**Evaluation Only. Created with Aspose.Words. Copyright 2003-2022 Aspose Pty Ltd.**

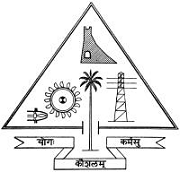
**PERSONALITY ANALYSIS FROM TWITTER**

**TWEETS**

Mini project submitted in partial fulfillmentof the requirements for the award of the degree of**Master of Computer Applications** of the**APJ Abdul Kalam Technological University**

submitted by

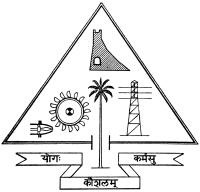
**ANITTA ANTONY (TCR19MCA029) DEVIKA K J (TCR19MCA015) JEEVAN JOSE (TCR19MCA018)**



**DEPARTMENT OF COMPUTER APPLICATIONS GOVERNMENT ENGINEERING COLLEGE THRISSUR - 680009**

DECEMBER 2021

**DEPARTMENT OF COMPUTER APPLICATIONS GOVERNMENT ENGINEERING COLLEGE, THRISSUR** THRISSUR, KERALA STATE, PIN 680009



**CERTIFICATE**

This is to certify that the mini project titled **” PERSONALITY ANALYSIS FROM TWITTER TWEETS”**is a bonafidework done by

**ANITTA ANTONY (TCR19MCA029) DEVIKA K J (TCR19MCA015) JEEVAN JOSE (TCR19MCA018)**

under my supervision and guidance, and is submitted in January 2020 in partial

fulfillment of the requirements for the award of the Degree of Master of Computer Applications from APJ Abdul Kalam Technological University(KTU).

**Project Guide Project Coordinator Head of the Dept.** Dr.Sminesh C N Husain Ahamed Dr.Sminesh C N

Place : THRISSUR Date : 30/12/2021

**DECLARATION**

I hereby declare that the main project named, **Personality Analysis from Twitter Tweets**, is my own work and that, to the best of my knowledge and belief, it contains no material previously published another person nor ma- terial which has been accepted for the award of any other degree or course of the university or any other institute of higher learning, except where due acknowledgement and reference has been made in the text.

Signature

**ANITTA ANTONY (TCR19MCA029) DEVIKA K J (TCR19MCA015) JEEVAN JOSE (TCR19MCA018)**

**ACKNOWLEDGEMENT**

WewouldliketothankComputerApplicationDepartmentofGECThris- sur, for giving us this opportunity to pursue this project and successfully complete it.

We are highly indebted to our guide, **Dr. Sminesh C N** for her guidance and constant supervision as well as for providing necessary information regarding the project and also for her support in completing the project.

We express our heart-felt gratitude to project coordinator, **Prof. Hussain Ahamed**, for her committed guidance, valuable suggestions, constructive criticisms and precious time that she invested throughout the work.

We also express special thanks to the Head of the Computer Applica- tions Department, **Dr. Sminesh C N**, for his keen support and consistent encouragement in our academic activities.

We sincerely thank all other faculties of MCA department for guiding through the processes involved in the project .

**ABSTRACT**

Thesocialmedianetworksareanonlineforumthatisusedtoimprovesocial relationships with others by allowing people to share their thoughts, emo- tions, and experiences, among other things. Social media’s ever-increasing use makes data about human behavior readily available. Twitter is one of these social media platforms. The tweets and profile information of users can be used to predict their personalities. Nowadays, many organizations have also started shortlisting the candidates based on their personality as thisincreasethee fficiencyoftheworkbecausethepersonisworkinginwhat heisgoodatthanwhatheisforcedtodo. Thisprojectisbasedonanalysisof personalityofaTwitteruserbasedonthewordsusedintweetspostedbythe user. The personality type is analysed based on Big Five Personality Model that outputs agreeableness, conscientiousness, openness, neuroticism, and extraversion as personality traits. The personality of a human plays a major role in his personal and professional life.

