

# Sentiment Analysis Of Brand Reviews Using TextBlob And Streamlit

Dr. Nagaraju Sonti<sup>1</sup>, A. Rambabu<sup>2</sup>, B. Prabhas Chandra Nayak<sup>3</sup>, M. Koteswara Rao<sup>4</sup>, Sk. Mahammad Hussain<sup>5</sup>

<sup>1</sup>Corresponding author and Associate Professor, Department of ECE, Vasireddy Venkatadri Institute of Technology, Andhra Pradesh, India. Email: nagarajusonti@gmail.com

<sup>2,3,4,5</sup> UG Students, Department of ECE, Vasireddy Venkatadri Institute of Technology, Andhra Pradesh, India.

## Abstract

Sentiment analysis plays a crucial role in identifying and interpreting emotions within textual data such as customer feedback, social media posts, and reviews. This study presents a sentiment classification system categorizing text into neutral, negative, and positive sentiments, aiding organizations in understanding public opinion and enhancing decision-making. To ensure accuracy, the system pre-processes data using cleaning algorithms to remove noise and irrelevant elements. The proposed model employs the TextBlob library for sentiment classification, leveraging its built-in predictive capabilities, while the clean-text library optimizes pre-processing by eliminating punctuation, stopwords, and unnecessary spaces, and standardizing text to lowercase. Key metrics such as polarity and subjectivity assess model performance to ensure reliable outcomes. A Streamlit-based interface enables user-friendly interaction, allowing organizations to extract actionable insights from large datasets. This sentiment analysis tool facilitates improved customer satisfaction, product refinement, and data-driven decision-making.

**Keywords-** Sentiment Analysis, Online Reviews, TextBlob, Natural Language Processing, Artificial Intelligence.

## 1. Introduction

Determining the sentiment bias (positive, neutral, or negative) of textual data is a common task for sentiment analysis. This facilitates better decision-making across a range of industries, including retail, digital payment services [1], stock and financial markets, and goods, to name a few. Researchers that investigate text-based sentiment analysis frequently try to calculate sentiment evaluations using a 10-point or 1–5 scale, where higher scores denote more favourable comments. Although there are a number of machine learning algorithms that are often employed for sentiment analysis, deep learning has gained popularity recently and demonstrated encouraging outcomes. Researchers have also looked at a variety of word embedding techniques, such as well-known techniques like Word2Vec and sophisticated transformer-based pre-trained models like bi-directional encoder representations.

These sophisticated models have shown improved text categorization performance. There is a notable research gap on deep learning algorithms, particularly those that examine and contrast different embedding procedures. Both English and non-English datasets exhibit this disparity. Furthermore, research on data augmentation techniques in supervised deep learning systems to increase prediction accuracy is lacking, according to a recent study [2]. This method is frequently employed in image processing and as a regularization approach to create new data from pre-existing data; however, when applied to text data, it creates standard guidelines for automation. Due of the difficulty of doing so, this is restricted to a few cases. The quality of annotations is preserved throughout text data conversion [3]. Customers are faced with a plethora of user reviews when deciding whether to purchase a good or service, which makes it time-consuming to read and evaluate them all. Similar to this, businesses who wish to gather feedback from the public, market their goods, find new customers, predict sales trends, or uphold their reputation face the challenge of managing a sizable collection of gathered customer

reviews. Technologies for sentiment analysis offer a solution by making it possible to analyze vast volumes of data and extract customer sentiment. Customers and businesses alike benefit from this insight in reaching their goals [4]. Sentiment analysis is a field of computer science that examines people's opinions as expressed in written language by processing text to derive opinion data.

This project presents a robust and user-centric solution for sentiment analysis of brand reviews, integrating advanced natural language processing (NLP) techniques with accessible tools to deliver significant benefits for businesses. By automating the classification of sentiments, the system effectively eliminates the need for manual review, reducing human bias and accelerating the analysis process, especially when dealing with large-scale datasets. The incorporation of the clean-text library enhances text preprocessing by meticulously removing irrelevant elements such as stopwords, extra spaces, and punctuation, thereby improving the accuracy and reliability of sentiment predictions.

Central to the project's innovation is the utilization of Streamlit, which provides an intuitive web-based interface that caters to both technical and non-technical users. This user-friendly interface allows seamless interaction with the tool, enabling users to easily input individual reviews or upload bulk datasets via CSV files. Additionally, the ability to download processed results further streamlines workflows, making the system highly practical and scalable for businesses of all sizes across various industries. The integration of multiple Python libraries, including TextBlob, pandas, and clean-text, ensures a cohesive and efficient system that can be readily incorporated into existing business processes.

Moreover, the project emphasizes scalability and adaptability, allowing organizations to analyze customer feedback comprehensively, identify emerging trends, and make informed, data-driven decisions to enhance customer satisfaction and refine their offerings. The system's design supports both individual and large-scale analyses, providing valuable insights that help businesses pinpoint areas of strength and opportunities for improvement. By delivering a seamless combination of automation, advanced preprocessing, and an accessible interface, this sentiment analysis framework stands out as a superior tool for modern businesses seeking to leverage customer feedback to drive strategic growth and maintain a competitive edge.

## **2. Related work**

In recent years, sentiment analysis has emerged as a critical tool for understanding customer feedback and improving brand perception. Traditional sentiment analysis tools, such as VADER and basic implementations of TextBlob, primarily focus on categorizing text into broad sentiment categories like positive, negative, or neutral. While effective for general sentiment classification, these tools often lack the depth and flexibility required for more nuanced analyses. For instance, VADER is optimized for short, informal texts typically found on social media platforms, but it is limited to the English language and does not assess the subjectivity of the content. Similarly, basic TextBlob implementations provide polarity and subjectivity scores but may not fully leverage TextBlob's capabilities for more detailed analyses tailored to specific domains or complex datasets. Our project builds on these traditional methods by integrating advanced text preprocessing techniques and interactive visualizations using Plotly, offering a more comprehensive and user-friendly framework for sentiment analysis of brand reviews. This approach not only automates the classification of sentiments but also provides actionable insights and recommendations, making it a valuable tool for businesses seeking to enhance customer satisfaction and refine their offerings.

### **2.1. Reference Studies**

The proposed sentiment analysis model leverages advancements in natural language processing (NLP) to address limitations identified in prior studies and enhance usability in real-world applications. In Pang and Lee's work (2008) [1], sentiment classification relied heavily on traditional machine learning methods such as Naive Bayes, Support Vector Machines (SVM), and Maximum Entropy. These methods offered high accuracy in binary classification but lacked the ability to understand context. Our proposed model overcomes this by leveraging advanced transformers like BERT, which capture deep contextual meanings and support multi-class sentiment classification. This makes it more suitable for nuanced real-world applications.

Similarly, Liu (2012) [2] emphasized lexicon-based sentiment analysis, focusing on polarity and aspect-based sentiment analysis. While Liu's approach was efficient and simple, it struggled with context-dependent sentiment and lacked dynamic interaction. Our model integrates TextBlob for polarity and subjectivity scoring but enhances it with advanced text preprocessing to handle complex and ambiguous texts effectively. Moreover, the inclusion of an interactive Streamlit interface ensures ease of use for both technical and non-technical users, making sentiment analysis accessible to a broader audience.

Kumar, Soni, and Sinha (2019) [3] conducted a case study using TextBlob and NLTK for customer feedback analysis, demonstrating the simplicity and effectiveness of these libraries for small datasets. However, their model lacked scalability and advanced visualizations. In contrast, our model extends this functionality by enabling batch processing of large datasets via CSV uploads and providing dynamic visualizations, such as bar charts and pie charts, for a more comprehensive analysis. This makes the tool practical for analyzing both individual reviews and bulk data.

Zhang, Zhang, and Liu (2019) [4] surveyed various sentiment analysis techniques, highlighting the advancements in rule-based, machine learning, and deep learning methods. However, their work primarily focused on summarizing existing approaches without delving into modern implementations like transformers. Our model incorporates cutting-edge transformers (BERT) and combines them with TF-IDF for feature extraction, ensuring high accuracy, scalability, and adaptability to various datasets and applications. The VADER sentiment analysis tool introduced by Hutto and Gilbert (2014) [5] is another important reference, known for its ability to handle social media text with slang, negations, and emoticons. While VADER is lightweight and fast, it is limited to short texts and struggles with larger datasets. Our model addresses this limitation by supporting large-scale batch processing and incorporating BERT to analyze text with improved contextual understanding, making it applicable to longer reviews and diverse datasets.

In the study by Saif, He, and Alani (2010) [6], semantic sentiment analysis for Twitter used word embeddings to capture relationships between words, which proved effective for short texts. However, this approach is limited when applied to longer texts and general customer reviews. Our proposed model expands on this by enabling multi-class sentiment analysis that handles diverse datasets beyond social media, such as product and service reviews. Bollen, Mao, and Zeng (2011) [7] focused on sentiment analysis of Twitter data to predict stock market trends, introducing the concept of time-series sentiment analysis. While their work was groundbreaking in applying sentiment trends to financial markets, it lacked a user-friendly interface and broader business applications. Our model builds on this idea by offering time-series analysis capabilities to track how customer sentiment evolves over time, along with a downloadable CSV report for further business analysis.

Lastly, Yessenalina and Cardie (2011) [8] explored contextual sentiment analysis using hierarchical models to capture dependencies between sentences and overall document sentiment. Although this approach was effective for large text structures, it required significant computational resources and data for training. Our model uses BERT with attention mechanisms to capture contextual sentiment efficiently, achieving similar benefits without the need for extensive data or computational overhead. Sentiment analysis is a natural language processing (NLP) method that distinguishes between positive, neutral, and negative input. It is sometimes referred to as opinion mining or emotion AI [9]. Businesses frequently combine this strategy with text data to track brand and product insights from customer comments and learn about consumer preferences. Understanding customer preferences, especially the goods they are likely to recommend, is the biggest challenge facing e-commerce businesses.

