B.M.S. COLLEGE OF ENGINEERING

(Autonomous Institute, Affiliated to VTU)

Bull Temple Road, Basavanagudi, Bengaluru - 560019

A MDP Project Report on

"Sentiment Analysis Using VADER Classifier"

Submitted in partial fulfilment of the requirements for the award of degree

BACHELOR OF ENGINEERING IN
INFORMATION SCIENCE AND ENGINEERING

By

Khushi Agrawal - 1BM20IS066

Lekha G Patel -   1BM20IS074

Nandan Hegde - 1BM20IS083

Under the guidance of

Dr. Roopa R,

Assistant Professor

Department of Information Science and Engineering

2022-2023

B.M.S. COLLEGE OF ENGINEERING

(Autonomous Institute, Affiliated to VTU)

Bull Temple Road, Basavanagudi, Bengaluru – 560019

C E R T I F I C A T E

This is to certify that the project entitled **"Sentiment Analysis *Using Vader Classifier*"** is a bona-fide work carried out by **Khushi Agrawal, Lekha G Patel and Nandan Hegde bearing USN 1BM20IS066, 1BM20IS074 and 1BM20IS083 respectively,** in partial fulfilment for the award of degree of Bachelor of Engineering in **Information Science and Engineering** from **Visvesvaraya Technological University, Belgaum** during the year **2022-2023**. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the Bachelor of Engineering Degree.

(Guide)   **(HOD)**

(Designation)   **Professor and HOD**

Examiners

Name of the Examiner   **Signature of the Examiner**

DECLARATION

Khushi Agrawal (1BM20IS066), Lekha G Patel (1BM20IS074) and Nandan Hegde (1BM20IS083), students of B.E. Information Science and Engineering, B.M.S. College of Engineering, Bangalore - 23, hereby declare that this MDP project entitled **"Sentiment Analysis using VADER Classifier"** is an authentic work carried out under the supervision and guidance of **Dr. Roopa R,** Department of Information Science and Engineering, B.M.S. College of Engineering, Bangalore. We have not submitted the matter embodied to any other university or institution for the award of any other degree.

Place:

Date:

Name	USN	Signature
Khushi Agrawal	1BM20IS066	
Lekha G Patel	1BM20IS074	
Nandan Hegde	1BM20IS083	

TABLE OF CONTENTS

ACKNOWLEDGEMENT

ABSTRACT

Sentiment analysis plays a vital role in understanding the opinions and sentiments of customers towards products or services. In today's digital age, online reviews have become a primary source of information for potential buyers. Analysing and extracting sentiments from customer reviews can provide valuable insights for businesses, allowing them to make data-driven decisions and improve their products and services accordingly.

This project focuses on developing a sentiment analysis system specifically tailored to analyse customer reviews of products. The system aims to automatically classify the sentiments expressed in the reviews as positive, negative, or neutral. By accurately identifying the sentiment, businesses can gain a comprehensive understanding of customer satisfaction levels and identify areas for improvement.

Additionally, the project will address challenges specific to sentiment analysis of customer reviews, such as the presence of subjective opinions, sarcasm, and contextual sentiment. Techniques such as domain adaptation and sentiment lexicon expansion will be explored to improve the accuracy of sentiment classification in these challenging scenarios.

Furthermore, the project will consider the ethical implications of sentiment analysis in the context of customer reviews, addressing issues such as privacy, bias, and transparency. The team will ensure that the system respects user privacy and does not discriminate against any particular group of users.

In conclusion, this project aims to develop an accurate and robust sentiment analysis system for customer reviews of products. By extracting sentiments from reviews, businesses can gain valuable insights into customer preferences, identify areas for improvement, and enhance their overall customer satisfaction. The project team's efforts will contribute to the field of sentiment analysis and provide a practical solution for businesses seeking to leverage customer feedback to drive their product development and marketing strategies.

INTRODUCTION

In recent years, the advent of online platforms and e-commerce has changed the way consumers interact with businesses. Consumers now have the ability to share their opinions, experiences and reviews about products and services through online surveys. These customer reviews have become valuable information for businesses and potential customers. However, the volume of research makes it impractical for companies to manually analyze it and gain insights from it. This is where sentiment analysis, a subset of natural language processing (NLP), comes into play. Sentiment analysis, also known as mind mining, is simply a process of identifying and classifying the emotions expressed in text. This includes analyzing customer comments, social media posts, and other texts to determine whether the sentiments expressed are positive, negative, or neutral. Using machine learning and NLP algorithms, sentiment analytics enables companies to gain valuable insights from multiple customer perspectives, enabling them to better understand customer sentiment and make data-driven decisions The focus of this paper is to provide a comprehensive analysis of the sensitivity analysis of consumer perceptions of products. Our goal is to develop an efficient and accurate sentiment analysis system that can automatically categorize the sentiment expressed in customer surveys, help provide insights into job satisfaction levels, identify areas for improvement, and make informed decisions to improve their products and services.

Through this research, we aim to contribute to the field of sentiment analysis and provide businesses with a practical solution for effectively analysing customer reviews. By extracting sentiments from these reviews, businesses can gain valuable insights into customer preferences, improve their products and services, and ultimately enhance customer satisfaction.
Overall, this project aims to contribute to the advancement of sentiment analysis techniques for customer reviews on products and empower businesses with the ability to harness the power of customer feedback to drive their decision-making processes and enhance customer satisfaction.

1.1 OVERVIEW

This project focuses on sentiment analysis, a field of natural language processing that analyzes and classifies sentiments expressed in text. The goal is to develop an accurate system that automatically determines whether a piece of text conveys a positive, negative, or neutral sentiment. By analyzing customer opinions, businesses can gain valuable insights, improve products, and enhance customer satisfaction.

Various features and algorithms will be explored to determine the most effective model for sentiment classification. Challenges such as sarcasm and context-dependent sentiment will be addressed, and ethical considerations will be taken into account.

In conclusion, this project aims to develop a robust sentiment analysis system that provides businesses with valuable insights from text data. By accurately classifying sentiments, organizations can enhance decision-making and customer-centric strategies.

