****

**Bharati Vidyapeeth’s College of Engineering for Women, Pune – 43**

**Department of Computer Engineering**

**Project Based Learning Report**

on

**“Sentiment Analysis of Social Media Reviews Using NLP Approach”**

**SUBMITTED BY**

Ms. Mitali Rajesh Chavan (2239)

Ms. Sai Tanaji Nagane (2241)

Ms. Chahal Ohri (2244)

Ms. Khushi Padhar (2245)

Ms. Jagruti Patil (2248)

**Guide: Prof. Dr. Shital Pawar**

****

**Bharati Vidyapeeth’s College of Engineering for Women, Pune – 43**

**Department of Computer Engineering**

***Certificate of Originality***

This is to certify, that the project Sentiment Analysis of Social Media Reviews Using NLP Approach submitted by us is an outcome of our own work. We have duly acknowledged all the sources from which the ideas and extracts have been taken. This project has not been copied from anywhere and all the data has been collected by us.

**SUBMITTED BY**

Ms. Mitali Rajesh Chavan (2239) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ms. Sai Tanaji Nagane (2241) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ms. Chahal Ohri (2244) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ms. Khushi Padhar (2245) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ms. Jagruti Patil (2248) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature of Guide

Prof. Dr. Shital Pawar

****

**Bharati Vidyapeeth’s College of Engineering for Women, Pune – 43**

**Department of Computer Engineering**

**CERTIFICATE**

This is to certify that,

Ms. Mitali Rajesh Chavan (2239)

Ms. Sai Tanaji Nagane (2241)

Ms. Chahal Ohri (2244)

Ms. Khushi Padhar (2245)

Ms. Jagruti Patil (2248)

have successfully completed the PBL titled

**“Sentiment Analysis of Social Media Reviews Using NLP Approach”**

during the academic year 2022-23 towards the fulfilment of the

**Second Year in the Department of Computer Engineering under Savitribai Phule Pune University.**

**Prof. Dr. Shital Pawar Prof. D. D. Pukale**

**(Guide) (HOD)**

****

**Bharati Vidyapeeth’s College of Engineering for Women, Pune – 43**

**Department of Computer Engineering**

**ACKNOWLEDGEMENT**

Primarily, we would like to thank God for enabling us to complete this project successfully. We would also like to thank our teachers Prof. Dr. Shital Pawar and Prof. Shital Jadhav whose valuable guidance and continuous encouragement and support has helped us complete this project. Their suggestions, instructions and teachings have served as major contributors to this project. We would also like to thank our mentor Prof. Dr. Shital Pawar for her constant and encouraging support. We would like to extend our heartfelt gratitude towards our Head of Department Prof. D. D. Pukale for guiding and encouraging us throughout the project. We also want to thank our Principal Dr. S. R. Patil for providing us a congenial environment to work in. We would like to thank our parents and friends who have helped us with their valuable suggestions. Their guidance has been helpful in various faces of the completion of the project. Last but not the least, we would like to thank our classmates who have helped us a lot.

Ms. Mitali Rajesh Chavan (2239)

Ms. Sai Tanaji Nagane (2241)

Ms. Chahal Ohri (2244)

Ms. Khushi Padhar (2245)

Ms. Jagruti Patil (2248)

**Abstract**

We always search for product reviews on various social media platforms before making a purchase. We carefully evaluate the reviews before deciding whether or not to actually make the purchase. That is a typical human tendency. Sentiment analysis can help businesses understand customer demands and monitor brand and product perception via customer feedback. Sentiment analysis is a natural language processing (NLP) tool that identifies the emotional states in human speech or text, in this case, the text that consumers produce as feedback and reviews. Businesses can analyze this text for positive, negative, or neutral sentiment regarding the brand using sentiment analysis software based on machine learning. Companies have the chance to respond effectively -- and with more empathy -- with the help of this information. The goal is to strengthen relationships with customers and increase their loyalty. Our model helps segregate data based on these emotions thus helping the organizations understand their strengths and weaknesses. Furthermore, visualization of data enhances the understanding of the computed result. We have also displayed the most frequented words along with their frequency count percentage which further helps in easy and better understanding of the data.

**INDEX**

1. Introduction ….....……...……………………………………………………...…09
2. Literature Survey ………………………………………………………………...10
3. Problem Statement ………………………………………………………………19
4. Motivation .……………………………………………………………………....20

4.1 Sentiment Analysis…………………………………….……………..21

1. Objectives………………………………………………………………………...22
2. Scope……………………………………………………………………………..23
3. Methodology/ Proposed System Block Diagram………………………………...24

7.1 SIPOC Diagram ………………………………………………….…..24

7.2 Modeling ……………………………………………………………..26

7.3 Block Diagram ……………………………………………………….28

7.4 System Architecture ………………………………………………….29

1. Software and Hardware Requirements …………………………………………..30
2. Implementation …………………………………………………………………..31

9.1 Coding ……………………………………………………………….31

9.2 Visualization of Data ………………………………………………...32

1. Results and Discussion …………………………………………………………..36
2. Conclusion and Future Scope ……………………………………………………40

11.1 Conclusion ………………………………………………………….40

11.2 Future Scope ………………………………………………………..40

1. References ………………………………………………………………….…….41

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE**  **NO.** | **FIGURE NAME** | **PAGE**  **NO.** |
| 2.1 | Steps to follow while carrying out a Literature Survey | 10 |
| 4.1 | Data Science | 20 |
| 4.2 | Python | 20 |
| 4.3 | NLP | 20 |
| 4.1.1 | University/College | 21 |
| 4.1.2 | People’s perception of our college | 21 |
| 7.1.1 | SIPOC Diagram for our project | 25 |
| 7.3.1 | Block Diagram | 28 |
| 7.4.1 | System Architecture Diagram | 29 |
| 8.1 | Software and Hardware Requirement for the Project | 30 |
| 9.2.1 | Positive Word Cloud | 32 |
| 9.2.2 | Negative Word Cloud | 32 |
| 9.2.3 | Neutral Word Cloud | 33 |
| 9.2.4 | Visualization of Data using Microsoft Power BI | 33 |
| 9.2.5 | Positive Word Cloud | 34 |
| 9.2.6 | Negative Word Cloud | 34 |
| 9.2.7 | Neutral Word Cloud | 35 |
| 9.2.8 | Visualization of Data using Microsoft Power BI | 35 |
| 10.1 | Positive Word Cloud | 36 |
| 10.2 | Negative Word Cloud | 37 |
| 10.3 | Visualization of Rotten Tomatoes Review | 38 |
| 10.4 | Visualization of College Reviews | 38 |
| 10.5 | Frequency of Negative Word | 39 |
| 10.6 | Frequency of Positive Word | 39 |

**1. INTRODUCTION**

We always search for product reviews on various social media platforms before making a purchase. We carefully evaluate the reviews before deciding whether or not to actually

make the purchase. That is a typical human tendency.

Similar to this, kids and their parents research colleges and universities in depth when looking for colleges or universities. They get in touch with relatives, friends, and family to ask about their opinions of the college. Most importantly, they scan multiple social media platforms for reviews and remarks from other users. They consider the various viewpoints and ideas of the people that are present and work to comprehend the advantages and disadvantages of each university.

As a result, an institution's status or reputation for quality is often more significant because it influences prospective students' selections when applying there. Students are keen on receiving an outstanding educational experience; thus, it is crucial to make sure they and their parents have all the details about the college's reputation.

