**ALAGAPPA CHETTIAR GOVERNMENT COLLEGE OF ENGINEERING AND TECHNOLOGY, KARAIKUDI** - **630 003**

**(An Autonomous Institution Affiliated to Anna University, Chennai)**



**A CORPUS FOR TAMIL SENTIMENT ANALYSIS OF TAMIL TWEETS**

**A PROJECT REPORT**

***Submitted by***

**SOWMIYAA J R**

# (1515305)

***in partial fulfillment for the award of the degree***

***of***

# BACHELOR OF ENGINEERING

# IN

# COMPUTER SCIENCE AND ENGINEERING

# APRIL 2019

## ALAGAPPA CHETTIAR GOVERNMENT COLLEGE OF ENGINEERING

## AND TECHNOLOGY, KARAIKUDI - 630 003

(An Autonomous Institution Affiliated to Anna University, Chennai)

**BONAFIDE CERTIFICATE**

## Certified that this project report “A CORPUS FOR TAMIL SENTIMENT ANALYSIS OF TAMIL TWEETS” is the bonafide work of SOWMIYAA.J.R. (1515305) who carried out the project work under my supervision.

**SIGNATURE SIGNATURE**

Dr. L.GANESAN M.E., Ph.D. Dr. C. UMARANI M.E., Ph.D.

**PROFESSOR & HEAD SUPERVISOR**

**ASSISTANT PROFESSOR**

DEPT OF COMPUTER SCIENCE AND DEPT OF COMPUTER SCIENCE AND

ENGINEERING, ENGINEERING,

AlagappaChettiar Government College of AlagappaChettiar Government College of

Engineering and Technology, Engineering and Technology,

Karaikudi - 630 003. Karaikudi - 630 003.

Submitted for (**15CSZ03**) Project Viva-Voce Examination held on:\_\_\_\_\_\_\_\_\_\_\_\_\_

**INTERNAL EXAMINAR EXTERNAL EXAMINAR**

# ACKNOWLEDGEMENT

I thank the Almighty for keeping me healthy and providing me good circumstances to finish the project.

I would like to express my deep sense of gratitude and respectful regards to our Principal **Dr. A.ELANGO,** Alagappa Chettiar Government College of Engineering and Technology, Karaikudi-630 003 for giving an opportunity to do this project.

I owe my sincere thanks to **Dr. L.GANESAN, M.E., Ph.D. Professor and Head, Department of Computer Science and Engineering,** Alagappa Chettiar Government College of Engineering and Technology, Karaikudi-630 003 for providing various facilities and also giving valuable guidance during the project work.

I thank my Mentor **Dr. C.UMARANI M.E., Ph.D. Assistant Professor**, Department of Computer Science and Engineering, Alagappa Chettiar Government College of Engineering and Technology, Karaikudi-630 003 for her full support in doing my project work.

I like to express my hearty thanks to all our beloved staff members, nonteaching staff and my friends for their valuable suggestion and co-operation for finishing my project.

**ABSTRACT**

The main objective of this project is to create a corpus for sentiment analysis of Tamil Tweets. The Tamil language still lacks sufficient language resources to enable the tasks of Sentiment Analysis. Sentiment Analysis is a highly subjective and challenging task. Its complexity further increases when applied to the Tamil Language, mainly because of the large variety of dialects that are unstandardized and widely used in the web, especially in social media. In this project, a corpus of Tamil tweets annotated for Sentiment Analysis. The sentiments of the tweets are analysed with both supervised and unsupervised algorithms. Many datasets have been released to train sentiment classifiers in Tamil. These dataset contain only shallow annotation only having the words blindly without categorized. This corpus consists mainly of tweets in Tamil. The corpus was manually annotated as sentences and words with positive, negative and neutral sentiments. This corpus contains tweets as sentences and words both in English and Tamil can be used for research as Tamil Tweets does not contain pure Tamil tweets.

**TABLE OF CONTENTS**

**CHAPTERS TITLE PAGE NO.**

**ABSTRACT iv**

**LIST OF FIGURES vii**

**1. INTRODUCTION 1**

1.1 Objective of the Project 1

1.2 Scope of the Project 1

1.3 Outline of the Project 1

**2. LITERATURE DESCRIPTION 2**

­­­­­­­­**3. SYSTEM ANALYSIS 6**

3.1 Problem Definition 6

3.2 Existing System 6

3.3 Proposed System 6

**4. SYSTEM SPECIFICATION 7**

4.1 Hardware Requirements 7

4.2 Software Requirements 7

4.3 Software Specification 7

**5. SYSTEM DESIGN 15**

5.1 System Architecture 15

5.2 Class Diagram 15 5.3 Usecase Diagram 18

5.4 Flow Diagram 19

**6. MODULES DESCRIPTION 22**

* 1. Module 22
  2. Twitter Data Set Collection 22
  3. Twitter Data Set Preprocessing 22
  4. Applying Classifiers 23

6.5 Analysing and Storing the results 23

6.6Creation of Corpus 23

**7. SYSTEM TESTING 24**

7.1 Testing 24

7.2 Unit Testing 24

7.3 Integration Testing 24

7.4 Validation Testing 25

**8. CONCLUSION AND FURTHER ENHANCEMENT 26**

**9. APPENDICES**

9.1 Source code 27

9.2 Screen shots 67

**10.**  **REFERENCES 75**

**LIST OF FIGURES**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **FIG.NO.** | **FIGURE NAME** | | | **PAGE NO.** |
| 5.1  5.2  5.3  5.4 | | ARCHITECTURE DESIGN  CLASS DIAGRAM  USECASE DIAGRAM  FLOW DIAGRAM | 16  17  18  21 | |