Higher Education Prediction By Using Data Mining

Litesh Poojari[1], Mehta Priyanka[2] , Sandip Chavan[3]
[1, 2] Dept of Computer Engineering
[3] Asst-Prof, Dept of Computer Engineering
[1, 2, 3] Bharti VidyaPeeth C.O.E
Mumbai,India

Abstract- *Choosing an appropriate career is a standout amongst the most vital decisions and, as the number of careers and opportunities increases, this decision has become rather difficult for students..This can prompt a wrong career choice and therefore work in a field that has not been designed for them, thus reducing the productivity of the human resource. Therefore, it is very important to make a correct decision regarding the race at an appropriate age to avoid the consequences due to the wrong selection of the race.*

Keywords- C5.0Algoritm; Adaptive boosting technique; Data mining; Personality traits;dataData mining.

I. INTRODUCTION

ProblemPosed

With the increase in research and exploration in various sectors, there are many new career opportunities in all fields. It will create more confusion for students studying in the higher grade to select a career option. The reasons for this confusion could be ignorance of one's talent and personality, ignorance of the various options available, equal interests in multiple fields, less exposure, market , social life, social pressure. , the student may choose a wrong career course and the consequences of this bad decision can be job dissatisfaction, bad performance, anxiety and stress, social , etc.

II. AIM OF THE PROJECT

The framework would prescribe the understudy, a profession choice dependent on their identity characteristic, intrigue and their ability to take up the course.

III. SCOPE OF THE PROJECT

Choosing an appropriate career is a standout amongst the most vital decisions and, as the quantity of careers and opportunities increases, this decision hadbecome tough for the student. Through the survey that have been conducted by the Council for Scientiific and Industrial Reseearch (CSSR), about 39.9% mainof students are confused about career options. This can prompt a wrong career choice and therefore work in a

field that has not been designed for them, thus reducing the productivity of the human resource. Therefore, it is very important to make a correct decision regarding the race at an appropriate age to avoid the consequences due to the wrong selection of the race.This framework is a web application that will help the understudy to choose there carrer.

IV. REVIEW OF LITERATURE

Domain Explanation:

The number of students entering higher education in Ethiopia is growing rapidly from time to time. These students choose a field of study of their interest and the universities that wish to learn after taking the admission exam. Based on their field of interest and aftereffect of the admission exam, they were placed in different universities. And the universities place students in the departments based on their field of interest (choice) and aftereffect of the exam obtained during the admission exam to higher education. Students are placed as per distinctive criteriathat are prepared nationally for all universities. Students choose course and given numbers based on their order of choice. Students who get higher scores in the exam, women and developing regions have the priority of being assigned based on their first choice up to the department's capacity for acceptance, while other students are also placed according to their second choice. and third option. Student data thatwill be the main data setfor the intel needed that will be process and will be interpreted in different ways so that universitywill predict or classify. Classification is the most commonly applied data mining technique, which uses a series of pre-ordained attributes to develop a model capable of classifying the population of records in general. This approach often uses the decision tree or classification algorithms based on neural networks.

Related Work:

There are several websites and Web applications on the Internet that help students learn about the right career path. But most of these systems used only personality traits as the sole predictor of race, which could lead to an inconsistent response. Similarly, there are few sites that suggest a career

based solely on student interests. The systems didn't use student ability to knew if they will survive in that field or fail. The document is by

[1] Beth Dietz-Uhler and Janet E. Hurn suggests the importance of learning analysis to predict and improve student performance, which clarifies the importance of interest, skills, strengths, etc. . of your student.

[2] Another article by Roshani Ade, P. R. Deshmukh did suggest an incremental set of classifiers in which the hypothesis of the number of classifiers will be tested and using the "majority voting rule", the final results were determined. The proposed joint algorithm provided an accuracy of 90.8%.

[3]. Mustafa Agaoglu's paper suggested that importance of different attributes is in evaluating professor performance. Furthermore, the suggestions provided by the system are very generalized and are not specific to a university or country / state. The course suggestion is also generalized. For example, the results of some systems were a group of courses such as data analyst, accountant, law, etc.

Therefore, if a student receives a recommendation of this kind, he can be confused again because the course specified above belongs to different currents.

Hardware and Software Specifications:

This project makes extensive use of several hardware devices and software. These includes:

Computer with minimum @GHz quad core processor and 4 GB RAM with at-least 20 GB memory space free.

Visual Studio Eclipse.

Windows/Linux operating system with support to MySQL.

