VFAST Transactions on Software Engineering
https://vfast.org/journals/index.php/VTSE@ 2025, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 13, Number 2, April-June 2025 pp: 56-67

Investigating the Role of LASSO in Feature Selection for Educational Data Mining (EDM) Applications

Mustafa Ahmed Khan [1*], **Khalid Mahboob** [1], **Urooj Yousuf** [1], **Muhammad Ramzan** [1], **Muhammad Taha Shaikh** [1], **Salman Akber** [1]

[1] College of Computer Science and Information Systems, Institute of Business Management, Karachi, Pakistan

Keywords: *Educational Data Mining (EDM), Feature Selection, LASSO Regression, Regression Analysis, Regularization Techniques, Academic Performance Prediction*

Journal Info:
Submitted: April 02, 2025
Accepted: April 29, 2025
Published: May 04, 2025

Abstract

Over the past few years, the digitalization trend has implied that educational activities have gained a bulk of rich information from the entire interaction between students, assessments, and learning management systems. Such a huge amount of information is now creeping into Educational Data Mining (EDM) to abstract meaning into action plans that aim to enrich the learning process and personalized learning experiences. However, one would face challenges where high dimensionality and redundancy are concerned regarding model accuracy, interpretability, and efficiency. In all the scenarios, the Least Absolute Shrinkage and Selection Operator LASSO emerges as one of the most powerful solutions, as it does both regression and feature selection by shrinking the insignificant coefficients to zero. The model works effectively in EDM to find relevant features such as attendance, quiz scores, and study habits while filtering away other noises. This paper provides a systematic approach to mathematical groundwork, geometric interpretation, and the use of LASSO in application. Experiments conducted on synthetic datasets and real datasets, such as the UCI Student Performance dataset, comprise evidence that highly advanced and complex models could improve predictive accuracy. Limitations are discussed along with practical considerations and future research directions regarding the usefulness of LASSO in next-generation educational analytics.

***Correspondence author email address:** mustafa.ahmed@iobm.edu.pk
DOI: 10.21015/vtse.v13i2.2111

1 Introduction

The imminent digital transformation will lead to unparalleled numbers and varieties of data generated by learners and usher in the redesigning of educational systems. Examples are user demographic information, course interaction logs with students, assessment records, attendance, Learning Management System (LMS) logging, and many other databases.

The broad purpose of Educational Data Mining (EDM) is to derive valuable actions from that data to improve teaching methods, predict students' academic performance, mark students as at-risk, and personalize learning experiences [1, 2]. However, when we consider educational datasets, we will find that the datasets typically contain a much larger number of features that aren't necessarily relevant or redundant. Therefore, high-dimensional feature sets make model interpretation complex and are computationally inefficient while degrading prediction performance. That is why feature selection is an essential preprocessing step within EDM pipelines. Feature selection would then help to identify the most relevant features, such as attendance or assignment submission behaviors, quiz scores, or forum participation, as those features would most likely stand out in predicting student success or engagement [3–5].

Many feature selection techniques are available, and the emergence of the Least Absolute Shrinkage and Selection Operator (LASSO) is crucial among them. The LASSO technique improves upon linear regression models by applying L1 regularization, which gives a penalty for the absolute size of the regression coefficients. This adds a sparsity aspect to regression, so irrelevant features are eliminated automatically by forcing their corresponding coefficient estimates to zero [6, 7].

LASSO is an easy way through EDM to interpretability and model simplicity and, therefore, a focus for educators and researchers on the most influential variables affecting students' outcomes. For instance, it's possible to determine with LASSO whether early assignment submission patterns or time on LMS pages are relatively more significant indicators of student achievement than historical Grade Point Average (GPA) or demographic attributes [8–10].

This paper comprehensively describes LASSO in EDM. It explicates the method's theoretical bases, provides geometric and mathematical intuitions, and illustrates with actual datasets from educational institutions. It also discusses experimental results showing what LASSO can do for feature selection in learning environments and for improving model generalization. The paper concludes with practical guidelines and future potential in applying LASSO in advanced EDM settings.

