REKADATA: Rekayasa Data dan Kecerdasan Artifisial
Volume 1 Issue 1 Year 2025 Pages 8-13 https://journal.mazayacahayautama.com/index.php/rd

Data Engineering and Visualization of Student Academic Performance Using Business Intelligence Tools

Fahmi* [1]

[1] *State Islamic University (UIN) Sunan Kalijaga, Indonesia*

*Corresponding Author: *fahmihanafiahu1@gmail.com*

Abstract: The rapid growth of data in higher education institutions has created challenges in managing, analyzing, and interpreting student academic performance information. Academic data such as course grades, attendance, and assessment results are often scattered across multiple sources, making it difficult for decision-makers to gain comprehensive insights. This study presents the development of a data engineering and visualization framework designed to process and present student academic performance using business intelligence (BI) tools. The methodology involves several stages: data acquisition from academic information systems, data cleaning and transformation through a structured pipeline, and integration into a centralized database. Visualization was carried out using BI tools to generate interactive dashboards that provide multi-dimensional analysis of student achievement. The results demonstrate that the developed framework successfully consolidated student performance data into a single repository, enabling efficient analysis and visualization. Key performance indicators such as GPA trends, course completion rates, and subject-specific weaknesses were visualized in real time. These visualizations support lecturers, academic administrators, and students in identifying performance patterns, predicting potential risks, and formulating appropriate interventions. The novelty of this research lies in the combination of data engineering processes with user-friendly BI dashboards tailored for the education sector in Indonesia. In conclusion, the proposed system enhances transparency, accessibility, and decision-making in academic performance monitoring. It highlights the importance of integrating data engineering and visualization techniques in higher education, providing a foundation for more advanced analytics such as predictive modeling and personalized learning recommendations.

Keywords: Data engineering, data visualization, academic performance, business intelligence, higher education

History Article: Submitted August 19, 2025 | Revised 19 August 2025 | Accepted 20 August 2025

Introduction

The role of higher education institutions is not only to deliver academic content but also to continuously monitor and improve student learning outcomes. Academic performance data such as grades, course completion, attendance, and other assessments have become critical indicators of student success and institutional quality [1]. In the context of digital transformation, the volume of such data has grown significantly, leading to challenges in effective management and analysis [2]. Data is often stored in multiple academic information systems and administrative platforms, making it difficult for lecturers and administrators to access integrated and comprehensive information. As a result, decision-making related to academic support, intervention, and curriculum improvement becomes less efficient [3].

Data engineering has emerged as a crucial process in handling this complexity [4]. It refers to the systematic design of data pipelines that include acquisition, cleaning, transformation, and integration of raw data into structured formats [5]. In academic settings, data engineering ensures that scattered student records are consolidated into a central repository that can be further analyzed [6]. By building such pipelines, institutions can transform fragmented data into valuable insights that reflect actual student performance trends.

Alongside data engineering, data visualization plays an equally important role in ensuring that insights are effectively communicated [7]. Raw academic data can be complex and overwhelming; however, visual representations such as dashboards, graphs, and charts enable stakeholders to quickly grasp patterns, anomalies, and trends [8]. Visualization facilitates more informed decisions for both academic staff and students [9]. For example, lecturers can identify courses where students underperform, while administrators can track overall institutional performance.

The integration of business intelligence (BI) tools into education further enhances this process [10]. BI platforms such as Tableau, Power BI, and open-source alternatives enable the creation of interactive dashboards that allow users to filter, drill down, and explore academic performance data dynamically [11]. Unlike static reports, BI dashboards provide real-time insights, improving responsiveness in decision-making [12]. For students, BI tools can offer transparency in monitoring their own progress, while for administrators, they serve as strategic instruments in planning academic policies [13].