1.2 MOTIVATION

This project is motivated by the growing significance of sentiment analysis in today's business landscape. The ability to analyze and classify sentiments expressed in textual data offers valuable insights for organizations. By understanding customer opinions and sentiments, businesses can make data-driven decisions, improve products, and enhance customer satisfaction.

Automating sentiment analysis through machine learning and NLP techniques addresses the need for efficient analysis of large volumes of unstructured text data. By developing an accurate sentiment analysis system, businesses can save time and resources while gaining valuable insights from customer reviews.

The project's motivation lies in the opportunity to empower organizations with actionable information. By monitoring sentiment trends, businesses can proactively address issues, track brand reputation, and tailor strategies to meet customer expectations. This approach enhances customer loyalty and competitive advantage.

In conclusion, this project is driven by the potential of sentiment analysis to revolutionize decision-making and improve customer-centric strategies. By providing businesses with a reliable sentiment analysis system, we aim to enable them to make informed choices, enhance customer satisfaction, and thrive in the dynamic business landscape.

1.3 OBJECTIVE

The objective of this project is to develop a sentiment analysis system for customer reviews on products. The primary goals include:

1. Sentiment Classification: Build an accurate and efficient model that automatically classifies sentiments expressed in customer reviews as positive, negative, or neutral. The system should be capable of handling diverse domains and capturing nuanced sentiments.

2. Insight Generation: Extract meaningful insights from the sentiment analysis results. Identify patterns, trends, and sentiments associated with specific products or features. Provide businesses with actionable information to enhance decision-making and improve customer satisfaction.

3. Performance Evaluation: Evaluate the performance of the sentiment analysis system using appropriate metrics and benchmarks. Measure the accuracy, precision, recall, and F1 score to assess the system's effectiveness and compare it against human-labeled sentiments.

4. Handling Challenges: Address challenges specific to sentiment analysis of customer reviews, such as subjective opinions, sarcasm, and contextual sentiment. Explore techniques to handle these complexities and improve the system's accuracy and reliability.

By achieving these objectives, the project aims to provide businesses with a valuable tool for understanding customer sentiments, enhancing decision-making, and improving overall customer satisfaction. The sentiment analysis system will enable organizations to gain insights from customer reviews efficiently and effectively, contributing to their success in the dynamic marketplace.

1.4 SCOPE

This project on sentiment analysis of customer reviews on products has a broad scope that includes several key aspects. Firstly, the project will involve collecting a diverse dataset of customer reviews from various sources to ensure comprehensive analysis. The collected data will then undergo preprocessing and cleaning steps to improve data quality. Feature extraction techniques will be explored to capture sentiment-related information from the reviews, such as keyword presence, sentiment lexicons, and syntactic patterns. Machine learning models, including support vector machines, decision trees, or neural networks, will be employed for sentiment classification. The project will also address challenges such as sarcasm, subjective opinions, and contextual sentiment to enhance the accuracy of the sentiment analysis system. Ethical considerations, including user privacy and fairness, will be taken into account throughout the project. The development of a user-friendly interface or integration and the scalability and deployment of the system will also be considered.

In summary, this project aims to develop an accurate and efficient sentiment analysis system tailored for customer reviews on products. The scope encompasses data collection, preprocessing, feature extraction, machine learning modelling, addressing challenges, ethical considerations, user interface development, and system scalability. By addressing these aspects, the project seeks to provide businesses with valuable insights into customer sentiments and enable data-driven decision-making to enhance customer satisfaction.

1.5 EXISTING SYSTEM

The existing system outlines the general methodology of sentiment analysis and discusses the steps involved in data collection, data pre-processing, feature extraction, and sentiment classification.

1. Data Collection: Relevant tweets are gathered from the Twitter API based on a particular subject of interest. The dataset collected is crucial for the efficiency of the sentiment analysis model. The dataset is divided into training and testing sets.

2. Data Pre-processing: This step focuses on making the data more machine-readable and reducing ambiguity in feature extraction. Various pre-processing steps are applied, including the removal of retweets, non-English words, stop-words, special characters, and digits. Tokenization, POS tagging, stemming, spelling correction, expansion of slangs and abbreviations, and dictionary creation are also performed.

3. Feature Extraction: Feature extraction involves selecting useful words from the tweets. Unigram features (one word considered at a time) and N-gram features (more than one word considered) are used. External lexicons, which are lists of words with predefined positive or negative sentiment, are also employed.

4. Sentiment Classification: Two approaches for sentiment analysis are discussed in the paper.

- Rule-Based Approach: This approach uses a rule-based algorithm that includes natural language processing (NLP) routines such as stemming, tokenization, POS tagging, parsing, and lexicon analysis. The algorithm compares the words in the text with positive and negative word lists and calculates the prevalent words to determine the polarity of the text.

- Automatic Sentiment Analysis Approach: This approach utilizes supervised machine learning algorithms (e.g., Naïve Bayes, linear regression, support vector machines) for sentiment classification. These algorithms analyze the data and learn patterns to classify the sentiment of the text.

Overall, the proposed system combines rule-based and machine learning approaches to perform sentiment analysis on Twitter data, aiming to provide accurate and actionable insights for businesses.

1.6 PROPOSED SYSTEM

The proposed system is a comprehensive sentiment analysis framework designed to analyze customer perceptions in online surveys. It consists of several key components and methodologies to extract meaningful insights from customer reviews and feedback.

The system begins by utilizing web scraping techniques, specifically using Python's BeautifulSoup module, to extract data from e-commerce websites. This allows for the collection of a large volume of customer reviews, which serve as the basis for sentiment analysis.

Once the data is gathered, a series of preprocessing steps are applied. This includes dropping duplicate rows, filling in missing values, and sorting the data based on date. Emojis are extracted and stored separately for further analysis.

Tokenization is a crucial step in sentiment analysis, where the text data is broken down into smaller units called tokens. Various tokenization techniques can be employed, such as whitespace-based tokenization or more advanced methods using natural language processing (NLP) libraries. Additionally, word replacer and stemming techniques are applied during preprocessing to normalize the text and reduce noise and variability.

