Name: Sachin Verma
Contact Number: 8930083406
Branch: Mechanical Engineering
Email Id: sachin_v@me.iitr.ac.in
Enrollment number: 22117122

Stock Sentiment Analysis Using Machine Learning

Objective

The objective was to determine if sentiment derived from news articles could predict stock price movements and provide actionable trading signals. Various machine learning models were applied to predict stock prices, and the effectiveness of sentiment-adjusted moving averages was analyzed.

Introduction

In financial markets, investor sentiment often influences stock prices. This study explores the potential of using sentiment analysis on news articles to predict stock price movements. The analysis involved collecting news articles and stock price data, calculating polarity scores, and applying machine learning models to predict stock prices and generate trading signals.

Data Collection and Preparation

News Articles Data

News articles were collected from the New York Times and Yahoo Finance. The articles were preprocessed to extract relevant text and metadata. From yahoo Finance I collected data of 6 companies and from New York Times I collected previous years data of Apple to make more effective analysis of Apple.

Stock Prices Data

Historical stock prices were obtained from Yahoo Finance, including open, close, high, low, and adjusted close prices.

Data Integration

The news articles and stock price data were integrated based on the dates to align news sentiment with corresponding stock price movements.

Data preprocessing

Collected news headlines were preprocessed-

- Converted to lower Case, removed punctuation, stop words ,Tokenised , Lemmatization, Vectorisation, Tf-Idf were applied.

- Date column was converted to date format.

Polarity Score Calculation and Labeling

Polarity scores were calculated for each news article using the Stock Intensity Analyzer. These scores indicate the sentiment of the news articles, with positive scores suggesting positive sentiment and negative scores indicating negative sentiment.

These Polarity Scores I have used as an input along with the headlines to make models understand the sentiments more effectively.

Positive Polarity - Indicates a positive sentiment in the news article.

Negative Polarity - Indicates a negative sentiment in the news article.

Zero Polarity – Indicates a neutral sentiment in the news article.

To label the dataset, the percentage change in adjusted close prices was calculated. Articles were labeled based on the change in stock price following the publication of the article.

```
headlines['label'] = headlines['pct_change'].apply(lambda x: 2 if x > 0 else (1 if x == 0 else 0))
headlines.head()
```

	Date	Headline	Stock Name	polarity_score	pct_change	label
0	2024-06-19	Bullish S&P 500 calls, weak China industrial o...	TSLA	-0.4404	0.222089	2
1	2024-06-19	Tesla's 'aging portfolio' is a bigger issue th...	TSLA	0.0	0.222089	2
2	2024-06-20	Warren Buffett's Berkshire Hathaway dumps BYD ...	TSLA	-0.4215	0.222089	2
3	2024-06-20	Trending tickers: Apple, Palantir, Tesla and W...	TSLA	0.0	0.222089	2
4	2024-06-20	Musk plans stock option grants to Tesla's high...	TSLA	0.2263	0.222089	2

Machine Learning Models

Different machine learning models were applied to predict stock prices based on the labeled dataset:

Gaussian NB, Multinomial NB, Bernoulli NB

Logistic Regression

SVM

Decision Tree

Random Forest

K Nearest Neighbors

AdaBoost

XG boost

Accuracy score, confusion matrix, Precision score, F1 score, Recall score were calculated for each model.

Results

```
Evaluating GaussianNB...
Cross-validation score: 0.888809
Test set accuracy: 0.852941
Precision: 0.913599
Recall: 0.852941
F1 Score: 0.864505
Confusion Matrix:
 [[292   0]
 [205 897]]
```

```
Evaluating MultinomialNB...
Cross-validation score: 0.904948
Test set accuracy: 0.911047
Precision: 0.908823
Recall: 0.911047
F1 Score: 0.906674
Confusion Matrix:
 [[ 197   95]
 [  29 1073]]
```

```
Evaluating BernoulliNB...
Cross-validation score: 0.879840
Test set accuracy: 0.866571
Precision: 0.893200
Recall: 0.866571
F1 Score: 0.873735
Confusion Matrix:
 [[255  37]
 [149 953]]
```

```
Evaluating LogisticRegression...
Cross-validation score: 0.893653
Test set accuracy: 0.905308
Precision: 0.902101
Recall: 0.905308
F1 Score: 0.901460
Confusion Matrix:
 [[ 198   94]
 [  38 1064]]
```

```
Evaluating SVC...
Cross-validation score: 0.989958
Test set accuracy: 0.982783
Precision: 0.982704
Recall: 0.982783
F1 Score: 0.982695
Confusion Matrix:
 [[ 276   16]
 [   8 1094]]
```

```
Evaluating DecisionTree...
Cross-validation score: 0.994260
Test set accuracy: 0.960545
Precision: 0.961405
Recall: 0.960545
F1 Score: 0.960858
Confusion Matrix:
 [[ 271   21]
 [  34 1068]]
```

```
Evaluating KNeighbors...
Cross-validation score: 0.928621
Test set accuracy: 0.969154
Precision: 0.969741
Recall: 0.969154
F1 Score: 0.969361
Confusion Matrix:
 [[ 276   16]
 [  27 1075]]
```

```
Evaluating RandomForest...
Cross-validation score: 0.995875
Test set accuracy: 0.976327
Precision: 0.976189
Recall: 0.976327
F1 Score: 0.976221
Confusion Matrix:
 [[ 272   20]
 [  13 1089]]
```

