



FINAL YEAR PROJECT REPORT

**TEXT SUMMARIZER USING MACHINE
LEARNING**

**In fulfillment of the requirement
For degree of
BS (COMPUTER SCIENCES)**

By

MARIA YOUNUS

48535 BSCS

MUHAMMAD OSAMA

48415 BSCS

AMMARAH NASEEM KHAN

48519 BSCS

SUPERVISED

BY

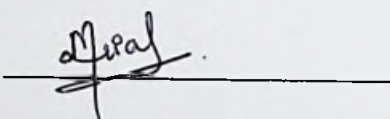
MISS FATIMA BASHIR

BAHRIA UNIVERSITY (KARACHI CAMPUS)

FALL-2020

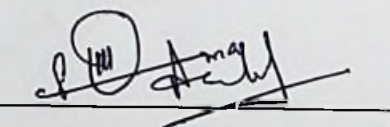
DECLARATION

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 : 

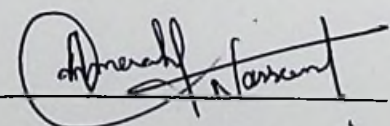
Name : Maria Younus

Reg No. : 02-134171-146

Signature :- 

Name : Mohammad Osama

Reg No. : 02-134171-018

Signature : 

Name : Ammarah Naseem Khan

Reg No. : 02-134171-129

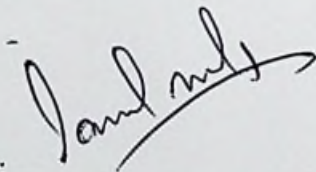
Date : 11/1/2021

APPROVAL FOR SUBMISSION

We certify that this project report entitled "TEXT SUMMARIZER USING MACHINE LEARNING" was prepared by Maria Younus, Mohammad Osama and Ammarah Naseem Khan have met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Computer Science (Honours) at Bahria University.

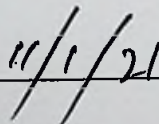
Approved by,

Signature



Supervisor : Miss Fatima Bashir

Date



ACKNOWLEDGEMENTS

We would like to thank everyone who had contributed to the successful completion of this project. We would like to express our gratitude to our research supervisor, Miss Fatima Bashir for her invaluable advice, guidance and her enormous patience throughout the development of the research.

Also, we would like to express our gratitude to our loving parent and friends who had helped and encouraged us. We have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. We would like to extend our sincere thanks to all of them.

We are highly indebted to Bahria University for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

We would like to express our gratitude towards our parents & course instruct at Bahria University for their kind co-operation and encouragement, which help us in completion of this project. Our thanks and appreciations also go to colleagues in developing the project and people who have willingly helped us out with their abilities.

TEXT SUMMARIZER USING MACHINE LEARNING

ABSTRACT

Text summarization refers to the technique of shortening long pieces of text. With such a big amount of data circulating in the digital space, there is a need to develop machine learning technique that can automatically shorten longer texts and deliver accurate summaries that can fluently pass the intended messages. The intention is to create a coherent and fluent summary having only the main points outlined in the document. Machine learning and natural language processing (NLP) will be used to automate text summarization. The objective of this project is to develop Text Summarizer using Machine Learning. It can overcome the grammar inconsistencies of the extractive method.

Text summarizer will convert the paragraph into sentences. First, Text summarizer split the paragraph into its corresponding sentences, then Text will be processed, the next step will be Tokenization. It will evaluate the weighted occurrence frequency of the words, and then it will substitute words with their weighted frequencies. All of the work will be done using R tool.

The main advantage of using this technique is that it provides the source text into a shorter version with semantics, it reduces the reading time, it expresses the main intent of the given document. Text summarization takes care of choosing the most significant portions of text and generates coherent summaries that express the main intent of the given document. Extraction based text summarization involves selecting sentences of high relevance (rank) from the document. The abstractive text summarization algorithms create new phrases and sentences that relay the most useful information from the original text just like humans do. The system first proceeds with the pre-processing of the given text, then tokenization and vectorization using Text Rank algorithm.

TABLE OF CONTENTS

DECLARATION	ii
APPROVAL FOR SUBMISSION	iii
DEDICATION	
viv	
ACKNOWLEDGEMENT	vi
ABSTRACT	vii
TABLE OF CONTENTS	
viii	
LIST OF TABLES	x
LIST OF FIGURES	xi
LIST OF SYMBOLS / ABBREVIATIONS	
xiii	

CHAPTER

1	INTRODUCTION	1
	1.1 Background	2
	1.2 Problem Statement	4
	1.3 Aims and Objectives	5
	1.4 Scope of Project	5
	1.5 Report Structure	6
2.	LITERATURE REVIEW	8
	2.1 Summarization Techniques	8
	2.1.1 Extractive Text Summarization	9
	2.1.2 Abstractive Text Summarization	9

		ix
2.2	Related Work	10
3.	DESIGN AND METHODOLOGY	16
3.1	Overview	16
3.2	Proposed Methodology	16
4	IMPLMENTATION	18
4.1	System Proposed	18
4.1.1	Reading Text from Text File.	19
4.1.2	Pre-Processing	19
4.1.3	Vector Representation	20
4.1.4	Similarity Matrix	21
4.2	TextRank Algorithm	22
4.2.1	Graph.	22
4.2.2	Top ranked Sentences	22
4.2.3	Generated Summary	23
5	RESULTS AND DISCUSSIONS	24
5.1	Extractive Summary By Textrank	24
5.2	Discussion	25
6	CONCLUSION AND RECOMMENDATIONS	26
	REFERENCES	28