VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Jnana sangama, Belgaum - 590018

PROJECT PHASE - II REPORT

on

PERSONAL CAREER RECOMMENDATION SYSTEM USING MACHINE LEARNING

Submitted in partial fulfillment of the

PROJECT

in

INFORMATION SCIENCE AND ENGINEERING

by

Goutham C **(1HK19IS026)**

Under the guidance of

Dr. A Syed Mustafa
Professor & Head

Prof. Anusha K V
Assistant Professor

Department of Information Science and Engineering,

2022-2023

HKBK COLLEGE OF ENGINEERING

22/1, Nagawara, Bengaluru 560045
E-mail: info@hkbk.edu.in**, URL: www.hkbk.edu.in**

HKBK COLLEGE OF ENGINEERING

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

PROGRAMME EDUCATIONAL OBJECTIVES (PEOs)

PEO 1: To Empower Students through Wholesome Education to achieve academic excellent education in the field of Information Science and Engineering**.**

PEO 2: To Provide Students with in-depth disciplinary knowledge in engineering fundamentals that require to succeed in Information Science and Engineering**.**

PEO 3: To Create Highly Qualified Professionals in multi-disciplinary areas with the knowledge of Information Technologies, Services Globally.

PEO 4: To Inculcate in Students a Professional and Ethical attitude with a strong character effective communication skills, teamwork skills, a multidisciplinary approach, and the ability to relate Engineering issues to broader social context**.**

PEO 5: To Provide Students with an academic environment aware of advanced technological growth leading to life-long learning through innovation and research with professional ethics that uplifts mankind.

PROGRAM SPECIFIC OUTCOMES(PSOs)

Professional Skills:
An ability to identify and analyze requirements, and in designing and implement well-tested technology solutions for rapidly changing computing problems and information system environments.

Problem-Solving Skills:
An ability to Design, develop and optimize solutions for information systems employing fundamentals of system hardware & software, graph theory, finite automata, data storage, and communication networks.

Collaborative Skills:
An ability to communicate and develop leadership skills and work effectively in team environments. They are capable of collaborating to design and implement well-tested solutions for rapidly changing computing problems and information system environments.

Successful Career and Entrepreneurship Skills:
An ability to adapt for innovation and changes and be successful in ethical professional careers along with the impact of computing on society, and platforms in creating innovative career paths to be an entrepreneur, and a zest for higher studies.

HKBK COLLEGE OF ENGINEERING

Accredited by NAAC

BENGALURU-560045

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

A

PROJECT PHASE – II REPORT

on

PERSONAL CAREER RECOMMENDATION SYSTEM USING MACHINE LEARNING

Submitted in partial fulfillment of

PROJECT

VIII Semester, Dept. of Information Science and Engineering 2022-2023

SUBMITTED BY

Goutham C **(1HK19IS026)**

ACKNOWLEDGEMENT

I would like to place our regards and acknowledgment to all who helped in making this project possible.I thank all those whose guidance served as a beacon of light and crowned our efforts with success.

First of all, I would take this opportunity to express our heartfelt gratitude to the management committee - Chairman **Mr. C.M. Ibrahim**, Director **Mr. C.M. Faiz Mohammed,** and the Principal **Dr. Tabassum Ara** for all the infrastructures provided to complete **Project Phase-II** in time.

I am deeply indebted to **Dr. A. Syed Mustafa,** HOD, Information Science and Engineering for the ineffable encouragement he provided for the successful completion of the project.

A special and earnest word of thanks to the project Co-ordinator **Dr.V. Balaji Vijayan** and project guides **Dr. A. Syed Mustafa & Prof. Anusha K V.,** for their constant assistance, support, patience, endurance, and constructive suggestions for the betterment of the project.

I am extremely thankful to the teaching and non-teaching staff of the Department of Information Science and Engineering for their valuable guidance and cooperation throughout our dissertation.

I thank my parents for their support and guidance provided to us to finish the project well ahead of time.I thank my friends who lent their support in every way possible to make sure the project has been completed. Last, but not least we would like to thank God for giving us this opportunity to do everythingin the appropriate time to finish this project.

Goutham C **(1HK19IS026)**

DECLARATION

I hereby declare that the entire work embodied in this Project work "**PERSONAL CAREER RECOMMENDATION SYSTEM USING MACHINE LEARNING**" has been carried out by us during the Seventh semester of Bachelorof Engineering in Information Science and Engineering at HKBK College of Engineering, Bengaluru affiliated to Visvesvaraya Technological University, Belagavi, under the guidance of **Dr. A Syed Mustafa**, Professor and HoD, Dept. of ISE, and our co-guide **Prof. Anusha K V,** Assistant Professor, Dept. of ISE, HKBK College of Engineering, Bengaluru. The work embodiedin this project work is original and it has not been submitted in part-time or full-time completion for any other degree in any other university.

GOUTHAM C **1HK19IS026**

ABSTRACT

Competition in today's society is heavily multiplying day by day. Especially since it is too heavy in the present day's technical world. So as to compete and reach the goal students need to be planned and organized from the initial stages of their education. So it is very important to constantly evaluate their performance, identify their interests and evaluate how close they are to their goal, and assess whether theyare on the right path directed towards their target. In order to cope with the changing education system andthe evolving new technologies, it is important for a student to identify his field of interest and select his bestamong the available wide range of courses. Many students opt the courses which are not of their interest asthey will not be having much knowledge about the courses of their interest. Most of the students in our society decide their future based on what their elders say or they rely on their friends or their family and dothe same course that they had done or doing. There will be no proper guidance for them on choosing their subjects or courses. This project is a part of the progress of education toward better course recommendations. We use a machine learning program that asks the client questions and recommends the better stream based on the skills and academic performance provided.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Nowadays, students are often facing a dilemma in deciding to choose a career in their life. There are several factors that influence students when choosing their career paths such as their personal aptitudes, educational achievement, and their environment. Upon completing their first degree or undergraduate students at the university, students are normally starting to consider a career path that may suit their skill and potential the best. However, many students made the wrong decision in selecting their career due to the lack of experience, help, and advice from friends and relatives, parents and lecturers, or career counseling. It is necessary to deliver career guidance in several ways likecourses, training, and seminars that offer group experiences in future career planning and group or individual counseling activities. The factor that may cause students not to be successful in their careers is the wrong choice of a job that suits them. It requires a decision-making process at an early stage. So this system gives a recommendation for students about their career based on their abilities.

1.1 About Career Prediction

As students are going through their academics and pursuing their interested courses, it is very important for them to assess their capabilities and identify their interests so that they will get to know in which career area their interests and capabilities are goingto put them in. This will help them in improving their performance and motivating their interests so that they will be directed towards their targeted career and get settledin that. Also recruiters while recruiting the candidates after assessing them in all different aspects, these kind of career recommender systems help them in deciding inwhich job role the candidate should be kept in based on his/her performance and otherevaluations.

1.2 Problem Statement

The main problem of difficulty in choosing the right career students is they do not

know how to make decisions and also may be due to a lack of knowledge and information about jobs and careers. In order to help students with career selection, it is essential to build Career Recommendation System with the capacity to meet all theneeds where it provides direction and guidance to students in choosing a career that suits with their skills and abilities.

Hence, we proposed the problem statement **"**Personal Career Recommendation System**"**.

1.3 Objectives

The main objectives of the project are the following:

- To recommend a suitable career option (job role) to the users.
- To make the recruitment system more secure and easier.
- To make the process of job hunting easier for freshers as well as experienced candidates.
- Help in analyzing their abilities.

1.4 Existing System

- There are certain models available in the market for the same problem, every model has its own advantage and disadvantage. Some of the existing systems are using these methods to take input.
- Prediction using Course grades Taking student academic course grades as input may not be efficient way as the individual student grades may depends on college, exam models, paper evaluation so on.
- Prediction using YES/NO questionnaire Asking questionnaire from various required fields and getting an answer in the form of yes/no may lead the student in confuse. Suppose in a particular course student may had only basic knowledge then he would not know whether to give yes/no. so, in this case he wouldn't provide input, and this would lead to wrong prediction.
- Not providing clear road map to the predicting career without showing clear road map how to proceed may put student in dilemma. Prediction with clear road map makes student clear about his/her career.

1.5 Proposed System

In earlier days student's career is forecasted by using questionnaire method. But it is atime-consuming process and it is very difficult to find the status of student's opinions.Various computing techniques are used to predict the career of the student's. In this research work Decision Tree, Random Forest, SVM and XG BOOST machine learning concepts are used to forecast the student's career. Compare to existing traditional methods new computing concepts like machine learning approaches produces better result.

CHAPTER 2

LITERATURE SURVEY

Qingwan,Linye.,[1] Career Recommendation for students based on deep learning and machine learning. – Hindawi.2022

In this paper, the author proposed Career Recommendations for students based on Deep Learning and Machine Learning. A hybrid CNN model is proposed for the employment suggestion of undergrads. And he concluded that by improving the activation function, pooling strategy, and loss function in the algorithm, the quality of model prediction is greatly improved.

Sushma K.,et.al.,[2] Cha Educational Career Recommendation System using Machine Learning. – 2021

In this paper, the author built an educational Career Recommendation System Using Machine Learning. The issues of the cold beginning, trust, and security are addressed in this approach to deal with utilizing the earlier data from every school or college. The suggestion framework is done using Python since it is simple and productive to place into impact calculations on select working frameworks.

Min N.,et,al.,[3] Career choice prediction based on campus big data, the potential behavior of college students- 2020

In this paper, the author created a career path Recommendation on the framework. A novel career pathframework for personalized jobs and skills suggest young, focusing on students and young professionals is presented. And he concluded that the framework can run on cloud infrastructure using apache spark or Hadoop cluster which can handle more number of jobs description and more users. To increase the system capability at any time, new machines can be added to the existing cluster without affecting the running application.

Bharat pat.,et.al.,_A career path recommendation framework-a novel career path framework.

In this paper, the author proposed a Student career prediction Student suggestion and it was proposed to help student career prediction in concluding their capacities which are solid

and weak. And he used the random forest as it contains various choice trees and different subsets of the given datasets and took the ordinary to deal with the accuracy of that datasets. It tends to be utilized for both Arrangements and Regression issues in ML[4]

Dr.Sadasivam R.,et.al., Student Career Prediction.-2022

In this paper, the author proposed a Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students and Proposed a prototypical cluster center generation approach to use the priori information from each college A novel regularization object was presented by overcoming any barrier between this present reality models and prototypical bunch.[5]

JOURNAL NAME & YEAR	**AUTHOR**	**TITLE& METHOD**	**REMARKS**
Hindawi.2022.	Qingwan, Linye	Career Recommendations for students based on Deep Learning and Machine Learning. A hybrid CNN model is proposed for employment recommendation of college students.	By improving the Activation function, pooling strategy, and loss function in the algorithm, the quality of model prediction is greatly improved.
International Journal of Advanced Research in Computer and Communication Engineering. 2021.	Sushma Koushik N, Chandana M S Lavanya V, Suhas Y, HarshithaV	Educational Career Recommendation System Using Machine Learning. The problems of cold start, trust, and privacy are solved in this approach	The recommendation system is built in Python because it is easy and efficient to put into effect algorithms on exclusive operating systems.
Multidisciplinary Digital Publishing Institute. 2020.	MinNie, Zhaohui Xiong Ruiyang Zhong, Wei Deng and Guowu Yang	Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students Proposeda Prototypical cluster center generation approach to use the priori information from each college	A noval regularizatio n Item was introduced by To bridge the gap between the real-world examples and prototypical cluster centers.

Research gate.2017	Bharat Patel, Varun Kakuste, Magdalini Eirinaki	A career path Recommendation on framework A novel career path framework for personalized job and skills Recommend youngs, focusing on students and young professionals is presented.	The system can runon cloud infrastructure using apache spark or Hadoop cluster which can handle more number of jobs description and more users. To increase the system capability at any time, new machines can be added to the existing cluster without affecting the running application.
International Journal of Health Science Journal of Health Sciences. 2022.	Dr. Sadasivam R, Paramasaivam S, Prakash raj N, Saravanam M	Student carrer prediction Student suggestion was proposed to help student career prediction in concluding there abilities in which they are solid and frail.	RF contains various choice threes different subsets of thegiven datasets and taken the normal to work on the precision of that datasets. It tends to be utilized for both Arrangements and Regression issue in ML

CHAPTER 3

3.1 SYSTEM REQUIREMENTS SPECIFICATION

3.1.1 Functional Requirements

The functional requirements for a system describe what the system should do. These requirements depend on the type of software being developed; the general approachtaken by the organization when writing requirements. The functional system requirements describe the system function in detail, its inputs and outputs, exceptions, and so on.

- The system first includes the collection of data
- Next, Pre-processing (cleaning) the training data and testing data.
- Based on the training and test data, the efficiency of the model will be analyzed.
- The system will ask the user to enter the necessary details.
- The system will predict the career.

3.1.2 Non-Functional Requirements

Non-functional requirements, as the name suggests, are requirements that are not directly concerned with the specific functions delivered by the system. They may relate to emergent system properties such as reliability, response time, and store occupancy. Alternatively, they may define constraints on the system such as the capabilities of I/O devices and the data representations used in system interfaces.

3.2 Hardware & Software Requirements

3.2.1 Hardware Requirements

- Processor-i3 4th gen or high.
- RAM - 4GB or high.
- HDD – 256GB.