2. Literature Review

Feature selection is an accepted paramount necessity in EDM, given that digital tools and learning management systems propel massive datasets into dimensions. Selecting and preserving the ximost-informative attributes in educational datasets directly affects predictive models intended for student performance analysis, early warning systems, and personalized learning interventions by ensuring their good measurement, transparency, and effectiveness [11].

Traditional feature selection techniques used in EDM fall into three categories: filter methods, wrapper methods, and embedded methods. Filter methods include statistical methods like correlation analysis, mutual information, and chi-square tests for ranking features based on statistical association with the dependent variable. However, despite their efficiency, they tend to neglect between-variable interactions. For example, forum participation and quiz grades may appear weakly correlated with student performance, but their interaction might have strong predictive power [12].

Wrapper methods, such as backward and forward selection and recursive feature elimination, evaluate subsets of features by training models repeatedly. These methods are very effective, but they are highly computation-intensive and quickly suffer from overfitting, especially in educational data that may be sparse or unbalanced [13].

Embedded methods, such as the decision tree and regularized regression feature selection algorithm, carry out feature selection and model training. Among these is LASSO regression, which became popular in EDM due to its unique capability of conducting regression and variable selection simultaneously by shrinking the coefficients for negligible variables to zeros [14].

Studies on applying LASSO have successfully identified specific behavioral indicators for final examination scores and dropout risk in LMS platforms, such as login sessions, attempts at assignments, forum activity, or time spent on a task. With Massive Open Online Courses (MOOCs), LASSO has reduced dimensionality in learner interaction logs to actionable insights. Other studies employed LASSO in formulating interpretable early warning systems that can inform instructors about real-time data of at-risk students [15].

It shows that LASSO also effectively models learning behavior by demographic groups. LASSO, for example, helps unearth which engagement metrics are more telling for first-generation students or which ones help different learners as appropriate in multi-institutional studies. As mentioned above, LASSO's applicability to high-dimensional data makes it a great candidate for analyzing intelligent tutoring systems, e-learning platforms, and educational sensor data (eye-tracking or keystroke dynamics) datasets [16].

Recent advances have seen Elastic Net Regularization proposed to mitigate correlated features (frequently occurring in educational datasets), combining LASSO's L1 penalty and Ridge's L2 penalty. Adaptive LASSO and Group LASSO allow for enhanced model interpretability and stability, making them appropriate for varying hierarchical structures of educational data [17, 18].

The Standard LASSO performs feature selection by shrinking some coefficients to zero, thereby promoting sparsity. However, it often selects one of the strongly correlated variables as the predictor and disregards the rest. In this case, the Elastic Net utilizes both penalties of L1 and L2, which provides grouping of selection within grouped correlated features; therefore, it has greater stability in such scenarios. Group LASSO takes this a step further by selecting groups of variables in their entirety based on some predefinition. This is important if features form natural clusters, such as behavioral or performance indicators in educational data. Hence, both Elastic Net and Group LASSO provide a better alternative to standard LASSO when correlated or grouped predictors are present [19].

The present article promotes findings in this area. It seeks to contextualize further the power and usefulness of LASSO in the context of EDM, mainly when applied to discerning latent predictors of academic success, learner engagement, and educational equity.

3. Mathematical Formulation and Geometric Interpretation (in Educational Data Mining)

3.1 Linear Regression in EDM

In EDM, linear regression methods are assumed to predict continuous educational variables such as final student grades, cumulative GPAs, or graduation time. These models upload specific educational attributes—including attendance, quiz marks, frequency of LMS activity, timely submission of assignments, and engagement in forums—as contributions to academic performance [20].

A standard form of the linear regression model is:

$$\hat{y}_i = \sum_{j=1}^{p} w_j x_{ij} + \varepsilon_i$$

Where:

- $\hat{y}_i$ is the predicted academic outcome (e.g., final grade) for student i,
- x_{ij} is the value of the j-th feature for student i (e.g., number of logins),
- w_j is the coefficient representing the importance of feature j,
- ε_i is the error term.

