

Academic Performance Prediction from Student–VLE Bipartite Interaction Graphs Using Centrality Features A Comparative Study with Classical Classifiers

Ai Irma Sumiati^{1)*}, Taqwa Hariguna²⁾, Azhari Shouni Barkah³⁾

¹⁾²⁾³⁾ Master of Computer Science, Universitas Amikom Purwokerto, Indonesia

¹⁾ 24MA41D017@student.amikompurwokerto.ac.id, ²⁾ taqwa@amikompurwokerto.ac.id,

³⁾ azhari@amikompurwokerto.ac.id

Submitted : Jan 10, 2026 | **Accepted** : Jan 17, 2026 | **Published** : Jan 21, 2026

Abstract: The rapid growth of digital learning platforms has increased the availability of student academic records and fine-grained interaction logs, creating opportunities for Educational Data Mining (EDM) to support early academic monitoring. However, many predictive models still rely mainly on individual tabular attributes and underutilize relational signals embedded in learning interactions. This study proposes a graph-mining feature approach for predicting student academic performance using a bipartite Student–VLE interaction graph. Centrality measures—degree, weighted degree, HITS hub, PageRank, and eigenvector centrality—are extracted to form a centrality feature set and combined with standard student information features. Using the public OULAD dataset, we compare three supervised classifiers: Random Forest, Support Vector Machine, and XGBoost. Experiments show that adding the centrality feature set consistently and substantially improves performance across all models compared to baseline tabular features. On the test set, XGBoost achieves the strongest results with accuracy 0.842, ROC-AUC 0.922, PR-AUC 0.902, and MCC 0.684, while Random Forest is close behind (accuracy 0.834, ROC-AUC 0.916, PR-AUC 0.894, MCC 0.672). The SVM model also benefits (accuracy 0.800, ROC-AUC 0.869, PR-AUC 0.811, MCC 0.599), confirming the robustness of the graph-derived signal. Scientifically, this study provides empirical evidence that a multi-centrality representation offers more systematic and transferable predictive value than relying on a single graph metric, across multiple classical model families under the same evaluation protocol. These findings indicate that graph-mining centrality features capture complementary structural information about learning engagement that is not represented by tabular attributes alone, and they offer a practical, interpretable enhancement to classic EDM pipelines for academic performance prediction.

Keywords: Educational Data Mining; Student Performance Prediction; Bipartite Graph; Centrality Features; XGBoost; Random Forest; SVM

INTRODUCTION

The rapid expansion of digital learning technologies has led higher-education institutions to accumulate increasingly rich academic data, ranging from grades and attendance to fine-grained interaction traces captured by learning management systems and virtual learning environments (VLEs). As these data become larger and more heterogeneous, predicting student academic performance is no longer a straightforward task and requires more appropriate analytical approaches to support reliable early identification of students who may need academic support (Thaher & Jayousi, 2020; Salim et al., 2024). In practice, institutions that can exploit educational data effectively are better positioned to design proactive interventions and improve educational quality through evidence-based decision making (Salim et al., 2024).

Educational Data Mining (EDM) provides a methodological foundation for extracting patterns from educational datasets at scale, enabling the development of predictive models that support monitoring, diagnosis, and targeted interventions. Prior studies report that classical machine learning algorithms—particularly Random

*name of corresponding author



Forest (RF) and Support Vector Machine (SVM)—often perform well for student performance prediction because they handle complex feature interactions and provide competitive accuracy with relatively moderate implementation cost (Nugroho, 2025; Sathe & Adamuthe, 2021). However, systematic evidence indicates that many performance prediction pipelines still rely predominantly on “individual-level” variables (e.g., demographics, prior grades, attendance) and underutilize relational signals such as learning interactions or collaboration patterns that emerge in digital learning ecosystems (Ismanto et al., 2022). This limitation is non-trivial, because learning interactions—at both individual and group levels—can meaningfully influence academic achievement and risk trajectories (Qiao, 2024).

Graph mining offers a principled way to represent and analyze relational structure in educational settings. In a graph representation, nodes may encode students and learning resources, while edges encode interactions (e.g., clicks, accesses, or engagement intensity). Centrality analysis—originally popularized in web and social networks—quantifies how “important” or “influential” nodes are within the network structure (Han et al., 2011). Among centrality measures, PageRank has been widely studied as a structural indicator of node influence, and its diffusion across domains—including educational applications—has been increasingly recognized (Yu & Yan, 2024). Recent work suggests that using PageRank-derived signals as additional features can improve predictive performance, particularly for identifying students at risk, compared with purely tabular baselines (Alamgir et al., 2024). Nevertheless, the broader methodological question remains open in many EDM studies: rather than relying on PageRank alone, can a richer centrality feature set extracted from a well-defined interaction graph provide a stronger and more consistent performance gain across multiple classical models?

This study addresses that gap by framing student engagement logs as a weighted bipartite student–VLE graph, where one node set represents students and the other represents VLE resources, and edge weights reflect interaction intensity. From this bipartite structure, we extract multiple centrality descriptors per student—such as degree, weighted degree, hub score (HITS), PageRank, and eigenvector centrality—and integrate them with student-level tabular attributes. We then evaluate and compare RF, SVM, and XGBoost under a consistent split protocol and unified preprocessing pipeline. This design directly responds to prior calls for more comprehensive EDM pipelines that combine systematic feature engineering, robust validation, and model comparison rather than focusing on a single algorithm or a single feature family (Salim et al., 2024; Rachmatika et al., 2020). While modern deep learning approaches (including GNN-based models) can capture interaction patterns automatically, they often involve higher computational cost and reduced interpretability; consequently, a centrality-driven graph mining pipeline paired with classical models remains a compelling and practical alternative for many institutional contexts (Qiao, 2024; Nazir et al., 2025).

