uses that knowledge to determine the optimal bound amongst all potential outputs. The SVM searches for the boundary dividing the two classes. Depending on the whether it is on the positive or negative side of the line. Indu kumar et al. [1] used SVM to predict the stock market; however, Wen Fenghua et al.[13] achieved superior results by decomposing using Singular Spectrum Analysis(SSA).

**2) Random Forest (RF) :** During training, Random Forest generates large number of decision trees and outputs the average prediction of each tree. Random Forest is a collaborative learning technique. Here’s how it works

**a) Training Phase:** During the training phase, RF constructs many decision trees. These decision trees are trained independently of one another using random subsets of training data and features.

**b) Forecast phase:** Every tress in the forest makes an independent forecast of the target variable (in this ,close prices) in order to make predictions. With the goal to generate a more reliable and accurate forecast, the final prediction is then achieved be averaging the predictions of all the trees( regression task). Large volumes of data may be handled by it effectively. The RF algorithm predicts outcomes more accurately than decision tree algorithm does. By lowering the number of estimators, the RF may be more quickly. Adding more trees is the most effective method of raising the model’s accuracy level.

**3) K-Nearest Neighbour (KNN):** Based on the group that the nearest data categorization technique that estimates the likelihood that a data point is a member of one or the other group. Feature similarity is used by Artificial Neural Network (ANN) Algorithms to forecast the value is assigned to each new point depending on how similarity is to existing points in the training set. When labelled data is too costly or unavailable, ANNs come in handy, allowing for high accuracy in arrange of prediction tasks.

**4) Naive Bayes :** The Bayes theorem applied via Naive bayes(NB) classification algorithm. It forecast each class’s likelihood of membership. The class with the highest probability is the most possible class. NB operates on categorical input variables as opposed to numerical ones. For data forecasting based on past performance, it is helpful. As a fast-learning classifier, we take into consideration NB for real-time prediction. As a result, real-time prediction is possible with it.

**5) Artificial Neural Networks (ANN), RNN, LSTM:**

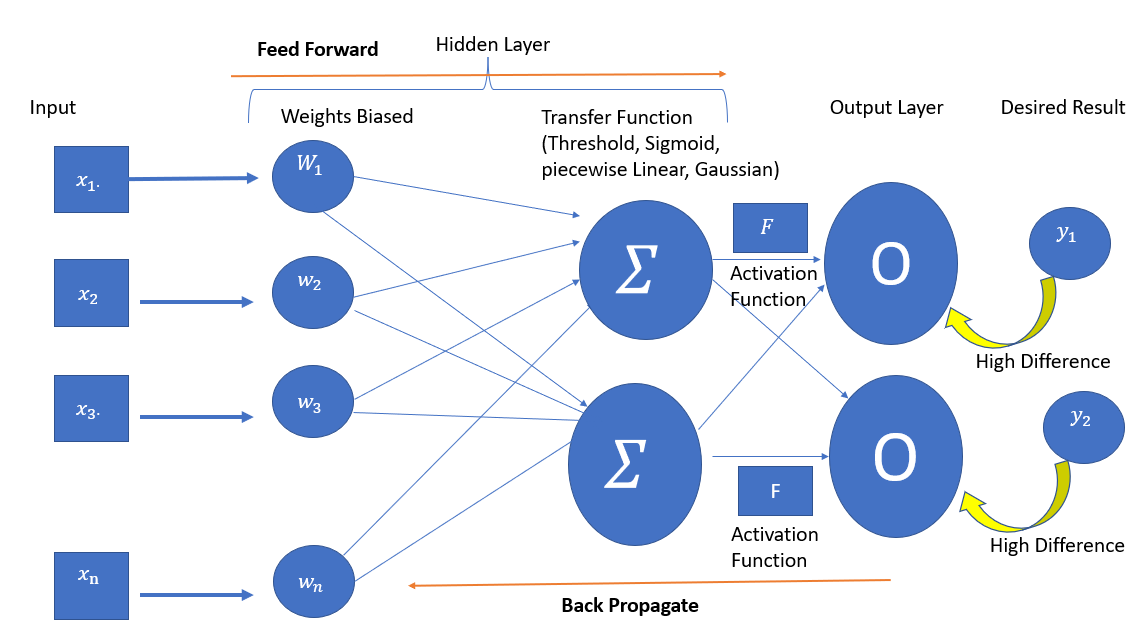
Among many other applications, Artificial Neural Networks(ANNs) can be utilized for sentiment analysis. Sentiment analysis, sometimes referred to as opinion mining, is the technique of examine textual data to ascertain the sentiment that is being expressed. This feeling could be neutral, hotline, or pleasant. Sentiment analysis tasks have seen the effective application of artificial neural network (ANNs), namely recurrent neural networks (RNNs) and their derivatives such as long Short-Term Memory (LSTM) network recurrent neural networks(RNNs) and their derivatives such as Long Short-Term Memory(LSTM) networks. In order to comprehend the sentiment conveyed in sentences or documents, these networks must be able to recognize dependencies and contextual information contained in text data. ANN emulates the capabilities of human mind. In human mind neurons which are also called nerves cells, contact to each other through sending messages.

