Stock Price Prediction using MachineLearning

Dr.K.Lakshminadh[1],KothamasuVamsiKarthikeya[2],Manda Durga Venkatesh[3],Markapudi Chaitanya Kumar[41]
Professor, [2, 3 & 4] Student
[1]drklmn7@gmail.com,[2] kothamasuvamsikarthikeya777@gmail.com,[3]venkateshmanda880@gmail.com,[4]bobbymarkapudi@gmail.com
Department of Computer Science and Engineering,
Narasaraopeta Engineering College, Narasaraopet, Andhra Pradesh, India

ABSTRACT- In the realm of financial markets, stock price prediction remains a critical subject, influencing risk management and investment strategies. This abstract offers an overview of current advancements, methodologies, challenges, and potential future directions in the domain of stock price forecasting.

It commences by delineating the evolution of stock price prediction techniques, transitioning from conventional statistical methods to more sophisticated machine learning and artificial intelligence models. Through harnessing vast volumes of financial data encompassing historical prices, trading volumes, sentiment analysis of news, and macroeconomic indicators, the adoption of data-driven methodologies has significantly bolstered predictive accuracy and resilience.

The introduction delves into pivotal methodologies for projecting stock values, encompassing time series analysis and machine learning algorithms such as decision trees and support vector machines. Moreover, there is a burgeoning interest in leveraging social media data, which offers a rich repository of textual content spanning from casual observations to intricate debates. This data holds promise in enhancing the precision of stock price predictions, thereby enabling more judicious investment decisions.

KEYWORDS: Stock price prediction using Machine Learni5ng,open, Close,volumes high, low, decision tree model,regression and volume

1. INTRODUCTION

"Machine learning" is a subfield of artificial intelligence that aims to improve accuracy over time by defining human learning via the use of diverse data and methods. As a result, machine learning (ML) is useful in developing the model using a variety of methods, after which it is evaluated and trained on this enormous volume of data. In order to build computers that can learn from data, machine learning (ML) employs complex statistical and probabilistic techniques. The dataset is a key component in machine learning. Because even a slight alteration in the data might have a significant effect on the result, the dataset should be as exact as feasible. This study uses supervised machine learning on a Netflix Finance dataset. This dataset has five variables: volume, open, close, low, and high. The terms "open, close, low, and high" relate to different, nearlyidentical bid prices for the stocks at different times. The quantity of shares that were exchanged during the transaction is referred to as the volume. The model is then tested using the test data.

Regression and decision tree models are employed separately in this work. Regression reduces errors, but decision trees store data and outcomes for a longer period of time. Moreover, graphs showing the regression-based model's projected and actual prices as well as the price change over time are given. There's also the k-nearest-neighbor method.

2. LITERATURE SURVEY

The review of the literature showed how widely machine learning techniques are applied globally to forecast stock market movements. Machine learning is more precise and effective in making predictions than conventional methods. Globally, there is a great deal of interest in studying this topic. Notably, studies by K. Raza [2] and M. Usmani, S. H. Adil, K. Raza, and S. S. A. Ali [1] examined several machine learning techniques and emphasized recent developments in this field.

While M. Billah, S. Waheed, and A. Hanifa [4] created their own training strategy to increase stock prediction accuracy using neural networks, H. Gunduz, Z. Cataltepe, and Y. Yaslan [3] employed deep neural network approaches to anticipate stock prices. Parallel to this, K. V. Sujatha and S. M. Sundaram [6] offered effective strategies for handling anomalous events that arise during system operation and often result in misleading projections or disturbances. An LSTM model was also employed in our work to predict stocks, which may indicate opportunities for increasing prediction accuracy. Additionally, studies and simulations were conducted by K.A. Althelaya, E.M. El-Alfy, and S. Mohammed [9] to assess the feasibility of using deep learning to predict stock values. Furthermore, ModAugNet, a ground-breaking stock forecasting framework, was introduced by Baek and Kim [10].marketplace

H. Y. Kim and C. H. Won, “Forecasting the volatility of stock price index: a hybrid model integrating LSTM with multiple GARCH-type models,” *Expert System Application*, vol. 103, pp. 25–37, 2018.B. Chen, J. Zhong, and Y. Chen, “A hybrid approach for portfolio selection with higher-order moments: empirical evidence from Shanghai Stock Exchange,” *Expert System Applications,* vol. 145, Article ID 113104, 2019.

