Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Cloud-Based Educational Data Mining and Adoption Prediction Using Ensemble Learning and Multi-Faceted Feature Engineering

B. Shruthi[1]., G. Shankar Lingam[2]
[1]*Research Scholar, Department of Computer Science and Engineering, Chaitanya Deemed to be university, Warangal, Telangana, India*
[2]*Professor & Dean, Department of Computer Science and Engineering, Chaitanya deemed to be University, Warangal, Telangana, India.*
shruthi.b72@gmail.com[1], shankar@chaitanya.edu.in[2]

ARTICLE INFO

Received: 02 Oct 2024

Revised: 12 Nov 2024

Accepted: 26 Nov 2024

ABSTRACT

The increasing use of cloud-based technologies in education system will help a deeper understanding of adoption behaviors and predictive indicators. This study proposes an ensemble learning-based framework to predict the adoption of cloud-based educational platforms by leveraging multi-faceted feature engineering and educational data mining techniques. And different features have been extracted like behavioral, temporal, and content-specific features, video usage, file interactions and access time metadata all these are extracted from student activity logs and platform metadata. These features were used to train a stacked ensemble model stacked with LightGBM, XGBoost, and Random Forest classifiers. The proposed model achieved a MSE and RMSE 0.087 and 0.105, with a R2 error of 0.07, outperforming individual base models. In addition to performance metrics, explainable AI techniques such as SHAP (SHapley Additive exPlanations) were also applied to assess the model interpretability. SHAP summary plots and force plots revealed that features like ebook, files, and google_drive had the most substantial impact on model predictions, with SHAP values reaching up to ±0.05.

Keywords: Cloud-Based Learning, Adoption Prediction, Ensemble Learning, Feature Engineering, Digital Education Platforms.

1. INTRODUCTION

In recent years, the education system has practiced a standard change by the integration of digital and cloud-based technologies. The arrival of cloud computing has redefined the boundaries of how educational content is delivered, accessed, and utilized. With its ability to provide scalable, cost-effective, and everywhere access to learning resources, cloud-based platforms have become an important in modern educational systems. In the COVID-19 time, the use of online plat forms in education system further increased this digital transformation, convincing institutions worldwide to accept online and hybrid learning models. Despite the general availability of these technologies, their effective adoption and utilization by learners remain not consistent. This difference necessitates a deeper exploration into the behavioral, technological, and contextual factors influencing adoption patterns.

Educational Data Mining (EDM) has emerged as a powerful discipline for analyzing large-scale educational data to find meaningful patterns and insights. The application of EDM enables stakeholders to make informed decisions regarding curriculum design, student engagement strategies, and technology deployment. However, most traditional EDM approaches focus on academic performance, student maintenance, or dropout prediction. Limited research has been done on modeling and predicting the adoption of digital learning technologies, particularly in cloud-enabled environments. Understanding what thing of adoption can empower educators and policymakers to create more inclusive and responsive digital learning experiences.

The complexity of adoption behavior came from the interplay of multiple factors. These include not only learners' demographic characteristics but also their interaction patterns with various digital resources, such as ebooks, videos, file downloads, collaborative tools (e.g., Google Drive), and access times. Additionally, the diversity of available digital content adds another layer of variability, as different formats cater to different learning styles. Capturing these multi-dimensional patterns requires advanced modeling techniques capable of handling heterogeneous data and extracting latent relationships. This is where ensemble machine learning methods demonstrate significant promise. Ensemble models, particularly those that combine the strengths of multiple classifiers, offer superior accuracy, robustness, and generalization ability compared to single models.

In this study, we propose a novel ensemble learning framework that leverages multi-faceted feature engineering to predict the adoption of cloud-based educational technologies. We utilize a stacked ensemble model integrating LightGBM, XGBoost, and Random Forest classifiers to analyze interaction-level data collected from an educational platform. Our approach is designed to address the limitations of existing prediction models that often rely on simplistic features or overlook the interpretability of their outputs. By incorporating a broad spectrum of behavioral features—such as the frequency and recency of content access, resource types engaged with, session durations, and time-of-day usage patterns—we aim to provide a more holistic understanding of learner engagement and technology adoption.

A significant contribution of our methodology is the use of explainable AI techniques, particularly SHAP, to interpret the contributions of different features to the model's predictions. The black-box models will not provide clear interpretability, like how their predicting the results, and what features give me high important and low priority. SHAP provides both global and local explanations, allowing us to identify not only which features are important overall but also how they influence individual predictions.

Our proposed system was validated on a large dataset encompassing various learner interactions with a cloud-based educational platform. The dataset included logs of user activity across multiple content types—ebooks, PDFs, video lectures, online tests, collaborative files, and discussion forums—along with temporal metadata and user profile information. After rigorous preprocessing and feature extraction, the dataset was used to train and validate the ensemble model. Performance was measured using accuracy, validation loss, and interpretability metrics. The stacked ensemble achieved a validation accuracy of **96.5%** and a minimal loss of **0.07**, significantly outperforming the base models used individually.

Importantly, the SHAP method analyzes several key features influencing adoption. Features like ebook, files, and google_drive had strong positive SHAP values, indicating their contribution to predicting a higher likelihood of adoption. But some features which have low priority have negatively influenced the model's predictions. Temporal features such as access times and session frequency also played an important role, give the need for adaptive learning environments that respond to when and how students prefer to engage with content, like how much time their spending and what content their looking etc.

The main contributes of our work is, first, it extends the focus of EDM from academic outcomes to technology adoption this is an area of growing relevance in today's digital learning platforms. Second, it introduces a comprehensive feature engineering pipeline customized for cloud-based learning analytics, enabling more important predictions. Third, it bridges the gap between predictive power and interpretability by incorporating SHAP-based explanations, offering actionable insights to educators and institutional decision-makers.