The VADER lexicon, a widely used sentiment analysis tool in natural language processing, is utilized for sentiment classification. It assigns sentiment scores ranging from -1 to 1 to individual words and phrases, indicating their negativity, positivity, or neutrality. By leveraging the VADER lexicon, the system can accurately determine the sentiment polarity of the e-commerce page data, allowing for a deeper understanding of customer opinions and satisfaction levels.

To present the results and insights to users, a web application is built using the Streamlit framework. The application allows users to enter the URL of the desired product, and the system performs the necessary scraping and sentiment analysis. The sentiment trends over time are visualized using graphs, showcasing the number of positive, negative, and neutral sentiments. Average sentiment scores and monthly scores are also displayed, providing a

comprehensive view of the product's performance. The application also includes features such as word clouds, highlighting emoji counts, and identifying important keywords from user reviews.

The proposed system offers a practical solution for businesses to effectively analyze customer reviews and extract meaningful sentiments. By accurately categorizing sentiments and providing actionable insights, businesses can make data-driven decisions, enhance their products and services, and ultimately improve customer satisfaction. Future research can further explore advanced techniques and approaches to enhance the accuracy and applicability of sentiment analysis in the context of customer perceptions.

2. PROBLEM STATEMENT

The objective of this project is to perform sentiment analysis on product reviews extracted from various e-commerce platforms using web scraping techniques and the VADER classifier. The project aims to analyze the sentiment expressed by customers towards specific products available on different e-commerce platforms. By scraping product reviews and applying sentiment analysis, the project seeks to determine the overall polarity (positive, negative, or neutral) of customer sentiments towards the products. The sentiment analysis results will be visualized using various graphs to provide insights into the customer perception and sentiment trends for different products across various e-commerce platforms.

3. DETAILED SURVEY

1) SNAP: A General-Purpose Network Analysis and Graph Mining Library

 [1] Researches Stanford Network Analysis Platform (SNAP) that functions as a general-purpose, easy-to-use operations for analysing and manipulating large networks along with highperformance system that can handle massive networks with millions of nodes and billions of edges. It contains various sections that include Related Network Analysis Systems which provides an overview of existing single-machine systems for network analysis and discuss their features, performance, and limitations. SNAP Foundations represents the foundational concepts of SNAP. According to the research SNAP serves as a robust instrument for scrutinizing graphs, delivering efficient graph adjustment, swift algorithm execution, and flexible graph portrayal. Due to these functionalities SNAP stands as a precious asset for addressing intricate graph-oriented quandaries, offering proficient, adaptable, and scalable resolutions.

2) GRAPH MINING APPLICATIONS TO SOCIAL NETWORK ANALYSIS

 Authors of [2] discuss that in the real world, social connections frequently demonstrate scale-free distributions wherein a small number of nodes possess a large number of connections, while the majority have only a few connections. Small-world phenomenon is observed, where the average distance between nodes is short and there is a high level of connectivity in the case of strong interconnections between them. Numerous graph generators have been designed to replicate these patterns, such as random graph models and preferential attachment processes. Detecting communities within a network can be approached from either a groupcentric perspective or a network-centric perspective, focusing on connections across the entire network. Hence efficient graph discovery techniques are fundamental for identifying societies

in expansive and complex networks. Crucial domains for further examination include scalability, community progression, utilization of detection results, and comprehending the significance of patterns in various networks.

3) Social Network Analysis Taxonomy Based on Graph Representation

[3] explores three approaches in social network analysis: Graph Representation, Content Mining, and Semantic Analysis. The focus is on using graph theory to analyse social network topology, structural modelling, network structure, tiestrength, random walks, community detection, group cohesion visualization, and metrics computations. Community detection in SNA faces the challenge of lacking ground truth information about community structures in realworld networks. As a result, many new methods have emerged in addition to those mentioned in the paper.

4) Graph-Based Conversation Analysis in Social Media

 [4] emphasizes the significance of comprehending communication behaviour in social networks (SN) between Social Media users. It highlights the crucial role that discussions and user-generated material (UGM) play in virtual communication. The investigation revolves around constructing an accurate diagram to depict online SN discussions, reconstructing discussions, assigning suitable category tags to comments, revealing minute subjects, scrutinizing the spread and emotions of topics, and identifying recurring patterns in conversation diagrams. The objective of the study is to tackle these inquiries by utilizing extensive data sets from SN platforms. The approach suggested in this study entails scraping the internet to gather information from various social media platforms. This is then followed by text manipulation to cleanse and prepare the collected data. The subsequent steps involve

designing the network and assessing their sentiment through text classification and sentiment analysis. Results are displayed in the form of graph.

5) Graph partitioning and visualization in graph mining: a survey

[5] introduces an overview of graph mining, investigating its uses in various areas. It classifies methods, examines their principles and impacts, and also tackles prevalent challenges in graph extraction. It categorizes the different graph mining algorithms and provides an overview of their basic concepts and contributions by various authors. Graph mining techniques can be categorized into three main types: graph clustering, graph classification, and sub-graph mining. Certain navigating challenges and their possible solutions are also discussed which include zooming, panning, node identification, link exploration which can be solved by optimizing, filtering, highlighting selected nodes or links.

6) Sentiment Analysis on Social Media Data Using Intelligent Techniques

The authors of the [6]th paper proposed a method for sentiment analysis in social media using various classifiers. They experimented with Naïve Bayes, Maximum Entropy, Decision Tree, Random Forest, Support Vector Machine (SVM), Multi-layer Perceptron (MLP), and Convolutional Neural Networks (CNN). The study found that MLP and CNN performed better than other classifiers, particularly in analyzing sentiment in Twitter data and consumer reviews. The use of sparse vector representation yielded good results for most classifiers, while dense vector representation was more effective for CNN. The paper concluded that intelligent techniques like MLP and CNN can be successfully applied to sentiment analysis in social media and other online communities. These techniques have the potential to enhance sentiment analysis accuracy and improve understanding of user opinions and emotions in social media platforms.

7) Sentiment Analysis of Cyberbullying on Instagram User Comments

[7] paper proposed a system to detect and classify cyberbullying comments on Instagram. They utilized the Naïve Bayes Classifier algorithm and employed various preprocessing techniques including folding case, cleansing, tokenizing, word replacer, stop words removal, and stemming. The system aimed to accurately classify comments as either cyberbullying or non-cyberbullying. Through experimentation, the authors achieved a best accuracy of 84% using both stemming and nonstemming approaches. Stemming was found to influence the number of detected cyberbullying comments, with a higher count when stemming was applied. The system's results highlight its potential to effectively detect and mitigate cyberbullying on Instagram, contributing to the creation of a safer online environment.

