



# Predicting Employability of a Student In R Programming

V.Manjula<sup>[1]</sup>, M.Devi Uma Sankari<sup>[2]</sup>, A.Periya Nayaki<sup>[3]</sup> R.Saranya<sup>[4]</sup>  
UG Students, Department of Computer Science and Engineering<sup>1,2</sup>,  
Assistant Professor, Department of Computer Science and Engineering<sup>3,4</sup>,  
Kamaraj College of Engineering and Technology, Virudhunagar, Tamilnadu.

*Abstract--* The students employability is a major concern for the institutions offering higher education and a method for early prediction of employability of the students is always desirable to take timely action. The company invests huge time & money (budget) on training fresh The engineering graduate to make them ready for work. Various classification techniques of data mining, like Bayesian methods, Multilayer Perceptrons and Sequential Minimal Optimization(SMO), Ensemble Methods and Decision Trees are available to predict the employability of Bachelor of Engineering (BE). The Random Forest algorithm which is best suited for predicting was found to be accurate rather than Naïve Bayes and Decision Tree . The data set is developed with the traditional parameters like academic performance along with AMCAT scores. Analysis was carried out to check impact of input variables like Aptitude, Communication, Technical & Logical factors which is obtained from the AMCAT scores. Threshold cutoff point is set for the grades in order to identify under which category of job the candidate belongs. A model was developed which predicts probability of Employability of engineering graduate students in campus placement based on skill set required in IT Sector. Personality and professional skill have a significant impact on college graduate career decision making.

*Index terms- Datamining, Optimization, Employability, Descision Tree, Random Forest.*

## I. INTRODUCTION

The big data revolution happening in and around 21st century has found a resonance with employment in companies. Big data activities reveals that they start with a strong core of analytics capabilities designed to address structured data, such

as basic queries, predictive modeling optimization and simulations.

India's substantial growth in recent years and globalization has resulted in a significant

increase in demand for technocrats. To meet requirements, All India Council of Technical Education (AICTE), New Delhi has approved new technical institutes as well as second shift within the same infrastructure & also diluted eligibility criteria of admission. It has been observed that the growth of Engineering and Technology educational Institutes increase in ten folds in India to fulfill requirement of future technocrats.

Educational institutions generate and collect huge amount of data. This may include students' academic records, their personal profile, observations of their behavior, their web log activities and also faculty profile. This large data set is basically a storehouse of information and must be explored to have a strategic edge among the Educational Organizations.

The classification experiment is performed using various Bayes algorithms like Decision Tree algorithm and Random Forest Algorithm to determine whether a graduate can be employed in a particular field or remains unemployed in an undetermined situation.

These methods helps to analyze the characteristics of graduates and match the characteristics with the occupations. Recently, several researchers have applied the latest intelligent algorithms and data mining methods to study the graduate career decision.

The issue of employability of Engineering Graduates is not only the concern of Technical Institutes but also the Industries and the Government at large. Interaction between Institute and the Industry is now widely recognized as an essential requirement to train and develop the right kind of



Technical manpower necessary to sustain and promote Industrial and Economic growth.

Predicting student employability can help identify the students who are at risk of unemployment and thus management can intervene timely and take essential steps to train the students to improve their performance

## II. LITERATURE SURVEY

[1]. Different classification algorithms are used to classify the students data. They are using the Random Decision Tree (RDT) algorithm, Naïve Bayes and Random Forest algorithm Educational data mining is a new emerging technique of data mining that can be applied on the data related to the field of education. It uses many techniques such as decision trees, neural networks, naive bayes, K-Nearest neighbour and many others. Using these techniques different kinds of knowledge can be discovered using association rules, classification and clustering. By using this we extract knowledge that describes students' performance in the end of the semester examination and all their details.

[2] Bennett et al. (1999) proposed a model of course provision in higher education which included five elements: disciplinary content knowledge ;disciplinary skills workplace awareness; workplace experience; and generic skills. This model goes some way towards including all the necessary elements to ensure a graduate achieves an optimum level of employability, but is still missing some vital elements .The USEM account of employability (Yorke and Knight, 2004; Yorke and Knight,2004) is probably the most well known and respected model in this field. USEM is an acronym for four inter-related components of employability: understanding ;skills; efficacy beliefs; and metacognition

Classification, applied to form the decision tree, is a new approach of data mining to find the hidden relations of the data. The decision tree can be given by using data mining to dig the information received from searching. The classical calculation method in data mining is the ID3 method which is based on dividing the minimum attributes in the information theory. The method is mainly as the described by Eesa et al.[3]

To construct the classifiers, we use the Waikato Environment for Knowledge Analysis (WEKA), an open-source data mining tool [12] which was developed at University of Waikato New Zealand. It provides various learning algorithm that can be easily implemented to the dataset. WEKA only accepts dataset in Attribute-Relation File Format (ARFF) format. Therefore, once the data preparation being done, we transform the dataset into ARFF file with extension of *.arff*.

