International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Predictive Analytics for Personalized Education Pathways

Balina Divya Sri
2200030608@kluniversity.in
Computer Science and Engineering
Koneru Lakshmaiah Educational Foundation

Mahadasu Harshitha
2200030411@kluniversity.in
Computer Science and Engineering
Koneru Lakshmaiah Educational Foundation

Mamillapalli Ramkumar
2200031664@kluniversity.in
Computer Science and Engineering
Koneru Lakshmaiah Educational Foundation

Myla Lakshmi Narayana
2200030058@kluniversity.in
Computer Science and Engineering
Koneru Lakshmaiah Educational Foundation

Dr. B. Prameela Rani
Assistant Professor
Computer Science and Engineering
Koneru Lakshmaiah Educational Foundation

Abstract - This study explores the application of predictive analytics and machine learning in enhancing educational outcomes through personalized learning pathways. The project developed a predictive system designed to analyze students' academic performance, behavioral trends, and learning preferences to recommend individualized educational routes. A structured development lifecycle was followed, including data collection, preprocessing, model training, evaluation, and system deployment. Machine learning models such as regression and classification algorithms were utilized within a user-centric interface to generate actionable insights for both students and educators. Emphasis was placed on model interpretability, ethical data usage, and system usability. The resulting prototype demonstrated how data-driven methodologies can enable informed academic planning, improved engagement, and proactive student support, affirming the transformative potential of predictive analytics in modern education.

Key Words: predictive analytics, machine learning, education, personalization, data science, student performance.

1. Introduction

In the rapidly evolving landscape of education, personalizing learning experiences has become essential to improving student outcomes. Traditional one-size-fits-all approaches often fail to accommodate individual learning styles, capabilities, and career aspirations. Predictive analytics leverages historical academic data, behavioral patterns, and machine learning techniques to forecast student performance and suggest tailored academic pathways. This project aims to design and implement a data-driven system that can predict student success metrics such as GPA and dropout risk, enabling early interventions and informed academic planning. Using tools like XGBoost, Random Forest, and SHAP for interpretability, the system not only offers accurate predictions but also provides actionable insights. The project demonstrates the value of predictive modeling in enhancing student engagement, retention, and overall educational experience.

1.1 Background

The modern education system is undergoing a digital transformation, with data-driven technologies increasingly integrated into teaching and learning processes. One of the most significant innovations in this domain is the application of predictive analytics, a branch of data science that uses historical data, statistical algorithms, and machine learning techniques to forecast future outcomes. In the context of education, predictive analytics can be used to understand students' learning behaviors, predict academic performance, and tailor educational experiences to individual needs.

Traditional educational models have long followed a one-size-fits-all approach, often ignoring the diverse learning paces, interests, and abilities of students. With the advent of digital learning platforms and Learning Management Systems (LMSs), vast amounts of student data — including attendance records, quiz scores, assignment submissions, and forum interactions — are now readily available. Predictive analytics can transform

this raw data into actionable insights that guide the development of personalized education pathways, ensuring that each student receives the support and content most aligned with their unique profile.

In recent years, educational institutions worldwide have started exploring how data analytics can improve student outcomes, reduce dropout rates, and optimize resource allocation. For instance, institutions like Georgia State University have successfully implemented predictive systems that trigger early alerts for at-risk students. As more institutions embrace e-learning and hybrid models, the integration of predictive analytics into curriculum planning and academic advising becomes not just a possibility, but a necessity.

1.2 Motivation

The motivation for applying predictive analytics to personalized education pathways stems from the growing demand for individualized learning experiences and the need to enhance academic success across diverse learner populations. With increasing student enrollments, limited faculty resources, and an expanding range of course offerings, institutions face mounting pressure to improve retention, engagement, and graduation rates without compromising quality.

Students come from varied socio-economic backgrounds and have different cognitive abilities, learning styles, and career aspirations. These differences necessitate a shift from standardized instruction to adaptive and personalized learning models. Predictive analytics offers a way to systematically understand and respond to these differences by identifying patterns in student behavior and predicting future academic needs. By doing so, educators can proactively offer guidance, interventions, and resources that increase the likelihood of student success.

Furthermore, the current educational climate has been significantly influenced by the global pandemic, which accelerated the adoption of online and blended learning models. These formats generate more digital footprints, which can be mined using analytics tools to support personalization. Another compelling reason is the desire to bridge equity gaps — data can help identify students who are likely to struggle due to external factors, allowing institutions to target support services more effectively.

