Report for Stock Movement Prediction Using Machine Learning and Deep Learning Models

Table of Contents

1. Introduction

Provide an overview of the project, its objectives, and the importance of stock movement prediction in financial markets. Briefly describe the methodologies employed, including data scraping, feature engineering, and model evaluation.

2. Data Collection: Web Scraping Process

- **Process Description**: Explain the steps of the scraping process, including tools used (e.g., Python libraries like PRAW usually for Reddit), the target website(s), and data extraction logic.
- **Challenges Encountered**:
    - Website restrictions (e.g., CAPTCHA, dynamic content).
    - Handling missing or noisy data.
- **Resolutions**:
    - Use of headless browsers or APIs.
    - Implementation of retry logic for failed requests.
    - Preprocessing techniques like handling null values or standardizing data formats.

3. Feature Extraction and Relevance

- **Features Extracted**:
  Describe the features obtained from the scraped data and why they are relevant to stock movement predictions. For example:
    - **Volume**: Indicates market activity.
    - **Sentiment Scores** (if applicable): Correlation between public sentiment and price movement.
    - **Price Trends**: Past trends often influence future prices.
- **Relevance**:
    - Discuss how these features provide signals for price prediction and their correlation with stock movements.

4. Model Evaluation and Performance Insights

- **Metrics Overview**:
  Present metrics for each model, such as:
    - Accuracy, Precision, Recall, F1-Score, ROC AUC, Matthews Correlation Coefficient (MCC).
- **Performance Analysis**:
    - Highlight best-performing models (e.g., Random Forest, Gradient Boosting).
    - Discuss underperformers (e.g., Deep Neural Network) and possible reasons (e.g., overfitting, insufficient data).
    - Include tables or plots for clarity.

Model	Accuracy	Precision	Recall	F1-Score	ROC AUC	MCC
Random Forest	1.000	1.000	1.000	1.000	1.000	1.000
Logistic Regression	0.988	0.987	0.989	0.988	0.990	0.970
Support Vector Machine	0.559	0.577	0.445	0.502	0.500	0.350

- **Visualization**:
  Include confusion matrices, ROC curves, or bar plots to support discussions.

5. Recommendations for Improvement

- **Model Improvements**:
    - Tuning hyperparameters for SVM, KNN, and DNN models.
    - Experimenting with advanced architectures for DNN.
- **Data Improvements**:
    - Include more diverse features (e.g., economic indicators, macro trends).
    - Address imbalances in the dataset using oversampling techniques.

6. Future Expansions

- **Integrating Multiple Data Sources**:
    - Scrape data from news articles, social media platforms (e.g., Twitter sentiment), and financial reports.
- **Real-Time Prediction**:
    - Build pipelines for real-time stock movement predictions.
- **Improved Accuracy**:
    - Use ensemble techniques to combine the strengths of different models.
    - Incorporate advanced models like Transformers or LSTMs for sequential data.

7. Conclusion

Summarize the key findings, emphasizing the strengths and weaknesses of the models. Conclude with the significance of the results and how the methodology can contribute to better stock movement predictions in the future.

Appendices (Optional)

Include detailed tables, code snippets, or screenshots of visualizations that support the analysis but are too detailed for the main report.