

## Decision Tree Classification of Students Placement Factors Using R Data Miner

Mrs. Dhanamma Jagli, Mrs. Meenakshi Garg

Assistant professor, Department Of MCA V.E.S. Institute of Technology Mumbai, India.

---

**Abstract:** Student campus placement attitudes have largely been assessed using college students are relatively small samples. In addition, no systematic framework has been used to investigate the origins or influencing social factors of such attitudes and how these interact with demographic characteristics. Using Google forms an online survey was conducted, to examine placement-related attitudes and behaviors among 3,340 adults. We examined the students' behavior in Technical and nontechnical interviews. Geographical availability and gender of the student which could affect placements are not investigated in this research stream. Findings also reveal differences in both attitudes and behaviors across the technical and nontechnical interviews. This survey was conducted to improve the placement process at the college.

**Keywords:** Analytics, Decision Tree, Data Mining, data Analysis, Placement Analytics, R Tool, R Data Miner.

---

### I. Introduction

Today campus placements have become one of the crucial factors for judging the quality of a college education provides. Efforts of students and college staff are judged by campus placement. Many Multinational Companies, big and small companies, visit the colleges for hiring new employees. Campus placement should provide good opportunities to the students they out into the industry looking for a job for them. Campus placement is one of the important aspects of the student while taking admission in the college... Each year, about 800 organizations hire more than 100,000 graduates from 3500+ colleges in India. While the primary goal of campus hiring is to bring in fresh talent, it also helps to build a brand name among the top colleges in the country.

This paper examines the various aspects of students behavior for getting placed through campus. In this paper we conducted an online survey for the students, to analyze various factors that had an impact on their placement. These factors included various technical and non-technical aspects of students personality. The focus is to explore the measurement, collection, analysis, and reporting of data as predictors of student success and drivers of the departmental process and program curriculum. Behavioral analytics in higher education is used to predict student success by examining how students behave and react to succeed in job placement. The paper examines what is being done to support students, whether or not it is effective, and if not why, and what educators can do. The paper also examines how these data can be used to create new metrics and inform a continuous cycle of improvement. It presents examples of working models from a sample of institutions of higher education: Finally, the paper identifies considerations and recommendations for using behavioral analytics and offer suggestions for future improvement.

### II. Workflow

The focus of the placement cell is to strike a match between recruiter expectations and student aspirations. The goal of the placement cell is to provide a unique platform to the students by acting as a link between the industries and the academia. To achieve its goal, the Training & Placement Cell works towards recognizing the core competencies of students and enhancing them to give their best to the world and beyond.

To achieve its goals placement cell follows the listed below procedure –

- 1) Placement cell registers students for getting placed.
- 2) In addition to that placement, cell invites various core and non-core companies for hiring fresh talent.
- 3) Companies specify eligibility criteria so that students can meet their expectations.
- 4) Students who are eligible for the company take participate in placement process of that company.
- 5) Depending upon their requirements companies conduct written test.
- 6) After clearing written test students to appear for interviews.
- 7) Companies may have one or two interviews.
- 8) Once the student clears all interviews they get offer letter from the company.