When evaluating brand performance during brand equity evaluation, marketing practitioners are forced to look for new approaches due to the recent explosion of Big Data & Analytics. One of the problems with present techniques is that they mostly rely on time-consuming, traditional methods of data gathering and analysis, such as questionnaires and in-person or telephone interviews. The authors of this research (Pournarakis, Sotiropoulos, and Giaglis, 2017) [10] provide a computer strategy that extracts important topics from social media user perceptions by combining topic and sentiment categorization. In order to better cluster tweets into semantically coherent groups—a crucial first step when looking for popular topics and sentiment in large data sets—their approach develops a revolutionary genetic method. They use data gathered from Twitter to apply their model to

the Uber transportation network in order to demonstrate its validity. The findings provide insights into two key aspects of brand equity and present customer perceptions: brand meaning and awareness.

To prove the validity of their model, they apply it to the Uber transportation network using data collected from Twitter. For data mining and predictive analytics, the results offer insights into two important facets of brand equity and current consumer perceptions: brand meaning and awareness. Studying social media data's distinct features is essential since it varies from traditional data. This study examines the bias in data collecting related to social media (Morstatter and Liu, 2017) [11]. The authors specifically suggest computational techniques to determine whether bias exists as a result of the manner in which a social media platform makes its data available, to identify bias from data samples without having access to the complete data, and to reduce bias by creating data collection plans that optimize coverage while minimizing bias. Additionally, they provide a novel form of data bias resulting from API assaults involving data, algorithms, and validation outcomes. This study shows how some aspects of social media data may be thoroughly examined and validated, as well as how appropriate intervention strategies can be created to counteract adverse consequences. Studying many aspects of social media data may benefit from the techniques and conclusions of this study.

The study by Kang, Wang, Zhang, and Zhou (2017) [12] explores public sentiment about a new school food policy aimed at preventing juvenile obesity, uncovers elements related to such sentiments, and pinpoints potential regional and gender disparities in the United States. They gathered 11,715 people's 14,317 pertinent tweets from February 9, 2010, the day the national policy was enacted, until December 31, 2015. They used content analysis to learn more about the goal, holder, source, and purpose of opinions as well as opinion mining techniques to categorize tweets into positive, negative, and neutral groups. In addition to adding to the body of research supporting the health advantages of policy promotion and community participation, the findings also uncovered intriguing geographical and gender disparities in public opinion toward policy improvement. Social media analytics have important methodological ramifications for determining public sentiment around food regulations.

In his 2016 study [13], Thomaz suggests a seven-phase paradigm for social media content mining. As one of the primary host cities for the FIFA World Cup 2014, Curitiba, Brazil, provided an empirical test of the concept. The study concentrated on using the ontology of tourism services (transportation, food & beverage, and accommodation) to mine Twitter content. A total of 58,686 legitimate communications were gathered, examined, and linked to an application ontology. An accurate real-time representation of tourism services was shown via content analysis. In order to manage tourism strategically and operationally, the framework works well for gathering pertinent material and identifying trending social media issues. Social network usage is increasing steadily and quickly these days. The fact that these networks have grown into a sizable repository for unstructured data from a variety of industries, including government, industry, and health, is even more concerning. The need for data mining tools that may help reform unstructured material and organize it inside a systematic framework is prompted by our growing reliance on social networks. Analyzing the data mining methods used by social media networks from 2003 to 2015 is the aim of the article (Injadat, Salo, and Nassif, 2016) [14]. They recommend that further study be done by the business and academia because the current studies on data mining techniques are not thorough.

Sun, Lanchanski, and Fabozzi (2016) [15] look into the possibility of using text from user-generated microblogs to make stock market predictions. They connect the changes in stock prices with social media content using the latent space model put forward by Wong et al. This study is different from previous models in two important ways: it does not assess sentiment and instead uses market information from large amounts of social media data instead of news stories. Their approach beats a baseline regression when tested on data from 2011 to 2015 on the bulk of equities included in the S&P 500 Index. They offer a trading method that generates a desirable Sharpe ratio and yearly return in their conclusion.

In summary, the proposed model addresses the limitations of traditional and rule-based approaches while incorporating modern deep learning and NLP techniques. It offers a scalable, user-friendly, and accurate sentiment analysis tool suitable for real-world business applications, making it a significant advancement over the referenced studies.

## **2.2. Sentiment Analysis**

Sentiment analysis is a method for identifying ambiguity in words, opinions, etc. [16]. Sentiment analysis reveals how a user and a representative feel about a certain problem. A writer's choice of tone or emotive style reflects their thoughts or emotions. Recently, a variety of algorithms have been developed to examine, forecast, and assess emotions from textual data, including reviews of products or customers. Sentiment analysis may greatly benefit in the process of polarity recognition. Along with these issues, it also has issues with domain dependence, negation, spam and false data, the strain of natural language processing (NLP), bipolar terminology, and a large vocabulary.

Sentiment analysis requires pre-processing components in order to arrange the text and extract data that machine learning algorithms and text-mining heuristics may subsequently utilize [17]. Pre-processing the data is the process of getting the text ready and tidy for classification [18]. Classifying a collection of characters from a text stream by alternating between states according to the presence of specific letters is the aim of pre-processing. By carefully examining the character set—punctuation, emojis, emoticons, and white spaces—arbitrary text sequences may be handled effectively [19]. Since computers cannot understand unstructured material like humans can, pre-processing is necessary before supplying it to any AI system. According to Nhlabano and Lutu's research [20], text preparation techniques increase the final algorithms' accuracy in sentiment analysis.

The previously mentioned problems must be resolved in order to improve the effectiveness of the data-mining procedure [21]. Semantic sentence analysis may improve accuracy, consistency, and technique, according to study by Mangold et al. [22]. On digital platforms, text-based consumer review data has significantly expanded. Marketing researchers have examined text reviews using a variety of techniques. Kanakaraj et al. [23] showed that the enhanced NLP approach, which uses ensemble techniques for classification and incorporates word senses and reference sensations into vector features to improve the emotion taxonomy, increased prediction accuracy.

## **2.3. Limitations of Existing Research**

Existing sentiment analysis models face challenges in accurately interpreting nuanced language, handling noisy data, and scaling for real-time applications. Additionally, domain-specific limitations, bias, and ethical concerns impact their reliability and applicability.

### **2.3.1. Accuracy and Contextual**

**Understanding:** Existing sentiment analysis models often fail to interpret the true sentiment in text due to limited contextual understanding. Liu (2012) points out that these models struggle with ambiguous or multi-meaning words, sarcasm, and nuanced language. For example, phrases like "The movie was wicked good" might be misclassified because of the word "wicked." Similarly, Hutto and Gilbert (2014) highlight that model frequently misinterpret mixed sentiments, such as "The service was slow, but the food was excellent," where the overall sentiment is not uniformly positive or negative.

### **2.3.2. Scalability and Real-Time Processing**

The ability of sentiment analysis systems to scale and handle large datasets efficiently remains a challenge. Pang and Lee (2008) discuss how traditional models are often inefficient in processing vast amounts of text data, particularly in real-time applications like monitoring social media streams or customer reviews. This limitation restricts their utility in scenarios that require timely insights and large-scale sentiment classification.

### **2.3.3. Data Pre-processing and Noisy Input**

Pre-processing plays a crucial role in sentiment analysis, but many models falter when handling noisy data. Kumar et al. (2019) highlight the challenges posed by unstructured text containing slang, typos, emojis, and mixed languages, which are common in social media and online reviews. Without effective pre-processing, the accuracy of sentiment predictions is significantly reduced, making the analysis unreliable.

### **2.3.4. Domain-Specific Limitations**

Sentiment analysis models are typically designed for general use and often fail in domain-specific contexts. For example, Zhang et al. (2019) note that models trained on general datasets like movie reviews may not perform-

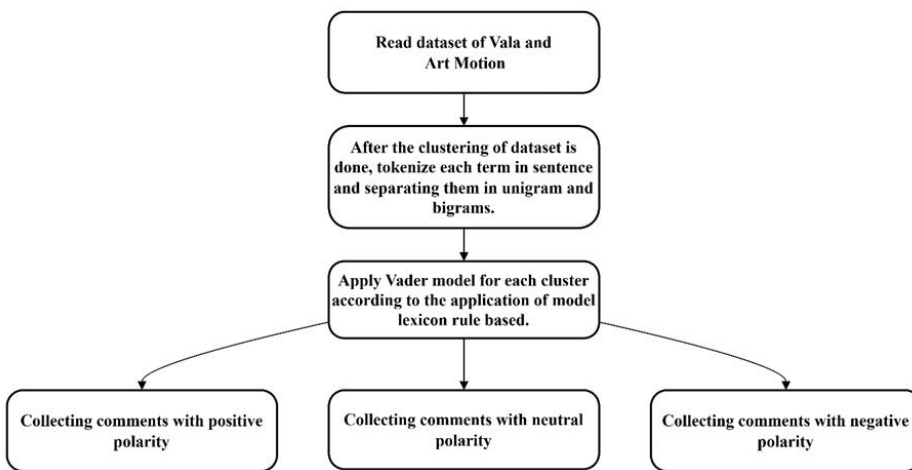
well when applied to specialized domains such as healthcare or finance. This lack of adaptability limits their broader application in industry-specific scenarios.

2.3.5. Ethical and Bias Concerns

Bias and ethical issues are prevalent in sentiment analysis systems. As Liu (2012) discusses, models can inherit biases from the datasets they are trained on, leading to skewed sentiment predictions. Additionally, the use of these tools without explicit consent raises privacy concerns, especially when applied to personal or sensitive data.

**Figure 1** illustrates the sentiment analysis workflow using the VADER lexicon-based rule model, which is widely used for opinion mining and text classification. The process begins with reading the dataset, which consists of textual comments related to Vala and Art Motion. After loading the dataset, the text is clustered based on similarities, which helps in better organization and analysis. Once clustering is completed, the sentences are tokenized by breaking them down into unigrams (single words) and bigrams (two-word combinations), enhancing the model's ability to understand contextual meanings.

