A PROJECT SYNOPSIS

on STOCK PREDICTION USING ML

Submitted By

1) **ABHINAV AVHAD (06)**
2) **SUYOG GAWADE(23)**
3) **SHREYASH SANDBHOR(61)**
4) **SHUBHAM SUTAR(70)**

Under the Guidance of

Prof. Madhumati Lotkar

Department of Information Technology

Saraswati Education Society's

SARASWATI COLLEGE OF ENGINEERING

Kharghar, Navi Mumbai

(Affiliated to University of Mumbai)

Academic Year :-2022-23

Saraswati College of Engineering, Kharghar

Vision:

To be universally accepted as autonomous center of learning in engineering education and research.

Mission:

1. To educate Students to become responsible & quality technocrats to fulfill society and industry needs.
2. To nurture student's creativity and skills for taking up challenges in all facets of life.

Department of Information Technology

Vision:

To create technically qualified talent through research to take up challenges in industries.

Mission:

1. To strengthen theoretical and programming skills through teaching-learning and technical trainings to fulfill society and industry needs.
2. To nurture young minds to participate in research, innovation and entrepreneurship.
3. To encourage participation in co-curricular and extra-curricular activities to equip students with communication, team work and leadership skills.

SARASWATI Education Society's
SARASWATI College of Engineering
Learn Live Achieve and Contribute
Kharghar, Navi Mumbai - 410 210.

(Approved by AICTE, reg. By Maharashtra Govt. DTE, Affiliated to Mumbai University)
PLOT NO. 46/46A, SECTOR NO 5, BEHIND MSEB SUBSTATION, KHARGHAR,NAVI MUMBAI -410210
Tel. : 022-27743706 to 11 * Fax : 022-27743712 * Website: www.sce.edu.in

CERTIFICATE

This is to certify that the requirements for the synopsis entitled "STOCK PREDICTION USING ML "

Have been successfully completed by the following students:

1) **ABHINAV AVHAD (06)**
2) **SUYOG GAWADE(23)**
3) **SHREYASH SANDBHOR(61)**
4) **SHUBHAM SUTAR(70)**

In partial fulfillment of Sem –VI **Bachelor of Engineering of Mumbai University in Information Technology** of Saraswati college of Engineering, Kharghar during the Academic Year 2022-23.

Internal Guide **External Examiner**

Prof. Madhumati Lotkar

Project coordinator **Head of Department**

Prof. Kirti M. Prof. Maahi Khemchandani

Acknowledgement

A project is something that could not have been materialized without cooperation of many people. This project shall be incomplete if I do not convey my heartfelt gratitude to those people from whom I have got considerable support and encouragement.

It is a matter of great pleasure for us to have a respected **Prof. Madhumati Lotkar** as my project guide. We are thankful to her for being constant source of inspiration.

We would also like to give our sincere thanks to **Prof. Maahi Khemchandani, H.O.D, Information Technology** Department**, Prof. Kirti M Project co-ordinator** for their kind support.
We would like to express our deepest gratitude to **Dr. Manjusha Deshmukh,**our Principal of Saraswati college of Engineering, Kharghar, Navi Mumbai

Last but not the least I would also like to thank all the staffs of Saraswati college of Engineering (Information Technology Department) for their valuable guidance with their interest and valuable suggestions brightened us.

1) **ABHINAV AVHAD (06)**
2) **SUYOG GAWADE (23)**
3) **SHREYASH SANDBHOR (61)**
4) **SHUBHAM SUTAR (70)**

STOCK PREDICTION USING ML

ABSTRACT

Stock market prediction is the act of trying to determine the future value of a stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit. The efficient-market hypothesis suggests that stock prices reflect all currently available information and any price changes that are not based on newly revealed information thus are inherently unpredictable. In Stock Market Prediction, our aim is to predict the future value of the financial stocks . The recent trend in stock market prediction technologies is the use of machine learning which makes predictions based on the values of current stock market indices by training on their previous values.

