2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS) | 978-1-6654-0521-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICACCS51430.2021.9441978

An Intelligent Career Guidance System using Machine Learning

Vignesh S, Shivani Priyanka C, Shree Manju H, Mythili K
Information Technology,
Sri Krishna College Of Technology
Kovaipudur, Coimbatore, India
17tuit152@skct.edu.in, 17tuit132@skct.edu.in, 17tuit134@skct.edu.in, k.mythili@skct.edu.in

Abstract*— Most of the students across the world are always in confusion after they complete higher secondary and the stage where they have to choose an appropriate career path. At the age of 18, the students don't have adequate maturity to accurately know about what an individual has to follow in order to choose a congenial career path. As we pass through the stages, we realize that every student undergoes a series of doubts or thought processes on what to pursue after 12th which is the single tallest question. Then comes the next agony whether they have essential skills for the stream they've chosen. Our computerized career counselling system is used to predict the suitable department for an individual based on their skills assessed by an objective test. If one completes their online assessment which we have created in our system, then automatically they will end up in choosing an appropriate course which will also reduce the failure rate by choosing a wrong career path.

***Keywords*—Career Counselling, Machine Learning Algorithms, K Nearest Neighbour, Classification**

I. INTRODUCTION

When it comes to choosing a career it's not only on what course you choose, it's more than what you want to become after your graduation. Career counselling is more about knowing and understanding about yourself and your capabilities and abilities. It is this time every student gets a lot of guidance from various circles (parents, teachers, other educational specialists, etc.) and accordingly the student decides about which course they want to join. Many times, we have come across a situation where a student opts for a course/stream and later repents for having chosen the one. To quote an example, there is a myth that one who does very well and scores highest marks in 12th grade chemistry will tend to choose chemical engineering because they are good in chemistry, however in reality that is not the case. We had multiple rounds of deliberation with students who are currently doing their engineering and students who are currently in 11th and 12th grade. Then we came up with an idea of providing an objective assessment of one's skill set and calibre that recommend a right stream to choose and hence we picked this as our problem statement and started thinking through how we can help the students in addressing this question.

As a first step we came up with broader skill sets which are strongly essential for each department in engineering such as Computer science and engineering, Electronics and Communication Engineering, Electrical and Electronics Engineering, Mechanical Engineering, etc. Depending upon one's mark in the objective assessment we have created, we will analyse their skill sets and predict which department is suitable for an individual. If one uses this functional chart to answer all these questions, the failure rate will drastically reduce in picking up the wrong choice. Our pointed questions will identify the core strength of the student's particular skill sets.

II. PROPOSED SYSTEM

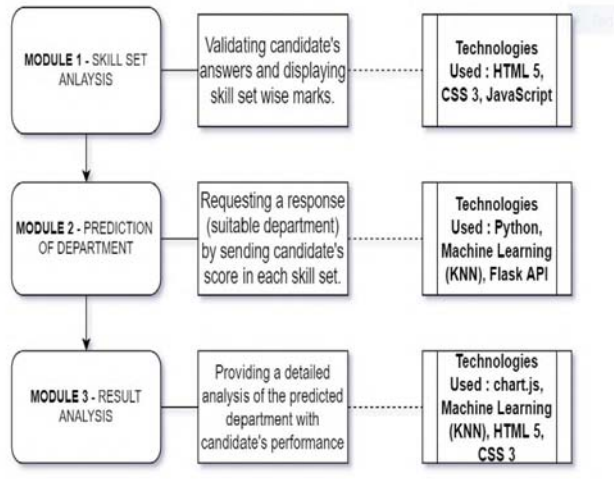

Fig. 1 Architecture Diagram of the recommender system

III. OVERVIEW OF THE SYSTEM

The framework basically analyses the skillset of the candidate which then will be used to map with the core skills of a specific department. The department which has its core skill where the candidate has scored more is then predicted and recommended to the candidate. It not only analyses with the help of core skills but also uses sub skills to map with the appropriate department. After the prediction part is done, the result analysis part will be present in the framework. In the result analysis part, the candidate can have a walkthrough of his overall performance in the skillset analysis part. The