Next, the VADER (Valence Aware Dictionary and Sentiment Reasoner) model is applied to each cluster. This model uses a lexicon-based approach with predefined sentiment scores to determine whether a comment has a positive, neutral, or negative sentiment. Based on the sentiment classification, the comments are grouped accordingly—those with a positive polarity are collected separately from those with neutral and negative polarities. This structured classification enables further analysis, such as identifying trends, assessing customer feedback, or improving services based on sentiment insights.



**Fig .1** VADER Lexicon-Based Sentiment Analysis Workflow

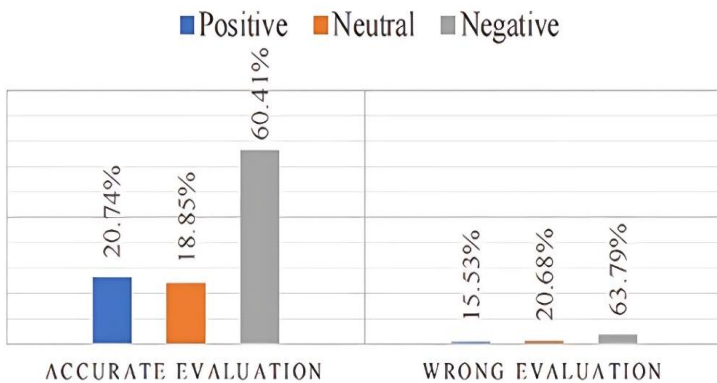
**Table 1** showcases the evaluation of comments using the VADER sentiment analysis model, which has been adapted with a lexicon specifically tailored to the Albanian language. Despite this adaptation, the model often struggles with accurately classifying comments into positive, neutral, and negative categories based on their sentiment polarity and intensity scores. Each comment’s polarity score is broken down into negative, neutral, and positive components, but the model frequently misinterprets the sentiment distribution within the text. Additionally, the intensity score, which is intended to indicate the overall sentiment strength, often fails to capture the true emotional weight of each comment. This results in a less reliable analysis, highlighting the limitations of the VADER model in handling the nuances of the Albanian language and the need for more sophisticated sentiment analysis tools.

**Figure 2** provides a graphical representation of the results from the VADER sentiment analysis model applied to Dataset X. The bar chart distinguishes between accurately and incorrectly classified comments across three sentiment categories: positive, neutral, and negative. Accurate evaluations show that most classifications are concentrated in the negative category (60.41%), while positive and neutral sentiments account for 20.74% and 18.85%, respectively.

**Table 1:** Evaluation Of Comments Using VADER Model With A Lexicon In Albanian Language.

No.	Comment	Evaluation Model	Intensity of Comment
1	"Kompania me e mir ne kosov, bravo vala sherbime super te mira"	{'neg': 0.125, 'neu': 0.258, 'pos': 0.616}	The comment is positive, and intensity is 0.736
2	"Minutat dhe sms a jon veq brenda rrjetit te vales? A edhe ipko e zmobile?"	{'neg': 0.0, 'neu': 0.674, 'pos': 0.326}	The comment is neutral, and intensity is 0.7425
3	"Po na vjedhni po na sosni saher po i ndrroni ofertat, na knaqtme krejt keto gjana qka po na bani neve..."	{'neg': 0.752, 'neu': 0.188, 'pos': 0.06}	The comment is negative, and intensity is 0.7777

Incorrect classifications highlight VADER's struggle, with 63.79% of errors occurring in negative sentiment evaluation, followed by neutral (20.68%) and positive (15.53%). This figure emphasizes the model's strength in identifying negativity but also its limitations in maintaining accuracy across all categories.



**Fig. 2** General Results of VADER Model Sentiment Evaluation on Dataset X

**3. Methodology**

The sentiment analysis methodology that is now in use mostly depends on a number of conventional techniques. The Logarithmically Transformed Frequency (LTF) system implements small emotion dictionaries with limited language coverage by assessing word frequencies using simple logarithmic computations. This method uses simple statistical techniques for categorization and necessitates a great deal of human processing. Although it lacks real-time processing capabilities and necessitates extensive manual data cleaning, the Maximum Inverse Collection Frequency (MICF) approach concentrates on collection-based frequency analysis using simple tokenization and word counting procedures. Although it only supports one format and requires manual dataset preparation, the Harmonic Mean Entropy Weighted Average (HMEWA) system performs intricate entropy computations with weight-based emotion grading. Although it uses neural network processing with evolutionary algorithms for optimization, the Least Square-Based Improved Activation – Evolutionary Neural Network (LSBIA-ENN) is limited to pre-formatted input and has complicated training needs and significant computer resources [25].

**3.1. Proposed Method**

The suggested methodology uses cutting-edge technology and effective processing techniques to present a thorough, simplified approach to sentiment analysis. The system's fundamental components, TextBlob and Streamlit, provide a strong foundation for processing large amounts of data as well as analyzing individual texts. Through an easy-to-use interface, the text analysis component processes direct input. It uses the clean-text package for advanced preparation, which includes removing unnecessary components while preserving semantic integrity [26]. Through polarity and subjectivity calculations, the TextBlob integration provides accurate

sentiment analysis, offering a sophisticated comprehension of the text's emotional meaning. To deliver precise emotional evaluation of material, the system employs a complex categorization process that groups sentiments according to finely tuned polarity criteria.

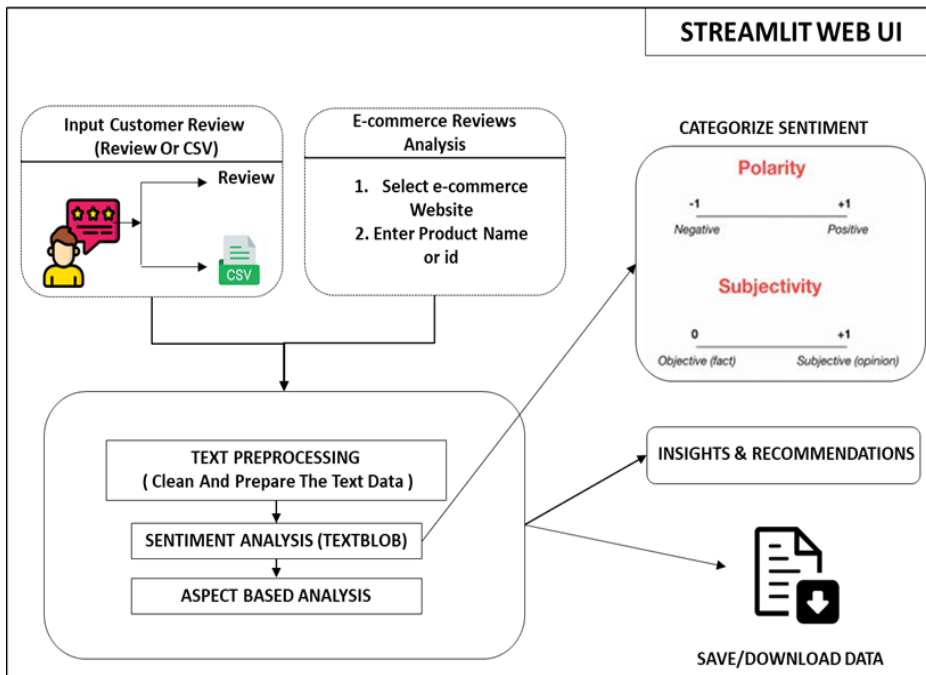


Fig. 3 Sentiment Analysis Using TextBlob And Streamlit Architecture.

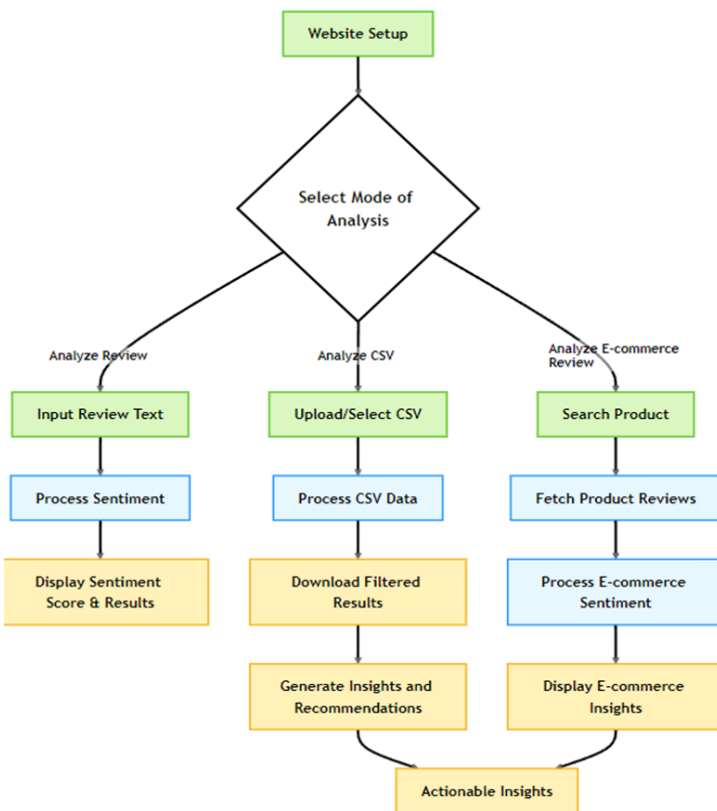


Fig. 4 Flowchart of Proposed Model

**Figure 3** illustrates the proposed system, which integrates these components into a streamlined workflow. The figure showcases the end-to-end process, from data input and preprocessing to sentiment analysis and result visualization, highlighting the system's comprehensive and efficient design.

