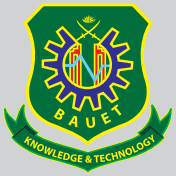
BAUET JOURNAL
Published by
Bangladesh Army University of Engineering & Technology
Journal Homepage: http://journal.bauet.ac.bd



Analysis of University Students' Performance Using WEKA to Enhance the Education Quality

Md. Tofail Ahmed[1*], Md. Jalal Uddin[2], Md. Humaun Kabir[3]

[1]Department of Information and Communication Engineering, Faculty of Engineering and Technology, Pabna University of Science and Technology, Pabna-6600, Bangladesh
[2]Department of Information and Communication Engineering, Faculty of Engineering, University of Rajshahi, Rajshahi-6205, Bangladesh
[3]Department of Computer Science and Engineering, Faculty of Engineering, Bangamata Sheikh Fojilatunnesa Mujib Science & Technology University, Jamalpur-2012, Bangladesh

Abstract: Universities of a country perform a significant role in our society to develop every sector as well as build the backbone of a nation by providing higher, specialized and contemporary education to the young generation. At present, around 163 universities, and 31 specialized colleges exist in Bangladesh under the supervision of University Grants Commission of Bangladesh (UGC). Recently, Bangladesh government and UGC have taken the biggest challenges to ensure the quality of education to create productive graduates. But there are some common problems of fresher students in most universities such as they are not enough familiar with university education system, accommodation crises, political unrest, etc. In recent years, the rate of failure and thus re-admission is reaching alarmingly high. In this research, it has analyzed and classified the published 1st-year result data of the last 4 years (2017-2020) of the Department of Information and Communication Engineering at Pabna University of Science and Technology, Bangladesh. The analytical data critically to extract valuable information and tried to find out the reasons behind failure in the semester final examination. The classification algorithms J48 and JRIP have been used in WEKA to predict the result of a student. If someone is found not going to do well in the upcoming exam, the student can be traced and kept in special care and encouraged to reduce the rate of fail/ readmission as well as save valuable resources and accelerate the production of a more well-educated young generation.

Keywords: *Data Mining; WEKA; J48; KDD; GNU; IREP and JRIP.*

Introduction: Data mining is one of the popular core processes for analyzing, classifying and extracting valuable and desire hidden knowledge from huge amount of data [1]. The main goal of data mining is to discover knowledge from huge amount of data and transforms it to an understandable structure for further use [2]. In real life activities, huge amount of data are stored in the databases. It is difficult job for an organization or researchers to discover knowledge from those huge databases [3]. Currently the huge amount 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 [4]. The purpose of classification is to correctly predict the class label for each case in the data. The dataset in which the class assignment is known is needed for classification working process. A classifier is generated by classification algorithm [5]. This classifier classifies those instances whose class label is not known. Data mining is also known as Knowledge Discovery from Data, or KDD which refers to extracting or mining the knowledge from large amount of data. The database of data mining may be a logical rather than a physical subset of data warehouse which can supply the additional resource base on demands. Data can be stored anywhere in various databases such as World Wide Web, external sources. If it is not the case, then you will use with a separate data mining database. Now a day's in real world the huge amount of data are available in the variety field of education sector, medical sector, industry and many other areas. Using such amount of data, it can extract some sort of information and knowledge which can be helps us in important decision making to solve a particular problem. For example, you can find out Alumni student in any university database, with the help of shopping database, you can identify how many items sold out. Data can be examined/ analyzed, summarized, understand and meet to challenges [6]. With the help of data mining it can not only analyzes the data but also process to discovery interesting pattern from the huge amount of data. The generated interesting hidden patterns by applying some of the data mining techniques are easy to understand, identified, applicable, and possibly useful. The data mining objectives refer to efficiently retrieve of data, information, knowledge discovery recognize unseen patterns and those patterns which are previously not considered/ explored, to reduce the level of complexity, time saving, etc from the databases [7].

Software and Classifier Algorithm Requirements: Waikato Environment for Knowledge Analysis (WEKA) is a popular data mining system that uses for implementing data mining algorithms. It can be run on Windows, Linux suite of machine learning software written in Java, developed at the University of Waikato, New Zealand [8]. It is free software licensed under the General Public License (GNU) and used for research, and Mac [9]. There are different types of classifiers in WEKA [10-11]. In this work decision tree classifier that is J48 and the rules classifier that is JRIP have been used.

Article history:
Received 14 May, 2022
Received in revised form 10 June,2022
Accepted 14 October, 2022
Available online 02 November, 2022

Corresponding author details: Md. Tofail Ahmed
E-mail address: 10205006@uap-bd.edu
Tel: +1 765 285 8650

J48: J48 is the most popular and powerful decision tree classifier. The WEKA (Quinlan 1986) classifier package has its version of C4.5 known as J48. It is a most popular and simple classifier technique which is used to make a decision tree [12]. The decision trees generated by J48 can be used for the classification of provided data to take a decision. J48 builds decision trees from a set of labeled training data using the concept of information entropy. It uses the fact that each attribute of the data can be used to decide by splitting the data into smaller subsets. J48 examines the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data [13]. To make the decision, the attribute with the highest normalized information gain is used. Then the algorithm recurs on the smaller subsets. The splitting procedure stops if all instances in a subset belong to the same class. Then a leaf node is created in the decision trees telling to choose that class. But it can also happen that none of the features given any information gain.