**CONTENTS**

**List of Figures vii Nomenclature viii 1 INTRODUCTION 1 2 ENVIRONMENTAL STUDY 3**

1. System Configuration . . . . . . . . . . . . . . . . . . . . . . . . 3
2. Hardware Requirements . . . . . . . . . . . . . . . . . . 3
3. Software Requirements . . . . . . . . . . . . . . . . . . 3
4. Software Specification . . . . . . . . . . . . . . . . . . . . . . . 4
5. Python . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
6. Windows 10 . . . . . . . . . . . . . . . . . . . . . . . . . 4
7. HTML . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
8. JavaScript . . . . . . . . . . . . . . . . . . . . . . . . . . 5
9. Python 3 . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
10. Flask . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
11. Sklearn . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
12. NLTK . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

**3 SYSTEM ANALYSIS 8**

1. Requirements Analysis . . . . . . . . . . . . . . . . . . . . . . . 8
2. Existing System . . . . . . . . . . . . . . . . . . . . . . . . . . . 9
3. Proposed System . . . . . . . . . . . . . . . . . . . . . . . . . . 9
4. List of Modules with description . . . . . . . . . . . . . 9
5. Web Application . . . . . . . . . . . . . . . . . 10
6. Data Collection . . . . . . . . . . . . . . . . . . 10
7. Data Preprocessing . . . . . . . . . . . . . . . 10
8. Analyse personality and plot result . . . . . . 11
9. Data Flow Diagram . . . . . . . . . . . . . . . . . . . . . 11
10. Level 0 dfd . . . . . . . . . . . . . . . . . . . . 12
11. Level 1 dfd . . . . . . . . . . . . . . . . . . . . 12
12. Feasibility Study . . . . . . . . . . . . . . . . . . . . . . . . . . . 13
13. Technical Feasibility . . . . . . . . . . . . . . . . . . . . 13
14. Economical Feasibility . . . . . . . . . . . . . . . . . . . 13
15. Behavioral Feasibility . . . . . . . . . . . . . . . . . . . 14

**4 SYSTEM DESIGN 15**

1. Architecture . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
2. Input Design . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16

4.2.1 Output Design . . . . . . . . . . . . . . . . . . . . . . . 16

**5 SYSTEM IMPLEMENTATION 17**

1. Dataset . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
2. Twitter Data collection . . . . . . . . . . . . . . . . . . . . . . . 17
3. Twitter Developer account . . . . . . . . . . . . . . . . . 17
4. Twitter authentication . . . . . . . . . . . . . . . . . . . 18
5. Tweepy library . . . . . . . . . . . . . . . . . . . . . . . 18
6. Tweets from user(Harry Kane- English footballer) . . . 18
7. Pre-processing . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
8. Tweets after removing emojis . . . . . . . . . . . . . . . 20
9. Tweets after removing url . . . . . . . . . . . . . . . . . 20
10. Tweets after removing hashtags and usernames . . . . 21
11. Tweets after stemming and tokenization . . . . . . . . . 21
12. Language translation . . . . . . . . . . . . . . . . . . . . . . . . 21
13. Wordfrequencycountofeachpersonalitycategoriesintwitter

tweets . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22

1. LIWC . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
2. LIWC Dictionaries . . . . . . . . . . . . . . . . . . . . . 24
3. Calculating LIWC Scores . . . . . . . . . . . . . . . . . 25
4. Average of each traits of user . . . . . . . . . . . . . . . . . . . 27
5. Matplotlib . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28
6. Ploting the results . . . . . . . . . . . . . . . . . . . . . . . . . . 28
7. sklearn . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
8. Analysing personality traits by verifying with threshold values 30
9. Analysis based on job description . . . . . . . . . . . . . . . . . 31
10. Testing . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31
11. Black box testing . . . . . . . . . . . . . . . . . . . . . . 31
12. Whitebox Testing . . . . . . . . . . . . . . . . . . . . . . 32

**6 RESULTS AND DISCUSSION 33 7 CONCLUSION 34 8 SCOPE FOR THE FUTURE ENHANCEMENT 35 9 SCREENSHOTS 36 BIBLIOGRAPHY 40**

**LIST OF FIGURES**

1. Level 0 dfd . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12
2. Level 1 dfd . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

