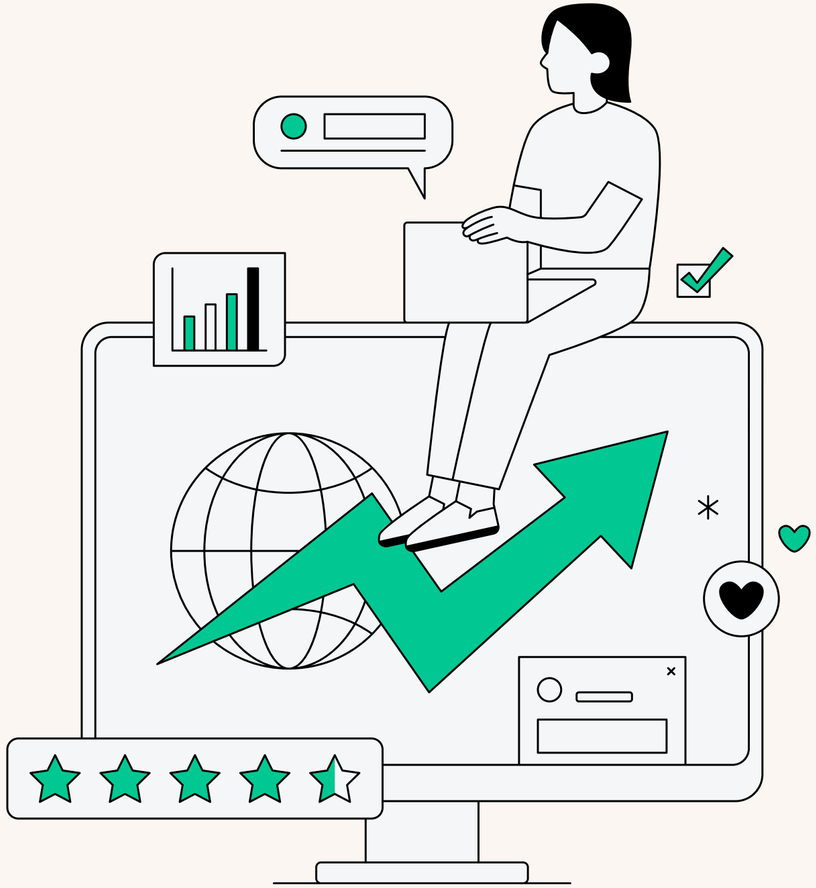
Presented by Stemtoklem

STOCK PREDICTIONS

Integrating News Sentiment and Technical Indicators

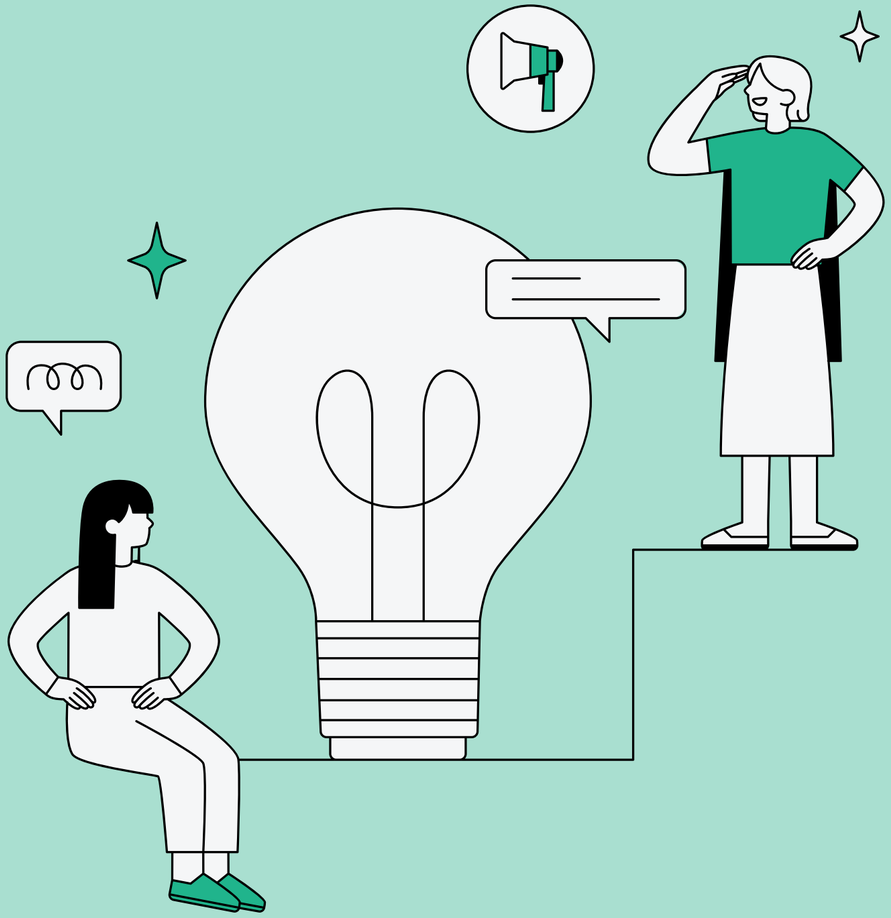
Date: 24-11-2024



