CAREER GUIDANCE SYSTEM USING MACHINE LEARNING

Hiren Patel Jash Sanghavi Saurin Shah Sachet Thacker Sandeep Mishra Student Student Student Student Project Guide

Department of Electronics and Telecommunication Engineering

K.J Somaiya Institute of Engineering and Information Technology,University of Mumbai,Maharashtra,India

Abstract—A Career Guidance System for Science students studying in 11th and 12th standard. It will help these students select a suitable career post 12th, which will fit along with their Aptitude, Emotional Quotient and Personality and Traits alongside their interests. This project consists of two main parts: one is the Web Portal which is an interface between the student and the system and which is designed using HTML, CSS, Bootstrap 4 at the Front End, and MySQL and PHP at the back end; the other is the Recommendation Engine which has been built using Machine Learning and implemented using Python, so as to recommend suitable career options to the student using the system. There are 3 Machine Learning Algorithms which will be used: Naive Bayes, K-Nearest Neighbour and Random Forest Classifier. The student, on coming to the portal, will be asked to undergo a series of questions, bifurcated into different tests, each to test the student's IQ, EQ and Personality. Upon collecting this data entered by the student, this data will be tested against an existing dataset and suitable career choices will be recommended to the student. The Career Guidance System, when employed, will help provide proper guidance to the students regarding which career choices would be best suited to them as per their inner qualities and their interest, thus helping them pursue a proper career and achieve success in the same.

Index Terms—Web development ;Machine learning

I. INTRODUCTION

In today's day and age, there are unprecedented developments taking place in all spheres of life, ranging from art to technology, and everything in between. More and more fields are coming up, more and more technologies are being discovered, more and more advancements are taking place in the existing technologies. An increase in technologies, in turn, leads to an increase in the demand of professionals who, in their careers, could effectively work with those technologies. And for these professionals to be successful in their respective fields, it is imperative that they make conscious and wise decisions while selecting their careers.

Students pursuing Science almost always face a dilemma as to which field should they study in after completing 12th. This is due to the plethora of career options available to them coupled with them having little to no knowledge about those careers, their future scope, as well as a misalignment between their interests, knowledge, and strengths and skills.

This, more often than not, leads to students making such life altering decisions just on the basis of their whims and fancies, as well as due to being influenced by the decisions of their peers. This Career Guidance System has been thus developed keeping the dire needs and requirements of such students in mind, so as to help them choose a suitable career option wisely. Through this project, students studying in Science will get a relatively clear idea as to which sub-fields from Engineering, Medical or Life Sciences could they pursue and achieve success in.

II. THEORY

1. WEB PORTAL

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, PHP, MYSQL 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, his/her EQ and his/her Personality traits. The portal will also collect information about the interest of the user. The MCQs in the IQ questionnaire will have only one correct option, whereas MCQs in the EQ and Personality questionnaires will not have any single correct option as 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

PERSONALITY TRAIT TEST

on screen. Bootstrap is a front-end framework used to create our web portal. PHP is deployed on a web server and is used to have links between the data set and the web portal to get the answers from the student coming to our portal. XAMPP to treat a localhost like a remote host by connecting using an FTP client.

Using the above said technologies, we have made a web portal . Underneath we have attached a picture which gives you all an idea about how it would look.

Fig 1: Web Portal home page.

It will be an introduction page which gives the user an idea about the portal and it works and a gist about us. It consists of a nav-bar, about us para and a footer.

APTITUDE TEST PAGE

The next page will be a test page which comprises an aptitude test. It consists of a nav-bar, form and a footer. The question will be displayed with the help of HTML language and will be styled with respect to CSS and the formation will be done by bootstrap respectively. The aptitude test will have a small introduction about itself and its working. Then it will have questions and four options for each question. The selection and the result will be displayed with the help of PHP which will be used as a back-end language. Underneath we have attached a picture which gives you all an idea about how it would look.
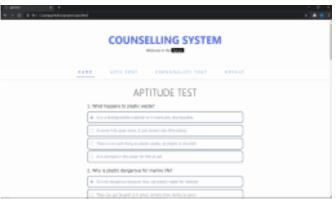

Fig 1: Aptitude test page.

The next page will be a test page which comprises a personality trait test. It consists of a navbar, form and a footer. The question will be displayed with the help of html language and will be styled with respect to CSS and the formation will be done by bootstrap respectively. The personality test will have a small introduction about itself and its working. Then it will have questions and four options for each question. The selection and the result will be displayed with the help of PHP which will be used as a backend language. Underneath we have attached a picture which gives you all an idea about how it would look.
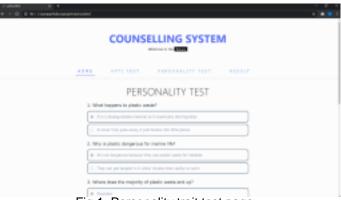

Fig 1: Personality trait test page.

RESULT PAGE

The next page will be a results page that will display the result based on the option you select, processed by the machine learning algorithm.

Fig 1: Result page page.

CONTACT US PAGE

The next page will be the contact us page. It consists of a navbar, contact-us form, details about the institute and a footer The page consists of a contact us form for any technical issues and also a number and email of a carrier consulting institute. Underneath we have attached a picture which gives you all an idea about how it would look.

