Babylonian Journal of Artificial Intelligence
Vol. **(2023)**, 2023, **pp**. 43–54
DOI: https://doi.org/10.58496/BJAI/2023/008 ISSN: 3006-5437
https://mesopotamian.press/journals/index.php/BJAI

Research Article

Employing Data Mining Techniques and Machine Learning Models in Classification of Students' Academic Performance

Hussein Alkattan[1,2,*], Alhumaima Ali Subhi[3], Oluwaseun A. Adelaja [4], Mostafa Abotaleb[1], Maad M. Mijwil[5], Pradeep Mishra [6], Denis Sekiwu[7], Wilson Bamwerinde[8], Benson Turyasingura[8]

[1] *Department of System Programming, South Ural State University, Chelyabinsk, Russia*

[2] *Directorate of Environment in Najaf, Ministry of Environment, Najaf, Iraq*

[3] *Electronic and Computer Center, University of Diyala, Baqubah MJJ2+R9G, Iraq*

[4] *Information Communication and Technology Department, Lagos State University, Lagos, Nigeria*

[5] *Computer Techniques Engineering Department, Baghdad College of Economic Sciences University, Baghdad, Iraq*

[6] *College of Agriculture, Rewa, JNKVV, (M.P.) India*

[7] *Department of Foundations of Education, Kabale University, Kabale, Uganda*

[8] *Department of Environment and Nature Resources, Kabale University, Kabale, Uganda*

ARTICLE INFO

Article History

Received 06 May 2023
Accepted 07 Jul 2023
Published 02 Aug 2023

Keywords

Data Mining

Machine Learning

Academic Performance

Decision Trees

Education Analysis

ABSTRACT

The study deals with the use of data mining techniques to build a classification model to predict students' academic performance. The research indicates that the use of machine learning models and data mining methods can reveal hidden patterns and relationships in big data, making them indispensable tools in the field of education analysis. Special emphasis was placed on the use of algorithms such as decision trees. The study includes an analysis of factors that affect students' academic performance such as previous academic achievement in educational activities, as well as social and psychological factors. Classification models were applied using the KNIME platform and the WEKA tool to analyse students' performance in three courses: database technology, artificial intelligence, and image processing in the ICT degree program. The results showed that the use of decision trees can effectively classify students' performance and determine the success and failure rate. The cruel outright mistakes, RMS error and relative supreme mistake all showed 0% whereas the kappa esteem got from the analysis extended between 0.991 and 1.00 which significantly concurs with most statistical values.

1. INTRODUCTION

The concept of predicting students' academic performance using classification models based on data mining techniques is attracting increasing attention from educational researchers and practitioners. Modern information technologies and the abundance of data open up new opportunities for analyzing and predicting learning outcomes. In recent years, interest in the application of machine learning and data mining methods in the educational field has increased significantly [1], [2]. The purpose of this work is to study the effectiveness of using various classification methods to predict students' academic performance in the Information, Communication and Technology (ICT) program. In particular, special attention is paid to the use of algorithms such as decision trees, random forests, support vector machines and neural networks [3], [4]. Data mining methods can reveal hidden patterns and dependencies in large amounts of data, making them indispensable tools in the field of educational analytics [5], [6]. For example, classification algorithms can be used to predict the probability of a student successfully completing a course based on his or her previous academic achievements and learning behavior [7], [8]. Recent research shows that the use of machine learning methods in the educational field can significantly improve the accuracy of predictions of students' academic performance [9], [10]. In particular, the combination of different classification methods and the use of ensemble methods allows achieving high accuracy and reliability of predictions [11], [12]. An

*Corresponding author. Email: alkattan.hussein92@gmail.com

important aspect of predicting academic performance is the selection and processing of raw data. The main factors affecting the results include students' academic achievement, class attendance, participation in learning activities, socio-economic status, and psychological characteristics [13], [14]. Proper data processing and selection of important features are important stages in building a classification model [15], [16].
Research on learning analytics also highlights the importance of using adaptive and personalized learning methods [17], [18]. The use of machine learning methods makes it possible to develop individual learning paths and recommendations for each student, which helps to increase his motivation and academic performance [19], [20]. The development and implementation of models for predicting academic performance requires an interdisciplinary approach and cooperation between specialists from different fields: pedagogy, education management, psychology, information technology, and statistics [21], [22]. This makes it possible to consider various aspects of the educational process and find comprehensive solutions to improve the quality of education [23], [24]. Another important aspect is the ethical aspect of using data and machine learning algorithms in the educational field. It is necessary to ensure the confidentiality and protection of students' personal data, as well as considering the potential risks and limitations associated with the use of automated decision-making systems [25], [26]. Prospects for further research include the development of new methods and algorithms, as well as their integration into educational platforms and learning management systems [27], [28]. The implementation of such solutions can help create more effective and adaptive learning environments that can support and develop each student [29], [30]. The J4.8 decision tree classifier and the decision tree learner workflow node were designed in the analysis stages using WEKA and KNIME to classify students into passers and failures. The confusion matrix network was constructed based on the results of three courses: database technology, artificial intelligence, and image processing. In addition, bar graph plots were used to analyze the dimensions of failed and successful students, which helped teachers and students improve their teaching and learning methods separately.

2. RELATED WORKS

In the field of predicting students' academic performance, many studies have been conducted that focus on exploring different methods and techniques. Previous work has focused on using machine learning and data analysis techniques to build classification models capable of predicting academic performance based on a variety of factors. One of the common methods for predicting students' academic performance is the use of decision trees, which have proven effective due to their ability to process both numerical and categorical data [31], [32]. For example, in the study [33], the use of decision trees was demonstrated to analyze students' academic performance data and identify the key factors that influence their success. Random forests are another effective technique used in predicting academic performance. This technique allows for improving the accuracy of predictions by creating and merging multiple decision trees [34], [35]. Studies show that random forests can be used to analyze large amounts of data and uncover important patterns [36].

