Sentimental Analysis of

Customer Reviews

An Internship Project Report

Submitted to

DLITHE CONSULTANCY SERVICES PRIVATE LIMITED

By

RAKESH KUMAR

Under the guidance of

Ms. Medini B V

Robotics & Design Engineer

ACKNOWLEDGEMENT

I would like to extend my heartfelt gratitude to **DLithe Consultancy Services Private Limited** for providing us with the invaluable opportunity to undertake our internship at their esteemed institution. Their support and guidance were instrumental in shaping my project, and I am truly grateful for the experience.

I would also like to express our deepest thanks to **Ms. Medini**, **Robotics & Design Engineer** who served as an exceptional guide throughout this internship journey. Her expertise, encouragement, and mentorship were pivotal in making this experience rewarding and enriching.

Additionally, I am grateful to **Mite Institute of Technology & Engineering** for giving us the chance to pursue this internship at DLithe Consultancy Services Private Limited. The college's support and encouragement have been crucial in enabling us to gain practical industry experience and further my learning.

Finally, I want to acknowledge all the individuals who assisted me in various ways, providing valuable insights and contributing to our growth and knowledge during the internship period. Thank you all for your unwavering support.

ORGANIZATIONAL INFORMATION

DLithe is a technology-based product company that has been serving IT companies and academic institutions since the year 2018. The company is led by industry professionals with two decades of experience. For IT companies, DLithe offers services such as technology consultancy, project development, IT recruitment, staffing, competency development, and content development. On the other hand, the company serves academic institutions by providing competency development services in niche technologies like artificial intelligence, internet of things, robotics, cybersecurity, augmented reality, and more. DLithe has also developed the arm-based Cortex M3 series microcontroller and the ioCube product in the embedded and IoT domain.

During my enriching internship with the Artificial Intelligence and Machine Learning domain, I had the privilege of being a part of an exceptional program under the guidance of this renowned organization. Throughout the internship, I gained comprehensive insights into diverse industry verticals, spanning from understanding project requirements to the final deployment phase.

DLithe’s internship program provided me with a valuable opportunity to immerse myself in real-world scenarios, gaining exposure to industry best practices and learning how to implement AI and ML solutions within an agile project life cycle. The supportive environment and dedicated mentors at the organization ensured that I could explore practical use cases for AI and ML implementation, enabling me to grow and learn during insightful post-mentoring sessions.

Overall, this AI and ML internship has been a transformative experience, equipping us with not only technical skills but also a deeper understanding of how AI and ML technologies play a vital role across various industries.

CONTENTS

ABSTRACT

The "Sentimental Analysis of Customer Reviews " project focuses on leveraging Natural Language Processing (NLP) techniques, specifically the Vader sentiment analysis module, to extract valuable insights from customer reviews. In today's digital age, online reviews play a pivotal role in shaping consumer decisions and brand perception. Understanding the sentiment expressed in these reviews can provide businesses with actionable insights for product improvement, customer satisfaction enhancement, and overall brand management. This project begins by collecting a diverse dataset of customer reviews from various sources such as e-commerce websites, social media platforms, and review aggregators. We then preprocess and clean the text data to prepare it for analysis. The Vader sentiment analysis module, a pre-trained NLP tool, is employed to assign sentiment scores to each review. Vader not only identifies the polarity (positive, negative, or neutral) of the sentiment but also provides a compound score that represents the overall sentiment intensity. Furthermore, ethical considerations and data privacy are considered to protect user information and prevent potential misuse. The primary objective of this project is to offer a user-friendly interface where individuals or businesses can submit text data, and the application will analyze the sentiment of that text using the Vader SentimentIntensityAnalyzer module. Additionally, it discusses the challenges faced during the development process, such as handling ambiguous queries, maintaining context, and addressing potential biases in the training data. The Flask web application establishes a RESTful API endpoint ("/sentiment") that accepts POST requests containing text data to be analyzed. It employs preprocessing techniques to enhance the quality of the input text data. This includes converting the text to lowercase, removing digits, and eliminating common stopwords from the English language. Once the text is preprocessed, the Vader SentimentIntensityAnalyzer is employed to calculate sentiment scores, including positive, negative, neutral, and compound scores.

INTERNSHIP OBJECTIVES

The primary objectives of the AI and ML internship was designed to equip us with a comprehensive skill set and practical knowledge in various areas of Artificial Intelligence and Machine Learning. The key objectives included:

1. Learning Python Basics: The internship aimed to provide a strong foundation in Python programming, as it is one of the most widely used languages in AI and ML. Participants were introduced to Python syntax, data structures, and essential libraries used in AI and ML development.

2. Understanding ML Algorithms: The internship focused on making us understand fundamental ML algorithms such as Linear Regression, Binary Classification, and Decision Trees. These algorithms form the building blocks for more advanced techniques and are crucial for understanding the basics of supervised learning.

3. Exploring Neural Networks: We delved into the world of Neural Networks, understanding their architecture and how they mimic the human brain's functioning. Topics covered included Activation Functions and Forward Propagation, which are essential concepts for building and training neural networks.

4. Applied CNN on MNIST: Convolutional Neural Networks (CNNs) are widely used for image recognition tasks. The internship included hands-on experience in applying CNNs on the popular MNIST dataset for digit recognition, providing practical exposure to image classification.

