# EXAMINING BRAND PERCEPTION CHALLENGES: VADER APPROACH VS. TRADITIONAL ANALYSIS FOR IMPROVED UNDERSTANDING

**MINI PROJECT REPORT**

***Submitted by***


| | |
|---|---|
| **Vidhiya S B** | **210701306** |
| **Vaisharli S** | **210701296** |


***in partial fulfillment for the award of the degree of***

**BACHELOR OF ENGINEERING**

***in***

**COMPUTER SCIENCE AND ENGINEERING**


**RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI**

**ANNA UNIVERSITY:: CHENNAI 600 025**


**APRIL 2024**

# RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

## BONAFIDE CERTIFICATE

Certified that this Report titled "**Examining Brand Perception Challenges: VADER Approach vs. Traditional Analysis for Improved Understanding**" is the bonafide work of **"Vidhiya S B (210701306), Vaisharli S (210701296)"** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

**SIGNATURE**

**Dr. V. Karthick., M.E., Ph.D.,**
**Associate Professor,**
Department of Computer Science and Engineering,
Rajalakshmi Engineering College,
Chennai – 602015

Submitted to Mini Project Viva-Voce Examination held on ____________________

**Internal Examiner** **External Examiner**

## ABSTRACT

In today’s digital era , Social media plays a key role in Brand’s reputation and perception. Using a Data driven approach, It is easier to analyze the brand’s public review and where it stands among all the organizations. It helps the brand to understand and analyze their position and helps them improve their brand. Sentiment Analysis is one of the NLP techniques used to analyze positivity, Negativity or neutrality of a data, which helps in accurate brand perception. It is impossible for an individual to check lakhs and lakhs of comments on social media and it is difficult to analyze the reputation of the brand among the public. This paper focuses on VADER (Valence Aware DIctionary and Sentiment Reasoner). classification which is a tool used for Sentiment Analysis within NLTK. VADER is a lexicon oriented sentiment analysis tool which can be used to analyze the customer reviews which are all over on the social media which can be uploaded as a csv to the User Interface and the user can understand how customers perceive a particular brand. The objective is to analyze and know how a brand is grasped among the people. This approach harnesses the power of VADER which is a part of NLTK ,which is used to scrutinize and analyze social media conversations. These help the business in targeted marketing , personalized brand improvement.

# ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Mr. S.MEGANATHAN, B.E, F.I.E.**, our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN**, **Ph.D.,** for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, Ph.D.**, Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide, **Karthick V** Professor, Department of Computer Science and Engineering. Rajalakshmi Engineering College for his valuable guidance throughout the course of the project.

**Vidhiya S B - 210701306**

**Vaisharli S - 210701296**

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| **VADER** | Valence Aware Dictionary and sEntiment Reasoner |
| **SVM** | Support Vector Machines |
| **NLTK** | Natural Language Toolkit |
| **NLP** | Natural Language Processing |
| **UI** | User Interface |
| **CSV** | Comma Separated Values |

# CHAPTER 1
# INTRODUCTION

## 1.1 GENERAL

Tools for sentiment analysis are crucial for determining public opinion and identifying the emotional undertone of textual material. Businesses, researchers, and politicians who need to analyze massive amounts of text, including social media postings, reviews, and survey replies, may find these tools very helpful. These solutions speed up trend identification, detect consumer contentment or displeasure, and track brand perception. The vast volume of textual data generated every day necessitates the use of sentiment analysis technologies. This data is too large and too complicated to analyze manually. Sentiment analysis tools classify text as positive, negative, or neutral using natural language processing (NLP) approaches. They can frequently offer more subtle insights into feelings and viewpoints.

