Educational Data Mining: Student Placement Prediction in Education Sector using Classification

A. Priyadharsini[1] Mrs. B. Azhagusundari[2]
[1]M. Phil. Scholar [2]Assistant Professor
[1,2]Department of Computer Science
[1,2]N.G.M, College, Pollachi, India

Abstract— Data mining is based on collecting knowledge from databases or data warehouses and the information collected that had never been known before, it is valid and operational. In recent years, the biggest challenges that educational institutions are dealing with the explosive growth of educational data and to use this data to improve the quality of managerial decisions. One of the biggest challenges is to improve the quality of the educational processes so as to enhance student's performance in academic and placement. Thus, it is extremely important decision to set new strategies and plans for a better management of the current processes. Educational data mining model helps to declare student's future learning outcomes using data sets of outgoing students. Prediction of student's placement in educational environments is important as well. In this paper, J48 classification algorithm is used to predict the placement for the final year students.

Key words: Educational Data Mining, Decision Tree, Classification, Student placement, J48

I. INTRODUCTION

Data mining is considered as the process of extracting important patterns from a given database and it always act as a valuable tool for converting data into usable information. Data mining has a wide range of applications in different areas that include marketing field, banking sector, educational research, surveillance, telecommunications fraud detection, and scientific discovery (Han & Kamber, 2008). More specifically, data mining can discover hidden information to support decision-making in various domains. The education data mining is one of these domains in which the primary concern is the evaluation and, in turn, enhancement of organizations based on education domain.

Data mining techniques are used to discover hidden information patterns and relationships of Educational data, which is helpful in decision making. Data mining can be applied to wide variety of applications in the educational sector for the purpose of improving the performance of students as well as the status of the educational institutions. Educational data mining is rapidly developing as a key technique in the analysis of data generated in the educational domain A single data contains valuable information. The type of information produced by the data and it decides the processing method of data. A collection of data that can produce valuable information, in education sector contains that information needed for mining, which helps the education sector to capture and compile low cost information. For this type of information and communication, technology is used. Now-a-days usage of educational database is increased rapidly because of the large amount of data that can be stored and analyzed.

Educational Data Mining (EDM) is an emerging field exploring data in educational context by applying different Data Mining (DM) techniques/tools.

Purpose of Study: This study has aims to implement several prediction techniques in data mining to assist educational institutions with predicting their student's placement. If students are predicted to have low academic performance or less chance to get the placement, then extra efforts can be made to improve their academic performance and placement activity.

II. LITERATURE SURVEY

A number of reviews pertaining to not only the diverse factors like personal, socio-economic, psychological and other environmental variables that influence the performance of students but also the models that have been used for the performance prediction are available in the literature and a few specific studies are listed below for reference.

Brijesh Kumar Baradwaj et al., describes the main objective of higher education institutions is to provide quality education to its students. One way to achieve highest level of quality in higher education system is by discovering knowledge for prediction regarding enrolment of students in a particular course, detection of abnormal values in the result sheets of the students, prediction about students' performance and so on, the classification task is used to evaluate student's performance and as there are many approaches that are used for data classification, the decision tree method is used here.

Mohammed M. Abu Tair and Alaa M. El-Halees (2012) applied the educational data mining concerns with developing methods for discovering knowledge from data that come from educational domain. Used educational data mining to improve graduate students' performance, and overcome the problem of low grades of graduate students and try to extract useful knowledge from graduate students data collected from the college of Science and Technology.

Abeer Badr El Din Ahmed, Ibrahim Sayed Elaraby (2014), currently the amount huge of data stored in educational database these database contain the useful information for predict of students performance. The most useful data mining techniques in educational database is classification. In this paper, the classification task is used to predict the final grade of students and as there are many approaches that are used for data classification, the decision tree (ID3) method is used here.

III. EDUCATIONAL DATA MINING USING CLASSIFICATION

Educational Data mining refers to extracting or "mining" knowledge from large amounts of educational data. Data mining techniques are used to operate on large volumes of

data to discover hidden patterns and relationships helpful in decision making. Currently, the data are stored in educational database, these database contain the useful information to predict students performance. The most useful data mining techniques in educational database is classification. In this paper, the classification task is used to predict the final grade of students and as there is many approaches that are used for data classification.

