

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi-590 018



A Project Work on

“CAREER GUIDANCE SYSTEM”

A Dissertation work submitted in partial fulfillment of the requirement
for the award of the degree

Bachelor of Engineering
In
Information Science & Engineering

Submitted by

Ms. Rakshita P R	1AY16IS087
Ms. Shreya M Nadkarni	1AY16IS097
Ms. Varshini R	1AY16IS116
Mr. Rajan Raj Das	1AY16IS128

Under the guidance of

Prof. Arpitha N S
Assistant Professor



DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING
ACHARYA INSTITUTE OF TECHNOLOGY

Affiliated to Visvesvaraya Technological University, Belagavi, Approved by AICTE and Accredited
by NAAC, New Delhi)

Acharya Dr. Sarvepalli Radhakrishnan Road, Soldevanahalli, Bangaluru-560107

2019-20

DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING
ACHARYA INSTITUTE OF TECHNOLOGY

(Affiliated to Visvesvaraya Technological University, Belagavi, Approved by AICTE and Accredited
by NAAC, New Delhi)

Acharya Dr. Sarvepalli Radhakrishnan Road, Soldevanahalli, Bengaluru-560107
2019-2020



Certificate

Certified that the Project work entitled **“CAREER GUIDANCE SYSTEM”** is a bonafide work carried out by **Ms. Rakshita P R (1AY16IS087), Ms. Shreya M Nadkarni (1AY16IS097), Ms. Varshini R (1AY16IS116) and Mr. Rajan Raj Das (1AY16IS128)** in partial fulfillment for the award of the degree of **Bachelor of Engineering in Information Science and Engineering** of the **Visvesvaraya Technological University**, Belgaum during the year 2019-20. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the Report deposited in the departmental library. The Project has been approved as it satisfies the academic requirements in respect of Project work prescribed for the Bachelor of Engineering Degree.

Prof. Arpitha N S
Guide

Prof. C K Marigowda
HOD-ISE

Principal

DECLARATION

We, Rakshita P R (1AY16IS087), Shreya M Nadkarni (1AY16IS097), Varshini R (1AY16IS116) and Rajan Raj Das (1AY16IS128) Students of B.E, Information Science and Engineering, Acharya Institute of Technology, Bangalore – 107 hereby declare that the project entitled “Career Guidance System “is an authentic record of our own work carried out under the supervision and guidance of Prof. Arpitha N S, Assistant Professor, Department of Information Science and Engineering, Acharya Institute of Technology, Bangalore. We have not submitted the matter embodied to any other university or institution for the award of any other degree .

Name	USN	Signature
Rakshita P R	1AY16IS087	-----
Shreya M Nadkarni	1AY16IS097	-----
Varshini R	1AY16IS116	-----
Rajan Raj Das	1AY16IS128	-----

Place: Bangalore

Date:

ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of this project would be incomplete without the mention of the people who made it possible through constant guidance and encouragement.

We would take this opportunity to express our gratitude to **Sri. Premnath Reddy**, Chairman, Acharya Institutes and **Dr Prakash M R**, Principal, Acharya Institute of Technology for providing the necessary infrastructure to complete this project.

We wish to express our deepest gratitude and thanks to **Prof. C K Marigowda**, Head of The Department, Information Science and Engineering.

We wish to express sincere thanks to our guide **Prof. Arpitha N S**, Assistant Professor, Department of Information Science and Engineering for helping us throughout and guiding us from time to time.

A warm thanks to all the faculty of Department of Information Science and Engineering, who have helped us with their views and encouraging ideas.

ABSTRACT

Guidance systems have widespread applications in both academics and industry. Traditionally the performance of a guidance system has been measured by precision. The proposed system is focused on specializations that can be undertaken by engineers through reliable and flexible online tests that are adaptive in nature. The test will start with aptitude questions related to the field and then based on the inputs from the user and accuracy of the input, the user will be provided with the next set of questions and this adaptive nature continues throughout the test. On completion of the test, the users shall get a report of their performance based on which the system will provide a career path in which the user is more likely to succeed. The adaptive nature and the end result report using tableau are the two major components that will be the highlight of our project.

TABLE OF CONTENTS

ACKNOWLEDGEMENT	i
ABSTRACT	ii
LIST OF FIGURES	vi
LIST OF TABLES	vii

CHAPTERS	PAGE NO.
CHAPTER 1 INTRODUCTION	1
1.1 Overview	1
1.2 Factors that influence career choice	1
1.3 System Introduction	2
CHAPTER 2 DETAILED SURVEY.....	4
2.1 Paper Explanation	4
2.1.1 Extracting regular behaviors from social media	4
2.1.2 Feature selection based on behavior	4
2.1.3 Career guidance and employee management	5
2.1.4 Personal Interests on Social Media Content	5
2.1.5 Understanding What Affects Career Progression Using Twitter	5
2.1.6 Development of a Web-based Building Profession Career Portal	6
2.1.7 A game based approach to career guidance	6
2.1.8 Career counselling using data mining	7
2.1.9 Online career Guidance System	7
2.1.10 Career prediction model using Linear classification	7

2.2 Summary of papers.....	8
CHAPTER 3 PROBLEM DEFINITION.....	12
3.1 Motivation of the project.....	12
3.2 Problem statement	12
3.3 Existing system	12
3.3 Proposed system.....	13
3.3 Objectives of proposed system.....	13
3.3 Advantages of the system.....	14
CHAPTER 4 SYSTEM ANALYSIS	15
4.1 Introduction To System Requirements Specification.....	15
4.2 Functional Requirements.....	15
4.3 Non-Functional Requirements	16
4.3.1 Performance.....	16
4.3.2 Availability	16
4.3.3 Reliability	16
4.3.4 Usability.....	16
4.3.5 Portability	16
4.3.6 Flexibility	17
4.4 Specific requirements.....	17
4.5 External interface requirements	18
4.5.1 User Interfaces	18
4.5.2 Hardware interfaces	19
4.6 Hardware requirements	19
4.7 Software requirements	19

CHAPTER 5 OBJECTIVES AND METHODOLOGY OF THE PROPOSED SYSTEM

.....	20
5.1 Proposed system.....	20
5.2 Objectives of the proposed system.....	21
5.3 Advantages of the proposed system.....	21
5.4 Methodology of the proposed system	22
5.4.1 Phases of career prediction	22
5.4.2 Development Process	23
5.4.3 Datasets used by the system	24
5.5 Methodology used in the system.....	28
5.5.1 PyQt.....	28
5.5.2 Tweepy	29
5.5.3 Importance of personality types	30
5.5.4 Naïve Bayes Classification.....	31
5.5.5 NLP for cleaning tweets	33

CHAPTER 6 DETAILED DESIGN35

6.1 System Overview	35
6.2 High level design.....	36
6.3 System data flow	37
6.4 Sequence diagram	39

CHAPTER 7 IMPLEMENTATION.....40

7.1 List of Modules identified along with their functions.....	40
7.2 Control flow between each unit	40
7.3 Aptitude Questionnaire	41
7.3.1 The launch GUI.....	41

7.3.2 User Registration Page	42
7.3.3 UI for aptitude	44
7.3.4 Result obtained from test	47
7.4 Twitter Analysis	52
7.4.1 UI to enter twitter handle	52
7.4.2 Getting twitter id's of who user follows	54
7.4.3 UI for getting career using machine learning	56
7.4.4 Analyzing tweets from users	58
7.5 Myers-Briggs Personality test	62
7.5.1 UI for personality input	62
7.6 Final Result	64
7.6.1 UI for final result	64
CHAPTER 8 RESULTS (SNAPSHOTS)	67
CHAPTER 9 TESTING	74
9.1 Testing Principle	74
9.2 Testing Methods	74
9.3 Levels of Testing	74
9.3.1 Unit Testing	75
9.3.2 Integration Testing	79
9.3.3 System Testing	80
CONCLUSION AND FUTURE ENHANCEMENT	82
REFERENCES.....	83

LIST OF FIGURES

FIGURE NO.		PAGE NO.
Figure 5.1	Proposed system	20
Figure 5.2	Development Process	24
Figure 5.3	Dataset for technical questionnaire	25
Figure 5.4	Dataset for mapping	26
Figure 5.5	Positively labelled dataset	27
Figure 5.6	Negative labelled dataset	27
Figure 5.7	Personality types and careers	30
Figure 5.8	Classification of tweets	34
Figure 6.1	System overview	35
Figure 6.2	High level design	36
Figure 6.3	System data flow	38
Figure 6.4	Sequence diagram	39
Figure 8.1	Front Page	67
Figure 8.2	Registration page	67
Figure 8.3	Instructions page	68
Figure 8.4	Questionnaire page	68
Figure 8.5	Result	69
Figure 8.6	Twitter based occupation	69
Figure 8.7	Personality test	70
Figure 8.8	Personality test url directed	70
Figure 8.9	Type of personality	71
Figure 8.10	Final Result	71
Figure 8.11	Career analysis using machine learning	72

Figure 8.12	Positive sentiment using twitter ID	72
Figure 8.13	Negative sentiment using text analyzer	73
Figure 8.14	Positive sentiment using text analyzer	73

LIST OF TABLES

TABLE NO.		PAGE NO.
Table 2.1	Summary of literature survey	11
Table 5.1	Tweets labelled with careers	33
Table 5.2	Training data	33
Table 5.3	Output of classification	33
Table 9.1	Unit testing	79
Table 9.2	Integration Testing	80
Table 9.3	System Testing	81

Chapter 1

INTRODUCTION

1.1 Overview

Students face a lot of problems in choosing their career path that matches their strengths, values, personality and skills. In most cases, unawareness of the opportunities holds them back from opting an appropriate career. Thus in general, choosing a career can prove to be daunting. Anything from fear to lack of awareness can toss up a roadblock on the path to a career decision. Some students believe that only one perfect career exists for an individual. However, there are many careers that share the same characteristics. Henceforth, there is a need for a system that guides the students to overcome the hurdles while opting a career path. Guidance systems have widespread applications in both academics and industry. Traditionally the performance of a guidance system has been measured by precision.

Supporting students in making well informed choices about subjects can lead them to have a more optimistic outlook on life, sense of purpose and greater level of contribution that they make to their families and society. Career education and guidance play an important role in curriculum that supports:

- Students' interests, strengths and aspirations
- Students' achievements
- Students at risk of poor outcomes
- Students making informed decisions about their subject choices and pathways.

1.2 Factors that influence career choices

There are multiple factors that influence career choices , but among the most important factors are :

- Aptitude tests: Aptitude tests are used to determine an individual's ability/potential to succeed in a certain task, with no prior knowledge or training. Aptitude tests can be used

in school exams and are frequently used as part of a pre-employment assessment. An aptitude test measures your ability to reason and learn new skills.

- **Personality:** Several recent studies take a different approach by suggesting that understanding your personality traits and characteristics are what is key to career fulfilment and success. Using a Myers Briggs or Holland Code test to discover your strongest personality traits. These tests tend to define you as for example, a ‘realistic’ type or an ‘artistic’ type. Based on your tendency to lean towards a certain type the tests then provide a list of careers that might suit you best.
- **Social media interest:** Personality has been found to significantly correlate with a number of real-world behaviours. It correlates with music taste: popular music tends to be significantly liked by extroverts, while people with a tendency to be less open to experience tend to prefer religious music and dislike rock music. Personality also impacts the formation of social relations: friends tend to be, to a very similar extent, open to experience and extroverts. Personality also influences how people interact online.

1.3 System Introduction

A Career Guidance System is a comprehensive system which helps students to choose an appropriate career based on their personalities and intelligence capabilities. It predicts various career options best suitable for an individual. This system boosts the confidence in a student by providing him the career option where he can excel and succeed. The system predicts the appropriate careers to the students through an aptitude test, also takes into consideration the personality and social media interests of the user. An aptitude test is an exam used to determine an individual's propensity to succeed in a given activity.

Aptitude tests assume that individuals have inherent strengths and weaknesses, and have a natural inclination toward success or failure in specific areas based on their innate characteristics. Some schools administer aptitude tests to students beginning in elementary school. Along with intelligence tests and achievement tests that measure student mastery of academic content, aptitude tests may be used to determine placement in gifted and talented programs or other specific educational tracks. For example, the Modern Language Aptitude Test

measures a student's potential for successfully mastering foreign languages. Aptitude tests can also help determine if a student needs special education services. Students who score well in tests measuring speed, accuracy and problem solving might decide to choose coursework in computer science, finance, or other fields requiring attention to detail .The aptitude test that is designed for our system consists of certain basic questions from various career choices such as mathematics, English, science, sports, chef, engineering, medicine and so on .The user is expected to complete these series of questions before moving on to the next step.

The next deciding factor is the personality, according to Isabel Briggs Myers there are 16 personalities and each of these personalities are associated with a set of careers. The users can get to know their personality by undertaking a personality test .A personality test is a method of assessing human personality constructs. Research on the importance of personality and intelligence in education shows evidence that when others provide the personality rating, rather than providing a self-rating, the outcome is nearly four times more accurate for predicting grades. The user is expected to enter his/her personality type into our system. Once the personality type is entered the system maps the personality with its respective set of associated careers. The third and the final deciding factor is the social media interests. Studies have proven that a person's content , the comments he posts and most importantly the people he follows helps in analysing the interests of the person.

The social media platform that is used to analyze the interests in our system is Twitter .Twitter has proven to be one of the best platforms for personality and behavioral analysis of people. Our system communicates with twitter using API calls to the twitter server that are made using the unique twitter key . The users will have to enter their respective twitter handle. Once the twitter handle is obtained the tweets of the person will be analysed to get the interests of the student. After the user goes through all the three modules of the system a set of appropriate careers will be predicted to the user by mapping the results obtained from the aptitude test, personality type and analysing their interests through twitter. Our system has another version to predict careers based on NLP and Naive Bayes algorithm on the tweets of the user.

Chapter 2

DETAILED LITERATURE SURVEY

2.1 Paper Explanation

2.1.1 [2011] Extracting regular behaviours from social media networks

Authors: Leiming YAN and Jinwei WANG

Social media network analysis has become very popular in recent years. In order to extract behaviors occurring regularly to reveal the microscopic evolving properties in social networks, the evolving process of networks is modeled as stochastic states transition, and the evolving behaviors are described as topological structure changes of a series of subgraphs. Then, based on Maximal Frequent Subgraph mining technology, RBMiner (Regular Behaviors Miner) algorithm is proposed to identify such regular behaviors in network dataset. The empirical evaluation using both synthetic and real dataset verifies that the proposed algorithm is valid, and the regular behavior patterns show more dynamic information hidden in evolving social networks than normal frequent subgraph patterns.

2.1.2 [2012] Feature Selection Based on Audience's Behavior for Information Filtering in

Online Social Networks

Author: AruneeRatikan and MikifumiShikda

This paper aims to prepare necessary features used in filtering information to get consistent information with the audience's situation. It studies behavior of audience in several aspects by analysis characteristic of post. The result indicates that three features are essential for the audience's decision to allow a post to be displayed on the audience's social network page, which are n1 (topic of post), n0 (audience's current situation), and n2 (type of relationship between audience and creator).

2.1.3 [2014]Career Guidance and Employment Management System

Author:Kasem Seng and Akram M. Zeki

This paper presents a design of a web based system for Career Guidance and Employment Management System (CGEMS). Basically, CGEMS is tries to help a user who is looking for career advice, guidance, or looking for their opportune job. CGEMS tries to make one place where student, job seeker, career advisor or consultant, and various company or organization can meet and help. In order to help the users to determine their best career choice, this system also offers some tests or quizzes that are related to such career and the user personality which will be useful for the career path. Furthermore, users of CGEMS such as company or organization can post a job description, when they look for employees.

2.1.4 [2017] FINDEREST: Identifying Personal Skills and Possible Fields of Study Based on Personal Interests on Social Media Content.

Authors: NapatsakornPisalayon, Jiraphan Sae-Lim, NoppakitRojanasit, PawitraChiravirakul

Students of our generation are having a hard time matching their real interests and preferences with skills needed in each field of study. High school students usually do not understand the application of these skills with the respective career fields. This study aimed to define personal skills from social media use behavior; in particular, Facebook feed content.

2.1.5 [2017] Understanding What Affects Career Progression Using LinkedIn and Twitter Data

Authors: Yiming Pan, Xuefeng Peng, Tianran Hu, and Jiebo Luo

The main objective of this paper is to determine how factors such as personality, industry and education background impact one's career path, and the highest career stage one could reach. This can be done by associating individuals' career paths with their education backgrounds, unique thinking styles, interests, and personalities by analyzing extensive users from Social Media. Finally, machine learning techniques are employed to predict career progression with a promising accuracy.

2.1.6 [2017]Development of a Web-based Building Profession Career Portal as a Guidance Information System for Secondary School Students.

Author: Afolabi, Adedeji Ojelabi, Rapheal, Amusan, Lekan Adefarati, Fred

Career selection is one of many difficult task students are faced with when leaving secondary schools. The aim of the study was to develop a web-based building profession career portal as a guidance information system for secondary school students with an aim to bring more awareness to the profession of building/building technology. The study utilized the use of a questionnaire and the system design of a web-based career portal. Statistical tools of charts and mean scores were used. In conclusion, a web-based building profession career portal as a guidance information system for secondary school students was developed. The study recommended that less subscribed professions should find ways to educate secondary school students, teachers and parents. This can be actualized through the use of career web-based portals that have detailed information about the course.