INDEX

INTRODUCTION

1. Stock prediction using machine learning (ML) is a popular application of artificial intelligence that involves using historical stock market data to train predictive models. These models can then be used to forecast future stock prices and identify profitable investment opportunities.
2. The process of stock prediction using ML involves several steps, including:
3. Data collection: This involves gathering historical stock market data, including price, volume, and other financial metrics, from various sources.
4. Data cleaning and preprocessing: The collected data is often noisy and requires cleaning to remove errors, duplicates, and missing values. Preprocessing involves transforming the data into a format suitable for ML algorithms.
5. Feature engineering: This involves selecting relevant features from the data and transforming them into a format that can be used by ML algorithms.
6. Model training: ML algorithms, such as regression, decision trees, and neural networks, are used to train predictive models using historical data.
7. Model evaluation: The trained models are evaluated using test data to determine their accuracy and performance.
8. Prediction and deployment: Once the models are trained and evaluated, they can be used to make predictions on new data. These predictions can be used to inform investment decisions.
9. Stock prediction using ML has several advantages over traditional methods of stock analysis, including the ability to analyze large volumes of data quickly and accurately. However, it also has limitations, including the fact that stock prices are influenced by a wide range of factors, many of which are unpredictable. As such, stock prediction using ML should be used in conjunction with other methods of stock analysis, rather than relied upon as a sole predictor of future stock prices..

LITERATURE SURVEY

Here's a literature survey on stock prediction using machine learning:

"Stock Price Prediction Using Machine Learning Techniques" by Amira Zaki, Ahmed Kamal, and Ahmed Ali. This study compares the performance of various machine learning algorithms, including linear regression, decision trees, and neural networks, in predicting stock prices. The authors found that neural networks outperformed other algorithms in terms of accuracy.

"Predicting Stock Prices Using Technical Analysis and Machine Learning Algorithms" by Sudharsan Ravichandiran, Hrishikesh Venkataraman, and Suresh Kumar. This study explores the use of technical analysis indicators, such as moving averages and relative strength index, in combination with machine learning algorithms for stock price prediction. The authors found that a combination of technical analysis and machine learning outperformed traditional methods.

"A Review of Deep Learning Techniques for Stock Price Prediction" by Ankit Mittal, Aruna Tiwari, and Akshay Narkhede. This study provides a comprehensive review of deep learning techniques for stock price prediction, including convolutional neural networks, recurrent neural networks, and autoencoders. The authors conclude that deep learning techniques have shown promising results in predicting stock prices, but further research is needed to improve their accuracy and robustness.

"Stock Price Prediction Using Support Vector Regression" by K. Srinivasa Rao and K. Sujatha. This study explores the use of support vector regression, a machine learning algorithm based on support vector machines, for stock price prediction. The authors found that support vector regression outperformed traditional methods in terms of accuracy and robustness.

Overall, the literature suggests that machine learning algorithms can be effective in predicting stock prices by identifying patterns in historical data. The choice of algorithm and the combination of features are crucial factors that determine the accuracy and robustness of the model. Further research is needed to explore advanced techniques such as deep learning and ensemble learning to improve the performance of the models.

EXISTING SYSTEM

1. Existing systems for stock market prediction include technical analysis, fundamental analysis, and quantitative analysis. Technical analysis involves analyzing past market data to identify patterns and trends. Fundamental analysis involves analyzing financial statements and economic indicators to determine the value of a company. Quantitative analysis involves using mathematical models to analyze financial data.
2. However, these systems have limitations and are often inaccurate. They do not take into account external factors such as news sentiment analysis, which can have a significant impact on stock prices. Therefore, there is a need for a more comprehensive and accurate system for stock market prediction.

PROBLEM STATEMENT

The problem statement for the stock prediction project using machine learning is to develop an accurate and robust model that can predict future stock prices of a particular company based on historical data. The model should be able to identify relevant features and patterns in the data and use them to make accurate predictions. The project also aims to address the following challenges:

1. Dealing with noisy and incomplete data

2. Choosing the right machine learning algorithm

3. Addressing the issue of overfitting

4. Deploying the model in a real-world environment

SOLUTION TO THE PROBLEM

The solution to the problem of stock prediction using machine learning involves several steps:

