

# Recommendation System for Career Path using Data Mining Approaches

Pratik Devidas Dusane, Nikita Vinayak Bhosale, Vaishnavi Avinash Avhad, Pratik Kailas Naikwade

Dept of Information technology  
Sanjivani College of Engineering, Kopargaon  
Kopargaon, India

nikitabhosale2298@gmail.com, vaishnaviavhad1998@gmail.com, pratikknaikwade@gmail.com, dusanepratik0@gmail.com

**Abstract** – Nowadays, students are often facing a problem to choose a career in their life. There are several factors that influenced the students while choosing their career path such as their personal aptitudes, educational achievement and their environment. After completing secondary and higher secondary education, students are normally starting to find a career path which may suit their skill and potential the best. However, many students made wrong decision on selecting their career due to the lack of experience, help and advice from friends and relatives, parents and lecturers, or career counseling. So, in order to tackle this situation we are developing such system which will help student to select their career path based on their interests.

**Keywords**– Data Mining, Recommendation System, Decision System, Regression.

## I. INTRODUCTION

Nowadays choosing a career has become a hectic and problematic job, especially within the new candidates. Choosing a career can simply be defined as a leading decision one takes at a very young age. It's never easy to understand what individuals want to do as they are not that mature to take a knowledgeable decision. Sometimes the hurdles they face discourage them and they try to abandon their choices. Lack of professional guidance, financial issues, and the fear of failure leads to a number of problems while choosing a proper career.

So it is very important that the students are provided the knowledge they need to choose a career that suits them and their talent. A recent survey on career awareness and options in career among the Indian students showed that 92% people ageing between 14-21 were aware of only 7-8 career option and there are a total around 260 career options available. So it is very high time to encourage the students to look at proper career path. For this reason we are developing a web application the will help the students to select and choose proper career paths according to their skills and talents.

## II. LITERATURE SURVEY

For this project that is the career recommendation system we studied different papers of different authors and found out various outcomes that will help our project. From [1] we indicated that each algorithm covers the clustering analyses weaknesses of other algorithms for some customers. The integrated data of all algorithms analyses brings detailed results from customers, behavioral method

and its relation with shopping basket as well. So, by using integrated collective data, it can be determined marketing policies and customer satisfaction appropriate to all customers' clustering and their orientation which finally lead to increased productivity and incomes.

In [2] Neelamadhab Padhy1 and RasmitaPanigrahi briefly reviewed the various data mining prediction techniques in literature. This research definitely helpful to other researcher to impart the several of data mining prediction issues. It is really very difficult to predict and it is a complex. Actually no approaches or tools can guarantee to generate the accurate prediction in the organization.

In this paper, Neelamadhab Padhy1 and RasmitaPanigrahi have analyzed the different algorithm and prediction technique. In spite the fact that the least median squares regression is known to produce better results than the classifier linear regression techniques from the given set of attributes. As comparison we found that Linear Regression technique which takes the lesser time as compared to Least Median Square Regression.

Predictive analysis was studied by Kavyaand Arumugam in [4]. According to them Predictive analytics is the future of data mining. This study focus towards the predictive analytics, regression techniques and forecasting in knowledge discovery domain. Business intelligence is used in predictive analytics for modeling and forecasting. Predictive analytics are more efficient in choosing marketing methods and helpful in social media analytics

### III. ARCHITECTURAL DESIGN

This section is dedicated to the system architecture. Here the process model the breakdown structure of system is been analyzed.

#### 1. Process Model

In this system, Student will first do the registration after which he/she can login and give the preliminary test after which a particular field will be recommended by the system based on the result of first test. Now, again the field specific test will appear and after giving that test, its result will depict the excellence of the student in that recommended field. Admin will be able to login and view, update and manage the records of the student after which the final records will be stored in the database. Students will be able to give the feedback about the system. When student will give the feedback, those feedbacks will be stored in the database which will then be fetched and monitored by the chat bot. Chat bot will give the responses to the queries of user if any.

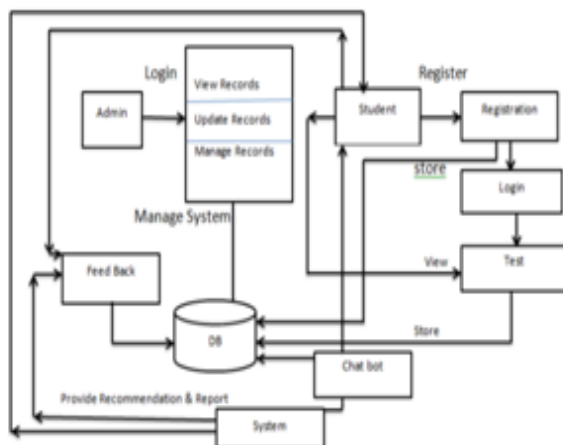


Fig.1. Process Module.

#### 2. Breakdown Structure

**2.1 Admin:** - In this module information access is provided to user by admin.

Admin keeps track of all records which are filled by the user.

Admin can View, Update, Delete any Student Record.

Main role of admin is to make changes in the current application if any updation is found.

If user gives any feedback, then According to their feedback admin can make changes in the application.

**3. Student:** - In this module student can perform various activities as follows

Firstly, Student Registration is done. After registration Student Login in System.

They need to provide their Personal Information and Educational details.

After that Test is conducted for student Based on that test result report is generated.

If user wants to give any suggestions about the Application then user can do it.

**3.Recommendation:**-Here, Actual processing happens. Analyze the test result and applying Algorithm.

This module gives the recommendation to user for selecting there career path.