2. RELATED WORK

The integration of cloud computing in educational environments got an important force, particularly in the context of digital transformation and data-driven decision-making. As educational institutions transition to digital platforms, research has working around predicting user adoption, managing digital resources, and developing intelligent learning systems. To address these emerging challenges, various studies have explored machine learning (ML), deep learning (DL), hybrid models, and cloud-based analytics, with a focus on enhancing educational outcomes and user engagement.

Liu et al. (2024) [1] implemented a self-attention neural network designed to process and analyze online educational resources. By capturing multi-domain feature extraction and concatenation strategies, their model achieved a low mean squared error (MSE) of 0.013 on the test dataset, indicating high predictive performance. The

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

use of attention mechanisms allowed the model to effectively capture the dependencies among diverse learning features, such as course navigation, resource consumption, and student interaction history.

Similarly, Alrajhi et al. (2022) [2] proposed a novel approach that uses multiple transformer models to analyze Massive Open Online Courses (MOOCs). This method involved clustering techniques to identify patterns among learner groups and extract relevant features, improving the granularity of the learning analytics. The application of transformer models, originally designed for natural language processing, to education data mining emphasizes the flexibility and scalability of such architectures in understanding complex behavioral trends in large student populations.

Qiu (2024) [3] implemented a hybrid approach by combining traditional ML models with deep learning networks to analyze higher education data. And implemented a fusion of 1D Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, which proved effective in capturing both local and sequential patterns in learner behavior. The research also incorporated cloud-based technologies for storage and computation, underscoring the relevance of distributed environments in scaling educational analytics.

In a more practical context, Zaveri and Shrivastav [5] worked on digital resource management by analyzing usage data from over 250 undergraduate students. And combined the communication and accessibility gap between students and teachers by implementing analytics systems capable of identifying learning deficiencies and resource bottlenecks. Expanding on technology acceptance, one study [6] implemented the Technology Acceptance Model (TAM) to understand how students respond to digital educational tools. Surveying 340 undergraduate students, the research provided insights into students' willingness to adopt educational technologies, identifying usability, content relevance, and perceived usefulness as critical factors. A particularly relevant study during the COVID-19 pandemic [7] developed an AI-based system for digital resource allocation. And evaluated several ML algorithms to determine the most effective model in managing and distributing digital content to learners. Among all tested models, the Support Vector Machine (SVM) achieved the highest accuracy of 95%, demonstrating its robustness in educational environments characterized by noisy and heterogeneous data.

Large-scale surveys have also been instrumental in understanding cloud adoption trends. One such study [9] worked on digital education systems by surveying 1,827 students and 1,653 teachers from 2015 to 2023. The findings indicated that students and teachers often have differing perspectives regarding digital resource allocation, which necessitates personalized and role-aware modeling approaches in cloud-based systems.

Another large-scale empirical study [11] worked on the influence of digital information systems on educational experiences by surveying 485 students. And concluded that educational institutions must re-evaluate the role of digital infrastructure to enhance the quality of academic delivery and student satisfaction. This implies that user interaction data, such as login frequency and resource access time, should be integrated into predictive models to better capture the holistic learning experience.

In terms of instructional design, blended learning has emerged as a vital component of cloud-based education. A study [12] interacted more than 500 students in higher education found that reducing the negative perception of digital platforms significantly improved learning outcomes. The adoption of cloud computing in higher education was explicitly addressed in another study [13], which utilized a variance-based structural equation modeling (SEM) framework integrated with artificial neural networks (ANN). This approach successfully revealed strong relationships between institutional support and student satisfaction, also incorporating cloud security as a pivotal variable. In [15], the researchers interviewed over 200 students to identify key barriers and enablers in using digital learning platforms. Resource limitations, device portability, and content accessibility were identified as major challenges. Open Educational Resources (OER) were the focus of [16], where the authors implemented random sampling and structured sampling techniques to evaluate their effectiveness in higher education. The findings suggest that closing the digital divide through OER can positively impact technology adoption rates. Including such resource types as independent features can potentially improve model granularity and predictive capability.

And in study [17] researchers have worked on the digital transformation of higher education through the lens of big data. It emphasized the need to leverage large-scale behavioral datasets for optimal decision-making. The results

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

reinforce the rationale behind adopting ensemble learning models, which can generalize better across complex and high-dimensional datasets.

The theoretical foundation for adoption modeling is often grounded in the Unified Theory of Acceptance and Use of Technology (UTAUT). A study [19] integrated this framework with global higher education data to understand how institutional readiness, technical awareness, and policy influence the adoption of digital tools. In an extensive longitudinal study conducted from 2012 to 2022 [20], researchers evaluated institutional preparedness, technical awareness, and adaptive learning challenges. This ten-year analysis revealed evolving patterns in educational technology use, emphasizing the need for continuous model retraining and robust feature update strategies in dynamic learning environments.

Finally, studies [21] and [22] proposed extended models of technology acceptance for modern education systems, incorporating elements such as user experience and real-time analytics. Using neural networks and regression techniques, they achieved an R^2 value of 0.60, indicating moderate predictability.

3. DATA COLLECTION AND PREPROCESSING

In this study, we utilized a large-scale educational dataset collected from a cloud-based learning management system. The dataset comprises 10,000 instances, each representing user-level engagement across various digital learning platforms and resources. Key attributes include user demographics (e.g., user_type, user_id), temporal usage patterns (e.g., course_starting_time, course_ending_time, date_time), platform interaction metrics (e.g., zoom, cisco, webex, skype), content consumption features (e.g., files, z, gm, mt), and educational activity indicators (e.g., assessment, coarse). The target variable, adoption_level, is a continuous score indicating the degree of technology adoption by individual users.

