ECOMMERCE REVIEWS SENTIMENT ANALYSIS BOT

A PROJECT REPORT

Submitted by

HARINI A (220701084)

in partial fulfilment for the course

OAI1903 - INTRODUCTION TO ROBOTIC PROCESS AUTOMATION

for the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

RAJALAKSHMI NAGAR

THANDALAM , CHENNAI – 602 105

NOVEMBER 2024

RAJALAKSHMI ENGINEERING COLLEGE CHENNAI - 602105

BONAFIDE CERTIFICATE

Certified that this project report "ECOMMERCE REVIEWS SENTIMENT ANALYSIS" is the Bonafede work of "HARINI A (220701084)" who carried out the project work for the subject OAI1903-Introduction to Robotic Process Automation under my supervision.

Mrs. J. Jinu Sophia

SUPERVISOR

Assistant Professor (SG)

Department of

Computer Science and Engineering

Rajalakshmi Engineering college

Rajalakshmi Nagar

Thandalam

Chennai-602 105

Submitted to Project and Viva Voce Examination for the subject OAI1903- Introduction to Robotic Process Automation held on _.

TABLE OF CONTENTS

CHAPTER NO.	A DC		GE NO
	ABSTRACT LIST OF TABLES		5 6
	LIST OF TABLES LIST OF FIGURES		8
	LIST OF ABBREVIATIONS		9
1.	INTRODUCTION		12
	1.1	GENERAL	
	1.2	PROBLEM STATEMENT	
	1.3	OBJECTIVE	
2.	LITI	ERATURE REVIEW	15
	2.1	INTRODUCTION TO SENTIMENT ANALYSIS	S
	2.2	TOOLS & TECHNOLOGIES	
	2.3	EXISTING RESEARCH & STUDIES	
	2.4	KEY INSIGHTS FROM LITERATURE	
3.	SYS	ΓEM DESIGN	20
	3.1	GENERAL	
		3.1.1 SYSTEM FLOW DIAGRAM	
		3.1.2 SEQUENCE DIAGRAM	
4.	CONCLUSIONS		29
	4.1	GENERAL	
	APPENDICES-SCREENSHOTS OF OUTPUT		
5.	REF	ERENCES	35

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman Thiru. S.Meganathan, B.E., F.I.E., our Vice Chairman Mr. M.Abhay Shankar, B.E., M.S., and our respected Chairperson Dr. (Mrs.) Thangam Meganathan, M.A., M.Phil., Ph.D., for providing us with the requisite infrastructure and sincere endeavouring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N.Murugesan**, **M.E.**, **Ph.D.**, our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P.Kumar**, **M.E.**, **Ph.D.**, Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our scincere deepest gratitude to our interna guides, **Mrs. J. Jinu Sophia**, **M.E.**, **(Ph.D)**, Assistant Professor(SG), Department of Computer Science and Engineering, Rajalakshmi Engineering College for the valuable guidance throughout the course of the project. We are very glad to thank our Project Coordinators.

Dr. N. Durai Murugan, M.E., Ph.D., Associate Professor, and

Mr. B. Bhuvaneswaran, M.E., Assistant Professor (SG), Department of Computer Science and Engineering for their useful tips during our review to build our project.

ABSTRACT

Processing (NLP) that plays a vital role in extracting insights from textual data. It involves classifying text into distinct emotional categories, such as positive, negative, and neutral. The increasing volume of textual data in various forms—such as social media posts, customer feedback, and product reviews—has made sentiment analysis an essential tool for businesses and organizations. However, manual sentiment categorization is time-consuming, error-prone, and often impractical for large datasets. This project seeks to address these challenges by developing an automated sentiment categorization system using UiPath, a leading Robotic Process Automation (RPA) platform.

The objective of this project is to automate the process of sentiment categorization by classifying text into three distinct categories: **Happy**, **Sad**, and **Neutral**. The system aims to provide businesses and individuals with accurate, efficient, and scalable solutions for sentiment analysis, reducing the time and effort required to process and analyze large amounts of text. By integrating **UiPath** with sentiment analysis techniques, the automation system processes raw text, classifies it according to predefined sentiment categories, and outputs the results into an **Excel file** for easy analysis and further use.

The core methodology involves using RPA to automate the collection and preprocessing of text data from multiple sources. This data is then processed by a **sentiment analysis model**, which assigns sentiment labels to the text. The automation workflows built in UiPath ensure that the data is processed without human intervention, significantly improving the speed and accuracy of the sentiment categorization process. The final

output—an Excel file—contains categorized data, which can be used for further analysis, trend identification, and decision-making.

