Journal of Artificial Intelligence and Computational Technology

Journal homepage: https://ojs.omgfzc.com/index.php/JECT

Designing A Model to Analyze the Impact of Applied and Mathematical Subjects on Student Academic Achievement

(Faculty of Computer Science, Omdurman Islamic University and West Kordofan University)

Babiker Gsmallah ELshiekh Mustafa [*1], Khalid Mohammed Osman Saeed[1], Mohamed Mohammed Othman[1], Mohammed Fadual Almola Abbas[2]

[1]Faculty of Computer Science and Information Technology Omdurman Islamic University ,Khartoum, Sudan

[2]Faculty of Computer Science Al-Mughtaribeen University, Khartoum, Sudan

*Correspondence: E-mail: babikergsmallh@oiu.edu.sd

Article Info

Article History:

Submitted/Received 9 October 2025

Revised in revised format 20 October 2025

Accepted 15 November 2025

Available 15 November 2015

Publication Date 15 November 2025

Keyword:

Data Mining ,Classification , Association Rule,, Weka , JRIP algorithm

Cite this article: Mustafa, B. G. E., Saeed, K. M. O., Othman, M. M., & Abbas, M. F. A. (2025). Designing a model to analyze the impact of applied and mathematical subjects on student academic achievement. Journal of Artificial Intelligence and Computational Technology, 2(2), 2025.

ABSTRACT

Educational institutions need to analyze their data to improve the educational process, which faces numerous challenges, such as the difficulty of measuring learning outcomes and the factors that influence them, and the lack of knowledge about the causes of student academic decline. This research aims to provide a comprehensive study of the use of data mining techniques in educational institution data. This study utilizes classification technology, one of the most important data mining techniques, and applies it to a sample of students from the College of Computer Studies and Information Technology at the Universities of Omdurman Islamic University and West Kordofan University.

The data were analyzed using their grades. Five classification models were constructed and compared in terms of the accuracy of the results, model construction time, and error rate. The goal was to ensure the best results. The JRIP algorithm was chosen, achieving the best results among the algorithms based on the specified factors. This enabled the researcher to clearly present the results, analyze, and understand the extent of the impact of practical subjects on student academic achievement compared to theoretical subjects, and propose appropriate solutions to be presented to relevant educational authorities to assist them in decision-making.

1. Introduction

Educational data mining (EDM) is a recent trend in educational data mining that focuses on using techniques, tools, and research designed to extract useful information and patterns from massive amounts of data generated by or related to student learning activities in an educational context.

With the significant increase in the volume of educational data available within universities, it has become necessary to use data mining techniques to extract useful knowledge from this data. Educational data mining aims to analyze student data to identify hidden patterns and relationships that can help improve academic performance and support decision-making. [1]

Knowledge discovery in databases (KDD) is a method for discovering and extracting hidden patterns from large data repositories. Educational data mining (EDM).

a recent trend in educational data mining that focuses on using techniques, tools, and research designed to extract useful information and patterns from massive amounts of data generated by or related to student learning activities in an educational context. This term draws on a variety of literature, including data mining (DM), machine learning (ML), psychometrics and statistics, information visualization, and computational modeling. It also relates to the social sciences, examining student behavior from social and cultural perspectives. Figure (1) below illustrates the use of DM in educational settings. It represents an iterative process of generating, testing, and improving hypotheses, through which systems can be tailored to meet the needs of each individual within the educational community. The practice of electronic data management transforms raw data from educational institutions into useful information that can significantly impact educational research and practice. [2]. The sequences of steps identified in extracting knowledge from data are shown in Figure 1.

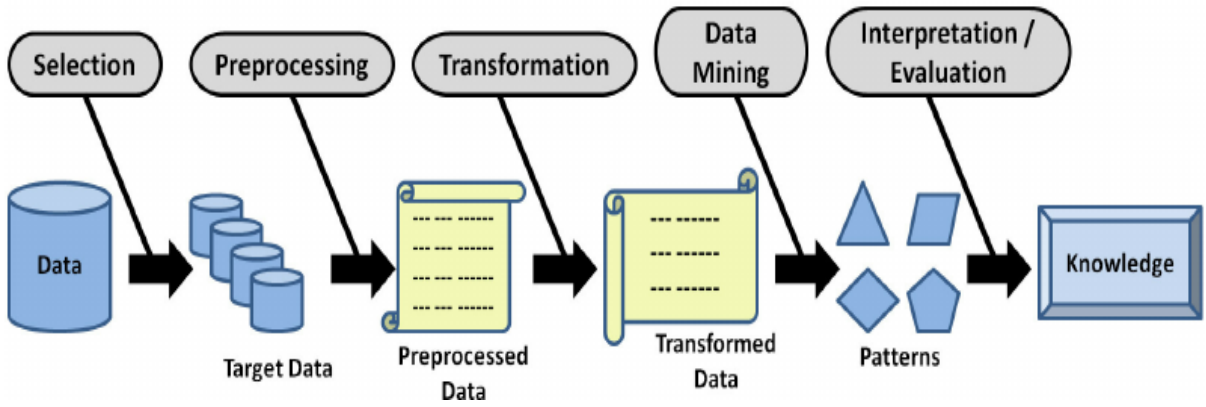


Figure 1: The steps of extracting knowledge from data

Deferens algorithms and techniques like Classification, Clustering, Regression , Association Rules, Nearest Neighbor method etc. are used for knowledge discovery from databases. These techniques and methods in data mining need brief mention to have better Understanding.[3]

A. Classification

DOI: https://doi.org/10.70274/jaict.2025.2.2.65 .
p- ISSN 34562-3478 e- ISSN 5443-1243

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large.

Classification is a task in data mining that involves assigning a class label to each instance in a dataset based on its features. The goal of classification is to build a model that accurately predicts the class labels of new instances based on their features.

There are two main types of classification: binary classification and multi-class classification. Binary classification involves classifying instances into two classes, such as "spam" or "not spam", while multi-class classification involves classifying instances into more than two classes.

algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier.

The process of building a classification model involves the following steps:

Data Collection

The first step in building a classification model is data collection. In this step, the data relevant to the problem at hand is collected. The data should be representative of the problem and should contain all the necessary attributes and labels needed for classification. The data can be collected from various sources, such as surveys, questionnaires, websites, and databases.

Data Pre-processing

The second step in building a classification model is data pre-processing. The collected data needs to be pre-processed to ensure its quality. This involves handling missing values, dealing with outliers, and transforming the data into a format suitable for analysis. Data pre-processing also involves converting the data into numerical form, as most classification algorithms require numerical input.

Handling Missing Values: Missing values in the dataset can be handled by replacing them with the mean, median, or mode of the corresponding feature or by removing the entire record.