Support vector machines (SVMs) have also found wide application in educational analytics and have shown high effectiveness in classification tasks related to predicting academic performance [37], [38]. For example, the study [39] describes the use of SVM to predict students' academic performance based on their academic achievements and social factors. Neural networks and deep learning are powerful tools for analyzing complex and nonlinear data. These techniques have shown high accuracy in academic performance prediction tasks, especially when large amounts of data are involved [40], [41]. In particular, the study [42] demonstrated the use of neural networks to predict student success based on a variety of educational data. The combination of different machine learning techniques, such as clustering methods, combines their advantages to achieve higher accuracy and reliability in predictions [43], [44]. For example, the study [45] presents an approach based on the combination of random forests and support vector machines to predict students' academic performance. A key aspect in predicting academic performance is the selection of important features and data processing, as these elements help ensure that the predictive models are both accurate and relevant to the specific context of the academic environment.

In the study [46], feature selection methods and their impact on the accuracy of prediction models are discussed. Data processing by methods such as normalization and handling of missing values are also considered very important [47]. Studies also emphasize the importance of psychological and social factors in predicting academic performance. In the study [48], the impact of student motivation and engagement on academic performance is investigated. Socio-economic factors also play a significant role in student success, as shown in the study [49]. In addition, adaptive and personalized educational methods, based on data analysis, can significantly contribute to increasing students' motivation and academic performance [50]. In the study [51], the use of recommendation systems to provide individualized learning paths is investigated. The ethical aspects of using data and machine learning algorithms in the educational field are also important. In the study [52], issues of privacy and protection of student data are investigated, as well as potential risks and limitations. Future research prospects include the development of new methods and algorithms and their integration into educational platforms and learning management systems [53], [54]. These solutions can contribute to the creation of more effective and adaptive learning environments that support each student individually [55].

3. CLASSIFICATION AS A PROCEDURE IN DATA MINING

Classification may be a prescient data mining method, where modern substances are relegated to pre-existing classes by carefully analyzing the highlights of these substances [56]. Classification is utilized to foresee the results of concealed cases based on Previous choices [57, 58]. Classification may be a directed learning handle that makes a difference anticipate the lesson of objects that have not however been classified, utilizing labeled information to prepare the show. The most reason of performing classification errands is to find particular connections between input qualities and yield lesson, and to extricate valuable cognitive patterns in foreseeing the lesson of modern obscure objects [59, 60]. The learning handle in classification strategy includes employing a preparing set to construct the show; taken after by an estimation step to decide the precision of the model employing a test set; and at last, a utilization step employment the relationship between input and output highlights of the demonstrate Classification to assist foresee concealed data within the future [61, 62]. These three fundamental steps of classification appeared in Figure 1.

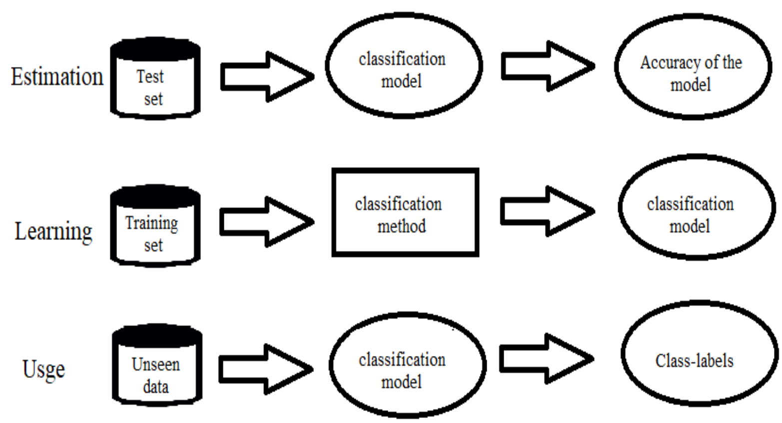

Fig. 1. Shows the Classification for the Model.

4. J4.8 ALGORITHM DECISION TREE CLASSIFICATION

The J4.8 decision tree classification algorithm is one of the most popular algorithms used in the field of machine learning. This algorithm is based on creating a predictive model based on dividing the data into sub-branches, where each branch represents a decision based on the value of a specific attribute [63]. The algorithm starts by selecting the attribute that achieves the greatest distinction between different classes of data. This is done using metrics such as Information Gain or Gain Ratio [64].
The gain information is calculated as follows:

$$Entropy\ (S_v) \times \frac{|\ vS}{|\ S}\ {}_{v\in\ Values\ (A)}\Sigma - Entropy\ (S) = IG(S,A) \qquad (1)$$

$IG(S,A)$ is the information gain of the attribute, $Entropy(S)$ is the entropy of the group and vS It is the subset of S in which attribute A has value v.

$$p_i \log_2 (p_i)_{i=1}^{c}\Sigma -= \text{Entropy } (S) \qquad (2)$$

iP is the relative probability of class i in the set S.
The data is divided into subsets based on the value taken by the selected attribute. A new node is created in the tree representing this attribute [65], [66].
Iterating the process:

- The process of selecting the attribute and dividing the data is repeated on each subset until one of the following conditions is reached:
- All samples in the subset reach the same class.
- There are no other attributes to divide.
- The tree reaches a certain depth that was previously specified [67].

- Dealing with dead branches:

Dead branches or nodes that do not add new information are dealt with by pruning the tree, which helps reduce complexity and increase the generalization ability of the model [68].