This project also aims to showcase the integration of **UiPath** with sentiment analysis algorithms, demonstrating the power of **Robotic Process Automation** to streamline business operations and data analysis. It

The results of this project demonstrate the feasibility and effectiveness of automating sentiment analysis using **UiPath**. The solution developed can handle large datasets, categorize sentiments accurately, and provide a reliable output for business intelligence applications. Future work may involve enhancing the model's accuracy with more advanced machine learning techniques, integrating additional sentiment categories, and extending the solution to handle multi-language text.

In conclusion, this project successfully automates the sentiment categorization process, offering an efficient and scalable solution for businesses to analyze text data quickly and accurately. By integrating **UiPath** with **Natural Language Processing**, the project showcases the power of **Robotic Process Automation** in improving the efficiency of business operations and data analysis, laying the groundwork for future developments in this field.

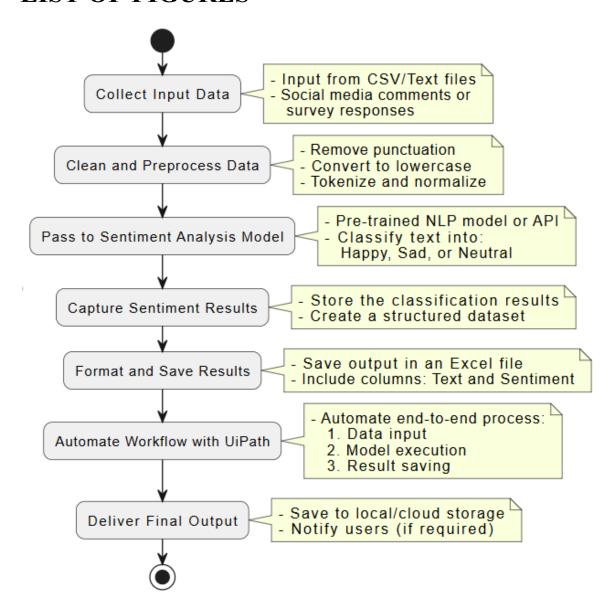
LIST OF TABLES

Field Name	Data Type	Description
productreviews.xlsx	xlsx	The "productreviews.xlsx"
		contains the reviews of
		amazon website.
Sentiments	xlsx	The "Sentiments.xlsx"
		Contains the empty sheet
		with heading value as
		"UNIQID ,TIME,
		REVIEW TITLE"
		Which helps to categorize
		as happy, sad and neutral
		reviews.

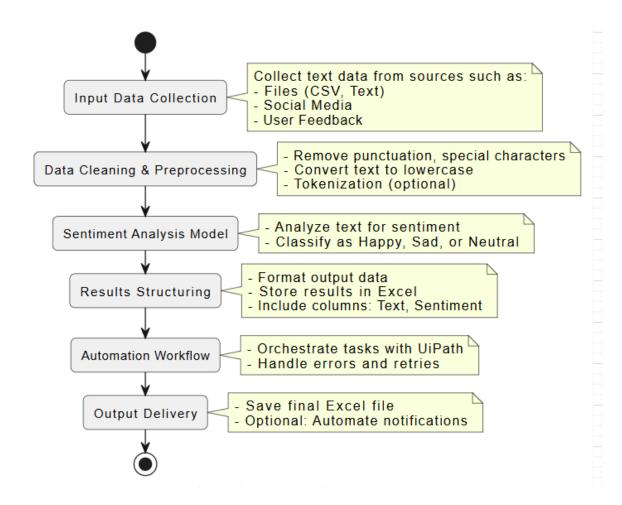
DESCRIPTION:

- Represents the data and contains details about the product reviews of real time data of amazon website available from kaggle.
- Represents and storage of input data into Sentiments.xlsx file. Contains information where all the reviews are analysed and categorized as 'happy','sad','neutral' reviews in the separate respective sheets.

LIST OF FIGURES



WORKFLOW OF SENTIMENT ANALYSIS OF ECOMMERCE REVIEWS



DATA PIPELINE FLOW OF SENTIMENT ANALYSIS

LIST OF DETAILS ABOUT SYMBOLS, ABBREVIATIONS AND

NOMENCLATURE

■ NLP: Natural Language Processing

RPA: Robotic Process Automation

API: Application Programming Interface

CSV: Comma-Separated Values

ML: Machine Learning

UI: User Interface

JSON: JavaScript Object Notation

HTTP: Hypertext Transfer Protocol

OCR: Optical Character Recognition

XAML: Extensible Application Markup Language

NLP(Natural Language Processing):

A branch of Artificial Intelligence that focuses on enabling machines to understand, interpret, and process human language. In this project, NLP techniques are used for analyzing text data and identifying sentiments.