978-1-6654-0521-8/21/$31.00 ©2021 IEEE

framework helps the candidate to recognise where he/she specifically is lacking sufficient skills, so that the candidate can improve those skills. The candidate can also view their performance skill wise in the skillset analysis part with various graphical representations in a dynamic nature. And the most important part in the result analysis part is that the system also recommends secondary and tertiary suitable departments according to the skillset analysis. With all the above-mentioned features, the students can easily assess themselves to know their strengths and weaknesses in specific areas of skill sets.

IV. MODULES

The framework totally consists of three modules where the whole process takes place. The first module is the skill set assessment module. In this module, the candidate takes up an assessment which will be having a combination of psychological and core skills-oriented questions.

At the end of the assessment, the candidate can know their scores in each and every skill set separately. The second module is the prediction module. In this module, with the help of the scores obtained by the candidate in the first module, the prediction takes place with the help of a machine learning algorithm running at the back-end of the web application.

The final result in the second module will be the prediction of the suitable department for the candidate. The third and final module is the result analysis module. In this module, a detailed analysis of the candidate's performance will be represented in various formats to provide a quality understanding for the candidate.

A. Skill Assessment Module

This module is designed and developed with the help of various web technologies such as HTML 5, CSS 3 and JavaScript. Hyper Text Markup Language 5 (HTML version 5) is a markup language used to design documents to get displayed in a web browser. It gives a skeletal structure to the document and so the document will be static if we only use HTML. Cascading Style Sheets 3 (CSS version 3) is a style sheet language which is especially used for adding styles and designs to the HTML document, so that the representation becomes better. JavaScript is a client-side scripting language which is mainly used for adding interactivity to the HTML web page. This makes the HTML page more of a dynamic page.

HTML 5 and CSS 3 plays a vital role in the front-end development and JavaScript plays a vital role in back-end development. Each and every question will be displayed separately with multiple choices. The validation part will be done with the help of JavaScript where each and every choice in a question will be having different weightage according to the best suitable answer. The validation will be done in a skill-wise manner where the final result will be displayed in a skill-wise manner respectively.

B. Prediction Module

The prediction module is the primary and core module among all the other modules. This module is built with various technologies such as machine learning algorithms, API and datasets for training the machine learning model. All the implementations are done with the help of python programming language. Python is a general-purpose programming language which can be used in almost every part of the implementation. The framework is predominantly developed with the help of python in the prediction module.

K-Nearest Neighbors is the machine learning algorithm used for classification purposes. K-Nearest Neighbors is a supervised machine learning algorithm used to classify the target values with the help of determining the distance between each neighbor using any formulae like Euclidean distance, Minkowsky, cosine similarity measure, chi square and correlation. K-Nearest Neighbor is well suited for classification problems. Even though there are many classification algorithms like Support Vector Machine, Random Forest, Naives Bayes etc., K-Nearest Neighbor is the one and only algorithm which has an accuracy of more than 90% with the dataset we have created of our own through various methodologies.

K-Means Clustering is the machine learning algorithm used for clustering purposes. K-Means Clustering is an unsupervised machine learning algorithm used to partition n observations into K clusters in which each value combines with the cluster with the nearest mean. In this framework, K-Means Clustering is specifically used to group the departments which are most appropriately mapped with the candidate's performance and to provide secondary and tertiary recommendations.

Flask API (Application Programming Interface) is used in the framework for having a communication between the front end and back end. An application programming interface is used to send requests and receive responses to and from the application. In our application, the scores obtained from the skillset assessment module is passed to the machine learning model via the flask API. The request in this scenario is to receive a suitable department with respect to the candidate's performance.

The response in this scenario is to provide the predicted department with respect to the candidate's performance. The implementation of this application programming interface part is done with the help of python.