ANN is highly capable and efficient data driven model which is widely used in complex non linear behaviour of data. ANN can be used in Hand Writing Recognization, Pattern Recognization, Face Identification, text Translation, Medical Diagnosis, Speech recognization, Credit card Fraud Detection, Stock Market Prediction etc.

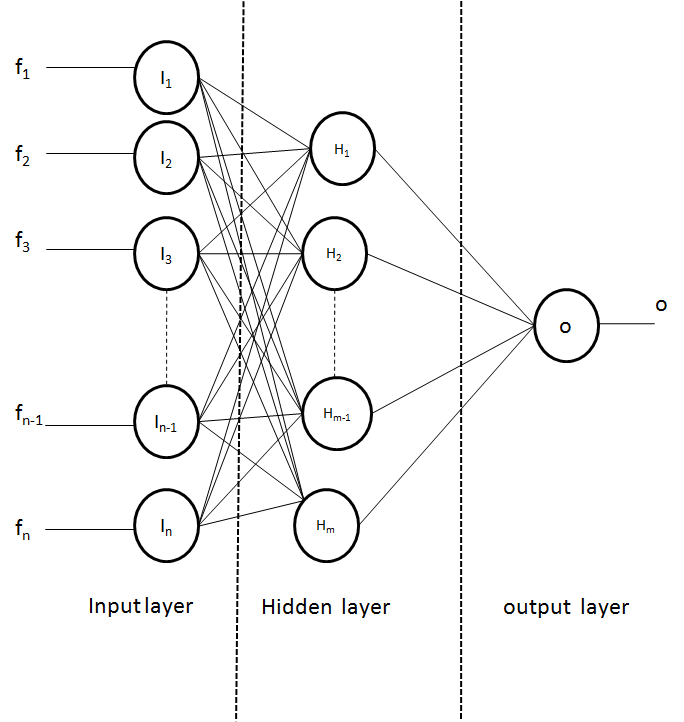
**Recurrent Neural Networks (RNNs):**

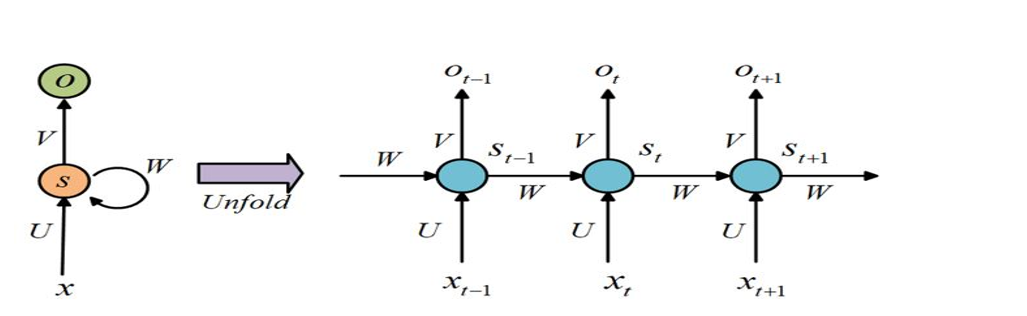
RNNs are a type of deep learning model that are specifically designed to handle sequential data, such as time series, text, and speech data. Unlike ANNs, RNNs have a feedback loop in their structure, which allows them to take into account the context from previous time steps when making predictions.

