Student's Employability Prediction Using Data Mining

Madhavi Girase, Suchita Lad, Prerna Pachpande

Abstract— Student's 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 paper uses Decision Trees as classification techniques of data mining to predict the employability of Undergraduate Engineering students and find the algorithm which is best suited for this problem. For this purpose, a data set is developed with the traditional parameters like socioeconomic conditions, academic performance and some additional emotional skill parameters. 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.

Index Terms— Data mining, Decision Trees, Employablity Prediction

1 Introduction

Higher education plays a key role in strengthening a nation's economy as it is an industry in itself and it supports the rest of the industry by providing a trained workforce. Earlier, the major concern for these Institutions were the decrease in the student success rate, decrease in retention of students, increase in students moving to other competitive institution and lack of counselling to students in subject selection. However, with education becoming more and more employment oriented, employment of students, graduating from any Institution has become a major factor in building the reputation of the Institution and hence a major concern. 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 objective is to study the level of employability skills among the Engineering students and identify the attributes looked upon by the companies for undergraduates engineering students. This model is applicable for colleges to improve their student's employability skills and also Industries can use this model to recruit students from different colleges through Campus Drive.

2 Data Mining

Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems.[1] It is an essential process where intelligent methods are applied to extract data patterns.[1][2] It is an interdisciplinary subfield of computer science.[1][3][4] The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.[1] Aside from the raw analysis step, it involves database and data management aspects, data preprocessing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.[1]

The actual data mining task is the semi-automatic or automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection), and dependencies (association rule mining, sequential pattern mining). This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system.

3 Decision Tree

A decision tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node. The following decision tree is for the concept that whether the student is employable or not. Each internal node represents a test on an attribute. Each leaf node represents a class.

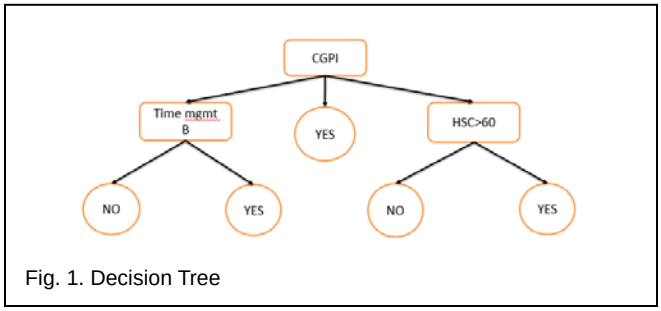

Fig. 1. Decision Tree

4 Analysis and Design

4.1 Feasibility Study

A feasibility study is carried out to select the best system that

meets performance requirements. The main aim of this activity is to determine whether it would be financially and technically feasible to develop the product. The feasibility study activity involves the analysis of the problem and collection of all relevant information relating to the product such as the different data items which would be input to the system, the processing required to be carried out on these data, the output data required to be produced by the system as well as various constraints on the behavior of the system. Technical feasibility is concerned with specifying equipment and software that will successfully satisfy the user requirement. The technical needs of the system may vary considerably, but might include: The facility to produce outputs in a given time, Response time under certain conditions, Ability to process a certain volume of transaction at a particular speed, Facility to communicate data to distant locations etc. In examining technical feasibility, configuration of the system is given more importance than the actual make of hardware. Operational feasibility is dependent on human resources available for the project and involves projecting whether the system will be used if it is developed and implemented. Operational feasibility is a measure of how well a proposed system solves the problems, and takes advantage of the opportunities identified during scope definition and how it satisfies the requirements identified in the requirements analysis phase of system development.

4.2 Pre-Design Analysis

In present competitive world an overall development of student is a key factor in empowering him to be employable. The employability is an over all outcome of different skills learned by the student as shown in Fig 2. Thus it becomes necessary to include data other than academic statistics to predict the employability.

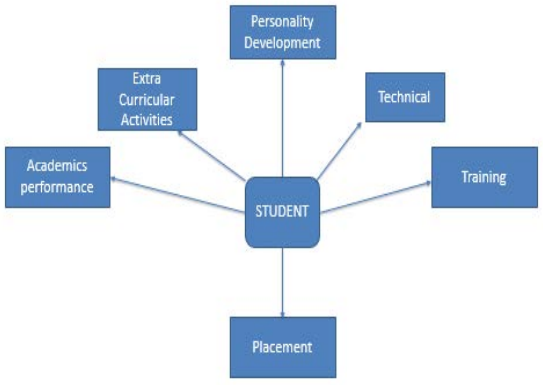

Fig. 2. Student's Skill Development Linkage

5 Methodology

- Collection of Student Data from TPO.
- Admin Login.
- Uploading Student Data.
- Filtering Student Data.
- Result.

