A Comparative Study of Engineering Students Pedagogical Progress

Khalid Mahboob

Dept. of Computer / Software Engineering
Sir Syed University of Engineering & Technology
Karachi, Pakistan

Danish Ur Rehman Khan

Dept. of Electronics Engineering
N.E.D University of Engineering & Technology
Karachi, Pakistan

Syed Abbas Ali

Dept. of Computer & Information System Engineering
N.E.D University of Engineering & Technology
Karachi, Pakistan

Fayyaz Ali

Dept. of Software Engineering
Sir Syed University of Engineering & Technology
Karachi, Pakistan

Abstract—**Students' pedagogical progress plays a pivotal role in any educational institute in order to pursue imperative education. Educational institutes, Universities, Colleges implement various performance measures in order to keep analyzing and tracking progress of students to cultivate benefits of education in a better way. There are several data mining techniques to apply on education in order to build constructive educational strategies and solutions. This study aims to analyze and track engineering under graduate student's records to judge quality education, student motivation towards learning, and student pedagogical progress to maintain education at high quality level and predicting engineering student's forthcoming progress. Different engineering discipline students' (of three different cohorts) data have been analyzed for tracing current as well as future pedagogical progress based on their sessional (pre-examination) marks. In this research, the classification techniques by k-nearest neighbor, Naïve Bayes and decision trees are applied to evaluate different engineering technologies student's performance and also there are different methodologies that can be used for data classification.**

Keywords—*Pedagogical progress; classification; k-nearest neighbor; Naïve Bayes; decision trees; engineering students*

I. INTRODUCTION

Following higher education is a challenging stage for students as well as educational institutes to deal with huge amount of data. There are various applications used in educational environment in the form of archives, images, blogs, audios, videos, artifacts, scientific documents, meta-data or online hyperlinks or in various other formats [1]. Educational institutes with variety of educational data like student attendance records, examination records, fees records, personal information, etc., entails to be managed and tracked time to time. Therefore, data mining techniques have been used to discern and extract certain patterns that are potentially expedient in the domain of education at any level. Thus, educational data mining can be regard as an interdisciplinary field that assists methods of extracting useful information from enormous sets of data [2]. The advancements in the field of educational data mining have made it possible to perform academic analysis in an innovative ways to focus on educational institutes' efficacy and to reduce student retention [3].

For the successful adoption of educational data mining, it is very necessary to have wide-ranging of pedagogical data so that the various data mining techniques can be applied on that to enhance learning process after analyzing students' academic data to monitor pedagogical progress with improvements, to predict and improve retention of student at an early stage, and to analyze the chances of failures or mistakes by the students in a learning system [4]. Predicting and enhancing students learning has become more challenging in order to improve student's grades and thus benefit educational institutes in adaptation of different learning strategies [5].

Educational data mining focuses on quantitative and qualitative analysis of data necessary to use many techniques based upon multiple disciplines like machine learning, artificial intelligence, expert system, pattern matching, decision support system and it involves the uses of these techniques in an effective ways like if a student/learner may intend to improve learning skills by acquiring e-learning method. Similarly, if a teacher/instructor may require identifying students learning performance by analyzing student academic or prelim records [3]. Any educational institute or university may uses data mining methodologies to determine that how the results of students can be enhanced along with taking attention to reduce student retentions [3]. There have been already many data mining methods applied in the field of education at multiple levels. So many parameters to analyze and track student pedagogical progress have been undertaken using data mining techniques at various levels: at a training system level to predict whether selected or specific knowledge/skills are mastered, at a course level or degree level to predict whether a student will be capable to pass a course or a degree, or to predict their marks [5]. The present study aims to initially identify and investigate the students score based upon pre-final or prelim marks throughout the academic session and finding ways to improve student performance.

Educational data mining deals with several data classification techniques such as Decision Trees, Naïve Bayes,

K-Nearest Neighbor, Neural Networks, Support Vector Machines, Quadratic classifiers, and many others [6], [7]. The information derived from applying these techniques can be used for tracking and monitoring student's pedagogical progress in various disciplines to probe students' performance in different courses [6]. This paper is planned as follows: Section II explains the related works. Section III briefly describes the different data mining classification techniques. Section IV defines the data preprocessing and methodology. The following section i.e. Section V presents the results and discussion used in this case study. The conclusion elaborates on useful findings, discourses them and presents future works.

II. RELATED WORKS

Much of the work has been done on educational data mining applying different techniques on different levels of education using different tools. This section contributes on the related work in the domain of education using data mining techniques:

Diego Buenaño Fernández et al. [1] have compared three open source tools (Weka, Rapid Miner, and Knime) on different academic records of students. Analysis based on three engineering programs (Network and Telecommunications Engineering, Computer and Information Systems Engineering and Electronics and Information Networks Engineering.) of a University. Most of the relevant tasks were consisted on the data pre-processing. Particularly, the K-Means algorithm is deployed on different attributes of academic data with the applicability of four specific algorithms: ChiSquaredAttributeEval, FilteredSubsetEval, GrainRatio-AttributeEval and OneRAttributeEval. The result shows that the three tools were very similar in working in terms of precision. Results obtained after detailed cluster analysis that can help teachers or instructors to guide students in a course with suitable measures on time.

Raheela Asif et al. [4] presented a case study based on predicting student academic performance at the end of university degree of the degree program at early stage which can help universities to emphasis on skilled students and to initially detect with low educational accomplishment and find effective ways to support them. The four academic cohorts' data comprising 347 undergraduate students have been extracted by using different classifiers. Artificial neural networks, decision trees, k-nearest Neighbor, naive Bayes, and rule induction classification techniques have been used in a study. The study has shown the possibility of predicting fourth year graduation performance at university with pre-university marks and initial university two year marks. The accuracy on analyzing the datasets was satisfactory. Further five courses were evaluated to predict the performance of graduation that did not lead to a better accuracy.

