International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

ACADEMIC PERFORMANCE PREDICTION BASED ON MULTI SOURCE, MULTI-FEATURE BEHAVIORAL DATA

[1]Mrs. G. ARCHANA, [2]B. PRANAVI, [3]K. SAMAVED, [4]G. SONITA MISRA

[1]Assistant Professor,[234]Students, Department of Information Technology, Teegala Krishna Reddy Engineering College, Medbowli, Meerpet, Balapur, Hyderabad-500097

ABSTRACT

Academic performance prediction has become an important research area in modern educational systems due to the rapid growth of educational data and digital learning platforms. Traditional evaluation methods mainly depend on examination scores and attendance records, which provide only limited insights into student learning behavior. However, student success is influenced by multiple factors such as study habits, engagement levels, participation in academic activities, and interaction with online learning systems. This project proposes an intelligent academic performance prediction system based on multisource, multi-feature behavioral data to improve prediction accuracy and support proactive educational decision-making. The proposed system integrates academic records, behavioral characteristics, and digital learning activities collected from multiple data sources. Data preprocessing techniques such as cleaning, normalization, handling missing values, and feature selection are applied to improve data quality and consistency. Feature fusion methods are then used to combine multiple attributes into a unified dataset for effective analysis. Machine learning algorithms including Linear Regression, Random Forest, and Neural Networks are employed to analyze learning patterns and predict academic performance. The system classifies students into different performance categories such as high-performing, medium-performing, and at-risk students. It also provides personalized recommendations, learning guidance, and early intervention support to improve student outcomes. Continuous monitoring of student activities enables institutions to track performance trends and identify academic risks at an early stage. The proposed framework enhances prediction reliability, supports personalized learning strategies, improves student retention, and promotes data-driven educational management. Overall, the system contributes to the development of intelligent and adaptive learning environments capable of improving academic success and institutional effectiveness.

Keywords: Academic Performance Prediction, Machine Learning, Educational Data Mining, Behavioral Data Analysis, Random Forest, Neural Networks, Predictive Analytics, Personalized Learning, Student Performance Monitoring, Data-Driven Education

I. INTRODUCTION

The education sector has experienced significant transformation due to the integration of digital technologies, online learning platforms, and data-driven methodologies. Modern educational institutions continuously generate large volumes of data related to student performance, attendance, participation, engagement, and learning behavior. Analyzing this educational data provides valuable

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

opportunities to improve teaching quality and student success. Academic performance prediction has therefore emerged as an important research area because it enables educational institutions to identify learning difficulties at an early stage and implement timely intervention strategies [1]. Traditional academic evaluation systems mainly rely on examination scores and attendance records to assess student performance [2]. Although these indicators provide useful information, they fail to capture important behavioral and engagement-related factors that influence academic success [3]. Student learning outcomes are affected by multiple factors including study habits, classroom participation, motivation, consistency, and interaction with digital learning platforms [4]. The rapid growth of e-learning environments and virtual classrooms has increased the availability of behavioral data such as login frequency, resource utilization, assignment submission patterns, and participation in online discussions [5]. However, many conventional systems do not effectively utilize this information for performance analysis [6]. Educational data mining and machine learning technologies have created new possibilities for extracting meaningful insights from large educational datasets [7]. Machine learning algorithms can identify hidden relationships between different student attributes and academic outcomes [8]. Techniques such as regression analysis, decision trees, random forests, and neural networks are widely used for predictive modeling in educational environments [9]. Predictive analytics allows institutions to monitor student progress continuously and identify at-risk students before academic failure occurs [10]. Early identification helps educators provide academic guidance, counseling, and personalized support to improve student performance [11]. The integration of academic, behavioral, and digital learning data enhances prediction accuracy and supports adaptive learning systems [12]. Furthermore, intelligent prediction systems contribute to improved academic planning, curriculum enhancement, and institutional performance evaluation [13]. Educational institutions are increasingly adopting data-driven approaches to improve learning efficiency and student retention [14]. Continuous monitoring of student activities also helps in identifying learning trends and engagement patterns [15].

