iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 3 **Issue:** III **Month of publication:** March 2015

**DOI:**

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

# Student Placement Prediction Using ID3 Algorithm

Namita Puri [#1], Deepali Khot [#2], Pratiksha Shinde [#3], Kishori Bhoite [#4], Prof. Deepali Maste [#5]
[#12345]Department of Computer Engineering, Atharva College of Engineering, Malad,Maharashtra,India.

***Abstract -Data mining is a technology which is used in many fields now a day. In educational field data mining is used to understand different perspectives in student data. The main aim of this research work is to identify relevant attributes based on academics, skills and curricular of final year student and design a model which can predict placement of the student using a classification technique based on decision tree. This model can be useful for faculties, university and students to more emphasize on those which are not eligible for placement according to this model.***
***Keywords: Data Mining, Classification, Prediction, ID3 Algorithm, Decision Tree.***

## I. INTRODUCTION

Here is the paper which focuses on the classification of the student for the placement process in an engineering college. Campus placement of a student plays very important role in a college. Campus placement is a process where companies meet colleges and identify students which are talented and qualified, before they complete their graduation. So this paper suggests a system which makes the work of prediction of placement of student easy. For this we suggest to use ID3 algorithm- data mining classification algorithm.

### *A. Data Mining [1]*

Data mining refers to extracting or mining useful patterns from large database. It is knowledge discovery in large amount of data. Data mining is a new technology which helps organizations to process data through algorithms to uncover meaningful patterns and correlations from large databases that may otherwise be not possible with standard analysis and reporting. Data mining has various applications in numbers of sectors. Data mining provides tools that can be used to understand the business requirements better through knowledge discovery. Data mining field finds its application in market analysis and management like customer relationship management, market segmentation. Other example can be of banking where ETL tools available in data mining can be used for overall bank management. Data mining functionalities are used to specify the kind of patterns to be found in data mining task. In general, data mining task can be classified into two categories descriptive and predictive. Descriptive mining tasks characterize the general properties of the data in the database. Predictive mining task perform process current data in order to make predictions.

### *B. Classification [1]*

Classification is a data mining technique used for systematic arrangement of group membership for data. For example, suppose you want to decide whether you will go out for playing depending on the climate. Classification is the process of finding a model that describes and distinguishes data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown .The derived model is based on the analysis of a set of training data. Classification and prediction may need to be preceded by relevance analysis, which attempts to identify attributes that do not contribute to the classification or prediction process. These attributes can then be excluded.

## II. LITERATURE REVIEW

Sweta Rai, Priyanka Saini, Ajit Kumar Jain [2] performed comprehensive study where they showed classification algorithm ID3 can be used to predict dropout students. Romero and Ventura [3] performed comprehensive study of Educational data mining from 1995 to 2005 .Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, and M. Inayat
Khan [4] applied K-mean clustering to analyze learning behavior of students which will help the tutor to improve the performance of students and reduce the dropout ratio to a significant level. Romero [5] studied on the factors that predict failure and non-retention in college courses. Ajay Kumar Pal, Saurabh Pal [6] performed prediction of student placement using different classification algorithm. D. Saravana Kumar, T. Jeevalatha [7] performed analysis on the student selection for placement using decision tree algorithms.



©IJRASET 2015: All Rights are Reserved

## III. ID3 ALGORITHM

The prediction is the main goal behind use of ID3 algorithm. ID3 is mainly use for decision making. **ID3 (Iterative Dichotomiser 3)** is an algorithm invented by Ross Quinlan used to generate a decision tree from a dataset [2]. It builds the decision tree from top to down. No backtracking is there. Information Gain and entropy is calculated in implementation.

*A. Entropy*

It is a measure in the information theory, which characterizes the impurity of an arbitrary collection of examples. If the target attribute takes on *c* different values, then the entropy S relative to this c-wise classification is defined as

$$\text{Entropy}(s) = \sum - P_i \log_2 P_i \text{ [8]}$$

Where *pi* is the proportion/probability of S belonging to class *i*. Logarithm is base 2 because entropy is a measure of the expected encoding length measured in bits.
For e.g. if training data has 14 instances with 6 positive and 8 negative instances, the entropy is calculated as

Entropy ([6+, 8-]) = 22 − (6 /14) log (6 /14) − (8/14) log (8/14) = 0.985

A key point to note here is that the more uniform is the probability distribution, the greater is its entropy.

*B. Information gain*

It measures the expected reduction in entropy by partitioning the examples according to this attribute. The information gain, Gain(S, A) of an attribute A, relative to the collection of examples S, is defined as

Gain(S,A)=Entropy(S)-∑( |Sv|/ |S|) Entropy(Sv)[8]

Where *Values* (*A*) is the set of all possible values for attribute A, and *Sv* is the subset of S for which the attribute A has value *v*. We can use this measure to rank attributes and build the decision tree where at each node is located the attribute with the highest information gain among the attributes not yet considered in the path from the root.