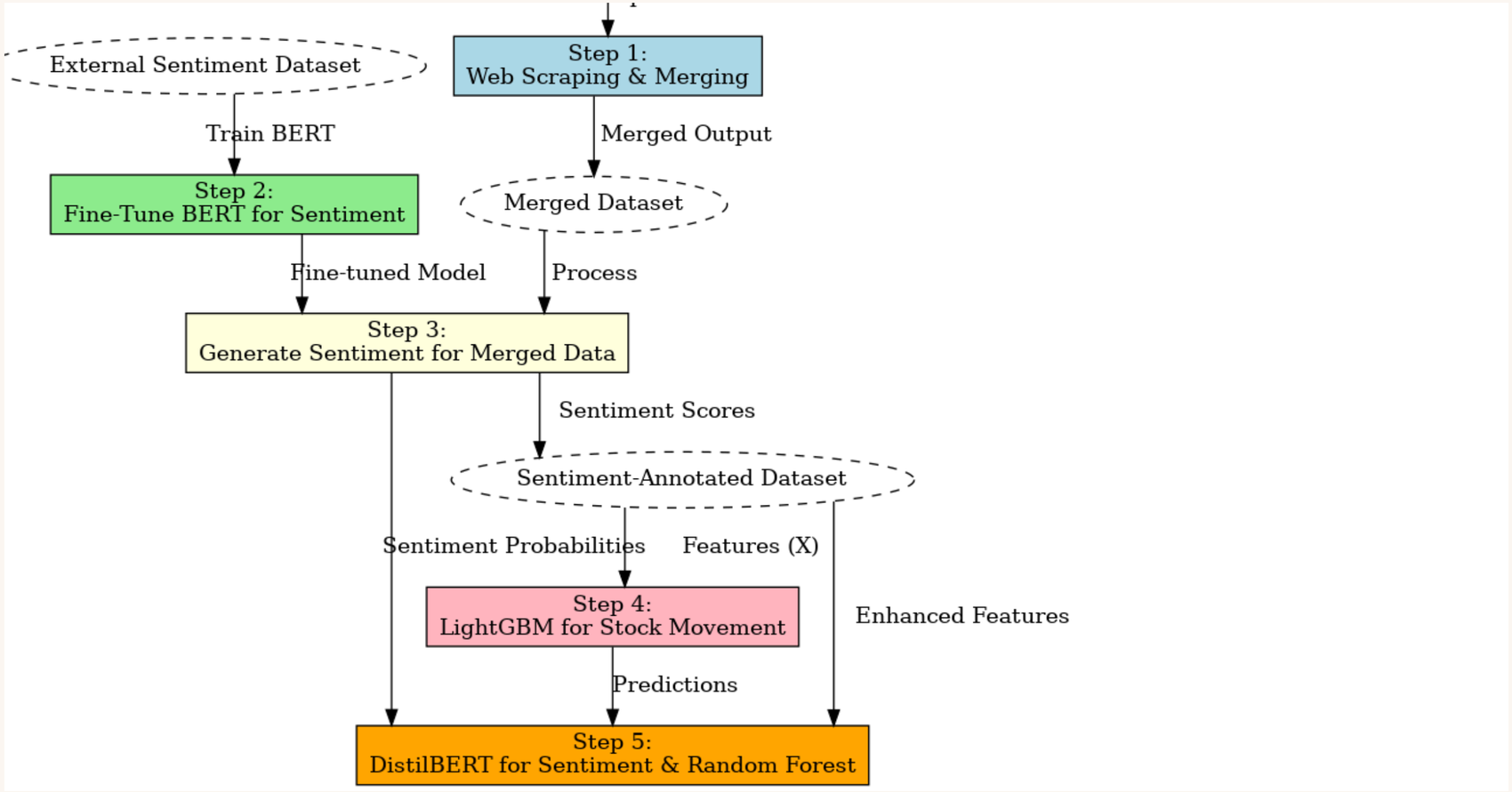
Project Overview

- **Objective**: Predict stock market movement by combining:
  - Sentiment analysis of news headlines.
  - Technical stock attributes (e.g., Open, High, Low, etc.).

