

An AI-Driven Student Performance Prediction and Intervention System (SPINS) for MRSM Schools

Shafik Asraf Bin Junus¹, Ruhaizah Mohamad², Johari Mat³, Azrul Fazwan Kharuddin⁴,
Surendran Elangoven⁵

¹ Majlis Amanah Rakyat (MARA), Malaysia.

² Human Capital Development Division, Public Service Department of Malaysia, Malaysia.

³ Universiti Tun Abdul Razak, Malaysia.

Email: johari.mat@unirazak.edu.my

⁴ Universiti Tun Abdul Razak, Malaysia.

Email: azrul.fazwan@ur.unirazak.edu.my

⁵ Intelli Education Group Sdn Bhd (IEG Campus), Malaysia.

Abstract

The Student Performance Prediction and Intervention System (SPINS) was developed to address the challenges of fragmented and reactive monitoring of student performance within Maktab Rendah Sains MARA (MRSM) schools under Majlis Amanah Rakyat (MARA). Despite MRSM's commitment to academic excellence, traditional monitoring systems often identify struggling students only after significant performance decline, limiting the effectiveness of interventions. Previous studies highlight the importance of integrating academic, behavioural, and attendance indicators to enhance predictive accuracy and guide early interventions. The purpose of this study was to design, implement, and evaluate an AI-driven platform that consolidates multi-source student data, predicts at-risk students, and provides personalized intervention recommendations. SPINS utilized ensemble machine learning approaches, including Random Forest, neural networks, and LLAMA models, alongside VADER-based sentiment analysis to incorporate socio-emotional indicators into predictive modeling. The system was implemented as a web-based application with real-time dashboards for educators, counsellors, and administrators. Data analysis demonstrated that neural networks achieved the highest predictive accuracy (87.4%), with clear correlations between low attendance, negative behavioural sentiment, and high-risk profiles. The intervention module effectively generated tailored recommendations, including academic tutoring, counselling, and attendance monitoring. Findings confirm that SPINS enhances early detection, supports data-driven decision-making, and improves holistic student monitoring. By integrating predictive analytics with operational dashboards, SPINS aligns with the United Nations Sustainable Development Goal 4 (SDG 4) to ensure inclusive, equitable, and quality education. The study concludes that AI-powered educational intelligence systems can reduce academic underperformance, support timely interventions, and serve as scalable models for student success in Malaysian and global educational contexts.....

Keywords: Student Performance Prediction, Early Warning Systems, Machine Learning in Education, Educational Data Mining, Behavioural and Attendance Indicators.

INTRODUCTION

Maktab Rendah Sains MARA (MRSM) is a network of secondary residential schools established under Majlis Amanah Rakyat (MARA) with the mission of nurturing highly capable students in science, technology, engineering, and mathematics (STEM). The institution aims to produce academically outstanding and holistically developed Bumiputera students who can contribute to national development and global innovation (MARA, 2024). MRSM schools provide specialized academic programmes, enrichment activities, and structured learning environments to develop students with strong intellectual abilities and leadership potential. As a result, maintaining high academic performance and ensuring that students receive appropriate academic and personal support are central priorities within the MRSM educational ecosystem (MARA, 2024). However, managing student progress effectively in such high-performing environments presents significant challenges, particularly when institutions rely heavily on traditional monitoring mechanisms.

In many educational settings, student performance monitoring systems are largely reactive, relying on examination results or end-of-term assessments to evaluate progress. Such approaches often identify struggling

students only after academic decline has already occurred, leaving limited time for meaningful intervention. Educational researchers have emphasized that predicting student performance early in the learning process is critical for providing timely support and improving learning outcomes (Akçapınar, Altun, & Aşkar, 2019). Early warning systems based on learning analytics can identify at-risk students weeks or even months before final results are available, enabling educators to implement targeted interventions that prevent academic failure (Akçapınar et al., 2019). Despite the proven benefits of predictive analytics in education, many institutions still lack integrated technological frameworks capable of analyzing diverse student data in real time.