API (Application Programming Interface):

A set of protocols and tools that allow different software applications to communicate. In this project, APIs might be used to integrate external sentiment analysis models or services into the UiPath workflow.

RPA (Robotic ProcessAutomation):

A technology for automating repetitive, rule-based tasks using software robots. UiPath, as an RPA tool, orchestrates the automation pipeline in this sentiment analysis project.

Sentiment Categories:

- **Happy:** Represents positive sentiment or emotions.
- **Sad:** Represents negative sentiment or emotions.
- **Neutral:** Represents text that expresses neither positive nor negative sentiments.

Workflow:

The sequence of steps or processes automated using UiPath for data preprocessing, analysis, and output generation.

Excel Output Format:

The structured format in which the sentiment results are stored, including columns for original text and the corresponding sentiment label.

Error Handling:

Strategies used within the automation workflow to manage unexpected failures, ensuring robustness and reliability.

Data Preprocessing:

The cleaning and preparation of raw text data by removing irrelevant elements like special characters, normalizing text, and converting it to a format suitable for analysis.

Chapter 1: INTRODUCTION

1.1 GENERAL

Sentiment analysis, also known as opinion mining, is an essential technique in Natural Language Processing (NLP) that enables the classification of text into various emotional categories such as positive, negative, and neutral. In recent years, sentiment analysis has become indispensable for businesses, especially in customer feedback management, brand monitoring, and social media analytics. It helps in gaining valuable insights into customer opinions, which can drive product development, marketing strategies, and customer relationship management.

This project focuses on automating the sentiment categorization process, particularly for large datasets, using UiPath, a popular Robotic Process Automation (RPA) tool. UiPath's ability to automate data handling, including extraction, processing, and categorization, makes it an ideal platform for developing efficient and scalable sentiment analysis solutions. The automation system developed in this project streamlines the traditionally labor-intensive process, enabling businesses and individuals to categorize sentiment from text data accurately and efficiently.

In this approach, the key objective is to integrate NLP techniques with UiPath's automation workflows to categorize text into three emotional classes: Happy, Sad, and Neutral. This project also integrates the output into an Excel file, making it easy for users to view, analyze, and take action on sentiment-driven insights.

1.2 Problem Statement

One of the significant challenges in sentiment analysis is the manual classification of large volumes of text data. Businesses and organizations dealing with feedback from customers, social media posts, product reviews, and surveys often find themselves overwhelmed by the sheer amount of data that requires categorization. The manual process is not only labor-intensive but also prone to human error, making it difficult to maintain accuracy and consistency across large datasets.

For instance, customer feedback in the form of text might include a mixture of complex sentences, idioms, and emotional tones that are challenging to categorize manually. As businesses increasingly rely on data-driven decision-making, there is a pressing need to automate sentiment analysis to provide faster, more accurate insights. Automating this process can also help businesses save time, reduce costs, and improve the overall customer experience by quickly addressing sentiments identified through text analysis.

The project addresses this problem by automating the sentiment classification process using UiPath. The solution will provide an efficient, scalable, and error-free method to process large volumes of text and categorize the data into the desired sentiment labels. This will ultimately help businesses and individuals in extracting actionable insights from unstructured data with minimal effort.

1.3 Objectives

The main objectives of this project are:

- Develop a Sentiment Classification System: To create a system that categorizes text data into three primary sentiment categories: Happy, Sad, and Neutral. The system will use natural language processing (NLP) techniques to accurately classify textual data based on the emotions conveyed.
- 2. Integrate UiPath Automation: To utilize UiPath as an automation tool for processing and categorizing text data. UiPath will help automate the data extraction from various sources, data preprocessing, sentiment categorization, and report generation. The system will be able to handle large datasets without manual intervention.
- 3. Generate Categorized Outputs in Excel: To ensure the categorized data is outputted in a user-friendly format. The system will create an Excel file where each piece of text is labeled with its corresponding sentiment. This will enable businesses or individuals to analyze the data further and make informed decisions based on sentiment trends.

In summary, this project seeks to automate sentiment analysis and provide categorized outputs that can be easily interpreted and used for business intelligence, customer service improvement, and various other applications.