## III. DATA MINING TECHNIQUES

Different types of data mining techniques are used to predict the student's employability. The data mining techniques are:

- Classification
- Regression Methods

Classification is a process of identifying the category of objects based on their characteristics. For example, we can use classification models to classify the skills of the students and whether they are suitable for a particular role. Random Forest, Naïve Bayes, Decision tree and many other algorithms are used in classification models.

Regression methods are discussed in section IV. Association Rule mining is a data mining technique which is used find the pattern in the dataset by using the frequency of relationship between data.

## IV. METHODOLOGY

A decision tree uses the traditional tree structure from your second year data structures module.It starts with a single root node that splits into multiple branches, leading to further nodes, each of which may further split or else terminate as a leaf node.Associated with each nonleaf node will be a test that determines which branch to follow.The leaf nodes contain the decisions.

### Decision tree topologies

There are variations to the basic decision tree structure for representing knowledge.



Some approaches limit trees to two splits at any one node to generate a binary decision tree. In the decision tree on the previous slide the decision variables are real valued and one real number is used to generate the decision split. For categoric data a binary decision may involve partitioning the levels of the variable into two groups. Another approach is to have a branch corresponding to each of the levels of a categoric variable to generate a multiway tree.

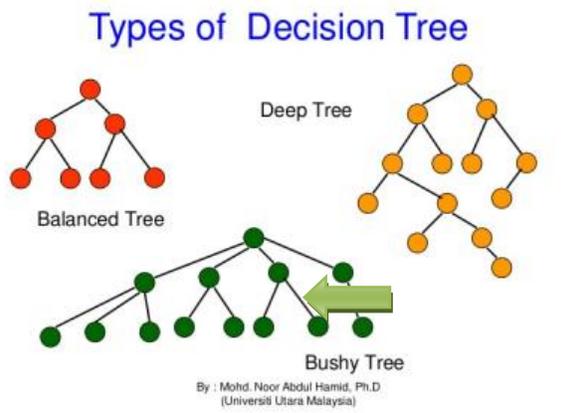


Fig 1 : decision tree

### 1. Initial exploration of data set:

The given data set is separated into two parts of 70% and 30% for training and testing respectively. Original test is done on the actual data set. Fig 2 represents the training data, testing data and their attributes in the students academic and AMCAT scores dataset.[2].

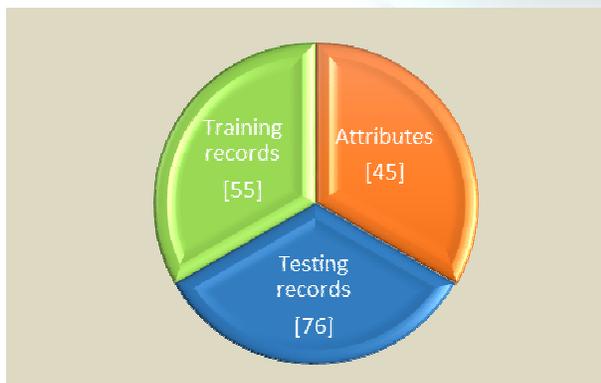


Fig 2: Exploration of data

### 2. Analysis

Attributes are classified as categorical and non-categorical as shown in Fig 3. A categorical variable is a variable that can take on one of a limited, and usually fixed, number of possible values, assigning each individual or other unit of observation to a particular group or nominal category on the basis of some qualitative property

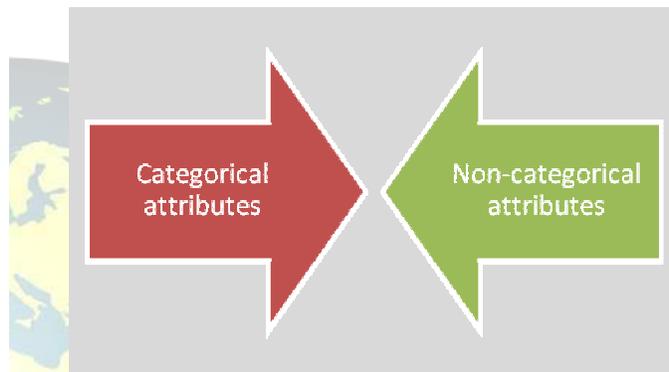


Fig 3: Types of attributes.

### 3. Initial Hypothesis

The relationship between the attributes are analyzed using correlation and chi-square method as shown in fig 4.

Correlation analysis is a method for measuring the covariance of two random variables in a matched data set. Covariance is usually expressed as the correlation coefficient of two variables X and Y. The correlation coefficient is a unit less number that varies from -1 to +1. The Chi Square statistic is commonly used for testing relationships on categorical variables. The null hypothesis is that no relationship exists on these categorical variables in the population; they are independent.

