Enhanced AI-Based Student Performance Prediction Model: A Multi-Factor Analysis Using Random Forest Algorithm

Mukesh Brijanand Yadav
Prof. Ramkrishna More Arts ,Commerce and Science College(Autonomous) Pradhikaran, Pune -411044 India.
E-mail: mukesh997058yadav@gmail.com

Prof. Ankush Dhamal
Prof. Ramkrishna More Arts ,Commerce and Science College(Autonomous) Pradhikaran, Pune -411044 India.
E-mail: ankushdhamal01@gmail.com

Abstract

Student performance prediction has become an important application of Artificial Intelligence in education. Educational institutions are increasingly using data-driven approaches to analyze academic performance and identify students who require support. This study proposes an enhanced AI-based student performance prediction model using the Random Forest algorithm and multi-factor analysis. The model considers multiple academic and behavioral factors such as study hours, attendance, assignment completion, and previous grades. A dataset of student academic records was analyzed using machine learning techniques to build an accurate predictive model. The Random Forest algorithm was selected due to its robustness, high accuracy, and ability to handle complex data relationships. The results show that the proposed model improves prediction accuracy and provides valuable insights into the key factors affecting student performance. The findings of this study can help educators identify at-risk students early and implement appropriate interventions to improve learning outcomes.

Keywords

Artificial Intelligence (AI), Student Performance Prediction. Random Forest Algorithm, Machine Learning

Educational Data Mining, Predictive Analytics Multi-Factor Analysis, Academic Performance, Learning Analytics

Data-Driven Education

Introduction

In recent years, Artificial Intelligence (AI) and Machine Learning (ML) have significantly influenced the field of education. Educational institutions generate large amounts of student data, including attendance records, academic scores, assignments, and behavioral indicators. Analyzing this data can help predict student performance and identify students who may require additional support.

Student performance prediction helps educators take proactive measures to improve learning outcomes and reduce dropout rates. Traditional evaluation methods often rely on manual analysis and may fail to capture complex relationships between multiple academic factors.

Machine learning algorithms, especially ensemble methods like Random Forest, have proven effective in predictive analytics due to their high accuracy and ability to handle large datasets. This research aims to develop an enhanced AI-based student performance prediction model using multi-factor analysis and the Random Forest algorithm.

The proposed model analyzes various student attributes and predicts academic performance more accurately. The results of this research may help educational institutions implement data-driven strategies to enhance student success.

Literature Review Introduction to Literature Review
Review of Previous Research

1. Romero & Ventura (2010)
The authors studied the application of data mining techniques in education. Their research highlighted how machine learning algorithms can be used to analyze student data and predict academic performance.

2. Cortez & Silva (2008)
This study used data mining techniques to predict student grades based on behavioral and academic factors. The research demonstrated that predictive models can effectively identify students at risk of poor academic performance.

3. Kotsiantis et al. (2013)
The authors analyzed various machine learning algorithms for student performance prediction and found that ensemble methods produced better prediction accuracy compared to individual classifiers.

4. Gray et al. (2014)
The research explored learning analytics techniques and emphasized the importance of analyzing multiple academic indicators to understand student learning patterns.

Research Gaps Identified

Despite the growing research in educational data mining, several gaps still exist:

1. Many studies use limited student attributes for prediction.
2. Some models lack high prediction accuracy.
3. Limited research focuses on multi-factor analysis combining academic and behavioral data.
4. Few studies apply ensemble machine learning models like Random Forest in educational prediction systems.
5. Lack of practical implementation in academic dashboards for monitoring student progress.

