International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-14 Issue-6, March 2026

Using Data Mining to Predict Secondary School Student Performance for Zambia

Mainess Kandah Namuchile, Christopher Mulwanda

***Abstract:** Predicting student performance remains a challenge in many education systems, especially in developing countries like Zambia, where robust predictive tools are scarce. This study shows how data mining methods can be utilised to improve the accuracy of performance prediction by leveraging mock examination results to meet the needs of school management. For this study, a dataset containing 1,170 instances and 17 attributes was constructed and analysed using four classification algorithms (J48, PART, BayesNet, and Random Forest). The findings indicate that although each classifier produced results with high accuracy above 99%, Random Forest performed best, delivering perfect predictions with 100% accuracy. These results emphasise the importance of data mining in generating reliable forecasts of student performance, enabling early detection of at-risk learners and timely interventions by school managers, teachers, and parents. The study recommends adopting Random Forest as the most suitable classifier for predicting student performance. By incorporating predictive analytics into educational management, schools can strengthen decision-making, refine teaching approaches, and ultimately improve learning quality.*

***Keywords:** Data Mining: Performance Prediction: Classification Algorithms: Random Forest: Zambia Secondary School.*

Nomenclature:
DM: Data Mining

I. INTRODUCTION

Data Mining (DM) has emerged as a crucial area within modern education systems. The field focuses on developing techniques for examining the distinctive data types produced within the education system. The primary aim for using DM in an educational setting is to enable the various stakeholders to gain deeper insights into students and their learning contexts [1]. Data mining offers substantial possibilities, such as analysing how students perform [2] and forecasting outcomes in the learning system to help reduce dropout rates [3]. When stakeholders, such as teachers, school managers, learners, decision-makers, and civil society organisations, apply data mining approaches to student information. These give stakeholders a platform to extract meaningful insights from the vast datasets at their disposal.

Manuscript received on 29 December 2025 | First Revised Manuscript received on 18 February 2026 | Second Revised Manuscript received on 03 March 2026 | Manuscript Accepted on 15 March 2026 | Manuscript published on 30 March 2026.
*Correspondence Author(s)

Mainess Kandah Namuchile, Department of Computer Science, School of Engineering and Technology, Mulungushi University, Zambia. ORCID ID: 0009-0002-0091-1413

Dr. Christopher Mulwanda*, Researcher, Centre for Environmental Justice, Plot 37741, Pitta Road, off Twin-Palm Road, Ibex Hill, Lusaka, Zambia. Email ID: mulwandachristopher0@gmail.com, ORCID ID: 0000-0002-1216-2483

The DM approach helps stakeholders uncover previously hidden details about students, including factors that might influence their future academic success [4]. The promising potential of DM in educational settings sparks widespread research across different educational levels. This study focuses on using various classifiers to predict student performance [2]. The study also aims to integrate DM into educational systems, which is expected to enhance both the effectiveness and quality of educational management in schools [5].

The incorporation of DM into the educational system is believed to support both school managers and students by improving key processes, such as instruction and learning [6]. This approach strengthens the teaching and learning strategies and future planning efforts [2]. Considering that student academic success is performance-based, the use of DM will provide school managers with insights to plan for student outcomes based on current student performance [7]. On the other hand, DM allows school managers to examine past and present student data to estimate future performance [7]. The ability to predict student achievement through data mining represents a significant shift from traditional approaches and advances the utilisation of information communication skills in educational environments [8].

However, in developing nations such as Zambia, the advantages of DM have not been fully appreciated or utilised. This lack of utilisation stems primarily from limited awareness of predictive approaches currently employed to identify the most effective forecasting methods for student performance. Zambia's educational structure, like many others, positions secondary education as preparation for higher education (colleges and universities). Throughout this preparatory stage, students must complete a curriculum covering at least 8 subjects. During these stages of basic education, students undertake various assessments. These assessments provide a platform for teachers and school managers to evaluate students' comprehension and knowledge development and prepare them for advanced academic challenges and opportunities ahead. These assessments take multiple forms, including homework, tests, and assignments [5]. In secondary schools, the major evaluations (mock and final examinations) hold particular significance. They are used for demonstrating student progress and preparedness. Schools supported by the examination body (Examinations Council of Zambia) administer these annual evaluations to examination classes (grades 9 and 12). In this context, the dedication and work ethic of teachers and administrators often show in final examination results, as student achievement is considered among the most vital issues in the education system [9]. However, student performance can be affected by numerous variables, such as a lack of technology [10], especially in many Zambian schools. This, as noted, may affect early detections and the prediction of student performance from assessment information. Assessment feedback plays a crucial role in shaping learning and success [10], yet producing precise feedback remains challenging in educational environments [11].