- **Key Steps**:
  a. Web scraping and dataset creation.
  b. Sentiment analysis using finetuned-BERT and DistillBERT.
  c. Stock movement prediction with Random Forest and LightGBM.



FLOW CHART

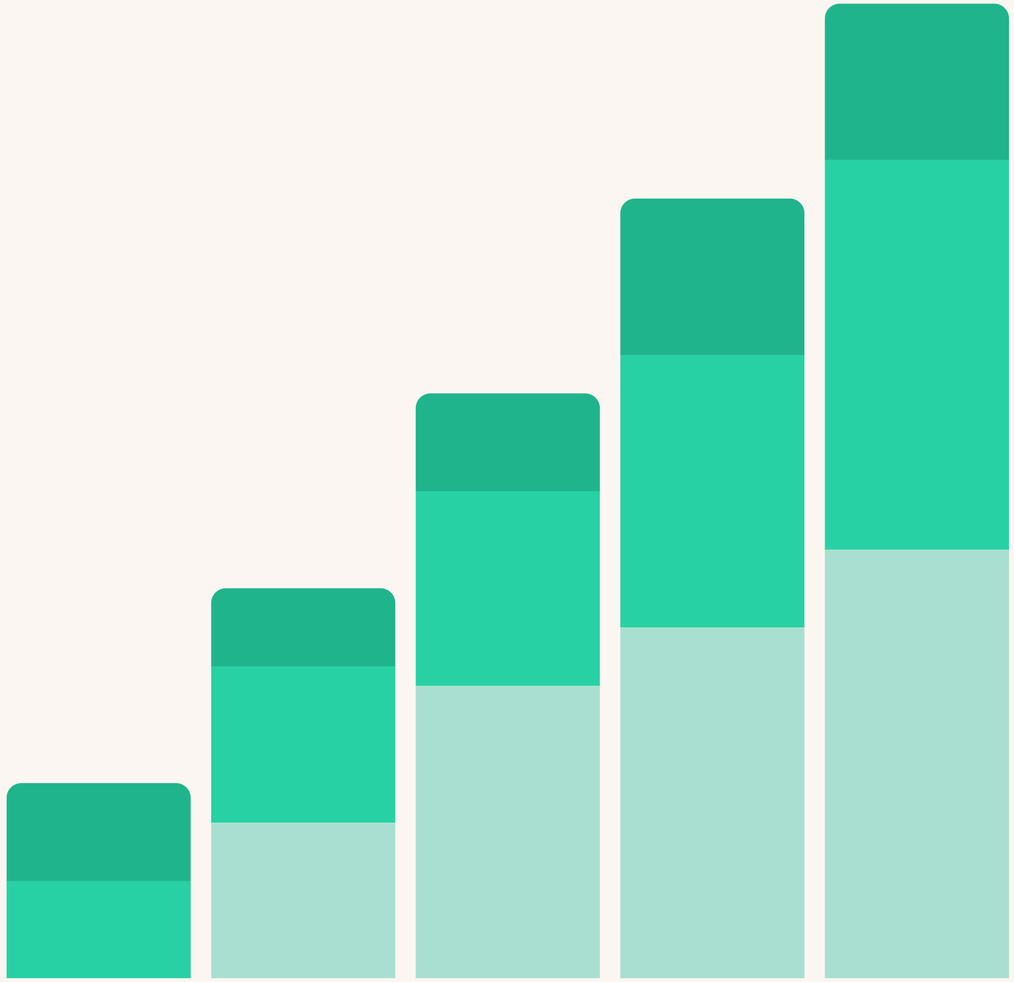

External Sentiment Dataset
Step 1:
Web Scraping & Merging
Train BERT
Merged Output
Step 2:
Fine-Tune BERT for Sentiment
Merged Dataset
Fine-tuned Model
Process
Step 3:
Generate Sentiment for Merged Data
Sentiment Scores
Sentiment-Annotated Dataset
Sentiment Probabilities
Features (X)
Step 4:
LightGBM for Stock Movement
Enhanced Features
Predictions
Step 5:
DistilBERT for Sentiment & Random Forest

WORKFLOW

1. Web Scraping & Merging:
   - Scraped daily news summaries and merged with Nifty50 stock data.

2. Sentiment Analysis:
   - Fine-tuned BERT for binary sentiment analysis (1/0).
   - Used distill bert for sentiment prediction (negative/positive) and sentimentprobabilties.

3. Stock Prediction Models:
   - Random Forest and LightGBM to predict stock movement using:
     - Sentiment probabilities.
     - Stock trading indicators.