Nawal Ali Yassein et al. [5] have proposed patterns in the available student and courses records to predict students' performance. Statistical package for social sciences (SPSS) and data mining tool (clementine) were used for experimentation purpose. The research comprises of two parts: first to find out the factors related to course success rate, second to determine predictors based on student performance. Both classification and clustering techniques were used to analyze different features that can affect student performance in a course(s).

Brijesh Kumar Baradwaj et al. [6] the classification – decision trees induction method is utilized to evaluate students' academic performance at the end of semester examination to identify early dropouts and to find out students' who might need attention and counseling analyzing diverse academic parameters like student attendance, assignments, class tests, seminar, etc.

T.Archana et al. [8] surveyed on focusing student's performance prediction by improving performance and by increasing student retention in order to increase the quality of education and can be valid in different environments.

Kalpesh Adhatrao et al. [9] have rendered a system to predict the performance of students based on their preceding performances under the concept of data mining classification. ID3 (Iterative Dichotomiser 3) and C4.5 classification algorithms have been applied to predict the performance of fresh students generally and individually.

Ashish Dutt et al. [10] provided the applicability and usability of a clustering method in the context of educational data mining consisting over three decades systematic literature review. The crucial benefit of clustering algorithm towards data analysis is that it provides relatively an explicit schema of learning ways of students specified a number of variables like completing learning tasks on time, groups learning, class learner behavior, classroom decoration and student learning motivation. Clustering can provide relevant intuitions to variables that are applicable in splitting the clusters.

Raheela Asif et al. [11] have predicted student performance using different data mining techniques based on pre-university marks and examination marks of initial years at university. Two cohorts of Civil Engineering technology data were analyzed and various classifiers were applied on that data with a reasonable accuracy. The decision trees were used as an indicator to detect the courses with low performance in order to give warning to students earlier in the degree program.

Raheela Asif et al. [12] performance of students' progress has been investigated by analyzing the data of two immediate cohorts and applying k-means clustering. Each student was characterized by 4-tuple with his/her average remain the same, or either increase or decrease when comparing to their preceding years. Different ranges of accuracies obtained by using different classifiers.

Raheela Asif et al. [13] two aspects of under-graduate students have been studied were: firstly to predict the students' academic accomplishment of four year study programme at the end and secondly analyzing typical progresses and combining them with the results of predictions. The results drawn were the possibility of prediction of graduation performance using pre-university and initial two years university marks only with a reasonable accurateness. Few courses were also put into focus as an indicator regarding good or bad performances with respect to low, intermediate and high marks.

Surjeet Kumar Yadav et al. [14] Decision tree algorithms have been applied on students' previous academic data to

produce a model that can be helpful to predict the students' academic performance in order to detect early drop outs of students. CART algorithm among others classification algorithms disclosed the best results for data classification.

Raheela Asif et al. [15] investigated the academic performance of students by applying X-means clustering technique analyzing the data of two immediate cohorts. Shifting of marks from high to low or low to high (or vice versa) in different academic years for both cohorts has been observed. It has been reported that there is a possibility of using one cohort in order to predict the performance of the succeeding cohort with varying accuracies using different classifiers.

III. Data Mining Approaches and Techniques

Data mining is a computational study of data processing which has been successfully functional in many areas that intend to extract useful knowledge from that data [5]. There are various techniques of data mining that are operate-able on large volumes of data to find out hidden patterns and their relationships helping in decision making for different applications such as Artificial Intelligence, Business Management, Decision Support, Machine Learning, Market Analysis, and Statistical and Database Systems [6], [9]. Likewise, several data mining algorithms and techniques are used in knowledge discovery from large databases such as Association Rules, Classification, Clustering, Decision Trees, Genetic Algorithm, Nearest Neighbor methods, Regression, etc. [6].

Among others, classification is a data mining technique, particularly, which plots data into predefined classes or groups [5], [9]. Classification is specifically used for predicting the unknown class label of data objects [16]. This is considered as the most commonly applied technique [6]. This approach often employs Classification (IF-THEN) Rules, Decision Trees or .Neural Networks [2]. In classification, the accuracy of the classification rules are estimated by using training data sets [6]. The classification can works on different training data sets by constructing a model or classifier. Building a classifier or model is the initial step in the learning phase. The classification algorithms are used in building the classifier with the set of parameters essential for proper discrimination [2], [6].

There are many classifiers to implement data mining methods in order to perform in a better way can also apply in the education domain [4]. There are some classifiers which outperforms better than others. Here is the summary of three famous classification techniques, i.e. decision trees, Naïve Bayes and k-nearest neighbor have been opted for this study.

A. Decision Trees

A decision tree is a non-cyclic tree structure which consists of root node, connecting branches and internal nodes (leaf nodes) [2], [4]. Each leaf node corresponds to an attribute denotes a test on it and holds the class label whereas each branch from a sequential path denotes the test outcome. The node at the topmost of the tree called the root node which represents the entire datasets [2], [4]. The tree always starts with the single node containing training datasets [16]. If the tuples in a dataset belongs to the same class then the node turns into a leaf, labeled with that corresponding class [16]. Otherwise, an attribute selection method is used to determine the splitting criterion. Such a method may use a heuristic or statistical measure (e.g., information gain or gini index) to select the best way to separate the tuples into individual classes [16].