5. DEVELOPMENT OF THE CLASSIFICATION MODEL

The consider was conducted utilizing one of the most data mining apparatuses, to be specific the J4.8 choice tree classifier. This classifier was connected to construct a classification show utilizing the WEKA instrument and the decision trees workflow hub within the KNIME analytics stage for analyzing students' scholarly execution. WEKA software, developed at the College of Waikato, could be a machine learning calculation composed in Java outlined for information mining [61]. KNIME (Konstanz Data Digger) program is simple to download, adaptable and free, giving clients with instruments for information preparing; it is instinctive and continually overhauled with modern highlights. Workflow ventures and data science components made in KNIME are reusable and open to all clients. Most of the data mining and machine learning components are integrated through a measured information handling pipeline in KNIME [69]. A choice tree may be a structure taking after a flowchart, where an inside hub speaks to a test on a property; a department outlines the test result; and the leaf node contains the class label [58, 70]. A choice tree is valuable for information disclosure since it does not require parameter tuning or master information to make models that are simple to examined and translate by humans [72]. The inside hub of a choice tree contains parts and part properties [71]. These strategies are utilized in executing data mining and can be utilized to form classification rules for data sets [73]. A choice set incorporates terminal or leaf nodes that begin from the root hub [74]. The precision of a choice tree is tried utilizing test information, whereas preparing information is utilized to construct the choice tree [74].

6. RESULTS AND DISCUSSIONS

The study analyzed the results of 242 students across three courses using the KNIME platform. The file reader, color manager, decision tree training, and decision tree to image converter nodes were used to analyze the data. The training process showed that the datasets could be classified using decision trees to determine the class labels of the samples. PMML readers were used to connect the models and create the workflow. The PMML node was connected to the decision tree training node and the decision tree to image converter node. Figure 2: shows the decision tree generation workflow in KNIME. The model created decision trees for the three courses, showing the percentage of students who passed and failed the exam. Figure 3: For the Database Technology (DB_TECH) course with a maximum score of 70, 155 students passed the exam with a score of more than 34.3 (64%), and 87 students failed with a score of 34.3 or less (36%). Figure 4: For the Artificial Intelligence (AI) course with a maximum score of 70, 170 students passed the exam with a score of 34.3 or more (70.2%), and 72 students failed with a score of 34.3 or less (29.8%). Figure 5: For the Image Processing course with a maximum score of 70, 156 students passed the exam with a score of 34.5 or more (64.5%), and 86 students failed with a score of 34.5 or less (35.5%). The analysis showed that decision trees can effectively classify student performance across courses.

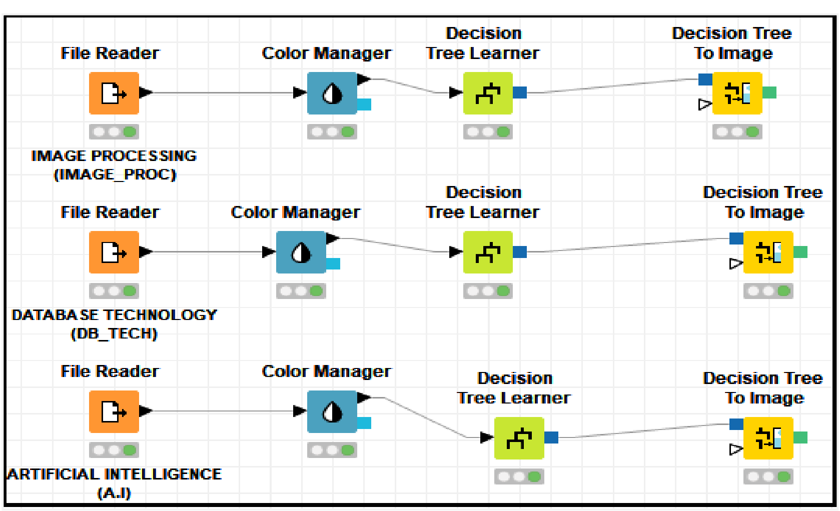

Fig. 2. shows the decision tree generation workflow in KNIME

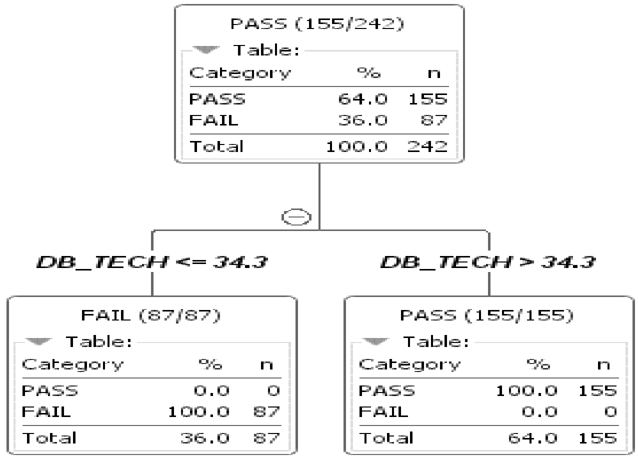

Fig. 3. For the Database Technology (DB_TECH) with a maximum score.

