



## **FINAL YEAR PROJECT REPORT**

# **TO FIND OUT THE QUALITY AND POPULARITY OF A PRODUCT BY USING USER COMMENTS**

**In fulfillment of the requirement  
For degree of  
BS (INFORMATION TECHNOLOGY)**

**By**

**ADNAN AHMED  
AYESHA SAJJAD  
WALEED AMIR**

**48870 BSIT  
48872 BSIT  
48903 BSIT**

**SUPERVISED**

**BY**

**MUHAMMAD SHAHID KHAN**

**BAHRIA UNIVERSITY (KARACHI CAMPUS)  
FALL-2020**

**TO FIND OUT THE QUALITY AND POPULARITY OF A PRODUCT BY  
USING USER COMMENTS**

**ADNAN AHMED  
WALEED AMIR  
AYESHA SAJJAD**

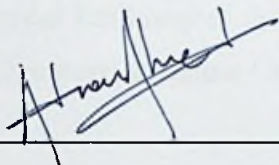
**A project report submitted in partial fulfilment of the  
requirements for the award of the degree of  
Bachelor of Information Technology (Honours)**

**Department of Computer Science  
Bahria University, Karachi Campus**

**December 2020**

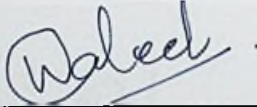
**DECLARATION**

We hereby declare that this project report is based on our original work except for citations and quotations which have been duly acknowledged. We also declare that it has not been previously and concurrently submitted for any other degree or award at Bahria University or other institutions.

Signature :  \_\_\_\_\_

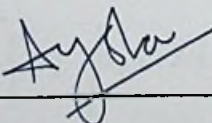
Name : ADNAN AHMED

Reg No. : 48870

Signature :  \_\_\_\_\_

Name : WALEED AMIR

Reg No. : 48903

Signature :  \_\_\_\_\_

Name : AYESHA SAJJAD

Reg No. : 48872

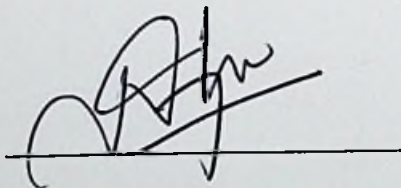
Date : December 22, 2020

**APPROVAL FOR SUBMISSION**

We certify that this project report entitled **“TO FIND OUT THE QUALITY AND POPULARITY OF A PRODUCT BY USING USER COMMENTS”** was prepared by **Adnan Ahmed, Waleed Amir and Ayesha Sajjad** has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Computer Science (Honours) at Bahria University.

Approved by,

Signature :



Supervisor: Mr. Muhammad Shahid Khan

Date :

11/01/2021

The copyright of this report belongs to Bahria University according to the Intellectual Property Policy of Bahria University BUORIC-P15 amended on April 2019. Due acknowledgement shall always be made of the use of any material contained in, or derived from, this report.

© 2019 Bahria University. All right reserved.

## ACKNOWLEDGEMENTS

We are grateful to ALLAH ALMIGHTY for enabling us to present this project as the outcome of our learning in this very form. We would like to thank everyone who had contributed to the successful completion of this project. We would like to express my gratitude to my research supervisor, Mr Muhammad Shahid Khan for his invaluable advice, guidance and his enormous patience throughout the development of the research.

In addition, we would also like to express my gratitude to our loving parent and friends who had helped and given me encouragement.

## TO FIND OUT THE QUALITY AND POPULARITY OF A PRODUCT BY USING USER COMMENTS

### ABSTRACT

With the growth of ecommerce applications, almost everything is being sold and purchased online, which for the most part is quite significant. These customers kind of leave their reviews about purchased product which basically are then reviewed by other customers before making a purchase in a subtle way. By using Natural Language Processing on these reviews, we for the most part have proposed a supervised model which can give the polarity i.e., for all intents and purposes contrary to popular belief negativity and positivity of these reviews which can be used to specifically identify whether the product quality generally is good or not, for all intents and purposes. We have collected a data set of smartphones from amazon, it was then labelled to train and test the model on which we did comparative analysis of Logistic Regression, Linear Regression, Decision Tree, Random Forest, and Support Vector Machine classifier and other pre-processing methods and techniques which includes tokenization, lemmatization, stop words removal, and Parts of Speech Tagging to filter the data set. Term Frequency-Inverse Document Frequency was used to convert words to vectors for the training of model. In comparative analysis, we found best results of logistic regression algorithm. The accuracy we found against logistic regression was 86%. So, logistic regression was then used for further processing and predictions i.e., for predicting the quality of a product and popularity of a product. The data set of 5000 reviews of a smartphone was loaded for predictions which our model approximately predicted correctly.

## TABLE OF CONTENTS

<b>DECLARATION</b>	i
<b>APPROVAL FOR SUBMISSION</b>	ii
<b>ACKNOWLEDGEMENTS</b>	v
<b>ABSTRACT</b>	vi
<b>TABLE OF CONTENTS</b>	vii
<b>LIST OF FIGURES</b>	ix
<b>LIST OF SYMBOLS / ABBREVIATIONS</b>	x

### CHAPTER

<b>1</b>	<b>INTRODUCTION</b>	1
	1.1 Background	1
	1.2 Problem Statements	1
	1.3 Aims and Objectives	2
	1.4 Scope of Project	3
<b>2</b>	<b>LITERATURE REVIEW</b>	4
<b>3</b>	<b>DESIGN AND METHODOLOGY</b>	9
	3.1 Data Set	10
	3.2 Pre-Processing	10
	3.2.1 Tokenizing	10
	3.2.2 Lemmatization	10
	3.2.3 Stop words Removal	10
	3.2.4 Parts of Speech	10
	3.3 Feature Extraction	11



3.3.1	TF-IDF	11
3.3.2	Bag of words	11
<b>4</b>	<b>IMPLEMENTATION</b>	<b>12</b>
4.1	Training Model:	14
4.1.1	Precision	16
4.1.2	Recall	17
4.1.3	F1-Score	17
4.1.4	Accuracy	17
<b>5</b>	<b>RESULTS AND DISCUSSIONS</b>	<b>20</b>
<b>6</b>	<b>CONCLUSION AND RECOMMENDATIONS</b>	<b>23</b>
6.1	Future Research	23
	<b>REFERENCES</b>	<b>25</b>