B. Naïve Bayesian Classifier

In terms of machine learning, Naive Bayes classifier is a kind of simple probabilistic model to solve problems controlled by strong independence assumptions [17]. It's highly scalable and fast to train data very efficiently in a supervised learning situation with high accuracy in numerous applications [4], [17]. We have a set of unknown tuples (instances), embodied by an n-dimensional vector, $X = (x_1, x_2, ..., x_n)$, with the probabilities of instances $P(C_i|X)$ (where i is possible outcome of a class). The posterior probability can be decomposed as:

$$P(C_i|X) = \frac{P(C_i)\ P\ (X|C_i)}{P(X)}$$

C. k-Nearest Neighbors Algorithm

The k-nearest neighbors' algorithm (k-NN) is regard as non-parametric method in the field of pattern recognition to classify records based on similarity measures learning [4], [18]. Two records or objects are measured by the distance between them based on the likeness of two records [4]. The output is considered as a class membership [18]. A record or object remains classified by a majority vote of its neighbors, or in other words, the k records with the minimum distance to the anonymous record with k is a positive integer and typically small [4], [18]. If k = 1, then the record or object is merely assigned to the class of that single nearest neighbor [18]. Hence, the k-NN algorithm is simplest among the machine learning algorithms.

IV. Data Preprocessing and Methodology

In the field of data mining, data preprocessing is a crucial step to deal with incomplete, noisy and inconsistent data [19]. Data preprocessing includes various tasks such as data cleaning, data integration, data transformation, data reduction, data discretization, etc. to continuously formulate data in a consistent and accurate style. For our case study, the data has been collected and preprocessed of three different cohorts (two class sections each) of three different engineering disciplines (Computer Engineering, Software Engineering, and Electronic Engineering) of the SSUET, Pakistan. The students' pedagogical progress is analyzed by taking single (core) course sessional marks of different technologies. The research focuses on three comparative studies in order to track and analyze engineering student's pedagogical progress are: comparative analysis of a performance of three different courses (use to teach in different engineering technologies), comparative analysis of gender wise performance in each technology and comparative study between two sections students' performance in a particular course.

Overall 290 undergraduate engineering students enrolled in academic batches 2011 Computer Engineering with 102 students (Section D and E), 2014 Software Engineering with 94 students (Section A and B) and 2017 Electronic Engineering

with 94 students (Section C and D) has been comparatively analyzed for academic progress in their particular course in different semesters. Only pre-examination (sessional) marks of students have been used in this study. Different variables as arbitrating parameters have been selected to measure students' academic progress. Different parameters and response variables varying according to technology and course are mentioned in a Table I for reference.

TABLE I.   SESSIONAL VARIABLES IN DATASET

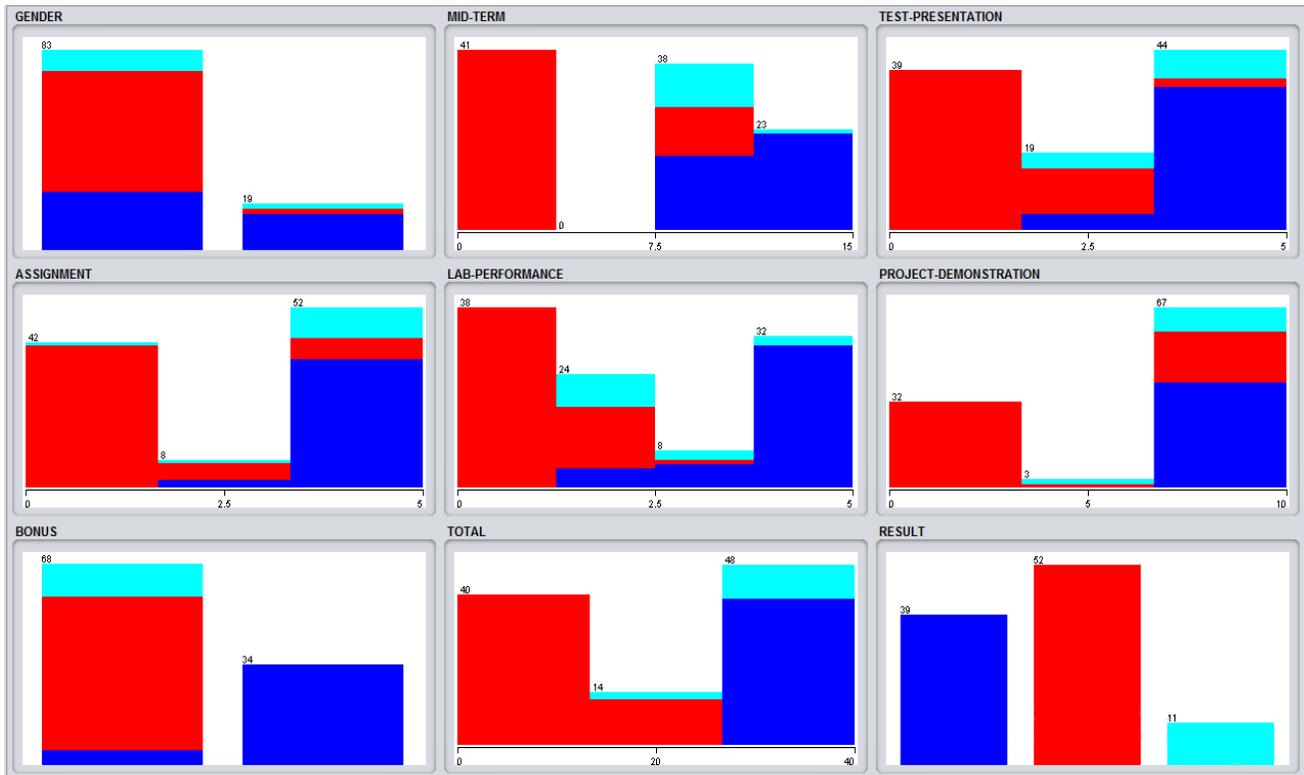| Variable	Values	Technology	Course	Possible Results
Gender	{Male, Female}	C.E, S.E, E.E	RDBMS, DS&A, OOP	
Mid-Term Marks	{12-15, 11-7, <=6}	C.E	RDBMS	{Excellent, Average, Poor}
	{15-20, 10-14, <=9}	S.E, E.E	DS&A, OOP	
Test / Presentation Marks	{4-5, 3, <=2}	C.E	RDBMS	{Excellent, Average, Poor}
Assignment Marks	{4-5, 3, <=2}	C.E	RDBMS	{Excellent, Average, Poor}
Lab Performance	{4-5, 3, <=2}	C.E	RDBMS	{Excellent, Average, Poor}
	{15-20, 10-14, <=9}	S.E, E.E	DS&A, OOP	
Project Demonstration	{8-10, 7-5, <=4}	C.E	RDBMS	{Excellent, Average, Poor}
Bonus Marks	{Yes, No}	C.E, S.E	RDBMS, DS&A	
Quiz/Test Marks	{7-10, 5-6, <=4}	S.E, E.E	DS&A, OOP	{Excellent, Average, Poor}
Total Marks	{32-40, 24-31, <=23}	C.E	RDBMS	{Excellent, Average, Poor}
	{35-50, 34-25, <=24}	S.E, E.E	DS&A, OOP	

