PHASE 3 PROJECT

GOOGLE STOCKS PREDICTION MODEL

MIRRIAM MUMBUA

INTRODUCTION

Objective

- In this project, my primary goal is to predict stock prices using machine learning. By analyzing historical data, I aim to build models that can classify whether the stock price will rise or fall. This is a crucial task for making informed investment decisions, as accurate predictions can help investors better manage risks and identify opportunities in the stock market.

Significance

- The significance of this project lies in its potential impact on financial decision-making. Predicting stock prices is a key component in developing effective investment strategies, allowing investors to make more informed choices. By leveraging machine learning models, we're not just relying on historical trends but also incorporating complex patterns and relationships within the data that might not be immediately obvious.
- For businesses, accurate stock price predictions can influence investment strategies, portfolio management, and risk assessment. On a broader scale, this project exemplifies how data-driven decision-making can reduce uncertainty in financial markets, potentially leading to better economic outcomes. Moreover, it highlights the growing importance of technology in finance, showcasing how machine learning can be applied to real-world problems to achieve tangible benefits.

Business Problem

Understandinge Stocks

- In the fast-paced world of stock trading, making informed decisions about buying and selling stocks is crucial. Investors and financial analysts constantly seek ways to predict stock price movements to maximize returns and minimize risks. The business problem at hand is to develop a reliable method to predict whether a stock's closing price will rise or fall on the following trading day. Accurate predictions can help investors make better decisions, optimize their portfolios, and ultimately increase profitability.

Goal:

The goal is to use machine learning to classify whether the stock price will increase or decrease, based on historical data. This classification is critical for making quick and informed decisions in a volatile market. By predicting the direction of price movements, investors can strategically decide when to buy or sell stocks, thereby enhancing their investment strategies.

DATA UNDERSTANDING

DATA SOURCES

I got this data set from Kaggle. It is a google stocks data set from 2010 - 2023. It has originally 7 features:

- **Date:** The date of the record (YYYY-MM-DD).
- **Open:** Opening price of the stock.
- **High:** Highest price of the stock for the day.
- **Low:** Lowest price of the stock for the day.
- **Close:** Closing price of the stock.
- **Adj Close:** Adjusted closing price.
- **Volume:** Number of shares traded.

DATA ANALYSIS

DATA CLEANING AND PREPROCESSING

Adding needed columns in the dataframe.

- I had to perform some basic maths on specific columns and create new columns to store the results so I can get useful data e.g percentage change, dayof the week, moving averages

Transforming Data Formats:

- Date Parsing: Transformed release dates into datetime format for time series analysis.

DATA ANALYSIS (EDA)PROCESS

- Created various plots (e.g., histograms, scatter plots) to identify patterns and trends.
- Computed summary statistics (mean, median, standard deviation) to understand data distribution.
- Examined relationships between all features.

MODELING

- **Model Selection and Tuning**: I utilized Gradient Boosting, a robust and widely-used machine learning algorithm, to address the classification problem. To optimize the model's performance, I conducted hyperparameter tuning using GridSearchCV. Through this process, I identified the best combination of parameters that balanced model complexity and accuracy:
- **Learning Rate**: 0.01
- **Max Depth**: 3
- **Min Samples Leaf**: 4
- **Min Samples Split**: 10
- **Number of Estimators**: 50
- **Subsample**: 1.0
- These parameters were carefully selected to enhance the model’s predictive capability while avoiding overfitting, ensuring that the model generalizes well to unseen data.

EVALUATION

The tuned Gradient Boosting model achieved an accuracy of 51.2% on the test set. The performance of the model can be broken down as follows:

Confusion Matrix:

[15 325

20 347

The model struggled to correctly predict class 0 (stock price drop) but was more effective in predicting class 1 (stock price rise).

Classification Report:

-Class 0 (stock price drop):

- Precision: 0.43
- Recall: 0.04
- F1-score: 0.08

- Class 1 (stock price rise):

- Precision: 0.52
- Recall: 0.95
- F1-score: 0.67

Overall accuracy: 51.2%

The precision for class 0 is low, indicating that the model rarely predicts a stock price drop correctly. However, the recall for class 1 is high, meaning the model is effective at identifying when the stock price will rise.

RECOMMENDATIONS

Further Tuning:

- Continue experimenting with hyperparameter tuning. The current model has been tuned, but further adjustments, such as trying different ranges for hyperparameters or using more advanced techniques like Bayesian optimization, might lead to better model configurations and performance improvements.

Thank you for reviewing this analysis. If you have any questions or need further information, feel free to ask. Your feedback is valuable, and I'm here to help with any additional insights or clarifications you may need.

Please feel free to ask any questions or request further details about the analysis and recommendations.

Email: *mirriam.mumbua@student.moringaschool.com*