Chapter 2: LITERATURE REVIEW

2.1 Introduction to Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a Natural Language Processing (NLP) technique used to determine the sentiment expressed in a text, categorizing it into different emotions such as positive, negative, or neutral. Over recent years, sentiment analysis has gained significant attention, especially in the fields of social media analysis, customer feedback processing, and brand management. It plays a crucial role in understanding public opinion, consumer sentiments, and emotional trends in large datasets of text.

Early sentiment analysis systems relied heavily on manual classification or rule-based systems that required predefined sets of keywords or lexicons. However, advancements in machine learning and deep learning techniques have made sentiment analysis more efficient and scalable. Today, a combination of supervised learning, unsupervised learning, and deep learning algorithms are employed to accurately classify sentiments in text.

2.2 Overview of Sentiment Categorization Techniques

Sentiment analysis can be categorized into several techniques based on the complexity and approach used. These include:

- Rule-Based Sentiment Analysis: In rule-based sentiment analysis,
 predefined rules or lexicons are used to determine the sentiment of
 a text. Sentiment lexicons, such as SentiWordNet, are employed to
 assign sentiment scores to words, and the overall sentiment is
 calculated based on these scores. However, this method may not be
 highly accurate in capturing context, irony, or sarcasm.
- Machine Learning-Based Sentiment Analysis: This technique involves training machine learning models using labeled data. Common algorithms used in sentiment analysis include Naive Bayes, Support Vector Machines (SVM), and Random Forests. These models learn from historical data to classify new, unseen texts into sentiment categories. Despite its effectiveness, it requires a large labeled dataset to train the model and is computationally intensive.
- Deep Learning-Based Sentiment Analysis: Deep learning techniques, particularly Recurrent Neural Networks (RNNs) and Short-Term Memory (LSTMs), Long Networks have revolutionized sentiment analysis capturing by complex relationships in text. Convolutional Neural Networks (CNNs) and Transformer models like BERT have also been employed for sentiment classification, offering more nuanced understanding and better performance on large datasets.

2.3 Tools and Technologies for Sentiment Analysis

Several tools and frameworks have been developed to aid in sentiment analysis, with varying features based on ease of use, performance, and scalability. Some of the most popular tools include:

- TextBlob: A Python library for processing textual data, TextBlob provides a simple API for performing sentiment analysis using pre-trained models. It categorizes text into positive, negative, and neutral sentiments based on a polarity score.
- VADER (Valence Aware Dictionary and sEntiment Reasoner):
 VADER is a lexicon and rule-based sentiment analysis tool optimized for social media text. It considers not only the words but also the context in which they are used, handling emoticons, slang, and informal language more effectively.
- Azure Text Analytics API: A cloud-based tool provided by Microsoft Azure, this API offers pre-built sentiment analysis capabilities that are robust and scalable. The API uses advanced machine learning models to evaluate the sentiment of the text and categorize it as positive, negative, or neutral. Azure's Text Analytics API provides significant advantages in terms of accuracy and ease of integration, making it a popular choice for enterprises looking to analyze customer feedback, product reviews, and social media comments.

IBM Watson NLP: Another powerful sentiment analysis tool, IBM
Watson offers advanced natural language processing services,
including sentiment analysis, emotion detection, and text
classification. It provides deep insights into the sentiments
expressed in large volumes of text, particularly useful for customer
service applications.

2.4 Existing Research and Studies

A number of studies have explored the effectiveness of sentiment analysis in various domains, ranging from product review analysis to financial forecasting. A key study by Pang and Lee (2008) analyzed various techniques for sentiment classification, comparing machine learning algorithms such as Naive Bayes and SVMs to rule-based approaches, finding that machine learning algorithms significantly outperformed the rule-based systems in terms of accuracy.

In social media analysis, sentiment analysis has been used extensively to monitor public opinion during events like **political elections** or **brand crises**. A study by Go, Bhayani, and Huang (2009) used Twitter data to analyze public sentiment about political candidates, leveraging machine learning to achieve a high level of accuracy in sentiment categorization.

Moreover, the application of deep learning models has been a breakthrough in sentiment analysis, as they can capture complex contextual relationships in text. Studies such as **Zhang et al. (2018)** highlight the use of LSTMs for more accurate sentiment classification, especially for long and complex text, such as customer service dialogues or product reviews.

2.5 Key Insights from Literature

• Lexicon-based approaches are often not enough on their own to

capture nuanced sentiments, particularly in short texts, sarcasm, or

context-dependent statements.

