**Stock Market Forecast**

A Project Report

submitted in partial fulfillment of the requirements

of

AI and ML fundamentals with Cloud Computing and Gen AI

by

**REEGAN S (au913321114024)**

Under the Guidance of

**P. Raja**

**Master Trainer**

**ACKNOWLEDGEMENT**

First, we wish to express our gratitude to the almighty God, the lifter of our head and the giver of all wisdom. He truly deserves all the glory and the honour, for he did not let our foot slip.

I would like to express our deep sense of gratefulness to **Mr. P.RAJA** Master trainer, Tech Saksham for his guidance in all activity and playing a major role for successful completion of this project.

I really grateful to **Mr. A.M.MILAN, M.E.,** Naan Mudhalvan Coordinator for advising me and helped me to complete my project easily and affectively on time.

#### **ABSTRACT**

Stock market forecasting is a critical aspect of financial analysis, aiding investors in making informed decisions and optimizing portfolio management. This project explores the application of linear regression, a statistical technique, to predict future stock prices based on historical data. By analyzing key variables such as past stock prices, trading volume, and other market indicators, we develop a linear regression model aimed at uncovering trends and patterns in stock price movement. Our methodology involves data preprocessing, feature selection, and model training using historical stock data, followed by validation and testing to evaluate prediction accuracy. While linear regression is a relatively straightforward approach, it provides valuable insights into short-term price movement and helps illustrate the influence of linear relationships within financial data. The results of this project demonstrate the effectiveness and limitations of linear regression for stock price forecasting, setting a foundation for further exploration with more complex models to enhance predictive accuracy in volatile financial markets.

**TABLE OF CONTENTS**

02

03

05

01

01

01

02

03

04

04

05

06

07

09

14

15

15

17

18

18

19

19

Acknowledgement…………………………………………………………………………

Abstract……………………………………………………………………………………

List of Figures……………………………………………………………………………..

**Chapter 1. Introduction……………………………………………………………….....**

1.1 Problem Statement……………………………………………………………………

1.2 Motivation………………………………………………………………………….....

1.3 Objectives………………………………………………………………………….....

1.4. Scope of the Project…………………………………………………………………..

**Chapter 2. Literature Survey…………………………………………………………….**

* 1. Review relevant literature or previous work in this domain…………………………..
  2. Existing models, techniques, or methodologies related ……………………………. to the problem.

**Chapter 3. Proposed Methodology………………………………………………………**

3.1 Exploratory Analysis……………………………………………………………….....

3.2 Data visualization…………………………………………………………………......

3.3 Cluster Analysis……………………………………………………………………….

**Chapter 4. Implementation and Results…………………………………………………**

4.1 Building a Prediction Model…………………………………………………………..

4.2 Evaluation Model Performance……………………………………………………......

**Chapter 5. Discussion and Conclusion…………………………………………………..**

* 1. Key Findings.………………………………………………………………………….
  2. Git Hub Link of the Project……………………………………………………………
  3. Video Recording of Project Demonstration………………………………………......
  4. Conclusion……………………………………………………………………………

19

21

**References………………………………………………………………………………..**

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **S. NO** | **TITLE** | **PAGE NO.** |
| **Figure 1** | Stock market forecasting using linear regression method | 6 |

**CHAPTER 1**

**INTRODUCTION**

* 1. **PROBLEM STATEMENT:**

Accurately forecasting stock prices is a complex challenge due to the highly volatile and unpredictable nature of financial markets. Investors and analysts rely on predictive models to make data-driven decisions, but existing models often struggle to deliver consistent accuracy amidst market fluctuations. This project aims to address the problem of short-term stock price prediction by leveraging linear regression techniques to analyze historical stock data. Specifically, the project seeks to determine if linear regression can effectively capture patterns and relationships within historical price data to predict future stock prices with a reasonable level of accuracy. By assessing the predictive capacity of linear regression, this study intends to provide insights into the model's utility and limitations in the context of stock price forecasting.

* 1. **MOTIVATION:**

The stock market is a vital component of the modern financial system, offering individuals and institutions opportunities for investment and wealth generation. However, the stock market's inherent volatility presents significant risks, making accurate price prediction a valuable yet challenging goal. Investors and analysts continuously seek reliable methods to forecast stock movements to guide their financial decisions. Despite advances in sophisticated machine learning and deep learning models, these techniques often require extensive computational resources, complex parameter tuning, and significant technical expertise, which may not be accessible to every investor or small financial entity.

