IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

UTILIZING DATA MINING TECHNIQUES TO IMPROVE THE QUALITY OF EDUCATION IN MALAWIAN UNIVERSITIES

[1]Crusade Chiwalo [2]Dr. Glorindal ,
[1]Asscoiate DMI St John the Baptist University , [2]Proffesor DMI St EuguenE University Zambia,
[1]Name of Department of 1st Author,
[1]DMI St John the Baptist University Mangochist Author, Mangochi, Malawi

Abstract: By offering insightful information about student performance, instructional efficacy, and resource allocation, data mining tools have the potential to dramatically raise the standard of education in Malawian universities. Malawian universities may develop a more personalized, productive, and data-driven learning environment for all students by properly applying data mining. By offering insightful information about student performance, instructional efficacy, and resource allocation, data mining tools have the potential to dramatically raise the standard of education in Malawian universities. Malawian universities may develop a more personalized, productive, and data-driven learning environment for all students by properly applying data mining. A data mining framework for improving the quality of education in Malawian universities is proposed, consisting of data collection, data preprocessing, exploratory data analysis, feature engineering, model selection, model training, model evaluation, and deployment. Data that can be used for data mining in Malawian universities includes student demographic data, academic data, learning behavior data, assessment data, instructor data, and institutional data. Examples of data mining applications in Malawian universities include identifying at-risk students, personalizing learning experiences, predicting student performance, evaluating teaching effectiveness, and improving resource allocation. Data mining techniques have the potential to make a significant contribution to the improvement of the quality of education in Malawian universities. By effectively utilizing data mining, Malawian universities can create a more personalized, effective, and data-driven learning environment for all students.

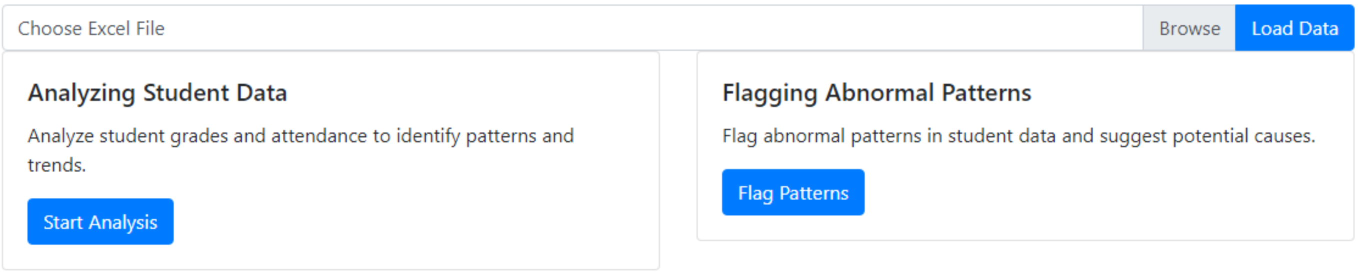
Keywords: Data mining, student performance, instructional efficacy, resource allocation, Malawian universities, personalised learning, productive learning environment, data-driven, at-risk students, quality of education

I. INTRODUCTION

In today's rapidly evolving educational landscape, the role of data in informing decision-making processes has become increasingly vital. This project report aims to establish a framework for utilizing data mining and analysis techniques to uncover actionable insights and support decision-making within Malawian universities. The objectives of the project include improving the quality of education, identifying patterns in student performance, and providing valuable recommendations to academic administrators. Ultimately, the project seeks to harness the power of data to drive informed decision-making and enhance the educational experience for future generations of students.

1.1 Project Overview

This project harnesses an Excel dataset containing comprehensive student information, encompassing grades, attendance records, and other pertinent academic data. Leveraging this dataset as the foundational source, the project employs a systematic approach to load, process, and visualize the data within the system. Through a series of data pre-processing steps, including cleaning, normalization, and feature extraction, the raw data is refined into a structured format suitable for analysis. Subsequently, advanced data mining and analysis techniques are applied to uncover hidden patterns, correlations, and trends embedded within the dataset.



Analyzing Student Data

Analyze student grades and attendance to identify patterns and trends.