4.1 Architecture diagram . . . . . . . . . . . . . . . . . . . . . . . . 15

1. Tweets from user . . . . . . . . . . . . . . . . . . . . . . . . . . 18
2. Tweets from user before preprocessing . . . . . . . . . . . . . . 19
3. Tweets after removing emojis . . . . . . . . . . . . . . . . . . . 20
4. Tweets after removing url . . . . . . . . . . . . . . . . . . . . . 20
5. Tweets after removing hashtags and usernames . . . . . . . . 21
6. Tweets after stemming and tokenization . . . . . . . . . . . . . 21
7. Word frequency count of openness category . . . . . . . . . . . 22
8. Word frequency count of conscientiousness category . . . . . . 22
9. Word frequency count of extraversion category . . . . . . . . . 23
10. Word frequency count of agreeableness category . . . . . . . . 23
11. Word frequency count of neuroticism category . . . . . . . . . 24
12. LIWC Dictionary . . . . . . . . . . . . . . . . . . . . . . . . . . 25
13. LIWC Score of openness category . . . . . . . . . . . . . . . . . 25
14. LIWC Score of conscientiousness category . . . . . . . . . . . . 26
15. LIWC Score of extraversion category . . . . . . . . . . . . . . . 26
16. LIWC Score of agreeableness category . . . . . . . . . . . . . . 27
17. LIWC Score of neuroticism category . . . . . . . . . . . . . . . 27
18. Average of each traits of user . . . . . . . . . . . . . . . . . . . 28
19. Graph of openness . . . . . . . . . . . . . . . . . . . . . . . . . 28
20. Graph of conscientiousness . . . . . . . . . . . . . . . . . . . . 29
21. Graph of extraversion . . . . . . . . . . . . . . . . . . . . . . . 29
22. Graph of agreeableness . . . . . . . . . . . . . . . . . . . . . . . 29
23. Graph of neuroticism . . . . . . . . . . . . . . . . . . . . . . . . 30
24. Analysing personality traits . . . . . . . . . . . . . . . . . . . . 30
25. Analysing user based on job description . . . . . . . . . . . . . 31
26. Main Page . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36
27. Enter twitter ID . . . . . . . . . . . . . . . . . . . . . . . . . . . 36
28. Plot results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 37

**NOMENCLATURE**

DFD Data Flow Diagram

HTML Hypertext Markup Language LIWC Linguistic Inquiry and Word Count NLP Natural Language processing NLTK Natural Language Toolkit

**CHAPTER 1 INTRODUCTION**

Nowadays, social media is the most popular environment among people. Most of the people use social media to share their emotions, daily life ac- tivities, ideas about several events (e.g., political, agenda topics) in the form of photos or texts. Twitter is one of the most widely used microblogging and social networking services that reflects people’s emotions as texts. If somebody is angry, happy or sad about an event, he or she generally prefers sharing these feelings by posting a tweet. It is a tool for people to show a reaction to events.

Personality is the status of human characteristics and habits in their daily lives. Detecting personality on Twitter is important because the data neededforanalysisarelargeandopen,andtraditionalpersonalitydetection processes such as interviews with psychologists are expensive and time- consuming. The results of personality detection can be used to improve the accuracy of the recommendation system, and make it easier for recruiters to place appropriate work positions.

This project aims to analysis the personality of a user from twitter tweets by using an intelligent personality assessment model. A personality model that is called Big Five Personality Model is used for the assessment of the proposedmodel.TheBigFivepersonalitytraitsisaclassificationfordi fferent personalities, which is widely accepted and used within scientific research inpsychology. Thefivepersonalitytraits,referredtoasOCEAN,correspond to Openness, Conscientiousness, Extraversion, Agreeableness and Neuroti- cism.



In this project, At first Data of a user is collected from Twitter using twitter API and then the data is preprocessed by eliminating user men- tions,emojis,url,stop words etc and also the NLP processes such as Tok- enization, Stemming, Lemmatization are done.Then we use LIWC to find word frequency and calculates the scores for each personality traits and plot the graph. Then a threshhold value is calculated by analysing personality traitsof30personsandanalysisthepersonalityofapersonaccordingtoeach personality traits. Also the bio of a person from his twitter account is also accessed and make an analysis from user job description which suggests in which personality traits that they should improve.



**CHAPTER 2 ENVIRONMENTAL STUDY**

1. **System Configuration**

System configuration describe the hardware and software requirement of the system for development.

1. **Hardware Requirements**

* Memory : 4 GB of RAM
* Processor : Intel Core i3 or equivalent CPU
* Speed : 2.4 GHz
* Proper Internet Connection

1. **Software Requirements**

* Operating system : Windows
* Front End :HTML, Javascript
* Back End : Flask
* IDE Used : Jupiter notebook, Google collab
* Data processing : LIWC
* Machine Learning Framework : Sklearn
* ToolKit used : Nltk



1. **Software Specification**
2. **Python**

Python is an interpreted, high-level and general-purpose programming lan- guage. Python can be used on a server to create web applications. Python works on di fferent platforms (Windows, Mac, Linux, Raspberry Pi, etc). Python runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick.Python can be treated in a procedural way, an object-oriented way or a functional way.