5. Emphasizing GitHub and LinkedIn Profile Maintenance: The internship recognized the importance of a strong online presence for aspiring AI and ML professionals. We were encouraged to maintain an active GitHub repository showcasing our projects and contributions, as well as a well-curated LinkedIn profile to showcase our skills and accomplishments.

6. Real-World Implementation To bridge the gap between theory and real-world application, the internship featured project called “Sentimental Analysis of Customer Reviews”. We worked on this practical use case, applying AI and ML techniques to design an application that could be used to analyze the user review’s.

WEEKLY OVERVIEW OF INTERNSHIP ACTIVITIES

Week 1:

- Objective: Understanding Python Fundamentals for AI & ML

- Activities:

  - Covered Python syntax and data structures

  - Explored essential libraries used in AI and ML

  - Worked on basic Python programming exercises and projects

Week 2:

- Objective: Learning Various Machine Learning Algorithms and Implementation in Python

- Activities:

  - Hands-on implementation of Binary Classification algorithms

  - Explored Decision Trees and their practical applications

  - Worked on Linear Regression and its use cases

  - Applied CNN on MNIST dataset for image classification

  - Understood concepts of Forward Propagation and Neural Networks

Week 3:

- Objective: Use Case Selection, Data Collection, Preprocessing, and Algorithm Exploration

- Activities:

  - Gathered relevant data for the Sentimental Analysis such as different reviews dataset.

  - Explored various Web scraping tools to further improve data gathering.

  - Pre-processed the data to remove null values, extra queries or responses and further found ways to annotate data.

  - Discussed potential challenges and approaches with the mentor during weekly sessions

Week 4:

- Objective: Model Training, Testing, and Sentiment Analysis

- Activities:

  - Data Labeling: Manually or automatically label the data with the corresponding sentiment categories (e.g., positive, negative, neutral). This labeled data will be used for training and testing the sentiment analysis model.
  - Data Splitting: Divide the labeled data into training, validation, and test datasets. The training set is used to train the sentiment analysis model, the validation set helps fine-tune hyperparameters, and the test set evaluates the model's performance
  - Feature Extraction: Convert the preprocessed text data into numerical features that can be used for model training. Common techniques include TF-IDF (Term Frequency-Inverse Document Frequency) and word embeddings (e.g., Word2Vec, GloVe).
  - Model Selection: Choose a suitable sentiment analysis model, such as the VADER (Valence Aware Dictionary and sEntiment Reasoner) module, known for its effectiveness in analyzing sentiment in text data. VADER is a lexicon and rule-based model specifically designed for sentiment analysis.
  - Model Training: Train the selected model using the labeled training dataset. This involves iterative optimization of model parameters to achieve high accuracy.
  - Model Evaluation: Evaluate the trained model's performance using the test dataset. Calculate metrics such as accuracy, precision, recall, and F1-score to assess its effectiveness.
  - Sentiment Analysis: Apply the trained model to unlabeled or new customer reviews to automatically classify them into sentiment categories (positive, negative, neutral).

Key Learnings:

- Gained proficiency in Python Programming language.
- Acquired knowledge and experience in applying diverse ML Algorithms using Python.
- Understood criticality of data preparation and data refactoring for use in Natural Language Processing techniques.
- Gained practical experience in training and evaluating NLP models.
- Developed proficiency in Natural Language Processing type of data and processing for the same.

CHALLENGES AND LEARNING OUTCOMES

Challenges:

1. Data Gathering and Preprocessing: Cleaning and preparing the conversational data can be challenging, as it may contain noise, spelling errors, or inconsistencies in language use. Ensuring data quality is crucial for training an effective chatbot. Moreover, the lack of data may also lend to irregular responses being generated.

2. Ambiguity and Context: Understanding and maintaining context in conversations is difficult, especially when dealing with ambiguous queries or multi-turn interactions. Customer reviews often contain sarcasm, irony, or ambiguous sentiments that can be challenging to detect accurately.

3. Feature Engineering: Selecting and engineering relevant features from text data can be challenging, as different features may work better for different types of reviews.

4. Real-time Analysis: Implementing real-time sentiment analysis for live feedback can be technically challenging, as it requires low latency and scalability.

5. Unrelated User Input: Users may pose questions or provide text that do not align with the project's primary focus, which is sentiment analysis of customer reviews. These inputs could be general queries, unrelated topics, or requests for information outside the project's expertise.

Learning Outcomes:

NLP Fundamentals: The project will provide a solid understanding of the underlying principles of natural language processing, including tokenization, part-of-speech tagging, named entity recognition, and sentiment analysis.

Data Preprocessing Techniques: Dealing with real-world conversational data will sharpen skills in data preprocessing, data cleaning, and feature extraction to prepare data for training.

Sentiment Mechanisms: I learned about sentiment mechanisms and their role in improving the chatbot's ability to focus on relevant parts of the input text, making my chatbot more context-aware and effective to generate a good judgement of the sentiment of the text.

Deployment and Integration: Integrating the chatbot into different platforms or applications will provide practical experience in deploying NLP-based solutions in real-world scenarios.