Strongly regarded colleges might have the potential to offer superior learning environments and have large alumni networks. Therefore, it aids in the overall development of the students while also providing the institution with a competitive advantage. However, we can only do this if we are aware of how outsiders or other individuals view our institute. We will have a better understanding of this through our project, which will benefit our institute's advancement.

**2. Literature Survey**

Any project's literature survey is crucial, and its relevance can be expanded upon by turning it into a critical part of the project. It benefits the validity of the project in a number of ways:

1. Establishes the consistency in knowledge and the applicability of existing materials by providing the interpretation of existing literature in light of current advancements in the field.
2. The research gaps that were initially examined are subsequently investigated to determine the most recent facts and theories to help in the advancement of the subject.
3. It shows where the current research stands in relation to the discipline's overall

framework.

We carried out a literature survey for our project as well. The overview of the literature

survey has been provided below in a tabulated format.

Fig 2.1 Steps to follow while carrying out a Literature Survey

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sr. No.** | **Author’s Name** | **Paper Name** | **Findings of Paper** | **Techniques or Methods used** | **Scope** |
| 1. | Reem Baragash, Hanan Aldowah | Sentiment Analysis based Machine Learning Approach: International Students’ Perception Towards  E-learning in Malaysian Universities | The adoption of the emerging field of sentiment analysis and opinion mining in educational systems has a great potential in improving the quality of teaching and learning processes in Higher Education institutions and assessing the situation in terms of performance, online materials, and services. | Support Vector Machine (SVM) , Naïve Baiyes (NB) | It can guide new and international students to select the appropriate university based on the review submitted by former students on social media sites. Further studies are required on the applications of sentiment analysis in higher education institutes to benefit from the massive amount of data on social media sites and various educational systems for future development. |
| 2. | Anas Abdelrazeq, Daniela Janssen, Christian Tummel, Sabina Jeschke | Sentiment Analysis of social media for Evaluating Universities | Twitter sentiment can potentially support the universities ranking system by analyzing posted statements and opinions of students and teachers in higher education institutions context. | Natural Language Processing (NLP) | There is a lot of room in improving the project in regards to the ﬁltering, features selection and chosen classiﬁer type. The sentiment classiﬁcation can be extended to include a third neutral class. Also, the tweets can be classiﬁed based on topics to give a better insight on what attracts Twitter users in the context of higher education institutes. |
| 3. | Vijay Bharath, Shanthini, Subbarayan | Sentiment Analysis on the Performance of Engineering Students in University Examination: A Non-parametric Approach Using Two-way Analysis of Variance Model | There is no significant difference in the pass percentage of students between the different categories of the courses offered and there is also no significant difference in pass percentage of students between the different engineering streams. | Two-way ANOVA Model | Sentiment analysis can help institutes to understand how people feel about them: positive, negative, or neutral. Monitoring, including sentiment analysis, is one of the most important ways to keep people engaged and interested. |
| 4. | Pravin Katkar, Hrushikesh Bachakar, Sandhya Gude, Prof. Amruta Patil | Sentiment Analysis of Top Colleges in Pune Using Social Media Data | Review of various college in Pune according to classification of user reviews and suggest the top colleges in Pune. | Support Vector Machine (SVM) , Naïve Baiyes (NB) | Data can still be classified further using word cloud for further analysis of the reviews. |
| 5. | Archana Rao, Kishore Baglodi | Role of Sentiment Analysis in Education Sector in the era of Big Data: A survey | Role of Sentiment Analysis in the context of teaching effectiveness, from students’ perspective, in the context of research optimization. | NRC Lexicon | The paper highlights the importance of sentiment analysis in the era of big data in the education domain. Large amounts of data sets can be obtained, mined and realized to analyze sentiments. |
| 6. | Mansur Alp Tocoglu and Aytug Onan | Sentiment Analysis on Students’ Evaluation of Higher Educational Institutions | They have presented a comprehensive analysis of conventional classifiers, ensemble methods and text representation schemes for sentiment analysis on students’ evaluation of higher educational institutions. | Naïve Baiyes (NB) | The empirical results indicate that machine learning based approach yields promising results on students’ evaluation of higher educational institutions. |
| 7. | Prateek Garg | Sentiment Analysis of Twitter Data using NLTK in Python | They classified twitter data into sentiments by using different supervised machine learning classifiers on data collected for different political parties to show which political party is doing their best for the welfare of the public. | Naïve Bayes (NB), MNB, BNB, Logistic Regression Classifier, Linear SVC, Nu SVC, Natural Language Toolkit (NLTK) | Use of parser can be embedded into the system to improve the results. We can also work on multiple languages like Hindi, Spanish and Arabic to provide sentiment analysis to local opinions in their mother tongue as well. |
| 8. | Anusuya Dhara, Arkadeb Saha, Sourish Sengupta, Pranit Bose | Sentiment Analysis of Product-based reviews using Machine Learning Approaches. | The ultimate outcome of this Training of Public reviews dataset is that, the  machine is capable of judging whether an entered sentence bears positive response or  negative response. | Naïve Bayesian classifier, Random Forest classifier, SVM, Logistic Regression | Future research shall explore sophisticated methods for opinion and product feature extraction, as well as new classification models that can address the ordered labels property in rating inference. Applications that utilize results from sentiment analysis  is also expected to emerge in the near future. |
| 9. | Zenun Kastrati ,Fisnik Dalipi ,Ali Shariq Imran ,Krenare Pireva Nuci and Mudasir Ahmad Wani | Sentiment Analysis of Students’ Feedback with NLP and Deep Learning: A Systematic Mapping Study | Identified a variety of challenges regarding the application of sentiment analysis to examine students’ feedback. Consequently, recommendations and future directions to address these challenges have been provided. | NLP and Deep Learning | This study’s results will inspire future research and development in sentiment analysis applications to further understand students’ feedback in an educational setting. In future work, the plan is to further deepen the analysis that was performed in this mapping study by conducting systematic literature reviews (SLRs) |
| 10. | G. Priyadarshini & Dr. D. Karthika | Sentiment Analysis in Higher Education: A Systematic Mapping Review based Deep Neural Network | The adoption of the emerging field of sentiment analysis and opinion-mining in educational systems has a great potential in improving the quality of teaching and learning processes in HE institutions. | Deep Neural  Network | Opinion mining can also guide new and international students to select the appropriate university based on the review submitted by former students on social media sites. |
| 11. | Drashti Panchal, Mihika Mehta, Aryaman Mishra, Saish Ghole, Mrs. Smita Dandge | Sentiment Analysis Using Natural Language Processing | In the business world, sentiment analysis can be used to track customer satisfaction levels, to gauge public opinion about a company or product, or to monitor employee morale | NLP | NLP is still not widely used and its complexity can be a barrier for the average person. This is where data visualization can be helpful. By presenting the data in an easy-to-understand format, such as graphs and charts, people can get a better grasp of the sentiment of a particular product or topic. |
| 12. | Zhou Gui Zhou | Research on Sentiment Analysis Model of Short Text Based on Deep Learning | It can be concluded that adding the above shallow learning features and profound learning features to the short text sentiment analysis. | Deep Learning | Theoretically, the semantic information of the LSTM model text is good, but from the practical point of view, the LSTM model still has deficiencies, and the BM-ATT-BiLSTM method in this paper combines the emotional part-of-speech features, location information features, and dependency features of words. |

**3. PROBLEM STATEMENT**

In today’s world, where it has become increasingly important to adapt with the fast-changing trends, the requirements of the people keep changing. Therefore, to be valued by the common masses, one must keep improving their product/ organization/ business based on the customer’s feedback and reviews. Doing so is of utmost importance if one does not want their business to become obsolete. In this project, we will implement a module that uses an NLP sentiment analysis algorithm. We will be classifying the online reviews into positive, negative or neutral sentiments. This will help people to analyze the masses’ response to their business and help them identify their areas of weakness, so that they can work on improving themselves as well as their product. Once we have classified the data set, we will also graphically represent them using various graphs and Word-clouds for better understanding of the computer results.