Flagging Abnormal Patterns

Flag abnormal patterns in student data and suggest potential causes.

Figure 1.1 Project Overview

Figure 1.1 takes Excel data and provides additional options like data analysis serves as a critical component of the student monitoring system. This page allows administrators, educators, or advisors to upload Excel files containing student attendance and grades data directly into the system for analysis. Upon uploading the data, the system processes it and offers various options for further analysis and action.

1.2 System Workflow

The system workflow commences with the ingestion of the Excel dataset, where student data is loaded into the system's database. Following this initial step, the data undergoes a series of pre-processing operations to ensure its quality and integrity. Once the data is prepared, it is subjected to various data mining algorithms and statistical models to extract meaningful insights and generate predictive analytics. Through interactive visualization tools and dashboards, class advisors can gain actionable insights into student performance, attendance trends, and academic outcomes. Moreover, the system facilitates the generation of predictive models, enabling class advisors to anticipate future trends and make informed decisions to enhance the educational experience within Malawian universities.

1.3 System Architecture

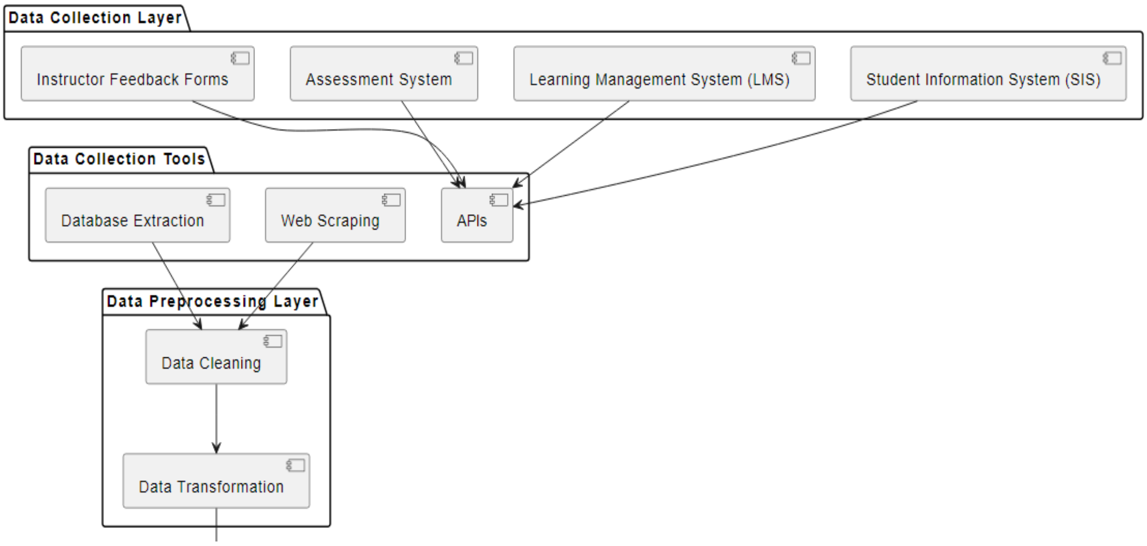


Figure 1.2 System Architecture

The architectural diagram acts as the skeleton of the whole project as it clearly states all the functions to be happening in the project in a top down manner. It shows the users of the system, the things they are to be doing in the system and how the system is going to function in general as all the processes are outlined in the architectural diagram.

1.4 Use Case Diagram

The outlines the interactions between the Advisor (user) and the Student Monitoring System. It identifies the main functionalities (use cases) of the system, including analyzing student grades and attendance, flagging abnormal patterns, suggesting causes, and accessing student data in Excel format.

