Campus Recruitment: Academic and Employability Factors Influencing Placement Using Data Mining

Sherin K Thomas
Department of Computer Application
Amal Jyothi College of Engineering
Kottayam, India
Sherinkthomas07@gmail.com

Nimmy Francis
Asst.prof Department of Computer Application
Amal Jyothi College of Engineering
Kottayam, India
nimmyfrancis@amaljyothi.ac.in

Abstract-- Educational Data Mining is the term for mining in the context of education. Educational Data Mining is focused with the development of innovative methods for discovering knowledge from educational databases and applying it to educational decision-making. In this study, we gathered data from students that included various details about their prior and current academic records, and then used several classification algorithm using Data Mining tools (WEKA) for analysis the student's academic performance for training and placement.

Keywords — Weka, Placement, Data mining, attributes, algorithms, accuracy

I. INTRODUCTION:

Nowadays, the most vital concern of a parent in getting his or her child enrolled into an educational institution in India is campus recruiting. Institutions engage in a variety of branding initiatives in order to attract MNC corporations for local and worldwide recruiting as well as academic partnerships. The education system is now regarded as one of India's most important economic partners. Because of the emphasis on placements, the educational system's value is dwindling. However, as a result of the epidemic, the whole educational system, as well as career opportunities, has been turned upside down. Students in both undergraduate and graduate programmes participate in the selection process in their last year of study.

The data mining process may be used to analyse pertinent information findings and generate multiple viewpoints in order to gain a better understanding of students' actions. Applying data mining techniques to the design of an educational environment uncovers relevant information that may be utilised in formative evaluation to help educators develop a pedagogical foundation for making key decisions.

II. DESCRIPTION

Many organizations have put their placement processes on hold because to mounting fears about the COVID-19 outbreak. India's colleges are in for a challenging year as a

result of the pandemic's impact on companies all across the world; they're having trouble recruiting students on campus. Most business and engineering institutions do campus placements every year between December and April. While many students have already been placed, others are still looking for work. Due to the protracted lockdown imposed to battle the COVID-19 epidemic, few corporations are postponing offers, and several have already retracted offers. There are still firms that honor job offers given to students and will not reduce their salaries. Let's have a look at them.

Key facts

- Every corporation makes every effort to hire as many freshmen as possible.
- To be considered for the placement recruiting, you must have a CGPA of at least 7.
- Communication fluency is a top priority for the company's campus recruitment.
- Your problem-solving assessment capability will increase your chances of being placed.
- The recruiters will ask you challenging questions in order to assess your problem-solving abilities. It will also give you a better chance of succeeding.

Data mining

Data mining is a method for finding patterns in huge data sets that uses approaches from machine learning, statistics, and database systems. Data mining is a computer science and statistics multidisciplinary topic with the ultimate objective of extracting information (using intelligent methods) from a data collection and transforming it into

an intelligible structure for further analysis their customers to expand greater effective advertising. Data mining processes are used to build machine learning models that power applications including search engine technology and website recommendation programs [3].

10.5281/zenodo.5109070
ISBN:978-93-5426-386-6@2021 *MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*

Commonly used functionalities include Data cleansing, Artificial intelligence (AI), Association rule learning, Clustering, Classification, Data analytics, Data warehousing, Machine learning, Regression.

III. PURPOSE OF THE STUDY

This paper examines the numerous factors that influence where students are assigned. The dataset utilised in this study comprises 14 characteristics, one target variable, and 215 occurrences, 148 of which were put and 67 of which were not. WE are using these data sets to train a dataset and determine the most accurate algorithms and their correctness, as well as construct a testable model.

IV. METHODOLOGY

It provides you a visualization tool to inspect the data.

V. IMPLEMENTATION

We're using three different algorithms on a particular data set to construct or extract an usable model for predicting the severity of coronary heart disease in a person in a certain setting.

Data source

This article discusses the process of analysing the academic performance of students pursuing a degree or a postgraduate degree, as well as the generation of illness patterns or behaviour using algorithms and data mining techniques. Waikato Environment for Knowledge Analysis was utilised as a development tool (weka).

Weka

Weka is a machine learning software package created at New Zealand's University of Waikato. The software is developed in the Java programming language. It comes with a graphical user interface and a set of visualisation tools and algorithms for data analysis and predictive modelling. Weka provides tools for data pre-processing, clustering, classification, regression, visualisation, and feature selection, allowing us to create machine learning approaches and apply them to real-world data mining situations.