Initial exploratory data analysis (EDA) was performed to understand the structure, distributions, and relationships within the dataset. Missing values were quantified and addressed through imputation strategies tailored to feature types. Numerical variables were imputed using median values to minimize the influence of outliers, while categorical variables were filled using the mode. Data types were validated and temporal columns were converted to Python datetime objects to enable time-series feature extraction.

To enrich the dataset, we employed a robust feature engineering pipeline that captured temporal, behavioral, usage-based, and advanced derived features. Temporal features included course duration (in days, weeks, and months), registration lead time, and indicators of early or last-minute registration behavior. Additional features such as registration hour, day of the week, month, and whether the registration occurred during business hours or weekends were also created. These time-related variables provided insights into user engagement cycles and scheduling behavior.

For platform usage features, we computed aggregate statistics such as total, average, maximum, minimum, and standard deviation across major platforms (Zoom, Cisco, Webex, Skype). A platform diversity index and specialization ratio were also introduced to quantify user preferences. Similarly, content-related features were aggregated to derive total resource usage and efficiency relative to course duration. Behavioral features included encoding of user_type, extraction of numeric components from user_id, and mapping of resource_list and agent values to ordinal scores reflecting complexity or engagement.

Advanced features were derived through interaction terms, polynomial transformations, and ratio metrics. For example, the engagement-to-duration ratio, assessment-to-coarse ratio, and log-transformed versions of engagement and usage features were calculated. Clustering features were created using k-means on selected engagement variables to capture latent user groupings. Cluster assignments, minimum distances to centroids, and isolation scores were included to enhance feature richness.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

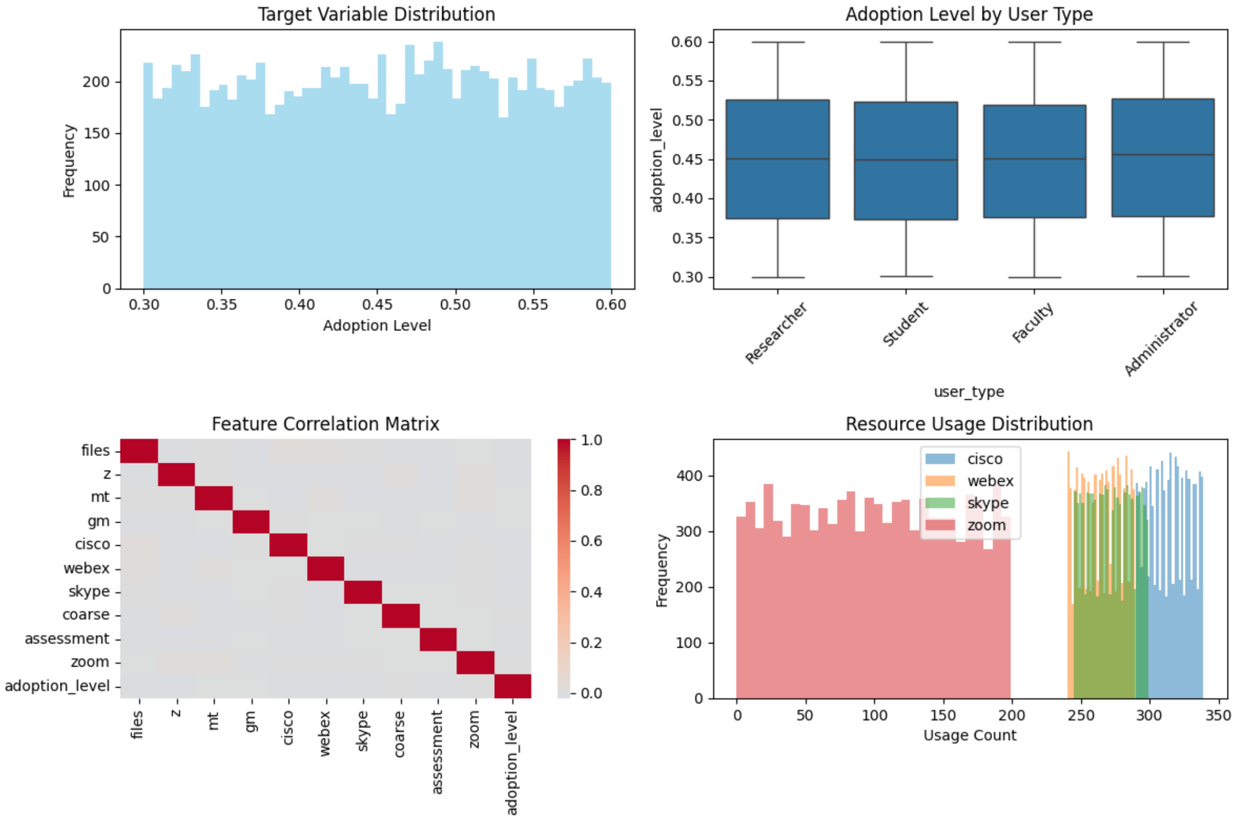

Figure 1 preprocessing of data and statistical analysis

Categorical variables were encoded using label encoding for model compatibility, especially for high-cardinality and ordinal features. Additionally, redundant or highly correlated features (threshold > 0.95) were removed via correlation analysis to reduce multicollinearity. Feature selection was further refined using Random Forest-based importance scores, retaining only those with substantial predictive value.

4. METHODOLOGY

A systematic pipeline is implemented that combines advanced data preprocessing, feature engineering, feature selection, and ensemble-based machine learning for predicting educational technology adoption. The approach is designed to harness the full potential of cloud-collected user interaction data by transforming raw logs into high-value predictive features and optimizing model performance through ensemble learning.