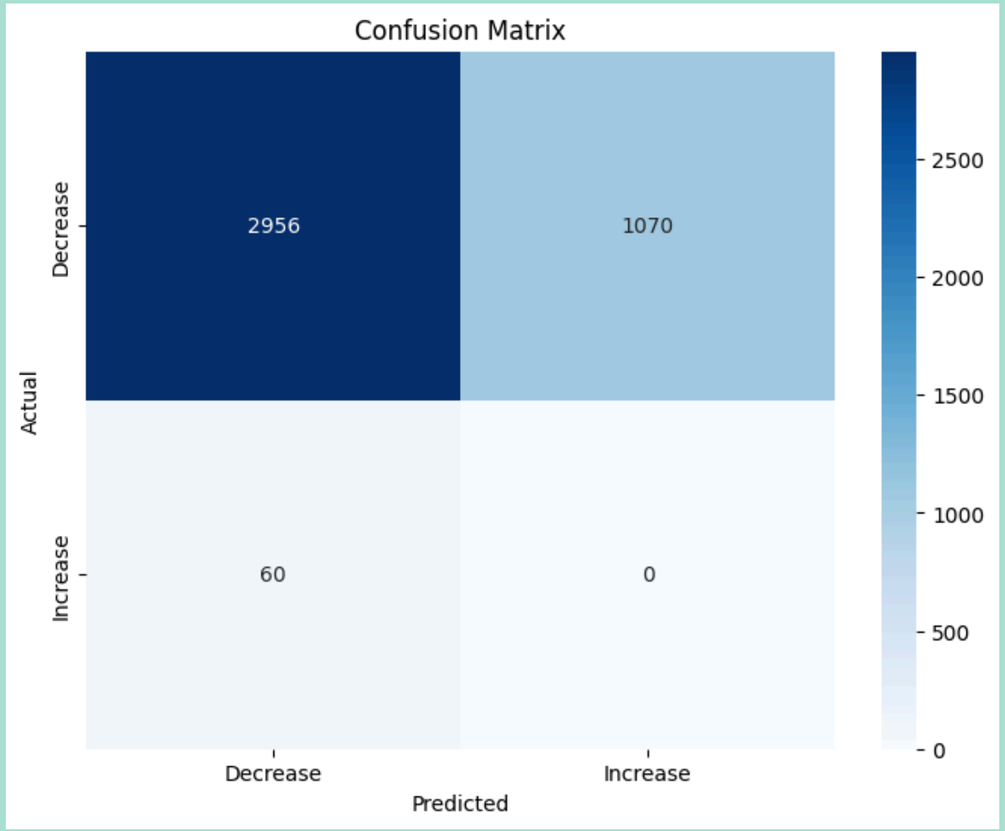
Data and Modeling

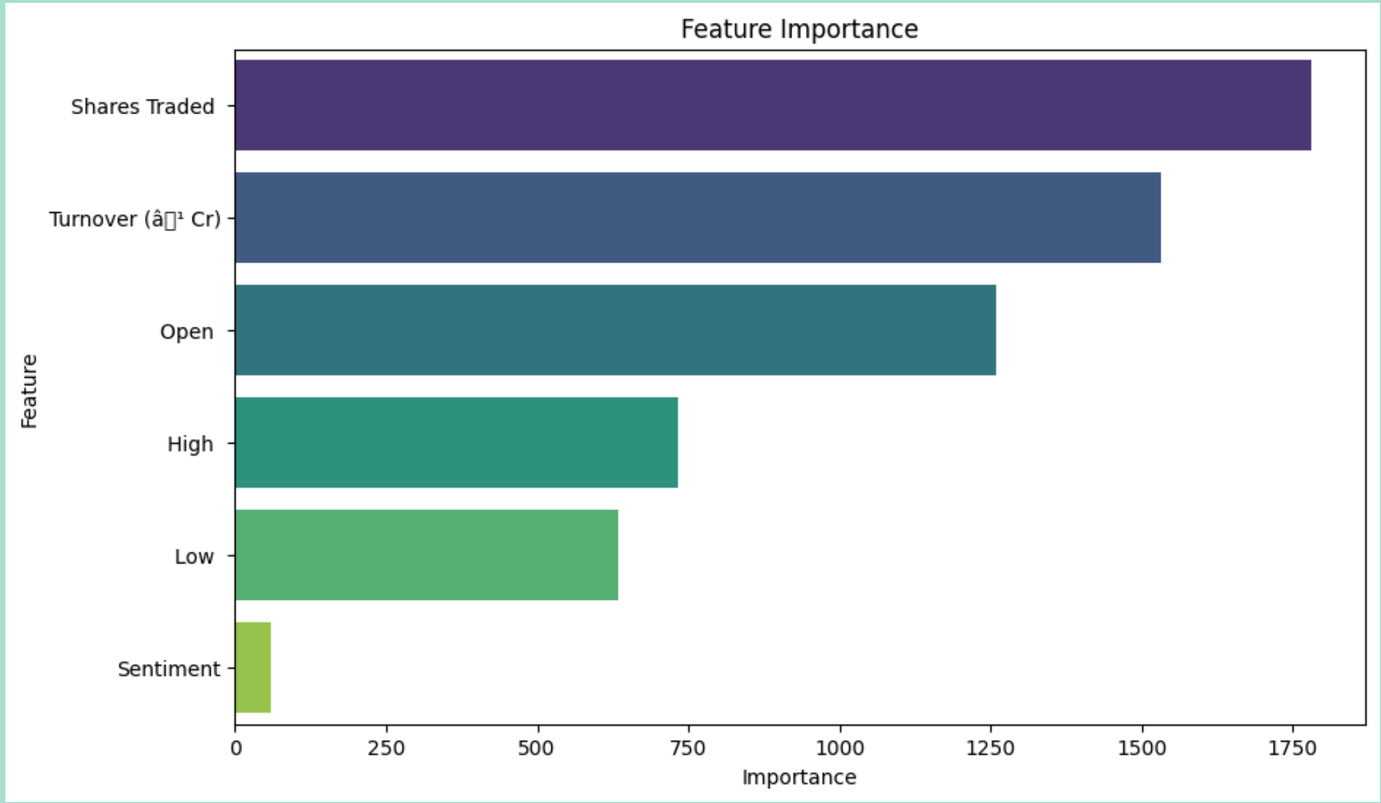
- Features:
  - Sentiments (predicted),Stock Movements.
  - Stock indicators: Open, High, Low, Turnover, Shares Traded,sentiment probability.
- Target Variables:
  - Sentiment classes (negative/neutral/positive).
  - Stock movement (1 = Increase, 0 = Decrease).

