A SURVEY ON INTELLIGENT CAREER GUIDANCE SYSTEM USING MACHINE LEARNING

Sneha H S[*1], Shreya R[*2], Sahana B C[*3], Soumya Sangalad[*4], Dr. Paramesha K[*5]

[*1,2,3,4]Student, Department Of Computer Science And Engineering, Vidyavardhaka College Of Engineering, Mysore, India.

[*5]Professor, Department Of Computer Science And Engineering, Vidyavardhaka College Of Engineering, Mysore, India.

ABSTRACT

After graduation, the majority of students around the world are always perplexed about which job route to pursue depending on their skills. Due to the lack of information about the different opportunities available they tend to end up in wrong sectors. Our computerized career counseling system is used to predict the suitable sector for an individual based on their skills. We consider various parameters to guide the students. Parameters such as Attendance status, extra circular activities, grade, technical skills, previous semester results, grasping capability, Aptitude grade, interaction with lecturers etc. are considered.

I. INTRODUCTION

When it comes to choosing our career path, it should be based on our interest and skill set that we possess. Knowing and understanding yourself, as well as your capabilities and abilities, is more important than ever in career counseling. 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 sector of job they want to join. Many times, we have come across a situation where a student opts for a job and later repents for having chosen the one. In our proposed system it is a "Web Enabled Application" which uses the Machine Learning technique for the extraction of useful information [prediction of student suitable career using classification rules]. People with similar characteristics and behaviors will tend to have similar preferences. Here they are sorted into groups and given advice based on their similarities. Proposed system recommends the career for the student based on the student characteristics and behavior. Proposed system is a tool for career advising.

II. LITERATURE SURVEY

One of the most pressing issues confronting higher education institutions today is how to increase student performance in leadership roles. [1]. One of the most effective ways to deal with quality improvement challenges is to provide new information related to educational processes and organizations in the management system. Information can be collected from operational and historical data on the educational organization's website using machine learning techniques. Data from prior students is stored in the system's usage database. Identification model is trained using this data and to assess the classification model. This paper introduces a recommendation program that predicts that students have one of the five conditions of placement, namely, Dream Company, Core Company, Mass Recruiters, Eligible and Interested in Placement. This model helps the placement cell within the organization to identify potential students and to focus on and improve their technical and interpersonal skills. Tech can also use this system to determine their individual placement status that they may benefit from. The system helps to improve the level of institutional placement as such serve as a key to improving the image of the institution. But this could be applied to only B. Tech students.

Exploratory Data Analysis (EDA) has provided visually analysis to extract the embedded depth [2].Job opportunities for graduates is what defines the success of any educational institution. Very good information can be obtained by all participants by carefully looking at this information. In this context, the process of The EDA procedure is used to assess student placement in a higher education institution. In this paper data about student placement is analyzed visually for producing hypotheses using mathematical models. EDA is the best way to find the distinct point from all other observations, but if not done right can lead us in the wrong direction. Random Forest is one of the good classification techniques and also it results in higher accuracy Decision Tree. Therefore for such classifications, Random Forest is recommended. Separation of databases and

database training are major aspects to get a good model. However, over fitting can be an issue in Random Forest. Trees may become unstable, and could result in a different tree even for small changes in data.

As Machine Learning (ML) algorithms became popular to solve complex and exciting real-world imaginative problems around us. [6]. this includes using well-known machine learning algorithms and testing them by solving simple guessing problems in the student community present in the education system. This paper proposes to solve the problem of predicting student placement using the linear regression model, K-neighbor regression model, decision tree regression model, XGBoost regression model, gradient boost regression model, light GBM regression model and random tree classifier model. This work is done in two phases. Phase 1 is done with a simple data set and Phase 2 is done with an extended data set with additional features about students. This research project examines the performance of these seven models in these two data sets and compares their results.

Using decision tree algorithms and clustering approaches, data mining was used to predict student placement. [7] Data mining is a design approach for testing data on non-changing search trends, and then confirm the findings using the patterns found in new sub-sets of data. The general policy of data mining is to extract the information from the database and convert it into a useful structure for continuous use. Here student performance is analyzed to predict the type of company with 95% accuracy and forecasting company name whether Microsoft, Samsung, Deloitte, etc. with 62% accuracy. The simplest method based on the decision tree algorithm is used to analyze student placement details of Thapar University Computer Department. This method will help the Placement Coordinator and Head of the Department in identifying a group of potential student's .Students will also be known what subjects they need to focus on for placements. This identification will help the Department and Placement Coordinator to design strategies to improve educational outcomes, planning skills and soft skills students. This model will play a critical part in improving the facility's overall installation.