Fig. 1.   Overall attributes visualization of RDBMS.

The course name acronym are RDBMS stands for Relational Database Management System, DS&A stands for Data Structures and Algorithm, and OOP stands for Object Oriented Programming. The preprocessing using Weka tool

[20] to analyze three different courses of different technologies has been presented below in the form of graphical visualization and statistical analysis of the attributes as mentioned in Table I. For the course RDBMS and technology Computer Engineering, 83 instances of male and 19 instances of females have been analyzed in the dataset. The preprocessing results clearly shows in Fig. 1 that 39 instances were Excellent progress (visualized in a dark blue bar), 11 were Average progress (visualized in a cyan bar), and 52 were Poor progress (visualized in a red bar) mined from the Result attribute.

For the course DS&A and technology Software Engineering, 59 instances of male and 35 instances of females have been analyzed in the dataset. The preprocessing results clearly shows in Fig. 2 that 69 instances were Excellent (visualized in a dark blue bar), 12 were Average (visualized in a cyan bar), and 13 were Poor (visualized in a red bar) mined from the Result attribute.
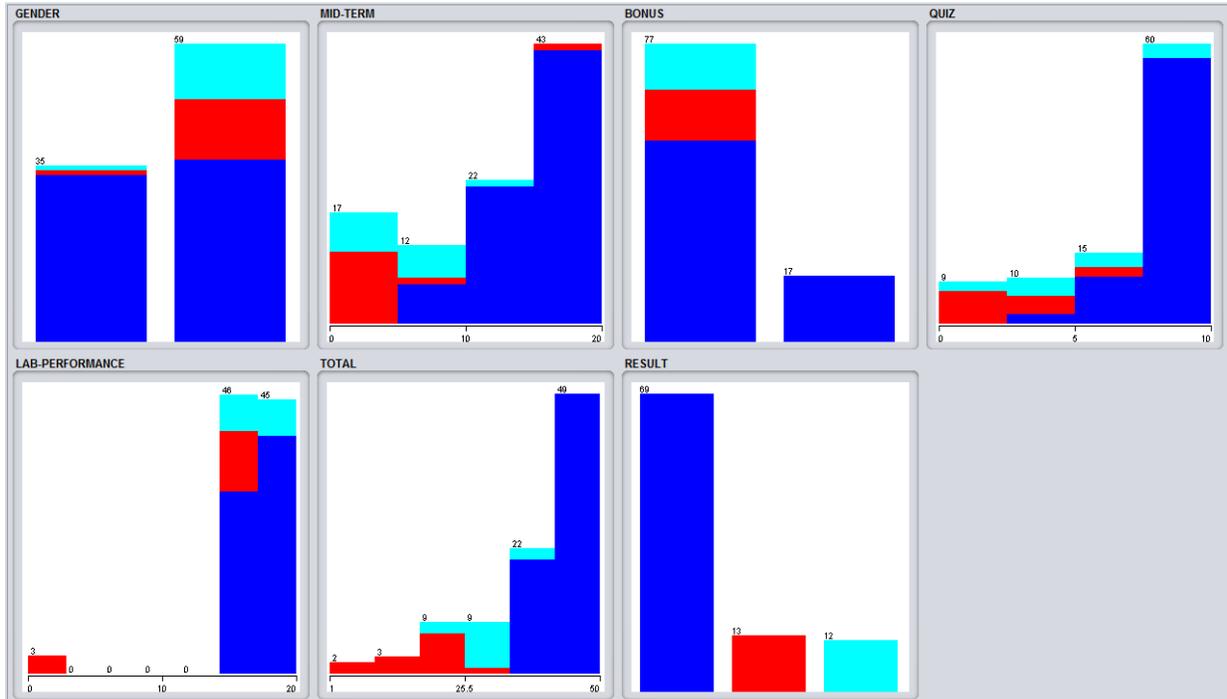