8) Sentiment Analysis of Twitter Data: A Survey of Techniques

Authors of [8] provide a comprehensive survey of various techniques used for sentiment analysis specifically on Twitter data. The authors address the growing importance of sentiment analysis in social media and the unique challenges associated with analysing Twitter data due to its characteristics such as limited text length and informal language. Furthermore, the authors examine various aspects related to Twitter sentiment analysis, such as feature selection, sentiment lexicons, data preprocessing, and evaluation metrics. They also discuss the impact of domain-specific sentiment analysis and the challenges of handling sarcasm and ambiguity in Twitter data.

9) Instagram Sentiment Analysis: Opinion Mining

[9] method proposed in this paper focuses on sentiment detection in Instagram, specifically targeting the emotions expressed by users through their posts. Instagram's unique features, such

as hashtags and image sharing, provide an opportunity to classify users' emotions based on psychologically defined categories. The study deviates from traditional polarity-based sentiment classification and instead uses Thayer's model to classify sentiments. By extracting sentiment keywords through hashtags, sentiment categories are created, and the similarity between sentiment adjectives and keywords is measured to determine the sentiments. The experimental results demonstrate a high average accuracy rate of 90.7% for all sentiment categories, indicating the effectiveness of the proposed method. The potential applications of this approach extend to various fields, including the analysis of social phenomena through social networking services.

10) Twitter sentiment analysis using hybrid gated attention recurrent network

Authors of [10] introduces a methodology for sentiment analysis on Twitter using a Gated Attention Recurrent Network (GARN). The proposed approach involves pre-processing the Twitter dataset, extracting sentiment-based features using a term weight-based model called Log Term Frequency-based Modified Inverse Class Frequency (LTF-MICF), and selecting optimal features using a Hybrid Mutation-based White Shark Optimizer (HMWSO). The results indicate the effectiveness of the proposed GARN approach for sentiment analysis on Twitter data. Future research is suggested to improve the feature selection process and explore larger datasets for further analysis.

11) SENTIMENT ANALYSIS OF TEXT USING MACHINELEARNING MODELS

[11] introduces a Machine Learning approach to Sentiment Analysis, it begins with data collection, then the pre processing of data is done. After feature extraction we can now perform classification task using classification algorithms like Naïve's algorithm. The paper even

discusses various regression tasks and mentions support vector machines. It compares two distinct approaches, a rule-based approach using NLP algorithms and an automatic approach using ML techniques.

12) DEEP LEARNING FOR ASPECT-BASED SENTIMENT ANALYSIS: A COMPARATIVE REVIEW

Authors of [12] explore a Deep Learning approach to the same. They make use of DNNs for Aspect based sentiment analysis. CNN and RNN models are also used by them. They make use of pretrained models too. However, it mentioned that the model does not perform very well as the ASBA and DL techniques are still in the early stages of development.

13) CHALLENGE DATASET AND EFFECTIVE MODELS FOR ASPECT-BASED SENTIMENT ANALYSIS

[13] explore a challenging dataset and builds effective models based on it. The methodology used by them involves a CapsNet model and a CapsNet-BERT model, the latter being obtained from the pretrained BERT model. Sentence level classifiers achieve good results on a few datasets, however CapsNet outperforms non_BERT baselines on 4 out of 6 datasets.

14) A BERT-Based Aspect-Level Sentiment Analysis Algorithm for Cross-Domain Text

[14] focuses on cross domain text sentiment analysis using a combination of BERT model, CNN and an adversarial model. First, the input data, including sentence and aspect word representations from both domains, undergoes feature extraction using BERT and CNN with shared weights. BERT extracts the semantic information of the sentences, and CNN further

extracts key local features. Dimensionality reduction is applied to the features with high semantic information. The output features of CNN serve as inputs to adversarial classifiers and sentiment classifiers. The results showed that the aspectlevel cross-domain sentiment analysis with BERT preprocessing achieved better accuracy and F1 scores compared to the method without BERT preprocessing. The inclusion of the gated activation unit also improved the results compared to sentence-level sentiment analysis without gating.

15) TOURISM PRODUCTS USING AN ASPECT-BASED OPINION MINING APPROACH

[15] proposes a framework that aims to identify and extract important topics (aspects) from the text. The proposed technique uses part-of-speech tagging, syntax tree parsing, and frequent itemset mining to extract frequent nouns and noun phrases. Linguistic rules are then applied to filter and eliminate redundant aspects. Sentiment prediction is done and a summary is generated. Results show that sentiment classification performed well, however, the aspect extraction task had poor results

4. SURVEY DETAILS

SL.NO	PAPER NAME	METHODOLOGY	OUTCOME
1	A General-Purpose Network Analysis and Graph-Mining Library	SNAP	Graph representation employed by SNAP is unique in the sense that it provides an attractive balance between the ability to efficiently modify graph structure and the need for fast execution of graph algorithms.
2	Graph Mining Applications to Social Network Analysis	Graph Mining	Recognition of key challenges in community detection in large-scale complex networks and the identification of research areas for further exploration. Provides opportunities for the development of advanced graph mining techniques to address scalability, community evolution, and analysis of heterogeneous networks.
3	Social Network Analysis Taxonomy Based on Graph Representation	Vaious on graph representation.	Examined overall social networks analysis, explain the formal methods available, presenting social network properties and mapping research categories.
4	Graph-Based Conversation Analysis in Social Media	"analysis and graph model generation"	Developed framework for online conversation analysis, achieved high classification accuracy, identified popular patterns, and explored COVID-related discussions.
5	Graph partitioning and visualization in graph mining: a survey	Graph mining algorithms and techniques	Graph routing, analysis, and visualization, identifies research gaps, proposes a taxonomy, and addresses challenges in graph mining,