The mentioned above author done the work on the similar work belongs to the price prediction on the different conditions like volumes,price momentum and RSI indicator and other various models used to predict the stock pirce

3. PROPOSEDSYSTEM

Our Model is proposed based on certain criteria as follows.

A. System Setup
B. Dataset Collection
C. Model Architecture
D. Deployment
E. Pre-processing

A. System Setup

The system configuration includes creating and analysing machine learning algorithms with Python 3.11 and the Scikit-learn module. Google Collaborator is the platform for the training and testing phases. Scalability and smooth process integration are achieved.

B. Dataset Collection

We gather three different companies data sets they are Amazon,Microsoft and Google, for each data set as follow colums like open,high,low,close as shown in below

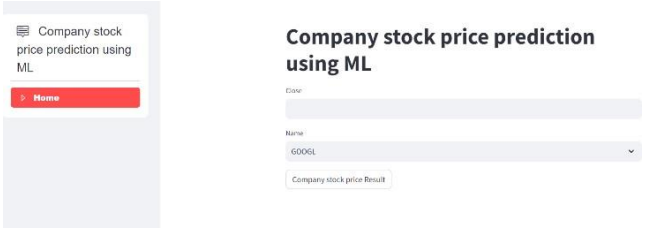
	Date	Open	High	Low	Adj Close	Close	Volume	Name
12852	11-06-2008	49.80	51.04	46.30	150.0	47.22	15491598	AMZN
12853	11-07-2008	47.76	49.79	47.01	150.0	49.21	7439842	AMZN
12854	11-10-2008	49.98	50.68	46.86	150.0	48.46	8986388	AMZN
12855	11-11-2008	47.10	48.06	44.81	150.0	46.30	9940491	AMZN
12856	11-12-2008	43.99	45.44	40.90	150.0	41.56	14773108	AMZN

C. Model Architecture

In this step, an optimal machine learning algorithm is chosen and applied to the preprocessed data. The model is trained, and its parameters are changed to reduce the difference between expected and actual outputs on the training dataset. The model's performance is then examined using a different validation dataset to determine its efficacy post-training.The nature of conflict and the performance requirements determine which assessment characteristics are used. Common assessment measures include accuracy, precision, cross-validation, and the F1 score. The model may be further enhanced based on the assessment results by modifying its parameters or utilising different methods. This phase is critical for improving the model's performance on newly collected, untested data. The model may be used to make predictions or decisions in a production setting after it has been trained and assessed.

D. Deployment

The final model is merged with the Python AWS module and streamlit for integration into an application. The programme merely accepts input and outputs it.



E. Pre-Processing

AMZN.info() : These gives the basic informaction about the dataset which consists of following columns

```
<class 'pandas.core.frame.DataFrame'>
Index: 648 entries, 12852 to 13499
Data columns (total 8 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Date       648 non-null    object
 1   Open       648 non-null    float64
 2   High       648 non-null    float64
 3   Low        648 non-null    float64
 4   Adj Close  648 non-null    float64
 5   Close      648 non-null    float64
 6   Volume     648 non-null    int64
 7   Name       648 non-null    object
dtypes: float64(5), int64(1), object(2)
memory usage: 45.6+ KB
```

Fig 2 information of data set

AMZN.isnull().sum(): Calculates the total of missing values in each column of the dataset. This aids in detecting missing data in columns.

```
AMZN.isnull().sum()

Date         0
Open         0
High         0
Low          0
Adj Close    0
Close        0
Volume       0
Name         0
dtype: int64
```

Fig 3 null values

after the preprocessing the data will be represent in the following way in a graphical representation.