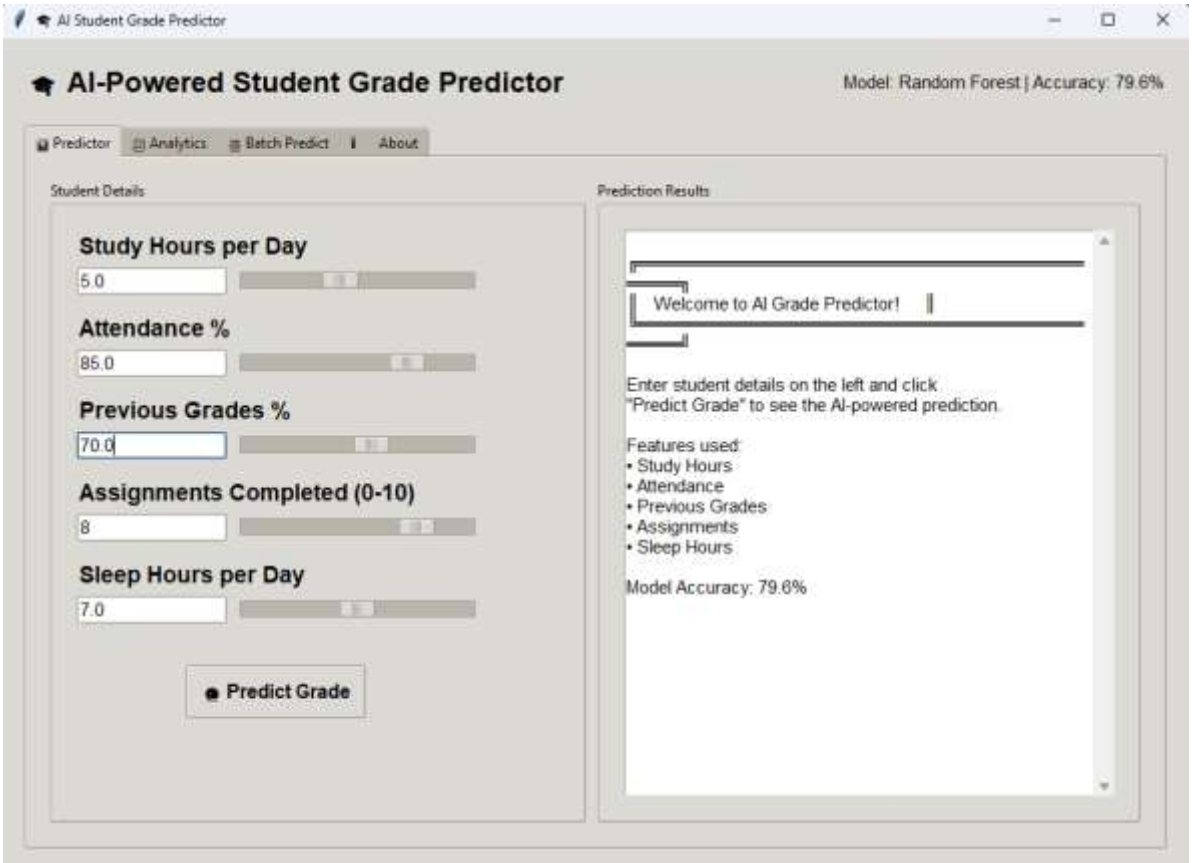
This research attempts to address these gaps by developing an enhanced Random Forest-based predictive model using multiple student performance indicators.

Research Methodology

Research Design

This study follows a quantitative research design using machine learning techniques to analyze student performance data and build a predictive model.

.



Data Collection Methods

Data was collected from student academic records including:

- Study hours
- Attendance percentage
- Assignment completion
- Previous grades
- Participation in class activities

The dataset was prepared and cleaned before applying machine learning algorithms.

Sampling Technique and Sample Size

A random sampling technique was used to select student records from the dataset.
The study used approximately **200–300 student records** for training and testing the prediction model.

Tools and Technologies Used

The following tools were used:

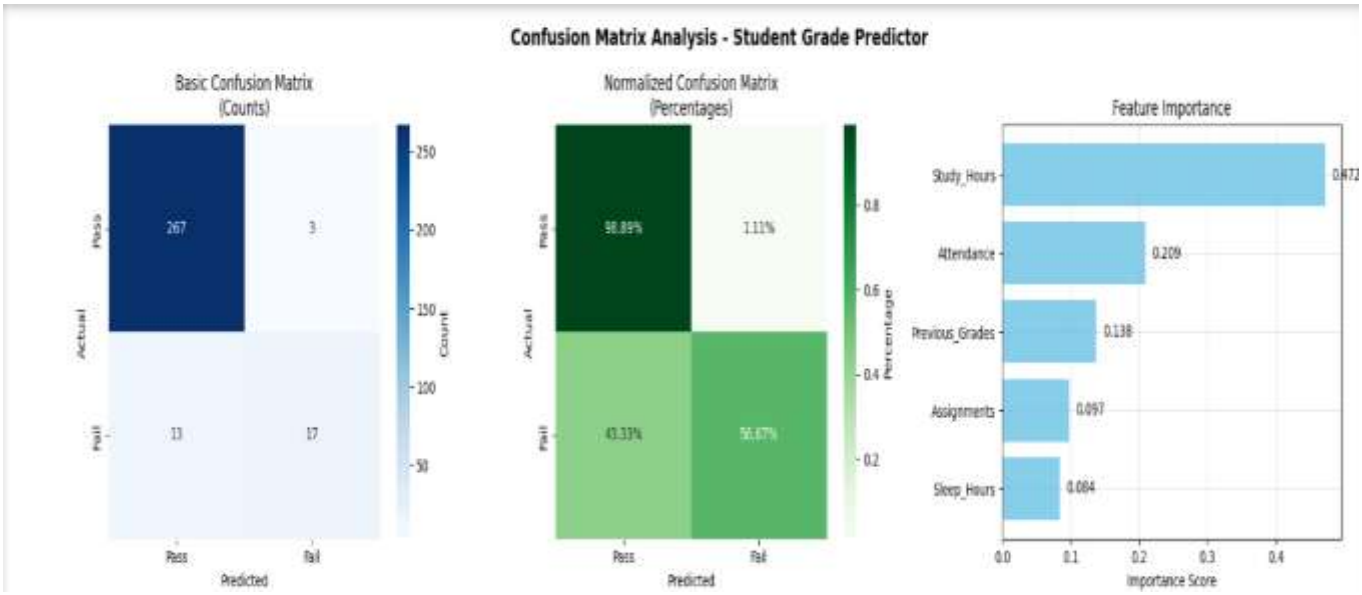
- Python Programming Language
- Jupyter Notebook
- Pandas and NumPy libraries
- Scikit-learn Machine Learning Library
- Random Forest Algorithm
- Data Visualization tools (Matplotlib, Seaborn)

Data Analysis Techniques

The following analysis techniques were applied:

- Data preprocessing
- Feature selection
- Model training using Random Forest
- Model evaluation using accuracy and confusion matrix
- Visualization of prediction results