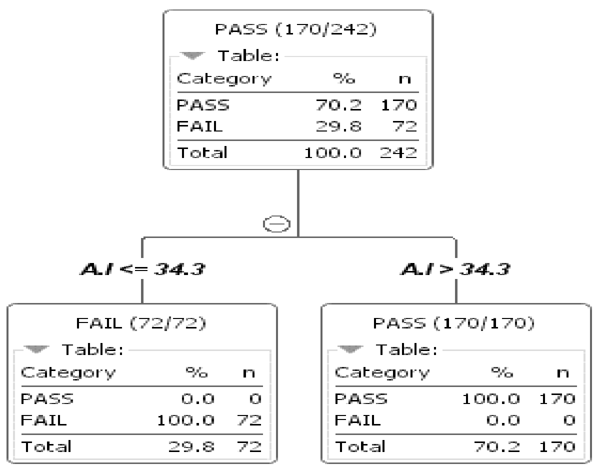

Fig. 4. For the Artificial Intelligence (AI) with a maximum score

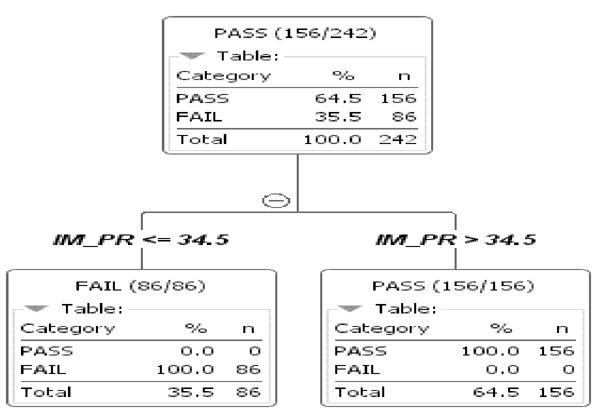

Fig. 5. For the Image Processing with a maximum score

The WEKA analytical platform used the J4.8 pruned tree classifier to generate decision trees and confusion matrices for three courses.
Figure 7: Decision tree for the Database Technology (DB_TECH) course, Figure 8: Decision tree for the Artificial Intelligence (AI) course, Figure 9: Decision tree for the Image Processing course.

WEKA J4.8 Decision Tree Classifier Output for Database Technology (DB_TECH), Artificial Intelligence (A.I) and Image Processing (IMAGE_PROC) Results. The DB_TECH, A.I and Image_Proc Result analysis. The J 4.8 tree generated in WEKA utilized ten (10) fold cross-validations with a full training set to obtain the classifier model.

Total Instances generated (DB_TECH, IMAGE_PROC & A.I): 242
Total Attributes/ Fields utilized: 6
DB_TECH
$A.I$
$IMAGE\_PROC$
$GRADE\_REMARKS\ (DB\_TECH)$
$GRADE\_REMARKS\ (A.I)$
$GRADE\_REMARKS\ (IMAGE\_PROC)$
$DB\_TECH\ <=\ 33.6: FAIL\ (87)$
$DB\_TECH\ >\ 33.6: PASS\ (155)$
$A.I\ <=\ 33.6: FAIL\ (72)$
$A.I\ >\ 33.6: PASS\ (170)$
$IMAGE\_PROC\ <=\ 34: FAIL\ (86)$
$IMAGE\_PROC\ >\ 34: PASS\ (156)$

TABLE I. SUMMARY OF STRATIFIED CROSS VALIDATION FOR DB_TECH, A.I AND IMAG_PROC TEST BASED ON SOME CRITERIA

Courses.	No of leaves produced	Tree size.	Time utilized to obtain model (seconds).	Instances classification based on the value of correctness.	Instances classification based on the value of incorrectness.	Kappa value obtained.	Mean Absolute (M.A) error Value.	Root Mean Squared (R.M.S) error value.	Relative Absolute (R.A) Error value.	Root Relative Squared (R.R.S) Error Value.
DB TECH	2	3	0.01 secs	242 (100%)	0 (0%)	1	0	0	0%	0%
AI	2	3	0.02 secs	242 (100%)	0 (0%)	1	0	0	0%	0%
IMAGE_PROC	2	3	0.05 secs	241 (99.5868%)	1 (0.4132%)	0.991	0.0041	0.0643	0.091%	13.4282%

a and b in the confusion matrix were classified as PASS and FAIL value respectively, a represents class for the number of students who PASSED while b is equal number of students who FAILED. The confusion Matrix obtained for the three courses in which the students were examined are displayed in a 2*2-dimensional order as illustrated below:

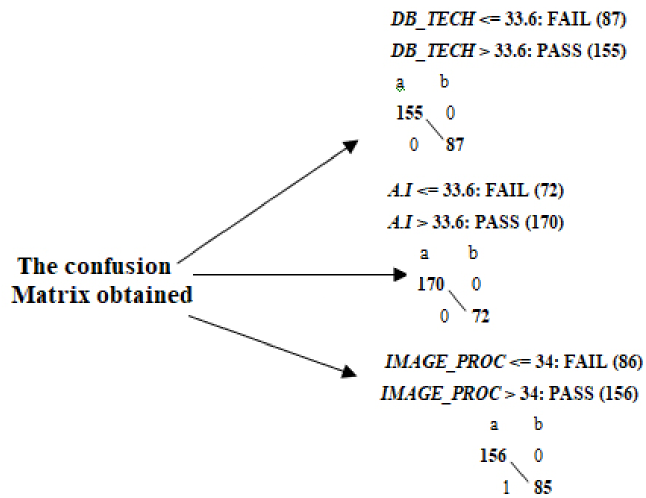

Fig. 6. Show the confusion matrix obtained.

TABLE II. THE DETAILED ACCURACY BY CLASS READINGS OBTAINED FOR THE DATABASE TECHNOLOGY (DB_TECH) AND ARTIFICIAL INTELLIGENCE (A.I) HAS SIMILAR VALUES AS SHOWN BELOW:

Values for the TP Rate	Values for the FP Rate	Precision Values obtained	Recall values	F-Measure Readings	MCC values obtained	ROC Area values obtained	PRC Area values obtained	Class
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	PASS
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	FAIL
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Remarks

TABLE III. THE DETAILED ACCURACY BY CLASS READINGS OBTAINED FOR THE IMAGE PROCESSING (IMAGE_PROC) HAD SLIGHT VARIATIONS DUE TO THE INCORRECTLY CLASSIFIED INSTANCES VALUES AND THE PERCENTAGE OBTAINED WHICH ARE 1 AND 0.4132% RESPECTIVELY.