2.1.7 [2017]A Game-Based Approach to Career Guidance

Author: Ian Dunwell, Petros Lameris, Sara de Freitas, Panos Petridis, Maurice Hendrix, Sylvester Arnab

Understanding the relationship between immediate decisions and long-term career prospects is a vital skill in a job market which is becoming increasingly globalised, competitive and dynamic. The application of digital technologies and multimedia to career-guidance offers the potential to go beyond the confines of the classroom and provide a depth of insight into career decisions and their consequences in interactive and engaging forms. This paper reports on an approach which uses a serious game as an approach to career guidance incorporating game-based elements such as character development and decision-making alongside a visual and user-centric approach to fostering engagement.

2.1.8 [2017]Career Counselling Using Data Mining

Authors:Nikita Gorad1 , IshaniZalte , Aishwarya Nandi , DeepaliNayak

Selecting an appropriate career is one of the most important decisions and with the increase in the number of career paths and opportunities, making this decision has become quite difficult for the students.. It may lead to wrong career selection and then working in a field which was not meant for them, thus reducing the productivity of human resource.rong career selection.This system is a web application that would help students studying in high schools to select a course for their career. The system would recommend the student, a career option based on their personality trait, interest and their capacity to take up the course

2.1.9 [2017]Online Career Guidance System

Authors: Crystal D'Mello, RiniAranha, Boni Gregory, VarshaShrivastava

It is an online career guidance system where students can choose from various courses after taking a quiz on their personalities.It also lists various colleges available where students can search colleges by their courses. Next the system allows users to give a test. The tests have two levels. The first level has questions which help us to identify which field the student is interested in and based on this the second level will accordingly ask questions from their field. The first level has the same questions for all the standards as it is not based on knowledge but on the interests. The second level questions will be from that field in which the student is interested the most, along with questions from other fields. After the successful completion of both the levels, the student will receive his result along with a detailed explanation and advice from us. The explanation will state the reason as to why we feel a certain career is better for him/her and why not the other.

2.1.10 [2018] Career Prediction Model Using Data Mining and Linear Classification

Authors :Rucha Hemant Rangnekar, Khyati Pradeep Suratwala, Sanjana Krishna, Dr. SudhirDhage.

At a certain stage of life, one has to choose a career and this choice can shape one's life entirely. Recently, more and more people have begun to re-evaluate their career decisions and change

careers at a later stage in life. This can be prevented by proper counseling of young teenagers before they begin their graduate studies. To solve this problem, data of existing students is used, where the personalities, aptitude and student general information is mapped with their careers. The Intuitive Career System uses a variety of questions that students have to answer to test their aptitude, personality as well as students background questions.

2.2 Summary of papers

SL.NO	Title of Paper	Problem Addressed	Authors Approach/ Method	Results
1.	Career Prediction Model Using Data Mining and Linear Classification	Re-evaluate career or change career at a later point of time.	The PCA, pre-processing and pattern recognition, KNN algorithm are used for mapping with cluster which was formed from predefined dataset.	The Intuitive Career System will present a graph focusing on strengths and weaknesses of different fields.
2.	FINDEREST: Identifying Personal Skills and Possible Fields of Study Based on Personal Interests on Social Media Content.	Define personal skills from social media use.	Facebook Graph API and Microsoft Cognitive Service API were used in order to identify terms from the feed contents and define related skills and fields by filtering those terms with predefined skills.	Participants evaluate their satisfaction with the analytical output to verify the precision of the system. 26.9% of participants highly satisfied other 73.1% of the participants felt neutral
3.	What Affects Career Decision Using Twitter Data	Determine how factors such as personality, industry and	By associating individuals' career paths with their education backgrounds, unique thinking styles, interests, and personalities	Interesting and significant correlations between

		education background impact one's career path.	by analyzing extensive users from Social Media.	personal traits and career progression are uncovered.
4.	Extracting regular behaviors from social media networks	Investigated to construct evolving behavior patterns in social networks	Investigation of properties of regular behavior using a methodology based on stochastic theory and Markov process and an efficient algorithm RBMiner is proposed to obtain regular evolving behaviors frequently occurring on social networks.	The regular evolving behaviours is obtained and thus the problem with static patterns of behaviours is solved.
5.	Career Guidance and Employment Management System	Lack of knowledge on various opportunities regarding career and dream jobs.	The Career Guidance and Employment Management System (CGEMS) offers some tests or quizzes that are related to such career and the user personality which will be useful for the career path.	Clarity regarding one's career and dreams knowing the respective strengths and weaknesses.
6.	Development of a Web-based Building Profession Career Portal as a Guidance Information System for Secondary School Students.	Career selection is one of the most difficult tasks faced by the students after secondary schooling. Most of professions are not known or given importance.	Building a web based profession career portal as a guidance information system using statistical tools of charts and mean scores.	The less subscribed professions are brought into awareness thus educating the students, parents and teachers.
7.	A Game-Based Approach to Career Guidance	Game-based enhancements to existing learning environments	character development and decision-making alongside a visual and user-centric approach to fostering engagement.	assertion that the overall approach proved effective in

		to be deployed, where serious games can be used to study the behavior of person		creating a gameplay model
8.	Feature Selection Based on Audience's Behavior for Information Filtering in Online Social Networks	Avoid overloading of irrelevant data on the user OSNs page.	Used 3 approach of information filtering that are : content-based approach , social filtering approach and hybrid collaborative approach. Waikato Environment for Knowledge Analysis (WEKA)	It analyzes the post that are created by the creator and decide whether to allow or not allow to the user OSNs page.
9.	Career Counselling Using Data Mining	Make it easier for the students they have proposed a system.	C5.0,AdaBoost is one of the boosting methods. For boosting of classifiers.	web application that can be used by any student who needs help in selecting the career path. The three set of questionnaires provided to the students based on personality traits, interests and capacity
10.	Online Career Guidance System	The guidance given to the students in not sufficient and the student's minds are always in a state of confusion to decide where their interest lies.	Guide the students to their appropriate career paths on understanding their current skill sets and also by understanding their mental abilities .	A Career Guidance System where students can see various career opportunities, the system shows various fields available after 10th, 12th for graduation

				and after graduation fields are also available.
--	--	--	--	--

Table 2.1 Summary of literature survey

Chapter 3

PROBLEM STATEMENT

In this section, we present detailed information about the project. First, motivation, problem statement, project objective, and existing system that worked against the problem and the proposed system is discussed.

3.1 Motivation for the project

The main motivation for choosing this project was the aftermath of numerous students discontinuing their studies after choosing a career path without a definite guidance and unawareness of their abilities matching a particular career path. According to the survey, one of the main reasons behind high college dropout rates was that many second-level students received no career guidance and others received it from unqualified teachers. Thus, the motivation to find the solution from the available resources can be implemented by building an user friendly system for all the students.

The main aim of this project is to design an effective, user friendly and an easily accessible application to help the students in becoming aware of the vast opportunity in choosing their career path based on their intellectual, behavioral and personal interests.

3.2 Problem Statement

The problem statement of the project is “Requisite for a career guidance system for high school students who lack exposure to different career paths based on their personality and skills.”

3.3 Existing system

Here are some details of the existing career guidance system:

- The existing systems do give general tests and consider the interests of the students, But fail to consider the personality aspects of the candidate.

- There are many online career guidance systems that predict the appropriate careers with respect to any one of the factors which include either through a series of tests , personal interests or based on the personality , but a collective system as a whole doesn't exist.
- The current system with career counsellors charge students with a fee.

3.4 Proposed system

The main objective of the career guidance system is that it helps students to choose an appropriate career based on their personalities and intelligence capabilities. It predicts various career options best suitable for an individual. This system boosts the confidence in a student by providing him the career option where he can excel and succeed. Since the system needs to provide service to a large number of students it needs a cloud infrastructure. Next the system allows users to take a series of tests consisting of various general questions. After test completion a score is calculated for each test. Based on those results the system manipulates and calculates the best career for that user. The system also monitors the behaviour of the student for long term and suggests different path. It allows for a robust and enhanced experience by handling a large number of visitors with efficiency.

3.5 Objectives of the proposed system

The main purpose of the system is to efficiently predict the career path through a fully automated system that not only clears the confusions but also gives accurate career options. It is a cost-effective method as it has no involvement of career counsellors and is a student friendly approach. The administrator of the system controls the contents of the questionnaire provided to the students. The students can login through the client computers with their username and can take the test. Along with the results of the test ,the behavioural aspects are mapped to give an accurate career options suiting one's personality whose interests are collected based on social media interactions.

3.6 Advantages of the proposed system

- The system predicts an appropriate career path.
- The system helps clear the confusions regarding the career aspects.
- The system helps the students figure out their personality and interests thus boosting their confidence.
- The system also predicts their careers through accessing the twitter ids , which helps the system analyse the interests of the student
- By looking at the different aspects such as the aptitude , personality type and the interests of the student the accuracy of the career predicted have a greater chance of being the correct career choice for the student .

Chapter 4

SYSTEM ANALYSIS

4.1 Introduction to System Requirements Specification

A software requirements specification (SRS) is a comprehensive description of the intended purpose and environment for software under development. The software requirements specification fully describes what the software will do and how it will be expected to perform.

A software requirements specification minimizes the time and effort required by developers to achieve desired goals and also minimizes the development cost. A good software requirements specification defines how an application will interact with system hardware, other programs and human users in a wide variety of real-world situations. Parameters such as operating speed, response time, availability, portability, maintainability, footprint, security and speed of recovery from adverse events are evaluated.

4.2 Functional requirements

In software engineering and systems engineering, a functional requirement defines a function of a system or its component, where a function is described as a specification of behavior between outputs and inputs. The desktop app given to the users will have the following requirements:

- Only an authorized person can access related details or can take a test and find the result.
- Users should be willing to allow access to their social media so that we can use the information for an accurate analysis and predict with precision.
- Analysis can be made depending on the different timeline at which the student approaches the system.

4.3 Non functional requirements

There are some non-functional requirements such as:

4.3.1 Performance

No. of terminals to be supported is dependent on the server that we will use at the time of deployment. The web application server used should provide good performance and ability to manage performance with techniques such as support for caching. After completing the career test the entire performance of the student will be evaluated as per the algorithms in less than a couple of seconds.

4.3.2 Availability

Online Career Guidance desktop app has 24/7 availability. It can be accessed 24 hours a day. For this UPS support must be on the server site with a backup of at least 8 hours in case of power failure. Students can take the test at any given point and can access other information at anytime. The resources for further actions after career analysis are available at all times to cater to the students' needs.

4.3.3 Reliability

It means the extent to which program performs with required precision. The website developed should be extremely reliable and secure so that information about the students profile will always remain confidential at any cost. The results are accurate and can be reliable.

4.3.4 Usability

The website should be user friendly and should require the least effort to operate. The web server used should provide services like session management to maintain sessions in the application.

4.3.5 Portability

The desktop application is flexible when used on multiple platforms or instantly accessed from the internet, desktop or network.

4.3.6 Flexibility

It is effort required to modify operational program. The whole website should be made using independent modules so that any changes done in one of the modules should not affect the other one and new modules can be added easily to increase functionality.

4.4 Specific Requirements

Since the student/users are the main target group of our site, we will only be concerned about some important functions for the user.

Students/Users:-

- The students can register for the application.
- Once the user has registered they can log in with their username and password , either to take up the tests .
- There are two categories of tests the user has to take up which includes aptitude test and the behavioural test .
- The aptitude tests include basic logical and critical thinking questions. Also includes English related questions.
- The behavioural tests includes personality based questions which help us classify the user to one of the sixteen personality types according to Carl Jung
- According to the scores in these tests the career paths shall be predicted.
- The social media details provided by the users will help the system understand the user's interests and will be mapped along with the results of the tests to predict the career paths.

- The users will be provided with a list of careers suitable for them after the tests are taken up by the students .
- The user can give their twitter Id to the fourth module of the project where negative and positive sentiment analysis is done to find the accuracy.

4.5 External Interface Requirements

It includes the following interfaces

- User Interfaces
- Hardware Interfaces

4.5.1 User Interfaces

The interface must be easy to understand. The user interface includes

- **Screen formats/organization**

The introductory screen will be the first to be displayed which will allow the users to do the member registration.

- **Window format/organization**

When the user chooses some other option, then the information pertaining to that choice will be displayed in a new window which ensures multiple windows to be visible on the screen and the users can switch between them.

- **End messages**

When there are some exceptions raising error like entering invalid details, then error messages will be displayed on the screen for the user.

4.5.2 Hardware Interfaces

Server side hardware

- Hardware recommended by all the software needed.
- Communication hardware to serve client requests

Client side hardware

- Hardware recommended by respective client's operating system and web browser.
- Database tools: Microsoft Excel
- Compatible operating system: Windows 10

Client side software

- Web browsers.

4.6 Hardware Requirements

System: Intel core I3 2.4 GHz and above.

Monitor: 15 VGA colour

Ram: 8 GB

Hard Disk: 1 TB

Mouse: normal

4.7 Software Requirements

Operating system: Windows 7/8/10.

Coding Language: JavaScript, Python

Software: PyQt

Data Base: MongoDB

Chapter 5

OBJECTIVES AND METHODOLOGY OF THE PROPOSED SYSTEM

5.1 Proposed system

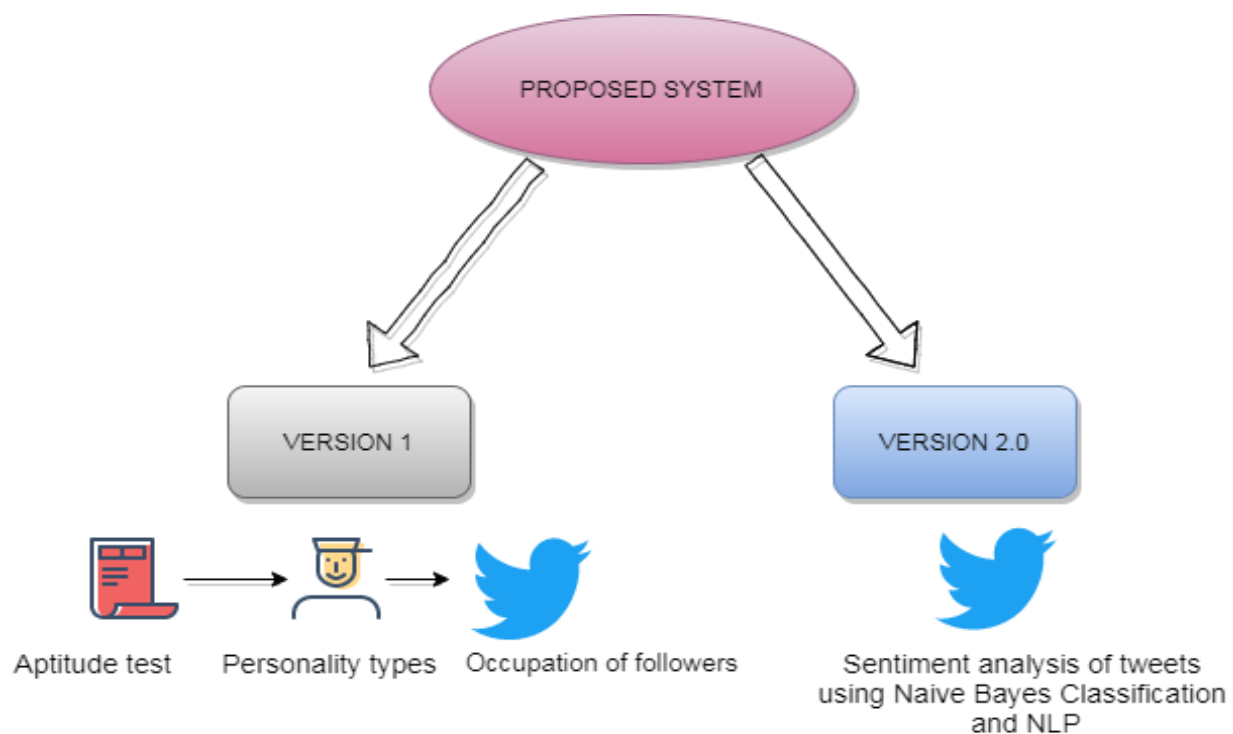


Fig 5.1 Proposed system

The main objective of the career guidance system is that it helps students to choose an appropriate career based on their personalities and intelligence capabilities. It predicts various career options best suitable for an individual. This system boosts the confidence in a student by providing him the career option where he can excel and succeed. The system allows users to take a series of tests consisting of various general questions. After test completion a score is calculated for each test. Based on those results the system manipulates and calculates the best career for that user. The proposed system consists of two versions. Version 1.0 includes aptitude

or technical tests followed by personality type and the twitter analysis .Twitter analysis in version 1 includes finding out the occupations of the people followed by the user.The careers from all the three factors will be mapped to give a list of appropriate careers .In version 2.0 only sentiment analysis of the tweets of the user will be used to predict the suitable career for the user. The system also monitors the behavior of the student for long term and suggests different path.It allows for a robust and enhanced experience by handling a large number of visitors with efficiency.