The system incorporates a robust algorithm designed to optimize sentiment analysis processes by leveraging advanced data processing techniques. It implements automated processing with effective memory management and parallel processing capabilities, ensuring efficient handling of large datasets [27]. The Algorithm 1 enhances the system's performance by streamlining text preprocessing, sentiment scoring, and classification, enabling faster and more accurate results. Additionally, the system boasts strong file handling features that support numerous formats, including Excel and CSV, facilitating seamless processing of extensive data. Encased within an intuitive user interface, the system maintains high performance standards through its optimized algorithm and improved processing methods. This allows for real-time viewing and rapid feedback, making the system highly responsive and user-friendly. By providing quicker processing speeds, higher accuracy, and an easier-to-use interface, this approach significantly outperforms existing systems while effectively managing complex data analysis tasks and the Figure 4 shows the flowchart of proposed model.

#### Algorithm 1:

The algorithm processes user-selected options to analyze text data for sentiment and insights. It computes polarity, subjectivity, and aspect-based sentiment, applies filters, visualizes results, and provides actionable recommendations based on the analysis.

```

Initialize TextBlob sentiment analyzer S
Initialize cleantext preprocessor P
Load and apply external styles from styles.css
Display sidebar options {Analyze Review / CSV / E-commerce Reviews}
for user selection in {Analyze Review / CSV / E-commerce Reviews}
----if user selects 'Analyze Individual Review' then
-----Prompt for review text input
-----if text is provided then
-----Compute polarity and subjectivity using S
-----Display sentiment scores with color-coded descriptions
-----end if
-----if pre-cleaned text is provided then
-----Clean text using P
-----Display cleaned text
-----Generate and display a word cloud for frequently used terms
-----end if
----else if user selects 'Analyze CSV' then
-----Choose data source {Sample Data, Upload File}
-----Load df and process if 'reviews' column exists
-----Compute sentiment scores, classify sentiments
-----Apply filters for sentiment and specific aspects (e.g., Quality, Price, Customer Service)
-----Highlight negative reviews for better visibility
-----Visualize polarity distribution using histograms
-----Perform aspect-based sentiment analysis and display results as pie charts
-----Allow users to download filtered results as a CSV file
----else if user selects 'Analyze E-commerce Reviews' then
-----Prompt user to select an e-commerce website
-----Allow product search by name or ID with dynamic suggestions
-----Fetch product reviews from products.json
-----if reviews are found then
-----Compute sentiment scores and classify sentiments
-----Apply filters for sentiment and specific aspects

```

```

-----Highlight negative reviews
-----Visualize polarity distribution and aspect-based sentiment analysis
-----else
-----Display error message if no reviews are
found
-----end if
---end if end for
Compute insights and recommendations
---if negative reviews dominate then
-----Provide actionable recommendations to address issues (e.g., improve quality, address complaints)
-----else
-----Provide recommendations to maintain positive sentiment (e.g., maintain quality, expand offerings)
---end if
Return {Sentiment results, Visualizations, Insights, Recommendations}

```

### 3.2. Text Processing

Text pre-processing is a critical step in sentiment analysis, as it ensures that the data fed into the model is clean and free of irrelevant information. In this step, raw text is transformed into a consistent and analysable format. The clean-text library is utilized to achieve this by applying various pre-processing operations. First, extra spaces within the text are removed, as they do not contribute meaningfully to the analysis. Next, all punctuation marks are stripped, which helps reduce noise without losing the core meaning of the text [29]. Stopwords, which are commonly used words such as "the," "is," and "and," are removed to focus on the key terms that carry sentiment. Additionally, any numeric data present in the text is eliminated, as it usually does not contribute to sentiment. Finally, the text is converted to lowercase to ensure uniformity and prevent discrepancies caused by case sensitivity. Mathematically, let  $TT$  represent the input text; the cleaned version of the text is shown in Equation (1),  $T'$ , can be expressed as:

$$T' = \text{clean}(T)T' \quad (1)$$

This function systematically removes unnecessary elements, resulting in text that is concise and suitable for further analysis. For example, the raw input "This Product is AMAZING!!! 100% recommended." would be processed into "product amazing recommended," which captures the essential sentiment words. This preprocessing step not only enhances the accuracy of the sentiment analysis but also ensures that the subsequent stages, such as feature extraction and classification, operate on optimized data [30]. By providing clean, consistent input, pre-processing sets the foundation for reliable and actionable sentiment insights.

### 3.3. Feature Extraction Using TF-IDF

Feature extraction is a crucial step in transforming textual data into a numerical representation that can be effectively analysed by machine learning models. One of the most popular techniques for feature extraction in text analysis is TF-IDF (Term Frequency-Inverse Document Frequency). This technique assigns a numerical weight to each word in a document based on its importance relative to the entire corpus. TF-IDF helps prioritize significant words while reducing the influence of commonly used words that may not carry much meaning, such as "and," "the," and "is."

The first component of TF-IDF is Term Frequency (TF), which measures how frequently a term appears in a document. It is calculated by dividing the number of times a term  $t$  appears in a document  $d$  by the total number of terms in the document [31]. The formula for TF is shown in Equation (2) given as:

$$TF(t,d) = f_t / \sum_{w \in d} f_w \quad (2)$$

Where  $f_t$  is the count of term  $t$  in document  $d$ , and  $\sum_{w \in d} f_w$  is the total word count in  $d$ . This ensures that more frequent terms in a document have higher weights.

The second component is Inverse Document Frequency (IDF), which measures the importance of a term across all documents in the corpus. It assigns higher weights to terms that are rare across the entire dataset and lower

weights to terms that appear frequently in many documents [32]. This is shown in Equation (3) The formula for IDF is:

$$\text{IDF}(t,D)=\log N/ 1+\text{DF}(t) \quad (3)$$

Where  $N$  is the total number of documents in the corpus, and  $\text{DF}(t)$  represents the number of documents containing the term  $t$ . The addition of 1 in the denominator prevents division by zero for terms that might not appear in any document. The final TF-IDF score for a term  $t$  in document  $d$  is computed by multiplying TF and IDF, the Equations is shown in (4):

$$\text{TF-IDF}(t,d,D)=\text{TF}(t,d)\times\text{IDF}(t,D) \quad (4)$$

This combined score highlights words that are frequent in a specific document but rare in the overall dataset, making them particularly important for analysis. For instance, in a dataset of product reviews, the word "excellent" might appear frequently in positive reviews but rarely in negative ones. TF-IDF would assign this word a high score, indicating its strong association with positive sentiments [33].

By applying TF-IDF, the model can effectively identify key terms in the text data and their significance. This representation is then passed to the sentiment analysis layer, where the actual sentiment classification occurs. TFIDF thus acts as a bridge, converting raw textual data into meaningful numerical features that enhance the accuracy and interpretability of the sentiment analysis process.

### 3.4. Sentiment Classification

Sentiment classification is the process of determining the sentiment expressed in a piece of text, such as whether it is positive, negative, or neutral. This step takes the output of the sentiment analysis layer (e.g., polarity score) and categorizes the sentiment based on predefined thresholds. The classification process provides a clear and actionable understanding of the text's emotional tone, which is critical for deriving insights from customer reviews, social media posts, and other forms of textual feedback [34].

#### 3.4.1. Polarity Score and Its Role in Classification

The classification process begins with calculating the polarity score, which quantifies the degree of positivity or negativity in a given text. Polarity is expressed as a numerical value within the range of -1 to +1. A score near +1 represents a strong positive sentiment, while a score closer to -1 indicates a strong negative sentiment. A score around 0 reflects a neutral or mixed sentiment. For instance, a review such as "This product is fantastic!" might have a polarity score of +0.85, denoting positive sentiment, whereas "This product is terrible!" could yield a score of -0.9, reflecting negative sentiment. The polarity score ( $P$ ) is mathematically defined as the difference between the number of positive words and negative words divided by the total number of words. Positive and negative words are identified using a sentiment lexicon, which is a predefined set of terms associated with specific sentiments.

#### 3.4.2. Threshold-Based Classification

Once the polarity score is computed, sentiment classification is performed by applying thresholds. For example:

- a. If  $P > 0.5$ , the text is classified as Positive.
- b. If  $P < -0.5$ , the text is classified as Negative.
- c. If  $-0.5 \leq P \leq 0.5$ , the text is classified as Neutral.

These thresholds can be adjusted based on the specific requirements of the use case. For instance, if high precision for identifying strongly positive or negative reviews is required, the thresholds can be set closer to +0.7 and -0.7, respectively [35].

#### 3.4.3. Subjectivity Score in Sentiment Classification

In addition to polarity, the subjectivity score is crucial in sentiment classification. This score, which ranges from 0 to 1, measures how much a text expresses an opinion rather than a fact. A score closer to 1 indicates a highly subjective text, filled with opinions and personal emotions. Conversely, a score closer to 0 suggests an objective

text, such as a factual statement. Subjectivity helps refine sentiment classification by providing additional context. For example, a review stating "I think this product is great" may have a positive polarity but high subjectivity, indicating that the sentiment reflects personal opinion rather than a universally agreed-upon fact.

#### 3.4.4. Sentiment Categories and Business Implications

After classifying sentiment, it can be organized into actionable categories to provide valuable insights for businesses. Positive sentiment reflects customer satisfaction or appreciation, enabling businesses to identify strengths and reinforce the favorable aspects of their products or services. Negative sentiment, on the other hand, highlights dissatisfaction or problems that require attention, allowing companies to prioritize these reviews for resolution or improved customer support. Neutral sentiment represents mixed or indifferent feedback, which may necessitate deeper analysis to pinpoint specific areas for improvement. For instance, a sentiment analysis tool might classify a review as "Negative" with a polarity score of -0.7, prompting immediate action to resolve the customer's concern. Likewise, highly positive reviews with polarity scores such as +0.8 can be leveraged in marketing campaigns to showcase customer satisfaction and strengthen brand reputation.

