IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 1, January 2019, ||V-4|| PP 07-09

Data Mining For Predicting and Suggesting Students Career

Mr.E.DILIPKUMAR[1] Ms.S.PRADEEPA[2]
[1]ASSISTANT PROFESSOR [2]PG STUDENT
DHANALAKSHMI SRINIVASAN COLLEGE OF ENGINEERING AND TECHNOLOGY, CHENNAI, TAMILNADU, INDIA

ABSTRACT: *Selecting an appropriate career is one of the most important decisions and with the increase in the number of career paths an d opportunities, making this decision have become quite difficult for the students. According to the survey conducted by the Council of Scientific and Industrial Research's (CSIR), about 40% of students are confused about their career options. This may lead to wrong career selection and then working in a field which was not meant for them, thus reducing the productivity of human resource. Therefore, it is quite important to take a right decision regarding the career at an appropriate age to prevent the consequences that results due to wrong career selection. This system is a web application that would help students studying in high schools to select a course for their career. The system would recommend the student, a career option based on their personality trait, interest a nd their capacity to take up the course.*

KEYWORDS: *Career prediction; Data mining; Personality traits; C5.0; Adaptive boosting.*

I. INTRODUCTION

With the increase in research and exploration in various domain, there are many new career opportunities in every field. This create more confusions to the students studying in tenth or twelfth grade to select one career option. The reasons for this confusion could be unawareness of self-talent and self- personality trait, unawareness of the various options available, equal interests in multiple fields, less exposure, market boom, assumed social life, peer-pressure etc. Due to theses confusions, the student may select a wrong career option and the consequences of this wrong decision could be work dissatisfaction, poor performance, anxiety and stress, social disregard etc. Thus, there should be proper counseling of the student‟s psychology, interest and their capacity to work in a particular field.

II. LITERATURE SURVEY

There are various websites and web apps over the internet which helps students to know their suitable career path. But most of those systems only used personality traits as the only factor to predict the career, which might result in an inconsistent answer. Similarly, there are few sites that suggest career based on only the interests of the students. The systems did not use the capacity of the students to know whether they would be able to survive in that field or not. The paper by [1]Beth Dietz-Uhler& Janet E. Hurn suggest the importance of learning analytics in predicting and improving the students performance which enlightens the importance of students interest, ability, strengths etc. in their performance. According to the paper by [2]Lokesh Katore, Bhakti Ratnaparkhi, Jayant Umale, the career prediction accuracy was

determined using 12 attributes of students and different classifiers with c4.5 having the highest accuracy of 86%. [3]Another paper by Roshani Ade, P.R.Deshmukh suggested incremental ensemble of classifiers in which the hypothesis from number of classifiers were experimented and by using „Majority voting rule‟, the final results was determined. The proposed ensemble algorithm gave an accuracy of 90.8%. [4]The paper by Mustafa Agaoglu suggested the importance of different attributes in evaluating the performance of faculty. It also showed the comparison of different classifiers proved that the most accurate classifier was c5.0 which has the maximum attribute usage compares to other classifiers like CART, ANN-Q2H, SVM etc. Also, the suggestions provided by the system are very much generalized and not specific to a university or country/state. The suggestion for course is also generalized. For example, the results of few systems were a group of courses like data analyst, accountant, law etc. Thus, if a student gets such a recommendation then he/she might again get confused as the above specified course belongs to different streams.

III. EXISTING SYSTEM

We perform the optimization by partitioning the dataset into a training and test dataset and following a standard validation procedure. During training, we keep k and λ fixed and optimize the remaining parameters of the model on the training dataset (80% of all data points). We then evaluate the performance

of the model on the test dataset (20% of all data points), by calculating the log-likelihood of the test data under the model produced during training. We repeat the procedure for arrange of values for k and λ to select an optimal configuration. A max-likelihood vector μ is computed once for each feature from the raw relative frequencies of observed values of that feature in the dataset. For fixed k and λ, the maximum-likelihood value of the remaining parameters can be computed with a standard expectation–maximization algorithm.

IV. PROPOSED SYSTEM

The project was to develop a web application that can be used by any student who needs help in selecting the career path. The system displays questionnaire to the students which the student will have to answer. The three set of questionnaires provided to the students based on personality traits, interests and capacity.

1. Personality traits:

A personality traits are characteristics that are distinct to an individual and are based on the psychology of a person. There have been many approaches to psychological traits theory but the one used in this project is the Myers-Briggs theory according to which there are four prominent pairs of personality characteristics which are: Introvert(I) vs Extrovert(E) Sensing(S) vs Intuitive(N) Thinking(T) vs Feeling(F) Judging(J) vs Perceiving(P) Based on these four pairs, a total of 16 types of personality types can be obtained by combining.

2. Interest:

Interest in this context implies that how much a student likes a subject and is keen to learn about it. Here the student will be first asked about the basic 3 streams i.e. Arts, Science and Commerce. The system then checks in which stream the student is more interested and then the further questions are comparison based questions which compare the subject of the selected stream and at the end determines one particular subject that the student is interested in.