This study holds that accurate feedback requires proper analysis, and without this foundation, prediction attempts fail and future data utilisation suffers [10]. Techniques that help teachers and school managers to predict student performance help identify struggling learners early, precisely [10], reduce dropout rates [12], and assist in recognising slow learners while evaluating the primary factors affecting their academic results [13]. The insights gained can help institutions modify teaching methods, enhance learning procedures [14], and improve decision-making and policy creation [15]. Consequently, many computer-assisted models for feedback and prediction have been developed internationally, though this development has not yet occurred in Zambian schools. Given this situation, the present study seeks to demonstrate how data mining with WEKA can analyse mock examination results and predict students' performance in final examinations. The research followed three goals: (1) to create a dataset for predicting student performance, (2) to evaluate various data mining classification methods for prediction, and (3) to suggest the most appropriate classification method for secondary school use.

II. METHODOLOGY

The research used student examination data from Chankwa Secondary School, located in Mufulira, Zambia's Copperbelt Province. This primary data included mock and final examination scores, which were used to create a dataset containing 1,170 records across 17 variables. The dataset comprises three mandatory subjects (English, Science, and Mathematics) for students who sat their national examinations between 2014 and 2016. This study purposefully sampled the examination records for the stated period because they were more complete, consistent, and reliable, providing a more valid basis for inclusion and testing formats.

To supplement the examination dataset, this study conducted an extensive literature review to triangulate the collected information. The supplementary information enabled the study to establish theoretical foundations, thereby providing a basis for understanding how researchers utilise data mining techniques in the educational environment. The combination of the actual examination records and data from the literature review provides a comprehensive framework that supports the research methodology and enhances the interpretation of the findings.

The school management at Chankwa Secondary School, who were the custodians of the examination records, were approached with an introductory letter from Copperbelt University. The school management granted access to examination records for the requested period to support the study. Confidentiality and anonymity were ideal, given the sensitivity of the records, especially since the 5-year embargo had not yet elapsed. The dataset excluded any information that could identify individual students, and all findings were presented using combined data rather than individual results. During the data (record) collection process, the researchers and research assistants obtained examination records by following proper institutional Channels and, most importantly, adhered to strict institutional ethical guidelines. The use of examination records can raise serious concerns due to the strict examination guidelines enforced by the Examination of Zambia. Following the laid-down guidelines ensured credibility and protected the school management and students' privacy rights.

A. Reviews of Related Literature

Data mining is the practice of examining data from multiple angles and transforming it into meaningful insights to reveal underlying relationships and patterns embedded in large datasets [16]. Its usefulness has been demonstrated across many domains, including fraud detection, targeted advertising, marketing analytics, credit and loan evaluation, and predictive modelling [17]. Within education, where digital records and learner-related data are growing rapidly, there is an increasing demand for approaches that can systematically analyse large volumes of information. In response to this need, EDM has gained prominence as a method for analysing learning data and forecasting student performance. Because education administrators typically aim to improve academic outcomes, continuous monitoring of learner performance is essential. By leveraging data mining, education stakeholders can make better-informed decisions through early identification of learning difficulties and timely academic support interventions [18].

Evidence of the value of data mining in educational management is well documented in the literature. [19], compared several algorithms, J48, Naïve Bayes, Bayes Net, Multilayer Perceptron, SVM, REPTree, and Random Forest, using two datasets designed for predicting student performance. They reported that J48 delivered notably strong classification performance in their evaluations. Other studies have equally examined determinants of student achievement by analysing both environmental and educational factors using data mining methods [20]. Their work applied C4.5 and ID3 decision tree techniques, combined with a ranker-based feature selection approach, to improve the prediction of student outcomes. Likewise, [21] analysed engineering student records using C4.5 and CART decision tree models to predict whether learners would pass, fail, or progress. Collectively, the studies and literature show that data mining approaches and techniques could improve the accuracy of student performance prediction and provide actionable evidence to support educational decision-making in schools.