JRIP: Repeated incremental pruning to produce error reduction (ripper) is a build an optimized algorithm. JRIP uses this algorithm it is the optimized version of Incremental Reduced Error Pruning (IREP) [14]. Repeated Incremental Pruning to Produce Error Reduction (RIPPER) is a rule-based learner that builds a set of rules that identify the classes while minimizing the amount of error. The error is defined by the number of training examples misclassified by the rules. RIPPER is a straightforward approach for generating classification rules. It is considered to be more efficient than decision trees on large and noisy datasets. It undergoes four phases such as Growth, Pruning, Optimization, and Selection [15]. In the growth phase, it produces a sequence of individual rules by adding predicates until the rule satisfies stopping criteria. The rules that reduce the performance of the algorithm are pruned in the second phase. In the optimization step, each rule is optimized by adding up attributes to the original rule or generating a new rule using phase 1 and phase 2. In the last stage, the best rules are retained and others are ignored from the model. It employs the description length function to calculate the description length of the rule.

Data Analysis Methods: Data analysis is an ongoing activity, which not only answers the question but also gives some directions for future data collection. Data analysis procedures (DAP) help arrive at the data analysis. The procedures put this research project in perspective and assist in testing the hypotheses.

Data Analysis: Data pre-processing is required for preparing the test data for our proposed method. Data cleaning removes noise and inconsistent data from a raw data source. Then the data relevant to the analysis task are retrieved from the dataset. In this research, it has been selected course marks, course GPA, Gender, Student Type, Attendance Status, Study, Understanding of Class, and Results for every student for 4 years of the First Semester and Second semester. The fig. 1 shows the data analysis tree of this research work.

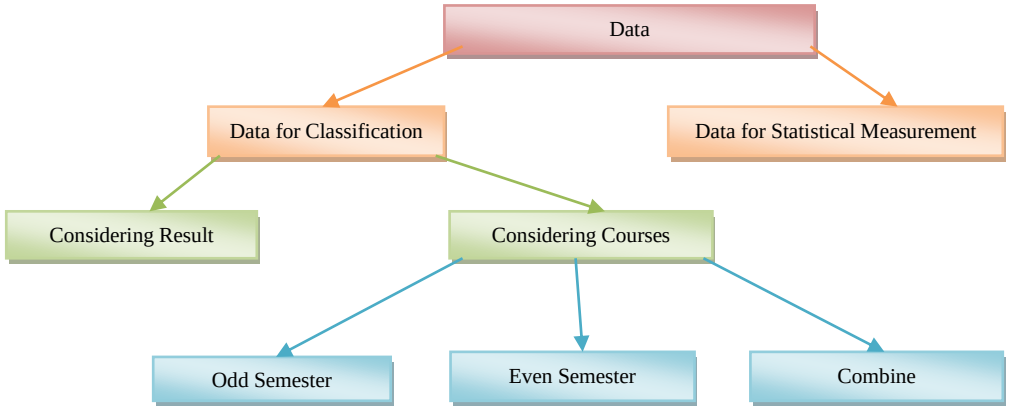


Fig. 1: Data Analysis Tree.

Data for Statistical Analysis: Statistical analysis and data Mining will be a useful resource to those solving practical problems, at the same time enabling them to benefit from ideas developed in other domains.

Major, Minor and Lab Courses: In the Table 1, it shows major theory courses, minor theory courses, and lab courses that are taken for this research work.

Table 1. Major, Minor and Lab Courses List.

Sl	Major Theory Courses	Minor Theory Courses	Lab Courses
1	ICE-1111	PHY-1111	PHY-1122
2	ICE-1121	MATH-1111	ICE-1212
3	ICE-1211	CHEM-1111	APEE-1292
4	-	APEE-1291	-
5	-	MATH-1211	-
6	-	STAT-1221	-
7	-	PHY-1221	-

Student Type, Gender and Attendance Status: There are two types of students who start their semester with two types of gender. At the end of the semester, the entire students are given their attendance record. There are three types of attendance status. The 60% attendance is considered as Dis-collegiate, 61% to 74% attendance is considered as non-collegiate and 75% to 100% attendance is considered as regular. Table 2 shows the summary of the data.

Table 2. Student Type, Gender and Attendance Status.

Sl	Student Type	Gender	Attendance Status
1	Regular	Male	Regular
2	Readmission	Female	Non-collegiate
3	-	-	Dis-Collegiate

Data Classification Method: The data classification method runs a classification algorithm with three types of data. Those are considering courses, student activity, and results. Here describe the data in two parts.

Classification Considering Courses: For generating a decision tree and decision rules using WEKA for the prediction of student's performance it needs to convert the numeric attribute of the result to a nominal attribute considering some definition. The definition of the result class has shown in Table 3.

Table 3. Result Class Definitions.

Sl	GPA Range	Class
1	0.00 to 1.99	Fail
2	2.00 to 3.24	Average
3	3.25 to 3.49	Good
4	3.50 to 3.74	Very Good
5	3.75 to 4.00	Excellent

For generating decision trees and decision rules using WEKA for the prediction of student's performance give category to the numerical data of major theory courses, minor theory courses, and lab with the definition. Table 4 shows the definition of the attribute.

Table 4. Category of Attribute Type 1.

Sl	Marks	Major Theory	Minor Theory	Lab
1	0 to 39	MJF-Fail	MNF-Fail	FL-Fail
2	40 to 64	MJF-Average	MNF-Average	FL-Average
3	65 to 69	MJF-Good	MNF-Good	FL-Good
4	70 to 74	MJF-VeryGood	MNF-VeryGood	FL-VeryGood
5	75 to 100	MJF-Excellent	MNF-Excellent	FL-Excellent

Classification Considering Student Activity: For generating decision rules using WEKA for the prediction of students' performance need to convert the numeric attribute of student activity to a nominal attribute considering some definition. Understanding class is a measure based on the class test marks of the student. And study data are collected from those students who fail the 1st year exam. The definition of attribute according to the understanding of class by getting marks 0 to 20 as shown in Table 5.