## 1.2 OBJECTIVE

This project aims to create an interactive Streamlit sentiment analysis application that lets users upload data, evaluate sentiment scores using the NLTK VADER tool, and visualize the outcomes. Through an intuitive interface, users can upload text data to be processed into dataframes using pandas and translated to strings. Every text item undergoes VADER sentiment analysis, which yields sentiment scores ranging from -1 (negative) to 1 (positive) and divides them into three categories: positive, negative, and neutral. The program uses a pie chart to show the overall sentiment distribution, which is calculated by aggregating individual reviews. Additionally, by classifying text data according to the month of creation and arranging it in a line, it shows the temporal distribution of sentiments. Furthermore, the most common terms are indicated by word clouds created from the text data that do not include stop words. Lastly, utilizing the text data analysis, the DistilGPT-2 model from Hugging Face Transformers is employed to produce recommendations for brand enhancement. In order to improve consumer happiness and brand reputation, this all-encompassing method offers insights into brand perception and enables data-driven decision-making.

### 1.3 EXISTING SYSTEM

Various techniques are employed by current sentiment analysis and brand perception systems to derive significant insights. Brandwatch uses natural language processing (NLP) and deep learning models to identify topics and analyze sentiment on social media. It provides configurable dashboards with extensive statistics. Brand mention tracking and sentiment analysis are made possible by Hootsuite Insights' integration of AI and NLP algorithms to track sentiment trends in real time across social media channels. Sprout Social offers social media management tools as well as a comprehensive picture of brand health through sentiment analysis based on natural language processing (NLP). Talkwalker monitors brand mentions across a variety of online sources using AI-powered sentiment analysis and image recognition to provide actionable insights and pinpoint influential users. Lexalytics uses machine learning and natural language processing to analyze sentiment and themes in big text collections.

### 1.4 PROPOSED SYSTEM

Then the sentiment analysis is performed through VADER SentimentIntensityAnalyzer to obtain a compound sentiment score and sentiment labels(positive , negative or neutral) are assigned according to the score. Then the data is aggregated to generate visual insights of sentiment distribution and trends using histograms and pie charts.

Word clouds are generated to visualize the most frequent positive and negative words where the stopwords are excluded.The top 5 positive and negative reviews are displayed on the Streamlit User Interface.The suggestions are also included by using Hugging face transformer distil-gpt2.

# CHAPTER 2

# LITERATURE SURVEY

Ankita Gandhi , Kinjal Adhvaryu , Soujanya Poria , Erik Cambria and Amir Hussain work on a project called Multimodal sentiment analysis[1]. 'Multimodal' refers to the multiple modes of communication. It also includes the various forms of medium like audio, video, and in the form of text to enhance interaction experience. Sentiment analysis grows much in the field of artificial intelligence for analyzing sentiment about a product or services of an organization. Opinions and ideas are growing towards sentiment analysis. For this purpose, multimodal sentiment analysis focuses on collecting data in the form of video instead of text alone.

Cheng Chen, Bin Xu , Jong-Hoon Yang, and Mi Liu did a project called "Sentiment analysis of Animated Film reviews using Intelligent Machine Learning"[2]. This project aims to analyze the sentiment of users towards the animated movie, the dataset which is used to train the model is collected from various users by analyzing their behavior. Even in a social media or e-commerce platform uses the information of the users browsing history, customers favorites, add-ons, most viewed content and liked content based on this the recommendation of a content for a user is done. This paper focused on the construction of several deep learning models, intelligent machine learning-based text sentiment classification. These models were compared based on their experimental results and methods. This analysis towards the movie only achieves the overall sentiment of the review data and fails to analyze the specific context in the movie, specifically plot, special effects and content of the movie.

Burns, S. (2019). Natural Language Processing: A Quick Introduction to NLP with Python and NLTK[3]. This research provides a beginner-friendly introduction to NLP concepts using the NLTK library in Python. It covers core NLP tasks like text processing, tokenization, stemming, and lemmatization.

Bai, K., & Tan, K. H. (May 2024). The Influence of Online Social and Physical Presence on User Consumption Decisions in TikTok Livestreaming: A Scoping Review[4]. This recent scoping review (May 2024) examines how online social presence (followers, comments) and physical presence (livestreamers themselves) influence user decisions while watching live streams on TikTok.

Vought, V., Vought, R., Lee, A. S., Zhou, I., Garneni, M., & Greenstein, S. A. (Mar. 2024). Application of sentiment and word frequency analysis of physician review sites to evaluate refractive surgery care[5]**.** This research (March 2024) showcases a practical application of sentiment analysis. It analyzes reviews on physician websites to assess patient sentiment towards refractive surgery care.