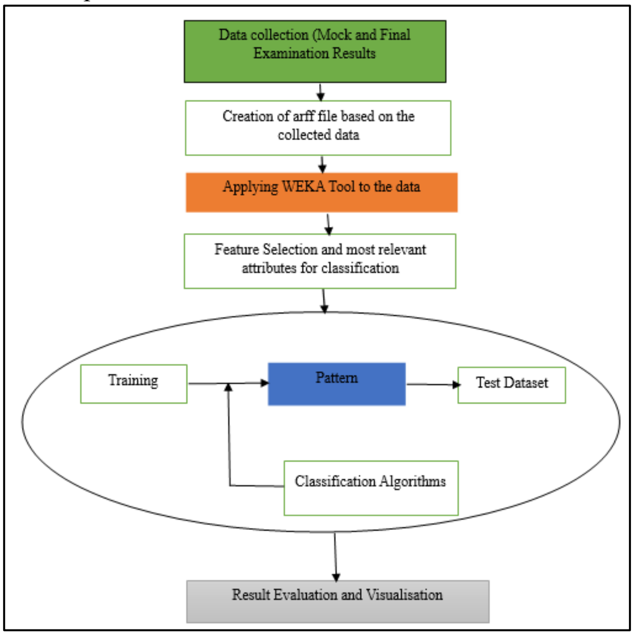
B. Feature Selection

In data mining, selecting the correct feature is a vital process. This is the process leading to data analysis and the identification of important and valuable attributes for inclusion. This process enables the study to eliminate attributes that are considered irrelevant, redundant, or contain noise in the dataset. Reducing the dimensions or attributes promotes computational efficiency, improves model generalisation, and minimises the risk of overfitting. Correlation-based methods of attribute evaluation are commonly employed to assess the contribution of different variables to improved predictions and their relationships with other variables, ensuring that only the most informative features are used in the analysis [22].

A five-structure approach was utilised to develop the dataset. These are (i) training, (ii) pattern identification, (iii) testing, (iv) result evaluation, and (v) knowledge visualisation. Having completed these steps, the study then applied several machine learning algorithms in WEKA to measure predictive accuracy. The machine learning algorithms were J48, PART, BayesNet, and Random Forest [23]. The combination of these machine learning algorithms and classification methods enhanced the quality of the input data. It boosted classification results, highlighting feature selection as an important component of successful data mining.

C. Dataset Model Processing

The study shows the process and the approach for forecasting student performance. The data obtained from Chankwa Secondary School (examination records) were considered unprocessed and used in this experiment, as presented in the methodology section. To generate the final dataset, all key steps were performed, including data preparation and data mining. Initially, the original records were prepared and saved as CSV files. The CSV files were then transformed to ARFF format and later imported into the modelling tool for analysis. Figure 1 shows the summarised and complete workflow.



[Fig.1: Dataset Processing Model]

D. Description of Algorithms

The WEKA's J48 algorithm was used to implement the C4.5 decision tree approach. This offered flexibility by generating both pruned and unpruned tree versions. This created a balance between the model's accuracy in predicting outcomes and the ease with which users can interpret its structure. On the other hand, PART takes a different strategy. The study used a divide-and-conquer approach to build partial decision trees multiple times, converting the most effective leaf nodes into IF-THEN rules. The process creates a decision list that merges concepts from decision tree creation and rule-based machine learning [24].

On the other hand, the BayesNet builds Bayesian network models by applying several search methods and evaluation criteria. This allows the algorithm to capture probabilistic connections and both conditional dependence and independence relationships between variables [25]. Random Forest is an ensemble method that generates numerous unpruned decision trees by bootstrapping the training data and randomly selecting features at each tree's construction. The method combines predictions from these varied trees to generally produce better accuracy and consistency while reducing the risk of overfitting [3].