The dataset used for the machine learning model is developed manually as there was no appropriate data available related to the core concept of this application. All the values present in the dataset are of only numerical values. The dataset basically contains more than 500 rows which means 500 unique values with several features and target variables. There are seven different features available in the dataset which consists of core skills and sub skills such as analytical skills, logical reasoning skills, mathematical skills, problem solving skills, programming skills, creativity skills and hardware skills. Among these seven skills, some are considered as core skills for a specific department while those same skills are also considered as sub skills for a different department. As this is a

multi-class problem, we have used multi class classification technique while developing the K-Nearest Neighbor model. Hence, there will be several target labels present in the dataset. In this dataset, there are five different target labels available each representing a specific department. For the machine learning model, 80% of the data present in the dataset is taken for training purposes and 20% of the data present in the dataset is taken for validation and testing purposes.

V. Experimental Results

A. Prediction results using classification

The performance of the machine learning model can be determined by a concept called confusion matrix. A confusion matrix is represented in the form of a table in which there will be four different values present in the matrix such as true positive, true negative, false positive and false negative. This can be used on the test dataset for which the true values are already known.

1. True positives (TP) are the value in which the examples are correctly determined as positive.
2. False positives (FP) are the value in which the examples are negative but are actually determined as positive.
3. True negatives (TN) are the value in which the examples are correctly determined as negative.
4. False negatives (FN) are the value in which the examples are positive but are actually determined as negative.

The confusion matrix determines the performance of the classification model by calculating precision, recall, accuracy, f-measure and error-rate. The following are the formula for each of the above-mentioned performance measures:

TP = TP/(TP+FN)
FP = FP/(FP+TN)
Precision = TP/(TP+FP)
Recall = TP/(TP+FN)
Accuracy = (TP+TN)/(TP+TN+FP+FN)
F-measure=2*(precision*recall)/(precision+recall)
Error- rate = 1-accuracy

TABLE I
PERFORMANCE MEASURES FOR CLASSIFICATION TECHNIQUES

Classification Technique	Accuracy	F - measure	Error - rate
KNN	0.9410	0.9213	0.01964
SVM	0.8632	0.9018	0.03154
Naive Bayes	0.8714	0.8835	0.06127

From the above inferred table, K-Nearest Neighbor algorithm is the classification technique which is determined as the efficient classifier for recommending suitable departments. The various F - measure values are mentioned below for various departments:

TABLE II
F-MEASURE FOR DEPARTMENTS

Classified Department	F - Measure
CSE (Computer Science Engineering)	0.9315
EEE (Electrical and Electronics Engineering)	0.9512
ECE (Electronics and Communication Engineering)	0.9637
MECH (Mechanical Engineering)	0.9849

B. Recommendation results using Clustering

The efficiency of the clustering can be determined by Within Cluster Sum of Squares (WCSS). WCSS can be calculated using the formula given below,

$$WCSS = \sum_{k=1}^{K} \sum_{i \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{kj})^2$$

While applying K-Means Clustering, there will be a maximum number of 5 iterations which in turn results with three different clusters. The success rate in each of the clusters is calculated and that will be used for department recommendation purposes where there will be a higher success rate and a low failure rate. The table below shows the success rate of each department:

TABLE III
SUCCESS RATE OF DEPARTMENTS

Department	Cluster 1	Cluster 2	Cluster 3
CSE	70.71 %	83.34 %	60.73%
EEE	82.65 %	64.71 %	71.43%
ECE	65.89 %	79.91 %	86.67%
MECH	78.15 %	89.12 %	67.85%

The table shows the presence of various success ratios in each department in each cluster. The higher the success rate, lower the chances of selecting a department by a candidate where he/she has higher chances of failure rate.