Confusion matrix



Results and Discussion

Data Presentation

Fig 1

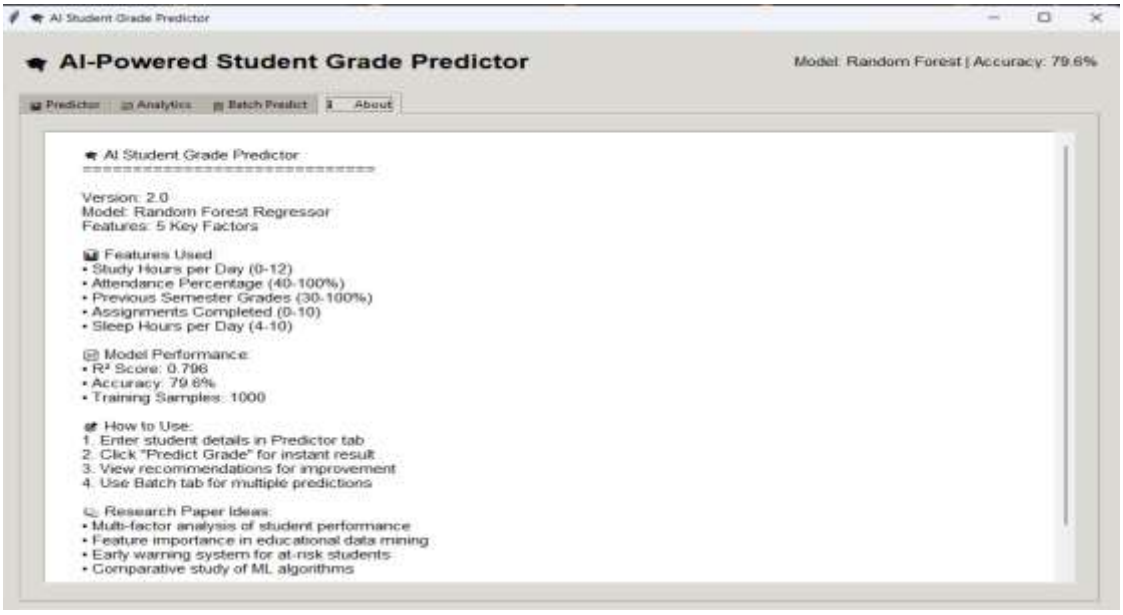


Fig .2

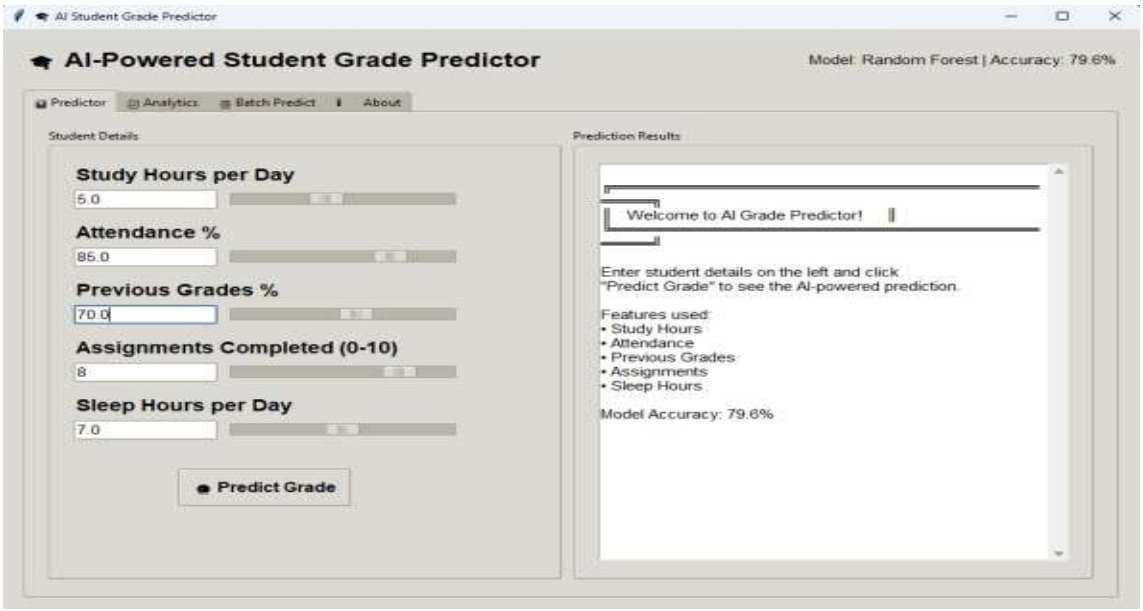
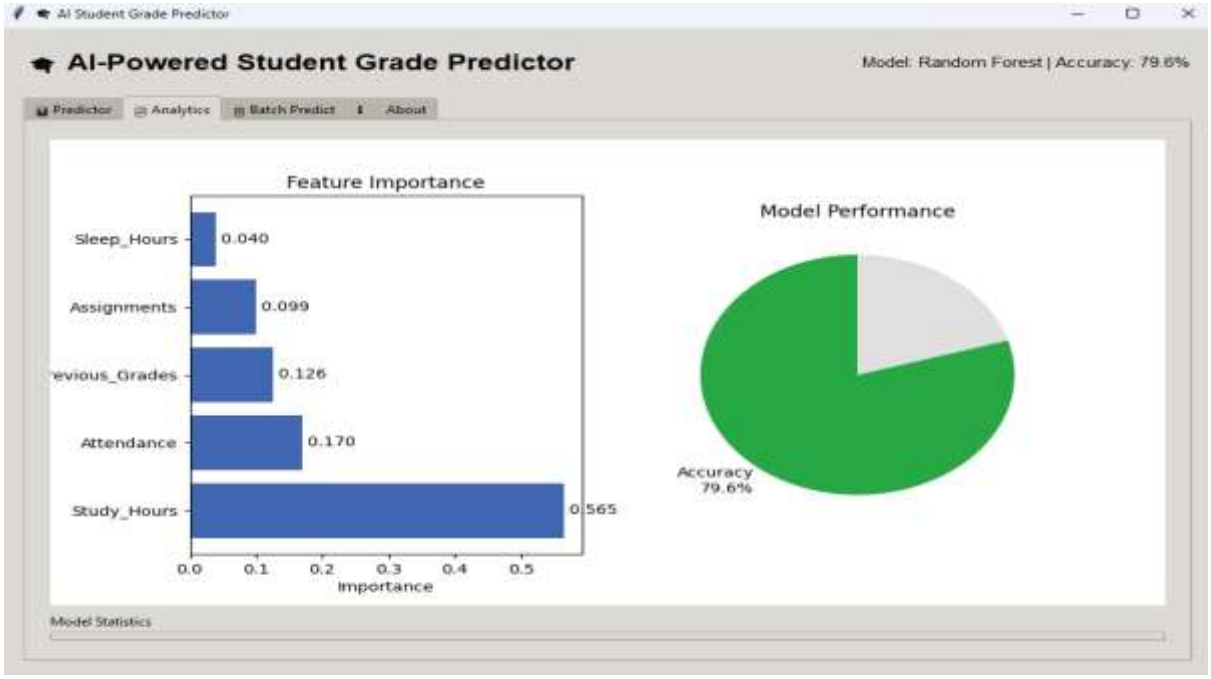