Chen, C., Xu, B., Yang, J.-H., & Liu, M. (Jul. 2022)[6]. Sentiment Analysis of Animated Film Reviews Using Intelligent Machine Learning. This book focuses on applying machine learning algorithms to automate sentiment analysis tasks. The example provided is analyzing reviews for animated films.

Bhattacharyya, S., Snasel, V., Hassanien, A. E., Saha, S., & Tripathy, B. K. (2020). Deep Learning: Research and Applications[7]. This book provides a comprehensive overview of deep learning concepts, a powerful subset of machine learning. It offers a theoretical foundation for those interested in the underlying principles.

Christodoulakis, Christina, et al. (2020). Pytheas[8]: pattern-based table discovery in CSV files. This paper introduces Pytheas, a tool that automates table extraction from CSV files by identifying patterns within the data. This can be useful for preparing data for NLP or machine learning tasks.

Obagbuwa, I.C., Danster, S., & Chibaya, O. C. (Jul. 2023)[9]. Supervised machine learning models for depression sentiment analysis. This research explores using machine learning to analyze text data and potentially detect signs of depression. It raises ethical considerations regarding data privacy, as training such models would require access to data labeled with depression diagnoses.

Titelman, G. (2024). Evaluating the Societal Impact of Large Language Models. arXiv preprint arXiv:2405.02253[10].This paper by Titelman (May 2024) explores a crucial aspect of NLP advancements: the societal impact of large language models (LLMs). LLMs are powerful AI systems trained on massive amounts of text data, capable of generating human-quality text, translating languages, writing different kinds of creative content, and answering questions in an informative way.

# CHAPTER 3
# SYSTEM DESIGN

## 3.1 DEVELOPMENT ENVIRONMENT

### 3.1.1 HARDWARE SPECIFICATIONS

This project uses minimal hardware but in order to run the project efficiently without any lack of user experience, the following specifications are recommended

**Table 3.1.1** Hardware Specifications

| | |
|---|---|
| **PROCESSOR** | Intel Core i5 |
| **RAM** | 4GB or above (DDR4 RAM) |
| **GPU** | Intel Integrated Graphics |
| **HARD DISK** | 6GB |
| **PROCESSOR FREQUENCY** | 1.5 GHz or above |

### 3.1.2 SOFTWARE SPECIFICATIONS

The software specifications in order to execute the project has been listed down in the below table. The requirements in terms of the software that needs to be pre-installed and the languages needed to develop the project has been listed out below.

**Table 3.1.2** Software Specifications

| | |
|---|---|
| **FRONT END** | Python(Streamlit) |
| **BACK END** | Python |
| **FRAMEWORKS** | Transformers, Streamlit |
| **SOFTWARES USED** | Visual Studio, Jupyter Notebook |