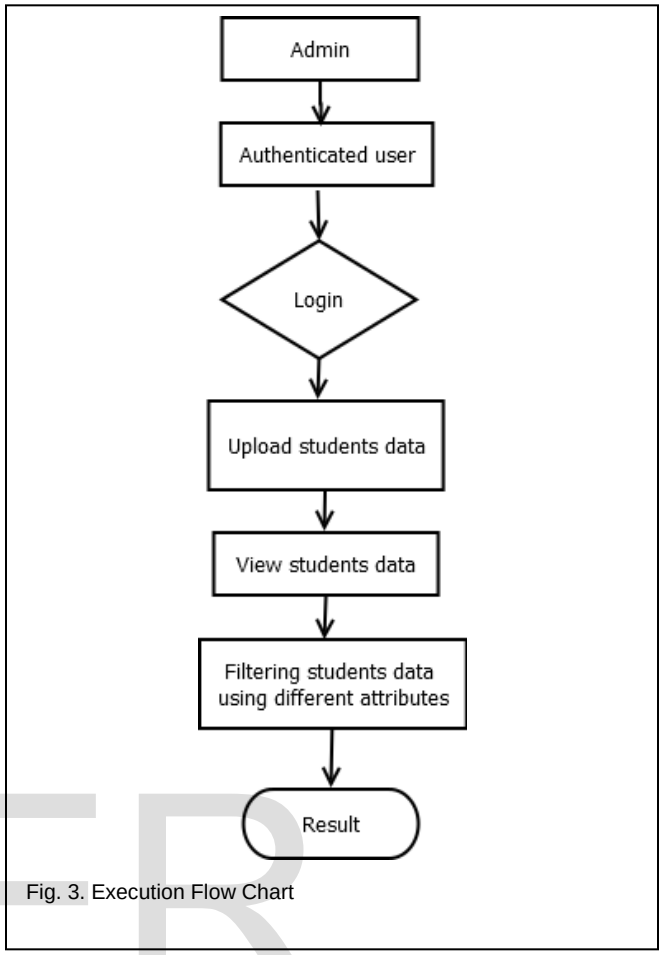

Fig. 3. Execution Flow Chart

6 Implementation

Data of various students is collected from available sources like Training and Placement cell and from the students itself. This data not only includes academic scores but also includes attributes and skills acquired by a particular student. The data

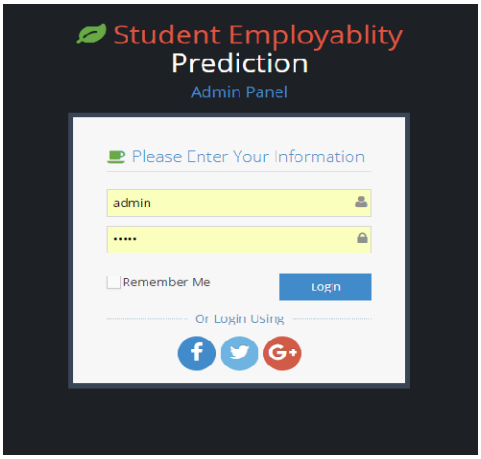

Fig. 4. Login Window

is collected in form of spreadsheet or also can be automatically synthesized from an online form which is to be filled by stu-

dents. The user of application can only login with proper credentials from a login window as shown in Fig 4. On successful login a user is asked to select the dataset on which the tool is to be used as shown in Fig 5. An upload successful popup window also appears as well as the data can be viewed in view students tab as seen in Fig 5.

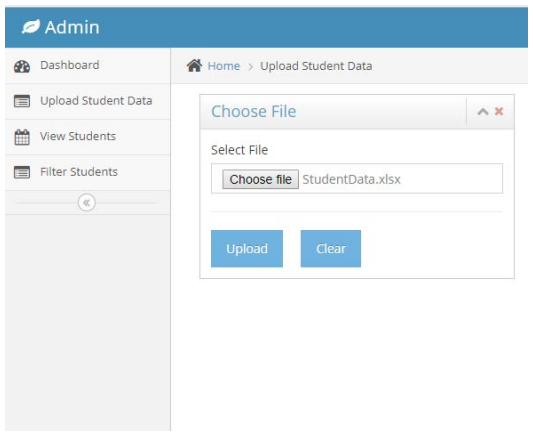

Fig. 5. Database Selection and Upload

As discussed earlier now the user has to apply filter based upon his requirement as shown in Fig.6. The different parameters available for filtering are Home Location, SSC Precentage, HSC Percentage, Average Pointer, number of KTs, Decision making, Leadership and so on. The parameters which cannot