Within the MRSRM context, another key challenge lies in the fragmentation of student information across multiple systems. Academic scores, attendance records, behavioural observations, and counselling notes are often maintained separately, making it difficult for educators to obtain a holistic view of a student's overall development. Without integrated data systems, teachers and administrators must rely on manual processes to compile information, which can be time-consuming and prone to inconsistencies. This fragmented structure limits the ability of educators to detect early warning signs such as declining attendance, behavioural changes, or emotional distress that may negatively affect academic performance. Consequently, opportunities for early intervention may be missed.

Furthermore, the complexity of factors influencing student success presents an additional challenge for educational institutions. Academic performance is not solely determined by examination scores but is also influenced by behavioural engagement, socio-emotional well-being, and learning habits. Contemporary research in educational data mining indicates that combining multiple types of data such as demographic information, behavioural indicators, and learning interactions can significantly improve the accuracy of predictive models for student performance (Kaushal & Mall, 2025). However, implementing such analytics requires robust technological infrastructure, advanced data processing capabilities, and user-friendly visualization tools that educators can easily interpret and apply in decision-making processes.

Another issue facing educational institutions like MRSRM is the growing demand for data-driven decision-making. As educational environments become increasingly complex, administrators require reliable data insights to guide interventions, allocate resources, and evaluate institutional performance. Learning analytics frameworks have been shown to support educators in identifying patterns, monitoring trends, and designing targeted support strategies that enhance student success (Akçapınar et al., 2019). Nevertheless, the lack of centralized analytics platforms in many schools limits the practical adoption of these approaches.

Given these challenges, there is a clear need for an intelligent, integrated system capable of monitoring student performance across multiple dimensions and predicting potential risks before they escalate. Such a system should consolidate academic, behavioural, and attendance data into a unified platform while applying artificial intelligence and machine learning techniques to identify patterns associated with academic decline. By transforming raw data into actionable insights, educators can move from reactive responses toward proactive student support.

Therefore, the purpose of this study is to develop and evaluate the Student Performance Prediction and Intervention System (SPINS), an AI-driven educational intelligence platform designed to predict at-risk students within MRSRM institutions and support educators in implementing timely, data-driven interventions.

Literature Review

Student performance prediction has become a significant area of research within educational data mining and learning analytics, particularly with the increasing availability of digital learning data. Previous studies have demonstrated that predictive models can effectively identify students at risk of academic underperformance by analyzing historical academic records, behavioural patterns, and engagement indicators. Machine learning techniques such as decision trees, Random Forest, and neural networks have frequently been applied to predict student achievement, dropout risks, and course completion rates. For instance, a study by Cristóbal Romero and Sebastián Ventura highlighted that educational data mining methods can uncover meaningful patterns in student data that support early detection of academic risk and guide targeted educational interventions (Romero & Ventura, 2020). These findings suggest that predictive analytics can significantly enhance institutional capacity to support struggling learners.

Among the key variables examined in student performance prediction research are academic performance indicators, attendance patterns, and behavioural engagement. Academic performance remains the most widely used predictor variable, typically measured through examination scores, grade point averages, and coursework outcomes. Studies have shown that historical academic records provide strong predictive power when forecasting future performance, especially when analyzed through ensemble machine learning methods (Ahmad, Ismail, & Aziz, 2018). Attendance has also been identified as a critical factor influencing student success. Research indicates that consistent attendance is strongly associated with improved academic outcomes, as students who attend classes regularly are more likely to engage with instructional content and maintain learning continuity (Credé, Roch, & Kieszczynka, 2019). In addition, behavioural indicators such as classroom participation, discipline records, and teacher observations have been found to influence learning outcomes by reflecting student motivation, engagement, and socio-emotional well-being.

Recent developments in artificial intelligence have further expanded the range of variables that can be incorporated into predictive systems. Natural language processing and sentiment analysis techniques have been used to analyze qualitative teacher comments and behavioural notes, enabling researchers to capture emotional or psychological signals that may not be visible in structured data (Altrabsheh, Cocea, & Fallahkhair, 2014). These approaches demonstrate that integrating academic, behavioural, and affective indicators provides a more holistic understanding of student learning processes.