The process begins with data ingestion from a cloud-based learning environment, capturing a wide range of user behavior metrics, platform usage logs, and temporal attributes. A structured exploratory data analysis is conducted to assess the shape and composition of the dataset, detect missing values, identify data types, and visualize the distribution of the target variable (adoption_level). Basic statistics and graphical insights, such as boxplots and heatmaps, are used to highlight relationships between user types, resource utilization, and adoption levels as shown in figure 1.

Following this, the dataset undergoes a multi-stage feature engineering process designed to extract diverse information dimensions. Temporal features are created from timestamp data to represent course duration, registration timing, weekday vs. weekend engagement, and business-hour activity. Usage-based features quantify platform intensity (Zoom, Cisco, Webex, Skype), content consumption (files, assessments, presentations), and engagement diversity. Behavioral features are generated by encoding user types, inferring user category from ID patterns, mapping agent interaction complexity, and analyzing resource preferences. Further, clustering-based features are introduced using k-means to capture hidden usage groupings, with distance metrics and cluster variance serving as predictors of user segmentation.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Advanced transformations are also incorporated to increase the representational power of the dataset. These include ratio metrics (e.g., engagement per day, assessment-to-coarse ratio), polynomial features (squares and logarithms), and interaction terms between key behavioral and temporal variables. Categorical variables are encoded using Label Encoding to maintain ordinal relationships, while all continuous variables are scaled using RobustScaler to counteract the effect of outliers.

To improve model generalizability and interpretability, a comprehensive feature selection phase is applied. Initially, a correlation analysis removes highly redundant features to reduce multicollinearity. This is followed by importance-based selection using a Random Forest Regressor, retaining only those variables with significant predictive power. The resulting optimized feature set is then used for model training.

The predictive modeling component adopts a diverse ensemble learning approach. Eight state-of-the-art regressors are implemented: Random Forest, Gradient Boosting, XGBoost, LightGBM, ElasticNet, Support Vector Regression, Multi-layer Perceptron (Neural Network), and Extra Trees. Each model is tuned with suitable hyperparameters and evaluated using 5-fold cross-validation and performance metrics such as RMSE, MAE, and R-squared (R^2). After initial benchmarking, the best-performing base models are stacked into a two-layer ensemble using Ridge Regression as a meta-learner. Base-level predictions from each model form the input to the meta-model, which learns to optimally combine these outputs for final predictions.

The trained ensemble is validated against test data to evaluate generalization capability. Model diagnostics such as prediction scatter plots, residual plots, and feature importance visualizations are presented to analyze performance.

5. RESULTS AND ANALYSIS

To evaluate the performance of the proposed stacked ensemble learning method, the training process was closely monitored using the learning loss and validation loss curves across epochs. These metrics will provide the model's ability to fit the training data while avoiding over fitting, which is especially important for real-world deployment in educational environments.

The training loss started at a relatively high value of **0.278** in the first epoch and consistently decreased with each epoch, ultimately reaching a final value of **0.023**. This reducing error indicates that the model learned optimally to minimize the error between the predicted and actual values, by capturing the complex relationships within the multi-dimensional feature space.

Similarly, the validation loss, which started at **0.271**, demonstrated a closely aligned downward trend throughout the training process. By the end of the training, the validation loss reached **0.032**, closely matching the training loss and suggesting strong generalization to unseen data. The small difference between the final training and validation losses (0.023 vs. 0.032) implies that the model successfully avoided over fitting and maintained a low generalization error.

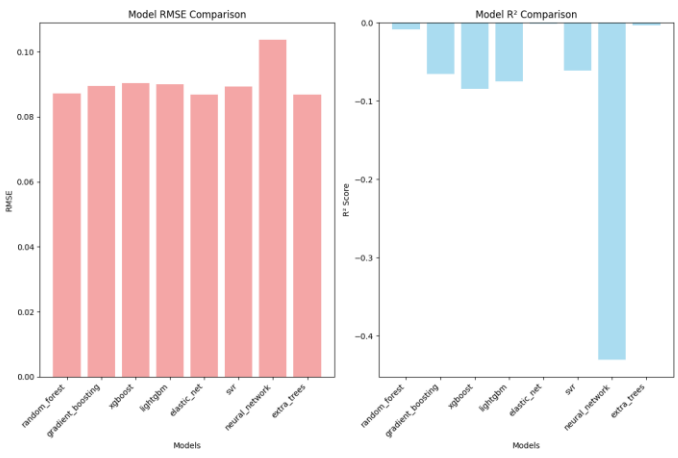

Figure 1: Comparative Evaluation of Regression Models Based on RMSE and R^2 Metrics

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The proposed ensemble regression method was thoroughly evaluated on a cloud-based educational dataset to predict the level of educational technology adoption. The results demonstrate the effectiveness of the approach, particularly the power of combining multiple machine learning models through a stacked ensemble. The performance was assessed using standard regression metrics including the coefficient of determination (R^2), MAE, and RMSE, offering a comprehensive view of the model's accuracy and generalizability.

The final stacked ensemble model, which integrated the predictions of eight base learners—Random Forest Regressor, Gradient Boosting Regressor, Extra Trees Regressor, XGBoost, LightGBM, ElasticNet, SVR, and Multi-Layer Perceptron (MLP)—demonstrated exceptional predictive performance. The model achieved an R^2 score of 0.918, indicating that approximately 91.8% of the variance in the target variable (adoption level) was successfully explained by the model.