Dealing with Outliers: Outliers in the dataset can be detected using various statistical techniques such as z-score analysis, boxplots, and scatterplots. Outliers can be removed from the dataset or replaced with the mean, median, or mode of the corresponding feature.

Data Transformation: Data transformation involves scaling or normalizing the data to bring it into a common scale. This is done to ensure that all features have the same level of importance in the analysis.

Feature Selection:

The third step in building a classification model is feature selection. Feature selection involves identifying the most relevant attributes in the dataset for classification. This can be done using various techniques, such as correlation analysis, information gain, and principal component analysis.

Correlation Analysis: Correlation analysis involves identifying the correlation between the features in the dataset. Features that are highly correlated with each other can be removed as they do not provide additional information for classification.

Information Gain: Information gain is a measure of the amount of information that a feature provides for classification. Features with high information gain are selected for classification.

Principal Component Analysis:

Principal Component Analysis (PCA) is a technique used to reduce the dimensionality of the dataset. PCA identifies the most important features in the dataset and removes the redundant ones.

Model Selection:

The fourth step in building a classification model is model selection. Model selection involves selecting the appropriate classification algorithm for the problem at hand. There are several algorithms available, such as decision trees, support vector machines, and neural networks. Decision Trees: Decision trees are a simple yet powerful classification algorithm. They divide the dataset into smaller subsets based on the values of the features and construct a tree-like model that can be used for classification.

Model Training

The fifth step in building a classification model is model training. Model training involves using the selected classification algorithm to learn the patterns in the data. The data is divided into a training

DOI: https://doi.org/10.70274/jaict.2025.2.2.65 .
p- ISSN 34562-3478 e- ISSN 5443-1243

set and a validation set. The model is trained using the training set, and its performance is evaluated on the validation set.

Model Evaluation

involves assessing the performance of the trained model on a test set. This is done to ensure that the model generalizes well.

B. Clustering

Clustering in data mining is a technique used to group similar data points into clusters, where objects within a cluster are more alike than those in other clusters. It's a fundamental unsupervised learning method for discovering hidden patterns and structures within data. Clustering helps in understanding data relationships, segmenting data, and identifying outliers.

C. Association Rule

Association rule mining finds interesting associations and relationships among large sets of data items. This rule shows how frequently a itemset occurs in a transaction. A typical example is a Market Based Analysis. Market Based Analysis is one of the key techniques used by large relations to show associations between items.It allows retailers to identify relationships between the items that people buy together frequently. Given a set of transactions, we can find rules that will predict the occurrence of an item based on the occurrences of other items in the transaction.

Rule Evaluation Metrics

- Support(s) - The number of transactions that include items in the {X} and {Y} parts of the rule as a percentage of the total number of transaction. It is a measure of how frequently the collection of items occur together as a percentage of all transactions.
- Support = $\sigma(X+Y) \div$ total - It is interpreted as fraction of transactions that contain both X and Y.
- Confidence(c) - It is the ratio of the no of transactions that includes all items in {B} as well as the no of transactions that includes all items in {A} to the no of transactions that includes all items in {A}.
- Conf(X=>Y) = Supp(X$\cup$Y) ÷ Supp(X) - It measures how often each item in Y appears in transactions that contains items in X also.
- Lift(l) - The lift of the rule X=>Y is the confidence of the rule divided by the expected confidence, assuming that the item sets X and Y are independent of each other .The expected confidence is the confidence divided by the frequency of {Y}.
- Lift(X=>Y) = Conf(X=>Y) ÷ Supp(Y) - Lift value near 1 indicates X and Y almost often appear together as expected, greater than 1 means they appear together more than expected and less than 1 means they appear less than expected .Greater lift values indicate stronger association.

2. Related Works

Educational data mining (EDM) has gained significant attention in recent years as a valuable tool for predicting student performance, identifying at-risk students, and uncovering patterns in educational data. Various studies have explored different techniques for analysing student data, including classification algorithms and association rule mining.

Lee and Kim (2020) employed machine learning techniques, particularly classification algorithms such as decision trees and logistic regression, to predict student performance in e-learning systems. Their study demonstrated that classification models could effectively predict students' academic success based on various factors like engagement and interaction with course materials. In this work, the authors highlighted the potential of these algorithms for providing early warnings of students at risk of failing, a concept that aligns with our goal of predicting student performance based on their academic behavior in the Computer Science faculty.[4]

DOI: https://doi.org/10.70274/jaict.2025.2.2.65 .
p- ISSN 34562-3478 e- ISSN 5443-1243

Similarly, Singh and Kumar (2022) conducted a comprehensive review of recent trends in educational data mining, focusing on advanced techniques such as deep learning and neural networks. While their review did not directly address classification or association rule mining, it underscored the growing role of machine learning methods in extracting meaningful patterns from student data. This review informs our approach, as it encourages the use of sophisticated algorithms like decision trees and association rule mining to analyse the grades and behaviors of students in higher education institutions.[5]

Ali and Ahmed (2021) adopted a hybrid data mining approach to predict student dropout in online learning environments. By combining decision tree classification with clustering techniques, they were able to identify students at risk of dropping out. This approach is conceptually similar to our study, where we employ classification algorithms to predict academic success and association rule mining to uncover relationships between various academic factors, such as participation, attendance, and assignment completion.[6]

Sharma and Pandey (2023) applied classification techniques to improve student retention in higher education. Their study used various classification algorithms to predict student retention, demonstrating that early identification of at-risk students could lead to effective interventions. This concept of identifying at-risk students is directly applicable to our study, as we aim to classify Computer Science students based on their academic performance and behavior to predict their success and retention.[7]

Al-Suqri et al. (2024) explored the use of association rule mining to develop personalized learning pathways in online education. Their research demonstrated how association rule mining could uncover hidden patterns in student data, allowing educators to tailor learning experiences to individual needs. In our study, we adopt a similar approach by using association rule mining to identify correlations between academic factors like participation, assignments, and final grades, providing insights into how these factors influence student performance in the Computer Science faculty.[8]

In summary, published studies highlight the effectiveness of classification and association rule mining algorithms in predicting student performance, identifying at-risk students, and uncovering valuable patterns in educational data. This study builds on these works by applying classification and association rule mining techniques to analyze student grades in the Faculty of Computer Science at Omdurman Islamic University, with the aim of improving our understanding of the factors influencing student success.

3. Methodology

Data mining is the process of transforming raw data into useful information. There are four basic steps in data mining. The steps are: data gathering, data preprocessing, data mining or applying classification algorithms, interpretation (Tan et al. 2016)[9]. Figure 1 shows the total methodology of our work.

