IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 8, Issue 11, November 2021

DOI: 10.17148/IARJSET.2021.81142

CAREER COUNSELING SYSTEM USING MACHINE LEARNING

Swasti Bhutani

Department of Information Technology, Maharaja Agrasen Institute of Technology, Delhi, India

Abstract: As more and more opportunities are originating in today's technical world engineering students are getting more options to choose their field of interest from. Hence, it becomes very important for them to be aware of their interests and capabilities at early stages of their career. This will help them start early, and target their efforts in improving their performance. They can assess themselves on all grounds and work on their weaknesses. Even the recruiters assess students in these aspects before recruiting them for a particular job role. This will help both candidates and recruiters to analyze and evaluate the candidate's performance in various areas and suggest the best job profile for him.This paper mainly concentrates on the career area prediction of computer science domain candidates.

keywords- Career Counseling, SVM, Decision Tree, Feature Selection, XG Boost

INTRODUCTION

In today's extremely ambitious technical world, students need to be ahead of their times to stay competitive. They need to have the forethought to plan their career. Organizing and designing their career path in early stages of their education can help them make targeted efforts towards their goals. Hence it becomes imperative to consistently evaluate their performance, identify their interests and analyze how close they are to their goals. This benefits them in improving themselves, motivating themselves to a better career path if their capabilities are not up to standard to reach their goal and evaluate themselves before going to the career peak point. Even the recruiters while recruiting candidates into their companies evaluate candidates on various parameters and draw a final conclusion to select an employee or not and if selected, finds a best suited role and career area for him. Some of the many types of roles for which recruiters look for candidates are Database administrator, Business Process Analyst, Developer, Testing Manager, Networks Manager, Data scientist and so on. All these roles require some essential knowledge to be placed in them. Recruiters analyze these skills, talents and interests and place the candidate in the right job role suited for them. Various third party performance evaluation portals like Co-Cubes, AMCAT are already using these career recommendation systems too. They only take into account factors like technical abilities and psychometry of students. This career prediction system also considers students' abilities in sports, academics and their hobbies, interests, competitions, skills and knowledge. After evaluating all the factors the total number of parameters that were considered as inputs are 36, and there are 15 job roles. As the input parameters and final classes of output are large in number, advanced machine learning algorithms like SVM, Random Forest decision tree, OneHot encoding, XG boost are used.

Figure 1: Overview of various Advanced Machine Learning Algorithms

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 8, Issue 11, November 2021

DOI: 10.17148/IARJSET.2021.81142

Machine Learning as the name suggests is about training machines to learn the given inputs and respond to new inputs or scenarios based on their knowledge of previous information automatically without explicitly providing any programme or human intervention. Computers are given the ability to learn and make decisions using statistical techniques. This aims at reducing human involvement in machine dependable problems and scenarios. It solves complex problems with ease and negligible human intervention. NLP, classification, prediction, image recognition, medical diagnosis, algorithm building, self-driving cars are various applications of machine learning. In this project classification and prediction algorithms are used. Majority of problems in machine learning can be solved using supervised and unsupervised learning. If the final class labels are previously known and all the other data items are to be assigned with one of the available class labels, then it is called supervised. And if the final output classes and sets are not known and it is done by identifying the similarity between data point and their characteristics and finally they are made into groups based on these characteristics then it is called unsupervised. Classification algorithms are supervised. Based on the properties of the provided input parameters a predefined class label is assigned to them. There are other alternatives like clustering and regression. After assessing the type of the problem the apt model is chosen. In this project algorithms like SVM, OneHot encoding, Decision tree and XG boost are used. After the data is trained and tested, most accurate results given by the algorithms used After training and testing the data with these we take into consideration the most accurate results given algorithm for our further processing. So, the initial task done is predicting the output using all algorithms proposed above and later analyzing the results and there on continued with the most accurate algorithm. So finally, this paper deals with various advanced machine learning algorithms that involve classification and prediction and are used to improve the accuracy for better prediction, reliability and analyzing these algorithms performance.

2. IMPLEMENTATION

2.1 Data Collection:

Collection of data is one of the major and most important tasks of any machine learning project. Because the input we feed to the algorithms is data. So, the algorithm's efficiency and accuracy depends upon the correctness and quality of data collected. So the data will be the output. For student career prediction many parameters are required like students academic scores in various subjects, specializations, programming and analytical capabilities, memory, personal details like relationship, interests, sports, competitions, hackathons, workshops, certifications, books interested and many more. As all these factors play a vital role in deciding a student's progress towards a career area, all these are taken into consideration. Data is collected in many ways. Some data is collected from employees working in different organizations, some amount of data is collected through LinkedIn api, some amount of data is randomly generated and other from college alumni databases. Totally nearly 20 thousand records with 36 columns of data are collected.