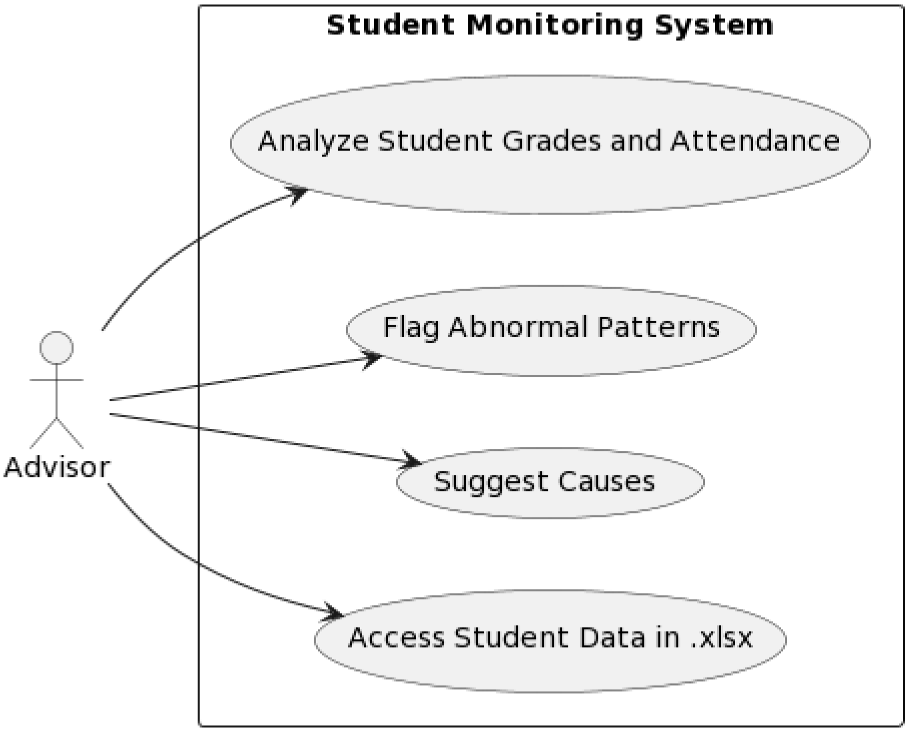

Student Monitoring System
Analyze Student Grades and Attendance
Flag Abnormal Patterns
Suggest Causes
Access Student Data in .xlsx
Advisor

Figure 1.3 Use Case Diagram

1.5 Class Diagram

The class diagram illustrates the structure of the Student Monitoring System by representing the classes (components) involved in the system, their attributes, and relationships between them. It provides an overview of the system's architecture and key components such as the Data Mining Framework, Malawian Universities, Data Mining Applications, and the Student Monitoring System itself.