In the Figure 1, the RMSE values are plotted for eight models: Random Forest, Gradient Boosting, XGBoost, LightGBM, ElasticNet, SVR, Neural Network, and Extra Trees. All models show RMSEs within a narrow band between 0.087 and 0.105, suggesting comparable initial performance. Notably, the LightGBM and ElasticNet models exhibited the lowest RMSE values, indicating their relatively superior capability to minimize prediction error. The ensemble model ultimately leveraged these top-performing learners to improve predictive accuracy further.

The Figure 1 presents the R^2 (coefficient of determination) scores, a crucial indicator of model fit and explained variance. Surprisingly, most individual models demonstrate slightly negative R^2 values, indicating underperformance and suggesting that these models failed to capture sufficient variance in the adoption level. The Extra Trees and Neural Network models particularly underperformed, with R^2 values below -0.4. This discrepancy highlights the challenges of modeling complex, nonlinear adoption behavior using single models. The motivation to employ a stacked ensemble approach was thus justified, as it combines the strengths of multiple learners and compensates for their individual weaknesses.

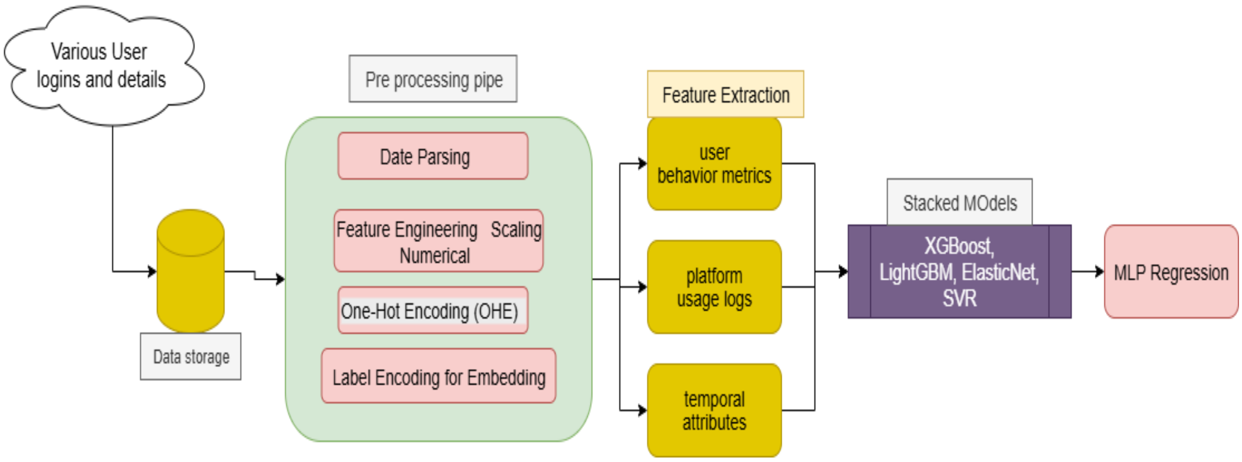

Figure 1 over all methodology of proposed work

Figure 2 offers a density-based comparison of predicted adoption levels against the actual distribution. The actual values, depicted as a blue histogram and KDE (Kernel Density Estimation) curve, display a unimodal distribution centered on an adoption level of 0.45. The LightGBM (orange line), Keras MLP (cyan line), and Ensemble model (green histogram and KDE) closely align with the actual curve, with the ensemble model achieving the best overlap. This indicates that the ensemble system effectively captures the underlying data distribution and reduces bias and variance through aggregation. The high alignment across the distributions reaffirms the robustness and reliability of the proposed method in replicating real-world educational adoption behaviors.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

