Project Based Learning Report

on

Check for reviews on mobile phone wheather it is positive or negative using NLP toolbox

Submitted in the partial fulfillment of the requirements
For the Project based learning in (**AI & Data Mining**)
in
Electronics & Communication Engineering

By

2014111071 SAURAV KUMAR
2014111073 GAURAV KUMAR
2014111074 HARSHITA KUMARI

Under the guidance of Course In-charge

Prof. Dr. Tanuja Dhope

Department of Electronics & Communication Engineering

Bharati Vidyapeeth
(Deemed to be University)
College of Engineering,
Pune – 4110043

Academic Year: 2023-24

Bharati Vidyapeeth
(Deemed to be University)
College of Engineering,
Pune – 411043

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

CERTIFICATE

Certified that the Project Based Learning report entitled**, "Check for reviews on mobile phone wheather it is positive or negative using NLP toolbox"**
work is done by

2014111071 SAURAV KUMAR
2014111073 GAURAV KUMAR
2014111074 HARSHITA KUMARI

in partial fulfillment of the requirements for the award of credits for Project Based Learning (PBL) in **AI & Data Mining** of Bachelor of Technology Semester VII, in Electronics and communication engineering.

Date:

Dr. Tanuja Dhope — **Course In-charge**

Dr. Arundhati A. Shinde — **Professor & Head**

INDEX

Chapter 1

Description of the project Based Learning:

Check for reviews on mobile phone wheather it is positive or negative is based on sentiment analysis. In this we use NLP toolbox.

SENTIMENT ANALYSIS:

Sentiment analysis is a technique used to determine the emotional tone or sentiment expressed in a text. It involves analyzing the words and phrases used in the text to identify the underlying sentiment, whether it is positive, negative, or neutral.

Sentiment analysis has a wide range of applications, including social media monitoring, customer feedback analysis, and market research.

One of the main challenges in sentiment analysis is the inherent complexity of human language. Text data often contains sarcasm, irony, and other forms of figurative language that can be difficult to interpret using traditional methods.

However, recent advances in **natural language processing** (NLP) and machine learning have made it possible to perform sentiment analysis on large volumes of text data with a high degree of accuracy.

Three Methodologies for Sentiment Analysis

There are several ways to perform sentiment analysis on text data, with varying degrees of complexity and accuracy. The most common methods include a lexicon-based approach, a machine learning (ML) based approach, and a pre-trained transformer-based deep learning approach. Let's look at each in more detail:

Lexicon-based analysis

This type of analysis, such as the NLTK Vader sentiment analyzer, involves using a set of predefined rules and heuristics to determine the sentiment of a piece of text. These rules are typically based on lexical and syntactic features of the text, such as the presence of positive or negative words and phrases.

While lexicon-based analysis can be relatively simple to implement and interpret, it may not be as accurate as ML-based or transformed-based approaches, especially when dealing with complex or ambiguous text data.

Machine learning (ML)

This approach involves training a model to identify the sentiment of a piece of text based on a set of labeled training data. These models can be trained using a wide range of ML algorithms, including decision trees, support vector machines (SVMs), and neural networks.

ML-based approaches can be more accurate than rule-based analysis, especially when dealing with complex text data, but they require a larger amount of labeled training data and may be more computationally expensive.

Pre-trained transformer-based deep learning

A deep learning-based approach, as seen with BERT and **GPT-4**, involve using pre-trained models trained on massive amounts of text data. These models use complex neural networks to encode the context and meaning of the text, allowing them to achieve state-of-the-art accuracy on a wide range of NLP tasks, including sentiment analysis. However, these models require significant computational resources and may not be practical for all use cases.

- **Lexicon-based analysis** is a straightforward approach to sentiment analysis, but it may not be as accurate as more complex methods.
- **Machine learning-based** approaches can be more accurate, but they require labeled training data and may be more computationally expensive.
- **Pre-trained transformer-based deep learning** approaches can achieve state-of-the-art accuracy but require significant computational resources and may not be practical for all use cases.

