STOCK SENTIMENT ANALYSIS USING MACHINE LEARNING

**Kripansh Mohit**

**21112065**

Steps:

1. Collection of Textual and Stock Data:

* Collected news headlines from financial times, economic times and New York times.
* Calculated quarter on quarter growth %, fetched PE ratio, DE ratio from yahoo finance.

1. Preprocess the Data:

* Preprocess the textual data by removing noise, tokenizing text into words or phrases, and applying techniques such as stemming and lemmatization to standardize text representations.

1. Calculate Sentiment Score:

* The model and tokenizer are loaded using AutoTokenizer and Auto Model For Sequence Classification from the transformers library.
* Sentiment analysis done using the cardiffnlp/twitter-roberta-base-sentiment-latest model.

1. Collect Stock Data: 1 year data is taken from Yahoo finance.
2. Label the Textual Data:

* Label the textual data with corresponding stock price movements (e.g., increase, decrease, or no change) over a specified time horizon to create a labeled dataset for supervised learning.

1. Train, Evaluate and Choose the Best Model:

* Trained and evaluated machine learning models, including classification algorithms such as logistic regression, support vector machines (SVM), random forests, and neural networks, to predict stock price movements based on textual sentiment.
* Perform model evaluation using appropriate metrics, such as accuracy, precision, recall, F1-score, and receiver operating characteristic (ROC) curve analysis, to assess the performance of the sentiment analysis model.

1. Visualization:

* Matplotlib is used to plot the required data for analysis.

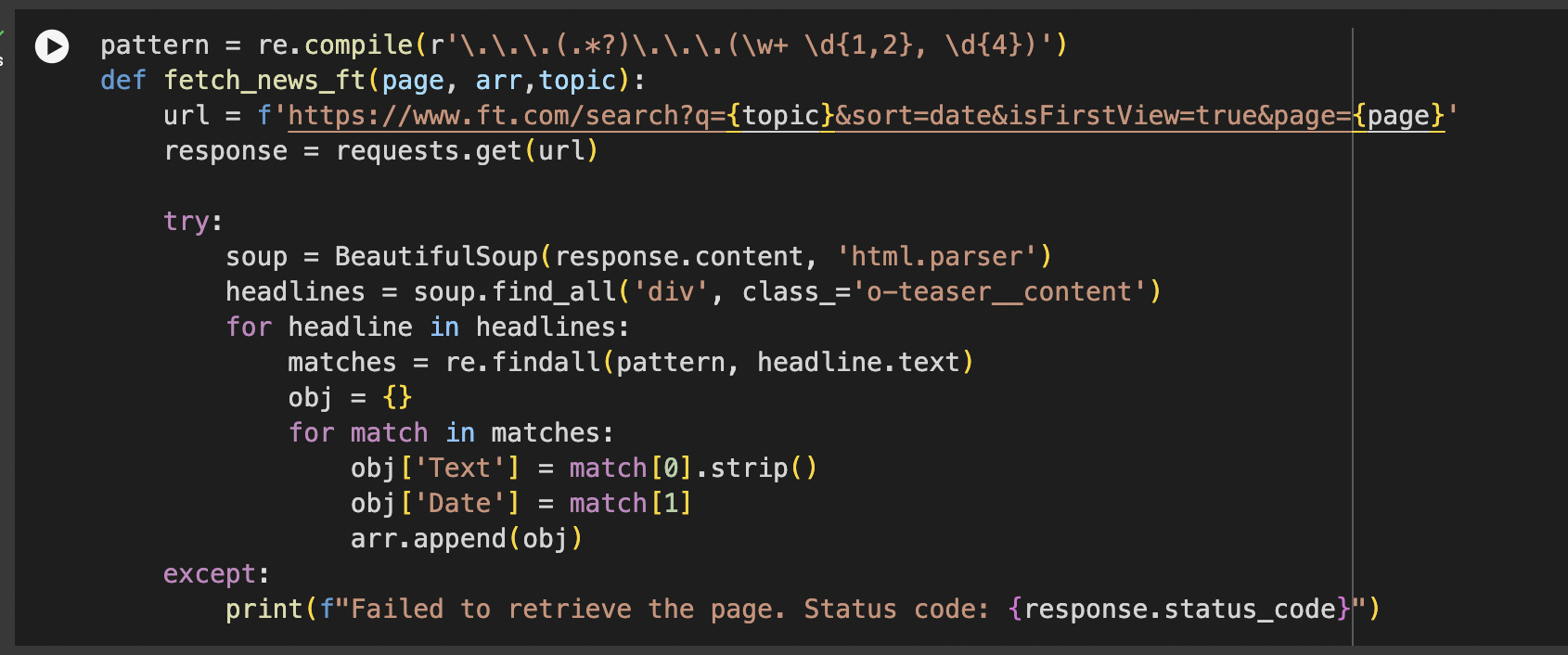
1. Analysis & Risk Calculations:

