

Career Guidance System using Artificial Intelligence

Anudnya Patil, Aditi Pednekar, Christina Manakkal, Sherly Mathias, Priya Chaudhari

Anudnya Patil, Department of Information Technology, St. Francis Institute of Technology

Aditi Pednekar, Department of Information Technology, St. Francis Institute of Technology

Christina Manakkal, Department of Information Technology, St. Francis Institute of Technology

Sherly Mathias, Department of Information Technology, St. Francis Institute of Technology

Priya Chaudhari, Department of Information Technology, St. Francis Institute of Technology

Abstract - Choosing a stream after 10th class is often very challenging for students. Students are not always aware of the various career fields available and might end up making the wrong choice because of lack of knowledge. It is very important to have knowledge about the career paths the student is interested in and how to achieve it. So it is necessary to enlighten students about various career options available and also it is very important to provide required guidance to a student who has a keen interest in becoming an engineer or astronaut but doesn't know which exams to give, how and where to apply for the course. Building a decent career is every person's dream. In order to make that dream come true, knowledge about getting there is extremely important. We have developed a website which gives information about the possible career options available after 10th class and our website also provides a career aptitude test for students who are confused about their career path and provides appropriate results by making the user answer a set of questions. Apart from this our website also provides famous professions like teacher, engineer, doctor, lawyer etc and provides detailed information like which exams to give, which degree you should have, exam pattern, syllabus etc to achieve a particular profession.

Key Words: Random Forest, Career Guidance Website, Aptitude Test, Artificial Intelligence, Machine Learning, Career Roadmap

1. INTRODUCTION

Making wrong career choices can negatively impact an individual's life. As career choices need to be made from a young age it can be challenging for students. Students are both confused as well as scared when it comes to selecting one's career path as it is the most crucial decision they will have to take. This decision decides their future. As the students do not have knowledge about the various career options they often end up taking the wrong career path and might later on regret it but it is generally too late by then. In such a situation, the best way to prevent this is to educate the students about all the possible career options which are available after 10th class and give them a proper insight

about all career fields and how to achieve a particular career. Making a choice is stressful for both the students as well as the parents. Although everything is available over the internet, all the information is scattered and students get confused. While choosing the right career it is necessary to think with respect to skills, areas of the interest, abilities and capabilities. This proposed work is the method to provide actual guidance to the students by recommending them careers after 10th class. The students are estimated based on the aptitude test and available career options are recommended accordingly. The main goal of our website is to make all this scattered information which is scattered over the internet available in one single platform. [11,12].

2.1. LITERATURE SURVEY

The authors in [1]. have studied that will help the students select a suitable career post 12th, which will fit along with their Aptitude, Emotional Quotient and Personality and Traits along-side their interests. The authors proposed two main parts of the system :one is the Web Portal which is an interface between the student and the system and which is designed using HTML and CSS with Bootstrap 4 at the Front End, and MySQL and PHP is used for the back end. Also they have used a recommendation engine which is implemented using python and built with the help of Machine Learning. This recommends students suitable career options with the engine. The Machine Learning Algorithms used here are Naive Bayes, K- Nearest Neighbour and Random Forest Classifier.

The authors in [2]. have analyzed the research gap where a design improvement can be introduced to help students to be more aware of the career decisions they are making, and in the process, try to remove some ambiguity because career is a very important part of an individual's life. The solution that was proposed is a mobile and web-based application with a chatbot being the main point of contact.

The authors in [3]. have analyzed previously that the career professionals used the opinion poll to identify the detail that is potentially influencing career decisions

among the students because career choices play a critical role in the life of a student. Then, they proposed a framework called as the Approach Cluster Centers Based On the XGBOOST (ACC BOX) framework to predict students' career decisions. As a result of multiple experiments conducted, it was observed that the approach of this paper is superior to all the other approaches to predict career choice.

The authors in [4] have studied all the challenges that are associated with a manual career guidance system and have developed a software application career guide information system for students who are in junior as well as for students who are in senior secondary school to get more access to the information regarding subject combinations for various career choices. They have proposed a Career Guide Information System (CGIS) which is an automated online platform wherein students can get real online guide on their future careers regardless of their location by methods like Unified Modeling language (UML) and by using PHP programming Language and MySQL along with XAMPP server which allows the application to use TCP/IP for intranet packet sharing and file/information transfers across system.