This project proposes an intelligent academic performance prediction system based on multisource, multi-feature behavioral data. The system integrates academic records, participation metrics, behavioral characteristics, and digital learning activities to provide a comprehensive understanding of student performance [16]. Data preprocessing techniques such as data cleaning, normalization, missing value handling, and feature selection are applied to improve data quality and consistency [17]. Feature fusion techniques combine multiple attributes into a unified dataset for effective analysis [18]. Machine learning models including Linear Regression, Random Forest, and Neural Networks are implemented to analyze student behavior and predict academic outcomes [19]. Random Forest algorithms improve prediction accuracy by handling complex datasets and minimizing overfitting problems [20]. Neural Networks are capable of identifying non-linear relationships between different educational attributes [21]. The proposed system classifies students into categories such as high-performing, medium-performing, and at-risk students [22]. This classification helps institutions provide timely intervention and personalized learning support [23]. The system also includes recommendation

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

mechanisms that suggest study improvements and engagement strategies for students [24]. Personalized learning enhances student motivation, participation, and knowledge retention [25]. Continuous monitoring of behavioral and academic data allows institutions to track performance changes over time [26]. Predictive insights generated by the system support academic counseling, curriculum development, and decision-making processes [27]. The framework also supports online learning environments by analyzing digital learning activities and engagement metrics [28]. The integration of machine learning with educational analytics improves academic performance prediction and enables proactive educational management [29]. Overall, the proposed system contributes to the development of intelligent, adaptive, and data-driven educational environments that improve student success and institutional effectiveness [30].

II. LITERATURE SURVEY

Educational data mining and academic performance prediction have become major research areas due to the increasing availability of educational datasets and advancements in machine learning technologies. Researchers have explored different methods to analyze student learning behavior and predict academic outcomes using academic, behavioral, and engagement-related data [1]. Romero and Ventura conducted extensive research in educational data mining and emphasized the importance of classification and prediction algorithms in identifying student learning patterns and academic trends [2]. Their work demonstrated that educational data analysis can improve institutional decision-making and learning effectiveness [3]. Thai-Nghe, Drumond, Horváth, and Schmidt-Thieme proposed predictive frameworks based on matrix factorization techniques to improve student performance prediction accuracy [4]. Their study highlighted the effectiveness of machine learning approaches in modeling complex educational relationships [5]. Baker and Yacef explored the application of educational data mining techniques for analyzing student interaction and engagement data [6]. Their research focused on detecting hidden learning patterns using student activity information collected from online learning systems [7]. Shahiri, Husain, and Rashid developed prediction systems using multiple machine learning algorithms such as decision trees, regression models, and neural networks [8]. Their comparative analysis concluded that combining multiple educational features significantly improved prediction reliability [9]. Cortez and Silva investigated the influence of behavioral factors such as study time, participation, and consistency on academic performance prediction [10]. Their findings showed that behavioral attributes strongly influence learning outcomes [11]. Huang and Fang proposed predictive systems integrating learning behavior and course-related factors to improve academic prediction accuracy [12]. Their research demonstrated that behavioral analysis produces better performance evaluation than traditional academic indicators alone [13]. Kotsiantis studied the application of machine learning algorithms in educational systems and highlighted the importance of feature selection and preprocessing techniques [14]. His research emphasized that appropriate preprocessing significantly enhances predictive model performance [15]. Recent studies have also explored deep learning techniques for educational analytics because of their ability to capture complex non-linear relationships among features [16]. Neural network models have shown

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

promising results in predicting student success in online learning environments [17]. Researchers have additionally focused on early risk identification systems that help institutions provide timely support to struggling students [18]. Such predictive systems improve student retention rates and reduce academic failure [19]. Educational researchers also highlighted the importance of continuous monitoring systems capable of analyzing student performance dynamically over time [20].