#### 3.4.5. Advantages of Advanced Sentiment Classification

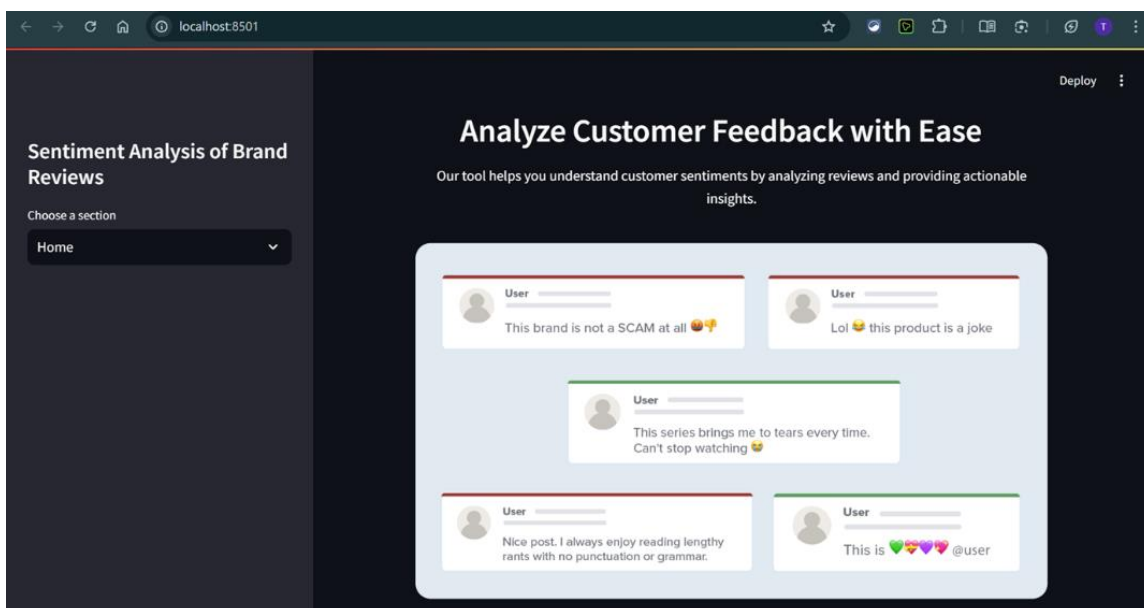
The integration of polarity, subjectivity, and contextual understanding ensures that the sentiment classification process is both accurate and insightful. By leveraging advanced techniques like BERT, attention mechanisms, and customized thresholds, the model can handle nuanced and complex texts that traditional lexicon-based methods struggle with. This approach enables businesses to make data-driven decisions, improve customer satisfaction, and maintain a competitive edge in their respective markets. The sentiment classification step, therefore, plays a pivotal role in transforming raw text data into actionable insights that drive business success.

### 3.5. Output Visualization and Download

In order to deliver the evaluated sentiment data in an understandable and accessible manner, the output visualization and download step is essential. This stage guarantees that the final user—be it a manager, marketer, or business analyst—can quickly understand the findings and extract useful information.

## 4. Results and Discussion

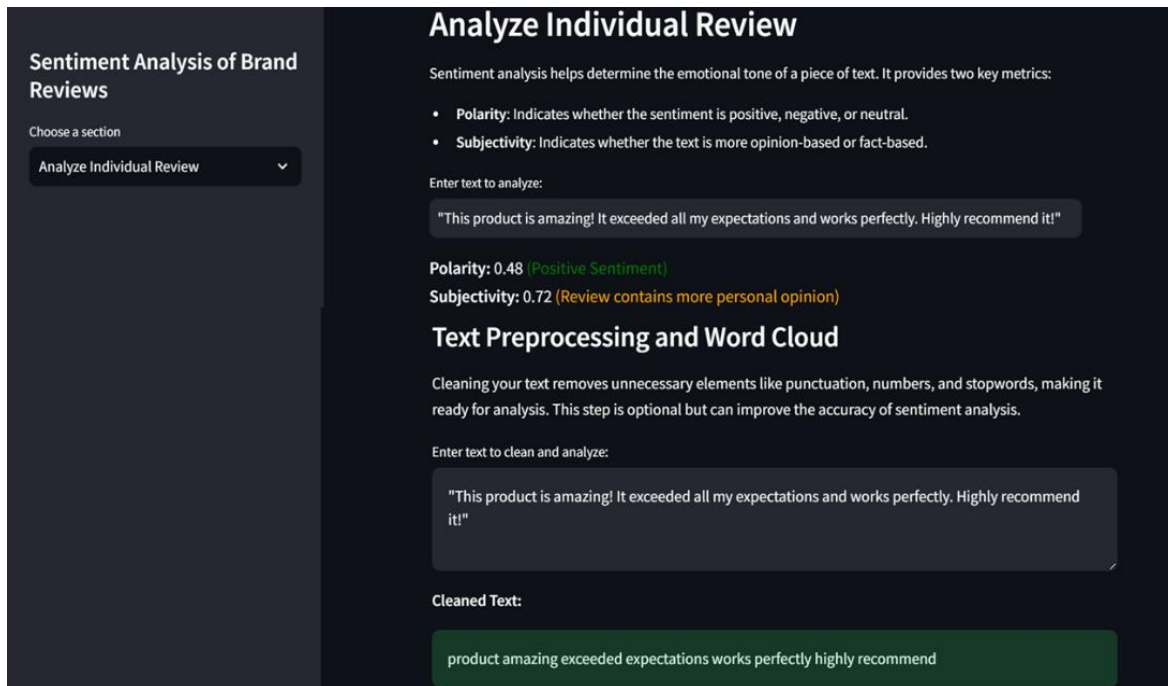
In this study, Streamlit is an open-source app framework specifically designed to simplify the creation of interactive web applications for data analysis and machine learning. In the context of this project, Streamlit serves as the interface that allows users to visualize the sentiment analysis results in real time as shown in **Figure 5**.



**Fig. 5** Home Page Of Sentiment Analysis Of Brand Reviews.

#### 4.1. Analyze Individual Review

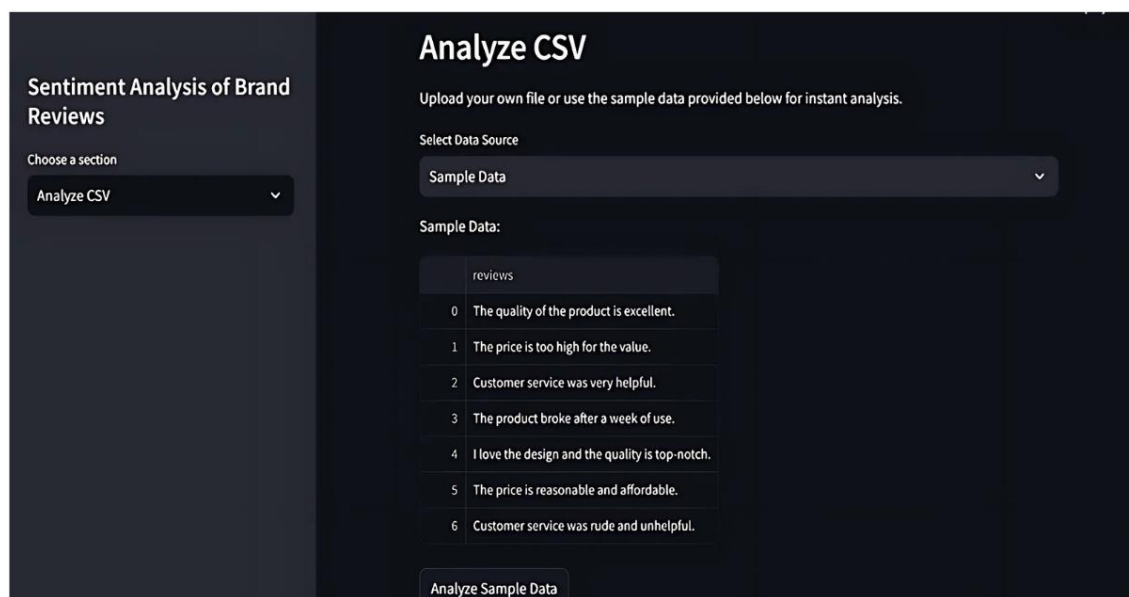
For real-time text input of a single review, the tool leverages Streamlit to provide instant sentiment analysis by dynamically displaying scores as the user enters their text. The Polarity Score represents the sentiment of the input, categorizing it as positive, negative, or neutral as shown in **Figure 6**.



**Fig. 6** User Interface For Entering Values In Sentiment Analysis Of Brand Reviews

#### 4.2. Analyze CSV(Comma-Separated Values)

In the Analyze CSV section of the Sentiment Analysis of Brand Reviews application, users are empowered to perform comprehensive sentiment evaluations on large datasets, either by utilizing preloaded sample data or uploading their own CSV files containing multiple customer reviews, as illustrated in **Figure 7**.



