NED UNIVERSITY
OF ENGINEERING & TECHNOLOGY

Open-Ended-Lab REPORT

Group Members: 1. Vinodh Kumar Nihallani (CS-21150)
2.Usman (CS-21111)
3. Muhammad Asim (CS-21058)

Department: *Computer & Information Systems*

Batch: *2021*

Course: *Machine Learning (CS-324)*

Course Instructor: *Miss Mahnoor Malik*

Table of Content: Page No

1) Problem Definition and Scope:

I. Objective:-

The objective of this project is to develop and evaluate machine learning models for predicting prices in High Frequency Trading (HFT) environments using historical market data.

II. Scope:

Focus on a specific set of financial assets, Predict intraday price movements, Implemented Logistic Regression, Decision Tree Classifier, and Random Forest Classifier for comparison.

III. Data Requirements:

Historical price data obtained from financial data providers, Include technical indicators (e.g., moving averages, RSI), market sentiment, and volume data.

2) Methodology:

I. Data Collection and Preprocessing:

Utilized APIs to fetch historical market data, Handled missing values and outliers, engineered features including technical indicators and sentiment analysis scores.

II. Model Selection:

Logistic Regression: Chosen for its simplicity and effectiveness in binary classification tasks.

Decision Tree Classifier: Selected for its interpretability and ability to handle non-linear relationships.

Random Forest Classifier: Implemented for improved robustness and accuracy through ensemble learning.

III. Model Development:

Implemented models using *scikit-learn* in Python. Optimized hyper-parameters using grid search. Validated models using N-fold cross-validation. Employed *matplotlib, seaborn* for *visualization* and *performance analysis.* Utilized *pandas, numpy* for data handling, preprocessing, and model implementation.

3) Evaluation:

We have evaluated overall model's performance in regards on: *Accuracy Score, Classification and Confusion Matrix.*

I. Comparison and Interpretation:

Logistic Regression: provided a good balance between performance and simplicity.

Decision Tree Classifier: offered insights into feature importance.

Random Forest Classifier: outperformed other models due to ensemble learning.

4) Conclusion:-

In conclusion, the project successfully developed and evaluated machine learning models for predicting prices in High Frequency Trading. The Random Forest Classifier emerged as the most promising model, demonstrating strong performance and resilience. By addressing identified risks and leveraging feature engineering insights, this project lays a foundation for further advancements in algorithmic trading strategies based on machine learning.

This endeavor underscores the potential of machine learning in enhancing decision-making processes within fast-paced financial markets, paving the way for more sophisticated and adaptive trading strategies in the future.