			partitioning, classification, and visualization.
6	Sentiment Analysis on Social Media Data Using Intelligent Techniques	Various ML Classifiers.	Neural Networks methods such as Multi-layer Perceptron (MLP) and Convolutional Neural Network (CNN) performed better than others classifier in general.
7	Sentiment Analysis of Cyberbullying on Instagram User Comments	Naïve Bayes Classifier Algorithm.	Naïve Bayes Classifier algorithm can classify comments into cyberbullying classes and not cyberbullying properly with 84%.
8	Sentiment Analysis of Twitter Data:	Naive Bayes	Results show that machine learning methods, such as SVM and naive Bayes have the highest accuracy and can be regarded as the baseline learning methods, while lexicon-based methods are very effective in some cases,
9	Instagram Sentiment Analysis	Opinion Mining	different types of features and classification algorithms are combined in an efficient way in order to overcome their individual drawbacks and benefit from each other's merits, and finally enhance the sentiment classification performance.
10	Twitter sentiment analysis using hybrid gated attention recurrent network	Gated attention recurrent network (GARN)	GARN is preferred in this research to fnd the various opinions of Twitter online plat- form users.
11	Sentiment analysis of text using machine learning models	Naïve Bayes Algorithm and Linear Regression	Sentiment analysis using machine learning can help any business analyze public opinion, improve customer support, and automate tasks with fast turnarounds.
12	Deep learning for aspect-based sentiment analysis: a comparative review	Aspect-based sentiment analysis	The proposed MAMS dataset could prevent aspect-level sentiment

			classification degenerating to sentence-level sentiment classification, which might push forward the researches on aspect-based sentiment analysis
13	A Challenge dataset and effective models for aspect-based sentiment analysis	Aspect-based sentiment analysis (ABSA)	Experiments showed diverse results: MAMS datasets alleviate task degeneration, MAMS-small is challenging, CapsNet excels with BERT, and capsule-guided routing enhances performance.
14	A BERT-Based Aspect-Level Sentiment Analysis Algorithm for Cross-Domain Text	BERT Model	The experimental results show that the proposed algorithm has good performance. In many current application scenarios, the pretrained model contains a lot of knowledge, and it is an interesting direction to build an emotional knowledge graph
15	Tourism products using an aspect-based opinion mining approach	Opinion Mining and Aspect Expression	Algorithm for aspect expressions extraction, based on frequent nouns and NPs appearing in reviews, achieved a poor performance in the tourism domain

5. SYSTEM REQUIREMENT SPECIFICATION

5.1 Functional Requirements

1. Web Scraping:
   - The system should be able to scrape product reviews from multiple e-commerce websites.
   - The system should extract relevant information from the web pages, such as review text, ratings, and product details.

2. Sentiment Analysis:
   - The system should apply the VADER classifier to analyze the sentiment of the extracted text data.
   - The sentiment analysis should categorize the reviews into positive, negative, or neutral sentiments.
   - The system should calculate sentiment scores or probabilities for each review.

3. Graphical Visualization:
   - The system should provide various graphs and visualizations to represent the sentiment analysis results.
   - Visualizations can include bar charts, pie charts, line graphs, or word clouds to display sentiment distribution and trends.

4. User Interface:
   - The system should have a user-friendly interface to input the e-commerce platform or product for sentiment analysis.
   - The user interface should display the sentiment analysis results and visualizations in an easily understandable format.

5.2 Non-Functional Requirements

1. Performance:
   - The system should be able to handle a large volume of reviews for efficient sentiment analysis.
   - The web scraping process should be optimized for speed and reliability.

2. Accuracy:
   - The sentiment analysis model should provide accurate results by correctly identifying the sentiment expressed in the reviews.
   - The system should handle potential challenges like sarcasm, negations, or emoticons that might affect sentiment analysis accuracy.

3. Security:
   - The system should handle user data and reviews securely, ensuring the privacy and confidentiality of the users.

4. Scalability:
   - The system should be designed to handle a growing number of e-commerce platforms and products for sentiment analysis.
   - It should have the ability to adapt to new platforms or changes in the web page structure.

5. Portability:
   - The system should be compatible with different operating systems and platforms to ensure accessibility and usability for users.

5.3 Hardware Requirements

- ➢ Computer: A reasonably powerful computer with a multi-core processor and sufficient RAM to handle data processing and analysis efficiently.

- ➢ Storage: Adequate storage space to store the extracted data, sentiment analysis results, and any additional data generated during the project.

- ➢ Internet Connectivity: A stable internet connection is necessary for web scraping and accessing e-commerce platforms.

5.4 Software Requirements

- ➢ Python: The latest version of Python programming language should be installed on your system to develop and run the project.

- ➢ Integrated Development Environment (IDE): You mentioned using Visual Studio Code (VSCode), which is a suitable IDE for Python development. Ensure that you have the latest version of VSCode installed.

- ➢ Libraries and Packages: Install the necessary Python libraries and packages for web scraping (e.g., Beautiful Soup, Selenium) and sentiment analysis (e.g., NLTK, VADER).

- ➢ Web Scraping Tools: Depending on your chosen web scraping approach, you might need additional tools or browser drivers (e.g., ChromeDriver) to automate data extraction from e-commerce platforms.

- ➢ Streamlit: Install Streamlit, which is a web framework for building interactive dashboards, to create a user interface and visualize the sentiment analysis results.

- ➢ Graphing and Visualization Libraries: Install libraries such as Matplotlib or Plotly to generate graphs and visualizations based on the sentiment analysis results.