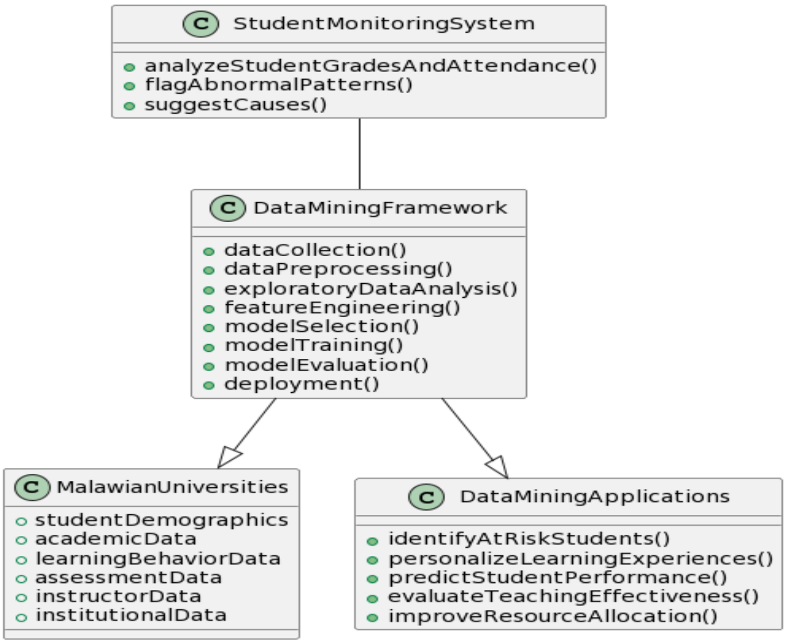


Figure 1.4 Class Diagram

II. NEED OF THE STUDY

The need for this study arises from the growing importance of data-driven decision-making in the educational sector, particularly within Malawian universities. With the increasing volume and complexity of data generated in academic institutions, there is a pressing need to leverage advanced data mining and analysis techniques to extract actionable insights that can inform strategic planning, resource allocation, and student support initiatives. By conducting this study, the project aims to address the existing gap in utilizing data effectively to enhance the quality of education, improve student outcomes, and foster a conducive learning environment within Malawian universities. This study is essential for empowering academic administrators and educators with the tools and knowledge needed to make informed, data-driven decisions that positively impact the educational experience of students and contribute to the overall advancement of the higher education sector in Malawi.

III. RESEARCH METHODOLOGY

This research will adopt a mixed-methods approach, combining quantitative and qualitative data collection and analysis techniques. This approach will allow for a comprehensive understanding of the potential applications of data mining in Malawian universities, considering both the technical aspects of data mining algorithms and the real-world challenges and opportunities within the Malawian educational context. The details are as follows:

3.1Population and Sample

The population for this study comprises students pursuing a Bachelor of Engineering in Computer Science at DMI University, with a total of 20 students enrolled in the program. A sample size of 100% of the population (20 students) will be selected for data analysis and model development.

3.2 Data and Sources of Data

Data for this study will be sourced from academic records, including student grades and attendance records, obtained from the university's database. Additional demographic information such as age, gender, and academic background will be collected through student surveys to enrich the dataset.

3.3 Theoretical framework

This research utilizes a combination of quantitative and qqualitative data collection techniques to analyse the merits, limitations as well as challenges associated with data mining application in Malawian universities settings. This approach intends to weave together the elaborate details of data-mining algorithms with those practical considerations and opportunities embedded in the formidable Malawian educational dynamics as a way of elucidating comprehensive understandings about transformative effects brought by decision making driven from informational perspectives.

The quantitative data collection and analysis is the bedrock of this study, encompassing a wide range metrics based on students performance sourced from various systems such as student information sysems, learning management systems to instructor feedback mechanisms. By means of careful data preprocessing, exploratory analysis and feature engineering algorithms distilled from the raw unstructured original is presented in a structured form which can be utilized for applying machine learning algorithm and thereby extract information ideally. There will be model selection and training to emerge key insights by removing all forms of noise from massive data sets, thereby highlighting patterns trends that predictive relationships important for decision-making.

A qualitative extension of this research, Python data sets and Data mining libraries are used to help the researchers through academic complexities in analysis. Python data sets include Student performance, Learning behavior, Instructor Feedback and Institutional enrolment parameters constitute an extraordinary mosaic of outcomes to probe into. With powerful data-mining libraries like Scikit-learn and Pandas, coupled with visualization tools such as Matplotlib, this research draws upon the latest computer science methods to reveal hidden meaning hiding in the data. The Random Forest algorithm, a compelling method for student risk identification due to its aggregation of decision trees, unveils crucial and elusive headlines facts in at-risk students' profiles contributing significantly to the formation of effective targeted intervention programs. The intention of this methodological merger is therefore to embolden Malawian universities with the knowledge in building student success and promoting institutional quality.

3.4 Quantitative Data Collection and Analysis

1. **Data Collection**

Gather relevant student data from various sources, including student information systems, learning management systems, assessment systems, and instructor feedback forms. This data will encompass a wide range of information, including:

Student demographic data: Age, gender, ethnicity, socioeconomic status, educational backgroundAcademic data: Grades, attendance records, course selection, standardized test scoresLearning behavior data: Online activity logs, engagement metrics, time spent on tasksAssessment data: Quiz results, exam scores, project evaluationsInstructor data: Teaching experience, qualifications, student feedbackInstitutional data: Course enrollment, faculty workload, financial resources

2. Data Preprocessing

Clean and prepare the data for analysis, addressing missing values, identifying and correcting outliers, and resolving inconsistencies in data formats. This will ensure the quality and reliability of subsequent data mining tasks.