The Natural Language Toolkit (NLTK) Library

The Natural Language Toolkit (NLTK) is a popular open-source library for natural language processing (NLP) in Python. It provides an easy-to-use interface for a wide range of tasks, including tokenization, stemming, lemmatization, parsing, and sentiment analysis.

NLTK is widely used by researchers, developers, and data scientists worldwide to develop NLP applications and analyze text data.

One of the major advantages of using NLTK is its extensive collection of corpora, which includes text data from various sources such as books, news articles, and social media platforms. These corpora provide a rich data source for training and testing NLP models.

Chapter 2

Configure in Jupyter Notebook:

To configure check for review on mobile phone firstly we have to open Jupyter Notebook after the jupter notebook open following are the steps:

1.Firstly we have to downloaded the dataset phone review from website kaggle.
2. we upload the dataset in jupyter notebook
3. we import the Library:

```
In [2]: import numpy as np
        import pandas as pd
        import seaborn as sns
        import matplotlib.pyplot as plt
        from nltk.sentiment.vader import SentimentIntensityAnalyzer
        sentiments = SentimentIntensityAnalyzer()
```

4. we retrive the data using pandas

```
In [3]: data = pd.read_csv("phone reviews.csv")
        print(data.head())
```

5. use to explore the data using different pandas function:

data exploration

```
n [4]: print(data.describe())
```

```
In [5]: data = data.dropna()
        data
```

```
In [12]: df = data.head(10)
         df
```

6. use to visualize the data using Seaborn Library and Matplotlib Library:

Data Visualization

```
[9]: plt.figure(figsize=(16,7))

     plt.subplot(2,2,1)
     sns.histplot(data['rating'],color='DeepPink')
     plt.subplot(2,2,2)
     sns.boxplot(data['rating'],color='yellow')

     plt.tight_layout()
     plt.show()
```

```
In [13]: ax = sns.countplot(x='helpfulVotes', data=df)
         ax.bar_label(ax.containers[0])
         plt.title("helpfulVotes", fontsize=20)
```

```
In [15]: g = sns.FacetGrid(df, col="helpfulVotes")
         g = g.map(plt.hist, "rating")
```

```
In [16]: sns.barplot(y='rating',x='name',data=df)
         plt.show()
```

In [17]:
```
f, ax = plt.subplots(figsize=(8, 6))
sns.pointplot(x="name", y="helpfulVotes", data=df)
plt.show()
```

3]:
```
sns.histplot(df['helpfulVotes'],color="r",bins=20,kde=True,)
plt.show()
```

In [5]:
```
ratings = data["rating"].value_counts()
numbers = ratings.index
quantity = ratings.values
```

In [6]:
```
custom_colors = ["skyblue", "yellowgreen", 'tomato', "blue", "red"]
plt.figure(figsize=(10, 8))
plt.pie(quantity, labels=numbers, colors=custom_colors)
central_circle = plt.Circle((0, 0), 0.5, color='white')
fig = plt.gcf()
fig.gca().add_artist(central_circle)
plt.rc('font', size=12)
plt.title("Product Ratings", fontsize=20)
plt.show()
```

7. Then analyse the sentiment data using NLTK library:

Sentiment Analysis

]:
```
# sentiments = SentimentIntensityAnalyzer()
data["Positive"] = [sentiments.polarity_scores(i)["pos"] for i in data["body"]]
data["Negative"] = [sentiments.polarity_scores(i)["neg"] for i in data["body"]]
print(data.head())
```

]:
```
x = sum(data["Positive"])
y = sum(data["Negative"])
```

]:
```
def sentiment_score(a, b):
    if (a>b) :
        print("Positive 😊 ")
    elif (b>a):
        print("Negative 😠 ")
```

]:
```
sentiment_score(x, y)
```

Chapter 3

Results and Analysis with photograph:

In [2]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from nltk.sentiment.vader import SentimentIntensityAnalyzer
sentiments = SentimentIntensityAnalyzer()
```

In [3]:
```
data = pd.read_csv("phone reviews.csv")
print(data.head())
```