5.2 Objectives of the proposed system

The main purpose of the system is to efficiently predict the career path through a fully automated system that not only clears the confusions but also gives accurate career options. It is a cost-effective method as it has no involvement of career counsellors and is a student friendly approach. The administrator of the system controls the contents of the questionnaire provided to the students. The students can login through the client computers with their username and can take the test. Along with the results of the test, the behavioural aspects are mapped to give an accurate career options suiting one's personality whose interests are collected based on social media interaction.

5.3 Advantages of the proposed system

- The system predicts an appropriate career path.
- The system helps clear the confusions regarding the career aspects.
- The system helps the students figure out their personality and interests thus boosting their confidence.
- The system also predicts their careers through accessing the twitter ids , which helps system analyse the interests of the student
- By looking at the different aspects such as the aptitude , personality type and the interests of the student the accuracy of the career predicted have a greater chance of being the correct career choice for the student .

5.4 Methodology of the Proposed System

5.4.1 Phases of career prediction

Our project comes in two versions:

- Version 1.0: In this version prediction is done using personality test, aptitude tests. Most importantly twitter analysis in this version is done by obtaining the occupation of the user's follower's .Here we predict up to 4 careers. This version of the project is completed in four phases, the first phase is the aptitude test or technical test which consists of basic questions from certain technical fields, the second phase is the twitter analysis and the third phase is the mapping the personality type with the career .The fourth and final phase is obtaining the results of each phase into a list and mapping the list with a created dataset to finally predict the most appropriate careers.
- a) Aptitude test: Based on the technical knowledge we would ask few questions which covers certain technical fields. Based on the below mentioned 16 technical fields we would ask the user 6 questions on each technical field. By analysing the answers from the users we would select 2 strong technical fields of the user.This stage would output technical fields. The fields include Chef ,Theatre,Hotel management, Politics, Law, Photography, Software Development, Medicine, Banker, Sports, Aviation, Civil Engineering, History, Electrical, English Mathematics.
- b) Twitter Analysis: Twitter analysis is done by obtaining the user's twitter handle which the user has to enter in the system. Using Tweepy we get the User ID of twitter to get the ID's of whom the user is following Ex: User is following Dhoni, Kohli, Salman, Yash ,Amitabh.We get the occupation of each twitter account user is following and group by the max followed occupation. We get the two major followed occupation.
- c) Personality type: The User would be allowed to get the personality test online to identify his personality. There are 16 personalities out of which there will one output of the personality. Personality test can be taken from <https://www.16personalities.com/free-personality-test> .The below are the possible

personality types of the user [ISTJ, ISFJ, INFJ, INTJ, ISTP, ISFP, INFP, INTP, ESTP, ESFP, ENFP, ENTP, ESTJ, ESFJ, ENFJ, ENTJ]. Once the user has got to know his personality type he can enter the personality type in our system . The output from this phase is the personality type.

- d) Obtaining the final set of careers : At every phase there is an output . Certain careers are obtained at every phase . These careers are then contained in a list along with the personality type and are mapped with a dataset created for the system . The system then gives out the final most appropriate choices .
- Version 2.0: This version includes prediction of careers only by twitter analysis. As stated earlier in the report social media can be a great platform for understanding the individual interests and understanding their personal preferences in this Version prediction is done using Natural Language Processing (NLP) with tweets. Using Naïve Bias algorithm the tweets which are posted by the use are scanned. These are considered as data sets. Suppose a user has posted 100 tweets and 100 tweets are scanned to get the key words suppose he/she has posted about pilot. Then the user gets the occupation as pilot. This version can accurately predict the occupation of the user and is very helpful in career guidance.

5.4.2 Development Process

The development of the project takes place in different stages . First all the input is taken from the user end , this includes the test results, twitter handle and personality types .The data is collected in the form of lists . The collected data is then mapped with the datasets to predict the appropriate careers .

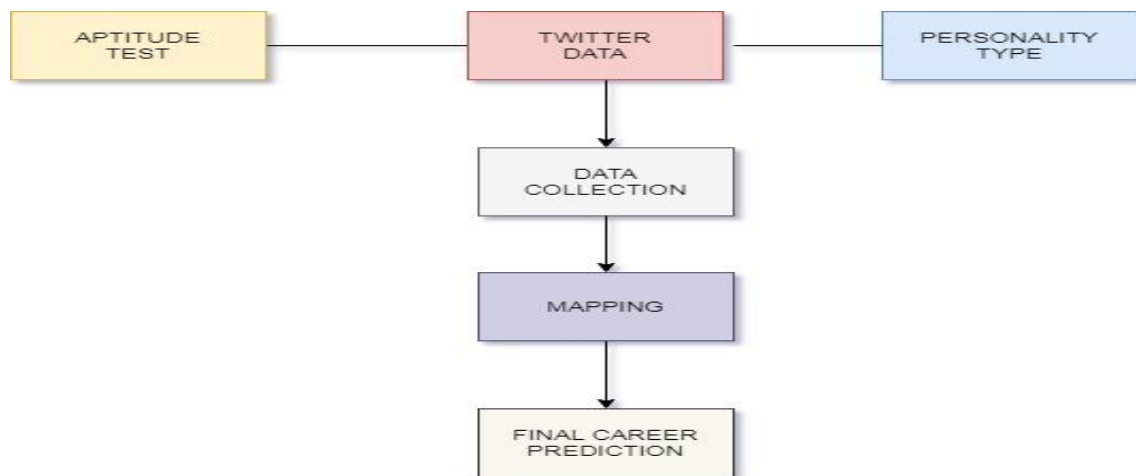


Fig 5.2 Development Process

5.4.3 Datasets used by the system

There are a number of datasets that are used by the system . Each of these datasets were created for the system usage.

- Dataset for the technical tests: This dataset consists of approximately 100 questions. The technical tests taken up by the users have different fields of careers . Each field consists of 6 questions. The dataset comprises the questions , followed by different options and the correct answer for each question. The questions designed are in the increasing level of complexity . The user will have to answer all the questions in order to go to the next step in their career analysis .

	A	B	C	D	E	F	G
1	Career	Questions	OptionA	OptionB	OptionC	OptionD	Correct Ans
2	Mathematics	Let A and B be two given events such that $P(A) = 0.6$, $P(B) = 0.2$ and $P(A/B) = 0.5$. Then $P(A'/B')$ is	1/10	3/10	3/8	6/7	C
3	Mathematics	Total number of possible matrices of order 2×3 with each entry 1 or 0 is:	6	32	36	64	D
4	Mathematics	The notation $f'(x)$ was used for derivative by ?	Leibniz	Einstein	Newton	Lagrange	D
5	Mathematics	An amount of Rs. 10,000 becomes Rs. 20,736 in 2 years. If the rate of interest is compounded half yearly, what is the annual rate of interest? (in percentage)	25%	20%	40%	30%	C
6	Mathematics	By what least number, $432/625$ be divided to get a perfect cube?	4/5	6/5	5/6	2/5	D
7	Mathematics	A fraction is such that if the numerator is multiplied by 3 and denominator is reduced by 3, we get $18/11$, but if the numerator is increased by 8 and the denominator is doubled, we get $2/5$. Then the fraction is	8/15	12/25	11/21	9/28	B
8	Mathematics	A hall, 20 m long and 15 m broad, is surrounded by a verandah of uniform width of 2.5 m. The cost of flooring the verandah at Rs. 3.50 per square metre is?	Rs. 500	Rs. 600	Rs. 700	Rs. 800	C
9	ENGLISH	Egregious (Synonym)	Joyful	Awful	Energetic	Friendly	B

Fig 5.3 Dataset for technical questionnaire

- Dataset used for mapping the final result : The output of the technical test gives us two strong areas or fields of the user . The next step which is the twitter analysis gives us the two most followed occupations . The third phase gives us the personality type of the user . All the results that are obtained from the user are then mapped to this dataset . This dataset comprises technical knowledge , followed by the various different occupations of twitter followers , personality type and finally the most suitable career for all the combinations. The dataset is created with thousands of entries in it , keeping in mind that no combinations of the attributes are missed .

	Technical Knowledge	Twitter Followers	Personality	Predicted Career
211	Medicine	Null	ENFJ	Photographer
212	Medicine	Null	ENFJ	Real estate broker
213	Medicine	police	ENFJ	Market research analyst
214	Medicine	hr	ENFJ	Market research analyst
215	Medicine	security	ENFJ	Market research analyst
216	Medicine	artist	ENFJ	Executive assistant
217	Medicine	actor	ENFJ	Market research analyst
218	Medicine	medical	ENFJ	Physical therapist
219	Medicine	writing	ENFJ	Real estate broker
220	Medicine	architect	ENFJ	Motivational speaker
221	Medicine	activist	ENFJ	Art director
222	Medicine	bank	ENFJ	Real estate broker
223	Medicine	lawyer	ENFJ	Sociologist
224	Medicine	enginner	ENFJ	Executive assistant
225	Medicine	politics	ENFJ	Motivational speaker
226	Medicine	sports	ENFJ	Real estate broker
227	Medicine	scientist	ENFJ	Real estate broker
228	Medicine	entrepreneur	ENFJ	Mediator
229	Medicine	teacher	ENFJ	Art director
230	Banker	Null	ENFJ	Real estate broker
231	Banker	Null	ENFJ	Editor
232	Banker	police	ENFJ	Real estate broker
233	Banker	hr	ENFJ	Sociologist

Fig 5.4 Dataset for mapping

- Datasets used in version 2.0: In this version we use two main datasets. One dataset contains the positively labelled tweets and the second contains negative labelled tweets. These tweets are further classified to predict the career. Every career that the system predicts needs to have two datasets. Collection of data is extremely exhaustive yet important in this scenario.

A1		@switchfoot http://twitpic.com/2y1zl - Awww, that's a bummer. You shoulda got David Carr of Third Day to do it. ;D	
A		B	
1	@switchfoot http://twitpic.com/2y1zl - Awww, that's a bummer. You shoulda got David Carr of Third Day to do it. ;D	-1	
2	is upset that he can't update his Facebook by texting it... and might cry as a result School today also. Blah!	-1	
3	@Kenichan I dived many times for the ball. Managed to save 50% The rest go out of bounds	-1	
4	my whole body feels itchy and like its on fire	-1	
5	@nationwideclass no, it's not behaving at all. i'm mad. why am i here? because i can't see you all over there.	-1	
6	@Kwesidei not the whole crew	-1	
7	Need a hug	-1	
8	@LOLTrish hey long time no see! Yes.. Rains a bit ,only a bit LOL, I'm fine thanks , how's you ?	-1	
9	@Tatiana_K nope they didn't have it	-1	
10	@twittera que me muera ?	-1	
11	spring break in plain city... it's snowing	-1	
12	I just re-pierced my ears	-1	
13	@caregiving I couldn't bear to watch it. And I thought the UA loss was embarrassing	-1	
14	@octolinz16 it counts, idk why I did either. you never talk to me anymore	-1	
15	@smarrison i would've been the first, but i didn't have a gun. not really though, zac snyder's just a doucheclown.	-1	
16	@iamjazzyfizzle I wish i got to watch it with you!! I miss you and @iamililnicki how was the premiere?!	-1	
17	Hollis' death scene will hurt me severely to watch on film wry is directors cut not out now?	-1	
18	about to file taxes	-1	
19	@LettyA ahh ive always wanted to see rent love the soundtrack!!	-1	
20	@FakerPattyPattz Oh dear. Were you drinking out of the forgotten table drinks?	-1	
21	@alydesigns i was out most of the day so didn't get much done	-1	
22	one of my friend called me, and asked to meet with her at Mid Valley today...but i've no time *sigh*	-1	
23	@angry_barista I baked you a cake but I ated it	-1	
24	this week is not going as i had hoped	-1	
25	blagh class at 8 tomorrow	-1	

Fig 5.5 Positively labelled dataset

A1		One of the Dominican Republic based Gulfstream IV came in from Narita this morning, cleared customs refueled and departed for low				
A		B	C	D	E	F
1	One of the Dominican Republic based Gulfstream IV came in from Narita this morning, cleared customs refueled and					
2	RT @TomPodolec: Nova Scotia suffered an unbelievable loss today. An RCMP officer killed along with 15 others today in a mass shooting. A loâ€					
3	@SteveMargenau @puttco Thanks					
4	Cathay Pacific 747-8F on short final Rwy 15 over the Knik Arm Inlet					
5	RT @breakingavnews: Saudia operates 40-hour flight to and from Los Angeles to repatriate 187 Saudi citizens. The trip included eight pilotsâ€					
6	RT @marklanebiz: Tornado GR4 performs low level loch pass in Scotland. Sound up! â€					
7	Elusive Air China Cargo 747-400F was on final Rwy 15 then winds changed, they were told to execute a missed					
8	RT @breakingavnews: An-225 carrying 150 tonnes of medical supplies landed earlier today at Paris-Vatry airport. It's scheduled to depart toâ€					
9	@AlexandriaboyVA @jetcitystar @flightradar24 @KalittaAir I just learnt last week that the owner also has a drag racing team and is former racing driver					
10	@puttco Yes, they are still coming through					
11	Kalitta Air 747-400F N706CK making condensation this morning as it departed for Incheon					
12	Perfect lighting for Polar Air 747-400F N450PA as it crosses 33 at Sierra and onto Yankee. It declared an emergency					
13	Western Global 747-400BCF rotating early for Narita					
14	I don't know what happened to this Everts Air Cargo DC-6..I no longer see it around					
15	RT @people: Winnie loves her mailman. 5Y" 5Y"4i: @winniethegoldentoller https://t.co/pGdXVXJIDw					
16	Kalitta Air 747-400F N716CK dancing on touchdown 7R inbound from ORD as another Kalitta heavy behind is on long					
17	@HawaiiAviation Seattle based Saltchuk owns Northern Aviation Services which runs both Aloha and Anchorage based Northern Air Cargo. It pops up here when cargo der					
18	Cathay Pacific fleet on the ramp this , some heading back to HKG and others lower 48 bound					
19	@DaveSnider Most probably					
20	@U4EIA No idea, there probably is an SOP on how they do it in formation. They were not behind us for more than a minute					
21	RT @breakingavnews: Powerful C-17 takeoff â€					
22	RT @breakingavnews: Chadian Su-25 mistakenly fires missile through a fuel tanker at Adji Kosseï airbase, killing at least five civilians. hâ€					
23	@rbyrnreilly It was an incentive flight for non flyers					
24	RT @jamjake01: @breakingavnews @Fiicaan It's now on its way back to Gander while another IL76 passed over me today in the UK inbound to Ganâ€					
25	Demo of chute dropping a load at the Malemute dropping zone on the Army side at Ft. Richardson where 4th Brigade					

Fig 5.6 Negative labelled dataset

5.5 Methodology used in the system

5.5.1 PyQt

- The user interface of the system is built using the PyQt toolkit.
- PyQt is a GUI widgets toolkit. It is a Python interface for Qt, one of the most powerful, and popular cross-platform GUI libraries. PyQt is a blend of Python programming language and the Qt library.
- It enables Python to be used as an alternative application development language to C++ on all supported platforms including iOS and Android. PyQt5 may also be embedded in C++ based applications to allow users of those applications to configure or enhance the functionality of those applications. The entire user interface in the project is built using the PyQt library .
- PyQt is free software developed by the British firm Riverbank Computing. It is available under similar terms to Qt versions older than 4.5; this means a variety of licenses including GNU General Public License (GPL) and commercial license, but not the GNU Lesser General Public License (LGPL). PyQt supports Microsoft Windows as well as various flavours of UNIX, including Linux and MacOS (or Darwin).
- PyQt5 can be installed by using ‘ pip install PyQt5 ‘
- PyQt implements around 440 classes and over 6,000 functions including a substantial set of GUI widgets, classes for accessing SQL databases, data aware widgets that are automatically populated from a database etc .
- Some of PyQt main components include :
 - The *QtCore* module contains the core non-GUI classes, including the event loop and Qt's signal and slot mechanism. It also includes platform independent abstractions for Unicode, threads, mapped files, shared memory, regular expressions, and user and application settings.
 - The *QtGui* module contains the majority of the GUI classes. These include a number of table, tree and list classes based on the model–view–controller design

pattern. Also provided is a sophisticated 2D canvas widget capable of storing thousands of items including ordinary widgets.

- The *QtNetwork* module contains classes for writing UDP and TCP clients and servers. It includes classes that implement FTP and HTTP clients and support DNS lookups. Network events are integrated with the event loop making it very easy to develop networked applications.
- The *QtDesigner* module contains classes that allow Qt Designer to be extended using PyQt.
- A few notable applications that use PyQt are :
 - Anki, a spaced repetition flashcard program
 - Calibre, an E-book management application
 - Dropbox, a file hosting service
 - Spyder, a Python data science IDE
 - TortoiseHg, a graphical interface for the Mercurial source management program (Hg)
 - Veusz, a scientific plotting application.