Troubleshooting and Bug Fixing: Debugging and addressing issues that arise during development will hone problem-solving skills in an NLP context.

PROJECT DETAILS

CHAPTER 1

INTRODUCTION

The "Sentiment Analysis of Amazon Reviews" project is an endeavor to extract valuable insights from the vast amount of user-generated content in the form of Amazon product reviews. Amazon, being one of the world's largest online marketplaces, accumulates a substantial volume of reviews and feedback from its customers. This project aims to harness the power of Natural Language Processing (NLP) and sentiment analysis techniques to automatically evaluate and categorize the sentiments expressed within these reviews.

The primary goal is to provide businesses, sellers, and Amazon itself with a deeper understanding of customer sentiments. By analyzing these reviews, the project seeks to identify the underlying positive, negative, and neutral sentiments. These insights can be invaluable for product development, customer service improvements, and overall brand reputation management.

Ethical considerations are also addressed, promoting responsible AI usage and mitigating potential biases in the app’s responses. Data visualization techniques, such as word clouds and sentiment analysis, provide insights into frequently used words and the emotional tone of user inputs, enhancing the chatbot's overall effectiveness.

In this project, text data from Amazon reviews is collected, processed, and subjected to sentiment analysis. Through the use of sentiment analysis tools and techniques, the system will determine whether the reviews express satisfaction, dissatisfaction, or neutral opinions. The analysis results will be presented in a meaningful and actionable format, offering stakeholders a clearer perspective on customer sentiment, which, in turn, can guide decision-making and customer-focused strategies.

The "Sentiment Analysis of Amazon Reviews" project holds great potential to empower businesses and sellers on the Amazon platform to improve their products and services, enhance customer satisfaction, and make data-driven decisions. It aligns the power of NLP and sentiment analysis with the immense wealth of user-generated data on the Amazon platform.

LITERATURE SURVEY

[1] "**Sentimental Analysis based Review Classification in E-Commerce**" - The literature survey for sentiment analysis in e-commerce reveals a comprehensive understanding of sentiment analysis techniques and their applications. Key papers such as Liu's "A survey of sentiment analysis and opinion mining" and Cambria's "A survey of opinion mining and sentiment analysis" offer insights into fundamental sentiment analysis methods, challenges, and broader applications. Other studies, like Mudambi and Schuff's investigation into user preferences in e-commerce product reviews, provide valuable insights into user behavior. Aspect-based sentiment analysis models, such as the one proposed by Mei et al., cater to e-commerce's need for granular product feedback analysis. Moreover, integrating sentiment analysis with recommendation systems, as explored by Lu et al., demonstrates the potential for improving product recommendations in e-commerce. Overall, this literature survey is instrumental in shaping a comprehensive perspective on sentiment analysis in the context of e-commerce.

[2] **"Natural Language Processing for Sentiment Analysis An Exploratory Analysis on Tweets by Wei Yen Chong, Bhawani Selvaretnam, Lay-Ki Soon [2014]":** This paper explores the application of Natural Language Processing (NLP) techniques for sentiment analysis of tweets. The study aims to determine sentiment based on subjects in tweets and involves subjectivity classification, semantic association, and polarity classification. The paper highlights the differences between sentiment analysis in tweets and traditional text, such as the limited length of tweets and the unique data characteristics of Twitter. It discusses previous research in sentiment analysis and outlines how NLP techniques can be used to improve sentiment analysis accuracy, especially in the context of short tweets..

[3] **"Sentiment analysis for Amazon.com reviews by Levent G¨uner, Emilie Coyne, and Jim Smit. [March 1, 2019]":** The paper investigates the feasibility of applying sentiment analysis techniques to product reviews from Amazon.com. The study involves comparing, training, and testing various machine learning algorithms on a dataset of 60,000 product reviews randomly selected from a larger dataset of 4 million reviews available on Kaggle. The paper compares the performance of three different algorithms: Multinomial Naive Bayes (MNB), Linear Support Vector Machine (LSVM), and Long Short-Term Memory network (LSTM). The results show that the LSTM model performed the best. The paper then applies this LSTM model to the remaining 3.94 million reviews from the Kaggle dataset and a newly scraped dataset from Amazon.com, achieving very accurate sentiment classification, particularly for furniture products. The study concludes that LSTM networks are suitable for classifying sentiment in product reviews across different categories, and further research is needed to explore multi-class classification, including a neutral class.

[4] **"Unfair Reviews Detection on Amazon Reviews using Sentiment Analysis with Supervised Learning Techniques by Elshrif Ibrahim Elmurngi and Abdelouahed Gherbi [11-05-2018]":** This paper discusses the significance of reputation and trust in e-commerce and addresses the challenge of identifying unfair reviews in online consumer feedback systems. It utilizes Sentiment Analysis (SA) and compares the performance of four machine learning algorithms to classify reviews. The results show that Logistic Regression (LR) is the most

effective classifier, not only for sentiment analysis but also for detecting unfair reviews. The paper emphasizes the importance of credible reviews in e-commerce and the role of reputation systems in building trust among users.

PROBLEM STATEMENT

In today's digitally connected world, businesses are inundated with vast amounts of customer feedback and reviews across various online platforms. These customer reviews are a valuable source of information, providing insights into the customer experience and satisfaction. However, the sheer volume and unstructured nature of this data make it challenging for businesses to extract meaningful insights.