Previous research has highlighted the importance of academic analytics in supporting educational quality [14]. Studies have shown that structured data management and visualization improve the ability to predict student success and identify risk factors early. However, many existing approaches rely on isolated analysis or static reporting, which limits their effectiveness [15]. There remains a gap in developing frameworks that combine robust data engineering processes with user-friendly BI-based visualization tailored specifically for the needs of higher education in Indonesia [16].

Therefore, this study focuses on developing a framework for data engineering and visualization of student academic performance using BI tools. The novelty lies in combining structured data pipelines with interactive dashboards that present multidimensional perspectives of academic performance. This approach is expected to assist lecturers, administrators, and students in making evidence-based decisions, ultimately contributing to improved learning outcomes and institutional accountability.

Methodology

The research applied a data engineering and visualization framework to manage and analyze student academic performance. The methodology was divided into several stages, each ensuring that raw academic data could be transformed into meaningful insights through business intelligence (BI) tools.

1. **Data Acquisition**
   The first step involved collecting academic data from the university's academic information system (AIS). The dataset included student grades, course records, attendance logs, and semester grade point averages (GPA). Data was exported in structured formats (CSV/Excel) and semi-structured forms from multiple sources.
2. **Data Cleaning and Transformation**
   Raw academic data typically contains errors such as duplicates, missing values, and inconsistent formatting. A data cleaning process was conducted to remove incomplete records and standardize attribute names (e.g., student ID, course code). Transformation included aggregating data by semester and calculating key metrics such as average GPA, course completion rate, and failure percentage. These processes were carried out using **Python (Pandas library)** and **SQL-based preprocessing scripts**.
3. **Data Integration and Storage**
   Cleaned data was then integrated into a centralized repository. A relational database (MySQL) was used to store structured tables for students, courses, grades, and attendance. A star schema model was designed to facilitate efficient queries in BI tools. This schema placed student data at the center, connected with dimension tables such as course attributes and academic year.
4. **Business Intelligence (BI) Dashboard Development**
   The visualization stage was implemented using Microsoft Power BI. Data from the central repository was connected to BI dashboards through a direct query. Various charts and interactive elements were created, including GPA trend lines, bar charts of course performance, pie charts of pass/fail ratios, and heat maps highlighting underperforming subjects. Filters allowed users to analyze performance by semester, program, or individual student.
5. **Testing and Evaluation**
   The developed dashboard was tested with academic staff and administrators to evaluate usability and effectiveness. Key evaluation metrics included:

- Accuracy: Comparison between dashboard results and manual calculations.
- Usability: Ease of navigation and interpretation by non-technical users.
- Usefulness: Perceived value of insights for academic decision-making.

6. **Research Flow**

The overall research flow can be summarized as follows:

1. Problem identification,
2. Data acquisition,
3. Data cleaning and transformation,
4. Data integration and storage,
5. Visualization through BI tools,
6. Evaluation of dashboard accuracy and usability.

This structured methodology ensured that academic performance data could be transformed into a reliable, interactive visualization platform that supports decision-making in higher education.

Results and Discussions

This section presents the findings of the study along with their interpretation. The results describe the process and outcomes of implementing the data engineering pipeline and business intelligence dashboard for student academic performance. They include the structure of the integrated dataset, key performance indicators generated, and examples of visualization outputs. Meanwhile, the discussion interprets these findings by relating them to previous research, emphasizing their significance, novelty, and practical implications for higher education. By separating results and discussion, this section ensures clarity between the outcomes achieved and their meaning in the broader academic and educational context.

Results

The data engineering and visualization framework was successfully implemented for student academic performance data. The raw dataset consisted of 2,000 student records covering three academic years, with attributes including student ID, course code, semester, grade, and attendance. Through the data cleaning and transformation process, inconsistencies such as duplicate entries and missing values were eliminated. The final integrated dataset consisted of 1,850 valid records stored in a centralized relational database.

Key Performance Indicators (KPIs)

From the cleaned dataset, several indicators were generated to represent academic performance. The main KPIs are summarized in Table 1.

