A Mini Project Report

On

EduTrack – Intelligent Career Guidance System

Submitted by

TY A-14 Megha Gaikwad

TY A-17 Suraksha Harwalkar

In partial fulfillment for the award of the degree of

Bachelor of Technology

IN

Computer Science Engineering

Pradnya Niketan Education Society, Pune.

NAGESH KARAJAGI *ORCHID* COLLEGE OF ENGGINEERING & TECHNOLOGY SOLAPUR.

2023-2024

Pradnya Niketan Education Society , Pune.
NAGESH KARAJAGI *ORCHID* COLLEGE OF ENGG. & TECH.,
SOLAPUR.

Gut No. 16, Solapur-Tuljapur Road, Tale Hipparaga, Solapur – 413 002
Phone: (0217) 2735001/02, Fax. (0217) 2735004

Certificate

This is to certify that

Ms. A-14 Megha Gaikwad

Ms. A-17 Suraksha Harwalkar

of class **TE-A** has satisfactorily completed the **Mini Project** work entitled **EduTrack**

– Intelligent Career Guidance System as prescribed by Dr.Babasaheb Ambedkar

Technological University Lonere, Maharashtra, India in the academic year 2023-24.

Date of Submission:

(Prof. P.D.Jawalkar)   (Dr. V. V. Bag)
Seminar Guide   **Head of Department**

Examiners: (Name with Signature & Date)

1 : _____

2 : _____

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 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. This report mainly concentrates on the career area prediction of computer science domain candidates.

INDEX

ACKNOWLEDGEMENT

I would like to express my gratitude and extend my sincere appreciation to everyone who contributed to the successful completion of this mini project report.

Firstly, I extend my heartfelt thanks to my mini project guide, Prof. P.D Jawalkar mam, for their unwavering support, guidance, and encouragement throughout this journey. Their expertise, patience, and guiding feedback have been important in shaping this report and increase its quality. I am deeply thankful for their mentorship and for y sharing their knowledge and expertise on the subject.

I am also very grateful to NKOCET-Solapur for granting me the opportunity to undertake this mini project and for providing the necessary resources and facilities. Access to the library, research materials, and databases has been very much beneficial for conducting thorough research and collecting information for this report.

My thankfulness also goes to my friends for their constant support and invaluable discussions. Their constructive feedback, brainstorming sessions, and exchange of ideas have greatly contributed to enhancing my understanding of the topic and guiding the direction of this report.

In conclusion, I would like to express appreciation to everyone mentioned above and to all who have directly or indirectly supported this mini project report. Your support, guidance, and encouragement have been crucial, and I am honored and privileged to have had the chance to learn and improve from this experience.

Thank you all.

A-14 Megha S. Gaikwad
A-17 Suraksha A. Harwalkar

CHAPTER ONE
INTRODUCTION

Competition in today's society is heavily multiplying day by day, especially in the present day's technical world. To compete and reach their goals, students need to be planned and organized from the initial stages of their education. It is very important to constantly evaluate their performance, identify their interests, and assess how close they are to their goals. This helps determine whether they are on the right path toward their targets. Continuous evaluation assists students in improving themselves and motivating them towards a better career path if their capabilities are not up to the mark for reaching their goals. Pre-evaluation is crucial before they reach their career peak point.

Recruiters, while hiring candidates, evaluate them on different parameters to make final decisions about their selection. If selected, they find the best-suited role and career area for the individual. There are many types of roles like Database Administrator, Business Process Analyst, Developer, Testing Manager, Networks Manager, Data Scientist, and so on. Each role requires specific prerequisite knowledge. Recruiters analyze candidates' skills, talents, and interests to place them in the right job role. Prediction systems facilitate these recruitment tasks by providing recommendations based on inputs. Various career recommendation systems and job role prediction systems are already used by third-party performance evaluation portals like Co-Cubes and AMCAT. These portals primarily consider technical abilities and psychometry of students to suggest suitable job roles based on their performance.

However, a more comprehensive approach considers additional factors such as abilities in sports, academics, hobbies, interests, competitions, skills, and knowledge. By including these factors, the total number of parameters considered as inputs is 36, with the final job roles fixed to 15. Given the large number of input parameters and output classes, typical programming and normal algorithms may not provide the best possible classification and prediction. Therefore, advanced machine learning algorithms like SVM, K-Nearest Neighbour, OneHot Encoding, are used for more accurate and effective outcomes.