To structure the investigation and ensure each question is answerable by the reported experimental results, this study addresses:

1. RQ1: To what extent does integrating a bipartite centrality feature set improve student performance prediction compared with a tabular-only baseline?
2. RQ2: Which model family benefits most from centrality-based graph mining features—RF, SVM, or XGBoost?
3. RQ3: Do centrality features improve both ranking quality (ROC-AUC, PR-AUC) and decision quality (accuracy, F1, MCC, balanced accuracy) under a consistent evaluation protocol?
4. RQ4: How does incorporating centrality features change the error profile (confusion matrix patterns) relative to baseline tabular models?

This paper makes the following contributions: a) reproducible approach to construct a weighted student–VLE bipartite graph from interaction logs and extract a multi-metric centrality feature set; b) an empirical evaluation demonstrating that centrality features can substantially enhance predictive performance beyond student-level tabular attributes, supporting prior observations about graph-feature utility in student risk detection; c) a controlled comparison of RF, SVM, and XGBoost under the same split protocol and metrics, clarifying which classifiers most effectively exploit graph-derived signals; and d) a practical modeling approach aligned with EDM best practices for institutional adoption—balancing interpretability, reproducibility, and predictive utility.

LITERATURE REVIEW

Educational Data Mining for Student Performance Prediction

The increased adoption of digital learning systems has expanded the availability of educational data, ranging from demographics and prior achievement to fine-grained behavioral logs from learning platforms. EDM emerged to extract patterns from such data to support early identification of at-risk students and evidence-based educational decisions (Thaher & Jayousi, 2020; Salim et al., 2024). In practice, EDM pipelines for performance prediction typically include data preprocessing, feature engineering, model training, and evaluation using classification metrics appropriate for the target label and class distribution (Han et al., 2011).

A recurring challenge in EDM is that many educational datasets are imbalanced (e.g., fewer students in the “at-risk” class), which can bias models toward the majority class and reduce the usefulness of predictions for

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

intervention. To address this, studies often integrate resampling strategies such as SMOTE into the learning pipeline (Thaher & Jayousi, 2020). However, the usefulness of any balancing strategy depends strongly on the evaluation protocol and the extent to which features reflect meaningful learning signals rather than noise or leakage. Therefore, reproducible and properly validated pipelines remain central concerns in EDM (Salim et al., 2024; Rachmatika et al., 2020).

Machine Learning Methods in EDM: RF, SVM, and Boosting

Classical supervised learning remains widely used in EDM due to its strong empirical performance on structured, mixed-type datasets and its relatively straightforward training and deployment. Random Forest has been highlighted for its robustness and reduced overfitting risk via ensemble learning, making it a common baseline for academic prediction tasks (Nugroho, 2025; Sathe & Adamuthe, 2021). SVM is similarly popular, particularly for high-dimensional feature spaces, where margin maximization and kernel methods can capture complex decision boundaries (Sathe & Adamuthe, 2021; Salim et al., 2024).

Despite their strengths, many ML-based EDM studies still rely primarily on individual-level features (e.g., demographics, grades, attendance), without systematically incorporating relational or interaction structure. Systematic and survey-oriented works report that this “feature-individual” focus is dominant, even though learning is shaped by peer influence, collaborative behaviors, and platform-mediated interactions (Ismanto et al., 2022; Qiao, 2024). This motivates the exploration of feature representations that encode interaction structure more directly.

Recent trends in EDM also point to a shift from single-model baselines toward stronger ensemble and boosting approaches as competitive predictors for tabular data (Pan et al., 2025; Salim et al., 2024). While deep learning methods can be effective in certain settings (Nazir et al., 2025), classical and boosting models remain attractive when interpretability, reproducibility, and efficient training are important, or when the primary goal is a clear comparison of feature representations rather than a complex end-to-end architecture.

Graph Mining for Learning Interaction Data: Bipartite Student–VLE Networks and Centrality Features

Graph mining provides a framework to represent educational interactions as networks and to derive structural measures that capture influence, prominence, or engagement patterns. In learning analytics contexts, interactions may be modeled as (i) student–student graphs (peer relations) or (ii) bipartite graphs, such as Student–VLE (Virtual Learning Environment) networks, where edges indicate a student’s interaction with learning resources. Bipartite modeling is particularly appropriate for platform logs because it preserves the natural two-type structure (students and learning materials) rather than forcing interactions into a student–student projection that may lose information or introduce artifacts (Alamgir et al., 2024).

Centrality measures are among the most widely used graph mining outputs for converting network structure into numerical features. PageRank is a prominent example, originally developed for ranking nodes by importance based on link structure and often used to quantify influence in networks (Yu & Yan, 2024). In educational settings, PageRank-derived features can represent how “centrally positioned” a student is within an interaction network, potentially reflecting engagement intensity or strategic resource usage. Prior work has shown that adding PageRank features to machine learning models can improve predictive performance compared to purely tabular baselines, including in studies using public datasets such as UCI Student Performance and OULAD (Alamgir et al., 2024).

However, PageRank is not the only meaningful structural signal. Network science and graph mining commonly employ a set of centrality measures—such as degree, weighted degree, HITS hub scores, and eigenvector-based centrality—to capture different aspects of a node’s role (Han et al., 2011). In educational interaction graphs, these measures can encode complementary information: for example, degree and weighted degree reflect breadth and intensity of engagement, while eigenvector-like measures can reflect connections to highly active resources or interaction patterns that are globally informative. This motivates using a centrality feature set rather than relying on a single graph metric.

At the same time, the literature also suggests that graph-based modeling choices matter. Some graph mining applications remain theoretical or focused on general social networks rather than learning-specific networks, meaning that methodological adaptation and practical validation in education contexts are still needed (Yu & Yan, 2024). Moreover, while deep graph models such as graph neural networks can capture dynamic interaction patterns, they may trade off simplicity and transparency for performance and are not always evaluated alongside strong classical baselines in a reproducible pipeline (Qiao, 2024).