Traditional linear regression minimizes the residual sum of squares (RSS), which is a measure of deviations between actual and predicted outcomes:

$$J(w) = \frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

3.2 LASSO Regression for Sparse Feature Selection

This then introduces L1 regularization within the framework and sets up a penalty based on the absolute values of the coefficients [21]. The following transformation makes the optimization objective:

$$\min_{w} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \alpha \sum_{j=1}^{p} |w_j| \right\}$$

The penalty term $\alpha \sum |w_j|$ promotes smaller coefficients; in many cases, it drives some coefficients precisely to zero, their further exclusion from the model. In EDM, the model will automatically identify and eliminate features with less impact, such as rarely used LMS tools or minimally accessed learning resources. By another formulation, LASSO can alternatively be expressed in a constrained manner:

$$\min_{w} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \right\}, \quad \text{subject to } \sum_{j=1}^{p} |w_j| \leq t$$

Where t controls the allowable total weight for all features, smaller t values lead to more zero coefficients.

3.3 Geometric Interpretation in EDM

From a visualization standpoint, the optimization process can be inspired by contours of cost functions and constraint regions. The cost function would contour ellipses, whereas the L1 constraint area would look like a diamond (L1-norm ball). So, the solution is optimal when the contour of the lowest cost function cuts through the constraint border [22].

In EDM contexts (discussed in a later section), this point of intersection mainly lies on one of the axes, which means that the corresponding coefficient (and feature) would be zero. For example, if it is learned that "student device type" doesn't have much predictive power, its weight is dropped in favor of more predictive variables such as "average quiz score" or "late assignment submission."

3.4 Regularization and Educational Insight

The regularization parameter α has been a crucial factor in balancing model sparsity with prediction accuracy. Higher α encourages the development of simpler models with fewer features, making the results interpretable to educators and stakeholders for making data-driven decisions [23].

Lastly, tuning α-as, routinely done via cross-validation, allows practitioners to confer on the model mere indicators of learners' behavior and performance that are very helpful in meaning and student support planning.

4. Implementation and Practical Considerations in Educational Data Mining

4.1 Importance of Feature Selection in Practical EDM Models

The quality of educational data is generally rich and diverse, but suffers from high dimensionality and feature redundancy. For instance, LMS logs capture dozens of activity metrics per student: login frequency, click patterns, time spent on different course materials, etc. Each feature is not necessarily informative or contributes to predicting learning outcomes such as course completion, performance, or dropout risk.

LASSO regression provides a rather utilitarian approach to automation and dimensionality reduction. It allows researchers to winnow datasets to highlight only features with the most potent predictive power, making the models more interpretable to educational stakeholders, such as instructors, counselors, or institutional policymakers [24].

4.2 Regularization Parameter Tuning

The effectiveness of LASSO in EDM relies heavily on setting a proper regularization parameter α that determines the penalty strength between the model coefficients. Higher values of α pull more coefficients into a null point, making the model more straightforward but possibly leading to a loss of accuracy if penalized excessively [25].

Cross-validation is the most widely used method for tuning α in the EDM context. In cross-validating α, the data is spanned into several folds, and the performance of the models computed on different α values can be found to derive the best compromise between sparsity and accuracy. Therefore, the only relevant predictors will remain in the final model, such as weekly LMS activity, attendance score, or formative assessment results.

4.3 LASSO Implementation in EDM Using Python (scikit-learn)

Given below the code block which is a standard implementation pipeline used in EDM research, making use of the Python package known as scikit-learn;

```
from sklearn.linear_model import LassoCV
from sklearn.Preprocessing import StandardScaler
from sklearn.pipeline import make_pipeline

# Example: Predicting final grade based
on student behavioral and demographic features
model = make_pipeline
(StandardScaler(), LassoCV(cv=10))
model.fit(X_train, y_train)
```

• `StandardScaler` guarantees that all attributes are on an equivalent scale (of paramount importance for LASSO).
• `LassoCV` performs hyperparameter tuning automatically through cross-validation.
Once the model is trained, selected features can be extracted using:

```
selected_features = X_train.columns[model.named
_steps['lassocv'].coef_ != 0]
```

Then, these selected features can be interpreted to understand which student behaviors or characteristics most influence performance, thereby supporting early intervention or course redesign [26, 27].