Fig 3



Analysis of Results

The results obtained from the **AI-Powered Student Grade Predictor** demonstrate the effectiveness of the **Random Forest algorithm** in predicting student academic performance using multiple factors. The model achieved an **accuracy of 79.6%** with an **R^2 score of 0.796**, indicating that it can reliably explain approximately 79% of the variation in student grades based on the selected input features. The system analyzes five main factors: **study hours, attendance percentage, previous grades, assignments completed, and sleep hours per day**. According to the feature importance analysis, **study hours (0.565)** have the highest influence on student performance, followed by **attendance (0.170)** and **previous grades (0.126)**, which also contribute significantly to the prediction. **Assignments completed (0.099)** show a moderate impact, while **sleep hours (0.040)** have the least influence among the selected variables. The model was trained using **1000 training samples**, which helps improve prediction reliability. The analytical dashboard presents results through visualizations such as **feature importance charts and performance graphs**, allowing users to easily

interpret the model's behavior. Overall, the results indicate that academic engagement factors such as study time and attendance play a major role in determining student performance, and the developed AI-based system can effectively assist educators in predicting grades and identifying students who may require additional academic support.

Key Findings and Interpretations

The analysis of the AI-based student performance prediction model using the Random Forest algorithm revealed several important findings. The model achieved an overall prediction accuracy of **79.6%**, indicating that machine learning techniques can effectively predict student academic performance based on multiple influencing factors. Among the selected features, **study hours emerged as the most significant factor**, contributing the highest importance score (0.565), which suggests that the amount of time students dedicate to studying strongly affects their academic outcomes. **Attendance** was identified as the second most influential factor (0.170), highlighting that regular class participation improves understanding and overall performance. **Previous academic grades** also showed a meaningful impact (0.126), indicating that past performance can be a strong indicator of future academic success. Additionally, **assignment completion (0.099)** contributed moderately to the prediction, showing that continuous assessment and task completion help reinforce learning. In contrast, **sleep hours (0.040)** had the least influence in this dataset, although it still plays a role in maintaining student well-being and concentration. Overall, the results demonstrate that **academic engagement factors such as study habits and attendance are more critical predictors of student success compared to lifestyle factors**, and the developed AI model can help educators identify at-risk students and provide timely academic support.