3.2.2 Software Requirements

- Jupyter notebook.
- Google collab
- Sikit Learn
- Matplotlib
- Python

CHAPTER 4

SYSTEM TOOLS

4.1 Machine Learning

Machine Learning is an application of "Artificial Intelligence" (AI). it's quick growing technology in numerous field which enable computers to find out instantly from the historical information. Machine Learning utilize many formulae for develop a prediction model, currently, this system more and more utilized in numerous fields likecare, image process, speech recognition, object detection, robotics, data processing, video games, text analysis finance and lots of additional. Machine Learning formula build a mathematical model supported sample information called "training data" so asto form a prediction or call while not being expressly programmed to perform the task. Being a field of science, Machine Learning deal with the ways that within which Machine Learning victimisation its previous expertise. With the arrival of recent techniques and approaches of Machine Learning, we have a tendency to presently possess a capability to hunt an answer to health connected issues like polygenic disorder and high blood pressure detection. A system victimisation Machine Learningtechnique that has the aptitude to discover whether or not the person has polygenic disorder or not are often developed. Moreover, detection the polygenic disorder Associate in Nursingd high blood pressure in an early stage, results in treating the patient before it become vital.

Features of Machine Learning

- It utilize the data to observe patterns in a dataset and modify program operation accordingly.
- It concentrate on the development of computer programs that can learn themselves to grow and switch when open to new data.
- It allow computers to detect concealed insights using sequential algorithms without being specific programmed.
- It automates analytical model buildings using statistical and Machine Learning algorithms.

4.1.1 How does Machine Learning work

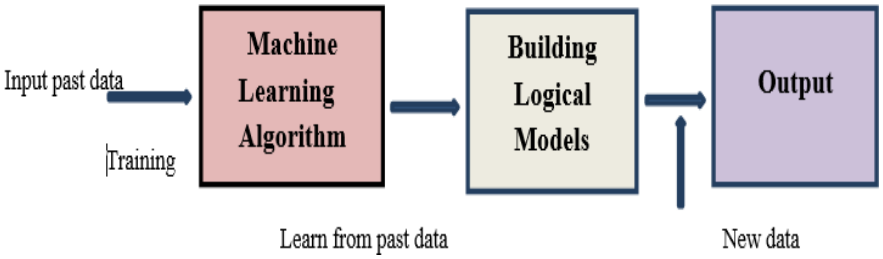


Figure 4.1.1: How Machine Learning Works

A Machine Learning system learns and improve from past input data, and construct the prediction model. Every time we input a new data to a prediction model, it predictsthe output for it as shown in Fig.4.1.1 The accuracy of predicted output rely on the large amount of data. Machine learning deals with huge amount of data and exploringdata to identifying numerous patterns in the given dataset and involves minimal humanintervention. Machine learning model learns to adopt to new example and produce a better accuracy.

4.2 Types of Machine Learning

4.2.1 Supervised learning

It is a kind of Machine Learning method which help to train models from labelled data,it permit to prognosticate the output for future or unseen data. This learning method handle labelled data where the output data design are known to the system. This learning technique is less complex, more accurate and reliable result.

It has classified into two categories: classification and regression.

- Classification is a supervised learning type with the output has predefinedlabels that have discrete values. The aim is to prognosticate discrete values that belongto a specific class and estimate on the basis of accuracy

- Regression is a supervised learning type the output has a continuous valuethat is unrestricted to defined separate values. Regression is a process to find the correlation between input and output variable. The task of regression algorithm is to find a mapping function to input variable with continuous output variable.

Some examples of supervised learning algorithm are:

- Support vector machine for classification problems.
- For regression problems using linear regression.
- For classification and regression problems using random forest algorithm.

4.2.2 Unsupervised learning

It is a kind of Machine Learning method used to train the Machine Learning algorithm using the data has no labels or unclassified and allows the algorithm to acton the data without guidance, the machine just looks for whatever pattern it can find.Unsupervised learning work with un-labelled data in which the output is just based on the collection of perceptions. This learning technique is more complex and moderately accurate but reliable results. Unsupervised learning can be used for anomaly detection and clustering.

It has classified into two categories: clustering and associative rule mining problems.

- **Clustering:** The method of grouping similar entities together is called clustering. The aim of unsupervised learning method is to detect similar objects are grouped into one cluster and are different from objects in further cluster.
- **Association rule:** It is type of unsupervised learning method detect the dependencies of one data item to further data item and merge them such that they can help profit better.

  Some examples of unsupervised learning algorithm are:

- k-means for clustering problem.
- Apriori algorithm for associative rule mining problem.

4.2.3 Reinforcement learning

It is a kind of Machine Learning where an agent learns to act in a environment by performing activity and seeing the results. Agent is reinforcement learning algorithmthat learn from trial and error, for each good action, the agent gets positive reward and for each negative action the agent gets punishment or penalty. The agent learns every time what action led to positive feedback and negative feedback. In this learning process there is no labelled data, so the agent learns by its own experience.

In this project, X-ray images obtained from two different sources were used for the diagnosis of COVID-19. A COVID-19 X-ray image database was developed by Cohen JP using images from various open access sources. This database is constantlyupdated with images shared by researchers from different regions. At present, there are 127 X-ray images diagnosed with COVID-19 in the database. Below Figure 2 shows a few COVID-19 cases obtained from the database and the findings of the experts.

- Sample an open source dataset of X-ray images for patients who have testedpositive for COVID-19
- Sample “normal” (i.e., not infected) X-ray images from healthy patients.
- Train a CNN to automatically detect COVID-19 in X-ray images via the dataset we created.
- Evaluate the results from an educational perspective.

There are some steps to be followed in this system i.e. are Preprocessing, Feature selection, Feature extraction, Classification & Staging.

4.3 Programming Language Used

4.3.1 Python

Python is an interpreted, object-oriented, high-level programming language withdynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for

use as a scripting or glue language to connect existing components together. Python's simple, easy-to-learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Pythoninterpreter and the extensive standard library are available in source or binary formwithout charge for all major platforms and can be freely distributed.

4.4 Algorithm Used

4.4.1 Decision Tree

A decision Tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart-like tree structure, where each internalnode denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.

A decision tree is a graphical representation of possible solutions to a problem based on certain conditions. It is a decision support tool that uses a tree-like modelof decisions and their possible consequences, including chance event outcomes, resource costs, and utility.

A decision tree can be used in a career prediction system to predict an individual's futurecareer path. The decision tree can be used to identify the most likely career paths based on the individual's skills, interests, and personality. The decision treecan also be used toidentify the most likely career paths based on the individual's current career path.

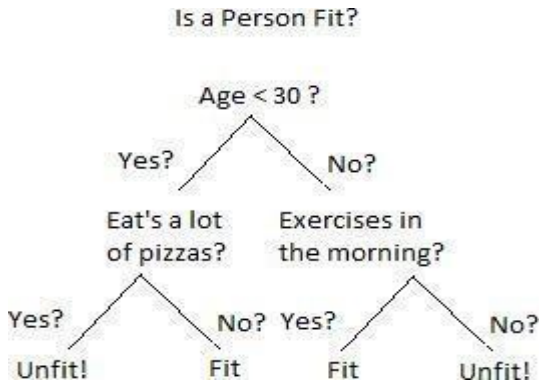


Fig No:4.4.1 Example of Decision Tree

4.4.2 Support Vector Meachine (SVM)

There are many different types of machine learning algorithms that could be used forcareer prediction, but support vector machines (SVMs) are a popular choice. SVMs area type of supervised learning algorithm, which means they learn from labeled training data. In the context of career prediction, this would mean using data about people's past careers (or other relevant information) to train the SVM to predict futurecareers.SVMs are known for being effective in high-dimensional spaces, which makes them well-suited for this task. They are also relatively robust to overfitting, meaning they can generalize well from the training data to new data.

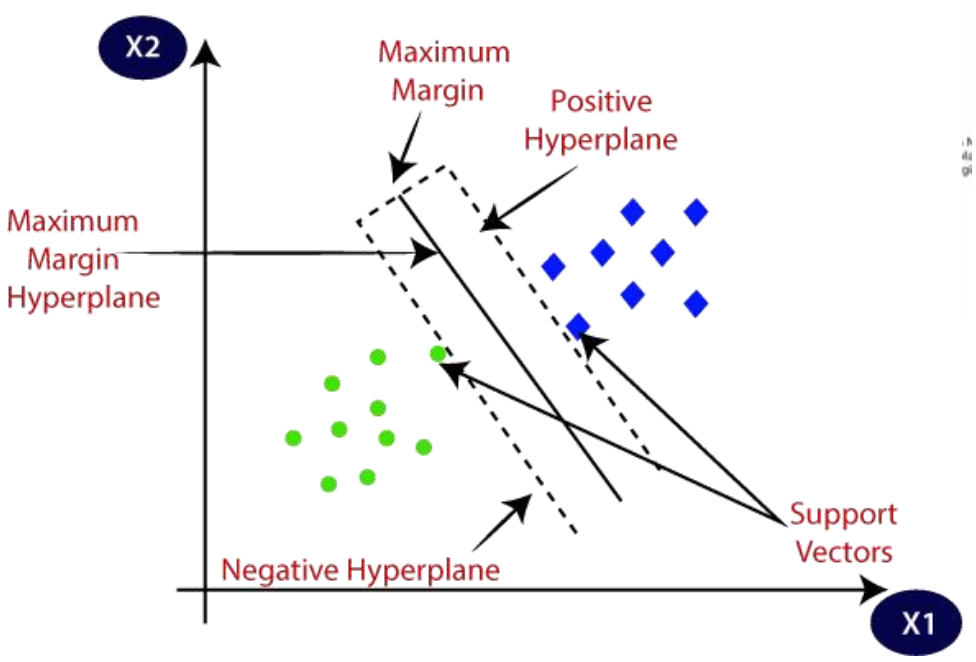


Fig No:4.4.2 Support Vector Meachine

4.4.3 Random Forest

One advantage of using a random forest in a career prediction system is that it can handle a large number of input features without requiring them to be individually scaled. This is because the random forest algorithm is based on a decision tree, which

is not sensitive to feature scaling. In addition, random Forests are also resistant to overfitting, meaning that they can still make accurate predictions on unseen data evenif the training data is not representative of the entire population.

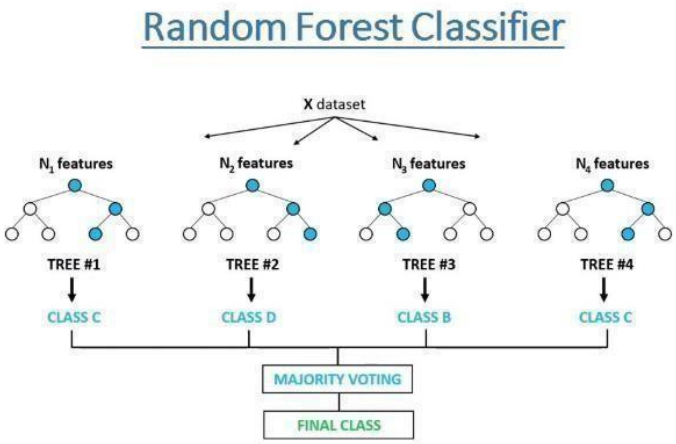


Fig No:4.4.3 Random Forest Classifier

4.4.4 XG BOOST

XGBOOST was introduced by TIANQI CHEN, a great machine learning library. It extends Grade Boosting Machines, which has been conceded as one of the most effective supervised literacy algorithms. XGBoost stands for “Extreme Gradient Boosting," according to the abstract. Grade Boosting Machines are one of the stylishperforming algorithms in supervised literacy, and XGBoost is an extension of them.XGBoost stands for “Extreme Gradient Boosting," according to the abstract. The model and parameters are the abecedarian aspects of XGBoost, as it's a supervised literacy strategy. The model is a fine model that's used to prognosticate issues grounded on input values, and the parameters must be learned from the data set**.**

CHAPTER 5

SYSTEM DESIGN

5.1 System Arcitecture

The system is developed in python. The system takes input from the datasetsand produces the result.

The system-building process consists of the following sequential steps:

1. Fetching the dataset
2. Cleaning the dataset
3. Selection of the features of a dataset
4. Train Model
5. Use the model to predict results.

Using real data from San Jose State University's Career Center, including job postings aimed at Computer Science and Engineering majors, and real resumes of students of the Computer Engineering department, we developed a proof-of- concept system prototype. For the implementation, we employed various tools from the Hadoop ecosystem, namely the Apache Tika parser library for the Text Mining module, the Apache Mahout recommendation algorithm libraries for the Recommendation Enginesmodule, and the NoSQL MongoDB database.

The recommendation engine runs every half an hour and generates new jobs and skills recommendation for users. The algorithm considers profiles of new user generated after last run and also considers already existing users' profiles. So, it updates the recommendations for existingusers and generates new recommendations for newly registered users every 30 minutes.