Machine learning models require substantial labeled datasets and

computational resources. but they offer significant can

improvements in accuracy when compared to traditional methods.

• Deep learning techniques, specifically LSTMs and BERT-based

models, have proven to be the most effective for analyzing

complex sentiments due to their ability to understand context and

sequence.

The integration of cloud-based APIs, such as Azure Text Analytics,

simplifies the implementation of sentiment analysis by offering

pre-built models, ensuring higher accuracy and reducing the need

for large-scale data training.

CHAPTER 3: SYSTEM DESIGN

Input Layer

This layer handles the collection of raw data from various sources,

such as:

19

• **Files**: CSV or text files containing user feedback, survey responses, or social media comments.

• **User Input**: Direct input through an interface.

The raw text data serves as the foundation for further processing.

Preprocessing Layer

This layer prepares the raw data for sentiment analysis. Key tasks include:

• **Data Cleaning**: Removing unwanted characters, punctuation, and extra spaces.

 Normalization: Converting text to lowercase to ensure uniformity.

• **Tokenization (Optional)**: Splitting sentences into words for advanced analysis.

UiPath automates these processes to ensure consistency and efficiency.

Sentiment Analysis Model

At the core of the architecture, the sentiment analysis model performs text classification.

• The **model** uses pre-trained NLP algorithms or APIs to categorize text into three sentiments:

Happy: Indicates positive emotions.

Sad: Reflects negative emotions.

• **Neutral**: Represents a lack of strong emotional tone.

The model processes the preprocessed data and assigns a sentiment label to each entry.

Output Structuring Layer

This layer structures the analysis results for easy interpretation and reporting.

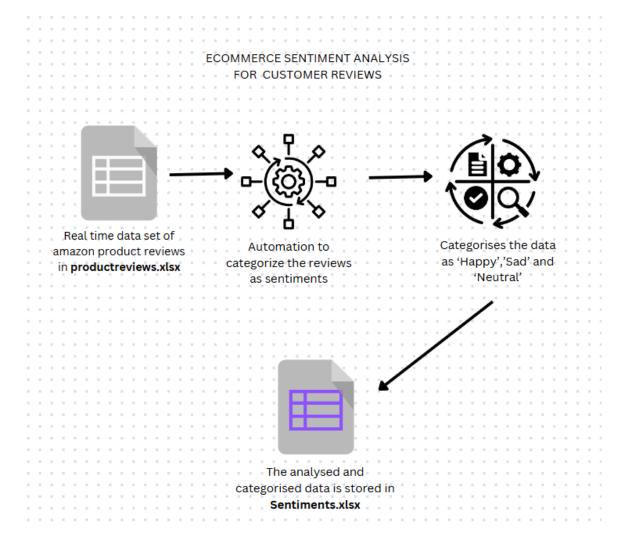
- The processed data is formatted and stored in a structured
 Excel file.
- The file contains two columns:
 - o **Text**: Original input text.
 - Sentiment: The assigned sentiment label

Output Delivery Layer

The final layer delivers the categorized sentiment results to the user.

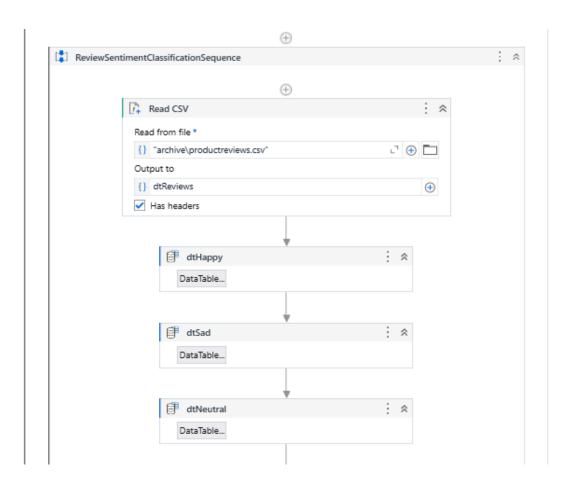
- Results are stored locally or in the cloud.
- Notifications or updates can be triggered automatically (if required).

