

SENTIMENT ANALYSIS ON AMAZON CUSTOMER'S REVIEW USING NLP

BY

Renesa Bente Maula

ID: 201-15-3101

This Report Presented in Partial Fulfillment of the Requirements for the
Degree of Bachelor of Science in Computer Science and Engineering.

Supervised By

Dr. S.M. Aminul Haque

Associate Professor

Department of CSE

Daffodil International University

Co-Supervised By

Md. Mahfujur Rahman

Lecturer (Senior Scale)

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY


DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Project/internship titled “SENTIMENT ANALYSIS ON AMAZON CUSTOMER’S REVIEW USING NLP”, submitted by Renesa Bente Maula ID: 201 – 15 – 3101 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 19th January 2023.

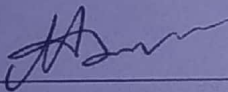
BOARD OF EXAMINERS



Dr. Touhid Bhuiyan
Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

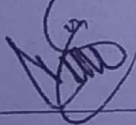
Chairman



Nazmun Nessa Moon
Associate Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

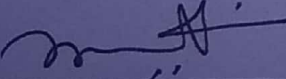
Internal Examiner



Md. Abbas Ali Khan
Assistant Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Mohammad Shorif Uddin
Professor

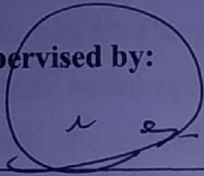
Department of Computer Science and Engineering
Jahangirnagar University

External Examiner

DECLARATION

I hereby declare that, this thesis base project has been done by me under the supervision of **Dr. S.M. Aminul Haque, Associate Professor of CSE Department, Daffodil International University.** I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Dr. S.M. Aminul Haque
Associate Professor
Department of CSE
Daffodil International University

Submitted by:



Renesa Bente Maula
ID: 201-15-3101
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Dr. S.M. Aminul Haque, Associate Professor of CSE Department, Daffodil International University.** Deep Knowledge & keen interest of our supervisor in the field of “Sentiment Analysis” to carry out this project. Our project is based on ”NLP”. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Md. Mahfujur Rahman,** Lecturer(Senior Scale), Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

The business environment of today has grown incredibly competitive and difficult. Growth of businesses now places a lot of emphasis on customer happiness. To understand and meet the demands of their clients, business organizations devote a significant amount of money and human resources to different techniques. But many businesses are failing to satisfy customers as a result of the manual analysis of customers varied wants being done in an imperfect way. As a result, they are losing their customers' trust and increasing their marketing expenses. Sentiment Analysis is a solution that we can use to resolve the issues. Machine learning (ML) and natural language processing are both included into the system (NLP). Analysis of people's feelings about certain topics, products, and services is called sentiment analysis, and it is used rather often to get insights into how the general public thinks about certain topics, products, and services. We are able to do that by using any data that is found online. In this article, we present two natural language processing strategies (Bag-of-Words and TF-IDF) as well as various machine learning classification techniques to perform sentiment analysis on a large dataset that is imbalanced and contains many classes of data.

TABLE OF CONTENTS

| CONTENTS | PAGE |
|--|-------------|
| Board of examiners | i |
| Declaration | ii |
| Acknowledgements | iii |
| Abstract | iv |
| Table of contents | v-vi |
| List of figures | vii |
| List of tables | viii |
| | |
| CHAPTER | |
| | |
| CHAPTER 1: INTRODUCTION | 1-3 |
| 1.1 Introduction | 1 |
| 1.2 Motivation | 1-2 |
| 1.3 Rationale of the Study | 2 |
| 1.4 Objectives | 2 |
| 1.5 Expected Outcome | 2 |
| 1.6 Report Layout | 3 |
| | |
| CHAPTER 2: BACKGROUND | 4-6 |
| 2.0 Terminologies | 4 |
| 2.1 Related Works | 4-6 |
| 2.2 Research Summary | 6 |
| 2.3 Scope of the problem | 6 |
| | |
| CHAPTER 3: RESEARCH METHODOLOGY | 7-24 |
| 3.0 Introduction | 7 |
| 3.1 Data Source | 7 |
| 3.2 Data collection and preparation | 7-24 |
| ©Daffodil International University | v |

| | |
|--|--------------|
| CHAPTER 4: EXPERIMENTAL RESULTS | 25-28 |
| 4.0 Fetch positive & negative words | 25-26 |
| 4.1 Upvote(s) prediction | 27-28 |
| CHAPTER 5: CONCLUSION AND FUTURE WORK | 29 |
| 5.0 Conclusion | 29 |
| 5.1 Future work | 29 |
| REFERENCES | 30 |

LIST OF FIGURES

| FIGURES | PAGE NO |
|---|----------------|
| Figure 3.1: Dataset | 8 |
| Figure 3.2: Importing libraries before preprocessing | 11 |
| Figure 3.3: Preprocessing | 11 |
| Figure 3.4: Adding & Defining column | 12 |
| Figure 3.5: EDA | 14 |
| Figure 3.6: Importing Seaborn | 15 |
| Figure 3.7: Heatmap | 16 |
| Figure 3.8: BoW | 17 |
| Figure 3.9: TF law | 18 |
| Figure 3.10: IDF law | 18 |
| Figure 3.11: TF-IDF | 19 |
| Figure 3.12: Imbalance data & Cross validation | 23 |
| Figure 4.1: RandomOverSampler | 24 |
| Figure 4.2: Positive Words | 25 |
| Figure 4.3: Negative Words | 26 |
| Figure 4.4: Upvotes Prediction | 27 |
| Figure 4.4: Upvotes Prediction after removing neutral votes | 28 |

LIST OF TABLES

| TABLE | PAGE NO |
|-------------------------------------|----------------|
| Table 3.1: Range of imbalanced data | 22 |

CHAPTER 1

Introduction

1.1 Introduction

Customer contentment measures how consumers feel about businesses, products, and services. Numerous research have found that customer contentment and product or service quality are the two most important elements influencing business performance. To remain competitive, businesses must carefully evaluate what their customers need and want from the products or services they offer. In addition, they must effectively manage their clientele by ensuring that their customers are pleased to conduct business with them. In social networks and other platforms, unstructured data predominates. It takes time to gather customer opinions and make the appropriate judgments using this data. Sentiment analysis is an easier method for determining the views of individuals.[4] The basic objective of sentiment analysis is to categorize the positivity, negativity, or neutrality of textual data. Using sentiment analysis tools, decision-makers may track shifts in public or consumer sentiment towards different entities, activities, goods, technologies, and services. Sentiment research may help a business swiftly improve its products and services, as well as a political party or social group generate high-quality work. Sentiment analysis facilitates rapid comprehension of the public's opinion. The vast bulk of the data used for sentiment analysis is collected from social media sites and stored in databases known as datasets.[7] Nonetheless, sentiment analysis becomes challenging when datasets are uneven, massive, multi-classified, etc. Types of Sentiment Analysis was best described by Mayur Wankhade, Annavarapu Chandra Sekhara Rao & Chaitanya Kulkarni. [2] Amazon provided me with a big, imbalanced, multi-classed, real-world dataset for my study. I've preprocessed and vectorized the data using NLP methods. The polarity of textual data was then identified using Machine Learning classification techniques.[8] Bag-of-Words and Term Frequency - Inverse Document Frequency are the used NLP approaches.

1.2 Motivation

- To get hands-on experience on sentiment analysis.
- Detect and understand customer feelings.
- Improve product quality

- To get experience in real life with raw data.