Fig. 2.   Overall attributes visualization of DS&A.
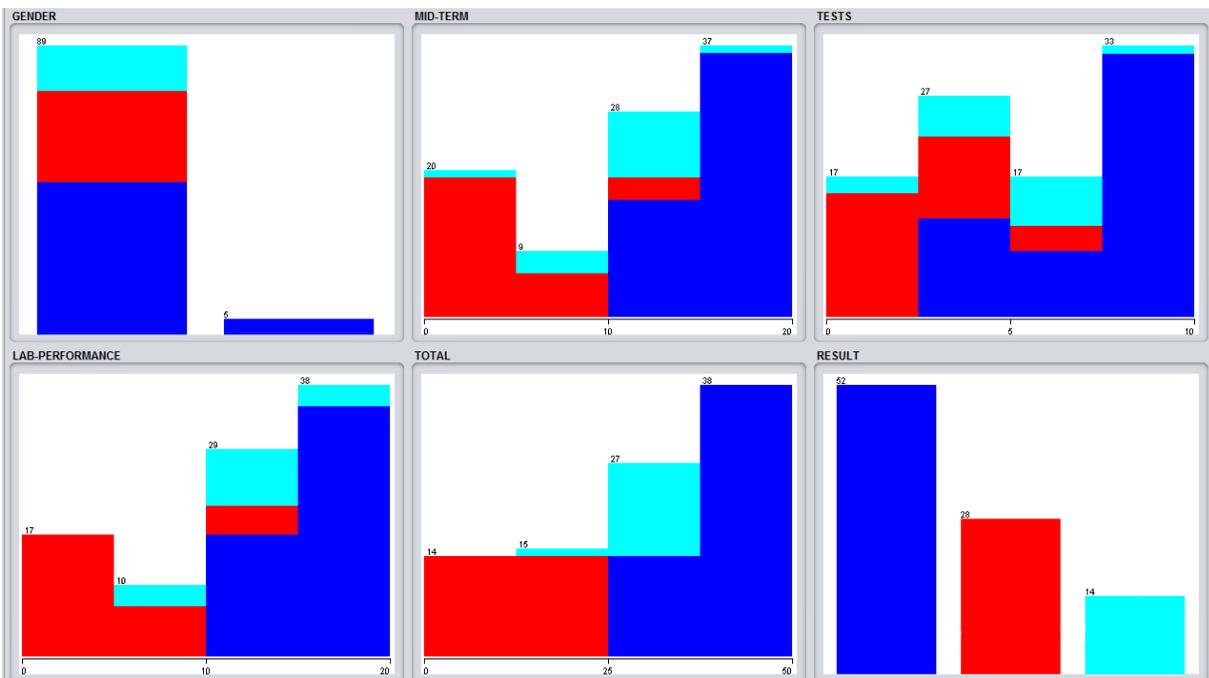

Fig. 3.   Overall attributes visualization of OOP.

For the course DS&A and technology Software Engineering, 59 instances of male and 35 instances of females have been analyzed in the dataset. The preprocessing results clearly shows in Fig. 2 that 69 instances were Excellent progress (visualized in a dark blue bar), 12 were Average progress (visualized in a cyan bar), and 13 were Poor progress (visualized in a red bar) mined from the Result attribute.

For the course OOP and technology Electronic Engineering, 89 instances of male and 5 instances of females have been analyzed in the dataset. The preprocessing results clearly shows in Fig. 3 that 52 instances were Excellent progress (visualized in a dark blue bar), 11 were Average progress (visualized in a cyan bar), and 52 were Poor progress (visualized in a red bar) mined from the Result attribute.

The Weka (Waikato Environment for Knowledge Analysis) tool (Knowledge Explorer) is the useful GUI (Graphical User Interface) tool with the collection of major machine learning algorithms coded in java [1], [14], [20]. It basically contains applications for data pre-processing, association rules, classification, clustering, regression and visualization [1]. Engineering under-graduate students of three different cohorts of three different technologies with three different courses have been studied comparatively to trace their pedagogical progress using sessional (pre-examination marks) in particular courses. The tool Weka was employed to carry out study using three different classification algorithms Decision trees J48, Naïve Bayes, and K-Nearest Neighbor [20]. The reason of particularly using these three algorithms for classification beside other algorithms in a study is that they give better results and represent rules which can be simply interpretable by humans and therefore can be used in making decision rules [4].

V. RESULTS AND DISCUSSION

The course wise results per technology and per gender of accuracy, kappa, mean absolute error, root mean squared error, relative absolute error, and root relative squared error with different classifiers are organized in Table II. The kappa statistic basically measures the agreement of prediction with the true class -- 1.0 indicates complete agreement. The Mean Absolute Error simply measures the average degree of the errors in a set of estimates, without considering their direction. It simply measures accuracy for continuous variables. The Root Mean Squared Error is a quadratic scoring rule which mainly measures the average degree of the error. Relative values are simply ratios, and have no units. The ratios are commonly expressed as fractions (e.g. 0.762), as percent (fraction x 100, e.g. 76.2%), as parts per thousand (fraction x 1000, e.g. 762 ppt), or as parts per million (fraction x 106, e.g. 762,000 ppm).