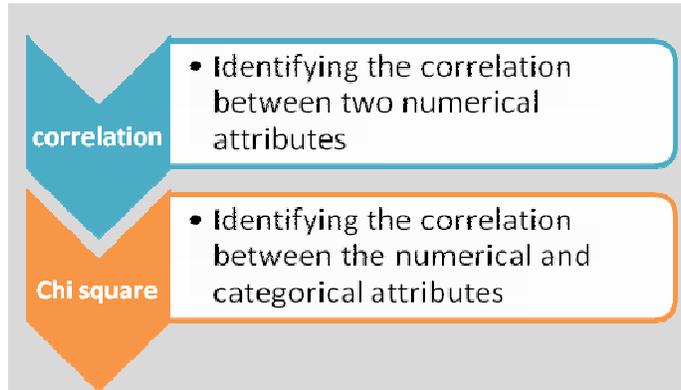


Fig 4 : Identifying relationship between attributes.

#### 4. Prediction:

The Student performance data set is collected and the data set contains 45 variables[2]. First we have to choose the attributes to predict the employability . The categories include BPO only, IT only, BPO & IT and Not Placed. These scores will result whether the student will be placed or not if so then in which category will he be placed .

The attributes used to predict the BPO category are:

- English(AMCAT)
- HS 2111
- HS 2161

The following attributes are used to predict the IT category. They are:

- Quants(Mathematics)
- Computer science theory
- Programming papers(lab)
- Automata module & Computer Programming module in AMCAT.

The following attributes are used to predict the Either BPO or IT category. They are:

- English(AMCAT)
- HS 2111
- HS 2161
- Quants(Mathematics)
- Computer science theory
- Programming papers(lab)
- Automata module & Computer Programming module in AMCAT.

If the student does not fall under either of the categories then he is considered to be under the category “NOT PLACED”.

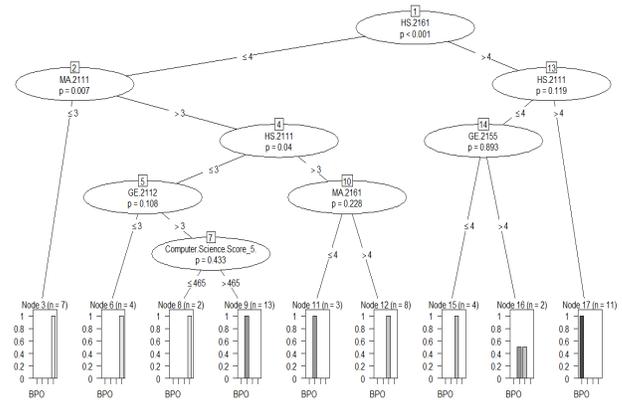


Fig 5: Relationship between Student marks and Employability.

#### V. CONCLUSION AND FUTUREWORK

This work has considered only Bachelor of Engineering (B. E) students, whereas Bachelor of Science (B. Sc) and MCA are also popular professional courses among employers. Future work will include the students of B. Sc. and MCA as well and will try to find out if there is a preference of B. S. / MCA over BE. Further employability has been defined as the students' ability to get employment during on Campus drives conducted in Vth Semester and does not take into account the pay package offered and the rating of the company in which he or she is placed. These aspects will be considered in our future work. Although academic performance is one of the criteria of student's consistent efforts and perseverance most of the companies are now concentrating upon employees current Technology updates. These parameters will be included in further research so as to infer if academically average students are also capable of getting good employment based on other skills and virtues.

#### VI. REFERENCES

[1]. “ K. Umamaheswari. & S.Niraimathi. “A Study on Student Data Analysis Using Data Mining Techniques , An International Journal of Advanced



Research in Computer Science and Software Engineering

[2]. “Lorraine Dacre Pool and Peter Sewell” The key to employability: developing a practical model of graduate employability

[3]. “Jianjun Wang and Li Li” Research on the College Graduate Employment Education Based on Data Mining Technology

[4]. “Tripti Mishra ,Dharminder Kumar and Sangeeta Gupta “Students’ Employability Prediction Model through Data Mining ,International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 4 (2016) ,

[5]. “Myzatul Akmam Sapaat, Aida Mustapha, Johanna Ahmad, Khadijah Chamili, Rahamirzam Muhamad” A Data Mining Approach to Construct Graduates Employability Model in Malaysia

[6]. “Vijay N. Kalbande, Dr. C.C. Handa” Developing a model to predict employability of engineering students in campus placement for IT sector. International Journal of Advance Research in Engineering, Science & Technology(IJAREST),

[7]. “Bangsuk Jantawan, Cheng-Fa Tsai” A Classification model on Graduate Employability using Beysian Approaches:A Comparison. International Journal of Innovative Research in Computer and Communication Engineering.