1.3 Rationale of the Study

Sentiment analysis can increase customer retention and loyalty by improving customer service and help outcomes.[2] To assess if a document is neutral, optimistic, or unbiased, Artificial Intelligence (AI) and Natural Language Processing are used (NLP).

Businesses may use sentiment analysis to produce more relevant messaging and better consumer communication.[3] Understanding the consumers' emotions helps us understand their experiences and improve customer service,[9] both of which eventually reduce customer turnover.

1.4 Objectives

The major goal of this study is to shed light on the specific emotions that underlie each review someone writes. We have taken real-world reviews from Amazon.com and used them in this. This paper provides a synopsis of the tools and strategies used in sentiment analysis as well as the issues that the discipline faces. This study also provides a thorough understanding of NLP. Its primary goal is to gather reviews from the web and forecast online users' preferences and opinions on a certain product, information that could be useful for economic or marketing research. Text analysis methods called sentiment analysis automatically identify the polarity of texts. For example, it aids in marketing when evaluating the performance of an advertising campaign or the introduction of a new product, identifying the most popular variations of a good or service, and even pinpointing the demographics who favor or disapprove of specific features. The word "opinion mining" refers to sentiment analysis as well.[1] Sentiment analysis aids in making decisions more intelligently by enabling users to click a mouse button to access more detailed and pertinent information about various goods and services.

1.5 Expected Outcome

- Get analyzed upvote(s) for different scores.
- Show the predicted score with respect to various customers, whatever the sentiment of the customer.
- Fetch positive and negative words of the developed NLP & ML model.

1.6 Report Layout

This paper is organized as follows:

- I. Background
- II. Research Methodology
- III. Experimental Results and Discussion
- IV. Summary, Conclusion, Recommendation and implication for future Research.
- V. Reference

CHAPTER 2

Background Study

2.0 Terminologies

The purpose of doing a sentiment analysis is to exactly glean the opinions of others from large amounts of unstructured review material and classify them as positive, negative, or neutral sentiment. At first, I have collected the dataset from Amazon Platform. Then, I have performed preprocessing on the collected Amazon dataset. After preprocessing, EDA has been applied to the data. Then I have applied Bag of Words (BoW) on the data. After that I've automated the NLP & ML model. Then applied the algorithms of NLP and Machine learning. Also, I've prepared the data modeling purpose.

2.1 Related Works

The use of sentiment analysis in social media monitoring allows businesses to understand more about how their customers feel about specific topics and discover important concerns in real time before they escalate. The other uses for sentiment analysis are:

- Analysis of customer support ticketing system
- Brand surveillance and reputation administration
- Pay attention to the voice of the customer (VoC)
- Pay attention to the voice of the employee
- Product evaluation
- Market & rivalry analysis

Analysis of Customer support ticketing system

Customer service management is particularly challenging because of the sheer number of requests, the range of themes, and the multiplicity of departments not to mention the urgency of each request inside a firm. In order to comprehend consumer demands, exactly like a person would.[6] Natural language understanding (NLU) and sentiment analysis look for meaning, emotion, tone, and other factors in spoken language.

Brand surveillance and reputation administration

Brand surveillance is one of the most popular applications of sentiment analysis in business. Negative online reviews tend to snowball, and the longer you ignore them, the more damage you'll do. You will receive fast notification when a brand is mentioned negatively thanks to sentiment analysis technologies.[7]

You may also track the reputation and image of your brand over time or at any particular time to measure your success. By monitoring news articles, blogs, forums, and social media for mentions of your business,[9] you may transform this data into useful information and analytics. You may rely on machine learning to keep up with developments, foresee outcomes, and take proactive rather than reactive action.

Pay attention to the voice of the customer (VoC)

You should aggregate and evaluate all of your customers' online, survey, chat, contact center, and email feedback.[8] To identify trends and recurring problems, you may arrange and categorize this data using sentiment analysis.

You may develop a customized customer experience by hearing what your consumers have to say and learning what approaches of communication work and what doesn't.[10]

Pay attention to your employees

You can enhance productivity, lower turnover, and better engage your workforce by examining the tone of employee feedback.[5]

Utilize sentiment analysis to evaluate employee surveys, emails, Slack conversations, Glassdoor reviews, and other data. Also, utilize unstructured data to comprehend the "why" behind the "who" and "what" and to identify the most frequent concerns and themes that will keep your employees satisfied and productive.

Product evaluation

As soon as new product is released , find out about public opinion. You may also examine feedback from years ago that you may have never seen. For a certain product feature (interface, UX, functionality), you can conduct a keyword search and use aspect-based

sentiment analysis to locate just the information you need. Determine how your target market perceives a product, what has to be altered about it, and what will please your most valued customers, each with sentiment analysis.

Market and rivalry analysis

Use sentiment analysis for market and rivalry analysis. Determine which of your competitors are being favorably mentioned and how your marketing efforts compare.

Examine the positive customer-facing language used by your rivals and integrate some of it into your brand messaging and style guide.

2.2 Research Summary

In this research, mainly I'm going to analyze a huge chunk of Amazon data and from the data, I'll extract some meaningful insights to analyze the customer's sentiments by performing several techniques of cleaning data processing on raw data.

2.3 Scope

A lot of research has been done and are ongoing on Sentiment Analysis. This is my attempt to understand the sentiments of the customers based on their reviews. I've analyzed the data which I've collected from the Amazon platform. Also, I've focused on the relevant papers that focused on Sentiment Analysis and NLP.

CHAPTER 3

Research Methodology

3.0 Introduction

The goal of the paper was to do a comparative study based on prediction model to understand the sentiments of the Amazon customers. For that, I've used a raw data set which has been collected from Amazon platform.

Also, I've used various plots and graphs that are available in python including Matplotlib and Seaborn. Before processing and vectorizing data, I used NLP methods. The polarity of textual information was then identified using classification approaches based on Machine Learning (ML). Bag-of-Words and Term Frequency - Inverse Document Frequency is the NLP method used.

3.1 Data source

I have experimented with a massive, imbalanced, multi-classed, and real-world dataset, which I've collected from Amazon.

3.2 Data Collection and Preparation Methodology

There are the below steps for my approaches:

- Data collection for ML classifier training & testing.
- Pre-processing the dataset in preparation for further processing.
- Apply Exploratory Data Analysis on data.
- Apply Bag of Words on data.
- Using NLP to convert textual data to vector format.
- Check Model Accuracy.
- Automate NLP and ML model.
- Preparing data for modeling purpose.
- Apply techniques of handling imbalanced data.

- Remove Noisy Data
- Tokenization
- Normalization

Remove Noisy Data

Examples of noisy data include the headers and footers of text files, as well as HTML, XML, and markup data.[13] It is necessary to delete certain types of noisy data because they are useless and don't supply any useful information.

Tokenization

Sentences in groups are tokenized in this process. Text segmentation or lexical analysis are other names for it.[12] In essence, it divides the data into small word chunks. For example, the sentence- “Julia122 rides a bicycle, can I ride?” After tokenization, this sentence will become- [‘Julia’, ‘122’, ‘rides’, ‘a’, ‘bicycle’, ‘,’’, ‘can’, ‘I’, ‘ride’, ‘?’]

Normalization

Before proceeding to normalization, we must first examine the tokenization result.[11] The goal of normalization is to extract more meaningful information from tokenized data.

Libraries used

The very first thing I need to do is load data preparation libraries. There are other libraries available, but the most popular and useful Python data libraries are Numpy, Matplotlib, and Pandas.