The authors in [5] have studied that most of the students across the globe are always in confusion after they complete higher secondary and therefore, the stage where they need to settle on an appropriate career path. Their system is a web based application which uses K-Nearest Neighbor algorithm and K-Nearest Neighbor algorithm which will help in choosing an appropriate course which is able to also reduce the failure rate by choosing a wrong career path.

The authors in [6] have studied how choosing a career has always been challenging for everyone since very long. The application acts as a solution for all the problems that students face while choosing a career. This application acts as an approach to provide actual guidance to the students by recommending them all the possible careers available after 10th and 12th grade using various ML(machine learning) algorithms.

The research work aims to develop a system titled, 'CareerBuddy', for helping students to make the correct career choice by providing information about all the career paths that the students can take after 10th class. The system helps the students to get an overview of all the career fields available after 10th class. Along with this, the system also provides services like aptitude tests which take input from the user in the form of questions and provides appropriate results by using the random forest machine learning algorithm.

2.2. SYSTEM ARCHITECTURE

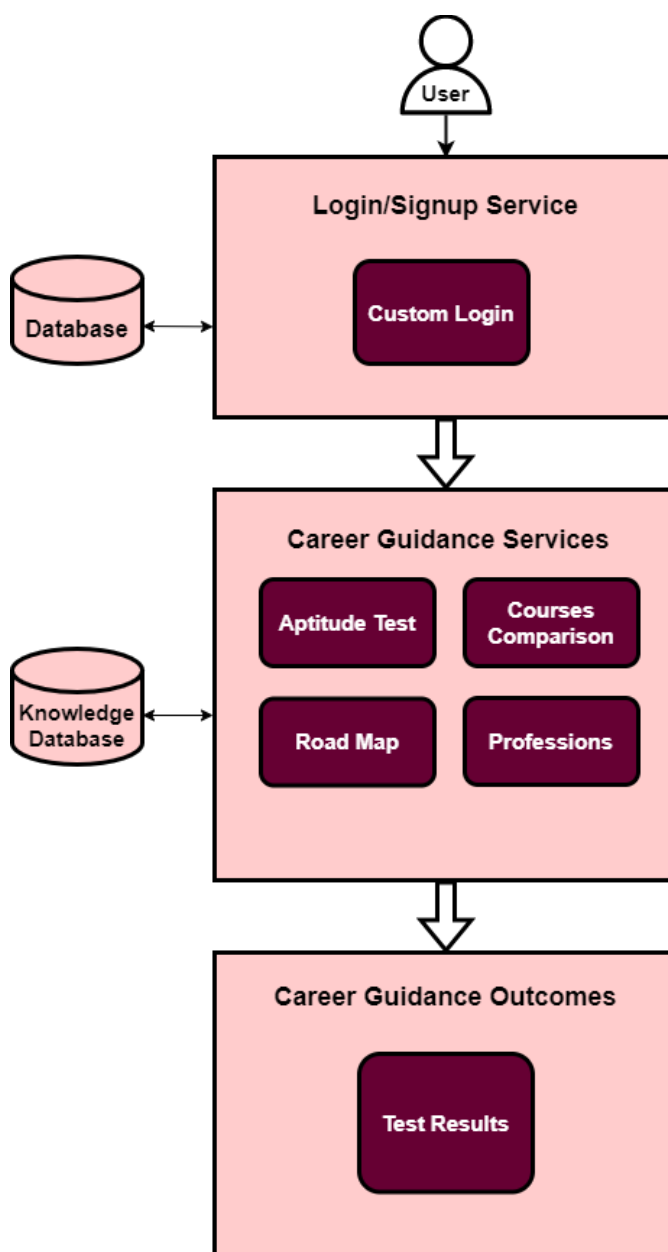


Fig. 2.1 Architecture of Career Guidance System

The architecture of the envisioned system is shown in fig. 2.1. It consists of four modules:

A. Aptitude Tests: This module is responsible for taking aptitude tests and providing results using the random forest machine learning algorithm. The answers given by the students are taken as an input which will be processed using the random forest algorithm. The aptitude test will be divided into two categories. The first test will ask for the general interests of the user and the second test will be more specific based on the results of the first test of the user pertaining to his interests.