We have Figure 1 that represents working methodology based on the framework. It is important to have a working methodology to govern our work before applying data mining techniques. The work methodology begins with problem definition, data collection and data preprocessing that includes data selection and data transformation and it precedes with data mining classification techniques with pruning which leads to discovering knowledge that is benefit to us.

Fig. 1: Data mining work methodology

A. *Problem Statement*

The data set used in this study is obtained from NGM College of Arts and Science, Pollachi. The placement and academic marks of previous students of UG- CS (Aided) and UG-CT (self finance) is collected from the college database. The personal, academic details of the second and third year student's details directly collected from the students through questionnaire. Both details are stored in other database and it is used to predict the placement status of the second and third year students.

B. *Data Collection*

We have collected students data of final year and pre final year students through questionnaire and the details of passed out students are collected from placement cell of N.G.M, college, Pollachi. The details of students placed are also collected from the Placement cell.

C. *Preprocess*

In present day's educational system, a student's placement performance is determined by the Gender, SSLC, HSc and UG marks. UG marks are based on the internal assessment and external exam mark. The internal assessment is carried out by the teacher based upon student's performance in educational activities such as class test, seminar, assignments, general proficiency, attendance and lab work. The end semester examination is one that is scored by the student in semester examination. Each student has to get minimum marks to pass a semester in internal as well as end semester examination.

D. *Data Selection*

In this step only those fields were selected which were required for data mining. A few derived variables were selected. While some of the information for the variables was extracted from the database. All the predictor and response variables which were derived from the database are given in Table I for reference.

Variables	Description	Possible Values
Gender	Student's Gender	{Male, Female}
Dept	Department	{CT,CS}
A Type	Admission Type	{Aided, Self Finance}
Board10	Board of Studies in 10th	{CBSE, STATE}
SSLC Marks	Percentage of Marks in SSLC	{35% to 100%}
Board12	Board of Studies in 12th	{CBSE, STATE}
HSC Marks	Percentage of Marks in HSC	{35 % to 100% }
Locality	Living Locality	{Village, Town, Taluk, District}
Bag log	Arrears in college academic	{Yes, No}
UGPerc	Percentage of marks in UG	{40% to 100%}
Job	Job after UG	{Yes, No}

Table 1: Description

E. *Attribute Selection*

Totally 101 instances with 14 attributes such as register no, name, gender, department, admission type, SSLC board, SSLC marks, HSc board, HSc mark, locality, baglog, UG percentage, job after UG and locality are passed to attribute selection process using Best Fit with CFS subset algorithm based on the placement class. After completion of the evaluation of training data, the selected attributes are Gender, SSLC mark, HSc mark and UG.

IV. CLASSIFICATION

Classification is supervised learning method. It consists of two steps: 1. Model is built by analyzing the data tuples from training data. 2. Test data is used to check accuracy. There are various classification techniques such as Decision Tree algorithm that include ID3, C4.5, Bayesian Network, Neural Network and Genetic algorithm etc can be used.J48 is a java implementation of c4.5 in tool. These techniques can be used to build the classification model

A. *J48 Classification*

J48 is an extension of ID3. The additional features of J48 are accounting for missing values, decision trees pruning, continuous attribute value ranges, derivation of rules, etc. J48 is the class for generating a pruned or unpruned C4.5 decision tree.

B. Pruning

Because of the outliers this is a significant step to the result. Some instances are present in all data sets which are not well defined and differ from the other instances on its neighborhood. The classification is performed on the instances of the training set and tree is formed. The pruning is performed for decreasing classification errors which are being produced by specialization in the training set. Pruning is performed for the generalization of the tree

This classification is used to predict the placement status for the second and third year students based on the academic performance and gender of the passed out students with UG Percentage, SSLC and HSC Marks.

