

TWITTER BOT FOR SUMMARIZING ARTICLES BASED ON TWITTER TRENDS

A Project Report

Submitted by

MS. SAMREEN REYAZ

MR. DISHANT BHATT

MR. GABRIEL FRANCIS

Under the guidance of

PROF. PANKTI DOSHI

in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY

COMPUTER ENGINEERING

At



**MUKESH PATEL SCHOOL OF TECHNOLOGY
MANAGEMENT AND ENGINEERING, NMIMS, MUMBAI**

MARCH 2020

DECLARATION

We, Samreen Reyaz, Dishant Bhatt and Gabriel Francis, Roll Nos. B005, B013 and B027 B.Tech (Computer Engineering), VIII semester understand that plagiarism is defined as anyone or combination of the following:

1. Un-credited verbatim copying of individual sentences, paragraphs or illustration (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.
2. Un-credited improper paraphrasing of pages paragraphs (changing a few words phrases, or rearranging the original sentence order)
3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who did wrote what. (Source: IEEE, The institute, Dec. 2004)
4. I have made sure that all the ideas, expressions, graphs, diagrams, etc., that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.
5. I affirm that no portion of my work can be considered as plagiarism and I take full responsibility if such a complaint occurs. I understand fully well that the guide of the seminar/ project report may not be in a position to check for the possibility of such incidences of plagiarism in this body of work.

Signature of the Student:

Name:	Samreen Reyaz	Dishant Bhatt	Gabriel Francis
Roll No.:	B005	B013	B027
Place:	Mumbai		
Date:			

CERTIFICATE

This is to certify that the project entitled “Twitter Bot For Summarizing Articles Based On Twitter Trends” is the bonafide work carried out by Samreen Reyaz, Dishant Bhatt and Gabriel Francis of B.Tech (Computer Engineering), MPSTME (NMIMS), Mumbai, during the VIII semester of the academic year 2019-2020, in partial fulfillment of the requirements for the award of the Degree of Bachelors of Engineering as per the norms prescribed by NMIMS. The project work has been assessed and found to be satisfactory.

Prof. Pankti Doshi

Internal Mentor

Examiner 1

Examiner 2

Dean

MPSTME, NMIMS University

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	LIST OF FIGURES	i
	LIST OF TABLES	iii
	LIST OF EQUATIONS	iv
	ABBREVIATIONS	v
	ABSTRACT	vi
1	CHAPTER I – INTRODUCTION	1
	1.1 Background	1
	1.2 Project Overview	2
	1.3 What is Text Summarization?	3
	1.4 Motivation of Problem	3
	1.5 Problem Statement	4
	1.6 Scope of the Problem	4
	1.7 Assumptions and Limitations	5
	1.8 Organization of Report	5
2	CHAPTER II - REVIEW OF LITERATURE	6
	2.1 Numerical Techniques to Represent Text Data	6
	2.1.1 1-Hot Encoding	6
	2.1.2 N-Gram Model	7
	2.1.3 Bag-of-Words	8
	2.1.4 Vector Semantics	9
	2.1.5 Word2Vec	10
	2.1.6 GloVe Language Model	11

2.2	Approaches to Text Summarization	12
2.3	Extractive Text Summarization	13
2.3.1	Word Frequency	14
2.3.2	Term Frequency – Inverse Document Frequency	14
2.3.3	TextRank	15
2.3.4	Latent Semantic Analysis	17
2.4	Evaluation of Summarization	18
2.4.1	Bilingual Evaluation Understudy (BLEU)	21
2.4.2	Recall-Oriented Understudy for Gisting Evaluation (ROUGE)	21
2.5	Web Scraping	27
2.5.1	Techniques for Web Scraping	24
2.6	Twitter	24
2.6.1	Fetching Trends from Twitter	27
2.6.2	Analysing Trends and Context	29
3	CHAPTER III – ANALYSIS AND DESIGN	30
3.1	Functional Requirements	30
3.2	Non-functional Requirements	30
3.3	Selection of Summarization and Evaluation Algorithms	31
3.4	Components	32
3.5	General Flow of Execution	35

4	CHAPTER IV - IMPLEMENTATION DETAILS	37
	4.1 Software Requirements	37
	4.2 Hardware Requirements	37
	4.3 Libraries and Functions	37
	4.4 Steps Followed for Individual Modules	39
	4.4.1 Summarization	39
	4.4.2 Web Scraping	44
	4.4.3 Twitter	46
5	CHAPTER V – RESULTS AND DISCUSSION	48
6	CHAPTER VI – CONCLUSION AND FUTURE SCOPE	55
	REFERENCES	56
	APPENDIX	59
	ACKNOWLEDGEMENT	61