ABSTRACT

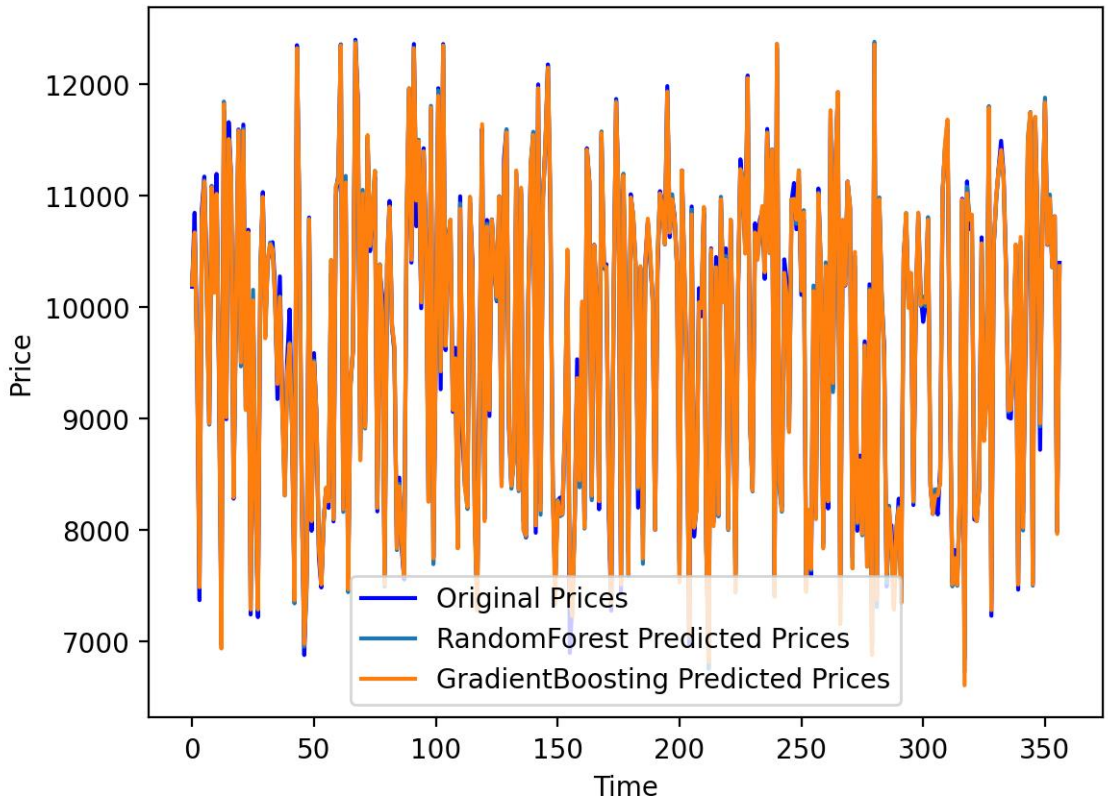
The document provided is a detailed report on stock market trend prediction using Random Forest (RF) and Long Short-Term Memory (LSTM) models. It includes an introduction, graph interpretations, model performance metrics (RMSE and MAE), and visual inspections of the models. Additionally, it contains a section on overall interpretation, focusing on model fit versus error metrics, and considerations for model selection. The final part of the document outlines prerequisites and instructions for installing necessary libraries and running the models using Streamlit.

Introduction

This project aims to predict stock market trends employing Random Forest and LSTM models, utilizing datasets from DJI, NASDAQ, NYSE, RUSSELL and S&P. Random Forest is adept at capturing complex, non-linear relationships, essential for stock price fluctuations. LSTM, known for its efficiency in handling sequential data, aligns well with the sequential nature of stock market data, accommodating the irregular time intervals characteristic of this data.

Graph Interpretation

1. **Random Forest Graph Interpretation:** The depicted graph compares actual stock prices with those predicted by Random Forest and Gradient Boosting models. The analysis of the graph reveals various insights:



- Blue Points signify actual stock prices, showing significant variability.
- The Green Line, representing Random Forest's predictions, seems less smooth, potentially indicating a model that captures more data noise.
- The Orange Line, indicative of Gradient Boosting's predictions, exhibits higher variance.
- Prediction Accuracy: Both models generally follow the trend of stock prices, albeit with noticeable noise.