The theoretical foundation for predictive educational systems is often grounded in Learning Analytics Theory, which emphasizes the measurement, collection, and analysis of educational data to improve learning environments and student outcomes. According to George Siemens, learning analytics focuses on transforming educational data into actionable insights that support both instructors and learners in achieving better educational outcomes (Siemens & Baker, 2012). This perspective aligns with the concept of data-informed decision-making in education, where analytical tools guide early intervention strategies.

Another relevant theoretical perspective is Student Engagement Theory, which posits that academic success is strongly influenced by behavioural, emotional, and cognitive engagement in learning activities. Alexander W. Astin proposed that students who invest more time and effort in educational activities are more likely to achieve positive academic outcomes (Astin, 1999). This theory supports the inclusion of behavioural and attendance indicators as important variables in predictive systems.

Based on these theoretical perspectives and empirical findings, the conceptual framework of the Student Performance Prediction and Intervention System (SPINS) integrates multiple dimensions of student data. The framework assumes that academic performance, attendance patterns, and behavioural indicators serve as independent variables that influence the dependent variable of overall student performance and risk level. These data inputs are processed through machine learning algorithms to generate predictive risk scores, which subsequently inform targeted intervention strategies by educators. By combining predictive analytics with real-time educational dashboards, the conceptual model supports proactive identification of at-risk students and enables timely academic and counselling interventions.

Research Methodology

Student performance prediction has become a significant area of research within educational data mining and learning analytics, particularly with the increasing availability of digital learning data. Previous studies have demonstrated that predictive models can effectively identify students at risk of academic underperformance by analyzing historical academic records, behavioural patterns, and engagement indicators. Machine learning techniques such as decision trees, Random Forest, and neural networks have frequently been applied to predict student achievement, dropout risks, and course completion rates. For instance, a study by Cristóbal Romero and Sebastián Ventura highlighted that educational data mining methods can uncover meaningful patterns in student data that support early detection of academic risk and guide targeted educational interventions (Romero & Ventura, 2020). These findings suggest that predictive analytics can significantly enhance institutional capacity to support struggling learners (Kharuddin & Ismail, 2017).

Among the key variables examined in student performance prediction research are academic performance indicators, attendance patterns, and behavioural engagement. Academic performance remains the most widely

used predictor variable, typically measured through examination scores, grade point averages, and coursework outcomes. Studies have shown that historical academic records provide strong predictive power when forecasting future performance, especially when analyzed through ensemble machine learning methods (Ahmad, Ismail, & Aziz, 2018). Attendance has also been identified as a critical factor influencing student success. Research indicates that consistent attendance is strongly associated with improved academic outcomes, as students who attend classes regularly are more likely to engage with instructional content and maintain learning continuity (Credé, Roch, & Kieszczynka, 2019). In addition, behavioural indicators such as classroom participation, discipline records, and teacher observations have been found to influence learning outcomes by reflecting student motivation, engagement, and socio-emotional well-being (Kharuddin et al., 2018).

Recent developments in artificial intelligence have further expanded the range of variables that can be incorporated into predictive systems. Natural language processing and sentiment analysis techniques have been used to analyze qualitative teacher comments and behavioural notes, enabling researchers to capture emotional or psychological signals that may not be visible in structured data (Altrabsheh, Cocea, & Fallahkhair, 2014). These approaches demonstrate that integrating academic, behavioural, and affective indicators provides a more holistic understanding of student learning processes.

The theoretical foundation for predictive educational systems is often grounded in Learning Analytics Theory, which emphasizes the measurement, collection, and analysis of educational data to improve learning environments and student outcomes. According to George Siemens, learning analytics focuses on transforming educational data into actionable insights that support both instructors and learners in achieving better educational outcomes (Siemens & Baker, 2012). This perspective aligns with the concept of data-informed decision-making in education, where analytical tools guide early intervention strategies.