Theoretical Meaning of Centrality in Learning Analytics

Learning analytics (LA) examines learner interaction data from educational technologies to understand and improve learning (Gašević et al., 2019; Beer et al., 2021). Because LA practice often advances faster than theory, “centrality” is used in several distinct ways rather than as a single unified concept (Virgili, 2019; Khalil et al.,

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

2022). In the literature, centrality can mean (i) structural prominence in a network representation of learning activity, (ii) pedagogical importance of particular tasks or learning-design elements, (iii) temporal importance of key moments or phases for intervention, and (iv) practical and institutional salience shaped by dashboards, sensemaking, and governance capacity (Knight et al., 2020; Peach et al., 2021; Knobbout & Stappen, 2020; Ifenthaler et al., 2024).

A key implication is that centrality metrics are only educationally meaningful when their representation and interpretation are tied to learning theory and validated in context. Structural measures depend on how the interaction graph is defined, while task and temporal views require linking analytics to learning goals and when learning changes occur. Sensemaking and organizational views further emphasize that what becomes “central” in practice depends on interpretability and institutional readiness, and algorithmic perspectives warn that model-based “importance” can reflect modeling choices as much as pedagogy (Poquet, 2024; Groß, 2024; Khalil et al., 2022).

Research Gap and Study Positioning

Across EDM research, three gaps appear consistently. First, despite increasing recognition that interaction structure affects learning outcomes, many predictive models still emphasize individual variables and underutilize relational data (Ismanto et al., 2022; Qiao, 2024). Second, studies that adopt graph mining often focus narrowly on PageRank alone, leaving open the question of whether a broader centrality feature set provides more robust or generalizable benefits (Yu & Yan, 2024; Han et al., 2011). Third, several reviews highlight that EDM pipelines integrating feature engineering, imbalance handling (e.g., SMOTE), and rigorous validation are not yet consistently applied—particularly in contexts that require clear comparability and reproducibility of results (Thaher & Jayousi, 2020; Salim et al., 2024).

Building on these observations, the present study positions itself as a comparative and reproducible evaluation of RF, SVM, and XGBoost under a consistent protocol, contrasting a baseline feature set against a graph-enhanced representation derived from a bipartite Student–VLE network. Instead of treating PageRank as the sole graph signal, the study emphasizes a centrality feature set (including PageRank alongside other centralities) to better capture complementary structural characteristics in student-resource interactions. This positioning aligns with prior evidence that PageRank-based features can improve prediction (Alamgir et al., 2024), while extending the feature design to reflect broader graph mining practice and addressing the need for stronger, better-controlled comparative pipelines in EDM (Salim et al., 2024; Han et al., 2011).

METHOD

Research Design

This study is designed as a supervised learning–based predictive analysis to classify students’ academic performance as a binary classification task. The main objective is to examine whether a graph mining approach that extracts centrality features from a Student–VLE bipartite graph can improve predictive performance compared to a conventional approach relying solely on tabular student attributes. To ensure a fair and consistent evaluation, the study compares three machine learning models commonly used in Educational Data Mining, namely Random Forest (RF), Support Vector Machine (SVM), and XGBoost, under two feature configurations: (i) a tabular baseline and (ii) tabular features augmented with centrality-based graph features. All experiments follow a consistent, reproducible protocol, including fixed random seeds, saved split indices, stored experiment configurations, and systematic saving of trained models and evaluation metrics into run-specific output directories.

Dataset and Prediction Target

The primary dataset used in this study is the Open University Learning Analytics Dataset (OULAD) because it provides both student demographic information and detailed traces of learning interactions on the Virtual Learning Environment (VLE). In implementation, the tabular baseline is constructed mainly from the studentInfo table, while the interaction signals used to build the bipartite graph are derived from the studentVle table. The prediction target is derived from the final_result attribute in studentInfo, which represents the final academic outcome at the course-instance level (a student enrolled in a specific module and presentation). This target is selected because final_result is among the cleanest and most consistent labels for representing academic success within OULAD, reducing potential bias that may arise when labels are constructed only from partial assessment components. Accordingly, the unit of analysis in this study is each student–module–presentation record.

Data Preparation

Data preparation begins by constructing a tabular dataset from studentInfo and ensuring that each record has a valid final_result label. Next, VLE interaction data from studentVle are filtered to match the same course context so that the resulting graph edges truly reflect a student’s activity within a given module and presentation rather than mixing interactions across different courses. Data cleaning also addresses duplicate records that can emerge

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

because some source tables are event-based. For tabular features, categorical variables are retained as categorical and later processed via one-hot encoding, while numerical variables are preserved as numerical and standardized when needed (particularly for SVM). Missing values are handled deterministically to support reproducibility, including explicit encoding of missingness for graph-derived features through a dedicated indicator feature.

Train–Test Split and Reproducibility

To maintain evaluation validity, this study uses a hold-out split with stratification by the binary label so that class proportions remain comparable between training and test sets. To guarantee full reproducibility, the row indices for training and testing are saved to a `split_idx.json` file, and all models across all feature configurations are trained and evaluated using the same split. This ensures that any performance differences across models or feature sets are attributable to the information content of the features and the learning capacity of the models, rather than sampling variability. In addition, each run stores experiment configuration files and model artifacts in a dedicated output folder, enabling experiments to be paused and resumed without changing the underlying settings.

Feature Engineering

The baseline configuration uses the tabular features available in `studentInfo`, which primarily include demographic attributes and study context variables. Categorical variables such as `code_module`, `code_presentation`, `gender`, `region`, `highest_education`, `imd_band`, `age_band`, and `disability` represent student background and course context. Numerical variables such as `num_of_prev_attempts` and `studied_credits` capture prior academic attempts and study load. This baseline reflects the common EDM approach of predicting performance from individual-level attributes without explicitly modeling relational structure in digital learning environments.

To incorporate relational information from learning activity, this study constructs a bipartite graph modeling the relationship between students and VLE learning resources. In this graph, nodes in the first partition represent students, while nodes in the second partition represent VLE resources (e.g., `id_site`). An edge is created when a student interacts with a particular resource, and the edge weight may represent interaction intensity such as `sum_click`. The graph is constructed separately for each (`code_module`, `code_presentation`) pair to ensure that the structure remains specific to each course instance rather than mixing across modules or semesters. This design allows the graph to capture connectivity patterns among students and shared resources, which often reflect common learning behaviors and levels of engagement with instructional materials.