NLP Model Used

I have used many NLP algorithms like Bag of Words (BoW), and Term-Frequency Inverse-Document-Frequency (TF-IDF) from intuition. And application point of view as well to convert my text features into the numerical feature.

Pandas

The vast majority of its applications are found in the domains of data science and machine

learning, and Pandas is an open-source Python library. Numpy, a separate package that enables the manipulation of multidimensional arrays, serves as the basis for its construction. This particular tool is the most effective one for importing and maintaining datasets. One of the data manipulation packages that sees the most use overall in the Python ecosystem is called Pandas. It works well with a wide variety of other modules that are used in data science, and it is typically included in all Python distributions.[13] This includes distributions that are sold by commercial vendors, such as ActiveState's ActivePython, as well as distributions that are included with your operating system. Pandas make it easy to complete a significant number of the laborious and recurring chores that are connected with dealing with data, including:

- Data cleaning
- Data normalization
- Data Visualization
- Statistical evaluation
- Analysis of the data
- Loading and updating data
- And even more

Pandas allows you to perform all of the tasks that have led top data scientists around the world to name it the best tool for data analysis and manipulation.

We can import pandas with a shortcut alias like: `import pandas as pd` .

```

In [1]: import pandas as pd
import numpy as np

In [2]: df=pd.read_csv(r'D:\nlp\Reviews.csv')
df.head()

Out[2]:

```

| | Id | ProductId | UserId | ProfileName | HelpfulnessNumerator | HelpfulnessDenominator | Score | Time | Summary | Text |
|---|----|------------|----------------|---------------------------------|----------------------|------------------------|-------|------------|-----------------------|---|
| 0 | 1 | B001E4KFG0 | A3SGXH7AUHU8GW | delmartan | 1 | 1 | 5 | 1303862400 | Good Quality Dog Food | I have bought several of the Vitality canned d... |
| 1 | 2 | B00813GRG4 | A1D87F6ZCVE5NK | dli pa | 0 | 0 | 1 | 1346976000 | Not as Advertised | Product arrived labeled as Jumbo Salted Peanut... |
| 2 | 3 | B000LQOCH0 | ABXLMWJXXAIN | Natalia Corres "Natalia Corres" | 1 | 1 | 4 | 1219017600 | "Delight" says it all | This is a confection that has been around a fe... |
| 3 | 4 | B000UA0QIQ | A395BORCF6GVXV | Karl | 3 | 3 | 2 | 1307923200 | Cough Medicine | If you are looking for the secret ingredient l... |
| 4 | 5 | B006KZZZ7K | A1UQRSLF8GWT | Michael D. Bigham "M. Wassir" | 0 | 0 | 5 | 1350776000 | Great taffy | Great taffy at a great price. There was a wid... |

```

In [3]: df.columns
Out[3]: Index(['Id', 'ProductId', 'UserId', 'ProfileName', 'HelpfulnessNumerator', 'HelpfulnessDenominator', 'Score', 'Time', 'Summary', 'Text'], dtype='object')

In [4]: df['Helpful%']=np.where(df['HelpfulnessDenominator']>0,df['HelpfulnessNumerator']/df['HelpfulnessDenominator'],-1)

In [5]: df.head()

```

Figure 3.2: importing libraries before preprocessing

```

In [5]: df.head()

Out[5]:

```

| | Id | ProductId | UserId | ProfileName | HelpfulnessNumerator | HelpfulnessDenominator | Score | Time | Summary | Text | Helpful% |
|---|----|------------|----------------|---------------------------------|----------------------|------------------------|-------|------------|-----------------------|---|----------|
| 0 | 1 | B001E4KFG0 | A3SGXH7AUHU8GW | delmartan | 1 | 1 | 5 | 1303862400 | Good Quality Dog Food | I have bought several of the Vitality canned d... | 1.0 |
| 1 | 2 | B00813GRG4 | A1D87F6ZCVE5NK | dli pa | 0 | 0 | 1 | 1346976000 | Not as Advertised | Product arrived labeled as Jumbo Salted Peanut... | -1.0 |
| 2 | 3 | B000LQOCH0 | ABXLMWJXXAIN | Natalia Corres "Natalia Corres" | 1 | 1 | 4 | 1219017600 | "Delight" says it all | This is a confection that has been around a fe... | 1.0 |
| 3 | 4 | B000UA0QIQ | A395BORCF6GVXV | Karl | 3 | 3 | 2 | 1307923200 | Cough Medicine | If you are looking for the secret ingredient l... | 1.0 |
| 4 | 5 | B006KZZZ7K | A1UQRSLF8GWT | Michael D. Bigham "M. Wassir" | 0 | 0 | 5 | 1350776000 | Great taffy | Great taffy at a great price. There was a wid... | -1.0 |

```

In [6]: df['Helpful%'].unique()

Out[6]: array([ 1., -1., 0.8, 0., 0.5, 0.66666667, 0.25, 0.89473684, 0.83333333, 0.75, 0.33333333, 0.3, 0.11111111, 0.42857143, 0.875, 0.85714286, 0.2, 0.26315789, 0.6, 0.71428571, 0.53846154, 0.57142857, 0.91489362, 0.86666667, 0.82352941, 0.78571429, 0.74074074, 0.4, 0.375, 0.28571429, 0.14285714, 0.77777778, 0.125, 0.9, 0.94117647,

```

Figure 3.3: Data pre-processing

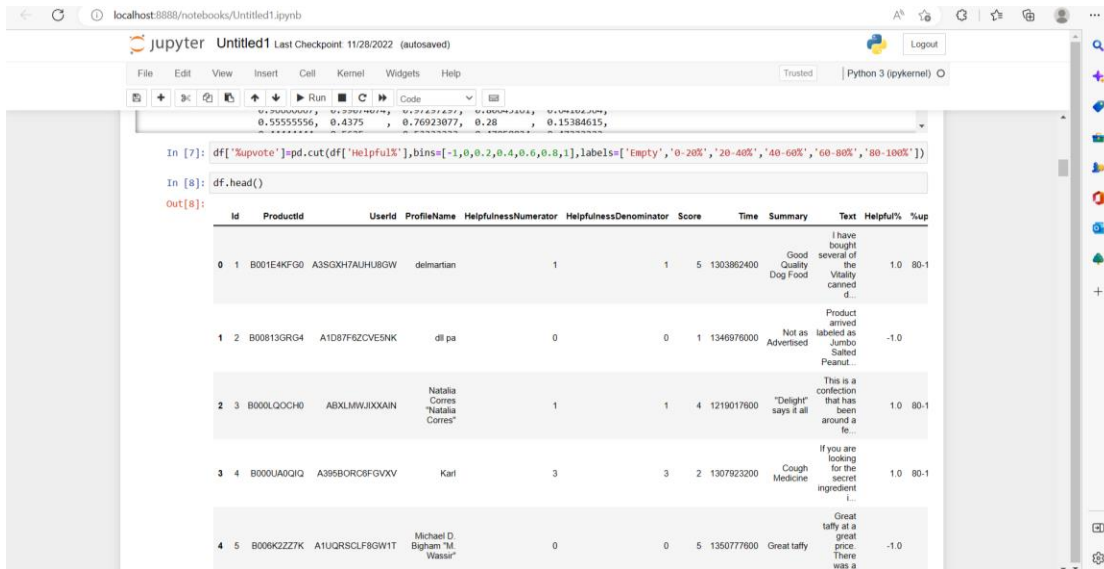


Figure 3.4: Adding & Defining column

After importing pandas and NumPy, I've added one column for upvoting the matrix. The column I've added is the division of the numerator and denominator. Then defined that column naming 'Helpful%'. This shows unique values ranging between -1 to +1.