Modern educational environments generate diverse forms of student data including academic records, behavioral metrics, digital interactions, attendance information, and online engagement activities [21]. Traditional systems mainly rely on examination scores and attendance records, which fail to provide a complete understanding of student learning behavior [22]. To overcome these limitations, researchers have proposed multisource data integration frameworks that combine academic, behavioral, and digital learning attributes [23]. Feature fusion techniques have been widely adopted to merge multiple educational attributes into a unified representation for effective analysis [24]. Random Forest algorithms are extensively used in educational prediction systems because of their high accuracy and ability to handle large datasets [25]. Regression models are useful for estimating academic scores, while neural networks capture deeper feature relationships [26]. Researchers have also applied ensemble learning approaches to improve prediction robustness and minimize overfitting [27]. Continuous behavioral analysis has become increasingly important in modern educational systems because online learning platforms generate valuable engagement-related data [28]. Studies indicate that factors such as login frequency, learning duration, resource usage, and assignment completion patterns strongly influence academic success [29]. Therefore, intelligent academic performance prediction systems based on multisource, multi-feature behavioral data provide more reliable, adaptive, and personalized learning support in modern educational environments [30].

III. PROPOSED SYSTEM

The proposed system introduces an intelligent and data-driven framework for predicting student academic performance using multisource, multi-feature behavioral data. Unlike traditional evaluation methods that rely only on examination marks and attendance records, the proposed system integrates academic information, behavioral characteristics, and digital learning activities to provide a comprehensive understanding of student learning behavior. Academic data such as internal marks, assignments, attendance, and subject performance are combined with behavioral attributes including study habits, participation levels, consistency, and engagement patterns. In addition, digital learning activities such as login frequency, time spent on learning platforms, interaction with study materials, and participation in online discussions are also analyzed. The collected data is first processed using preprocessing techniques such as data cleaning, missing value handling, normalization, and feature selection to improve data quality and consistency. Feature fusion methods are then applied to combine multiple attributes into a unified dataset suitable for machine learning analysis. This integrated approach enables the system to identify hidden relationships among different performance-related factors and improves prediction reliability. The proposed framework supports continuous monitoring of student performance and

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

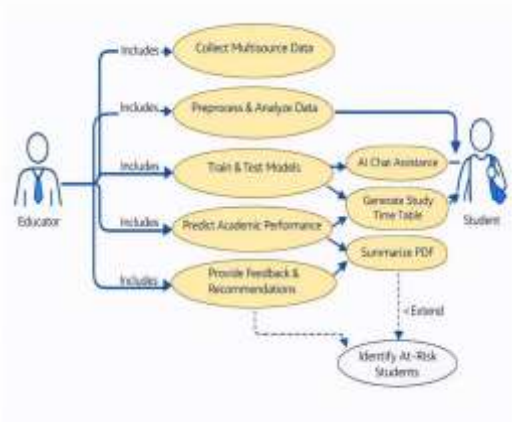
dynamically updates predictions based on newly available data.

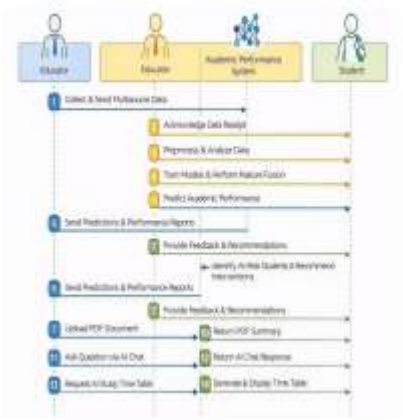
The core functionality of the system is implemented using machine learning algorithms such as Linear Regression, Random Forest, and Neural Networks. Linear Regression is used for basic academic score estimation, while Random Forest improves prediction accuracy by handling complex feature relationships and reducing overfitting problems. Neural Networks are capable of identifying non-linear patterns within educational datasets and generating advanced predictive insights. The trained models analyze student data and classify students into categories such as high-performing, medium-performing, and at-risk students. Early identification of at-risk students enables educational institutions to provide timely academic support, counseling, and intervention strategies. The system also includes a personalized feedback and recommendation module that suggests study improvements, better engagement strategies, and customized learning plans. Advanced features such as AI chatbot support, timetable generation, and PDF summarization further enhance the learning experience. By integrating machine learning with educational analytics, the proposed system improves academic prediction accuracy, supports personalized learning, enhances student retention, and promotes data-driven decision-making in educational institutions.