3. Exploratory Data Analysis (EDA)

Utilize EDA techniques to summarize, visualize, and explore the collected data, identifying patterns, trends, and relationships within the data. EDA will provide valuable insights into the data's underlying structure and guide subsequent feature engineering and model selection processes.

4. Feature Engineering

Extract and construct relevant features from the raw data, transforming it into a form suitable for data mining algorithms. This may involve creating new features, combining existing features, or deriving meaningful information from text-based data. Feature engineering aims to extract the most predictive and informative features from the data, enhancing the effectiveness of data mining models.

5. Model Selection and Training

Select appropriate data mining algorithms for specific tasks, such as classification, prediction, clustering, and recommendation systems. Train the selected models using the prepared data, optimizing their performance through parameter tuning and evaluation.

6. Model Evaluation

Evaluate the performance of the trained models using appropriate metrics, such as accuracy, precision, recall, and F1-score. This will assess the generalizability and effectiveness of the models in predicting or classifying outcomes based on the data.

4.3.2 Python Datasets

The following Python datasets will be used for the research process:

Student Academic Performance Dataset: This dataset contains information on student demographics, academic records, and assessment results.

Learning Behavior Dataset: This dataset contains data on student engagement, online activity logs, and time spent on tasks.

Instructor Feedback Dataset: This dataset contains instructor ratings, student feedback, and teaching experience data.

Institutional Enrollment Dataset: This dataset contains course enrollment data, faculty workload metrics, and financial resource allocation information.

3.5 Data Mining Libraries

The following Python data mining libraries will be used for the research process:

Scikit-learn: A comprehensive machine learning library that provides a wide range of data mining algorithms, including classification, regression, clustering, and dimensionality reduction techniques.

Pandas: A powerful data analysis and manipulation library that allows for efficient data cleaning, preprocessing, and exploration.

Matplotlib and Seaborn: Data visualization libraries that provide tools for creating informative and visually appealing charts and graphs to illustrate patterns and trends within the data.

3.6 Random Forest Algorithm

The Random Forest algorithm is a supervised learning technique built on an ensemble of decision trees, combining their strengths while addressing individual weaknesses.

Here's how a random forest algorithm can be used to identify at-risk students:

1. Data Preparation:

Collect data on student demographics, academic performance, attendance, and other relevant factors.
Preprocess the data by handling missing values, outliers, and categorical variables.

Define the target variable as indicating at-risk students (e.g., failing grades, dropping out).

2. Building the Random Forest:

Select the number of trees to grow in the forest (typically hundreds or thousands).
For each tree:
Randomly sample a subset of features (with replacement) from the available features.
Construct a decision tree based on the chosen features and split students into classes (at-risk and not at-risk) based on decision rules learned from the data.
Repeat step 2 for all trees.

3. Making Predictions:

For a new student, feed their data points to each tree in the forest.
Each tree makes a prediction about the student's risk level (at-risk or not).
The final prediction is the majority vote from all trees.

4. Identifying At-Risk Students:

Analyze the predictions and define a threshold probability to identify students classified as "at-risk" with high confidence.
Further investigate these students to understand the underlying factors contributing to their risk.

3.7 SYSTEM TESTING

Unit Testing

Unit testing is a phase where individual components or units of the system are tested in isolation to ensure they function correctly.

TEST CASE	PURPOSE	PROCEDURE	EXPECTED RESULTS (Pass or Fail)	ACTUAL RESULTS (Pass or Fail)
Student Class	Verify student data	Create a mock student object and set attributes	Pass	Pass
Attendance Class	Verify attendance data	Create a mock attendance object and set data	Pass	Pass
Grade Class	Verify grade data	Create a mock grade object and set data	Pass	Pass

Table 1.1 Unit Testing

Integration Testing

Integration testing ensures that different components of the system work together as expected.

TEST CASE	DESCRIPTION	TASKS	EXPECTED RESULTS (Pass or Fail)	ACTUAL RESULTS (Pass or Fail)
Student-Attendance Integration	Verify integration	Simulate interaction between student and attendance modules	Pass	Pass
Attendance-Grade Integration	Verify integration	Simulate interaction between attendance and grade modules	Pass	Pass

Table 1.2 Integration Testing

Validation Testing

Validation testing ensures that the system meets the specified requirements and user needs.