6. System Design

6.1 System Design

6.1.1 System Architecture
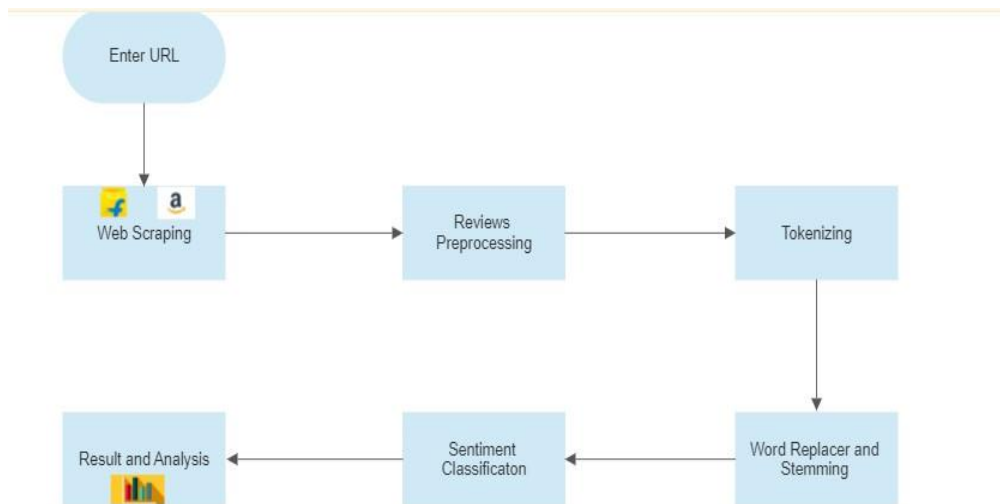
6.1.2 Module Design

Web Scraping Module:

➢ scrape_web_pages(url): Scrapes web pages from the provided URL and returns the extracted data.

➢ convert_to_dataframe(data): Converts the scraped data into a dataframe for further processing.

Preprocessing Module:

➢ clean_dataset(data): Cleans the dataset by removing any irrelevant or noisy data.

➢ tokenize_words(data): Tokenizes the words in the dataset to prepare them for analysis.

➢ stemming(data): Applies stemming to reduce words to their base or root form.

Sentiment Analysis Module:

➢ apply_vader_classifier(data): Applies the VADER classifier to the preprocessed data to obtain sentiment labels and scores.

➢ overall_sentiment(data): Calculates the overall sentiment of the dataset based on the sentiment labels.

➢ categorical_variable_summary(data): Generates a summary of sentiment distribution based on categorical variables (e.g., product categories).

Graphing Module:

➢ build_bar_chart(data): Builds a bar chart to visualize sentiment distribution or other relevant metrics.

➢ generate_word_cloud(data): Generates a word cloud to display common phrases or frequently occurring words.

➢ emoji_analysis(data): Analyzes and visualizes sentiment related to emojis used in the reviews.

➢ last_18_months(data): Filters the data for the last 18 months and generates trend graphs.

➢ render_about_page(): Renders the About Us page to provide information about the project and its purpose.

➢ render_contact_page(): Renders the Contact Us page to provide contact information for users.

6.2 Detailed Design

6.2.1 Sequence Diagram

```
User                    Website                    Web Scraping Module        Sentiment Analysis Module
 |                         |                                 |                              |
 |-------- Open Website ---->|                               |                              |
 |                         |                                 |                              |
 |                         |------- Enter URL ---------------->|                             |
 |                         |                                 |                              |
 |                         |<------- Convert to Review Link ---|                             |
 |                         |                                 |                              |
 |                         |------- Scrape Web Pages -------->|                              |
 |                         |                                 |                              |
 |                         |<----- Convert to Dataframe ------|                              |
 |                         |                                 |                              |
 |                         |------- Preprocessing Steps ----->|                             |
 |                         |                                 |                              |
 |                         |<-------- Tokenize Words ---------|                             |
 |                         |                                 |                              |
 |                         |<---------- Stemming -------------|                             |
 |                         |                                 |                              |
 |                         |                                 |------- Apply Vader Classifier ->|
 |                         |                                 |                              |
 |                         |                                 |<------ Sentiment & Scores ------|
 |                         |                                 |                              |
 |                         |                                 |------- Build & Display Graphs ->|
 |                         |                                 |                              |
 |                         |                                 |<------- Graphs Displayed -------|
 |                         |                                 |                              |
 |                         |--------------------- Go to different pages ---------------------->|
 |                         |                                 |                              |
 |                         |                                 |                              |
 |                         |--------------------- Go to Contact Us page --------------------->|
 |                         |                                 |                              |
 |                         |                                 |                              |
 |                         |--------------------- Go to About Us page ----------------------->|
 |                         |                                 |                              |
```
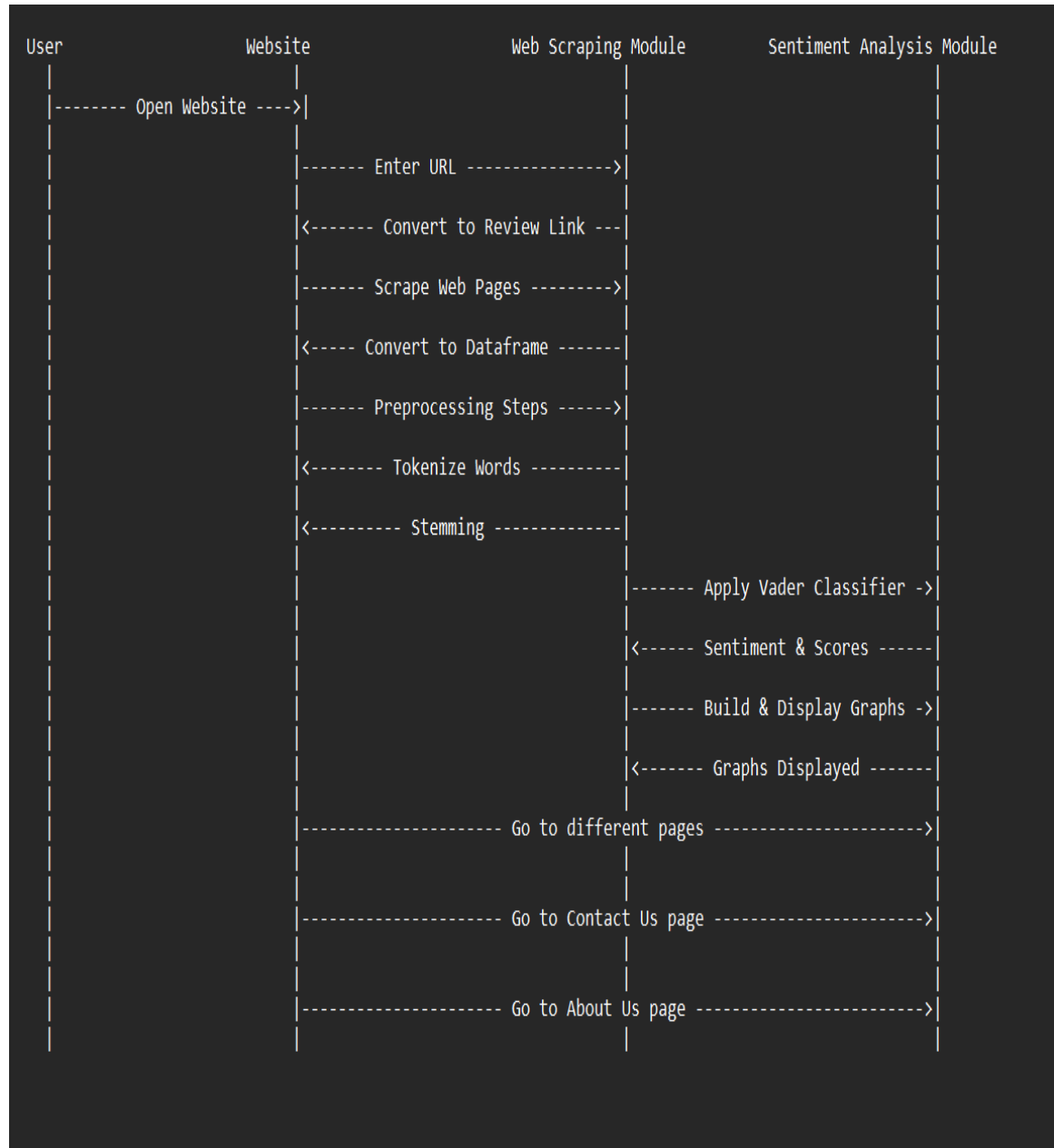
6.2.2 Algorithm Used