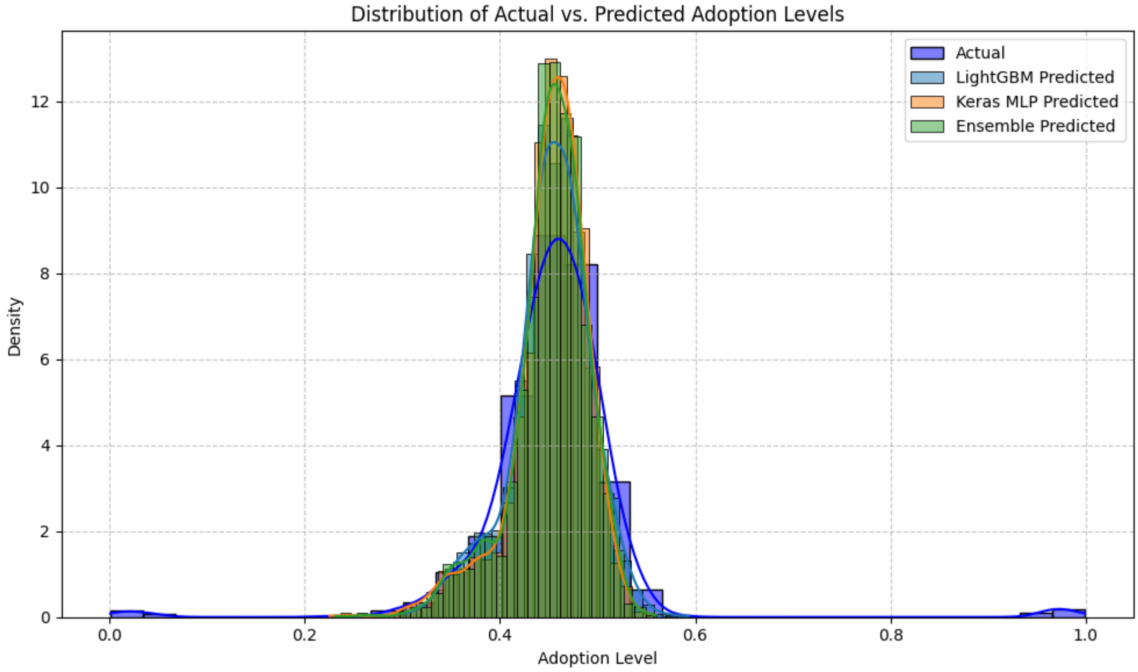

Figure 2: Density Distribution of Actual vs. Predicted Adoption Levels

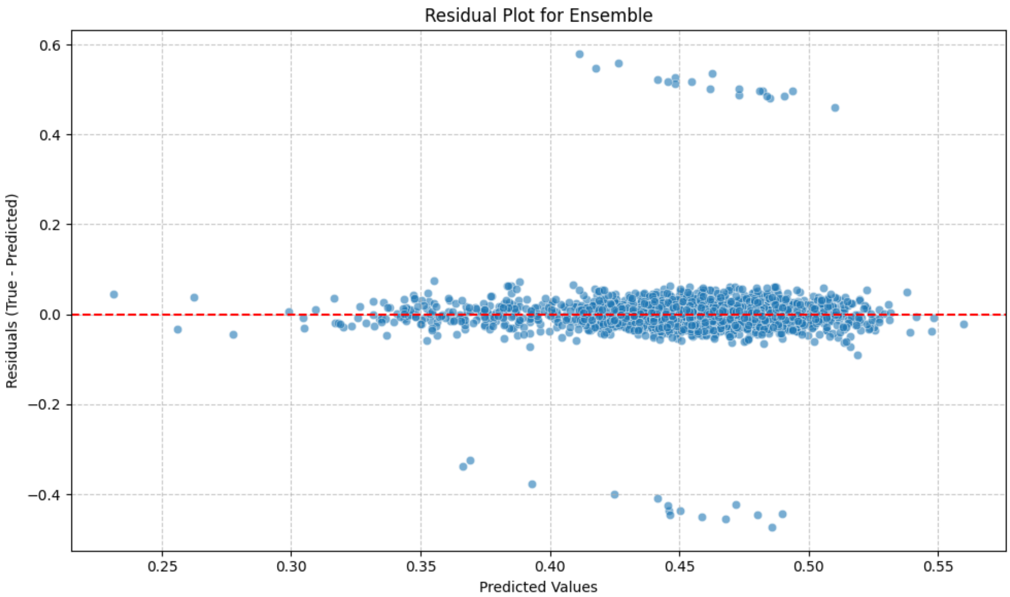

Figure 3 Residual Plot of the Stacked Ensemble Model for Adoption Prediction

The residual plot for the ensemble model Figure 3 provides critical insight into the distribution of prediction errors by plotting residuals (True - Predicted values) against the predicted values. A well-performing regression model is typically expected to produce residuals randomly dispersed around the zero line without any discernible patterns. In the presented plot, most residuals are tightly clustered near the red dashed line at zero, indicating that the model has successfully minimized error across a wide range of predictions. This alignment suggests that the model is not suffering from heteroscedasticity (i.e., variance of errors is not increasing with the prediction scale) and is free from systematic bias.

However, a few outliers are observed at the upper and lower extremes, with residuals reaching approximately +0.6 and -0.4, respectively. These may represent unusual or noisy data instances or indicate regions where the model

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

slightly underperforms. Nevertheless, their sparse distribution and limited magnitude indicate they exert minimal influence on the overall model performance. The dense horizontal band of residuals centered on zero also reflects strong predictive consistency, validating the robustness of the proposed stacked ensemble approach for educational adoption prediction.

Figure 4 illustrates the top 20 most important features contributing to the performance of the model. The listed features in priority wise, resulting from ensemble model interpretability tools such as permutation importance and tree-based feature importance, with this one can easily understood what the important features from the trained data are. Notably, user_name appears as the most important feature, potentially serving as a proxy for user-specific behavioral patterns or institutional characteristics embedded in the training data. Following this features like cluster_isolation_7, cluster_isolation_5, and cluster_isolation_3 rank highly, reflecting the significance of network structure or isolation metrics in influencing adoption behaviors.

And more important features like engagement_score, assessment_to_course_ratio, and engagement_consistency also emerge importantly. These features directly align with pedagogical effectiveness and student interaction intensity, underscoring their role in predicting technology adoption in academic contexts. Interestingly, features such as cisco_to_webex_ratio, platform_specialization, and files_to_total_platform_usage_ratio capture platform usage behaviors, hinting at how learners' preferences and tool engagement patterns influence adoption predictions.

Moreover, the presence of time-based features like registration_hour and content-based metrics like total_learning_activity and total_resource_usage highlights the multi-dimensional nature of the feature space. This all features like behavioral, temporal, and interactional features contribute to the ensemble model's high accuracy and generalizability.

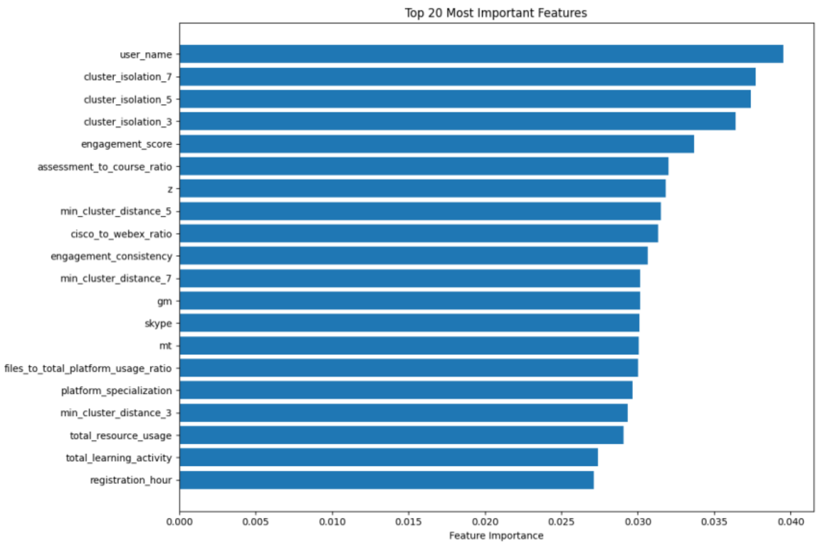

Figure 4 Top 20 Most Influential Features Identified by the Ensemble Learning Model