Table 5. Category of Attribute Type 2.

Sl	Class Test Marks	Understanding Class
1	0-7.99	UCBad
2	8-12.99	UCAvergae
3	13-13.99	UCGood
4	14-14.99	UCVeryGood
5	15-20	UCExcellent

For generating decision rules using WEKA for the prediction of students' performance need to convert the numeric attribute of student activity to a nominal attribute considering some definition. Study data are collected from those students who fail the 1st year exam. Table 6 shows the category of students according to the study of a student per day (hour).

Table 6. Category of Attribute Type 3.

Sl	Study Per Day(hour)	Study
1	0-1	SBad
2	1-2	SAverage
3	2-3	SGood
4	3 or more	SVeryGood

Classification Considering Result: Classification Considering Result runs WEKA for generating classification rules and a tree with the data of odd semester results, even semester results, and final results. To be promoted to a higher class a student must obtain the requirements of YGPA of 2.25 or higher and also a credit point of 30 or higher. The defined result with the following categories shows in Table 7 and Table 8.

Table 7. Result Class Definition 1.

Sl	YGPA Range	Credit point	Result
1	Greater than or equal to 2.25	Greater than or equal to 30	Pass
2	Less than or equal to 2.25	Less than or equal to 30	Fail

If the student will fulfill the requirement of pass, than he /she get promotion to the next year. If the student will not fulfill the requirement of pass that he/she is fail, than he/she will not get promotion to the next year and also get readmission to the same year.

Table 8. Result Class Definition 2.

Sl	Result	Final Result
1	Pass	Promotion
2	Fail	Readmission

Classification Method Procedure: The J48 algorithm has been used to generate a decision tree. The root of the tree is decided by calculating the information gain of every attribute [16]. The attribute with the highest information gain is selected as the root. The child of the root is the attribute with the second highest information gain and so on. The leaf nodes represent the class level with the lowest information gain. Eq. 1 is the information gain formula to calculate Entropy for training dataset S with C classes.

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i \quad \text{.....eq. 1}$$

where P_i is the probability of randomly picking an element of class *i* (i.e. the proportion of the dataset made up of class *i*). Then the JRIP Algorithm has used to generate decision rules. JRIP Proceeds by treating all the examples of a particular judgment in our training data as a class and finding a set of rules that cover all the members of that class. Therefore, it proceeds to the next class and does the same, repeating this until all classes have been covered.

Analysis of Pruned Tree and Decision Rules: After applying J48 algorithm implemented in WEKA to get pruned tree. The pruned tree contains a root node, intermediate node, and leaf node [17]. In pruned tree the attribute with high information gain is in the root node then the root node is split into sub-node. This type of node is called an intermediate node. When any node finds the pure class then it stops splitting. In the leaf node, there exists a class level. In this research work, a pruned tree has been

generated from the dataset. The pruned tree root node contains the “Final Minor Theory Courses", the intermediate node contains other attributes like “Final Lab Courses" etc, and in the leaf node there is a class level like “Excellent", “Very good" etc. JRIP algorithm in WEKA has been also run for the same dataset to produces classification rules. From the analysis of the dataset, JRIP generates some rules with 66% training and the rest are testing data sets.

Experimental Result and Analysis: The experimental setup and brief discussion of experimental output have shown in the first part of this section. At the end of the section, the analysis of the result has been shown.

Experimental Setup: The experiment has been conducted on a 2.1 GHz Intel(R) Celeron processor with 4GB RAM, running on Microsoft Windows 10. WEKA has been used as a tool for research work. The J48 algorithm and JRIP algorithm by WEKA is a java implementations of the C4.5 algorithm, and RIPPER algorithm respectively.

Result of Statistical Measurement: The final examination result of the ICE department has been collected from the authentic exam controller of Pabna University of Science and Technology (PUST). This exam result data has been collected for analysis using J48 and JRIP algorithms with the help of WEKA software. In this research work, Microsoft Excel has been used for statistical measurement of the exam result. All statistical measurement of data is shown in Table 9 to Table 16.

Table 9. Major Theory Courses Failure Rate.

Sl	Courses	Failure Rate
1	ICE-1111	4.67%
2	ICE-1131	10.74%
3	ICE-1211	9.38%

Table 10. Minor Theory Courses Failure Rate.

Sl	Courses	Failure Rate
1	PHY-1111	07.79%
2	MATH-1111	06.88%
3	CHEM-1111	12.84%
4	APEE-1291	05.72%
5	MATH-1211	22.91%
6	STAT-1221	03.64%
7	PHY-1221	04.69%

Table 11. Lab Courses Failure Rate.

Sl	Courses	Failure Rate
1	ICE-1122	01.83%
2	ICE-1132	00.45%
3	ICE-1212	02.08%
4	ICE-1222	01.56%

Table 12. Failure Rate According to Gender.

Sl	Gender	Failure Rate
1	Male	17.36%
2	Female	21.43%

Table 13. Failure Rate According to Attendance Status

Sl	Attendance Status	Failure Rate
1	Dis-collegiate	64.29%
2	Non collegiate	20.40%
3	Regular	12.90%

Table 14. Failure Rate According to Student Type for All Years.

Sl	Student Type	Failure Rate
1	Regular	15.87%
2	Readmission	31.03%

Table 15. Absent Percentage in Even Semester Exam.