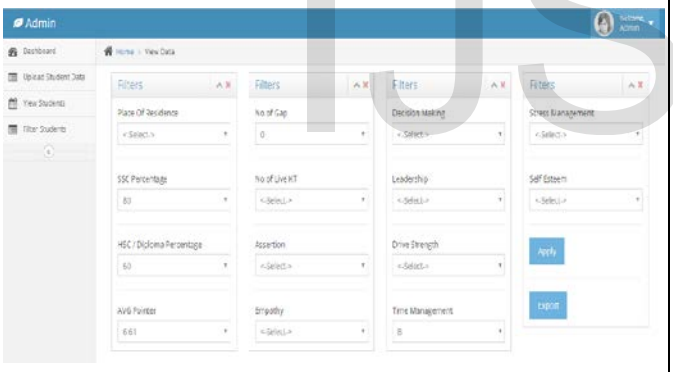

Fig. 6. Applying Filters to values of parameters

be assigned a particular values are given grades which are to be decided upon experiences and activities of every individual. As soon as a filter is applied to particular parameter the students having valid entries are display in Result section as shown in Fig.7 and can also be exported in an offline spreadsheet.

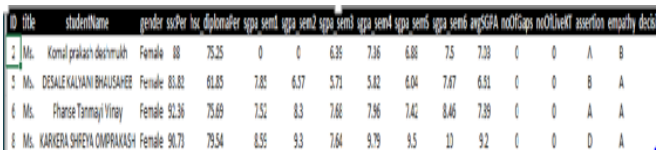

Fig. 7. Result

7 Conclusion

Data mining techniques are effective for implementation on educational data set. The result shows success in implementing educational data mining techniques to classify a decision tree model. The classification techniques have successfully predicted the number of students who are likely to pass their examination for work integrated learning placement. The research findings, shows that Decision Tree algorithm outperformed used data mining techniques in this study

Acknowledgment

We have great pleasure in presenting the paper on "Student's Employability Prediction Using Data Mining". We thank our honourable Principal Dr. M. J. Lengare for his inspiration. And we thank Prof. J. P. Patil our Head of the Department and also our project guide for the correct guidance and advice. We wish to express our gratitude for our project guide Prof. J. P. Patil, for his consistent encouragement in the successful performance of the project. We also thank our department staff who all encouraged and satisfied our needs to finish this project work. We are very happy to thank our lab assistants for giving a well-equipped lab for developing this project work. We extent our thanks and gratitude to our parents, friends and those who helped us directly and indirectly for the successful completion of this project work.

References

[1] H. Ba-Omar, I. Petrounias, and F. Anwar, "A framework for using web usage mining for personalize e-learning" *Proc. IEEE International Conference in Advanced Learning Technologies Seventh,* pp. 937938, 2007.

[2] D. Kabakchieva, "Predicting student performance by using data mining methods for classification, " *Cybernatics and Information Technologies*, vol. 13, no. 1, pp. 61-71, 2013.

[3] L. Affendey, I. Paris, N. Mustapha, M. Sulaiman, and Z. Muda, "Ranking of influencing factors in predicting students' academic performance, " *Information Technology Journal*, vol. 9 No. 4, pp. 832-837, 2010.

[4] P. Cheewaprakobkit, "Study of factor analysis affecting achievements of undergraduate," *International Multi Conference of Engineers and Computer Scientists, 2013.*

[5] A. A. Juan, T. Daradoumis, J. Faulin, and F. Xhafa, "SAMOS: a model for monitoring students' and groups' activities in collaborative e-learning," *International Journal of Learning Technology*, vol. 4, no. 1-2, pp. 53-72, 2009.

[6] Y. Yusoff, M. Omar, A. Zaharim, A. Mohamed, and N. Muhamad, "Employability skills, performance score for fresh engineering graduates in Malaysian industry, " *Asian Social Science*, vol. 18, no. 16, pp. 140-145, 2012.

[7] Q. A. AI-Radaideh, E. W. AI-Shawakfa, and M. I. AI-Najjar, "Mining student data using decision trees", *International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan, 2006.*

[8] Tanwani, Ajay Kumar, et al. "Guidelines to select machine learning scheme for classification of biomedical datasets." *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics.* Springer Berlin Heidelberg, 2009.

[9] B.K. Bharadwaj and S. Pal., "Data Mining: A prediction for performance improvement using classification", *International Journal of Computer Science and Information Security (IJCSIS)*, vol. 9, no. 4, pp. 136-140, 2011.