A PROJECT REPORT

on

"STOCK MARKET PRICE PREDICTION"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
INFORMATION TECHNOLOGY

BY

SANU VERMA	21051509
K TANVI ANANYA	21051523
UTKRIST JAISWAL	21051526

UNDER THE GUIDANCE OF
DR. MOHIT RANJAN PANDA

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAR, ODISHA - 751024

April 2024

A PROJECT REPORT

on

"STOCK MARKET PRICE PREDICTION"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN INFORMATION TECHNOLOGY

BY

SANU VERMA	21051509
K TANVI ANANYA	21051523
UTKRIST JAISWAL	21051526

UNDER THE GUIDANCE OF

DR. MOHIT RANJAN PANDA

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAE, ODISHA -751024

April 2024

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certify that the project entitled

"STOCK MARKET PRICE PREDICTION"

submitted by

SANU VERMA	21051509
K TANVI ANANYA	21051523
UTKRIST JAISWAL	21051526

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering OR Information Technology) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2024, under our guidance.

Date: / /

MOHIT RANJAN PANDA
Project Guide

Acknowledgements

We are profoundly grateful to **DR MOHIT RANJAN PANDA** of **Affiliation** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion. .....

SANU VERMA
K TANVI ANANYA
UTKRIST JAISWAL

ABSTRACT

This project focuses on predicting stock market prices using machine learning techniques. It employs historical stock data of companies like Google, Apple, and Amazon to analyze patterns and trends in their stock prices. The project utilizes Python libraries such as pandas, numpy, yfinance, and matplotlib for data manipulation, visualization, and analysis.

The project begins by collecting historical stock data from January 2019 to December 2020 for Google, Apple, and Amazon. It explores various aspects such as closing prices, sales volume, and daily returns to gain insights into the behavior of each company's stock. Machine learning models, specifically Random Forest Regressors, are trained and evaluated for each company to predict future stock prices. The evaluation metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) score to assess the performance of the models.

Keywords: Machine Learning, Stock Market Prediction, Historical Data Analysis, Python, Random Forest Regressor.

Contents

List of Figures

Chapter 1

Introduction

In the realm of finance, particularly in stock market trading, the ability to accurately predict stock prices is of paramount importance. This project endeavors to explore machine learning techniques to predict stock market prices, focusing on companies like Google (GOOG), Apple (AAPL), and Amazon (AMZN). The project's significance lies in its potential to offer insights into future stock trends, aiding investors, financial analysts, and traders in making informed decisions.

Figure 1.1: STOCK MARKET DATA VISUALS

Chapter 2

Basic Concepts/ Literature Review

Stock market price prediction is a critical area of research and application within the field of finance. It involves the use of various tools and techniques, including machine learning algorithms, to forecast the future prices of stocks traded in financial markets. The literature in this domain encompasses a wide range of methodologies and approaches aimed at enhancing the accuracy and reliability of stock price predictions.

2.1 Historical Data Analysis…….

One fundamental aspect of stock market price prediction involves the analysis of historical market data. Researchers often leverage historical stock price data obtained from sources such as Yahoo Finance to identify patterns, trends, and correlations that may influence future price movements. Exploratory data analysis (EDA) techniques, including data visualization and statistical analysis, are commonly employed to gain insights into the behavior of individual stocks and broader market trends.

2.2 Machine Learning Techniques…….

Machine learning techniques play a pivotal role in stock market price prediction due to their ability to uncover complex patterns and relationships within large datasets. Various machine learning algorithms, such as Random Forest Regression, Support Vector Machines (SVM), and Long Short-Term Memory (LSTM) networks, have been applied to predict stock prices based on historical market data. These algorithms offer different advantages and trade-offs in terms of predictive accuracy, computational efficiency, and interpretability.

2.3 Evaluation Metrics…….

The evaluation of stock price prediction models requires the use of appropriate performance metrics to assess their effectiveness. Commonly used evaluation metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2). These metrics provide quantitative measures of the models' predictive accuracy and help researchers compare the performance of different algorithms.

2.4 Challenges and Future Directions…….

Despite significant advancements in stock market price prediction, several challenges remain, including the presence of market noise, unpredictable events, and the inherent complexity of financial markets. Future research directions may involve the development of advanced machine learning models, the integration of alternative data sources (e.g., news sentiment analysis, social media data), and the exploration of ensemble learning techniques to improve prediction accuracy and robustness.

Chapter 3

Problem Statement / Requirement Specifications

The problem addressed in this project is the prediction of stock market prices for a selected set of companies. Stock market prediction is a complex and challenging task due to the inherent volatility and unpredictability of financial markets. The aim is to develop machine learning models capable of accurately forecasting future stock prices based on historical market data.

3.1 Project Planning

Steps to be followed while planning project development:

1. Define Objectives: Clearly define the objectives and goals of the project, including the scope, target outcomes, and deliverables.