Sl	Year	2017	2018	2019	2020
1	Absent Percentage in Odd Semester	03.89%	06.25%	06.79%	11.23%
2	Absent Percentage in Even Semester	04.16%	05.38%	05.17%	10.71%

Table 16. Number of Absent and Attended Student in Even Semester Who Fail in Odd Semester.

Sl	Condition	2017	2018	2019	2020
1	Absent in Exam	40.22%	42.85%	50.20%	49.20%
2	Participated in Exam	60.30%	57.14%	47.10%	50.12%

Result of Classification Method: In this work, two classification algorithms have been used which are J48 and JRIP. The output of six datasets run by J48 and JRIP by WEKA has shown in this section. For all datasets, the select test mode is Test mode: split 66.0% train, remainder test. Table 17 shows the dataset name and number of students that are selected for classification algorithms.

Table 17. Number of Instances for Dataset.

Sl	Dataset Name	Instances
1	Odd Semester	218.00
2	Even Semester	192.00
3	Combined Semester	192.00
4	Major Courses	037.00
5	Minor Courses	073.00
6	Study and Understanding Class	039.00

Output of J48 for Odd Semester: The output of J48 for the Odd Semester result shows in Fig. 2. The Number of Leaves 21 and the size of the tree 26 have been used to produce the result.

```
J48 pruned tree
------------------

FinalMajTheoC = MJFExcellent
|   FinalMinTheoC = MNFExcellent: Excellent (9.0/1.0)
|   FinalMinTheoC = MNFVeryGood: VeryGood (1.0)
|   FinalMinTheoC = MNFGood: VeryGood (2.0)
|   FinalMinTheoC = MNFAverage: Excellent (0.0)
|   FinalMinTheoC = MNFFail: Average (1.0)
FinalMajTheoC = MJFVeryGood: VeryGood (27.0/9.0)
FinalMajTheoC = MJFGood
|   FinalMinTheoC = MNFExcellent: VeryGood (2.0)
|   FinalMinTheoC = MNFVeryGood: VeryGood (14.0/4.0)
|   FinalMinTheoC = MNFGood: Good (13.0/2.0)
|   FinalMinTheoC = MNFAverage: Good (12.0/5.0)
|   FinalMinTheoC = MNFFail: Fail (1.0)
FinalMajTheoC = MJFAverage
|   FinalMinTheoC = MNFExcellent: Average (0.0)
|   FinalMinTheoC = MNFVeryGood: Good (4.0/1.0)
|   FinalMinTheoC = MNFGood: Good (20.0/4.0)
|   FinalMinTheoC = MNFAverage: Average (37.0/3.0)
|   FinalMinTheoC = MNFFail: Average (30.0/4.0)
FinalMajTheoC = MJFFail
|   FinalMinTheoC = MNFExcellent: Fail (0.0)
|   FinalMinTheoC = MNFVeryGood: Fail (0.0)
|   FinalMinTheoC = MNFGood: Good (2.0/1.0)
|   FinalMinTheoC = MNFAverage: Average (13.0/5.0)
|   FinalMinTheoC = MNFFail: Fail (30.0/1.0)
```

Fig. 2: J48 Pruned Tree for Odd Semester Result Analysis.

Analysis: The result evaluation of J48 Pruned tree for Odd Semester result shows that if the number of students makes Excellent results in the final major theory courses than the failure rate is comparatively low. If the number of students makes a Good result in final major theory courses than the failure rate is also low. If the number of students makes Average result in final major theory courses and fail in the minor theory courses than the failure rate is comparatively high. Finally, if the numbers of students fail in final major theory courses and also fail in the minor theory courses than the failure rate is also comparatively high. The class Excellent, VeryGood, Good, etc have been discussed in Table 3.

Output of JRIP for Odd Semester: Fig. 3 shows the Output of JRIP for the Odd Semester result. The Number of rules to generate the output of JRIP algorithm is 09.

```
JRIP rules:
===========

(FinalMinTheoC = MNFExcellent) => Class=Excellent (17.0/6.0)
(FinalMinTheoC = MNFVeryGood) => Class=VeryGood (27.0/9.0)
(FinalMinTheoC = MNFGood) and (FinalMajTheoC = MJFVeryGood) => Class=VeryGood (11.0/3.0)
(FinalMajTheoC = MJFExcellent) => Class=VeryGood (3.0/1.0)
(FinalMajTheoC = MJFFail) and (FinalMinTheoC = MNFFail) => Class=Fail (30.0/1.0)
(FinalMinTheoC = MNFGood) => Class=Good (35.0/7.0)
(FinalMajTheoC = MJFGood) => Class=Good (13.0/6.0)
(FinalMajTheoC = MJFVeryGood) => Class=Good (2.0/0.0)
 => Class=Average (80.0/12.0)

Number of Rules : 9
```

Fig. 3: JRIP Rules for Odd Semester Result Analysis.

Analysis: Fig. 3 shows that the number of student makes Excellent results in the final minor theory courses among them 17 student makes Excellent results in the odd semester final result. The numbers of students make VeryGood results in the final minor theory courses among them 27 students make VeryGood results in the odd semester final result. A number of the student makes Good results in the final minor theory courses and the student makes VeryGood results in the final major theory courses among them 11 students make VeryGood results in the Odd Semester final result. A number of the student makes Excellent results in the final major theory courses among them 3 students make VeryGood results in the Odd Semester final result. Several student Fail in the final minor theory courses and students also Fail in the final major theory courses among them 30 student Fail in the odd semester final result. A number of the student makes Good results in the final minor theory courses among them 35 students make Good results in the odd semester final result. A number of the student makes Good results in the final major theory courses among them 13 students make Good results in the Odd Semester final result. The number of student makes VeryGood result in the final major theory courses among them 2 student makes Good result in the Odd Semester final result. It has been discussed the class Excellent, VeryGood, Good, etc in Table 3.