To prevent information leakage, graph features are computed in a way that respects the train–test split. Centrality measures for the training subset are computed using interactions available within the training split, while centrality measures for the test subset are computed using interactions available within the test split. This splitwise procedure avoids a scenario where test interactions indirectly influence training-derived features, which could otherwise lead to overly optimistic evaluation. As a result, the model is evaluated under conditions that more closely represent generalization to unseen data.

The graph-enhanced configuration augments the tabular baseline with a targeted set of centrality measures chosen to capture complementary aspects of student engagement in a student–resource network. Degree centrality represents breadth of engagement by counting how many distinct resources a student accessed. Weighted degree represents intensity of engagement by summarizing total interaction volume. HITS hub score is included to capture a student’s tendency to connect broadly to resources that are themselves widely accessed, reflecting “hub-like” participation patterns in the bipartite structure. PageRank is used as a diffusion-style importance measure that accounts for both direct engagement and the global connectivity of the accessed resources under a random-walk interpretation. Eigenvector centrality is included as an alternative influence-based measure that emphasizes connection to globally “important” resources and provides a contrast to PageRank under a different recursive importance assumption. These five measures were selected because (i) they are well-established, computationally tractable on large course graphs, (ii) they cover distinct engagement mechanisms (breadth, intensity, hub-like connectivity, diffusion-based importance, and recursive influence), and (iii) they yield student-level numeric descriptors that integrate cleanly with classical ML pipelines.

Other centralities were not used in this study to keep the feature set compact and interpretable while avoiding measures that are either less stable or less aligned with the bipartite student–resource setting. For example, betweenness and closeness centrality typically require shortest-path computations that can be expensive at scale and may be sensitive in sparse, disconnected graphs, while also being harder to interpret as engagement in resource-access networks. By focusing on degree-family and influence-style centralities, the design prioritizes measures that directly map to engagement volume and structural prominence in a student–VLE interaction graph.

Because some students may have insufficient VLE traces in a given course instance, centrality features can be missing or undefined. These cases are treated as missing graph evidence rather than “low centrality.” Missing values are imputed deterministically (e.g., fixed value or training-set statistic), and the learning pipeline is expected to absorb the remaining signal through the combination of tabular and available graph-derived features. After

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

extraction, two aligned feature matrices are produced under identical split indices: a tabular baseline matrix and a tabular-plus-centrality matrix, enabling a controlled comparison where the only difference is the presence of centrality-based graph features.

Model Training and Hyperparameter Tuning

Model training is conducted using a consistent pipeline across algorithms, including preprocessing and hyperparameter search. Categorical variables are transformed using a OneHotEncoder with `handle_unknown='ignore'` to remain robust to unseen categories. Numerical variables are processed consistently, including standardization to support margin-based models such as SVM. Hyperparameters are selected using 3-fold cross-validation on the training set, with a primary selection focus on ranking metrics such as ROC-AUC to avoid dependence on a specific threshold during tuning. For Random Forest, tuning includes the number of trees, maximum depth, and regularization-related parameters such as minimum leaf size. For SVM, tuning includes the regularization parameter `C` and the optional `class_weight` setting to improve robustness in imbalanced or margin-sensitive settings. For XGBoost, tuning includes the number of estimators, learning rate, tree depth, and subsampling parameters to control overfitting. The best-performing model from cross-validation is then refit on the full training split and saved as an artifact for reuse without retraining.

Evaluation Metrics

Evaluation is performed on the held-out test set using both threshold-free and threshold-based metrics. Threshold-free metrics such as ROC-AUC and PR-AUC are reported to assess probability ranking quality, which is important when decision thresholds may vary across operational contexts. For threshold-based evaluation at the default threshold of 0.5, the study reports accuracy, F1-score, balanced accuracy, and the Matthews Correlation Coefficient (MCC). In addition, a confusion matrix is reported to provide detailed insight into error types (false positives and false negatives). Using multiple metrics is particularly important in EDM, since a model with high accuracy may still perform poorly for certain subgroups or error types, and therefore performance interpretation should be supported by complementary measures.

Implementation and Output Artifacts

All experiments are implemented in Python using standard libraries for data processing, machine learning, and graph analysis. To ensure that the study is reproducible and resumable, each run produces an output directory containing configuration files, split indices, best model artifacts, hyperparameter search results, and summary evaluation metrics. This output structure supports experiment auditing, transparent comparison across feature configurations, and future extensions such as adding new centrality measures, modifying graph construction rules, or introducing additional baseline models without altering the core evaluation protocol.

RESULT

Evaluation setup and reporting conventions

All experiments were evaluated on the same OULAD split with 26,074 students used for training and 6,519 students used for testing. The target distribution is nearly balanced, with a positive rate of 0.4720 in training and 0.4720 in testing. Model performance is reported on the test set using a fixed decision threshold of 0.50 for classification metrics (accuracy, F1, MCC, balanced accuracy, and the confusion matrix). In addition, we report threshold-free ranking metrics, namely ROC-AUC and PR-AUC, to reflect the quality of probability ranking independent of any single threshold. Two feature configurations are compared. The baseline configuration uses studentInfo-style tabular variables only. The centrality feature configuration augments the same tabular variables with graph-mining centrality features computed from the Student-VLE bipartite graph, including degree, weighted degree, HITS hub score, PageRank, and eigenvector centrality. The objective of the results section is to quantify the effect of adding these centrality features and to identify which classifier family benefits most under the same evaluation protocol.

Overall comparison across models on the test set

Table 1 reports model performance for Random Forest, SVM (linear), and XGBoost under both feature configurations. The baseline tabular setting yields modest performance across models, with accuracy values around 0.615–0.628 and ROC-AUC values around 0.654–0.670. When the centrality features are added, all three models improve substantially, and the improvements are consistent across metrics (accuracy, F1, MCC, balanced accuracy) and ranking metrics (ROC-AUC, PR-AUC). Among the centrality-augmented models, XGBoost achieves the strongest overall results, followed closely by Random Forest, while SVM also improves but remains below the tree-based methods.