5.5.2 Tweepy

- To be able to carry out twitter analysis we need to be able to get access to the user's content on twitter . This is achieved through accessing the Twitter API.
- The content includes all the tweets , retweets , comments and all the people followed by the user .
- Tweepy is the twitter library for python .
- Twitter allows us to mine the data of any user using Twitter API or Tweepy. The data will be tweets extracted from the user. The first thing to do is get the consumer key, consumer secret, access key and access secret from twitter developer available easily for each user. These keys will help the API for authentication.
- Tweepy is one of the libraries that should be installed using pip. Now in order to authorize our app to access Twitter on our behalf, we need to use the OAuth Interface.

Tweepprovides the convenient Cursor interface to iterate through different types of objects. Twitter allows a maximum of 3200 tweets for extraction.

- There are different types of data we can collect, with the obvious focus on the “tweet” object. Once we have collected some data, the possibilities in terms of analytics applications are endless.

5.5.3 Importance of personality type

- Myers and Briggs proposed that there were four key dimensions that could be used to categorize people:
- Introversion vs. Extraversion
- Sensing vs. Intuition
- Thinking vs. Feeling
- Judging vs. Perceiving
- Based on these categories 16 personalities were proposed and each of them associated with a set of careers .

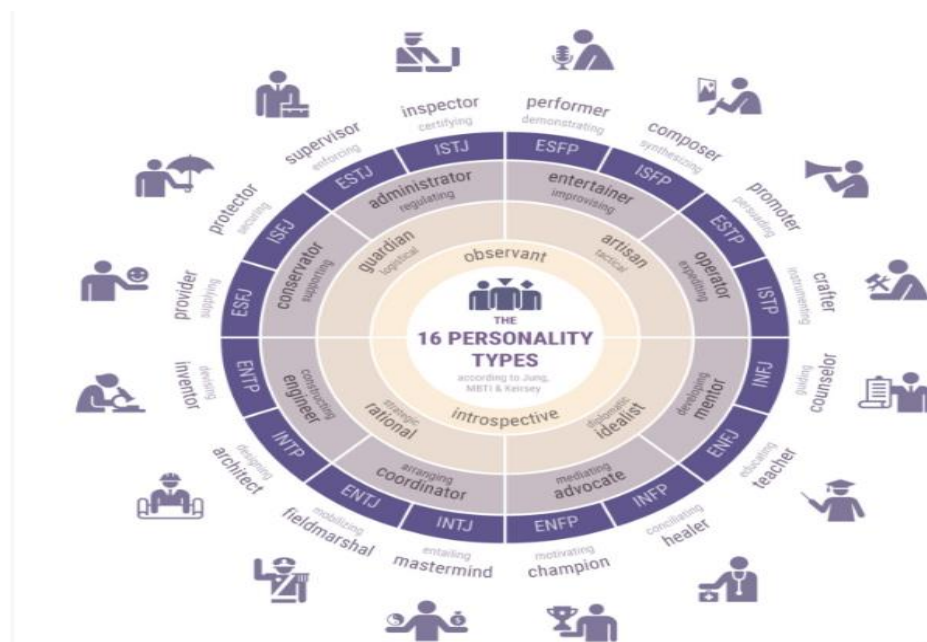


Fig 5.7 Personality types and associated careers

5.5.4 Naïve Bayes Classification

- In version 2.0 of our project we use naive bayes to analyze the tweets . We also use sentiment analysis to analyze the sentiments of our tweets that were extracted from the user's profile.
- We use two datasets in the project one which contains the positive tweets and the second which contains the negative tweets .
- The dataset contains the positively labelled and negatively labelled tweets.
- This sentimental data is classified in the project using the Naive Bayesian classifier.
- Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. There is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable.
- For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods.
- An advantage of naive Bayes is that it only requires a small number of training data to estimate the parameters necessary for classification
- Class Distribution: In Eq. 1, we have calculated the fraction of documents in each class π_c where the class c , word w at a word frequency f , $NN = \{nn_1, \dots, nn_n\}$ where N is the total number of words and n represents each word.

$$\pi_c = \frac{class_c}{\sum_{n=1}^N class_n}$$

- Probability of each word per class :For calculating our probability, we have found the average of each word for a given class. For class c and word w , the average is given in Eq. 2

$$P(w|c) = \frac{word_{wc}}{word_c}$$

- With low α value, where α alpha presents the smoothing value for the unseen words that don't appear in the training data. Where the V is an array of all the words in the vocabulary

$$P(w|c) = \frac{word_{wc} + \alpha}{word_c + |V| + 1}, \alpha = 0.001$$

- Multinomial Naive Bayes Classifier: Combining probability distribution of P with fraction of documents belonging to each class, given in Eq. 4, Eq. 5 and 6. One issue is that if a word appears once the probability of it appearing again increases, in order to smooth this we take the log frequency Eq.7.

$$\Pr(c) \propto \pi_c \prod_{w=1}^{|V|} \Pr(w|c)^{f_w}$$

$$\Pr(c) \propto \log(\pi_c \prod_{w=1}^{|V|} \Pr(w|c)^{f_w})$$

$$\Pr(c) = \log \pi_c + \prod_{w=1}^{|V|} f_w \log(\Pr(w|c))$$

- Even though the stop words have already been set to 0 for this specific use case, the IDF implementation is being added to generalize the function. As we can see, IDF has little effect as we removed the stop words. However, for the smoothing it makes the model more accurate. Hence, our optimal model is mathematically expressed in Eq. 10

$$\Pr(c) = \log \pi_c + \prod_{w=1}^{|V|} \log(1 + f_w) \log(\Pr(w|c)) \quad (10)$$

- Let us take Twitter's tweets and build a classifier based on the given tweets. This classifier will tell whether a tweet is under the category of "Politics or Sports" or not.
- The basic example of the tweet data will be classified based on the containing texts as the table below .

Tweet Id	Text	Category
294051752079159296	99 days to go until the start of #ct13. Did you know Chris Gayle is the only player in the event\u2019s history to be dismissed for 99?	Sports
291019672701255681	On Jan 10, PM #Abe received a courtesy call from Mr. Yoshihiro Murai, Governor of Miyagi Prefecture. \n http://t.co/EsyP40Gl	Politics
305581742104932352	Video of last week's hot topics: #2pack, #Draghi, pensions & #drug tests. @EuroparlTV video http://t.co/9GVBa315vM	Politics
291520568396759041	10 off the over, 10 required! Captain Faulkner to bowl the last over, in close discussion with veteran Warne. The final spot on the line #BBL02	Sports

Table 5.1 Tweets labelled with careers

- The table following table contains the training data

Tweet Id	Category	Text
306624404287275009	Sports	99 days to go until the start of #ct13. Did you know Chris Gayle is the only player in the event\u2019s history to be dismissed for 99?
306481199130505216	Sports	Tonight's Scottish First Division match between Dumbarton and Raith Rovers has been postponed due to a frozen pitch
304353716117590016	Politics	@GSANetwork raises awareness & stands up to stop #LGBT #bullying in school & online. http://t.co/FWIG5vVmi @glaad
304844614517547008	Politics	Blasts Deja Vu. How many times have we been in this *exact* moment? Failed or Ignored intel/no CCTV/blame game and innocents dead.

Table 5.2 Training data

- The output example of this algorithm is something like :

TweetId	Category
301733794770190336	Politics
301576909517619200	Politics
305057161682227200	Sports
286543227178328066	Politics

Table 5.3 Output of classification

5.5.5 NLP for cleaning the tweets

- Natural Language Processing is used for cleaning the data .
- The tweets that are obtained from the users and the tweets that are contained in the datasets , it will contain lots of characters that will not be meaningful to any machine learning algorithm. These characters include #, @ and punctuation marks.

- In order to remove these characters from the tweets and make classification easier we use the python re library this provides regular expression matching operations
- If there is an uppercase and lowercase versions of the same word found the algorithm will consider them as two different words .

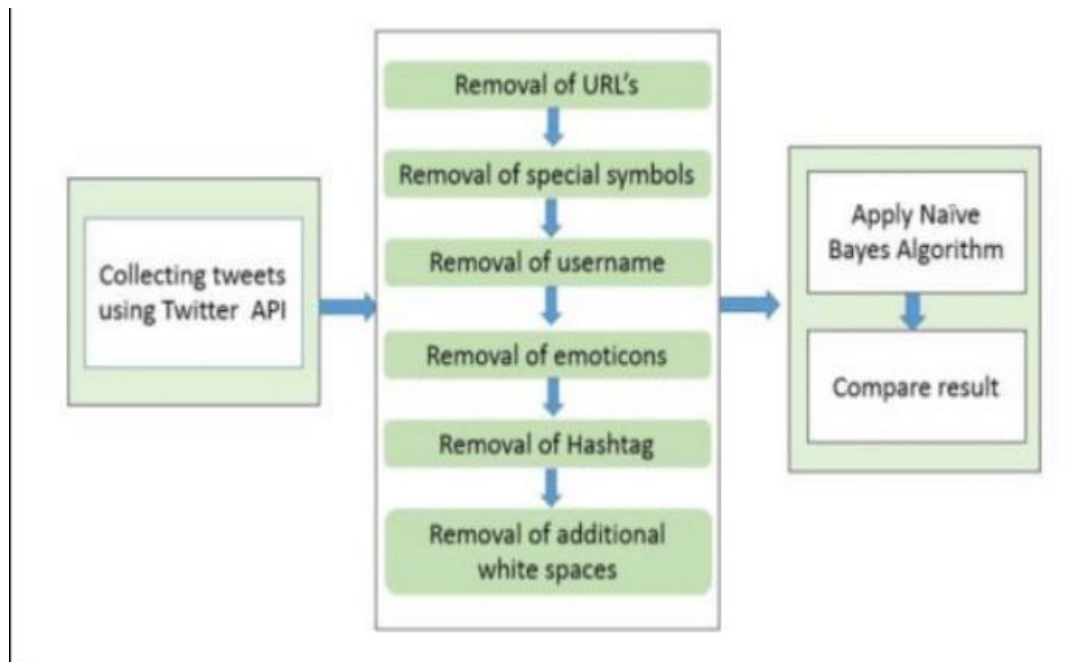


Fig 5.8 Classification of Tweets using Naive Bayes

Chapter 6

DETAILED DESIGN

6.1 System Overview

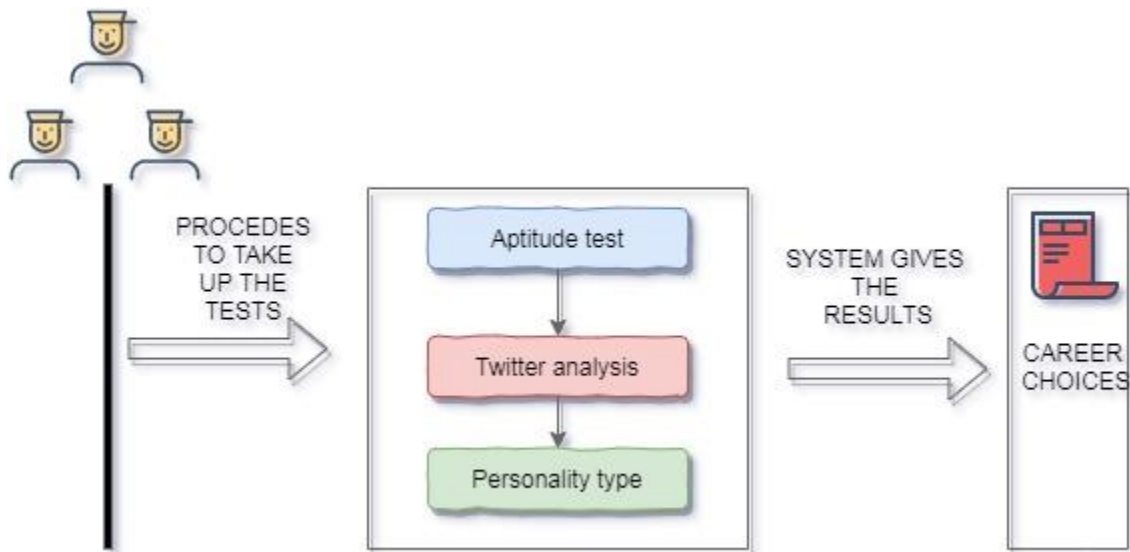


Fig 6.1 System Overview

The above figure represents the overview of the system. Every user logs into the system by providing the necessary information and then proceeds to take up the tests. The system comprises aptitude or technical test after the completion of this, the user is required to enter the twitter handle. Once that is done the user must enter the personality type. If he does not know the personality type then he has an option to take up the personality tests through the system portal itself. Once the user clears all the three steps. The system will output a list of careers that is appropriate and suitable for the user.

6.2 High level design

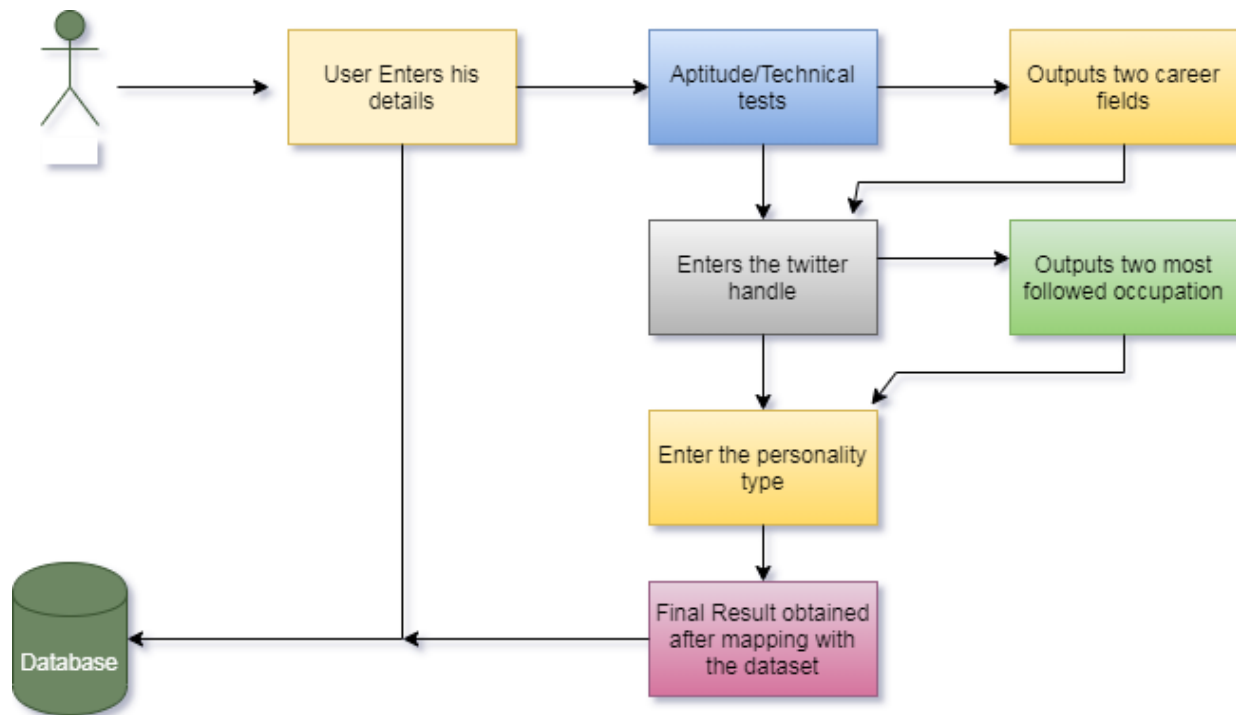


Fig 6.2 High level design

High-level design (HLD) explains the architecture that would be used for developing a software product. The architecture diagram provides an overview of an entire system, identifying the main components that would be developed for the product and their interfaces. In the high level design the main components are the Aptitude/Technical tests which the user has to take up. It outputs two career choices based on the scores that were awarded to the correct answers. The top two highest scored fields of careers will be the result. The next component is the twitter analysis, where the user enters his/her twitter handle through which the top two most followed occupations will be outputted. Final component is the personality type, where the user is required to enter the type of personality. The final result will be produced after mapping with the dataset. A maximum of four careers can be predicted in the final result. The user details along with their final career choices are stored in the database.

6.3 System Data Flow

- The data flow through the entire system is depicted in the figure below.
- The user will have to enter his personal details such as the name, email address, the entered details will be further stored in the database.
- The user is then navigated to the home page and then to the procedure briefing page.
- The next page which the user is navigated to is the technical questionnaire. This is Stage 1. Once the user finishes answering all the questions given to him. He will get the result of the two highest scored career fields.
- The next stage is the twitter analysis. This is stage 2. If the user has a twitter handle then the occupation of the followers will be obtained and the two most predominantly followed professions will be predicted as the output of this stage. If the user does not have the twitter id then he moves to the next step Stage 3.
- Stage 3 is entering the personality type. If the user is not aware of the type he can take up the personality test then enter the type. If he is already aware of the personality type then he can directly enter it.
- The results from all the three stages will be mapped and checked with the dataset to give the final career.
- The final result is then stored in the database.