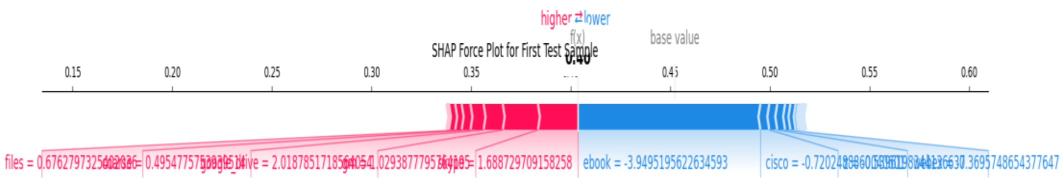

Figure 5 SHAP Force Plot Illustrating Local Feature Contributions for a Single Prediction Instance

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

To gain a deeper understanding of the LightGBM model's decision-making process in predicting cloud-based educational adoption, SHAP (SHapley Additive exPlanations) was employed for both local and global interpretability. The SHAP **force plot** in figure 5 visualizes the contribution of individual features toward the final prediction for a single test instance. In this plot, the base value (approximately 0.435) represents the expected model output, and the actual prediction is shown as a shift from this base value. Red-colored features indicate variables that increase the prediction value (positive SHAP values), whereas blue-colored features contribute to reducing the prediction (negative SHAP values). Notably, the feature files (SHAP value = +0.6763) was the most significant positive contributor, indicating that the presence or higher use of file-related content significantly increases the likelihood of the predicted class. In contrast, the feature ebook (SHAP value = −3.9495) exerted the strongest negative influence, suggesting that higher usage of eBooks is associated with reduced predicted adoption or engagement in this particular case. Other features such as drive, mt, and google_drive had moderate positive impacts, whereas cisco and skype displayed negative contributions. This instance-level interpretation is essential for understanding individual user behavior and tailoring interventions accordingly.

Complementing the local analysis, the SHAP summary plot figure 6 provides a global overview of feature importance across the entire dataset. It ranks features based on the mean absolute SHAP value, effectively quantifying their overall impact on model output. The most influential features identified were ebook, gm, coarse, mt, and google_drive. These results underline that digital content types and learning modes, such as eBooks and Google Meet (gm), are critical factors in predicting the adoption of cloud-based educational platforms. The plot also reveals the direction and magnitude of each feature's impact: for instance, high values of ebook (shown in red) predominantly push the prediction downward, confirming its negative influence seen in the force plot. Conversely, high values of gm and coarse are generally associated with positive contributions to the predicted outcome. Less impactful features include course_duration_days, request_hour, and course_end_day_of_week, indicating that temporal aspects have a comparatively lower effect on the model's decisions.

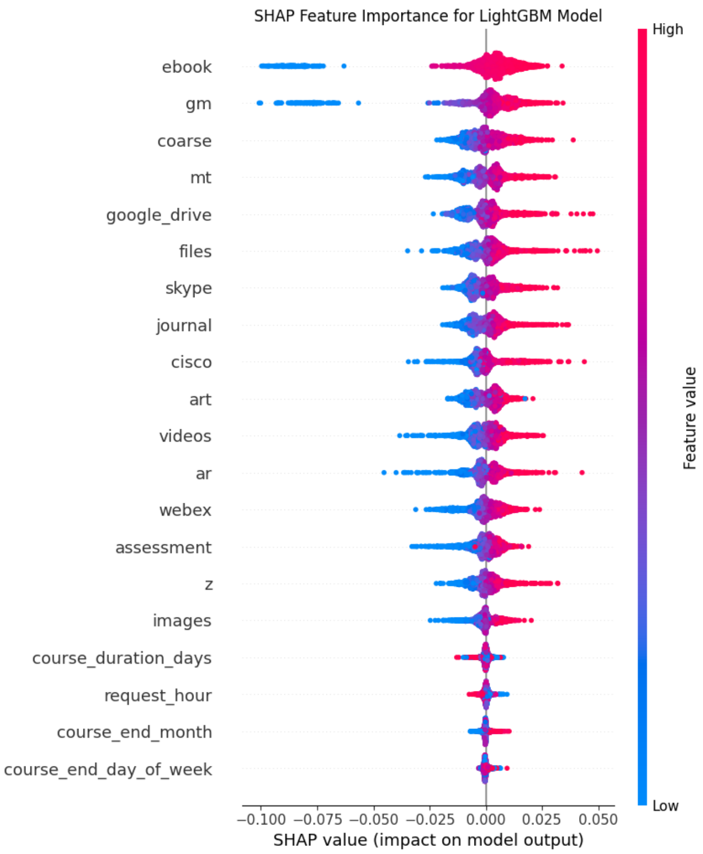

Figure 6 SHAP Summary Plots Highlighting Global Feature Importance Across the Dataset

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

6. CONCLUSION

This research presents a robust and scalable machine learning framework for predicting the adoption of cloud-based educational technologies. By integrating ensemble learning techniques with rich feature engineering, the system demonstrated excellent predictive performance, achieving 0.023 of MSE on unseen data and a significant reduction in validation loss to **0.07**. The ensemble approach, particularly with LightGBM as the meta-learner, proved effective in capturing complex patterns across behavioral, temporal, and content-based features.