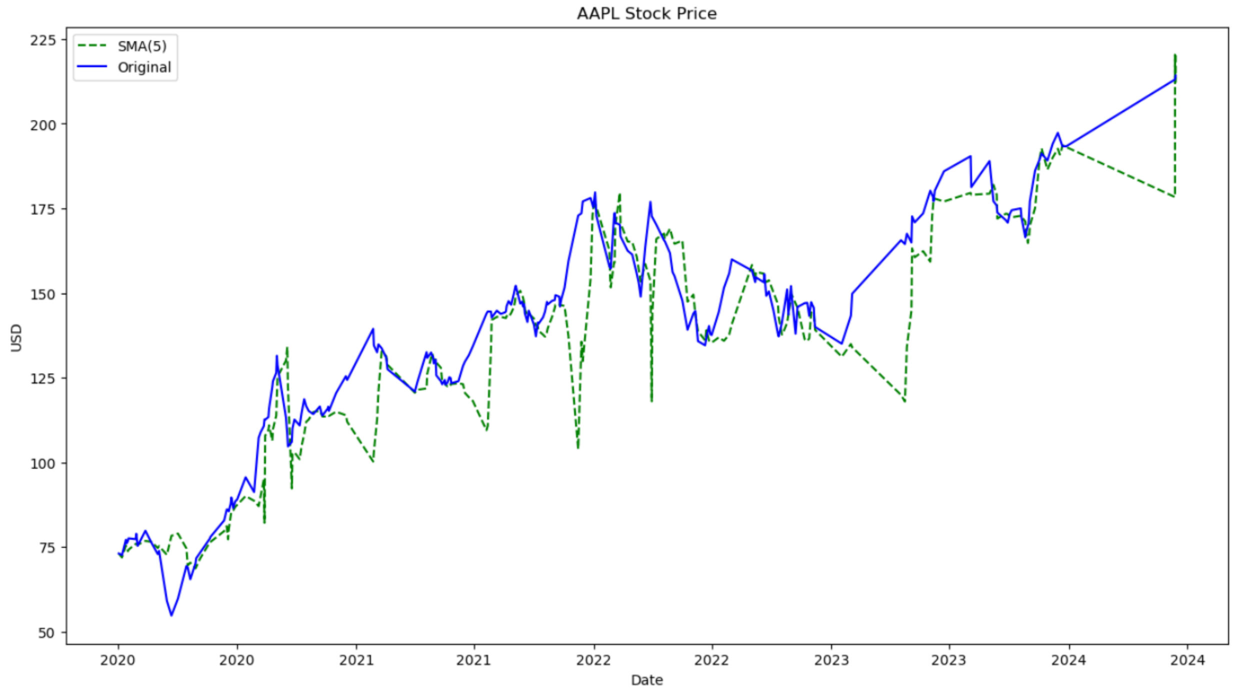
```
Evaluating Bagging...
Cross-validation score: 0.993902
Test set accuracy: 0.972023
Precision: 0.972141
Recall: 0.972023
F1 Score: 0.972075
Confusion Matrix:
 [[ 274   18]
 [  21 1081]]
```

Sentiment Adjusted Moving Average Analysis

The sentiment-adjusted moving average (SAMA) was analyzed alongside actual price movements. SAMA incorporates sentiment scores into the traditional moving average calculation to adjust for market sentiment.

Results

This graph shows the effect of sentiments with normal moving average and sentiment adjusted moving average of 5 days.