Tools Used:

pandas, BeautifulSoup, transformers, scikit-learn, lightgbm,Matplotlib.

key visualisations for LightGBM Model



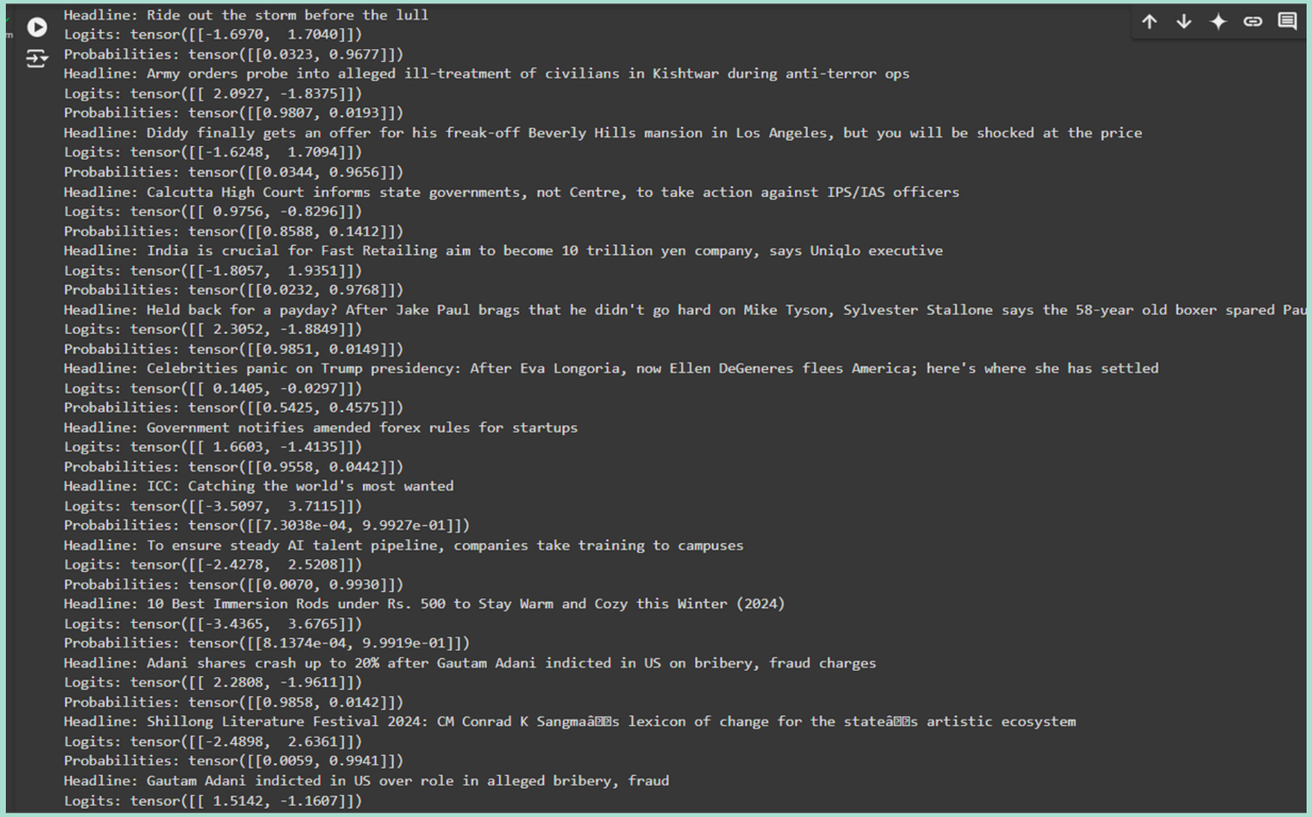


Results for LightGBM Model

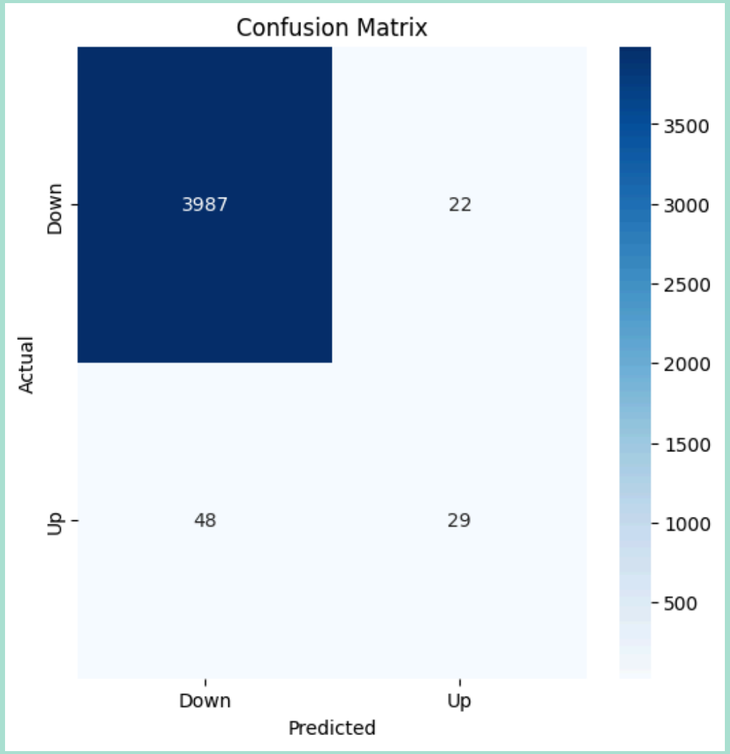
```
print("\nClassification Report:")
print(classification_report(y_test, y_pred))

accuracy = accuracy_score(y_test, y_pred)
print(f"\nAccuracy: {accuracy:.2f}")
```