DOI: https://doi.org/10.70274/jaict.2025.2.2.65 .
p- ISSN 34562-3478 e- ISSN 5443-1243

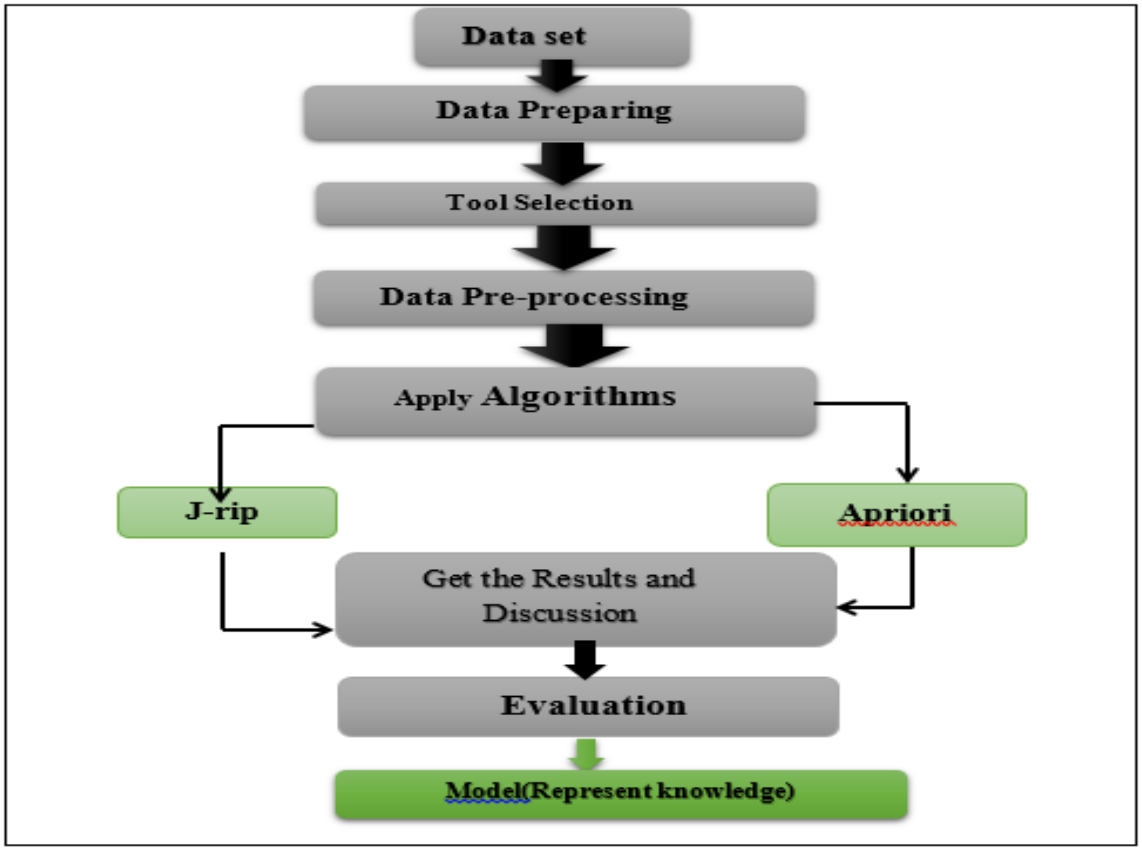


Figure 1 shows the total methodology of our work.

A. Data Mining Process

To evaluate students and determine their levels at Omdurman Islamic University and all Sudanese universities, there is an internal assessment and end-of-semester examination. This internal assessment is based on students' performance in educational activities, such as tests, semester work, assignments, seminars, and laboratory work. Students are evaluated by an examination for each subject in the relevant semester. Each student must obtain the minimum grade to pass that semester.

B. Data Preparations

The database we intend to apply some data mining techniques to is a sample of data from the Faculties of Computer Science and Information Technology at Omdurman Islamic University. The sample includes data from two semesters of students in the Information Technology Department, for full-time students at the university, during the 2015-2016 academic year, totaling 643 records. In this step, data stored in different tables was merged into a single table after eliminating merge errors.

C. Data and Transformation

Here, only the fields required for data extraction are selected. Some derived variables are selected, while some information about the variables is extracted from the database..

D. Pre-processing:

Academic data were collected from the College of Computer Science and Information Technology at Omdurman Islamic University. The data contains 643 records, as shown in Figure 2 Pre-processed data.

DOI: https://doi.org/10.70274/jaict.2025.2.2.65 .
p- ISSN 34562-3478 e- ISSN 5443-1243

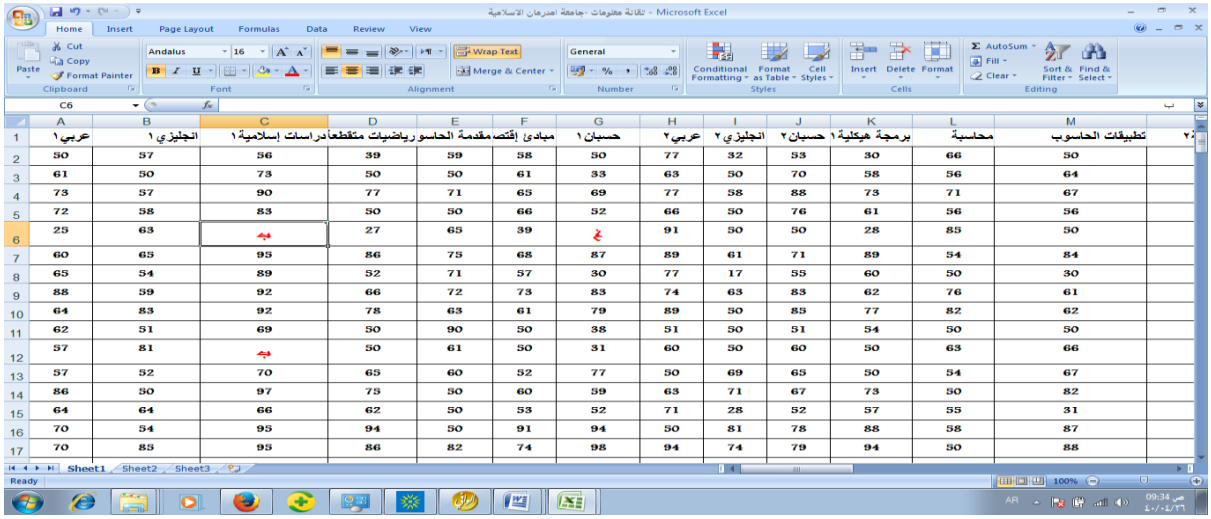


Figure (2)Sample of pre-processing data

E..Data Preparing:

In the process of data processing, we must work to ensure that our data is free from problems, as this data may have. Among the subscribers who may face researchers looking for data collected from different sources and choices are: - Outliers (noisy data), subscription prices (missing value), and not all of them (inconsistent data). A general problem generally occurs due to the absence of data in a statement dedicated to unique data during loading, or when no information about it is provided(not available).