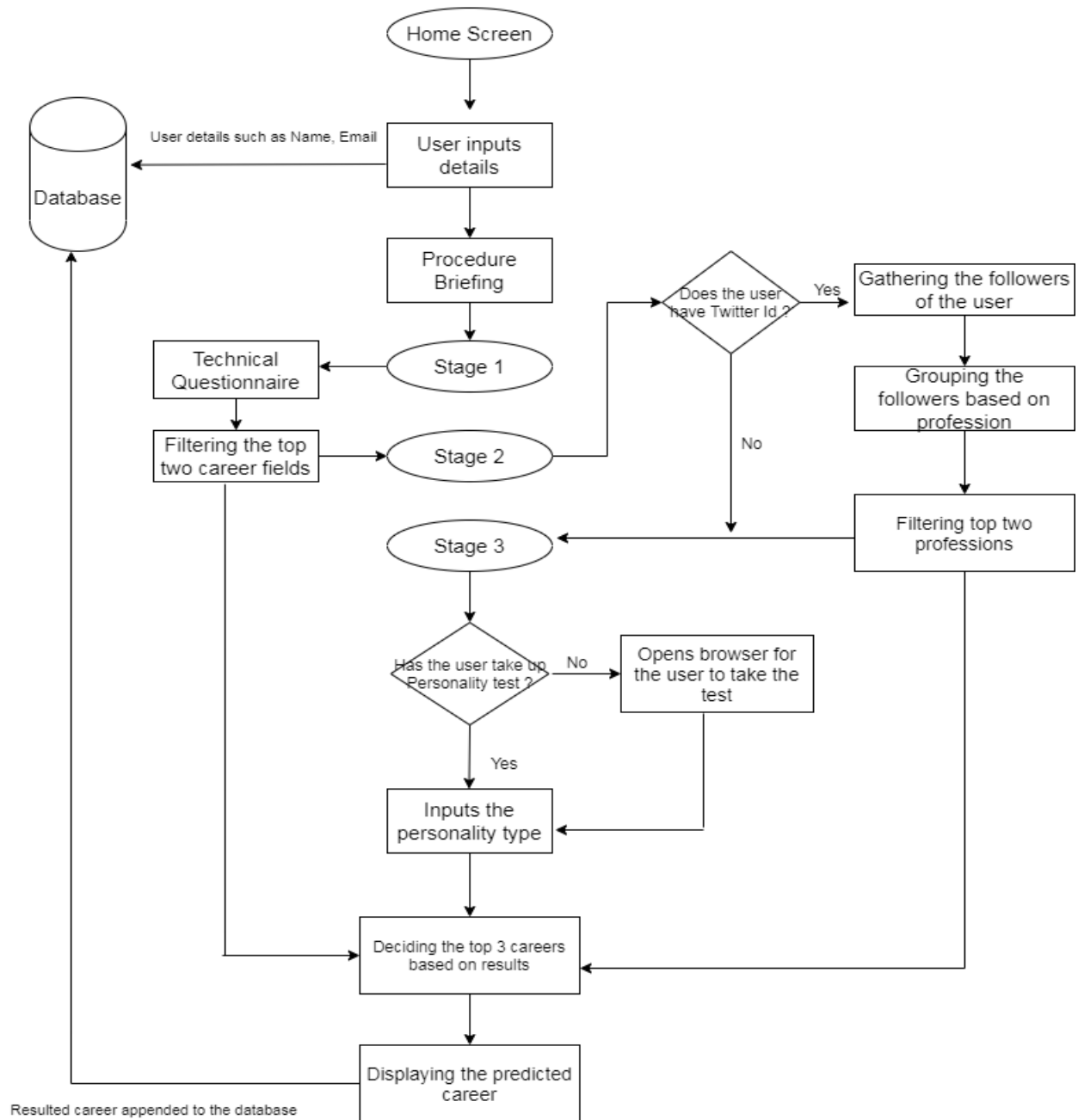


Fig 6.3 System data flow

6.4 Sequence diagram

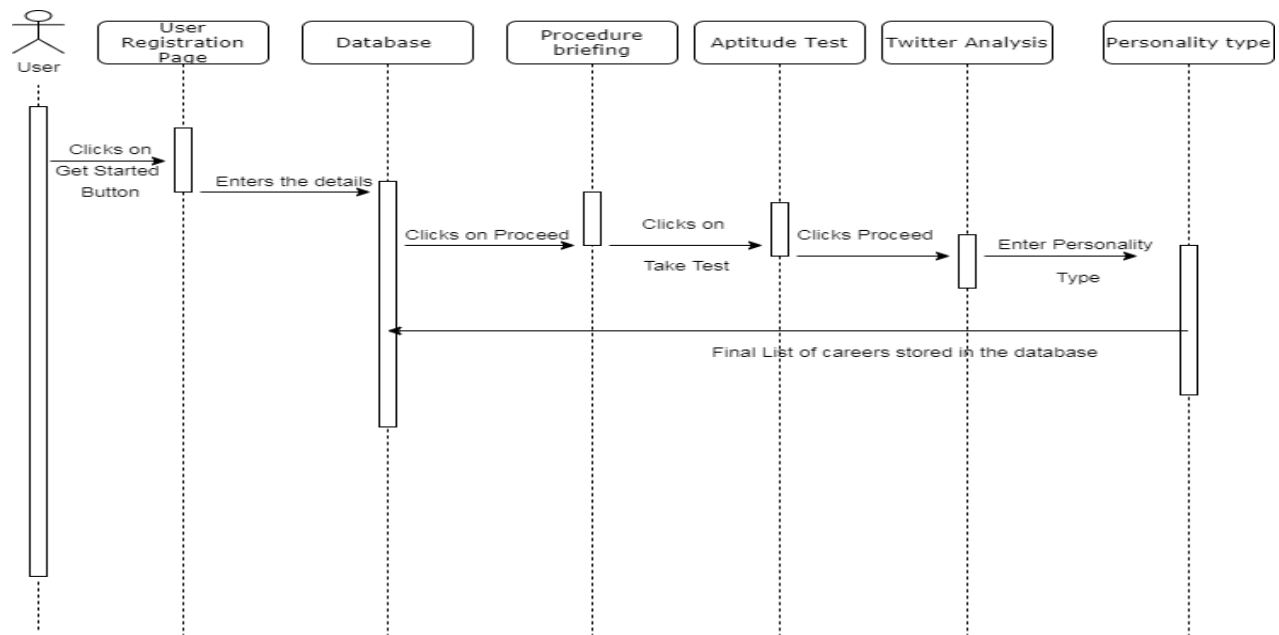


Fig 6.4 Sequence diagram

A sequence diagram simply depicts interaction between objects in a sequential order i.e. the order in which these interactions take place. We can also use the terms event diagrams or event scenarios to refer to a sequence diagram. Sequence diagrams describe how and in what order the objects in a system function. The above diagram represents the sequence of events that the user has to go through when he uses our portal for the career analysis. The objects of the diagram are user registration page, database, aptitude questionnaire, twitter analysis and personality type.

CHAPTER 7

IMPLEMENTATION

7.1 List of Modules identified along with their functions

- **Participants or Students** : The GUI helps to communicate information to the students
- **Admin** : The admin is responsible for the content that will be uploaded to the portal. This includes the questions for the different careers.
- **Curating the questions** : The admin is responsible for the addition of the aptitude questions pertaining to the different career fields.
- **Dynamic GUI** : An interactive and dynamic GUI has to be built using the python PyQt module that the user/student can take up the test on and view the results.
- **Gathering twitter data** : Tweets are collected to curate datasets that are used for the training purpose of the ML model using the python Tweepy module.
- **Getting occupation for the twitter user** : Creating a data which has the information about the occupation of the most popular twitter users the student may follow.
- **Career dataset based on Personality** : Creating a dataset that has the respective career suggestions for 16 types of personalities.
- **Final Mapping Dataset** : Creating a dataset with all the permutations of that are a result of the three different modules to give a final result.

7.2 Control flow between each of the units

- Step 1 : Launch the python application and register as a user to the portal.
- Step 2 : The user takes up an aptitude test for various different subjects and fields. A result is displayed for this step.
- Step 3 : The user inputs their Twitter ID from which we obtain the career the user is interested in by looking at the people they follow
- Step 4 : Using NLP and Multinomial Bayesian Classification to analyze the tweets and predict the interested career field.
- Step 5 : The user takes up the Myers Briggs 16 personality test and the corresponding career is mapped to the user.

- Step 6 : The final result is displayed on the GUI.

7.3 Aptitude Questionnaire

7.3.1 The launch GUI

The python application is developed using the PyQt module. The following code snippet shows the launching page.

```
from PyQt5.QtGui import *
from PyQt5.QtWidgets import *
from PyQt5.QtCore import *
from PyQt5 import QtCore, QtGui, QtWidgets
from PyQt5.QtGui import QMovie
from Instructions import Ui_Instructions
from RegisterUI import Ui_Register
from QuestionsUI import Ui_Questions
class Ui_FirstPage(object):
    def setupUi(self, FirstPage):
        self.run = QTimer()
        self.run.setInterval(200)
        self.run.timeout.connect(self.checkstatus)
        self.run.start()
        FirstPage.setObjectName("FirstPage")
        FirstPage.setEnabled(True)
        FirstPage.resize(583, 640)
        FirstPage.setFixedSize(QtCore.QSize(583,640))
        icon = QtGui.QIcon()
```

```
icon.addPixmap(QtGui.QPixmap("Images/icon.jpg"),
QtGui.QIcon.Normal, QtGui.QIcon.Off)

FirstPage.setWindowIcon(icon)

self.label = QtWidgets.QLabel(FirstPage)

self.label.setGeometry(QtCore.QRect(0, 0, 591, 641))

self.label.setText("")

self.mngif=QMovie("Images/Career Guidance3.gif")

self.label.setMovie(self.mngif)

self.mngif.start()

self.label.setScaledContents(True)

self.label.setObjectName("label")

self.pushButton = QtWidgets.QPushButton(FirstPage)

self.pushButton.setGeometry(QtCore.QRect(230, 540, 121, 41))

self.pushButton.setStyleSheet("*{\n"
"color: rgb(0, 0, 127);\n"
"font: 13pt \"century gothic\";\n"
"border-radius: 15px;\n"
"background-color: rgba(251,155,251,0.2);\n"
"}\n"
"*:hover{\n"
"color: rgb(0, 0, 0);\n"
"}")
```

7.3.2 The User Registration Page

This code snippet shows us the registration UI where the student has to enter their first name, last name and email address to register to the portal.

```

class Ui_Register(object):
    Registered=0
    email_reg='x'
    name_reg='x'
    def setupUi(self, Register):
        Register.setObjectName("Register")
        Register.resize(590, 640)
        self.groupBox = QtWidgets.QGroupBox(Register)
        self.groupBox.setGeometry(QtCore.QRect(10, 250, 571, 361))
        icon = QtGui.QIcon()
        icon.addPixmap(QtGui.QPixmap("Images/icon.jpg"), QtGui.QIcon.Normal,
        QtGui.QIcon.Off)
        Register.setWindowIcon(icon)
        font = QtGui.QFont()
        font.setFamily("MV Boli")
        font.setPointSize(10)
        font.setBold(False)
        font.setItalic(False)
        font.setWeight(50)
        self.groupBox.setFont(font)
        self.groupBox.setStyleSheet("color:rgb(0, 170, 0);\n"
        "border-color: rgb(255, 170, 0);\n"
        "font: 10pt \"MV Boli\";")
        def retranslateUi(self, Register):
            _translate = QtCore.QCoreApplication.translate

```

```
Register.setWindowTitle(_translate("Register", "Register"))

self.groupBox.setTitle(_translate("Register", "Enter Details"))

self.label.setText(_translate("Register", "Name:"))
align="center"><span style="color:#00007f;">First
Name:</span></p></body></html>"))

self.label_2.setText(_translate("Register", "Second
align="center"><span style="color:#00007f;">Second
Name:</span></p></body></html>"))

self.label_3.setText(_translate("Register", "Email
align="center"><span style="color:#00007f;">Email
ID:</span></p></body></html>"))

self.label_4.setText(_translate("Register", "Email
align="center"><span style="color:#00007f;">Email
ID:</span></p></body></html>"))

self.pushButton.setText(_translate("Register", "Proceed"))
```

7.3.3 The UI for Aptitude

The below code snippet shows the code for the Test Questions which is in the form of MCQs. The user had to choose the option for an answer in the checkbox provided. The test questions are divided into various subjects pertaining to each of the careers.

```
class Ui_Questions(object):

    travel_df = pandas.read_excel('Book1.xlsx')

    content = travel_df.to_dict()

    questcareerlist=list(content['Career'].values())

    questionlist=list(content['Questions'].values())

    optionAlist=list(content['OptionA'].values())

    optionBlist=list(content['OptionB'].values())

    optionClist=list(content['OptionC'].values())

    optionDlist=list(content['OptionD'].values())
```

```
current_ans_list=list(content['Correct Ans'].values())

avbl_careers=[]

for i in questcareerlist:

    if i not in avbl_careers:

        avbl_careers.append(i)

        scores=[0]*len(avbl_careers)

        print(scores)

        print(avbl_careers)

res={}

def setupUi(self, Questions,User_details):

    self.User_details=User_details

    self.Questions=Questions

    self.qindex=0

    self.final_result=['Null','Null']

    self.len=len(Ui_Questions.avbl_careers)

    Ui_Questions.res={ Ui_Questions.avbl_careers[i]:Ui_Questions.scores[i]

    for i in range(len(Ui_Questions.avbl_careers))}

    print(Ui_Questions.res)

    self.Questions.setObjectName("Questions")

    self.Questions.resize(590, 640)

    self.Questions.setFixedSize(QtCore.QSize(590,640))

    icon = QtGui.QIcon()

    icon.addPixmap(QtGui.QPixmap("Images/icon.jpg"),QtGui.QIcon.Normal,QtGui.QIcon.Off)

    self.Questions.setWindowIcon(icon)
```



```

self.label_2 = QtWidgets.QLabel(self.Questions)
self.label_2.setGeometry(QtCore.QRect(0, -1, 590, 641))
#self.label_2.setStyleSheet("selection-background-color: rgba(255, 255, 255, 50);")
self.label_2.setText("")
self.label_2.setScaledContents(True)
self.label_2.setObjectName("label_2")
self.gif4=QMovie("Images/bckg_q.png")
self.label_2.setMovie(self.gif4)
self.gif4.start()
self.label = QtWidgets.QLabel(self.Questions)
self.label.setGeometry(QtCore.QRect(20, 20, 551, 61))
if optionA+optionB+optionC+optionD==2:
    if optionA==2:
        print("A selected")
        selected='A'
    elif optionB==2:
        print("B selected")
        selected='B'
    elif optionC==2:
        print("C selected")
        selected='C'
    elif optionD==2:
        print("D selected")
        selected='D'
    else:

```

```
QMessageBox.information(None,'Warning!!','No Option selected , Select Any  
One Option')
```

7.3.4 The result obtained from the test

The result obtained from the aptitude test is stored in a list. We obtain this result from the number of questions that are correctly answered for each of the different subjects. A minimum of three questions have to be right for each of the subjects for that particular field to be considered for the final result. The resultant array from this stage of the project will have two career choices, the two with the most correctly answered question. In case there is a condition that arises where two career options have an equal number of questions correctly answered, the career fields are sorted in a way that is based on priority, more priority given to technical and literature fields based on popularity of the careers.

```
if res_value[sort_index[0]]>3:  
    self.final_result[0]=Ui_Questions.avbl_careers[sort_index[0]]  
if res_value[sort_index[1]]>3:  
    self.final_result[1]=Ui_Questions.avbl_careers[sort_index[1]]  
{'Mathematics': 0, 'English ': 0, 'Software Development': 0}  
C selected  
{'Mathematics': 1, 'English ': 0, 'Software Development': 0}  
[1, 0, 0]  
[0, 0, 1]  
[2, 1, 0]  
[Null, 'Null']  
No. of questions left:: 17  
D selected
```

{'Mathematics': 2, 'English ': 0, 'Software Development': 0}

[2, 0, 0]

[0, 0, 2]

[2, 1, 0]

['Null', 'Null']

No. of questions left:: 16

D selected

{'Mathematics': 3, 'English ': 0, 'Software Development': 0}

[3, 0, 0]

[0, 0, 3]

[2, 1, 0]

['Null', 'Null']

No. of questions left:: 15

D selected

{'Mathematics': 4, 'English ': 0, 'Software Development': 0}

[4, 0, 0]

[0, 0, 4]

[2, 1, 0]

['Software Development', 'Null']

No. of questions left:: 14

B selected

{'Mathematics': 5, 'English ': 0, 'Software Development': 0}

[5, 0, 0]

[0, 0, 5]

[2, 1, 0]

['Software Development', 'Null']

No. of questions left:: 13

C selected

{'Mathematics': 6, 'English ': 0, 'Software Development': 0}

[6, 0, 0]

[0, 0, 6]

[2, 1, 0]

['Software Development', 'Null']

No. of questions left:: 12

A selected

{'Mathematics': 6, 'English ': 0, 'Software Development': 0}

[6, 0, 0]

[0, 0, 6]