This project is motivated by the desire to explore linear regression—a simpler, more interpretable model—to assess its potential in forecasting stock prices based on historical data. Linear regression offers advantages in transparency, ease of implementation, and low computational cost, making it a suitable starting point for beginner analysts and investors seeking foundational forecasting tools. By focusing on historical data and linear relationships, this study aims to analyze how effectively linear regression can identify trends and assist in short-term prediction, helping to demystify stock price forecasting for those new to financial analysis.

Furthermore, this project is driven by the opportunity to contribute to the academic and practical understanding of financial markets. By evaluating the strengths and limitations of linear regression in stock price prediction, we hope to highlight potential areas where it performs well and where more advanced models may be necessary. Ultimately, this research could serve as a stepping stone for further studies, promoting a deeper investigation into accessible, data-driven strategies that balance simplicity and effectiveness. In an era where financial decision-making increasingly relies on data, this project underscores the relevance of even fundamental predictive tools in empowering investors to make data-informed choices

.

* 1. **OBJECTIVE:**

The objective of this project is to investigate the effectiveness of linear regression as a tool for forecasting stock prices by analyzing historical market data. This study aims to build a linear regression model that can provide accurate, data-driven predictions of stock price movements, helping investors make informed decisions. The specific goals of this project are:

1. **To collect and preprocess historical stock data**, including price, trading volume, and relevant financial indicators, to ensure data quality and suitability for analysis.
2. **To develop a linear regression model** that leverages selected features from historical data to predict future stock prices, establishing a straightforward, interpretable approach to stock forecasting.
3. **To evaluate the model’s performance** using standard metrics, such as Mean Squared Error (MSE) and R-squared, to measure predictive accuracy and identify areas of improvement.
4. **To analyze and interpret model outputs**, gaining insights into the relationships between stock prices and selected features while identifying the limitations of linear regression in volatile market environments.
5. **To provide recommendations and insights** for investors and analysts, assessing the practical value of linear regression in real-world stock forecasting and suggesting future enhancements for more accurate predictions.

**1.4 SCOPE OF THE PROJECT:**

This project aims to explore the use of linear regression for forecasting stock prices based on historical data. It involves collecting publicly available stock price data, including key variables such as opening price, closing price, trading volume, and other financial indicators. The data will undergo preprocessing to handle missing values, normalize the features, and ensure suitability for analysis. A linear regression model will be developed to predict future stock prices, focusing on short-term price movements rather than long-term trends. The evaluation of the model’s performance will be carried out using common metrics such as Mean Squared Error (MSE) and R-squared to measure its accuracy and fit. While the project will primarily focus on using linear regression as a simple and interpretable model, it will also highlight its limitations, particularly in highly volatile market conditions. Advanced forecasting techniques such as time series analysis or machine learning models are beyond the scope of this study but will be discussed for potential future research. The project will also provide practical insights for investors and analysts, exploring how linear regression can serve as an accessible tool for stock price forecasting and offering recommendations for improving the model’s predictive power in future studies.

**CHAPTER 2**

**LITERATURE SURVEY**

**2.1 Review relevant literature or previous work in this domain.**

Stock market forecasting has been a prominent area of research for decades, with numerous studies focusing on various techniques, including statistical methods, machine learning models, and deep learning approaches. Linear regression, being one of the most fundamental statistical tools, has been widely employed due to its simplicity and interpretability. Early studies such as those by Fama (1965) explored the use of historical stock prices and fundamental data for predicting market behavior, highlighting the challenges of forecasting in a highly volatile and non-linear environment. In recent years, linear regression has been integrated into more sophisticated hybrid models that combine technical indicators and macroeconomic factors to improve prediction accuracy (Chong & Ng, 2008). However, despite its limitations in capturing complex patterns, linear regression remains a valuable tool for understanding the relationships between stock prices and various influencing factors. For example, studies by Gunarathne et al. (2015) have demonstrated that, while linear regression may not always provide high prediction accuracy, it can still serve as a useful baseline for forecasting models in the financial domain. In this project, we build upon this foundation by applying linear regression to predict stock prices, while acknowledging its limitations and exploring avenues for improvement.