## 3.2 SYSTEM DESIGN

### 3.2.1 ARCHITECTURE DIAGRAM

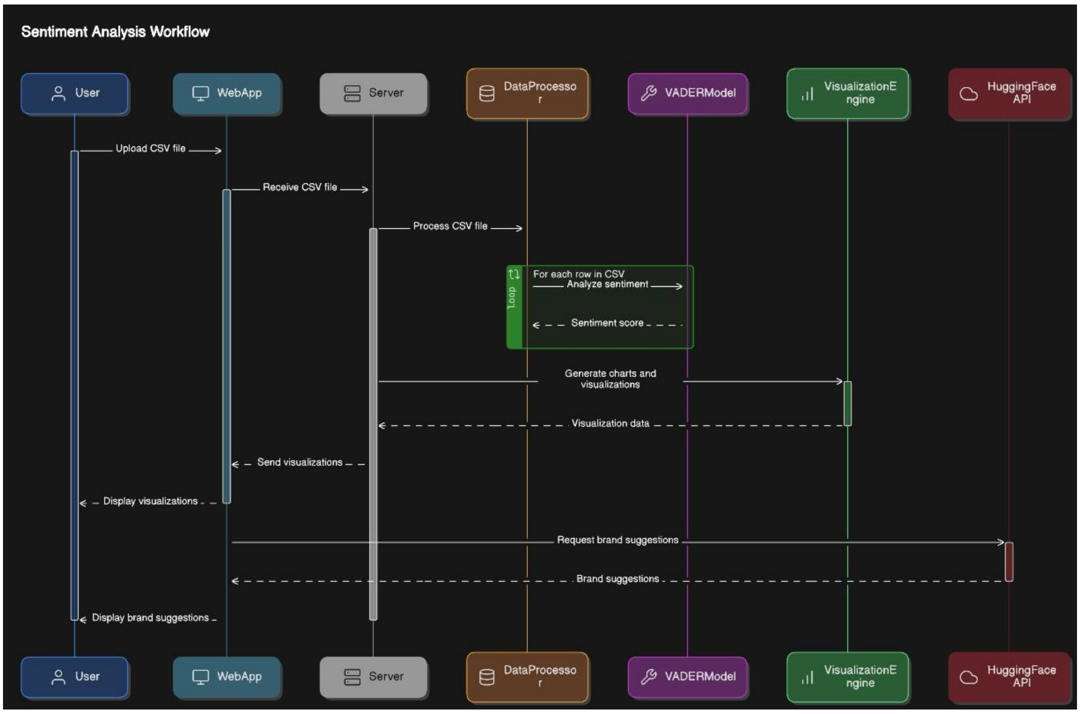


**Fig 3.2.1 Architecture Diagram**

**PRE-PROCESSING:**

The text data in the DataFrame is processed to ensure uniformity and compatibility for sentiment analysis. This includes converting the text columns to string type (**data["body"] = data["body"].astype("str")**)

**TRAINED MODEL:**

VADER Sentiment Analysis: An already-built sentiment analysis tool from the NLTK library is called VADER (Valence Aware Dictionary and Sentiment Reasoner). Because it makes use of a pre-defined sentiment lexicon, no additional training is needed for this application. Sentiment scores are calculated by explicitly applying the VADER model to the text input.

Hugging Face Transformers: The program performs two tasks using pre-trained models from Hugging Face Transformers:

Text Generation: In response to prompts, replies are generated using the distilgpt2 model. Without further training, this model can be used for text creation tasks because it has already been pre-trained on a sizable corpus of text data.

# CHAPTER 4

# PROJECT DESCRIPTION

## 4.1 MODULE DESCRIPTION

### 4.1.1 DATA PRE-PROCESSING:

This module involves uploading and preparing the data for analysis. Users can upload their data files (CSV or Excel) through the Streamlit UI. The data is then loaded into pandas DataFrames, and text columns are converted to strings to ensure compatibility with the sentiment analysis process.. Additionally, date columns are converted to datetime format for further temporal analysis.

### 4.1.2 SENTIMENT ANALYSIS:

The VADER sentiment analysis tool is used in this module to evaluate the sentiment of textual data. Sentiment scores, which range from -1 (negative) to 1 (positive), are produced by analyzingeach text entry. The text is categorized into Positive, Negative, and Neutral sentiments using thesescores. An overall sentiment distribution for the uploaded data is produced by combining the results.

### 4.1.3 DATA VISUALIZATION:

This module includes a number of visualizations designed to aid users in understanding the trends and sentiment distribution over time. Plotly can be used to create line charts that illustrate sentiment trends over predetermined time periods and pie charts that show the sentiment distribution for various brands. In order to display the sentiment distribution by count and percentage, histograms are also made. Furthermore, word clouds are created to illustrate the most commonly used terms in both positive and negative reviews.

### 4.1.4 SUGGESTION GENERATION:

This module uses pre-trained models from Hugging Face Transformers to generate improvement suggestions for brands based on negative reviews. The distilbart-cnn-12-6 model is used to summarize the negative reviews and make grammatical corrections. The distilgpt2 text generation model is then prompted with the updated summary to produce useful recommendations for enhancing the brand.