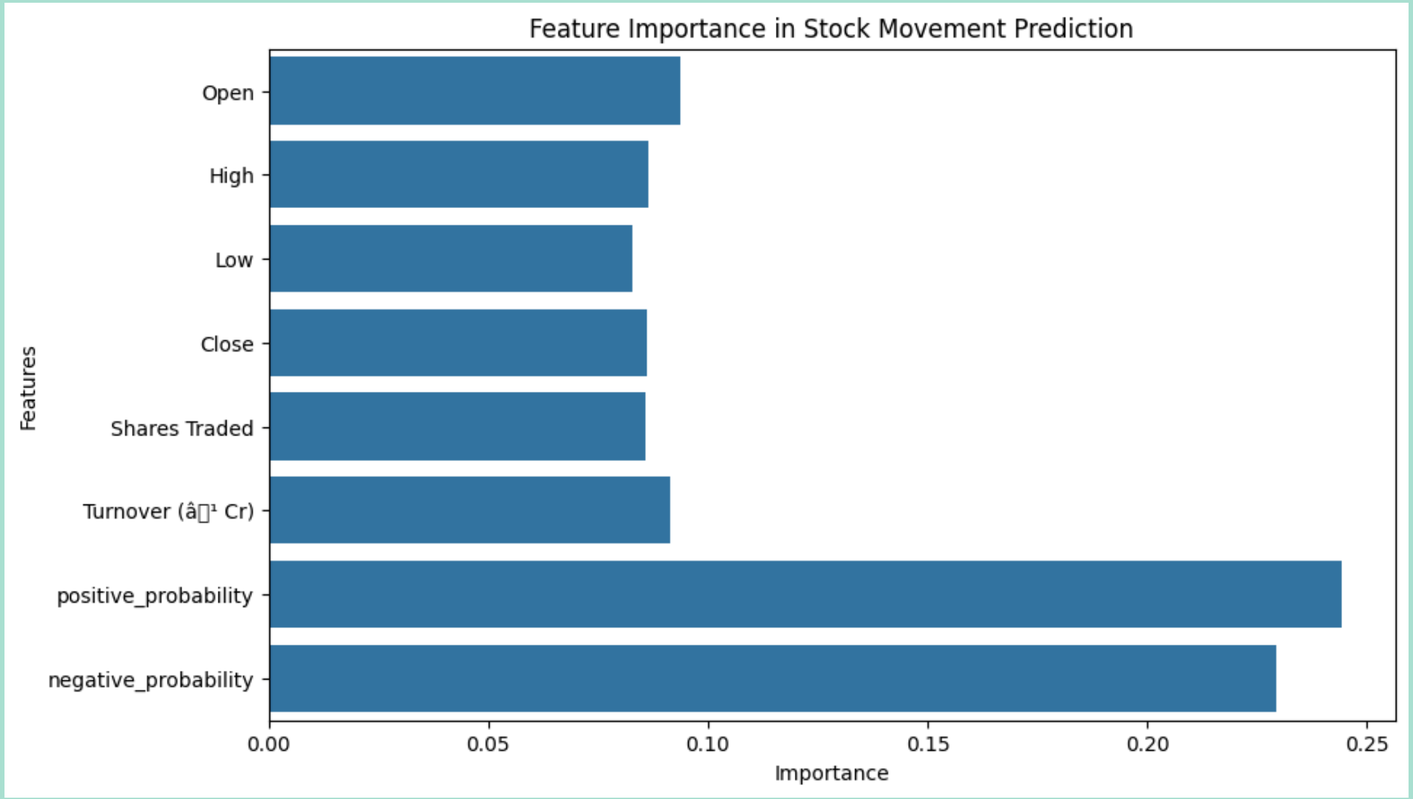
```
Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.73      0.84      4026
           1       0.00      0.00      0.00        60

    accuracy                           0.72      4086