The problem at hand is the need to efficiently analyze and interpret these customer reviews to gain actionable insights. Traditional manual methods are time-consuming and prone to human error, making them inadequate for handling the scale and complexity of modern customer feedback. Businesses require an automated and accurate solution that can:
Classify Sentiment: Automatically categorize customer reviews into positive, negative, or neutral sentiments based on the language and tone used in the text.
Extract Key Insights: Identify common themes, issues, or trends within customer reviews, allowing businesses to address pain points and capitalize on strengths.
Monitor Brand Reputation: Continuously monitor online sentiment to gauge brand reputation and customer sentiment over time.
Improve Decision-Making: Provide decision-makers with data-driven insights to inform product development, marketing strategies, and customer support efforts.

To address these challenges, we propose the development of a sentiment analysis system that leverages natural language processing (NLP) and machine learning techniques. This system will process and analyze customer reviews in real-time, generating sentiment scores and actionable insights to empower businesses in making data-informed decisions.

The successful implementation of this project will enable businesses to enhance customer satisfaction, identify areas for improvement, and maintain a positive brand image in an increasingly competitive digital landscape

PROJECT OBJECTIVE

Specifically, the project aims to achieve the following goals:

- Automated Sentiment Classification:
  This objective focuses on building a machine learning or natural language processing model capable of automatically categorizing customer reviews into predefined sentiment categories. The primary sentiment categories are typically "positive," "negative," and "neutral." By achieving this objective, you can save time and resources that would otherwise be spent manually reading and categorizing each review.

- Accurate Sentiment Analysis:
  Accuracy is a critical aspect of sentiment analysis. The objective is to ensure that the model correctly categorizes customer feedback. Achieving a high level of accuracy is essential to provide valuable insights into how customers perceive a product or service. This involves minimizing false positives and false negatives to provide a reliable assessment of sentiment.

- Performance Metrics:
  Defining and measuring key performance metrics is essential for evaluating the effectiveness of the sentiment analysis model. These metrics include:

  - Accuracy: The proportion of correctly categorized reviews.
  - Precision: The ratio of true positive predictions to the total predicted positives, which measures the model's ability to avoid false positives.
  - Recall: The ratio of true positive predictions to the total actual positives, which measures the model's ability to avoid false negatives.

The successful completion of this project will result in a user-friendly sentimental analysis app, capable of providing valuable information and assistance to users about the product reviews, enhancing their overall experience.

METHODOLOGIES

1. **Data Collection and Preprocessing**: In this project, data collection involves obtaining a dataset of customer reviews. These reviews could be from e-commerce platforms, social media, or any other source. Preprocessing includes cleaning the text data by removing special characters and non-ASCII characters. It also involves converting text to lowercase to ensure uniformity and handling missing or irrelevant data points, ensuring data quality.

2. **Exploratory Data Analysis (EDA):** EDA helps in understanding the dataset. For this project, it involves exploring the distribution of customer reviews. You might analyze the length of reviews, the frequency of positive and negative sentiments, and common keywords. Visualizations like word clouds can help identify frequently occurring words and phrases, while sentiment analysis can provide insights into user emotions and preferences within the dataset.

3. **Sentiment Analysis:** In this project, the sentiment analysis part entails using NLTK's SentimentIntensityAnalyzer to evaluate the sentiment of customer reviews. It assigns sentiment scores (positive, negative, or neutral) to understand the overall sentiments expressed in the reviews.

4. **Text Normalization**: Text normalization techniques, such as lemmatization and tokenization, are applied to ensure that the text data is consistent and standardized. This step helps in improving the quality of the text input.

5. **Feature Extraction - TF-IDF:** TF-IDF vectorization is used to convert the pre-processed text into numerical feature vectors. These vectors represent the importance of words or phrases within the reviews, allowing the model to compare and find similar matches for user queries.

6. **Building the Model**: The TF-IDF feature vectors are used to train a machine learning model. In this case, a cosine similarity model might be employed. This model helps find the most similar matches between user queries and the dataset. It is crucial for retrieving relevant responses.

7. **Evaluation and Metrics:** Evaluation metrics like perplexity, confidence scores, or similarity scores are used to assess the performance of the sentiment analyzer. These metrics measure the effectiveness of the system in providing relevant and appropriate responses to user queries.

8. **Real-time Interaction**: Optimizing response generation for low latency is vital for maintaining real-time conversational experiences and user engagement. It ensures that the system can provide quick responses, enhancing the user experience.

9. **Testing:** Extensive testing is conducted to validate that the sentiment analyzer meets the project objectives and desired functionalities. Various user scenarios are considered to

ensure that the system performs reliably.

By following this methodology, the project aims to build a sophisticated and query-aware conversational sentiment analyser that can provide sentiment scores (positive, negative, neutral) interactions, enhancing user engagement and satisfaction.

CHAPTER 6

IMPLEMENTATION

Introduction

In this project, a sentiment analysis system is implemented using TF-IDF word embeddings and cosine similarity. The system's primary goal is to analyze and categorize customer reviews into positive, negative, or neutral sentiments.The sentiment analysis in this project is based on the NLTK library, utilizing the SentimentIntensityAnalyzer and the VADER sentiment analysis model.