NumPy

NumPy is the core Python library for scientific computing. This Python library contains a multidimensional array object, a number of derivative objects (such as masked arrays and matrices), and a number of methods for executing array operations quickly. Among these procedures are discrete Fourier transforms, elementary linear algebra, fundamental statistical operations, and random simulation, among others. It is the library that is used for all mathematical purposes. Numpy's goal is to make Python capable of quick and effective numerical computation.

NumPy supports object-oriented programming in its entirety, starting with ndarray. For instance, the ndarray class offers several methods and attributes. Since many of its methods are duplicated by functions in the NumPy namespace's outer namespace, programmers may use any coding style. Due to its flexibility, the NumPy array dialect and NumPy ndarray class are Python's de facto multidimensional data transport languages.

NumPy can be imported using a shortcut alias: `import numpy as np`

Matplotlib

The library used to create charts is called Matplotlib. This is a comprehensive Python package for creating static, animated, and interactive visualizations. The Matplotlib package allows for both basic and sophisticated operations. It creates interactive figures that can zoom, pan, and update, as well as publishing-quality plots. Python allows you to create 2D visualizations of arrays. Additionally, the built-in capabilities of Matplotlib allow us to create a variety of charts.

EDA

EDA is a data analysis phenomenon used to acquire a better knowledge of data elements such as:

- data's primary characteristics,
- variables and their interrelationships,
- determining which factors are crucial to solving our issue.

EDA is used to examine the data and synthesize the most important information.

It will provide you with a fundamental comprehension of your data, its distribution, null values, and much more.

You may examine data using either graphs or Python methods.

There will be two distinct forms of analysis. Univariate and Bivariate. In univariate analysis, a single characteristic is analyzed. In contrast, bivariate analysis involves studying one attribute alongside the target attribute.

In the non-graphical method, you will use functions like `shape`, `summary`, `describe`, `isnull`, `info`, and `datatypes`.

In the graphical method, scatter, box, bar, density, and correlation plots will be used.

Utilizing summary statistics and graphical representations, exploratory data analysis is a vital stage in doing a preliminary study on data to identify anomalies, establish patterns, test hypotheses, and confirm assumptions. It is better to first comprehend the facts and then try to draw as many conclusions as you can from them. EDA concentrates on making sense

of the information currently accessible before getting filthy with the data. We may quickly use the Exploratory Data Analysis (EDA) method to condense the main characteristics of our data.[14] Additionally, it is an approach that uses graphs and charts to extract solutions from our unprocessed data. In order to glean insights from data, this technique employs statistics and data visualization designs. The main benefit of the EDA methodology is that it enables us to conduct further analysis on our raw data than just hypothesis testing.[15] More specifically, we can examine our data for in-depth solutions by using statistical models like averages, standard deviation, median, and quartiles, among others.

After preprocessing the data, I used exploratory data analysis to analyze certain data. It is an approach to data analysis that employs visual approaches. Utilizing statistical summaries and graphical representations, it is used to uncover trends and patterns, as well as to confirm assumptions.

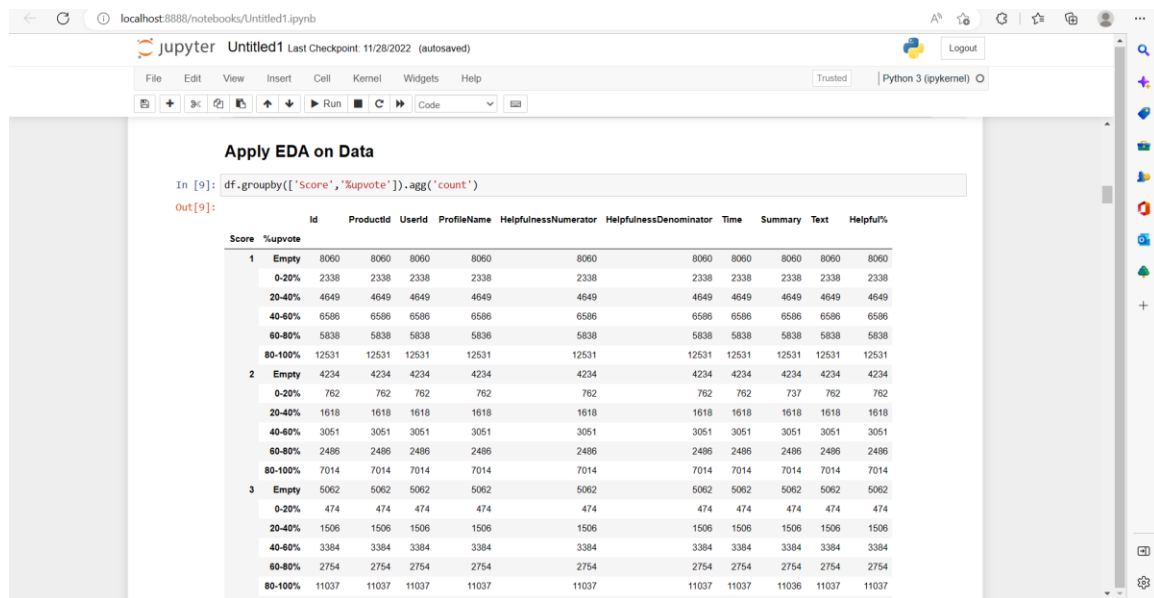


Figure 3.5: Exploratory Data Analysis

During EDA, I have analyzed upvotes for different scores. For analyzing upvotes, I have used a specific function named 'groupby'. Also, I've created pivot table and heatmap of the problem statement or dataset for better conclusion.

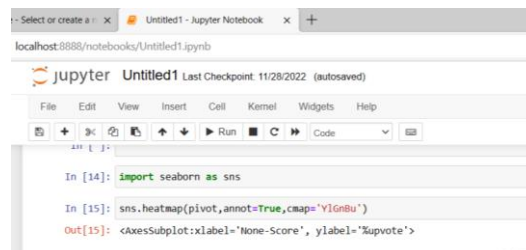
Pivot table

A pivot table is a powerful tool for calculating, compressing, and analyzing data that allows you to identify similarities, patterns, and trends in your data. It is a fun method to quickly summarize a lot of information. A Pivot Table enables us to assess numerical data comprehensively and deliver unexpected data-related responses. A pivot table is particularly suited for the job of querying vast quantities of data in many user-friendly methods. With the aid of a pivot table, we can draw many different conclusions. We can come with various inferences with the help of pivot table.

Seaborn

Seaborn is utilized to observe random distributions. A few graphs and patterns for data visualization can be found in Seaborn. It is based on Matplotlib and strongly integrated with Pandas data structures. It supports our understanding and data exploration.

Before creating a heatmap, first I imported a data visualization module named 'seaborn'. The alias of that module is 'sns'. With the help of that alias, the heatmap function has been called. After that, to make it more user-friendly, we can assign `annot = True`.



```

In [14]: import seaborn as sns
In [15]: sns.heatmap(pivot, annot=True, cmap='YlGnBu')
Out[15]: <AxesSubplot:xlabel='None-Score', ylabel='Kupvote'>
```

Figure 3.6: Importing Seaborn

Heatmap

A heat map is used to illustrate the degree of association between many variables by displaying the coefficients. It assists in determining characteristics that help develop machine learning models. The use of the heat map allows for the correlation matrix to be translated into color labeling. Heat maps assist viewers in concentrating their attention on

the aspects of data visualizations that are most significant by assisting in the better depiction of the number of places and events included within a dataset. Despite this, they are so much more than that.. Before creating a heatmap, I had imported a data visualization module named ‘seaborn’ translated into color labeling. Heat maps assist viewers in concentrating their attention on the aspects of data visualizations that are most significant by assisting in the better depiction of the number of places and events included within a dataset. Despite this, they are so much more than that.. Before creating a heatmap, I had imported a data visualization module named ‘seaborn’. I had assigned a color map for changing color.