Methodology

Overview

Data extraction involves extracting patterns from data stored or stored in an organization. Algorithm C5 provides missing data and noise recognition. The problem of excessive regulation and elimination of errors is solved with the C5 algorithm. The classification technique of the C5.0 classifier areparticipated which attributes can be relevant and which will not be relevant to the classification. 4.1.0 Reinforcement of the C5.0 classifier allows classifier reinforcement to improve accuracy. The improvement is an AI meta-calculation to for the most part lessen predisposition, and furthermore a variety

in managed learning and a group of AI calculations that debilitate understudies. [7] AdaBoost is one of the motivation strategies. It is called versatile in light of the fact that it utilizes various emphases to create a solitary solid composite understudy. Amid each instructional meeting, another frail of understudy will be added to the set and a weighting vector are acclimated to concentrate on e.g. that were erroneously grouped in post rounds.**To accomplish the target of this investigation the accompanying exploration systems were utilized**.

Data Source The source of data used to undertake this study was taken from the assistance any institution. The data set contains about more than 100 instances of placed students. The data source that is been obtained from the assistant registrar office is recorded on exel. The information of resources chose for this examination are Faculty of Natural and Computational Science, Faculty of Agriculture, and Faculty of Engineering.

Data Preparation This comprises selection, that defines the target classes , also helps in handling noisy data,and if accounting for missing data field, and preparing the processed data in a file form acceptable to the software. The data of the students is including name, id, total entrance mark, the departments to be selected and gender of each student. From these all information the sexual orientation, absolute imprint request of divisions of the understudies' decision and the office in which every understudy set is chosen for the trial that will find there meeting process and the segment of the information into preparing and dataset, cross-approval test set are utilized.

Data Selection Some of the attributes in the initial dataset that were not relevant (relevant) with respect to the goal of the data mining experiment were ignored. The name of the attribute, the registration number and the faculty will not be used because of they had no data extraction value to classify the students into different departments. The main attributes used for this study are sex, total mark, biology, biotechnology, chemistry, geology, mathematics, physics, sports science, statistics, plant science, animal production, veterinary pharmacy, water resources, agricultural economy, natural resources, natural resources, natural resources, electricity, civil, mechanical, construction and employment.

Data Integration and Transformation The data that is been used for this study for the collected from the same data and they will have the same format and year. Therefore, there is none data forintegration and so the transformations techniquesis been used for integrate and Application of DataMining to Predict Students course. Be that as it may,

since it has taken from three unique resources the information of understudies of these two resources are consolidated.

Tools and Techniques The tools used in this study are the MS Excel and Weka software which are used respectively for the pre-processing and data classification algorithms. Weka is and is afimilar software for the machine learning project that are written in object oriented programing language Java, it is been developed at the University of Waeikato. Weka are free of cost. Therefore, Weka is an excellent data mining tool that will be used forproject in education system, since it will be used for different classification technique. Data mining classification techniques help classify data according to certain rules [3]. For the classification algorithms of this study, such as c5.0, an adaptive reinforcement was applied to discover the distribution of students in different departments.

V. CONCLUSION

The comparison between the use of C5.0 with adaptive reinforcement and C5.0 in the data set with personality, interest and capacity was shown above. This shows that to select a career it is not only a student's personality trait that is important, but also the student's interest and the student's ability to take those courses. Using this system, the student only has to answer the question that the system shows and, depending on the answers, the system recommends a particular course together with the list of universities that offer these courses. Therefore, the effort required to search even in universities
reduced.

VI. ACKNOWLEDGEMENT

After having endured the experience, there were many who helped us in our project and we would like to thank everyone.

We are deeply indebted to our dear Director, Dr. MZ Shaikh, and our Head of Department (DR), Dr. DR Ingle, for giving us this precious opportunity to complete this project and sincerely thank you for your help, without which it would be possible It was difficult to correctly terminate this synopsis report.

We also thank our guide, Professor Sandeep Chavan, for her kind help and advice during the work, and also for this opportunity to work on this project on "Predicting access to higher education through data mining", which also It has helped us to do a lot of research, we have learned so many new things that we are really grateful to everyone.

It is a great pleasure to thank the help and suggestions we received from the computer engineering department.

REFERENCES

[1] UD Beeth, HE Jaanet, using Analytics for learning to predict student : a faculty perspective. 2015 online interactive learning magazine.

[2] KS Lokeesh, RS Bhaktii, . Professional new forecasting method and career recommendation for individuals through the analysis of personal traits using the C4.5 algorithm. IEEE communication technology (GCCT) on 3 December 2015.

[3] To Roshani, PR Deshmukh, an incremental set of classifiers as in the technique for predicting the student's choice. IEEE Networks & Soft Computing (ICNSC) on September 25th 2015.

[4] A Mustafer, Predictor of the performance of the instructor through the technique of data mining in higher education. IEEE 2016; 4: 2379-2387.

[5] C Ling, R Dymitr, et al. Big Data: opportunities for Big Data analysis is good for the research and development of the project.

[6] M Yannick, X Jie, et al. Predict grades for the weak students in the data base using data mining.

[7] M Jiri, S Jan, AdaBoost. Machine Perception Center help in the analyising the big data , Czech Technical University, Prague.