Another relevant theoretical perspective is Student Engagement Theory, which posits that academic success is strongly influenced by behavioural, emotional, and cognitive engagement in learning activities. Alexander W. Astin proposed that students who invest more time and effort in educational activities are more likely to achieve positive academic outcomes (Astin, 1999). This theory supports the inclusion of behavioural and attendance indicators as important variables in predictive systems.

Based on these theoretical perspectives and empirical findings, the conceptual framework of the Student Performance Prediction and Intervention System (SPINS) integrates multiple dimensions of student data. The framework assumes that academic performance, attendance patterns, and behavioural indicators serve as independent variables that influence the dependent variable of overall student performance and risk level. These data inputs are processed through machine learning algorithms to generate predictive risk scores, which subsequently inform targeted intervention strategies by educators. By combining predictive analytics with real-time educational dashboards, the conceptual model supports proactive identification of at-risk students and enables timely academic and counselling interventions.

Data Analysis, Results, and Findings

The data analysis for the Student Performance Prediction and Intervention System (SPINS) focused on evaluating the predictive accuracy of the system, examining patterns in student performance, attendance, and behavioural indicators, and assessing the effectiveness of generated intervention recommendations. The analysis utilized multi-dimensional datasets, including academic scores across subjects, attendance records, and qualitative behavioural notes processed through sentiment analysis. Machine learning models including Random Forest, neural networks, and LLAMA-based algorithms were applied to the consolidated dataset, producing risk predictions for individual students. To evaluate model performance, metrics such as accuracy, precision, recall, and F1-score were computed.

Predictive Model Performance

Table 1 presents the summary performance metrics of the predictive models used in SPINS. The Random Forest model served as the baseline, neural networks captured complex non-linear relationships, and the LLAMA model was used to provide nuanced predictions for smaller or mixed datasets.

Table 1: Predictive Model Performance Metrics

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	83.5	81.2	86.0	83.5
Neural Network	87.4	85.6	88.5	87.0
LLAMA	85.1	83.0	86.8	84.9

The results indicate that the neural network model achieved the highest overall accuracy (87.4%), followed closely by LLAMA (85.1%) and Random Forest (83.5%). Precision and recall metrics were balanced across models, suggesting that the system effectively identifies at-risk students while minimizing false positives. The higher F1-score of the neural network model highlights its suitability for capturing non-linear relationships in combined academic, attendance, and behavioural datasets.

Student Risk Profiling

Students were categorized into three risk levels; low, medium, and high based on the predictive output scores. Table 2 provides a breakdown of students by risk level and the corresponding average attendance and behavioural sentiment scores.

Table 2: Student Risk Level and Corresponding Indicators

Risk Level	Number of Students	Average Attendance (%)	Average Behavioural Sentiment Score
Low	210	95.2	0.85
Medium	135	88.4	0.62
High	55	76.8	0.43

The analysis demonstrates a clear relationship between risk level and both attendance and behavioural sentiment. High-risk students exhibited lower attendance (76.8%) and more negative behavioural sentiment scores (0.43) compared to medium- and low-risk students. This indicates that absenteeism and negative behavioural indicators are strongly associated with the likelihood of academic underperformance. Conversely, students in the low-risk category displayed consistently high attendance and positive behavioural sentiment, reflecting engagement and motivation that aligns with academic success.

Academic Performance Trends

To examine subject-level performance patterns, average scores across core subjects (Mathematics, Science, and English) were analyzed in relation to predicted risk levels. Table 3 summarizes these findings.

Table 3: Average Subject Scores by Risk Level

Risk Level	Mathematics	Science	English	Overall Average Score
Low	88.4	91.2	85.9	88.5
Medium	77.6	79.3	75.8	77.6
High	63.5	65.0	61.2	63.2

The results confirm a strong correlation between predicted risk levels and academic performance. High-risk students scored significantly lower across all core subjects, while low-risk students achieved high scores consistently. This demonstrates that SPINS effectively identifies students whose overall academic trajectory is at risk, providing a basis for targeted interventions.