In this study, two cases of missing data will be addressed. The first case is when a student is absent or has a substitute for one or more exams, as shown in Figure (3).

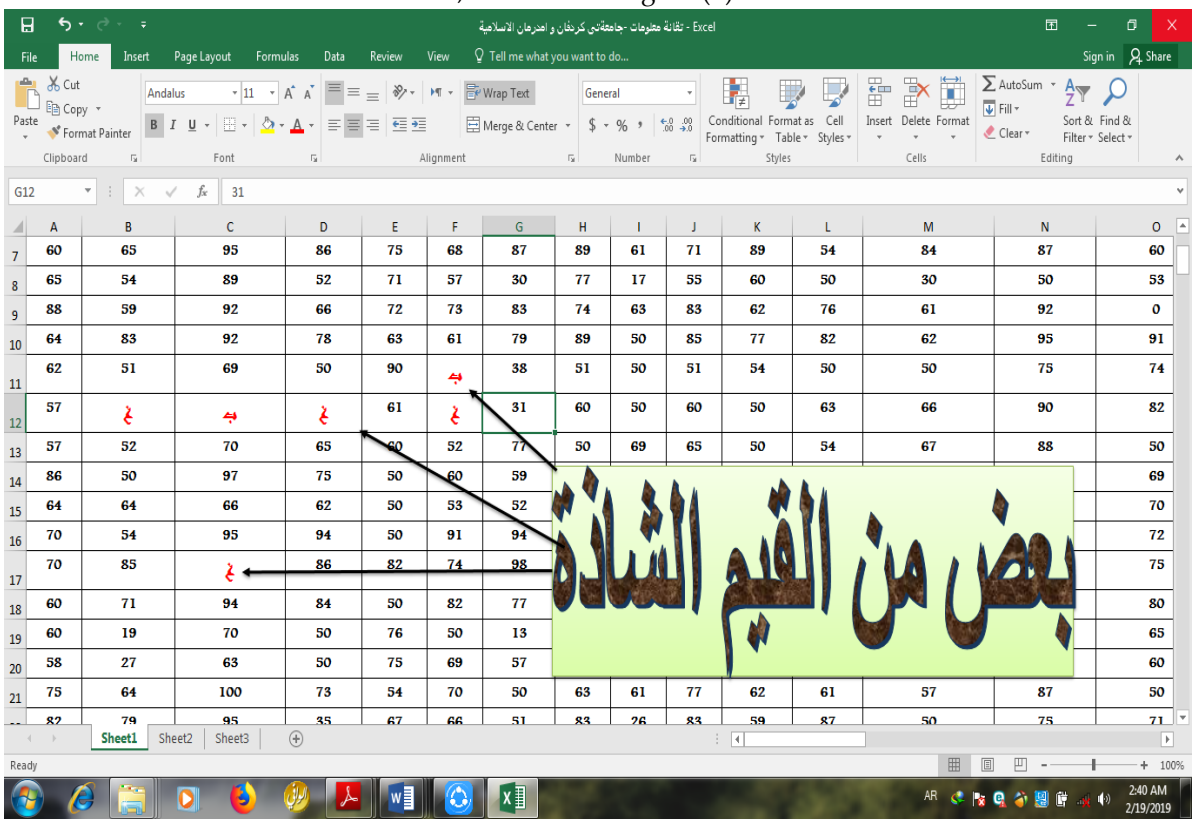


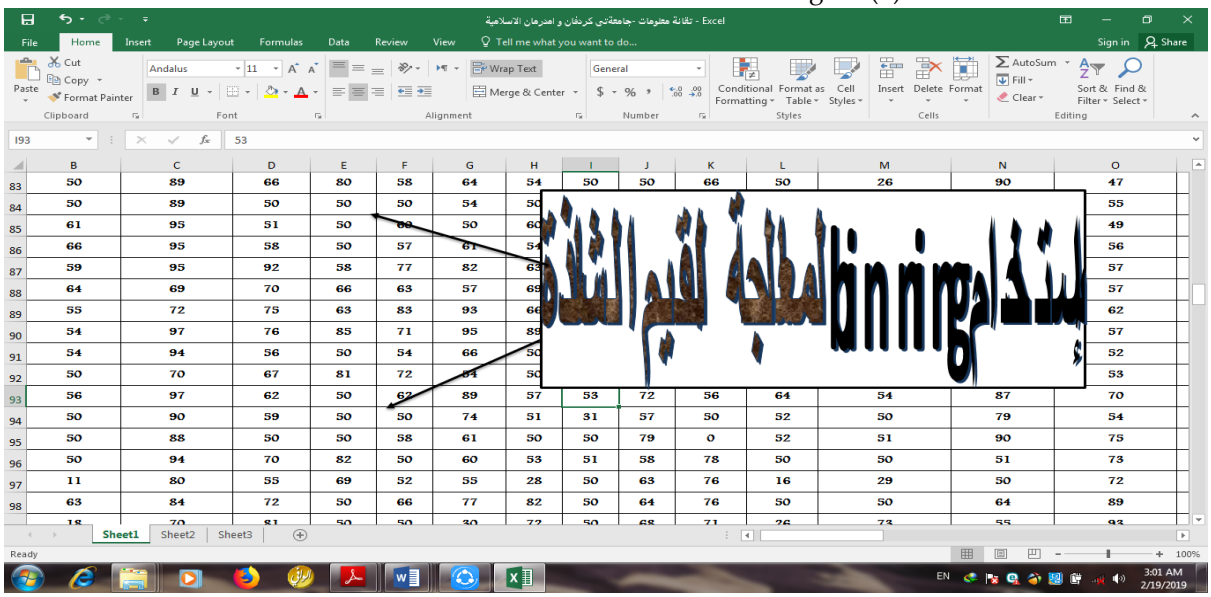
Figure **(3): Example of some missing and outlier values**

F. Data Cleaning using Bagging Method

DOI: https://doi.org/10.70274/jaict.2025.2.2.65 .
p- ISSN 34562-3478 e- ISSN 5443-1243

bagging is a method of smoothing out multiple values of a variable by identifying new, limited values of the same type by consulting adjacent or surrounding values.

By using the Binning Method, we dealt with determining new values of a limited number, so that they are of the same type, by consulting the values adjacent to them or surrounding them for all samples of abnormal values in each record due to absence or alternative as in Figure (4).



Figure(4) Data processing by Binning

G..Data Transformation:

The data or grades are converted into point values for each subject and then combined into forms suitable for data mining techniques as in Figure (5) Converting grades into point values.