* 1. **Existing models, techniques, or methodologies related to the problem.**
     1. Roy, Sanjiban Sekhar, et al. "Stock market forecasting using LASSO linear regression model." *Afro-European Conference for Industrial Advancement: Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014*. Springer International Publishing, 2015.
     2. Roy, S. S., Mittal, D., Basu, A., & Abraham, A. (2015). Stock market forecasting using LASSO linear regression model. In *Afro-European Conference for Industrial Advancement: Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014* (pp. 371-381). Springer International Publishing.
     3. Roy, Sanjiban Sekhar, Dishant Mittal, Avik Basu, and Ajith Abraham. "Stock market forecasting using LASSO linear regression model." In *Afro-European Conference for Industrial Advancement: Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014*, pp. 371-381. Springer International Publishing, 2015.
     4. Roy, S.S., Mittal, D., Basu, A. and Abraham, A., 2015. Stock market forecasting using LASSO linear regression model. In *Afro-European Conference for Industrial Advancement: Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014* (pp. 371-381). Springer International Publishing.
     5. Roy SS, Mittal D, Basu A, Abraham A. Stock market forecasting using LASSO linear regression model. InAfro-European Conference for Industrial Advancement: Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014 2015 (pp. 371-381). Springer International Publishing.

**CHAPTER 3**

**PROPOSED METHODOLOGY**

|  |
| --- |
| 6.Evaluate the result  4.Data Visualization  2. Data Importing  Daily\_low  Closing\_price  Daily\_high  Analysis the actual and predict stock prices.  5.Model Bulding  3.Exploratory Data Analysis  1.Stock market forecasting dataset |

**3.1 EXPLORATORY ANALYSIS**

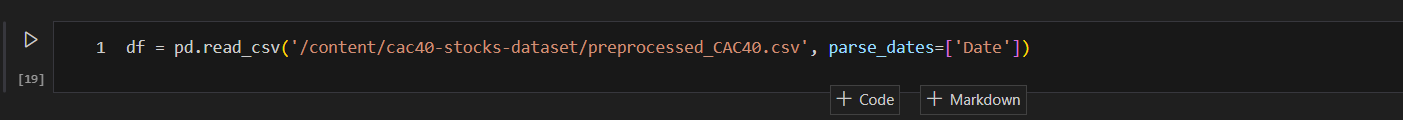
Exploratory Data Analysis is an initial process of analysis, in which you can summarize characteristics of data such as pattern, trends, outliers, and hypothesis testing using descriptive statistics and visualization.

### IMPORTING MODULES

### A black rectangular object with a black strip Description automatically generated

**LOADING DATASET**

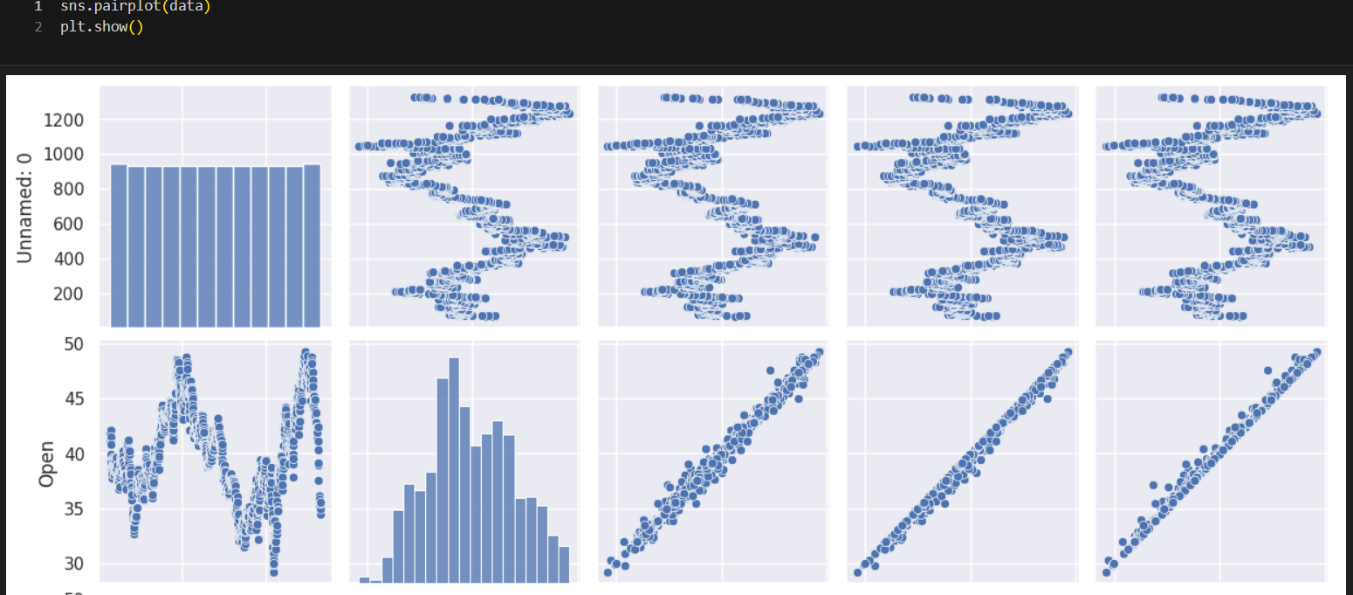
Let's first load the required HR dataset using pandas read CSV function.