4.4 Interpretability in Educational Settings

A key benefit of LASSO in EDM is its interpretability. In contrast to deep neural networks, which are black-box models, LASSO results in a sparse set of non-zero coefficients directly associated with definite variables [28]. This transparency allows:

- Educators to zero in on indicators that have an actual impact (e.g., low quiz scores or inconsistent LMS use)
- Institutions to improve their learning analytics dashboards
- Counselors to frame personalized support approaches based on the most highly influential factors on at-risk students.

4.5 Limitations and Extensions

LASSO can be seen as powerful but limited by the following:

- The algorithm may select only one variable from a group of correlated variables (e.g., login vs. time spent on the course)
- Model instability: Small changes in the training data could induce an entirely different selection of features
- May incur bias on coefficient estimation

To address these issues, it is usually advised to use Elastic Net Regularization, which combines L1 with L2 penalties, especially when correlated features are standard, as in EDM datasets. Group LASSO and Adaptive LASSO further improve the selection behavior for hierarchical or categorical educational features [29].

5. Experimental Results in Educational Data Mining

To prove the effectiveness of LASSO regression in feature selection for educational situations, we created two experiments: one on synthetic education datasets and the other on real-world student performance data. The aim is to show LASSO's capacity to eliminate irrelevant features that are not strong predictors of academic performance.

Before applying the LASSO model, preprocessing was done on the actual dataset. Missing values, where applicable, were imputed to render the dataset as complete as possible. Also, all features underwent normalization (or standardization) to put them on a comparable scale, an important step because the penalty of the LASSO term is sensitive to feature value magnitudes, which could skew the selection process if unscaled. Preprocessing, therefore, supported a balanced and efficient selection of features while enhancing the stability and readability of the results from the LASSO model.

5.1 Synthetic Educational Dataset Evaluation

Synthetically generated data, such as 500 students and 30 features, was created to replicate a typical learning analytics scenario. Included in its features were some common behavioral and demographic factors, like:

- In number of logins into LMS per week,
- Time spent on lecture videos,
- Quiz scores,
- Forum participation in the class,
- On-time submission of assignments,
- Device type used and
- Internet speed, etc.

For the final course grade as the target variable, 8 out of 30 features were designed to be directly related, while the remaining were to be used as noise or irrelevant information. LASSO was employed with varying regularization parameters α in radians, and results were hence recorded as [30]:

- Number of non-zero coefficients (selected features),
- Root Mean Squared Error (RMSE) on the test set.

Table 1. Impact of Alpha on Feature Selection and Prediction Accuracy

Alpha	Non-zero Coefficients	Test RMSE
0.01	28	5.6
0.1	17	4.4
0.5	10	3.8
1.0	8	3.9
2.0	5	5.1

The table illustrates how the selection of the regularization parameter α affects the number of chosen features while prediction accuracy changes in a student performance prediction model. Extremely low α values (0.01) attract most of the features (28) into the model, leading to a relatively high test RMSE of 5.6 due to overfitting caused by many irrelevant features. As the α value increases, LASSO gradually shrinks coefficients to zero, reducing the number of selected features. At the $\alpha = 0.5$ case, about 10 features are left, which has a test RMSE computation for the attaining minimum value of 3.8, which is considered a good compromise between model complexity and prediction performance. Raising the α further to 1.0 and 2.0 continues the decline of features selected but subsequently pushes the RMSE higher again, indicating underfitting from the absence of significant predictors. In EDM, it demonstrates the importance of tuning α carefully so as not to either overfit or underfit, thus maximizing the chances of the model to unearth crucial behaviors of students and performance indicators.

It has been shown that increasing the value of α boosts the number of features selected while minimizing the loss of prediction accuracy when the selected set closely resembles the truly informative features.

5.2 Real-World Educational Dataset: Student Performance

The real-world application was the so-familiar Student Performance Dataset from the UCI Machine Learning Repository with attributes [31] like:

- Parental education level,
- Study time per week,
- Failures in past courses,
- Internet access at home,
- Absences,
- School support programs,
- Gender, etc.

