

# **SENTIMENT ANALYSIS OF TWITTER DATA**

## **A PROJECT REPORT**

*Submitted by*

**M.SUTHAKARAN (810015104094)**

**J.MOHAMED RIYAZ FAROOK (810015104721)**

*in partial fulfillment for the award of the degree*

*of*

**BACHELOR OF ENGINEERING**

*in*

**COMPUTER SCIENCE AND ENGINEERING**



**UNIVERSITY COLLEGE OF ENGINEERING, BIT CAMPUS,**

**TIRUCHIRAPPALLI**

**ANNA UNIVERSITY::CHENNAI 600 025**

**APRIL 2019**

**UNIVERSITY COLLEGE OF ENGINEERING,  
BIT CAMPUS,**

**TIRUCHIRAPPALLI-620 024.**

**BONAFIDE CERTIFICATE**

Certified that this project report “**SENTIMENT ANALYSIS OF TWITTER DATA**” is the bonafide work of **Mr. SUTHAKARAN M (810015104094)** and **Mr. MOHAMED RIYAZ FAROOK J (810015104721)** who carried out Project work under my supervision.

**Dr. D.VENKATESAN**

Assistant Professor & Head

Department of CSE

Anna University, BIT Campus

University College of Engineering

Tiruchirappalli - 620 024

**Dr.I.SHAHANAZ BEGUM**

Assistant Professor

Department of CSE

Anna University, BIT Campus

University College of Engineering

Tiruchirappalli - 620 024

Submitted for the ANNA UNIVERSITY Viva-voce Examination held on

.....

**Internal Examiner**

**External Examiner**

## **DECLARATION**

We hereby declare that the work entitled “**SENTIMENT ANALYSIS OF TWITTER DATA**” is submitted in partial fulfillment of the requirements for the award of the degree in B.E.-Computer Science and Engineering, University College of Engineering, BIT Campus, Tiruchirappalli, is a record of our own work carried out by us during the academic year 2018-2019 under the guidance of **Dr. I.SHAHANAZ BEGUM**, Assistant Professor, Department of Computer Science and Engineering, University College of Engineering, BIT Campus, Tiruchirappalli. The extent and source of information are derived from the existing literature and have been indicated through the dissertation at the appropriate places. The matter embodied in this work is original and has not been submitted for the award of any other Degree, either in this or any other University.

**Signature of the Candidates**

**SUTHAKARAN M**

**MOHAMED RIYAZ FAROOK J**

I certify that the declaration made above by the candidate is true.

**Signature of the Project Guide**

**Dr. I. SHAHANAZ BEGUM**

**Assistant Professor**

Department of CSE,

University College of Engineering-BIT

Tiruchirappalli – 620 024

## ACKNOWLEDGEMENT

First and foremost of all, we would like to thank our beloved Parents and God Almighty for giving us the strength, knowledge, ability and opportunity to undertake this project study and to persevere and complete it satisfactorily.

We would like to convey our thanks to our honorable Dean **Dr. T. SENTHIL KUMAR**, Associate Professor, for having provided us with all required facilities not only to complete our project without hurdles but also for the entire course of study.

We extremely indebted to our Head of the Department **Dr. D. VENKATESAN**, Assistant Professor, Department of Computer Science and Engineering, for the devoted attention shown upon us and making the project a grant success.

We whole heartedly express our gratitude to our Internal Project Guide **Dr. I. SHAHANAZ BEGUM** Assistant Professor, Department of Computer Science and Engineering for her valuable guidance and help for the successful completion of the project.

We also express our sincere thanks to our Project Coordinators **Mr. P. KARTHIKEYAN** Assistant Professor, **Mr. C. SANKAR RAM** Assistant Professor, **Mr. C. SURESH KUMAR** Teaching Fellow Department of Computer Science and Engineering, for her constant inspiration and support.

We extend our thanks to all other Teaching and Non-Teaching Staff members for their constant encouragement and moral support. And we would like to thank all our well-wishers who have encouraged us to do this project.