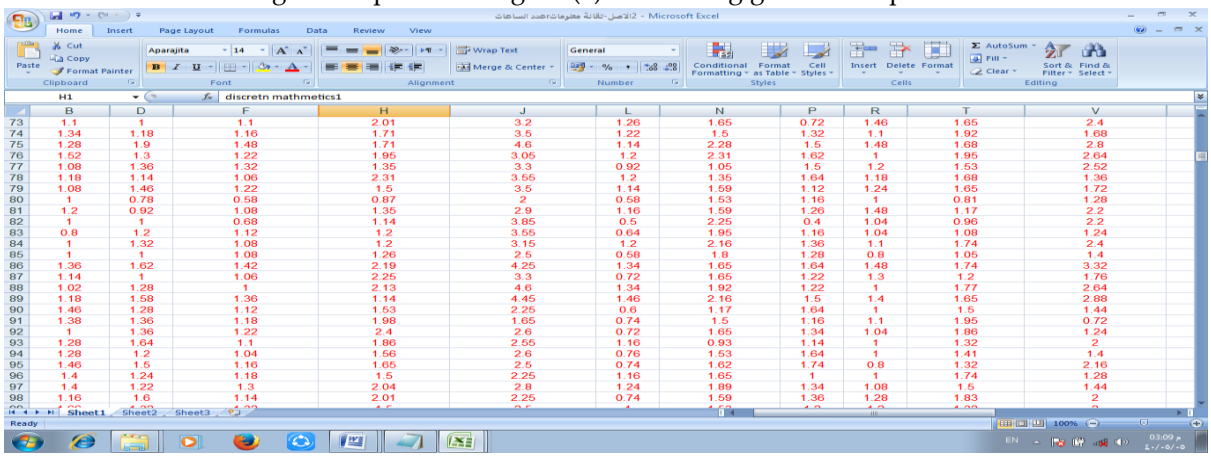


Figure (5) Converting scores to point values

In this study, the normalization technique is used by linking the scores to different ratios. The attribute data is scaled so that it falls within a small specified range, as shown in Table3.

Table 3: Normalization data by selected range

Degree of student	Symbol	Estimate of the Rate
3.50 –4.00	A	Excellent
3.00–3.49	B	Very Good
2.50–2.99	C	Good
2.00–2.49	D	Pass

DOI: https://doi.org/10.70274/jaict.2025.2.2.65 .
p- ISSN 34562-3478 e- ISSN 5443-1243

0–1.99	F	Fail

Finally, we classify it as in the table above according to the point value that the student achieved in the subject to appear as in Figure (6).

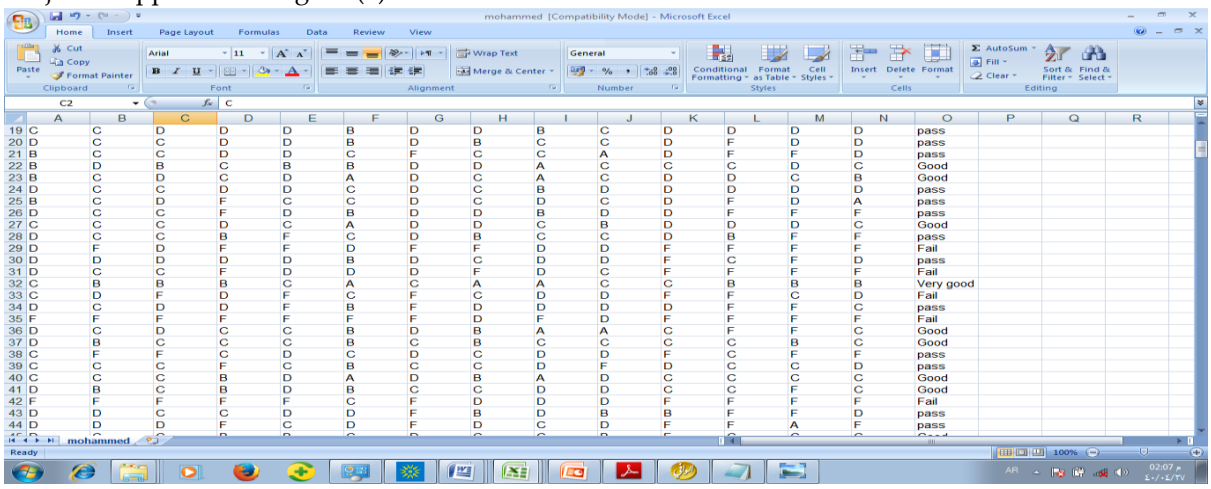


Figure (6) Data normalization

H. Preparing data for use in the Weka Tool:

At this stage, the researcher converted data from an Excel file into a CSV format to be displayed on the Weka interface, as shown in(7)

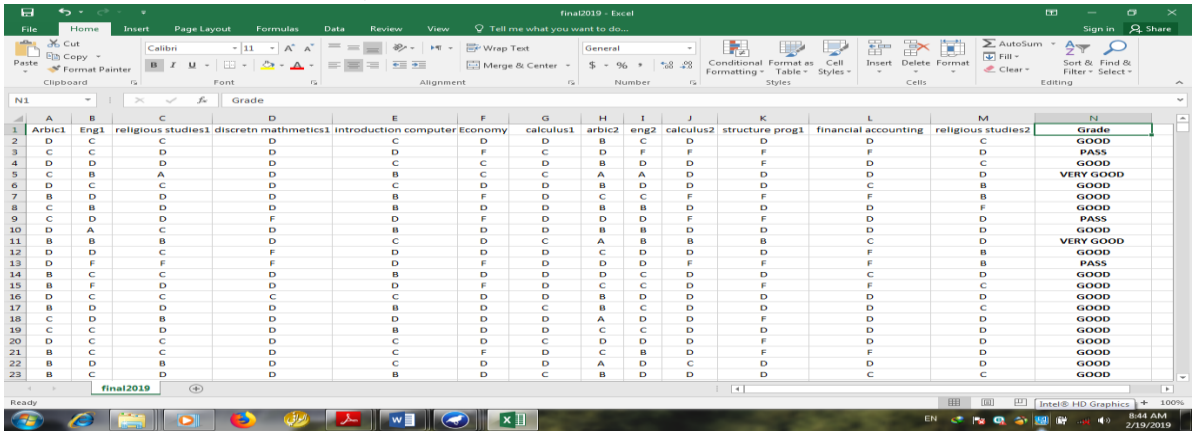
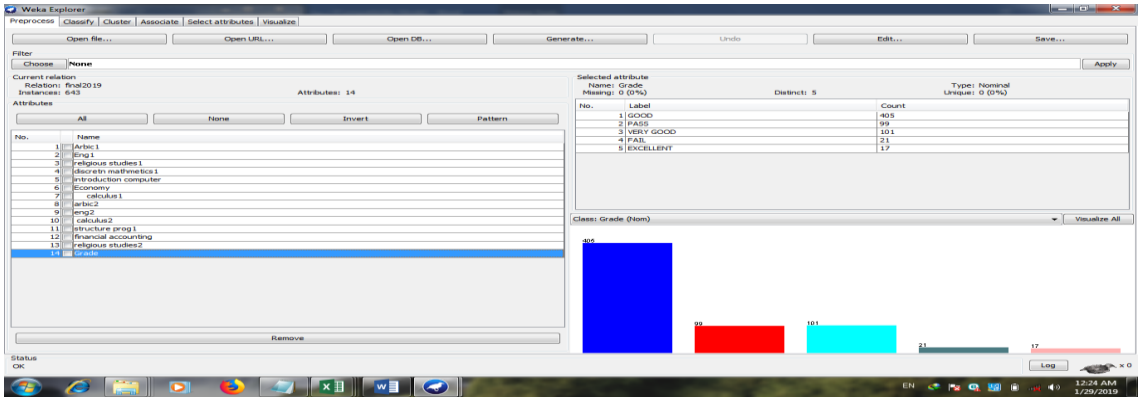