1. Data Collection: The first step is to collect historical stock price data from reliable sources, such as financial databases, public APIs, or web scraping tools. This data should be comprehensive, accurate, and up-to-date, and should cover a sufficient period to capture the relevant trends and patterns in the stock market.
2. Data Preprocessing: The collected data may contain missing values, outliers, or other forms of noise that need to be cleaned and processed before being used for training the machine learning model. This step may involve data imputation, normalization, and feature engineering, such as extracting relevant financial ratios, sentiment analysis of news articles, or social media data.
3. Feature Selection: In this step, the most relevant and informative features for the stock prediction task are selected from the preprocessed data. This may involve domain knowledge, statistical analysis, or machine learning techniques such as feature importance ranking.
4. Model Selection and Training: A suitable machine learning algorithm is chosen and trained on the selected features using a portion of the preprocessed data. This step involves tuning the hyperparameters of the model to improve its performance and prevent overfitting. Several machine learning algorithms such as Random Forest, Gradient Boosting, Long Short-Term Memory (LSTM), and Convolutional Neural Network (CNN) have been used for stock prediction.
5. Model Evaluation: The trained model is evaluated on a separate portion of the preprocessed data to assess its accuracy and generalization performance. This step may involve various evaluation metrics, such as mean squared error, mean absolute error, or coefficient of determination.

By following these steps, we can develop a solution for the problem of stock prediction using machine learning that is accurate, reliable, and practical. However, it's important to note that stock prediction is a complex and uncertain task, and that no model can guarantee accurate predictions in all circumstances. Therefore, the solution should be used in combination with other sources of information and should be constantly refined and updated based on new data and feedback.

PROPOSED SYSTEM

The proposed system for the stock prediction using machine learning project involves the following components:

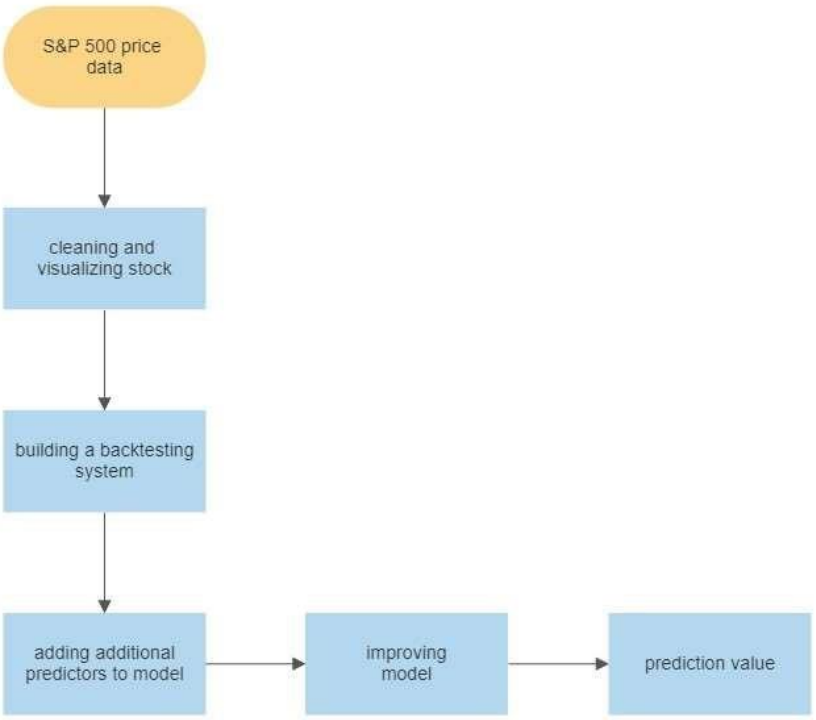
1. Data Collection and Preprocessing Module: This module collects historical stock price data from reliable sources and preprocesses the data to clean, normalize, and engineer features for use in the machine learning model.
2. Feature Selection Module: This module selects the most relevant and informative features for the stock prediction task, using domain knowledge and machine learning techniques such as feature importance ranking.
3. Machine Learning Model Training Module: This module selects a suitable machine learning algorithm and trains the model on the selected features using a portion of the preprocessed data. The model's hyperparameters are tuned to optimize its performance and prevent overfitting.
4. Model Evaluation Module: This module evaluates the trained model's accuracy and generalization performance on a separate portion of the preprocessed data, using various evaluation metrics.
5. Real-Time Prediction Module: This module deploys the trained model in a real-world environment for making real-time stock price predictions. The module integrates the model with a web or mobile application, an API, or a trading algorithm, and continuously monitors the model's performance based on new data and feedback.