```
         asin           name  rating               date  verified  \
0  B0000SX2UC          Janet       3   October 11, 2005     False
1  B0000SX2UC     Luke Wyatt       1    January 7, 2004     False
2  B0000SX2UC         Brooke       5  December 30, 2003     False
3  B0000SX2UC  amy m. teague       3     March 18, 2004     False
4  B0000SX2UC  tristazbimmer       4    August 28, 2005     False

                                      title  \
0                  Def not best, but not worst
1                  Text Messaging Doesn't Work
2                             Love This Phone
3                     Love the Phone, BUT...!
4  Great phone service and options, lousy case!

                                                body  helpfulVotes
0  I had the Samsung A600 for awhile which is abs...           1.0
1  Due to a software issue between Nokia and Spri...          17.0
2  This is a great, reliable phone. I also purcha...           5.0
3  I love the phone and all, because I really did...           1.0
4  The phone has been great for every purpose it ...           1.0
```

data exploration

In [4]:
```
print(data.describe())
```

```
              rating  helpfulVotes
count  67986.000000  27215.000000
mean       3.807916      8.229690
std        1.582906     31.954877
min        1.000000      1.000000
25%        3.000000      1.000000
50%        5.000000      2.000000
75%        5.000000      5.000000
max        5.000000    990.000000
```

In [5]:
```
data = data.dropna()
data
```

Out[5]:

	asin	name	rating	date	verified	title	body	helpfulVotes
0	B0000SX2UC	Janet	3	October 11, 2005	False	Def not best, but not worst	I had the Samsung A600 for awhile which is abs...	1.0
1	B0000SX2UC	Luke Wyatt	1	January 7, 2004	False	Text Messaging Doesn't Work	Due to a software issue between Nokia and Spri...	17.0
2	B0000SX2UC	Brooke	5	December 30, 2003	False	Love This Phone	This is a great, reliable phone. I also purcha...	5.0
3	B0000SX2UC	amy m. teague	3	March 18, 2004	False	Love the Phone, BUT...!	I love the phone and all, because I really did...	1.0
4	B0000SX2UC	tristazbimmer	4	August 28, 2005	False	Great phone service and options, lousy case!	The phone has been great for every purpose it ...	1.0
...	...	...	...	...	...	...	...	...