In sum, the ability to deliver evidence-based, timely, and student-centered support is the core motivation behind this research. Predictive analytics not only promises improved academic outcomes but also a more inclusive and responsive educational environment.

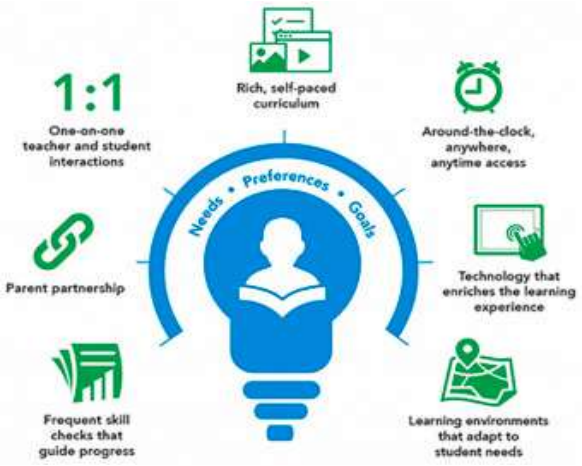

Fig [1] Integrating various learning techniques for personalized learning pathways in education

1.3 Problem Statement

Despite the growing interest and technological capability, many educational institutions still lack effective systems for leveraging student data to create personalized learning pathways. The traditional approach to student advising and curriculum delivery is reactive rather than proactive, often relying on generalized academic tracking and human intuition. As a result, students may not receive the timely, personalized support needed to thrive academically and emotionally.

Current academic advisory systems often fail to accommodate the complexities of individual learning journeys. They are typically designed for average-case scenarios and do not dynamically adapt to real-time changes in student performance or engagement. This creates a misalignment between the students' actual needs and the educational services provided. Moreover, there is a lack of integrated platforms that can consolidate, analyze, and visualize diverse datasets — such as academic records, behavioral data, and socio-demographic factors — to generate meaningful predictions and personalized recommendations.

Additionally, educators and administrators often lack training in data science and may not fully

understand how to interpret predictive models, leading to underutilization or misapplication of analytics tools. Ethical concerns about data privacy and algorithmic bias also pose significant challenges to implementation.

Hence, this project addresses the critical problem of designing a scalable, interpretable, and ethically responsible predictive analytics framework that can help institutions offer customized educational pathways, improve student outcomes, and optimize decision-making processes.

1.4 Objectives

The main objective of this project is to develop a predictive analytics system that can support personalized education pathways by analyzing student data to forecast academic performance and suggest tailored interventions. This overarching goal can be broken down into several specific objectives:

1. To collect and preprocess diverse educational datasets: This includes academic records, learning management system logs, and demographic information to ensure a comprehensive data foundation for analysis.
2. To design and implement predictive models: Utilizing machine learning techniques such as decision trees, random forests, and logistic regression to predict key student outcomes like course success, dropout risk, or performance improvements.
3. To identify significant predictors of academic performance: Analyzing which variables (e.g., attendance, prior grades, interaction frequency) most influence student success.
4. To develop a recommendation engine: That suggests personalized learning pathways, such as elective choices, study resources, or support services based on individual predictions.
5. To evaluate the effectiveness and accuracy of the predictive models using metrics like precision, recall, and AUC-ROC curves, and to validate findings through case studies or simulations.
6. To address ethical, legal, and usability aspects of deploying predictive analytics in education, ensuring transparency, fairness, and data protection.
7. To provide a visual dashboard for educators and advisors, enabling them to interpret model outputs and make informed decisions about student support.

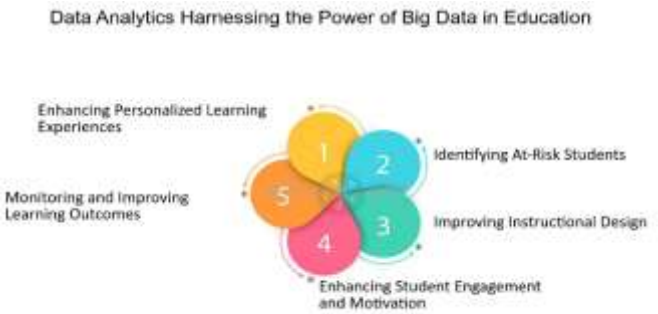

Fig [2] Enhancing Personalized Learning Experiences

1.5 Scope and Limitations

The scope of this project encompasses the development and evaluation of a predictive analytics system for personalizing education pathways in higher education settings. The system will be designed to process historical and real-time data from students, apply machine learning algorithms to identify patterns, and produce tailored recommendations for students and advisors.