Output of J48 for Even Semester: The output of J48 for the Even Semester result shows in Fig. 4. The Number of Leaves and the size of the tree have been used to produce the result.

```
J48 pruned tree
------------------

FinalMajTheoC = MJFExcellent
|   FinalMinTheoC = MNFExcellent
|   |   Lab = FLExcellent: Excellent (8.0)
|   |   Lab = FLVeryGood: VeryGood (3.0)
|   |   Lab = FLGood: Excellent (0.0)
|   |   Lab = FLAverage: Excellent (0.0)
|   |   Lab = FLFail: Excellent (0.0)
|   FinalMinTheoC = MNFVeryGood: VeryGood (8.0/1.0)
|   FinalMinTheoC = MNFGood: VeryGood (3.0)
|   FinalMinTheoC = MNFAverage: VeryGood (0.0)
|   FinalMinTheoC = MNFFail: Good (1.0)
FinalMajTheoC = MJFVeryGood
|   FinalMinTheoC = MNFExcellent: VeryGood (1.0)
|   FinalMinTheoC = MNFVeryGood: VeryGood (19.0)
|   FinalMinTheoC = MNFGood
|   |   Lab = FLExcellent: VeryGood (4.0/1.0)
|   |   Lab = FLVeryGood: VeryGood (4.0/1.0)
|   |   Lab = FLGood: VeryGood (0.0)
|   |   Lab = FLAverage: Good (3.0)
|   |   Lab = FLFail: VeryGood (0.0)
|   FinalMinTheoC = MNFAverage: Good (1.0)
|   FinalMinTheoC = MNFFail: Good (3.0/1.0)
FinalMajTheoC = MJFGood: Good (37.0/14.0)
FinalMajTheoC = MJFAverage: Average (77.0/18.0)
FinalMajTheoC = MJFFail
|   FinalMinTheoC = MNFExcellent: Average (0.0)
|   FinalMinTheoC = MNFVeryGood: Average (0.0)
|   FinalMinTheoC = MNFGood: Average (1.0)
|   FinalMinTheoC = MNFAverage: Average (10.0)
|   FinalMinTheoC = MNFFail: Fail (9.0)
```

Fig. 4: J48 Pruned Tree for Even Semester Result Analysis.

Analysis: Fig. 4 shows that if the number of the students makes Excellent results in the final major theory courses and the students fail in the minor theory courses than the failure rate is comparatively low. If the number of students make a VeryGood result in final major theory courses and students Fail in the minor theory courses than the failure rate is also low. Finally, if the number of students fails in final major theory courses and also fail in the minor theory courses than the failure rate is also comparatively high.

Output of JRIP for Even Semester: The output of JRIP for the Even Semester shows in figure 5. Here the number of rules 07 is used to produce the output.

```
JRIP rules:
===========

(FinalMinTheoC = MNFExcellent) and (Lab = FLExcellent) => Class=Excellent (10.0/1.0)
(FinalMinTheoC = MNFFail) and (FinalMajTheoC = MJFFail) => Class=Fail (9.0/0.0)
(FinalMajTheoC = MJFGood) => Class=Good (36.0/13.0)
(FinalMajTheoC = MJFVeryGood) and (Lab = FLAverage) => Class=Good (6.0/1.0)
(FinalMajTheoC = MJFVeryGood) => Class=VeryGood (28.0/4.0)
(FinalMajTheoC = MJFExcellent) => Class=VeryGood (15.0/2.0)
 => Class=Average (88.0/18.0)

Number of Rules : 7
```

Fig. 5: JRIP Rules for Even Semester Result Analysis.

Analysis: Fig. 5 shows that the number of student makes Excellent results in the final minor theory courses and student makes also Excellent result in the Lab, 10 students makes Excellent result in the Even Semester final. A number of the student Fail in the final minor theory courses and the student also Fail in the final major theory courses, 9 students Fail in the Even Semester final. A number of the student makes Good results in the final major theory courses, and 36 students make Good results in the Even Semester final. A number of the student makes VeryGood results in the final major theory courses and the student makes an average result in the Lab, 6 students make Excellent results in the Even Semester final result. The number of students makes VeryGood result in the final major theory courses, 28 students make VeryGood result in the Even Semester final. A number of the student makes Excellent results in the final major theory courses, and 15 students make VeryGood results in the Even Semester final.

Out of J48 for Combined Semester: Output of J48 for combined semester result shows in Fig. 6.

```
J48 pruned tree
------------------

FinalMinTheoC = MNFExcellent
|   Lab = FLExcellent: Excellent (9.0/1.0)
|   Lab = FLVeryGood: VeryGood (2.0)
|   Lab = FLGood: Excellent (0.0)
|   Lab = FLAverage: Good (1.0)
|   Lab = FLFail: Excellent (0.0)
FinalMinTheoC = MNFVeryGood: VeryGood (26.0/3.0)
FinalMinTheoC = MNFGood
|   FinalMajTheoC = MJFExcellent: Good (0.0)
|   FinalMajTheoC = MJFVeryGood: VeryGood (11.0/2.0)
|   FinalMajTheoC = MJFGood: Good (14.0/4.0)
|   FinalMajTheoC = MJFAverage: Good (15.0/4.0)
|   FinalMajTheoC = MJFFail: Good (1.0)
FinalMinTheoC = MNFAverage
|   FinalMajTheoC = MJFExcellent: Average (0.0)
|   FinalMajTheoC = MJFVeryGood: Good (3.0)
|   FinalMajTheoC = MJFGood: Good (7.0/2.0)
|   FinalMajTheoC = MJFAverage: Average (23.0/2.0)
|   FinalMajTheoC = MJFFail: Average (13.0)
FinalMinTheoC = MNFFail
|   FinalMajTheoC = MJFExcellent: Average (0.0)
|   FinalMajTheoC = MJFVeryGood: Good (2.0)
|   FinalMajTheoC = MJFGood: Average (5.0/1.0)
|   FinalMajTheoC = MJFAverage: Average (49.0/8.0)
|   FinalMajTheoC = MJFFail: Fail (11.0)
```