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Table 1. OULAD Model Performance

Model	Feature set	Accuracy	F1	MCC	Balanced Acc.	ROC-AUC	PR-AUC
Random Forest	Baseline tabular	0.6262	0.5839	0.2472	0.6224	0.6705	0.6321
Random Forest	Centrality features	0.8342	0.8323	0.6716	0.8362	0.9155	0.8941
SVM (Linear)	Baseline tabular	0.6282	0.5897	0.2515	0.6249	0.6696	0.6297
SVM (Linear)	Centrality features	0.7995	0.7924	0.5993	0.8001	0.8695	0.811
XGBoost	Baseline tabular	0.6154	0.585	0.2271	0.6132	0.6539	0.6081
XGBoost	Centrality features	0.8415	0.8367	0.684	0.8425	0.9224	0.9016

The bar charts in Figure 1 show that adding the centrality feature set consistently improves performance over the baseline tabular features for all evaluated classifiers. Accuracy rises from 0.626 to 0.834 for Random Forest, from 0.628 to 0.800 for SVM, and from 0.615 to 0.842 for XGBoost; the same pattern is observed for F1-score (from approximately 0.58–0.59 to 0.79–0.84) and MCC (from approximately 0.23–0.25 to 0.60–0.68), indicating substantially stronger class separation when relational information from the student–VLE interaction graph is included.

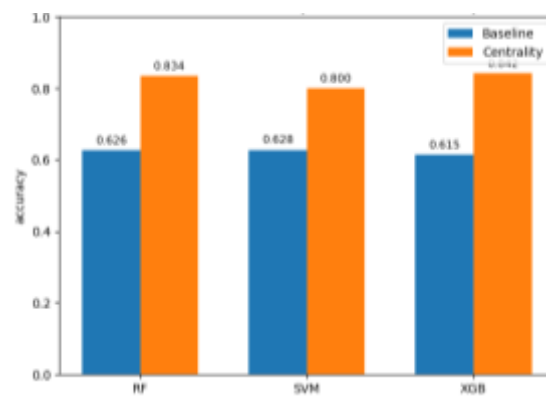


Figure 1. Comparison of Accuracy across Models (Baseline vs Centrality)

Figure 2 provides an error-level explanation of this improvement for the centrality-based models. All three models achieve a high number of correct positive predictions (true positives), namely 2682 (RF), 2494 (SVM), and 2647 (XGBoost), while maintaining moderate false negatives (395, 583, and 430, respectively). At the same time, false positives remain present (686 for RF, 724 for SVM, and 603 for XGBoost), suggesting that the main remaining errors are students predicted as positive when they are negative.

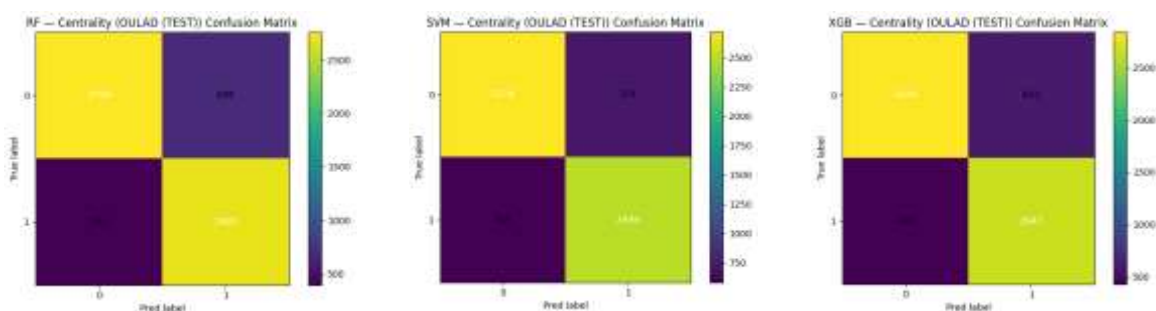


Figure 2. Comparison of Confusion Matrix across Models (Centrality)

Figure 3 demonstrates that the centrality feature set substantially improves the models' ranking ability, as indicated by the upward shift of the ROC curves toward the top-left corner and the larger AUC values compared with the baseline. This pattern holds for RF, SVM, and XGBoost, suggesting that the graph-based centrality information provides additional discriminative signal beyond standard tabular attributes.

*name of corresponding author



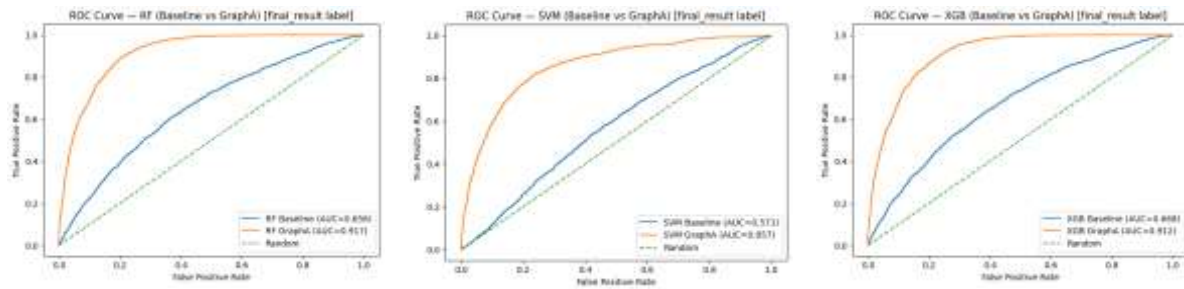


Figure 3. Comparison of ROC Curve across Models (Baseline vs Centrality)

Effect of adding centrality features relative to baseline

The impact of adding centrality features is large and uniform across the three classifiers. For Random Forest, accuracy increases from 0.6262 to 0.8342, and MCC increases from 0.2472 to 0.6716, indicating a major improvement in balanced, correlation-based classification quality rather than a superficial shift in threshold behavior. For SVM, accuracy increases from 0.6282 to 0.7995, and MCC increases from 0.2515 to 0.5993, showing that even a linear margin-based classifier benefits strongly once relational structure is encoded as centrality features. For XGBoost, the gains are the strongest overall: accuracy rises from 0.6154 to 0.8415 and MCC rises from 0.2271 to 0.6840, accompanied by a large increase in ROC-AUC from 0.6539 to 0.9224 and PR-AUC from 0.6081 to 0.9016. These consistent improvements support the interpretation that bipartite centrality features add predictive information that is not captured by the tabular baseline alone.