Properties of weka

- The new machine learning schemes can also be developed with this package.
- The raw data collected from the field may contain several null values and irrelevant fields.
- The data pre-processing tools provided in weka helps to cleanse the data.
- Save the pre-processed data in your local storage for applying ML algorithms.
- Next, depending on the kind of ML model that are trying to develop we can select one of the options such as Classify, Cluster, Associate etc.
- The Attributes Selection allows the automatic selection of features to create a reduced dataset.
- Under each category, WEKA provides the implementation of several algorithms. We can select an algorithm of our choice, set the desired parameters and run it on the dataset.
- Then, WEKA would give you the statistical output of the model processing.

We'll look at a dataset from the Kaggle repository that has 215 instances and 14 characteristics. Several criteria are used to forecast where a student will be placed.

The Kaggle repository data set provides student characteristics related to their academic success. This data collection serves as the classifier's training data. [4]

Table 1: Attributes and its values

Input Attributes	Function/ Values
age	age in years
sex	sex (1 = male; 0 = female)
Ssc_p	Secondary school percentage
Ssc_b	Secoundary school branch
Hsc_p	Higher school percentage
Hsc_b	Higher school branch
Degree_p	Degree Percentage
Degree_t	Degree trend
workex	Work Experience
etest	Entrance test
specialization	Specialization
Mba_p	Mba_percentage
status	1 = placed; 2 = not placed;
salary	Salary Expectation

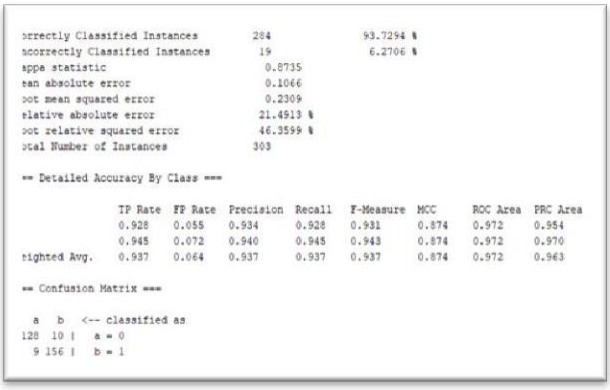
Algorithms

Decision Tree: A branch of supervised learning algorithms is the Decision Tree algorithm. The decision tree approach, unlike other supervised learning algorithms, may also be utilised to solve regression and classification issues.

By learning basic decision rules inferred from past data, the purpose of employing a Decision Tree is to develop a training model that can be used to predict the class or value of the target variable (training data). When using Decision Trees to forecast a record's class label, we start at

10.5281/zenodo.5109070
ISBN:978-93-5426-386-6@2021 *MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*

the top and work our way down the tree. The values of the root attribute and the record's attribute are compared. We follow the branch that corresponds to that value and leap to the node based on the comparison.



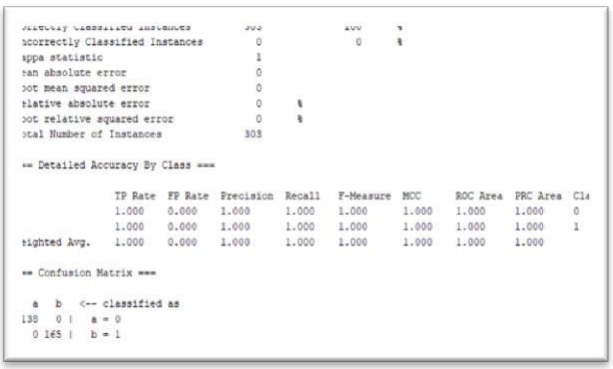
```
orrectly Classified Instances         284               93.7294 %
ncorrectly Classified Instances        19                6.2706 %
appa statistic                          0.8735
ean absolute error                      0.1066
oot mean squared error                  0.2309
elative absolute error                 21.4913 %
oot relative squared error             46.3599 %
otal Number of Instances               303

== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area
               0.928    0.055    0.934      0.928   0.931      0.874    0.972     0.954
               0.945    0.072    0.940      0.945   0.943      0.874    0.972     0.970
eighted Avg.   0.937    0.064    0.937      0.937   0.937      0.874    0.972     0.963

== Confusion Matrix ===

   a   b   <-- classified as
 128  10 |   a = 0
   9 156 |   b = 1
```

Figure1: Summary for test modal using Decision Tree

Random Tree: The random trees classifier, unlike the Support Vector Machine (SVM), can handle a mix of categorical and numerical variable. The Random Trees is also less sensitive to data scaling while SVM often required data to be normalized prior to the training/classification [7].

When the training set is short or uneven, however, SVM is said to perform better. Random Trees is a considerably less computationally costly classifier than SVM, and it works better and faster with large training sets.