CHAPTER TWO
LITERATURE REVIEW

[1] Personalized Career Path Recommender System for Engineering Students

Opting for a university specialization is a grueling decision for all academy scholars. Due to the lack of guidance and limited online resources, scholars' opinions often depend on the private comprehension of family and friends. This increases the threat of high university dropout rates, students changing their university disciplines, and choosing an irrelevant career path. To address these downsides, this exploration paper presents a Personalized Career-path Recommender System (PCRS) to provide guidance and help academy scholars choose their respective disciplines. The main idea of PCRS is to mimic the role of professional counselors who assist scholars in making this difficult decision by assessing their academic and personal interests. The design of PCRS is based on a fuzzy-sense intelligence with two main input parameters: academic performance and specific profile.

[2] Prediction of Student's Academic Performance Using Machine Learning Algorithms

Learning analytics and supportive learning are emerging research areas in today's era of big data, data mining, machine learning, and artificial intelligence to facilitate students' learning. Student education is crucial to the sustainable development of society as students acquire knowledge from schools and through extracurricular activities, developing abilities to contribute to the community. Many students progress to higher levels of education and earn degrees like Ph.D., while many graduate each year. However, some students marginally pass or fail their courses and are required to retake them. This paper proposes an improvised conditional network-based deep support vector machine (ICGAN-DSVM) algorithm. ICGAN addresses the issue of low data volume by mimicking new training datasets, whereas DSVM extends SVM from shallow learning to deep learning. DSVM leverages small datasets, making it more efficient in comparison to traditional deep neural networks.

[3] A Machine Learning-Based Approach for Recommending Courses at the Graduate Level

Students often face confusion during their academic careers regarding course selection and future scope, necessitating proper guidance, which is not always accessible. This paper proposes a system for recommending courses to students according to their benefits and

interests, focusing on higher or graduate-level study. It utilizes data mining and machine learning techniques to provide accurate recommendations. Factors considered include student performance, preferred interests, and skills. Machine learning methods like neural networks and various learning algorithms help in guiding students toward the best courses for their careers.

[4] Analysis Optimization of K-Nearest Neighbor Algorithm with Certainty Factor in Determining Student Career . A career is a series of developments or progressions in respective professions experienced by every human being. These careers manifest as job titles and work, related to professional and academic growth. The crucial element of a career is the development or progress at every life stage. K-Nearest Neighbor (KNN) is a method for data classification and organization, while the certainty factor is an uncertain decision-making method. This study uses datasets such as students' interests, talents, and exam scores to predict career-appropriate decisions. The student career prediction system combines KNN and the certainty factor. It is expected that this combined analysis provides better guidance for students in determining their careers with accurate results. The KNN-certainty factor method aims to predict student careers with high accuracy.

CHAPTER THREE
METHODOLOGY

3.1 Data Collection:

Collection of data is one of the major and most important tasks of any machine learning project. The input we feed to the algorithms is data, so the efficiency and accuracy of the algorithms depend upon the correctness and quality of the data collected. The output depends on the data quality. For student career prediction, many parameters are required, such as students' academic scores in various subjects, specializations, programming and analytical capabilities, memory, personal details like Database fundamentals, Computer Architecture, Networking, competitions, hackathons, workshops, certifications, books of interest, and many more. As all these factors play a vital role in deciding a student's progress towards a career area, all these are taken into consideration. Data is collected in many ways. Some data is collected from employees working in different organizations, some amount of data is collected through LinkedIn API, some data is randomly generated, and other data is from the college alumni database. In total, nearly 20 thousand records with 36 columns of data are collected.

3.2 Data Pre-processing:

Collecting the data is one task, and making that data useful is another vital task. Data collected from various means will be in an unorganized format and may contain many null values, invalid data values, and unwanted data. Cleaning all this data and replacing it with appropriate or approximate data, removing null and missing data, and replacing them with some fixed alternate values are the basic steps in pre-processing data. Even data collected may contain completely garbage values. It may not be in the exact format or way that it is meant to be. All such cases must be verified and replaced with alternate values to make data meaningful and useful for further processing. Data must be kept in an organized format.