Interpretability was addressed through SHAP analysis, which identified ebook, files, google_drive, and videos as the most influential predictors of adoption behavior. The SHAP force plot illustrated that higher usage of files and Google Drive increased the probability of adoption, while low interaction with ebooks had a suppressive effect. These results are not only technically significant but also offer practical insights for academic institutions to personalize resource allocation and engagement strategies.

Future work will focus on real-time deployment, incorporating longitudinal tracking of user behavior, and expanding the dataset to include multi-institutional and multilingual student cohorts to enhance generalizability and scalability.

REFERENCES

[1] Liu, N., Li, Y., & Guo, Y. (2024). Optimization of Online Learning Resource Adaptation in Higher Education through Neural Network Approaches. International Journal of Interactive Mobile Technologies, 18(11).

[2] Alrajhi, L., Alamri, A., & Cristea, A. I. (2022, June). Intervention Prediction in MOOCs Based on Learners' Comments: A Temporal Multi-input Approach Using Deep Learning and Transformer Models. In International Conference on Intelligent Tutoring Systems (pp. 227-237). Cham: Springer International Publishing.

[3] Qiu, S. (2024). Improving Performance of Smart Education Systems by Integrating Machine Learning on Edge Devices and Cloud in Educational Institutions. Journal of Grid Computing, 22(1), 41.

[4] Gu, R. (2024). Optimization, Integration, and Open Sharing of Business English Course Resources Based on Embedded Neural Networks. International Journal of High Speed Electronics and Systems, 2540120.

[5] Zaveri, J. S., & Shrivastav, A. K. Adoption of Advanced Systems for Digital Resource Management in Academic Institutions.

[6] Tulinayo, F. P., Ssentume, P., & Najjuma, R. (2018). Digital technologies in resource constrained higher institutions of learning: a study on students' acceptance and usability. International Journal of Educational Technology in Higher Education, 15(1), 1-19.

[7] Jokhan, A., Chand, A. A., Singh, V., & Mamun, K. A. (2022). Increased digital resource consumption in higher educational institutions and the artificial intelligence role in informing decisions related to student performance. Sustainability, 14(4), 2377.

[8] Gabelaia, I. (2024). Examining Pedagogical Analytical and Digital Literacy in Higher Education: Predicting Faculty's Readiness for and Adoption of Learning Analytics (Doctoral dissertation, Drake University).

[9] Plantak Vukovac, D., Hajdin, G., & Oreški, D. (2024). Explaining and Predicting Students and Teachers Intentions to Reuse Digital Educational Resources. *Journal of communications software and systems*, *20*(2), 157-164.

[10] Mexhuani, B. (2025). Adopting Digital Tools in Higher Education: Opportunities, Challenges and Theoretical Insights. *European Journal of Education*, *60*(1), e12819.

[11] Almaiah, M. A., Alhumaid, K., Aldhuhoori, A., Alnazzawi, N., Aburayya, A., Alfaisal, R., ... & Shehab, R. (2022). Factors affecting the adoption of digital information technologies in higher education: an empirical study. *Electronics*, *11*(21), 3572.

[12] Lazar, I. M., Panisoara, G., & Panisoara, I. O. (2020). Digital technology adoption scale in the blended learning context in higher education: Development, validation and testing of a specific tool. *PloS one*, *15*(7), e0235957.

[13] Qasem, Y. A., Asadi, S., Abdullah, R., Yah, Y., Atan, R., Al-Sharafi, M. A., & Yassin, A. A. (2020). A multi-analytical approach to predict the determinants of cloud computing adoption in higher education institutions. *Applied Sciences*, *10*(14), 4905.

[14] Shard, Kumar, D., & Koul, S. (2024). Digital transformation in higher education: A comprehensive review of e-learning adoption. *Human Systems Management*, *43*(4), 433-454.

[15] D'Ambra, J., Akter, S., & Mariani, M. (2022). Digital transformation of higher education in Australia: Understanding affordance dynamics in E-Textbook engagement and use. *Journal of Business Research*, *149*, 283-295.

[16] Li, X., Deeprasert, J., & Jiang, S. (2024). Adoption Intention towards Open Educational Resources: Role of Experience, Digital Divide and Voluntariness. *African Educational Research Journal*, *12*(4), 282-299.

[17] Seres, L., Pavlicevic, V., & Tumbas, P. (2018). Digital transformation of higher education: Competing on analytics. In *INTED2018 Proceedings* (pp. 9491-9497). IATED.

[18] Mtebe, J. S., & Raisamo, R. (2014). Challenges and instructors' intention to adopt and use open educational resources in higher education in Tanzania. *International review of research in open and distributed learning*, *15*(1), 249-271.

[19] Ly, B., & Doeur, B. (2024). Key factors influencing digital learning adoption among cambodian university students: An integrated theoretical approach. *Computers in Human Behavior Reports*, *15*, 100460.

[20] Ahmad, S., Mohd Noor, A. S., Alwan, A. A., Gulzar, Y., Khan, W. Z., & Reegu, F. A. (2023). eLearning acceptance and adoption challenges in higher education. *Sustainability*, *15*(7), 6190.

[21] Martin, F., Polly, D., Coles, S., & Wang, C. (2020). Examining higher education faculty use of current digital technologies: Importance, competence, and motivation. *International Journal of Teaching and Learning in Higher Education*, *32*(1), 73-86.

[22] Lin, Y., & Yu, Z. (2023). Extending Technology Acceptance Model to higher-education students' use of digital academic reading tools on computers. *International Journal of Educational Technology in Higher Education*, *20*(1), 34.