Applied Mathematics and Computational Intelligence
Volume 13, No. 4, 2024 [49-61]

Data Visualization of Student Academic Performance Analysis

Nur Farah Amira Shahidan[1], Alif Faisal Ibrahim[2*], Muhammad Nabil Fikri Jamaluddin[3]

[1,2,3]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM) Perlis Branch, Arau Campus, 02600 Arau, Perlis, Malaysia

[*] Corresponding author : aliffaisal@uitm.edu.my

Received: 14 October 2024
Revised: 23 October 2024
Accepted: 30 October 2024

ABSTRACT

Understanding and enhancing student academic performance has become increasingly crucial, yet educators often face data overload, making identifying and prioritizing factors influencing student performance difficult. This study seeks to create a comprehensive system for visualizing and analysing student academic performance. The research follows the waterfall model, with clearly defined phases to ensure an organized and complete development process. The study's main objectives include determining the requirements and techniques to analyse student academic performance and developing and designing a data visualization dashboard using Microsoft Power BI. The final phase consists of evaluating of the student performance dashboard using the Technology Acceptance Model (TAM). A total of 35 respondents was randomly chosen for evaluation and hands-on sessions. The dimensions encompassed Perceived Ease of Use, Perceived Usefulness, Attitude towards Using, and Behavioural Intention. The Attitude of Using received the highest mean score of 4.39, closely followed by Behavioural Intention with a score of 4.38. The results indicate that the participants have a high level of satisfaction with using the dashboard, considering it advantageous and indicating a desire for future implementation. Finally, the project exhibits potential for further advancement, encompassing predictive analytics and tailored learning recommendations, aiming to provide even more precise and actionable insights into student performance.

Keywords: big data, dashboard, data visualization, Technology Acceptance Model (TAM)

1 INTRODUCTION

The usage of data visualization tools has been prevalent in the generation of valuable and informative data. As highlighted by Asamoah [1], data visualization presently serves as the primary tool for data analysis and interpretation in several corporations and organizations. Within the field of education, the application of data visualization analysis can help practitioners understand complex educational trends. As a result, technology for data visualization is crucial to educational administration, teacher decision-making, and student learning [2]. All academic institutions place the utmost importance on education and student achievement [3]. In recent years, there has been a rising interest in evaluating such data to better understand student learning behaviour [4],[5]. The data offers valuable insights through the analysis of students' learning styles, strengths, and areas for improvement.

Visually representing and analysing data on student achievement can be challenging, despite the ample availability of data. A recurring issue for educators and administrators in the era of data-driven education is the abundance of information available which can result in data overload [6]. Additionally, additional elements such as study habits, school attendance, social activities, the students' family history, and others may have an impact on their performance [7]. Numerous studies have demonstrated that education dashboards can serve as valuable decision-making tools to assist educators in developing their lesson plans, assessing the level of knowledge in the class, and monitoring individual pupils [8],[9],[10],[11].

2 LITERATURE REVIEW

2.1 Factors Related to Student Performance

Numerous demographic factors, such as family structure, language background, race and ethnicity, gender, and student's academic achievement, are significantly influenced by socioeconomic status [12]. For example, minority students may encounter linguistic or cultural obstacles that hinder their ability to succeed academically. Additionally, the financial standing of the family may have a big influence on the students' access to resources and educational opportunities. Suleman et al. [13] confirmed that low socioeconomic status has a detrimental effect on academic achievement by raising stress levels at home and restricting access to necessary resources.

Social factors are crucial in determining how well students perform because they affect their motivation, academic engagement, and general well-being. Peer relationships, extracurricular activities, and community involvement are examples of social factors that significantly influence how well students perform. The next factor refers to school-related factors, which include things like extra paid classes, aspirations for higher education, and absences from school. For example, aspirations for a higher education urge students to participate actively in their academics, which leads to a deeper understanding and a drive for academic excellence. Attending class regularly is crucial for learning new material, developing new skills, and finishing assignments, all of which contribute to overall academic advancement.

