

SENTIMENTAL ANALYSIS ON TEXT BASED ON PRODUCT REVIEWS

A PROJECT REPORT

Submitted by

AKASH S A(22BTAD004)

HARISH S (22BTAD022)

KAILAS NATH (22BTAD028)

in partial fulfilment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE



Karpagam Academy of Higher Education

COIMBATORE – 641021

TAMILNADU INDIA

MARCH 2025

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA
SCIENCE**

This is to certify that the project entitled

**SENTIMENTAL ANALYSIS ON TEXT BASED ON
PRODUCT REVIEWS**

Is the bonafide record of project work done by

AKASH S A (22BTAD004)

HARISH S (22BTAD022)

KAILAS NATH (22BTAD028)

of B.Tech Artificial Intelligence and Data Science during the year 2023-2024

Dr.B.Lanitha
Project Guide

Dr. B. Arunkumar
Head of the Department

Submitted for the project Viva-Voce examination held on _____

Internal Examiner

Project Coordinator

ABSTRACT

Product review analysis using Python involves leveraging natural language processing (NLP) techniques to extract insights from customer feedback. By utilizing libraries such as TextBlob, NLTK, and spaCy, sentiment analysis can be performed to determine whether reviews are positive, negative, or neutral. Topic modeling algorithms like Latent Dirichlet Allocation (LDA) help identify recurring themes or issues within the reviews. Preprocessing steps, including tokenization, stopword removal, and stemming, are crucial for cleaning the text data. Machine learning models, such as logistic regression or support vector machines, can be trained to classify reviews based on sentiment or specific attributes. Visualization tools like Matplotlib and Seaborn enable the creation of intuitive graphs and charts to represent trends and patterns. Python's flexibility and extensive libraries make it an ideal choice for analyzing large datasets of product reviews efficiently. This approach helps businesses understand customer preferences, improve products, and enhance overall satisfaction. By automating the analysis process, companies can save time and make data-driven decisions. The integration of deep learning models, such as BERT, further enhances the accuracy of sentiment and emotion detection. Overall, Python-based product review analysis provides a comprehensive and scalable solution for extracting actionable insights from customer feedback. Additionally, aspect-based sentiment analysis can be employed to evaluate specific product features, such as performance, design, or usability, providing granular insights into customer opinions. Python's compatibility with cloud platforms like AWS and Google Cloud allows for scalable processing of large review datasets. Real-time analysis can be achieved by integrating APIs for continuous feedback monitoring, enabling businesses to respond promptly to customer concerns. Advanced techniques like word embeddings (e.g., Word2Vec, GloVe) improve the understanding of contextual relationships within the text.

ACKNOWLEDGEMENT

Any organized and systematic work calls for the co-operation and co-ordination of a team of people. My project is no exception to this. Hence, these pages find the space for thanking all those who have directly and indirectly contributed to the completion of this work in a successful manner.

I express my gratitude to **Dr.R.Vasanthakumar, B.E. (Hons), D.Sc.**, President, Karpagam Charity trust, for his encouragement and support in this project work.

I sincerely thank **Shri.K.Murugaiah, B.E**, CEO, Karpagam Educational Institutions, for his constant support and enduring encouragement for the successful completion of the dissertation.

I sincerely thank **Dr. B.Venkatachalapathy**, Vice Chancellor of Karpagam Academy of Higher Education, for his encouragement and support in this project work.

I sincerely thank **Dr.S.Ravi**, Registrar of Karpagam Academy of Higher Education, for his encouragement and support in this project work.

My special thanks to **Dr.P.Palanivelu**, Controller of Examinations, Karpagam Academy of Higher Education, for his timely help for the progress of this work for his valuable suggestion and timely help.

I express my heartfelt thanks to **Dr.A.Amudha**, Dean, Faculty of Engineering, Karpagam Academy of Higher Education, for his encouragement and support in this project work.

I express my heartfelt thanks to **Dr. B. Arun kumar**, Professor and Head, Department of Artificial Intelligence and Data Science, Faculty of Engineering, Karpagam Academy of Higher Education, for his encouragement and valuable guidance in this project work.

I would like to extend my heartfelt thanks to our guide **Dr.B.Lanitha, Associate Professor**, Department of Artificial Intelligence and Data Science, Karpagam Academy of Higher Education, for his encouragement in carrying out this project work.

I also express my thanks to my parents, my friends, well-wishers for their encouragement and best wishes in the successful completion of this dissertation.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	iii
	TABLE OF CONTENTS	v
	LIST OF FIGURES	viii
1	INTRODUCTION	1
	1.1 NATURAL LANGUAGE PROCESSING	2
	1.2 MACHINE LEARNING TECHNIQUES	3
	1.3 PANDAS	5
2	LITERATURE REVIEW	6
	2.1 PRACTICAL NATURAL LANGUAGE PROCESSING : A COMPREHENSIVE GUIDE TO BUILDING REAL-WORLD NLP SYSTEM	6
	2.2 HANDS-ON MACHINE LEARNING WITH SCIKIT-LEARN, KERAS AND TENSORFLOW BY AURELIEN GERON	8

	2.3 MACHINE LEARNING FOR ABSOLUTE BEGINNERS BY OLIVER THEOBALD	11
	2.4 PYTHON FOR DATA ANALYSIS : DATA WRANGLING WITH PANDAS, NUMPY AND JUPYTER BY WES MCKINNEY	14
3	METHODOLOGY	18
	3.1 EXISTING SYSTEM	18
	3.2 PROPOSED SYSTEM	18
	3.3 DATA COLLECTION AND PRE-PROCESSING	20
	3.3.1 Data Ingestion	20
	3.3.2 Data Cleaning	20
	3.3.3 Exploratory Data Analysis	21
	3.4 FEATURE EXTRACTION	21
	3.5 MODEL SELECTION	22
	3.6 RANDOM FOREST CLASSIFIER	22
	3.7 MODEL TRAINING	23
	3.8 MODEL ACCURACY AND SCORE	24

4	RESULTS AND DISCUSSION	26
	4.1 COMPARING THE EXISTING MODEL AND OUR MODEL	26
	4.2 STREAMLIT OVER HTML, CSS AND GUI FRAMEWORKS	28
5	CONCLUSION AND FUTURE WORK	44
	5.1 CONCLUSION	29
	5.2 FUTURE WORK	30
	APPENDIX A	31
	A.1 SOURCE CODE	31
	A.2 SCREENSHOTS	36
	REFERENCES	38

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE NO.
3.5.1	Logistic Regression vs Random classifier	22
3.8.1	Model Accuracy and score	25
3.8.2	Model Workflow Representation	25
4.1.1	Existing Model vs Our Model	27
4.2.1	Streamlit vs Other Framework	28
A.2	EDA Result	36
A.2.2	Basic prediction	36
A.2.3	General User Interface 1	37
A.2.4	General User Interface 2	37