Values for the TP Rate	Values for the FP Rate	Precision Values obtained	Recall values	F-Measure Readings	MCC values obtained	ROC Area values obtained	PRC Area values obtained	Class
1.000	0.012	0.994	1.000	0.997	0.991	0.994	0.994	PASS
0.988	0.000	1.000	0.988	0.994	0.991	0.994	0.993	FAIL
0.996	0.007	0.996	0.996	0.996	0.991	0.994	0.993	Remarks

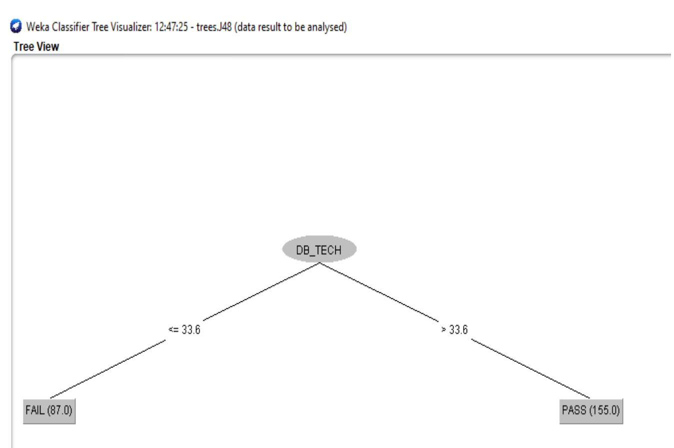

Fig. 7. J 4.8 Classifier generated for DB_Tech grade

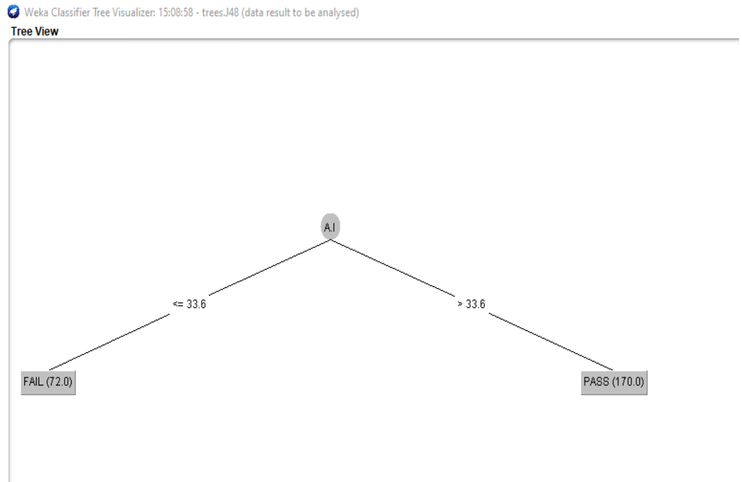

Fig. 8. J 4.8 Classifier generated for A.I grades

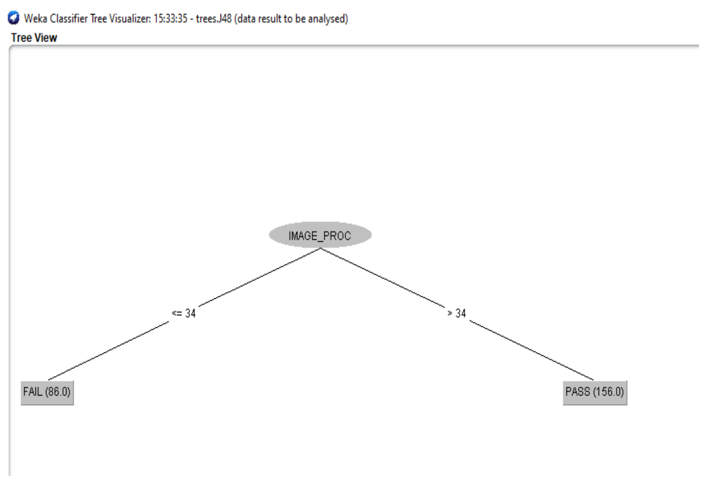

Fig. 9. Classifier generated for IMAGE_PROC grade.

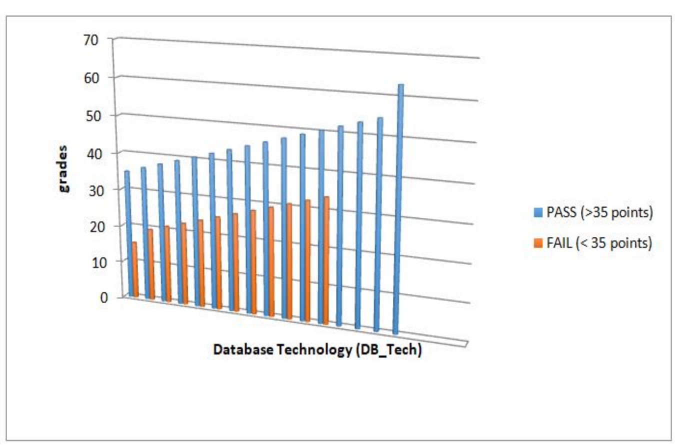

Fig. 10. Bar Chart line Plot for Database Technology grades.