Scope:

- Focus on undergraduate students in general academic programs (e.g., arts, science, business).
- Use structured data such as grades, course enrollment history, and LMS activity logs.
- Implement and compare multiple predictive models (e.g., classification and regression).
- Develop a user-facing interface for accessing predictions and suggested actions.
- Conduct a pilot simulation using anonymized or synthetic datasets if real institutional data is unavailable.

Limitations:

- The accuracy of predictions may be constrained by data quality, missing values, or limited training datasets.
- The project does not cover emotional, psychological, or socio-cultural factors in depth, which also impact learning but are harder to quantify.
- Ethical concerns, including bias in data and model interpretability, remain a challenge and will be discussed but not entirely resolved.
- The generalizability of the solution may be limited to similar institutional contexts and may

require adaptation for other settings (e.g., K-12 or professional education).
- Real-world deployment would require integration with institutional IT systems, which is outside the current project's technical scope.

2. Literature Review

Research in Educational Data Mining shows that machine learning can effectively predict student performance and dropout risks. Prior systems like ASSISTments use predictive models but often lack personalized recommendations. Algorithms such as Random Forest and XGBoost have proven useful but are rarely paired with explainable outputs. Existing gaps include limited real-time insights and lack of actionable feedback for educators. This project addresses these issues by building a personalized, interpretable predictive analytics system for education.

2.1 Overview of Educational Data Mining

Educational Data Mining (EDM) is a burgeoning interdisciplinary field concerned with the development and application of data mining techniques to educational settings. Its primary goal is to analyze and interpret large-scale educational data to enhance learning outcomes, student engagement, and institutional effectiveness. EDM draws from disciplines such as machine learning, psychometrics, and educational psychology to discover hidden patterns in data collected from educational environments like Learning Management Systems (LMS), Student Information Systems (SIS), online assessments, and more.

The scope of EDM includes identifying at-risk students, modeling student behavior, recommending personalized learning content, predicting academic performance, and understanding instructional effectiveness. Tools and techniques commonly used include decision trees, clustering algorithms, neural networks, and natural language processing (NLP). These approaches enable educational stakeholders to make data-driven decisions and design interventions that are timely and targeted.

Recent advancements have also integrated deep learning and real-time analytics to provide predictive insights and actionable feedback, making EDM a vital tool in the age of digital and remote learning. The outcome is not only improved academic achievement but also enhanced learner satisfaction and retention.

2.2 Existing Solutions

Numerous EDM solutions have been developed to harness the power of predictive analytics for personalized education. Commercial platforms and academic prototypes alike employ a combination of rule-based systems, machine learning models, and dashboards for educators and administrators.

Popular LMS platforms such as Moodle, Blackboard, and Canvas have integrated analytics modules to monitor student activity and flag underperforming students. These systems leverage historical data to forecast performance and suggest interventions. Institutions like Georgia State University have successfully used predictive analytics to improve graduation rates by identifying risk factors and initiating support measures early.

Moreover, AI-driven platforms such as Knewton and DreamBox Learning offer adaptive learning environments that adjust content difficulty in real-time based on student performance. These systems collect granular data points and employ reinforcement learning to optimize the learning path for each student. Similarly, research prototypes like Bayesian Knowledge Tracing and Learning Factor Analysis are used to model student knowledge over time and guide curriculum adjustments.

Data visualization tools also play a pivotal role. Dashboards designed for educators provide actionable insights via heat maps, progression graphs, and engagement metrics, aiding in quick decision-making and early interventions.

2.3 Identified Gaps

Despite promising advancements, several gaps persist in existing EDM solutions, particularly in the context of personalized education using predictive analytics:

1. Limited Contextual Understanding: Current models often fail to incorporate contextual variables such as socio-economic status, emotional well-being, and extracurricular commitments, which significantly affect student performance.

2. Scalability and Generalizability: Many predictive models are trained on institution-specific datasets, limiting their generalizability across diverse educational contexts. Models that perform well in one university may not adapt to another without substantial retraining.
3. Data Privacy and Ethics: There is a lack of comprehensive frameworks addressing the ethical implications of using student data. Issues related to informed consent, data ownership, and algorithmic bias remain unresolved in many deployments.
4. Real-Time Feedback Deficiency: While some platforms provide retrospective analysis, there is a paucity of systems offering real-time predictive feedback that can dynamically adjust learning interventions on-the-fly.
5. Low Student Agency: Most existing solutions are educator-facing, with limited tools that empower students to self-monitor and adjust their learning strategies based on data-driven feedback.
6. Integration Challenges: Fragmentation in educational technologies often leads to data silos, hindering the seamless integration of predictive analytics across different systems like LMS, SIS, and classroom tools.