**4. MOTIVATION**

When given the chance to work on a project, our first consideration was that it should be practical in nature and help in our development as computer engineers. We started by talking about the students' different areas of interest, which included data science, machine learning, and artificial intelligence.

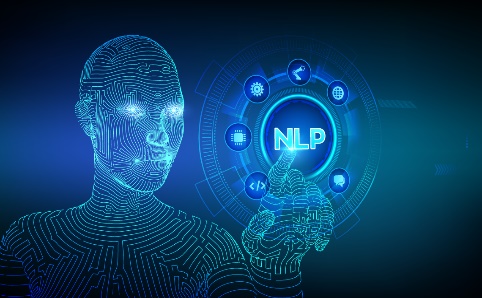


Fig 4.3 NLP

Fig 4.2 Python

Fig 4.1 Data Science

Additionally, we talked about various problems and looked for feasible solutions.

We sought a topic that would reflect the most recent developments in technology. Then, we evaluated which languages we felt most comfortable using. We chose Python as our programming language because of this. We were interested in using some of the buzzwords that are currently popular in the IT industry, which further motivated us to research machine learning and natural language processing (NLP).

This eventually led us to our topic, ‘SENTIMENT ANALYSIS’.

**4.1 Sentiment Analysis**

Sentiment Analysis commonly referred to as opinion mining, is a method for attempting to determine an object's positive, negative, and neutral qualities. Natural language processing (NLP) is one of the techniques that may be used for analyzing sentiments.

In our daily lives, we turn to a product's reviews to determine whether or not to purchase it. In a way, reviews assist in our decision-making process. As we began to consider this, we became interested in the potential benefits of applying NLP to perform statistical reviews of our college.

Fig 4.1.1 University/College

Our time at college has been wonderful. We began to wonder about how the outside world saw our college as we deliberated the application of NLP. People who haven't attended this college directly can only learn about it from reviews posted on various social media sites. We made the decision to work on this project at that point.

Recently, our college applied for NBA accreditation, which further encouraged us to take on this project so as to familiarize ourselves with an outsider’s point of view. In this way, we would also be able to alleviate people’s perceptions of our college.

**5. OBJECTIVES**

Fig 4.1.2 People’s perception of our college

* The objective of this project is to build a sentiment analysis system. Sentiment analysis can gather people’s opinions on many levels of feedback and give various institutions information to help them make wise decisions. In this study, the sentiment analysis of datasets is examined.
* In this research, we have developed an algorithm that automatically categorizes text as positive, negative or neutral using natural language processing and then visualizes the sentiments thus examined.
* For this, we gathered student reviews from a variety of platforms and then categorized the text as positive, negative or neutral using natural language processing. Feedback from people is crucial because it can help the associated stakeholder better understand how they are perceived.
* In this case, by giving feedback, students can let the professor know when they don't comprehend what the lecturer is trying to explain. Feedback from students can also be used to understand a variety of problems that students encounter, such as those related to not understanding the lecture. For teaching to be improved, feedback must be gathered.
* Students and professors would be able to handle teaching and learning concerns in the most advantageous and effective manner possible for the students. Students' opinions on the present teaching methods can be determined by analyzing student comments using sentiment analysis techniques.
* We have also used a bigger data set obtained from Kaggle. This will be used to demonstrate that our project is not limited to just small data i.e., it can be used for big data as well. Another important feature highlighted here is that our project is not restricted to any specific field. As long as we have the data set, our project can be implemented for it. This demonstrates the flexibility of our project.
* Our end goal is to achieve a model wherein sentiments can be checked for a given set of data. This will help in gauging the people’s emotions pertaining to a given situation or scenario.

**6. SCOPE**

We are utilizing student testimonials from various social media sites for this project, in addition to acquiring feedback from both alumni and presently enrolled students. This will add to our set of data. In order to get input from both our alumni and the students enrolled in the institute, we additionally designed a Google Form. The improvement of teaching, decision-making, and evaluation techniques used or adopted by the institution is the effect of sentiment analysis on educational procedures. The development of sentiment analysis technologies and the processing of student feedback by educational institutions are major investments aimed at eliciting the thoughts and insights of their students. In this work, the sentiment analysis of student feedback has been examined.

Since the data procured for the college reviews was not large enough, we have also made use of a pre-existing data set from Kaggle that consists of movie reviews from various critics.

Adopting sentiment analysis can help improve any sector. It gives a general overview of how ML and NLP work in sentiment analysis. In order for sentiment analysis to be used in the a space, it is necessary to explore trends and issues in the related field. Sentiment analysis offers the ability to extract people’s opinions with their sentiment orientation at the document level, phrase level, entity level, and aspect level. The document-level analysis examines a comment to discover whether the comments are generally good, negative, or neutral for a course. The sentiment is extracted from each sentence at the level of the sentence and aids in determining the strengths and disadvantages of a course. To give students' opinions, entity-level sentiment extraction combines entity and sentiment analysis.

The sentiment orientation of each data category is examined via aspect-based sentiment analysis, which analyses numerous data categories in a comment at a finer level. Data from various inputs will be examined at several levels based on an educational application.

We will classify the data in our project according to emotions, such as good, negative, and neutral. Pie charts and graphs will also be used to depict this data graphically.

**7. METHODOLOGY/ PROPOSED SYSTEM BLOCK DIAGRAM**

**7.1 SIPOC Diagram**

Before implementing a business process, it is helpful to visually document it by using a SIPOC (suppliers, inputs, process, outputs, customers) diagram. Due to their lack of profundity, SIPOC diagrams (pronounced "sigh-pock") are additionally referred to as high level process diagrams. SIPOC diagrams are helpful for narrowing down a debate, assisting team members in developing a shared language, and comprehending an iterative improvement process. To connect the value of the customer to a business process, some organizations employ COPIS diagrams that use SIPOC diagrams in the other direction.

Supplier: The "suppliers" for this project are students, alumni, and social media websites. Comparing tweets and Facebook postings would be simpler now that social media is the most common form of communication in modern society. Since any social media site can be accessed from any personal device connected to the university's wireless network, using social media is also completely free.

Input: For our project, we are using student feedback as well as reviews and comments from various social media platforms.

Process: One of the most popular NLP applications for determining human intent from reviews is sentiment analysis. Opinion mining is done in the educational industry to collect student comments and pedagogically improve learning and teaching protocols. For language processing, we applied the Vader algorithm since it provides better accuracy and efficiency.

Machine Learning methods can be supervised learning or unsupervised learning, and it applies the common ML algorithms and uses linguistic features. Sentiment analysis-based data refinement and NLP were used to categorize students' perspectives. Vader, a statistical and data mining tool, was used to assess the algorithm's precision and the sentiment of tweets. In our project, the data analysis is done using these two methodologies.