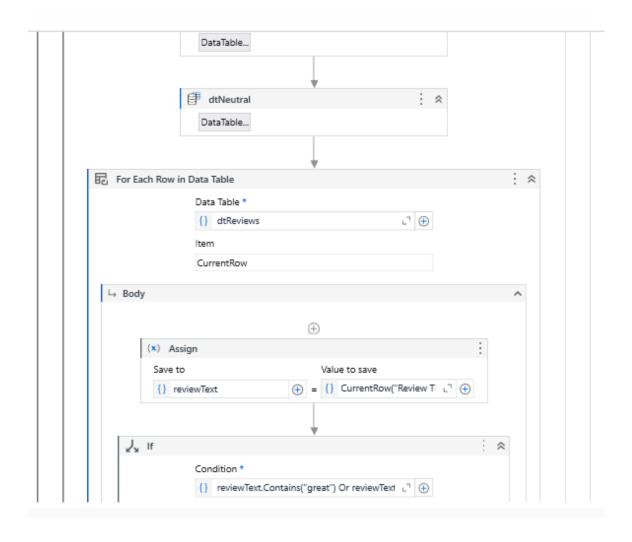
This layer ensures that stakeholders receive the output in an accessible and timely manner.

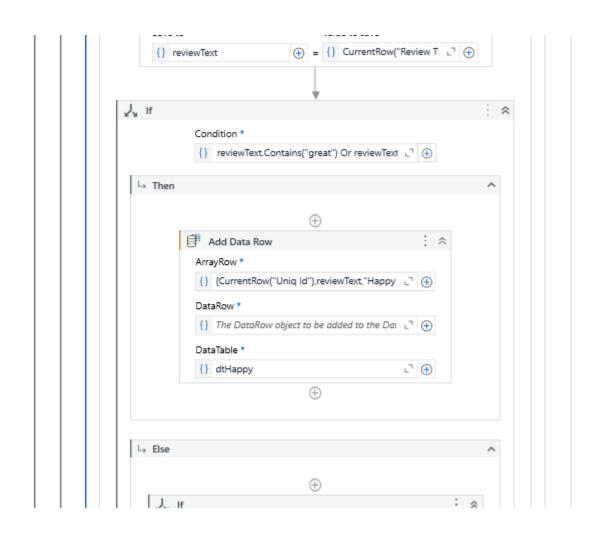


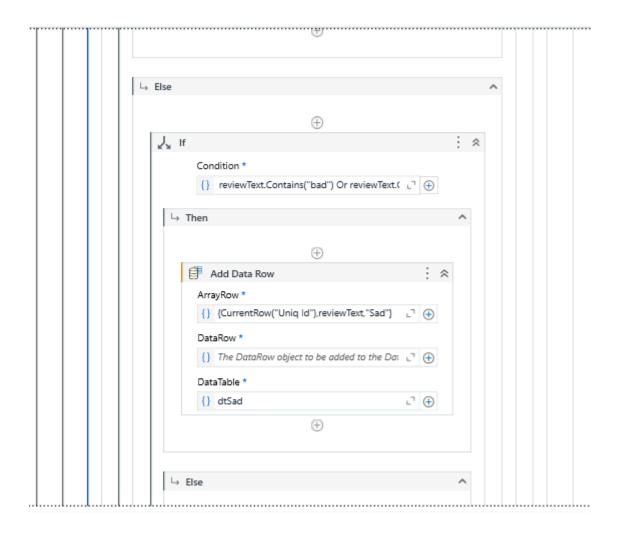
SYSTEM FLOW DIAGRAM

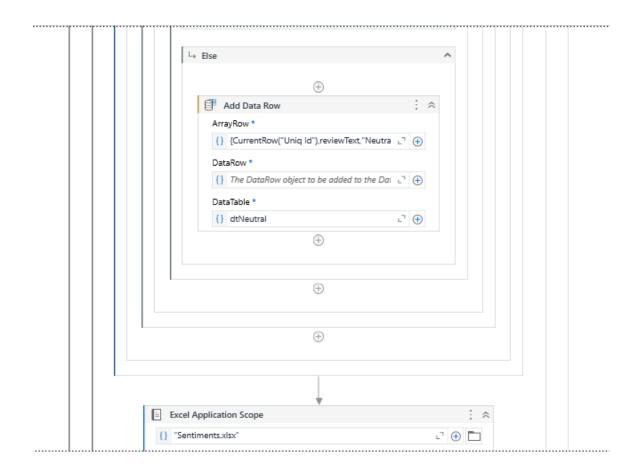
SEQUENCE DIAGRAM

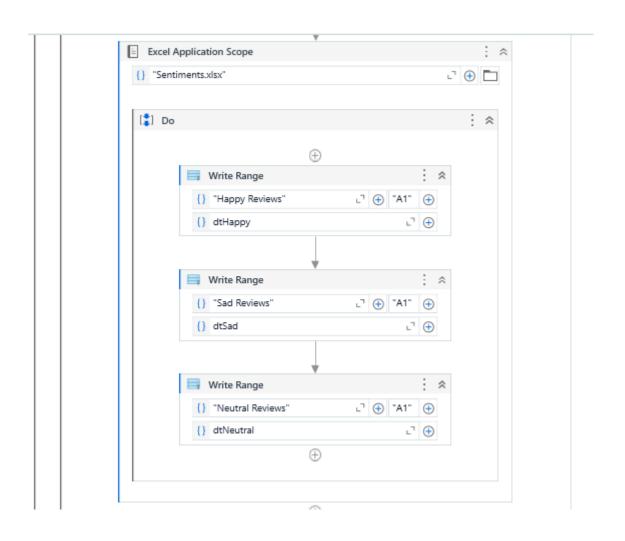












Chapter 4: CONCLUSION & FUTURE WORK

4.1 Conclusion

The Sentiment Categorizer project successfully demonstrated the application of Robotic Process Automation (RPA) combined with Natural Language Processing (NLP) techniques to automate the process of sentiment analysis. By integrating UiPath with NLP APIs such as Azure Text Analytics, the system was able to efficiently process large datasets of textual data, classify them into predefined sentiment categories (Happy, Sad, and Neutral), and output the results in an easily accessible format such as Excel. This automation reduced manual labor significantly and increased accuracy, offering valuable insights for various industries such as customer service, marketing, and social media monitoring.

The primary goal of automating sentiment categorization was achieved through an intuitive and streamlined workflow in UiPath, which was designed to read raw textual data, analyze sentiment, and store the output in structured formats. The project's ability to handle diverse datasets like customer feedback and product reviews proves its applicability in real-world scenarios. The integration of error-handling mechanisms ensured the reliability of the system, making it robust for large-scale sentiment analysis tasks.

The results of the project highlight the effectiveness of RPA in automating complex processes that traditionally required manual intervention. Moreover, it sets a strong foundation for future enhancements and scalability, particularly in more advanced sentiment analysis tasks.

4.2 Future Work

