Career Prediction System

V Madhan Mohan Reddy[1]

[1]Department of Computer Science and Engineering, Madanapalle Institute of Technology and Science, Andhra Pradesh, India

ABSTRACT

In today's world Choosing an appropriate career path is one of the most important decisions and with the increase in the number of career options and opportunities, making this decision even more difficult for the students. Different people suggest different career paths but at last, the student must select their career. 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 the students to wrong career selection and then working in an area which was not meant for them, this leads their career in wrong path, and this may not be good for their career as well as for the country. Therefore, it is quite important to take a right decision regarding the future career. This system would help students studying in under graduation or post- graduation to select a correct career path for their career. The system would recommend the student, a career option based on their skills in various fields and areas. This career prediction system mainly concentrates on the career prediction of computer science and engineering domain candidates. This system will help them in deciding in which job role the candidate should be fit in.

Keywords : Council of Scientific and Industrial Research, post-graduation, under graduation

I. INTRODUCTION

With the gradually increasing technology and research in various areas and fields, there are many new career options in every field. This creates more confusion to the students studying under graduation to select one career option. The reasons for this confusion could be unawareness of self- talent and self-personality, unawareness of the various options available etc. Because of these confusions, the student may select a wrong career path and the results of this wrong decision could be work dissatisfaction, poor performance, anxiety and mental stress etc.

Career Prediction System using Machine Learning is developed for Engineering Graduating Students who

are pursuing their final year (Computer Science), who are confused regarding which field/path to choose for their career. As there are already many options available to choose Careers, still taking the correct decision is a challenge. So, we have considered all the aspects which are important to choose a Career, aspects considered in our project are gradings in essential subjects/fields and interested course, which are very important for making any decision on Career. In our system we will take gradings in essential fields as input. We will give this input data to the system which was already trained based on the data set. Based on the input it will predict the best suitable career along with clear road map. There are many types of career paths like Database administrator, Business Process Analyst, Software Testing, UI Developer, Testing Manager, Networks Analyst Manager, Data scientist engineer and so on. All these roles require some prerequisite knowledge in required areas/fields. As the input parameters and final classes/outcomes of output as bulk in number typical programming and normal algorithms cannot produce the best possible output classification and prediction. So advanced machine learning algorithms like SVM, Random Forest decision tree, XG boost are used.

II. EXISTING SYSTEM

There are certain models available in the market for the same problem, every model has its own advantage and disadvantage. Some of the existing systems are using these methods to take input.

2.1 Prediction using Course grades
Taking student academic course grades[1,2] as input may not be efficient way as the individual student grades may depends on college, exam models, paper evaluation so on.

2.2 Prediction using YES/NO questionnaire
Asking questionnaire from various required fields and getting an answer in the form of yes/no may lead the

student in confuse. Suppose in a particular course student may had only basic knowledge then he would not know whether to give yes/no. so, in this case he wouldn't provide input, and this would lead to wrong prediction.

2.3 Not providing clear road map to the predicting career
Predicting career without showing clear road map how to proceed may put student in dilemma. Prediction with clear road map makes student clear about his/her career.

III. IMPLEMENTATION

3.1 Data Flow Diagram
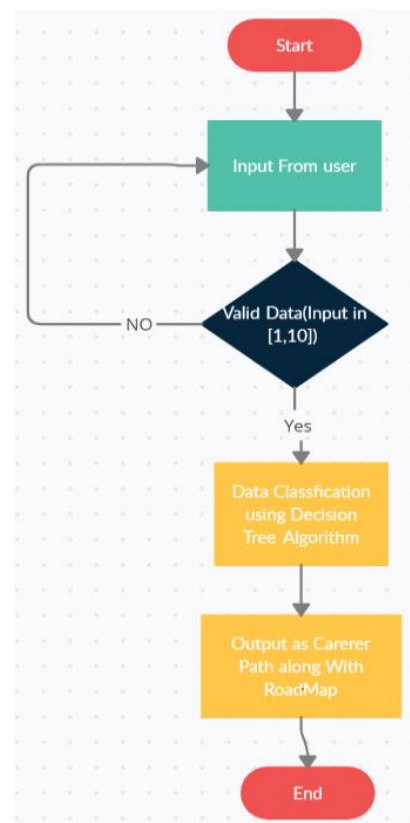
Data Collection:
Data Collection is one of the major and most important tasks of any machine learning projects. Because the input we collect from the users as data is going to provide to the algorithms to generate appropriate output. So, the algorithms accuracy and efficiency