Various questions are individual decisions. Answers can be used to perform the activity of the decision tree. And the target decision will be the final output based on majority voting and averaging as applicable. The users

can navigate through the system to view roadmaps, take the test and browse through various courses related to their interests.

The generation of the output will be performed using the following steps:

- 1) The user will select the options available to them for the questions on the frontend website.
- 2) The conversion of the selected options into appropriate format so that it can serve as an input data to the trained machine learning model in the backend of the system.
- 3) The inputs will be processed and the result will be fetched from the model and displayed on the webpage as final outcome or redirection to the second stage depending upon which quiz the user submitted.

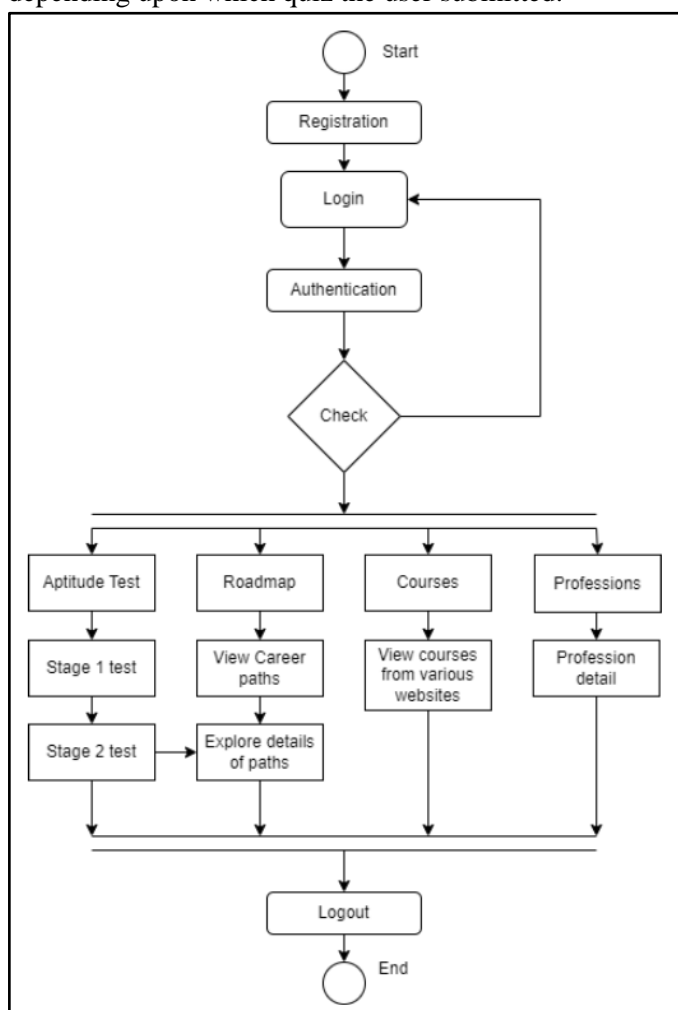


Fig. 2.2 System Design

B. Roadmap: This module is responsible for giving an overview of the possible career paths available after 10th class. The user is first redirected to register or login and further they can explore the courses available, go through the roadmap and view the career path available. The user can also take an aptitude test which is provided in two stages for better accuracy in the designated field.

C. Courses: In this section, the user can browse the web page for courses and register if they are interested. The links and short description are provided for the same.

D. Profession: In this section, the user can view the detailed information with respect to a certain career. For example, the qualifications, exams, if applicable, specific degrees required and so on.

2.4. EXPERIMENTAL EVALUATION

Implementation Details:

A. Dataset:

The dataset used in the ML algorithm for the recommendation consists of 31 columns and 841 entries which we have made for the aptitude test taking into consideration the likes and dislikes for a specific stream. The dataset contains numeric and categorical data. The parameters are the respective questions for a specific stream and according to the positive or negative answers the most feasible stream is suggested. There are a total of 5 datasets used for the respective 5 models i.e for Overall (Arts, Science, Commerce), in depth arts, science, commerce and for vocational streams.

