Flipkart Reviews Sentimental Analysis Using Python

Project Description

A natural language processing (NLP) technique used to determine whether data is positive, negative or neutral. Sentimental analysis (or opinion mining) is often performed on textual data to help business brand monitor and product sentiment in customer feedback and to understand customer needs.

Modules Used In The Project:

-nltk

-wordcloud

-pandas

-matplotlib

-seaborn

Project overview

Project objective

This project involves the use of python packages like pandas and nltk (natural language tool kit) as this task is of sentiment analysis. The goal of this project is to analyse the sentiment of flipkart review and to accurately extract people's opinions from a large number of unstructured review texts and classifying them into sentiment classes.

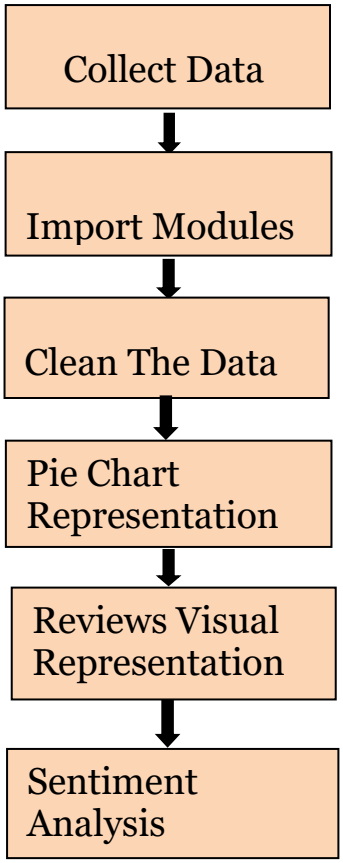
Project context

Flipkart is one of the most popular Indian companies. It is an e-commerce platform that competes with popular e-commerce platforms like Amazon. One of the most popular use cases of data science is the task of sentiment analysis of product reviews sold on e-commerce platforms.Here we will see how to analyze the sentiment of Flipkart reviews,I will walk you through the task of Flipkart reviews sentiment analysis using python.

Implementing the project will take you through the life of data analyst. From collecting data(in the form of sentences), cleaning the data, analyse the data and at last showing the results in visual representation using python modules.

Our task is to predict whether the review given is positive or negative.

Project stages



TASK 1

Collect Data:

We need to do data analysis! So for that we need to fetch the data and set up environment before moving into Data analysis.

The dataset I am using here for Flipkart reviews sentiment analysis is downloaded from Kaggle.

We need some requirements to do this project lets look into that:

-we will need a jupyter notebook installed in our computer.

-The data we are using should be csv file format.

TASK 2

Import the necessary modules:

Let's start this task by importing the necessary Python libraries and the dataset.

Go to the jupyter notebook ad import the modules.

```
In [3]: import pandas as pd
        import seaborn as sns
        import matplotlib.pyplot as plt
        from nltk.sentiment.vader import SentimentIntensityAnalyzer
        from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator

        data = pd.read_csv("https://raw.githubusercontent.com/amankharwal/Website-data
        print(data.head())
```

```
                                        Product_name  \
0  Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...
1  Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...
2  Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...
3  DELL Inspiron Athlon Dual Core 3050U - (4 GB/2...
4  DELL Inspiron Athlon Dual Core 3050U - (4 GB/2...

                                              Review  Rating
0  Best under 60k Great performanceI got it for a...       5
1                                Good perfomence...       5
2  Great performance but usually it has also that...       5
3          My wife is so happy and best product 👌😘       5
4  Light weight laptop with new amazing features,...       5
```

And here is the output, lets move into the next task....

TASK 3

Cleaning Data:

In this step we will see whether the is clean or not.

The dataset contains only three columns. Let's have a look at whether any of these columns contains missing values or not.

```
In [4]: print(data.isnull().sum())
```

```
Product_name    0
Review          0
Rating          0
dtype: int64
```

So, The dataset doesn't have any null values.As this is the task of sentiment analysis of flipkart reviews, I will clean and prepare the column containing reviews before heading to sentiment analysis:

```
In [5]: import nltk
        import re
        nltk.download('stopwords')
        stemmer = nltk.SnowballStemmer("english")
        from nltk.corpus import stopwords
        import string
        stopword=set(stopwords.words('english'))

        def clean(text):
            text = str(text).lower()
            text = re.sub('\[.*?\]', '', text)
            text = re.sub('https?://\S+|www\.\S+', '', text)
            text = re.sub('<.*?>+', '', text)
            text = re.sub('[%s]' % re.escape(string.punctuation), '', text)
            text = re.sub('\n', '', text)
            text = re.sub('\w*\d\w*', '', text)
            text = [word for word in text.split(' ') if word not in stopword]
            text=" ".join(text)
            text = [stemmer.stem(word) for word in text.split(' ')]
            text=" ".join(text)
            return text
        data["Review"] = data["Review"].apply(clean)
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\souja\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping corpora\stopwords.zip.
```

Here we've downloaded the 'stopwords' in english language.