Data Visualization

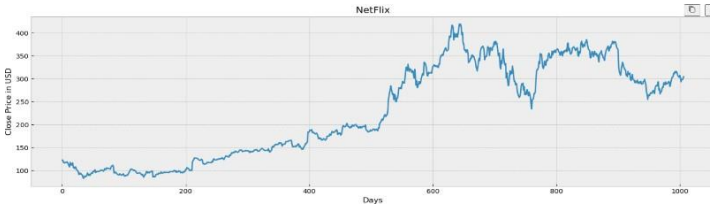


fig 4 data visualization after preprocessing

In graph x-axis represent days and y-axis represent price of amazon in USD.The graph represent the data of Amazon from past may years how it perform .By performing the operations now its completely zero error and no duplicates in the data

The graph displays Amazon's success across numerous years, with the x-axis denoting days and the y-axis displaying stock values in USD. The data is painstakingly filtered to ensure that there are no mistakes or duplications, resulting in an accurate depiction of the stock's past performance.

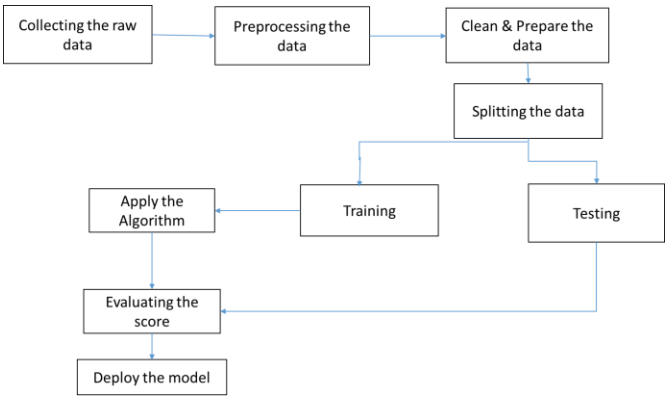


Fig 5 pre-processig steps

A decision tree was a hierarchical model decision-making model that represents various actions and they possible outcomes, such as costs, utility, and chance event results, using a tree-like structure. This method illustrates the decision- making processes of an algorithm through conditional control statements. In operations research, decision analysis in particular, decision trees are used to assist determine the most effective methods for achieving specific goals. They are frequently employed as a crucial machine learning tool as well.

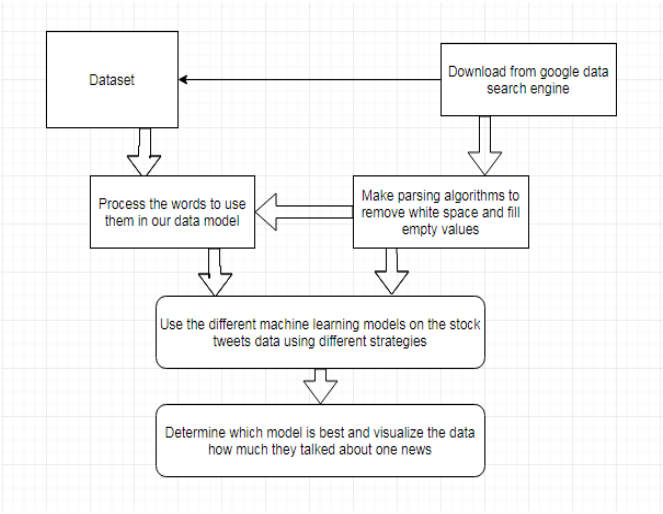


Fig 6 prepocessing steps

The dataset was collected from websites, and then it was processed by deleting duplicate data and empty columns one at a time. Preprocessing techniques were also applied, and the best method with the highest accuracy was identified in order to predict the best output, from which we could estimate the company price when it could move upward or downward.

3.METHODOLOGY

The stock market doesn't appear to be static, thus there are a lot of unanswered issues, which makes stock market forecasting look like a challenging subject at first. However, by suitably implementing machine learning techniques, one may train the system to learn from the data and make relevant inferences by contrasting the past and present data.

The information under examination, which included over a thousand records of stock prices and related metrics, was obtained from well-known corporations including Amazon, Google, and Microsoft. This dataset includes volume, low, high, open, close, date, and symbol divisions for daily stock price records made during the course of the year. Information on a single company was chosen for study and simulation. The Pandas module in Python was used to import the CSV dataset and convert it into a data frame.This made it possible to extract data for a single company by classifying the data based on the photo field. Following this acknowledgement as characteristic of The data was separate into training data and testing sets and processed using the Python sklearn module. For the test set, twenty percent of the entire dataset was kept. Although there are many machine learning models available, this study focuses on two of the most popular models and uses them to generate its predictions.

