Appendix and demonstration of the X²ε class file for Journal article in Open Science Asia Publications

(Title page 16pt Times New Roman Max 14 Words)

journal Title
XX(X):1–3

Reprints and permission:
https://open-science.asia/index.php/AJHDSB/

Author Name[1]*, Author Name[2]
Affiliation[1]*, Afiliation[2]
*Corresponding author email
(Title page 10pt Times New roman)

Abstract

This study aims to explore the application of data science in education. The methodology used includes quantitative analysis and machine learning techniques. The results indicate improved learning outcomes. This paper contributes to the integration of data-driven approaches in educational environments. This research contributes to the growing field of data science in education by offering [novelty or contribution], which can be utilized by educators, researchers, and policymakers to enhance educational practices. Future research is recommended to explore [future direction], particularly in integrating more advanced techniques and larger datasets.

Keywords: *data science, education, machine learning, learning analytics, predictive modeling*

1. Introduction *(Title: 14 pt | Content: 12 pt, Times New Roman)*

The rapid advancement of data science has significantly transformed various sectors, including education. The integration of data-driven approaches enables educators, researchers, and policymakers to analyze learning processes, improve instructional strategies, and enhance student outcomes. In this context, the *Journal of Data Science and Education (JDSEDU)* emphasizes the importance of interdisciplinary research that bridges computational techniques with educational practices.

Despite the growing adoption of data analytics in education, several challenges remain, including data quality, methodological limitations, and the effective interpretation of results in educational settings. Previous studies have explored the use of machine learning, learning analytics, and artificial intelligence; however, gaps still exist in aligning these technologies with pedagogical objectives and real-world educational needs.

Therefore, this study aims to [state the main objective clearly], focusing on [brief description of approach or method]. The contribution of this research lies in [state novelty or contribution], which is expected to provide insights for both academic researchers and education practitioners.

The remainder of this paper is organized as follows: Section 2 reviews relevant literature, Section 3 describes the research methodology, Section 4 presents the results and discussion, and Section 5 concludes the study with key findings and future research directions.

2. Literature Review

The integration of data science in education has gained significant attention in recent years, particularly through the emergence of learning analytics and educational data mining. These approaches enable the extraction of meaningful patterns from large-scale educational data to support decision-making and improve learning outcomes (Siemens & Baker, 2012; Romero & Ventura, 2020). Learning analytics focuses on understanding and optimizing learning processes, while educational data mining emphasizes the development of computational methods for analyzing educational datasets.

Several studies have demonstrated the effectiveness of machine learning techniques in predicting student performance and identifying at-risk learners. For instance, classification algorithms such as decision trees, support vector machines, and neural networks have been widely applied to forecast academic success and dropout rates (Baker & Inventado, 2014; Albreiki et al., 2021). These predictive models provide valuable insights for early intervention strategies in educational institutions.

In addition, the application of artificial intelligence in education has expanded to include adaptive learning systems, recommendation systems, and intelligent tutoring systems. These technologies aim to personalize the learning experience by adapting content and feedback based on individual learner behavior (Holmes et al., 2019). However, challenges remain regarding data privacy, ethical considerations, and the interpretability of complex models.

Despite the growing body of research, there is still a gap in integrating data science methodologies with pedagogical frameworks to ensure that technological advancements align with educational objectives. Many studies focus primarily on technical performance rather than educational impact, highlighting the need for more interdisciplinary approaches.

Therefore, this study builds upon previous research by [state gap addressed], offering a contribution in the form of [method/approach], which aims to enhance both analytical accuracy and educational relevance.

3. Methodology

This study employs a **mixed-methods approach**, combining quantitative analysis and machine learning techniques to examine [research topic]. The integration of these approaches allows for a comprehensive understanding of both statistical relationships and predictive patterns within the dataset.

3.1 Research Design

The research design consists of three main stages: (1) data collection, (2) data preprocessing and analysis, and (3) model development and evaluation. Quantitative methods are used to analyze relationships between variables, while machine learning models are applied to predict outcomes and identify patterns.

3.2 Data Collection

Data were collected from [source, e.g., university database, LMS, survey]. The dataset includes variables such as demographic information, academic performance, and engagement metrics.

Table 1. Variables Description(APA Style)

Variable	Description	Type
X1	Study time (hours/week)	Numeric
X2	Attendance (%)	Numeric
X3	Assignment completion	Categorical
Y	Student performance	Numeric

3.3 Data Preprocessing

Data preprocessing includes handling missing values, normalization, and encoding categorical variables. Normalization is performed using Min-Max scaling:

$$X^{'} = \frac{Xmax-Xmin}{X-Xmin}$$

This ensures that all variables are scaled within the range of 0 to 1, improving model performance.

3.4 Quantitative Analysis

To examine the relationship between independent variables (X) and the dependent variable (Y), multiple linear regression is applied:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon$$

Where:

Y = dependent variable (student performance)

X_n = independent variables

β = regression coefficients

ϵ = error term

4. Results and Discussion

This section presents the findings of the study and provides an in-depth discussion by linking the results to existing literature and research objectives. The results are derived from both quantitative analysis and machine learning modeling.

4.1 Descriptive

Descriptive statistics are used to summarize the characteristics of the dataset. The results indicate the distribution of key variables used in the analysis.