Output: System designers and educational officials should gain a comprehensive understanding of how students feel about the e-learning system in order to take the required interventions and actions to improve the situation. Students comment on their educational experiences on social media networks. A sample of data was analyzed using sentiment analysis to learn what the students thought and understood about the educational methods used in our institute. As people from all over the world communicate with one another on a single social media platform to debate diverse topics, social media is a potent source of data.

Since instructors and administrators can better understand students' areas of concern by analyzing sentiment and satisfaction parameters in their posts, comments, and feedback, this result indicates that sentiment analysis has a significant potential to enhance teaching and learning processes in universities. Our study highlighted the fact that social media sentiment analysis can offer precise and helpful data for future planning and assist developers in improving the services of institutions, i.e., it can offer suggestions to improve the institution. For easier understanding, we have graphical representations of the analyzed data.

Customer: It is possible to determine whether students have positive or negative feelings towards the institute by analyzing their feedback using sentiment analysis tools. So, solving institutional challenges is advantageous for institutions.

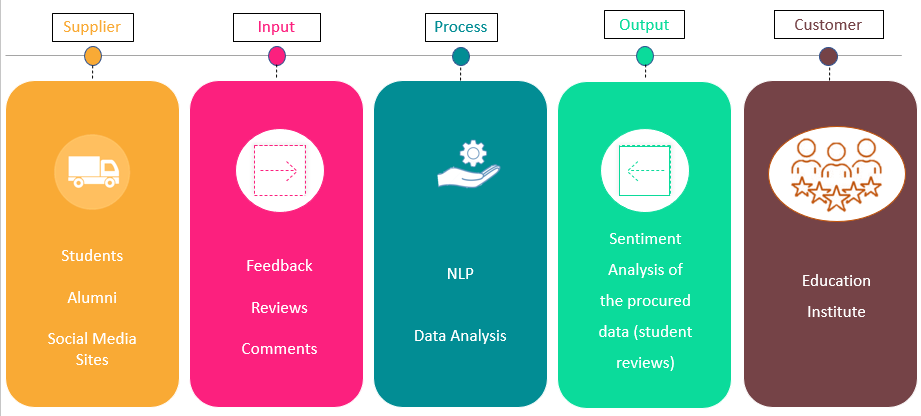


Fig 7.1.1 SIPOC Diagram for our project

**7.2 Modeling**

* Word Cloud

A word cloud is an example of a visual representation. It's a collection of words that are shown in different sizes; the larger and bolder the term, the more frequently it appears in a document and the more significant it is.

Data visualisation, text data, font colours, word frequency analysis, and individual word images are all included in text clouds. These methods, sometimes known as tag clouds or text clouds, are excellent for obtaining the most pertinent data from databases.

This data visualisation assisted us in understanding the keywords of the entire database that we acquired. Keywords allowed us to better classify and handle information by allowing us to understand the students' true feelings about the institute.

* NLP

The field of computer science known as "natural language processing" (NLP) is more particularly the field of "artificial intelligence" (AI), which is concerned with providing computers the capacity to comprehend written and spoken words in a manner similar to that of humans.

NLP powers software that translates text between languages, reacts to spoken commands, and summarises massive amounts of text in real time. In the form of voice-activated GPS systems, digital assistants, speech-to-text dictation software, customer care chatbots, and other consumer conveniences, we have engaged with NLP. However, NLP also plays a significant role in developing solutions that aid in streamlining corporate operations, boosting worker productivity, and streamlining business procedures.

## Working of NLP

Before implementing any kind of business case in natural language processing, there are a few phases or pre-processing processes that we need to take care of. It differs from pre-processing for machine learning.

* Cleaning data: This involves lowercasing of data, removing punctuations, removing numbers, and much more so that data should be easy to read by the program.
* Tokenization: It is used to split paragraphs and sentences into smaller units that can be more easily assigned meaning.
* NLTK (Natural Language Toolkit): It is one of the most powerful NLP libraries which contains packages to make machines understand human language and reply to them in an appropriate desired response.

For our project, we used Jupyter Notebook, a web-based interactive computing platform.

* Jupyter-Notebook: It is the original web application for producing and sharing computational documents. It provides a straightforward, simplified, and document-focused environment.
* Libraries:

There are various libraries which have been used as:

1.matplotlib.pyplot: This library is used for visualization and plotting.

2.WordCloud

3. Stopwords

4.nltk.sentiment

5. Vader

Python was used for the analysis's implementation of the code.

* Polarity

It is used for calculating the polarity. Polarity is the perspective of the stated emotion determined by the element’s sentiment, which determines whether the text communicates the user’s positive, negative, or neutral feelings toward the entity in question. Two new columns of subjectivity and polarity are added to the data frame. Polarity is the output that lies between [-1,1], where -1 refers to negative sentiment and +1 refers to positive sentiment.

* Data Aggregation

In order to reach to a conclusion or result, we collected the data that fuelled our problem statement. Data collection is one of the important tasks which should be performed at earliest. We identified various sources from which we gathered our data. Those sources are:

* Twitter
* Facebook
* Shiksha.com
* Collegedunia.com
* Careers360.com
* Getmyuni.com
* Collegedekho.com
* Google Forms
* Alumni Feedback

Data collection was a cluttered process, given that the data did not come from an established source. Our dataset's construction required utilising a range of tools and approaches as well as working with multiple sources.

**7.3 Block Diagram:**

Data Collection

Processing of Data

NLP Algorithm

Sentiment Classification

Presentation of Results

**[[**

Fig 7.3.1 Block Diagram

**7.4 System Architecture**

Online Reviews

Google FormResponses

College Website

Exploratory Data Analysis

Various WH -Questions

Filtering of Repeated Words

Special Character Treatment

Emojis

Positive

Negative

Classification using NLP

Fig 7.4.1 System Architecture Diagram

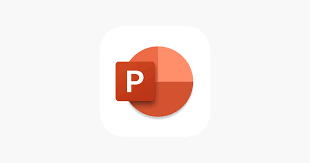
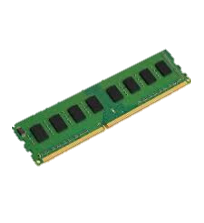
**8. SOFTWARE AND HARDWARE REQUIREMENTS**

**Hardware Requirements:**

* Processor : 11th Gen Intel(R) Core (TM) i7-1195G7 @ 2.90GHz 2.92 GHz
* Installed RAM : 16.0 GB (15.8 GB usable)

**Software Requirements:**

* Google Colab
* Jupyter Notebook
* Microsoft Word
* Microsoft Excel
* Microsoft Power Point
* Power BI
* Pandas, NumPy, Vader Libraries
* NLTK Corpus Library
* Count Vectorizer
* Sklearn Library
* String library
* Matplotlib Library



* Wordcloud

Fig 8.1 Software and Hardware Requirement for the Project

**9. IMPLEMENTATION**

**9.1 Coding:**

We had 3 different files for the data set as we collected it from different resources. Therefore, we used the data collected via these three data sets for our program code. We first import the csv files as dataframes. Then we name the columns as c1, c2 and so on and delete all other columns except for the one containing the data (reviews). After that, using the Vader library, we categories the sentiments of the given data into positive, negative and neutral. Once this is done, we combine all three dataframes and convert it all to lowercase. After that, we make the data pure by removing the stopwords. We then export the final csv file and display the so procured dataset with the help of the wordcloud.