[2, 1, 0]

['Software Development', 'Null']

No. of questions left:: 11

A selected

{'Mathematics': 6, 'English ': 0, 'Software Development': 0}

[6, 0, 0]

[0, 0, 6]

[2, 1, 0]

['Software Development', 'Null']

No. of questions left:: 10

A selected

{'Mathematics': 6, 'English ': 0, 'Software Development': 0}

[6, 0, 0]

[0, 0, 6]

[2, 1, 0]

['Software Development', 'Null']

No. of questions left:: 9

A selected

{'Mathematics': 6, 'English ': 1, 'Software Development': 0}

[6, 1, 0]

[0, 1, 6]

[2, 1, 0]

['Software Development', 'Null']

No. of questions left:: 8

A selected

{'Mathematics': 6, 'English ': 1, 'Software Development': 0}

[6, 1, 0]

[0, 1, 6]

[2, 1, 0]

['Software Development', 'Null']

No. of questions left:: 7

A selected

{'Mathematics': 6, 'English ': 1, 'Software Development': 0}

[6, 1, 0]

[0, 1, 6]

[2, 1, 0]

['Software Development', 'Null']

No. of questions left:: 6

C selected

{'Mathematics': 6, 'English ': 1, 'Software Development': 1}

[6, 1, 1]

[1, 1, 6]

[2, 1, 0]

['Software Development', 'Null']

No. of questions left:: 5

B selected

{'Mathematics': 6, 'English ': 1, 'Software Development': 2}

[6, 1, 2]

[2, 1, 6]

[2, 0, 1]

['Software Development', 'Null']

No. of questions left:: 4

B selected

{'Mathematics': 6, 'English ': 1, 'Software Development': 3}

[6, 1, 3]

[3, 1, 6]

[2, 0, 1]

['Software Development', 'Null']

No. of questions left:: 3

A selected

{'Mathematics': 6, 'English ': 1, 'Software Development': 4}

[6, 1, 4]

[4, 1, 6]

[2, 0, 1]

['Software Development', 'Mathematics']

No. of questions left:: 2

B selected

{'Mathematics': 6, 'English ': 1, 'Software Development': 5}

[6, 1, 5]

[5, 1, 6]

[2, 0, 1]

['Software Development', 'Mathematics']

No. of questions left:: 1

A selected

{'Mathematics': 6, 'English ': 1, 'Software Development': 6}

[6, 1, 6]

[6, 1, 6]

['Software Development', 'Mathematics']

7.4 Twitter Analysis

7.4.1 UI for the User to enter their Twitter Handle

The below code snippet shows the code for building a UI that will let the user enter their twitter handle so that we can analyse their tweets and follow list to deduce the people they interact with on social media to find out the field they might be interested in.

```
class Step2_Action_Window(object):  
    def setUpUi(self,Form,final_result_1,User_details):  
        self.User_details=User_details  
        self.final_result_1=final_result_1  
        self.Form=Form  
        self.Form.setObjectName("Form")  
        self.Form.resize(590, 640)  
        self.button="Choose"  
        self.label = QtWidgets.QLabel(self.Form)  
        self.label.setGeometry(QtCore.QRect(0, 0, 590, 640))  
        self.label.setText("")  
        self.label.setPixmap(QtGui.QPixmap("Images/back.jpg"))  
        self.label.setScaledContents(True)  
        self.label.setObjectName("label")  
        self.label_2 = QtWidgets.QLabel(self.Form)  
        self.label_2.setGeometry(QtCore.QRect(220, 50, 151, 131))  
        self.label_2.setText("")  
        self.label_2.setPixmap(QtGui.QPixmap("Images/Logo_t.png"))  
        self.label_2.setScaledContents(True)  
        self.label_2.setObjectName("label_2")  
        self.label_3 = QtWidgets.QLabel(self.Form)  
        self.label_3.setGeometry(QtCore.QRect(170, 440, 281, 41))
```



```
self.label_3.setStyleSheet("font: 10pt \"MV Boli\";")
self.label_3.setAlignment(QtCore.Qt.AlignCenter)
self.label_3.setObjectName("label_3")
self.comboBox = QtWidgets.QComboBox(self.Form)
self.comboBox.setGeometry(QtCore.QRect(260, 490, 73, 31))
self.comboBox.setStyleSheet("background-color: rgb(0, 85, 255);\n"
"color:rgb(235, 235, 235);\n"
"font: 8pt \"MV Boli\";")
```

7.4.2 Getting the Twitter IDs of the people the user follows

The following code snippet shows how we can use the Tweepy python module to get a list of IDs the student follows on Twitter. This is done using the Twitter Developer options support provided by Twitter. The consumer key, consumer secret, access token and access token secret are used to authenticate the developer using OAuth Authentication. Once a minimum number of usernames have been extracted using API calls, we then compare them with our dataset to deduce the occupation of all these IDs. Then two of the frequently followed occupations are given as the result of this step.

```
from tweepy import OAuthHandler
from tweepy import API
from tweepy import Cursor
from datetime import datetime, date, time, timedelta
from collections import Counter
import sys
import time
import tweepy
import csv
```

```

import data

consumer_key = '3twKJ34s9NWj2kGV8WCyi5gCe'

consumer_secret = 't3ZJbDfvOYEwi2ioX7O4D8S2Z8xNBGuKqIgLIQrm1gcki5RZon'

access_token = '1222580048634343424-K9GQ0zVkfWQ31RnDHZWafUPgBl3Row'

access_token_secret = '71XSMKZzUNzC6x9NNc0xJmGpYdQ7lprLnnDznXUttGnW1'

def Occupation_count(id_name):

    auth = tweepy.OAuthHandler(consumer_key, consumer_secret)

    auth.set_access_token(access_token, access_token_secret)

    api = tweepy.API(auth, wait_on_rate_limit=True, wait_on_rate_limit_notify=True,
    retry_count=3, retry_delay=60)

    ids = []

    page_count = 0

    for page in tweepy.Cursor(api.friends_ids, id=id_name, count=50).pages():

        page_count += 1

        print ('Getting page { } for friends ids'.format(page_count))

        ids.extend(page)

        if page_count==2:

            break

    print(ids)

    ids=str(ids)

    ids=ids[1:-1]

    print(ids)

Usernames ['kaajalvashisht', 'joydatta', 'AnupamPKher', 'realpreityzinta', 'TheShilpaShetty',
'iPoonampandey', 'Akon', 'Maheshbhupathi', 'ajaydevgn', 'Priyu_B', 'NadeeshCherian',
'irrfank', 'MirzaSania', 'MTVRoadies', 'japna_p', 'virendersehwag', 'bhogleharsha', 'IPL',
'SriSri', 'ShaneWarne', 'rpsingh', 'sreesanth36', 'SPFleming7', 'rohet1', 'akshaykumar',

```

```
'LalitKModi', 'ImRo45', 'ImRaina', 'BeingSalmanKhan', 'GautamGambhir', 'parthiv9',  
'KP24', 'scottbstyris', 'priyankachopra', 'aamir_khan', 'flintoff11', 'kartikmurali',  
'juniorbachchan', 'henrygayle', 'msdhoni', 'sidmallya', 'harbhajan_singh', 'YUVSTRONG12',  
'ImZaheer', 'TheVijayMallya', 'markb46', 'sachin_rt', 'ABdeVilliers17', 'geneliad',  
'allaboutcricket']
```

followsa actor

followsa actor

follows a sports

follows a sports

followsa actor

follows a sports

['sports', 'actor']

7.4.3 UI for getting career from twitter using Machine Learning

The below code snippet shows the code to generate a UI where the user can either input their Twitter ID or input a sentence for analysis.

```
class Ui_MainWindow(object):  
  
    def setupUi(self, MainWindow):  
  
        MainWindow.setObjectName("MainWindow")  
  
        MainWindow.setFixedSize(800,600)                                ## FIXED  
        to a particular size  
  
        icon = QtGui.QIcon()  
  
        icon.addPixmap(QtGui.QPixmap("pyqt5/resources/twitter_black_logo.png"),  
                        QtGui.QIcon.Normal, QtGui.QIcon.Off)  ## ICON setter using image  
  
        MainWindow.setWindowIcon(icon)  
  
        self.centralwidget = QtWidgets.QWidget(MainWindow)
```

```
self.centralwidget.setObjectName("centralwidget")
self.bgimage = QtWidgets.QLabel(self.centralwidget)
self.bgimage.setGeometry(QtCore.QRect(-10, 0, 821, 561))
self.bgimage.setFrameShape(QtWidgets.QFrame.Box)
self.bgimage.setText("")
self.bgimage.setPixmap(QtGui.QPixmap("pyqt5/resources/simplebg.jpg"))
self.bgimage.setScaledContents(True)
self.bgimage.setObjectName("bgimage")
self.toptitle = QtWidgets.QLabel(self.centralwidget)
self.toptitle.setGeometry(QtCore.QRect(100, 110, 600, 91))
font = QtGui.QFont()
font.setFamily("Ebrima")
font.setPointSize(20)
font.setBold(True)
font.setWeight(75)
self.toptitle.setFont(font)
self.toptitle.setLayoutDirection(QtCore.Qt.LeftToRight)
self.toptitle.setTextFormat(QtCore.Qt.AutoText)
self.toptitle.setScaledContents(False)
self.toptitle.setAlignment(QtCore.Qt.AlignCenter)
self.toptitle.setObjectName("toptitle")
self.twitterlogo = QtWidgets.QLabel(self.centralwidget)
self.twitterlogo.setGeometry(QtCore.QRect(310, -10, 180, 180))
self.twitterlogo.setText("")
self.twitterlogo.setPixmap(QtGui.QPixmap("pyqt5/resources/twitter_blue.png"))
```

```
self.twitterlogo.setScaledContents(True)

self.twitterlogo.setObjectName("twitterlogo")

self.twitterID = QtWidgets.QLabel(self.centralwidget)

self.twitterID.setGeometry(QtCore.QRect(150, 200, 170, 30))

font = QtGui.QFont()

font.setFamily("Dubai")

font.setPointSize(18)

font.setBold(True)

font.setWeight(75)

self.twitterID.setFont(font)

self.twitterID.setObjectName("twitterID")
```

7.4.4 Analyzing the tweets from the user

The below code snippet shows the implementation of Multinomial Bayesian Network used to analyze the tweets from a particular Twitter handle. The Tweepy library is used once again to collect tweets to form datasets pertaining to each career by gathering Tweets from famous personalities which are in the same career field. This dataset is then cleaned using NLP and then fed to the algorithm as train and test datasets. The algorithm then generates a .pkl model file. This file is then used to predict the career the user is most interested in. The tweets from the user are then fed to the algorithm and then the prediction is made. Apart from this, there is provision also to check the career inclination of single tweets in a similar fashion.

```
def clean_tweet(tweet):

    return ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t])|(\w+:\w+\S+)", " ", tweet).split())

def strip_all_entities(text):

    entity_prefixes = ['@', '#']

    words = []
```

```

        for word in text.split():
            word = word.strip()
            if word:
                if word[0] not in entity_prefixes:
                    words.append(word)
        return ' '.join(words)

    with open('neg.csv', 'a', encoding='utf-8', newline = '') as f:
        writer = csv.writer(f)
        for i in out_tweets:
            writer.writerow([i])
        tweets.append(i)
        personality.append(val)
    print("Done " + screen_name)
    return out_tweets

    pilotcsv = open('pilotcsv.csv', 'r')
    reader = csv.reader(pilotcsv)
    for row in reader:
        get_all_tweets(str(row[0]), val=1)
    pilotcsv = open('notpilot.csv', 'r')
    reader = csv.reader(pilotcsv)
    for row in reader:
        get_all_tweets(str(row[0]), val=-1)
    x = 0

    with open('dataset140.csv', 'r', encoding='utf-8') as f:
        reader = csv.reader(f)

```

```

for i in reader:
    tweets.append(i[0])
    personality.append(-1)
    if x == 10000:
        break
    x+=1
print(len(tweets))
nb = MultinomialNB()
vectorizer = CountVectorizer(stop_words='english')
#print(tweets)
train_features = vectorizer.fit_transform(tweets)
print(len(vectorizer.get_feature_names()))
print('Training starting.....')
nb.fit(train_features, [int(r) for r in personality])
print('Trained.....')
test_features = vectorizer.transform(tweets[:500])
joblib.dump(nb, 'personality5.pkl')
test = ['@jacobsommers Sorry tell them mea culpa from me and that I really am sorry.']
test = vectorizer.transform(test)
print(nb.predict(test))
test = ["@F1Smasher Yes. As we will have been away from flying for a while we will have
to go in the simulator for a refresher."]
test = vectorizer.transform(test)
print(nb.predict(test))
def getCareerbyTweet(tweet):
    detectModel = joblib.load('personality5.pkl')

```

```

vectorizer = CountVectorizer(stop_words='english')

vocab = analyser.vocabLoader()

vectorizer.fit_transform(vocab)

text = vectorizer.transform([tweet])

#print("text : ", text)      #CLEANING FUNCTION

if csr_matrix.getnnz(text) == 0:

    return 'Invalid Statement'

depVal = detectModel.predict(text)

#print(depVal)

if depVal[0] == -1:

    return 'Negative Sentiment'

else:

    return 'Pilot personality based Sentiment'

# FUNCTION TO GET THE Career VALUE BASED ON NUMEROUS TWEETS FROM
A TWITTER HANDLE

def getCareerbyId(ID):

    detectModel = joblib.load('personality5.pkl')

    vectorizer = CountVectorizer(stop_words='english')

    vocab = analyser.vocabLoader()

    vectorizer.fit_transform(vocab)

    dPos ,dNeg = 0, 0

    tweets = getTweets.get_all_tweets(ID)

    if tweets == None:

        return 'Invalid'

        x = 0

    for tweet in tweets:

```



```
text = vectorizer.transform([str(tweet)])  
depVal = detectModel.predict(text)  
if depVal == [-1]:  
    dNeg += 1  
else:  
    dPos += 1  
depVal = detectModel.predict(text)
```

7.5 Myers-Briggs Personality Test

7.5.1 UI for the personality input

The code snippet given below gives us to code to build a GUI that lets the users input their personality type out of the 16 personalities that are given under the Myers Briggs personality test.

```
class Step3_Action_Window(object):  
    def setupUi(self, Form, final_result_1, final_result_2, User_details):  
        self.User_details = User_details  
        self.final_result_1 = final_result_1  
        self.final_result_2 = final_result_2  
        self.Form = Form  
        self.Form.setObjectName("Form")  
        self.Form.resize(590, 640)  
        self.button = "Choose"  
        self.label = QtWidgets.QLabel(self.Form)  
        self.label.setGeometry(QtCore.QRect(0, 0, 590, 640))  
        self.label.setText("")  
        self.label.setPixmap(QtGui.QPixmap("Images/P_bck.jpg"))
```

```

self.label.setScaledContents(True)

self.label.setObjectName("label")

self.label_2 = QtWidgets.QLabel(self.Form)

self.label_2.setGeometry(QtCore.QRect(0, 0, 600, 200))

self.label_2.setText("")

self.label_2.setPixmap(QtGui.QPixmap("Images/Personality.png"))

self.label_2.setScaledContents(True)

self.label_2.setObjectName("label_2")

self.label_3 = QtWidgets.QLabel(self.Form)

self.label_3.setGeometry(QtCore.QRect(10, 220, 580, 41))

self.label_3.setStyleSheet("font: 10pt \"MV Boli\";")

self.label_3.setAlignment(QtCore.Qt.AlignCenter)

self.label_3.setObjectName("label_3")

self.comboBox = QtWidgets.QComboBox(self.Form)

self.comboBox.setGeometry(QtCore.QRect(260, 270, 73, 31))

self.comboBox.setStyleSheet("QComboBox::hover\"\"{ \"\"background-color:
rgb(126,126,126);\n\"\"} ")

def check(self):

    ans=self.comboBox.currentText()

    if self.button=="Choose"andans=="Yes':

        self.label_3.setText("Please choose your Personality")

        self.comboBox.hide()

        self.comboBox2.show()

        self.button="Next"

        self.pushButton.setText("Get Career!")