I’ve created a heatmap of the problem statement for a better conclusion.

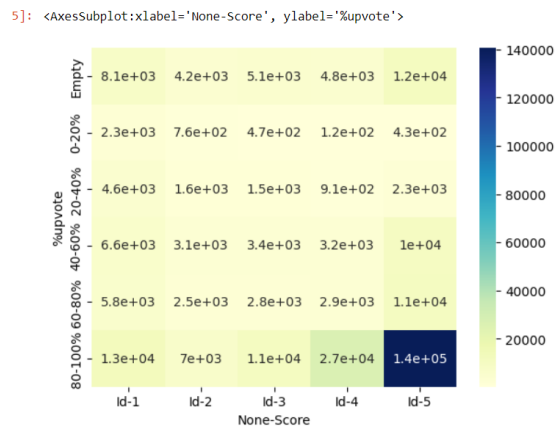


Fig 3.7: Heatmap

If we inference from the heatmap, we can see reviews are basically skewed towards positive. Data in darker density which inferences that, many people agree with the score 5 reviews.

Apply Bag of Words on data

It is the fundamental model employed in natural language processing (nlp). To explain it formally, it is a technique for extracting features from text data. A word's vector representation needs to be made. In order to construct a vector representation of a word, BoW marks the term's appearance in each review from the dictionary. It is utilized in natural language processing and information retrieval.

By applying Bag of Words on data, what I've done is:

- Get all the unique scores.
- Eliminate all the neutral score which is 3.
- Extract all the dependent data (y) and independent data (x).
- Convert text into vector using NLP.

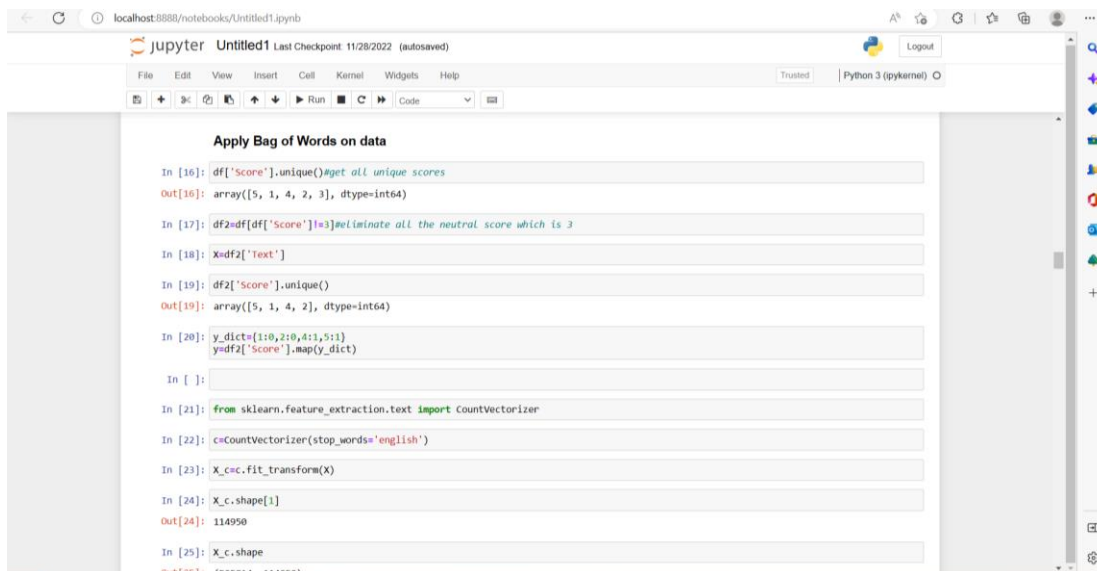
Whenever the score is less than 3, then I've mapped it with 0 and whenever the score is greater than 3, then I've mapped it with 1. Because, I've to convert the problem statement into some classification problem.

For converting the problem statement into classification problem, I've declared the dictionary as 'y_dict' and applied that on Score. Then mapped the dictionary. Finally, stored the data on 'y' variable.

Because any information about the arrangement or structure of the words inside the text is discarded, the text is called a "bag" of words. This paradigm ignores syntax and even word order while maintaining the plurality of words inside a text. The model simply cares about recognized terms being in the document, not where they appear.

For performing bag of words (bow), I've imported 'CountVectorizer' class. Then created an object of the CountVectorizer and set stop_words as 'english' which has returned an object and stored in c.

By performing bag of words, all the unique scores of the dataset has been obtained.



```
Apply Bag of Words on data

In [16]: df['Score'].unique()#get all unique scores
Out[16]: array([5, 1, 4, 2, 3], dtype=int64)

In [17]: df=df[df['Score']!=3]#eliminate all the neutral score which is 3

In [18]: x=df2['text']

In [19]: df2['Score'].unique()
Out[19]: array([5, 1, 4, 2], dtype=int64)

In [20]: y_dict={1:0,2:0,4:1,5:1}
y=df2['Score'].map(y_dict)

In [ ]:

In [21]: from sklearn.feature_extraction.text import CountVectorizer
In [22]: c=CountVectorizer(stop_words='english')
In [23]: X_c=c.fit_transform(x)
In [24]: X_c.shape[1]
Out[24]: 114950
In [25]: X_c.shape
Out[25]: (733814, 114950)
```

Figure 3.8: Bag of Words

Convert Text into Vector using NLP

Vectorization is the name of this procedure. It is a natural language processing (NLP) technique that converts words or phrases from a lexicon to a matching vector of real numbers, which is then used to calculate word predictions and word similarity/semantics.

There are several ways to transform string data into numerical data:

- Bag of Words
- TF-IDF
- Word2Vec
- Text Preprocessing

Term Frequency

Term frequency refers to the number of times a word appears in a text or corpus. It is feasible that a word will appear in a sentence more frequently in a lengthy one than in a shorter one since no two phrases have precisely the same length.

Term frequency can be defined as-

$$TF = \frac{\text{No of time word appear in the document}}{\text{Total no of word in the document}}$$

Figure 3.9: TF law

Inverse Document Frequency

Another approach for determining the significance of a word is inverse document frequency. It is based on the idea that words that are less common but more informational and significant. IDF is represented by the formula-

$$IDF = \log_{10} \frac{\text{Number of Document}}{\text{Number of document in which word appear}}$$

Figure 3.10: IDF law

TF-IDF

Term Frequency - Inverse Document Frequency (TF-IDF) is a measurement of a term's importance within a corpus or dataset. It simply involves multiplying two tables. It lowers the values of frequent words that are utilized in various documents.



```
Apply TF-IDF on data
In [69]: from sklearn.feature_extraction.text import TfidfVectorizer
In [70]: tf=TfidfVectorizer()
In [71]: X_catf.fit_transform(x)
```

Figure 3.11: TF-IDF

Natural Language Processing

Whether written or spoken, it processes and applies to all human languages. The primary idea behind NLP is this. NLP refers to text fragments, which might comprise nouns, verbs, and adjectives. Sentences can be broken down into smaller words using different parts of speech, and each every word has its own unique meaning. Nouns, pronouns, verbs, adverbs, adjectives, conjunctions, prepositions, and interjections are examples of minor parts of speech in the English language that may be employed in several circumstances and achieve a range of functions.