# CHAPTER 5
# IMPLEMENTATION AND RESULTS

## 5.1 IMPLEMENTATION

### 5.1.1 VADER:

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool designed to efficiently analyze and quantify the sentiment expressed in social media contexts. Part of the Natural Language Toolkit (NLTK) library, VADER excels in handling the casual and emotive language often found in online communications, such as social media posts, product reviews, and comments. VADER operates by leveraging a comprehensive dictionary where words are associated with predefined sentiment intensity scores, indicating whether they convey positive, negative, or neutral sentiment. This lexicon-based approach is complemented by rule-based heuristics that enhance its accuracy in sentiment detection. These heuristics account for textual nuances such as punctuation, capitalization, degree modifiers (words that amplify or dampen sentiment intensity like "very" or "extremely"), and conjunctions that shift sentiment focus (e.g., "but" indicating a contrast). VADER generates four sentiment metrics: positive, negative, neutral, and compound scores. The compound score is a normalized, weighted sum of the individual sentiments, ranging from -1 (most negative) to +1 (most positive), providing a holistic view of the overall sentiment of the text. VADER's robustness, speed, and ease of use make it a popular choice for sentiment analysis in various applications, especially those involving large datasets where rapid processing is essential. VADER's capacity to produce discrete scores for positive, negative, and neutral sentiments within a text in addition to a composite sentiment score is one of its key advantages. For summarizing overall sentiment, the compound score—a normalized measure that goes from +1 (most positive) to -1 (most negative)—is particularly helpful. VADER's usefulness goes beyond social media; it works well in a variety of contexts where prompt and precise sentiment analysis is crucial, such as product reviews, news articles, and customer feedback. Because of its efficiency and simplicity, it can be used for both commercial and scholarly applications. Furthermore, VADER is intended to function immediately out of the box without requiring a large amount of training data, which makes it a sensible option for tasks requiring trustworthy and timely sentiment analysis.

### 5.1.2 DISTILGPT2 TRANSFORMER:

DistilGPT2 is a streamlined and efficient version of the OpenAI GPT-2 model, optimized for natural language generation tasks through a process called knowledge distillation. This smaller, faster, and lighter variant maintains the powerful transformer architecture of GPT-2, which comprises multiple layers of self-attention mechanisms and feedforward neural networks. These layers enable the model to understand and generate text by capturing intricate dependencies within the data. The distillation process involves training the smaller model (DistilGPT2) to replicate the behavior and performance of the larger, original GPT-2 model, thus retaining much of its capability while reducing computational requirements. DistilGPT2's compact size and enhanced efficiency make it particularly suitable for applications where resource constraints are a concern, such as on-device processing or scenarios requiring real-time text generation. By leveraging the transformer architecture's strengths, DistilGPT2 can generate coherent and contextually relevant text, making it a versatile tool for tasks ranging from conversational AI and content creation to automated summarization and beyond. The Hugging Face Transformers library provides easy access to DistilGPT2, facilitating its integration into various natural language processing workflows and enabling developers to harness its capabilities with minimal overhead. DistilGPT2 has been adopted in a wide range of applications beyond standard text generation due to its performance and versatility. For example, it is utilized in customer service to automate responses and support, in educational technology to deliver individualized learning experiences, and in content creation tools to produce creative writing. The model is a useful tool in any field where producing and comprehending natural language is essential because of its capacity to produce human-like text that is coherent and aware of context. DistilGPT2, which strikes a balance between power and efficiency, is a big step toward improving the usability and practicality of advanced NLP models. DistilGPT2 has been adopted in a wide range of applications beyond standard text generation due to its performance and versatility. For example, it is utilized in customer service to automate responses and support, in educational technology to deliver individualized learning experiences, and in content creation tools to produce creative writing. The model is a useful tool in any field where producing and comprehending natural language is essential because of its capacity to produce human-like text that is coherent and aware of context. DistilGPT2, which strikes a balance between power and efficiency, is a big step toward improving the usability and practicality of advanced NLP models.