67978	B081H6STQQ	Rock Edge	5	July 16, 2019	False	Candy bar phone is back!	Update 8/14/19 5 stars now! I've been using th...	12.0
67979	B081H6STQQ	Cindy Cowles	3	July 17, 2019	False	Updated review	Update: If you like sending and receiving pict...	2.0
67980	B081H6STQQ	Los Kositos	5	October 7, 2019	False	From iPhone to Android and loving it	I love my new phone. I've been dying to do a r...	1.0
67981	B081H6STQQ	jande	5	August 16, 2019	False	Awesome Phone, but finger scanner is a big mis...	I love the camera on this phone. The screen is...	1.0
67982	B081H6STQQ	2cool4u	5	September 14, 2019	False	Simply Amazing!	I've been an Xperia user for several years and...	1.0

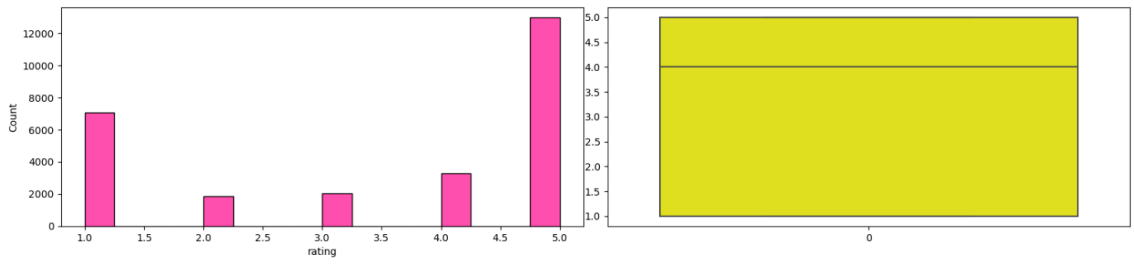
In [12]:
```
df = data.head(10)
df
```

Out[12]:

	asin	name	rating	date	verified	title	body	helpfulVotes
0	B0000SX2UC	Janet	3	October 11, 2005	False	Def not best, but not worst	I had the Samsung A600 for awhile which is abs...	1.0
1	B0000SX2UC	Luke Wyatt	1	January 7, 2004	False	Text Messaging Doesn't Work	Due to a software issue between Nokia and Spri...	17.0
2	B0000SX2UC	Brooke	5	December 30, 2003	False	Love This Phone	This is a great, reliable phone. I also purcha...	5.0
3	B0000SX2UC	amy m. teague	3	March 18, 2004	False	Love the Phone, BUT...!	I love the phone and all, because I really did...	1.0
4	B0000SX2UC	tristazbimmer	4	August 28, 2005	False	Great phone service and options, lousy case!	The phone has been great for every purpose it ...	1.0
6	B0000SX2UC	the cell phone store owner	5	April 16, 2004	False	Wanna cool Nokia? You have it here!	Cool. Cheap. Color: 3 words that describe the ...	2.0
7	B0000SX2UC	Matt	4	April 3, 2004	False	Problem with 3588i universal headset	The 3599i is overall a nice phone, except that...	2.0
8	B0000SX2UC	Charles Cook	5	November 24, 2003	False	cool phone!!!!!!!	I've never owned a Nokia phone before, so this...	7.0
9	B0000SX2UC	Amazon Customer	3	February 2, 2004	False	Pissed off-a little bit	ok well im in school and i need the text messa...	3.0
10	B0000SX2UC	habblie	4	December 25, 2004	False	works great, but don't dropt it	I've had this phone for over a year and I real...	1.0

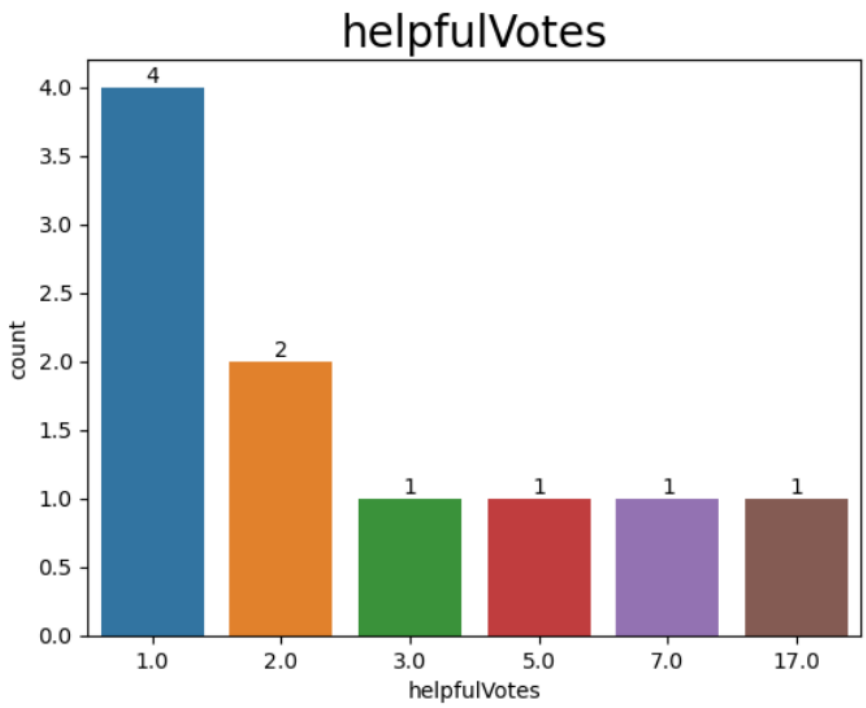
Data Visualization

9]:
```
plt.figure(figsize=(16,7))