N Gram

To put it another way, an n-gram is a collection of tokens. Tokens may represent individual characters or subsets of characters, but in the field of computational linguistics, they are most often words. The n only stands for the total number of tokens. An example of an n-gram may be the word sequence in and of itself, or it could be a prediction model that assigns it a probability. The term "gram" refers to a grouping of letters; "n gram" means to break a statement up into words; for instance, "I eat meat only" has 4 gram, whereas "rose is red" contains 3 gram. Named-entity recognition (NER) is a system used to break up tweets or comments into smaller chunks, with each chunk having two words.

NLP Approach

The NLP approach follow the following discrete steps:

- Text Preprocessing/Tokenization
- Lexical Analysis
- Syntactical Analysis
- Semantic Analysis

Measure accuracy of Sentiment Analysis

The field of sentiment analysis seems to have made tremendous progress for the last two years, while it is still very far from being perfect. It is essential to have a method for evaluating both the performance and the accuracy of the sentiment API providers that you employ, whether those providers be IBM, Google, or Azure. Measuring accuracy was extensively reviewed by Alexander Pak & Patrick Paroubek.[1] The data would be categorized as either positive, neutral, or negative according to the majority of sentiment analysis algorithms. As a consequence of this, a good rule of thumb is to check and see whether the system's classification of the data is in agreement with the user's gut instinct. It is impossible to arrive at a definitive answer to this question using standard mathematical methods since it is so subjective and complex. According to study, human analysts tend to agree around 80–85% of the time when assessing the emotion (positive, negative, or neutral) of a particular text content.

There are three extremely essential statistics that go into establishing how effectively a sentiment analysis system functions.

- Accuracy
- Recall
- F1 Score

Accuracy

The frequency with which a sentiment evaluation is accurate. [Number of Correct Queries/Total Queries] - This would be used to determine the overall accuracy of the system.

Recall

A measurement of the number of words containing emotion that were considered to have sentimental value. This might be seen as a measure of how correctly the system decides whether or not something is neutral.

F1 Score

The F1-Score combines accuracy and recall. Because it reveals how efficiently your system is running, this metric is among the most important ones that can be taken. The following equation may be used to get an individual's F1 Score: $F1 = [2 * (Precision * Recall) / (Precision + Recall)]$

The score falls between 0.0 and 1.0, with 1.0 representing a flawless performance. The F1 Score provides us with a single statistic that ranks a system based on both its accuracy and its recall, which is a really valuable feature.

Prepare data for modeling purpose

For data preparation, we have to follow following steps:

- Data Collection
- Data Processing
- Data analysis
- Data visualization
- Register & Create Project
- Link/Upload & Process Data
- Visualize data
- Train the model without coding

Apply techniques of handling imbalanced data

The term "imbalanced data" refers to a classification data collection that has proportions of classes that are not balanced with one another. The classes in a data set that account for the vast majority of the total entries are referred to as majority classes. Classes with a lesser share of the population are minority classes.

The table below shows the range of imbalanced data:

| Degree of imbalance | Proportion of Minority Class |
|---------------------|------------------------------|
| Fair | 20-40% of the data set |
| Medium | 1-20% of the data set |
| Intense | <1% of the data set |

Table 3.1: Range of imbalanced data[3]

More ways to handle imbalanced data

- Use the accurate assessment metrics
- Take another sample from the training set
- Use K-fold Cross-Validation appropriately
- Concatenate Diverse Resampled Datasets
- Resample with various ratios
- Group the several class
- Design Models

Resample the training set

In addition to applying unique assessment criteria, one might also endeavor to get distinct datasets. Under-sampling and oversampling are two techniques for creating a balanced dataset from an unbalanced one.

Under-sampling & Over-sampling

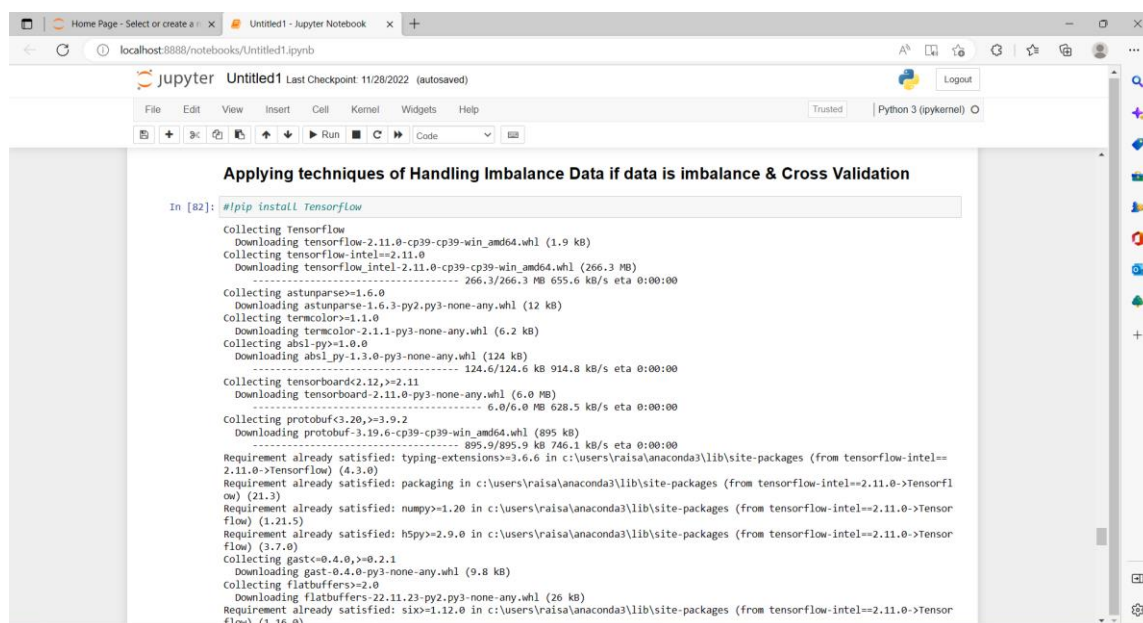
By decreasing the size of the most prevalent class, under-sampling helps to create a more even distribution throughout the dataset. When there are sufficient amounts of data, this method is used. To recover a fresh balanced dataset for future modeling, It is essential to maintain all of the samples from the uncommon class and randomly choose an equal number of samples from the plentiful class.

On the other contrary, oversampling is used in situations in which there is an inadequate amount of data. It does this by boosting the size of uncommon samples in an effort to create a more balanced dataset.

There is no technique of resampling that has an unquestionably significant advantage over the others. Both the specific use case and the dataset itself will determine which of these two strategies will be applicable in a given situation. In many cases, the best results may be achieved by using a strategy that combines both over- and under-sampling.