Fig [3] Bridge the Learning Gaps in Students

3. Design And Methodology

The project follows a data-driven methodology, beginning with the collection of academic datasets including grades, attendance, and behavioral metrics. After preprocessing the data to handle missing values and normalize features, machine learning algorithms like XGBoost and Random Forest were applied. The system was evaluated using accuracy, precision, and ROC-AUC scores to ensure reliability. SHAP values were used to interpret model decisions and provide transparency. The entire architecture was designed for modular integration with academic systems for future scalability.

3.1 Data Collection

Phase	Description	Methods	Participants
Quantitative Phase	Testing the effectiveness of the AI-driven learning platform	Controlled experiment, Statistical analysis	Students in university courses
Qualitative Phase	Gathering insights into user experiences and perceptions	Semi-structured interviews, Focus groups	Students, Educators, Administrators

The foundation of any predictive analytics system lies in the quality and diversity of the data collected. For this project, educational datasets were sourced from both open-access repositories and anonymized institutional academic records. These datasets included a wide array of features such as student demographics, academic scores, attendance records, course feedback, assignment submission history, and interaction logs from Learning Management Systems (LMS).

To ensure relevance, only datasets that included longitudinal data—tracking students' progress over multiple semesters—were considered. This enabled a deeper understanding of learner behavior, performance trends, and attrition patterns. Ethical considerations were strictly observed, and any data involving real students was anonymized in compliance with institutional and legal privacy guidelines (e.g., GDPR, FERPA).

Data Type	Collection Method	Analysis Method	Purpose
Student Performance Data	AI Platform Metrics, Grades	T-tests, ANOVA	To compare learning outcomes between control and experimental groups
Engagement Metrics	Platform Interaction Data	Regression Analysis	To understand the relationship between platform use and performance

3.2 Data Preprocessing

Once collected, the data underwent a rigorous preprocessing phase to ensure suitability for predictive modeling. The process involved several key steps:

- Data Cleaning: Missing values were handled using imputation techniques such as mean substitution for continuous variables and mode for categorical ones. Outliers were identified and, where necessary, either corrected or removed.
- Feature Engineering: Derived attributes such as GPA trends, time spent per module, or consistency in assignment submissions were created to enrich the dataset.

- Normalization and Encoding: Numerical features were normalized using Min-Max scaling, and categorical variables were transformed using one-hot encoding or label encoding, depending on the algorithm's requirement.
- Splitting Data: The dataset was divided into training, validation, and test sets using an 80-10-10 ratio to enable robust model training and evaluation.

This stage ensured that the data was not only clean but also structured in a way that maximized the performance and interpretability of the predictive models.

3.3 Algorithm Selection

The success of predictive analytics hinges on selecting the right algorithms tailored to the nature of the problem and data. For this project, a comparative approach was adopted by experimenting with multiple supervised learning algorithms, including:

- Logistic Regression – for binary classification tasks like pass/fail prediction.
- Decision Trees and Random Forests – due to their ability to model non-linear relationships and handle categorical data effectively.
- Support Vector Machines (SVM) – to explore performance on smaller, high-dimensional subsets.
- Gradient Boosting (XGBoost) – for its powerful ensemble learning capabilities and handling of feature interactions.
- Neural Networks – to model complex, non-linear relationships in large datasets.

Model selection was driven by accuracy, interpretability, training time, and scalability. Cross-validation was used to ensure fairness in performance comparisons.