Fig. 6: J48 Pruned Tree for Combined Semester Result Analysis.

Analysis: The result evaluation of J48 Pruned tree for the Combined Semester analysis result shows that if the number of students makes Good results in the final minor theory courses and the students Fail in the major theory courses than the failure rate is comparatively low. If the number of students makes Average result in final minor theory courses and students Fail in the major theory courses than the failure rate is comparatively high. Finally, if the number of students fails in final minor theory courses and also fail in the major theory courses than the failure rate is also high.

Output of JRIP for Combined Semester: The output of JRIP for the Combined Semester result shows in Fig.7. Here the number of rules 09 has been used to generate the output of WEKA.

```
JRIP rules:
===========

(FinalMinTheoC = MNFExcellent) => Class=Excellent (12.0/4.0)
(AttendanceStatus = Noncollegiate) and (FinalMajTheoC = MJFFail) and (FinalMinTheoC = MNFFail) => Class=Fail (11.0/0.0)
(FinalMinTheoC = MNFVeryGood) => Class=VeryGood (26.0/3.0)
(FinalMinTheoC = MNFGood) and (FinalMajTheoC = MJFVeryGood) => Class=VeryGood (11.0/2.0)
(FinalMinTheoC = MNFGood) and (Lab = FLVeryGood) => Class=Good (16.0/2.0)
(FinalMajTheoC = MJFGood) and (Lab = FLVeryGood) => Class=Good (6.0/1.0)
(FinalMinTheoC = MNFGood) and (Lab = FLGood) => Class=Good (10.0/3.0)
(FinalMajTheoC = MJFVeryGood) => Class=Good (5.0/0.0)
 => Class=Average (95.0/14.0)

Number of Rules : 9
```

Fig. 7: JRIP Rules for Combined Semester Result Analysis.

Analysis: In Fig. 7, it is shown that the number of student makes Excellent results in the final minor theory courses, 10 students make Excellent results in the 1st year final exam. The number of students whose Attendance Status is Non-collegiate, students Fail in the final major theory courses, and students also Fail in the final minor theory courses, 11 students Fail in the 1st year final exam. The number of students makes VeryGood result in the final minor theory courses, 26 students make VeryGood result in the 1st year final exam. A number of the student makes Good results in the final minor theory courses, and the student makes VeryGood results in the final major theory courses 11 students make VeryGood result in the 1st year final exam. A number of the student makes Good results in the final minor theory courses, the student makes VeryGood results in the lab, and 16 students make Good results in the 1st year final exam. A number of the student makes Good results in the final major theory courses, and the student makes VeryGood results in the lab, 6 students make Good results in the 1st year final exam. A number of the student makes Good results in the final minor theory courses, the student makes Good results in the lab, and 5 students make Good results in the 1st year final exam. The number of students makes an Average result in the final major theory courses, 5 students make Good results in the 1st year final exam.

Output of J48 for Major Courses: Fig. 8 shows the output of J48 for Major Courses which is used to analyze the performance of students. Here the number of leaves 6 and the size of the tree 11 have been used to generate the figure.

```
J48 pruned tree
------------------

MajICE1111 = Pass
|   MajICE1211 = Pass: Pass (12.0)
|   MajICE1211 = Fail
|   |   MajICE1131 = Pass: Pass (7.0)
|   |   MajICE1131 = Fail: Fail (8.0)
MajICE1111 = Fail
|   MajICE1131 = Pass
|   |   MajICE1211 = Pass: Pass (2.0)
|   |   MajICE1211 = Fail: Fail (3.0)
|   MajICE1131 = Fail: Fail (5.0)
```

Fig. 8: J48 Pruned Tree for Major Courses Analysis

Analysis: Fig.8 shows that if the number of students passes in ICE-1111, fail ICE-1211, and fail ICE-1131 then the number of students who fail the major courses is 8. If the number of students pass ICE-1111, fail in ICE-1211, and pass ICE-1131 then the number of students who fail in the major courses is 7. If the numbers of students fail in ICE-1111, pass ICE-1131, and fail in ICE-1211 then the number of students who fail in the major courses is 3. If the numbers of students fail in ICE-1111, pass ICE-1131, and pass in ICE-1211 then the number of students who fails in the major courses is 2.

Output of JRIP for Major Courses: The output of JRIP for Major Courses shows in Fig. 9. Here, it has been used the number of rules is 03.

```
JRIP rules:
===========

(MajICE1111 = Fail) => Class=Fail (10.0/2.0)
(MajICE1211 = Fail) and (MajICE1131 = Fail) => Class=Fail (8.0/0.0)
 => Class=Pass (19.0/0.0)

Number of Rules : 3
```

Fig. 9: JRIP Rules for Major Courses Analysis.