3.3 OneHot Encoding:

OneHot Encoding is a technique by which categorical values present in the data collected are converted into numerical or other ordinal formats so that they can be provided to machine

learning algorithms and yield better prediction results. Simply, OneHot encoding transforms categorical values into a form that best fits as input to various machine learning algorithms. This algorithm works fine with almost all machine learning algorithms. A few algorithms, like random forest, handle categorical values very well. In such cases, OneHot encoding is not required.

The process of OneHot encoding may seem difficult, but most modern-day machine learning algorithms take care of that. The process is easily explained here: For example, in a data set, if there are values like "yes" and "no," an integer encoder assigns values to them like 1 and 0. This process can be followed as long as we continue using fixed values for "yes" as 1 and "no" as 0. As long as we assign or allocate these fixed numbers to these particular labels, this is called integer encoding. But consistency is very important because if we invert the encoding later, we should get back the labels correctly from those integer values, especially in the case of prediction. The next step is creating a vector for each integer value. Let us suppose this vector is binary and has a length of 2 for the two possible integer values. The "yes" label encoded as 1 will then be represented with the vector [1, 1], where the zeroth index is given the value 1. Similarly, the "no" label encoded as 0 will be represented as [0, 0], which indicates that the first index is represented with the value 0.

For example, [pillow, rat, fight, rat] becomes [0, 1, 2, 1]. This imparts an ordinal property to the variable, i.e., pillow < rat < fight. This ordinal characteristic is usually not required and desired, so OneHot encoding is required for the correct representation of distinct elements of a variable. It makes the representation of categorical variables more expressive.

CHAPTER FOUR
LANGUAGE USED

1. EduTrack The Career Guidance System is a Recommendation System, the output of which is given on a web portal. The technical aspects of this project are HTML, CSS, BOOTSTRAP, Python. In this project, we build a Web Portal. The student arrives at the Web Portal, wherein he/she will have to attempt MCQ based questionnaires each testing his/her IQ i.e. the Aptitude, The portal will also collect information about the interest of the user they are subject to the person's traits and differ from one person to the other. After submitting all Questionnaires, the answers will be stored in a database, from where they will be used as the test case against an existing dataset so as to recommend career choices to the student.

On the Machine Learning front, firstly we have collected data for our dataset by circulating questions, gathered from professionals, amongst people from different branches, streams and careers. We then built a Recommendation Engine after trying and testing our Machine Learning model with different algorithms, and selected the algorithm which provides best accuracy and results. To compensate for the less amount of data collected, we will be combining 2 methods to get the desired optimum result. We shall test our sample data entry from the student against our training dataset, and display the top career choices on the Web Portal. We have used the following to build the portal: HTML (Hypertext Mark-up Language) part to structure a web page and its content. CSS stands for Cascading Style Sheets. It describes how HTML elements should be displayed

2.Python : Python as a language, and give introduction about Data science, Machine learning and then describe packages that are popular in the Data science and Machine learning sectors, such as NumPy, SciPy, TensorFlow, Keras ,Matplotlib etc.

Essential Libraries and Tools :

Understanding what scikit-learn is and how to use it is important, but there are a few other libraries that will enhance your experience. scikit-learn is built on top of the NumPy and SciPy scientific Python libraries. In addition to NumPy and SciPy, we will be using pandas and matplotlib. In scikit-learn, the NumPy array is the fundamental data structure. scikit-learn takes in data in the form of NumPy arrays. Any data you're using will have to be con- verted to a

NumPy array. The core functionality of NumPy is the ndarray class, a multidimensional (n-dimensional) array**.**

import pandas as pd :

import pandas as pd. Simply imports the library that current namespace, but rather than using the name pandas , it's instructed to use the name pd instead. This is just so you can do pd. whatever instead of having to type out pandas. whatever all the time if you just do import pandas.

import numpy as np

 NumPy is an open-source numerical Python library. NumPy contains a multi-dimensional array and matrix data structures. It can be utilized to perform a number of mathematical operations on arrays such as trigonometric, statistical, and algebraic routines. NumPy is an extension of Numeric and Numarray.

import Random

import random imports the random module, which contains a variety of things to do with random number generation. Among these is the random() function, which generates random numbers between 0 and 1.

import matplotlib.pyplot as plt

Pyplot is a collection of functions in the popular visualization package Matplotlib. Its functions manipulate elements of a figure, such as creating a figure, creating a plotting area, plotting lines, adding plot labels, etc.