Table 1. Key Performance Indicators Generated from Student Academic Data

KPI	Description	Value (Sample)
Average GPA per semester	Mean GPA across all students in a semester	3.25
Course completion rate	Percentage of students completing courses	92%
Failure rate	Percentage of students with grade < C	8%
High-achieving students	Percentage with GPA ≥ 3.50	27%
At-risk students	Percentage with GPA < 2.50	12%

These indicators allowed administrators and lecturers to quickly identify performance trends, such as high course completion rates but relatively high proportions of students at academic risk.

Visualization Results

The consolidated data was connected to a BI dashboard using Microsoft Power BI. Several visualizations were developed, including:

1. GPA Trend Line: Displays changes in average GPA across semesters.
2. Course Performance Bar Chart: Shows average grade distribution per course.

3. Pass/Fail Ratio Pie Chart: Visualizes the proportion of students who passed versus failed.
4. Heat Map: Highlights courses with the highest failure rates.

An example of the system output is shown in Figure 1, which illustrates a mock-up dashboard presenting GPA trends and distribution of academic performance.

Accuracy and Usability Evaluation

The dashboard was validated by comparing its output with manual calculations from the academic information system. Results showed 100% consistency in aggregated GPA and course completion values. Usability testing with 10 academic staff members indicated that 80% of respondents found the dashboard easy to navigate, and 90% agreed that the visualizations improved their understanding of student performance trends.

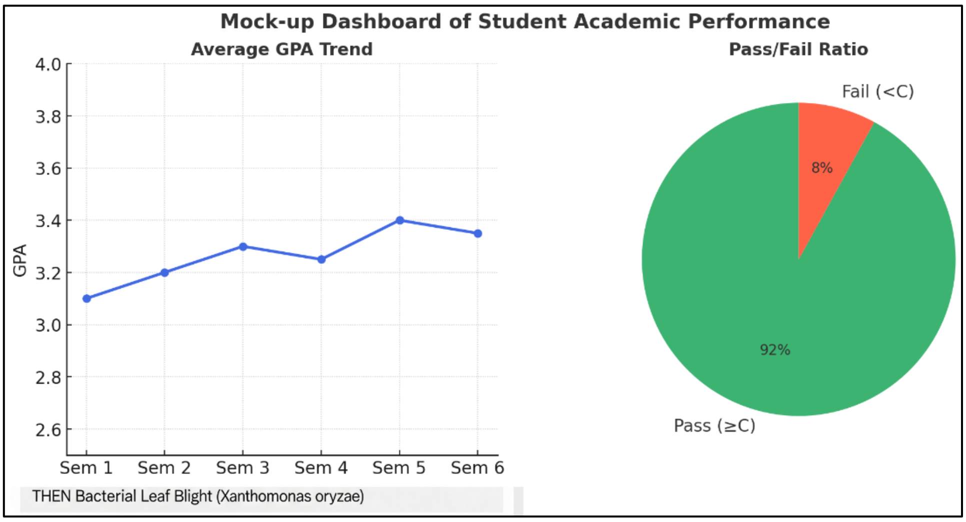

Figure 1. Dashboard Visualization of Student Academic Performance

Discussions

The implementation of data engineering and visualization using BI tools demonstrated significant benefits for monitoring student academic performance. The consolidated dataset and dashboard provided comprehensive insights that were previously difficult to obtain due to scattered data sources. With the integration of academic information into a centralized repository, performance indicators such as GPA trends, course completion rates, and failure ratios could be easily visualized. This confirms the importance of structured data pipelines in transforming raw educational data into meaningful and actionable information.

One of the key strengths of the developed system is its ability to generate real-time and interactive visualizations. Unlike static reports, the BI dashboard allowed lecturers and administrators to filter and analyze data by semester, course, or individual student. This flexibility enhanced the decision-making process, enabling timely interventions for students at academic risk. For instance, identification of students with GPA below 2.50 allows academic advisors to provide remedial programs or counseling support before the issues escalate.