   macro avg       0.49      0.37      0.42      4086
weighted avg       0.97      0.72      0.83      4086


Accuracy: 0.72
```

PROBABILITIES AND TENSORS FROM DISTILBERT



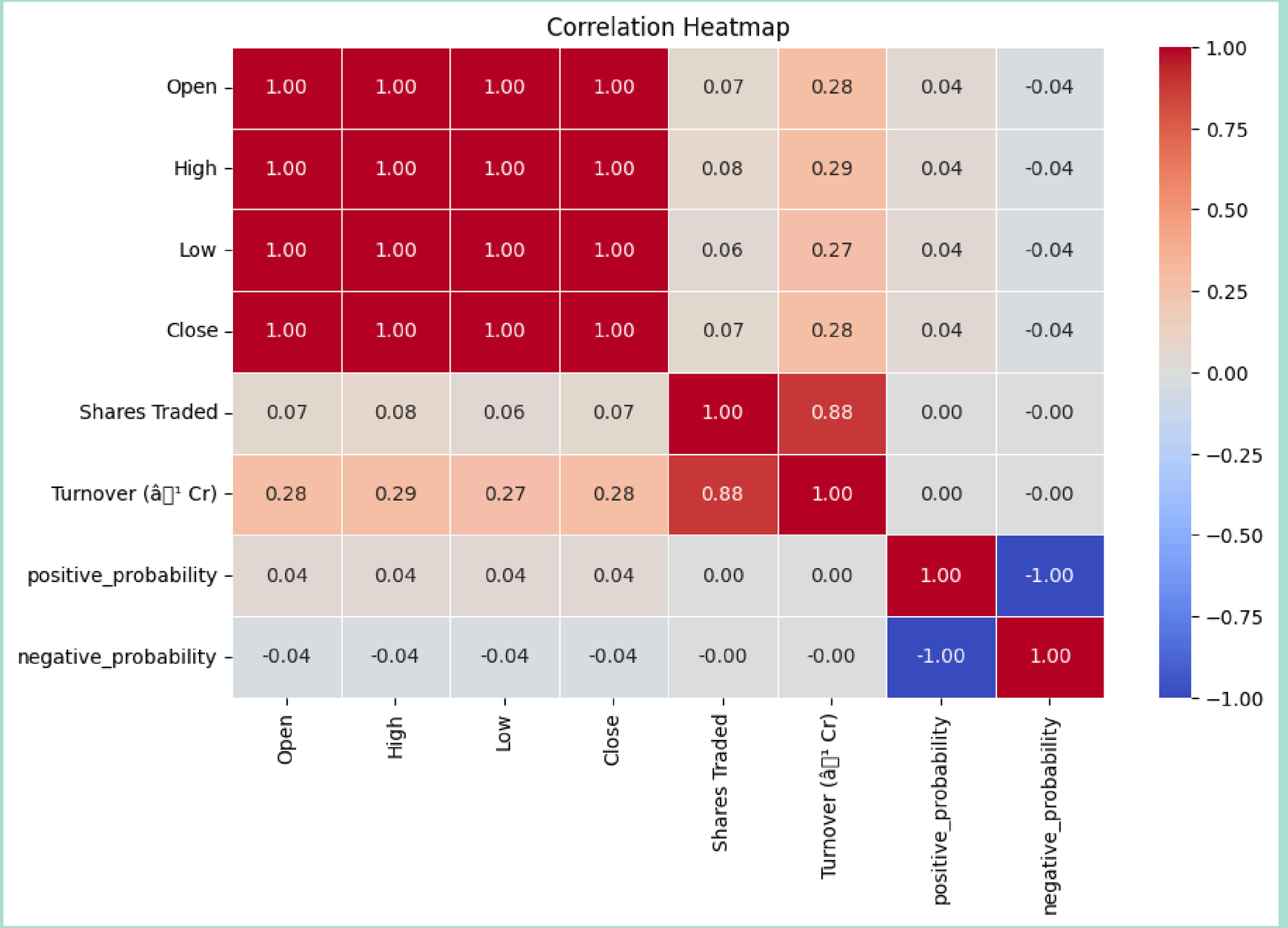
key visualisations for Randomforest with sentiment probabilty