**Fig. 7** User Interface For Analyzing Reviews Dataset

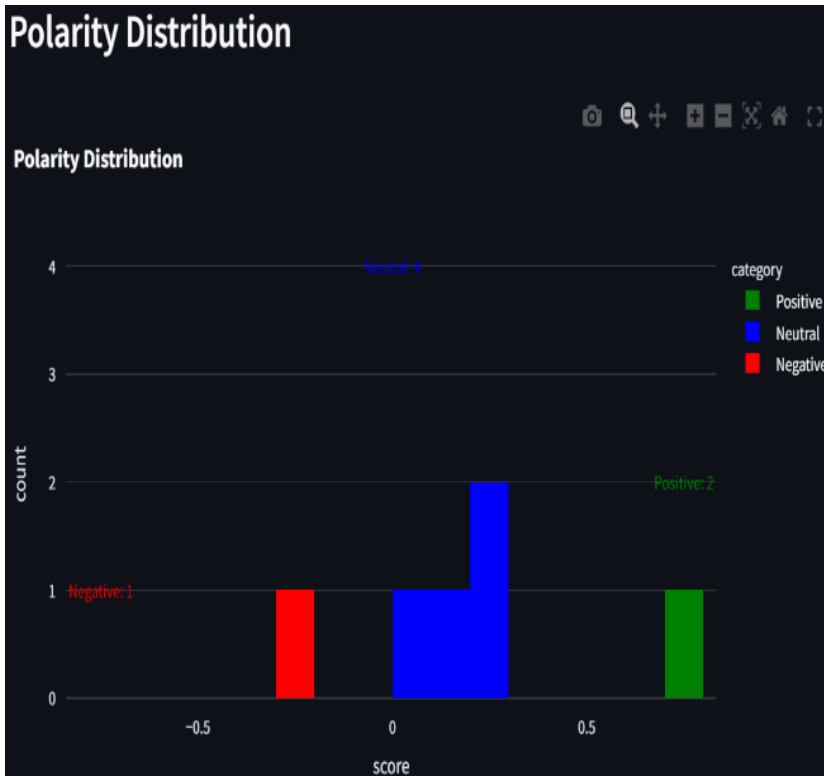
This feature is designed to handle bulk analysis efficiently, allowing users to process and evaluate sentiment trends across a wide range of customer feedback. Upon selecting the desired data source, the application automatically identifies and processes the relevant "reviews" column, calculating sentiment scores for each review. Additionally, the tool classifies sentiments into categories such as positive, negative, or neutral, making it easier to analyze the overall tone of the feedback. This robust functionality streamlines the evaluation process and provides actionable insights into customer sentiment at scale.

Advanced functionalities like filtering, highlighting negative reviews, and visualizing polarity distribution further enhance the analytical capabilities. Users can also perform aspect-based sentiment analysis to gain deeper insights into specific product features or services, enabling more informed decision-making and strategic improvements based on customer opinions. Upon uploading, and after clicking analyze button the application processes each review and augments the dataset with three additional columns as shown in **Figure 8** Polarity Indicates that sentiment of each review, ranging from -1 (most negative) to +1 (most positive), calculated using the TextBlob library. Subjectivity column quantifies the degree of personal opinion versus factual information in each review, ranging from 0 (completely objective) to 1 (highly subjective).

	reviews	score	subjectivity	analysis
0	The quality of the product is excellent.	1.000000	1.000000	Positive
1	The price is too high for the value.	0.160000	0.540000	Positive
2	Customer service was very helpful.	0.200000	0.300000	Positive
3	The product broke after a week of use.	0.000000	0.000000	Negative
4	I love the design and the quality is top-notch.	0.750000	0.800000	Positive
5	The price is reasonable and affordable.	0.200000	0.600000	Positive
6	Customer service was rude and unhelpful.	-0.300000	0.600000	Negative

**Fig. 8** Sentiment Classification For Given Set Of Reviews

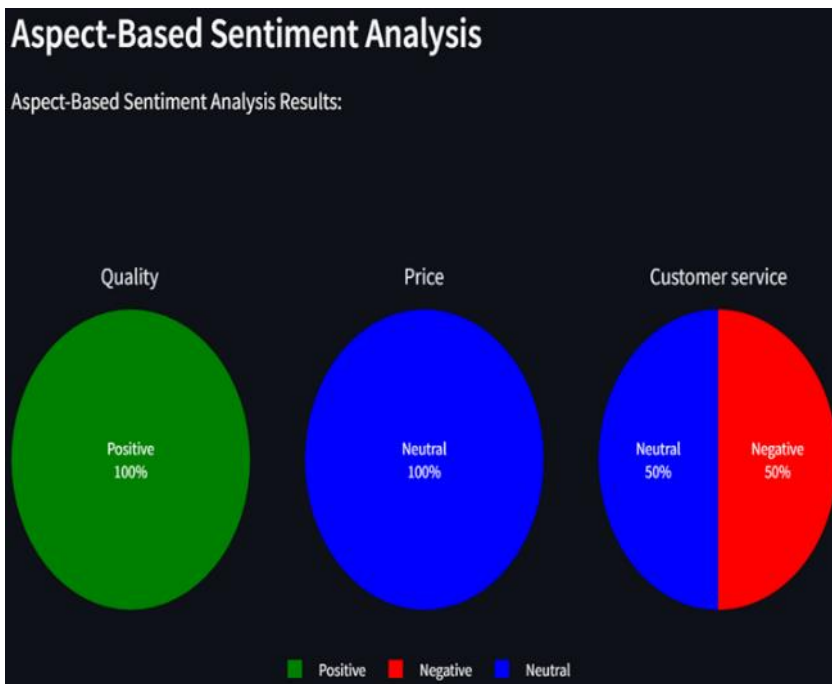
Sentiment Category Classifies each review as Positive, Negative, or Neutral based on predefined polarity thresholds (threshold\_pos = 0.5 for Positive and threshold\_neg = 0 for Negative). For example, uploading a CSV file with 100 reviews results in a table where each row displays the original review, its polarity score, and its sentiment classification. This structured format enables users to efficiently scan and analyze the distribution of sentiments, identifying trends such as the ratio of positive to negative feedback also added features like select sentiment and select category for easy accessing of desired reviews effectively.



**Fig. 9** Polarity Distribution For Obtained Classified Reviews

4.2.1. Visualization for Bulk Analysis

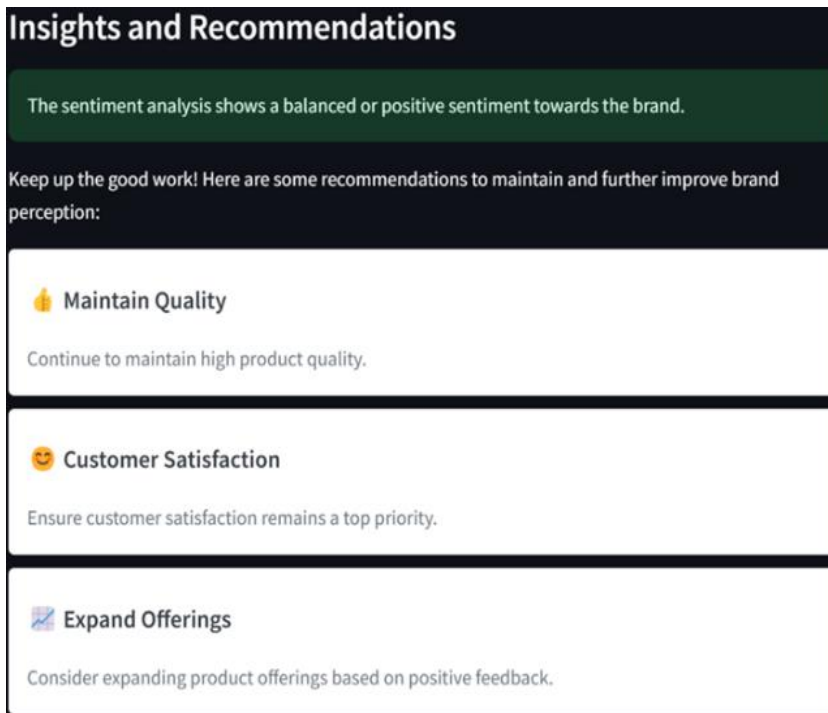
Complementing the tabular data, the Analyze CSV section offers dynamic visualizations to enhance data interpretation. One key visualization is the Polarity Distribution, where a histogram displays the distribution of sentiment polarity scores across all reviews. This visualization highlights the frequency of positive, negative, and neutral sentiments, providing a clear overview of overall sentiment trends as shown in **Figure 9**.



**Fig. 10** Aspect-Based Sentiment Analysis On Reviews Dataset

Additionally, this histogram allows users to quickly identify the predominant sentiment within the dataset, making it easier to spot any significant shifts or patterns in customer feedback. By visualizing the data in this manner, users can gain deeper insights into the sentiment landscape, facilitating more informed decision-making and strategic planning.

Furthermore, the visual representation of sentiment distribution helps in identifying outliers and anomalies, which might indicate specific issues or areas of excellence. These insights can be crucial for businesses aiming to improve their products or services based on customer feedback. Aspect-Based Sentiment Analysis: Pie charts represent the distribution of sentiments (Positive, Negative, Neutral) for specific aspects such as quality, price, and customer service as shown in **Figure 10**. This analysis offers granular insights into particular areas of customer feedback, helping identify strengths and areas needing improvement.



**Fig. 11** Insights And Recommendations For Classified Reviews

These visualizations are generated using Plotly, ensuring interactive and high-quality graphical representations. They enable users to quickly comprehend sentiment patterns and derive actionable insights to enhance brand perception and customer satisfaction.

#### 4.2.2. Insights and Recommendations

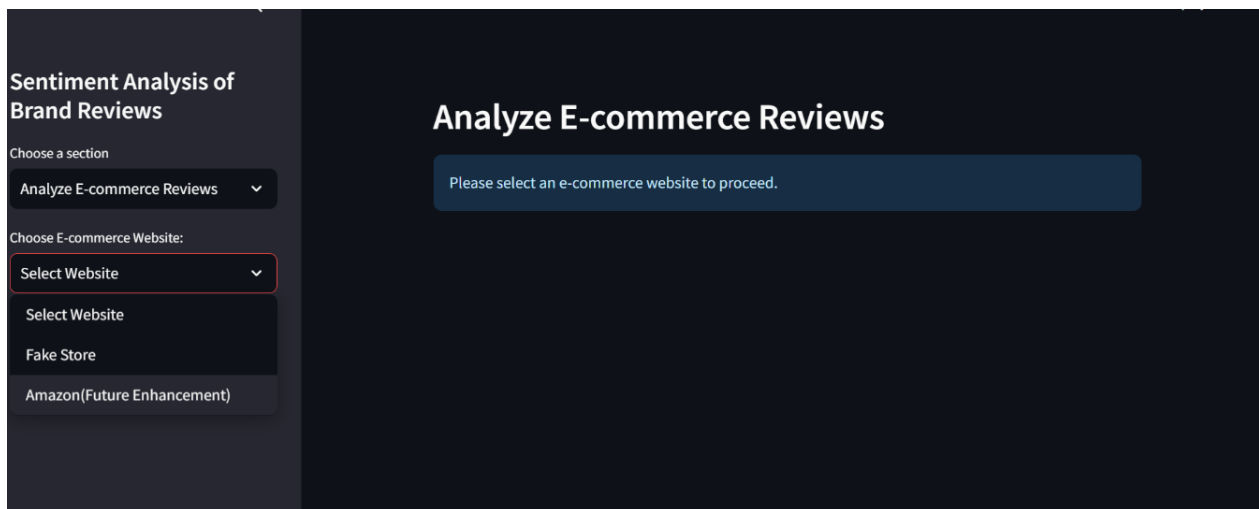
Based on the analyzed data, the application provides actionable insights and recommendations to enhance brand perception, as illustrated in **Figure 11**. For instance, if a considerable number of negative reviews are linked to specific aspects such as pricing or customer service, the application advises reevaluating pricing models or improving customer support systems.