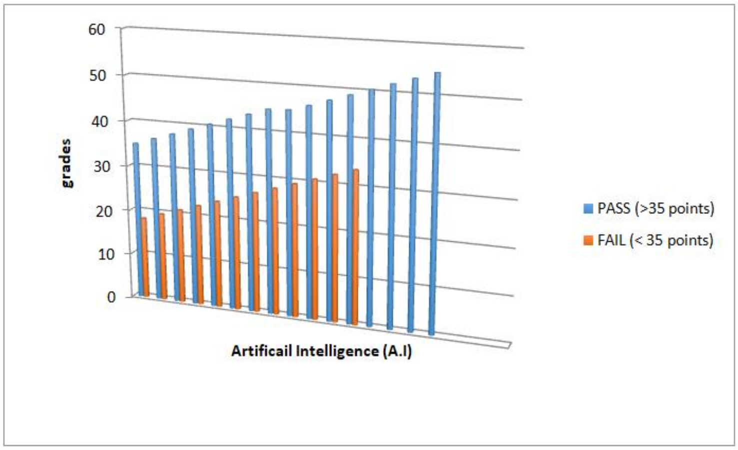

Fig. 11. Bar Chart line Plot for Artificial Intelligent grades

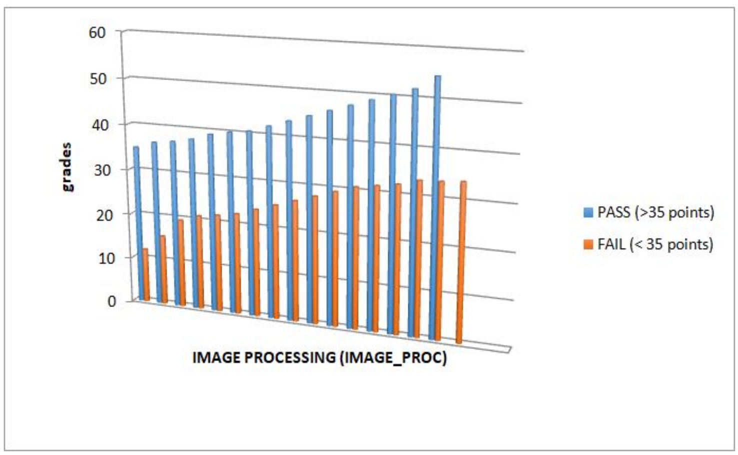

Fig. 12. Bar Chart line Plot for Image Processing grades.

7. CONCLUSIONS

The study shows that the use of machine learning and data mining techniques can significantly improve the accuracy of predictions of students' academic performance. These models contribute to the development of more effective and adaptive learning environments to the needs of each student, increasing their motivation and improving their academic performance. This conforms to recent research in education which shows that educational data mining has become an effective tool for exploring the hidden relationships in educational data and predicting students' academic achievements [76]. It is important that education institutions adopt these tools to reinforce teaching and learning in an age of Big data and artificial intelligence [77]. The process of developing and implementing these models requires a multidisciplinary approach that includes experts from the fields of education, psychology, information technology, and statistics, ensuring that the models are comprehensive, data-driven, and effectively tailored to address the diverse factors influencing student outcomes. In addition, ethical aspects related to the confidentiality and protection of student data must be considered when using these techniques. Future research prospects include the development of new methods and algorithms and their integration into educational platforms and learning management systems to improve the quality of education and support each student individually.

Funding
The author's paper explicitly states that the research was self-funded and no support was received from any institution or sponsor.

Conflicts of Interest

The paper states that there are no personal, financial, or professional conflicts of interest.

Acknowledgment

The author acknowledges the assistance and guidance received from the institution in various aspects of this study.