Buy and Sell Signal Prediction

The Predicted labels were used to develop a trading strategy using Moving Averages to predict buy and sell signals. Buy Signal (1) and Sell Signal (-1)

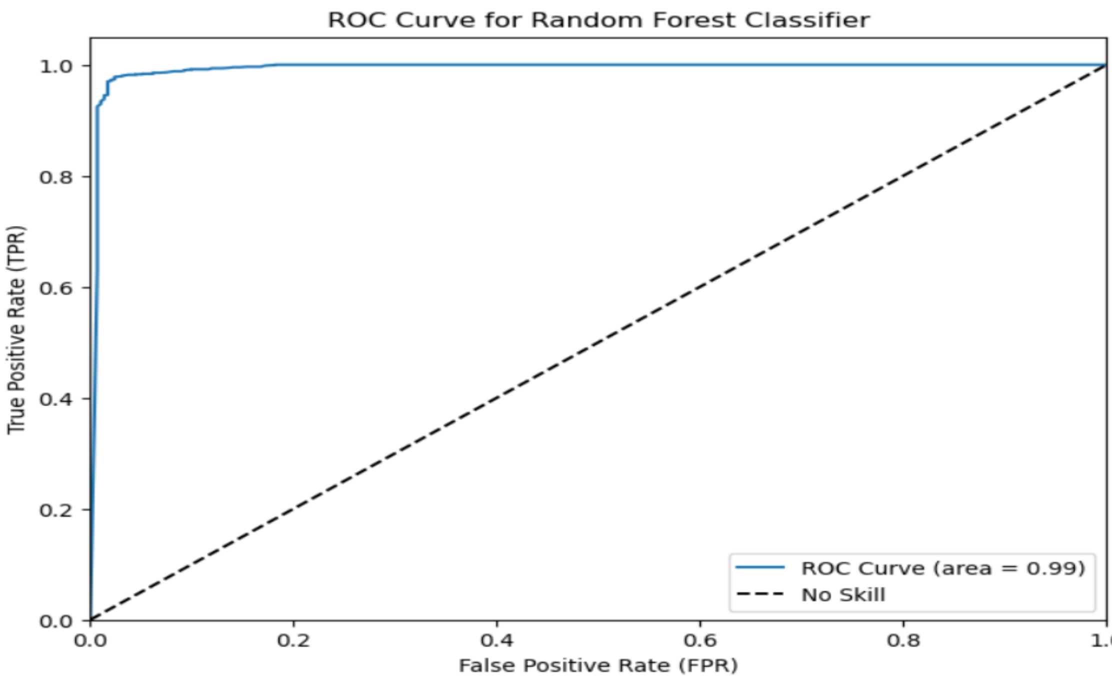
Calculate portfolio values.

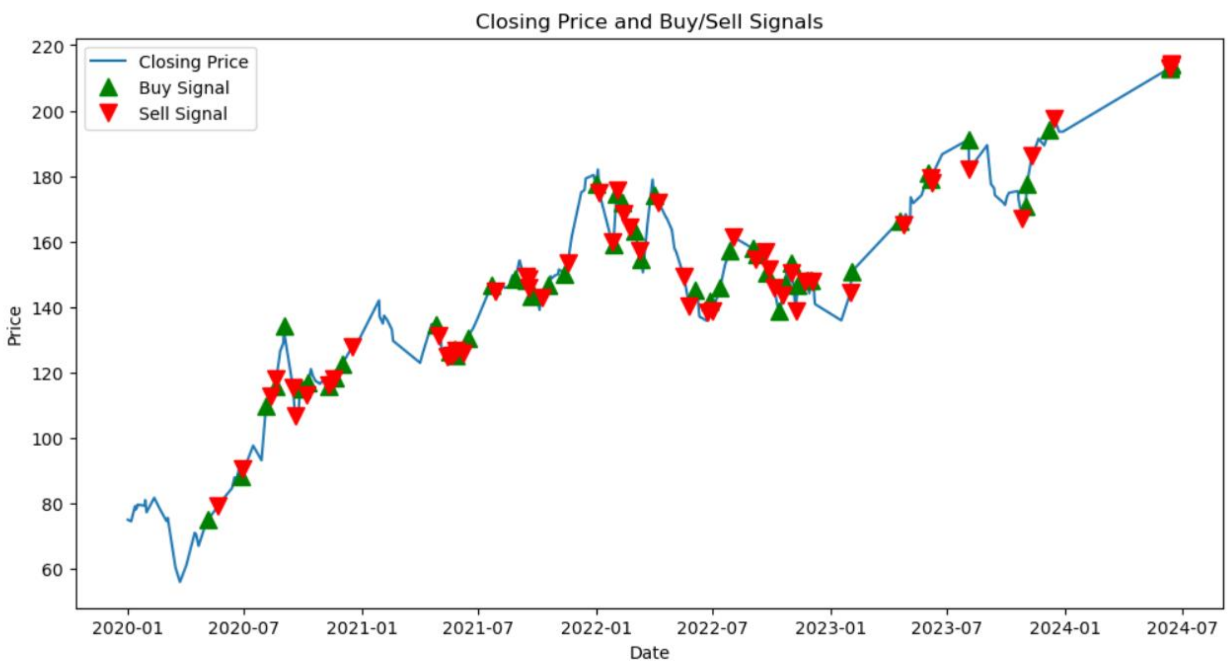
Number of trades executed.