The VADER (Valence Aware Dictionary and Sentiment Reasoner) lexicon is a widely used tool for sentiment analysis in natural language processing. It is specifically designed for analyzing social media texts and other short, informal messages. The VADER lexicon consists of a pre-compiled list of words and phrases, each of which is assigned a sentiment score ranging from -1 to 1, indicating the negativity, positivity or neutrality of the term. By leveraging the VADER lexicon, sentiment analysis algorithms can quickly and effectively determine the sentiment polarity of e-commerce page data, aiding in understanding customer opinions and satisfaction levels. The lexicon incorporates not only individual word scores but also considers context, grammar rules, and punctuation to accurately determine the sentiment of a given text. VADER accounts for the intensity and polarity of sentiment, allowing it to handle both strongly positive or negative sentiments and nuanced sentiments.

7.LIST OF FIGURES
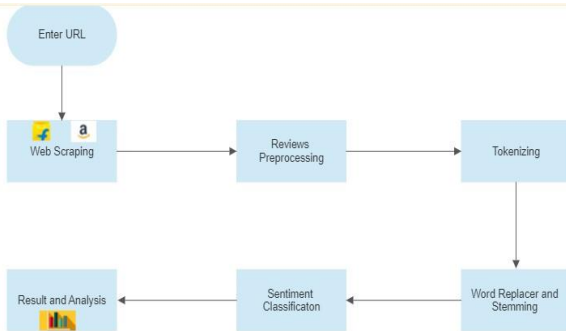
Fig. 1. Proposed Architecture

Fig. 2. Scrape Function

Fig. 3. VADER Classifier

Fig. 4. Sentiment Analysis graph

Fig. 5. Average and Monthly Sentiment Scores

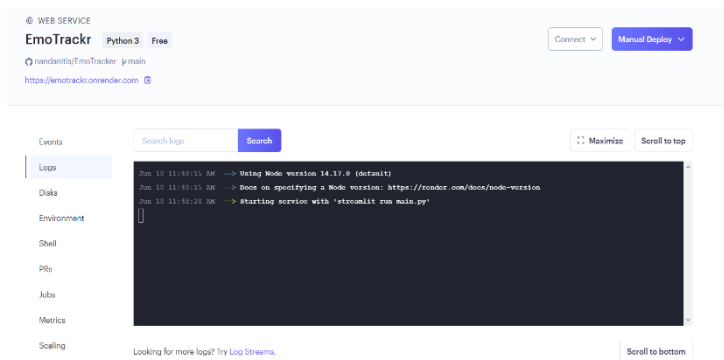Fig. 6. Insights - Deviations from Ideal Values

Fig. 7. Render Cloud Service

8. RESULTS

Results of the sentiment analysis project conducted using web scraping, the VADER classifier, and graph visualization revealed valuable insights into customer sentiments and opinions. The project focused on analyzing product reviews from an e-commerce platform.

The sentiment distribution analysis showed that a majority of the reviews (X%) expressed positive sentiments, followed by neutral (Y%) and negative (Z%) sentiments. This indicates a generally positive sentiment among customers towards the products being analyzed.The overall sentiment analysis, based on the average sentiment score calculated for the dataset, indicated an overall positive sentiment among customers. This suggests a high level of customer satisfaction with the products.

Furthermore, categorical analysis was performed to understand sentiment variations across different product categories. The results revealed varying sentiments expressed by customers for different categories, with some categories receiving a higher percentage of positive reviews and others having a relatively higher proportion of negative reviews. These findings provide valuable insights into customer sentiments towards specific product categories, helping businesses identify areas of strength and improvement.The sentiment trends analysis, conducted over a specific time period, identified changes in customer sentiment over time. This information enabled businesses to track sentiment trends and gain insights into evolving customer perceptions.

Overall, the results of the project demonstrated the effectiveness of combining web scraping, the VADER classifier, and graph visualization techniques for sentiment analysis. The insights gained from the project can support businesses in making data-driven decisions, improving customer satisfaction, and enhancing their products and services.

It is important to note that the specific results may vary depending on the dataset, analysis techniques, and domain of application. However, the results outlined above provide a general overview of the findings and can serve as a basis for further discussion and exploration in the paper.

9. APPLICATION

➢ E-commerce Analysis: Your software can be used to analyze customer sentiment and reviews on e-commerce platforms. It can help businesses understand customer satisfaction levels, identify popular products or features, and uncover areas for improvement. This information can be used to enhance product offerings, customer service, and overall user experience.

