

Efficient Stock Prediction Technique using ML

Hemachanderen R P¹, Akash S², Sanjay Roshan R³, Ms.Sangeerani Devi A⁴ and Ms.S.Hemavathi⁵

¹⁻³B.E CSE, Department of Computer Science And Engineering, Sri Sairam Engineering College, Chennai, Tamil Nadu, India
Email: e8cs022@sairamtap.edu.in , e8cs143@sairamtap.edu.in , e8cs030@sairamtap.edu.in

⁴⁻⁵Assistant Professor, Department of Computer Science And Engineering, Sri Sairam Engineering College, Chennai, Tamil Nadu, India

Email: sangeerani.cse@sairamtap.edu.in , hemavathi.cse@sairamtap.edu.in

Abstract—Stock market prediction helps people to determine the future movement of the stock value of a financial exchange. The Predicting method should have high accuracy and be effective in predicting the stock. Most of the Prediction System will not have high accuracy as they use some unsuitable Algorithms. Our Goal is to make a Stock predicting System which predicts the stock with high accuracy and in an effective way. We have used a Machine learning Algorithm called Support Vector Machine (SVM). This Algorithm gives higher accuracy in predicting stocks. We will Collect the stock Datasets of previous years and then the Pre-processing will be done. Data cleaning, Data transformation, Data selection comes under the Pre-Processing. After this process inputs will be passed to the Support Vector Machine (SVM) Algorithm and finally the prediction will be done.

I. INTRODUCTION

Nowadays, the stock market is becoming a new trend. The number of people following stocks has increased drastically. To help them we have developed this project to predict the stock of a particular company. This project will be helpful for them to get an insight into stock movement. A significant profit could be yielded by successfully predicting the future stock price. Support Vector Machines (SVM) algorithm is widely used for the prediction of stock prices and their movements. Our domain is mainly on machine learning. We wanted the model to understand and recognize the datasets. We concentrated on pre-processing the data, training the model and then predicting the stock value using the algorithm.

II. EXISTING METHODOLOGY

The existing work of prediction systems will be implemented using various machine learning algorithms. This system is a vital pattern for the Stakeholders (shareholders of a specific company) which gives the predicted stock value as this value helps the Stakeholder to make a decision when to buy the stock to attain the maximum profit.

III. PROPOSED METHODOLOGY

In our prediction system first, the data collection is done. The data collection sets have many companies stock price for the last five years. From this dataset then data is sent to data pre-processing. The pre-processing is carried out to get the proper dataset. After this process, the test train split is done. Then the model is trained. After this process SVM algorithm is applied. Then the Prediction will be carried out and the results will be

given. The Accuracy is slightly improved compared to the existing models. Because Support Vector Machine algorithm are models that are supervised learning that can analyse data for classification. Other machine learning algorithms have less than 70% accuracy but SVM algorithm gives more accuracy than any model.

IV. LITERATURE REVIEW

We have taken five papers as literature review for our references. First “Evaluating future stock value asset using machine learning” by R.Soujanya, AkshithGoud, AbishekBhandwalkar, AnilKumar. In this Prediction of stock is done by using Linear regression algorithm. Second “Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market” by Xianghui Yuan, Jin Yuan, Tianzhao Jian, Quart UL Ain. In this Prediction of stock is done by using Random Forest, ANN ect. Third paper “Predicting Stock Market Trends using Machine Learning Algorithms Via Continuous and Binary Data” by Mojtaba, Nabipour, Pooyan, Navyeri, Amir Mosavi. In this prediction system they have used many machine learning algorithm which includes random forest, knn, ann etc to predict the stock value. Forth paper “Efficacy of News Sentiment for Stock Market Prediction” by Sneh Kalra, Jay Shankar Prasad. In this 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. Fifth paper “Indices prediction of Bangladeshi stock by using time series forecasting and performance analysis” by Muhibur Rahman, Majumder, Imran Hossain, Mohammed d Kamrul Hassan. In this prediction of stock price is a complicated process due to randomness of stock price. In forecasting the price of stock, Researchers has used various algorithms.

V. SYSTEM ARCHITECTURE

The datasets are collected from the source. The Pandas Library helps us to pre-process the collected data. Therefore, we get the proper dataset. Then Scikit-learn to do the classification and regression procedures. After this process, the data is split and then it is trained. After the trained data is given to the model. Then the Support vector Machine (SVM) is applied and the prediction is done. Then the evaluation is done followed by this process. Finally, the Predicted results are shown. This is the System Architecture of our model.

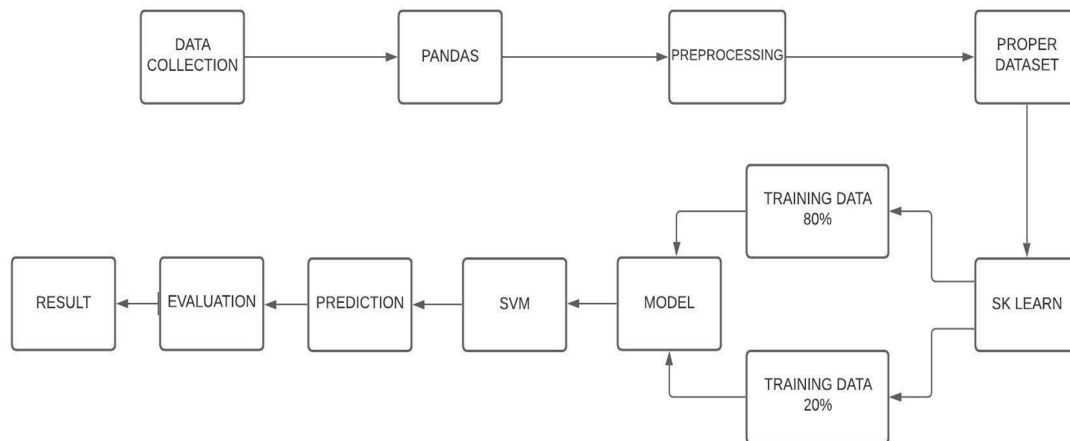


FIG-1

VI. DATASET COLLECTION

Data Collection is an important task for building a machine learning model. We collect the specific dataset based on requirements from the internet. The dataset contains some unwanted data also. So first we need to pre-process the data and obtain perfect data set for the algorithm.