The LASSO regression model was employed on 10-fold cross-validated optimal values of α to arrive at the model, which shortlisted 6 out of 16 identifiable features, such as:

- Study time,
- Number of past failures,
- Attendance (absences),
- Parental education,
- School support,
- Internet access.

Table 2 compares the performance of LASSO and Ordinary Least Squares (OLS) regression models applied to educational datasets on tasks such as predicting student final grades or course completion [32].

Table 2. Performance Comparison of LASSO and OLS Regression in Student Performance Prediction

Metric	Value
Optimal Alpha	0.15
Selected Features	6 of 16
Test RMSE (LASSO)	4.7
Test RMSE (OLS)	5.3

The table above summarizes the performance of LASSO regression with cross-validation-based automatic alpha tuning under an educational data mining context. The alpha value finally chosen was 0.15, which allowed only 6 of the original 16 features to be retained. This shows how LASSO can effectively eliminate irrelevant or redundant variables, such as unimportant metrics on LMS interactions, while keeping only the most predictive ones: quiz scores, assignment scores, and attendance. In addition, the Test RMSE calculated from the LASSO model is 4.7, substantially lower than the 5.3 obtained from the classical Ordinary Least Squares (OLS) regression technique. Thus, LASSO not only narrows down features, simplifying the model, but also increases the odds of accurate prediction. In EDM, this means that LASSO facilitates the identification of factors that matter most concerning student performance for instructors and administrators, thereby making a case for data-driven interventions and policy decisions.

It was interesting to observe that the model's performance and interpretability would improve when a focused set of actionable features was identified through LASSO rather than conventional regression models.

5.3 Discussion

These experiments validate that LASSO can:

- Select only the most impactful educational features,
- Reduce model complexity,
- Improve prediction accuracy,
- Provide better interpretability for stakeholders in education.

Additionally, an automation tool's ability to eliminate features is particularly valuable for institutions involved in large-scale learning analytics, where manual feature engineering is neither scalable nor feasible.

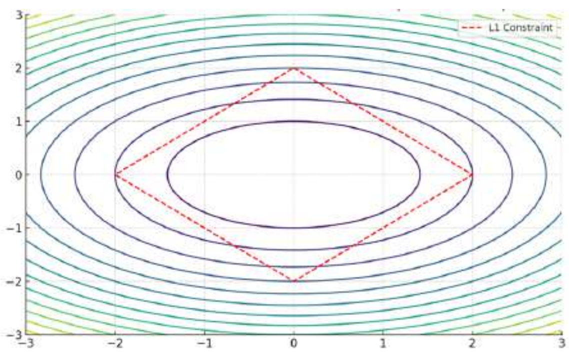

Figure 1. LASSO Cost Contours and L1 Constraint Geometry

Figure 1 visually explains the core concept behind LASSO feature selection. The concentric elliptical lines represent the level curves of the loss function in a typical linear regression setting — in this case, it could represent the error of predicting students' final grades using multiple educational features such as attendance, quiz scores, and forum activity. Each contour corresponds to a set of coefficient values (w1, w2) that result in the same model error. The red diamond-shaped region illustrates the L1 constraint imposed by LASSO, which limits the total absolute sum of the coefficients. What makes this constraint special is its sharp corners aligned with the coordinate axes. This geometric structure encourages the optimal solution — the point where the most minor possible elliptical contour touches the constraint boundary — to often lie directly on the axes. Practically, one or more coefficients will be forced to zero. In EDM, this leads to automatic feature selection, where less critical variables (e.g., 'device type' or 'internet speed') are excluded. In contrast, the most relevant features (e.g., 'study time' or 'quiz performance') are retained. The intersection between the ellipse and the diamond is the mathematical representation of this feature selection process.