This research employs few key machine learning models to make predictions. They are

A. Decision Tree Model

Within machine learning, a decision tree stands as a technique capable of... generate both classification and regression models. The decision tree gets its name from the fact that it begins at the base, much like an upside-down tree, then branches off to display alternative outcomes.

A decision tree is a hierarchical decision-making model that illustrates a number of options and their possible consequences, including chance event outcomes, resource costs, and utility, using a tree-shaped framework. Algorithms with conditional control statements are demonstrated. Not only are decision trees extensively utilized in operations research, specifically in decision analysis to determine the best paths to take in order to achieve objectives, but they are also a key component of machine learning.

Decision trees, as well as their closely related impact diagrams, are useful visual and analytical aids for supporting decision-making processes. They assist in determining the anticipated values of various choices under consideration.

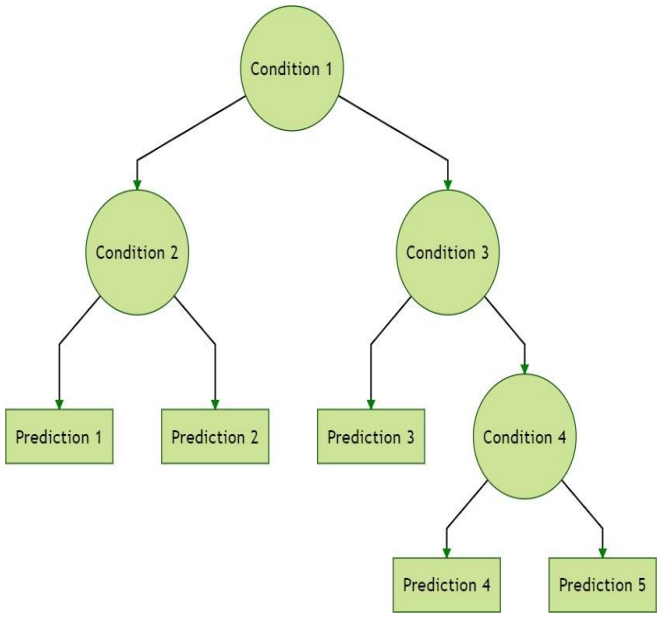


Fig 7 decision tree

Similar to a flowchart, a decision tree has leaf nodes that indicate algorithm outputs, branches that represent rules, and centre nodes that represent features. This is one of the most potent supervised machine learning approaches available, as it can be applied to both regression and classification problems. It also plays a crucial role in Random Forest, where it is taught using several training data sets, solidifying its standing as one of the most essential machine learning techniques. The decision tree can be transformed into a decision rule, in which the criteria along the route form an if clause and the contents of the leaf node decide the outcome. Theseregulations often have the following structure:

Once requirements 1, 2, and 3 are satisfied, the outcome is known. By building association rules and placing the target variable on the right-hand side, decision rules can be produced. These rules could potentially be causal or temporal relationships.

B. Model Based on Linear Regression

Regression models are often used to predict continuous values from a set of independent variables [5]. These models use a specified linear function to estimate continuous values. For example, one may anticipate the quantity of soil erosion depending on rainfall. Regression models explain the relationship between dependent and independent variables. Variables are determined by fitting a line through observed data points.