In an RNN, the hidden state from the previous time step is passed along with the input to the current time step, allowing the network to capture information from the entire sequence of inputs. The hidden state is updated with each time step, allowing the network to maintain information about the entire sequence.

**ANN**

**Multi Perciption Model (ANN)**

****

**RNN**

**IV. The Execution Operations of RF, KNN:**

**System Setup:**

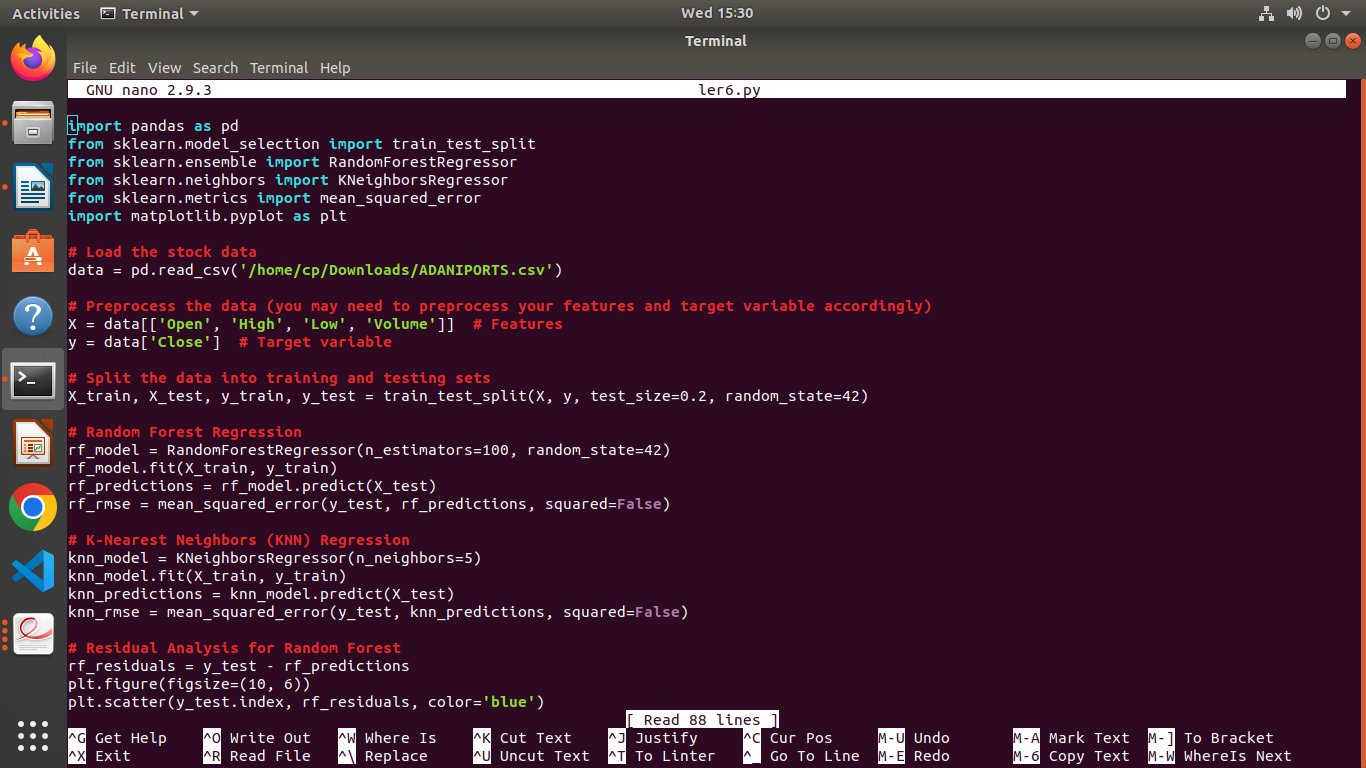
a) Data Collection

Here Kaggle’s historical inventory repository TESLA dataset was used. With a simple click, kaggle affords a training and trying out dataset that may be used to retrieve a organization’s stock rate for any time variety. This eliminates the want for manual records mining through different manner. We used stock data from the Tesla Corporation for this prediction, which spans exactly 10 years from 2010 to march 2020. Date, open, low, high, close, adj near, and volume are some of the 2193 values inside the dataset.