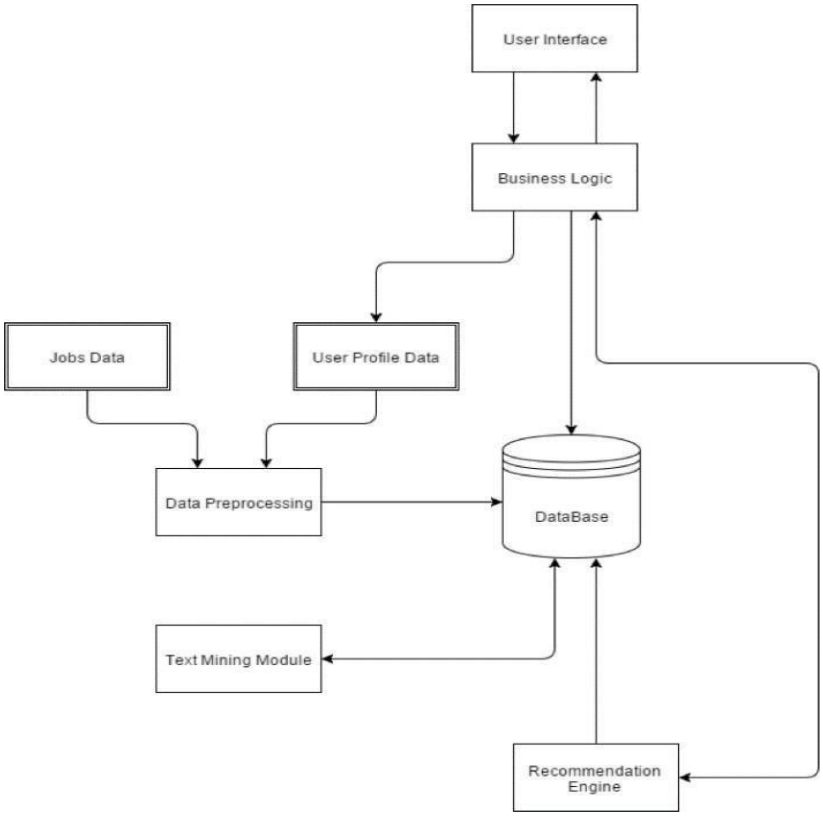


Fig No:5.1 System Arcitecture

5.2 Use Case Diagram

An interaction between a user and a system is described by a use case diagram. Use case diagrams to describe what a system does from the standpoint of an external observer. The emphasis is on what a system does rather than how. Use case diagrams are closely connected to scenarios. A scenario is an example of what happens when someone interacts with the system. A use case diagram is a collection of actors, use cases, and their communications. For initial development, we can use this use case. Inthis use case diagram, we can see the following use cases and actors. Use cases are self-explanatory and they represent the main functions of the Career PredictionSystem.

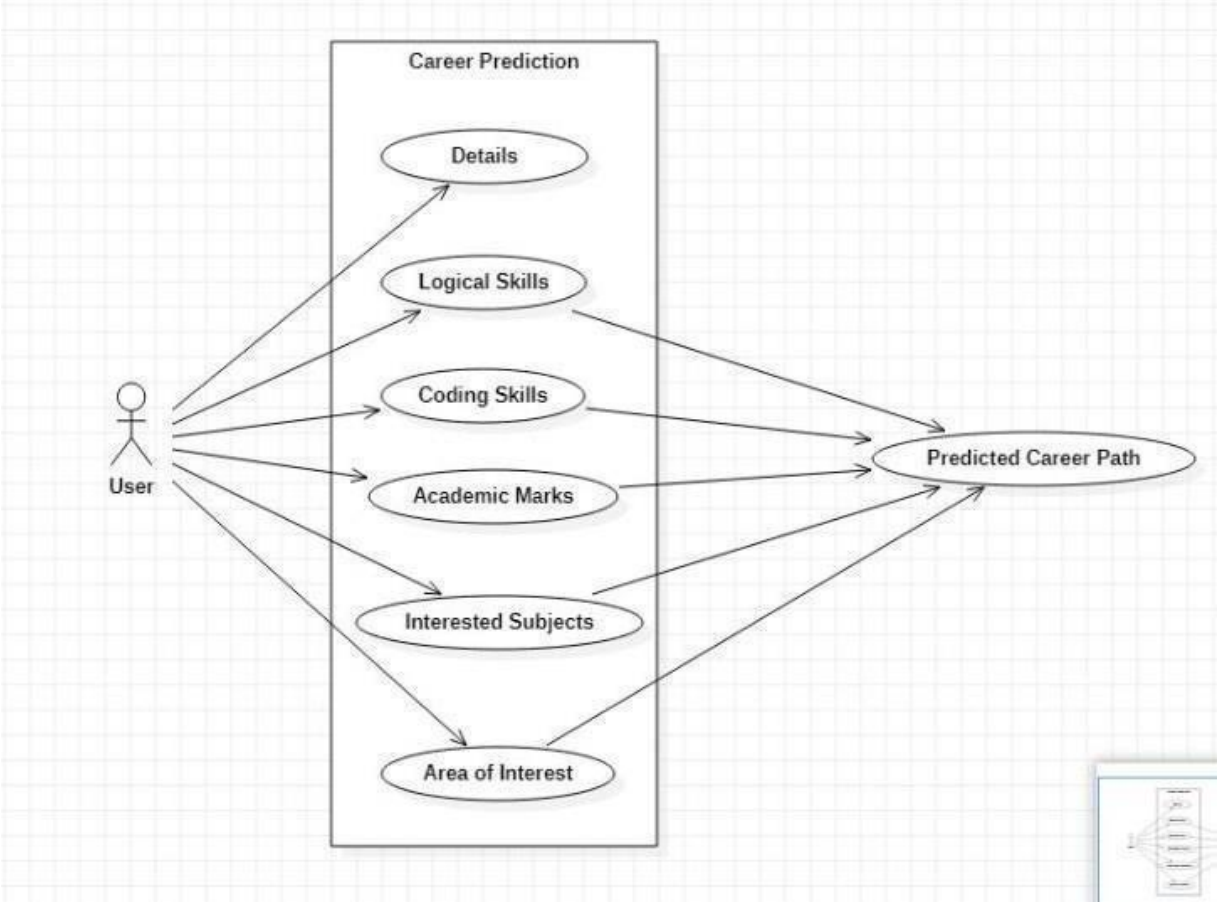


Fig No :5.2 Use Case Diagram

5.3 Data Flow Diagram

The Figure shows the process flow when we were choosing the model for statistical analysis. First, we import the data, and then we process and clean the data. Corresponding tothe data we fit a statistical model, and then we evaluate the model andcross-validate the modelto ensure its functionality. We implement the selected modelupon the data set. Finally, we produce a report based on the result.

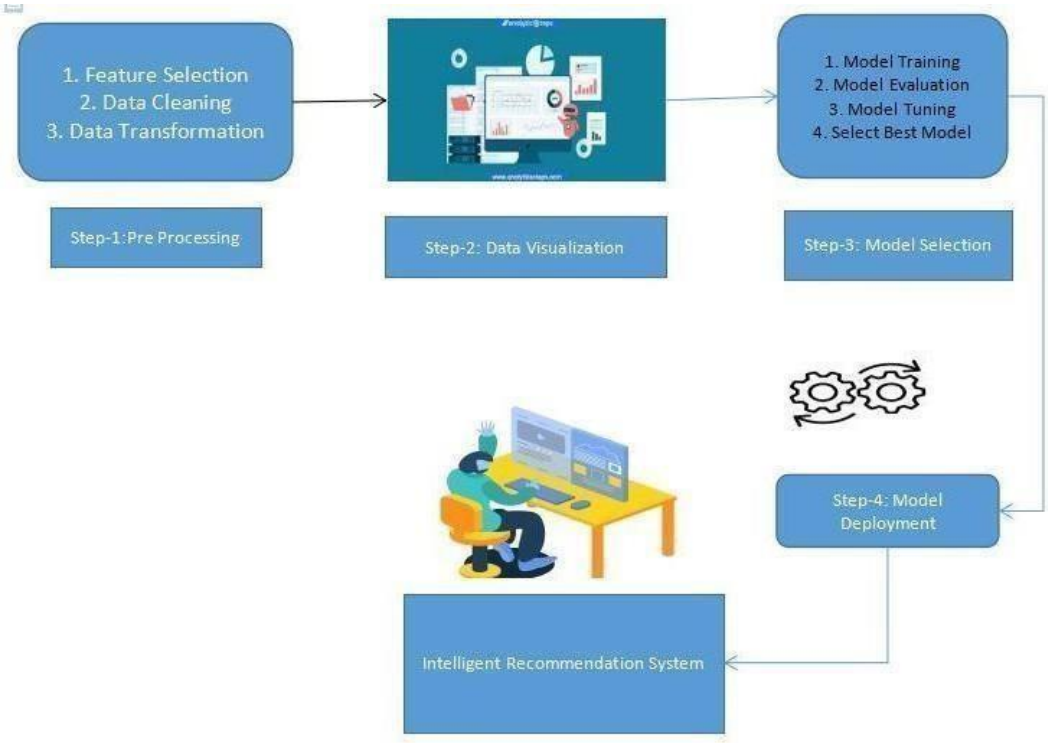


Fig No:5.3 Data Flow Diagram

CHAPTER 6

IMPLEMENTATION

6.1 Module Description

6.1.1 Introduction

In software industries, Object Oriented Development process is widely used. Object-Oriented Programming has heavily contributed toward a standardized method of modeling known as the Unified Modeling Language (UML). UML has become a synonym for software modeling. UML is commonly used to model the software architecture as per the requirements and it includes a set of graphic notation techniques to create visual models of software-intensive systems. With the help of different UML diagrams for building the software, source code can beeasily generated. The correctness of the source code depends on the UML specification which needs to be standard, complete, precise, and unambiguous. A good UML specification leads to clearly defined semantics and an efficient code can be generated. The project is based on predicting career.

6.1.2 Methodology

The system is developed in python. The system takes input from the datasets andproduces The result. The system-building process consists of following sequentialsteps.

- Fetching the dataset
- Cleaning the dataset
- Selection of the features of datasets
- Train Model
- Use the model to predict results

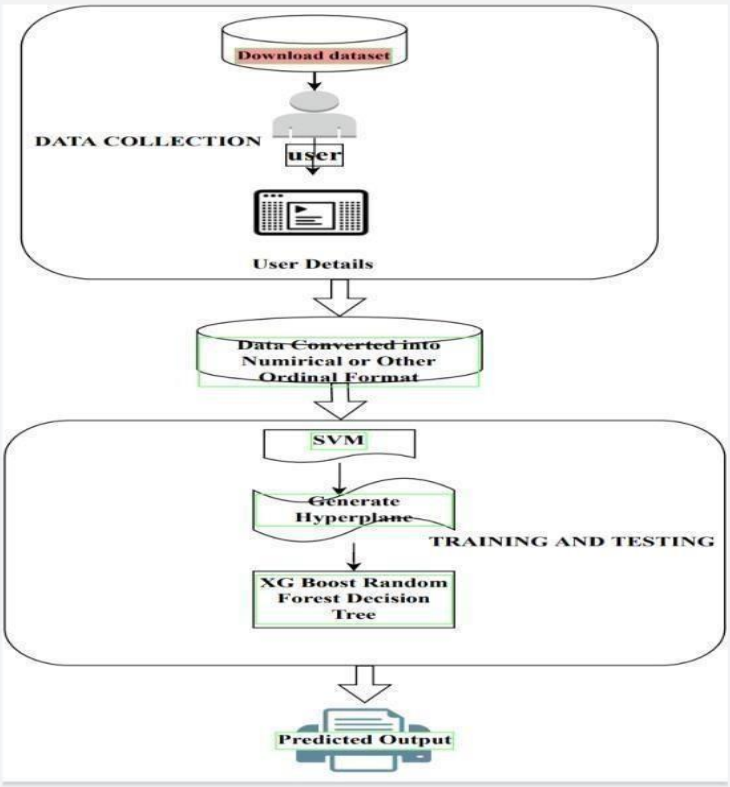


Fig 6.1.2 Methodology

6.1.3 Data Extraction

The first step in our project is to collect data on students' marks, interests, and skills. We have collected the data from students from the Kaggle website and storedit in CSV files.

6.1.4 Data Processing and Cleaning

The next step in the project is the processing of the data extracted. Data collected from various means will be in an unorganized format and there may be a lot of null values, invalid data values, and unwanted data. Cleaning all these data and replacingthem with appropriate or approximate data and removing null and missing data and replacing them with some fixed alternate values are the basic steps in pre-processingdata. The dataof students' details are directly sent for data cleaning. We read the data

from all the columns, then combine and store it in a common file for easy processingand access.

6.1.5 Encoding

Encoding is a technique of converting categorical variables into numerical values thatcould be easily fitted to a machine learning model.

A machine learning algorithm needs to be able to understand the data it receives. For example, categories such as “small”, “medium”, and “large” need to be converted into numbers. To solve that, we can for example convert them into numeric labels with “1” for small, “2” for medium, and “3” for large.