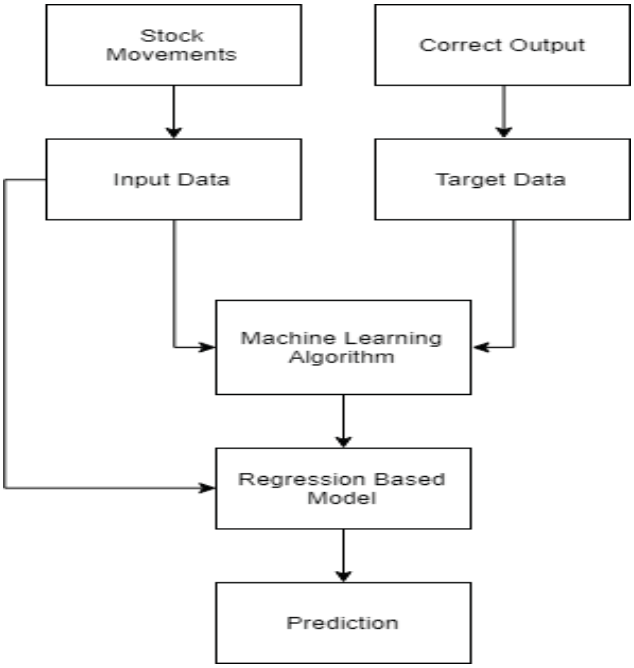


Fig. Flow Chart for Regression Based Model

Regression analysis as a method to find the value of a dependent variable, such as price prediction related to a specific stocks, based on given levels of independent values. It involves fit corectly a line to the gathered data to illustrate the association between these values.

One of the most important methods in regression analysis is linear regression, also referred to as linear least squares or ordinary least squares. When every independent variable varies by one unit, it calculates the average change in the dependent variable. Polynomials can also be utilized to combine interaction effects and simulate curvature.

Regression models are often used to predict continuous values from a set of independent variables [5]. These models use a specified linear function, generally written as V = a + bk + error, to predict continuous outcomes.
In this work, the gradient descent linear regression approach, shown in Figure 1, is used to predict accurate values by minimising the error function. Accurate predictions are generated after assessing the data with Linear Regression [6] using the specified equation. Volume, close, open, high, and low are among the variables included in the regression analysis. The generated projections are visually displayed to demonstrate the patterns of stock market values over time, and the confidence score is evaluated using the R-squared confidence test.

C. K-Nearest Neighbors (K-NN) Based Model

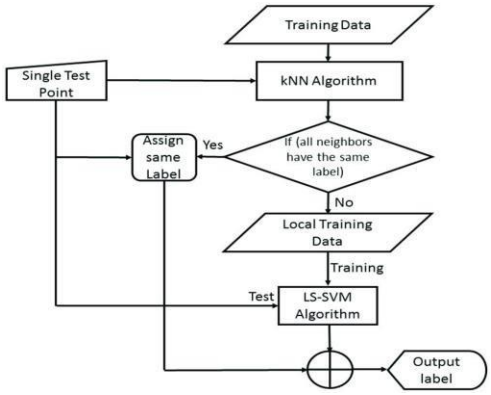


Fig. 5 K-NN Layers

One of the most advanced kinds of model networks is K-Nearest Neighbours (K-NN). In contrast to regression models, K-NN takes long-term dependencies into account since it keeps data from previous states. Its capacity to evaluate the correlation between recent and present data points is what sets it apart. This emphasises how crucial it is to comprehend how data changes over time in the context of K-NN.

The first motivation for employing this model in stock forecasting is because the forecasts raly heavily on data and often take into account the market's lengthy history [6]. In order to help the RNNs by keeping information for earlier stages, the LSTM controls error and improves prediction accuracy [7]. demonstrating its superior reliability over alternative approaches.

Because the stock market needs processing massive volumes of data, the learning rate may be reduced, which could result in relatively minor gradients with respect to the weight matrix.[8] This is the same as the problem with the Vanishing Gradient. K-NN prevents this from happening. The K-NN is composed of an input gate, an output gate, a forget gate, and a remembering cell. The gate-controlled value is retained by the cell for long-term propagation. In this study, two K-NN layers are stacked on top of one another to generate a sequential model with an output value of 256.

The input layer is composed of two layers: layer [0] and layer [1]. To minimize overfitting of the data and expedite the training process, nodes will be stopped at a predetermined value of 0.3 during the assessment stage. At last, the dense core layer is introduced, where each data is connected to all the other neurons in the layer that follows.

RESULTS AND DISCUSSION

Training and assessing the proposed method are based on the Amazon dataset. The dataset is split into training and testing sets once different machine learning techniques are chosen. Based on these divides, the program then returns the following results.