VI. Conclusion

In the system, we have designed and developed a web-based application for a career guidance system which provides suitable recommendations for a candidate in choosing an appropriate department. The recommendation provided in the proposed system is more accurate than the existing career guidance system. We have used the K-Nearest Neighbour algorithm to classify the skill sets of the candidate and predict a suitable department with respect to the performance of the candidate and we have also used K-Means Clustering algorithm and the clusters formed is by splitting the students' scores of the particular skill set and determining the rate of success for various departments in every cluster. The rate of success in each of the clusters is calculated and that will be used for department recommendation purposes where there will be a higher success rate and a low failure rate. In this project, the career guidance system has been researched thoroughly and then designed and developed a web-based application with expected outputs. In the near future, the framework's accuracy rate will be enhanced and additional features will be used for recommending a suitable department and also the outliers of the framework will be removed gradually.

References

[1]. A.M. El-Halee's, and M.M. Abu Tair, "Mining educational data to improve students' performance: A case study," International Journal of Information and Communication Technology Research, 2011, pp. 140-146.M. tech Er. Rimmy Chuchra. " Use of data mining techniques for the evaluation of student performance: a case study ". International Journal of Computer Science and Management Research, 1(3):425–433, October 2012.

[2]. Muhfuza Haque Md. Hedayetul Islam Shovon. " Prediction of student academic performance by an application of k- means clustering algorithm ". International Journal of Advanced Research in Computer Science and Software Engineering, 2(7):353–355, July2012.

[3]. S.Karthik M.Sukanya, S.Biruntha and T.Kalaikumaran, "Mining Data mining: Performance improvement in education sector using classification and clustering algorithm," ICCCE In Proceedings of the International Conference on Computing and Control Engineering, 2012.

[4]. Brijesh Kumar Bhardwaj and Saurabh Pal. Data mining:" A prediction for performance improvement using classification & (IJCSIS) International Journal of Computer Science and Information Security, 9(4), April,2011. [17]

[5]. H. Anandakumar and K. Umamaheswari, "An Efficient Optimized Handover in Cognitive Radio Networks using Cooperative Spectrum Sensing," Intelligent Automation & Soft Computing, pp. 1–8, Sep. 2017. doi:10.1080/10798587.2017.1364931

[6]. Haldorai, A. Ramu, and S. Murugan, "Social Aware Cognitive Radio Networks," Social Network Analytics for Contemporary Business Organizations, pp. 188–202. doi:10.4018/978-1-5225-5097-6.ch010

[7]. H. Anandakumar and K. Nisha, "Enhanced multicast cluster-based routing protocol for delay tolerant mobile networks," International Journal of Information and Communication Technology, vol. 7, no. 6, p. 676, 2015.

[8]. Md. Hedayetul Islam Shovon and Mahfuza Haque." An approach of improving student's academic performance by using k-means clustering algorithm and decision tree". (IJACSA) International Journal of Advanced Computer Science and Applications, 3(8):146–149, August,2012.

[9]. T. M. Mitchell. Machine Learning. McGraw-Hill Companies, New York, USA, 1997, ISBN 0-07-042807-7.

[10]. Jhan and M. Kamber.Data Mining. Concepts and Techniques, Simon Fraser University, Morgan Kaufmann publishers, ISBN 1-55860-489-8,2001.

[11]. ShiNa, Shi Na, Liu Xumin, and Guan Yong. Research on k-means clustering algorithm: An improved k-means clustering algorithm. In Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on, pages 63–67, 2010.

[12]. Powes, D.M.W,"Evaluation: From Precision, Recall and F-measure To Roc, Informedness, Markedness & Correlation " , Journal of Machine Learning Technologies, ISSN: 2229-3981 & ISSN: 2229-399X, Volume 2, Issue 1, pp-37- 63, 2011.

[13]. Holmes, Sr., O. W. (1858). The autocrat of the breakfast-table. Cambridge, MA: Houghton, Mifflfflfflin.

[14]. Horan, J. J. (2010). The Virtual Counseling Center: Its niche, resources, and ongoing research and development activity. Journal of Career Assessment, 18, 328–335.

[15]. Hornyak, D. A. (2007). Utilizing cognitive information processing theory to assess the effectiveness of DISCOVER on college students' career development. Dissertation Abstracts International Section A: Humanities and Social Sciences,68(6-A), 2319.Iaccarino, G. (2001).