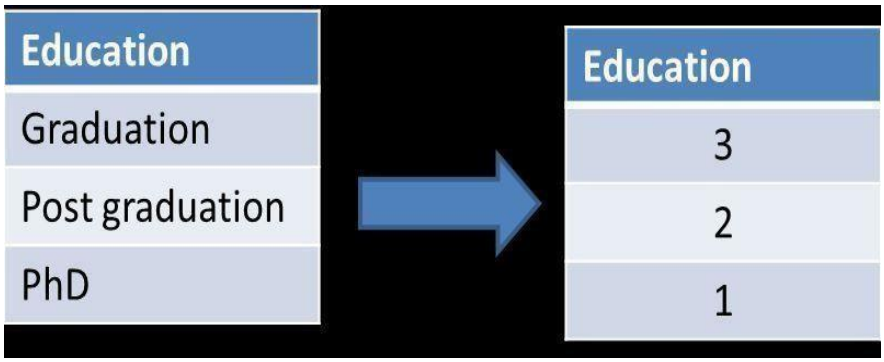


Fig No:6.1.5 Encoding

6.2 Sample Code

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import pickle

import time
```

```
import streamlit as st

from db import *


pickleFile=open("weights.pkl","rb")

regressor=pickle.load(pickleFile)



df = pd.read_csv('./data/mldata.csv')

df.head()


df['workshops'] = df['workshops'].replace(['testing'],'Testing')

df.head()


print(df.columns.unique)


n = df['Suggested Job Role'].unique()

print(len(n))
```

```
print('The shape of our training set: %s professionals and %s features'%(df.shape[0],df.shape[1]))
```

Binary Encoding for Categorical Variables

```
newdf = df
newdf.head(10)
```

```
cols = df[["self-learning capability?", "Extra-courses did","Taken inputs from seniors or elders", "worked in teams ever?", "Introvert"]]
for i in cols:
    print(i)
    cleanup_nums = {i: {"yes": 1, "no": 0}}
    df = df.replace(cleanup_nums)
```

```
print("\n\nList of Categorical features: \n" , df.select_dtypes(include=['object']).columns.tolist())
```

Number Encoding for Categorical

```
mycol = df[["reading and writing skills", "memory capability score"]]

for i in mycol:

    print(i)

    cleanup_nums = {i: {"poor": 0, "medium": 1, "excellent": 2}}

    df = df.replace(cleanup_nums)


category_cols = df[['certifications', 'workshops', 'Interested subjects', 'interested career area ', 'Type of company want to settle in?',

                    'Interested Type of Books']]

for i in category_cols:

    df[i] = df[i].astype('category')

    df[i + "_code"] = df[i].cat.codes


print("\n\nList of Categorical features: \n" , df.select_dtypes(include=['object']).columns.tolist())


## Dummy Variable Encoding


print(df['Management or Technical'].unique())
```

```
print(df['hard/smart worker'].unique())


df = pd.get_dummies(df, columns=["Management or Technical", "hard/smart worker"], prefix=["A", "B"])

df.head()


df.sort_values(by=['certifications'])


print("List of Numerical features: \n" , df.select_dtypes(include=np.number).columns.tolist())



category_cols = df[['certifications', 'workshops', 'Interested subjects', 'interested career area ', 'Type of company want to settle in?', 'Interested Type of Books']]

for i in category_cols:

  print(i)


Certifi = list(df['certifications'].unique())

print(Certifi)
```

```
certi_code = list(df['certifications_code'].unique())

print(certi_code)


Workshops = list(df['workshops'].unique())

print(Workshops)

Workshops_code = list(df['workshops_code'].unique())

print(Workshops_code)


Certi_l = list(df['certifications'].unique())

certi_code = list(df['certifications_code'].unique())

C = dict(zip(Certi_l,certi_code))


Workshops = list(df['workshops'].unique())

print(Workshops)

Workshops_code = list(df['workshops_code'].unique())

print(Workshops_code)

W = dict(zip(Workshops,Workshops_code))


Interested_subjects = list(df['Interested subjects'].unique())
```

```
print(Interested_subjects)

Interested_subjects_code = list(df['Interested subjects_code'].unique())

ISC = dict(zip(Interested_subjects,Interested_subjects_code))


interested_career_area = list(df['interested career area '].unique())

print(interested_career_area)

interested_career_area_code = list(df['interested career area _code'].unique())

ICA = dict(zip(interested_career_area,interested_career_area_code))


Typeofcompany = list(df['Type of company want to settle in?'].unique())

print(Typeofcompany)

Typeofcompany_code = list(df['Type of company want to settle in?_code'].unique())

TOCO = dict(zip(Typeofcompany,Typeofcompany_code))


Interested_Books = list(df['Interested Type of Books'].unique())

print(Interested_subjects)

Interested_Books_code = list(df['Interested Type of Books_code'].unique())

IB = dict(zip(Interested_Books,Interested_Books_code))
```

```
Range_dict = {"poor": 0, "medium": 1, "excellent": 2}

print(Range_dict)



A = 'yes'

B = 'No'

col = [A,B]

for i in col:

if(i=='yes'):

   i  =  1

  print(i)



f =[]

A = 'r programming'

clms = ['r programming',0]

for i in clms:

  for key in C:
```

```
    if(i==key):
        i = C[key]
        f.append(i)
print(f)


C = dict(zip(Certifi,certi_code))


print(C)


import numpy as np
array = np.array([1,2,3,4])
array.reshape(-1,1)


def inputlist(Name,Contact_Number,Email_address,
    Logical_quotient_rating, coding_skills_rating, hackathons,
    public_speaking_points, self_learning_capability,
    Extra_courses_did, Taken_inputs_from_seniors_or_elders,
    worked_in_teams_ever,Introvert, reading_and_writing_skills,
```

memory_capability_score, smart_or_hard_work, Management_or_Techinical,

Interested_subjects, Interested_Type_of_Books,certifications, workshops,

Type_of_company_want_to_settle_in, interested_career_area):

Afeed = [Logical_quotient_rating, coding_skills_rating, hackathons, public_speaking_points]

input_list_col = [self_learning_capability,Extra_courses_did,Taken_inputs_from_seniors_or_elders,worked_in_teams_ever,Introvert,reading_and_writing_skills,memory_capability_score,smart_or_hard_work,Management_or_Techinical,Interested_subjects,Interested_Type_of_Books,certifications,workshops,Type_of_company_want_to_settle_in,interested_career_area]

```
feed = []
K=0
j=0
for i in input_list_col:
  if(i=='Yes'):
    j=2
    feed.append(j)
```

```
    print("feed 1",i)

  elif(i=="No"):
    j=3
    feed.append(j)

    print("feed 2",j)

  elif(i=='Management'):
    j=1
    k=0
    feed.append(j)
    feed.append(K)

    print("feed 10,11",i,j,k)

  elif(i=='Technical'):
    j=0
```

```
    k=1
    feed.append(j)
    feed.append(K)

    print("feed 12,13",i,j,k)

  elif(i=='Smart worker'):
    j=1
    k=0
    feed.append(j)
    feed.append(K)

    print("feed 14,15",i,j,k)

  elif(i=='Hard Worker'):
    j=0
    k=1
    feed.append(j)
    feed.append(K)
```

```
    print("feed 16,17",i,j,k)


  else:
    for key in Range_dict:
      if(i==key):
        j = Range_dict[key]
        feed.append(j)


        print("feed 3",i,j)


    for key in C:
      if(i==key):
        j = C[key]
        feed.append(j)


        print("feed 4",i,j)


    for key in W:
      if(i==key):
```

```
        j = W[key]
        feed.append(j)

        print("feed 5",i,j)

    for key in ISC:
      if(i==key):
        j = ISC[key]
        feed.append(j)

        print("feed 6",i,j)

    for key in ICA:
      if(i==key):
        j = ICA[key]
        feed.append(j)

        print("feed 7",i,j)
```

```
    for key in TOCO:
      if(i==key):
        j = TOCO[key]
        feed.append(j)

        print("feed 8",i,j)

    for key in IB:
      if(i==key):
        j = IB[key]
        feed.append(j)

        print("feed 9",i,j)

  t = Afeed+feed
  output = regressor.predict([t])
```

```
 return(output)


def main():



 html1="""
  <div style="text-align:center;">
    <h1>🎓 Personal Career Recommendation System </h1>
  </div>
    """
 st.markdown(html1,unsafe_allow_html=True)




 col1, col2, col3 = st.columns(3)


 with col1:
    st.image("./images/image1.png")
```

```
with col2:
    st.image("./images/image2.png")

with col3:
    st.image("./images/image3.png")



st.sidebar.title("Your Information")

Name = st.sidebar.text_input("Full Name")

Contact_Number = st.sidebar.text_input("Contact Number")

Email_address = st.sidebar.text_input("Email address")

if not Name and Contact_Number and Email_address:
  st.sidebar.Error("Please fill out your name and EmailID")
```

```
if Name and Contact_Number and Email_address:
  st.sidebar.success("Thanks!")


Logical_quotient_rating = st.slider(
  'Rate your Logical quotient Skills', 0,10)
st.write(Logical_quotient_rating)


coding_skills_rating = st.slider(
'Rate your Coding Skills', 0,10)
st.write(coding_skills_rating)


hackathons = st.slider(
  'Enter number of Hackathons participated',0,10)
st.write(hackathons)


public_speaking_points = st.slider(
'Rate Your Public Speaking', 0,10,1)
st.write(public_speaking_points)
```

```
self_learning_capability = st.selectbox(
  'Self Learning Capability',
  ('Yes', 'No')
  )


Extra_courses_did = st.selectbox(
  'Extra courses',
('Yes', 'No')
)


Taken_inputs_from_seniors_or_elders = st.selectbox(
  'Took advice from seniors or elders',
  ('Yes', 'No')
  )


worked_in_teams_ever = st.selectbox(
```

```
  'Team Co-ordination Skill',
  ('Yes', 'No')
  )


Introvert = st.selectbox(
  'Introvert',
  ('Yes', 'No')
  )

reading_and_writing_skills = st.selectbox(
  'Reading and writing skills',
  ('poor','medium','excellent')
  )
st.write('You selected: *{}*' .format(reading_and_writing_skills))

memory_capability_score = st.selectbox(
  'Memory capability score',
  ('poor','medium','excellent')
```

```
  )
 st.write('You selected: *{}*' .format(memory_capability_score))


 smart_or_hard_work = st.selectbox(
  'Smart or Hard Work',
  ('Smart worker', 'Hard Worker')
  )
 st.write('You selected: *{}*' .format(smart_or_hard_work))


 Management_or_Techinical = st.selectbox(
  'Management or Techinical',
  ('Management', 'Technical')
  )
 st.write('You selected: *{}*' .format(Management_or_Techinical))


 Interested_subjects = st.selectbox(
  'Interested Subjects',
  ('programming', 'Management', 'data engineering', 'networks', 'Software Engineering', 'cloud computing', 'parallel computing', 'IOT', 'Computer Architecture', 'hacking')
```

```
    )
  st.write('You selected: *{}*' .format(Interested_subjects))

  Interested_Type_of_Books = st.selectbox(
    'Interested Books Category',
    ('Series', 'Autobiographies', 'Travel', 'Guide', 'Health', 'Journals', 'Anthology', 'Dictionaries', 'Prayer books', 'Art', 'Encyclopedias', 'Religion-Spirituality', 'Action and Adventure', 'Comics', 'Horror', 'Satire', 'Self help', 'History', 'Cookbooks', 'Math', 'Biographies', 'Drama', 'Diaries', 'Science fiction', 'Poetry', 'Romance', 'Science', 'Trilogy', 'Fantasy', 'Childrens', 'Mystery')
    )
  st.write('You selected: *{}*' .format(Interested_Type_of_Books))

  certifications = st.selectbox(
    'Interested_Type_of_Books',
    ('information security', 'shell programming', 'r programming', 'distro making', 'machine learning', 'full stack', 'hadoop', 'app development', 'python')
    )
  st.write('You selected: *{}*' .format(certifications))
```

```
workshops = st.selectbox(
  'Workshops Attended',
  ('Testing', 'database security', 'game development', 'data science', 'system designing', 'hacking', 'cloud computing', 'web technologies')
  )
st.write('You selected: *{}*' .format(workshops))

Type_of_company_want_to_settle_in = st.selectbox(
  'Type of Company You Want to Settle In ',
  ('BPA', 'Cloud Services', 'product development', 'Testing and Maintainance Services', 'SAaS services', 'Web Services', 'Finance', 'Sales and Marketing', 'Product based', 'Service Based')
  )
st.write('You selected: *{}*' .format(Type_of_company_want_to_settle_in))

interested_career_area = st.selectbox(
  'Interested Career Area',
  ('testing', 'system developer', 'Business process analyst', 'security', 'developer', 'cloud computing')
  )
```

```
st.write('You selected: *{}*' .format(interested_career_area))


result=""


if st.button("Predict"):


result=inputlist(Name,Contact_Number,Email_address,Logical_quotient_rat
ing, coding_skills_rating, hackathons,
                public_speaking_points,
self_learning_capability,Extra_courses_did,
                Taken_inputs_from_seniors_or_elders,worked_in_teams_ever,
Introvert,
                reading_and_writing_skills,memory_capability_score,
smart_or_hard_work,
                Management_or_Techinical,Interested_subjects,
Interested_Type_of_Books,
                certifications,                workshops,
Type_of_company_want_to_settle_in, interested_career_area)


  my_bar = st.progress(0)
```

```
    for percent_complete in range(100):
        time.sleep(0.05)
        my_bar.progress(percent_complete + 1)
st.success("Predicted Career Option : "
              "{}".format(result))



    create_table()
    add_data(Name,Contact_Number,Email_address,Logical_quotient_rating,
coding_skills_rating, hackathons,
          public_speaking_points, self_learning_capability,Extra_courses_did,
          Taken_inputs_from_seniors_or_elders,worked_in_teams_ever,
Introvert,
          reading_and_writing_skills,memory_capability_score,
smart_or_hard_work,
          Management_or_Techinical,Interested_subjects,
Interested_Type_of_Books,
          certifications, workshops, Type_of_company_want_to_settle_in,
interested_career_area)
```

```
 html3="""