III. RESULTS

Table I: Dataset Creation and Description

Attribute	Description	Values
GE	Gender	M/F
GR	Grade	9/12
AG	Age	15/18
SSA	Students' School Attends	Private/Public
SSL	Students' School location	Urban/Peri-urban/Rural
SSMT	Students' School Mode of Transport	Bus/Foot/Bicycle
PVN	Province	Copperbelt
YR	Year	2014/2015/2016
MKSGRADE	Marks Senior Grade	1-75% to 100%, 2- 70% to 74%, 3- 65% to 69%, 4- 60% to 64%, 5- 55% to 59%, 6- 50% to 54%, 7-40% to 49%, 8- 35% to 39%, 9- <35%
MKJRGRADE	Marks Junior Grade	1-75% to 100%, 2-60% to 74%, 3-50% to 59%, 4- 40% to 49%, F- <40
TOS	Type of School	Co-Ed, Boys/Girls
SJ	Subjects	English, Maths, Science
RT	Calculated from marks	Ass, Fall
SCD School Code 5054	SCD School Code 5054	SCD School Code 5054

A. GE: Gender is used to classify the students as either male or female. This factor is quite important because it helps determine the effect of students' gender on prediction.
B. GR: Grade is used to identify the student as either nine (9) or twelve (12).
C. SSMT: Mode of transportation to school. It determines the student's mode of transportation. The possible values are: by foot, by bicycle, and by bus.
D. MKSRGRADE: Marks/Grade obtained at senior secondary level, and it is declared as the response variable. It is also split into nine class values: 1 – 75% to 100%, 2 – 70% - 74%, 3 – 65% - 69%, 4– 60% - 64%, 5 – 55% - 59%, 6 – 50% - 54%, 7 – 40% - 49%, 8 – 35% - 39%, 9- < 35%.
E. MKJRGRADE: Marks/Grade obtained at junior secondary level, and it is declared as the response variable. It is also split into nine class values: 1 – 75% to 100%, 2 – 60% to 74%, 3 – 50% to 59%, 4 – 40% to 49%, F- < 40%.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

F. TOS: Type of school. This determines the type of secondary school the student attended. It includes the possible values, co-education, boys, and girls.

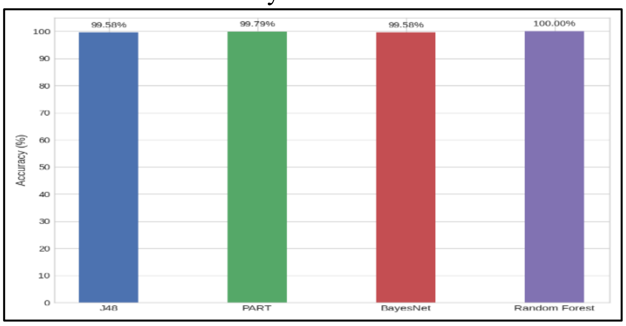
Table 1 presents the attributes and descriptions of the dataset used to predict student performance. The dataset from examination records contained 1,170 instances with 17 attributes. The attributes selected for this study were academic, demographic, and institutional factors. Among the key variables in the dataset were Type of School and location (SSA, SSL, TOS), grade level (GR), gender (GE), age (AG), and mode of transport (SSMT). Academic achievement was captured using junior and senior secondary marks (MKJRGRADE, MKSGRADE), which served as the response variables. Contextual fields, such as year (YR) and province (PVN), were retained to capture regional variation and anchor the records to a specific time frame.

For data modelling, the grading variables were encoded into predefined bands, such as MKSGRADE: 1 = 75–100%, 9 = <35%. This approach enabled classification algorithms to learn from performance categories rather than raw scores, making the dataset suitable for machine learning, especially classification-based analysis.

Table II: Comparing the Performance of the Four Classifiers

Model	Accuracy		
	Dataset		
	Correctly Instances	**Classified**	**Incorrectly Classified Instances**
J48	99.5763%		0.4237%
PART	99.788%		0.2118%
Bayes Net	99.5763%		0.4237%
Random Forest	100%		0%

To evaluate the dataset's utility, we tested it against four classification algorithms: J48, PART, BayesNet, and Random Forest. Table 2 presents the results, which were notably strong across the board. J48 and BayesNet both reached 99.58% accuracy, while PART edged slightly higher at 99.79%. Random Forest stood out entirely, classifying every instance correctly at 100%. These results suggest that the dataset is well-structured and sufficiently rich in relevant features for the algorithms to distinguish between student performance levels reliably.