Cross-Validation

Cross-validation is a machine learning approach that divides the training data into a training set and a test set. The training set is used to construct the model, whereas the test set is used to assess the model's performance in production. It helps to validate our model's efficacy and precision using unseen data. It is a method for testing Machine Learning models using subsets of the original data set.



```
In [82]: #!pip install TensorFlow

Collecting TensorFlow
  Downloading tensorflow-2.11.0-cp39-cp39-win_amd64.whl (1.9 kB)
Collecting tensorflow-intel==2.11.0
  Downloading tensorflow_intel-2.11.0-cp39-cp39-win_amd64.whl (266.3 MB)
-----
266.3/266.3 MB 655.6 kB/s eta 0:00:00
Collecting astunparse==1.6.0
  Downloading astunparse-1.6.3-py2.py3-none-any.whl (12 kB)
Collecting termcolor==1.1.0
  Downloading termcolor-2.1.1-py3-none-any.whl (6.2 kB)
Collecting absl-py==1.0.0
  Downloading absl_py-1.3.0-py3-none-any.whl (124 kB)
-----
124.6/124.6 kB 914.8 kB/s eta 0:00:00
Collecting tensorboard<2.12, >=2.11
  Downloading tensorboard-2.11.0-py3-none-any.whl (6.0 MB)
-----
6.0/6.0 MB 628.5 kB/s eta 0:00:00
Collecting protobuf<3.20, >=3.9.2
  Downloading protobuf-3.19.6-cp39-cp39-win_amd64.whl (895 kB)
-----
895.9/895.9 kB 746.1 kB/s eta 0:00:00
Requirement already satisfied: typing-extensions>=3.6.6 in c:\users\raisa\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->TensorFlow) (4.3.0)
Requirement already satisfied: packaging in c:\users\raisa\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->TensorFlow) (21.3)
Requirement already satisfied: numpy>=1.20 in c:\users\raisa\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->TensorFlow) (1.21.5)
Requirement already satisfied: h5py>=2.9.0 in c:\users\raisa\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->TensorFlow) (2.7.0)
Collecting gast<=0.4.0, >=0.2.1
  Downloading gast-0.4.0-py3-none-any.whl (9.8 kB)
Collecting flatbuffers>=2.0
  Downloading flatbuffers-22.11.23-py2.py3-none-any.whl (26 kB)
Requirement already satisfied: six>=1.12.0 in c:\users\raisa\anaconda3\lib\site-packages (from tensorflow-intel==2.11.0->TensorFlow) (1.16.0)
```

Figure 3.12: Apply techniques of handling imbalance data & cross validation


```
ow-estimator, tensorboard-data-server, rsa, protobuf, opt-einsum, oauthlib, keras, grpcio, google-pasta, gast, cachetools, astu
nparse, absl-py, requests-oauthlib, google-auth, google-auth-oauthlib, tensorboard, tensorflow-intel, tensorflow
Successfully installed tensorflow-2.11.0 absl-py-1.3.0 astunparse-1.6.3 cachetools-5.2.0 flatbuffers-22.11.23 gast-0.4.0 google
-auth-2.14.1 google-auth-oauthlib-0.4.6 google-pasta-0.2.0 grpcio-1.50.0 keras-2.11.0 libclang-14.0.6 oauthlib-3.2.2 opt-einsum
-3.3.0 protobuf-3.19.6 requests-oauthlib-1.3.1 rsa-4.9 tensorboard-2.11.0 tensorboard-data-server-0.6.1 tensorboard-plugin-wit-
1.8.1 tensorflow-estimator-2.11.0 tensorflow-intel-2.11.0 tensorflow-io-gcs-filesystem-0.28.0 termcolor-2.1.1

In [87]: !pip3 install -U imbalanced-learn

Collecting imbalanced-learn
  Downloading imbalanced_learn-0.9.1-py3-none-any.whl (199 kB)
  -----
  199.3/199.3 kB 1.5 MB/s eta 0:00:00
Requirement already satisfied: joblib>=1.0.0 in c:\users\raisa\anaconda3\lib\site-packages (from imbalanced-learn) (1.1.0)
Requirement already satisfied: numpy>=1.17.3 in c:\users\raisa\anaconda3\lib\site-packages (from imbalanced-learn) (1.21.5)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\raisa\anaconda3\lib\site-packages (from imbalanced-learn) (2.2.
0)
Collecting scikit-learn>=1.1.0
  Downloading scikit_learn-1.1.3-cp39-cp39-win_amd64.whl (7.6 MB)
  -----
  7.6/7.6 MB 738.7 kB/s eta 0:00:00
Requirement already satisfied: scipy>=1.3.2 in c:\users\raisa\anaconda3\lib\site-packages (from imbalanced-learn) (1.9.1)
Installing collected packages: scikit-learn, imbalanced-learn
  Attempting uninstall: scikit-learn
    Found existing installation: scikit-learn 1.0.2
    Uninstalling scikit-learn-1.0.2:
      Successfully uninstalled scikit-learn-1.0.2
ERROR: Could not install packages due to an OSError: [WinError 5] Access is denied: 'C:\Users\Raissa\anaconda3\lib\site-pac
kages\*-klearn\feature_extraction\_hashing_fast.cp39-win_amd64.pyd'
Consider using the "--user" option or check the permissions.

In [72]: from imblearn.over_sampling import RandomOverSampler

In [73]: os=RandomOverSampler()

In [74]: X_train_res,y_train_res= os.fit_sample(X_c, y)
```

Figure 3.13: Importing RandomOverSampler

CHAPTER 4

Experimental Results

4.0: Fetch positive & negative words

After checking the model accuracy, I've fetched top 20 positive words and top 20 negative of the NLP and ML model.

```
it[41]:
```

| | Word | Coefficient |
|--------|------------|-------------|
| 80590 | pleasantly | 3.978480 |
| 39067 | downside | 3.033060 |
| 5865 | addicting | 2.828222 |
| 94653 | skeptical | 2.808813 |
| 55024 | hooked | 2.459712 |
| 39209 | drawback | 2.433279 |
| 96339 | soothing | 2.368762 |
| 96126 | solved | 2.303026 |
| 111503 | welcome | 2.301913 |
| 87956 | resist | 2.300825 |
| 68452 | met | 2.283705 |
| 35721 | delish | 2.202883 |
| 43217 | excellent | 2.153508 |
| 113120 | worries | 2.141052 |
| 35686 | delighted | 2.124230 |
| 54125 | hesitant | 2.118523 |
| 24743 | calm | 2.110912 |
| 91051 | saves | 2.078592 |
| 87163 | relaxing | 2.072605 |
| 10993 | awesome | 2.070006 |

Figure 4.1: Top 20 positive words

```
In [42]: coef_df.tail(20)
```

```
Out[42]:
```

| | Word | Coefficient |
|---------------|----------------|--------------------|
| 108627 | useless | -2.284180 |
| 106878 | unappealing | -2.330753 |
| 46078 | flavorless | -2.380995 |
| 102830 | terrible | -2.416637 |
| 21286 | blech | -2.451505 |
| 41113 | embarrassed | -2.521665 |
| 37558 | disappointment | -2.523858 |
| 24946 | cancelled | -2.547662 |
| 88340 | returnable | -2.593668 |
| 67890 | mediocre | -2.604430 |
| 89851 | ruined | -2.624542 |
| 65057 | lousy | -2.664902 |
| 88934 | ripoff | -2.724250 |
| 37523 | disappointed | -2.830562 |
| 62394 | lacked | -2.913340 |
| 37555 | disappointing | -3.188990 |
| 113146 | worst | -3.211218 |
| 34984 | deceptive | -3.402158 |
| 106836 | unacceptable | -3.468227 |
| 107366 | undrinkable | -3.694207 |

Figure 4.2: Top 20 negative words

4.1: Data Preparation for predicting the upvote(s)

localhost:8888/notebooks/Untitled1.ipynb

Jupyter Untitled1 Last Checkpoint: 11/28/2022 (autosaved) Python 3 (ipykernel)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

Data preparation for predicting the upvotes

```
In [61]: data=df[df['Score']==5]
data.head()
```