It provides all required documents to the student.

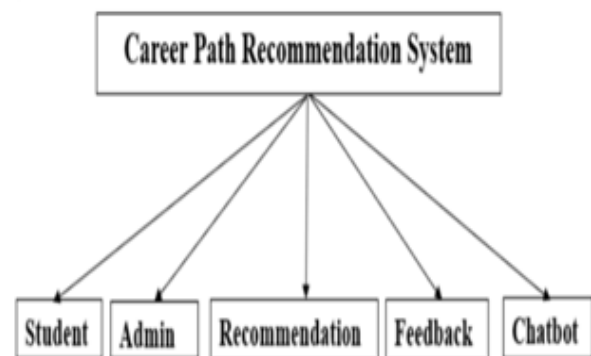


Fig.2. Breakdown Structure.

**4. Feedback:**-To support performances optimization at runtime, this module implemented.

User will give feedback so that Admin can analysis whether application is helpful or not.

If there is some requirements which should get added by admin side it can also analysis so that it will help admin to raise their awareness of strengths and areas for improvement, and identify actions to be taken to improve performances

**5. Chatbot:** - This module provides interaction between user and admin. This module is used to solve the query of user and provide Suggestions.

#### 5. Algorithm

##### 5.1 Decision Tree Algorithm

Decision Tree Algorithm tries to solve the problem by using tree representation. Each internal node of the tree corresponds to an attribute, and each leaf node corresponds to a class label.

Following are the steps for this algorithm:

- Start
- Register in System
- Login Successfully
- Give the First Test set Root Node Career category
- Give the Second test set Sub Node Career Branch
- Ready to give Third MCQ Test g. Generate Result
- View the Report
- Give the Feedback
- Logout from System
- Stop

## 6. Linear Regression Algorithm

A regression is a statistical analysis assessing the association between two variables. It is used to find the relationship between two variables. et.al [2] NeelamadhabPadhy, and Rasmita defined that is a one kind of predictive model which provides the prediction about the unknown data values by using the known data.

There are so many techniques are available like Classification, Regression, Time series analysis, Prediction etc. If a set of random data  $(x_1, y_1)$ ,  $(x_2, y_2)$ ,  $(x_n, y_n)$  for two numerical variables X and Y, where X is a cause of Y. In this linear regression analysis, the distribution of the random data appears as a straight line in X, Y space when X and Y are perfectly related linearly. This captures a relationship between two variables. This line function can be given as [3].

$$A = ax + b$$

Here, the linear regression model is used to extract the texture features from the correlation in the frequency channel pairs. The energy values of two frequency channels of one of the channel pair in the top ten list are taken from the channel energy matrix M and consider these energy values as the random data  $(x_1, y_1)$ ,  $(x_2, y_2)$ , .....  $(x_n, y_n)$  for two variables X and Y represent a straight line in X, Y space.

Steps :

Import all required file including mean to define x value and y value.

Find the slope from the accepted values i.e. m

Calculate slope intercept by using  $Y = mx + b$

Determine best-fit value using

$$m = \frac{(x \cdot y) - x \cdot y}{(x \cdot x) - (x \cdot x)}$$

Find intercept of best-fit using  $b = y - mx$

Then by using round method and Rsquared method of prediction.

(like :  $\text{str}(\text{round}(m, 2)) + "x + " + \text{str}(\text{round}(b, 2))$ )

xprediction = value

$$yprediction = (m * xprediction) + b$$

## IV. CONCLUSION AND FUTURE WORK

From the result of this research there are several things that can be concluded.

- Created a recommendation system for the students
- Conducted several tests and analysed student's performance.
- Generated result in the form of chart.

## REFERENCES

- [1]. Farhad Soleimani Gharehchopogh, Tahmineh Haddadi Bonab and Seyyed Reza Khazae, "A Linear Regression Approach to Prediction of Stock Market Trading Volume: A Case Study", International Journal of Managing Value Supply Chains (IJMVSC), September 2013
- [2]. Neelamadhab Padhy and Rasmita Panigrahi, "Data Mining: A prediction Technique for the workers in the PR Department of Orissa (Block and Panchayat)", International Journal of Computer Science, Engineering and Information Technology (IJCEIT), Vol.2, No.5, October 2012.
- [3]. VSRD-IJCSIT, Vol.2 (3), 2012 pages no -210-222. "Data warehousing and its OLAP, MRDM tech. for Decision Support in Business Organization of 21st century", 2001.
- [4]. Kavya V1, Arumugam S2, "A review on predictive analytics in data mining", International Journal of Chaos, Modeling and simulation (IJCCMS), September 2016.
- [5]. [1] "Career Path Recommendation System for UiTM Perlis Students Using Fuzzy Logic", Tajul Rosli Razak, Muhamad Arif Hashim, IEEE 2014.
- [6]. [2] "Evaluation of Student Performance in Laboratory Applications using Fuzzy Decision Support System Model", Zehra Yıldız, A. Fevzi Baba, 2014 IEEE Global Engineering Education Conference (EDUCON), 3-5 April 2014.
- [7]. [3] "Advanced forecasting of career choices for college students based on campus big data", Min Nie, Lei Yang, Jun Sun, Han Su, Hu Xia, Defu Lian, Kai Yan, Higher Education Press and Springer-Verlag Berlin Heidelberg 2017.
- [8]. [4] "Automatically constructing grade membership functions of fuzzy rules for students' evaluation", Shih-Ming Bai, Shyi-Ming Chen, 2008, Pg:1408-1414.