Error profile changes based on confusion matrices

The confusion matrices show that the centrality feature set reduces both false negatives and false positives, which explains why improvements appear simultaneously in balanced accuracy, MCC, and the ranking metrics. In the XGBoost baseline tabular model, the test confusion matrix is (TN=2245, FP=1197, FN=1310, TP=1767), which implies a positive-class recall of $1767/(1767+1310) \approx 0.574$ and a negative-class recall of $2245/(2245+1197) \approx 0.652$. After adding centrality features, the confusion matrix becomes (TN=2839, FP=603, FN=430, TP=2647), increasing positive-class recall to $2647/(2647+430) \approx 0.860$ and negative-class recall to $2839/(2839+603) \approx 0.825$. This indicates that the model becomes substantially better at identifying both classes, not merely shifting toward one class. A similar pattern is observed for Random Forest, where false negatives decrease from 1367 to 395 and false positives decrease from 1070 to 686 when centrality features are added. For SVM, the baseline confusion matrix (TN=2353, FP=1089, FN=1335, TP=1742) improves to (TN=2718, FP=724, FN=583, TP=2494) under centrality features, again reflecting reductions in both error types. Taken together, these results indicate that centrality features improve separability in the feature space in a way that benefits different model families.

DISCUSSIONS

Interpretation of results

This study examined whether graph-mining features derived from a bipartite Student–VLE interaction graph can improve student academic performance prediction beyond standard tabular student information, and whether the effect is consistent across different classical classifiers. Across the OULAD test split using the final_result label, the findings show a coherent pattern: enriching studentInfo features with a centrality feature set consistently strengthens predictive performance, implying that relational structure in learning interactions carries signal that is not captured by tabular attributes alone. This observation aligns with the broader Educational Data Mining perspective that learning outcomes are shaped not only by individual characteristics, but also by interaction patterns within digital learning environments. (Thaher & Jayousi, 2020; Ismanto et al., 2022; Salim et al., 2024)

When comparing tabular-only baselines against models augmented with centrality features, the magnitude of improvement is substantial and consistent across all evaluated model families. Random Forest improves from 0.626 to 0.834 in test accuracy, while ROC-AUC increases from 0.670 to 0.916 and PR-AUC from 0.632 to 0.894. SVM shows the same direction of change, with accuracy rising from 0.628 to 0.800, ROC-AUC from 0.670 to 0.869, and PR-AUC from 0.630 to 0.811. XGBoost also improves strongly, with accuracy increasing from 0.615 to 0.842, ROC-AUC from 0.654 to 0.922, and PR-AUC from 0.608 to 0.902. The gains extend beyond accuracy and AUC-style metrics; decision-oriented measures such as MCC also increase sharply (for example, XGBoost rises from 0.227 to 0.684), supporting the interpretation that centrality features improve class separability rather than merely shifting outcomes via thresholding. (Han et al., 2011; Salim et al., 2024)

The relative comparison across model families indicates that tree-based approaches extract the most value from the centrality-enhanced representation. Under the centrality feature set, XGBoost achieves the strongest overall

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

results, reaching 0.842 accuracy with MCC 0.684, ROC-AUC 0.922, and PR-AUC 0.902. Random Forest follows closely with accuracy 0.834, MCC 0.672, ROC-AUC 0.916, and PR-AUC 0.894, suggesting that ensemble trees broadly exploit the nonlinear interactions between demographic attributes and multiple network-derived signals effectively. SVM benefits substantially as well (accuracy 0.800; MCC 0.599), but remains below the tree-based models, particularly on ranking metrics (ROC-AUC 0.869; PR-AUC 0.811). This ordering is consistent with prior comparative findings that boosting methods are often more flexible in capturing complex feature interactions than linear-margin classifiers, especially when multiple engineered features jointly encode complementary aspects of behavior and engagement. (Sathe & Adamuthe, 2021; Salim et al., 2024)

The improvements are not confined to either ranking quality or decision quality; instead, both aspects improve together under the same evaluation protocol. In XGBoost, ROC-AUC increases from 0.654 to 0.922 and PR-AUC from 0.608 to 0.902, while accuracy improves from 0.615 to 0.842, F1 from 0.585 to 0.837, MCC from 0.227 to 0.684, and balanced accuracy from 0.613 to 0.843. Random Forest and SVM show the same coupled behavior, with ranking gains accompanied by consistent increases in accuracy, F1, MCC, and balanced accuracy. This joint movement across metric families strengthens the conclusion that the added features improve discrimination in a fundamental way, rather than producing superficial gains tied to a single metric type. (Han et al., 2011; Salim et al., 2024)

Changes in the error profile further clarify how the centrality feature set alters model behavior. Under tabular-only baselines, all models display heavy confusion between classes, reflected in large counts of both false positives and false negatives. For XGBoost, baseline errors are substantial (TN=2245, FP=1197, FN=1310, TP=1767), whereas the centrality-enhanced model becomes markedly more balanced with fewer errors on both sides (TN=2839, FP=603, FN=430, TP=2647). Random Forest shows a similar shift, moving from (TN=2372, FP=1070, FN=1367, TP=1710) to (TN=2756, FP=686, FN=395, TP=2682), and SVM moves from (TN=2353, FP=1089, FN=1335, TP=1742) to (TN=2718, FP=724, FN=583, TP=2494). The parallel rise in balanced accuracy—such as XGBoost increasing from 0.613 to 0.843—indicates that these reductions do not come from favoring one class at the expense of the other, which is especially relevant in intervention-oriented settings where both missed-risk cases and false alarms can carry real institutional costs. (Rachmatika et al., 2020; Salim et al., 2024)