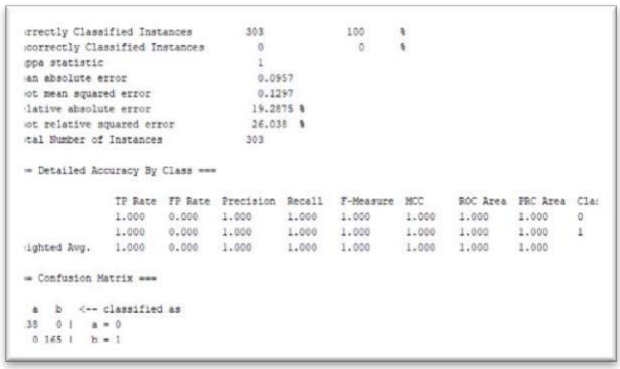
Many versions of the Random Trees algorithm exist. Object Analyst uses the Opens implementation which use the Gini Impurity index to determine what is a good split point for a node on the classification tree and the minimum number of samples, the maximum tree depth and the accuracy of the trees as stopping criteria.



```
ncorrectly Classified Instances         0                 0      %
appa statistic                          1
ean absolute error                      0
oot mean squared error                  0
elative absolute error                  0      %
oot relative squared error              0      %
otal Number of Instances              303

== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Cla
               1.000    0.000    1.000      1.000   1.000      1.000    1.000     1.000     0
               1.000    0.000    1.000      1.000   1.000      1.000    1.000     1.000     1
eighted Avg.   1.000    0.000    1.000      1.000   1.000      1.000    1.000     1.000

== Confusion Matrix ===

   a   b   <-- classified as
 138   0 |   a = 0
   0 165 |   b = 1
```

Figure2: Summary for test modal using Random Tree

Random Forest: Random forest is a learning method that is supervised. It creates a "forest" out of an ensemble of decision trees, which are commonly trained using the "bagging" approach. The bagging method's basic premise is that combining several learning models improves the final output.



```
rrectly Classified Instances         303               100      %
correctly Classified Instances         0                 0      %
ppa statistic                          1
an absolute error                      0.0957
ot mean squared error                  0.1297
lative absolute error                 19.2875 %
ot relative squared error             26.038  %
tal Number of Instances              303

= Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Cla
               1.000    0.000    1.000      1.000   1.000      1.000    1.000     1.000     0
               1.000    0.000    1.000      1.000   1.000      1.000    1.000     1.000     1
ighted Avg.    1.000    0.000    1.000      1.000   1.000      1.000    1.000     1.000

= Confusion Matrix ===

   a   b   <-- classified as
138   0 |   a = 0
   0 165 |   b = 1
```

Figure3: Summary for test modal using Random Forest

Table 2: Result before filtration

Techniques	Accuracy	Time
Decision Algorithm (j48)	95.7294 %	0.01
Random Tree	100 %	0.08
Random Forest	100 %	0.07

We were able to raise the accuracy of the decision tree to 96.0594 percent after applying some filter to the attribute, which is just approximately 0.38 percent increase in values. Since it has a value close to 95 percent, we can utilise this approach, but the other two algorithms outperform the decision tree approach.

Table 3: Result after filtration

Techniques	Accuracy	Time
Decision Algorithm (j48)	96.0594 %	0.01
Random Tree	100 %	0.08
Random Forest	100 %	0.07

10.5281/zenodo.5109070
ISBN:978-93-5426-386-6@2021 *MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*

VI. CONCLUSION

Data mining tactics aid in the discovery of hidden features in the recruitment of students from diverse fields, allowing them to focus on the crucial notion of placement. It will concentrate on a variety of issues that students should be aware of.

The use of specific algorithms and mixes of multiple extremely essential features for placement prediction using data mining is the focus of this article. Using 14 characteristics, the Random Tree and Decision Forest Tree outperformed with 100 per cent accuracy.

Moreover the algorithm Random Forest having accuracy 100% as well as the performance timing of 0.07 gives a chance over the Random Tree. The algorithms are implemented with the default parameters only.

VII. REFERENCES

[1] https://www.researchgate.net/publication/282422075_Analysis_of_Campus_Recruitment_Parameters_in_an_Indian_Context
[2] https://www.researchgate.net/publication/282422 075 https://www.who.int/news
[3] https://www.investopedia.com/terms/d/dataminig
[4] Bruce Basta, D' Archy Becker, P.Jane Staly Basma Saleh1, Ahmed Saedi2, Ali alAqbi3, Lamees Salman4
[5] TimesJobs.com Bureau, "Train campus recruits early to build broader talent base"(Online) Available: http://articles.economictimes. indiatimes.com/2013-05-06/news/39065498_1_campus-recruitment-surveyed-organisations-timesjobs-com(May 6,2013)
[6] Gokuladas (2010), "Technical and non-technical education and the employability of engineering graduates: an Indian case study", International Journal of Training and Development, Vol. 14 No.2,pp.130-143

10.5281/zenodo.5109070
ISBN:978-93-5426-386-6@2021 *MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*