2.2 Big Data Characteristics

Big data characteristics are commonly characterized by the "3Vs": growing data volume, high data velocity, and variety of data [14]. Which, when combined, offer organizations an unprecedented opportunity and challenge. Innovative storage solutions and distributed processing power are required to handle the ever-increasing volume. To uncover its hidden insights, the wide variety of data necessitated adaptable tools and techniques, while velocity demanded real-time analytics and agile decision-making. In a data-driven world, mastering these 3Vs enables organizations to gain a competitive edge, optimize operations, and find new growth opportunities.

The variety of big data in education, stemming from diverse sources like student demographics, academic records, learning styles, and behavioral patterns, holds significant implications for student performance. Proper analysis of this varied data can enable a comprehensive understanding of individual learning needs, allowing educators to tailor teaching methodologies to match diverse student requirements. As a result, institutions can identify correlations between various factors, such as socioeconomic background, extracurricular activities, or engagement levels. Such a comprehensive perspective facilitates the design of customized educational experiences, focused interventions, and the formulation of inclusive approaches that address the unique abilities and difficulties of every student.

2.3 Big Data Characteristics

The technique of visualizing data is known as data visualization, it involves interpreting the results of an analysis in various ways to provide a more efficient decision-making process [15]. Visual representations are made of the data insights to improve the findings' readability and accessibility. Dedicated platforms for visualization such as Tableau, Microsoft Power BI, Google Data Studio, or programming languages like Python (Matplotlib, Seaborn), R (ggplot2), or JavaScript libraries (D3.js) are widely used to create comprehensive and interactive visualizations. Examples of related works include a study by Mohamad Ghazali and Saad [16] which utilized multidimensional data visualization methods in online student performance system, demonstrating how artificial intelligence and visualizations can be combined to enhance primary school education. In addition, Leo et al. [6] proposed an academic dashboard for monitoring key performance indicators (KPIs) based on data from the PDDIKTI feeder or Higher Education Database, which contains data accumulated from various universities in Indonesia.

3 METHODOLOGY

The Waterfall method, comprising five important phases: planning, analysis, design, development, and testing. These phases form the cornerstone of a research project, guiding its systematic and meticulous evolution. Every stage of this research methodology waterfall hybrid is crucial to maintaining the validity, reliability, and effectiveness of the research process. The waterfall method's sequential design ensures that each stage builds on the preceding one, resulting in believable and effective results.

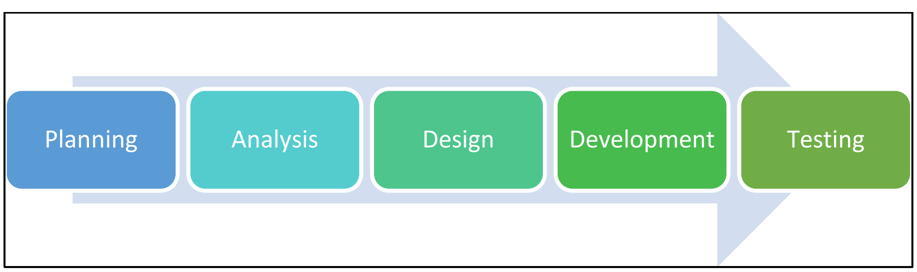

Figure 1: Methodology (based on Prototyping Model)

The first phase is a thorough assessment of scholarly resources drawn from prestigious academic databases, and repositories. The goal is to gain a complete understanding of existing research, identify gaps, and evaluate the relevance of the project's issue statement by meticulously studying scholarly articles, theses, and conference papers from many sources. The analysis phase involves data collection and a comprehensive literature review to gather relevant information and insights for the project. The key job in the early phase of academic performance analysis is to identify appropriate data sources that offer relevant information for the study.

During the design phase, wireframes for both the dashboard and website were developed as well as other project requirements. Within the development phase, the design from the previous phase was implemented into an actual project. Among the activities within this phase include setting up a Data Warehouse, a centralized repository designed to store and manage large volumes of structured and unstructured data. Powerful tools for dealing with massive datasets in this context, combined with a distributed file system, allow for the storage of massive data across numerous servers while simultaneously ensuring scalability and fault tolerance. The use of appropriate tools in the Data Warehouse activity lays the groundwork for a successful Extract, Transform, Load (ETL) process. After the ETL process is completed, the output is used for building the dashboard. Lastly, the testing process was conducted by end users to evaluate the developed dashboard's usability.