plt.subplot(2,2,1)
sns.histplot(data['rating'],color='DeepPink')
plt.subplot(2,2,2)
sns.boxplot(data['rating'],color='yellow')

plt.tight_layout()
plt.show()
```

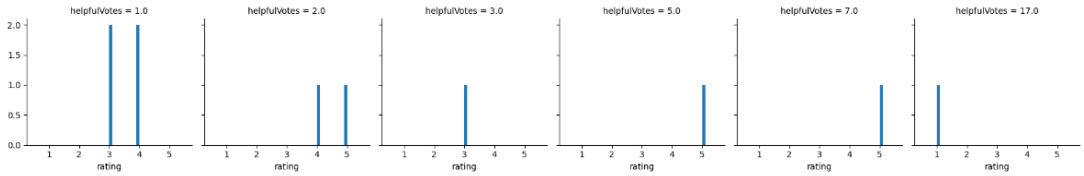


In [13]:
```
ax = sns.countplot(x='helpfulVotes', data=df)
ax.bar_label(ax.containers[0])
plt.title("helpfulVotes", fontsize=20)
```

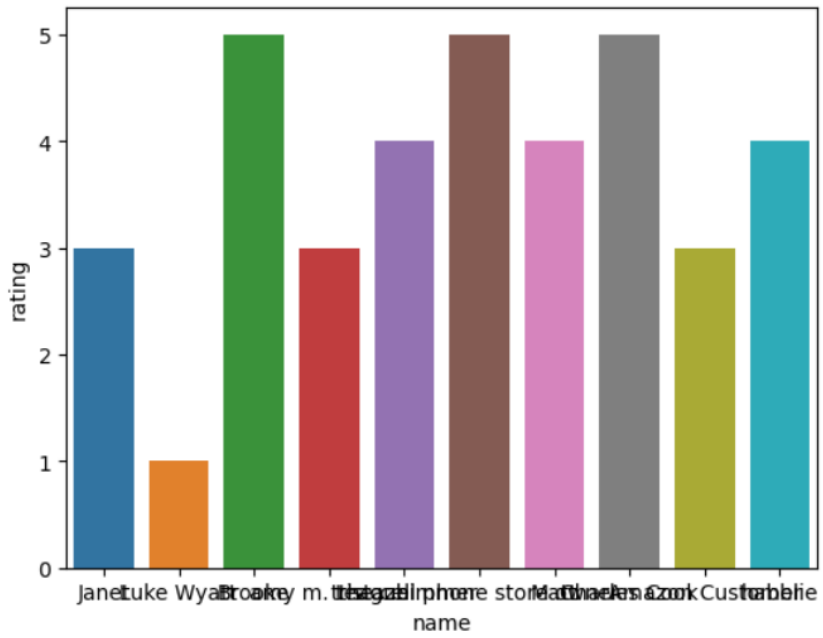
Out[13]: Text(0.5, 1.0, 'helpfulVotes')



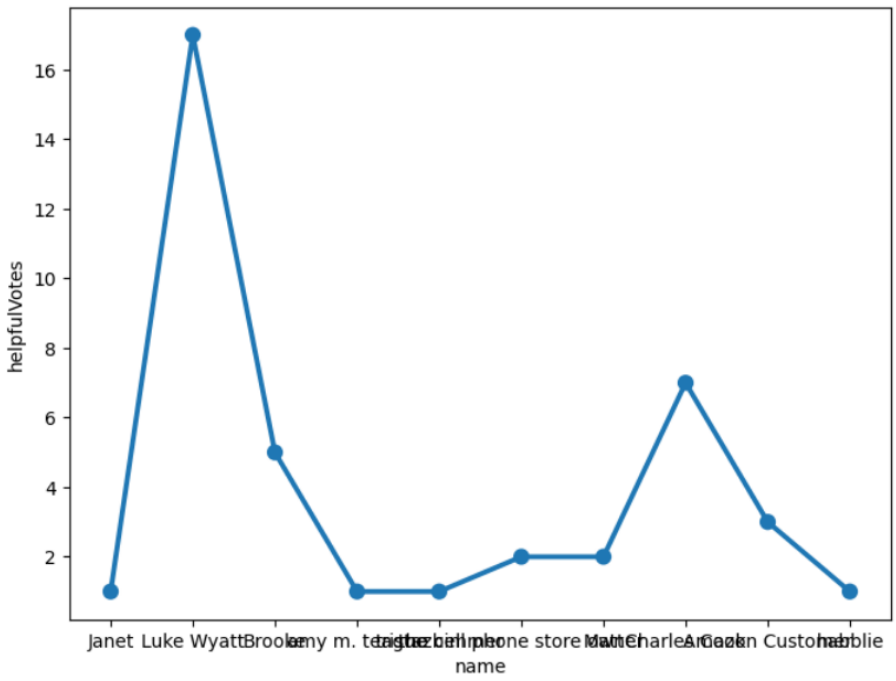
In [15]:
```
g = sns.FacetGrid(df, col="helpfulVotes")
g = g.map(plt.hist, "rating")
```



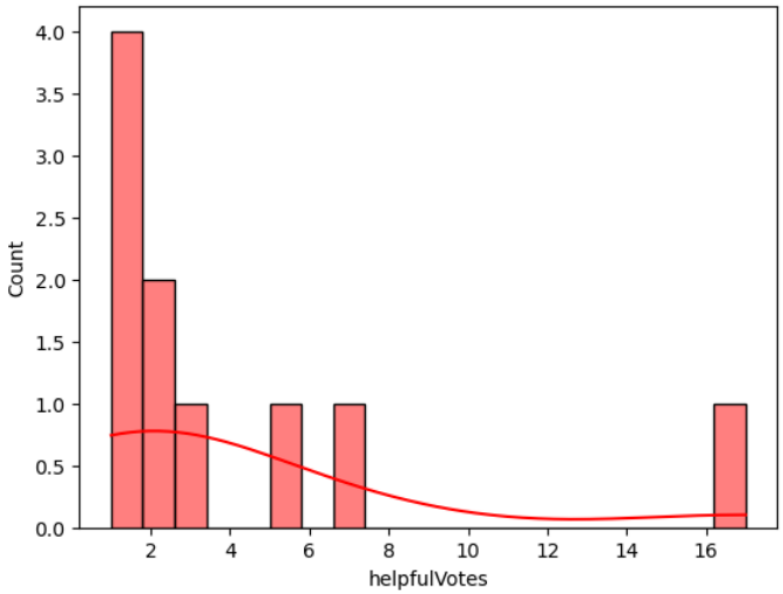
In [16]:
```
sns.barplot(y='rating',x='name',data=df)
plt.show()
```



In [17]:
```
f, ax = plt.subplots(figsize=(8, 6))
sns.pointplot(x="name", y="helpfulVotes", data=df)
plt.show()
```