Similarly, areas that receive predominantly positive feedback, such as product quality, are highlighted to encourage businesses to emphasize these strengths in marketing campaigns and product development initiatives. In scenarios where the sentiment distribution is balanced or moderately positive, the application suggests maintaining current strategies while identifying opportunities for further enhancements based on positive feedback trends.

#### 4.3. Analyze E-commerce product Reviews

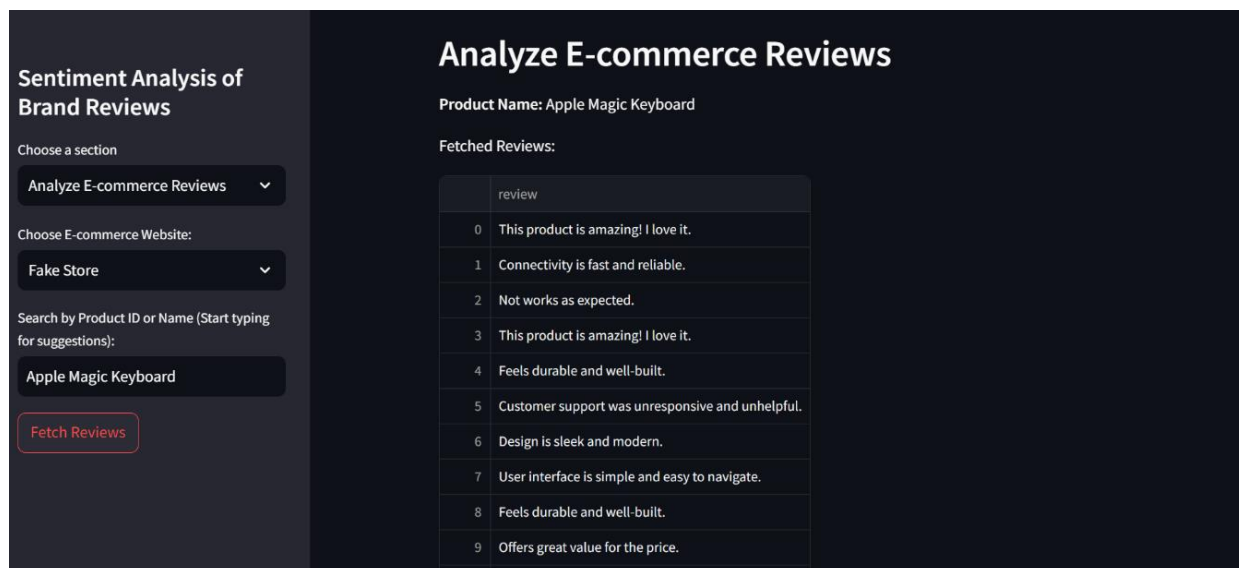
This sentiment analysis tool includes a dedicated module designed specifically for analyzing e-commerce product reviews as shown in Figure 12. This module enables users to evaluate customer feedback from online stores by

processing and categorizing reviews based on sentiment polarity (positive, negative, or neutral). Instead of relying on third-party platforms, we have created our own fake e-commerce store, which serves as a controlled environment for collecting, storing, and analyzing product reviews. By having our own e-commerce platform, we can generate, simulate, and fetch customer reviews directly from the store's database.



**Fig. 12** User Interface For Analyzing E-commerce Product Reviews

The reviews collected from the fake store undergo preprocessing and NLP-based sentiment analysis, helping us assess customer opinions about different products. This approach mimics the way businesses analyze real customer feedback from e-commerce platforms like Amazon, Flipkart, or eBay, making it an effective model for understanding and improving user satisfaction.



**Fig. 13** E-commerce Product Searching and Fetching Reviews Section

Upon selecting an e-commerce website it fetches reviews for the selected product and displays them in a table as shown in the **Figure 13**. Sentiment classification of e-commerce product reviews follows a similar approach to the analysis of CSV-based review datasets. The resulting visualizations typically include diagrams such as Sentiment Classification for a Given Set of Reviews, Polarity Distribution for the Classified Reviews, Aspect-Based Sentiment Analysis of the Reviews Dataset, and Insights and Recommendations Based on Classified Reviews.

### 4.3. A Comparative Analysis of Traditional Sentiment Analysis vs. the TextBlob-Based Approach

Sentiment analysis has evolved as a crucial tool in extracting meaningful insights from textual data, especially in customer feedback and social media monitoring. Traditional sentiment analysis methods, such as TextBlob and VADER, have been widely adopted for their simplicity and effectiveness in classifying text into broad sentiment categories like positive, negative, or neutral. While these tools are valuable for quick sentiment classification, they often fall short in offering detailed insights or handling complex, domain-specific datasets. For instance, VADER is designed to work best with short, informal texts, primarily in English, and doesn't consider subjectivity, limiting its application in more nuanced scenarios.

In contrast to these traditional methods, the approach employed in this study using TextBlob offers a more refined and flexible framework for sentiment analysis. By integrating aspect-based sentiment analysis, this approach is capable of not only determining the overall sentiment of a text but also identifying sentiment trends linked to specific aspects, such as product quality, customer service, and pricing. This level of granularity enables businesses to better understand customer opinions and address specific areas of concern. Moreover, the inclusion of Streamlit as an interactive user interface enhances usability, allowing users to easily switch between analyzing individual reviews and processing larger datasets, such as those provided in CSV file formats. This user-friendly interface simplifies the analysis process, making it accessible to a wider range of users, including those with limited technical expertise.

**Table 2** highlights the key differences between traditional sentiment analysis tools like TextBlob and VADER, and the enhanced approach used in this project.

**Table 2:** Traditional Sentiment Analysis vs. the TextBlob-Based Approach

Feature	Traditional Sentiment Analysis Tools (TextBlob, VADER)	TextBlob-based Approach (Project)
<b>Sentiment Categories</b>	Broad categories (positive, negative, neutral)	Aspect-based analysis (quality, price, customer service, etc.)
<b>Subjectivity Detection</b>	VADER lacks subjectivity analysis; TextBlob provides basic subjectivity score	Advanced subjectivity analysis integrated with sentiment polarity
<b>Visualization</b>	Basic visualizations (e.g., bar charts, pie charts)	Advanced visualizations using Plotly, showcasing sentiment distributions and aspect-specific trends
<b>Customizability</b>	Limited customizability (e.g., polarity threshold, basic sentiment classification)	Customizable sentiment thresholds, aspectspecific analysis, and flexible handling of complex datasets
<b>Interactive Interface</b>	No interactive interface	Interactive UI using Streamlit, allowing for seamless review and bulk dataset analysis
<b>Actionable Insights</b>	Basic insights on overall sentiment	Actionable insights for business strategy, such as improving customer service and brand perception
<b>Use Case</b>	General sentiment classification for social media, news, reviews	Tailored sentiment analysis for customer feedback, reviews, and industry-specific use cases

This comparison demonstrates how the TextBlob-based approach in this study enhances traditional sentiment analysis tools by offering more detailed, customizable, and actionable insights. The use of advanced visualizations, multilingual support, and aspect-based analysis ensures that businesses and researchers can extract deeper insights from customer feedback, thus improving decision-making and overall service quality.

## 5. Conclusion

This project effectively showcases the application of sentiment analysis using TextBlob and Streamlit to process and analyze brand reviews. By automating the classification of sentiments, businesses can derive valuable insights into customer emotions, which can significantly enhance their decision-making processes. The integration of advanced text preprocessing techniques not only boosts the accuracy of sentiment predictions but also ensures reliable and precise analysis, making the system highly effective for real-world applications [47].

The simplicity and versatility of the tool make it accessible to both technical and non-technical users. With its intuitive interface, users can easily input or upload customer reviews, perform sentiment analysis, and download the processed results for further evaluation. This streamlined workflow adds immense value to the tool, transforming it into a practical and scalable solution suitable for businesses of any size and across industries. Beyond its current capabilities, this project emphasizes the growing importance of sentiment analysis in modern business strategies. With broader applications in fields like customer experience, marketing, and product management, this project highlights the potential of data-driven decision-making, setting the stage for future innovations in sentiment analysis and natural language processing [48].

While the current sentiment analysis system is robust and capable of delivering valuable insights, several additional features could significantly enhance its functionality and provide deeper analytical capabilities. One such feature is Sentiment Trends Over Time, which would involve implementing a time-series analysis to track how customer sentiment evolves over weeks or months. This would allow businesses to observe patterns in customer feedback, such as how sentiment changes after a product update or a promotional campaign, helping them make timely and informed decisions.

Another enhancement is Keyword Analysis, which involves identifying the most frequently used words in positive, negative, or neutral reviews. This feature would help businesses pinpoint specific aspects of their products or services that customers appreciate or dislike. For example, if "durable" frequently appears in positive reviews, the company can highlight this in their marketing. Conversely, if "delay" is a recurring term in negative reviews, it may signal issues with shipping or service delivery that need to be addressed. Finally, comparison across products would enable users to analyze sentiment scores for multiple products or services within a single interface. This feature would be especially useful for companies offering a wide range of products, as it would allow them to quickly identify which products are performing well and which ones require improvement. By integrating these features, the system would provide even deeper insights, making it an indispensable tool for businesses seeking to improve customer satisfaction, refine their offerings, and gain a competitive edge in their industry.