4 DEVELOPMENT

4.1 Extract

Data collection is an organized process that consists of three major stages: extraction, transformation, and loading (ETL) into a centralized repository. During the extraction phase, data is gathered from a variety of sources, including databases, files, APIs, and external systems. Once gathered, the data is transformed, polished, standardized, and cleansed to meet the central repository's criteria. The modified data is then loaded into the repository, ensuring that it is efficiently organized, indexed, and easily available for analysis and decision-making needs. The first phase of data collection is the extraction process, in which data is acquired from multiple sources to provide the essential information for analysis. Data for this project is gathered via Kaggle. The student performance dataset is available from the Kaggle repository.

4.2 Transform

Initially, the data is loaded into Jupyter Notebook, which offers a flexible platform for data manipulation. To improve data organization and identification, additional columns are added, and column names have been transformed for clarity and understanding, ensuring ease of interpretation for future studies. Next, numeric values in certain columns are transformed and replaced with relevant labels using predefined mapping dictionaries. This process seeks to standardize the data and make it more understandable. Furthermore, the datasets are vertically concatenated, resulting in a single unified dataset, which is then divided into separate data frames to focus on specific components like student information, subject details, and academic success metrics. This segmentation enables targeted analysis and provides deeper insights into the dataset. Finally, the modified data frames are saved as CSV files, retaining the processed data for later study and reference. This thorough transformation process guarantees that the data is optimal for analysis, resulting in more informed decision-making and useful insights. Figure 2 shows an example of transformation process of data using Jupyter Notebook.

```python
import pandas as pd

# Assuming combined_df contains the combined data from the tables Student, Subject, and AcademicPerformance

# Splitting data into separate DataFrames
student_df = combined_df[['student_id', 'gender', 'age', 'address', 'famsize', 'parent_status',
                          'Medu', 'Fedu', 'Mjob', 'Fjob', 'guardian', 'school',
                          'family_support', 'family_relationship']].drop_duplicates()

subject_df = combined_df[['subject_id', 'subject_name']].drop_duplicates()

academic_performance_df = combined_df[['student_id', 'subject_id', 'traveltime', 'studytime',
                                       'failures', 'school_support', 'extra_class',
                                       'extracurricular', 'nursery', 'higher_education',
                                       'internet', 'freetime', 'goout', 'day_alcohol',
                                       'weekend_alcohol', 'health', 'absences', 'G1', 'G2', 'G3']]
```

Figure 2: Splitting Combined Data into Separate Data Frames

4.3 Loading

The dataset stored in the data warehouse (Hive) was prepared for further operations. This involved using Beeline to connect to HiveServer2 and execute SQL queries on the Hive data warehouse within the VMware Workstation 16 virtual machine.

```
0: jdbc:hive2://localhost:10000> CREATE EXTERNAL TABLE academicperformance (  student_id INT,
. . . . . . . . . . . . . . . .>     subject_id STRING,
. . . . . . . . . . . . . . . .>     traveltime STRING,
. . . . . . . . . . . . . . . .>     studytime STRING,
. . . . . . . . . . . . . . . .>     failures INT,
. . . . . . . . . . . . . . . .>     school_support STRING,
. . . . . . . . . . . . . . . .>     extra_class STRING,
. . . . . . . . . . . . . . . .>     extracurricular STRING,
. . . . . . . . . . . . . . . .>     nursery STRING,
. . . . . . . . . . . . . . . .>     higher_education STRING,
. . . . . . . . . . . . . . . .>     internet STRING,
. . . . . . . . . . . . . . . .>     freetime STRING,
. . . . . . . . . . . . . . . .>     goout STRING,
. . . . . . . . . . . . . . . .>     day_alcohol STRING,
. . . . . . . . . . . . . . . .>     weekend_alcohol STRING,
. . . . . . . . . . . . . . . .>     health STRING,
. . . . . . . . . . . . . . . .>     absences INT,
. . . . . . . . . . . . . . . .>     G1 INT,
. . . . . . . . . . . . . . . .>     G2 INT,
. . . . . . . . . . . . . . . .>     G3 INT
. . . . . . . . . . . . . . . .>
. . . . . . . . . . . . . . . .>
. . . . . . . . . . . . . . . .> )ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
. . . . . . . . . . . . . . . .> LOCATION '/user/cloudera/student_data/new_academic_performance.csv' ;
```

Figure 3: Creating a Table of academic performance

4.4 Interactive Dashboard

The final stage of the development concentrated on creating the interactive dashboard. The Student Academic Performance Dashboard effectively uses Gestalt principles to improve visual perception and usability. Charts on each page include uniform designs and colour schemes, allowing for easy recognition and interpretation.