While the current iteration of the Sentiment Categorizer is effective, several areas for improvement and future work are identified:

- 1. Support for Multilingual Datasets: Currently, the system is optimized for English-language texts. Future versions could incorporate multilingual sentiment analysis by integrating more advanced NLP tools that support a wider range of languages. This would broaden the system's applicability in diverse markets globally.
- 2. Offline Sentiment Analysis: The project depends on cloud-based NLP APIs, which can be limited by internet connectivity or API usage restrictions. A significant enhancement would be to develop an offline version of the sentiment analysis tool, where sentiment models can be trained locally, and data is processed without needing constant API calls.
- 3. Real-time Sentiment Tracking: To make the system even more valuable for businesses, implementing a real-time dashboard that tracks sentiment trends over time would be beneficial. This could be used for live monitoring of social media posts, customer feedback, and other sources of textual data.
- 4. Deep Learning Model Integration: Incorporating advanced deep learning models like BERT or GPT for more nuanced sentiment understanding could greatly improve accuracy, especially in handling complex phrases, sarcasm, or context-based sentiments.

5. Scalability for Larger Datasets: The ability to handle even larger

datasets would make the system more scalable. This could involve

optimizing the UiPath workflows or integrating with big data

technologies like Hadoop or Apache Spark to process large

volumes of text data in parallel.

In conclusion, while the Sentiment Categorizer meets its current

objectives, there is vast potential for future improvements that would

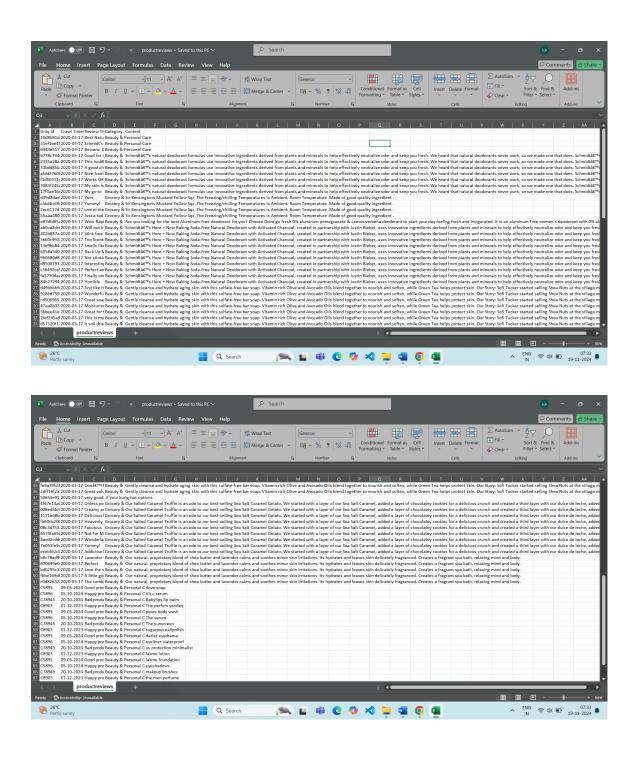
make it an even more powerful tool for sentiment analysis in diverse