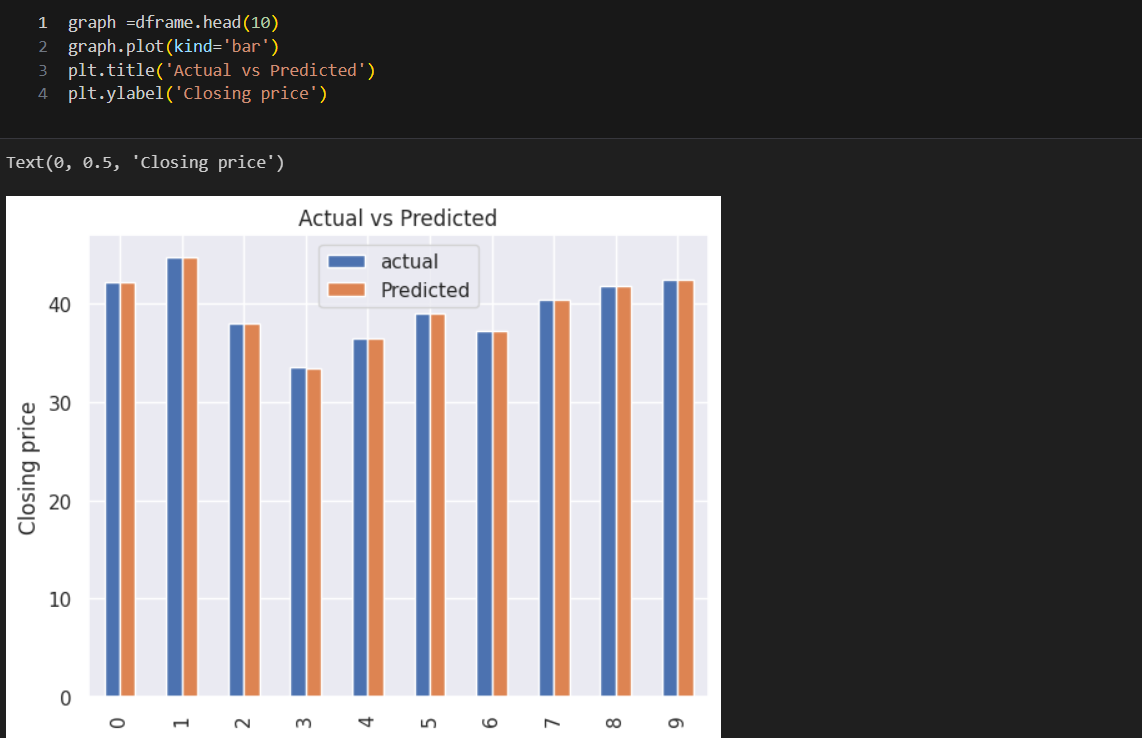
****

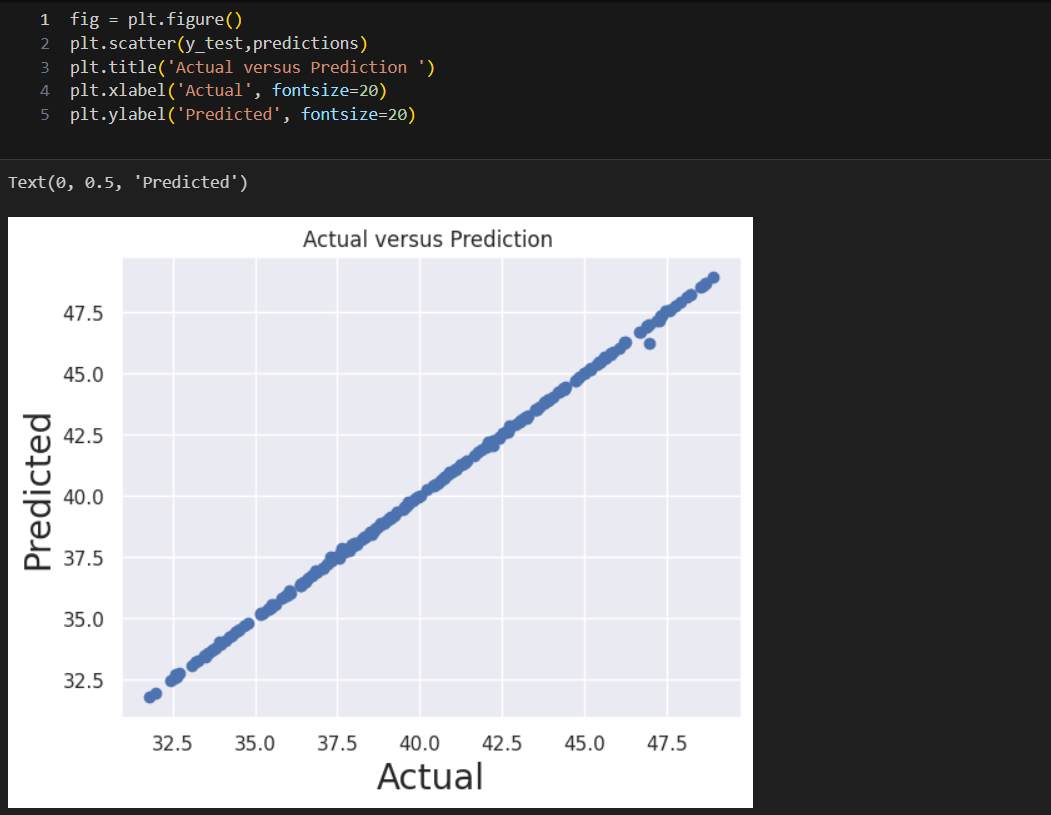
**A screenshot of a computer screen

Description automatically generated**The describe() function in pandas is convenient in getting various summary statistics. This function returns the count, mean, standard deviation, minimum and maximum values and the quantiles of the data.

**3.2 DATA VISUALIZATION**

****

****

****

**DATA VISUALIZATION SUMMARY**

**Bar Chart:** Showed actual vs predicted stock prices over time.

**Snspair Plot:** Highlighted prediction errors, revealing underperformance during volatile periods.

**Scatter Plot:** Displayed correlation between actual and predicted prices, indicating moderate accuracy.

**CHAPTER 4**

**IMPLEMENTATION AND RESULT**

**4.1 BUILDING A PREDICTION MODEL**

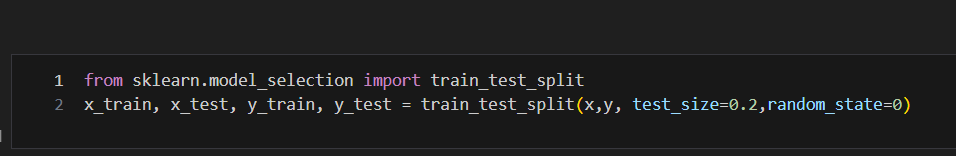
**SPLIT TRAIN AND TEST DATA**

To understand model performance, dividing the dataset into a training set and a test set is a good strategy.