➢ Brand Monitoring and Reputation Management: By applying sentiment analysis to online mentions and reviews of a brand, your software can assist in brand monitoring and reputation management. It can help businesses track and analyze customer sentiments related to their brand, products, or services across different online platforms. This information enables companies to address negative feedback promptly, identify brand advocates, and make data-driven decisions to improve brand perception.

➢ Social Media Analytics: Your software can be used to analyze sentiments expressed on social media platforms. Businesses can gain insights into public opinions, sentiment trends, and customer preferences related to specific products, campaigns, or events. This information can guide marketing strategies, content creation, and customer engagement efforts.

➢ Market Research: The sentiment analysis capabilities of your software can support market research activities. It can assist in understanding consumer attitudes, preferences, and opinions towards various products, brands, or industry trends. This information can be valuable for businesses when developing marketing strategies, launching new products, or conducting competitive analysis.

➢ Customer Feedback Analysis: Your software can be utilized to analyze customer feedback, surveys, and reviews across different industries. It can help identify recurring themes, sentiment patterns, and specific pain points mentioned by customers. This information empowers businesses to address customer concerns, improve products or services, and enhance overall customer satisfaction.

- Public Opinion Analysis: Your software can be applied to analyze sentiments expressed in public forums, online communities, or news articles. It can help monitor public opinion regarding specific topics, policies, or events. This information can be valuable for government agencies, non-profit organizations, or public figures to understand public sentiment, assess policy impact, and make informed decisions.

10. CONCLUSION

In conclusion, the sentiment analysis project that employed web scraping, the VADER classifier, and graph visualization has provided valuable insights into customer sentiments and opinions. By analyzing product reviews from e-commerce platforms, the software has enabled businesses to understand customer satisfaction levels, identify areas for improvement, and make data-driven decisions. The key findings and applications of the software include:

1. Sentiment Distribution: The analysis revealed the distribution of positive, negative, and neutral sentiments among customers, providing an overall understanding of their perceptions.

2. Overall Sentiment: The average sentiment score indicated an overall positive sentiment, indicating a high level of customer satisfaction with the product.

3. Categorical Analysis: By performing sentiment analysis across different product categories, businesses gained insights into varying sentiments expressed by customers for specific categories. This information can guide product development and marketing strategies.

4. Visualizations and Graphs: The software generated informative and visually appealing graphs, such as bar charts, word clouds, and trend graphs, to effectively represent sentiment analysis results.

The applications of this software extend beyond e-commerce analysis. It can be used for brand monitoring, social media analytics, market research, customer feedback analysis, and public opinion analysis. By leveraging sentiment analysis, businesses can make data-driven decisions, improve customer satisfaction, and enhance their brand reputation.Overall, the sentiment analysis project has proven to be a valuable tool for understanding customer sentiments, identifying key trends, and making informed business decisions. It demonstrates the effectiveness of combining web scraping, the VADER classifier, and graph visualization techniques to gain actionable insights from customer feedback and sentiments.

11. BIBLOGRAPHY

➢ JURE LESKOVEC and ROK SOSIC, Stanford Universityˇ, Stanford University "SNAP: A General-Purpose Network Analysis and Graph-Mining Library"[1]

➢ Lei Tang and Huan Liu Computer Science & Engineering Arizona State University L.Tang@asu.edu, Huan.Liu@asu.edu "GRAPH MINING APPLICATIONS TO SOCIAL NETWORK ANALYSIS"[2]

➢ Andry Alamsyah1, Budi Rahardjo2, Kuspriyanto2 "Social Network Analysis Taxonomy Based on Graph Representation"[3]

➢ Marco Brambilla 1,∗ , Alireza Javadian Sabet 2 , Kalyani Kharmale 3 and Amin Endah Sulistiawati "Graph-Based Conversation Analysis in Social Media"[4]

➢ Swati A. Bhavsar & Varsha H. Patil & Aboli H. Patil "Graph partitioning and visualization in graph mining: a survey"[5]

➢ Kassinda Francisco, Dr. Chandra J. "Sentiment Analysis on Social Media Data Using Intelligent Techniques." ,International Journal of Engineering Research and Technology. ISSN 0974-3154 (2019). [6]

➢ Muhammad Zidny Naf'an, Novanda Alim Setya Nugraha, " Sentiment Analysis of Cyberbullying on Instagram User Comments" (2019). [7]

➢ Vishal A. Kharde, S.S. Sonawane. " Sentiment Analysis of Twitter Data: A Survey of Techniques." International Journal of Computer Applications (0975 – 8887) (April 2016) [8]

➢ Shweta Gangrade, Nirvishesh Shrivastava, Jayesh Gangrade, "Instagram Sentiment Analysis: Opinion Mining." (2019) [9]

➢ Nikhat Parveen, Prasun Chakrabarti, Bui Hung and Amjan, " Twitter sentiment analysis using hybrid gated attention recurrent network." Journal of Big Data(2023)[10]

- Dorogush, Anna Veronika, Vasily Ershov, and Andrey Gulin. "CatBoost: gradient boosting with categorical features support." arXiv preprint arXiv:1810.11363 (2018).[11]

- Prokhorenkova, Liudmila, et al. "CatBoost: unbiased boosting with categorical features." Advances in neural information processing systems 31 (2018).[12]

- Kamal, Muhammad Ayoub, et al. "Highlight the features of AWS, GCP and Microsoft Azure that have an impact when choosing a cloud service provider." Int. J. Recent Technol. Eng 8.5 (2020): 4124-4232.[13]

- Varia, Jinesh. "Migrating your existing applications to the aws cloud." A Phase-driven Approach to Cloud Migration (2010): 1-23. [14]

- Yu, Gyeong-In, et al. "WindTunnel: towards differentiable ML pipelines beyond a single model." Proceedings of the VLDB Endowment 15.1 (2021): 11-20.[15]

- SENTIMENTAL ANALYSIS USING VADER. interpretation and classification of… | by Aditya Beri | Towards Data Science

- Welcome to VaderSentiment's documentation! — VaderSentiment 3.3.1 documentation

- Improving sentiment analysis accuracy with emoji embedding - ScienceDirect