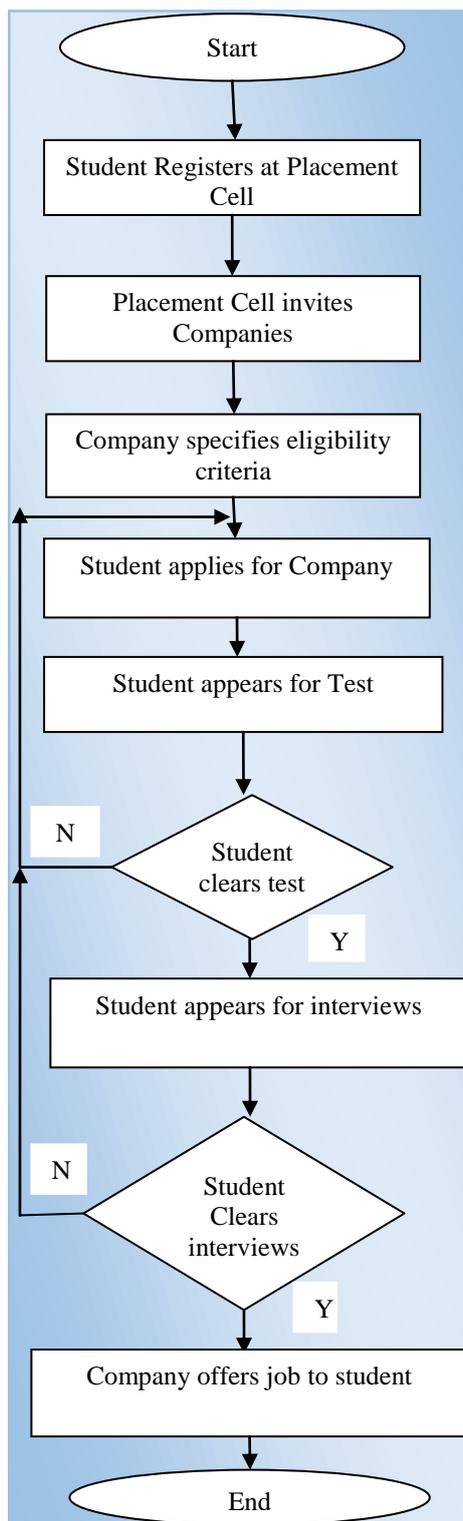


FIG: 1 Placement Procedure Workflow

### III. Literature Survey

The method proposed in this paper for student job placement belongs to the process of Knowledge Discovery and classification after Mining data. A number of methods are available to assist in extracting patterns that when interpreted provide valuable, possibly previously unidentified, insight into the stored data. This information can be predictive or descriptive in nature. Data mining, the pattern extraction phase of KDD, can take on many forms, the choice is dependent on the required results. KDD is a multi-step process that facilitates the conversion of data to useful information.

The process of finding useful patterns in data is known by a variety of names like knowledge extraction, information discovery, information harvesting, data archaeology, and data pattern processing. The term data mining has mostly been used by statisticians, data analysts, and the management information systems (MIS) communities. It has also gained popularity in the database field. After applying data mining techniques on data, Decision Tree was created. A decision tree is a machine learning algorithm that is broadly used by researchers to infer rules for improving the system. Decision Trees predict an output by tracking the decisions in the tree from the root node down to a leaf node according to the outcome of the tests along the path. Decision Trees consist of nodes (tests for the value of a certain attribute), that have exactly one incoming edge and leaves that predict the outcome and act as terminal nodes. A leaf node attribute produces a homogeneous result (all in one class), which does not require additional classification testing. The complexity of the tree is controlled by the stopping criteria and the pruning method applied and is measured by the total number of nodes, a total number of leaves, tree depth, and a number of attributes utilized. For generating Decision tree ID3 algorithm is used in which a set of non-categorical attributes are represented as  $NC_1, NC_2, \dots, NC_n$ , the categorical attribute  $C_1, C_2, \dots, C_n$ , and training set as T.

The algorithm used is given below Function ID3 (K: a set of non-categorical attributes, see: the categorical attribute, T: a training set returns a decision tree;

Begin

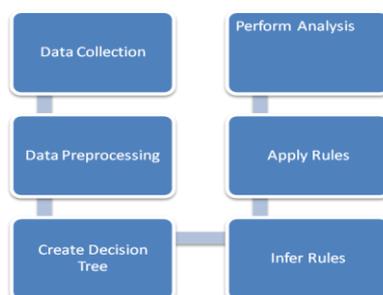
- ✓ Step 1: If T is empty, return a single node with value Failure;
  - ✓ Step 2: If T consists of records all with the same value for the categorical attribute, return a single node with that value;
  - ✓ Step 3: If K is empty, then return a single node with as value the most frequent of the values of the categorical attribute that are found in records of T; [note that then there will be errors, that is, records that will be improperly classified];
  - ✓ Step 4: Let A be the attribute with largest Gain (A, T) among attributes in K;
- Let  $\{A_j | j=1, 2, \dots, m\}$  be the values of attribute A;  
 Let  $\{T_j | j=1, 2, m\}$  be the subsets of T consisting respectively, of records with value  $A_j$  for attribute A;  
 Return a tree with root labeled A and arcs labeled  $a_1, a_2, a_m$  going respectively to the trees  $ID3(K-\{A\}, C, T_1)$ ,  $ID3(K-\{A\}, C, T_2)$ ,  $ID3(K-\{A\}, C, T_m)$ ; end ID3;