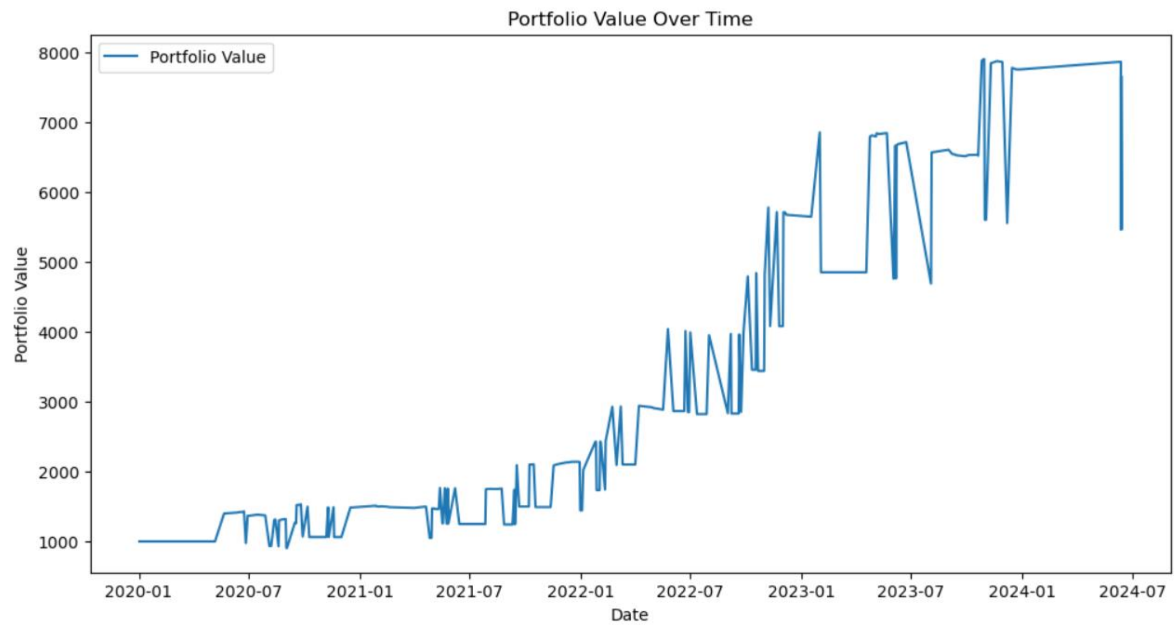
Sharpe Ratios.

Win Ratio.

Results


ROC Curve for Random Forest Classifier
1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate (TPR)
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate (FPR)
ROC Curve (area = 0.99)
No Skill


Closing Price and Buy/Sell Signals
Closing Price
Buy Signal
Sell Signal
220
200
180
160
140
120
100
80
60
Price
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07
2024-01
2024-07
Date



Results and Discussion

The results of the machine learning models, sentiment-adjusted moving average, and buy/sell signal predictions were evaluated. Key findings include:

- The accuracy of the machine learning models in predicting stock price movements.

- The correlation between sentiment scores and actual stock price movements.

- The profitability of the buy and sell signals generated based on sentiment analysis.

```
Final Portfolio Value: $7652.35
Sharpe Ratio: 1.35
Number of Trades Executed: 119
Win Ratio: 0.18
Maximum Drawdown: 0.00
```

Conclusion

The study demonstrates the potential of using sentiment analysis to predict stock price movements. Machine learning models and sentiment-adjusted moving averages provided valuable insights into market behavior. The buy and sell signals based on polarity scores showed promise for trading strategies.

Future Scope

Future work could explore:

- Enhancing the accuracy of sentiment analysis by incorporating more advanced natural language processing techniques.

- Testing the models on a broader range of stocks and news sources.

- Refining the trading strategy to include additional factors such as volume and volatility.

References

1. New York Times news articles.

2. Yahoo Finance stock prices.

3. Yahoo Finance Tweets

4. Stock Intensity Analyzer documentation.