Data Understanding

The project's dataset contains customer reviews, which serve as the foundation for training and testing the sentiment analyzer.

Data Visualization

Sentiment analysis is performed on the reviews using the VADER tool, allowing the project to understand the emotional context of the reviews. Word clouds are generated to visualize the most frequent words and phrases, aiding in identifying common topics and sentiments in the dataset.

Text Normalization

Text normalization techniques are applied to both customer reviews. These techniques ensure that the text data is clean and consistent. The normalization steps include:

- Lowercasing to maintain uniformity.
- Removal of special characters and punctuations for cleaner text.
- Elimination of non-ASCII characters for improved compatibility.
- Elimination of alphanumeric characters for better analysis.

Sentiment Analysis and Important Sentences:

Sentiment analysis is performed on the reviews, and the system identifies important sentences based on their sentiment scores.

The sentences are ranked based on their compound, positive, negative, and neutral scores, providing insights into the most emotionally charged interactions in the dataset.

CHAPTER 7

RESULT AND FUTURE SCOPE

The application allows users to input text, analyze its sentiment, and receive the sentiment score and category (positive, negative, or neutral) as output. Here are the expected results and the future scope for this project:

Result:

User-Friendly Sentiment Analysis: The Streamlit application provides an easy-to-use interface for users to input text and instantly analyze the sentiment. Users receive a sentiment category (positive, negative, or neutral) and sentiment scores (positive, negative, neutral, and compound) for the input text.

Visualization: The application displays an image and uses a simple, clean layout to present the sentiment analysis results, making it easy for users to understand the sentiment of the text they input.

VADER Sentiment Analysis: The project utilizes the VADER sentiment analysis tool, which is pre-trained and well-suited for analyzing sentiment in text data.

Future Scope:

Data Integration: Extend the project to handle real-time data from various sources, such as social media, product reviews, or customer feedback. This will enable businesses to monitor and analyze sentiment as it evolves over time.

Custom Sentiment Analysis Models: While VADER is a good starting point, more advanced sentiment analysis models, including machine learning-based models, can be incorporated to enhance accuracy, especially for specific domains.

Topic-Based Sentiment Analysis: Enhance the application to perform sentiment analysis based on specific topics or aspects within the text. This will allow for a more granular understanding of sentiment in reviews or social media posts.

Scalability: Optimize the application for scalability to handle large volumes of data and real-time sentiment analysis.