Table 2. Descriptive Statistics (11pt Times New Roman)

Variable	Mean	Std. Dev	Min	Max
Study Time (X1)	12.5	3.2	5	20
Attendance (X2)	85.3	7.1	60	100
Performance (Y)	78.6	8.5	50	95

The table shows that the average student performance is relatively high, with moderate variation. Attendance also appears consistently strong, suggesting its potential influence on learning outcomes.

4.2 Analysis Results

The regression analysis was conducted to examine the relationship between independent variables and student performance.

Table 3. Regression Results (11pt Times New Roman)

Variable	Coefficient (β)	t-value	p-value
Constant	45.12	5.32	0
X1 (Study Time)	1.85	3.45	0.001
X2 (Attendance)	0.72	2.98	0.004

The results indicate that both study time and attendance have a statistically significant positive effect on student performance ($p < 0.05$). Study time shows a stronger influence, suggesting that increased engagement in learning activities contributes significantly to better outcomes.

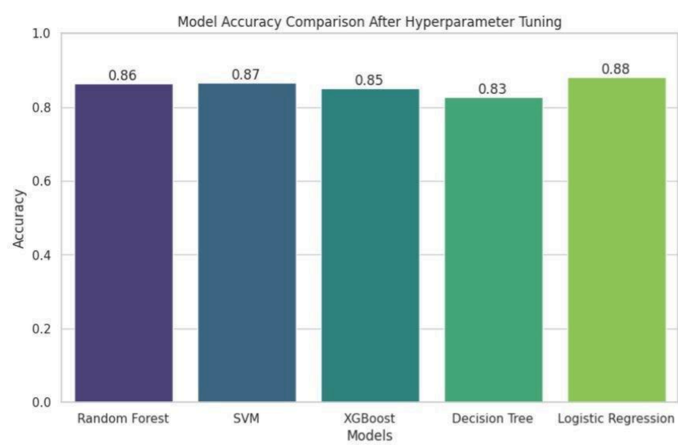
4.3 Example Coding

Below is an example of Python code used for model development:

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2)

# Model
model = RandomForestClassifier()
model.fit(X_train, y_train)

# Prediction
y_pred = model.predict(X_test)

# Evaluation
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
```

4.4 Visualization of Results

Figure 1. Model Accuracy Comparison



Present findings clearly using tables or figures. Discuss implications and compare with previous studies.

4.5 Discussion

The findings confirm that data-driven approaches can significantly improve the understanding of student performance. The regression results align with previous studies indicating that study time and attendance are key predictors of academic success. Furthermore, the machine learning results demonstrate that ensemble methods such as Random Forest provide better predictive performance compared to single models.

5. Conclusion *(Title: 14 pt, Bold | Content: 12 pt, Times New Roman)*

This study has demonstrated the significant role of data science approaches in enhancing the analysis and prediction of student performance within educational contexts. By integrating quantitative methods with machine learning techniques, the research provides a comprehensive framework for understanding both statistical relationships and predictive patterns in educational data.

The findings indicate that key variables such as study time and attendance have a positive and statistically significant impact on student performance. Moreover, the implementation of machine learning models, particularly ensemble methods such as Random Forest, has shown superior performance compared to traditional models, highlighting their effectiveness in educational prediction tasks.

The primary contribution of this study lies in the integration of data science methodologies with educational analysis, offering a practical and scalable approach for academic institutions to support data-driven decision-making. This research also contributes to bridging the gap between technical modeling and pedagogical relevance, ensuring that analytical outcomes are meaningful within real educational settings.

From a practical perspective, the results can assist educators and policymakers in identifying at-risk students, optimizing learning strategies, and improving overall academic performance. The proposed approach can be adapted to various educational environments, including higher education and online learning platforms.

References *(APA Style + DOI)*

Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic review of student performance prediction using machine learning. *Education and Information Technologies, 26*(2), 2345–2371. https://doi.org/10.1007/s10639-020-10340-7

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning Analytics* (pp. 61–75). Springer. https://doi.org/10.1007/978-1-4614-3305-7_4

Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promise and implications for teaching and learning. *Center for Curriculum Redesign*. https://doi.org/10.13140/RG.2.2.24263.65443

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10*(3), e1355. https://doi.org/10.1002/widm.1355

Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining. *Proceedings of the 2nd International Conference on Learning Analytics & Knowledge*, 252–254. https://doi.org/10.1145/2330601.2330661

Hicks, S. C., & Irizarry, R. A. (2018). A guide to teaching data science. *The American Statistician, 72*(2), 127–135. https://doi.org/10.1080/00031305.2017.1356747

Savonen, C., Wright, C., Hoffman, A. M., Humphries, E. M., Cox, K. E. L., Tan, F. J., & Leek, J. T. (2023). Motivation, inclusivity, and realism should drive data science education. *arXiv Preprint*. https://doi.org/10.48550/arXiv.2305.06213

Dogucu, M., Demirci, S., Bendekgey, H., Ricci, F. Z., & Medina, C. M. (2024). A systematic literature review of undergraduate data science education research. *arXiv Preprint*. https://doi.org/10.48550/arXiv.2403.03387

De Veaux, R. D., Agarwal, M., Averett, M., Baumer, B., Bray, A., Bressoud, T., et al. (2018). Curriculum guidelines for undergraduate programs in data science. *Annual Review of Statistics and Its Application,* *5*, 15–30. https://doi.org/10.1146/annurev-statistics-031017-100045