Results for Randomforest with sentiment probabilty

```
Classification Report:
              precision    recall  f1-score   support

           0       0.99      0.99      0.99      4009
           1       0.57      0.38      0.45        77

    accuracy                           0.98      4086
   macro avg       0.78      0.69      0.72      4086
weighted avg       0.98      0.98      0.98      4086

Accuracy Score:
0.982868330885952
```


Correlation Heatmap
Open
High
Low
Close
Shares Traded
Turnover (â¹ Cr)
positive_probability
negative_probability
1.00 1.00 1.00 1.00 0.07 0.28 0.04 -0.04
1.00 1.00 1.00 1.00 0.08 0.29 0.04 -0.04
1.00 1.00 1.00 1.00 0.06 0.27 0.04 -0.04
1.00 1.00 1.00 1.00 0.07 0.28 0.04 -0.04
0.07 0.08 0.06 0.07 1.00 0.88 0.00 -0.00
0.28 0.29 0.27 0.28 0.88 1.00 0.00 -0.00
0.04 0.04 0.04 0.04 0.00 0.00 1.00 -1.00
-0.04 -0.04 -0.04 -0.04 -0.00 -0.00 -1.00 1.00
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

Stock Movement Prediction Models

Light BGM results

- Accuracy: 0.72%
  Key Metrics:
- Precision, Recall, F1-score visualized in a classification report.

Random Forest with sentiment probability results:

- Accuracy:0.98 %
  Key Metrics:
- Precision, Recall, F1-score visualized in a classification report.

Work

Previous Work

- Initial Models: Early methods of stock market prediction relied heavily on machine learning algorithms like SVM. These models required significant feature engineering and failed to capture **sequential dependencies** inherent in time-series data

- StockFormer: A transformer-based model combining news sentiment (via FinBERT) and stock price data. It outperforms traditional models like SVMs by improving prediction accuracy

Ongoing Work

- MASTER (Market-Guided Stock Transformer): A newer transformer model that focuses on intra-stock and inter-stock aggregation. It uses multi-head attention to capture dynamic correlations between stocks, improving the accuracy of stock forecasting

- Recent Developments: The latest research emphasizes multi-modal data integration. Models now incorporate news, financial metrics, and social media data to improve forecasting. This multi-source approach aims to create a more holistic prediction model

Novelty

- We initially focused on using volatile metrics such as the Volatility Index (VXN) to identify volatility patterns. However, our model's performance was suboptimal, yielding an accuracy of 55%, with recall and precision scores close to zero, indicating a significant data imbalance.
- Upon further analysis, we concluded that more data was needed to improve the model, but this added complexity. As a result, we decided to switch our approach. (reference link)
- Instead of modelling with only sentiment score, we modelled it with sentiment probability scores using **DistilBERT**
- We introduce softmax layer to get **sentiment probabilities .**
- We modelled them by random forest.

Thank you very much!

STEMTOKLEM