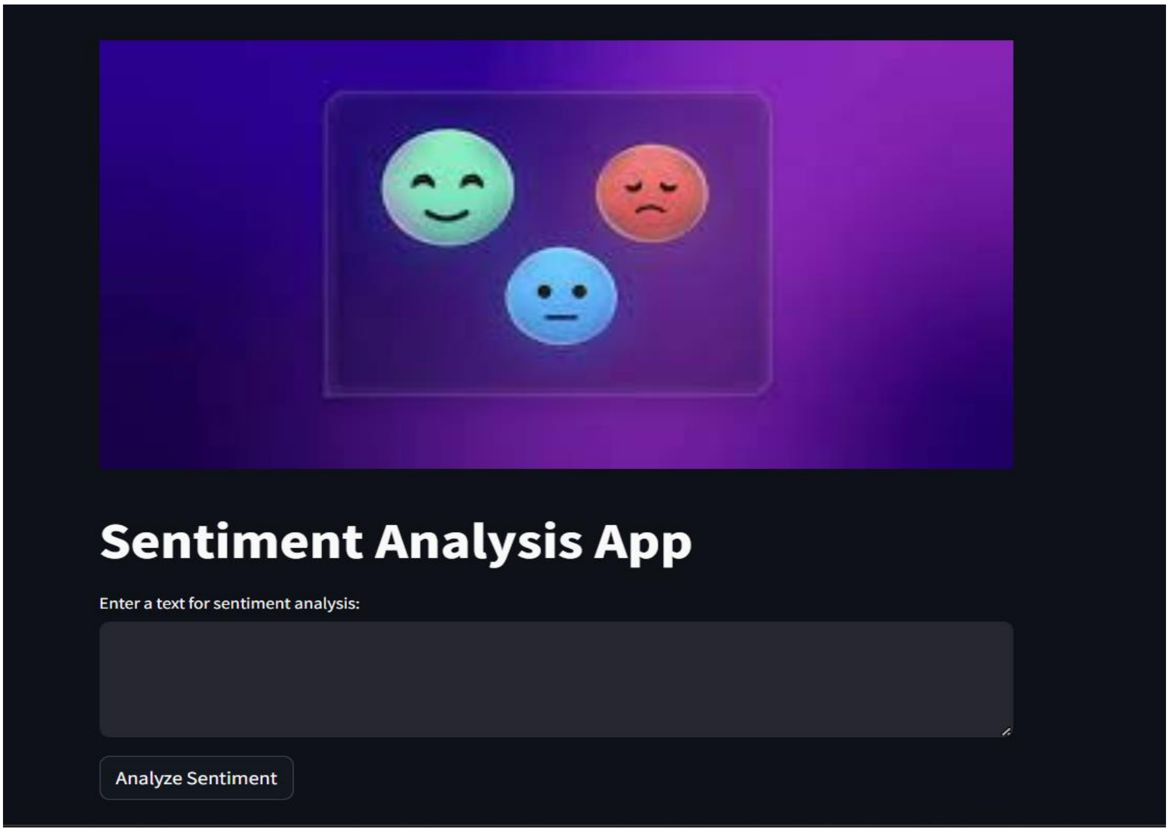

Figure 1.Cover Page

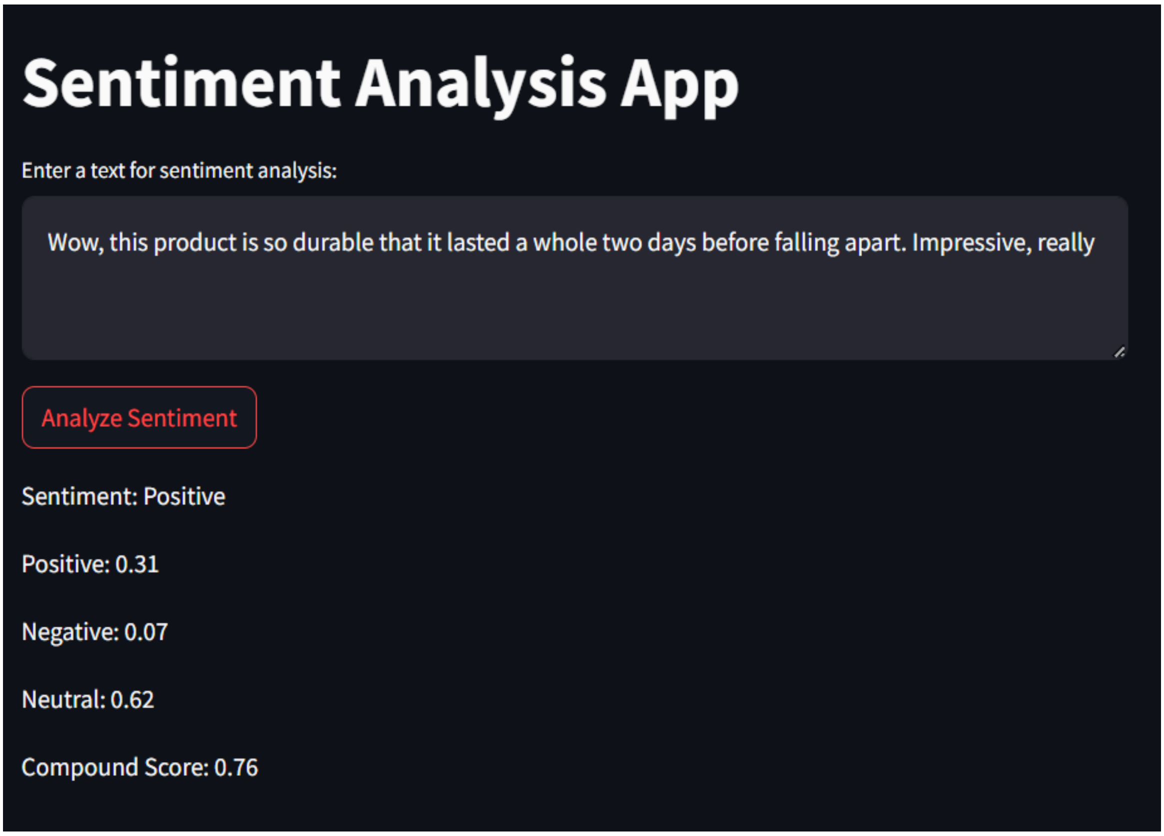

Figure 2. Positive Input Statement