depend upon the correctness and quality of data collected. For student career prediction many parameters are required like students' knowledge in various subjects, specializations, programming and analytical capabilities, hackathons, workshops, certifications, interested course and many more. As all these factors play vital role in predicting correct career. To train the model the data set has been taken with over 30 thousand records with including all the fields like subjects, programming, interested course and many more.

For input data is collected in the form of ratings in various required fields of computer science student. Asking to rate himself/herself in required fields is the correct way to get exact knowledge of students.


3.2 Data Pre-processing:

Making the collected data useful is very vital task in any project of machine learning. Data collected may contain some garbage values and empty values. It may contain some unwanted data as well. All such cases should be checked properly and replaced with suitable and alternate values to make data more meaningful and useful for further processing. Data must be kept in an organized format in order to train the model effectively. To train the model we are using machine learning algorithms like XG Boost and decision tree.

### 3.3 Machine Learning Algorithms

3.3.1 XG Boost

XGBoost is an implementation of gradient boosted decision trees designed to improve execution speed and model performance of a system. XGBoost stands for eXtreme Gradient Boosting. It belongs to a broader collection of tools under the umbrella of the Distributed Machine Learning Community or DMLC. XGBoost implementation has the features of scikit-learn and R implementations and have a newly added features like regularization. The main best features that the implementation of the algorithm provides: Automatic handling of missing values with sparse aware implementation, the fresh data. Gradient boosting is a technique where new models are made that can predict the errors/mistakes or remains of previous models of system and then added together to make the final prediction. In the training part generally, an objective function is defined.
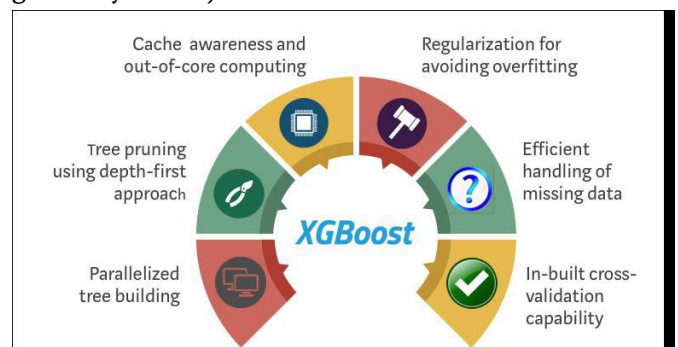
3.3.2 Decision Tree:

Decision tree is a Supervised learning technique and the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart which looks like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label. Decision trees laid foundation for many advanced algorithms like bagging, gradient boosting, and random forest. The XG Boost algorithm discussed above is the advanced version of decision tree.

The scenario that a decision tree follows is first selecting a root node. Calculate information gain or entropy for each of the nodes before the split. Select the node that has more information gain or less entropy. Further split the node and reiterate the process. The process is iterated until there is no possibility to split or the entropy is minimum. Entropy is the metric to measure uncertainty or randomness in the information being processed. The higher the entropy value, the harder it is to find any conclusions from that information. Information gain is the metric that measures how much entropy is reduced before to after split.

Note:- A is parent node of B and C.
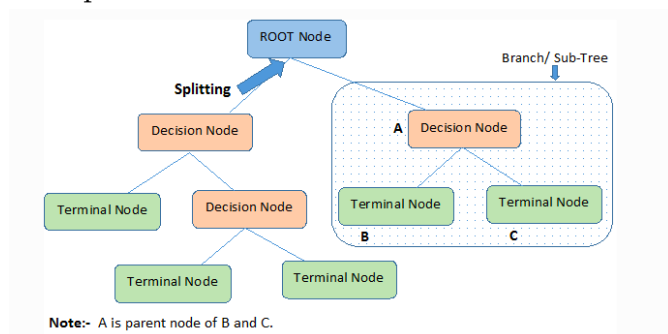
Entropy