#### IV. System Architecture

In order to improve the placement process, the structured approach was adopted to provide a comprehensive solution which used R- Data Miner to infer rules. The proposed architecture is illustrated in Fig2. This is a generalized architecture which can be applied to any placement procedure to improve it. The proposed architecture consists of six steps as follows:

- a) *Data Collection* - This stage consists of collecting data. We used a google form for our survey. For sample data, we collected data from students who got jobs through campus placement in the year 2015-16.
- b) *Data Preprocessing* - At this stage we prepared data so decision tree can be created. We preprocessed data to have logical and categorical attribute values.
- c) *Decision Tree Creation* - At this stage, we fed analyzed data to Data Miner of R tool to get decision tree.
- d) *Inference rules* - After decision tree was created, we inferred rules from decision tree, to get the attributes which are highly affected the campus placement of students.
- e) *Apply rules*- As we got rules from the previous step, we are still in the procedure to apply those rules to the current system.
- f) *Perform Analysis*- We will be performing analytics for session 2016-17 to check if we could really improve the system.

The proposed system is expected to improve the placement of the student. Here Knowledge is discovered which was hidden in student's placement procedure. This analysis helped in providing predictive knowledge which could lead to better campus placement for students.



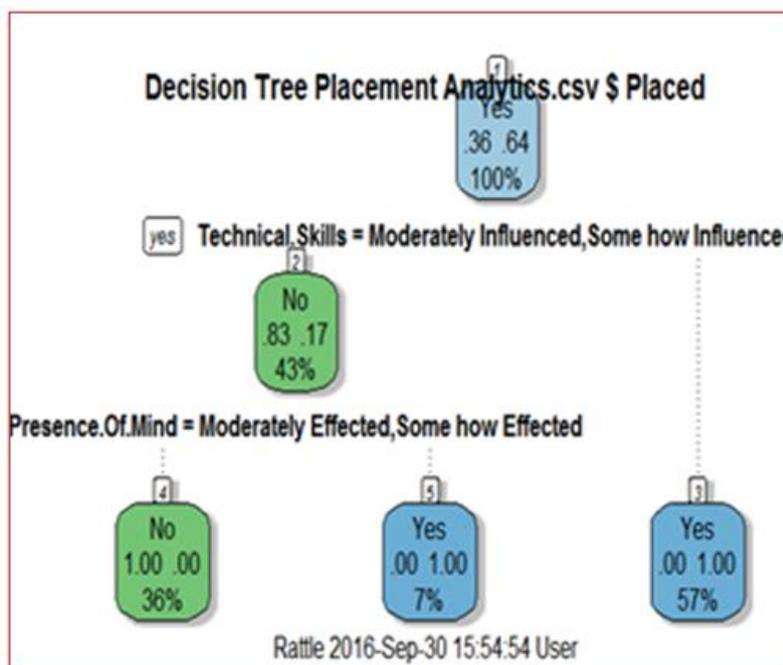
**FIG 2:** System Architecture

### V. Results And Discussion

In an educational institution, students placements have been playing an important role become very crucial for existence. The academic skills of the students very important to achier their target. But many times it's happened that only academic skills will not help students to get into the placements. Based on the analysis of t students placements carried out in the past years, crucial attributes related to placement have been found. They are nothing but attributes related to the overall personality of students including technical skills, attitude, the presence of mind, dressing behavior, Presence Of Mind, Communication skills, Presentation Skills, Dressing Sense, Personal Behavior, Attitude, usage of social networks. All these are the attributes which are influencing the placement of students. The data have been collected from previously placed students through survey system shown in the below Table 1. And constructed Decision Tree, and then derived rules, which can be used as key rules to the future predictions.