<div align="center">

CHAPTER FIVE

PROPOSED SYSTEM

</div>

Dataset is usually split into following

1.SPLITTING OF DATASET -

Training data and test data The output variable along with other variables are included in the training set . The model learns the data and tries to generate some pattern . The other part of the dataset serves as a test set to validate our model's prediction. The scikit library has a function called train_test_split to divide our data . Test_size is the parameter which gives us the percentage of data that should belong to the test set. Train size stores the remaining part as the training at a set , either of which should be specified.

<div align="center">
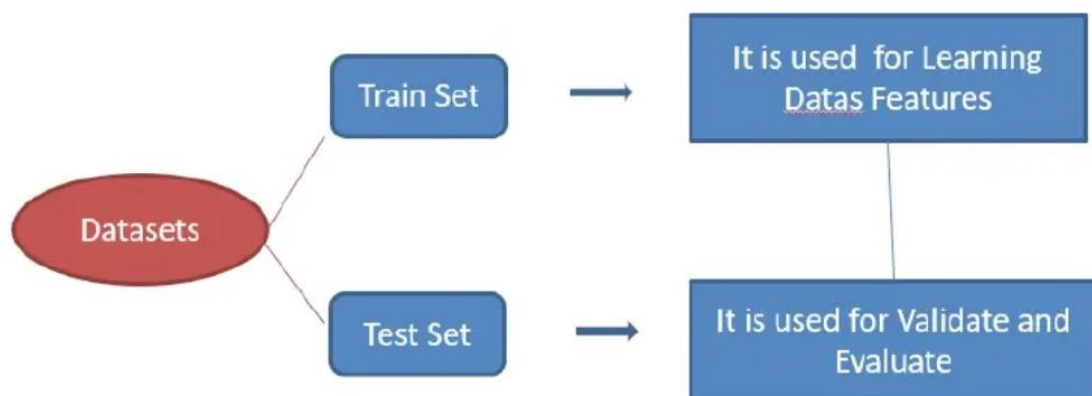
Figure 5.1 Splitting of Dataset

</div>

2. DATASET PRE-PROCESSING MODULE -

Data pre-processing is a cleaning technique which is used to convert / transform the raw data into a clean and properly structured dataset suitable for further analysis. Data is usually collected and gathered from various sources , so it should be good enough and in some specific format before the model learns or gets trained with the data. This will help in achieving better and accurate results with valuable information. The basic steps in pre- processing involve filling up missing values and null values , getting rid of possible outliers and normalization.

Figure 5.2 Data Preprocessing

3.TRAINING WITH ALGORITHM :

K-Nearest Neighbor (KNN) algorithm is an effortless but productive machine learning algorithm. It is effective for classification as well as regression. However, it is more widely used for classification prediction. KNN groups the data into coherent clusters or subsets and classifies the newly inputted data based on its similarity with previously trained data. The input is assigned to the class with which it shares the most nearest neighbors. Though KNN is effective, it has many weaknesses.
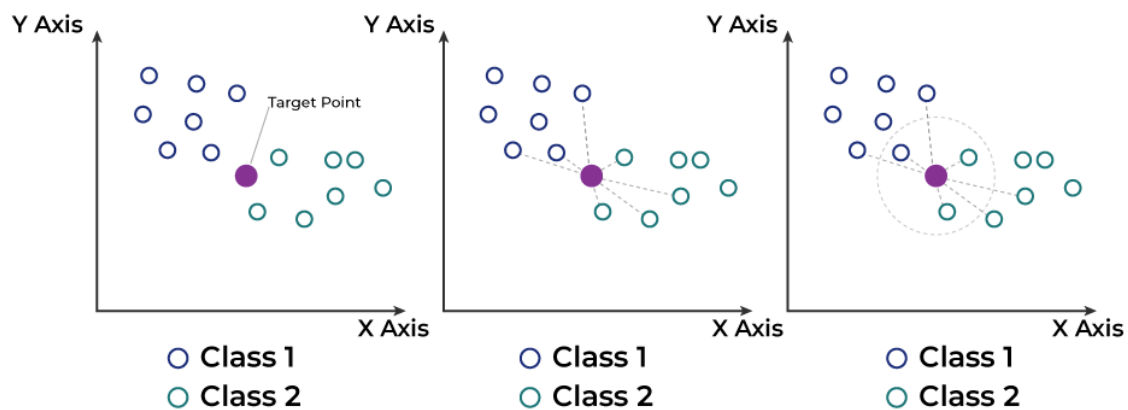
Figure 5.3 KNN Algorithm