```

```
elif self.button=="Choose"andans=='No':  
    print("Please take up the test!!")  
    self.comboBox.hide()  
    self.label_3.setText("Please take up your personlity test and then click on done")  
    self.button="Test"  
    self.pushButton.setText("Done")  
    self.driver=webdriver.Chrome(executable_path="./chromedriver.exe")  
    self.driver.get('https://www.16personalities.com/free-personality-test')  
    self.driver.maximize_window()  
elif self.button=="Test" :  
    try:  
        self.driver.close()  
    except:  
        print("already closed")  
        self.label_3.setText("Please choose your Personality")  
        self.comboBox.hide()  
        self.comboBox2.show()  
        self.button="Next"  
        self.pushButton.setText("Get Career")
```

7.6 Final Result

7.6.1 UI for the final result

The following code snippet shows the code on how to build the GUI to display the final result of all the three career predicting stages. The final result predicted as a combination of top two careers suitable for the student. We arrive at this result as a consequence of the three

stages, aptitude test, twitter analysis and personality test. The top two careers are predicted using a dataset formulated using the results from the previous stages.

```
class Result_Page(object):
```

```
    def setUpUi(self, Result_Page, Result_list, User_details):
```

```
        self.User_details=User_details
```

```
        self.Result_list=Result_list
```

```
        self.Result_Page=Result_Page
```

```
        self.Result_Page.setObjectName("Result_Page")
```

```
        self.Result_Page.resize(590, 640)
```

```
        self.groupBox = QtWidgets.QGroupBox(self.Result_Page)
```

```
        self.groupBox.setGeometry(QQtCore.QRect(10, 250, 570, 361))
```

```
        icon = QtGui.QIcon()
```

```
        icon.addPixmap(QtGui.QPixmap("Images/icon.jpg"),QtGui.QIcon.Normal,QtGui.QIcon.Off)
```

```
        self.Result_Page.setWindowIcon(icon)
```

```
        font = QtGui.QFont()
```

```
        font.setFamily("MV Boli")
```

```
        font.setPointSize(10)
```

```
        font.setBold(False)
```

```
        font.setItalic(False)
```

```
        font.setWeight(50)
```

```
        self.groupBox.setFont(font)
```

```
        self.groupBox.setStyleSheet("color:rgb(0,0,0);\n"
```

```
"border-color: rgb(255, 170, 0);\n"
```

```
"font: 15pt \"MV Boli\";")
```

```
self.groupBox.setAlignment(QtCore.Qt.AlignCenter)
self.groupBox.setFlat(False)
self.groupBox.setCheckable(False)
self.groupBox.setObjectName("groupBox")
self.frame = QtWidgets.QFrame(self.groupBox)
self.frame.setGeometry(QtCore.QRect(0, 0, 570, 500))
self.frame.setFrameShape(QtWidgets.QFrame.StyledPanel)
self.frame.setFrameShadow(QtWidgets.QFrame.Raised)
self.frame.setObjectName("frame")
self.label = QtWidgets.QLabel(self.frame)
self.label.setGeometry(QtCore.QRect(0, 40, 570, 30))
self.label.setText("The Predicted Careers matching your profile is:")
```

CHAPTER 8

RESULTS

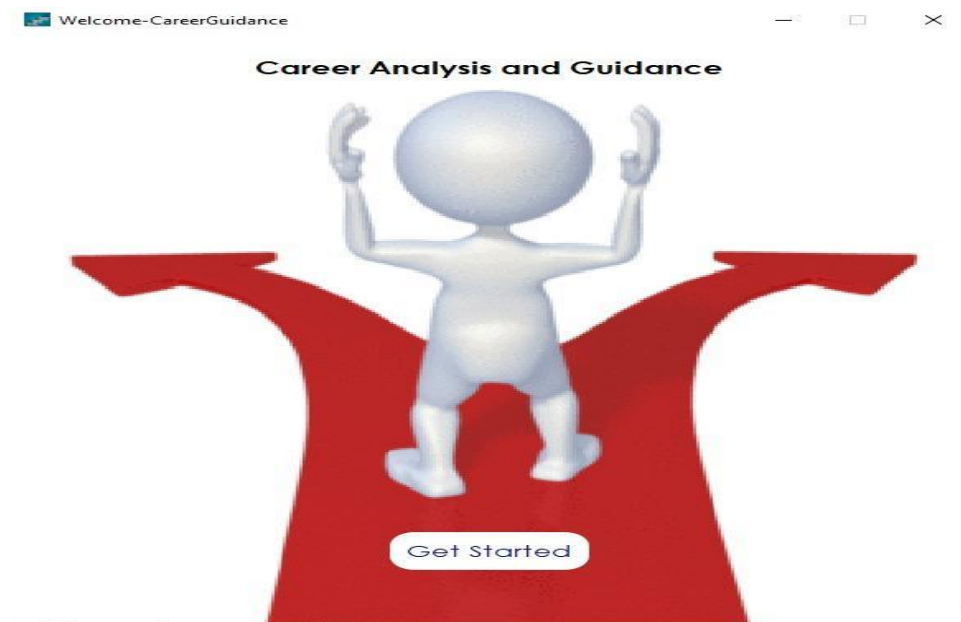


Figure 8.1 : Front page

The figure 8.1 depicts the front page with a get started button ,using which the user can get started.

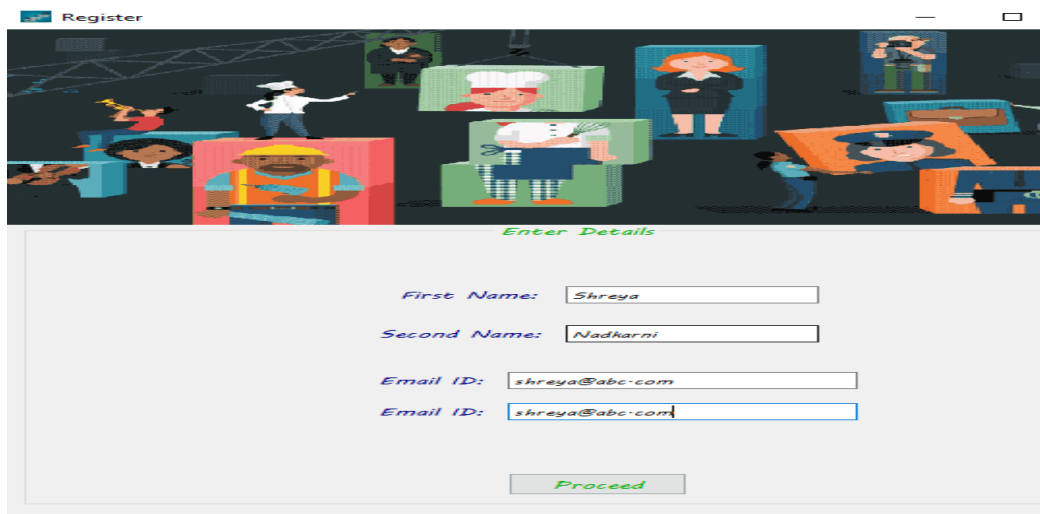


Figure 8.2 : Registration page

Figure 8.2 shows the registration page of the application for the users to register .

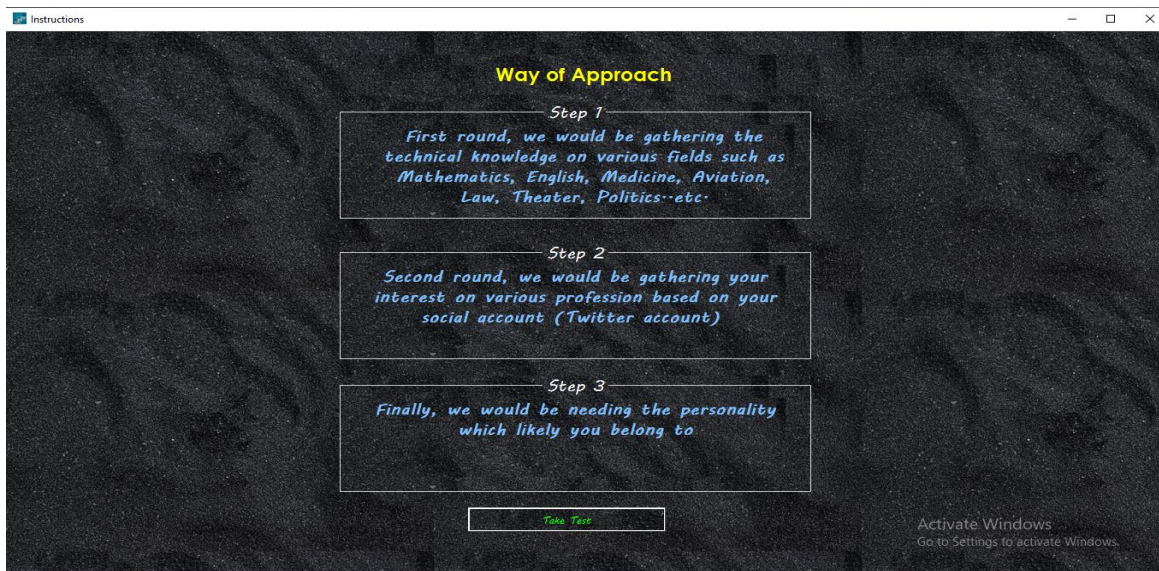


Figure 8.3 : Instruction page

Figure 8.3 shows the instruction page of the Career Guidance Application.

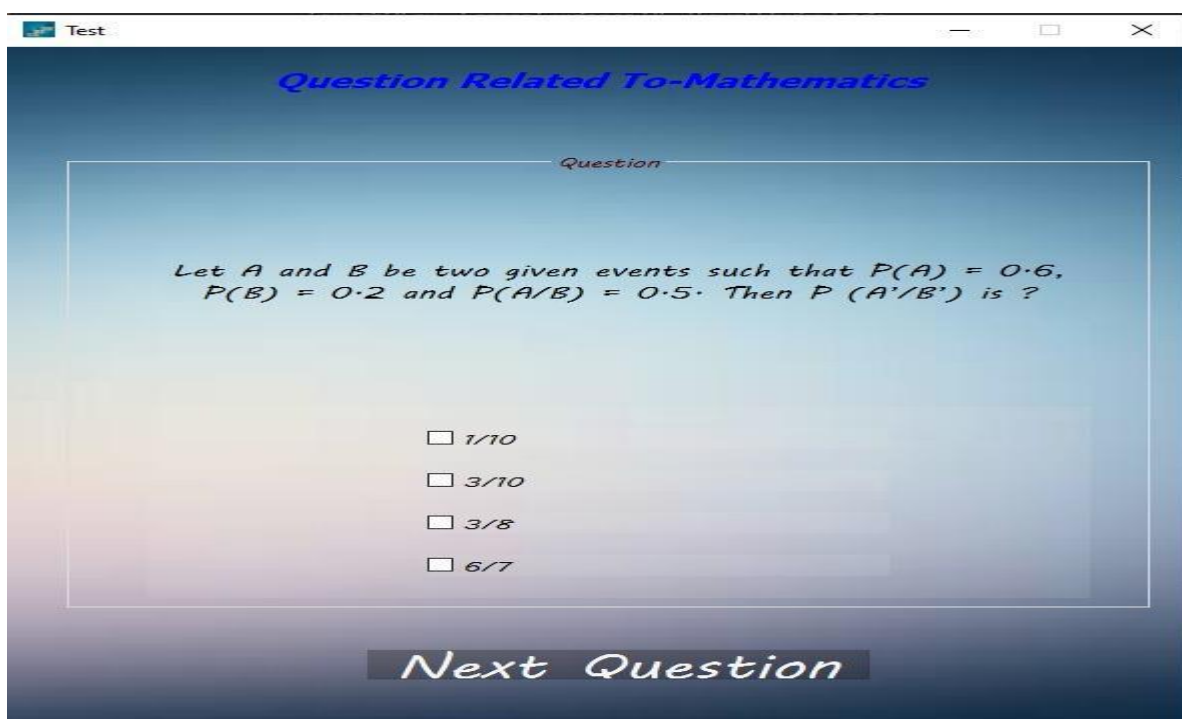


Figure 8.4 : Questionnaire page

Figure 8.4 shows the questionnaire page with questions related different subjects.

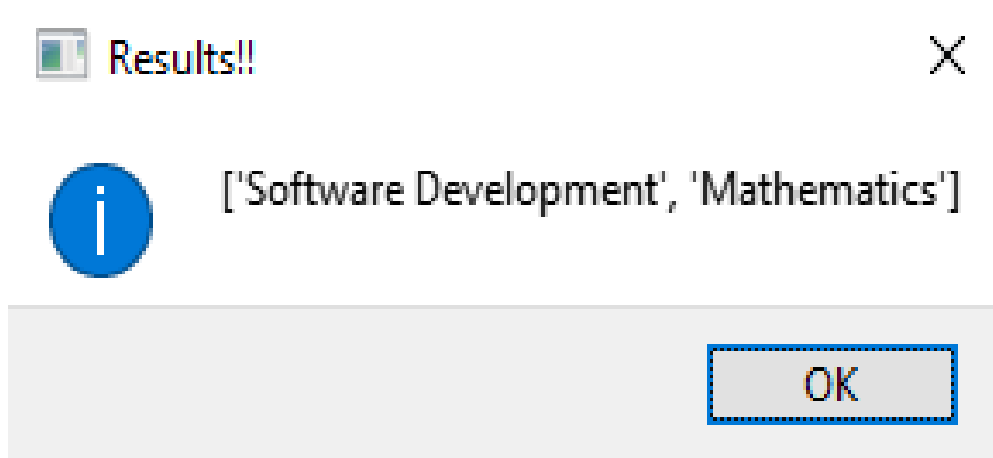


Figure 8.5 : Result

Figure 8.5 shows the results of the candidates who took up the aptitude test

A screenshot of a 'Form' window with a blue background. It features a Twitter logo icon, the text 'Enter Your Twitter ID', a text input field with the example 'Eg: narendramodi', the question 'Do you have a twitter account?', a 'Yes' dropdown menu, and a 'Next' button.

Figure 8.6 : Twitter based occupation result

Figure 8.6 shows the twitter based occupation result where the users first asked if they have a twitter account or not. If yes ,then the hidden text area for entering the twitter handle of the respective user appears on the screen.



Figure 8.7 : Personality test

Figure 8.7 is the page built to ask the users if they have taken up the personality test or not.

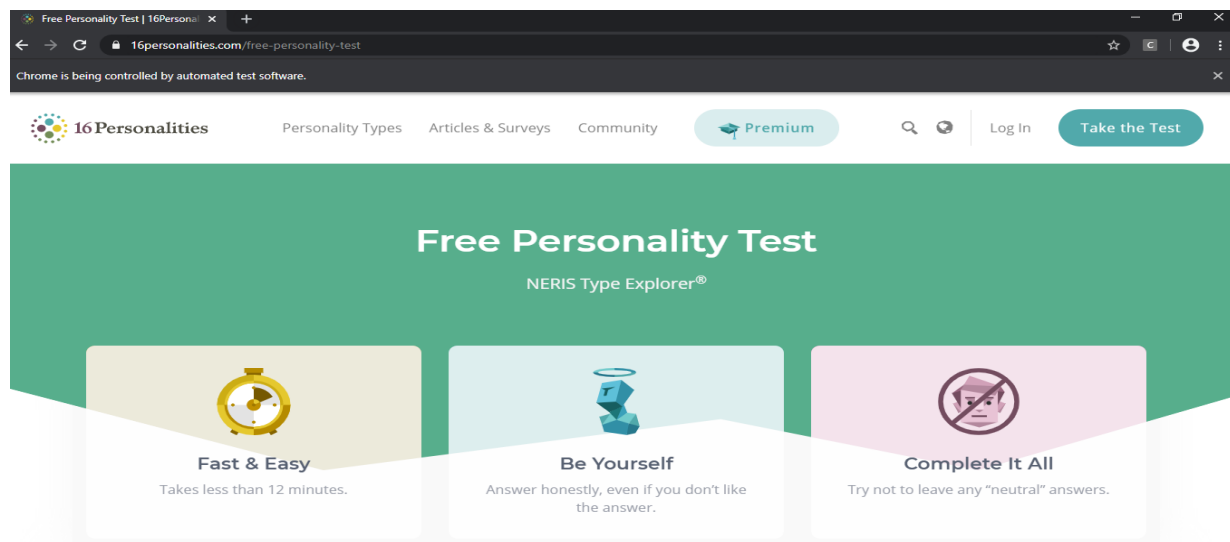


Figure 8.8 : Personality test url directed page

If the user hasn't taken the personality test and chooses No as explained in figure 8.7, then the user is directed to the 16 personality test website to take up the personality test. After taking the test the students are redirected back to the application.



Figure 8.9: Types of personality

Figure 8.9 shows the different types of personality where a candidate can choose one of the personality, which was a result of the personality test taken. Once the personality type is chosen, the user can click on Get Career to find out the result.