   <div style="color:red; margin:80px; text-align:center;">
    Developed by TEAM-76
   </div>
    """

 st.markdown(html3,unsafe_allow_html=True)


if _name=='main_':
      main()
```

CHAPTER 7

SYSTEM 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.

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

7.1 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-boxtest, a tester will interact with the system's user interface by providing inputs and examining outputs without knowing how and where the inputs are workedupon.

7.2 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 knowthe 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.

7.3 Unit Testing:

Unit testing is a software development process in which the smallest testable parts of an application, called units, are individually and independently scrutinized for proper operation. Unit testing is often automated but it can also be done manually. The goal of unit testing is to isolate each part of the program and show that individual parts are correct in terms of requirements and functionality. Test cases and results are shown in the Tables.

Table 7.3.1 Module Tested: User Access

Test Case	1
Name of the Test	Working of Login Credentials
Module being tested	User access
Sample input	Login Credentials
Expected output	Should open home page
Actual output	Home page is accessed
Remarks	Pass

In this test case, we test the login credentials of the user by providing login ID and password. If the access is provided for the credentials, then the test case is passed, if the access is denied, then the test case fails.

Table 7.3.2 Module Tested: Career Detection

Test Case	2
Name of the Test	Detecting CAREER
Module being tested	Test for different results
Sample input	Different rating
Expected output	CAREER detected
Actual output	CAREER detected
Remarks	Pass

In this test case, we are providing different result as an input. If the given ratings is CAREER detected, then the test case is pased.

7.4 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 important because of the following reasons:

- 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.

CHAPTER 8

RESULTS

8.1 Visualization

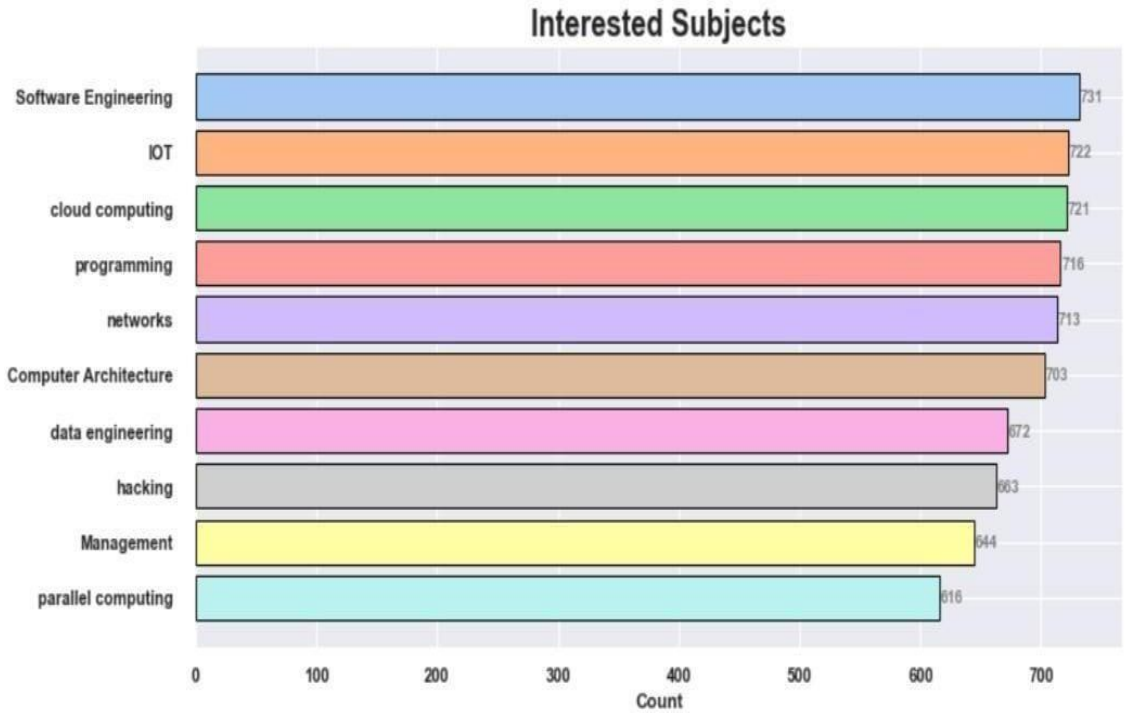


Fig No:8.1.1 Intersted Subjects

Visualising the interested subjects from our dataset to understand and analyse it better.

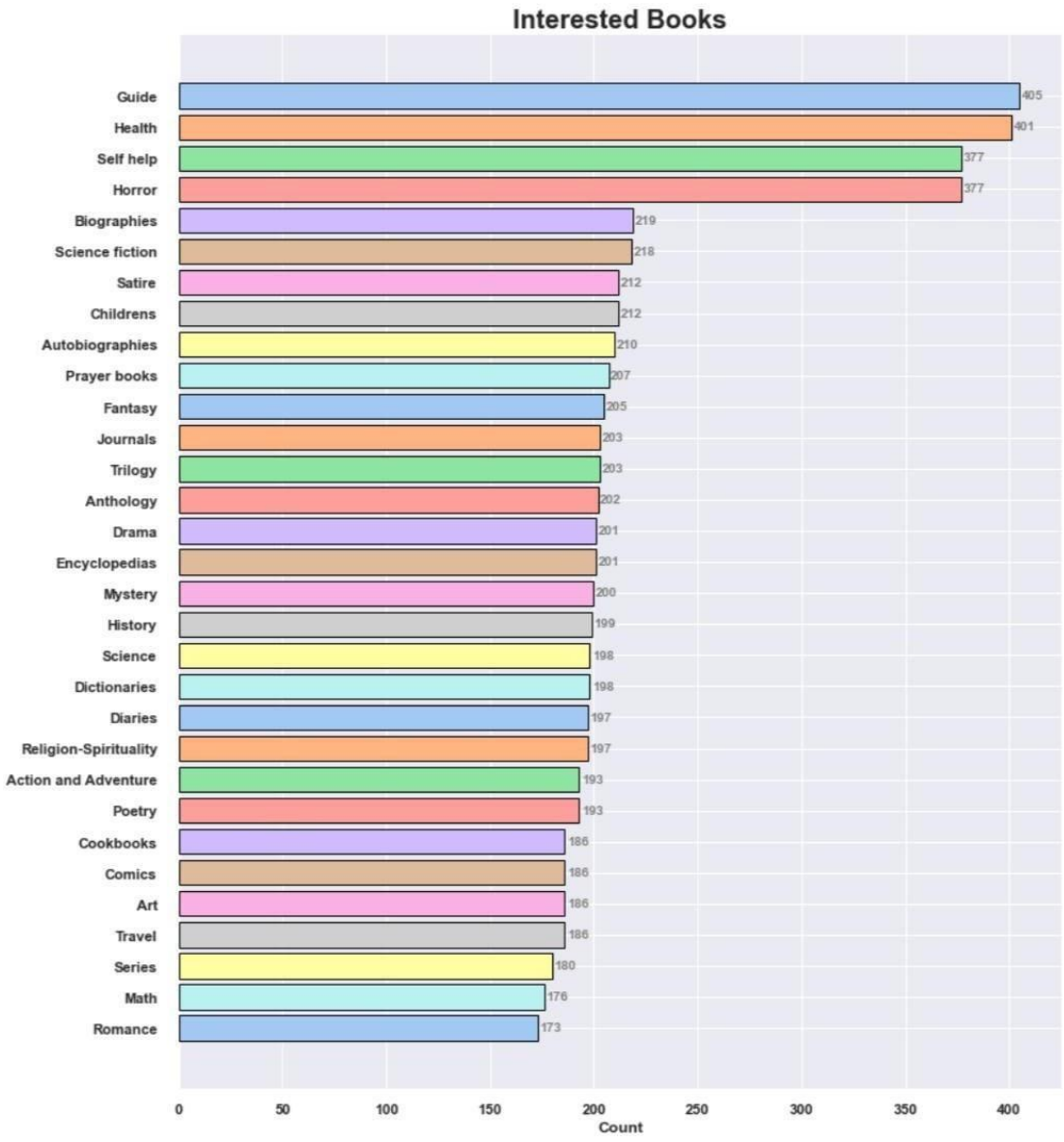


Fig No:8.1.2 Interested Books

Visualising the interested books from our dataset, to understand andanalyse it better.

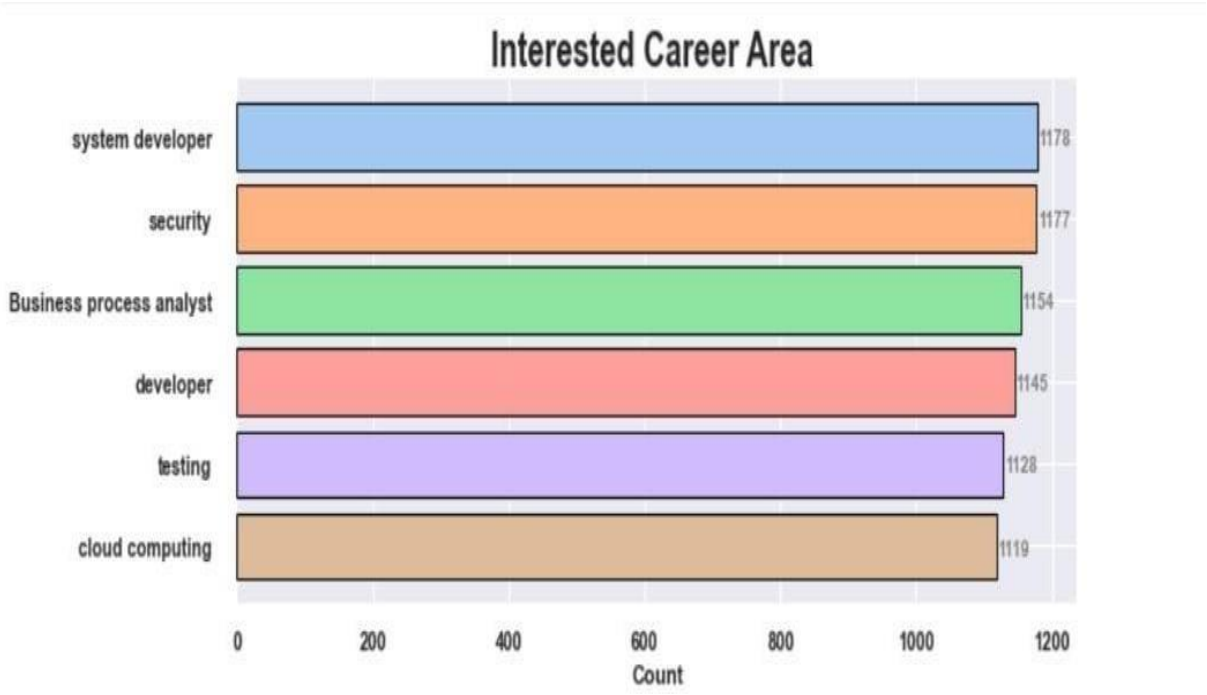


Fig No:8.1.3 Interested Career Area

Visualising the interested career area from our dataset, to understand andanalyse it better.

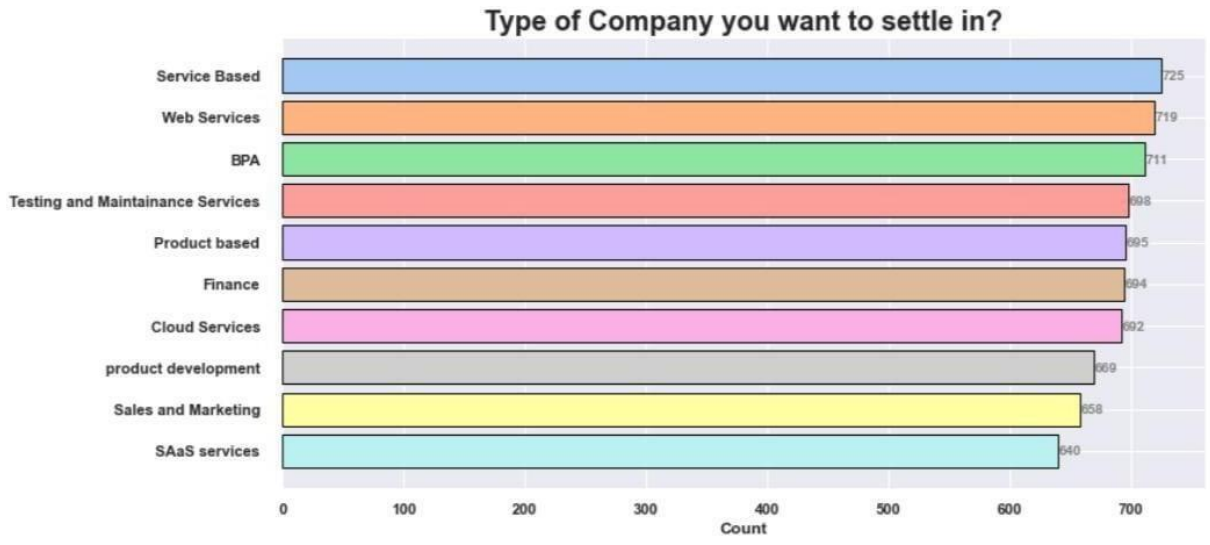


Fig No:8.1.4 Type of Company you want to settle in

Visualising the type of company you want to settle in from our dataset, tounderstand and analyse it better

Fig No:8.1.5 Workshops Attended

Visualising the workshops attended from our dataset, to understand andanalyse it better

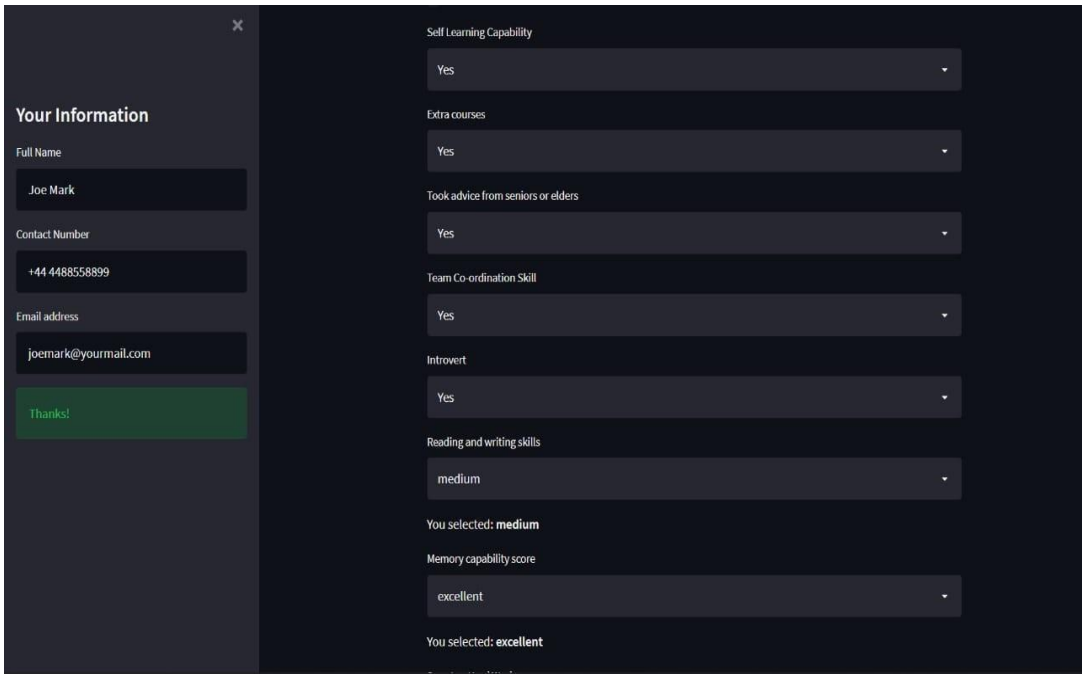
Fig No:8.1.6 Certifications

Visualising the certification in from our dataset, to understand and analyseit better

8.2 Web Page Result

Fig No:8.2 Web Page Result

A web page is created as a front end, so that is would be user friendly. The outcome is as follows.



8.3 Final Result

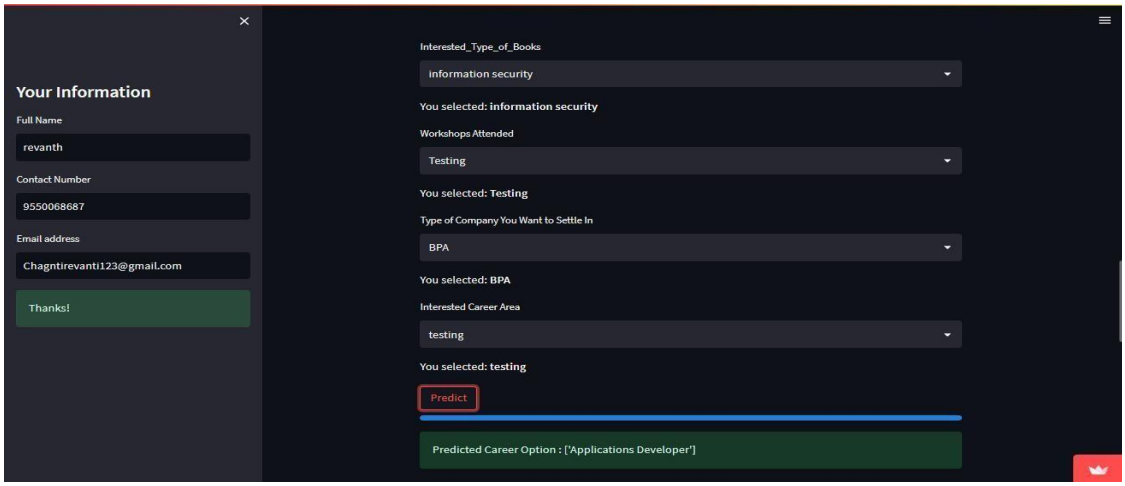


Fig No:8.3 Final Result

At finally Predicted Career Options.

CONCLUSION

In conclusion, Indecisive prospective students are to be guided to select study programs. Most of the study programs are closely related and recommending specific programs are difficult. Further, existing solutions lacking in the use of students' skills, interests, and academicperformance towards filtering a possible recommendation. Hence, the proposed system enables students to choose from the recommended study field. The design process considered the factors such as the academic performance of the student's school results; his/her skills from activities and interests. The main benefit of SCS is the prospect to recommend a field of study rather than an SP.In this way, the prospective students can select a specific SP on their own within a recommended study field. Every day new programs are introduced and continuously increasing. So further decomposition may be needed in order to have more specialized fields of study. In the future, the students' dropouts must be studied to know the reason. Such research helps in designing better recommendation systems for the education domain.

FUTURE WORK

Future research can be conducted by including other parameters as input variables and adding other machine learning algorithms to the modelling process. In addition, it is necessary to harness the effectiveness of DM methods to investigate students' learning behaviors, address their problems, optimize the educational environment, and enable data-driven decision making.

REFERENCES

1. https://towardsdatascience.com/building-a-job-recommender- for-non- technical-business-roles-via-nlp-and-machine- learning-626c4039931e
2. https://www.researchgate.net/publication/358791179_Career_Recommendation_ for

   _College_Students_Based_on_Deep_Learning_and_Machine_Learning
3. https://www.researchgate.net/publication/269272063_Career_path_recomm endatio n_sy stem_for_UiTM_Perlis_students_using_fuzzy_logic
4. https://ieeexplore.ieee.org/document/6869553
5. https://www.semanticscholar.org/paper/Student-Career- Prediction-Using-Decision- Tree-and-VidyaShreeram-Muthukumaravel/28490efaca5b4cd1c215fa79dec03f14f58c17c4
6. https://ieeexplore.ieee.org/document/9137992
7. https://www.ijraset.com/research-paper/career-recommendation-usinann
8. Dr. Mahendra Tiwari ,Manmohan Mishra, "Accuracy Estimation of Classification Algorithmswith DEMP Model", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 11, November 2018.
9. Ms. Roshani Ade,Dr. P. R. Deshmukh, "An incremental ensemble of classifiers as a techniquefor prediction of student's career choice", 2014 First International Conference on Networks &Soft Computing
10. Rutvija Pandya Jayati Pandya , "C5.0 Algorithm to Improved Decision Tree with Feature Selection andReduced Error Pruning", International Journal of Computer Applications (0975 –8887) Volume 117 –No. 16, May 2019.
11. Y. Lu, S. E. Helou, D. Gillet, "Recommender System for Job Seeking and RecruitingWebsite", WWW 2018 Companion, May 13–17, 2018, Rio de Janeiro, Brazil.
12. W. Hong, S. Zheng, H. Wang. "A Job Recommender System Basedon User Clustering".Journal of Computers, vol. 8, no. 8, August 2020

13. Q. Nguyen, T. Huynh, T.N. Hoang, “Adaptive methods for job recommendation basedon user clustering**”,** in Proceedings of the 3rd NationalFoundation for Science and Technology Development Conferenceon Information and Computer Science (NICS), 2018.
14. A. A. Bakar and Choo-Yee Ting, "Soft skills recommendation systems for ITjobs: ABayesian network approach," *Data Mining and Optimization (DMO),2011 3rd Conference on*, Putrajaya, 2019, pp. 82-87
15. A. Gupta and A. Aijaz. “Recommendations@LinkedIn,” presented at theHadoop WorldConference, Germany, 2021.
16. R. Aylward. “Glassdoor Expands Its Integrated Social Platform With Neo4j”Internet,retrieved fromhttp://info.neotechnology.com/rs/neotechnology/images/Glassdoor.pdf, October 2019.
17. P. Melville, R. J. Mooney, R. Nagarajan. “Content-Boosted Collaborative Filtering forImproved Recommendations”, in Proc. AAAI- 02, 2021.
18. C. F. Chien and L. F. Chen, “Data mining to improve personnel selection and enhancehuman capital: A case study in high-technology industry,” Expert Systems with Applications, vol. 34(1), pp. 280-290, 2019.
19. W. M. Al-Adrousy, H. A. Ali, T. T. Hamza. “A Framework for Career-Education Hybrid Recommender System using a Selective Path Delta-SimRank Algorithm”. International Journal of Computer Application(0975-8887), Vol. 90, No. 2, March 2018.
20. N. D. Almalis, G. A. Tsihrintzis, N. Karagiannis and A. D. Strati, "FoDRA — A new content-based job recommendation algorithm for jobseeking and recruiting," *Information,Intelligence, Systems and Applications (IISA), 2015 6th International Conference on*, Corfu, 2020.
21. Y. Zhang, C. Yang, Z. Niu. “A Research of Job Recommendation System Based on Collaborative Filtering”, presented at Seventh International Symposium on ComputationalIntelligence and Design (ISCID), Hangzhou, China, 2016.

APPENDIX-1

Paper Published in Journal of IJARSCT

Publication Details:

"PERSONAL CAREER RECOMMENDATION SYSTEM" in Journal of Emerging Technologies and Innovative Research, March 2023,Volume-3, Issue-2.

Website link- http://ijarsct.co.in/

Fwd: Acceptance Notification for your paper IJARSCT-2023-03-8100
1 message

Lokesh S B <lokeshreddyloki0125@gmail.com> Sat, 1 Apr, 2023 at 4:05 pm
To: deepaknsri@gmail.com

---------- Forwarded message ---------
From: **IJARSCT Journal** <ijarscted@gmail.com>
Date: Mon, 20 Mar, 2023, 11:31 pm
Subject: Acceptance Notification for your paper IJARSCT-2023-03-8100