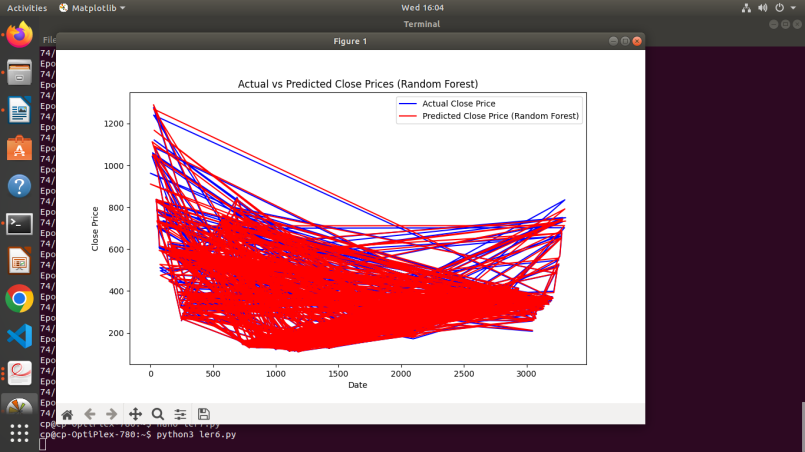
b) platform ubuntu o/s and install necessary libraries in python environment:

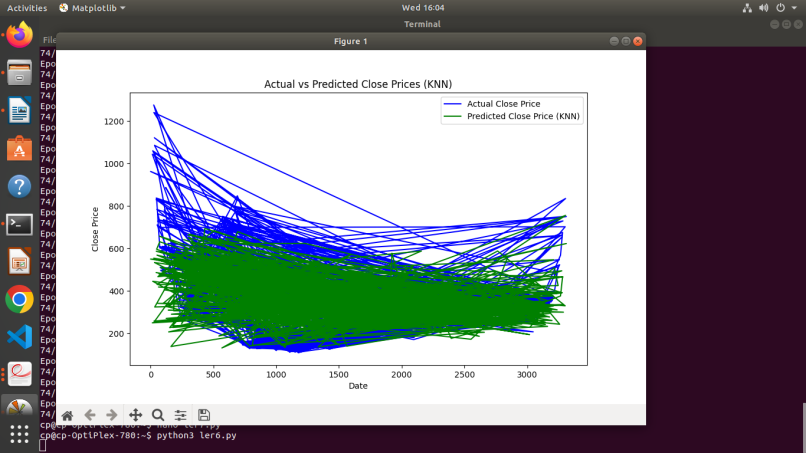
1. we may use JupyterLab , it is an interactive pocket book, code and records development surroundings this is handy through the web and here python will undoubtedly be compared to the numerous other domain-particular open source and commercial prgramming languages and equipment in extensive use for statistics analysis and interactive, exploratory computing and statistics visualization, inclusive of R, MATLAB, SAS, and others. In recent years, Python’s growing library guide (particularly pandas) has made it a great choice for records processing workloads.
2. Numpy, Pandas, Matplotlib, Chart Studio :