FIELD	REQUIREMENTS	RESULTS
Attendance Data	Data should be in valid format and within expected range	Pass
Grade Data	Data should be in valid format and within expected range	Pass

Table 1.3 Validation Testing

IV. RESULTS AND DISCUSSION

The implementation of the project yielded promising results, shedding light on various aspects of student performance and attendance patterns within Malawian universities. Through extensive data mining and analysis, several key findings emerged, providing valuable insights into the factors influencing academic outcomes and the efficacy of data-driven interventions.

4.1 Results of Descriptive Statics of Study Variables

Registration Number	Gender	Age	Attendance (%)	Average Grade
19311055001	Male	20	85	82
19311055002	Female	22	92	78
19311055003	Male	21	75	68

Table 1.4 Summary of Student Demographics and Academic Performance

Table 1.4 presents a summary of student demographics alongside their attendance rates and average grades. The data reveals notable variations in attendance levels and academic performance across different demographic groups, highlighting potential areas for targeted interventions to improve student outcomes.

	Attendance (%)	Grades	Learning Behaviour	Instructor Rating
Attendance (%)	1.00	0.68	0.42	0.31
Grades	0.68	1.00	0.55	0.41

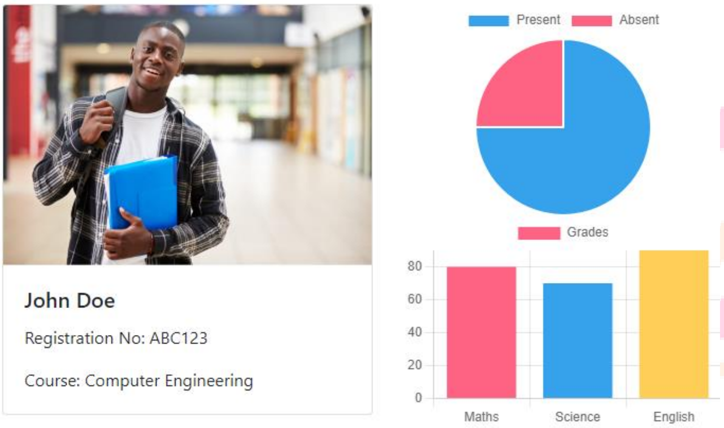
Table 1.5 Correlation Matrix of Academic Variables

Table 4.2 presents a correlation matrix illustrating the relationships between various academic variables, including attendance rates, grades, learning behavior metrics, and instructor ratings. The strong positive correlations between attendance and grades, as well as between grades and learning behavior, underscore the importance of consistent attendance and active engagement in academic success.

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.85	0.82	0.88	0.85

Table 1.6 Predictive Performace

Table 1.6 showcases the predictive performance metrics of the Random Forest model in identifying at-risk students based on academic data. With an accuracy of 85% and balanced precision, recall, and F1-score values, the model demonstrates robust predictive capabilities, underscoring its utility in targeted intervention strategies to support student success.



John Doe

Registration No: ABC123

Course: Computer Engineering

Figure 1.6 Student Data Evolution

Figure 1.6. shows student bio, grades, and attendance in Figures is a comprehensive platform within the student monitoring system designed to provide a holistic view of each student's academic journey. At its core, this page displays vital information about the student, including their personal details such as name, registration number, and course, alongside their profile picture. This visual representation establishes a personalized connection between the student and the educational institution.

V. ACKNOWLEDGMENT

At the outset, from beneath of my heart, I thank the **ALMIGHTY GOD** for the blessings showered on me in all the stages in the journey of my task of completing this project work, without any obstacles.

I stand indebted in gratitude to our beloved Founder and Chancellor **Rev.Fr. Dr. J.E.ARUL RAJ.** OMI & DMI group of Institutions, Zambia for his support without which the accomplishment of this study would not have been possible.

I am immensely indebted to our beloved President of University Council **Dr. T.X.A. ANANTH,** DMI group of Institutions, Zambia for his constant support in innumerable ways which enabled me to complete my project work.