A. Decision tree

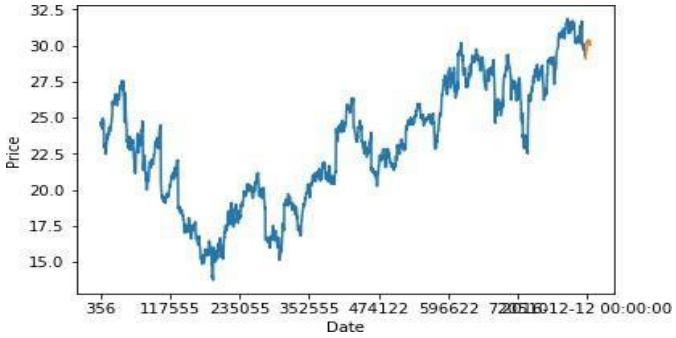


Fig 6 graphic representation of decision tree model

The graph (Fig. 6) above displays the data with 90 epochs and a batch size of 512. The score from the R-square test was 0.8625.which is confidence score

Based on above parameters the decision tree algorithm gives the best output in predicting the stock price whether in upside or downside in stock momentum it gives the 85% of accurecy in predicting of stock momentum

B. Regression model

The real trend is displayed in blue, and the prediction is displayed in red. These two lines' close closeness indicates,
how effective the model based on regression is.

When a hige period of time has increased, the prediction approximates the actual trend.The result of the model was T-score as 0.0106 and Mse gives (0.03 rmse) and test result is 0.0875. Less accurate than SVM models were models based on regression.

From the above the we compare the each and every method gives the various of formats of outputs and each will define various of conditions based on the algorithm performance this will be consider for the cross verification for forther valuation.

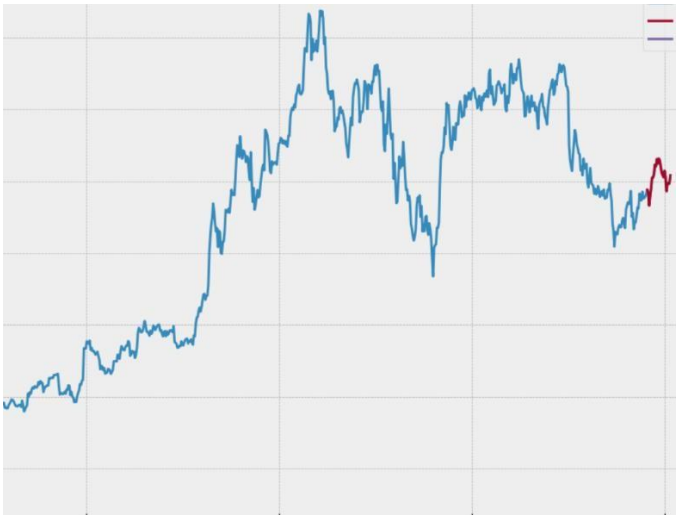


Fig.7 grapic representation of regression

C.K-Neighbors Regression

Nearest Neighbors-based regression is applicable when dealing with continuous data labels, as opposed to discrete variables. It determines the label for a query point by computing the mean of the labels belonging to its closest neighbors.

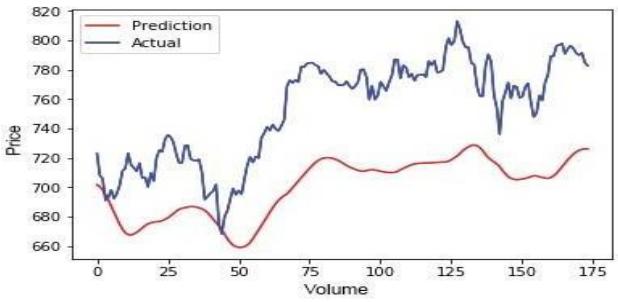


Fig7 graphic representation of k-neighbor regression

The K-NN technique is a basic, easy to done supervised learning algorithm that can tackle classification and regression model issue

In the k-nn algorithm as shown in the figure the blue line represent the how the price is moving and the red line represent the prediction of the algorithm which gives output based on our data set

Compare to the other algorithms the k-nn also gives the best accurecy while predicting the stock price it will cover the all vertice and key parameters while working on the data sets

Table-1 Accuracy of different algorithms

ALGORITHAMS	ACCURACY
K-NN	81%
Linear Regression	72%
K-neigh bours Regression	80%
Decision Tree	85%

The table represent the all the algorithams present in this project compare to each and every algoritham
Decision Tree gives the higest accurancy compare to other algorithm so we consider it and used to project the values.