## 5.2 OUTPUT SCREENSHOTS

**Sentiment Distribution based on Phone Reviews:** A pie chart shows the distribution of positive and negative sentiment in the reviews.

**Reviews count according to the brand:** There's a table that shows the number of reviews for Nokia and Huawei. It also shows the sentiment distribution for each brand – the number of positive reviews and negative reviews.

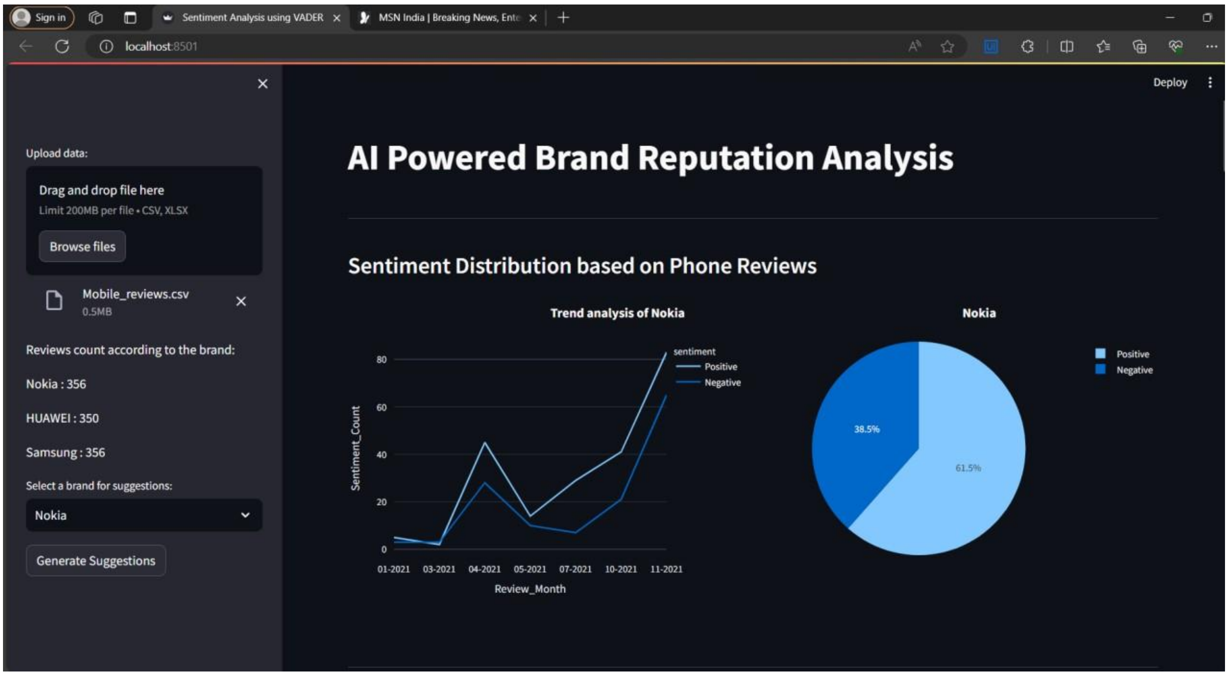


**Table 5.2.1 Sentiment Distribution**

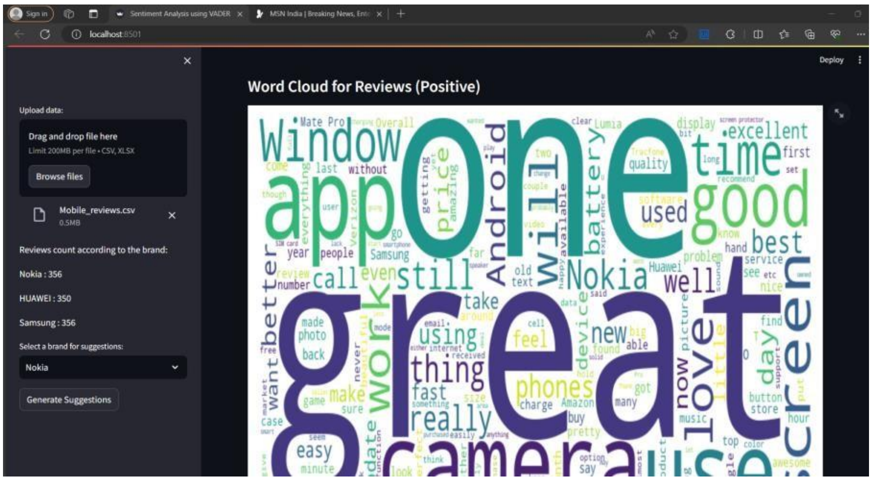


**Fig 5.2.2 Word Cloud**

**Word Cloud for Reviews (Positive):** This section displays a word cloud, a visual representation of the most frequent words used in positive reviews. The larger a word appears in the cloud, the more often it showed up in the reviews. In this case, some of the prominent words are "great", "phone", "battery", "camera", and "price".

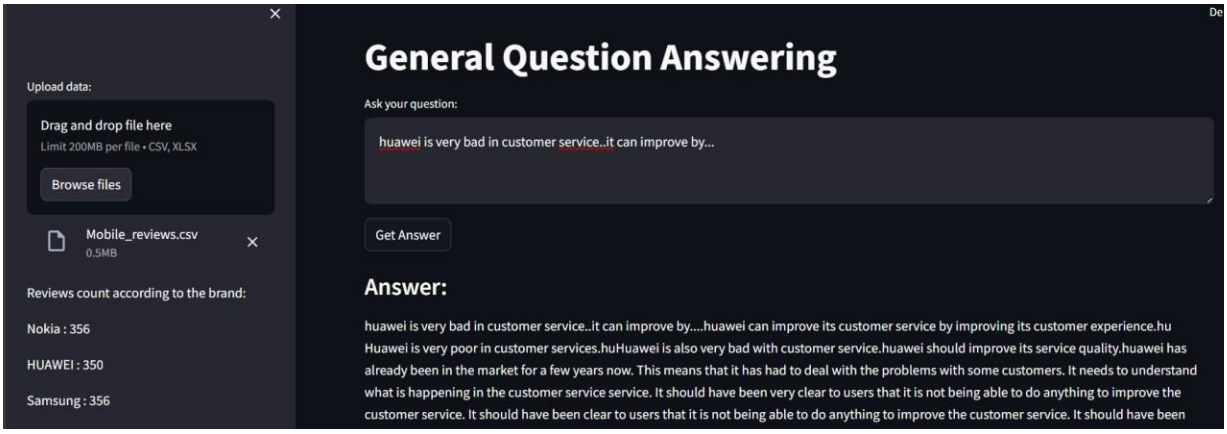


**Fig 5.2.3 Suggestions**

Distilgpt 2 provides response for the asked question . It is displayed in the Answer Section.

# CHAPTER 6
# CONCLUSION AND FUTURE ENHANCEMENTS

## 6.1 CONCLUSION

An important step forward in comprehending and reacting to customer feedback in real-time is the integration of DistilGPT2 for text generation and VADER for sentiment analysis within a single platform. This project shows how to classify reviews, analyze sentiments from textual data, and produce insightful insights for brand reputation management using sophisticated natural language processing (NLP) tools. Through the utilization of DistilGPT2's human-like text generation capability and VADER's lexicon-based approach for precise sentiment scoring, the system offers a comprehensive solution for companies looking to better understand their clientele and tailor their offerings. The data is made easier to understand by the visualizations, which include word clouds for frequently used terms and pie charts for sentiment distribution. This makes it possible to derive actionable insights from the data.

This project underscores the importance of leveraging cutting-edge NLP technologies to transform raw textual data into strategic insights. By automating the sentiment analysis process and providing detailed visual representations, businesses can gain a comprehensive understanding of customer perceptions and sentiments at scale. The system's ability to process and analyze large volumes of data efficiently allows companies to stay attuned to customer feedback in real-time, identify emerging trends, and address potential issues proactively. Furthermore, the integration of DistilGPT2 for generating brand improvement suggestions based on negative reviews adds a proactive dimension to customer relationship management, enabling businesses to take targeted actions to enhance customer satisfaction and loyalty. This holistic approach to sentiment analysis and feedback management demonstrates how technology can be harnessed to foster closer customer relationships, drive product and service enhancements, and ultimately strengthen brand reputation.