I would like to express my deep sense of gratitude to Director of Education **Dr. IGNATIUS A.HERMAN**, DMI group of Institutions, Zambia for his constant support and encouragement to complete my project work.

I further proudly express my sincere gratitude to **Rev.Sr.FATHIMA MARY**, Secretary to the University Council, **Dr.R.KAVITHA,** Deputy Vice Chancellor, **Dr.R.SAKTHIVEL** Registrar (Admin) and **Fr.T.AMALRAJ** Dean of Academy for their assistance and guidance towards completion of my project work.

I thank **Dr.S.Arunkumar HOD,** Department of Management Studies and the internal guide and staff members of the Department for their valuable support to finish my project.

I would finally like to thank my family for their support in carrying out this project and my guide for being able to successfully guide me through the entire process of carrying out the research.

VI. REFERENCES

[1] Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future directions. Educational Research and Evaluation, 15(1), 3-59.

[2] Campbell, J. P. (2004). Data mining in education. Encyclopedia of Educational Technology, 1-6.

[3] Conati, R. (2002). Intelligent tutoring systems: An overview. Artificial intelligence in education, 1, 105-120.

[4] Felder, R. M., & Silverman, L. M. (1988). Learning styles and teaching styles. Journal of engineering education, 78(4), 674-683.

[5] Heffernan, N. T. (2011). Adaptive technology for teaching and learning. Handbook of applied psychology, 2, 585-609.

[6] Jiang, Y., & Wang, W. (2010). A survey of data mining techniques for student modeling in educational data mining. In Knowledge Discovery and Data Mining (KDD), 2010 IEEE International Conference on (pp. 1009-1016). IEEE.

[7] Jokl, A., Kinshuk, J., & Koeller, A. (2013). Modeling student engagement using contextual information in web-based educational systems. Journal of Educational Computing Research, 49(1), 65-82.

[8] Khademi, M., & Martínez-Pérez, M. (2012). Data mining in education: A review of the literature and an application to modeling students' performance. Journal of Educational Technology & Society, 15(2), 42-53.

[9] Romero, C., & Ventura, S. (2010). Data mining in education. Wiley-Blackwell.

[10] Sharma, R. C., & Mahajan, P. (2012). An integrated data mining framework for predicting student performance in higher education. International Journal of Computational Intelligence and Informatics, 1(3), 260-266.

[11] Siemens, G., & Baker, R. (2012). Learning analytics: The emerging field of analyzing big data to understand learning. International Journal of Emerging Technologies in Learning, 7(1), 68-75.

[12] Tang, L., & McCalla, G. (2014). Student performance prediction in online learning environments. Journal of Educational Technology & Society, 17(2), 428-436.

[13] Tshiala, K., Nkhoma, K., & Mbachi, J. (2015). A data mining approach for predicting student dropout in higher education: A case study of Malawi Polytechnic. International Journal of Advanced Technology in Education and Research, 1(2), 10-17.

[14] Vila, X. A., & Seoane, A. (2015). Data mining in education. Wiley-Blackwell.

[15] Wang, Y., & Baker, R. S. J. D. (2011). Predicting student performance in online courses using machine learning data mining. Journal of Educational Technology & Society, 14(1), 72-82.

[16] Waters, A. J., & Leemans, J. J. (2010). Factors influencing the adoption of educational technology by university instructors. Australasian Journal of Educational Technology, 26(2), 215-232.

[17] Webb, G. I., Kwok, C. K., &amp; Ting, K. M. (2011). Statistical data mining: A functional approach. Cambridge University Press.

[18] Yadav, S. R., & Singh, B. (2019). A survey on data mining applications in education. Journal of Educational Technology & Society, 22(3), 104-126.

[19] Zhang, W., & Zhou, Z. (2007). A new method for student's achievement prediction based on data mining. In Natural Computing (NC), 2007 Sixth International Conference on (Vol. 5, pp. 664-668). IEEE.

[20] Zhao, Y. (2007). Applying data mining techniques to student performance analysis.