B. Algorithm:

The Machine Learning algorithm which is used for recommendation is one of the highly important aspects of this system. There are different types of algorithms which generate different models and different expected results. We went through various algorithms like KNN, Naive Bayes, Decision Trees and then Random Forest, which is an ensemble technique. Single decision tree might be easy to visualize and implement but it mostly contains high variance. Considering this limitation, we are using an ensemble method. It merges the predictions of the several independent models formed using the training dataset for better results.

Ensemble methods are broadly categorized into two types - Sequential techniques and Parallel techniques. In sequential techniques, results from previous models are used for next models in a continuous input and output process. It aims to combine weak learners into strong learners such that the final model in the sequence has the highest accuracy. However, in parallel technique, models are generated independently and their result is aggregated for the best result. Random forest uses this bagging technique to give us the output. Different training subsets with replacements are created for this purpose. The Random Forest algorithm uses the parallel ensemble technique and base learners are generated independently and parallelly.

Random forest is a combination of various decision trees. In the Random Forest algorithm, firstly, a number of records are selected randomly. These records are used to create individual decision trees. This process is truly random in the sense that not all attributes are considered while designing an individual tree. Each tree is diverse. Now, each of these trees will generate an output. The final output will be decided by combining these outputs. This can be done using Majority Voting for classification of the result. Hyperparameters can be used to enhance or control the performance of the model. These include the total number of trees built by the algorithm, maximum number of features considered before splitting, minimum number of leaves to split an internal node, cross validation and many more.

We will use two tests with varied amounts of questions and Random Forest algorithm is being used for the purpose of recommendation based on overall scores. The most important features of the Random Forest Algorithm is that it can handle the data sets containing continuous and categorical variables.[8].

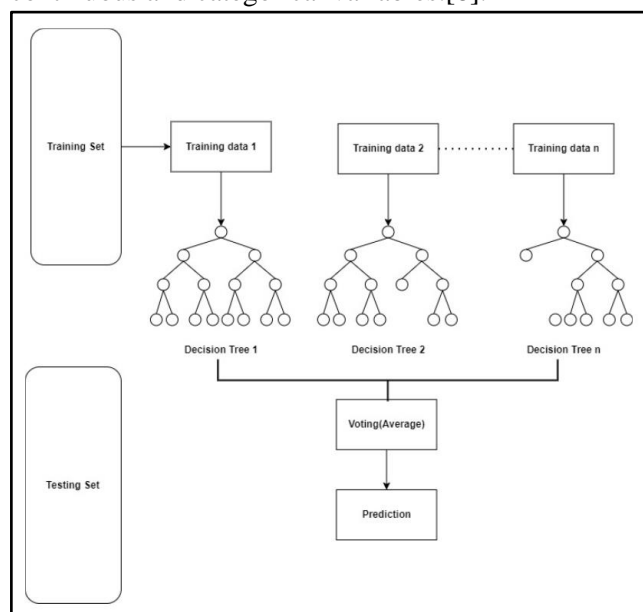


Fig. 2.3 Random Forest Algorithm

User needs to login to access the services.

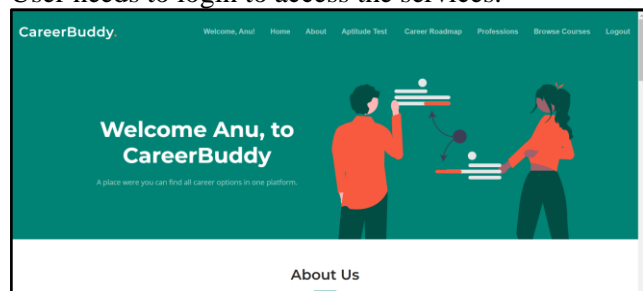


Fig. 2.4 Home Page

The user has the choice to take the first two-staged test if really confused or directly test for 12th, Vocational or Diploma Courses if they prefer either.