ME: Moderately Affected, SI: Strongly Influenced,  
 MI: Moderately Influenced, SE: Strongly Effected,  
 SHI: Some How Influenced, SHE: Some How Effected

Txt	Common Soft Skills	Skills	Presentation Skills	Dressing Sense	Personal Behavior	Attitude	Placed
T1	SE	SI	SI	SI	SI	MI	Yes
T2	SE	SI	SI	SI	SI	MI	Yes
T3	SE	SI	SI	SI	SI	MI	Yes
T4	SE	MI	SI	MI	MI	MI	Yes
T5	SE	MI	SI	MI	MI	MI	Yes
T6	ME	SI	MI	MI	SI	SI	Yes
.							
T22	ME	SI	MI	SI	SI	SI	Yes
T23	SE	SI	SHI	SHI	SI	SI	Yes
T24	SE	SI	SI	SI	SI	MI	Yes
T25	SE	SI	MI	MI	SI	SI	Yes



**Tree as rules:**

- ✓ Rule number: 3 [Placed=Yes cover=8 (57%) prob=1.00] Technical Skills=Strongly Influenced.
- ✓ Rule number: 5 [Placed=Yes cover=1 (7%) prob=1.00] Technical. Skills = Moderately Influenced, Somehow Influenced  
 Presence.Of.Mind=Strongly Effected
- ✓ Rule number: 4 [Placed=No cover=5 (36%) prob=0.00] Technical Skills=Moderately Influenced, Somehow Influenced, Presence. Of. Mind= Moderately Effected, Some how Effected.

## VI. Conclusions

In this paper, some practical approach has been implemented to improve the student's placement by adding skill enhancements. It has proved that only academics skills are not enough to get into the present industry. So that focus should be done on student's holistic growth along with communication Skills and technical skills and personal skills. By using R data miner decision tree constructed and framed rules to prepare students community in the future for employable in an effective manner.so that students need to improve the continues learning skills in order to retain in the industry.

## References

- [1] Kanno, H., Koeske, G.F.: MSW students' satisfaction with their field placements: The role of preparedness and supervision quality. *J. Soc. Work Educ.* 46, 23–38 (2010).
- [2] Patrón, R.: Can more education be bad? Some simple analytics on financing better education for development. 1091–1100.
- [3] Baker, R.S.J.D., Yacef, K., Slade, S., Prinsloo, P., Bültmann, M., Chatti, M.A., Schroeder, U., Prinsloo, P., Slade, S.: Learning analytics for online discussions: a pedagogical model for intervention with embedded and extracted analytics. *Am. Behav. Sci.* 15, 2–5 (2012).
- [4] Lefebvre, A., Spruit, M., Ota, W.: Towards reusability of computational experiments capturing and sharing Research Objects from knowledge discovery processes. *Proc. 7th Int. It. Conf. Knowl. Discov. Knowl. Eng. Knowl. Manag. (IC3K 2015)*. 1, 456–462 (2015).
- [5] Riegler, M., Pogorelov, K., Lux, M., Halvorsen, P., Griwodz, C., De Lange, T., Eskeland, S.L.: an Explorative hyperbolic-tree-based clustering tool for unsupervised knowledge discovery. *Proc. - Int. Work. Content-Based Multimedia. Index.* 2016–June, (2016).
- [6] Lefebvre, A., Spruit, M., Ota, W.: Towards reusability of computational experiments capturing and sharing Research Objects from knowledge discovery processes. *Proc. 7th Int. It. Conf. Knowl. Discov. Knowl. Eng. Knowl. Manag. (IC3K 2015)*. 1, 456–462 (2015).
- [7] Grillo, A., Spring, A., Gazarek, J., Morphew, A.: Innovative student placements at Northern Health. 6, 3076 (2013).
- [8] Calvet Liñán, L., Juan Pérez, Á.A.: Educational Data Mining and Learning Analytics: differences, similarities, and time evolution. *RUSC. Univ. Knowl. Soc. J.* 12, 98 (2015).
- [9] Williamson, B.: Digital education governance: data visualization, predictive analytics, and “real-time” policy instruments. *J. Educ. Policy.* 0, 1–19 (2015).
- [10] Keser, S.B.: A Case Study of Optimal Decision Tree Construction for RfKON Database. (2016).
- [11] Guleria, P., Sharma, A., Sood, M.: web -based data mining tools : performing feedback analysis and. 5, 35–44 (2015).