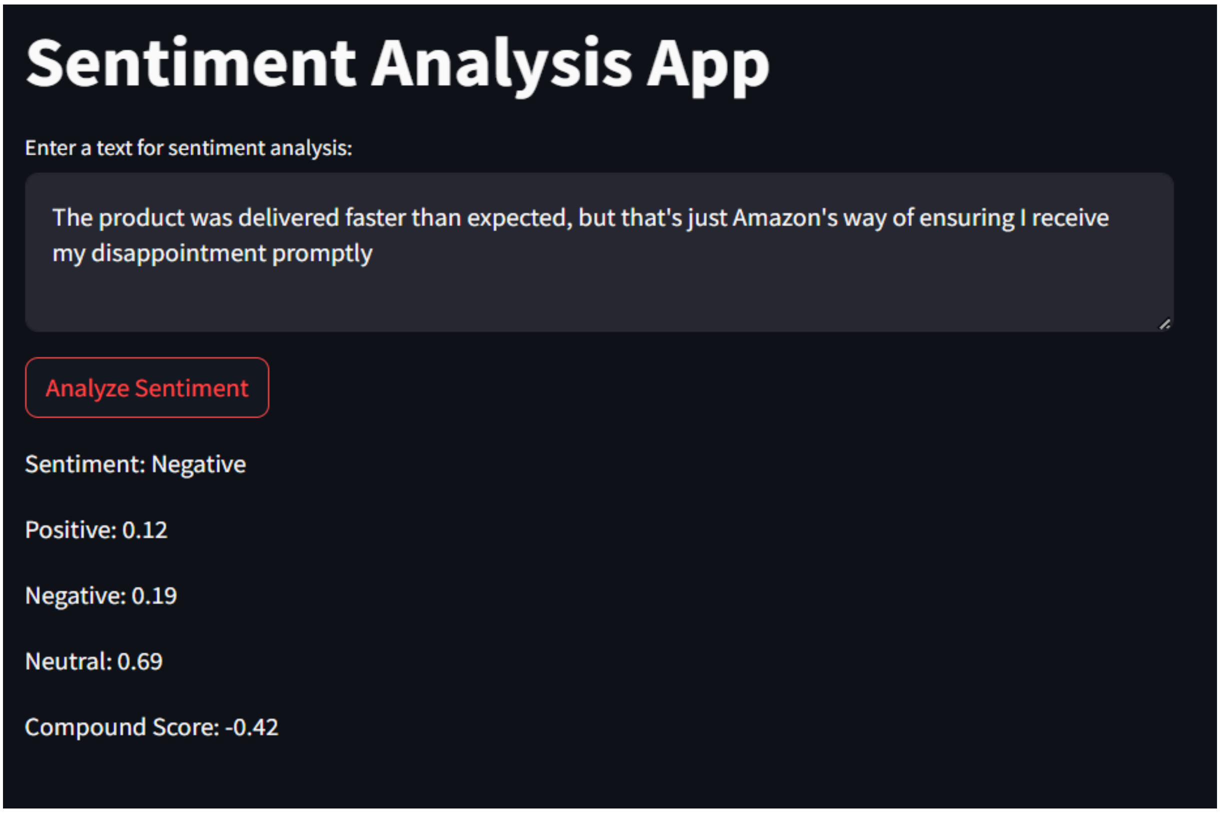

Figure 3. Negative Input Statement

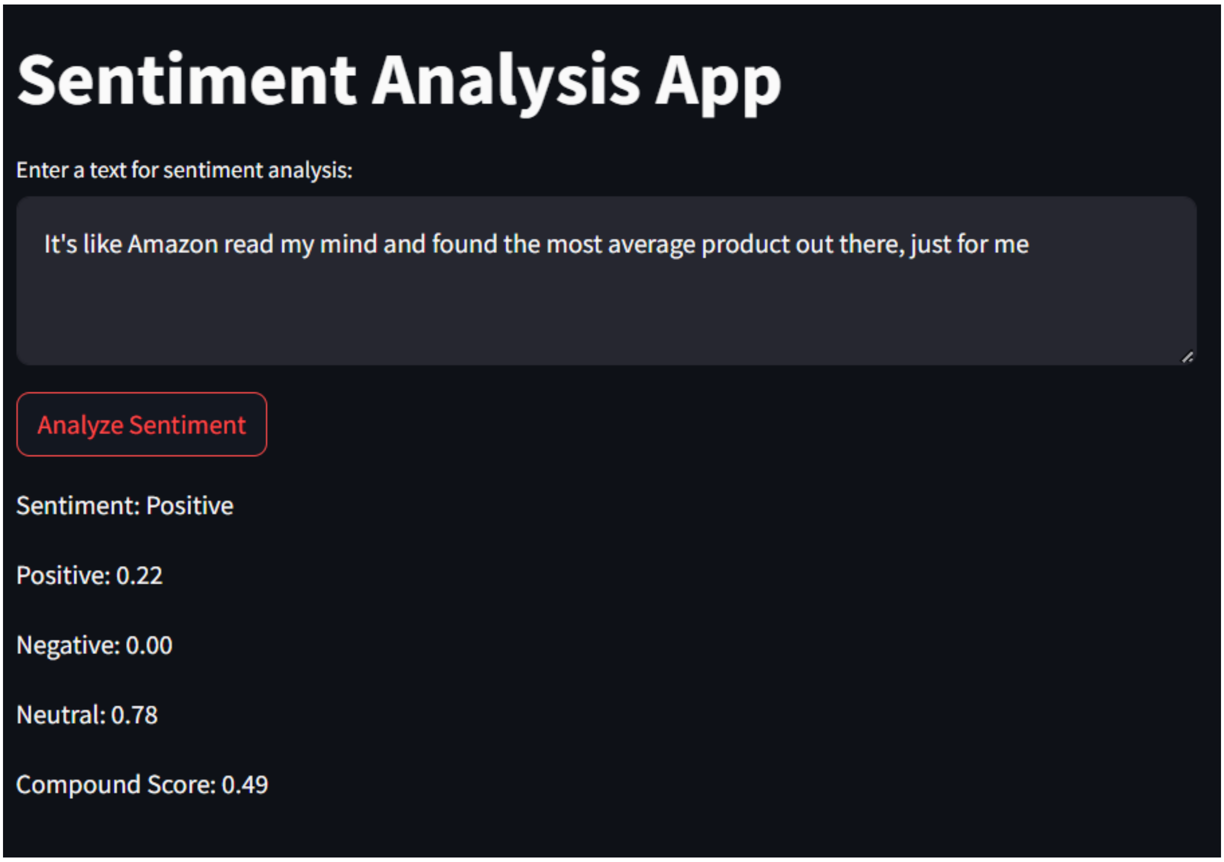

Figure 4.Neutral Input String

APPENDIX

PROJECT CODE:

1.app.py:

- Description: This Python script contains the code for a Streamlit web application for sentiment analysis. It allows users to input text and performs sentiment analysis using the VADER Sentiment Intensity Analyzer. The app displays the sentiment (positive, negative, or neutral) along with sentiment scores.
- Components:
  - ✓ Import necessary libraries, including Streamlit and NLTK's SentimentIntensityAnalyzer.
  - ✓ Download necessary NLTK resources (VADER lexicon).
  - ✓ Initialize the SentimentIntensityAnalyzer (SIA) to analyse sentiment.
  - ✓ Define a function analyze_sentiment(text) that takes text input, analyses sentiment, interprets sentiment scores, and returns the sentiment label and scores.
  - ✓ Create the Streamlit user interface (UI):
    - Set the app title as "Sentiment Analysis App."
    - Add a text area where users can enter text for sentiment analysis.
    - Include a "Analyse Sentiment" button.
  - ✓ When the "Analyse Sentiment" button is clicked:
    - Check if the user has entered text.
    - Analyse sentiment using the analyze_sentiment function.
    - Display the sentiment label, positive score, negative score, neutral score, and compound score in the Streamlit app.
    - Provide a warning if no text is entered.

2. Sentiment-analysis.py :

- Data Collection: Obtain Amazon reviews data from a suitable source, such as Amazon's API, web scraping, or a pre-compiled dataset.
- Data Preprocessing: Clean the text data by removing special characters, HTML tags, and irrelevant information. Tokenize the text into words or phrases. Remove stop words and perform stemming or lemmatization to standardize text.
- Sentiment Lexicon: Use a sentiment lexicon or dictionary like VADER (Valence Aware Dictionary and sentiment Reasoner) to score the sentiment of the text. Alternatively, you can train your own sentiment analysis model.

3. Sentiment Scoring: The SentimentIntensityAnalyzer in VADER assigns sentiment scores to each piece of text. These scores typically include a compound score (indicating the overall sentiment), positive score, negative score, and neutral score.

4. Data Preprocessing: Before using VADER, you should pre-process your text data to clean and prepare it. This might include removing special characters, stemming or lemmatizing words, and transforming the text into lowercase.

5. Scoring and Labelling: Apply the SentimentIntensityAnalyzer to your text data, and it will assign sentiment scores and labels to each text entry. For example, a text entry might be labelled as "Positive" with a positive score of 0.5, "Negative" with a negative score of -0.3, or "Neutral" with a neutral score close to 0.

6. Data Visualization: Once you have sentiment scores and labels, you can use data visualization techniques to explore and communicate the sentiment patterns in your data. Here's how VADER and SentimentIntensityAnalyzer can be used in data visualization:

- Bar Charts: Create bar charts to visualize the distribution of sentiment labels (e.g., Positive, Negative, Neutral) in your dataset. This provides an overview of the sentiment balance.
- Time Series Plots: If your data has a temporal aspect (e.g., tweets over time), you can create time series plots to show how sentiment changes over different time periods.
- Word Clouds: Generate word clouds to display the most frequent words associated with positive, negative, or neutral sentiments. This helps identify key themes and keywords related to sentiment.
- Histograms: Use histograms to visualize the distribution of sentiment scores. This can provide insights into the overall sentiment tendency in the dataset.
- Box Plots: Box plots can help show the spread of sentiment scores and identify outliers.
- Heatmaps: Create heatmaps to visualize sentiment scores in a grid format, which can be useful for identifying patterns or correlations.
- Correlation Plots: Investigate correlations between sentiment scores and other variables in your dataset, such as user demographics or product attributes.

For access to the complete code and detailed implementation, please find the project's GitHub repository at the following link:

https://github.com/rakesh-mijar/aiml_intern_dlithe/blob/main/Final_Sentimental_Analysis.py

BIBLIOGRAPHY

1. S. Lakshmi. (2018). Sentimental Analysis based Review Classification in E-Commerce. International Journal for Research in Applied Science and Engineering Technology, 6, 825–828. doi:10.22214/ijraset.2018.4139

2. Chong, Wei & Selvaretnam, Bhawani & Soon, Lay-Ki. (2014). Natural Language Processing for Sentiment Analysis: An Exploratory Analysis on Tweets. 212-217. 10.1109/ICAIET.2014.43. Evaluating Chatbots: A Systematic Review.

3. Coyne, Emilie & Smit, Jim & Güner, Levent. (2019). Sentiment analysis for Amazon.com reviews. 10.13140/RG.2.2.13939.37920.

4. lmurngi, E. I. & Gherbi, A. (2018). Unfair Reviews Detection on Amazon Reviews using Sentiment Analysis with Supervised Learning Techniques. Journal of Computer Science, 14(5), 714-726. https://doi.org/10.3844/jcssp.2018.714.726

5. Maitama, J.Z., Idris, N., Abdi, A., Shuib, L. and Fauzi, R., 2020. A systematic review on implicit and explicit aspect extraction in sentiment analysis. IEEE Access, 8, pp.194166 194191.