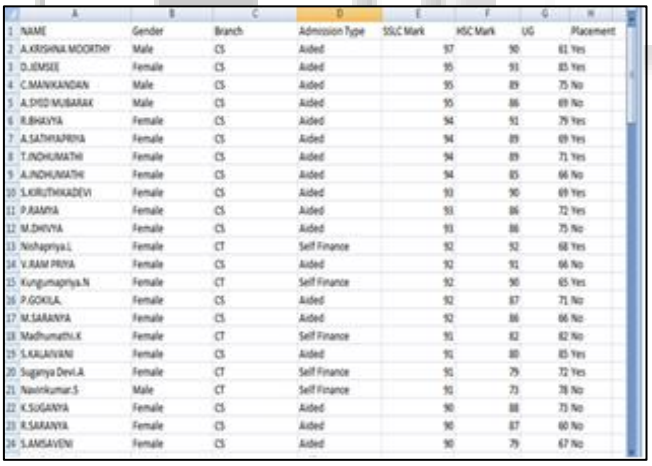
V. EXPERIMENTAL RESULT

A. Weka Tool

In the WEKA data mining tool, J48 is an open source Java implementation of the C4.5 algorithm. The WEKA tool provides a number of options associated with tree pruning. In case of potential over fitting pruning can be used as a tool for précising. In other algorithms the classification is performed recursively till every single leaf is pure, that is the classification of the data should be as perfect as possible. This algorithm it generates the rules from which particular identity of that data is generated. The objective is progressively generalization of a decision tree until it gains equilibrium of flexibility and accuracy.

B. Dataset

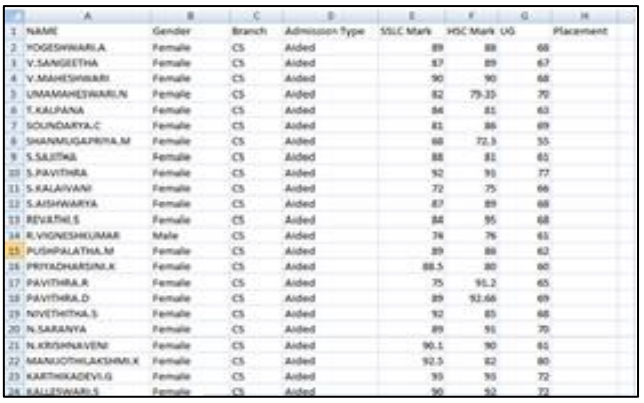
Passed out student details collected from placement cell are used as training set to evaluate. Figure 2 specify the dataset set used.



NAME	Gender	Branch	Admission Type	SSLC Mark	HSC Mark	UG	Placement
A.KRISHNA MOORTHY	Male	CS	Aided	97	90	61	Yes
D.JEMSEE	Female	CS	Aided	95	91	85	Yes
C.MANIKANDAN	Male	CS	Aided	95	89	75	No
A.SYED MUBARAK	Male	CS	Aided	95	86	69	No
R.BHAVYA	Female	CS	Aided	94	91	79	Yes
A.SATHYAPRIYA	Female	CS	Aided	94	89	69	Yes
T.INDHUMATHI	Female	CS	Aided	94	89	71	Yes
A.INDHUMATHI	Female	CS	Aided	94	85	66	No
S.KIRUTHIKADEVI	Female	CS	Aided	93	90	69	Yes
P.RAMYA	Female	CS	Aided	93	86	72	Yes
M.DHIVYA	Female	CS	Aided	93	86	75	No
Nishapriya.L	Female	CT	Self Finance	92	92	68	Yes
V.RAM PRIYA	Female	CS	Aided	92	91	66	No
Kungumapriya.N	Female	CT	Self Finance	92	90	65	Yes
P.GOKILA	Female	CS	Aided	92	87	71	No
M.SARANYA	Female	CS	Aided	92	86	66	No
Madhumathi.K	Female	CT	Self Finance	91	82	62	No
S.KALAIVANI	Female	CS	Aided	91	80	85	Yes
Sugarya Devi.A	Female	CT	Self Finance	91	79	72	Yes
Navinkumar.S	Male	CT	Self Finance	91	73	78	No
K.SUGANYA	Female	CS	Aided	90	88	71	No
R.SARANYA	Female	CS	Aided	90	87	60	No
S.AMSAVENI	Female	CS	Aided	90	79	67	No

Fig. 2: Passed out Student's Data Set – Training File

Data collected from final year students and pre final year students through questionnaire are used for test data evaluation. Figure 3 show the dataset used for testing.