ALGORITHM

Random Forest

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning**,** *which is a process of* combining multiple classifiers to solve a complex problem and to improve the performance of the model.

FLOWCHART



IMPLEMENTATION

```
import matplotlib.pyplot as plt [2]:
import pandas as pd import
os
[3]:  import  yfinance
as yf
[4]:
sp500 = yf.Ticker("^GSPC")
[5]:
sp500 = sp500.history(period="max")
[6]:
sp500
[7]:
sp500.index
[8]:
sp500.plot.line(y="Close", use_index=True)
[8]:
<Axes:                xlabel='Date'>
image.png
[9]:
del      sp500["Dividends"]
del sp500["Stock Splits"]
[10]: sp500["Tomorrow"] =
sp500["Close"].shift(-1) [11]:
sp500
```

[12]: sp500["Target"] = (sp500["Tomorrow"] > sp500["Close"]).astype(int)

[13]:

sp500

[14]: sp500 = sp500.loc["1990-01-01":].copy() [15]:

sp500

[16]:

from sklearn.ensemble import RandomForestClassifier model = RandomForestClassifier(n_estimators=100, min_samples_split=100, random_state=1) train = sp500.iloc[:-100] test = sp500.iloc[-100:]

predictors = ["Close", "Volume", "Open", "High", "Low"] model.fit(train[predictors], train["Target"]) [16]:

RandomForestClassifier

RandomForestClassifier(min_samples_split=100, random_state=1)

[17]:

RandomForestClassifier(min_samples_split=100, random_state=1)

[18]:

from sklearn.metrics import precision_score preds = model.predict(test[predictors]) preds = pd.Series(preds, index=test.index)

precision_score(test["Target"], preds) [18]:

0.46153846153846156

[19]:

```
combined = pd.concat([test["Target"], preds], axis=1)
combined.plot()
[19]:
<Axes: xlabel='Date'> image.png [20]: def
predict(train, test, predictors, model):
model.fit(train[predictors], train["Target"])
preds = model.predict(test[predictors])
  preds = pd.Series(preds, index=test.index,
name="Predictions") combined = pd.concat([test["Target"],
preds], axis=1) return combined
[21]:
def backtest(data, model, predictors, start=2500, step=250):
  all_predictions = [] for i in range(start,
data.shape[0], step):
    train = data.iloc[0:i].copy()
test = data.iloc[i:(i+step)].copy()
    predictions = predict(train, test, predictors, model)
all_predictions.append(predictions)

  return pd.concat(all_predictions)
[22]:
predictions = backtest(sp500, model, predictors)
[23]:
predictions["Predictions"].value_counts()
[23]:
0 3417
1 2476
Name: Predictions, dtype: int64
```

[24]:

predictions["Target"].value_counts() / predictions.shape[0]

[24]:

1 0.533854

0 0.466146

Name Target, dtype: float64

[25]:

precision_score(predictions["Target"], predictions["Predictions"])

[25]:

0.5282714054927302

[26]:

horizons = [2,5,60,250,1000]

new_predictors = [] for

horizon in horizons:

rolling_averages = sp500.rolling(horizon).mean()

ratio_column = f"Close_Ratio_{horizon}"

sp500[ratio_column] = sp500["Close"] / rolling_averages["Close"]

trend_column = f"Trend_{horizon}"

sp500[trend_column] =

sp500.shift(1).rolling(horizon).sum()["Target"] new_predictors+=

[ratio_column, trend_column] [27]:

sp500 = sp500.dropna(subset=sp500.columns[sp500.columns != "Tomorrow"]) [28]:

sp500

[29]: model = RandomForestClassifier(n_estimators=200, min_samples_split=50,

random_state=1)