Analysis: It is shown in Fig.9 that the number of students who fail in the ICE-1111 then 10 students fails in the major theory courses. Several students fail in the ICE-1211 and fail in ICE-1131 then 8 students fail in the major theory courses.

6.3.9. Output of J48 for Major Courses: Fig. 10 represents the Output of J48 for Minor Courses analysis. Here, the number of leaves is 8, and the size of the tree is 15.

```
J48 pruned tree
------------------

MinSTAT1211 = Pass
|   MinPHY1111 = Pass
|   |   MinCHEM1111 = Pass: Pass (36.0/4.0)
|   |   MinCHEM1111 = Fail
|   |   |   MinMATH1211 = Pass
|   |   |   |   MinAPEE1291 = Pass: Pass (10.0)
|   |   |   |   MinAPEE1291 = Fail: Fail (2.0)
|   |   |   MinMATH1211 = Fail: Fail (8.0)
|   MinPHY1111 = Fail
|   |   MinMATH1211 = Pass
|   |   |   MinCHEM1111 = Pass: Pass (3.0)
|   |   |   MinCHEM1111 = Fail: Fail (2.0)
|   |   MinMATH1211 = Fail: Fail (7.0)
MinSTAT1211 = Fail: Fail (5.0)
```

Fig. 10: J48 Pruned Tree for Minor Courses Analysis.

Analysis: Fig. 10 shows that if the number of students passes STAT-1211, pass PHY-1111 and pass CHEM-1111 1211 then the number of students are failed minor courses are 36. If the number of students pass in STAT-1211, pass in PHY-1111, fail in CHEM-1111, pass in MATH-1211, and fail in APEE-1271 then the number of students is failed in minor courses are 2. If the numbers of students pass in STAT-1211, pass in PHY-1111, fail in CHEM-1111, pass in MATH-1211, and pass in APEE-1271 then the number of students are fail in minor courses are 10. If the numbers of students pass in STAT-1211, pass in PHY-1111, fail in CHEM-1111, and fail in MATH-1211 then the number of students are failed in minor courses is 8. If the numbers of students pass in STAT-1211, fail in PHY-1111, pass in MATH-1211, and pass in CHEM-1111 then the number of students are fail in minor courses is 3. If the numbers of students pass in STAT-1211, fail in PHY-1111, pass in MATH-1211, and fail in CHEM-1111 then the number of students are failed in minor courses is 2. If the numbers of students pass in STAT-1211, fail in PHY-1111, and fail in MATH-1211 then the number of students are failed in minor courses is 7.

Output of JRIP for Major Courses: The output of JRIP for Minor Courses shows in Fig. 11. Here, the number of rules is 04.

```
JRIP rules:
===========

(MinCHEM1111 = Fail) and (MinMATH1211 = Fail) => Class=Fail (13.0/0.0)
(MinAPEE1291 = Fail) => Class=Fail (9.0/3.0)
(MinPHY1111 = Fail) => Class=Fail (9.0/3.0)
 => Class=Pass (42.0/3.0)

Number of Rules : 4
```

Fig. 11: JRIP Rules for Minor Courses Analysis.

Analysis: Fig. 11 shows the number of students fails the CHEM-1111 and students also fail MATH-1211 then 13 students fail the minor theory courses. The number of students fail in the APEE-1291 then 9 students fails in the minor theory courses. The number of students fail in the PHY-1111 then 9 students fails in the minor theory courses.

Output of J48 for Study and Understanding Class: The Output of J48 for Study and Understanding Class analyze in Fig. 12. Here, the number of leaves is 13, and the size of the tree is 16.

```
J48 pruned tree
------------------

Study = SExcellent: Bad (0.0)
Study = SVeryGood: Bad (0.0)
Study = SGood: Average (4.0)
Study = SAverage
|   UnderstandingClass = UCExcellent: Bad (0.0)
|   UnderstandingClass = UCVeryGood: Bad (0.0)
|   UnderstandingClass = UCGood: Average (1.0)
|   UnderstandingClass = UCAverage: Average (4.0)
|   UnderstandingClass = UCBad: Bad (6.0)
Study = SBad
|   UnderstandingClass = UCExcellent: Bad (0.0)
|   UnderstandingClass = UCVeryGood: Bad (0.0)
|   UnderstandingClass = UCGood: Average (4.0)
|   UnderstandingClass = UCAverage: Bad (10.0)
|   UnderstandingClass = UCBad: Bad (10.0)
```

Fig. 12: J48 Pruned Tree for Study and Understanding Class Analysis.

Analysis: The result evaluations of the J48 Pruned tree for Study and Understanding Class analysis result show that most of the students fail in the exam study category, average or bad. Also, most of the students fail in the exam whose capability of understanding class category is average or bad. It has been discussed in the attribute Study and Understanding Class in Tables 5 and 6.

Output of JRIP for Study and Understanding Class: The output of JRIP for Study and Understanding Class shows in Fig. 12 to analyze the result. Here, the number of rules is 04.

```
JRIP rules:
===========

(Study = SGood) => Class=Average (4.0/0.0)
(UnderstandingClass = UCGood) => Class=Average (5.0/0.0)
(UnderstandingClass = UCAverage) and (Study = SAverage) => Class=Average (4.0/0.0)
 => Class=Bad (26.0/0.0)

Number of Rules : 4
```

Fig. 13: JRIP Rules for Study and Understanding Class Analysis.

Analysis: Fig.13 of JRIP Rules for Study and Understanding Class Analysis shows that most of the students who fail in the exam whose study category are average or bad. Also, most of the students who fail in the exam whose capability to understanding class category are average or bad. The attribute study and understanding class has been shown in table 5 and 6.