The low pass rate of Science students has triggered an alert at South African Higher Education Institutions. [3]. The Admission Point Score (APS) that normally enrolls students seems to be well-versed in assessing students' ability, but the fact is that students who meet the APS requirements for the Science program have failed to meet the requirements for passing. This report seeks to build a recommendation engine that will inform students of their academic track record of the selected program. In addition, the debate was conducted on the method used to nominate students to be admitted to the program, on the grounds that they had no knowledge. However, finding the factors that influenced a student's future success was not an easy task. Attributes considered were Background, Individual and Pre-University categories. Six classification models from different paradigms were used in the data to predict which of the following three classes a student falls into: QualMin (eligible for a minimum of three years); Qualified and Failed. Models have proven that a combination of these values can predict a student's outcome within 67% - 70% accuracy. Overall, the Naive Bayes model provided outstanding accuracy, precision, and recall values. Interestingly, the feature of Pre-University had very little impact. However this model faced many limitations such as synthesis of data. It may not capture real-world and visual patterns that distort results, but it does allow mimicking the theoretical environment in which conceptual evidence can be constructed.

Analyzing Placement Test of Computer Science Students using Association Rule Mining [8] .Here FP-Growth algorithm was used for finding association rules of the questions in the placement test that associated with grades in a programming fundamental course. 174 student's details belonging to computer science branch were obtained to analyze. Basic knowledge or skills of students who get good grades were also analyzed. It was found that students who get the best grades have great skills and logical thinking. These skills may not be available to teach directly but can be developed through time and familiarity.

A job performance prediction model was built based on AQ (Adversity Quotient) level and career desire. A job performance is nothing but completion of work given by lecturers. Questionnaire of AQ Related Personality Traits and Job Performance Questionnaire was given to students of UTHM (University Tun Hussein Onn Malaysia) [5]. The response was analyzed using various aspects such as total, mean, frequency, percentage, and the Multiple Linear Regression test. Job performance is predicted by the variables Adversity Quotient (AQ) and two types of personality career desire, according to the Multiple Linear Regression analysis (Social(S) and

Entrepreneur(E)), that is Job Performance . Education Advisor and unit counseling can help and advice students to plan, as well as joining an appropriate intervention program to improve their AQ level.

According to studies, when rookie programmers are enrolled in an introductory course with students who have prior programming skills, they confront extra difficulties [9]. It is suggested to split an introductory programming course for novice and experienced students. This is known as the best practice because students could learn according to their expertise level. The introductory course at Michigan Tech has two starting points, one for newbie programmers and the other for students with prior programming experience. The separate programs for 2 types of students resulted in a positive impact for novice programmers.

With the advancement in internet technology, online job hunting plays an important role in job search [4]. For job hunters, relying entirely on keywords is tough. An item-based collaborative filtering algorithm is used to recommend jobs for the job seekers. Algorithm has been improved by combining position descriptions and resume information. Experimental tests on real dataset have shown that have a great improvement in the work recommendation results. This model largely outperformed the original algorithm. But this could not be applied to Campus Selection Process.

III.　METHODOLOGY

Machine Learning

Machine learning is a branch of artificial intelligence that allows machines to learn and develop without explicitly programmed. Machine learning is concerned with the design of software programs that can acquire data and learn for themselves.

Supervised Learning Technique

It is a speculative model used for tasks where it involves predicting one value using other values in a set of available data. Supervised reading will contain pre-defined labels. Separates an item based on one of the specified labels' parameters. K Nearest Neighbor, Naive Bayes, Decision Tree, ID3 algorithm, Random Forest Algorithm, Support Vector Machine, drop-down methods, and many other supervised learning algorithms are available. Depending on the requirements, labels, parameters and set of available data we select the appropriate prediction algorithm. When there is ambiguity in the data, algorithms are employed to develop a model that provides predictions based on the evidence. For this prediction project, we use the "Bayesian Classifier or K Nearest Neighbor algorithm" which is an effective and efficient algorithm for all different types of parameters, as it provides correct results.