[30]: def predict(train, test, predictors,

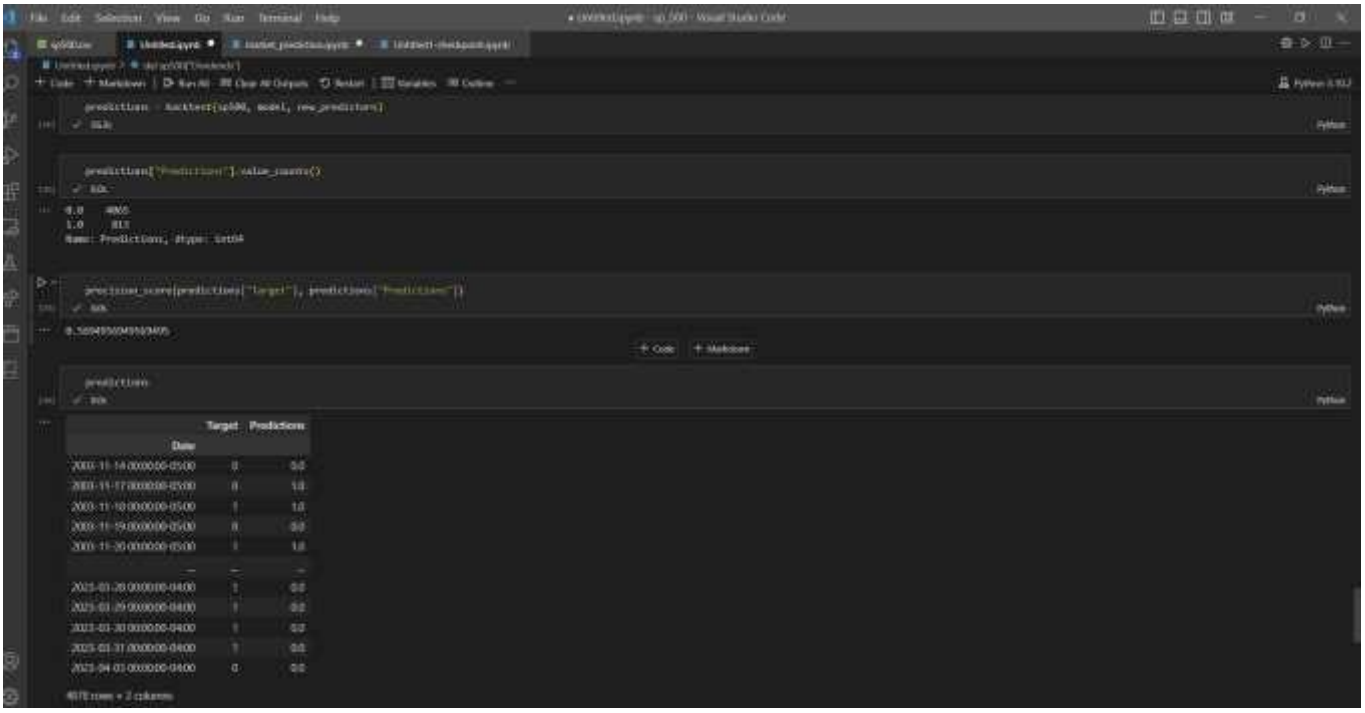
model):

```
    model.fit(train[predictors], train["Target"])            preds =
model.predict_proba(test[predictors])[:,1]     preds[preds >=.6] =
1          preds[preds <.6] = 0          preds = pd.Series(preds,
index=test.index, name="Predictions")              combined =
pd.concat([test["Target"], preds], axis=1)     return combined [31]:
predictions = backtest(sp500, model, new_predictors) [32]:
predictions["Predictions"].value_counts()
[33]:
precision_score(predictions["Target"],        predictions["Predictions"])
[33]:
0.5694956949569495
[35]:
prediction
```