Out[61]:

| | Id | ProductId | UserId | ProfileName | HelpfulnessNumerator | HelpfulnessDenominator | Score | Time | Summary | Text | Helpful% | %upv |
|---|----|------------|----------------|-------------------------------|----------------------|------------------------|-------|------------|--|---|----------|-------|
| 0 | 1 | B001E4KFG0 | A3SGXH7AUHU8GW | delmartian | 1 | 1 | 5 | 1303862400 | Good Quality Dog Food | I have bought several of the Vitality canned d... | 1.0 | 80-10 |
| 4 | 5 | B006K2ZZ7K | A1UQRSCLF8GW1T | Michael D. Bigham "M. Wassir" | 0 | 0 | 5 | 1350777600 | Great taffy | Great taffy at a great price. There was a wid... | -1.0 | N |
| 6 | 7 | B006K2ZZ7K | A1SP2KVKFXRU1 | David C. Sullivan | 0 | 0 | 5 | 1340150400 | Great! Just as good as the expensive brands! | This saltwater taffy had great flavors and was... | -1.0 | N |
| 7 | 8 | B006K2ZZ7K | A3JRGQVEQN31IQ | Pamela G. Williams | 0 | 0 | 5 | 1336003200 | Wonderful, tasty taffy | This taffy is so good. It is very soft and ch... | -1.0 | N |
| | | | | | | | | | | Right now I'm mostly | | |

Figure 4.3: Upvotes prediction

pdf x Home Page - Select or create a r x Untitled1 - Jupyter Notebook x +

ocalhost:8888/notebooks/Untitled1.ipynb

Jupyter Untitled1 Last Checkpoint: 11/28/2022 (autosaved) Python 3 (ipykernel)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Code

Out[62]: ['80-100%', NaN, '60-80%', 'Empty', '40-60%', '20-40%', '0-20%']
 Categories (6, object): ['Empty' < '0-20%' < '20-40%' < '40-60%' < '60-80%' < '80-100%']

In [63]: *#get rid of the neutral votes*
 data2=data[data["%upvote"].isin(['80-100%', '60-80%', '20-40%', '0-20%'])]

In [64]: data2.head()

Out[64]:

| ProductId | UserId | ProfileName | HelpfulnessNumerator | HelpfulnessDenominator | Score | Time | Summary | Text | Helpful% | %upvote |
|------------|----------------|------------------------|----------------------|------------------------|-------|------------|---|---|----------|---------|
| 3001E4KFG0 | A3SGXH7AUHU8GW | delmartian | 1 | 1 | 5 | 1303862400 | Good Quality Dog Food | I have bought several of the Vitality canned d... | 1.0 | 80-100% |
| B000E7L2R4 | A1MZY09TZK0BBI | R. James | 1 | 1 | 5 | 1322006400 | Yay Barley | Right now I'm mostly just sprouting this so my... | 1.0 | 80-100% |
| B0001PB9FE | A3HDKO7OWQNK4 | Canadian Fan | 1 | 1 | 5 | 1107820800 | The Best Hot Sauce in the World | I don't know if it's the cactus or the tequila... | 1.0 | 80-100% |
| B0009XLVG0 | A2725IB4YY9JEB | A Poeng "SparkyGoHome" | 4 | 4 | 5 | 1282867200 | My cats LOVE this "diet" food better than thei... | One of my boys needed to lose some weight and ... | 1.0 | 80-100% |
| B001GVISJM | A2MUGFV2TDQ47K | Lynrie "Oh HELL no" | 4 | 5 | 5 | 1268352000 | Strawberry Twizzlers - Yummy | The Strawberry Twizzlers are my guilty pleasur... | 0.8 | 60-80% |

to search 64°F Haze

Figure 4.4: Predicting the upvotes after removing the neutral vote

CHAPTER 5

Conclusions and Future Work

5.0 Conclusion and Discussion

Sentiment analysis plays a significant role in natural language processing (NLP). It is the intersection of human emotional awareness with machine learning technology. In this research, In analyzing sentiment, Several Machine Learning classification methods and Natural Language Processing approaches have been used to a huge, unbalanced, multi-classified, real-world dataset. Also, methods of sentiment analysis based on certain customers review and related methodologies have been proposed. Future study will include continually increasing the comparison area with new findings.

5.1 Future work and Further Development

I wish to enhance this research-based project's usefulness. In the future, I'll look at how to improve the performance of prediction. I am also planning to update this study with more analyses and real-world datasets.

REFERENCES

- [1]. Alexander Pak, Patrick Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", January 2010.
- [2]. Mayur Wankhade, Annavarapu Chandra Sekhara Rao, Chaitanya Kulkarni, "A survey on sentiment analysis methods, applications, and challenges", IEEE explore, India 07 February 2022.
- [3]. KD nuggets available at <kdnuggets.com> last accessed on 11 January 17:30 pm.
- [4]. Adil Rajput, "Natural Language Processing, Sentiment Analysis and Clinical Analytics", February 2019.
- [5]. Aryaman Sharma, "Natural Language Processing And Sentiment Analysis", ResearchGate J., vol. 08, November 2021.
- [6]. Md. Taufiqul Haque Khan Tusar, Md. Touhidul Islam, "A Comparative Study of Sentiment Analysis Using NLP and Different Machine Learning Techniques on US Airline Twitter Data", ResearchGate, Bangladesh 02 October 2021.
- [7]. Wei Yen Chong, Bhawani Selvartarenam, Lay-Ki Soon, "Natural Language Processing for Sentiment Analysis", Conference, 10.1109/ICAIET, 2014.
- [8]. Neha Gaur, Dr. Neetu Sharma, "Sentiment Analysis in Natural Language Processing", J. vol. 3, May-June 2017.
- [9]. Fazeel Abid, Muhammad Alam, Muhammad Yasir, Chen li, "Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter", China 12 December 2018.
- [10]. Norbert Blenn, Christian Doerr, "Context-Sensitive Sentiment Classification of Short Colloquial Text", ResearchGate Conference, 30 May 2014.
- [11]. Tetsuya Nasukawa, Jeonghee Yi, "Sentiment analysis: Capturing favorability using natural language procesisng", ResearchGate Conference, 10 March 2015.
- [12]. Fuad Alattar, Khaled Shaaln, "Using Artificial Intelligence to Understand What causes Sentiment Changes on Social Media", IEEE, 29 April 2021.
- [13]. Vishal. A. Kharde, Prof. Sheetal. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques", 22 April 2016.
- [14]. Ravi Parikh, Matin Movassate, "Sentiment Analysis of User-Generated Twitter Updates using Various Classification Techniques", 4 June 2009.
- [15]. Bo Pang, Lillian Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts", 29 September 2004.

SENTIMENT ANALYSIS ON AMAZON CUSTOMER'S REVIEW USING NLP

ORIGINALITY REPORT

23%

SIMILARITY INDEX

18%

INTERNET SOURCES

6%

PUBLICATIONS

16%

STUDENT PAPERS

PRIMARY SOURCES

| | | |
|---|--|-----|
| 1 | dspace.daffodilvarsity.edu.bd:8080 Internet Source | 10% |
| 2 | Submitted to Daffodil International University Student Paper | 2% |
| 3 | Submitted to Liverpool John Moores University Student Paper | 1% |
| 4 | web.archive.org Internet Source | 1% |
| 5 | Submitted to Coventry University Student Paper | 1% |
| 6 | Submitted to Aligarh Muslim University, Aligarh Student Paper | 1% |
| 7 | Submitted to Melbourne Institute of Technology Student Paper | 1% |
| 8 | core.ac.uk Internet Source | <1% |