RandomForest Metrics (RUSSELL):

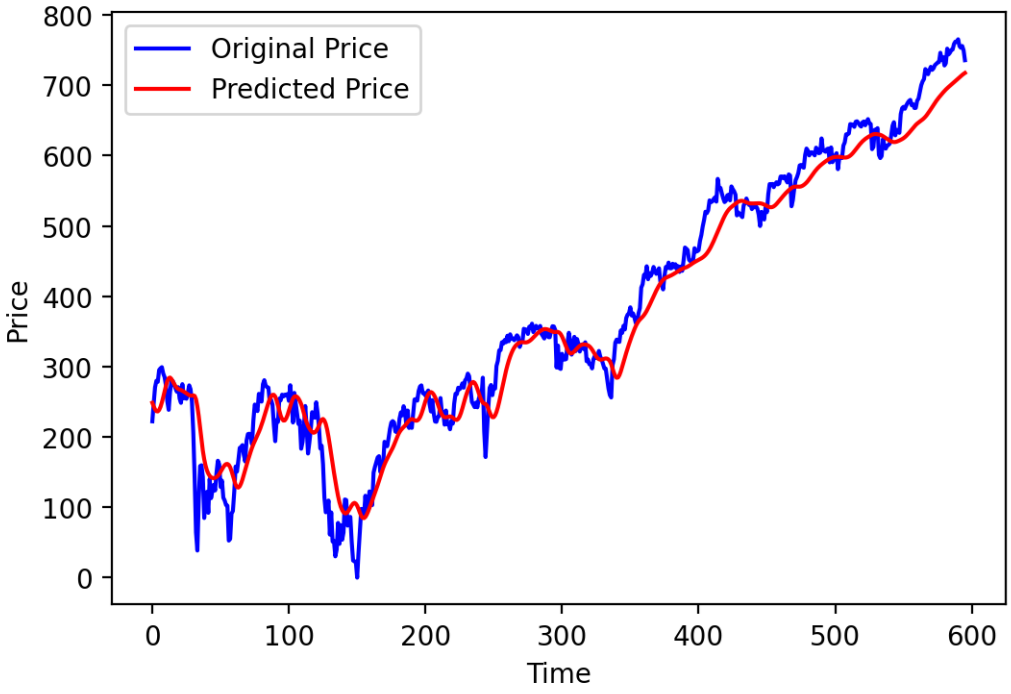
- RMSE of 13.19 suggests reasonable accuracy, with predictions deviating from actual prices by this margin.
- MAE of 9.85 indicates fewer large-scale errors, beneficial for stock price prediction.

GradientBoosting Metrics:

- Slightly improved RMSE and MAE over RandomForest, hinting at a marginally better data fit.

Visual Inspection: Despite adequate numerical metrics, the visual analysis shows significant noise in both models' predictions, possibly indicating an inability to accurately capture the underlying trend.

2. **LSTM Graph Interpretation:** This section interprets a graph comparing actual and LSTM-predicted stock prices, revealing:



- An upward trend in both actual and predicted prices.
- Close tracking of actual prices by the LSTM model, suggesting high predictive accuracy.

LSTM Metrics (RUSSELL):

- Higher RMSE and MAE than RandomForest and GradientBoosting, indicating larger average deviations from actual prices.

Visual Inspection:

- The LSTM model closely mirrors the actual price trend, albeit with scaling issues, suggesting its effectiveness in capturing general stock price movements.

Overall Interpretation

- **Model Fit vs. Error Metrics:**

  RandomForest and GradientBoosting show decent data fit but may capture more noise than trend. The LSTM model, despite higher error metrics, appears to capture the price trend effectively.

- **Considerations for Model Selection:**

  The choice of model depends on the specific requirements – LSTM for overall trend capture, RandomForest, and GradientBoosting for more precise short-term predictions.

Pre-Requisites

1. **Installation of libraries:**

   pip3 install -r requirements.txt

   includes pandas, numpy, sklearn, keras, streamlit, matplotlib.

2. **For executing models:**

   Use streamlit run streamlit_RF.py for Random Forest and streamlit run streamlit_LSTM.py for LSTM.