## References

- [1] Pang, B.; Lee, L. Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval* 2008, 2, 1–135.  
Liu, B. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.
- [2] Kumar, V.; Soni, P.; Sinha, S. Sentiment Analysis on Customer Feedback: A Case Study Using TextBlob and NLTK. *International Journal of Computer Science and Information Security* 2019, 17, 133–138.
- [3] Zhang, L.; Zhang, Y.; Liu, B. A Survey on Sentiment Analysis: Approaches, Applications, and Challenges. *International Journal of Computer Science and Network Security* 2019, 19, 12–22.
- [4] Hutto, C.J.; Gilbert, E.E. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the 8th International Conference on Weblogs and Social Media (ICWSM 2014)*, 2014.
- [5] Saif, H.; He, Y.; Alani, H. Semantic Sentiment Analysis of Twitter. *Proceedings of the 22nd International Conference on Computational Linguistics* 2010, 1–8.
- [6] Bollen, J.; Mao, H.; Zeng, X.J. Twitter Mood Predicts the Stock Market. *Journal of Computational Science* 2011, 2, 1–8.
- [7] Yessenalina, A.; Cardie, C. Contextual Sentiment Analysis: On the Utilization of Context in Sentiment Classification.
- [8] Ghiassi, M., Skinner, J. and D. Zimbra, 2013, Twitter Brand Sentiment Analysis: A Hybrid System
- [9] Using N-Gram Analysis and Dynamic Artificial Neural Network. *Expert Systems with Applications*, 40, Issue 16, 6266–6282.
- [10] He, W., Zha, S. and L. Li, 2013, Social Media Competitive Analysis and Text Mining: A Case Study in The Pizza Industry. *International Journal of Information Management*, 33, Issue 3, 464–472.

- [11] Injadat, M., Salo, F. and Nassif A.B. (2016). Data Mining Techniques in Social Media: A Survey. *Neurocomputing*. Vol.214, 654-670. [12] Kang, Y., Wang, Y., Zhang, D. and Zhou, L. (2017). The Public's Opinions on A New School Meals Policy for Childhood Obesity Prevention in the U.S.: A Social Media Analytics Approach. *International Journal of Medical Informatics*. Vol. 103, 83-88.
- [13] Liu, B., 2008, *Web Data Mining Exploring Hyperlinks, Contents, and Usage Data*. Springer, 7, 411.
- [14] Morstatter, F. and Liu, H. (2017). *Discovering, Assessing, and Mitigating Data bias in Social Media*. *Online Social Networks and Media*. Vol.1, 1-13.
- [15] Mostafa, M.M., 2013, *More Than Words: Social Networks' Text Mining for Consumer Brand Sentiments*. *Expert Systems with Applications*, 40, Issue 10, 4241-4251.
- [16] Naylor, R.W., Lamberton, C.P. and P.M. West, 2012, Beyond the "Like" Button: The Impact of Mere Virtual Presence on Brand Evaluations and Purchase Intentions in Social Media Settings. *Journal of Marketing*, 76, No. 6, 105-120.
- [17] Noone B.M., McGuire, K.A. and K.V. Rohlfs, 2011, Social Media Meets Hotel Revenue Management: Opportunities, Issues and Unanswered Questions. *Journal of Revenue and Pricing Management*, 10, Issue 4, 293-305.
- [18] Pournarakis, D.E., Sotiropoulos, D.N. and Giaglis, G.M. (2017). *Decision Support Systems*. Vol.93, 98-110.
- [19] Russell, M.A., 2013, *Mining the Social Web*. O'reilly Media Inc., 5, 45.
- [20] Scarfi, M., 2012, *Social Media and the Big Data Explosion*. Forbes.
- [21] Sun, A., Lanchanski, M. and Fabozzi, F.J. (2016). Trade the Tweet: Social Media Text Mining and Sparse Matrix Factorization for Stock Market Prediction. *International Review of Financial Analysis*. Vol. 48, 272-281.
- [22] Tan, P.N., Steinbach, M. and V. Kumar, 2006, *Introduction to Data Mining*, Pearson, 3.
- [23] Thomaza, G.M., Bizb, A.A., Bettonic, E.M., Filhod, L.M. and Buhalise, D. (2016). Content Mining Framework in Social Media: A FIFA World Cup 2014 Case Analysis. *Information & Management*. Article in Press.
- [24] Younis, E.M.G., 2015, Sentiment Analysis and Text Mining for Social Media Microblogs Using Open Source Tools: An Empirical Study. *International Journal of Computer Applications*, 112, 5.
- [25] Zafarani, R., Abbasi, M.A. and H. Liu, 2014, *Social Media Mining: An Introduction*. Cambridge University Press, 16.
- [26] Ma, L.; Wang, Y. Constructing a semantic graph with depression symptoms extraction from twitter. In *Proceedings of the 16<sup>th</sup> IEEE International Conference on Computational Intelligence in Bioinformatics and Computational Biology*, Siena Tuscany, Italy, 9-11 July 2019; IEEE: New York, NY, USA; pp. 1-5.
- [27] Bhowmik, N.R.; Arifuzzaman, M.; Mondal, M.R.H.; Islam, M. Bangla text sentiment analysis using supervised machine learning with extended lexicon dictionary. *Nat. Lang. Process. Res.* 2021, 1, 34-45. [CrossRef]
- [28] Wu, J.J.; Chang, S.T. Exploring customer sentiment regarding online retail services: A topicbased approach. *J. Retail. Consum. Serv.* 2020, 55, 102145. [CrossRef]
- [29] Xu, F.; Pan, Z.; Xia, R. E-commerce product review sentiment classification based on a Naïve Bayes continuous learning framework. *Inf. Process. Manag.* 2020, 57, 102221. [CrossRef]
- [30] Kabir, A.I.; Ahmed, K.; Karim, R. Word Cloud and Sentiment Analysis of Amazon Earphones Reviews with R Programming Language. *Inform. Econ.* 2020, 24, 55-71. [CrossRef]
- [31] Balakrishnan, V.; Lok, P.Y.; Rahim, H.A. A semi-supervised approach in detecting sentiment and emotion based on digital payment reviews. *J. Supercomput.* 2021, 77, 3795-3810. [CrossRef]
- [32] Yang, L.; Li, Y.; Wang, J.; Sherratt, R.S. Sentiment analysis for E commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE Access* 2020, 8, 23522-23530. [CrossRef]
- [33] Carosia, A.E.; Coelho, G.P.; Silva, A.E. Investment strategies applied to the Brazilian stock market: A methodology based on sentiment analysis with deep learning. *Expert Syst. Appl.* 2021, 184, 115470. [CrossRef]
- [34] Zad, S.; Heidari, M.; Jones, J.H.; Uzuner, O. A survey on concept level sentiment analysis techniques of textual data. In *Proceedings of the 2021 IEEE World AI IoT Congress (AIIoT)*, Seattle, WA, USA, 10-13 May 2021; IEEE: New York, NY, USA; pp. 0285-0291.
- [35] Jing, N.; Wu, Z.; Wang, H. A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Syst. Appl.* 2021, 178, 115019. [CrossRef]
- [36] Keikhosrokiani, P.; Pourya Asl, M. (Eds.) *Handbook of Research on Opinion Mining and Text Analytics on Literary Works and Social Media*; IGI Global: Pennsylvania, PA, USA, 2022.
- [37] Fang, X.; Zhan, J. Sentiment analysis using product review data. *J. Big Data* 2015, 2, 5. [CrossRef]
- [38] Mukherjee, S.; Bhattacharyya, P. Feature specific sentiment analysis for product reviews. In *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*, New Delhi, India, 11-17 March 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 475-487.
- [39] Yadav, A.; Jha, C.K.; Sharan, A.; Vaish, V. Sentiment analysis of financial news using unsupervised approach. *Procedia Comput. Sci.* 2020, 167, 589-598. [CrossRef]

- [40] Pota, M.; Ventura, M.; Catelli, R.; Esposito, M. An effective BERT-based pipeline for twitter sentiment analysis: A case study in ITALIAN. *Sensors* 2021, 21, 133. [CrossRef] [PubMed]
- [41] Qurat, T.A.; Mubashir, A.; Amna, R.; Amna, N.; Muhammad, K.; Babar, H.; Rehman, A. Sentiment Analysis Using Deep Learning Techniques: A Review. *Int. J. Adv. Comput. Sci. Appl.* 2017, 8, 424–433. [CrossRef]
- [42] Shorten, C.; Khoshgoftaar, T.M.; Furht, B. Text data augmentation for deep learning. *J. Big Data* 2021, 8, 101. [CrossRef] [PubMed]
- [43] Shuai, Z.; Lina, Y.; Aixin, S. Deep Learning based Recommender System: A Survey and New Perspectives. *ACM J. Comput. Cult. Herit.* 2017, 52, 1–38.
- [44] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* 2017, 60, 84–90. [CrossRef]
- [45] Kobayashi, S. Contextual augmentation: Data augmentation bywords with paradigmatic relations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), New Orleans, LA, USA, 1–6 June 2018; Volume 2, pp. 452–457.* [CrossRef]
- [46] Duong, H.T.; Nguyen-Thi, T.A. A review: Preprocessing techniques and data augmentation for sentiment analysis. *Comput. Soc. Netw.* 2021, 8, 1. [CrossRef]
- [47] Birjali, M.; Kasri, M.; Beni-Hssane, A. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowl.-Based Syst.* 2021, 226, 107134. [CrossRef]
- [48] Haque, T.U.; Saber, N.N.; Shah, F.M. Sentiment analysis on large scale Amazon product reviews. In *Proceedings of the 2018 IEEE International Conference on Innovative Research and Development (ICIRD), Bangkok, Thailand, 11–12 May 2018.*