To: Lokesh S B <lokeshreddyloki0125@gmail.com>

Paper ID:	Paper Title	Author Name
IJARSCT-2023-03-8100	Survey on Personal Career Recommendation System using Machine Learning	Dr. A Syed Mustafa Prof.Anusha K V Deepak N Goutham C Karthika Hiremath Lokesh S B

Dear Author(s),

We are pleased to inform you that your manuscript entitled above has been **ACCEPTED** for publication in International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), Volume 3, Issue 1, and March 2023. You are requested to send payment details to ijarscted@gmail.com. Your prompt response will be appreciated. Thank you again for your submission to IJARSCT. **Kindly send the WORD file in template available at the website if not sent already.**

- **Payment Details: Rs.500 (Five Hundred) (Soft Copy of Certificate) without DOI**
- **Payment Details: Rs. 700 (Seven Hundred) (Soft Copy of Certificate) with DOI**
- **Payment Details: Rs. 999 (Nine Hundred Ninety Nine) (Soft and Hard Copy of Certificate and Paper) with DOI**

NOTE: For more than 5 authors, Rs. 100 extra per author

Note: Each Author will Individual Certificate

Bank Details:
Name of Person: Naksh Solutions (Sapna Agrawal)
Bank Name: Bank of Maharashtra
Account No.: 60358034672

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY

IJARSCT

CERTIFICATE OF PUBLICATION

INTERNATIONAL STANDARD SERIAL NUMBER
ISSN NO: 2581-9429

THIS IS TO CERTIFY THAT

Dr. A Syed Mustafa

HKBK College of Engineering, Bangalore, Karnataka, India

HAS PUBLISHED A RESEARCH PAPER ENTITLED

Personal Career Recommendation System

IN IJARSCT, VOLUME 3, ISSUE 2, MARCH 2023

Certificate No: 032023-A428
www.ijarsct.co.in

Crossref
DOI: 10.48175/IJARSCT-8675
www.doi.org
www.crossref.org

7.301
Scientific Journal Impact Factor
www.sjifactor.com

Editor-in-Chief

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY

IJARSCT

CERTIFICATE OF PUBLICATION

INTERNATIONAL STANDARD SERIAL NUMBER
ISSN NO: 2581-9429

THIS IS TO CERTIFY THAT

Prof. Anusha KV

HKBK College of Engineering, Bangalore, Karnataka, India

HAS PUBLISHED A RESEARCH PAPER ENTITLED

Personal Career Recommendation System

IN IJARSCT, VOLUME 3, ISSUE 2, MARCH 2023

Certificate No: 032023-A429
www.ijarsct.co.in

Crossref
DOI: 10.48175/IJARSCT-8675
www.doi.org
www.crossref.org

7.301
Scientific Journal Impact Factor
www.sjifactor.com

Editor-in-Chief


INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY
IJARSCT
CERTIFICATE OF PUBLICATION
INTERNATIONAL STANDARD SERIAL NUMBER ISSN NO: 2581-9429
THIS IS TO CERTIFY THAT
Deepak N
HKBK College of Engineering, Bangalore, Karnataka, India
HAS PUBLISHED A RESEARCH PAPER ENTITLED
Personal Career Recommendation System
IN IJARSCT, VOLUME 3, ISSUE 2, MARCH 2023
Certificate No: 032023-A430
www.ijarsct.co.in
Crossref
DOI: 10.48175/IJARSCT-8675
www.doi.org
www.crossref.org
7.301
Journal Impact Factor
www.sjifactor.com
Editor-in-Chief


INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY
IJARSCT
CERTIFICATE OF PUBLICATION
INTERNATIONAL STANDARD SERIAL NUMBER ISSN NO: 2581-9429
THIS IS TO CERTIFY THAT
Goutham C
HKBK College of Engineering, Bangalore, Karnataka, India
HAS PUBLISHED A RESEARCH PAPER ENTITLED
Personal Career Recommendation System
IN IJARSCT, VOLUME 3, ISSUE 2, MARCH 2023
Certificate No: 032023-A431
www.ijarsct.co.in
Crossref
DOI: 10.48175/IJARSCT-8675
www.doi.org
www.crossref.org
7.301
Journal Impact Factor
www.sjifactor.com
Editor-in-Chief


INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY
IJARSCT
CERTIFICATE OF PUBLICATION
INTERNATIONAL STANDARD SERIAL NUMBER ISSN NO: 2581-9429
THIS IS TO CERTIFY THAT
Karthika Hiremath
HKBK College of Engineering, Bangalore, Karnataka, India
HAS PUBLISHED A RESEARCH PAPER ENTITLED
Personal Career Recommendation System
IN IJARSCT, VOLUME 3, ISSUE 2, MARCH 2023
Certificate No: 032023-A432
www.ijarsct.co.in
Crossref
DOI: 10.48175/IJARSCT-8675
www.doi.org
www.crossref.org
7.301
Journal Impact Factor
www.sjifactor.com
Editor-in-Chief


INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY
IJARSCT
CERTIFICATE OF PUBLICATION
INTERNATIONAL STANDARD SERIAL NUMBER ISSN NO: 2581-9429
THIS IS TO CERTIFY THAT
Lokesh SB
HKBK College of Engineering, Bangalore, Karnataka, India
HAS PUBLISHED A RESEARCH PAPER ENTITLED
Personal Career Recommendation System
IN IJARSCT, VOLUME 3, ISSUE 2, MARCH 2023
Certificate No: 032023-A433
www.ijarsct.co.in
Crossref
DOI: 10.48175/IJARSCT-8675
www.doi.org
www.crossref.org
7.301
Journal Impact Factor
www.sjifactor.com
Editor-in-Chief

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

Personal Career Recommendation System

Dr. A Syed Mustafa[1], Prof. Anusha KV[2], Deepak N[3], Goutham C[4], Karthika Hiremath[5], Lokesh SB[6]
Professor and Head of Department, Department of Information Science and Engineering [1]
Assistant Professor, Department of Information Science and Engineering[2]
Students, Department of Information Science and Engineering[3,4,5,6]
HKBK College of Engineering, Bangalore, Karnataka, India

Abstract: *As students are going through their academics and pursuing their interested courses, it is very important for them to assess their capabilities and identify their interests so that they will get to know in which career area their interests and capabilities are going to put them. This will help them in improving their performance and motivating their interests so that they will be directed towards their targeted career and get settled in that. Also, recruiters while recruiting candidates after assessing them in all different aspects, these kinds of career recommender systems help them in deciding which job role the candidate should be kept in based on his/her performance and other evaluations. This paper mainly concentrates on the career area prediction of computer science domain candidates.*

Keywords: Carrer, SVM, XGBoost, Machine Learning

I. INTRODUCTION

Nowadays, students are often facing a dilemma in deciding to choose a career in their life. There are several factors that influenced students when choosing their career path as their Personal aptitudes, educational achievement, and their environment. Upon completing their first degree or undergraduate students at the university, students are normally starting to consider a career path that may suit their skill and potential the best. However, many students made the wrong decision in selecting their career due to the lack of experience, help, and advice from friends and relatives, parents and lecturers, or career counseling. It is necessary to deliver career guidance in several ways like courses, training, and seminars that offer group experiences in future career planning and group or individual counseling activities. The factor that may cause students not to be successful in their careers is the wrong choice of a job that suits them. It requires a decision-making process at an early stage. So this system gives a recommendation for students about their careers based on their academic result and their abilities.

II. ALGORITHMS

2.1 Support Vector Machine

Support Vector Machine is an acronym. It is a supervised machine-learning approach that is typically applied to both classification and regression-type issues. Various classification issues are where this is primarily applied. The standard algorithmic step is to plot each data point in an n-dimensional space, where n is the number of features and a feature's value is the value of a specific coordinate. The following step is to classify by obtaining the hyper-plane that sharply divides the two classes. Practically speaking, kernels are used to implement SVM algorithms. There are three different types of SVM, and the linear SVM hyperplane is calculated or discovered by using linear algebra to the problem. The realization is that SVM can be rephrased by using the inner product of observations.

2.2 XG BOOST

Extreme Gradient Boosting is referred to as XGBoost. The implementation of gradient boosting algorithms is called XGBoost. It is accessible in a variety of tools mats, including tools, a library, etc. It primarily concentrates on computational efficiency and model performance. It significantly cuts down on the amount of time and significantly improves the model. Its implementation includes recent additions like regularisation in addition to capabilities seen in sci-kit-learn and R versions. Gradient boosting using both L1 and L2 type regularizations is referred to as regularised gradient boosting. The following are the key benefits that the algorithm's implementation offers: Automatic handling of

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

missing values with sparse aware implementation, block structure to facilitate parallel tree construction, and ongoing training to support further enhancing a model that has already been fitted on the new data.

2.3 Decision Tree