Fig. 4 clearly indicates the results of comparative study of three courses and per gender per three different technologies in terms of accuracy using three different classifiers for data classification. From Table II and Fig. 4, J48 and k-NN achieves the highest accuracies for the course RDBMS whereas k-NN achieves the highest accuracy when analyzing the results based on gender in the same course. In the same manner, k-NN has maximum accuracy for the course DS&A as compared with the others two classifiers and for gender analysis as well. Likewise, J48 and k-NN attains highest accuracies for the course OOP whereas k-NN was best in accuracy than the rest of the two classifiers for gender analysis.
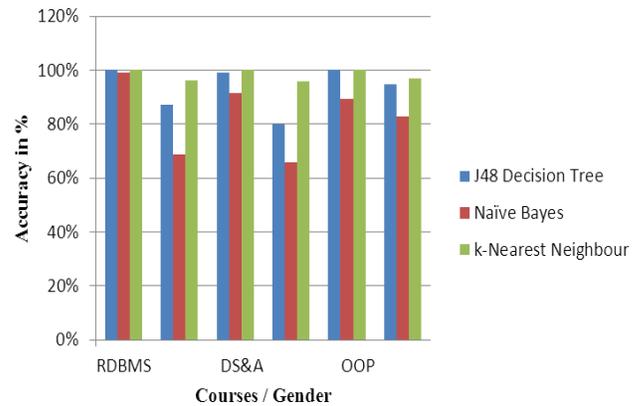

Fig. 4. Comparison of classifiers accuracy.

For J48 algorithm, the decision trees are obtained based on above study results shown in Fig. 5 to 9.

By observing the decision trees, the students' progress in different courses of different engineering technologies can be analyzed and tracked in order to find out academic strengths and weakness. The tree in Fig. 5 indicates that the students who scored above 31 marks have excellent progress with 39 students, those who scored above 23 and less than equal to 31 have average progress with 11 students, and those who scored below and equal to 23, they have a poor progress with 52 students in a course RDBMS.

TABLE II. CORRELATION RESULTS COURSE WISE PER TECHNOLOGY PER GENDER

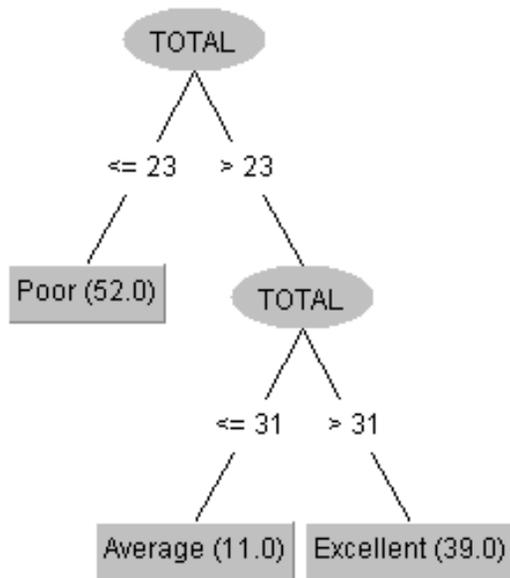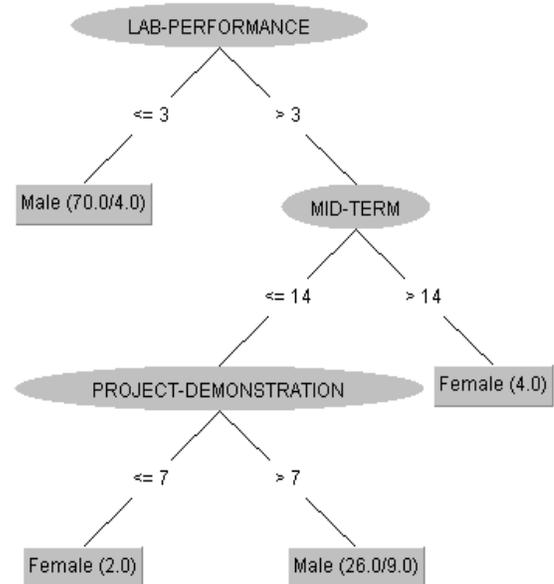| Course / Gender	Classifier	Accuracy		Kappa statistic		Mean absolute error		Root mean squared error		Relative absolute error		Root relative squared error	
RDBMS (CE)	J48 Decision Tree	100 %	87.2549 %	1	0.4289	0	0.1893	0	0.3077	0 %	61.6848 %	0 %	79.019 %
	Naïve Bayes	99.0196 %	68.6275 %	0.9832	0.3401	0.0297	0.3081	0.1081	0.5389	7.6244 %	100.3648 %	24.5272 %	138.3908 %
	k-Nearest Neighbour	100 %	96.0784 %	1	0.8592	0.0089	0.0542	0.0108	0.1553	2.2722 %	17.6433 %	2.4624 %	39.8842 %
DS&A (SE)	J48 Decision Tree	98.9362 %	79.7872 %	0.975	0.589	0.0131	0.2934	0.0809	0.383	4.5336 %	62.6884 %	21.4719 %	79.2323 %
	Naïve Bayes	91.4894 %	65.9574 %	0.8116	0.3542	0.0576	0.354	0.1833	0.5338	19.9468 %	75.6293 %	48.6552 %	110.4211 %
	k-Nearest Neighbour	100 %	95.7447 %	1	0.911	0.0114	0.0637	0.0127	0.1671	3.96 %	13.6103 %	3.3748 %	34.5606 %
OOP (EE)	J48 Decision Tree	100 %	94.6809 %	1	0	0	0.1007	0	0.2244	0 %	92.3716 %	0 %	99.9141 %
	Naïve Bayes	89.3617 %	82.9787 %	0.8277	0.272	0.0611	0.1778	0.178	0.3557	15.6583 %	163.018 %	40.3704 %	158.3742 %
	k-Nearest Neighbour	100 %	96.8085 %	1	0.558	0.0092	0.0382	0.0114	0.1266	2.3657 %	35.0721 %	2.5962 %	56.3469 %

Fig. 5. Decision tree produced for RDBMS.