[Fig.2: Comparison of Classifiers]

Figure 2 shows a comparison of the predictive accuracy obtained from the four applied classifiers: J48, PART, BayesNet, and Random Forest. Having applied these classifiers to the dataset, the results show consistently excellent performance across all four models, with accuracies above 99%. Specifically, both J48 and BayesNet achieved 99.58%, while PART produced a slightly higher value of 99.79%. The Random Forest achieved the strongest performance, reaching 100% accuracy and correctly classifying all instances. The findings of this study indicate that the dataset was accurately prepared and appropriate for predictive modelling, allowing each algorithm to classify student performance reliably.

IV. DISCUSSION

The results demonstrate that a comprehensive dataset was assembled by merging student background information, including gender, age, school type, and transportation methods, with academic performance data from both practice and final examinations. This integrated approach provides a holistic perspective on the factors influencing student performance in the selected subjects, highlighting the multifaceted nature of academic achievement (Table 1). The analysis focused on three core subjects required in Zambian schools: Mathematics, Science, and English. These subjects are essential for determining students' eligibility for advancement to higher education institutions, such as colleges and universities. The dataset's validity and reliability were enhanced by maintaining complete records over three years (2014-2016), thereby minimising data gaps that could affect the modelling process. Furthermore, categorical performance levels were used instead of exact numerical scores, enabling machine learning models to identify significant achievement patterns and improving the interpretability of results for teachers, educators, and school management. This approach is consistent with other educational prediction studies that categorise grades to facilitate classification tasks [26].

The analysis indicates that the selected algorithms achieved high accuracy in meeting the prediction objectives. Tree-based and rule-based methods, specifically J48 and PART, as well as the probability-based BayesNet model, all produced highly reliable predictions, with accuracies exceeding 99% (Table 2 and Figure 2). These results are consistent with previous educational data mining research, which has demonstrated strong predictive capabilities using similar classifiers and methodologies [27]. Notably, the Random Forest classifier achieved perfect classification with no errors. This exceptional accuracy can be attributed to the ensemble method's ability to generate multiple decision trees from random samples of data and feature sets, thereby increasing stability, reducing variance, and limiting overfitting, particularly when multiple factors influence student outcomes.

As indicated, the results of this study are significant for school managers and other stakeholders, including education officials, civil society organisations, and donors in the sector. This is because the results suggest a realistic approach to transitioning from merely reporting past results to actively supporting students in schools. When schools have the localised capacity to predict final exam performance from mock exam results with nearly perfect accuracy, they can spot struggling students early and offer remedial support. This enables school managers to focus interventions more

DOI: 10.35940/ijrte.E8329.14060326
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

effectively, improve individual student performance and allocate teaching and counselling resources where they will have the greatest impact before final exams commence. Similarly, other studies on performance prediction confirm that these models generally provide value not only through accuracy, but also by enabling early interventions that help students stay in school and complete their courses [28]. In this study, the results and available literature indicate that incorporating Random Forest-based predictive tools into school performance tracking systems in learning settings can improve teaching methods and decision-making, and provide focused feedback, ultimately leading to better learning outcomes across the education system.

A. Recommending the Optimal Classifiers

The comparison among the four classifiers indicates that all models performed well. The accuracy of all classifiers was above 99% (Figure 2). This also demonstrates that the dataset used in this study was well-organised and suitable for predictive modelling, enabling each algorithm to classify student performance with high reliability. However, the results show that the Random Forest classifier is best suited to achieve perfect accuracy, with zero misclassified cases. The Random Forest classifier is an ensemble learning mechanism that builds many decision trees and aggregates their predictions, enabling it to capture complex patterns while improving generalisation and limiting overfitting. Most importantly, even when other models were tuned by adjusting parameters, none achieved results comparable to Random Forest. Therefore, this study aligns with prior studies reporting Random Forest accuracy above 99% and suitability for educational performance prediction tasks in learning institutions [3].