$$E(S) = \sum_{x \in X} -p(x) \log_2 p(x)$$

Information Gain

$$IG(A, S) = E(S) - \sum_{t \in T} p(t)E(t)$$

Training the machine is the most important task in this project. The better the machine trains the better the results it produces. Training is where performance of the algorithm, quality of data, and required output all appears out. Training is the process of making the machine to learn and giving it the capability to make further predictions based on the training it took. Whereas testing means already having a predefined data set with output also previously labelled and the model is tested whether it is working fine or not and is giving the right prediction or not. If maximum number of predictions are correct, then model will have a good accuracy percentage and is reliable to continue with otherwise train the machine in better way with

appropriate data set with better algorithm. Further new set of inputs and the predictions made by the model will be keep on adding to the dataset which makes dataset more powerful and accurate.

IV. Result Analysis

The machine is trained and tested with different data with two algorithms. The developed system takes input from the user, calculates the output using Decision Tree Algorithm and again displays the results back to the user. The displayed result shows the clear road map of resultant career. The user needs to follow the road map shown in the result to achieve better career in that area. In the road map user can find courses and tasks he/she needs to learn and complete to excel in that career.
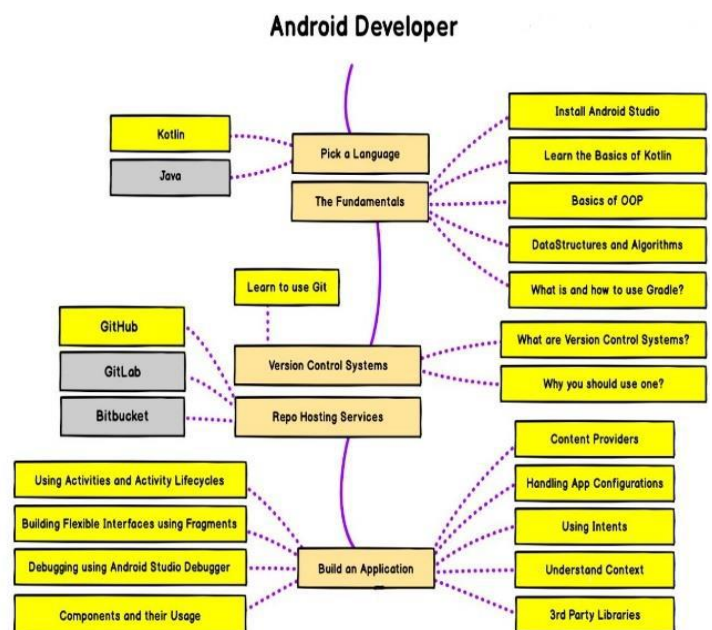
V. Conclusion

Career Prediction System developed using machine learning is very effective in predicting correct career based on the student skills possessed in required fields. The result shows the correct career along with clear road map which directs the student to excel in that career. Students needs to follow the courses and tasks in the road map to secure better future in that career.

VI. Future Scope

Today's education system is developing gradually to cope with that student needs to change their learning process to gain the knowledge. With the revolutionary changes occurring in Engineering Fields, numerous Engineering Fields are available, we will expand our Scope by providing Guidance to other Fields such as Mechanical, Electrical, Civil etc. including CS/IT. Chatbot will be used for interaction with human beings. Additional materials and course related latest videos could also provide to the students.

VII. REFERENCES

[1]. D. M. D. Angeline, Association rule generation for student performance analysis using apriori algorithm, The SIJ Transactions on Computer Science Engineering & its Applications (CSEA) 1 (1) (2013) p12–16.

[2]. M. M. Quadri, N. Kalyankar, drop out feature of student data for academic performance using decision tree techniques, Global Journal of Computer Science and Technology 10 (2).

[3]. Khasanah, Annisa Uswatun. "A review of student's performance prediction using educational data mining techniques." Journal of Engineering and applied sciences 13: 5302-5307.

[4]. OTOBO Firstman Noah, BAAH Barida and Taylor Onate Egerton, "Evaluation of student performance using data mining over a given data space", International Journal of Recent Technology and Engineering (2013):4

Cite this article as :