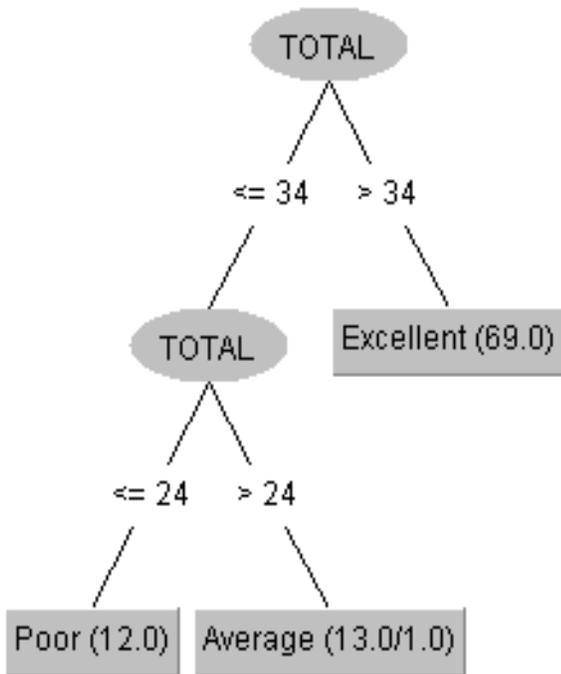Fig. 6. Decision tree produced for gender (in lab performance and project demonstration).

Fig. 7.   Decision tree produced for DS&A.

Fig. 8.   Decision tree produced for gender (in a quiz and lab performance).

The tree in Fig. 6 indicates that particularly in lab-performance and in a mid-term, female students' progress is better than male students whereas male students were comparatively better than female students in a project demonstration of Computer Engineering technology.

The tree in Fig. 7 indicates that the students who scored above 34 marks have excellent progress with 69 students, those who scored above 24 and less than equal to 34 have average progress with 13 students (one misclassified record), and those who scored below and equal to 24, they have a poor progress with 12 students in a course DS&A.

The tree in Fig. 8 indicates that particularly in a quiz and mid-term, female students' progress is comparatively better than male students whereas male students were better than female students in a lab performance and bonus marks of Software Engineering technology.

The tree in Fig. 9 indicates that the students who scored above 34 marks have excellent progress with 52 students, those who scored above 20 and less than equal to 34 have average progress with 14 students and those who scored below and equal to 20, they have a poor progress with 28 students in a course OOP.

The tree obtained for gender wise comparison in a course OOP was an incomplete tree with 1 number of leave(s) and size of a tree for only male(s) students consisting of 5 incorrectly classified instances for all female students due to false positives instances classified in a confusion matrix. False positives can be defined as a class of instances with number of instances predicted positive that are actually negative.
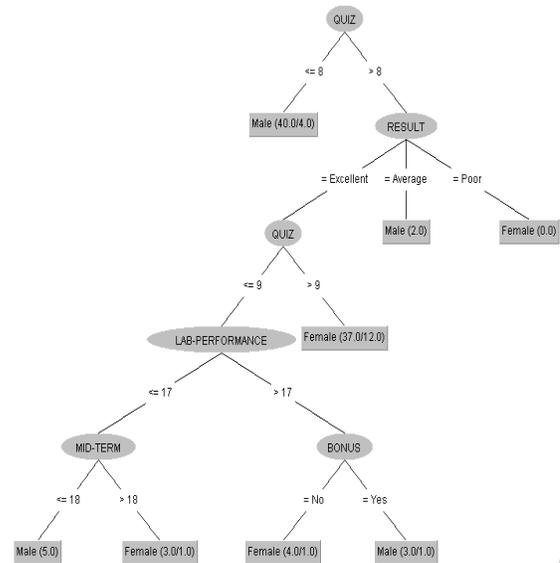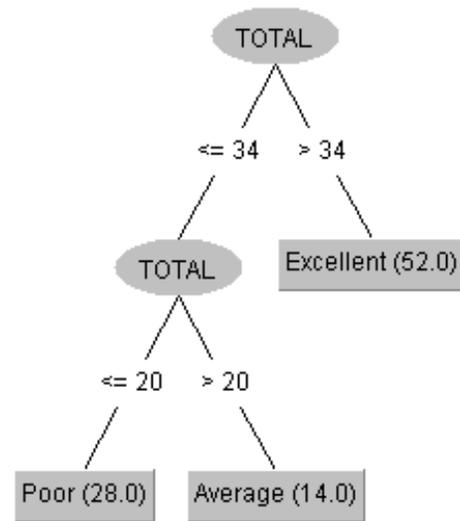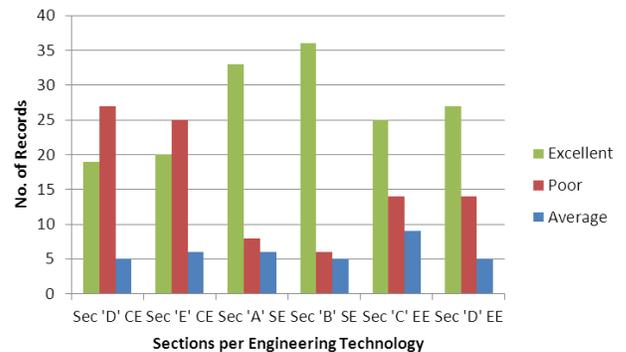

Fig. 9.   Decision tree produced for OOP.

Fig. 10. Comparison of section wise per engineering technology.

Fig. 10 shows the results of comparative study of two sections per engineering technology with respect to excellent, poor or average progress. In a computer engineering technology, majorly students' progress displays poor in both sections in above figure but section 'D' shows further poor progress than section 'E'. In a software engineering technology, major students' progress contribution is excellent but section 'B' shows more excellent progress than the other section as shown in a figure. In an electronic engineering technology, most of the students are shown excellent progress but section 'D' is comparatively improved than section 'C'.