Since the data set obtained from our college was not as large as we had hoped, we were unable to test our algorithm on a big dataset. For this, we decided to use a pre-existing dataset from Kaggle that consisted of the movie reviews from film critiques. In the critiques’ dataset each record represents a critic review published on Rotten Tomatoes, with the URL used for the scraping, critic name, review publication, date, score, and content. We used these reviews as our data set so that we could test our code for big data sets as well.

Although it took us almost an entire day to optimize our code in order to shorten the execution time, we successfully managed to use our code for this big data as well. This helped us realize the importance of code optimization. Not only that, it also helped us test and realize that our project is not only limited to our college, but as long as we have the proper data set, it can be used in any industry. It has helped us to reduce the execution time considerably.

**9.2 Visualization of Data:**

**[I] College Data**

Fig 9.2.1 Positive Word Cloud



Fig 9.2.2 Negative Word Cloud



Fig 9.2.3 Neutral Word Cloud

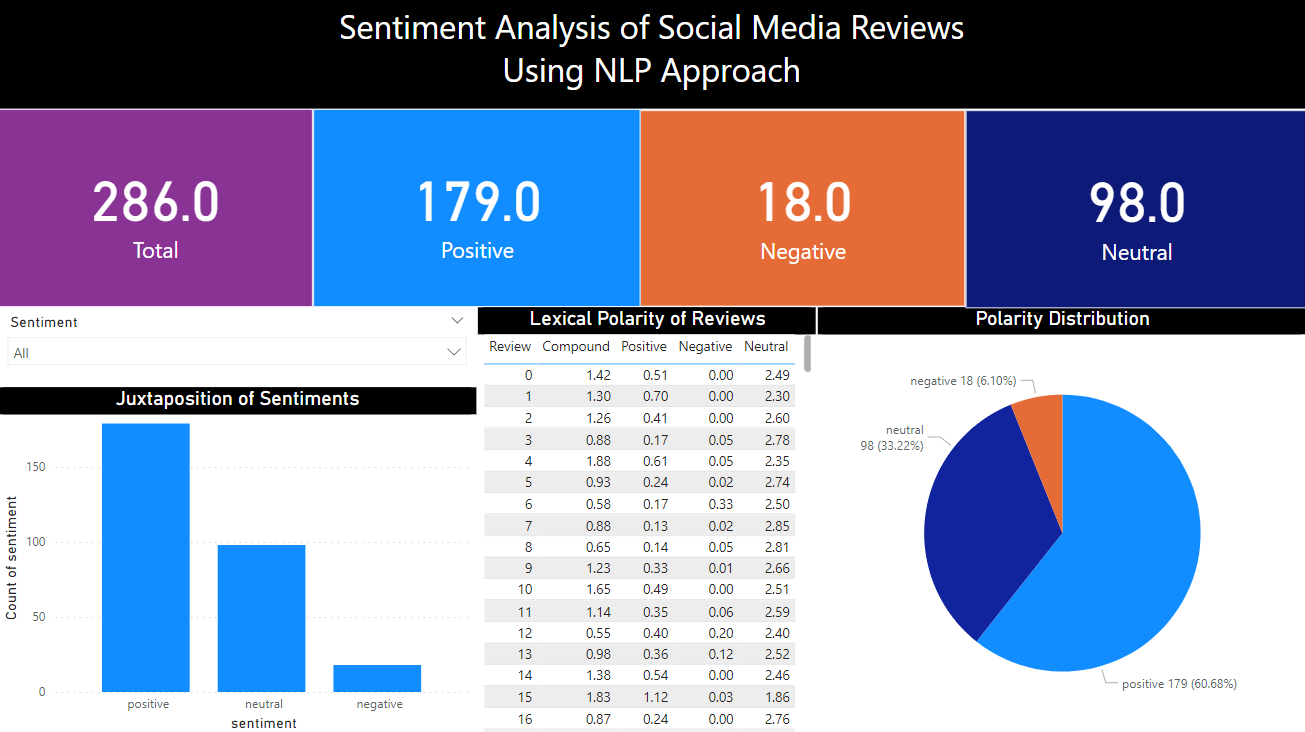
******

Fig 9.2.4 Visualization of Data using Microsoft Power BI

**[II] Rotten Tomatoes Movie Reviews data set**

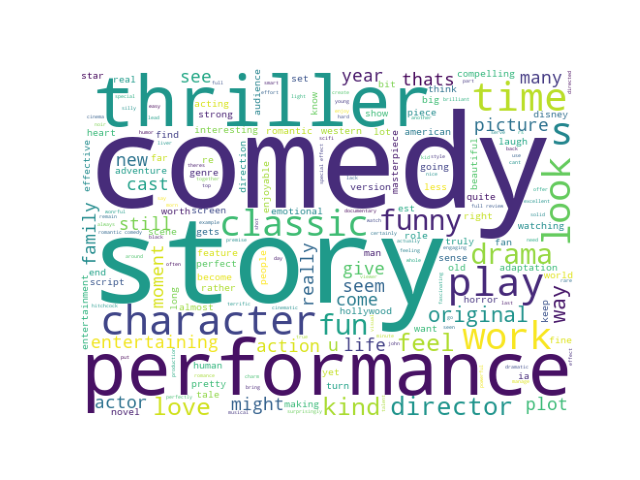


Fig 9.2.5 Positive Word Cloud



Fig 9.2.6 Negative Word Cloud



Fig 9.2.7 Neutral Word Cloud

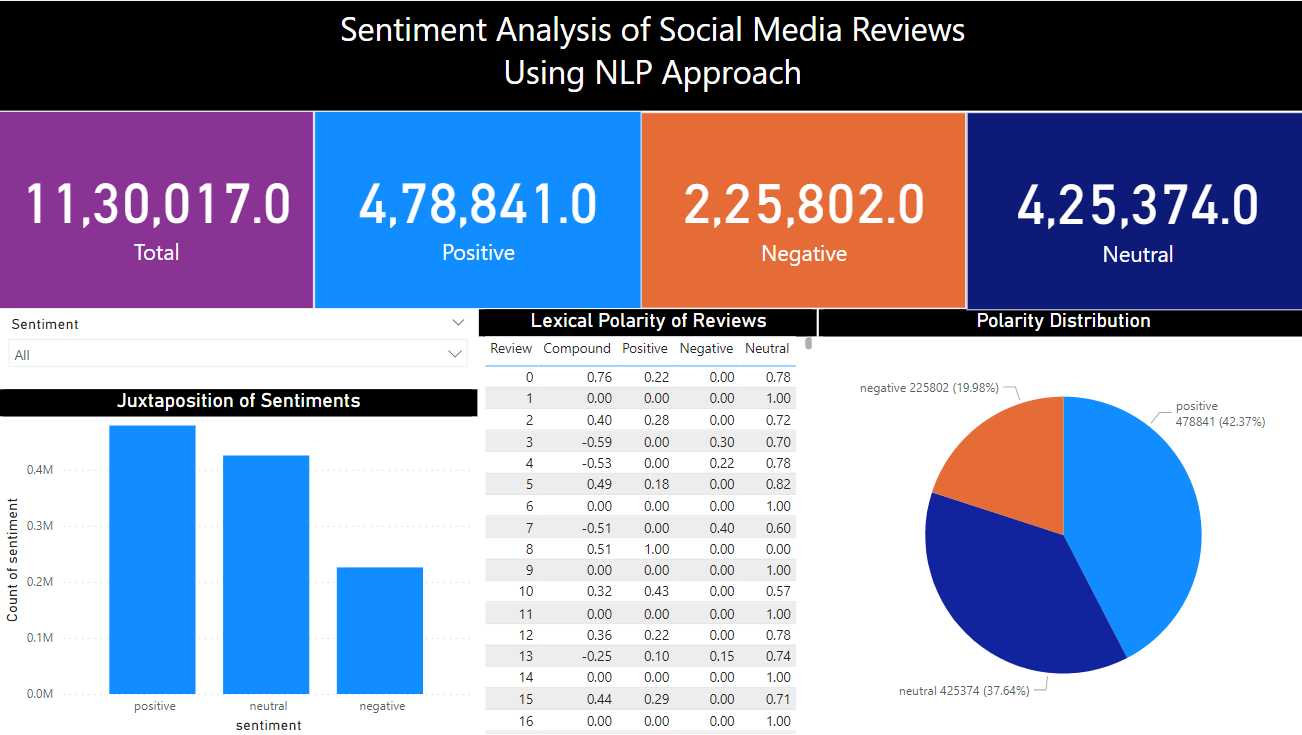
******

Fig 9.2.8 Visualization of Data using Microsoft Power BI

**10. RESULTS AND DISCUSSION**

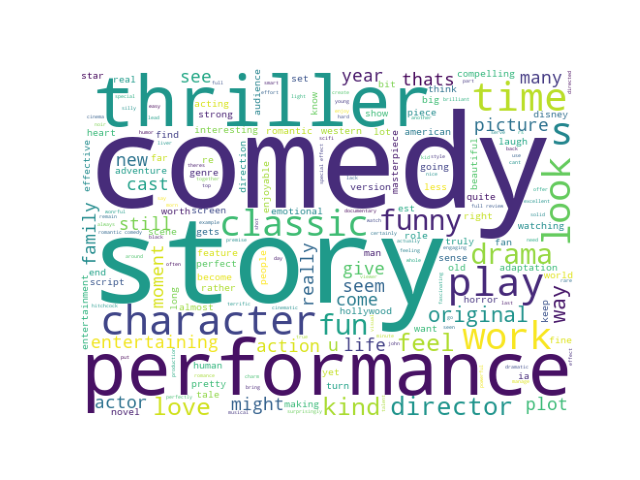
We gathered the data set for analysing the college reviews. Since the data set was small, we weren't able to check our algorithm for big data, so we used a pre-existing data set from Kaggle that consisted of the reviews from 'rotten tomatoes' and we were successfully able analyse it. We also represented it