

Summary of the model project and model validation scores.

Objective:

The aim of the Stock Market Price and Sentiment Analysis Project is to forecast stock price movements using historical stock market data and sentiment data obtained from financial news headlines and social media.

Data Collection:

1. Stock Data: The "FinanceDataSet.ipynb" notebook retrieves historical data for stock symbols (ONCO, CNEY, TNXP, APLD, KTTA) from January 1, 2022, to September 26, 2024, using the `yfinance` library, capturing details like Open, Close, High, Low, and Volume.
2. Sentiment Data: The "finviz_news_sentimentData.ipynb" notebook scrapes financial news headlines from Finviz for sentiment analysis using VADER, classifying news as positive, negative, or neutral.

Analysis Process:

The "StockMarketPrice_SentimentAnalysis.ipynb" notebook merges the sentiment data with stock data, creating a target variable indicating whether the closing price increased compared to the previous day.

Model Evaluation:

A Random Forest Classifier was used to classify stock price movements. Key metrics include:

1. Class 0 (No Price Increase): Precision 0.76, Recall 0.74, F1-score 0.75.
2. Class 1 (Price Increase): Precision 0.42, Recall 0.44, F1-score 0.43.
3. Overall Accuracy: 65%, indicating moderate performance, with better results for predicting "No Price Increase."

Time Series Forecasting:

The project explored stock price forecasting using ARIMA models:

1. ARIMA Model Results:

Predicting constant future values. For instance:

ONCO: Predicted a constant price of 1992.0.

APLD: Forecasted around 3.39, ignoring upward trends.

CNEY: Forecasted prices increase gradually but remain relatively close to the final forecasted value.

TNXP: The forecasted values range from approximately 12.00 to 13.47, indicating a consistent increase.

APLD: All forecasted values are approximately 3.39.

KTTA: The forecasted prices stabilize around 23.20

Key Observations:

The classification model performed better for "No Price Increase" predictions. The forecasting models showed limitations due to low complexity, often predicting constant values, underscoring the necessity for robust techniques to capture stock market fluctuations effectively.