* Final Portfolio and Total Returns
* Daily mean and Daily Standard Deviation of Prices
* Sharpe Ratio
* Maximum drawdowns
* Number of trades
* Win Ratio

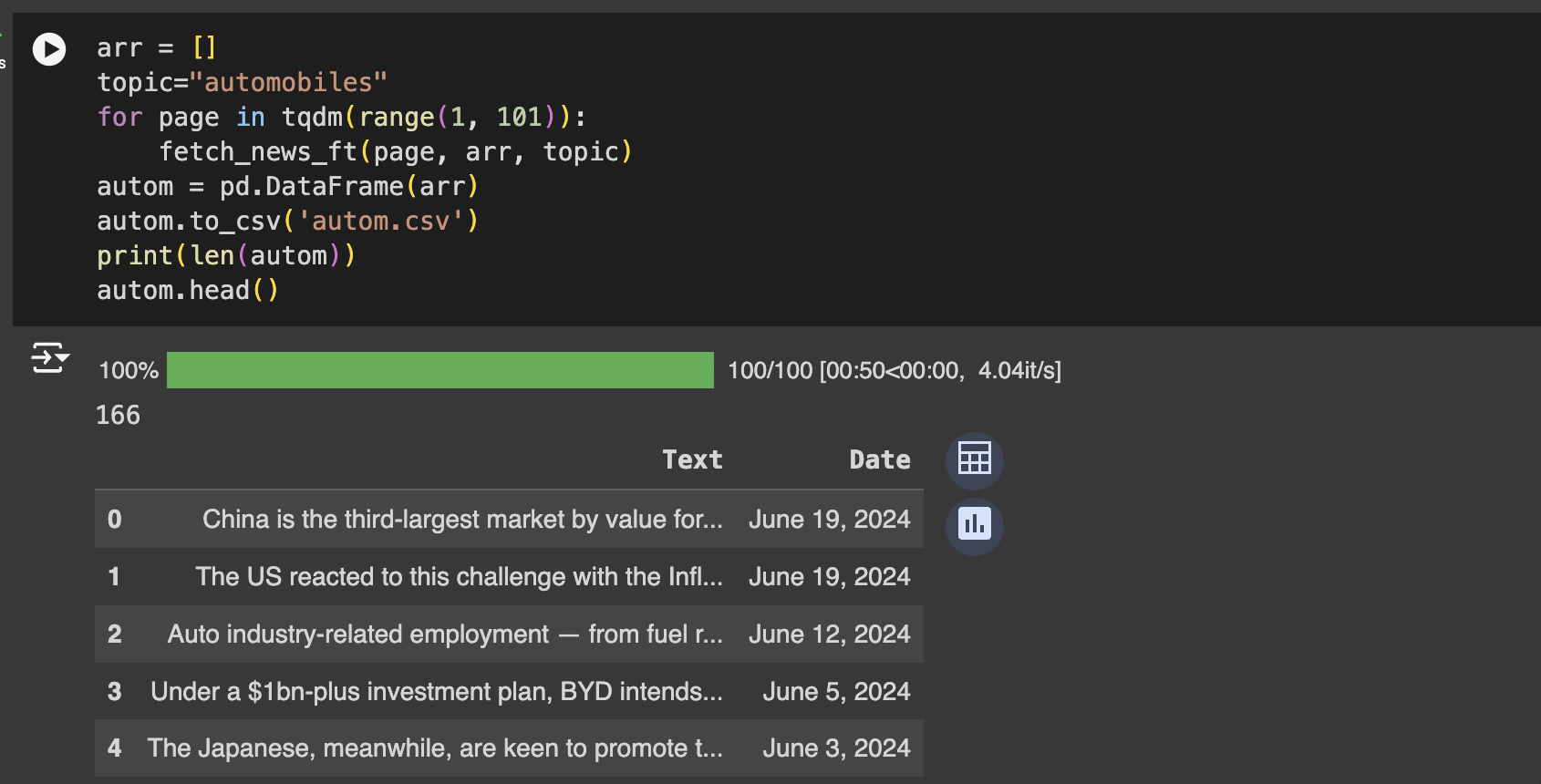
Code:

1. Collect Textual Data:

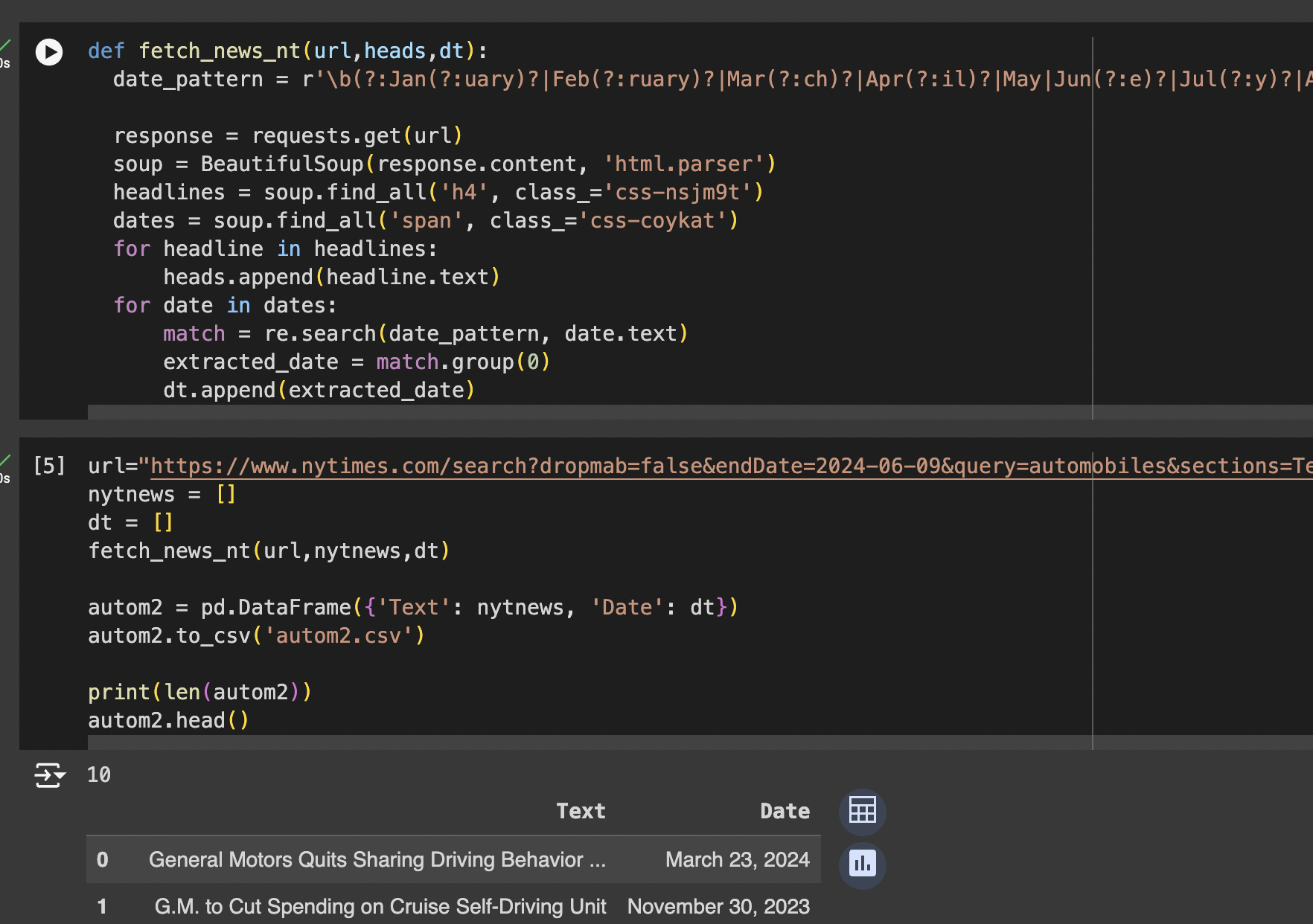
Fetch news from financial times



Topic: Automobile



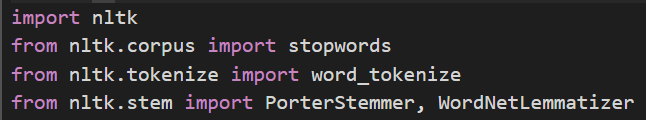
Fetch news from New York times



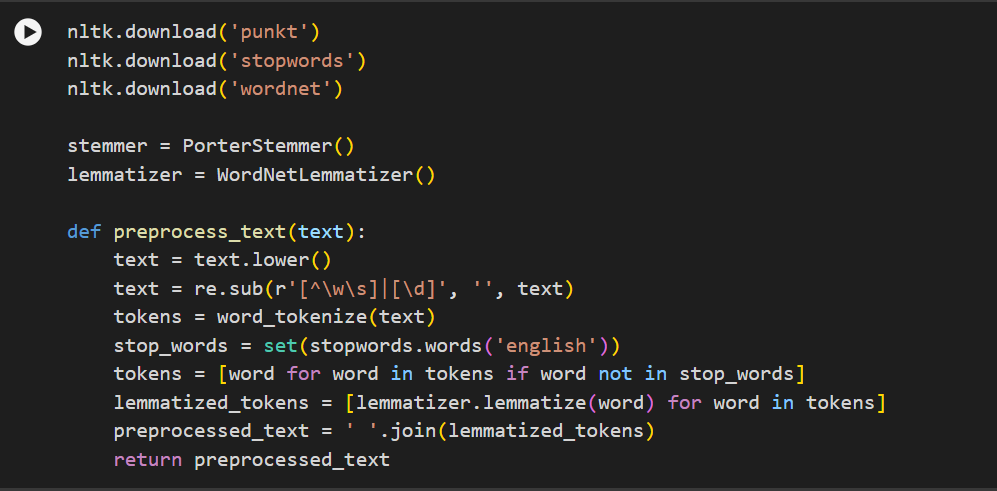
Fetch news from economic times



1. Preprocess the Data:



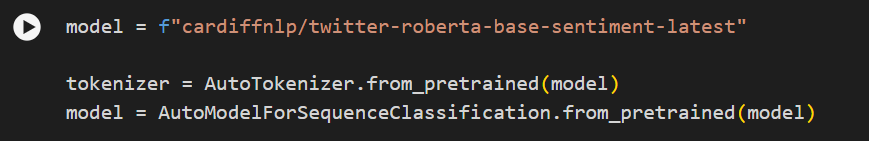
Removing stop words, tokenizing text into words, and reducing words to their base forms using lemmatization.

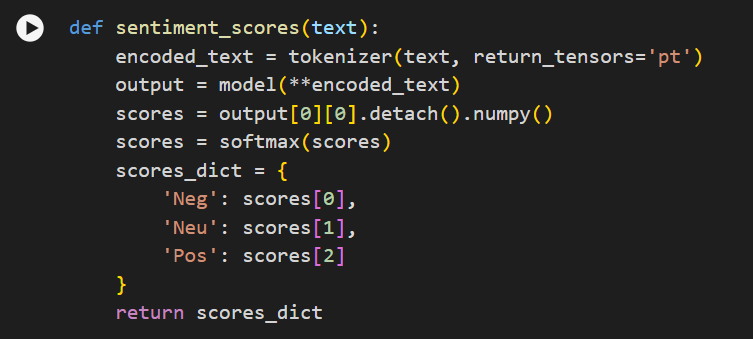


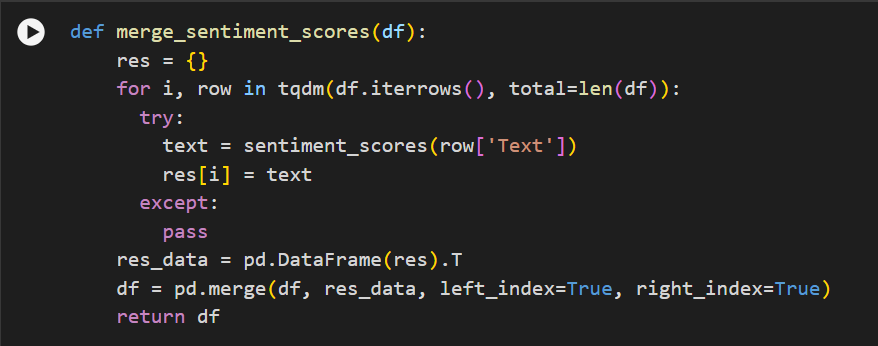
* Converts the text to lowercase
* Removes non-alphanumeric characters and digits.
* Tokenizes the text into individual words.
* Removes stopwords from the tokens.
* Lemmatizes each token using the WordNetLemmatizer.
* Joins the lemmatized tokens back into a preprocessed string.
* Returns the preprocessed text.

1. Extract and merge Sentiment Scores:









1. Collect Stock Data:

Imported Yahoo Finance for stock data.

A screen shot of a computer

Description automatically generated

Added these growth metrics base data frame of news, and named them respectively.

A screenshot of a computer

Description automatically generated

Collected open, close prices to label the day as green or red.

A screenshot of a computer

Description automatically generated

1. Label the Data:

If close price is greater than Open price then label it as 1, else 0.

A screen shot of a computer code

Description automatically generated

1. Train, Evaluate and Choose Best Model:

For training General Motors and Tesla prices are used, tested on Ford price.

A black background with orange text

Description automatically generated

Import various models like RandomForest, SVC, Logistic Regression, AdaBoost, XGBoost and CatBoost.

A screen shot of a computer

Description automatically generated

XGBoost gave best F1 score

A screen shot of a computer

Description automatically generated

1. Visualization:

A graph with lines and dots

Description automatically generated

1. Key metrics for testing on Ford price chart:

A screen shot of a black and white screen

Description automatically generated

Risk free rate taken: 4.7%

1. Conclusion:

Market news does affect stock price and can be clubbed with quantitative data to obtain stock specific buy/sell signals which tend to be more robust than just textual data.

However, the low win ratio indicates that most of the news is either stale or trap, still overall profitability and decent sharpe of 1.16 shows market reacts strongly to some news wher huge profits can be made.