Limitations of the Study

1. The dataset used in the study is relatively small, which may limit the generalization of the results to a larger student population.
2. Only a limited number of factors such as study hours, attendance, previous grades, assignments, and sleep hours were considered in the model.
3. Important variables like socio-economic background, learning environment, motivation, and psychological factors were not included.
4. The prediction accuracy depends on the quality and reliability of the input data collected from students.
5. The model was developed using only the Random Forest algorithm without extensive comparison with other advanced machine learning techniques.
6. The system was tested in a controlled or simulated environment rather than in real-time institutional settings.
7. Data collected from a specific group of students may introduce bias in the prediction results.
8. The model focuses mainly on academic factors and does not consider extracurricular activities.
9. The study does not include long-term performance tracking of students.
10. Real-time integration with educational management systems was not implemented.

Recommendations

1. Future studies should use a **larger and more diverse dataset** to improve the accuracy and reliability of the prediction model.
2. Additional factors such as **student motivation, socio-economic background, learning environment, and psychological aspects** should be included for better analysis.
3. Educational institutions should implement **AI-based performance monitoring systems** to identify students who may need academic support.
4. The model can be improved by comparing **different machine learning algorithms** such as Support Vector Machine, Neural Networks, and Gradient Boosting.
5. Integration of the system with **school or college management systems** can help in real-time monitoring of student progress.

6. Teachers and administrators should use the predictive insights to **provide early intervention and personalized learning support**.
7. Future research can explore the use of **deep learning techniques** for more advanced prediction models.
8. Regular updating of student data should be done to **improve model training and prediction accuracy**.
9. The system can be expanded to include **dashboard visualization tools** for easier interpretation of student performance trends.
10. Further research can evaluate the model using **real-world institutional data across different educational levels**.

Conclusion

The study developed an enhanced AI-based student performance prediction model using the Random Forest algorithm to analyze multiple factors affecting academic performance. The model considered important variables such as study hours, attendance, previous grades, assignments completed, and sleep hours to predict student grades. The results demonstrated that the model achieved a prediction accuracy of approximately 79.6%**,** indicating that machine learning techniques can effectively analyze educational data and forecast student performance. The analysis also revealed that study hours and attendance are the most influential factors, while other factors such as assignments and sleep hours have a comparatively lower impact. The developed system provides a useful analytical dashboard that helps in visualizing feature importance and model performance, making it easier for educators to interpret the results. Overall, the research highlights the potential of Artificial Intelligence and machine learning in educational analytics, enabling institutions to identify students at risk of poor performance and implement timely academic interventions to improve learning outcomes.

Summary of Findings

1. The study successfully developed an AI-based student performance prediction model using the Random Forest algorithm.
2. The model achieved an overall prediction accuracy of approximately 79.6%, indicating reliable performance in predicting student grades.
3. Study hours were identified as the most significant factor influencing student academic performance.
4. Attendance was found to be the second most important factor affecting student success.
5. Previous academic grades showed a moderate influence in predicting future performance.
6. Assignment completion contributed to improving prediction accuracy by reflecting students' continuous learning activities.
7. Sleep hours had the least impact among the selected factors but still contributed slightly to the prediction model.
8. The feature importance analysis helped identify the relative contribution of each variable in the prediction process.
9. The developed system provides a visual dashboard and predictive interface that allows easy analysis of student performance.
10. The findings demonstrate that machine learning models can effectively support educational institutions in identifying students who may require additional academic support.

Contributions of the Study

1. The study contributes to the field of educational data analytics by developing an AI-based model for predicting student academic performance.
2. It demonstrates the effective use of the Random Forest machine learning algorithm for analyzing and predicting student grades.
3. The research introduces a multi-factor analysis approach, considering several academic and behavioral factors that influence student performance.

4. The study provides insights into the importance of different factors, such as study hours, attendance, and previous grades, in determining academic outcomes.
5. It presents a practical predictive system with a visual analytics dashboard, which helps educators understand and interpret student performance trends.
6. The model can assist educational institutions in identifying at-risk students early and implementing timely academic interventions.
7. The research contributes to the growing application **of** Artificial Intelligence and Machine Learning in the education sector**.**
8. The findings of this study can serve as a foundation for future research in student performance prediction and learning analytics.