CHAPTER SIX
OUTPUT
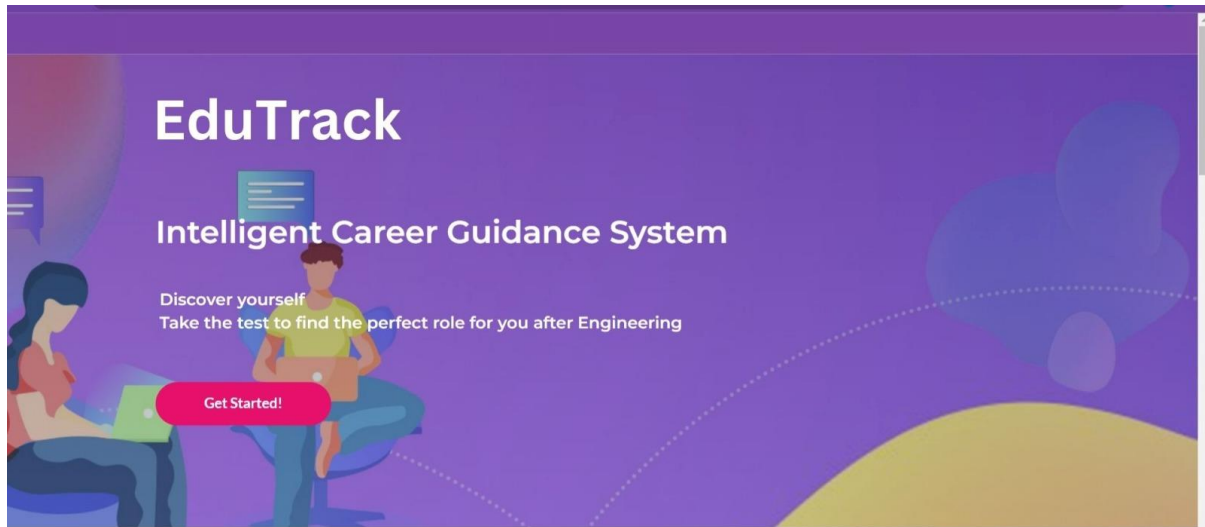
Figure 6.1

This is the homepage of our EduTrack – Intelligent Career Guidance System. Here you can see one button is displayed as "Get Started", by clicking on this button the page will redirect to the main page for the Evaluation based on the some basic analyzing questions for students/users.

Figure 6.2

In this , there are questions regarding to the knowledge of basic computer science subjects and after that students have to select one of the options given according to their level of understanding.
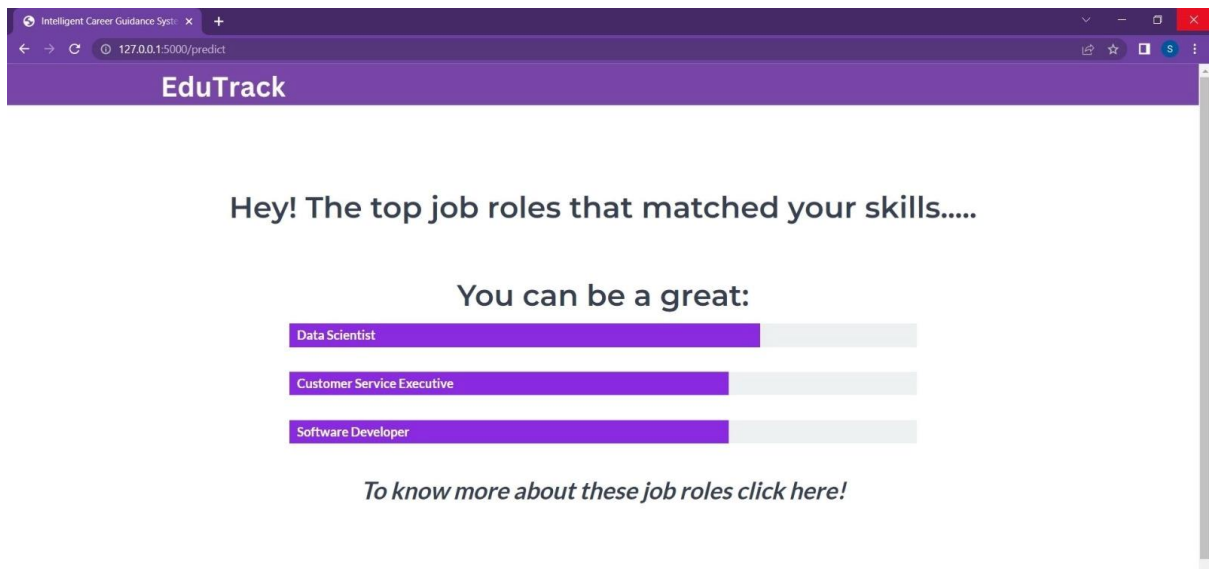
Figure 6.2