VI. Conclusion and Future Directions

The present study aims the significance, scope and techniques of data mining in the domain of education is addressed in a multiple education disciplines and technologies at higher education level interestedly. One of the useful and widely used data mining methodologies is classification. There are many classification algorithms but three mostly used classification algorithms have been used in this study. In this study, only a single (core) course sessional or pre-examination marks of three different cohorts belonging to computer, software and electronic engineering have been analyzed and followed to determine students pedagogical progress in their corresponding engineering fields and their learning attitude towards particular course(s) for the preparation of final examination based on their pre-examination marks. Further, gender wise and section wise progress has also been studied for each stated engineering technologies. The highly influencing sessional variables have been identified as a criterion of awarding and judging students' academic progress in a course(s) and applying data mining classification techniques to implement high potential data mining applications at higher education level, referring to the optimal manipulation of data mining approaches and techniques to deeply analyze and track the engineering student's pedagogical progress throughout the academic session. The study basically employs the classification techniques with three different classifiers: J48 Decision Trees, Naïve Bayes and K-NN to classify attributes affecting students' progress in their core course(s) for the betterment of academic stakeholders' assistance and regulation in order to improve academic progress which is the main objective of study. Essentially, as mentioned earlier, these three algorithms have been selected for this study because they perform better than any other technique and are easily interpretable. Among the three algorithms, J48 and K-NN gives better results in terms of accuracy.

In future, some more courses with sessional or pre-examination marks as well as final examination marks of more engineering technologies for students' academic evaluation can be deemed in order to refine and embody continuously smooth pedagogy and learning process. Thus, the study helps and guide students to improve their academic performance and reduce failure in a course(s) by taking appropriate actions and to increase retention for the semester examination.

References

[1] Fernandez DB, Lujan-Mora S, "Comparison of applications for educational data mining in Engineering Education," 2017 IEEE World Engineering Education Conference (EDUNINE). 2017.

[2] www.tutorialspoint.com. (2018). Data Mining Tutorial. [online] Available at: https://www.tutorialspoint.com/data_mining/. [Accessed 21 Feb. 2018].

[3] Essay UK Free Essay Database. (2018). Essay: Educational Data Mining (EDM) - Essay UK Free Essay Database. [online] Available at: http://www.essay.uk.com/essays/information-technology/essay-educational-data-mining-edm/. [Accessed 21 Feb. 2018].

[4] Asif R, Merceron A, Pathan MK, "Predicting Student Academic Performance at Degree Level: A Case Study. International Journal of Intelligent Systems and Applications," 2014Aug;7(1): pp. 49–61.

[5] Yassein NA, Helali RGM, Mohomad SB. "Predicting Student Academic Performance in KSA using Data Mining Techniques," Journal of Information Technology & Software Engineering. 2017;07(05).

[6] Kumar B, Pal S, "Mining Educational Data to Analyze Students Performance," International Journal of Advanced Computer Science and Applications. 2011;2(6).

[7] En.wikipedia.org. (2018). Statistical classification. [online] Available at: https://en.wikipedia.org/wiki/Statistical_classification#Algorithms/. [Accessed 23 Feb. 2018].

[8] Archana T, Gandhi UD, "Prediction of Student Performance in Educational Data Mining - A Survey. International," Journal of Pharmacy & Technology. 2016Sep;8(3): pp. 17757–63.

[9] Adhatrao K, Gaykar A, Dhawan A, Jha R, Honrao V, "Predicting Students Performance Using ID3 and C4.5 Classification Algorithms," International Journal of Data Mining & Knowledge Management Process. 2013;3(5):39–52.

[10] Dutt A, Ismail MA, Herawan T, "A Systematic Review on Educational Data Mining," IEEE Access. 2017Jan17;5: pp. 15991–6005.

[11] Asif R, Hina S, Haque SI, "Predicting Student Academic Performance using Data Mining Methods," International Journal of Computer Science and Network Security. 2017May;17(5):187–91.

[12] Asif R, Merceron A, Pathan MK, "Investigating Performances' Progress of Students," pp.116–23.

[13] Asif R, Merceron A, Ali SA, Haider NG, "Analyzing undergraduate students performance using educational data mining," Computers & Education. 2017;113: pp. 177–94.

[14] Yadav SK, Bharadwaj B, Pal S, "Data Mining Applications: A comparative Study for Predicting Student's performance," INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY & CREATIVE ENGINEERING. 2012Feb;1(12): pp. 13–9.

[15] Asif R, Merceron A, Pathan MK, "Investigating performance of students," Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK 15. 2015.

[16] Agarwal S, "Data Mining: Data Mining Concepts and Techniques," 2013 International Conference on Machine Intelligence and Research Advancement. 2013.

[17] En.wikipedia.org. (2018). Naive Bayes classifier. [online] Available at: https://en.wikipedia.org/wiki/Naive_Bayes_classifier [Accessed 13 March. 2018].

[18] En.wikipedia.org. (2018). K-nearest neighbors algorithm. [online] Available at: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm [Accessed 13 Mar. 2018].

[19] En.wikipedia.org. (2018). Data pre-processing. [online] Available at: https://en.wikipedia.org/wiki/Data_pre-processing [Accessed 05 Apr. 2018].

[20] Cs.waikato.ac.nz. (2018). Weka 3 - Data Mining with Open Source Machine Learning Software in Java. [online] Available at: http://www.cs.waikato.ac.nz/ml/weka/ [Accessed 07 Apr. 2018].