Finance Club
Open Project Summer 2024

Stock Sentiment Analysis Using Machine Learning Techniques (#FC24OPS3)

Overview:

This project delves into the creation of a stock trading strategy based on sentiment analysis powered by machine learning. The goal is to harness valuable insights gleaned from textual data, such as news articles and social media posts, to predict stock price movements and guide informed trading decisions. By collecting and analyzing this data, we aim to uncover the underlying sentiment of the market and individual stocks. This sentiment analysis, coupled with machine learning algorithms, can potentially identify patterns and predict future price trends, ultimately empowering investors to make more strategic choices in the ever-evolving stock market.

Flow of Project:

Data Collection :

> First we have to collect the data that is the variations of stock prices and news articles . for news headlines we are using new York times api key and then stored the headlines into a csv file . for the stock prices we will use a python library yfinance

Data Preprocessing:

> Now the data has to be made suitable for machine learning algorithms that is it need to be preprocessed and it goes through cleaning ,tokenizing , removing stop words and lemmatization using respective python libraries

Data Importing:

Since we have made two files one of 20 years for training and one of 6 months for testing and saved into csv files now we will import these files and store the data using read function into some variables df1 and df2

Making a Stock File:

Now we have to make a files which has variations of stocks during the same 20 years and 6 months we will using yfinance for the same ticker is name of stock which can be changed to check variations of different stocks

Adding Subjectivity and Polarity :

We have defined subjectivity and polarity function in code above using the library we will use it here to get the values which can be used afterwards to train the model
where subjectivity is Indicates whether the text expresses personal opinions, beliefs, or feelings and polarity measures overall sentiment of the text ranging from negative to positive

Finding sentiment score :

This code analyzes sentiment of news headlines in a DataFrame. It uses VaderSentiment to get compound, negative, neutral, and positive scores for each headline. These scores are stored in separate lists for further use in your stock strategy. We will follow same steps for the other file of 6 months
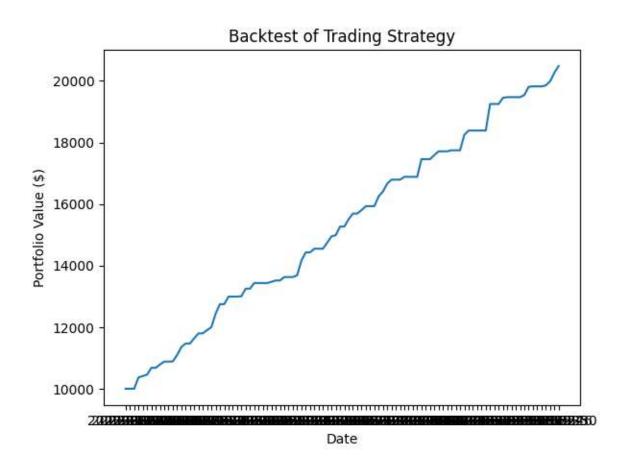
Train the model:

First I have trained the model using random forest model but the precission score using this is low so I am directly using the lineardiscriminantanalysis model which has better precision and after that we will use sklearn and take a classification report

Testing:

Now we have a model so we use another files of 6 months and test on it now we require a trading strategy for backtesting

My stratergy is if the label goes as follows 001 then I will buy or else I will sell the stock

We will initiate a portfolio value of 10000 and calculate portfolio value after each transaction so that we can get daily returns and drawdowns and also sharpe ratio

Backtest of Trading Strategy

```
Buy Signals: ['2023-01-05', '2023-01-12', '2023-01-18', '2023-01-25',
'2023-01-30', '2023-02-06', '2023-02-09', '2023-02-14', '2023-02-21',
'2023-02-24', '2023-03-01', '2023-03-07', '2023-03-10', '2023-03-17',
'2023-03-22', '2023-03-28', '2023-04-05', '2023-04-12', '2023-04-17',
'2023-04-21', '2023-04-26', '2023-05-04', '2023-05-09', '2023-05-16',
'2023-05-23']
Sell Signals: ['2023-01-11', '2023-01-17', '2023-01-24', '2023-01-27',
'2023-02-03', '2023-02-07', '2023-02-13', '2023-02-15', '2023-02-23',
'2023-02-27', '2023-03-06', '2023-03-08', '2023-03-16', '2023-03-21',
'2023-03-24', '2023-04-03', '2023-04-06', '2023-04-13', '2023-04-19',
'2023-04-24', '2023-04-28', '2023-05-05', '2023-05-11', '2023-05-19']
peak =20484.681887664174

max_drawdown=1.354122122483514e-16

sharpe ratio = 0.7063809289886621

win_ratio=24/25
```

https://colab.research.google.com/drive/15Q6cNo1t5-v_nZl2xEzWWEZVad3tpcSA?usp=sharing