NAME	Gender	Branch	Admission Type	SSLC Mark	HSC Mark	UG	Placement
HOGESHWARI.A	Female	CS	Aided	89	88	68	
V.SANGEETHA	Female	CS	Aided	87	89	67	
V.MAHESHWARI	Female	CS	Aided	90	90	68	
UMAMAHESWARI.N	Female	CS	Aided	82	79.33	70	
T.KALPANA	Female	CS	Aided	84	81	63	
SOUNDARYA.C	Female	CS	Aided	81	86	69	
SHANMUGAPRIYA.M	Female	CS	Aided	88	72.3	55	
S.SALITHA	Female	CS	Aided	88	81	61	
S.PAVITHRA	Female	CS	Aided	92	91	77	
S.KALAIVANI	Female	CS	Aided	72	75	66	
S.AISHWARYA	Female	CS	Aided	87	89	68	
REVATHI.S	Female	CS	Aided	84	95	68	
R.VIGNESHKUMAR	Male	CS	Aided	76	76	61	
PUSHPALATHA.M	Female	CS	Aided	89	86	62	
PRIYADHARSINI.K	Female	CS	Aided	88.5	80	60	
PAVITHRA.R	Female	CS	Aided	75	91.2	65	
PAVITHRA.D	Female	CS	Aided	89	92.66	69	
NIVETHITHA.S	Female	CS	Aided	92	85	68	
N.SARANYA	Female	CS	Aided	89	91	70	
N.KRISHNAVENI	Female	CS	Aided	90.1	90	61	
MANJUTHILAKSHMI.K	Female	CS	Aided	92.5	82	80	
KARTHIKADEVI.G	Female	CS	Aided	93	93	72	
KALEESWARI.S	Female	CS	Aided	90	92	72	