Figure (7): Preparing data for use in the Weka Tool

I . Uploading a data file to Weka:

Select Explore from the four options on the right side of the figure, and the explorer interface will appear. Then, select Open File and select the file format you want to upload and work with. Here, the file is in .csv format. After uploading, it will appear as shown in Figure (8).



DOI: https://doi.org/10.70274/jaict.2025.2.2.65 .
p- ISSN 34562-3478 e- ISSN 5443-1243

Figure (8) Loading the data file into weka

in fact, the figure shows what is displayed after loading the data. The first six lists represent the basic operations supported by the explorer. Currently, we are in the preprocessing phase. After selecting the .CSV file and uploading it to the program, we specify the attributes we want to perform operations on. There is also the option to select all of them and run experiments on them. Then, using the graph, we can choose to visualize all, so the program displays the grades for all subjects and the average.

J. Algorithms used in the research

The data was classified into five different categories based on the student's grade in the average as shown in Table (3-1), and then the grades of each subject were displayed separately and classified. Then comes the role of training and testing on the data in order to build classification models that depend on a number of data, the most important of which is the algorithm used. To obtain the best results, the results obtained from some selected algorithms were compared, which represented previous studies in more than one study previously mentioned.

Five algorithm models were built: JRIP, Naive Bayes Decision Table, J48, and ZeroR. These algorithms were selected after conducting several initial tests on the training dataset using several algorithms provided by the Wika tool. These algorithms were found to be the best in terms of results, and then compared in terms of accuracy and error rate. Cross-validation testing method were used to train, test, and build classification models.

Table 3: Students' grades in the average

Number of students	degree	Category number
17	Excellent	1
101	Very Good	2
405	Good	3
99	Pass	4
21	Fail	5

K. Building Model

The classification algorithm (learning) is trained on a training data set containing records of known class and The model's performance is evaluated using classification algorithms, with the goal of achieving the highest accuracy and lowest error rate when applied to the test data. This results in constructing the best model.

The extracted model is executed on the application data set.

Choose classify the loaded data:

Includes five models of classification algorithms:

DOI: https://doi.org/10.70274/jaict.2025.2.2.65 .
p- ISSN 34562-3478 e- ISSN 5443-1243

1) ZeroR algorithm:

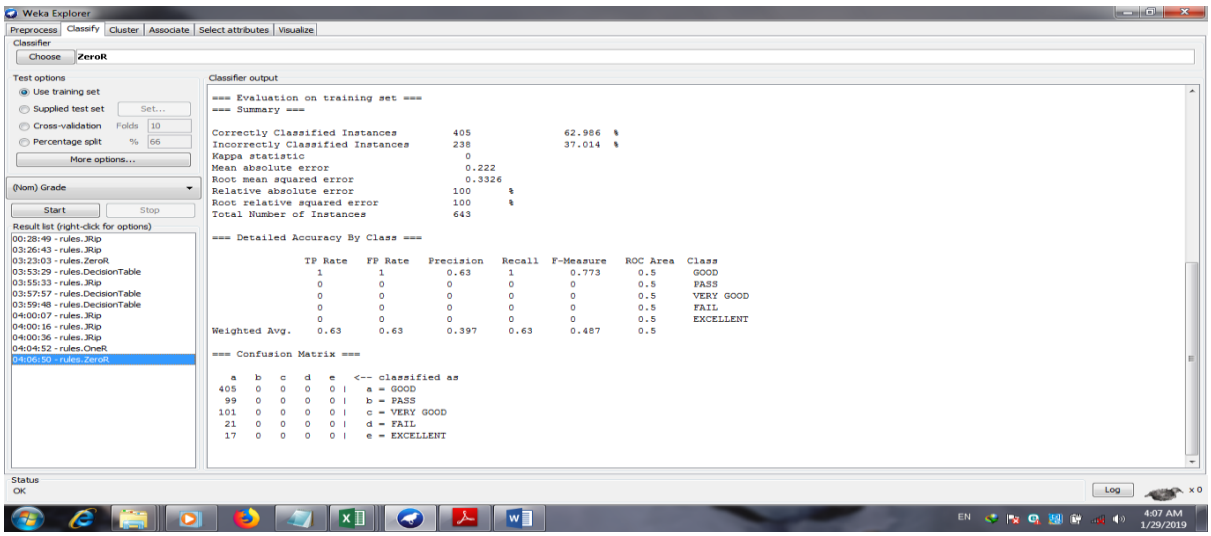


Figure (9) zero algorithm result

2) Decision table algorithm:

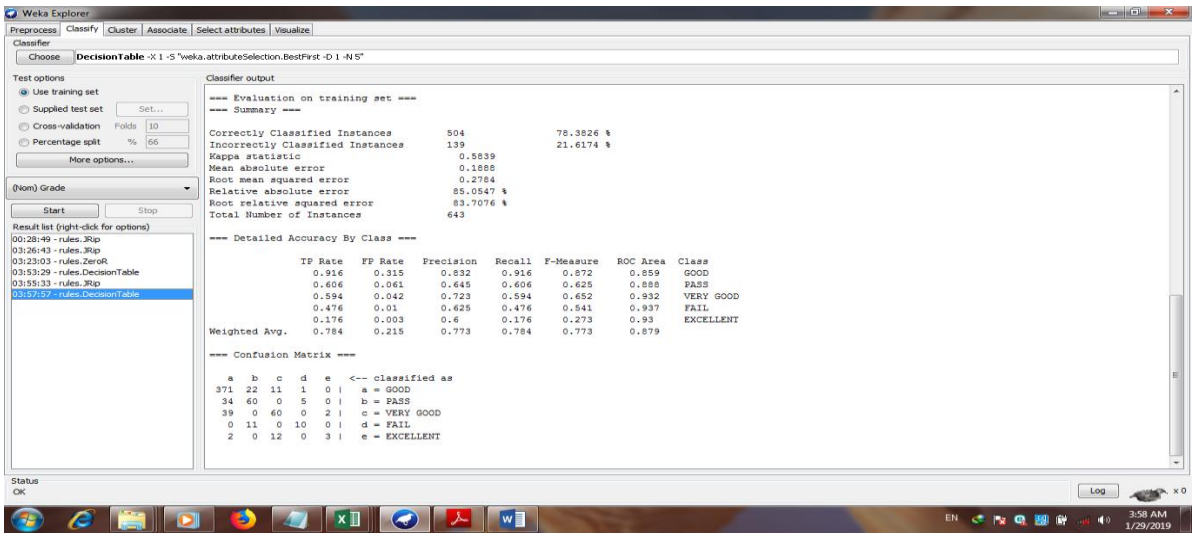
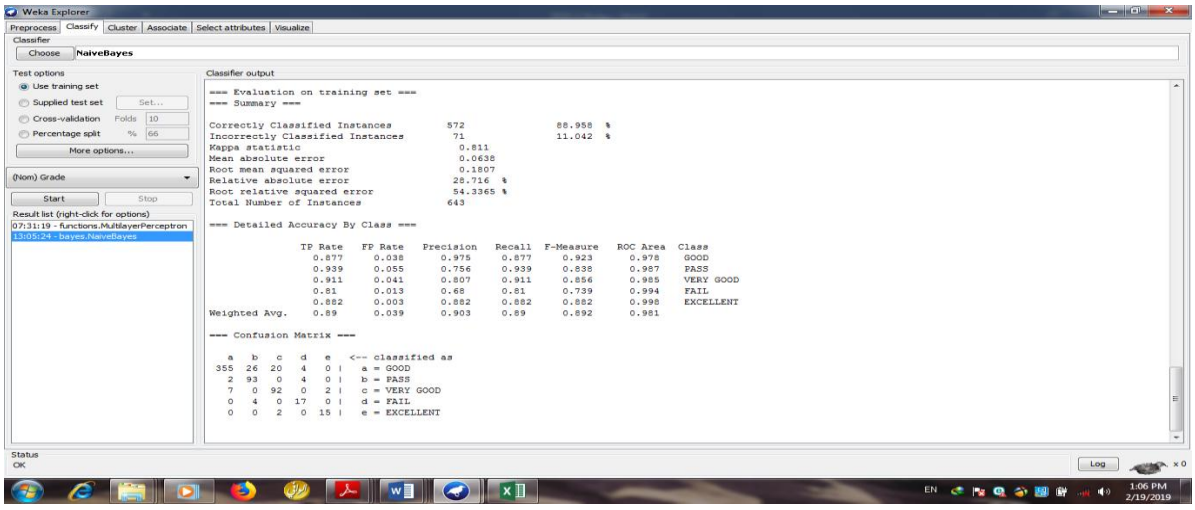


Figure (10) decision table algorithm result

3) Naive Base algorithm:



DOI: https://doi.org/10.70274/jaict.2025.2.2.65 .
p- ISSN 34562-3478 e- ISSN 5443-1243

Figure (11) Naive Base algorithm result

4) J48 algorithm:

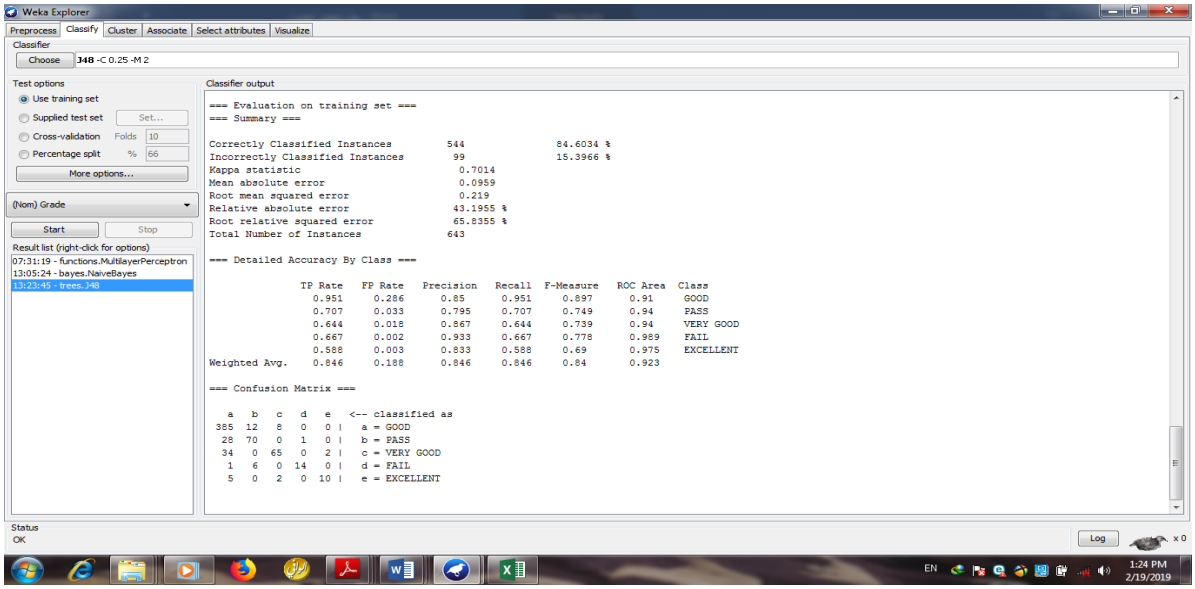


Figure (12) J48 algorithm result

5) Jrip algorithm

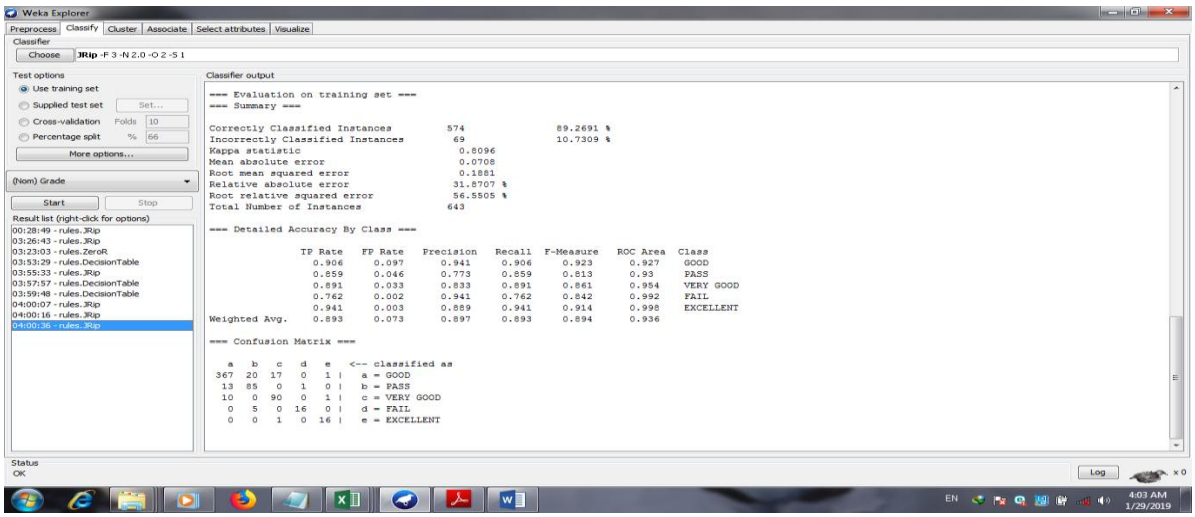


Figure (12) Jrip algorithm result

4. Results and Discussion

After conducting numerous experiments, the researcher arrived at the optimal algorithm for conducting the study in terms of accuracy and execution time. This is the JRIP algorithm, which was the best among the five algorithms selected for the study. After conducting numerous experiments on this algorithm, which is included in the Weka data mining tool, a large number of rules were obtained. After uploading these rules, it was found that some of them were classified as intuitive, while others had varying degrees of logic. Therefore, the researcher designed a model to analyze all the obtained results. We selected the rules that demonstrate the strength of the correlation between the subjects and clearly identify the subjects that have the greatest impact on the GPA, in terms of their impact on the

DOI: https://doi.org/10.70274/jaict.2025.2.2.65 .
p- ISSN 34562-3478 e- ISSN 5443-1243

student's academic level. This helps in making appropriate decisions to develop and improve the educational level of students by increasing academic achievement above what it is currently at our Sudanese university in particular. This can be done by following a specific methodology, such as increasing the number of hours for a particular subject or decreasing the number of hours for a particular subject, depending on its impact on the university student's GPA. Recommendations are made after analyzing each year's results, comparing them to the results of the previous year, and identifying the main influences on Student's academic achievement..

The researcher selected the best results for discussion, which are represented by students' grades in practical and theoretical subjects. This helps in making appropriate decisions to develop and improve the educational process, discussing a portion of it, as shown in Table (4), which illustrates the set of experiments the researcher conducted to select the optimal algorithm:

These algorithms were tested to determine which one was the best in practice, as follows:

Number-wise, based on the criteria, the results were as shown in Table (1-4), the results of the algorithm accuracy.

Table (4) results of algorithm accuracy

algorithm	Determine the error rate of the classifier or model used	Determine the accuracy rate of the classifier or model used	Number of records misclassified	Number of records whose classification was affected
ZeroR	% 37.014	% 62.986	238	405
Decision table	% 21.6174	% 78.3826	139	504
JRIP	% 10.7309	% 89.2691	69	574
NaïveBase	% 11.042	% 88.958	71	572
j48	15.3966 %	% 84.6034	99	544

Here, the researcher chose the JRIP algorithm, which was the best in results among all the algorithms under study, as shown in the table above, based on the accuracy criteria and execution time.The detailed results of the algorithm were as follows:

1.Accuracy by Class:

Table (4-2) Classifier Accuracy

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.906	0.097	0.941	0.906	0.923	0.927	GOOD
0.859	0.046	0.773	0.859	0.813	0.93	PASS

DOI: https://doi.org/10.70274/jaict.2025.2.2.65 .
p- ISSN 34562-3478 e- ISSN 5443-1243

	0.891	0.033	0.833	0.891	0.861	0.954	VERY GOOD
	0.762	0.002	0.941	0.762	0.842	0.992	FAIL
	0.941	0.003	0.889	0.941	0.914	0.998	EXCELLENT
Weighted Avg	0.893	0.073	0.897	0.893	0.894	0.936	

2.Confusion Matrix:

A	B	C	D	E	classified as
367	20	17	0	1	A=GOOD
13	**85**	0	1	0	B=PASS
10	0	**90**	0	1	C=VERY GOOD
0	5	0	**16**	0	D=FAIL
0	0	1	0	**16**	E= EXCELENT

The arrows indicate the number of records that the classifier correctly classified.

5. Conclusion

After reviewing and analyzing the research results, the importance of a student's academic record and its impact on the academic levels of the students under study became apparent. Consequently, the study highlighted the importance of analyzing student data to benefit from it in decision-making. Conducting numerous experiments in this field enables educational institutions to develop their plans and increase efficiency. Data mining is a modern, highly efficient method for analyzing, compiling, and classifying data. By searching for the best algorithm for data classification, we can analyze student results and grades in each subject, analyze their impact on the student's GPA, and establish relationships that increase the system's control and provide the best possible outcomes. The results revealed that the JRIP algorithm delivered the desired results with a high degree of accuracy, making it the best among the algorithms selected in the study. It is also possible to improve student performance in university by identifying students who struggle in practical subjects. University administrations develop effective plans to increase their achievement in these and similar subjects, and provide additional guidance to students upon enrollment at university.

DOI: https://doi.org/10.70274/jaict.2025.2.2.65 .
p- ISSN 34562-3478 e- ISSN 5443-1243

Reference:

[1] Papadogiannis, M. Wallace, and G. Karountzou, "Educational data mining: A foundational overview," Encyclopedia, vol. 4, no. 4, pp. 1644–1664, 2024

[2] X. Du et al., "Educational data mining: A systematic review of research and emerging trends," Information Discovery and Delivery, vol. 48, no. 4, pp. 225–236, 2020

[3] S. Wang and W. Shi, "Data mining and knowledge discovery," in Springer Handbook of Geographic Information, Berlin, Heidelberg: Springer, 2012, pp. 49–58
[4] J. S. Lee and M. K. Kim, "Predicting Student Performance with Machine Learning: A Data Mining Approach," Journal of Educational Data Mining, vol. 12, no. 2, pp. 90-102, 2020.

[5] P. B. Singh and V. Kumar, "Educational Data Mining: A Comprehensive Review of Recent Trends and Techniques," International Journal of Educational Technology in Higher Education, vol. 21, no. 1, pp. 1-21, 2022.

[6] M. T. Ali and S. U. Ahmed, "A Hybrid Data Mining Approach for Predicting Dropout in E-Learning Systems," Computers in Education, vol. 153, pp. 1-12, 2021.

[7] R. Sharma and S. Pandey, "Application of Data Mining in Improving Student Retention in Higher Education Institutions," International Journal of Educational Data Mining, vol. 16, no. 3, pp. 210-225, 2023.

[8] N. A. F. Al-Suqri, F. S. Alghamdi, and A. M. Alharthi, "Data Mining Techniques for Personalized Learning Pathways in Online Education," Education and Information Technologies, vol. 29, no. 2, pp. 999-1014, 2024.

DOI: https://doi.org/10.70274/jaict.2025.2.2.65 .
p- ISSN 34562-3478 e- ISSN 5443-1243