using word cloud so that anyone who sees this can easily understand the highlights of a given data. Since our data set consists of critic reviews for Hollywood movies, our analysis is helpful for the Hollywood industry to understand their audience’s preferences. It provides the overall idea of the movies’ standing in the Hollywood industry.

Fig 10.1 Positive Word Cloud

The positive word cloud helps us realize what the audience likes, what they expect to see and what to keep working on. For example, the words like **comedy, thriller, action, and love** are highlighted on this word cloud. This shows that this interests the audience and they enjoy content relat4ed to this in the movies. This tells us that the common masses enjoy movies with themes relating to love, comedy and action.



Fig 10.2 Negative Word Cloud

The negative word cloud helps us realize what the audience dislikes, what they do not like to see and what to improve on. For example, the words like **war, horror, drama, and plot** are highlighted on this word cloud. This shows that these genres are not received that well by the audience and that there is scope for improvement. It means that there is still scope for the improvement for various things like the plot of the movies, the horror or dramatic elements of the movies and so on.

People’s emotions are directly proportional to the revenue collected by the film makers who eventually pay taxes. Therefore, good quality movies being produced will directly help improve the economy. Since these are critic reviews, it often helps shape the opinion of the masses. It will thus help in the overall development of the quality of the Hollywood industry.

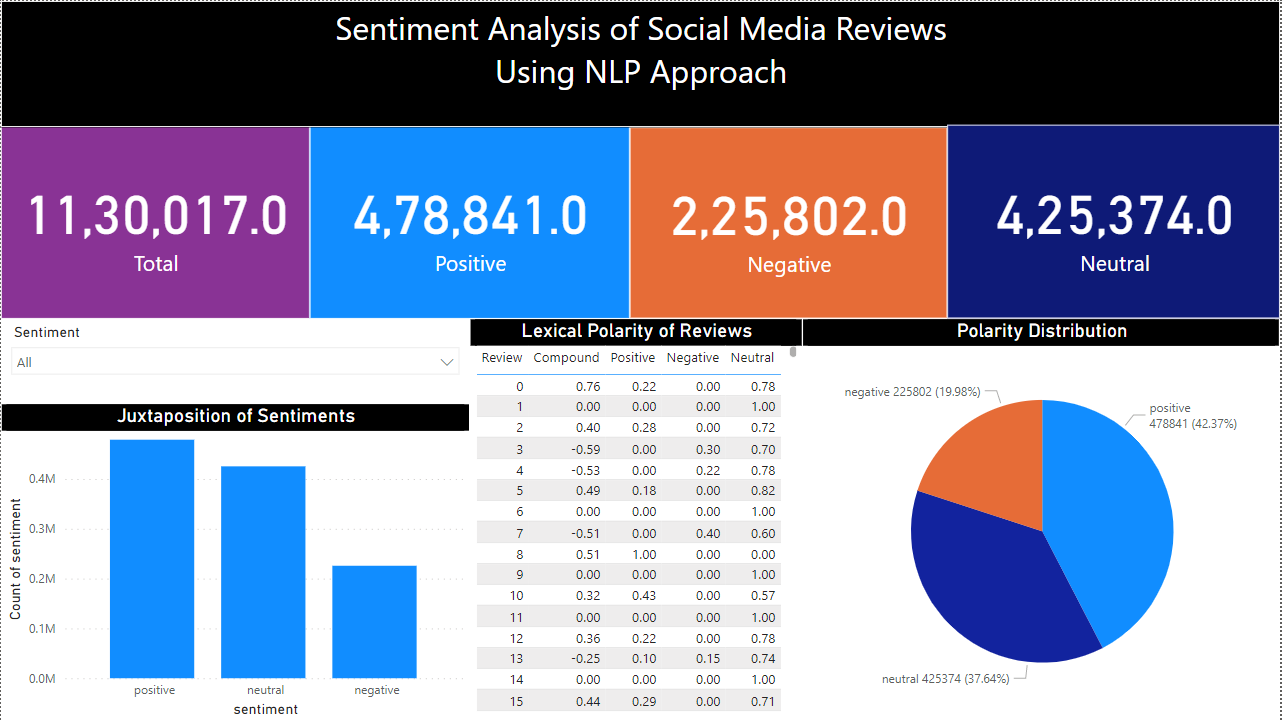
The picture below is the visualization of the result that has been compiled for our data set obtained from Kaggle using the Microsoft Power BI software. It makes our data more readable and easier to understand.