IV. SYSTEM DESIGN

The system design of the proposed academic performance prediction framework consists of multiple interconnected modules that work together to collect, process, analyze, and predict student academic performance. The first module is the Data Collection Module, which gathers multisource data related to student academic records, behavioral characteristics, and digital learning activities. Academic data includes examination scores, assignments, attendance, and subject-wise performance, while behavioral data includes study habits, consistency, participation levels, and engagement metrics. Digital learning data includes login frequency, learning duration, interaction with online resources, and virtual classroom activities. The collected data is stored in a structured database system for efficient management and retrieval. The next component is the Data Preprocessing Module, which transforms raw data into a clean and structured format suitable for analysis. This module performs operations such as removing duplicate records, handling missing values, normalizing feature scales, and converting categorical attributes into numerical representations. Feature selection techniques are also applied to identify the most relevant attributes influencing academic performance. After preprocessing, the Feature Fusion Module combines multiple educational attributes into a unified dataset, enabling comprehensive analysis of academic, behavioral, and digital learning factors simultaneously.

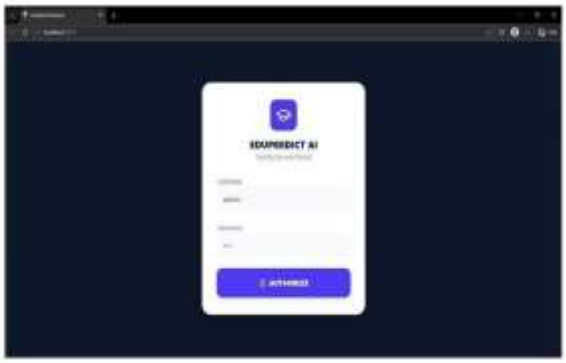
International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

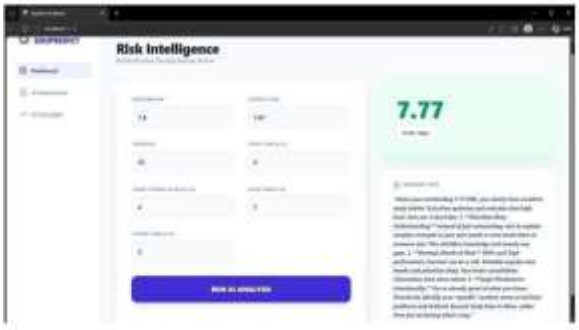




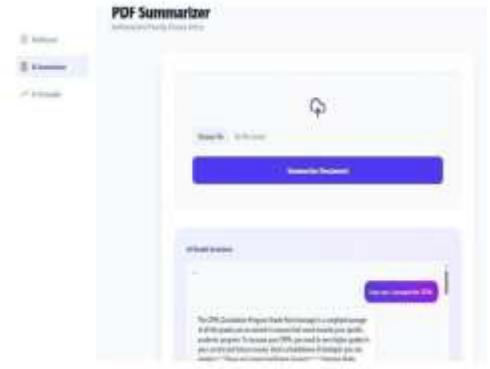
The Machine Learning Module forms the core analytical component of the system. This module applies machine learning algorithms such as Linear Regression, Random Forest, and Neural Networks to learn patterns from historical student data and generate predictive models. The dataset is divided into training and testing subsets using data splitting techniques, and model performance is evaluated using metrics such as R^2 Score and Mean Absolute Error (MAE). Random Forest algorithms are particularly effective because they improve prediction accuracy and handle complex educational datasets efficiently. The Performance Prediction Module uses the trained machine learning models to estimate future student performance and classify students into categories such as high-performing, medium-performing, and at-risk students. The system also includes a Feedback and Recommendation Module that generates personalized suggestions based on prediction outcomes. Students receive recommendations such as improving study consistency, focusing on weak subjects, increasing engagement, and following personalized study schedules. The user interface presents predictions and recommendations in an understandable format using dashboards and reports. Overall, the system design ensures efficient data processing, accurate prediction generation, continuous monitoring, and intelligent academic support for improving educational outcomes.