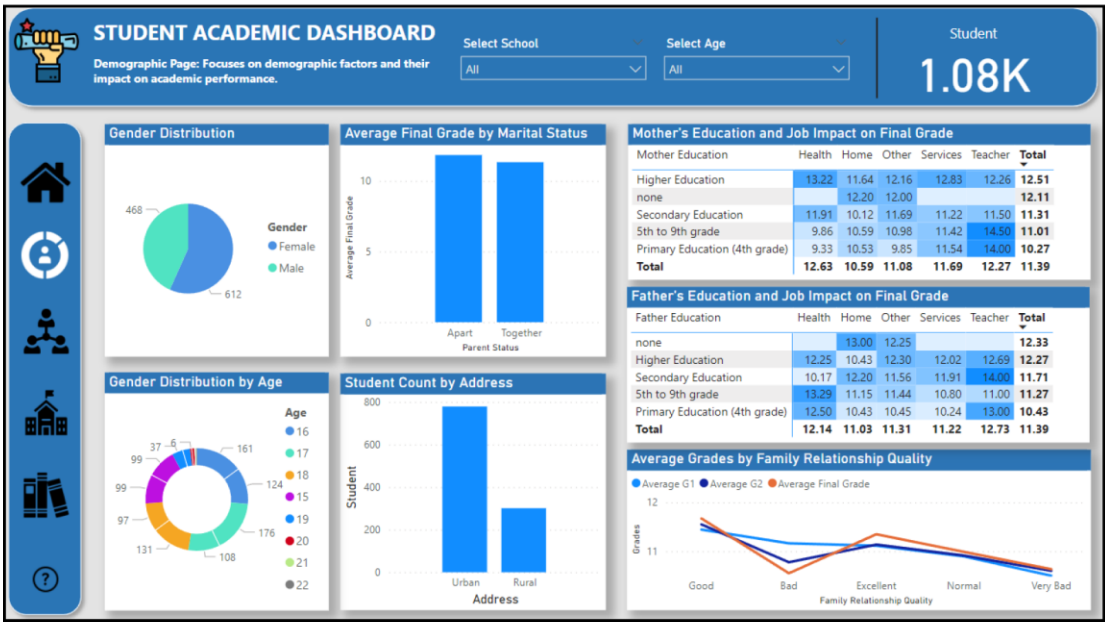

Figure 4: The Demographic Page

The Demographic Page in Figure 4 examines demographic factors and their impact on academic success. It includes numerous charts, such as a pie chart illustrating the distribution of students by gender and a bar chart comparing students' final scores based on their parents' marital status. A doughnut chart represents the distribution of students by age and gender, while a bar chart contrasts the number of students from urban and rural locations. The matrix shows how the mother and father's education and jobs affect students' grades, darker blue shades mean that a student has a higher grade. A line chart illustrates how the quality of family relationships influence student grades. This website also has slicers for selecting specific schools and age groups, which allow users to filter the data.