This study recommends Random Forest as being the most suitable classifier for predicting student performance in secondary schools. The value of using Random Forest is not limited to learning institutions or to predicting student results, but extends to multiple stakeholders. For learning institutions such as schools and management, Random Forest offers a more reliable platform for tracking learner progress across subjects and identifying students at risk of underperforming early. This provides a more flexible guide for institutional planning, management, and more efficient use of resources. For individual teachers, the predictive outputs can help strengthen instructional decisions in everyday academic life by pinpointing learners who need additional support. This will also enable the teachers to provide the needed interventions to learners who are underperforming promptly. For students in schools or learning facilities, accurate result predictions will increase awareness of their academic position, further motivate them early on, and encourage proactive engagement with learning tasks. And for parents, predictive insights will enable them to monitor their children's progress and provide timely counselling, encouragement, and academic support at home [9]. In addition, policymakers can utilise data mining methods, such as Random Forest, in education systems to improve evidence-based decision-making, support quality assurance efforts, and help lower dropout rates across the country's schools.

V. CONCLUSION

This study demonstrates that applying data mining techniques can enhance the prediction of student performance in secondary schools and other educational institutions. A comprehensive dataset from Chankwa Secondary School in the Mufulira district of Zambia, focusing on socio-demographic attributes, was utilised. By integrating mock examination indicators in Mathematics, English, and Science, the study established a robust foundation for evaluating several classification techniques, including J48, PART, BayesNet, and Random Forest. The results indicate that decision tree–based models and probabilistic approaches achieved high accuracy rates (above 99%). Notably, Random Forest was identified as the most effective classifier for predicting student performance. This exceptional predictive performance underscores the advantages of ensemble learning, which improves generalisation and reduces the risk of overfitting.

These findings hold significant value for a range of educational stakeholders. For school managers and policymakers, the predictive models offer a means to identify at-risk learners early and facilitate the timely implementation of targeted interventions. For subject teachers, performance prediction can inform more responsive instructional planning and enable personalised support where it is most needed. Additionally, parents can benefit from clearer performance forecasts, improved monitoring of student progress, and the ability to provide earlier support at home. Although the study was conducted at a single school in the Mufulira district, the proposed framework is generalisable and can be adapted to other schools, districts, and subject areas nationwide.

VI. ACKNOWLEDGEMENT

The authors gratefully acknowledge the Administrators at Chankwa Secondary School for their invaluable support in providing access to the data used in this study. Their cooperation and commitment made it possible to create the dataset that underpins the research findings presented in this paper.

DECLARATION STATEMENT

Some of the cited references are older and are explicitly noted as [16]. However, these works remain significant for the current study, as they are pioneering in their fields.

As the article's author, I must verify the accuracy of the following information after aggregating input from all authors.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been funded by any organizations or agencies. This independence ensures that the research is conducted objectively and without external influence.
- **Ethical Approval and Consent to Participate:** The content of this article does not necessitate ethical approval or consent to participate with supporting documentation.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Author's Contributions:** Each author has individually contributed to the article. Mainess Kandah Namuchile; Conceptualised, Prepared Methodology, Data Collection, and wrote the original manuscript draft. While Dr. Christopher Mulwanda supervised, edited, and reviewed the Manuscript.