IV.　FEATURE SELECTION

Bayesian Classifier Algorithm Steps

Step 1: Training datasets extracted from the database (SQL Server) and supervised learning algorithm applied for processing.
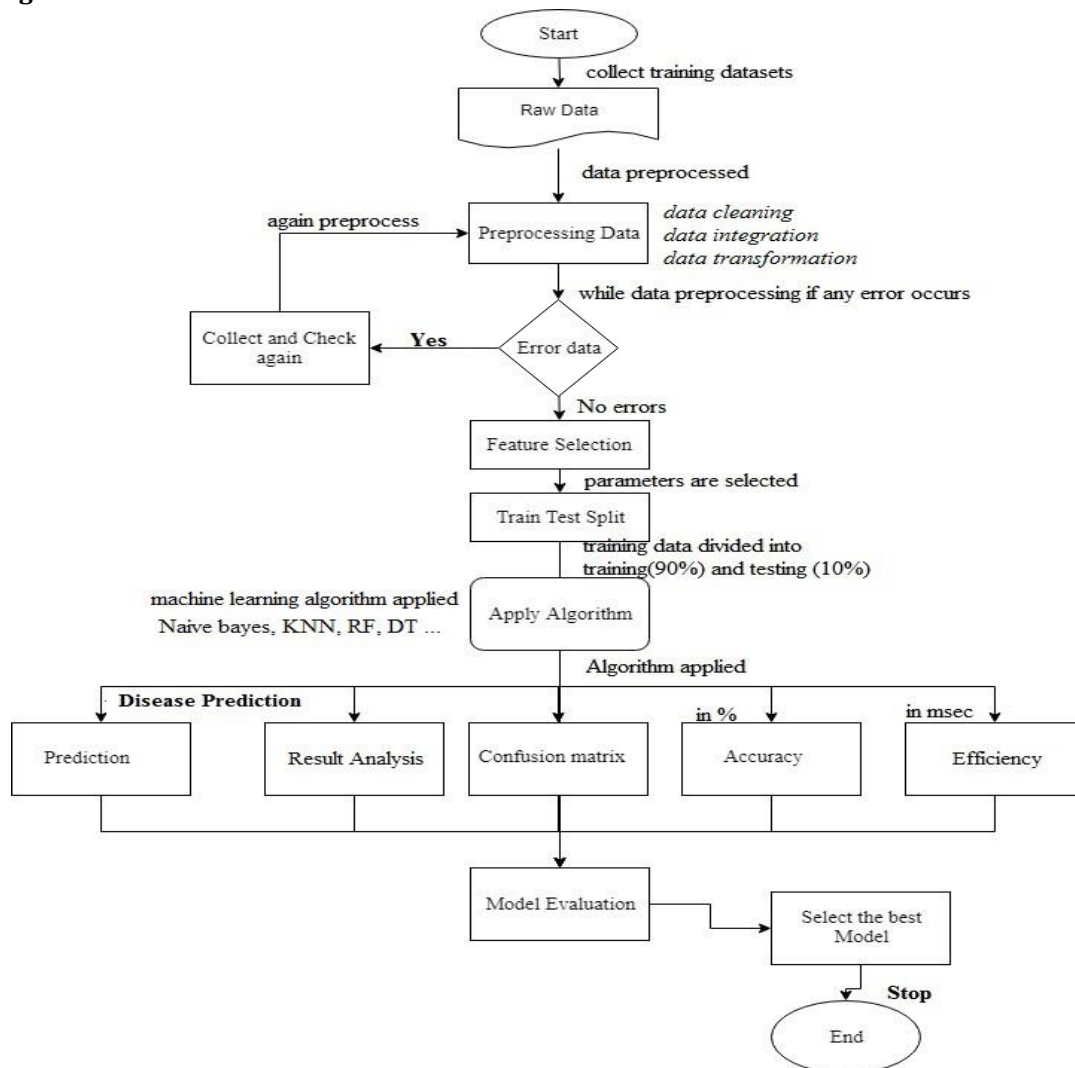
Step 2: Then we find the probability for all parameters used for crop yield prediction.

Step 3: We apply the formulae $[P = n\_c + (m * p)/ (n + m)]$

Step 4: Then probability values multiplied by p.

Step 5: Finally all possible outcomes probabilities are compared and highest probability is predicted as the final result.

Flow Diagram

V.   RESULTS AND DISCUSSION

Outcome of the literature review

Limitation:

- Accuracy rate is less
- Not applicable different branches
- Parameters considered are less
- Not suitable for real time

Solutions in the proposed system:

Our proposed solution is a real-time "Web Enabled Application" project. And it is applicable to different branches. Parameters such as Attendance status, extra circular activities, grade, technical skills, previous semester results, grasping capability, Aptitude grade interaction with lecturers etc. are all considered hence increases accuracy 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 Neighbor 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 success rate in each cluster is calculated, and the results will be utilized to make department recommendations, with a higher success rate and a lower 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.

VII.   REFERENCES

[1] Senthil kumar Thanga vel, Divya Bharath P, Abijith Sankar (2017), "Student Placements Analyzer: A Recommendation System using machine learning"

[2] Jumuna Nagaria, Senthil Velan Sc (2020), "Utilization Exploratory Data Analysis for the Predition of campus placements for educational institutions"

[3] Tasneem Abed, Ritesh Ajoodha, Ashwini Jadhav (2020), "A Prediction on model to improve student placements at a south African higher education on institution"

[4] Peng Yi, Cheng yang, Chen Li, Yingya Zhang (2016),"A Job Recommendation method optimization by position descriptions and resume information"

[5] M. F. Lee, N. F. Mat Nawi & C. S. Lai (2017), "Engineering students job performance predition model based on adversity quotient & career interest "

[6] Tadi Aravind, Bhimavarapu Sasidhar Reddy, Sai Avinash, Jeyakumar G (2019), "A comparative study on machine learning algorithm for predicting the placements information of under graduation student."

[7] Karan Pruthi, Dr. Parteek Bhatia (2015),"Application of data mining in prediting placements of students"

[8] Peerasak Pianprasit Parinya Seesai and Sunisa Rimcharoen (2017), "Association rule minning for Analysing Placements test of computer science student"

[9] Glen Archer, Briana Bettin, Leonard Bohmann, Allison Carter, Christopher Cischke, Linda M.Ott, leo Ureel (2017), "The Impact of Placements Strategies on the success of student in introductory computer science".