TASK 4

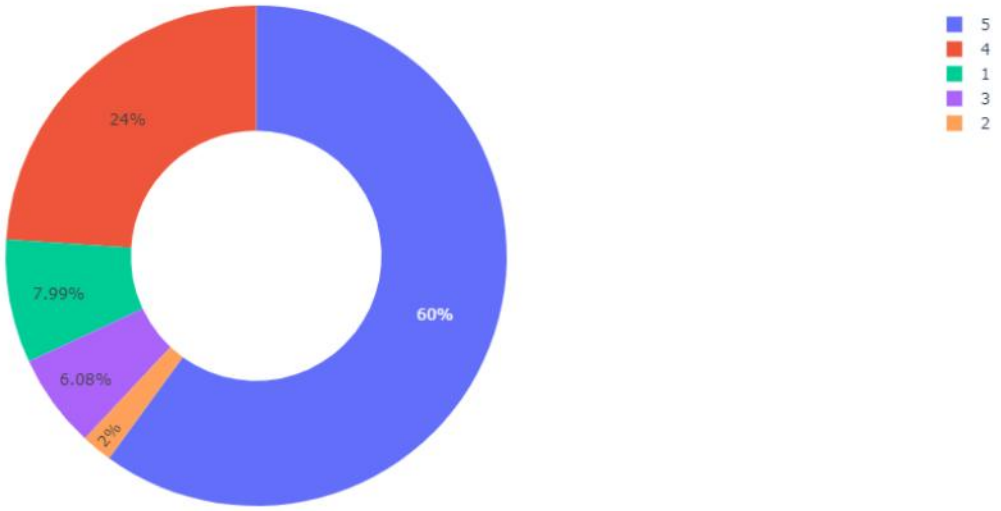
Pie Chart Representation

The Rating column of the data contains the ratings given by every reviewer. So let's have a look at how most of the people rate the products they buy from Flipkart:

```
In [6]: ratings = data["Rating"].value_counts()
        numbers = ratings.index
        quantity = ratings.values

        import plotly.express as px
        figure = px.pie(data,
                        values=quantity,
                        names=numbers,hole = 0.5)
        figure.show()
```

Output:



So 60% of the reviewers have given 5 out of 5 ratings to the products they buy from Flipkart. Now let's have a look at the kind of reviews people leave. For this, I will use a word cloud to visualize the most used words in the reviews column.

TASK 5

Reviews Visual Representation:

```
In [7]: text = " ".join(i for i in data.Review)
stopwords = set(STOPWORDS)
wordcloud = WordCloud(stopwords=stopwords,
                      background_color="white").generate(text)
plt.figure( figsize=(15,10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```

Output:

TASK 6

Sentiment Analysis Of Flipkart Reviews:

Now I will analyze the sentiments of Flipkart reviews by adding three columns in this dataset as Positive, Negative, and Neutral by calculating the sentiment scores of the reviews:

```
In [8]: nltk.download('vader_lexicon')
sentiments = SentimentIntensityAnalyzer()
data["Positive"] = [sentiments.polarity_scores(i)["pos"] for i in data["Review
data["Negative"] = [sentiments.polarity_scores(i)["neg"] for i in data["Review
data["Neutral"] = [sentiments.polarity_scores(i)["neu"] for i in data["Review"
data = data[["Review", "Positive", "Negative", "Neutral"]]
print(data.head())
```

```
[nltk_data] Downloading package vader_lexicon to
[nltk_data]     C:\Users\souja\AppData\Roaming\nltk_data...

                                              Review  Positive  Negative  \
0  best  great performancei got around  backup bi...     0.395     0.101
1                                        good perfom     0.744     0.000
2  great perform usual also game laptop issu batt...     0.277     0.000
3                          wife happi best product 👌😘     0.512     0.000
4  light weight laptop new amaz featur batteri li...     0.000     0.000

   Neutral
0    0.504
1    0.256
2    0.723
3    0.488
4    1.000
```

Now let's see how most of the reviewers think about the products and services of Flipkart:

```
In [9]: x = sum(data["Positive"])
        y = sum(data["Negative"])
        z = sum(data["Neutral"])

        def sentiment_score(a, b, c):
            if (a>b) and (a>c):
                print("Positive 😊 ")
            elif (b>a) and (b>c):
                print("Negative 😠 ")
            else:
                print("Neutral 🙂 ")
        sentiment_score(x, y, z)
```

Output:

```
Neutral 🙂
```

So most of the reviews are neutral. Let's have a look at the total of Positive, Negative, and Neutral sentiment scores to find a conclusion about Flipkart reviews:

```
In [10]: print("Positive: ", x)
         print("Negative: ", y)
         print("Neutral: ", z)
```

Output:

```
Positive:  923.5529999999985
Negative:  96.77500000000013
Neutral:  1283.6880000000006
```

CONCLUSION:

So, most people give Neutral reviews, and a small proportion of people give Negative reviews. So we can say that people are satisfied with Flipkart products and services.
In this analysis we've achieved;
-Collection of data
-Cleaning data
-Sentiment analysis of reviews.

Applications of sentiment analysis:

Sentiment analysis is the automated process of analyzing text to determine the sentiment expressed (positive, negative or neutral). Some popular sentiment analysis applications include social media monitoring, customer support management, and analyzing customer feedback.

Using natural language processing techniques, machine learning software is able to sort unstructured text by emotion and opinion.

Sentiment analysis is one of the most popular ways to analyze text, such assurvey responses, customer support issues, online reviews, and live chats, because it can help companies stay on top of customer satisfaction. .