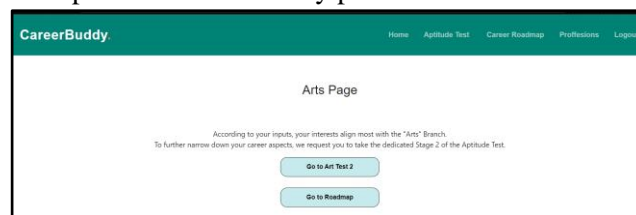


Fig. 2.5 Arts stream recommended after aptitude test

Fig. 2.5 is an example of one of the outputs of the two-stage test. The user has two options - take the recommended second stage test or go to the roadmap.



Fig. 2.6 Career roadmap

Several courses are mentioned in the roadmap page. Each of the links will redirect you to a more detailed webpage for that specific choice.

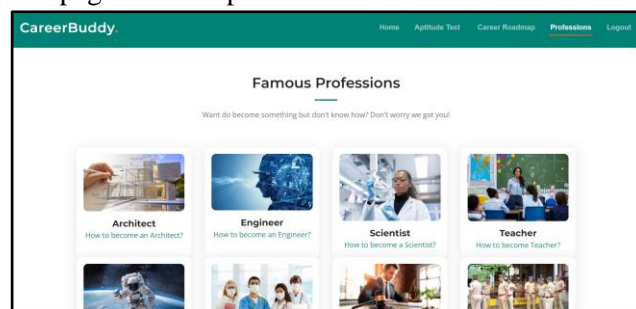


Fig. 2.7 Professions tab

In Fig. 2.7, we can see the famous professions webpage, from there if we click on Engineer for example, we are able to see a website showing types, exams, steps and other details related to engineering.

In all the above figures, we have displayed some parts from all of the extensive features available in the website.

A. Experimental Results

The performance of the suggested classification system is evaluated on the basis of the accuracy. It calculates out of the total number of predictions, how many labels the model got right. This can be seen as the percentage of predictions that were accurate.

The Accuracy score is calculated by dividing the number of correct predictions by the total number of predictions.

$$\text{Accuracy} = \left(\frac{\text{Number of Correct Predictions}}{\text{Total number of Predictions}} \right) \times 100$$

This module was evaluated against 252 entries and several algorithms were taken into consideration. Here, we are demonstrating a few among several comparisons alongside the actual values in the form of a graph. It can be seen in the diagram that the algorithms KNN, Decision Tree and Random forest gave accuracy scores of approximately 88 percent, 92 percent and 96 percent respectively. Random forest displayed the highest accuracy as expected.

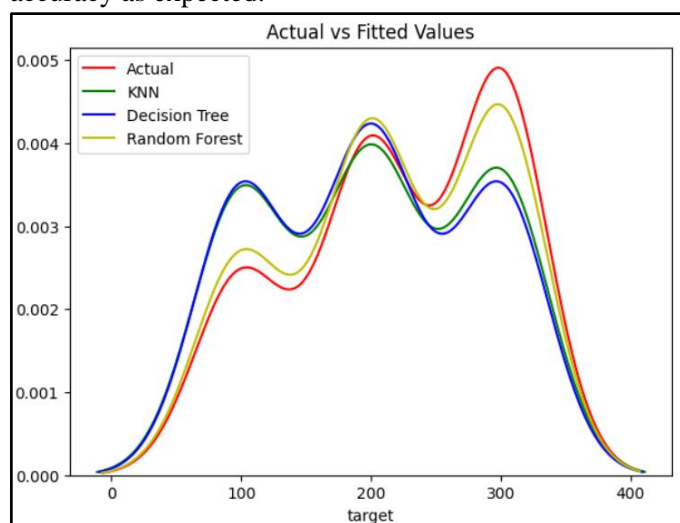


Fig. 2.8 Actual vs Predicted Values

TABLE I: MODEL PERFORMANCE COMPARISON

Algorithm	Accuracy
K Nearest Neighbors	88 %
Decision Tree	92 %
Random Forest	96 %

3. CONCLUSION

CareerBuddy is a career guidance website where the main purpose is to integrate all the scattered processes and information regarding career interests into a single platform. It provides students an overview of all the

available career paths that they can choose from which helps in reducing confusion in students. It also provides an aptitude test that the student can take and careerBuddy recommends the most suitable career path to the student based on the answers the student has given to the questions. The system uses a random forest machine learning algorithm and takes the answers given by the user as an input to provide a clear and appropriate result in the form of a graph. This proposed work tries to give a nice beginning to the career of a student which will benefit the students. This proposed work will help students to explore and plan for future careers with minimum confusion as all information is provided on a single platform[13].

