Sentiment Analysis for Stock Market

By Nidhi Worah & Aditeya Prajapti

Team 50
Kinetic Kernels

02

Problem Statement

Develop and demonstrate a working prototype of a stock trading platform that incorporates NLP capabilities to help investors make more informed and efficient trading decisions.

Motivation

News headlines can have a significant impact on the stock market as they can influence investors' perceptions and decision making. Additionally, news headlines can also affect investor sentiment, which can further influence stock prices. The stock market can also be affected by the credibility and reliability of the source of the news. Therefore, it is important to analyze news headlines critically and understand their potential impact on the stock market.

Approach 1

- For the first Solution to the given problem statement we decided to go with stock prediction using NLTK Library
- After pre-processing and creating a bag of words count vectorizer was used for text to vector conversion
- The data was split into train and test. Predictions were made using Logistic Regression and Random Forest regressor but the accuracy achieved was very low(~50% for both algorithms)
- Also the predictions were very unsatisfactory

Solution 1- drawback

```
# Accuracy, Precision and Recall
from sklearn.metrics import accuracy_score, precision_score, recall_score
score1 = accuracy_score(y_test, lr_y_pred)
score2 = precision_score(y_test, lr_y_pred)
score3 = recall_score(y_test, lr_y_pred)
print("---- Scores ----")
print("Accuracy score is: {}%".format(round(score1*100,2)))
print("Precision score is: {}".format(round(score2,2)))
```

```
---- Scores ----
Accuracy score is: 50.79%
Precision score is: 0.51
```

Fig:Logistic Regression model performance

```
# Accuracy, Precision and Recall
score1 = accuracy_score(y_test, rf_y_pred)
score2 = precision_score(y_test, rf_y_pred)
score3 = recall_score(y_test, rf_y_pred)
print("---- Scores ----")
print("Accuracy score is: {}%".format(round(score1*100,2)))
print("Precision score is: {}".format(round(score2,2)))
print("Recall score is: {}".format(round(score3,2)))
```

```
---- Scores ----
Accuracy score is: 50.79%
Precision score is: 0.51
```

Fig: Random Forest Regressor model performance

Methodology

Sentiment Analysis

Subjectivity and Polarity refer to the degree of subjectivity and sentiment expressed in a piece of text.Subjectivity refers to whether a statement is a fact or an opinion. Polarity refers to the sentiment expressed, whether it is positive, negative, or neutral

Algorithm

LINEAR DISCRIMINANT MODEL

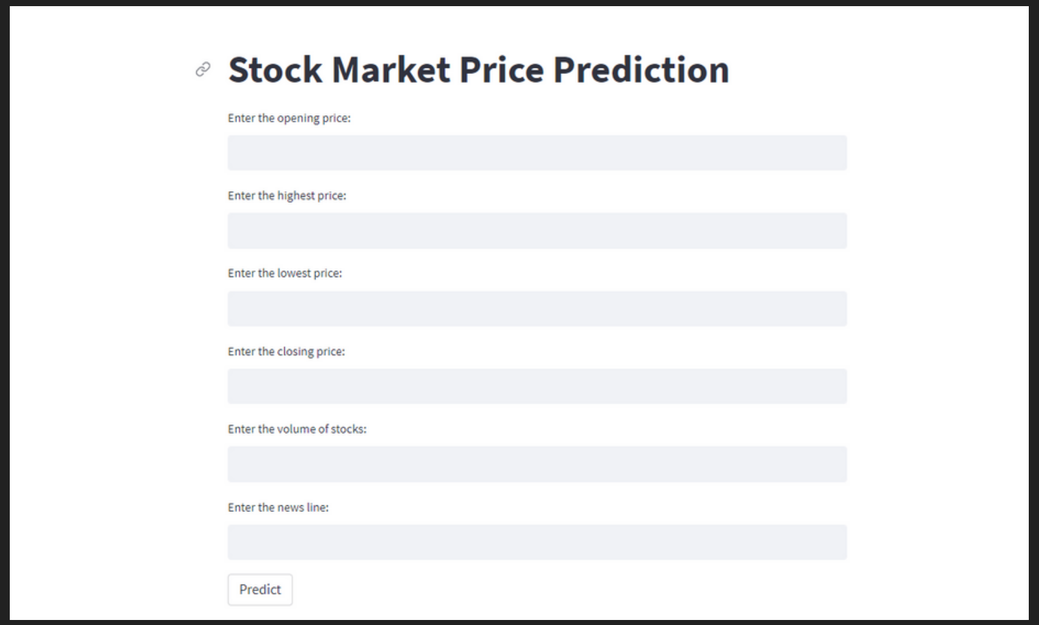
Linear Discriminant Analysis (LDA) is a supervised learning algorithm that is used for dimensionality reduction and classification tasks. It is a variation of the general linear model and is also known as Fisher's linear discriminant. The goal of LDA is to find the linear combination of features that separates different classes of data with the largest margin. It does this by computing the mean vectors for each class and the covariance matrix of the entire dataset, and then projects the data onto a lower-dimensional space where the classes are well-separated.

	precision	recall	f1-score	support
0	0.98	0.87	0.92	190
1	0.89	0.98	0.94	208
accuracy			0.93	398
macro avg	0.94	0.93	0.93	398
weighted avg	0.93	0.93	0.93	398

Results

The classification report after training the model

Application interface



Future Scope

Reinforcement learning (RL) is a type of machine learning that allows an agent to learn by interacting with its environment and receiving feedback in the form of rewards or penalties. It can be used for sentiment analysis by training an RL agent to take actions, such as classifying text as positive or negative, based on a set of predefined rewards or penalties. The agent can then learn to optimize its actions based on the feedback it receives. This approach can be useful in situations where the sentiment of text is not clearly defined or where the agent needs to learn to adapt to new data over time. However, it requires a lot of data and computational resources to train the agent effectively.