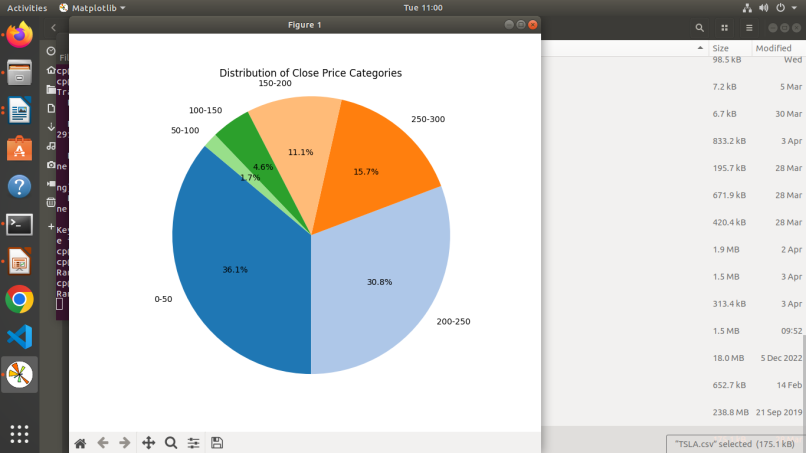
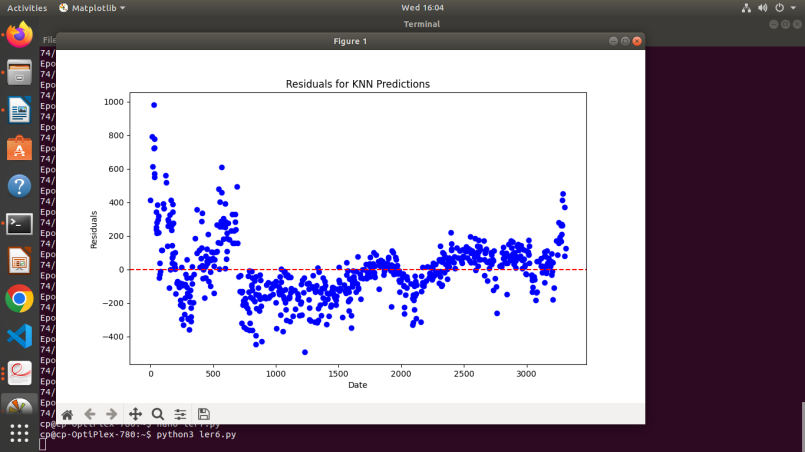
Numpy is a scientific computing environment based totally on Python. Pandas is a python open-source library that provides efficient and person- pleasant facts shape and evaluation skills. Matplotlib is a python toolkit that permits you to create graphs, charts, and different records visualization tools. We will want the Chart Studio package deal to apply it, which includes utilities for running with Plotly’s Chart studio provider



Here after executed trained by algorithms RF and KNN, we got below prediction observations and charts.







Based on the provided output and assuming that the Random Forest and KNN models were trained and evaluated correctly, we can draw the following conclusions:

|  |  |
| --- | --- |
| **Algorithm Used** | **Accuracy of RMSE** |
| Random-Forest | **2.9760530477364284** |
| **K-Nearest Neighbor** | **82.60991988263267** |

Additionally, it's noted that the length of is 484, while the length of the Data Frame index is 2416. This discrepancy indicates an issue with the KNN model's predictions, likely due to misalignment between the input data and the model's expectations. Given these results and the issue with the KNN predictions' length mismatch, the Random Forest model appears to be the better performing algorithm for stock price prediction in this scenario.

However, it's crucial to ensure that both models are trained and evaluated under similar conditions and that any discrepancies in the data or modelling process are addressed before making a final determination. Additionally, further analysis, experimentation, and tuning may be necessary to optimize the performance of both models.

**Algorithm formulas used here in my research**

**the Root Mean Squared Error (RMSE)** for both KNN and Random Forest algorithms.

The RMSE for a dataset with n samples can be calculated using the following formula:

RMSE=n1i=1∑n(Yactual,i-Ypredicted,i)2

Where:

* Yactual,i is the actual value for the ith sample.
* Ypredicted,i is the predicted value for the ith sample.
* n is the total number of samples.