Fig 10.3 Visualization of Rotten Tomatoes Reviews

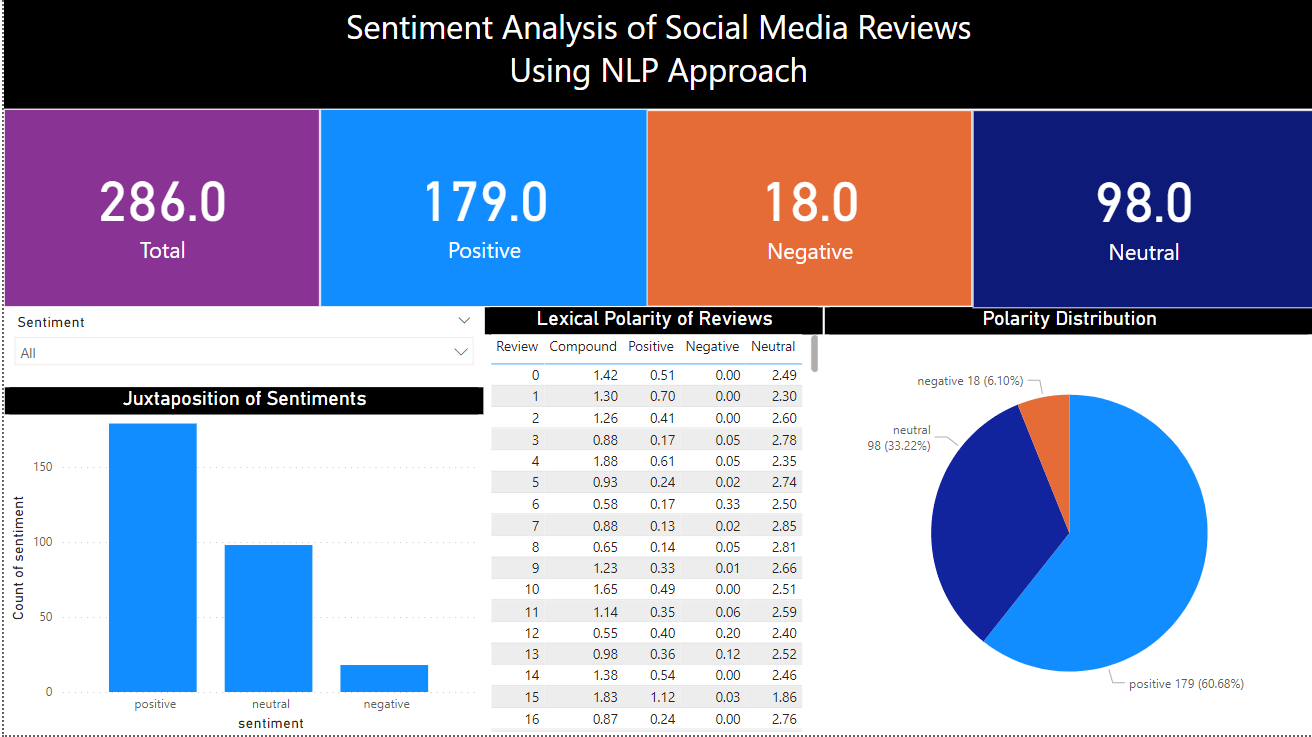
Similarly, we have visualized the results for the data set collected from our college as well using the various tools available in Microsoft Power BI for an easy and enhanced grasping of the computed results.

Fig 10.4 Visualization of College Reviews

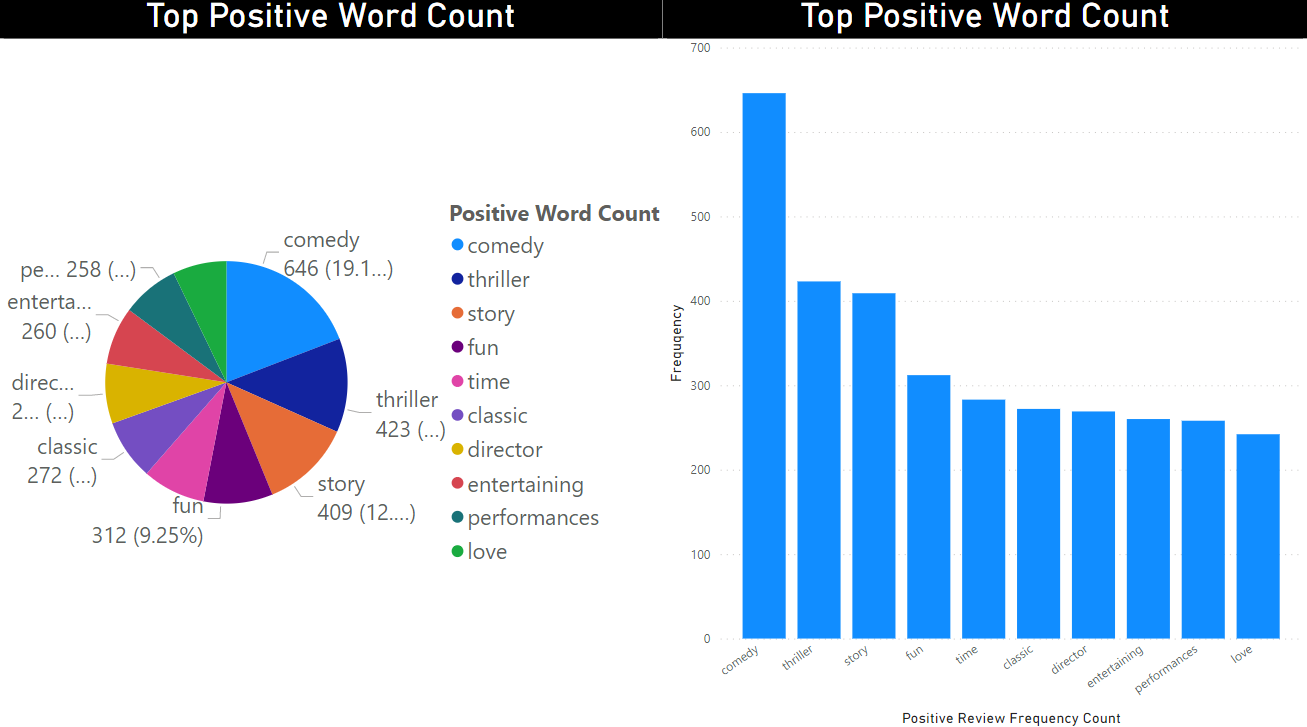
In addition to the visualization of the entire result, we also used separate visualizations to present the most frequently occurring words along with their frequency percentage. This type of representation tells us about which word has the highest frequency along with its word frequency count. This helps us better understand our strengths and weaknesses.