A popular and straight-for e-learning categorization challenge is the use of decision trees. For several sophisticated algorithms, like bagging, gradient boosting, and random forest, decision trees provided the fundamental building blocks. The improved form of the previously stated XG Boost method this broad decision tree CART, C4.5, C5, andID3 are the three most popular decision trees. In the event that the variable is numerical, a node represents a split on the input variable (X). An output variable (y) that is essential for prediction is present in the leaf, also known as the tree's terminal nodes. Selecting a root node is the first step in the usual decision tree process. Before the split, figure out each node's information gain or entropy. All paragraphs must be indented. All paragraphs must be justified, i.e. both left- justified and right-justified.

III. IMPLEMENTATION

3.1 Data Collection

One of the biggest and most crucial jobs of every machine learning project is data collection. as a result the input data is fed to the algorithms. Therefore, the accuracy and efficiency of the algorithms depend on how well the data is collected and how accurate it is. The result will be the same as the data. Numerous factors, including academic performance in multiple disciplines, specialties, programming, analytical skills, memory, relationship status, interests in sports, contests, hackathons, workshops, and books, among others, are needed to predict a student's career. All of these characteristics are taken into account since they are crucial in determining how far a student will advance in a certain vocational field. Many methods exist for gathering data. Some information is gathered from employees of various organizations, LinkedIn, and Colleges

3.2 Data Pre-Processing

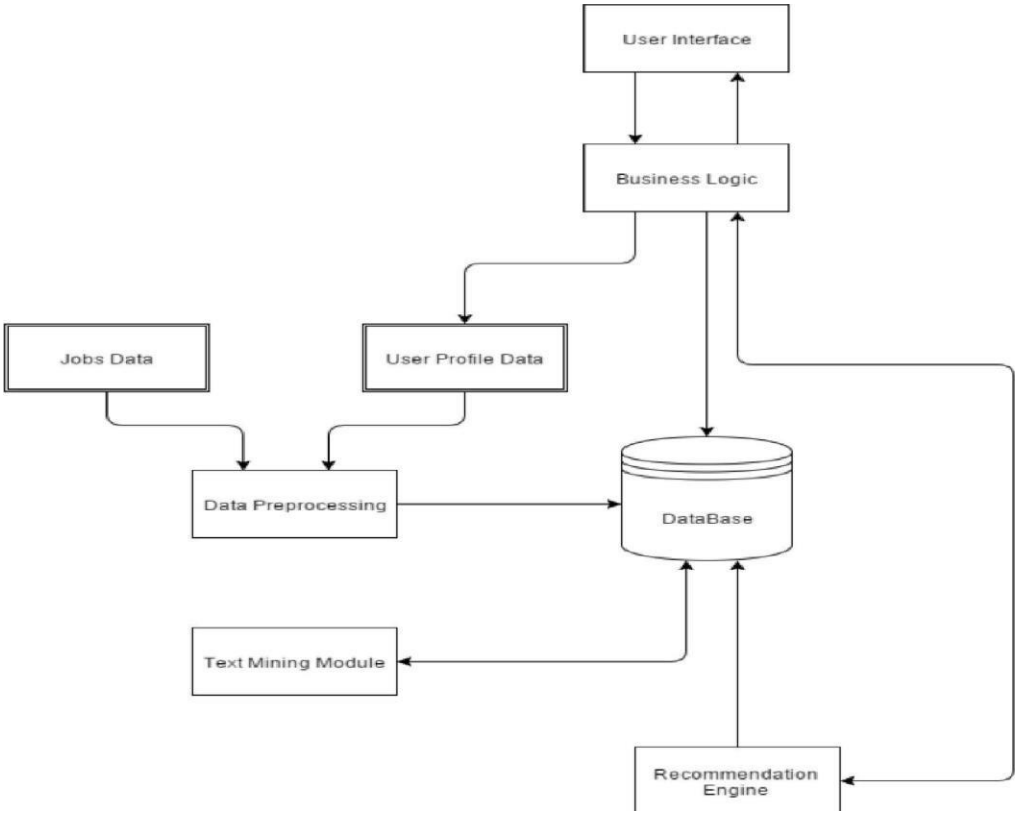
Making the data meaningful is a crucial effort that goes hand in hand with gathering the data. There may be many null values, invalid data values, and undesirable data in the data obtained through various methods because it will be in a disorganized manner. removing all of this data and substituting them with more accurate or suitable data. The fundamental processes in preprocessing data include identifying null and missing data, eliminating it, and replacing it with some predefined alternate values. Even acquired data could have entirely useless values. It might not be in the precise shape or manner in which it was intended. To make the meaning of the data intelligible and helpful for future processing, all such cases must be validated and replaced with alternative values. The storage of data must beorganized.

3.3 One Hot Encoding

OneHot Encoding is a method for providing categorical values found in the collected data to machines by converting them to numerical or other ordinal formats. Improving prediction outcomes via learning algorithms. Categorical values are transformed using the OneHot encoding technique into a format that is most suitable for feeding into different machine learning algorithms. Nearly any machine learning algorithm is compatible with this algorithm. Only the a few algorithms, like random forest, effectively handle categorical values. In these circumstances, one hot encoding is not necessary. Although the OneHot encoding process may appear challenging, most contemporary machine learning algorithms handle that. This article clearly explains the procedure: In a data set, for example, if there are yes and no values, integers respectively 1 and 0.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

IV. ARCHITECTURE DIAGRAM



V. LITERATURE SURVEY

5.1 Personal Factor

Personal factors including student behavior, such as feelings, activities, as well as the finances of the pupils, are important components of understudy academic performance. age and sexual orientation, which impact how they perform. The most often employed parameters for prediction are age and gender since they are considered internal causes of variability and are simple to define and assess. The analysts looked at how frequently psychometric factors might affect how well students performed.

5.2 Academic Factors

Academic factors are descriptors used to describe variables that explain a student's performance on the academic track at a university. Given its enormous influence on the future of training, the cumulative grade point average (CGPA) is the primary factor that has been used most frequently student author took into account the student GPA. The most significant factors in forecasting a student's future academic performance, according to the authors, are their past academic performance and their parents' educational backgrounds. Others investigated how past academic success affected how well pupils will do going forward.

5.3 Financial Factors

Financial aspects Related elements imply that the parents' financial capacity to support their children's education and guide their future careers. A few specialists looked at the relationship between educational programming, parental instruction, and salary.

5.4 Family Factors

Family variables have an impact on an individual's educational background, ability to help their children with their education, and ability to create an environment that is conducive to learning. Results showed that regardless of whether students' school has an impact on students' performance, parents' efforts have a crucial role in predicting grades. The

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

most important factors in predicting an understudy's future scholastic presentation in the art hour were determined to be their former academic performance and their parents' teaching background. Studies often examine how parents have an impact. education background, income, and the impact of students' interpretation on academic success.

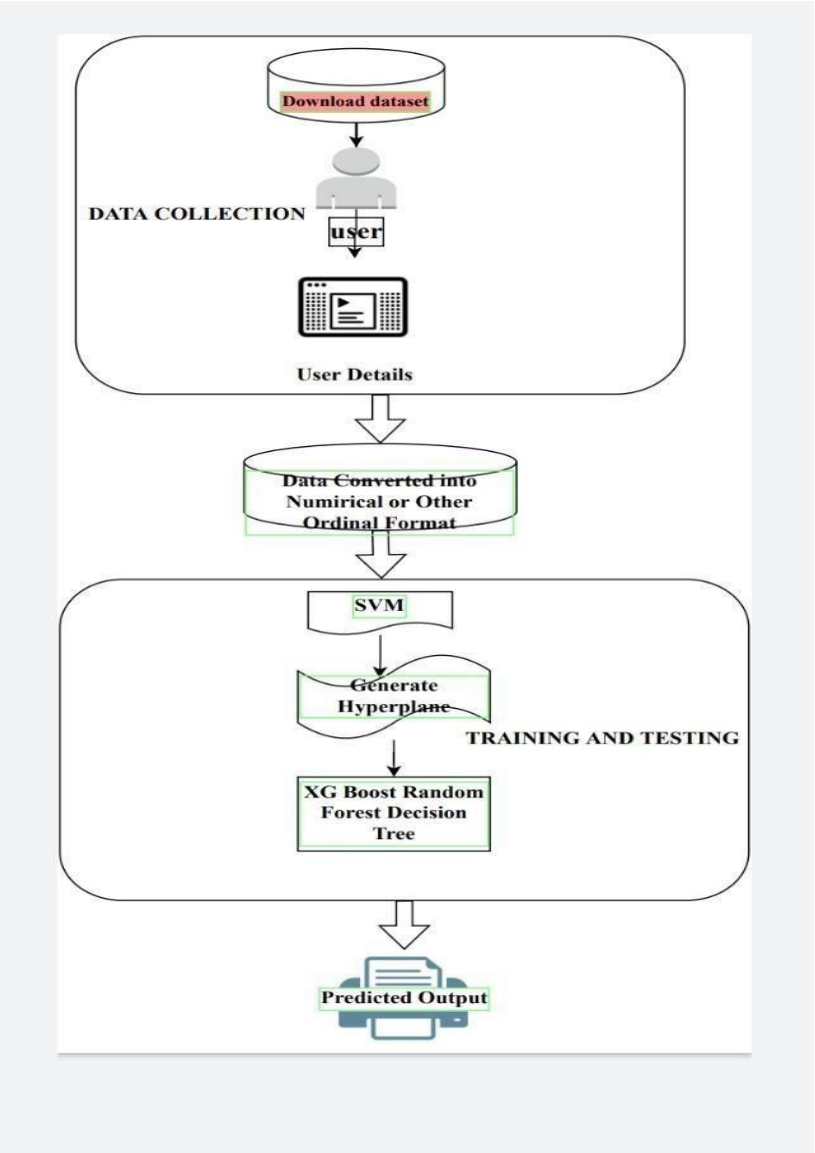
5.5 Institution Factor

The factors that affect this classification have to do with the academic program and resources that the foundation provides for the highest academic performance of its trainees. The authors looked at how psychometric factorsfrequently affect how candidates present themselves.

JOURNAL NAME & YEAR	AUTHOR	TITLE& METHOD	REMARKS
Hindawi.2022.	Qingwan, Linye	Career Recommendations for students based on Deep Learning and Machine Learning. A hybrid CNN model is proposed for employment recommendation of college students.	By improving the Activation function, pooling strategy, and loss function in the algorithm, the quality of model prediction is greatly improved.
International Journal of Advanced Research in Computer and Communication Engineering. 2021.	Sushma Koushik N, Chandana M S Lavanya V, Suhas Y, HarshithaV	Educational Career Recommendation System Using Machine Learning. The problems of cold start, trust, and privacy are solved in this approach	The recommendation system is built in Python because it is easy and efficient to put into effect algorithms on exclusive operating systems.
Multidisciplinary Digital Publishing Institute. 2020.	MinNie, Zhaohui Xiong Ruiyang Zhong, Wei Deng and Guowu Yang	Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students Proposeda Prototypical cluster center generation approach to use the priori information from each college	A noval regularization Item was introduced by To bridge the gap between the real-world examples and prototypical cluster centers.
Research gate.2017	Bharat Patel, Varun Kakuste, Magdalini Eirinaki	A career path Recommendation on framework A novel career path framework for personalized job and skills Recommend youngs, focusing on students and young professionals is presented.	The system can run on cloud infrastructure using apache spark or Hadoop cluster which can handle more number of jobs description and more users. To increase the system capability at any time, new machines can be added to the existing cluster without affecting the running application.
International Journal of Health Science Journal of Health Sciences. 2022.	Dr. Sadasivam R, Paramasaivam S, Prakash raj N, Saravanam M	Student carrer prediction Student suggestion was proposed to help student career prediction in concluding there abilities in which they are solid and frail.	RF contains various choice threes different subsets of the given datasets and taken the normal to work on the precision of that datasets. It tends to be utilized for both Arrangements and Regression issue in ML

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

VI. METHODOLOGY



The system was created using Python. The system produces after receiving input from the datasets. As a result The steps that make up the system-building process are done in order.

Obtaining the dataset

Cleaning the Dataset

Selecting for feature dataset

Build a model

Predict outcomes using the model

VII. RESULT