## 6.2 FUTURE ENHANCEMENTS

With an eye toward future improvements, there are a number of ways to increase this project's capabilities. First off, by better comprehending the context, adding more NLP models—like BERT (Bidirectional Encoder Representations from Transformers) or RoBERTa (Robustly optimized BERT approach)—could increase the precision and depth of sentiment analysis. Furthermore, by including a multilingual sentiment analysis module, the system would be able to handle reviews and feedback in multiple languages, increasing its marketability in international marketplaces. The creation of a stronger feedback loop, in which the system continuously picks up new information and applies it to improve accuracy and relevance over time, could be another improvement.

**REFERENCES**


[1] S. Burns, *Natural Language Processing: A Quick Introduction to NLP with Python and NLTK*. 2019.

[2] K. Bai and K. H. Tan, “The Influence of Online Social and Physical Presence on User Consumption Decisions in TikTok Livestreaming: A Scoping Review,” *Cyberpsychol. Behav. Soc. Netw.*, May 2024, doi: 10.1089/cyber.2023.0526.

[3] S. Raghavendra, *Beginner’s Guide to Streamlit with Python: Build Web-Based Data and Machine Learning Applications*. 2023.

[4] V. Vought, R. Vought, A. S. Lee, I. Zhou, M. Garneni, and S. A. Greenstein, “Application of sentiment and word frequency analysis of physician review sites to evaluate refractive surgery care,” *Adv Ophthalmol Pract Res*, vol. 4, no. 2, pp. 78–83, Mar. 2024.

[5] F. Millstein, *Natural Language Processing With Python: Natural Language Processing Using NLTK*. Frank Millstein, 2020.

[6] S. Poria, A. Hussain, and E. Cambria, *Multimodal Sentiment Analysis*. 2018.

[7] C. Chen, B. Xu, J.-H. Yang, and M. Liu, “Sentiment Analysis of Animated Film Reviews Using Intelligent Machine Learning,” *Comput. Intell. Neurosci.*, vol. 2022, p. 8517205, Jul. 2022.

[8] S. Bhattacharyya, V. Snasel, A. E. Hassanien, S. Saha, and B. K. Tripathy, *Deep Learning: Research and Applications*. Walter de Gruyter GmbH & Co KG, 2020.

[9] Christodoulakis, Christina, et al. "Pytheas: pattern-based table discovery in CSV files." *Proceedings of the VLDB Endowment* 13.12 (2020): 2075-2089.

[10] I. C. Obagbuwa, S. Danster, and O. C. Chibaya, “Supervised machine learning models for depression sentiment analysis,” *Front Artif Intell*, vol. 6, p. 1230649, Jul. 2023.

[11] M. Falter *et al.*, “Using natural language processing for automated classification of disease and to identify misclassified ICD codes in cardiac disease,” *Eur Heart J Digit Health*, vol. 5, no. 3, pp. 229–234, May 2024.

[12] R. Li, J. Wang, J. Li, and M. Kou, “Long, short, and medium terms wind speed prediction model based on LSTM optimized by improved moth flame

optimization algorithm," *Environ. Sci. Pollut. Res. Int.*, May 2024, doi: 10.1007/s11356-024-33580-8.

[13] S. V. Balshetwar, A. Rs, and D. J. R, "Fake news detection in social media based on sentiment analysis using classifier techniques," *Multimed. Tools Appl.*, pp. 1–31, Mar. 2023.

[14] A. Schroeder, C. Mayer, and A. M. Ward, *The Book of Dash: Build Dashboards with Python and Plotly*. No Starch Press, 2022.

[15] V. S. Dsouza *et al.*, "A sentiment and content analysis of tweets on monkeypox stigma among the LGBTQ+ community: A cue to risk communication plan," *Dialogues Health*, vol. 2, p. 100095, Dec. 2023.