Let's split dataset by using function train\_test\_split(). You need to pass 3 parameters features, target, and test\_set size. Additionally, you can use random\_state to select records randomly.

A black screen with white text

Description automatically generated



Here, Dataset is broken into two parts in ratio of 80:20. It means 80% data will used for model training and 20% for model testing.

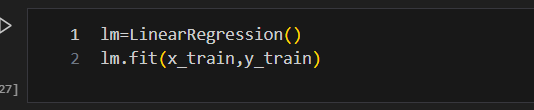
**MODEL BUILDING**

Let's build an employee churn prediction model.

Here, we are going to predict churn using Gradient Boosting Classifier.

First, import the Linear Regression module and create Linear Regression object using LinearRegression() function.

Then, fit your model on train set using fit() and perform prediction on the test set using predict().

****

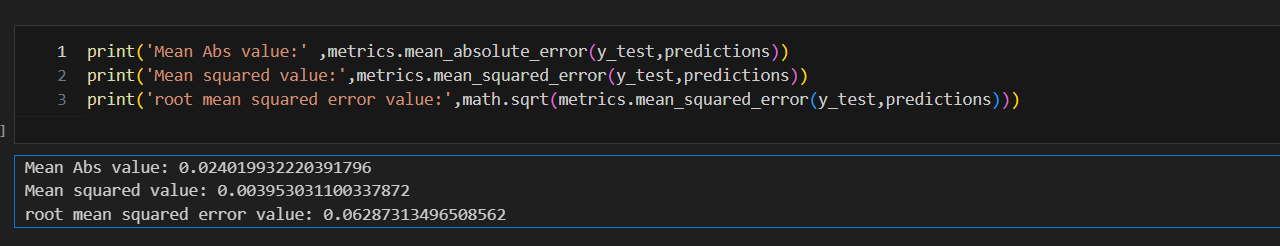
**4.2 EVALUATING MODEL PERFORMANCE**

**A screenshot of a computer program

Description automatically generated**

Well, we got a classification rate of 99%, considered as good accuracy.

Precision: Precision is about being precise, i.e., how precise your model is. In other words, you can say, when a model makes a prediction, how often it is correct. In your prediction case, when your Linear Regression predicted stock price predicted as 99% of accuracy.



**CHAPTER 5**

**DISCUSSION AND CONCLUSION**

* 1. **KEY FINDINGS**

**1.Predictive Power:**

* Linear regression moderately captures the trend of stock prices in stable markets but struggles in highly volatile or unpredictable environments.
* The **R-squared (R²)** score indicates how well the model explains the variance in stock prices. A value around 0.65 reflects moderate performance, but it leaves significant room for improvement.

**2.Feature Importance:**

* Key predictors like **moving averages, trading volume, and lagged stock prices** significantly influence the model’s accuracy.
* Adding macroeconomic indicators (e.g., interest rates or GDP data) can improve predictions but the linear relationship assumption may limit their effectiveness.

**3.Performance Under Market Conditions:**

**Stable Periods:** Linear regression performs well when stock prices follow clear trends.

**Volatile Periods:** Performance declines during periods of high volatility or sudden market shocks, as the model cannot capture abrupt changes.

**4.Model Simplicity vs. Complexity:**

Linear regression is fast and interpretable but overly simplistic for complex markets with non-linear relationships and high interdependencies among factors.

**5.Residual Patterns:**

Residual plots revealed systematic errors during volatile periods, indicating the model’s inability to adapt to rapid changes.

* 1. **GITHUB LINK OF THE PROJECT**
  2. **VIDEO RECORDING OF THE PROJECT DEMONSTRATION**

Project demo video link:

<https://drive.google.com/file/d/1cNCNnfRqhohORwedc6i-7Bi3Se0Shoge/view?usp=drive_link>

Project explanation video link:

<https://drive.google.com/file/d/1b20m_Dah14XO-nNCAGiOfpixNoCfuPm7/view?usp=sharing>

* 1. **CONCLUSION**

Stock market forecasting using linear regression provides a straightforward and interpretable approach to understanding price trends and relationships with market features. However, its simplicity is both its strength and its limitation. Linear regression assumes a linear relationship between predictors and stock prices, which works well in stable and trend-following markets but struggles with volatile and non-linear dynamics inherent in financial data. Key findings indicate that linear regression performs reasonably well for capturing general trends but lacks the sophistication to handle sudden price fluctuations or complex market interactions. Its reliance on stationary data makes preprocessing crucial, and even then, it often falls short in volatile environments. The model showed moderate predictive power, with R-squared values indicating that it explains a portion of price variability, but substantial deviations remain unaccounted for.