Finally , according to the input given by the user, predictions are made by the machine learning algorithm and result is displayed in the form of specific 'Domain name' and bars representing how much possibility of each of the domain displayed on screen.
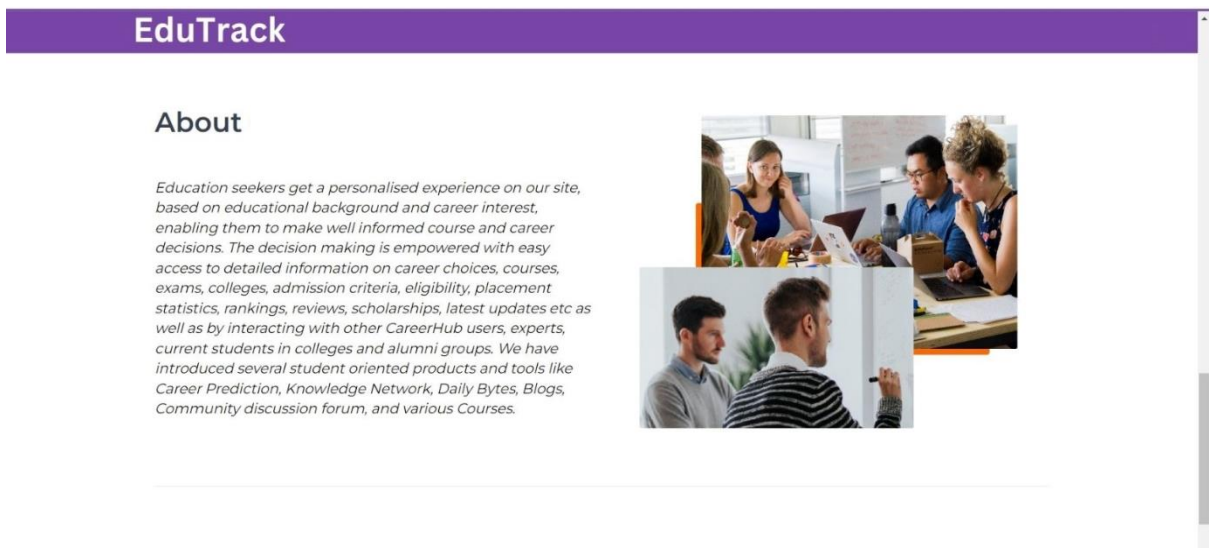