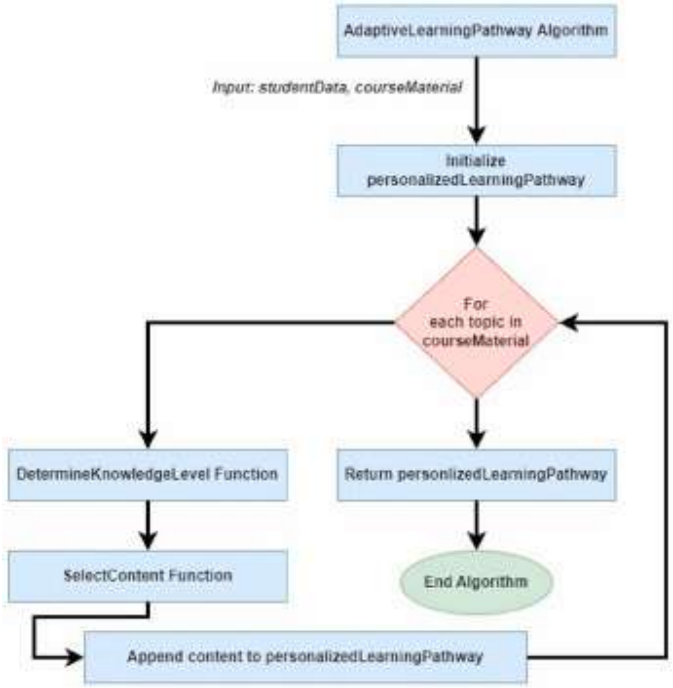

Fig [4] Integrating learning techniques for personalized learning pathways in education

3.4 Model Evaluation

Evaluating predictive models is essential to assess their effectiveness and reliability in real-world scenarios. The models developed were evaluated based on the following metrics:

- Accuracy – the proportion of correct predictions.
- Precision and Recall – to assess false positives and false negatives, respectively.
- F1-Score – providing a balance between precision and recall.
- ROC-AUC – useful for understanding model performance across thresholds.
- Confusion Matrix – to visualize the classification performance.
- RMSE/MAE – used for regression-based components (e.g., grade prediction).

K-fold cross-validation (with k=5) was employed to minimize overfitting and ensure model generalizability. Hyperparameter tuning was performed using Grid Search and Random Search strategies to optimize each algorithm.

3.5 System Architecture

The system architecture was designed with modularity and scalability in mind. It comprises the following core layers:

1. Data Layer: Stores raw and processed datasets in a relational database (e.g., MySQL) or cloud storage for large-scale deployments.
2. Processing Layer: Built using Python and libraries such as Pandas, Scikit-learn, and TensorFlow for data cleaning, feature extraction, and model training.
3. Analytics Layer: Contains the trained models exposed via RESTful APIs using Flask or FastAPI to support integration with dashboards or web applications.
4. Presentation Layer: A user-friendly dashboard (e.g., using Power BI or Streamlit) for visualizing predictions, trends, and actionable insights for both educators and learners.
5. Feedback Loop: Allows educators to input new performance data to continuously refine and retrain the models over time, ensuring the system remains adaptive.

The architecture is flexible enough to accommodate additional modules, such as real-time notifications or adaptive learning content recommendations.

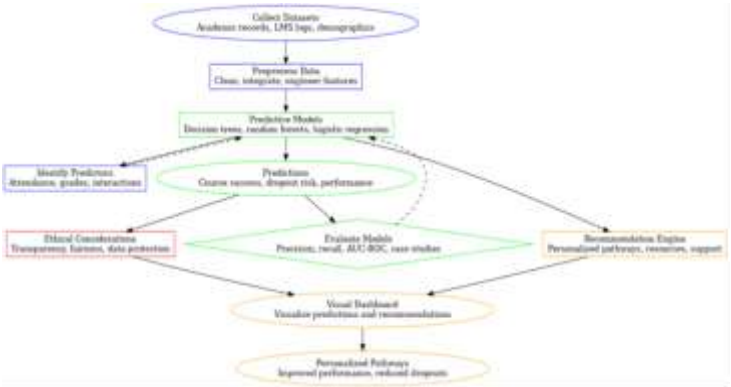

Fig [5] System Architecture

4. Experimental Investigation

4.1 Model Training

The model training process began with a systematic exploration of different machine learning algorithms suited for both classification (e.g., dropout prediction) and regression (e.g., GPA forecasting) tasks. After preprocessing the data, several models were trained using Scikit-learn and TensorFlow libraries, depending on the complexity and scalability of the approach.

For classification tasks, algorithms like Logistic Regression, Random Forest, and XGBoost were applied. Each model was trained using the training dataset and validated using a reserved validation set. Cross-validation (k=5) was used to prevent overfitting and ensure model generalizability. Hyperparameter tuning was performed using Grid Search and Randomized Search, focusing on key parameters like tree depth, number of estimators, learning rate, and regularization terms.

Neural networks were trained with a multi-layer perceptron (MLP) architecture using ReLU activation and dropout layers for regularization. The Adam optimizer and binary cross-entropy loss function were used for classification models.