applications.

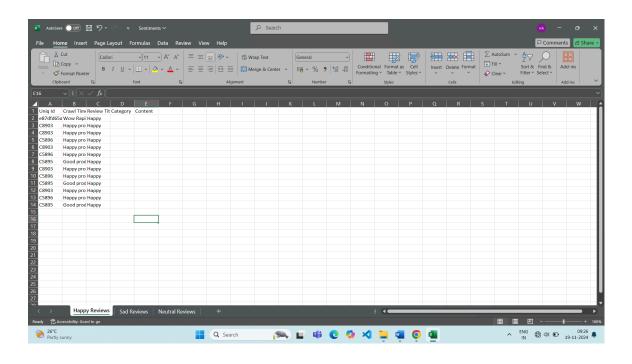
CHAPTER 5: SCREENSHOTS:

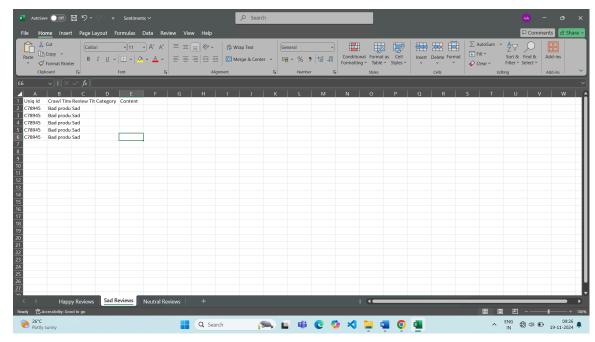
Appendix 1: Sample Input Dataset

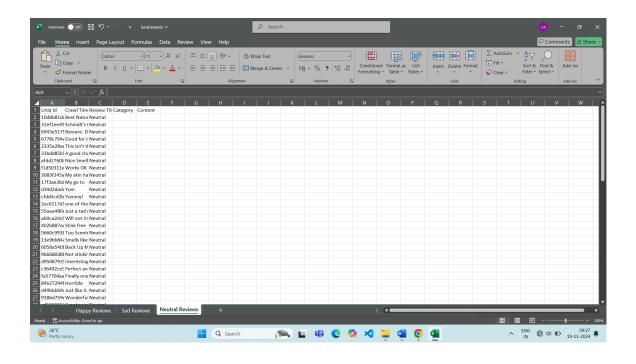
31

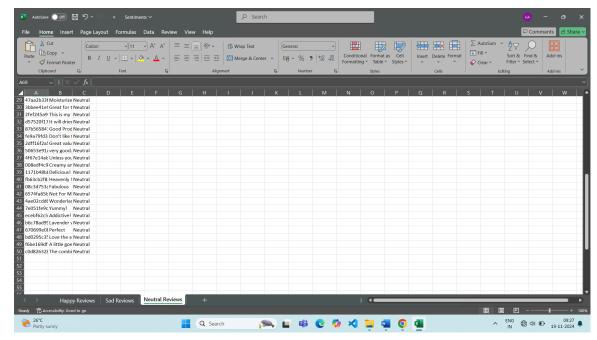


Appendix 2: Output Sample File









REFERENCES

- 1. Ariponnammal, S. and Natarajan, S. (1994) 'Transport Phonomena of Sm Sel X Asx', Pramana Journal of Physics Vol.42, No.1, pp.421-425.
- 2. Barnard, R.W. and Kellogg, C. (1980) 'Applications of Convolution Operators to Problems in Univalent Function Theory', Michigan Mach, J., Vol.27, pp.81–94.
- 3. Shin, K.G. and Mckay, N.D. (1984) 'Open Loop Minimum Time Control of Mechanical Manipulations and its Applications', Proc.Amer.Contr.Conf., San Diego, CA, pp. 1231-1236.