REFERENCES

1. Fuseini, I. & Missah, Y. M. 2024. A critical review of data mining in education on the levels and aspects of education. Quality Education for All, 1(2), 41–59. DOI: https://doi.org/10.1108/QEA-01-2024-0006.
2. Ahmed, A. M., Rizaner, A. & Ulusoy, A. H., 2016. Using Data Mining to Predict Instructor Performance. Procedia Computer Science, Volume 102, p. 137–142. DOI: https://doi.org/10.1016/j.procs.2016.09.380
3. Hussain, S., Abdulaziz Dahan, N., Ba-Alwi, F. M. & Ribata, N., 2018. Educational Data Mining and Analysis of Students' Academic Performance Using WEKA. Indonesian Journal of Electrical Engineering and Computer Science, 9(2), pp. 447-459. DOI: https://doi.org/10.11591/ijeecs.v9.i2.pp447-459
4. Amra, I. A. A. & Maghari, A. Y. A., 2017. Students' performance prediction using KNN and Naïve Bayesian. 2017 8th International Conference on Information Technology (ICIT), p. 909–913. DOI: https://doi.org/10.1109/ICITECH.2017.8079967
5. Kumar, M. & Singh, A. J., 2017. Evaluation of Data Mining Techniques for Predicting Student's Performance Predicting Student's Performance. I.J. Modern Education and Computer Science, Volume 8, pp. 25-31. DOI: https://doi.org/10.5815/ijmecs.2017.08.03
6. Bonde, S. N. & Kirange, D. K., 2018. Survey on Evaluation of Students' Performance in Educational Data Mining. 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), p. 209–213. DOI: https://doi.org/10.1109/ICICCT.2018.8473228
7. Mohammad, A., 2018. Measuring Academic Performance of Students in Higher Education Using Data Mining Techniques. s.l.:Loughborough University. URL: https://repository.lboro.ac.uk/articles/thesis/Measuring_academic_performance_of_students_in_higher_education_using_data_mining_techniques/9470215.
8. Yassein, N. A. M., Helali, R. G. & Mohomad, S. B., 2017. Predicting Student Academic Performance in KSA using Data Mining Techniques. Journal of Information Technology & Software Engineering, 7(5). DOI: https://doi.org/10.4172/2165-7866.1000220
9. Soni, A., Kumar, V., Kaur, R. & Hemavathi, D., 2018. Predicting Student Performance using Data Mining Techniques. International Journal of Pure and Applied Mathematics, 119(12), pp. 221-227. URL: https://acadpubl.eu/jsi/2018-119-12/articles/12a/221.pdf.
10. Kaur, G. & Singh, W., 2016. Prediction Of Student Performance Using the Weka Tool. An International Journal of Engineering Sciences, Volume 17, pp. 8-16. URL: http://ijoes.vidyapublications.com/paper/Vol17/02-Vol17.pdf.
11. Mamoon-Al-Bashir, A., Rezaul Kabir, A. & Rahman, I., 2016. The Value and Effectiveness of Feedback in Improving Students' Learning and Professionalising Teaching in Higher Education. Journal of Education and Practice, 7(16), pp. 38-41. URL: https://files.eric.ed.gov/fulltext/EJ1105282.pdf.
12. Ashraf, A., Anwer, S. & Khan, M. G., 2018. A Comparative Study of Predicting Students' Performance by Use of Data Mining Techniques. American Scientific Research Journal for Engineering, Technology, and Sciences, 44(1), pp. 122-136. DOI: https://doi.org/10.5281/zenodo.1406000
13. Adefemi, K. O., Mutanga, M. B. & Jugoo, V. 2025. Hybrid Deep Learning Models for Predicting Student Academic Performance. Mathematics and Computational Applications, 30(3), 59. DOI: https://doi.org/10.3390/mca30030059.
14. Amrieh, E. A., Hamtini, T. & Aljarah, I., 2016. Mining Educational Data to Predict Students' Academic Performance using Ensemble Methods. International Journal of Database Theory and Application, 9(8), pp. 119-136. DOI: https://doi.org/10.14257/ijdta.2016.9.8.13.
15. Manazir, S. H. 2023. Reimagining public policy formulation and analysis: A comprehensive theoretical framework. Discover Global Society, 1(16). DOI: https://doi.org/10.1007/s44282-023-00018-4.
16. Kumari, P. M., Nabi, A. & Priyanka, P., 2014. Educational Data Mining and Its Role in the Educational Field. International Journal of Computer Science and Information Technologies, 5(2), pp. 2458-2461. URL: http://www.ijcsit.com/docs/Volume%205/vol5issue2/ijcsit20140502100.pdf. The works remain significant, see the declaration