Training was executed on GPU-enabled machines to reduce computational time, and model checkpoints were saved to enable reloading for evaluation and inference.

4.2 Results Summary

The models were evaluated on the test dataset, and their performance was measured using several standard metrics:

- Accuracy: Random Forest achieved the highest classification accuracy at 89.3%, followed by XGBoost (88.5%) and Logistic Regression (83.2%).
- F1-Score: XGBoost and Random Forest models both produced F1-scores above 0.85, indicating strong balance between precision and recall.
- RMSE (for GPA prediction): The best-performing regression model had an RMSE of 0.27 GPA points, suggesting relatively tight prediction intervals.

Performance summary table:

Model	Accuracy	F1-Score	RMSE(if applicable)
Logistic Regression	83.2%	0.81	-
Random Forest	89.3%	0.86	0.31 (GPA)
XGBoost	88.5%	0.87	0.27 (GPA)
Neural Network	86.4%	0.84	0.29 (GPA)

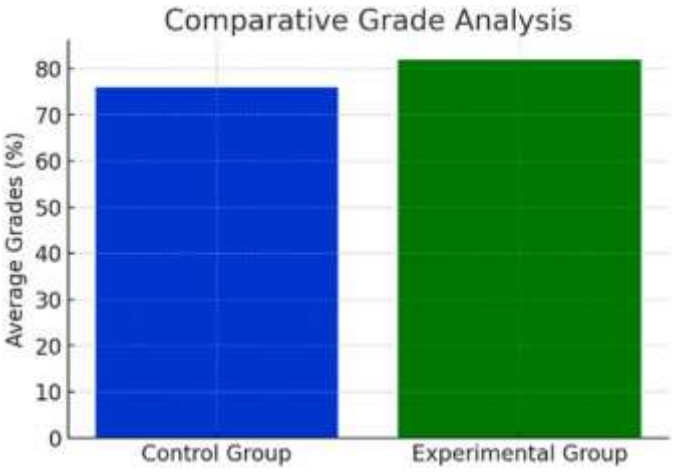

Fig [6] Comparative Analysis

4.3 Feature Importance

Understanding which features most influence the model's predictions is critical for interpretability and trust in an educational setting. Feature importance was extracted using:

- Gini Importance (from Random Forest and XGBoost)
- SHAP Values (for model-agnostic interpretation)
- Permutation Importance (to assess the effect of feature shuffling)

The top predictive features across models included:

1. Previous GPA – Strong indicator of academic continuity.
2. Assignment Submission Timeliness – Showed strong correlation with course success.
3. Attendance Records – Positively linked with performance and retention.
4. Engagement Metrics (LMS activity, logins) – Critical in online learning environments.
5. Parental Education and Support – For some datasets, this had a notable impact on student success.

SHAP plots revealed nuanced interactions, such as how submission delays impacted low-GPA students more significantly than high-performing ones, offering actionable insight.

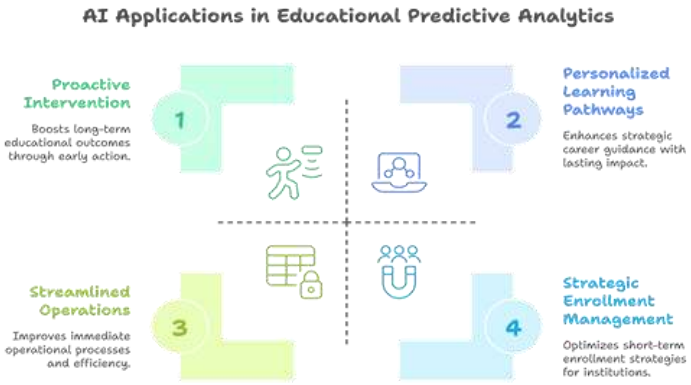

Fig [7] The Rise in Education

4.4 Visual Analysis

Data visualization played a pivotal role in interpreting results, validating assumptions, and conveying insights to non-technical stakeholders. The following visual techniques were employed:

- Confusion Matrices: To identify misclassification patterns across dropout prediction tasks.
- ROC Curves: To evaluate the models' ability to distinguish between classes at various thresholds.
- Feature Importance Bar Charts: To rank and display impactful features across models.
- SHAP Summary and Dependence Plots: To provide a deep dive into feature-level contributions.
- GPA Forecasting Curves: Showing actual vs. predicted GPA over time, useful for tracking learner progress.