1. **Windows 10**

Windows 10 is a series of personal computer operating systems produced by Microsoft as part of its Windows NT family of operating systems. It is the successor to Windows 8.1, and was released to manufacturing on July 15, 2015, and broadly released for retail sale on July 29, 2015.Windows 10 receivesnewbuildsonanongoingbasis,whichareavailableatnoadditional cost to users, in addition to additional test builds of Windows 10 which are available to Windows Insiders. The latest stable build of Windows 10 is Version 1909 (November 2019 Update). Devices in enterprise environments can receive these updates at a slower pace, or use long-term support mile- stones that only receive critical updates, such as security patches, over their ten-year lifespan of extended support.

Windows 10 received mostly positive reviews upon its original release in July 2015. Critics praised Microsoft’s decision to provide a desktop- orientedinterfaceinlinewithpreviousversionsofWindows,contrastingthe tablet-oriented approach of 8, although Windows 10’s touch-oriented user interface mode was criticized for containing regressions upon the touch- oriented interface of Windows 8.



1. **HTML**

Hypertext Markup Language (HTML) is the standard markup language for documents designed to be displayed in a web browser. It can be assisted by technologies such as Cascading Style Sheets (CSS) and scripting languages such as JavaScript. Web browsers receive HTML documents from a web server or from local storage and render the documents into multimedia web pages. HTML describes the structure of a web page semantically and orig- inally included cues for the appearance of the document. HTML elements are the building blocks of HTML pages. With HTML constructs, images and other objects such as interactive forms may be embedded into the rendered page.

1. **JavaScript**

JavaScript , often abbreviated as JS, is a programming language that con- forms to the ECMAScript specification.JavaScriptis high-level, often just in time compiled, and multi-paradigm. It has curly-bracket syntax, dynamic typing, prototype-based object-orientation, and first-class functions. Along side HTML and CSS, JavaScript is one of the core technologies of the World Wide Web. JavaScript enables interactive web pages and is an essential part of web applications. The vast majority of websites use it for client-side page behavior, and all major web browsers have a dedicated JavaScript engine.

1. **Python 3**

PythonwasdevelopedbyGuidovanRossumintheearly1990sanditslatest version is 3.7.1, we can simply call it as Python3. Python 3.0 was released in 2008. andisinterpretedlanguagei.eit’snotcompiledandtheinterpreterwill checkthecodelinebyline. ThisarticlecanusedtolearnverybasicsofPython programminglanguage.Pythonisageneral-purposeinterpreted,interactive, object-oriented, and high-level programming language.Like Perl, Python source code is also available under the GNU General Public License (GPL). Python is named after a TV Show called eMonty¨ Python´ıs Flying Circus´ı



and not after Python-the snake.

1. **Flask**

Flask is a micro web framework written in Python. It is classified as a mi- croframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other compo- nents where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tool.

1. **Sklearn**

Scikit-learn(Sklearn)isthemostusefulandrobustlibraryformachinelearn- ing in Python. It provides a selection of e fficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. This library, which is largely written in Python, is built upon NumPy, SciPy and Mat- plotlib.

1. **NLTK**

The Natural Language Toolkit, or more commonly NLTK, is a suite of li- braries and programs for symbolic and statistical natural language process- ing (NLP) for English written in the Python programming language. It was developed by Steven Bird and Edward Loper in the Department of Com- puter and Information Science at the University of Pennsylvania. NLTK includes graphical demonstrations and sample data. It is accompanied by a book that explains the underlying concepts behind the language processing taskssupportedbythetoolkit,plusacookbook.NLTKisintendedtosupport research and teaching in NLP or closely related areas, including empirical linguistics, cognitive science, artificial intelligence, information retrieval, and machine learning. NLTK has been used successfully as a teaching tool, as an individual study tool, and as a platform for prototyping and build- ing research systems. NLTK consists of the most common algorithms such as tokenizing, part-of-speech tagging, stemming, sentiment analysis, topic segmentation, and named entity recognition. NLTK helps the computer to analysis, preprocess, and understand the written text.



**CHAPTER 3 SYSTEM ANALYSIS**

**This document was truncated here because it was created in the Evaluation Mode.**