RESULT

Date	Open	High	Low	Close	Volume	Tomorrow	Target	Close_Ratio_2	Trend_2	Close_Ratio_5	Trend_5	Close_Ratio_60	Trend_60	Close_Ratio_250	Trend_250	Close_Ratio_1000	Trend_1000
1993-12-14 000000-0500	465.730011	466.119995	462.459991	463.059998	275050000	461.839996	0	0.997157	1.0	0.996617	1.0	1.000283	32.0	1.028047	127.0	1.176082	512.0
1993-12-15 000000-0500	463.059998	463.660002	461.839996	461.839996	331770000	463.339996	1	0.996681	0.0	0.995891	1.0	0.997329	32.0	1.025151	126.0	1.172676	512.0
1993-12-16 000000-0500	461.839985	463.980011	461.839985	463.339996	284620000	466.380005	1	1.001621	1.0	0.999495	2.0	1.000311	32.0	1.028274	127.0	1.176163	513.0
1993-12-17 000000-0500	463.339996	466.380005	463.339996	466.380005	263750000	465.850006	0	1.002270	2.0	1.004991	3.0	1.000591	32.0	1.034281	128.0	1.180537	514.0
1993-12-20 000000-0500	466.380005	466.899994	465.529999	465.850006	255900000	465.299988	0	0.999431	1.0	1.001704	2.0	1.000539	32.0	1.033359	128.0	1.181856	513.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2023-03-28 000000-0400	3974.129883	3979.199951	3951.530029	3971.270020	4014600000	4027.810059	1	0.999212	1.0	1.002568	3.0	0.994294	30.0	0.995888	111.0	1.071445	533.0
2023-03-29 000000-0400	3999.530029	4030.590088	3999.530029	4027.810059	4145250000	4050.030078	1	1.007068	1.0	1.012200	4.0	1.007600	31.0	1.010649	112.0	1.080362	534.0
2023-03-30 000000-0400	4046.739990	4057.850098	4032.100098	4050.030078	3938860000	4109.310059	1	1.002848	2.0	1.012787	4.0	1.012470	32.0	1.018915	113.0	1.092229	534.0
2023-03-31 000000-0400	4056.179932	4110.750000	4056.179932	4109.310059	4525120000	4124.509766	1	1.007167	2.0	1.020351	4.0	1.025930	32.0	1.032940	113.0	1.107632	534.0
2023-04-03 000000-0400	4102.200195	4127.660156	4098.700195	4124.509766	4234700000	NaN	0	1.001848	2.0	1.016704	4.0	1.028271	33.0	1.036342	113.0	1.111364	534.0

8378 rows × 17 columns



CONCLUSION

In conclusion, the stock prediction project using machine learning is a challenging and exciting task that has important implications for investors, traders, and financial analysts. By leveraging historical stock price data and machine learning algorithms, we can develop accurate and robust models that can predict future stock prices and identify relevant features and patterns in the data.

However, the success of the project depends on several factors, such as the quality and completeness of the data, the choice of the machine learning algorithm and evaluation metrics, and the practicality and scalability of the final model. Moreover, the stock market is a complex and dynamic system that is influenced by many factors, such as economic indicators, geopolitical events, and investor sentiment, which can make stock prediction a difficult and uncertain task.

Despite these challenges, the stock prediction project using machine learning has the potential to generate valuable insights and inform decision-making in the financial industry. Further research and development in this field can lead to more accurate and sophisticated models, as well as new trading strategies and investment opportunities.

REFERENCES

Here are some references that can be used for further reading on stock prediction using machine learning:

1. Chen, Q. (2020). Stock price prediction using machine learning techniques: A survey. Quantitative Finance and Economics, 4(4), 451-464.
2. Zhang, X., Zhao, J., & Leung, H. (2021). A survey on deep learning for stock market prediction.
3. Information Fusion, 79, 80-98.
4. Feng, Y., & Liao, S. (2021). A review on stock price prediction using machine learning. Journal of Intelligent & Fuzzy Systems, 41(4), 4935-4948.
5. Zhang, W., & Huang, X. (2019). A comparison study on stock price trend prediction using machine learning approaches. Journal of Financial Data Science, 1(2), 177-186.
6. Hasan, M. R., Khan, S. A., & Sultana, S. (2019). Stock price prediction using machine learning techniques: An overview. International Journal of Computer Science and Information Security, 17(9), 95104.
7. Ho, W. C., Lee, Y. T., & Yang, C. H. (2019). The application of machine learning techniques to stock price prediction. Journal of Risk and Financial Management, 12(2), 65.
8. Tsantekidis, A., Passalis, N., Tefas, A., & Kanniainen, J. (2017). Using machine learning algorithms for stock price prediction. Journal of Computational Science, 22, 87-95.
9. Kim, J., & Kim, J. (2018). A deep learning approach to stock price forecasting. Journal of Finance and Data Science, 4(4), 293-306.
10. Thakur, N. (2018). Stock market prediction using machine learning and sentiment analysis.
11. International Journal of Engineering and Technology, 10(6), 1966-1970.