V. RESULTS





International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



VI. CONCLUSION

Academic performance prediction has become an essential requirement in modern educational environments due to the increasing complexity of factors influencing student learning outcomes. Traditional academic evaluation systems mainly depend on examination scores and attendance records, which provide only limited insights into student behavior and learning patterns. The proposed system overcomes these limitations by introducing a multisource, multi-feature behavioral data analysis framework that integrates academic records, behavioral characteristics, and digital learning activities. Through the application of data preprocessing techniques, feature fusion methods, and machine learning algorithms such as Linear Regression, Random Forest, and Neural Networks, the system is capable of generating accurate and reliable academic performance predictions. The framework supports continuous monitoring of student activities and enables early identification of at-risk students, allowing educational institutions to implement timely intervention and personalized learning strategies. The inclusion of recommendation systems, AI chatbot support, timetable generation, and learning assistance further enhances the effectiveness of the platform. By leveraging educational data mining and predictive analytics, the system improves academic decision-making, student retention, curriculum planning, and institutional performance evaluation. The proposed approach also supports adaptive and intelligent learning environments suitable for modern online and blended education systems. Overall, the project demonstrates that integrating machine learning with educational analytics can significantly improve academic performance prediction, promote proactive educational management, and enhance student success. Future enhancements may include real-time analytics, deep learning optimization, cloud integration, and advanced adaptive learning mechanisms to further improve prediction efficiency and personalized educational support.

References

1. 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.
2. Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009. *Journal of Educational Data Mining*, 1(1), 3–17.
3. Thai-Nghe, N., Drumond, L., Horváth, T., & Schmidt-Thieme, L. (2011). Matrix

factorization methods for predicting student performance. *Educational Data Mining Journal*, 3(1), 1–15.

4. Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72, 414–422.
5. Cortez, P., & Silva, A. (2008). Using data mining to predict secondary school student performance. *Proceedings of Future Business Technology Conference*, 5–12.
6. Huang, S., & Fang, N. (2013). Predicting student academic performance in engineering dynamics courses. *International Journal of Engineering Education*, 29(2), 461–469.
7. Kotsiantis, S. B. (2012). Use of machine learning techniques for educational proposes: A decision support system for forecasting students' grades. *Artificial Intelligence Review*, 37(4), 331–344.
8. Al-Radaideh, Q. A., Al-Shawakfa, E. M., & Al-Najjar, M. I. (2006). Mining student data using decision trees. *International Arab Conference on Information Technology*, 1–6.
9. Pandey, M., & Taruna, S. (2014). Towards the integration of multiple classifier pertaining to the Student's performance prediction. *Perspectives in Science*, 8, 364–366.
10. Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
11. Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
12. Aggarwal, C. C. (2015). *Data mining: The textbook*. Springer.
13. Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
14. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
15. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
16. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., & Thirion, B. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
17. Chollet, F. (2018). *Deep learning with Python*. Manning Publications.
18. Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media.
19. Raschka, S., & Mirjalili, V. (2019). *Python machine learning*. Packt Publishing.
20. Larose, D. T. (2014). *Discovering knowledge in data: An introduction to data mining*. Wiley.
21. Siemens, G., & Baker, R. S. J. D. (2012). Learning analytics and educational data mining. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 252–254.
22. Romero, C., Ventura, S., & García, E. (2008). Data mining in course

management systems. *Computers & Education*, 51(1), 368–384.

23. Pena-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, 41(4), 1432–1462.
24. Kumar, V., & Chadha, A. (2011). Learning from data: Predictive analytics in education. *International Journal of Computer Science Issues*, 9(3), 1–8.
25. Romero, C., Ventura, S., & Pechenizkiy, M. (2010). Handbook of educational data mining. *CRC Press*.
26. Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*, 23(1), 537–553.
27. Dutt, A., Ismail, M. A., & Herawan, T. (2017). A systematic review on educational data mining. *IEEE Access*, 5, 15991–16005.
28. Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). The prediction of students' academic performance using classification data mining techniques. *Applied Mathematical Sciences*, 9(129), 6415–6426.
29. Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure. *International Journal of Computers & Applications*, 39(1), 1–9.
30. Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies*, 13(1), 61–72.