ACKNOWLEDGEMENT

We extend our heartfelt gratitude and acknowledge the support and guidance provided by our mentor Prof. Priya Chaudhari for valuable inputs, continuous encouragement and cooperation throughout this research work. We also wish to thank the librarians at the SFIT Library for their assistance in finding the necessary research materials. Finally, we express our gratitude to our families and friends for their constant support and encouragement.

REFERENCES

1. H.Patel, J.Shah, S.Thacker, S.Mishra "Career Guidance System Using Machine Learning". International Journal Of Creative Research Thoughts, Vol. 9, no.5, pp. a57-a60, May 5, 2021.
2. A. Biala "Digital System and Application for Career Guidance for students".B.S. Thesis,Indian Institute of Technology Guwahati, Assam, Assam, 2021
3. M. Nie, Z. Xiong, R. Zhong, W. Deng, and G. Yang, "Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students." Applied Sciences, vol. 10, no. 8, pp. 2841-2854, Apr. 20, 2020, doi: 10.3390/app10082841.
4. O.Akpofure, T.Tete,"Design and Implementation of a Career Guide Information System for Secondary School Leavers An Object Oriented Approach ". International Journal of Science and Research, Vol. 4, no. 4, pp. 2635 -2641, April, 2015.
5. S. Vignesh, C. Shivani Priyanka, H. Shree Manju and K. Mythili, "An Intelligent Career Guidance System using Machine Learning," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 987-990, doi: 10.1109/ICACCS51430.2021.9441978.
6. R. Ghatugade, G. Chormunge, V.Kumar, R. Nikhare, "Web Based Career Guidance System," International Research Journal of Engineering and Technology (IRJET), vol. 08, no. 03, pp. 2490-2494.
7. V. K. Gupta, A. Gupta, D. Kumar and A. Sardana, "Prediction of COVID-19 confirmed,death, and cured cases in India using random forest model", Big Data Mining and Analytics,vol. 4, no. 2, pp. 116-123, Feb. 2021, doi: 10.26599/bdma.2020.9020016.

8. K. V. Reddy, M. Reddy, V. Kaur, and G. Kaur, "Career Guidance System Using Ensemble Learning.", SSRN Electronic Journal, 2022, pp. 33-39, doi: 10.2139/ssrn.4157249.
9. G. Khanvilkar and P. Deepali Vora, "Sentiment Analysis for Product Recommendation Using Random Forest." International Journal of Engineering & Technology, vol. 7, no. 3, pp. 87-89, 2018, doi: 10.14419/ijet.v7i3.3.14492.
10. N. VidyaShreeram and D. A. Muthukumaravel, "Student Career Prediction Using Decision Tree and Random Forest Machine Learning Classifiers." Proceedings of the First International Conference on Computing, Communication and Control System, I3CAC 2021, 7-8 June 2021, Bharath University, Chennai, India, 2021, doi: 10.4108/eai.7-6-2021.2308621.
11. S. Thamke, A. Talankar, V. Wadhale, S. Baghile, and T. Manekar, "Web Based Scientific Career Counseling System" International Journal Of Creative Research Thoughts(IJCRT), vol. 10, no. 5 May 2021, pp. 202-205.
12. M. Singh, N. Bhurra, N. Chittodiya, and N. Sonis, "Career Guidance System." International Research Journal of Modernization in Engineering Technology and Science, vol. 4, no. 11, pp. 1413-1415, Nov. 2022, doi: 10.56726/irjmets34704.
13. L. Prasanna and D. Haritha, "Smart Career Guidance and Recommendation System." International Journal of Engineering Development and Research, vol. 7, no. 3, pp. 633-638, Feb. 2021.