*C. Implementation for Id3 Algorithm [2]*

Step 1: compute classification entropy.
Step 2: for each attribute, calculate information gain using classification attribute.
Step 3: select attribute with highest information gain.
Step 4: remove node attribute, for future calculation.
Step 5: repeat steps 2-4 until all attribute have been used.

```
Function ID3 (Input attribute, Output attribute, Training data)
{
If (Training data is empty)
{
Return a single node with Failure;
}
If (all records in training data have positive value)
{
Return a single node with level positive.
}
If (all records in training data have negative value)
{
The single node with level negative;
}
If (input attribute is empty)
{
Return a single node with the value of the most frequent value;
```



©IJRASET 2015: All Rights are Reserved

}
Otherwise
{
Compute information gain for each attribute;
Split the attribute with highest information gain value;
Return a tree with root node X and arcs X1, X2…, Xm;
Recursively call the ID3 function until all attribute have been used.
}}

## IV. ATTRIBUTES FOR PLACEMENT PREDICTION SYSTEM

Placement Prediction system gives those factors which are useful for teachers and students to improve success percentage of placement. Placement system takes many attributes as a input related to student information and company information. Out of those only some attributes are used for prediction of placement.

TABLE I.
STUDENT ATTRIBUTE TABLE

| **Attributes** | **Values** |
|---|---|
| $10^{th}$ % | First class<br>Second class<br>Third class |
| $12^{th}$ % or Diploma | First class<br>Second class<br>Third class |
| Degree Aggregate | First class<br>Second class<br>Third class |
| Sem Performance | Poor<br>Average<br>Good |
| Communication Skill | Poor<br>Average<br>Good |
| Work on Projects | Poor<br>Average<br>Good |
| Internship | Yes / No |
| Education Gap | Yes / No |
| Backlog | Yes /No |
| Placement | Yes/ No |

## V. WORKING WITH ALGORITHMS

Various decision tree algorithms are present in data mining such as ID3, C4.5, and CHAID etc.

### *A. C4.5 (Successor of ID3)*

C4.5 is a popular algorithm which is used for the generation of a decision trees. It is an advanced level of the ID3 algorithm which is designed to overcome its limitations. The decision trees generated by this algorithm are used for prediction and it is a classifier of statistical type.

### *B. CHAID*

CHAID is the short version of CHi-squared Automatic Interaction Detector. When computing classification trees it is used for performing multi-level splits. CHAID is usually used for prediction and also classification, and even for detection of interaction



©IJRASET 2015: All Rights are Reserved

between different variables [9]. The main advantage of CHAID over the others is CHAID is non-parametric. CHAID tries to stop growing the tree before the occurring of over fitting.

According to calculations done by T. Jeevalatha and N. Ananthi [9] accuracy of the ID3 compare to other algorithms is as follows:

*Precision and recall [7]:*

Precision: Percentage of selected items that is correct

Recall: Percentage of correct items that are selected

$$\text{Precision} = \frac{\text{true positive}}{\text{True positive + false negative}}$$

$$\text{Recall} = \frac{\text{true positive}}{\text{True positive + false negative}}$$

$$\text{Accuracy} = \frac{\text{true positive + true negative}}{\text{True positive + false positive + false negative + true negative}}$$

TABLE II.
ACCURACY OF THE ALGORITHMS [7]

| Algorithm | Total number of students | Student whose result was wrongly predicted | Accuracy |
|---|---|---|---|
| ID3 | 1342 | 68 | 95.33 |
| C4.5 | 1342 | 81 | 95.05 |
| CHAID | 1342 | 90 | 94.18 |

## VI. CONCLUSION AND FUTURE WORK

Here we conclude that ID3 is a best algorithm which can be used for classification and prediction of student's placement in a engineering college. Result indicates that ID3 decision tree algorithm is best classifier with 95% accuracy. This study will also work to identify those students which needed special attention and guidance to increase their chances of placement.

The generated knowledge will be quite useful for management of placement process to develop policies and strategies for better planning and implementation of educational program like improving soft skills and infrastructure to increase the placed student rate in University.