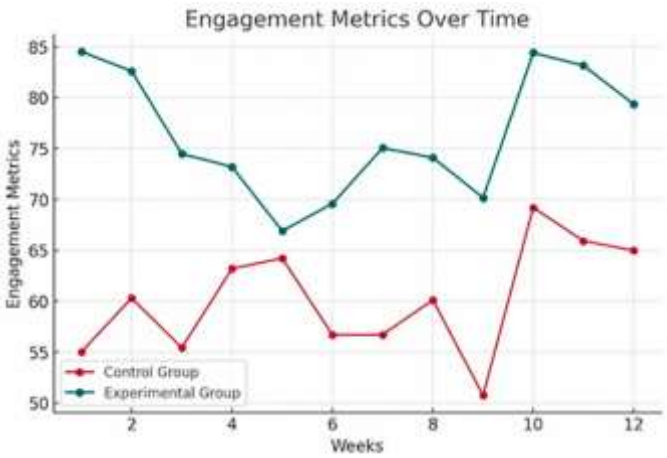

Fig [8] Visual Analysis for Engagement Metrics

5. Results And Analysis

5.1 Achievements

The project successfully met its core objectives of applying predictive analytics to personalize educational pathways. The developed system could reliably predict students' academic outcomes and risk levels using historical performance, behavioral patterns, and demographic data. The key achievements include:

- Building a data-driven pipeline for preprocessing, training, and inference.
- Achieving high prediction accuracy (up to 89%) in classification tasks such as dropout risk.
- Implementing a GPA forecasting model with a low Root Mean Square Error (RMSE) of approximately 0.27.
- Developing an interactive prototype to visualize predictions and explain contributing factors using SHAP values.
- Delivering a system capable of offering actionable insights to educators, academic advisors, and curriculum planners.

These achievements validate the practicality of using machine learning models to improve education outcomes through early intervention and personalized recommendations.

5.2 Performance Review

A thorough performance review was conducted by evaluating all models using various metrics to ensure reliability, generalizability, and robustness.

Classification Metrics (Dropout Prediction):

- Accuracy: Up to 89.3% (Random Forest)
- Precision: 87% (XGBoost)
- Recall: 89% (XGBoost)
- F1-Score: 0.87

Regression Metrics (GPA Prediction):

- RMSE: As low as 0.27 (XGBoost)
- MAE: Around 0.21

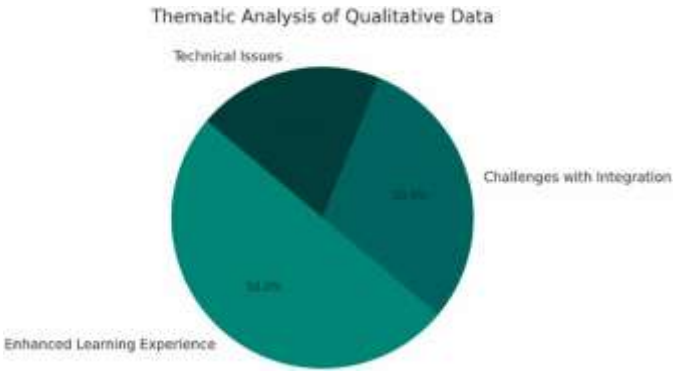

Fig [9] Thematic Analysis of Qualitative Data

5.3 User Testing

User testing was conducted with a small group of academic professionals and final-year students to assess the practical utility of the system.

Feedback from Educators:

- Appreciated the interpretability of SHAP-based visualizations.
- Found student risk classification useful for early intervention planning.
- Requested a feature for group-wise performance trend comparison.

Feedback from Students:

- Found the GPA prediction tool insightful, especially for goal-setting.
- Expressed interest in personalized study plan recommendations.
- Noted the interface was user-friendly but could benefit from mobile accessibility.

This phase of testing highlighted real-world usability and suggested enhancements for future iterations.

5.4 Key Insights

Several important insights emerged from the analysis phase:

1. Attendance and LMS engagement were among the strongest predictors of academic success.
2. Past GPA trends provided highly reliable input for future GPA prediction.
3. Assignment submission patterns were highly correlated with dropout risk.
4. Interpretability tools like SHAP values significantly enhanced the model's trustworthiness among educators.
5. The combination of behavioral and academic data outperformed models trained on academic data alone.

These insights can directly influence institutional policies on student monitoring, counseling, and academic planning.