Fig. 3: Final year's student's test file

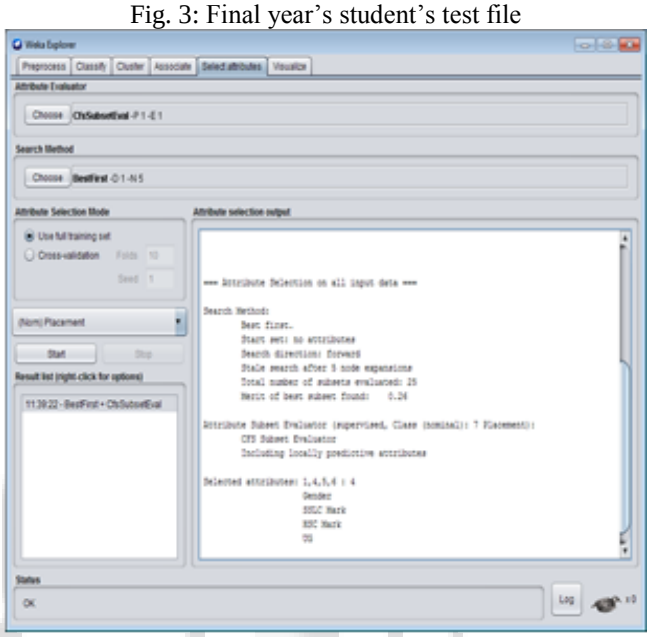


Fig. 4: Attribute selection for passed out student's dataset J48 Pruned Tree

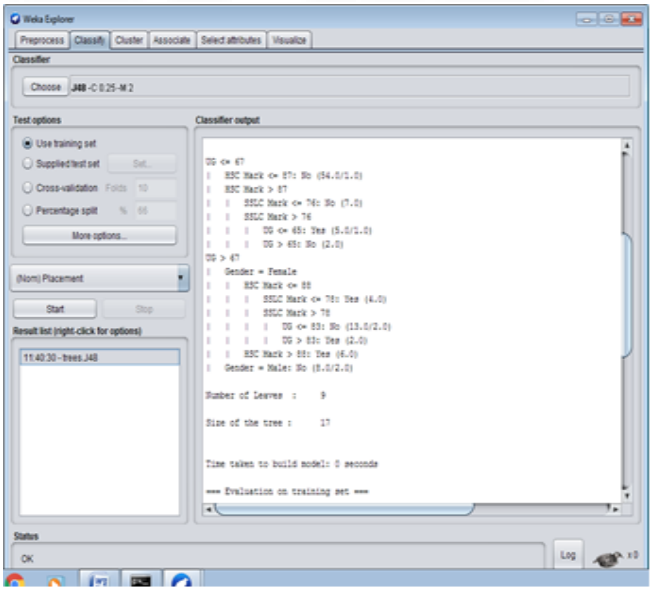


Fig. 5: J48 Pruned Tree for Passed out student's preprocessed data set

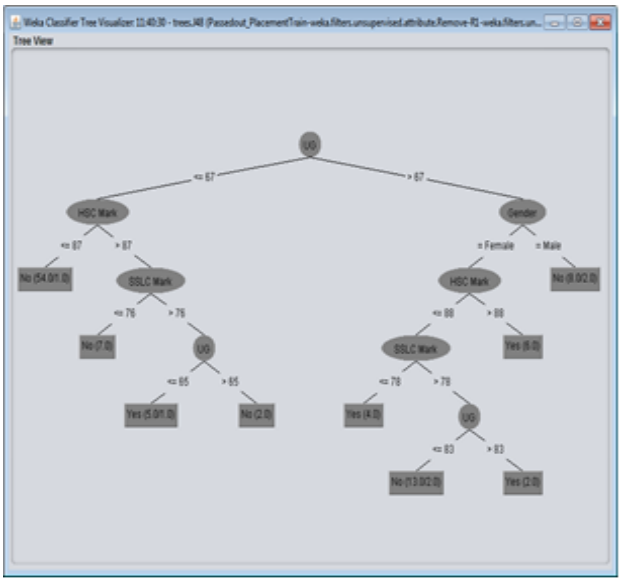


Fig. 6: J48 Tree for Placement Prediction

On evaluating the data set under J48 classifier the result generated with the correctly classified instance by 94.05 %, F-Measure value 0.938 and Recall value 0.941.

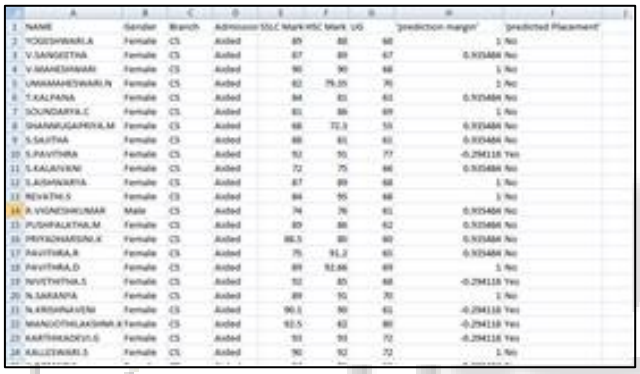


Fig. 6: Predicted placement result file J48 pruned tree

VI. CONCLUSION

In this paper, the classification rule is used predicts the placement performance of the second and third year students based on the passed out student's academic and placement performance.

As there exist many approaches that are used for data classification, the J48 algorithm is used here. Information's like Board of studies in SSLC and HSC, Marks obtained in SSLC and HSC, Attendance, Seminars, Assignments, Paper presentation, bag logs, category, living locality, percentage were collected from the passed out student's database and also current students.

This study will help to the students and the professors to improve the placement performance of those who are at the risk of less chance in campus interview selection. This study will also work to identify those students who needed special attention to placement interviews and taking appropriate action for the placement related activity.

REFERENCES

[1] Han, J. Kamber. M., "Data Mining: concepts and techniques. 2nd Edition, Morgan Kaufmann publishers (2008).
[2] Brijesh Kumar Baradwaj, Saurabh Pal, "Data mining: machine learning, statistics, and databases", 1996.
[3] Mohammed M. Abu Tair, Alaa M. El-Halees, "Mining Educational Data to Improve Students' Performance: A Case Study", 2012.
[4] Abeer Badr El Din Ahmed, Ibrahim Sayed Elaraby, "Data Mining: A prediction for Student's Performance Using Classification Method", World Journal of Computer Application and Technology 2(2): 43-47, 2014.
[5] Udeni Jayasinghe, Anuja Dharmaratne, Ajantha Atukorale, "Students' Performance Evaluation in Online Education System Vs Traditional Education System", IEEE 2015 12th International Conference on Remote Engineering and Virtual Instrumentation (REV).
[6] Jiawei Han, Micheline Kamber, "Data Mining: Concepts and Techniques, 2nd edition", 2006.
[7] P. Ajith, M.S.S.Sai, B. Tejaswi, "Evaluation of Student Performance: An Outlier Detection Perspective", 2013.
[8] Varun Kumar, Anupama Chadha, "An Empirical Study of the Applications of Data Mining Techniques in Higher Education", 2011.
[9] Hongjie Sun, "Research on Student Learning Result System based on Data Mining", 2010.