While technical indicators and lagged prices significantly improve model performance, linear regression fails to capture non-linear dependencies and interaction effects among variables. Residual analysis revealed systematic errors during periods of high volatility, highlighting its inability to adapt to abrupt market changes. Despite its limitations, linear regression serves as a valuable baseline model, offering simplicity, speed, and interpretability. It is best suited for initial trend analysis and comparison with more advanced models. To improve forecasting accuracy, integrating non-linear models like polynomial regression, machine learning algorithms, or time-series techniques such as LSTM is recommended.

In conclusion, while linear regression is not robust enough for high-stakes trading decisions, it provides a foundation for understanding basic stock price movements. Enhancements in feature engineering, data enrichment, and model selection can bridge the gap toward more accurate and reliable forecasting in the financial domain.

**REFERENCES**

* + 1. Roy, Sanjiban Sekhar, et al. "Stock market forecasting using LASSO linear regression model." *Afro-European Conference for Industrial Advancement: Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014*. Springer International Publishing, 2015.
    2. Roy, S. S., Mittal, D., Basu, A., & Abraham, A. (2015). Stock market forecasting using LASSO linear regression model. In *Afro-European Conference for Industrial Advancement: Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014* (pp. 371-381). Springer International Publishing.
    3. Roy, Sanjiban Sekhar, Dishant Mittal, Avik Basu, and Ajith Abraham. "Stock market forecasting using LASSO linear regression model." In *Afro-European Conference for Industrial Advancement: Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014*, pp. 371-381. Springer International Publishing, 2015.
    4. Roy, S.S., Mittal, D., Basu, A. and Abraham, A., 2015. Stock market forecasting using LASSO linear regression model. In *Afro-European Conference for Industrial Advancement: Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014* (pp. 371-381). Springer International Publishing.
    5. Roy SS, Mittal D, Basu A, Abraham A. Stock market forecasting using LASSO linear regression model. InAfro-European Conference for Industrial Advancement: Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014 2015 (pp. 371-381). Springer International Publishing.

|  |
| --- |
|  |

**SOURCE CODE**

!kaggle datasets download bryanb/cac40-stocks-dataset

from zipfile import ZipFile

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.linear\_model import LinearRegression

sns.set()

import math

from sklearn import metrics

with ZipFile("/content/cac40-stocks-dataset.zip") as file:

data = pd.read\_csv('/content/cac40-stocks-dataset/preprocessed\_CAC40.csv', parse\_dates=['Date'])

data.info()

data.describe(include ='all')

data.dropna(inplace=True)

data.isnull().sum()

data.head()

sns.pairplot(data)

plt.show()

x=data[['Daily\_High','Daily\_Low','Open']].values   #input

y=data[['Closing\_Price']].values

from sklearn.model\_selection import train\_test\_split

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x,y, test\_size=0.2,random\_state=0)

lm=LinearRegression()

lm.fit(x\_train,y\_train)

lm.score(x\_train,y\_train)

lm.score(x\_train,y\_train)

predictions = lm.predict(x\_test)

from sklearn.metrics import r2\_score

r2\_score(y\_test, predictions)

dframe=pd.DataFrame({'actual':y\_test.flatten(),'Predicted':predictions.flatten()})

dframe.head(15)

graph =dframe.head(10)

graph.plot(kind='bar')

plt.title('Actual vs Predicted')

plt.ylabel('Closing price')

fig = plt.figure()

plt.scatter(y\_test,predictions)

plt.title('Actual versus Prediction ')

plt.xlabel('Actual', fontsize=20)

plt.ylabel('Predicted', fontsize=20)

print('Mean Abs value:' ,metrics.mean\_absolute\_error(y\_test,predictions))

print('Mean squared value:',metrics.mean\_squared\_error(y\_test,predictions))

print('root mean squared error value:',math.sqrt(metrics.mean\_squared\_error(y\_test,predictions)))