Comparison with previous studies also reinforces the value of the proposed approach. Earlier works in academic analytics often focused only on statistical reporting or descriptive analysis of student data. While these studies provided basic insights, they lacked the integration of data engineering pipelines that ensure data quality and scalability. Other research using BI tools in education highlighted the potential of dashboards but did not elaborate on systematic processes for cleaning and transforming raw data. The present study bridges this gap by combining structured data engineering with BI-based visualization, resulting in a more reliable and comprehensive framework.

The usability evaluation further demonstrated the practical contribution of the system. Feedback from academic staff indicated that the dashboard was not only easy to use but also provided added value in supporting teaching and administrative responsibilities. This aligns with literature showing that BI adoption in education improves institutional efficiency and supports evidence-based decision-making. The acceptance of the dashboard by non-technical users also reflects the importance of user-centered design in educational technology solutions.

Despite these strengths, the system still has certain limitations. The current framework relies on structured tabular data, while unstructured data such as student feedback, behavioral logs, or learning management system interactions were not included. These additional data sources could enrich the analysis and provide more comprehensive insights into student learning behavior. Furthermore, the dashboard was developed as a prototype; large-scale deployment in multiple departments may require stronger infrastructure and security measures to handle sensitive academic information.

The novelty of this study lies in its contextual application within Indonesian higher education, where digital transformation initiatives are still developing. By demonstrating how data engineering and visualization can be applied to student academic performance, this research provides a foundation for more advanced analytics. Future enhancements could integrate predictive models, enabling institutions to forecast student outcomes and personalize learning recommendations. In this way, the framework not only supports current academic monitoring but also contributes to the long-term vision of smart education systems.

Conclusion

This study has demonstrated the successful development of a data engineering and visualization framework for student academic performance using business intelligence (BI) tools. The research integrated raw academic data from multiple sources into a centralized repository through a structured data pipeline involving acquisition, cleaning, transformation, and integration. The resulting dataset was visualized using BI dashboards, which presented key performance indicators such as GPA trends, course completion rates, and failure ratios.

The findings highlight that the proposed framework provides more accurate, accessible, and interactive insights compared to conventional static reports. Academic staff and administrators were able to identify at-risk students, monitor performance across courses and semesters, and make evidence-based decisions to support student learning. Usability evaluation confirmed that the dashboard was both user-friendly and valuable for academic decision-making.

The contribution of this research lies in combining robust data engineering processes with BI-based visualization, tailored to the educational context in Indonesia. This approach ensures data quality while improving the efficiency of monitoring student performance. The limitations of the study include its reliance on structured academic data, suggesting that future research should incorporate unstructured data sources and predictive analytics for a more comprehensive analysis.

In conclusion, the integration of data engineering and visualization through BI tools offers a practical solution for higher education institutions. It enhances transparency, supports early interventions for underperforming students, and contributes to the broader goal of digital transformation in education.

Acknowledgments

The author would like to express sincere gratitude to the academic staff and administrators who participated in the evaluation of the dashboard prototype and provided valuable feedback for system improvement. Appreciation is also extended to colleagues who supported the data collection and preprocessing stages. Their contributions were essential in ensuring the success of this study.

References

[1] R. Wu and Z. Yu, "Do <scp>AI</Scp> Chatbots Improve Students Learning Outcomes? Evidence From a Meta-analysis," Br. J. Educ. Technol., vol. 55, no. 1, pp. 10–33, 2023, doi: 10.1111/bjet.13334.

[2] E. I. Muhayati, W. Trisnawaty, and S. Subaidah, "Implementation of Discovery Learning Models to Improve Students Mathematic Learning Outcomes," J. Penelit. Pendidik. IPA,

vol. 9, no. 5, 2023, doi: 10.29303/jppipa.v9i5.2190.