2. MACHINE LEARNING

Machine Learning is at the core of the project. It is being extensively used to help predict and recommend suitable career choices to the user at the portal. Our project is basically a recommendation system, and various machine learning algorithms can be used to build recommendation systems, depending on the output requirement.

Recommendation systems work on recommendation engines. A recommendation engine filters the data using different algorithms and recommends the most relevant items to users. It captures the liking and inclination of the user and based on that recommends options which match the users inclination.

Recommendation engines work on filtering models. There are 2 main types of filtering models: Content Based Filtering and Collaborative Filtering. Content based filtering works when the previous preferences and likes and dislikes are known so as to filter what product could be liked in the future. In the case of the Career Guidance System, this doesn't work. Thus we turn towards Collaborative Filtering.

Collaborative Filtering further is of two types: User-User Based Filtering and Item-Item Based Filtering. User-User Collaborative Filtering first finds the similarity score between users. Based on this similarity score, it then picks out the most similar users and recommends products which these similar users have liked or bought previously. Item-Item Collaborative Filtering finds similarity between each pair of items, and then recommends items liked in the past.

While the data was being collected, we tested various machine learning algorithms on Movie Recommendation Dataset. Support Vector Machine (SVM) gave an accuracy of 0.68, while Naive Bayes algorithm gave an accuracy of 0.75.

We have decided to use Naive Bayes, K-Nearest Neighbour and Random Forest Classification Algorithms.

For collecting the dataset, we needed to decide which models to use for gauging the personality of the person, for which we researched the following models

- Holland Codes: Holland Codes are one of the most popular models used for career tests today. Holland argued that the choice of a vocation is an expression of personality. There are six personality types in Holland's model and most people will fit into a few of the categories:
- The Big Five: The Big Five come from the statistical study of responses to personality items. Using a technique called factor analysis researchers can look at the responses of people to hundreds of personality items and ask the question "what is the best way to summarize an individual?".
- MBTI: The underlying assumption of the MBTI is that we all have specific preferences in the way we construe our experiences, and these preferences underlie our interests, needs, values, and motivation. The MBTI sorts psychological differences of people into four opposite pairs, or "dichotomies", with a resulting 16 possible psychological types.

We needed to implement the Personality Test in our google form so as to gather data. In Big 5 Personality Test, the personality of the person is calculated using a set of 5 mathematical formulae which is much more easy to implement using excel sheets as compared to Holland Code and MBTI. Thus we selected the Big 5 Personality Test, for both testing the user as well as for collecting our data.

III. RESULT AND DISCUSSION

The student arriving at our portal will first select the options of their interest and then will go for the questionnaires presented on their screen The answers that they have selected will be then directed to the DATABASE of our system which will become our testing data. We needed to implement the Personality Test in our google form so as to gather data. In Big 5 Personality Test, the personality of the person is calculated using a set of 5 mathematical formulae which is much more easy to implement using excel sheets as compared to Holland Code and MBTI. The machine learning requires the data for testing and training The answers selected by students will be used for the training part of machine learning. Thus after the model trains on our training data and our input from the user is taken as test data from which we get the recommended career as our output The accuracy through the various algorithms differs from each other.

IV. CONCLUSION

A portal has been designed for taking inputs from the applicant, testing his/her aptitude as well as his/her interests and personality. These inputs are used as the test data for the Machine Learning model which then gives the recommended career as the output. It is also observed that the accuracy predicted of various methods differ from each other. It is predicted that the Random Forest Classifier will be providing highest accuracy and the Naive Bayes Algorithm will be providing least accuracy.

ML Algorithm	Accuracy
Naive Bayes	0.57
K-Nearest Neighbour	0.64
Random Forest Classifier	0.81

Table 1: Machine Learning Algorithm Performance

REFERENCES

[1] "Automated Career Counseling System for Students using CBR and J48 Maha Nawaz, Anum Adnan, Unsa Tariq, Jannat Fatima Salman, Rabia Asjad, and Maria Tamoor. J. Appl. Environ. Biol. Sci., 4(7S)113-120, 2014 ISSN: 2090-4274 Journal of Applied Environmental and Biological Sciences

[2] Sam, Gladence Karthi, M Anu, Maria. (2015). A statistical comparison of logistic regression and different bayes classification methods for machine learning. ARPN Journal of Engineering and Applied Sciences. 10. 5947-5953.

[3] Sam, Gladence Karthi, M Anu, Maria. (2015). A statistical comparison of logistic regression and different bayes classification methods for machine learning. ARPN Journal of Engineering and Applied Sciences. 10. 5947-5953.

[4] Sam, Gladence Karthi, M Anu, Maria. (2015). A statistical comparison of logistic regression and different bayes classification methods for machine learning. ARPN Journal of Engineering and Applied Sciences. 10. 5947-5953.

[5] Smart Career Guidance and Recommendation System, IJEDR 2019 — Volume 7, Issue 3 — ISSN: 2321-9939

[6] Student Career Prediction Using Advanced Machine Learning Techniques, International Journal of Engineering Technology Website: www.sciencepubco.com/index.php/IJET

4