Why the centrality feature set improves performance on OULAD

A key reason the centrality feature set improves performance is that studentInfo variables alone provide a limited snapshot of the learner, while the OULAD environment encodes rich behavioral interaction through the VLE. When students and learning materials/activities are represented as a bipartite graph, centrality measures become proxies for access, connectivity, and intensity of engagement within the learning ecosystem. For example, degree and weighted degree reflect how broadly and how intensely a student interacts with learning resources, while HITS hub-like signals can emphasize students who connect across multiple important activities. PageRank and eigenvector-based measures further capture recursive notions of importance, where being connected to influential resources or high-activity regions of the graph can matter. These signals are not strictly “social” in the sense of peer-to-peer ties, but they are relational in that they summarize position within the learning interaction structure; this is precisely the type of information that tabular demographic or static attributes cannot express (Han et al., 2011; Yu & Yan, 2024; Qiao, 2024). The ROC curves reinforce this mechanism. Under baseline features, ROC curves remain relatively close to the diagonal, indicating limited separability, while the centrality-enhanced models bend sharply toward the top-left, indicating much higher true-positive rates at low false-positive rates. This shape change is consistent with the hypothesis that interaction structure introduces a strong discriminative signal about academic outcomes (Alamgir et al., 2024; Salim et al., 2024).

Why tree ensembles outperform SVM under the centrality feature set

While SVM benefits significantly from centrality features, tree-based models (Random Forest and especially XGBoost) achieve higher ROC-AUC and PR-AUC. A plausible explanation is that centrality measures interact nonlinearly with demographic attributes and prior academic attempts. For example, the predictive meaning of high weighted degree may differ across course presentations or demographic segments, and such conditional interactions are naturally captured by trees and boosting. XGBoost can further refine these interactions by sequentially fitting residual errors, which often improves ranking performance and calibration in structured tabular problems (Sathe & Adamuthe, 2021; Salim et al., 2024). In contrast, a linear SVM seeks a single separating hyperplane, which can underfit when the boundary depends on complex feature interactions unless more expressive kernels are used, which may increase computational cost and tuning complexity.

Theoretical implications for EDM and learning analytics

These findings have theoretical implications for Educational Data Mining (EDM) because they strengthen the view that academic performance is shaped not only by static, individual attributes but also by a learner’s structural position in the digital learning ecosystem. The results support the idea that interaction logs contain relational signal

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

that is qualitatively different from conventional tabular variables, and that this signal can be captured through a multi-centrality representation without abandoning classical, deployable model families. Conceptually, the contribution is not merely “adding features,” but providing empirical evidence that a set of complementary centrality measures operationalizes multiple dimensions of engagement—breadth, intensity, and structural prominence—that are not reducible to demographics or study-history indicators. In this sense, the work advances EDM theory around what constitutes meaningful evidence of learning behavior: not only how much activity occurs, but how activity is organized within the student–resource interaction structure.

For learning analytics, the study clarifies how “centrality” can be interpreted as a theoretically meaningful construct when tied to an explicit representation of learning activity and validated through explainability. In a Student–VLE bipartite network, degree and weighted degree can be read as proxies for the breadth and intensity of resource engagement, while PageRank, hub scores, and eigenvector-based measures reflect global prominence that depends on connecting to structurally influential resources. The permutation and SHAP analyses help bridge a persistent LA gap: centrality metrics are often used, but their role in model decisions is rarely made explicit. By identifying which centralities consistently drive predictive changes and how their values push model outputs, the study supports a more accountable, theory-aware use of network measures—treating “centrality” not as a generic metric, but as an interpretable family of constructs whose educational meaning depends on the graph definition and the model’s learned attributions. This also provides a stronger conceptual basis for LA-driven interventions, because monitoring can be aligned with the specific engagement dimensions the model evidences as most consequential rather than relying on coarse activity totals alone.

Practical implications for early monitoring and academic support

From an application perspective, the centrality feature set can be interpreted as a compact summary of learning engagement structure that complements studentInfo. If deployed responsibly, such signals can support earlier identification of students who may need academic assistance, because interaction patterns often reveal disengagement or uneven resource usage before final outcomes are recorded. However, the model should be used as decision support rather than as an automated determinant of academic consequences. Institutions should combine model outputs with contextual judgment, confirmatory evidence, and transparent intervention policies to avoid unfair or opaque decision-making (Salim et al., 2024; Ismanto et al., 2022).

Limitations and future work

This study has several limitations. First, the centrality feature set depends on how the bipartite graph is constructed, including how edges are defined and weighted; alternative definitions may yield different centrality values and thus different predictive performance. Second, while the results are strong on the public OULAD split, generalization to other institutions requires re-validation because learning platforms, course structures, and student populations vary. Third, the current approach summarizes interactions into static centrality values; it does not explicitly model temporal trajectories of engagement, which can be important for earlier prediction. Future work can address these limitations by exploring temporal centrality sequences, community-structure features, or hybrid approaches that combine classical ML with more advanced graph-based models when interpretability and data scale allow (Qiao, 2024; Pan et al., 2025; Nazir et al., 2025).

CONCLUSION

This research demonstrates that incorporating graph-mining features from a bipartite Student–VLE interaction graph can significantly strengthen academic performance prediction compared with using student tabular attributes alone. Across Random Forest, SVM, and XGBoost, the centrality feature set consistently improves ranking metrics (ROC-AUC and PR-AUC) as well as decision-oriented metrics (accuracy and MCC), indicating better class separability and more balanced error behavior. Among the evaluated models, XGBoost provides the best overall performance, followed closely by Random Forest, while SVM remains competitive but generally lower on ranking metrics. The results support the view that PageRank should be treated as one component within a broader centrality family, where combining multiple centrality measures yields a richer and more reliable representation of engagement structure than relying on a single measure. Practically, the proposed approach offers an interpretable and computationally feasible enhancement for EDM systems intended to support academic monitoring and targeted interventions. Future work should validate the pipeline across additional cohorts or institutions, test alternative graph construction and weighting schemes, and explore temporal extensions that model changes in student engagement over time.