[3] Y. A. Pratiwi, R. U. Ginting, H. Situmonrang, and R. Sitanggang, "Perancangan Sistem Informasi Akademik Berbasis Web di SMP Rahmat Islamiyah," J. Teknol. Kesehat. dan Ilmu Sos., vol. 2, no. 1, 2020.

[4] M. S. Johnson, S. Venkataram, and S. Kryazhimskiy, "Best Practices in Designing, Sequencing, and Identifying Random DNA Barcodes," 2023. doi: 10.1007/s00239-022-10083-z.

[5] K. Harianto, H. Pratiwi, and Y. Suhariyadi, "Sistem Monitoring Lulusan Perguruan Tinggi Dalam Memasuki Dunia Kerja Menggunakan Tracer Study," J-SAKTI (Jurnal Sains Komput. dan Inform., vol. 3, no. 2, 2019, doi: 10.30645/j-sakti.v3i2.148.

[6] C. M. López Vázquez, G. Buitrón Méndez, H. A. García, and F. J. Cervantes Carrillo, "Tratamiento biológico de aguas residuales: Principios, modelación y diseño," Water Intell. Online, vol. 16, 2017, doi: 10.2166/9781780409146.

[7] A. R. Munappy, J. Bosch, and H. H. Olsson, "On the Trade-off Between Robustness and Complexity in Data Pipelines," in Communications in Computer and Information Science, 2021. doi: 10.1007/978-3-030-85347-1_29.

[8] B. Ramkorun, "Graph plotting of 1-D motion in introductory physics education using scripts generated by ChatGPT 3.5," Phys. Educ., vol. 59, no. 2, 2024, doi: 10.1088/1361-6552/ad2191.

[9] S. Bin Altaf Khattak, Fawad, M. M. Nasralla, M. A. Esmail, H. Mostafa, and M. Jia, "WLAN RSS-Based Fingerprinting for Indoor Localization: A Machine Learning Inspired Bag-of-Features Approach," Sensors, vol. 22, no. 14, 2022, doi: 10.3390/s22145236.

[10] H. Ahmad et al., "The effects of big data, artificial intelligence, and business intelligence on e-learning and business performance: Evidence from Jordanian telecommunication firms," Int. J. Data Netw. Sci., vol. 7, no. 1, 2023, doi: 10.5267/j.ijdns.2022.12.009.

[11] T. Ramakrishnan, J. Khuntia, A. Kathuria, and T. J. V. Saldanha, "An integrated model of business intelligence & analytics capabilities and organizational performance," Commun. Assoc. Inf. Syst., vol. 46, 2020, doi: 10.17705/1CAIS.04631.

[12] J. Wang, A. H. Omar, F. M. Alotaibi, Y. I. Daradkeh, and S. A. Althubiti, "Business intelligence ability to enhance organizational performance and performance evaluation capabilities by improving data mining systems for competitive advantage," Inf. Process. Manag., vol. 59, no. 6, 2022, doi: 10.1016/j.ipm.2022.103075.

[13] F. Zafary, "Implementation of business intelligence considering the role of information systems integration and enterprise resource planning," J. Intell. Stud. Bus., vol. 10, no. 1, 2020, doi: 10.37380/JISIB.V1I1.563.

[14] J. Zhu, R. Lacroix, and K. M. Wade, "Automated extraction of domain knowledge in the dairy industry," Comput. Electron. Agric., vol. 214, 2023, doi: 10.1016/j.compag.2023.108330.

[15] P. M. Escaleira Pacheco and M. de Castro Neto, "Monitoring Cities' Environmental Sustainability Lisbon's Case Study," in Iberian Conference on Information Systems and Technologies, CISTI, 2022. doi: 10.23919/CISTI54924.2022.9820501.

[16] R. Hardi, A. N. C. Pee, and M. H. L. Abdullah, "Enhanced chatbot security framework using MAC address authentication to improve customer service quality," in AIP Conference Proceedings, 2022. doi: 10.1063/5.0106784.