References

[1] Batool, J. Rashid, M. W. Nisar, J. Kim, H.-Y. Kwon, and A. Hussain, "Educational data mining to predict students' academic performance: A survey study," *Education and Information Technologies*, vol.28, pp.905–971, 2023.
[2] Shu and Y. Ye, "Knowledge Discovery: Methods from data mining and machine learning," *Social Science Research*, vol.110, pp.102817, February 2023.
[3] Huang, H. Xiong, S. Chen, Z. Lv, J. Huang, Z. Chang, and F. Catani, "Slope stability prediction based on a long short-term memory neural network: comparisons with convolutional neural networks, support vector machines and random forest models," *International Journal of Coal Science & Technology*, vol.10, no.18, pp.1-14, April 2023.
[4] Ramadhan, B. A. Kumar, I. Bala, M. M. Mijwil, M. Abotaleb, H. Alkattan, and Z. Albadran, "IoT-Integrated Multi-Sensor Plant Monitoring and Automated Tank-Based Smart Home Gardening System," In BIO Web of Conferences, vol.97, pp.1-11, April 2024.
[5] Dol and P. M. Jawandhiya, "Classification Technique and its Combination with Clustering and Association Rule Mining in Educational Data Mining — A survey," *Engineering Applications of Artificial Intelligence*, vol.122, pp.106071, 2023.
[6] Miao, Y. Lv, K. Yu, L. Liu, and J. Jiang, "Research on coal mine hidden danger analysis and risk early warning technology based on data mining in China," *Process Safety and Environmental Protection*, vol.171, pp.1-17, 2023.
[7] McIver and M. A. Friedl, "Using prior probabilities in decision-tree classification of remotely sensed data," *Remote Sensing of Environment*, vol.81, no.2, pp.253-261, 2002.
[8] Mahbobi, S. Kimiagari, and M. Vasudevan, "Credit risk classification: an integrated predictive accuracy algorithm using artificial and deep neural networks," Annals of Operations Research, vol.300, pp.609–637, 2021.
[9] Albreiki, N. Zaki, and H. Alashwal, "A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques," *Education Sciences*, vol.11, no.5, pp.552, 2021.
[10] Alsariera, Y. Baashar, G. Alkawsi, A. Mustafa, A. A. Alkahtani, and N. Ali, "Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance," *Computational Intelligence and Neuroscience*, vol.2022, no.1, pp.1-11, 2022.
[11] . Laradji, M. Alshayeb, and L. Ghouti, "Software defect prediction using ensemble learning on selected features," *Information and Software Technology*, vol.58, pp. 388-402, 2015.
[12] Lin, H. Zheng, B. Han, Y. Li, C. Han, and W. Li, "Comparative performance of eight ensemble learning approaches for the development of models of slope stability prediction," *Acta Geotechnica*, vol.17, pp.1477–1502, 2022.
[13] Azhar, S. Nadeem, F. Naz, F. Perveen and A. Sameen, "Impact of Parental Education and Socio-Economic Status on Academic Achievements of University Students," *European Journal of Psychological Research*, vol.1, no.1, pp.1-9, 2014.
[14] Gorard and B. H. See, "The impact of socio-economic status on participation and attainment in science," *Studies in Science Education*, vol.45, no.1, pp.93-129, 2009.
[15] Mijwil, G. Ali, and E. Sadıkoğlu, "The Evolving Role of Artificial Intelligence in the Future of Distance Learning: Exploring the Next Frontier," *Mesopotamian Journal of Computer Science*, vol.2023, pp:98-105, May 2023.
[16] Gourari, M. Skouri, Raoufi M., and F. Ouatik, "The Future of the Transition to E-learning and Distance Learning Using Artificial Intelligence," In 2020 Sixth International Conference on e-Learning (econf), pp.1-6, 06-07 December 2020.
[17] Fernández-Morante, B. Cebreiro-López, M. Rodríguez-Malmierca, and L. Casal-Otero, "Adaptive Learning Supported by Learning Analytics for Student Teachers' Personalized Training during in-School Practices," *Sustainability*, vol.14, no.1, pp.124, 2021.
[18] Mijwil, K. K. Hiran, R. Doshi, M. Dadhich, AH. Al-Mistarehi, and I. Bala, "ChatGPT and the Future of Academic Integrity in the Artificial Intelligence Era: A New Frontier," *Al-Salam Journal for Engineering and Technology*, vol. 2, no. 2, pp116-127, April 2023.
[19] Da Silva, L. P. S. Dias, S. Rigo, J. L. V. Barbosa, D. R. F. Leithardt, and V. R. Q. Leithardt, "A Literature Review on Intelligent Services Applied to Distance Learning," *Education Sciences*, vol.11, no.11, pp.666, 2021.
[20] Ilić, V. Mikić, L. Kopanja, and B. Vesin, "Intelligent techniques in e-learning: a literature review," *Artificial Intelligence Review*, vol.56, pp.14907–14953, 2023.
[21] Chen, D. Zou, G. Cheng, and H. Xie, "Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of Computers & Education," *Computers & Education*, vol.151, pp.103855, 2020.
[22] Breen and R. Lindsay, "Different Disciplines Require Different Motivations for Student Success," *Research in Higher Education*, vol.43, pp.693–725, 2002.
[23] Nazaretsky, M. Ariely, M. Cukurova, and G. Alexandron, "Teachers' trust in AI-powered educational technology and a professional development program to improve it," *British Journal of Educational Technology*, vol.53, no.4, pp.914-931, 2022.
[24] Muhamedyev, K. Yakunin, Y. A. Kuchin, A. Symagulov, T. Buldybayev, S. Murzakhmetov, and A. Abdurazakov, "The use of machine learning "black boxes" explanation systems to improve the quality of school education," *Cogent Engineering*, vol.7, no.1, pp.1769349, 2019.
[25] Tan and C. W. Cheah, "Developing a gamified AI-enabled online learning application to improve students' perception of university physics," *Computers and Education: Artificial Intelligence*, vol.2, pp.100032, 2021.