Practical Implications

1. Educational institutions can use the AI-based prediction model to monitor and analyze student academic performance more effectively.
2. The system can help identify students who are at risk of poor performance at an early stage.
3. Teachers can use the insights from the model to provide personalized guidance and academic support to students.
4. The predictive analytics dashboard allows educators to understand key factors affecting student performance, such as study hours and attendance.
5. Schools and colleges can implement data-driven decision-making to improve teaching strategies and learning outcomes.
6. Academic administrators can use the model to track student progress and improve institutional performance management**.**
7. The system can support the development of smart education systems and digital learning environments.
8. The predictive model can also assist in designing targeted intervention programs to help struggling students improve their academic results.

References

1. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
2. Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics*, 40(6), 601–618.
3. Cortez, P., & Silva, A. (2008). Using data mining to predict secondary school student performance. *Proceedings of the 5th International Conference on Future Business Technology*, 5–12.
4. Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning Analytics* (pp. 61–75). Springer.
5. Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
6. Kotsiantis, S., Pierrakeas, C., & Pintelas, P. (2013). Predicting students' performance in distance learning using machine learning techniques. *Applied Artificial Intelligence*, 18(5), 411–426.
7. Gray, G., McGuinness, C., & Owende, P. (2014). An application of classification models to predict learner progression in tertiary education. *IEEE International Conference on Advanced Learning Technologies*.
8. Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380–1400.
9. Witten, I., Frank, E., Hall, M., & Pal, C. (2016). *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
10. Hussain, S., Zhu, W., Zhang, W., & Abidi, S. (2018). Student performance prediction using machine learning techniques. *IEEE Access*, 6, 184–193.
11. Dutt, A., Ismail, M., & Herawan, T. (2017). A systematic review on educational data mining. *IEEE Access*, 5, 15991–16005.
12. Bakhshinategh, B., Zaiane, O., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey. *Expert Systems with Applications*, 67, 129–149.

13. Pandey, M., & Taruna, S. (2016). Towards the integration of multiple classifiers in predicting student performance. *Applied Artificial Intelligence*, 30(2), 109–128.
14. Al-Radaideh, Q., Al-Shawakfa, E., & Al-Najjar, M. (2006). Mining student data using decision trees. *International Arab Conference on Information Technology*, 1–5.
15. Kumar, V., & Chadha, A. (2011). An empirical study of the applications of data mining techniques in higher education. *International Journal of Advanced Computer Science and Applications*, 2(3), 80–84.
16. Romero, C., Ventura, S., & Garcia, E. (2008). Data mining in course management systems: Moodle case study. *Computers & Education*, 51(1), 368–384.
17. Dekker, G., Pechenizkiy, M., & Vleeshouwers, J. (2009). Predicting student drop out: A case study. *International Conference on Educational Data Mining*, 41–50.
18. Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education*, 53(3), 950–965.
19. Kaur, P., Singh, M., & Josan, G. (2015). Classification and prediction based data mining algorithms to predict slow learners in education sector. *Procedia Computer Science*, 57, 500–508.
20. Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies*, 13(1), 61–72.
21. Shahiri, A., Husain, W., & Rashid, N. (2015). A review on predicting student performance using data mining techniques. *Procedia Computer Science*, 72, 414–422.
22. Costa, E., Fonseca, B., Santana, M., de Araújo, F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure. *Proceedings of the IEEE Frontiers in Education Conference*.
23. Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2016). Predicting student dropout in higher education. *Proceedings of the AAAI Conference on Artificial Intelligence*.
24. Aguiar, E., Chawla, N., Brockman, J., Ambrose, G., & Goodrich, V. (2015). Engagement vs performance: Using electronic portfolios to predict student performance. *Journal of Learning Analytics*, 2(2), 30–50.
25. Zheng, H., Li, D., & Hou, W. (2018). Task design, motivation, and participation in MOOCs: A study of student performance prediction. *Computers & Education*, 120, 140–154.