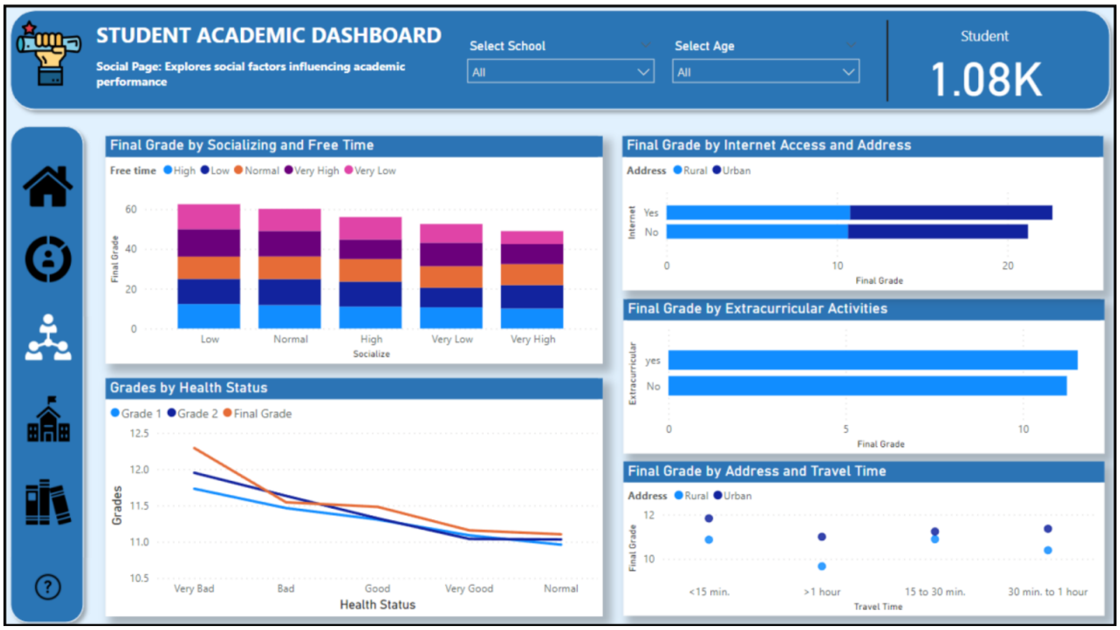

Figure 5: The Social Factor Page

The Social Factor Page in Figure 5 examines how social factors influence student achievement. It contains a stacked bar chart illustrating the relationship between socializing, free time, and final grades, as well as a line chart illustrating how health status corresponds with grades. A bar chart contrasts final scores according to internet availability and whether students live in urban or rural areas. Another bar chart depicts the effect of extracurricular activities on final grades, while a scatter plot depicts the association between travel time to school, address type, and final grades. Slicers on this page allow visitors to filter the data by school and age group, and a card shows the total number of students evaluated.

5 RESULTS AND DISCUSSIONS

5.1 Technology Acceptance Model (TAM)

The Technology Acceptance Model (TAM) is well-known for its structured assessment of four important dimensions: perceived ease of use (PEU), perceived usefulness (PU), attitude (ATT), and intention to use (BI). The evaluation process included administering structured questionnaires to selected 35 participants using Google Forms. These surveys were methodically developed and divided into three sections: the first collected demographic information, the second collected responses on various elements of TAM and the third one asked for suggestions and comments on the dashboard. This technique ensured a thorough examination of the user's attitudes and intentions toward the technology under consideration. Table 1 shows the questions within the survey instrument.

Table 1: TAM questionnaire items

Description	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Average Score
Perceived ease of use (PEOU)						
I found the various functions in the student academic performance dashboard were well integrated	0 (0%)	0 (0%)	4 (11.4%)	15 (42.9%)	16 (45.7%)	4.34
I would imagine that most people would learn to use the student academic performance dashboard very quickly	0 (0%)	0 (0%)	6 (17.1%)	17 (48.6%)	12 (34.3%)	4.17
I felt very confident using the student academic performance dashboard	0 (0%)	0 (0%)	6 (17.1%)	15 (42.9%)	14 (40%)	4.23
I thought the student academic performance dashboard was easy to use	0 (0%)	0 (0%)	7 (7%)	14 (40%)	14 (40%)	4.20
Perceived Usefulness (PU)						
The student academic performance dashboard would enhance my understanding of academic progress and areas needing improvement	0 (0%)	0 (0%)	6 (17.1%)	14 (40%)	15 (42.9%)	4.26
Using the student academic performance dashboard would improve my effectiveness in tracking	0	0	6	10	19	4.37