[26] Karpov, “Confidentiality and Protection of Student Personal Data,” Journal of Information Technologies, vol. 22, pp. 89-99, 2021.
[27] Volkov, “Development of New Data Analysis Methods in Education,” *Journal of Higher Education*, vol. 24, pp. 123-134, 2022.
[28] S. Savelev, “Integration of Analytical Solutions into Educational Platforms,” *Pedagogy and Psychology*, vol. 23, pp. 78-89, 2020.
[29] Lebedeva, “Creating Effective Educational Environments Using Data Analysis,” *Journal of Educational Research*, vol. 25, pp. 67-79, 2021.
[30] Tarasov, “Adaptive Educational Systems Based on Machine Learning,” *Journal of Pedagogy*, vol. 26, pp. 101-113, 2022.
[31] Ivanov, "Application of Machine Learning Methods in the Educational Field," *Education and Science*, vol. 23, pp. 45-56, 2020.
[32] Petrov, "Data Analysis in Education: Current Trends," *Journal of Higher Education*, vol. 17, pp. 123-134, 2021.
[33] V. Vasiliev, "Decision Trees for Predicting Student Performance," *Journal of Educational Technologies*, vol. 15, pp. 78-89, 2019.
[34] N. Nikolaev, "Random Forests and Their Application in Educational Analytics," *Information Systems in Education*, vol. 18, pp. 101-112, 2022.
[35] E. Egorov, "Data Mining Methods in Educational Systems," *Educational Informatics*, vol. 20, pp. 89-99, 2020.
[36] L. Lebedev, "Big Data Analysis in Education," *Questions of Pedagogy*, vol. 22, pp. 56-67, 2021.
[37] I. Ivanova, "Predicting Academic Performance Based on Previous Achievements," Journal of Pedagogy, vol. 19, pp. 34-45, 2020.
[38] S. Sidorov, "Application of Classification Algorithms for Student Data Analysis," *Journal of Educational Research*, vol. 16, pp. 112-123, 2022.
[39] A. Alekseev, "Machine Learning and Its Role in Educational Analytics," Science and Education, vol. 21, pp. 78-89, 2019.
[40] K. Kuznetsov, "Application of Machine Learning to Improve Prediction Accuracy in Education," *Questions of Education*, vol. 24, pp. 99-110, 2021.
[41] M. Mikhailov, "Ensemble Methods in Educational Analytics," *Journal of Information Technologies*, vol. 17, pp. 89-101, 2020.
[42] O. Orlova, "Combining Algorithms for Predicting Academic Performance," *Journal of Higher Education*, vol. 18, pp. 123-134, 2022.
[43] D. Dmitriev, "Analysis of Factors Affecting Student Performance," *Pedagogy and Psychology*, vol. 23, pp. 45-56, 2020.
[44] T. Timofeev, "Socio-Economic and Psychological Aspects of Student Performance," Journal of Education, vol. 19, pp. 78-89, 2021.
[45] I. Ilyin, "Feature Selection for Building Classification Models," Journal of Educational Technologies, vol. 16, pp. 34-45, 2020.
[46] P. Pavlov, "Data Processing and Its Impact on Prediction Accuracy," Information Systems in Education, vol. 21, pp. 101-112, 2022.
[47] V. Voronova, "Adaptive Learning Using Machine Learning Methods," Journal of Pedagogy, vol. 20, pp. 56-67, 2021.
[48] A. Anisimov, "Personalized Approaches in Learning Based on Data Analysis," Educational Informatics, vol. 19, pp. 78-89, 2020.
[49] S. Sokolova, "Recommendation Systems in Educational Analytics," Science and Education, vol. 22, pp. 112-123, 2021.
[50] E. Emelyanov, "Individual Educational Trajectories Using Machine Learning," Questions of Education, vol. 24, pp. 99-110, 2022.
[51] N. Nazarov, "Interdisciplinary Approach to Predicting Academic Performance," Journal of Higher Education, vol. 21, pp. 89-101, 2020.
[52] A. Antonov, "Collaboration of Specialists from Different Fields to Improve Education Quality," Questions of Pedagogy, vol. 18, pp. 56-67, 2021.
[53] O. Olkhovskaya, "Comprehensive Solutions for Improving Education Quality," Journal of Educational Research, vol. 20, pp. 123-134, 2022.
[54] I. Ignatov, "Integration of Data Analysis Methods into Educational Systems," Educational Informatics, vol. 23, pp. 45-56, 2020.
[55] E. Erofeeva, "Ethical Aspects of Data Usage in Education," Journal of Pedagogy, vol. 19, pp. 34-45, 2020. 2/2.
[56] M.S. Mythili and A.R. Mohamed Shanavas (2014), “An analysis of students” performance using classification algorithms”, IOSR, Journal of Computing Engineering, vol 16, Issue 1.
[57] Adelaja Oluwaseun Adebayo and Mani Shanker Chaubey (2019). Data Mining Classification Techniques on the analysis of student’s performance. Global Scientific Journals. pp 79 – 95.
[58] S.Neelamegam and E. Ramaraj, “Classification algorithm in Data mining: An Overview”, volume 3, is-sue 5, September 2013, pp. 1-5.
[59] Qasem A. Al-Radaideh and Eman Al Nagi, “Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance”. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No. 2, 2012, pp. 144-151.
[60] Lim T.S, Loh W.Y and Shih Y.S 1997, revised 1998. An empirical comparison of decision trees and other classification methods. Technical report 979, Department of Statistics, University of Wisconsin. [available from http://www.stat.wisc.edu/p/stat/ftp/pub/loh/treeprogs/quest1.7.techrep. zip].
[61] Witten I. Frank E., and Hall M. (2011). Data Mining: Practical Machine Learning Tools and Techniques 3rd Edition, Morgan Kaufmann Publishers.
[62] Geoffrey I.W, “Efficient search for association rules”. In Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 99-107, 2000.
[63] J. R. Quinlan, "C4.5: Programs for Machine Learning," Morgan Kaufmann, 1993.

[64] L. Breiman, J. Friedman, R. Olshen, and C. Stone, "Classification and Regression Trees," Wadsworth, 1984.
[65] T. Mitchell, "Machine Learning," McGraw-Hill, 1997.
[66] I. H. Witten and E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques," 2nd ed., Morgan Kaufmann, 2005.
[67] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," 2nd ed., Morgan Kaufmann, 2006.
[68] R. Kohavi and J. Quinlan, "Decision Tree Discovery," in Handbook of Data Mining and Knowledge Discovery, W. Kloesgen and J. Zytkow, Eds., Oxford University Press, 2002.
[69] Open for Innovation KNIME Software, URL: https://www.knime.com/knime-software Electric URL source: “Decision Tree learning” http: www.ke.tudarmstadt.de/lehre/archive ws0809/mldm/dt.pdf.
[70] DavinderKaur, Rajeev Bedi and S.K Gupta, “Review of decision tree data mining algorithms: ID3 and C4.5”, pp. 5-8, July 2015.
[71] Kabra R.R and Bichkar R.S, “Performance prediction of engineering students using decision tree”, International Journal of Computer Applications. Vol.36. No 11, December 2011, pp. 8-12.
[72] Jantan, H., Hamdan, A.R. and Othman, Z.A. (2010b). “Human Talent Prediction in HRM using C4.5 Classification Algorithm”, International Journal on Computer Science and Engineering, 2(08-2010), pp. 2526- 2534.
[73] Cherkauer K. J. and Shavlik J.W, “Growing Simpler Decision Trees to facilitiate Knowledge Discovery. In Proceedings of the Second International Conference of Knowledge Discovery and Data Mining (KDD-96), Portland, OR, August, 1996.
[74] L. Rokach and O. Maimon, “Decision Trees”, Data mining and Knowledge discovery handbook, 166-192.
[75] AI-Radaideh Q. A, AI-Shawakfa E. W, and AI-Najjar M. I, “Mining student data using decision trees”, International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan, 2006.
[76] Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 11.
[77] Luan, H., Geczy, P., Lai, H., Gobert, J., Yang, S. J. H., Ogata, H., Baltes, J., Guerra, R., Li, P., & Tsai, C.-C. (2020). Challenges and future directions of big data and artificial intelligence in education. *Frontiers in Psychology, 11*, 580820.