Performance Comparison of J48 and JRIP: It is very difficult to make a comparison between J48 and JRIP. In the measurement accuracy of J48 and JRIP, J48 has better accuracy than JRIP. But if we consider understanding the knowledge from both algorithms, then JRIP produces far more easy rules than J48 because most of the time J48 produces a long tree whereas JRIP generates comparatively short rules which are easy to understand. Table 18 presents the accuracy of both algorithms.

Table 18. Accuracy Comparison between J48 and JRIP.

Sl	Dataset	J48	JRIP
1	Odd Semester	78.3784%	74.3243%
2	Even Semester	66.1538%	61.5385%
3	Combined Semester	80.0000%	80.0000%
4	Major Courses	92.3077%	92.3077%
5	Minor Courses	88.0000%	80.0000%
6	Study and Understanding Class	97.6524%	96.3258%

Conclusions: In this competitive world, the educational setting also uses Data Mining tools to explore and analyze student performance, predict their results to prevent drops out and focus on both good and academically poor performers, feedback for the faculties and instructors, visualization of data and to have a better assessment of learning process. Evaluating students' performance is a complex issue. Educational Data Mining is a process that extracts useful knowledge from a large amount of data in an educational institute. Different types of information extraction methods for discovering different types of knowledge mining

are available. The classification algorithm generates a system that classifies the data accurately based on the training dataset. For classification algorithms, different approaches such as ID3, C4.5, Ripper, J48, JRIP, Naive Bayesian, etc are available in the Data Mining field. In this research, it has been used J48 and JRIP classification algorithms for the measurement of student performance and to make a prediction. Additionally, it has handled some statistical analysis of failure rate corresponding to Gender, Major-minor Courses, Attendance status, etc. It has been discovered some information that will be helpful for the department to make academically related matters for the betterment of students. This research work has tried to find out the reason why students fail and pass exams using classification algorithm implemented in WEKA. It helps predict student performance as well as give a recommendation to the students on which way they should go for making good results. We accomplished some analysis of finding in which courses the students are cutting a sorrow figure in the exam and failure rate corresponding to Odd Semester and Even semester exam. The failure percentage of males and females has been found with the help of analyzing the last four years' data.

References:

[1] Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, ViPin Kumar, Introduction to Data Mining, Wesley, 2005.

[2] Dorina Kabakchieva, Predicting Student Performance by Using Data Mining Methods for Classification, Cybernetics and Information Technologies, Volume 13, No 1, 2013.

[3] Jiawei Han, Micheline kamber, Jian Pei, Data Mining Concepts and Techniques, Morgan Kaufmann Publisher Inc., USA, 2011.

[4] Gaurav L. Agarwal, Prof. Hitesh Gupta, Optimization of C4.5 Decision Tree Algorithm for Data Mining Application, International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 3, March 2013.

[5] 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.

[6] Xingquan Zhu , Ian Davidson, Knowledge Discovery and Data Mining: Challenges and Realities, ISBN 978-1-59904-252, Hershey, New York, 2007.

[7] Joseph, Zernik, Data Mining as a Civic Duty – Online Public Prisoners Registration Systems, International Journal on Social Media: Monitoring, Measurement, Mining, Volume 1, No 1, pp. 84-96, September 2010.

[8] Eshwari Girish Kulkarni, Raj B. Kulkarni, WEKA Powerful Tool in Data Mining, International Journal of Computer Applications (0975 – 8887), National Seminar on Recent Trends in Data Mining (RTDM) 2016.

[9] Aaditya Desai, Dr. Sunil Rai, Analysis of Machine Learning Algorithms using WEKA, Proceedings published in International Journal of Computer Applications® (IJCA), International Conference & Workshop on Recent Trends in Technology, (TCET) 2012.

[10] Cristbal Romero, Sebastin Ventura, Pedro G. Espejo, Csar Hervs, Data Mining Algorithms to Classify Students, International Journal of Computer Science and Information Technologies, Volume 6 (4), 2015.

[11] Trilok Chand Sharma, Manoj Jain, Weka Approach for Comparative Study of Classification Algorithm, International Journal of Advanced Research in Computer and Communication Engineering Volume 2, Issue 4, April 2013.

[12] Anil Rajput, Ramesh Prasad Agarwal, Meghna Dubey, S.P. Saxena, Manmohan Raghuvanshi, J48 and JRIP Rules for E-Governance Data, International Journal of Computer Science and Security (IJCSS), Volume 5, Issue 2, 2011.

[13] Hameetha Begum, S. Begum, Data Mining Tools and Trends – An Overview, International Journal of Emerging Research in Management and Technology, Corpus ID: 212533554, 2013.

[14] Driyani Rajeshinigo, J. Patricia Annie Jebamalar, Educational Mining: A Comparative Study of Classification Algorithms Using WEKA, International Journal of Innovative Research in Computer and Communication Engineering, Volume 5, Issue 3, March 2017.

[15] Gaganjot Kaur, Amit Chhabra, Improved J48 Classification Algorithm for the Prediction of Diabetes, International Journal of Computer Applications (0975 – 8887), Volume 98, No 22, July 2014.

[16] M. Thangaraj, C.R.Vijayalakshmi, Performance Study on Rule based Classification Techniques across Multiple Database Relations, International Journal of Applied Information Systems (IJAIS), Corpus ID: 17551009, New York, USA Volume 5, No 4, March 2013.

[17] R. Aruna devi, Dr. K. Nirmala, Construction of Decision Tree : Attribute Selection Measures, International Journal of Advancements in Research & Technology, ISSN 2278-7763,Volume 2, Issue 4, April 2013.