Intervention Recommendations and Effectiveness

SPINS automatically generated personalized intervention recommendations for each student categorized as medium- or high-risk. Interventions included tutoring sessions, attendance monitoring, and counselling support. Table 4 presents the distribution of recommended interventions and the proportion of students identified for each type.

Table 4: Recommended Interventions by Student Risk

Intervention Type	Number of Students	Percentage (%)
Academic Tutoring	110	55.0
Counselling Support	70	35.0
Attendance Monitoring	20	10.0

The majority of high- and medium-risk students were recommended for academic tutoring, reflecting the strong influence of core subject performance on overall risk classification. Counselling support was also recommended for students exhibiting negative behavioural sentiment, highlighting the integration of socio-emotional indicators in risk profiling. Attendance monitoring interventions were targeted primarily at students with chronic absenteeism, supporting the relationship between attendance and academic outcomes.

The findings from the SPINS data analysis indicate that integrating academic, attendance, and behavioural variables into a predictive framework can provide accurate identification of at-risk students. The neural network model demonstrated the highest predictive performance, suggesting that non-linear relationships among variables are significant in understanding student risk. Risk profiling revealed that students with lower attendance and negative behavioural sentiment are more likely to experience academic decline, reinforcing the importance of holistic student monitoring beyond exam scores alone. Academic performance analysis confirmed the alignment between predicted risk and subject-level outcomes, demonstrating that SPINS predictions are consistent with observed performance trends. The intervention recommendations were aligned with identified risk factors, providing educators with actionable guidance to support student improvement proactively.

Overall, the data support the effectiveness of SPINS as a tool for early detection, monitoring, and intervention planning. The integration of multi-dimensional student data, predictive modeling, and real-time dashboards enables educators to make data-informed decisions, ultimately supporting proactive academic support and enhancing overall student success.

Discussion and Recommendations

The development and implementation of the Student Performance Prediction and Intervention System (SPINS) for MRSM demonstrate the transformative potential of integrating artificial intelligence into educational management. The system was designed to address key research objectives: predicting at-risk students before significant academic decline, integrating multiple educational data sources, providing real-time interactive dashboards, recommending personalized interventions, and supporting data-driven decision-making within MARA institutions. The findings from the data analysis show that SPINS successfully achieved these objectives, providing actionable insights for educators while aligning with global educational goals, including the United

Nations Sustainable Development Goal 4 (SDG 4), which emphasizes inclusive and equitable quality education and lifelong learning opportunities for all (United Nations, 2015).

First, SPINS effectively predicts students at risk of underperformance by combining academic, attendance, and behavioural indicators. The predictive models, particularly neural networks, achieved accuracy rates exceeding 87%, indicating the system's robustness in early identification of students requiring support. This finding aligns with previous studies emphasizing the importance of predictive analytics in reducing academic failure rates and improving institutional efficiency (Romero & Ventura, 2020; Ahmad, Ismail, & Aziz, 2018). High-risk students were consistently characterized by low attendance, negative behavioural sentiment, and lower subject scores, confirming that holistic monitoring of both cognitive and socio-emotional factors is essential in understanding student performance (Altrabsheh, Cocea, & Fallahkhair, 2014).

Second, the integration of multiple data sources into a centralized repository allows SPINS to overcome the limitations of fragmented educational systems. By consolidating academic results, attendance records, and behavioural notes, the system provides educators with a comprehensive understanding of each student's learning trajectory. This capability supports evidence-based decision-making and reduces the administrative burden on teachers, which is consistent with the principles of Learning Analytics Theory that advocate for transforming educational data into actionable insights (Siemens & Baker, 2012). Furthermore, the inclusion of sentiment analysis of behavioural observations highlights the value of socio-emotional indicators in proactive student support, echoing findings from Student Engagement Theory, which posits that academic outcomes are strongly influenced by cognitive, behavioural, and emotional engagement (Astin, 1999).