2. Gather Requirements: Collect and analyze the requirements of the users or stakeholders. Define the functionalities and features to be developed.

3. Create Project Plan: Develop a detailed project plan outlining the tasks, timelines, resources, and dependencies. Use project management tools such as Gantt charts or Kanban boards to organize and track progress.

4. Allocate Resources: Identify and allocate the necessary resources, including human resources, budget, equipment, and software tools required for project execution.

5. Risk Assessment: Identify potential risks and challenges that may impact project success. Develop mitigation strategies and contingency plans to address these risks effectively.

6. Define Milestones: Establish key milestones and checkpoints to monitor progress and ensure timely completion of project phases.

7. Communication Plan: Define a communication plan to facilitate effective communication among team members, stakeholders, and project managers.

Establish regular meetings and reporting mechanisms.

8. Quality Assurance: Implement quality assurance processes to ensure that project deliverables meet the specified standards and requirements. Conduct regular reviews and testing to identify and rectify any issues or defects.

9. Documentation: Maintain comprehensive documentation throughout the project lifecycle, including project plans, requirements, design documents, test cases, and user manuals.

10. Review and Feedback: Seek feedback from stakeholders at various stages of the project to validate requirements, address concerns, and incorporate changes as needed.

3.2 Project Analysis

After the requirements are collected or the problem statement is conceptualized, the project analysis involves:

- Requirement Analysis: Conduct a detailed analysis of the gathered requirements to ensure clarity, completeness, and consistency. Identify any ambiguities, contradictions, or missing information that may impact project execution.

- Feasibility Study: Evaluate the feasibility of implementing the project based on factors such as technical feasibility, economic viability, and organizational capabilities.

- Risk Analysis: Assess potential risks and uncertainties associated with the project and develop strategies to mitigate these risks effectively. Consider factors such as technology risks, market risks, and regulatory risks.

- Resource Analysis: Analyze the availability and allocation of resources required for project implementation, including human resources, financial resources, and infrastructure.

- Impact Analysis: Evaluate the potential impact of the project on stakeholders, business processes, and organizational goals. Identify potential benefits, challenges, and trade-offs associated with project implementation.

- Cost-Benefit Analysis: Conduct a cost-benefit analysis to assess the financial implications of the project, including upfront costs, ongoing expenses, and expected returns on investment.

expected returns on investment.

- Stakeholder Analysis: Identify key stakeholders involved in or affected by the project and analyze their interests, expectations, and influence levels. Develop strategies to engage and manage stakeholders effectively.

3.3 System Design

3.3.1 Design Constraints

- Software Environment: The project will be developed using Python programming language and relevant libraries such as pandas, numpy, scikit-learn, and matplotlib for data analysis, machine learning modeling, and visualization.

- Hardware Environment: The project can be executed on standard computer hardware with sufficient computational resources, memory, and storage capacity.

- Experimental Setup: The project does not require any specific experimental setup or environmental setup beyond standard computing equipment.Software Environment: The project will be developed using Python programming language and relevant libraries such as pandas, numpy, scikit-learn, and matplotlib for data analysis, machine learning modeling, and visualization.

3.3.2 System Architecture **OR** Block Diagram

The system architecture will consist of data collection, preprocessing, feature engineering, model development, model evaluation, and deployment stages. Data flows sequentially through these stages, with each stage performing specific tasks to process and analyze the data.

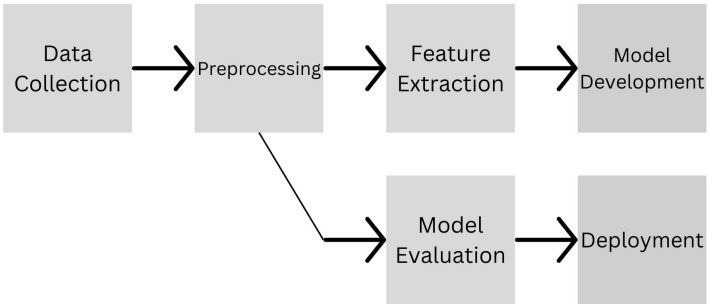

Figure 3.1: BLOCK DIAGRAM

Chapter 4

Implementation

4.1 Methodology OR Proposal

Details about the steps adopted for competing the project work.For the stock market price prediction project, the following methodology was adopted:

1. Data Collection: Historical stock market data for companies like Google (GOOG), Apple (AAPL), and Amazon (AMZN) was collected using the Yahoo Finance API (yfinance).

2. Data Preprocessing: The collected data was preprocessed to handle missing values, perform data transformation, and ensure consistency across the datasets.

3. Exploratory Data Analysis (EDA): Exploratory data analysis techniques were applied to gain insights into the data, understand its distribution, identify patterns, and visualize key metrics such as closing prices and trading volumes.