Let's apply this formula for both KNN and Random Forest:

**1. For KNN:**

RMSE KNN n1i=1∑n(Ytest,i−YKNN,i)2

**2. For Random Forest:**

RMSE RF n1i=1∑n(Ytest,i−YRF,i)

Where:

* Ytest,i is the actual closing price for the ith sample in the test set.
* YKNN,i is the predicted closing price by KNN for the ith sample.
* YRF,i is the predicted closing price by Random Forest for the ith sample.
* n is the total number of samples in the test set.

We have used these formulas to calculate the RMSE for both algorithms in Python code.

**V. Conclusion and future work:**

Predicting stock market returns is a tough task because inventory values are constantly converting and are dependent on multiple parameters that shape complicated styles. The historical dataset available on the enterprise's website consists of only some functions inclusive of high, low, open, close, adjoining near price of inventory prices, quantity of shares traded, and so on, which can be insufficient. to improve the accuracy of the expected fee, new variables had been created via combining current variables. The modelling of the sampling process, which was a very difficult topic in stock market prediction, was the most important contribution that we made. The process of picking the testing and training datasets was another difficult challenge. When developing the model that is used to forecast what would happen in the stock market, the LSTM algorithm is employing as the basis for the construction of the model.

**VI. References**

[1] Kumar, I., Dogra, K., Utreja, C., & Yadav, P. (2018, April). A comparative study of supervised machine learning algorithms for stock market trend prediction.

In 2018 Second International Conference on Inventive Communication and Computational Technologies(ICICCT) (pp.1003-1007). IEEE.

[2] Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory.PloS one, 12(7), e0180944.

[3] Long, J., Chen, Z., He, W., Wu, T., & Ren, J. (2020). An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market. Applied Soft Computing, 91, 106205.

[4] Stock market prediction using deep learning algorithms Somenath Mukherjee 1 Soumil De 2

[5] 1. Statman, M., Fisher, K.L.: Consumer Confidence and Stock returns. SSRN journal. <https://doi.org/10.2139/ssrn.317304>

[6]Kaul,V.https://www.deccanherald.com/specials/sunday‐spotlight/whats‐in‐store‐for‐the‐indian‐economy‐in‐2020‐789545.html

[7] S. Vazirani, A. Sharma and P. Sharma, "Analysis of various machine learning algorithm and hybrid model for stock market prediction using python," 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics

(ICSTCEE), 2020, pp. 203-207, doi: 10.1109/ICSTCEE49637.2020.9276859.

[8] B. Panwar, G. Dhuriya, P. Johri, S. Singh Yadav and N. Gaur, "Stock Market Prediction

Using Linear Regression and SVM", 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2021, pp. 629-631,

doi: 10.1109/ICACITE51222.2021.9404733.

[9] D. Wang, X. Liu and M. Wang, "A DT-SVM Strategy for Stock Futures Prediction with

Big Data," 2013 IEEE 16th International Conference on Computational Science and

Engineering, 2013, pp. 1005-1012, doi: 10.1109/CSE.2013.147.

[10] M. S. Roobini, K. Babu, J. Joseph and G. Ieshwarya., "Predicting Stock Price using Data

science technique," 2022 Second International Conference on Artificial Intelligence and

Smart Energy (ICAIS), 2022, pp. 1013-1020, doi: 10.1109/ICAIS53314.2022.9742772.

[11] Misra, Meghna, Ajay Prakash Yadav and Harkiran Kaur. “Stock Market Prediction using Machine Learning Algorithms: A Classification Study.” 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE) (2018): 2475-2478

[12] Vats, P. and Samdani, K., 2019, March. Study on Machine Learning Techniques In Financial Markets. In 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN) (pp. 1-5). IEEE.

[13] Fenghua, W. E. N., Jihong, X. I. A. O., Zhifang, H. E., & Xu, G. O. N.G. (2014). Stock price prediction based on SSA and SVM. Procedia Computer Science, 31, 625-631.