Fig 10.5 Frequency of Negative Words

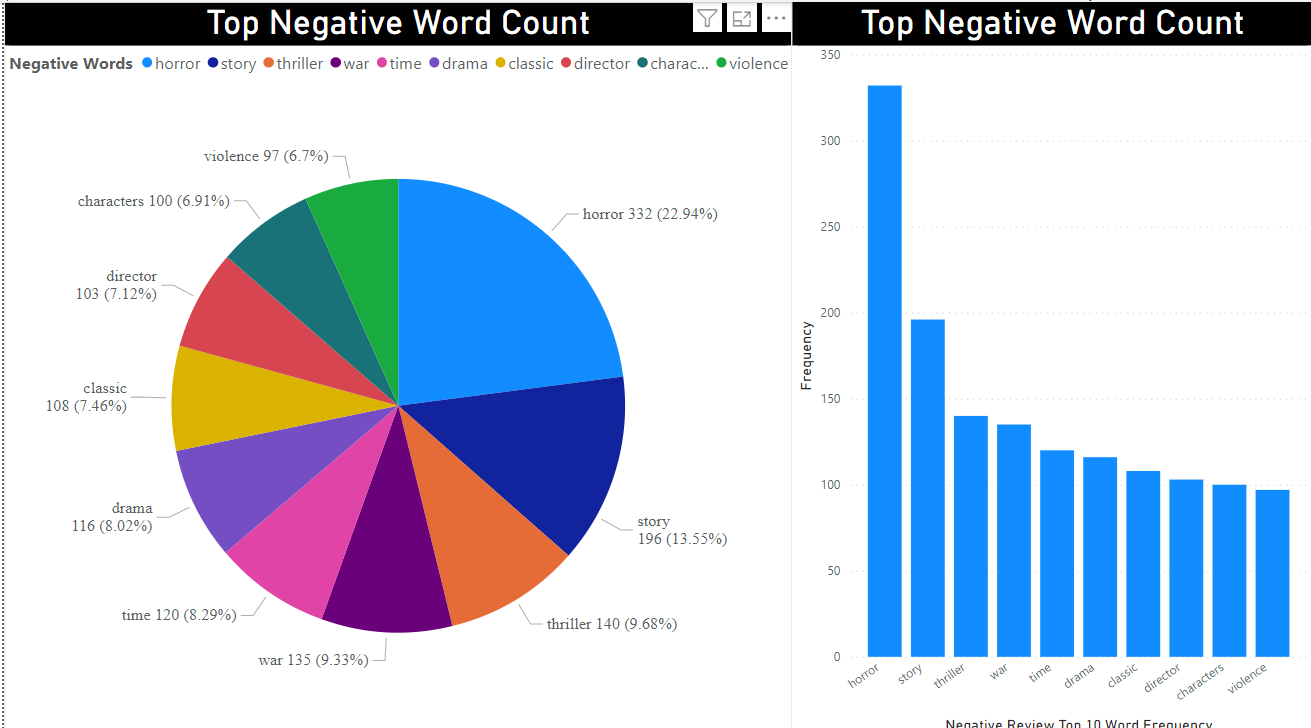


Fig 10.6 Frequency of Positive Words

**11. CONCLUSION AND FUTURE SCOPE**

**11.1 Conclusion**

When we first showed our project based on the college data set collected to our mentor, it was brought to our notice that the data set was too small. Hence, we were advised to use a larger data set. This led to all of the group members brainstorming about the various possible sources for data sets. While doing so, we came across the rotten tomatoes critic reviews data set on Kaggle. Since it perfectly fit our requirement criteria, we decided to use this data set to test our algorithm for big data.

In our project we have analysed the given data set into positive, negative and neutral sentiments. Therefore, it was successfully analysed that our algorithm worked for the big data set and we could differentiate it into positive, negative and neutral sentiments. We also represented the frequently occurring data with the help of word cloud to highlight the important sentiments of the people.

**11.2 Future Scope**

In our project we have used Vader for analysing the data. There are various such algorithms which can be used for sentiment analysis. We can use different algorithms to improve the accuracy of our model depending upon the data set used. In doing so we can make a comparative study of all the algorithms thus enabling ourselves to choose the best algorithm for each data set.

We can further develop our algorithm to recognize the sentiments expresses by various emoticons, emojis and other special characters that are often used by the masses while expressing their views.

Another future aspect of our algorithm we can work on would be to develop a system such that it scans the facial expressions of the customers inside a store, shop or mart while shopping, exiting or billing. After scanning, we can translate the scanned expressions into their corresponding emotions. This will help us in better understanding of the sentiments of the public.

**12. REFERENCES**

1. Reem Baragash and Hanan Aldowah. Sentiment analysis in higher education: a systematic mapping review in the Journal of Physics Conference Series Ser. 1860 012002, DOI 10.1088/1742-6596/1860/1/012002
2. Archana Rao P. N. and Kishore Baglodi. The Role of Sentiment Analysis in the Education Sector in the Era of Big Data: A Survey in the International Journal of Latest Trends in Engineering and Technology special Issue SACAIM 2017, pp. 022-024 e-ISSN:2278-621X.
3. Aytug Onan and Mansur Alp Toçoğlu. Sentiment Analysis on Students’ Evaluation of Higher Educational Institutions in the book “Intelligent and Fuzzy Techniques: Smart and Innovative Solutions” (pp.1693-1700), DOI:10.1007/978-3-030-51156-2\_197.
4. Anas Abdelrazeq, Daniela Janssen, Christian Tummel, Sabina Jeschke. Sentiment Analysis of Social Media for Evaluating Universities. 2nd International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC 2015), Dubai, UAE, 16-18 December 2015.
5. Drashti Panchal, Mihika Mehta, Aryaman Mishra, Saish Ghole, Mrs. Smita Dandge. Sentiment Analysis Using Natural Language Processing in the International Journal for Research in Applied Science and Engineering Technology Volume 10, 5th Issue of May 2022.
6. Vijay Bharath, Shanthini, Subbarayan. Sentiment Analysis on the Performance of Engineering Students in University Examination: A Non-parametric Approach Using Two-way Analysis of Variance Model in Webology, Volume 18, Special Issue on Artificial Intelligence in Cloud Computing, April 2021.
7. Pravin Katkar, Hrushikesh Bachakar, Sandhya Gude, Prof. Amruta Patil. Sentiment Analysis of Top Colleges in Pune Using Social Media Data in International Journal of Innovative Research in Science, Engineering and Technology.
8. Prateek Garg. Sentiment Analysis of Twitter Data using NLTK in Python in a research article from Thapar University.
9. Anusuya Dhara, Arkadeb Saha, Sourish Sengupta, Pranit Bose. Sentiment Analysis of Product-based reviews using Machine Learning Approaches in a Research paper from RCC institute of technology, West Bengal.
10. Zenun Kastrati, Fisnik Dalipi ,Ali Shariq Imran, Krenare Pireva Nuci and Mudasir Ahmad Wani. Sentiment Analysis of Students’ Feedback with NLP and Deep Learning: A Systematic Mapping Study an open access journal, published on 28 April 2021.
11. G. Priyadarshini & Dr. D. Karthika. Sentiment Analysis in Higher Education: A Systematic Mapping Review based Deep Neural Network from London Journal’s Press.
12. Zhou Gui Zhou. Research on Sentiment Analysis Model of Short Text Based on Deep Learning from a research paper of Computer College of Huanggang Normal University, Huanggang, Hubei 438000, China.
13. Pooja, Bhalla, R. A Review Paper on the Role of Sentiment Analysis in Quality Education. SN COMPUT. SCI. 3, 469 (2022). https://doi.org/10.1007/s42979-022-01366-9SA-E: Sentiment Analysis for Education by Nabeela Altrabsheh, Mohamed Medhat Gaber, Mihaela Cocea.
14. Thanveer Shaik, Xiaohui Tao, Christopher Dann, Haoran Xie, Yan Li, Linda Galligan. Sentiment analysis and opinion mining on educational data: A survey. Natural Language Processing Journal. Department of Computing and Decision Sciences, Lingnan University, Tuen Mun, Hong Kong.