ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report **"SENTIMENTAL ANALYSIS ON CONSUMER REVIEWS IN AUTOMOBILE INDUSTRY BASED ON LINGUISTIC ALGORITHMS"** is the bona fide work of **SHRIRAM B. (312313205100)** and **TANAY PARDESHI (312313205115)** who carried out the project work under my supervision, for the partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Information Technology.

SIGNATURE

Dr. V. Muthulakshmi M.E., Ph.D.,

Associate Professor

HEAD OF THE DEPARTMENT (LAB AFFAIRS)

Department Of Information Technology

St. Joseph's College of Engineering Old Mamallapuram Road

Chennai-600119

SIGNATURE

Dr. Lilly Raamesh M.E.,

Ph.D.,

Associate Professor

HEAD OF THE DEPARTMENT (STUDENT AFFAIRS)

Department Of Information Technology

St. Joseph's College of Engineering

Old Mamallapuram Road

Chennai-600119

Submitted for the Viva-Voce held on _____

INTERNAL EXAMINER **EXTERNAL EXAMINER**

CERTIFICATE OF EVALUATION

College Name : St. Joseph's College of Engineering

Branch & Semester : Information Technology (VIII)

S.NO	NAMES OF STUDENTS	TITLE OF THE PROJECT	NAME OF THE SUPERVISOR
1.	SHRIRAM B (312313205100)	SENTIMENTAL ANALYSIS ON CONSUMER REVIEWS IN AUTOMOBILE BASED ON LINGUISTIC ALGORITHMS	Dr.Lilly Raamesh
2.	TANAY PARDESHI (312313205115)		

The report of the project work submitted by the above student in partial fulfilment for the award of Bachelor of Technology degree in Information Technology of Anna University were evaluated and confirmed to be report of the work done by the above student.

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

The contentment and elation that accompany the successful completion of any work would be incomplete without mentioning the people who made it possible.

Our sincere thanks and gratitude to our respected Chairman and Managing Director **Dr. B. Babu Manoharan M.A., M.B.A., Ph.D.,** and our beloved Principal **Dr. Vaddi Seshagiri Rao M.E., M.B.A., Ph.D.,** for having encouraged us to do our under graduation in Information Technology in this esteemed college.

We also express our sincere thanks and heartfelt gratitude to our eminent Head of the Department – Lab Affairs **Dr. V. Muthulakshmi M.E., Ph.D.,** for having extended her helping hand at all times.

It is with deep sense of gratitude that we acknowledge our indebtedness to our beloved supervisor **Dr. Lily Raamesh M.E., Ph.D.,** a perfectionist for her expert guidance and connoisseur suggestion.

Last but not the least; we thank our family members and friends who have been the greatest source of support to us.

ABSTRACT

The system proposes a domain independent supervised learning methodology that uses machine learning techniques to build models or discriminators for the different classes such as positive or negative reviews using a large corpus for an automobile organization to increase the overall productivity. The training data consists of a set of training examples of the product reviews of automobiles that is segregated based on various models and years of manufacturing. This training dataset is fed into the system which analyses commonly occurring data patterns and identifies the polarity of each review provided by the user. The analysed data is then used to predict the nature of possible outcomes from previous data and provide recommendations to improve efficiency and assist in examining the effects of vehicular emissions on climate change. This is done with the help of the Natural Language Processing library (NLP) in Python that helps in dividing a sentence into positive, negative or neutral feedback based on the polarity of the sentence. This analysis can be further used in correlating the data to vehicular emissions of various automobile manufacturers and hence enabling them to identify the areas that require improvement and take corresponding measures to reduce emissions and address the issues of climate change.

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE NO
	ABSTRACT	**v**
	LIST OF FIGURES	**vi**
	LIST OF ABBREVIATIONS	**vii**
1	**INTRODUCTION**	**1**
	1.1 SYSTEM OVERVIEW	1
	1.2 SCOPE OF THE PROJECT	2
2	**LITERATURE SURVEY**	**3**
3	**SYSTEM ANALYSIS**	**10**
	3.1 EXISTING SYSTEM	10
	3.1.1 Disadvantages	10
	3.2 PROPOSED SYSTEM	11
	3.2.1 Advantages	11
	3.3 REQUIREMENT SPECIFICATION	12
	3.3.1 Hardware Requirements	12
	3.3.2 Software Requirements	12
	3.4 LANGUAGE SPECIFICATION	12
4	**SYSTEM DESIGN**	**15**
	4.1 SYSTEM ARCHITECTURE	15
	4.2 ACTIVITY DIAGRAM	16

	4.3 DATA FLOW DIAGRAM	18
	4.4 USE CASE DIAGRAM	20
	4.5 COLLABORATION DIAGRAM	21
	4.6 SEQUENCE DIAGRAM	22
5	**SYSTEM IMPLEMENTATION**	**23**
	5.1 MODULES	23
	5.1.1 Edmund's Developer API	23
	5.1.2 MongoDB	23
	5.1.3 Apache PySpark Environment	25
	5.1.4 NLTK library	26
6	**CONCLUSION AND FUTURE ENHANCEMENT**	**35**
	6.1 CONCLUSION	35
	6.2 FUTURE ENHANCEMENT	36
	APPENDIX 1 - SAMPLE CODING	**37**
	APPENDIX 2 - SCREEN SHOTS	**43**
	REFERENCES	**46**

LIST OF FIGURES

FIGURE NO.	NAME OF THE FIGURE	PAGE NO.
4.1	Architecture diagram	15
4.2	Activity Diagram	16
4.3	Data Flow Diagram	18
4.4	Use Case Diagram	20
4.5	Collaboration Diagram	21
4.6	Sequence Diagram	22

LIST OF ABBREVIATIONS

API	Application Programming Interface
REST	Representational State Transfer
HDFS	Hadoop File System
ASP	Active Server Page
OS	Operating System
JDK	Java Development Kit
JSON	JavaScript Object Notation
SQL	Structured Query Language
NLP	Natural Language Processing
NLTK	Natural Language Tool Kit
RDD	Resilient Distributed Datasets