and analysing academic performance trends	(0%)	(0%)	(17.1%)	(28.6%)	(54.3%)	
I would not need to use other tools or platforms because I could access all necessary academic information through this dashboard	0 (0%)	1 (2.9%)	11 (31.4%)	10 (28.6%)	13 (37.1%)	4.00
The dashboard would contribute to improving overall academic awareness and strategies for achieving academic goals	0 (0%)	0 (0%)	6 (17.1%)	16 (45.7%)	13 (37.1%)	4.20
Attitude of Using (ATT)						
I like the idea of using the student academic performance dashboard to explore and understand various aspects of academic performance	0 (0%)	0 (0%)	5 (14.3%)	11 (31.4%)	19 (54.3%)	4.40
I believe it is a good idea to use the student academic performance dashboard to enhance awareness and understanding of academic performance among students and educators	0 (0%)	0 (0%)	4 (11.4%)	11 (31.4%)	20 (57.1%)	4.46
I find it enjoyable to navigate and discover information using the student academic performance dashboard	0 (0%)	0 (0%)	4 (11.4%)	13 (37.1%)	18 (51.4%)	4.40
I am satisfied with the user interface and design of the student academic performance dashboard	0 (0%)	0 (0%)	5 (14.3%)	14 (40%)	16 (45.7%)	4.31

Behavioral Intention (BI)						
If I have access to the academic performance dashboard, I intend to use it to monitor and enhance student academic progress	0 (0%)	0 (0%)	3 (8.6%)	12 (34.3%)	20 (57.1%)	4.49
Using the academic performance dashboard would be my preferred methods to track student performance, identify areas for improvement and support educational planning	0 (0%)	0 (0%)	5 (14.3%)	15 (42.9%)	15 (42.9%)	4.29
To promote the use of the academic performance dashboard, I would recommend sharing its availability with teachers, educators and school administrators	0 (0%)	0 (0%)	3 (8.6%)	17 (48.6%)	15 (42.9%)	4.34
I will actively support the implementation of the academic performance dashboard if I see it benefitting student learning outcomes and educational strategies	0 (0%)	0 (0%)	4 (11.4%)	13 (37.1%)	18 (51.4%)	4.40

Table 2: Mean score for each dimension

Dimension	Mean Score
Perceived Ease of Use	4.24
Perceived Usefulness	4.21
Attitude of Using	4.39
Behavioral Intention	4.38
TOTAL	4.3

Table 2 shows the mean score of each dimension in the technology acceptance model. Users perceive the dashboard as notably simple to use, with a mean score of 4.24 out of 5, indicating a strong sense of usability and accessibility. Furthermore, the dashboard has a perceived usefulness score of 4.21, indicating that it is effective in providing valuable academic information and resources. Users also have a positive attitude toward using the dashboard, as demonstrated by a mean score of 4.39, showing that they are aware of its benefits and features. Furthermore, with a mean score of 4.38 for behavioral intention, users exhibit a high desire to continue using the dashboard, indicating their confidence in its continuous utility.

Overall, the dashboard received positive comments on all questions. The high average scores, particularly for the desire to use the dashboard to monitor and improve student progress, show that users value and benefit from the dashboard's capabilities. Despite a somewhat lower average rating for using the dashboard as a preferred method of tracking performance, the findings show substantial support for the dashboard's function in educational planning and student progress monitoring. Users are driven to recommend and support its deployment, demonstrating their belief in its potential benefits to student learning outcomes and instructional practices.

6 CONCLUSION

The study "Data Visualization of Student Academic Performance Analysis" takes an innovative method of educational research through the development of an interactive dashboard that visualizes student performance based on demographic, school, and social factors. By using combination of big data analysis and data visualization techniques, unprocessed data can be transformed into usable information with the help of visualization tools. Furthermore, the dashboard contains data categorised by variables that impact student achievement, such as demographic, school, social, and subject-related aspects. By consolidating and displaying these complicated data points in a user-friendly interface, the study enables decision-makers to discover patterns, discrepancies, and opportunities for focused actions. Furthermore, this was further supported by the favourable comments received from the evaluation session.

However, just like any information system, it may encounter problems and limitations in operation that must be addressed. Among a few problems and limitations is that the dashboard includes the source database, which is limited to two schools only, which could lead to biased insights and may not accurately represent broader trends or patterns in student performance across different schools and districts. In addition, students' data is limited because of privacy concerns.

REFERENCES

[1] D. Asamoah, "Improving Data Visualization Skills: A Curriculum Design," 2022. [Online]. Available: https://files.eric.ed.gov/fulltext/EJ1345404.pdf.