CONCLUSION

This research was an attempt to use machine learning techniques to more reliably and accurately predict future stock values of a firm. The researchers' main contribution is the new K Neighbors Regressor Model application, which they used to calculate stock prices.

Both the K Neighbors Regressor and Regression model have yielded positive outcomes, showcasing enhanced efficiency. These techniques have shown an uptick in prediction accuracy, fueling optimism about the potential for machine learning to anticipate stock market movements with greater precision and effectiveness. Despite this, we are also considering the Decision tree model, given its superior accuracy rate of 85%.

By this we can conclude that based on the old data of the various stocks in listed market we can collect them and remove the waste and null values then by applying the various pre-processing techniques we can decrese the complexity of the data while handling with methodology.Then we start applying the methods on it we consider here decision tree which gives the higest accuracy in predicting the stock whether in upside or down side and based on the it we can conclude whether to invest or avoid such stocks while investing bby this we can filter most of the bad stocks for investing

The stock current price is taken by the user and predict the future price based on the model

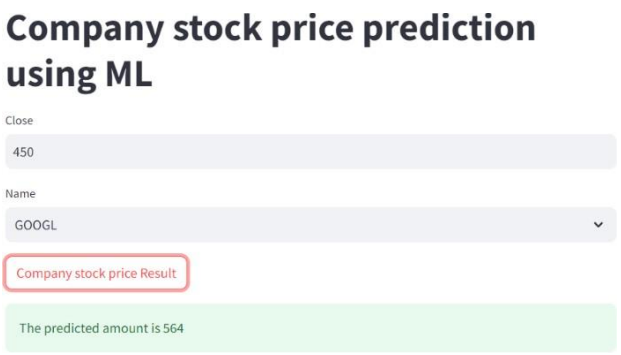


Fig 8 deployment

REFERENCES

1] "Stock market prediction using machine learning techniques," M. Usmani, S. H. Adil, K. Raza, and S. S. A. Ali, 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, 2016, pp. 322-327.

[2] et performance via the use of machine learning approaches," 2017 International Conference on Innovations in Computational and Electrical Engineering

[3] Y.Yaslan (2017) "Stock market direction prediction using deep neural networks," in Proceedings of the 25th Signal Processing and Communications Applications Conference (SIU), Antalya, pp. 1-4

[4] S. Waheed and A. Hanifa, "Utilising an enhanced neural network training algorithm for stock market prediction," pp. 1-4 in the 2nd International Conference on Electrical, Computer, and Telecommunication Engineering (ICECTE), Rajshahi, 2016.

[5] "Regression techniques for the predictio of stock price trend," by H. L. Siew and M. J. Nordin, 2012 International Conference on Statistics in Science, Business, and Engineering (ICSSBE), Langkawi, 2012, pp. 1-5

[6] V. Sujatha and S. M. Sundaram, "Stock index prediction using regression and neural network models under non normal conditions," INTERACT-2010, Chennai, 2010, pp. 59-63.

[7] Industrial Electronics and Applications (ICIEA), Wuhan, 2018, pp.2787-2790

[8] Chai and Y. Liu, "Applying long short term memory neural conditions," INTERACT-2010, Chennai, 2010, pp. 59-63.
networks for predicting stock closing price," 2017 8th IEEE International

[9] Beijing, 2017, pp. 575–578 in Conference on Software Engineering and Service Science (ICSESS). Bidirectional long-term stochastic volatility tree model for short- and long-term stock market prediction, K. A. Althelaya, E. M. El-Alfy, and S. Moha, 2018's 9th International Conference on Information and Communication Systems, Irbid, 2018,pp.151–156