3. Capacity:

Capacity implies that how efficiently a student can learn their interested subject and survive in that particular career path. For this purpose, the student will get questions that they had in their school curriculum and each question will have 4 options and a timer associated with it. Here the system asses not only the correctness of the answer but also the speed of the student to answer the question. This helps in knowing the memory, ability to solve and grasping capacity of the student. From the answers obtained, the system predicts a particular course for the student. The prediction is done using one of the Decision tree algorithms which is C5.0 on the personality traits of the student. To further enhance the accuracy of the algorithm Adaptive boosting was applied on the C5.0 algorithm and then C5.0 was applied on the dataset which included interests and capacity of the student also. The algorithm was implemented using the C5.0 package in R. From the rule-set and decision tree obtained, the personality combinations for various courses was made. Then the interest and capacity results were also integrated. Based on these combinations the web application was developed.

V. SYSTEM IMPLEMENTATION

Implementation is the process that actually yields the lowest-level system elements in the system hierarchy (system breakdown structure). The system elements are made, bought, or reused. Production involves the hardware fabrication processes of forming, removing, joining, and finishing; or the software realization processes of coding and testing; or the operational procedures development processes for operators' roles. If implementation involves a production process, a manufacturing system which uses the established technical and management processes may be required.

The purpose of the implementation process is to design and create (or fabricate) a system element conforming to that element‟s design properties and/or requirements. The element is constructed employing appropriate technologies and industry practices. This process bridges the system definition processes and the integration process.

System Implementation is the stage in the project where the theoretical design is turned into a working system. The most critical stage is achieving a successful system and in giving confidence on the new system for the user that it will work efficiently and effectively. The existing system was long time process.

The existing system caused long time transmission process but the system developed now has a very good user- friendly tool, which has a menu-based interface, graphical interface for the end user. After coding and testing, the project is to be installed on the necessary system. The executable file is to be created and loaded in the system. Again the code is tested in the installed system. Installing the developed code in system in the form of executable file is implementation.

It includes a code editor with features such as syntax highlighting, brace matching, and automatic indentation, and is also capable of compiling and uploading programs to the board with a single click. A program or code written for Arduino is called a sketch. Arduino programs are written in C or C++. The Arduino IDE comes with a software library called "Wiring" from the original Wiring project, which makes many common input/output operations much easier. Users only need define two functions to make a runnablecyclic executive program

The open-source Arduino Software (IDE) makes it easy to write code and upload it to the board. It runs on Windows, Mac OS X, and Linux. The environment is written in Java and based on Processing and other open-source software.

VI. CONCLUSION

The comparison of using C5.0 with adaptive boosting and C5.0 on dataset with personality, interest and capacity was shown before. This shows that for selecting a career not only the personality trait of a student is important, but also the interest of the student and the capacity of the student to take that courses is also important. Using this system, the student just needs to answer the question displayed by the system and based on the answers the system recommends a particular course along with the list of colleges providing those courses. Thus, the effort required to search the colleges will also reduce.

REFERENCES

[1]. Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban Computing: Concepts, Methodologies, and Applications," ACM Transaction on Intelligent Systems and Technology, vol. 5, no. 3, pp. 38:1–38:55, 2014.

[2]. "World Urbanization Prospects, the 2014 Revision: Highlights," United Nations, Department of Economic and Social Affairs, Population Division, New-York, Tech. Rep., 2014.

[3]. J. Eisenstein, A. Ahmed, and E. P. Xing, "Sparse Additive Genera- tive Models of Text," in ICML, Seattle, WA, 2011, pp. 1041–1048.

[4]. J. Cranshaw, J. I. Hong, and N. Sadeh, "The livehoods project: Utilizing social media to understand the dynamics of a city," in ICWSM, 2012, pp. 58–65.

[5]. A. X. Zhang, A. Noulas, S. Scellato, and C. Mascolo, "Hoodsquare: Modeling and recommending neighborhoods in location-based social networks," in ASE/IEEE SocialCom, 2013, pp. 69–74.

[6]. V. Frias-Martinez and E. Frias-Martinez, "Spectral clustering for sensing urban land use using Twitter activity," Engineering Applications of Artificial Intelligence, vol. 35, pp. 237–245, 2014.

[7]. G. Le Falher, Gionis Aristides, and M. Mathioudakis, "Where Is the Soho of Rome? Measures and Algorithms for Finding Similar Neighborhoods in Cities," in ICWSM, Oxford, 2015.

[8]. J. O. Berger and D. Sun, "Objective priors for the bivariate normal model," The Annals of Statistics, pp. 963–982, 2008. [9] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman, "Indexing by latent semantic analysis," JAsIs, vol. 41, no. 6, pp. 391–407, 1990.

[9]. N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, "Discovering Urban Functional Zones Using Latent Activity Tra- jectories," Knowledge and Data Engineering, IEEE Transactions on, vol. 27, no. 3, pp. 712–725, 2015. 2332-7790 (c) 2016 IEEE.

[10]. Jichang Zhao, Ruiwen Li, X. Liang, and K. Xu, "Segmentation and evolution of urban areas in beijing: A view from mobility data of massive individuals," in Proceedings of the 12th International Conference on Service Systems and Service Management (ICSSSM), 2015.

[11]. J. L. Toole, M. Ulm, M. C. Gonzalez, and D. Bauer, "Inferring Land ´ Use from Mobile Phone Activity," in UrbComp, New York, NY, USA, 2012, pp. 1–8.