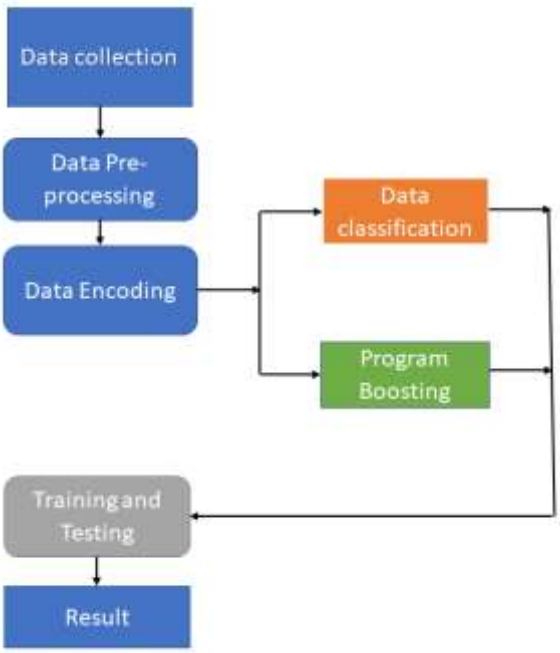

Figure 2: Process Flow Diagram of proposed system

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 8, Issue 11, November 2021

DOI: 10.17148/IARJSET.2021.81142

2.2 Data Pre-processing:

Collecting the data is one task and making that data useful is another vital task. Data collected from various means will be in an unorganized format and there may be a lot of null values, invalid data values and unwanted data. Cleaning all these data and replacing them with appropriate or approximate data and removing null and missing data and replacing them with some fixed alternate values are the basic steps in pre-processing of data. Even data collected may contain completely garbage values. It may not be in the exact format or way that is meant to be. All such cases must be verified and replaced with alternate values to make data meaningful and useful for further processing. Data must be kept in an organized format.

2.3 OneHot Encoding:

OneHot Encoding is a technique by which categorical values present in the data collected are converted into numerical or other ordinal format so that they can be provided to machine learning algorithms and get better results of prediction. Simply OneHot encoding transforms categorical values into a form that best fits as input to feed to various machine learning algorithms. This algorithm works fine with almost all machine learning algorithms. Few algorithms like random forest handle categorical values very well. In such cases OneHot encoding is not required.

Process of OneHot encoding may seem difficult but most modern day machine learning algorithms take care of that. The process is easily explained here: For example in a data if there are values like yes and no., integer encoder assigns values to them like 1 and 0.This process can be followed as long as we continue the fixed values for yes as 1 and no as 0. As long as we assign or allocate these fixed numbers to these particular labels this is called integer encoding. But here consistency is very important because if we invert the encoding later, we should get back the labels correctly from those integer values espe-cially in the case of prediction. Next step is creating a vector for each integer value. Let us suppose this vector is binary and has a length of 2 for the two possible integer values. The 'yes' label encoded as 1 will then be represented with vector [1,1] where the zeroth index is given the value 1. Similarly 'no' label encoded as '0' will be represented like [0,0] which represents the first index with value 0.

For example [pillow, rat, fight, rat] becomes [0,1,2,1]. This is here imparting an ordinal property to the variable , i.e. pillow < ret < fight. As this is ordinally characteristic and is usually not required and desired and so OneHot encoding is required for correct representation of distinct elements of a variable. It makes representation of categorical variables to be more expressive.

2.4 Feature Selection:

Only a few variables from the wide dataset are required to build a machine learning model and the rest of the data is redundant. Feature selection is a way of selecting the subset of the most relevant features from the original features set by removing the redundant, irrelevant, or noisy features. There are three types of feature selection methods: filter method, wrapper method, embedded method. The filter method selects features on the basis of statistics measures. Chi square test is one of the filter method techniques used to determine the relationship between various features and the target variable by calculating the chi square value between them. The formula for calculating chi square value is given as:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

where χ 2 = Chi-Square value, Oi = Observed frequency, Ei = Expected frequency

A bar chart of the feature importance scores for each input feature is created, and features with lower importance scores are removed.

RESULT:

We have plotted a bar chart showing the feature importance scores, and eliminated the features of less importance. The important features will be further used to predict the best possible job profile for the student. We will be using complex machine learning algorithms like SVM and Decision tree to come to the final conclusion. The existing data will be classified, trained and tested and new inputs would also be taken.

REFERENCES:

[1] P.KaviPriya, "A Review on Predicting Students' Academic Performance Earlier, Using Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering

[2] Ali Daud, Naif Radi Aljohani, "Predicting Student Performance using Advanced Learning Analytics", 2017 International World Wide Web Conference Committee (IW3C2).

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 8, Issue 11, November 2021

DOI: 10.17148/IARJSET.2021.81142

[3] Marium-E-Jannat,SaymaSultana,Munira Akther, "A Probabilistic Machine Learning Approach for Eligible Candidate Selection", International Journal of Computer Applications (0975 – 8887)Volume 144 – No.10, June 2016

[4] Sudheep Elayidom, Dr. Sumam Mary Idikkula, "Applying Data mining using Statistical Techniques for Career Selection", International Journal of Recent Trends in Engineering, Vol. 1, No. 1, May 2009.

[5] Dr. Mahendra Tiwari ,Manmohan Mishra, "Accuracy Estimation of Classification Algorithms with DEMP Model", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 11, November 2013.

[6] Ms. Roshani Ade,Dr. P. R. Deshmukh, "An incremental ensemble of classifiers as a technique for prediction of student's career choice", 2014 First International Conference on Networks & Soft Computing

[7] Nikita Gorad ,Ishani Zalte, "Career Counselling Using Data Mining", International Journal of Innovative Research in Computer and Communication Engineering.

[8] Bo Guo , Rui Zhang, "Predicting Students Performance in Educational Data Mining",2015 International Symposium on Educational Technology