All three algorithms are tested and trained on the data, and SVM topped all others in terms of accuracy, scoring 90.3 percent, followed by XG Boost at 88.33 percent. As SVM All subsequent data predictions are picked to be followed by SVM because it provided the highest accuracy. Thus, a web application is created to provide the student's input parameters and the end A prediction is produced and shown. SVM is indeed the background method in use, and fresh predictions are continuously added to the dataset to increase accuracy.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

VIII. CONCLUSION

We looked at how college students choose their careers based on their professional expertise, behavior consistency, and other associated behaviors The study has also provided a number of crucial insights for enhancing the model.

We have suggested a prototype approach for creating cluster centers that takes advantage of the past knowledge from each college. In order to bridge the gap between real-world instances and a prototypical number of clusters, we introduced a unique normalization item, which is motivated by the cluster assumption that samples in the same cluster should have the same label. Multiple experiments' findings show that our method is superior to other methods for predicting professional choice.

REFERENCES

[1]. https://towardsdatascience.com/building-a-job-recommender-for-non-technical- business-roles-via-nlp-and-machine-learning-626c4039931e

[2]. https://www.researchgate.net/publication/358791179_Career_Recommendation_for_College_Students_Based_on_Deep_Learning_and_Machine_Learning

[3]. https://www.researchgate.net/publication/269272063_Career_path_recommendation_system_for_UiTM_Perlis_students_using_fuzzy_logic

[4]. https://ieeexplore.ieee.org/document/6869553

[5]. https://www.semanticscholar.org/paper/Student-Career-Prediction-Using-Decision-Tree-and-VidyaShreeram-Muthukumaravel/28490efaca5b4cd1c215fa79dec03f14f58c17c4

[6]. https://ieeexplore.ieee.org/document/9137992 https://www.ijraset.com/research-paper/career-recommendation-using-ann

[7]. P.KaviPriya, "A Review on Predicting Students' Academic Performance Earlier, Using Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering.

[8]. Ali Daud, Naif Radi Aljohani, "Predicting Student Performance using Advanced Learning Analytics", 2017 International World Wide Web Conference Committee (IW3C2).

[9]. Nikita Gorad ,Ishani Zalte, "Career Counselling Using Data Mining", International Journal of Innovative Research in Computer and Communication Engineering.

APPENDIX-2

CONFERENCE

Publication Details:

We attended the 8th National Conference on Innovative Computing Techniques in "**DR.MGR Educational and Research Institute "**

NC23CS043

Dr.M.G.R.
Educational and Research Institute
(DEEMED TO BE UNIVERSITY)
(An ISO Certified Institution)
University with Graded Autonomy Status
Maduravoyal , Chennai - 600 095

ABET

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

8th National Conference
on
Innovative Computing Techniques

NCICT'23 - REBOrN
(Robotics, Edge-Computing, Bioinformatics, Ongoing research, Nano Technology)

Certificate of Participation

This certificate is presented to ..MR DEEPAK N,UG STUDENT,.....

of .....HKBK COLLEGE OF ENGINEERING, BANGALORE, INDIA.....for

presenting the paper titled .SURVEY ON PERSONAL CAREER RECOMMENDATION SYSTEM USING MACHINE LEARNING.....

on ..4th... April 2023.

Participant ID: NC23CS043

HoD-CSE

Registrar

NC23CS043

Dr.M.G.R.
Educational and Research Institute
(DEEMED TO BE UNIVERSITY)
(An ISO Certified Institution)
University with Graded Autonomy Status
Maduravoyal , Chennai - 600 095

ABET

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

8th National Conference
on
Innovative Computing Techniques

NCICT'23 - REBOrN
(Robotics, Edge-Computing, Bioinformatics, Ongoing research, Nano Technology)

Certificate of Participation

This certificate is presented to ..MR .GOUTHAM C,. UG STUDENT,.....

of .....HKBK COLLEGE OF ENGINEERING, BANGALORE, INDIA.....for

presenting the paper titled .SURVEY ON PERSONAL CAREER RECOMMENDATION SYSTEM USING MACHINE LEARNING.....

on ..4th... April 2023.

Participant ID: NC23CS043

HoD-CSE

Registrar

NC23CS043

Dr.M.G.R.
Educational and Research Institute
(DEEMED TO BE UNIVERSITY)
(An ISO Certified Institution)
University with Graded Autonomy Status
Maduravoyal , Chennai - 600 095

ABET

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

8th National Conference
on
Innovative Computing Techniques

NCICT'23 - REBOrN
(Robotics, Edge-Computing, Bioinformatics, Ongoing research, Nano Technology)

Certificate of Participation

This certificate is presented to MR.KARTHIKA HIREMATH, UG STUDENT,

of HKBK COLLEGE OF ENGINEERING, BANGALORE, INDIA for

presenting the paper titled SURVEY ON PERSONAL CAREER RECOMMENDATION SYSTEM USING MACHINE LEARNING

on 4th April 2023.

Participant ID: NC23CS043

HoD-CSE

Registrar

NC23CS043

Dr.M.G.R.
Educational and Research Institute
(DEEMED TO BE UNIVERSITY)
(An ISO Certified Institution)
University with Graded Autonomy Status
Maduravoyal , Chennai - 600 095

ABET

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

8th National Conference
on
Innovative Computing Techniques

NCICT'23 - REBOrN
(Robotics, Edge-Computing, Bioinformatics, Ongoing research, Nano Technology)

Certificate of Participation

This certificate is presented to MR. LOKESH S B, UG STUDENT,

of HKBK COLLEGE OF ENGINEERING, BANGALORE, INDIA for

presenting the paper titled SURVEY ON PERSONAL CAREER RECOMMENDATION SYSTEM USING MACHINE LEARNING

on 4th April 2023.

Participant ID: NC23CS043

HoD-CSE

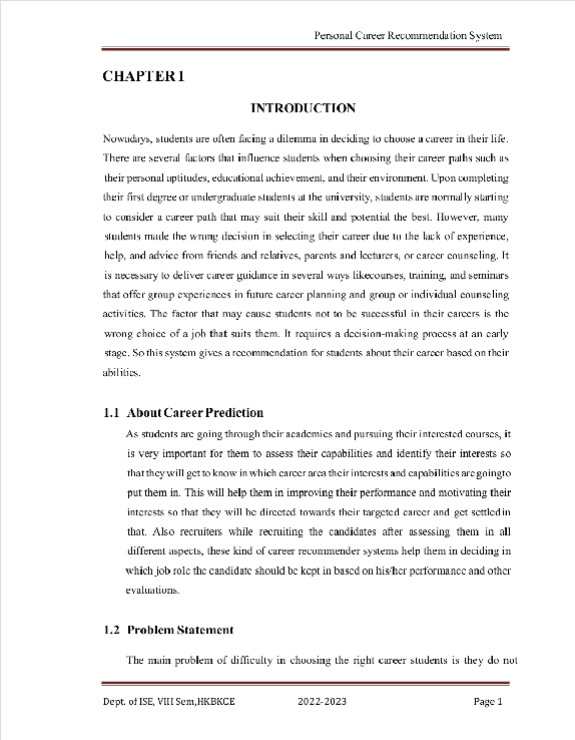
Registrar

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Turnitin Official
Assignment title:	Plagiarism Freedom -8310977655
Submission title:	Personal Career Recommendation System
File name:	Personal_Career_Recommendation_System.pdf
File size:	1.6M
Page count:	58
Word count:	7,528
Character count:	46,749
Submission date:	03-May-2023 07:24AM (UTC-0500)
Submission ID:	2083004053



Personal Career Recommendation System

CHAPTER 1

INTRODUCTION

Nowadays, students are often facing a dilemma in deciding to choose a career in their life. There are several factors that influence students when choosing their career paths such as their personal aptitudes, educational achievement, and their environment. Upon completing their first degree or undergraduate students at the university, students are normally starting to consider a career path that may suit their skill and potential the best. However, many students made the wrong decision in selecting their career due to the lack of experience, help, and advice from friends and relatives, parents and lecturers, or career counseling. It is necessary to deliver career guidance in several ways likecourses, training, and seminars that offer group experiences in future career planning and group or individual counseling activities. The factor that may cause students not to be successful in their careers is the wrong choice of a job that suits them. It requires a decision-making process at an early stage. So this system gives a recommendation for students about their career based on their abilities.

1.1 About Career Prediction

As students are going through their academics and pursuing their interested courses, it is very important for them to assess their capabilities and identify their interests so that they will get to know in which career area their interests and capabilities are goingto put them in. This will help them in improving their performance and motivating their interests so that they will be directed towards their targeted career and get settled in that. Also recruiters while recruiting the candidates after assessing them in all different aspects, these kind of career recommender systems help them in deciding in which job role the candidate should be kept in based on his/her performance and other evaluations.

1.2 Problem Statement

The main problem of difficulty in choosing the right career students is they do not

Dept. of ISE, VIII Sem,HKBKCE 2022-2023 Page 1

Personal Career Recommendation System

by Turnitin Official

Submission date: 03-May-2023 07:24AM (UTC-0500)
Submission ID: 2083004053
File name: Personal_Career_Recommendation_System.pdf (1.6M)
Word count: 7528
Character count: 46749

Personal Career Recommendation System

ORIGINALITY REPORT

6%	1%	1%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ijariit.com Internet Source	1%
2	Submitted to Harare Institute of Technology Student Paper	1%
3	Submitted to Higher Education Commission Pakistan Student Paper	1%
4	Submitted to Asian Institute of Technology Student Paper	1%
5	Submitted to Liverpool John Moores University Student Paper	<1%
6	Submitted to Kampala International University Student Paper	<1%
7	Submitted to Sogang University Student Paper	<1%
8	Submitted to The NorthCap University, Gurugram Student Paper	<1%

9	Submitted to Yakın Doğu Üniversitesi Student Paper	<1%
10	Submitted to Ghana Technology University College Student Paper	<1%
11	Submitted to Kingston University Student Paper	<1%
12	Submitted to Midlands State University Student Paper	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		