17. Kaur, H. & Bathla, E. G., 2018. Student Performance Prediction Using Educational Data Mining Techniques. International Journal on Future Revolution in Computer Science & Communication Engineering, 4(12), p. 93 – 97. URL: https://www.ijfrcsce.org/index.php/ijfrcsce/article/view/1819.
18. Asif, R., Hina, S. & Haque, S. I., 2017. Predicting Student Academic Performance using Data Mining Methods. International Journal of Computer Science and Network Security, 17(5), pp. 187-191. URL: http://paper.ijcsns.org/07_book/201705/20170527.pdf.
19. Anoopkumar, A. M. J. & Rahman, Z., 2018. Model of Tuned J48 Classification and Analysis of Performance Prediction in Educational Data Mining. International Journal of Applied Engineering Research, 13(20), pp. 14717-14727. URL: http://www.ripublication.com/ijaer18/ijaerv13n20_82.pdf.
20. Kumar, A. D. & Radhika, V., 2016. Mining Educational Data to Predict Higher Secondary Students' Performance. International Journal of Computational Intelligence and Informatics, 6(2), pp. 124-129. URL: http://ijcii.org/administrator/components/com_jresearch/files/publications/IJCII-2016-6-2-124.pdf.
21. Sarbaini, S., & Ulfa, F. 2024. Student Graduation Prediction Using the Decision Tree Method with C4.5 Algorithm. Jurnal Diferensial, 6(1), 9–15. DOI: https://doi.org/10.35508/jd.v6i1.12287.
22. Mehboob, B., Liaqat, R. M. & Abbas, N., 2017. Student Performance Prediction and Risk Analysis by Using a Data Mining Approach. Journal of Intelligent Computing, 8(2), pp. 49-57. URL: http://jic.org.pk/issues/2017/Mehboob.pdf.
23. Sana, Siddiqui, I. F. & Arain, Q. A., 2019. Analysing students' academic performance through education data mining. 3C Tecnología. Glosas sobre la innovación aplicada a la pyme. Edición especial, pp. 402-421. DOI: https://doi.org/10.17993/3ctecno.2019
24. Khasanah, A. & Harwati, U., 2017. A Comparative Study to Predict Students' Performance Using Educational Data Mining Techniques. IOP Conference Series: Materials Science and Engineering, 215, 012036. DOI; https://doi.org/10.1088/1757-899X/215/1/012036.
25. Mat, A. et al., 2019. A Review of Predictive Modelling Technique for Student Academic Performance Monitoring. MATEC Web of Conferences, EAAI Conference 2018. DOI: https://doi.org/10.1051/matecconf/201922504012
26. Yağcı, M. 2022.Educational data mining: Prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9, 11. DOI: https://doi.org/10.1186/s40561-022-00192-z.
27. Meghji, A. F., Mahoto, N. A., Unar, M. A., & Shaikh, M. A. 2019. Predicting student academic performance using data generated in higher educational institutes. 3C Tecnología. Glosas de innovación aplicada, 366–383. DOI: https://doi.org/10.17993/3ctecno.2019.specialissue2.366-383.
28. Nachouki, M., Mohamed, E. A., Mehdi, R., & Abou Naaj, M. 2023. Student course grade prediction using the random forest algorithm: Analysis of predictors' importance. Trends in Neuroscience and Education, 33, 100214. DOI: https://doi.org/10.1016/j.tine.2023.100214

AUTHOR'S PROFILE

Mainess Kandah Namuchile holds a master's degree in computer science. She is a Computer Science lecturer in the Department of Computer Science at Mulungushi University, School of Engineering

and Technology. Her research interests include data mining, cybersecurity, Artificial intelligence, and machine learning. She is currently pursuing her PhD, with a focus on advancing the application of machine learning techniques to climate change, predictive analytics and classification problems. Her work aims to develop practical, scalable AI-driven solutions to improve decision-making and service delivery.

Dr. Christopher Mulwanda is a researcher and academic with interests in ecology and natural resources management. He holds a PhD and MSc in Natural Resources Management, and a BA-Ed in Geography. He serves as Head of Research and Studies at the Centre for Environment Justice and lectures at Mulungushi University. His work spans supervising computer science research in areas intersecting with environmental data analysis and sustainable resource management. He also contributes to international scholarship as a reviewer for academic journals in biodiversity, conservation, climate change, agriculture, and sustainability.