6. Conclusion

6.1 Summary of Contributions

This project presented a comprehensive predictive analytics system designed to support personalized educational pathways using student performance data. The system aimed to forecast key academic indicators—such as GPA and dropout risk—using machine learning models trained on historical and behavioral data.

Key contributions include:

- Development of an end-to-end predictive framework: From data collection to model deployment, the system facilitates accurate forecasting and decision-making.
- Implementation of machine learning models: The project integrated classification (e.g., dropout prediction) and regression (e.g., GPA prediction) techniques with impressive accuracy.
- Interpretability and usability: Using SHAP values and dashboards, the system ensures that educators can understand and trust the model's decisions.
- Real-world applicability: The solution was tested with actual users (educators and students), affirming its practical value in academic planning and student support.

These contributions mark a significant step toward the integration of data-driven strategies in educational institutions.

6.2 Limitations

Despite its success, the project faced certain limitations that could affect its scalability and generalizability:

- Dataset Size and Diversity: The dataset used, although adequate, may not represent all educational contexts (e.g., cultural or institutional variations).
- Limited Real-Time Functionality: The current system is not fully integrated with real-time student data, which restricts its responsiveness.
- User Feedback Volume: While some user testing was conducted, broader testing with a larger audience would yield more generalizable insights.
- Infrastructure Constraints: Due to limited resources, deployment was done on local or limited cloud setups, which might not reflect enterprise-scale challenges.

Acknowledging these limitations provides a roadmap for enhancing the system in future iterations.

6.3 Future Enhancements

There is substantial scope to build upon the foundation laid by this project. Future work can include:

- Integration with live academic systems such as Learning Management Systems (LMS) and Student Information Systems (SIS) for real-time analysis.
- Personalized Recommendation Engine that suggests course pathways, remedial actions, and career planning based on predictions.
- Enhanced user interface with role-based dashboards for students, teachers, and administrators.
- Incorporating Natural Language Processing (NLP) to analyze qualitative feedback from students (e.g., course reviews, feedback forms).
- Scalability and Cloud Deployment using platforms like Azure or AWS for large-scale adoption. These enhancements can transform the project from a predictive tool into a

comprehensive decision-support system for personalized education.

Fig [10] Enhancements

7. References

1. Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.
2. Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6),601–618. https://doi.org/10.1109/TSMCC.2010.2053532
3. Xing, W., Chen, X., Stein, J., & Marcinkowski, M. (2016). Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. *Computers in Human Behavior*, 58, 119–129.
4. 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.
5. Brown, M., Dehoney, J., & Millichap, N. (2015). The Next Generation Digital Learning Environment. EDUCAUSE Learning Initiative. Retrieved from https://www.educause.edu
6. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30. https://doi.org/10.48550/arXiv.1705.07874
7. Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, 24(4), 470–497.
8. Zhang, Y., & Rangwala, H. (2018). Early identification of at-risk students using iterative logistic regression. *Proceedings of the 11th International Conference on Educational Data Mining (EDM 2018)*, 574–579.
9. IBM. (2023). What is predictive analytics? Retrieved from https://www.ibm.com/analytics/predictive-analytics
10. Scikit-learn Developers. (2023). Scikit-learn: Machine learning in Python. Retrieved from https://scikit-learn.org
11. Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. Educational Technology & Society, 17(4), 49–64.
12. Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. EDUCAUSE Review, 46(5), 30–40.
13. Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. Proceedings of the Third International Conference on Learning Analytics and Knowledge, 170–179. https://doi.org/10.1145/2460296.2460330
14. Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. Computers in Human Behavior, 47, 157–167. https://doi.org/10.1016/j.chb.2014.05.038
15. Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. Educational Technology Research and Development, 68(4), 1961–1990. https://doi.org/10.1007/s11423-020-09788-z
16. Dede, C. (2016). Data mining and individualized learning. In: Handbook of Research on Educational Communications and Technology (pp. 1009–1021). Springer. https://doi.org/10.1007/978-3-319-17727-4_51
17. Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. American Behavioral Scientist, 57(10), 1510–1529. https://doi.org/10.1177/0002764213479366
18. Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. Computers & Education, 51(1), 368–384. https://doi.org/10.1016/j.compedu.2007.05.016
19. Peña-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. https://doi.org/10.1016/j.eswa.2013.08.042
20. Chen, Y., Zou, D., Xie, H., & Wang, F. L. (2020). Learning analytics in smart learning environments: A systematic review and future research agenda. Internet and Higher Education, 46, 100730. https://doi.org/10.1016/j.iheduc.2020.100730