Figure 6.3

The above page gives us information about the EduTrack and it's various courses. We introduced several student oriented products and tools like career prediction, knowledge network, Blogs and various courses.
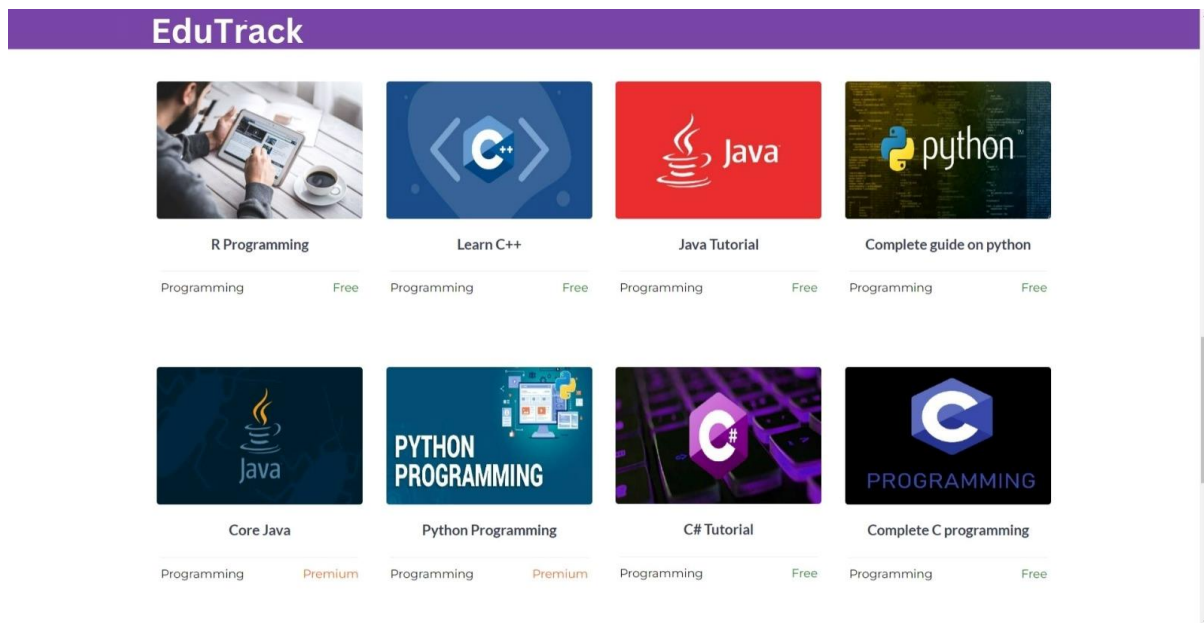
Figure 6.4

Here are the variety of courses which students can access through our websites free of cost and they are majorly related to the Computer science field.
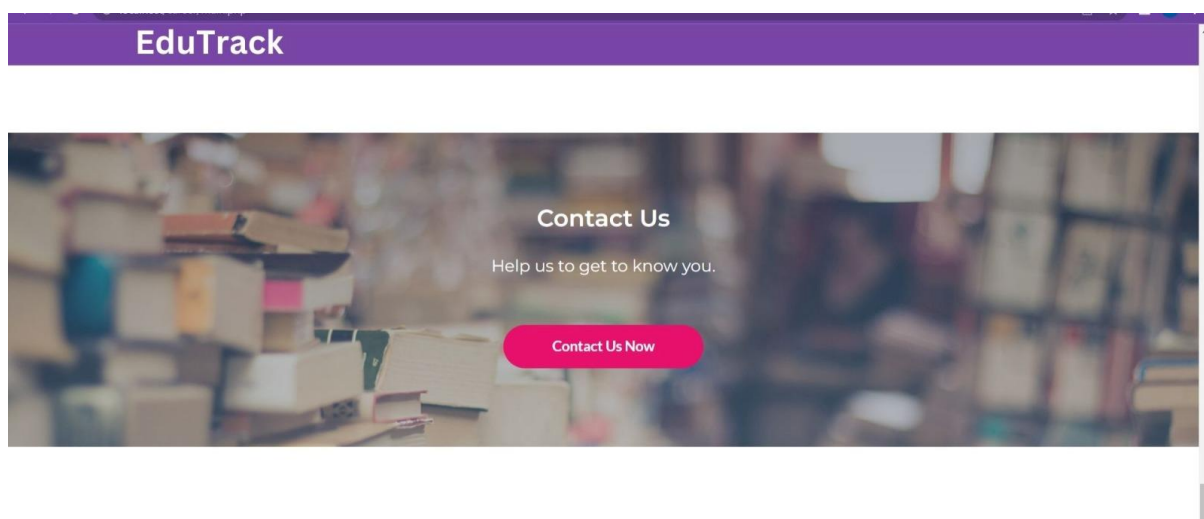

Figure 6.5

The 'Contact Us Now' button provided to get more information or to ask any queries related to the EduTrack website.

CHAPTER SEVEN
ADVANTAGES

1. Personalized Advice

Machine learning can analyze your unique skills, interests, and background to give you career suggestions that fit you specifically.

2. Future Predictions

These systems can look at current trends and predict which jobs will be in demand in the future. This helps you prepare for careers that will have good opportunities down the road.

3. Handles Many Users

Machine learning can handle advice for many people at once without slowing down. This is great for schools or companies with lots of people needing guidance.

4. Based on Real Data

The recommendations come from analyzing real-world data about jobs, salaries, and industry growth. This means the advice you get is grounded in what's actually happening in the job market.