VII. DATA PRE-PROCESSING

It is the gathering of task-related information based on some targeted variables to analyse and produce some valuable outcomes. However, some of the data may have to be processed to minimize data redundancy. Hence, it

```
#Get the stock data
#df = quandl.get("WIKI/AMZN")
df = web.DataReader('AAPL', data_source='yahoo', start='2015-10-01', end='2021-04-16')
print(df.head(100))
```

Date	High	Low	Open	Close	Volume	Adj Close
2015-09-30	27.885000	27.182501	27.542500	27.575001	265892000.0	25.400600
2015-10-01	27.405001	26.827499	27.267500	27.395000	255716400.0	25.234791
2015-10-02	27.752501	26.887501	27.002501	27.594999	232079200.0	25.419024
2015-10-05	27.842501	27.267500	27.469999	27.695000	208258800.0	25.511141
2015-10-06	27.934999	27.442499	27.657499	27.827499	192787200.0	25.633194
...	...	...	...	...	...	...
2016-02-17	24.552500	24.037500	24.167500	24.530001	179452800.0	22.815565
2016-02-18	24.722500	24.022499	24.709999	24.065001	156084000.0	22.383070
2016-02-19	24.190001	23.950001	24.000000	24.010000	141496800.0	22.331917
2016-02-22	24.225000	23.980000	24.077499	24.219999	137123200.0	22.527231
2016-02-23	24.125000	23.637501	24.100000	23.672501	127770400.0	22.018002

[100 rows x 6 columns]

FIG-2

is a must to process the data before analysing it and coming to the results. Data pre-processing is carried out by processes like data cleaning, data transformation, data selection. The data cleaning process includes filling in the null values, bringing the data to a similar scale, identify or remove outliers, and resolve inconsistencies in the data. The data transformation process includes aggregation, generalization, transformation which improves the accuracy and quality of the data. Data selection includes some methods or functions which allow us to select useful data for our system.

```
In [7]: ### Create the independent data set (X) #####
# Convert the dataframe to a numpy array
X = np.array(df.drop(['Prediction'],1))

#Remove the Last 'n' rows
X = X[:-forecast_out]
print(X)

[[ 27.57500076]
 [ 27.39500046]
 [ 27.59499931]
 ...
 [134.99000549]
 [133.94000244]
 [137.38999939]]

In [8]: ### Create the dependent data set (y) #####
# Convert the dataframe to a numpy array (All of the values including the NaN's)
y = np.array(df['Prediction'])
# Get all of the y values except the Last 'n' rows
y = y[:-forecast_out]
print(y)

[ 29.02750015  28.93000031  28.08499908 ... 124.76000214 120.52999878
 119.98999786]

In [9]: # Split the data into 80% training and 20% testing
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=44)

In [10]: # Create and train the Support Vector Machine (Regressor)
svr_rbf = SVR(kernel='rbf', C=1e3, gamma=0.1)
svr_rbf.fit(x_train, y_train)

Out[10]: SVR(C=1000.0, gamma=0.1)
```

FIG-3

VIII. PREDICTION

```
# Print support vector regressor model predictions for the next 'n' days
svm_prediction = svr_rbf.predict(x_forecast)
print(svm_prediction)

[122.2276258 121.27198433 127.71500119 132.22114352 133.78634306
 132.35107192 132.52894392 121.4144295 126.2495383 125.26145423
 114.3128709 113.27688818 111.40932296 131.59599445 131.60717948
 131.4614242 111.65294884 132.00253043 132.3725033 131.67697717
 122.74089096 131.58251303 132.52252542 131.98896537 131.58801765
 120.56821219 111.81944493 113.33925249 131.91498159 132.51356465]
```

FIG-4

IX. CONCLUSION

One of the required solutions is to predict the stock efficiently. A mistake could make a huge loss in predicting stocks, so predicting the stocks should be done efficiently. As a solution for this, we have used Support vector Machines (SVM) that predict with high accuracy and efficiency. Even though SVM predicts the stock efficiently, the user should have some basic idea about the stock market. The prediction may not be exact but the user can get more insights using our model.

REFERENCES

- [1] R.Soujanya, AkshithGoud, AbishekBhandwalkar, AnilKumar “Evaluating future stock value asset using machine learning”, materials today: proceedings volume 33, part 7, 2020
- [2] Xianghui Yuan, Jin Yuan, Tianzhao Jian, Quart UL Ain “Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market”, IEEE Access, 24 January 2020
- [3] Mojtaba, Nabipour, Pooyan, Navyeri, Amir Mosavi “Predicting Stock Market Trends using Machine Learning Algorithms Via Continuous and Binary Data”, IEEE Access, 12 August 2020
- [4] Sneha Kalra, Jay Shankar Prasad “Efficacy of News Sentiment for Stock Market Prediction”, International conference on Machine learning, Big Data, Cloud Computing and Parallel Computing, 2019
- [5] Muhibur Rahman, Majumder, Imran Hossain, Mohammed d Kamrul Hassan “Indices prediction of Bangladeshi stock by using time series forecasting and performance analysis”, International Conference on Electrical, Computer and Communication Engineering, 2019.