## **ABSTRACT**

The average of customer ratings on a product, which call a reputation, is one of the key factors in online purchasing decisions. The trustworthiness of a reputation is important since it can be manipulated rather easily. False reputation has to be handled the problem of a reputation being manipulated by unfair ratings and Designing a general framework that provides trustworthy reputations. As e-commerce is growing and becoming popular day-by-day, the number of reviews received from customer about the product grows rapidly. This creates difficulty for the potential customer to read them and to make a decision whether to buy or not the product. Problems also arise for the manufacturer of the product to keep track and to manage customer opinions. And also additional difficulties are faced by the manufacturer because many other merchants sites may sell the same product at good ratings and the manufacturer normally produces many kinds of products. So, many fake reviews posted on that product to increasing the sales of one product. In this work manipulated reviews and false reputation are identified. In this work, this problem is overcome. Here, the fake reviews are found out through the decision tree using CUMSUM algorithm and also detect the false reputation through computing the confidence of rating based on user activity, user objectivity, user consistency.

## **TABLE OF CONTENTS**

<b>CHAPTER NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
	<b>ABSTRACT</b>	
	<b>LIST OF FIGURES</b>	
	<b>LIST OF ABBREVIATIONS</b>	
<b>1.</b>	<b>INTRODUCTION</b>	
	1.1 Opinion Mining	4
	1.2 Sentiment Analysis and Data Mining	5
	1.3 Techniques	6
	1.4 Information Demand on Sentiment	6
	1.5 Opinion orientation	6
	1.6 Opinion Search Engine	7
	1.7 Objectives	7
<b>2.</b>	<b>LITERATURE SURVEY</b>	<b>8</b>
<b>3.</b>	<b>SYSTEM ANALYSIS</b>	
	3.1 Existing System	14
	3.2 Proposed System	15
<b>4.</b>	<b>SYSTEM DESIGN</b>	
	4.1 General	18
	4.2 Use Case Diagram	18

	4.3	Class Diagram	19
	4.4	System Architecture	20
	4.5	Data Flow Diagram	21
<b>5.</b>		<b>SYSTEM CONFIGURATION</b>	
	5.1	General	25
	5.2	Hardware Requirement	25
	5.3	Software Requirement	26
<b>6.</b>		<b>SYSTEM IMPLEMENTATION</b>	
	6.1	Modules Name	27
	6.1.1	Weblogs	
	6.1.2	Sentiment Analysis	
	6.1.3	Product Reviews	
	6.1.4	Summarization	
	6.1.5	Sentiment Classification	
<b>7.</b>		<b>SOFTWARE DESCRIPTION</b>	
	7.1	Java Virtual Machine	31
	7.2	Class Library	31
<b>8.</b>		<b>SOFTWARE TESTING</b>	

	8.1	Developing Methodology	33
	8.2	Unit Testing	33
	8.3	Functional Testing	33
	8.4	System Testing	34
	8.5	Performance Testing	34
	8.6	Integration Testing	34
	8.7	Validation	35
<b>9</b>		<b>CONCLUSION AND FUTURE WORK</b>	<b>36</b>
		<b>APPENDIX</b>	<b>37</b>
		<b>EVALUATION</b>	<b>40</b>
		<b>REFERENCES</b>	<b>42</b>



## LIST OF FIGURES

FIGURE NO.	DESCRIPTION	PAGE NO.
<b>Figure 1.1</b>	Data Mining Knowledge	2
<b>Figure 4.1</b>	Use Case Diagram	18
<b>Figure 4.2</b>	Class Diagram	19
<b>Figure 4.3</b>	System Architecture	20
<b>Figure 4.4</b>	Data Flow Diagram	21
<b>Figure 4.5</b>	Sequence Diagram	24

## LIST OF ABBREVIATIONS

<b>DBMS</b>	Database Management System
<b>KDD</b>	Knowledge Discovery in Database
<b>ICF</b>	Iterative Computation Framework
<b>EWGA</b>	Entropy Weighted Genetic Algorithm
<b>SCL</b>	Structural Correspondence learning
<b>UML</b>	Unified Modelling Language
<b>DFD</b>	Data Flow Diagram
<b>HCL</b>	Hardware Compatibility List
<b>JVM</b>	Java Virtual Machine
<b>JIT</b>	Just-in-time