REFERENCES

- Alamgir, Z., Akram, H., Karim, S., & Wali, A. (2024). Enhancing Student Performance Prediction via Educational Data Mining on Academic Data. *Informatics in Education*, 23(1), 1–24. <https://doi.org/10.15388/infedu.2024.04>

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

- Beer, C., Jones, D., & Lawson, C. (2021). The Challenge of Learning Analytics Implementation: Lessons Learned. *Ascilite Publications*. <https://doi.org/10.14742/apubs.2019.5>
- Gašević, D., Tsai, Y., Dawson, S., & Pardo, A. (2019). How Do We Start? An Approach to Learning Analytics Adoption in Higher Education. *International Journal of Information and Learning Technology*. <https://doi.org/10.1108/ijilt-02-2019-0024>
- Groß, R. (2024). Stochastic Contingency Machines Feeding on Meaning: On the Computational Determination of Social Reality in Machine Learning. *Ai & Society*. <https://doi.org/10.1007/s00146-024-02079-8>
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Ifenthaler, D., Gibson, D., & Dobozy, E. (2024). The Synergistic and Dynamic Relationship Between Learning Design and Learning Analytics. *Ascilite Publications*. <https://doi.org/10.14742/apubs.2017.752>
- Ismanto, E., Ghani, H. A., Saleh, N. I. M., Al Amien, J., & Gunawan, R. (2022). Recent systematic review on student performance prediction using backpropagation algorithms. *Telkomnika (Telecommunication Computing Electronics and Control)*, 20(3), 597–606. <https://doi.org/10.12928/TELKOMNIKA.v20i3.21963>
- Khalil, M., Prinsloo, P., & Slade, S. (2022). The Use and Application of Learning Theory in Learning Analytics: A Scoping Review. *Journal of Computing in Higher Education*. <https://doi.org/10.1007/s12528-022-09340-3>
- Knight, S., Gibson, A., & Shibani, A. (2020). Implementing Learning Analytics for Learning Impact: Taking Tools to Task. *The Internet and Higher Education*. <https://doi.org/10.1016/j.iheduc.2020.100729>
- Knobbout, J., & der Stappen, E. van. (2020). A Capability Model for Learning Analytics Adoption: Identifying Organizational Capabilities From Literature on Learning Analytics, Big Data Analytics, and Business Analytics. *International Journal of Learning Analytics and Artificial Intelligence for Education (Ijai)*. <https://doi.org/10.3991/ijai.v2i1.12793>
- Nazir, M., Noraziah, A., Rahmah, M., Fakherldin, M., & Khawaji, A. (2025). Transforming Education with Deep Learning: A Systematic Review on Predicting Student Performance and Critical Challenges. *Fusion: Practice and Applications*, 18(2), 79–99. <https://doi.org/10.54216/FPA.180207>
- Nugroho, M. W. (2025). *Analisis Performa Algoritma Random Forest dalam Mengatasi Overfitting pada Model Prediksi*.
- Pan, J., Zhao, Z., & Han, D. (2025). Academic Performance Prediction Using Machine Learning Approaches: A Survey. *IEEE Transactions on Learning Technologies*, 18, 351–368. <https://doi.org/10.1109/TLT.2025.3554174>
- Peach, R. L., Greenbury, S. F., Johnston, I. G., Yaliraki, S. N., Lefèvre, D., & Barahona, M. (2021). Understanding Learner Behaviour in Online Courses With Bayesian Modelling and Time Series Characterisation. *Scientific Reports*. <https://doi.org/10.1038/s41598-021-81709-3>
- Poquet, O. (2024). A Shared Lens Around Sensemaking in Learning Analytics: What Activity Theory, Definition of a Situation and Affordances Can Offer. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.13435>
- Qiao, S. (2024). The Analysis and Prediction Mode of Students' Academic Performance and Social Behavior in the Background of Big Data. *Proceedings of 2024 3rd International Conference on Artificial Intelligence and Education (ICAIE)*, 648–652. <https://doi.org/10.1145/3722237.3722350>
- Rachmatika, R., Bisri, A., Puspipetek, J. R., Pamulang, K., & Tangerang Selatan, K. (2020). JEPIN (Jurnal Edukasi dan Penelitian Informatika) Perbandingan Model Klasifikasi untuk Evaluasi Kinerja Akademik Mahasiswa. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*.
- Salim, M., Al-Din, N., & Al Abdulqader, H. A. (2024). Students' Academic Performance Prediction Using Educational Data Mining and Machine Learning: A Systematic Review. *International Journal of Research and Innovation in Social Science*. <https://doi.org/10.47772/IJRISS>
- Sathe, M. T., & Adamuthe, A. C. (2021). Comparative study of supervised algorithms for prediction of students' performance. *International Journal of Modern Education and Computer Science*, 13(1), 1–21. <https://doi.org/10.5815/ijmecs.2021.01.01>
- Thaher, T., & Jayousi, R. (2020). Prediction of Student's Academic Performance using Feedforward Neural Network Augmented with Stochastic Trainers. *14th IEEE International Conference on Application of Information and Communication Technologies (AICT)*. <https://doi.org/10.1109/AICT50176.2020.9368820>
- Torrás Virgili, M. E. (2019). Learning Analytics: Online Higher Education in Management. *Sociology and Anthropology*. <https://doi.org/10.13189/sa.2019.070202>
- Yu, D., & Yan, Z. (2024). Knowledge diffusion trajectories of PageRank: A main path analysis. *Journal of Information Science*, 50(1), 273–287. <https://doi.org/10.1177/01655515231167388>