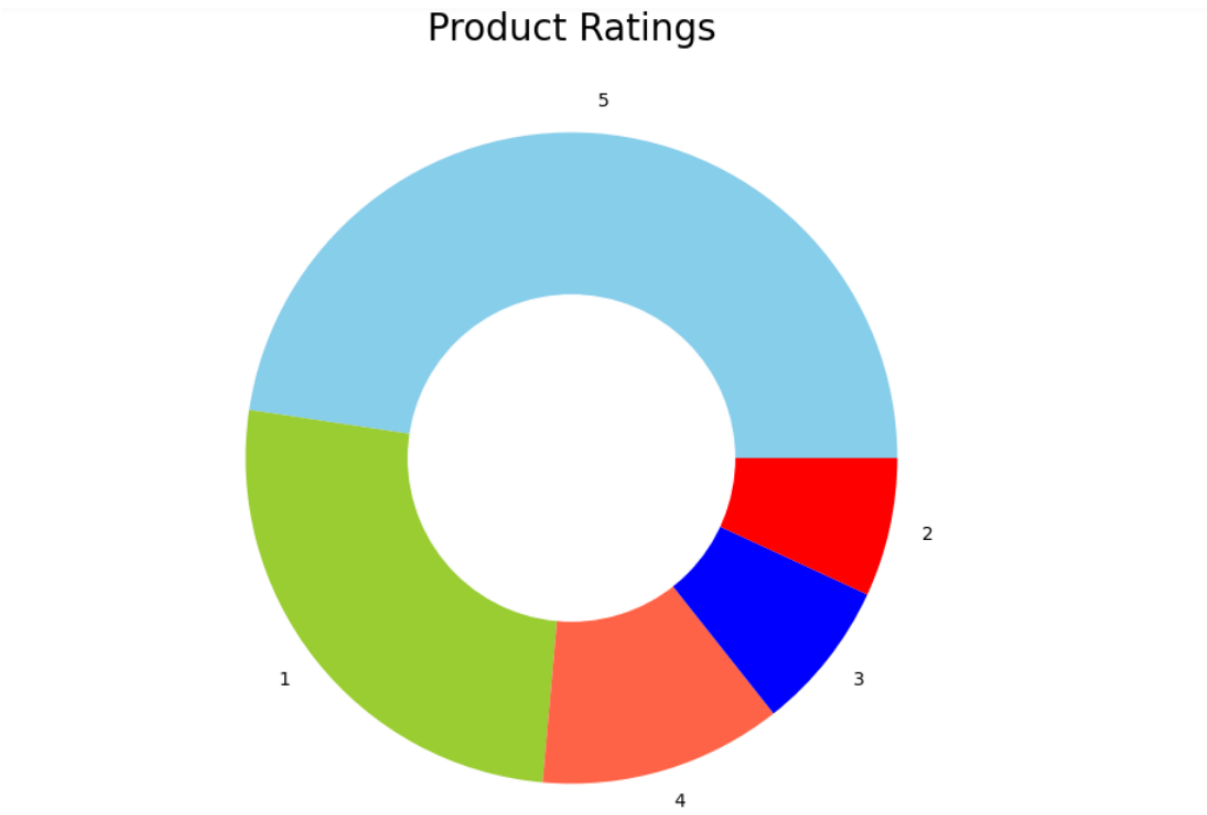


In [23]:
```
sns.histplot(df['helpfulVotes'],color="r",bins=20,kde=True,)
plt.show()
```



In [5]:
```
ratings = data["rating"].value_counts()
numbers = ratings.index
quantity = ratings.values
```

In [6]:
```
custom_colors = ["skyblue", "yellowgreen", 'tomato', "blue", "red"]
plt.figure(figsize=(10, 8))
plt.pie(quantity, labels=numbers, colors=custom_colors)
central_circle = plt.Circle((0, 0), 0.5, color='white')
fig = plt.gcf()
fig.gca().add_artist(central_circle)
plt.rc('font', size=12)
plt.title("Product Ratings", fontsize=20)
plt.show()
```



Sentiment Analysis

In [ ]:
```
# sentiments = SentimentIntensityAnalyzer()
data["Positive"] = [sentiments.polarity_scores(i)["pos"] for i in data["body"]]
data["Negative"] = [sentiments.polarity_scores(i)["neg"] for i in data["body"]]
print(data.head())
```

In [8]:
```
x = sum(data["Positive"])
y = sum(data["Negative"])
```

In [13]:
```
def sentiment_score(a, b):
    if (a>b) :
        print("Positive 😊 ")
    elif (b>a):
        print("Negative 😠 ")
```

In [14]:
```
sentiment_score(x, y)
```

```
Positive 😊
```

In [15]:
```
print("Positive: ", x)
print("Negative: ", y)
```

```
Positive:  5022.301000000003
Negative:  1756.803999999994
```

In [ ]:

Chapter 4

Outcome:

This project satisfy the CO4 outcome.in this we have to understand the basic concept of data mining and its functionality.

Conclusion:

This topic understand the data collection , data preprocessing, sentiment Analysis how we collect the data preprocess using different library and analyse sentiment data using NLTK library.