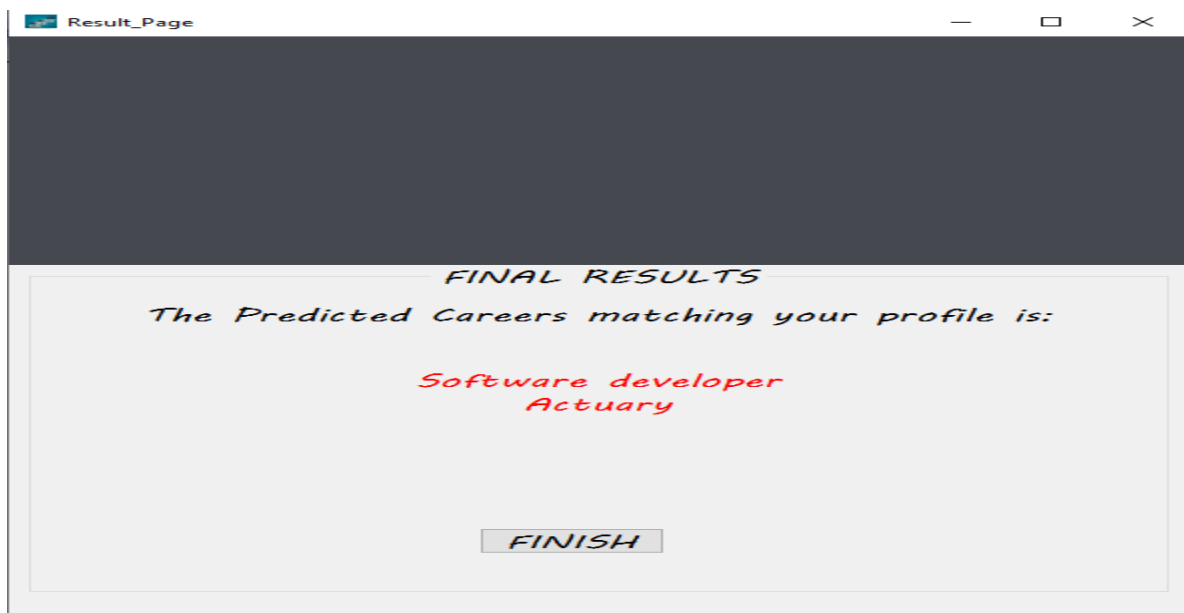
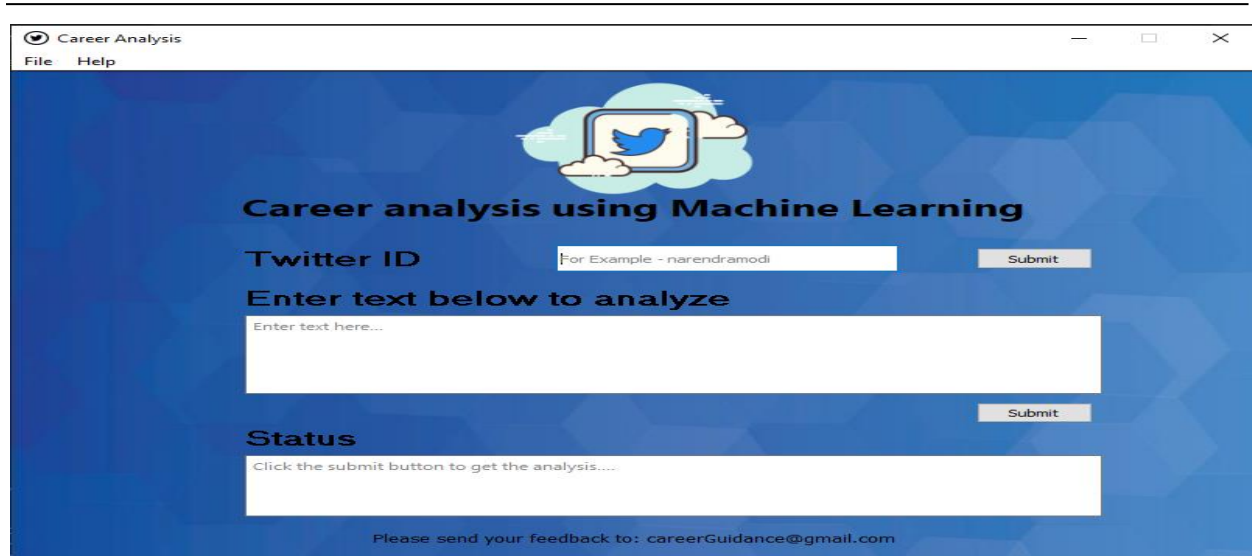


Figure 8.10 : Final result

Figure 8.10 shows the final result or the appropriate career paths to the users.



The screenshot shows a web application window titled "Career Analysis" with a menu bar containing "File" and "Help". The main content area has a blue background with a hexagonal pattern. At the top center is a logo featuring a blue bird (Twitter) inside a white rounded rectangle, surrounded by white clouds. Below the logo, the text "Career analysis using Machine Learning" is displayed in bold. Underneath, there is a section for "Twitter ID" with a text input field containing "For Example - narendramodi" and a "Submit" button. Below this is a section titled "Enter text below to analyze" with a large text input field containing "Enter text here..." and a "Submit" button. At the bottom, there is a "Status" section with a text input field containing "Click the submit button to get the analysis....". At the very bottom of the page, a footer text reads "Please send your feedback to: careerGuidance@gmail.com".

Figure 8.11 : Career analysis using Machine Learning

Figure 8.11 shows the Career analysis using Machine Learning which helps one to figure out their careers by either entering text in the text area or by entering the twitter ID.



This screenshot shows the same web application window as Figure 8.11, but with different input and output. The "Twitter ID" input field now contains "DaveWallsworth". The "Status" input field now displays "Pilot Career detected!". All other elements, including the logo, title, and buttons, remain the same as in Figure 8.11.

Figure 8.12: Positive sentiment using the twitter ID

Figure 8.12 shows the positive sentiment where the status shows the pilot career is detected for the entered twitter ID.

The screenshot shows a web application titled "Career Analysis" with a menu bar containing "File" and "Help". The main content area has a blue background with a hexagonal pattern. At the top, there is a logo of a Twitter bird inside a cloud. Below the logo, the text "Career analysis using Machine Learning" is displayed. There are two input sections: "Twitter ID" with a text box containing "For Example - narendramodi" and a "Submit" button; and "Enter text below to analyze" with a text box containing "Let us go watch a movie" and a "Submit" button. Below these, the "Status" section shows "Negative Sentiment". At the bottom, a feedback message says "Please send your feedback to: careerGuidance@gmail.com".

Figure 8.13 : Negative sentiment using text analyser

Figure 8.13 shows negative sentiment with respect to the pilot career by taking the input from the user in the text area.

The screenshot shows the same "Career Analysis" application. The "Twitter ID" section remains the same. The "Enter text below to analyze" section now contains the text "Landing in Manali after a 14 hour flight for the first time as co-pilot!". The "Status" section now shows "Pilot personality based Sentiment". The feedback message at the bottom remains the same.

Figure 8.14 : Positive sentiment using text analyzer

Figure 8.14 shows positive sentiment with respect to the pilot career by taking the input from the user in the text area.

CHAPTER 9

TESTING

Testing is the process of evaluating a system or its component(s) with the intent to find whether it satisfies the specified requirements or not. Testing is executing a system in order to identify any gaps, errors, or missing requirements in contrary to the actual requirements.

9.1 Testing Principle

Before applying methods to design effective test cases, a software engineer must understand the basic principle that guides software testing. All the tests should be traceable to customer requirements.

9.2 Testing Methods

There are different methods that can be used for software testing. They are

- **Black-Box Testing:** The technique of testing without having any knowledge of the interior workings of the application is called black-box testing. The tester is oblivious to the system architecture and does not have access to the source code. Typically, while performing a black-box test, a tester will interact with the system's user interface by providing inputs and examining outputs without knowing how and where the inputs are worked upon.
- **White-Box Testing:** White-box testing is the detailed investigation of internal logic and structure of the code. White-box testing is also called glass testing or open-box testing. In order to perform white-box testing on an application, a tester needs to know the internal workings of the code. The tester needs to have a look inside the source code and find out which unit/chunk of the code is behaving inappropriately.

9.3 Levels of Testing

There are different levels during the process of testing. Levels of testing include different methodologies that can be used while conducting software testing. The main levels of software testing are:

- **Functional Testing:** This is a type of black-box testing that is based on the specifications of the software that is to be tested. The application is tested by providing input and then the results are examined that need to conform to the functionality it was intended for. Functional testing of software is conducted on a complete, integrated system to evaluate the system's compliance with its specified requirements. There are five steps that are involved while testing an application for functionality.
 - The determination of the functionality that the intended application is meant to perform.
 - The creation of test data based on the specifications of the application.
 - The output based on the test data and the specifications of the application.
 - The writing of test scenarios and the execution of test cases.
 - The comparison of actual and expected results based on the executed test cases.

- **Non-functional Testing:** This section is based upon testing an application from its non-functional attributes. Non-functional testing involves testing software from the requirements which are non-functional in nature but important such as performance, security, user interface, etc. Testing can be done in different levels of SDLC. Few of them are

9.3.1 Unit testing

Unit testing is the testing of each module and the integration of the overall system is done. Unit testing becomes verification efforts on the smallest unit of software design in the module. This is also known as 'module testing'. The modules of the system are tested separately. This testing is carried out during the programming itself. In this testing step, each model is found to be working satisfactorily as regard to the expected output from the module. There are some validation checks for the fields. For example, the validation check is done for verifying the data given by the user where both format and validity of the data entered is included. It is very easy to find error and debug the system.

Test Case ID	Test Scenario	Test Steps	Test data/Fields	Expected Results	Actual Results	Pass/Fail
UT01	Application gets started when LaunchUi.py is executed	run python3 LaunchUi.py command	Input- Enter Command	Launching of Welcome window	Launching of Welcome window	Pass
UT02	Button Click event navigating the window to respective register page	Click on Get started given on the Welcome window.	Input- Button Click	Navigate to register window	Navigate to register window	Pass
UT03	Enter the empty first name on Register window	Click on Proceed on without typing anything on first name box	Input - button click	Popup warning window to alert user to enter valid first name	Popup warning window to alert user to enter valid first name	Pass
UT04	Enter the empty Second name on Register window	Click on Proceed on without typing anything on Second name box	Input - button click	Popup warning window to alert user to enter valid Second name	Popup warning window to alert user to enter valid Second name	Pass
UT05	Enter the empty email id on Register window	Click on Proceed on without typing anything	Input - button click	Popup warning window to alert user to enter valid email id	Popup warning window to alert user to enter valid email id	Pass

		on email id box				
UT06	Enter invalid Email ID	Click on proceed by entering invalid Email ID	Input - button click	Popup warning window to alert user to enter valid email id	Popup warning window to alert user to enter valid email id	Pass
UT07	Enter the re-empty email id on Register window	Click on Proceed on without typing anything on email id box	Input - button click	Popup warning window to alert user to enter valid email id	Popup warning window to alert user to enter valid email id	Pass
UT08	Enter different email ids on register window	Give the different input on email and re-enter other email	Input- Text Input	Popup warning window to alert user to email mismatch	Popup warning window to alert user to email mismatch	Pass
UT09	Enter all valid input on register window to proceed to next window	Give valid input on each field and click proceed	Input - Text Input and Button Click	Popup window to reflect successful data entry	Popup window to reflect successful data entry	Pass
UT10	Button Click event navigating the window to respective Test window	Click on the Take Test button given on the Instruction window.	Input-Button Click	Navigate to Test window	Navigate to Test window	Pass

UT11	Select no option in text window	Click on next Question without selecting any option	Input-Button Click	Popup of warning window to suggest no option is selected	Popup of warning window to suggest no option is selected	Pass
UT12	Button transition in last question	Click on all the answer of each question till last one	Input-Button Click	Changing to Next Question to Get Result	Changing to Next Question to Get Result	Pass
UT13	Button Click event navigating the window to respective Form window	Click on ok button of aptitude popup window	Input-Button Click	Navigate to Form window	Navigate to Form window	Pass
UT14	Select Yes for twitter account	Click on drop down button and choose yes	Input - Button Click	Visibility of Textbox to enter twitter id	Visibility of Textbox to enter twitter id	Pass
UT15	Select No for twitter account	Click on Dropdown button and choose no	Input-Button Click	Navigate to personality test Window	Navigate to Personality test Window	Pass
UT16	Select No for personality test to take up personality test	Click on Dropdown button and choose no	Input-Button Click	Navigate to chrome Window with personality test link	Navigate to chrome Window with personality test link	Pass

UT17	Select Yes when personality test is taken up	Click on Dropdown button and choose Yes	Input-Button Click	Transition to choose personality type	Transition to choose personality type	Pass
UT18	Select any Personality to get your career	Click on any personality and click on get career button	Input-Button Click	Navigate to result window	navigate to result window	Pass

Table 9.1 Unit testing

9.3.2 Integration testing:

Data can be lost across an interface, one module can have an adverse effect on the other sub function, when combined, may not produce the desired major function. Integrated testing is systematic testing that can be done with sample data. The need for the integrated test is to find the overall system performance. There are two types of integration testing. They are:

- Top-down integration testing.
- Bottom-up integration testing.

Test Case ID	Test Scenario	Test Steps	Test data/Fields	Expected Results	Actual Results	Pass/Fail
Integration Testing : Bottom Up Approach						
IT01	Transition from Welcome module to Register module	Click on “Get Started” Button	Event Listener Triggered for loading Register module	Successful Loading of Register module	Successful Loading of Register module	Pass
IT02	Transition from Register module to	Fill all the necessary field and click	Event Listener Triggered for loading	Successful Loading of Instruction module	Successful Loading of Instruction module	Pass

	Instruction module	proceed	Instruction module			
IT03	Transition from Instruction module to Test Module	Click on “Take Test” Button	Event Listener Triggered for loading Test Module	Successful Loading of Test module	Successful Loading of Test module	Pass
IT04	Transition from Test module to Form Module for Twitter API module	Click on ok button of aptitude popup window	Event Listener Triggered for loading Form module	Successful Loading of Twitter module	Successful Loading of Twitter module	Pass
IT05	Transition from Twitter Module to Personality Module	Click on ok button of Twitter window	Event Listener Triggered for loading Personality module	Successful Loading of Twitter module	Successful Loading of Twitter module	Pass
IT06	Transition from Personality to Result module	Click on Get career button after specific personality type	Event Listener Triggered for loading Result module	Successful loading of Result module	Successful loading of Result module	Pass

Table 9.2: Integration Testing

9.3.3 System Testing

System testing of software or hardware is testing conducted on a complete, integrated system to evaluate the system's compliance with its specified requirements. System testing falls within the scope of black-box testing, and as such, should require no knowledge of the inner design of the code or logic. System testing is the first step in the Software Development Life Cycle, where the application is tested as a whole.

- The application is tested thoroughly to verify that it meets the functional and technical specifications.
- The application is tested in an environment that is very close to the production environment where the application will be deployed.
- System testing enables us to test, verify, and validate both the business requirements as well as the application architecture.

Test Case ID	Test Scenario	Test Steps	Test data/Fields	Expected Results	Actual Results	Pass/Fail
ST01	Software System Test System	1.Collection of User data 2.Displaying with question and getting answer	1. Aptitude 2.Personality 3. Twitter Analysis	Display of Result with different Career Choices	Display of Result with different Career Choices	Pass
ST02	Twitter API: Testing Bidirectional communication with twitter server	1. Collection of Twitter Id 2. Get Access of Twitter API Key 3 Transfer username to the twitter server	Text with valid username	Fetching of post and comments of username provided	Fetching of post and comments of username provided	Pass
ST03	NLP System: Access of Keyword of the Posts	1.Collection of Post and training the model on these	Text without any special characters	Provided with respective career inclination	Provided with respective career inclination	Pass

Table 9.3 System Testing

CONCLUSION AND FUTURE ENHANCEMENT

Career Guidance System is an innovative idea. The opportunities provided by this e-medium are immense and many students can make use of this medium to choose a career more appropriate to their skills. In today's competitive and technology driven world, with innumerable options available, the student is generally confused on choosing the right or more suited career. The world these days is moving towards "information streams". The information is thrown to the user rather than the user learning about the information. Keeping the above in view, it is felt that the proposed system has the ability to connect to various students and help them connect to most suited career. To conclude, the objective of designing this system is to lend a helping hand to the students aiming for such a career. By using our system, one will be guided towards a career to pursue and how to work towards it. The proposed system currently deals with guiding students in a direction that is right for them i.e. to select a proper career path depending upon the present skill sets and mental abilities. The plan for future is to expand this system by including vocational subjects as well. It may be reiterated that at this juncture our objective is entirely on academic subjects and not on vocational subjects as these arise from an individual's liking and would need further improvisation on the model proposed above.

In future we can create effective web application that can gather information by evaluating and examining. Analytical, Memory Based, Technical, Logical, Hobbies, interests in Technical/Non-Technical, Performance of the student from the child hood and skill based tests can be conducted and information collected can be used to improve the accuracy. The Dataset can be built from several thousands of student's data. We can try to use the clustering methods for better understanding. We can also implement the techniques like Deep Neural Networks and Time series Analysis.

REFERENCES

- [1] Development of a Web-based Building Profession Career Portal as a Guidance Information System for Secondary School Students
- [2] Career Prediction Model Using Data Mining and Linear Classification
- [3] FINDEREST: Identifying Personal Skills and Possible Fields of Study Based on Personal Interests on Social Media Content.
- [4] Understanding What Affects Career Progression Using LinkedIn and Twitter
- [5] Data Career Guidance and Employment Management System [6] A Game-Based Approach to Career Guidance
- [6] Feature Selection Based on Audience's Behavior for Information Filtering in Online Social Networks
- [7] Career Counselling Using Data Mining
- [8] Online Career Guidance System
- [9] Youtube API : <https://developers.google.com/youtube/v3>
- [10] Facebook API: <https://developers.facebook.com/>
- [11] <https://medium.com/@Mandysidana/machine-learning-types-of-classification-9497bd4f2e14>
- [12] Reddit API: <https://www.reddit.com/dev/api/>
- [13] Twitter API: <https://developer.twitter.com/en/docs>