4. Feature Engineering: Features were engineered from the raw data to capture relevant information that could be used to train machine learning models. This involved calculating daily returns, computing moving averages, and extracting other relevant financial indicators.

5. Model Selection: Machine learning models, particularly Random Forest Regressors, were chosen for their ability to handle complex relationships in the data and make accurate predictions.

6. Model Training and Evaluation: The selected models were trained on the historical data and evaluated using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) to assess their performance.

7. Hyperparameter Tuning: Hyperparameters of the models were tuned using techniques like RandomizedSearchCV to optimize their performance.

4.2 Result Analysis OR Screenshots

```
Train Results:
==================
Mean Absolute Error (MAE) on train set: 26.99594568323207 %
Mean Squared Error (MSE) on train set: 16.131631890668796 %
Root Mean Squared Error (RMSE) on train set: 40.164202831213764 %
R-squared (R2) on train set: 99.94441719820409 %

Test Results:
==================
Mean Absolute Error (MAE) on test set: 57.078343553289514 %
Mean Squared Error (MSE) on test set: 69.92261292614363 %
Root Mean Squared Error (RMSE) on test set: 83.61974224197515 %
R-squared (R2) on test set: 99.76398968397044 %
```

```
Train Results:
==================
Mean Absolute Error (MAE) on train set: 31.932229454134713 %
Mean Squared Error (MSE) on train set: 24.62566492162726 %
Root Mean Squared Error (RMSE) on train set: 49.62425306402834 %
R-squared (R2) on train set: 99.91066230186667 %

Test Results:
==================
Mean Absolute Error (MAE) on test set: 69.65337253250715 %
Mean Squared Error (MSE) on test set: 100.0 %
Root Mean Squared Error (RMSE) on test set: 100.0 %
R-squared (R2) on test set: 99.64612496213788 %
```

```
Train Results:
==================
Mean Absolute Error (MAE) on train set: 29.36139151490768 %
Mean Squared Error (MSE) on train set: 17.85401921700812 %
Root Mean Squared Error (RMSE) on train set: 42.254016633934484 %
R-squared (R2) on train set: 99.95620229034223 %

Test Results:
==================
Mean Absolute Error (MAE) on test set: 66.82362699606182 %
Mean Squared Error (MSE) on test set: 94.43691391860098 %
Root Mean Squared Error (RMSE) on test set: 97.17865707993755 %
R-squared (R2) on test set: 99.76478789046298 %
```

Chapter 5

Conclusion and Future Scope

5.1 Conclusion

In this project, we explored the use of machine learning algorithms for predicting stock market prices based on historical data of three major tech companies: Google, Apple, and Amazon. We began by collecting and preprocessing the stock market data, including features such as open, high, low, close prices, adjusted close prices, and trading volume. Exploratory data analysis was conducted to understand the distributions and relationships between different variables. We observed correlations between stock returns and closing prices, which provided valuable insights into market trends.

Following the data exploration, we implemented machine learning models, specifically Random Forest Regressors, to predict the closing prices of the three companies. The models were trained on historical data and evaluated using various metrics such as mean absolute error, mean squared error, root mean squared error, and R-squared score. The results indicated promising performance of the models, achieving high R-squared scores on both training and testing datasets.

5.2 Future Scope

Despite the satisfactory performance of the machine learning models, there are several areas for future exploration and improvement:

- Feature Engineering: Experiment with additional features derived from technical indicators, news sentiment analysis, or macroeconomic indicators to enhance the predictive power of the models.

- Model Tuning: Explore hyperparameter tuning techniques such as grid search or Bayesian optimization to fine-tune the model parameters and

improve performance further.

- Ensemble Methods: Investigate the use of ensemble methods such as stacking or boosting to combine multiple models and potentially improve prediction accuracy.

- Time-Series Analysis: Implement advanced time-series analysis techniques such as ARIMA, SARIMA, or Prophet to capture seasonality and trends in stock market data more effectively.

- Real-Time Prediction: Develop a real-time prediction system that continuously updates models with the latest data to provide timely and accurate stock price forecasts.

- Risk Management: Integrate risk management strategies into the prediction models to quantify and mitigate potential financial risks associated with stock market investments.

References

[1]S. M. Metev and V. P. Veiko, Laser Assisted Microtechnology, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[2] Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[3] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," IEEE Electron Device Lett., vol. 20, pp. 569–571, Nov. 1999.

[4] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in Proc. ECOC'00, 2000, paper 11.3.4, p. 109.

[5] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.

[6] (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

[7] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/

[8] https://pypi.org/project/yfinance/- Library API

[9] https://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics -Packages

[10] https://scikit-learn.org/stable/modules/classes.html#module-sklearn.model_selection -Packages

[11] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html#sklearn.ensemble.RandomForestRegressor -Ensemble