[2] H. Li, W. Li, Z. Zhang, H. Yuan, and Y. Wan, "Machine learning analysis and inference of student performance and visualization of data results based on a small dataset of student information," 2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), vol. 15, pp. 117–122, Dec. 2021.

[3] S.-S. M. Ajibade, N. B. Ahmad, and S. M. Shamsuddin, “An Heuristic Feature Selection Algorithm to Evaluate Academic Performance of Students,” 2019 IEEE 10th Control and System Graduate Research Colloquium (ICSGRC), Aug. 2019, doi: 10.1109/icsgrc.2019.8837067.

[4] K. Casey and D. Azcona, “Utilizing student activity patterns to predict performance,” International Journal of Educational Technology in Higher Education, vol. 14, no. 1, Jan. 2017, doi: 10.1186/s41239-017-0044-3.

[5] A. F. Meghji, N. A. Mahoto, M. Ali Unar, and M. Akram Shaikh, “Predicting student academic performance using data generated in higher educational institutes,” 3C Tecnología_Glosas de innovación aplicadas a la pyme, pp. 366–383, May 2019.

[6] A. Leo, A. R. Kurniawan Maranto, F. Fanjaya, and J. Supriyadi, “Academic Dashboard For Monitoring KPI Based Using Data Feeder Dikti”, bit-Tech, vol. 4, no. 3, pp. 139–145, Jun. 2022.

[7] C.-C. Kiu, “Data mining analysis on student’s academic performance through exploration of student’s background and social activities,” 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA), vol. 21, pp. 1–5, Oct. 2018. doi:10.1109/icaccaf.2018.8776809

[8] I. Molenaar and C. a. N. K.-V. Campen, “How Teachers Make Dashboard Information Actionable,” IEEE Transactions on Learning Technologies, vol. 12, no. 3, pp. 347–355, Jun. 2018, doi: 10.1109/tlt.2018.2851585.

[9] C. Schifter, U. Natarajan, D. J. Ketelhut, and A. Kirchgessner, “Data-Driven Decision-Making: Facilitating Teacher Use of Student Data to Inform Classroom Instruction,” Learning & Technology Library (LearnTechLib), Dec. 01, 2014. https://www.learntechlib.org/p/148436

[10] F. Xhakaj, V. Aleven, and B. M. McLaren, “How Teachers Use Data to Help Students Learn: Contextual Inquiry for the Design of a Dashboard,” in Lecture notes in computer science, 2016, pp. 340–354. doi: 10.1007/978-3-319-45153-4_26.

[11] S. Michaeli, D. Kroparo, and A. Hershkovitz, “Teachers’ Use of Education Dashboards and Professional Growth”, IRRODL, vol. 21, no. 4, pp. 61–78, Jul. 2020.

[12] G. Ghaleb A. El Refae, A. Kaba, and S. Eletter, “The Impact of Demographic Characteristics on Academic Performance: Face-to-Face Learning Versus Distance Learning Implemented to Prevent the Spread of COVID-19”, IRRODL, vol. 22, no. 1, pp. 91–110, Mar. 2021.

[13] Q. Suleman, I. Hussain, and Z.-U. Nisa, “Effects of Parental Socioeconomic Status on the Academic Achievement of Secondary School Students in District Karak (Pakistan),” International Journal of Human Resource Studies, vol. 2, no. 4, p. 14, Oct. 2012, doi: 10.5296/ijhrs.v2i4.2511.

[14] D. Fuller, R. Buote, and K. Stanley, “A glossary for big data in population and public health: discussion and commentary on terminology and research methods,” Journal of Epidemiology & Community Health, vol. 71, no. 11, pp. 1113–1117, Sep. 2017, doi: 10.1136/jech-2017-209608.

[15] M. Hajirahimova and M. Ismayilova, “Big Data Visualization: Existing Approaches And Problems,” Problems of Information Technology, vol. 09, no. 1, pp. 65–74, Jan. 2018, doi: 10.25045/jpit.v09.i1.07.

[16] N. D. S. Mohamad Ghazali and A. F. Saad, “Online Student Performance System integrating Multidimensional Data Visualization and Chatbot for Primary School”, International Journal of Artificial Intelligence, vol. 9, no. 2, pp. 63-73, Dec. 2022.