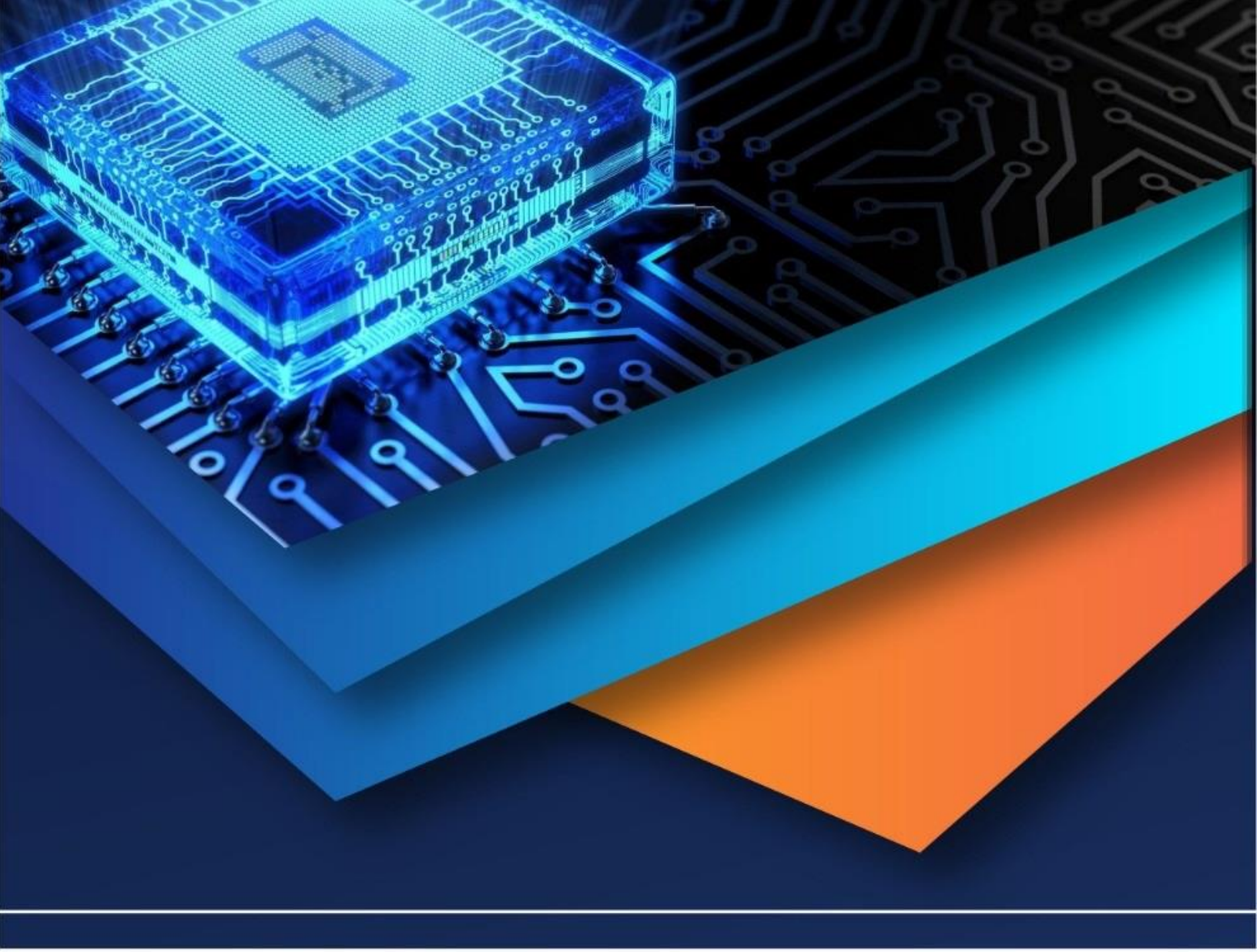
The present research can be extent with the help of neural network algorithm, genetic algorithms.

## REFERENCES

[1] Reference book on "Data Mining".
[2] Sweta Rai, Priyanka Saini, Ajit Kumar Jain "Model for Prediction of Dropout Student Using ID3Decision Tree Algorithm" International Journal of Advanced Research in Computer Science & Technology, Vol. 2 Issue 1 Ver. 2 Jan-March 2014
[3] Romera, C. and Ventura, S.," Educational Data Mining: A Survey from 1995 to 2005."Expert Systems with Applications33, 125-146, 2007
[4] S. Ayesha, T. Mustafa, A.R. Sattar, and M.I.Khan, Data Mining Model for Higher Education System, European Journal of ScientificResearch, Vol.43, No.1, pp.24-29, 2010.
[5] Romero, C., Ventura, S., Eapejo, P.G. and Hervas, C. "Data Mining Algorithms to Classify Students" In Proceedings of the 1st International Conference on Educational Data Mining, pp. 8-17, 2008.
[6] Ajay Kumar Pal, Saurabh Pal "Classification Model of Prediction for Placement of Students" I.J.Modern Education and Computer Science, 2013, 11, 49-56.
[7] D. Saravana Kumar, T. Jeevalatha "Performance Analysis of Undergraduate Students Placement Selection using Decision Tree Algorithms" International Journal of Computer Applications (0975 – 8887) Volume 108 – No 15, December 2014.
[8] Anand Bahety "Extension and Evaluation of ID3 – Decision Tree Algorithm"
[9] V.Ramesh, P.Parkavi, P.Yasodha"Performance Analysis of Data Mining Techniques for Placement Chance Prediction" International Journal of Scientific & Engineering Research Volume 2, Issue 8, August-2011 1 ISSN 2229.



©IJRASET 2015: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)