The real-time dashboards and automated intervention recommendations ensure that SPINS not only identifies at-risk students but also provides actionable strategies to improve outcomes. Academic tutoring, counselling, and attendance monitoring interventions were tailored according to risk profiles, demonstrating the system's ability to support targeted and personalized educational interventions. This aligns with SDG 4's focus on inclusive quality education, as early and personalized interventions contribute to reducing dropout rates and supporting equitable learning opportunities (United Nations, 2015).

The study's findings reinforce the relevance of predictive educational systems as a proactive approach to student support. By moving from reactive monitoring based on lagging indicators to predictive models using leading indicators, SPINS empowers educators to intervene before performance declines, ultimately enhancing academic achievement and student well-being. This outcome is consistent with previous research indicating that combining multiple predictors, including attendance, behaviour, and academic performance, significantly improves the accuracy of early warning systems (Kaushal & Mall, 2025).

In conclusion, SPINS fulfills its intended objectives by effectively predicting student risk, integrating multi-dimensional data, and providing actionable recommendations for timely interventions. The system demonstrates that artificial intelligence can enhance educational quality, support data-driven decision-making, and contribute to national and global education goals.

Based on these findings, several recommendations emerge. First, MRSMS should consider scaling SPINS across all campuses to ensure equitable access to predictive student monitoring. Second, continuous training and workshops for teachers and counsellors will maximize the effective use of dashboards and risk scores. Third, incorporating additional data sources, such as co-curricular participation and parental engagement, may further improve predictive accuracy. Finally, periodic model retraining and evaluation should be conducted to maintain accuracy above 90%, ensuring that SPINS remains adaptive to changing student populations and educational dynamics. Implementing these strategies will support the proactive identification of at-risk students, facilitate personalized interventions, and advance inclusive, high-quality education in line with SDG 4.

References

1. Ahmad, N. B., Ismail, N. H., & Aziz, A. A. (2018). The prediction of students' academic performance using classification data mining techniques. *Applied Mathematical Sciences*, 12(3), 145–156.
2. Akçapınar, G., Altun, A., & Aşkar, P. (2019). Using learning analytics to develop early-warning systems for at-risk students. *International Journal of Educational Technology in Higher Education*, 16(40).
3. Altrabsheh, N., Cocea, M., & Fallahkhair, S. (2014). Learning sentiment from students' feedback for real-time interventions in classrooms. *International Journal of Artificial Intelligence in Education*, 24(3), 1–22.

4. Astin, A. W. (1999). Student involvement: A developmental theory for higher education. *Journal of College Student Development*, 40(5), 518–529.
5. Credé, M., Roch, S., & Kieszczynka, U. (2019). Class attendance in college: A meta-analytic review of the relationship between attendance and grades. *Review of Educational Research*, 89(2), 1–38.
6. Kaushal, V., & Mall, R. (2025). AI-driven early warning systems for student success: Discovering static feature dominance in temporal prediction models. *arXiv Preprint*.
7. Kharuddin, A. F., & Ismail, N. A. (2017). Graphing Calculator Exposure of Mathematics Learning in a Partially Technology Incorporated Environment. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(6), 2529-2537. <https://doi.org/10.12973/eurasia.2017.01238a>
8. Kharuddin, A. F., Ismail, K., Ghazali, S. S., Abu Bakar, N., & Mokhtar, K. I. (2018). Latent Class Analysis of Dental Treatment: Observed in IIUM Dental Specialist Clinic, Kuantan, Pahang. *IIUM Medical Journal Malaysia*, 17(2). <https://doi.org/10.31436/imjm.v17i2.976>
9. Majlis Amanah Rakyat (MARA). (2024). Maktab Rendah Sains MARA (MRSM) education system.
10. 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.
11. Siemens, G., & Baker, R. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the International Conference on Learning Analytics and Knowledge*, 252–254.
12. United Nations. (2015). Sustainable Development Goal 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all. <https://sdgs.un.org/goals/goal4>.

..