5. Gets Better Over Time

The more the system is used, the smarter it gets. It learns from new data and feedback, constantly improving the accuracy and helpfulness of the advice it gives.

CHAPTER EIGHT
LIMITATIONS

1. The Accuracy of the our model can be improved.

2. We can also make our website more Dynamic and user- Friendly using React and Bootstrap.

3. Database technologies such as MongoDB etc. can be integrated to store more Education fields which can be later trained into our prediction model

4. The effectiveness of the system depends on the quality and amount of data it has. If the data is incomplete, outdated, or biased, the recommendations may not be accurate or reliable.

5. While machine learning can process a lot of data, it may not account for all personal factors, such as individual aspirations or changing personal circumstances, which are important in career decisions.

6. The system's performance heavily relies on the algorithms used. If the algorithms are not well-designed or not suitable for the specific context, the guidance provided may be less effective.

7. Using personal data for machine learning can raise privacy and ethical issues. Ensuring that data is used responsibly and that users' privacy is protected is crucial but can be challenging.

CHAPTER NINE
FUTURE SCOPE

The utility of this website extends its advantages to both students and parents alike. It provides a platform that caters to the educational needs and aspirations of the younger generation. In our ongoing efforts to enhance user experience, we can explore the application of technologies like React and the Bootstrap Framework. These technological enhancements are aimed at creating a user-friendly interface that is both intuitive and engaging.

Furthermore, we recognize the importance of refining the prediction model. One key avenue for improvement is to augment the questionnaire with additional questions. By doing so, we aim to achieve a higher degree of accuracy in predicting the most suitable career paths. In this endeavor, we also plan to introduce a broader range of aptitude questions that are designed to not only gauge academic strengths but also delve into the user's genuine interests and passions.

Expanding the questionnaire and collecting a more extensive set of responses will contribute to the enrichment of our dataset. A more substantial dataset is invaluable in generating more refined and precise forecasts. It enables the prediction model to capture a wider spectrum of user profiles and preferences, resulting in optimal career recommendations that align closely with individual aspirations and capabilities.

In the future, Edutrack can explore advanced machine learning techniques, such as natural language processing (NLP) and deep learning, to further personalize career recommendations. By analyzing not only users' skills and preferences but also their natural language inputs, the system can offer more tailored and nuanced guidance, taking into account individual strengths, aspirations, and learning styles.

In summary, this website serves as a versatile and beneficial resource for students and parents alike. Through the integration of advanced technologies and the expansion of our questionnaire, we are committed to providing an increasingly user-friendly and insightful platform. The ultimate goal is to offer accurate and tailored career guidance that empowers individuals to make informed decisions about their educational and professional futures.

CHAPTER TEN
CONCLUSION

The purpose of this studies become to design and broaden a website for profession prediction which predicts becoming options for a candidate in deciding on a appropriate field. The Options anticipated within the proposed gadget are greater legitimate & particular than the prevailing profession steerage structures inside the field. We have used the KNN set of rules to categorise the ability units of the candidate and predict a appropriate subject with the assist of the solutions of MCQs which the candidate crammed as comments and the K-Means Clustering algorithm is used to form the clusters with the aid of dividing the scholar responses for a specific ability set and predicting the rate of success for the respective fields in each cluster. For subject particular prediction purposes, the achievement fee in every of the clusters are determined; higher achievement charges and lower failure quotes were predicted. In this undertaking, diverse profession prediction systems had been studied very well for constructing an internet-based totally software with predicted outcomes .More studies is needed to better apprehend the framework's accuracy charge and introduction of additional capabilities and the removal of outliers inside the framework.

CHAPTER ELEVEN
REFERENCES

[1] Min Nie , ZhaohuiXiong , RuiyangZhong , Wei Deng and Guowu Yang (2020), " Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students" , Journal of Appl. Sci. pp. 1-14.

[2] K. B. Eashwar, R. Venkatesan and D. Ganesh (2017), "Student Performance Prediction Using SVM", International Journal of Mechanical Engineering and Technology (IJMET) Vol. 8, No. 11, pp. 649–662.

[3] AnkitaKadambande, Snehal Thakur, AkshataMoholand A.M.Ingole (2017), "Predict Student Performance by Utilizing Data Mining Technique and Support Vector Machine", International Research Journal of Engineering and Technology (IRJET), Vol. 04, No. 05, pp 2818-2821.