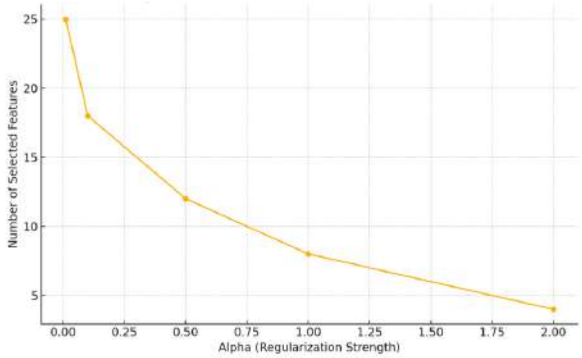

Figure 2. Effect of Alpha on Number of Selected Features

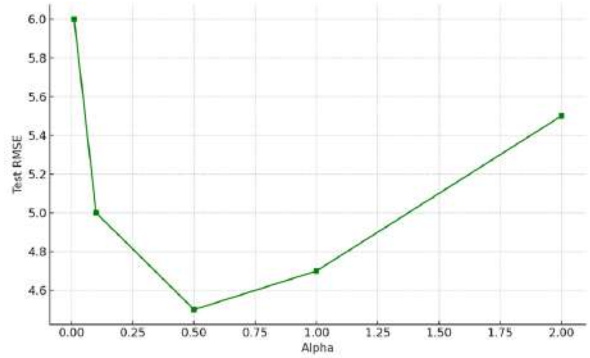

Figure 3. Relationship Between Alpha and Prediction Error (RMSE)

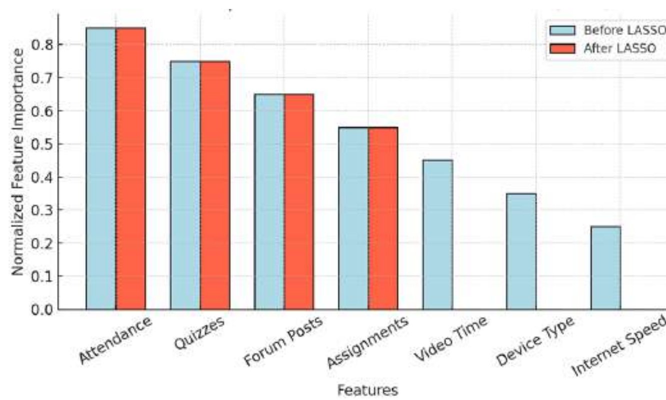

Figure 4. Feature Importance Before and After LASSO in EDM

Figure 2 depicts the direct relationship established between the regularization parameter and the number of features selected by LASSO regression modeling in educational data mining. The x-axis displays steadily increasing values of , which measures the intensity of the penalty inflicted upon the model's coefficients. At the same time, the y-axis unveils the number of features selected (i.e., features that benefited from non-zero coefficients after applying the LASSO effect). A clear indication can be derived that for higher values of , the fewer the selected features will be. When it is approximately zero, most features, including many arbitrary ones, are retained. Still, at higher values of , LASSO reduces weights of values less important towards zeros, concerning student performance prediction, at greater values of , only the most relevant predictors (assignment scores, quiz performance, and attendance) remain. In contrast, useless or noisy predictors (like device type or session duration) are discarded. This quality adds incredible weight to LASSO for yielding straightforward yet interpretable models concentrated on essential learning indicators.

Similarly, Figure 3 depicts a relationship between the regularization parameter and the prediction error measured using RMSE for some educational data mining tasks. As the plot notes, the curve shows a rapid decrease in RMSE as increases from 0 to approximately 0.5, indicating that moderate regularization helps improve generalization and decreases the degree of overfitting in the model. At this point, LASSO has set aside the noisy or irrelevant features, such as infrequent forum visits or minor login activities. At the same time, the model may focus more on the most salient predictors of student performance, like assignment scores, attendance, or study time. However, as increases past 0.5, we receive an increase in RMSE. This represents the standard compromise between regularization and underfitting: imposing too much regularization also suppresses influential features, which deteriorates the model's ability to predict. Applying such force on regularization in educational data mining terms could mean eliminating some pertinent learning behaviors if is highly penalized. The plot indicates that an intermediate (approximately 0.5 in this case) offers the best trade-off in balancing model complexity versus prediction accuracy; hence, it is a critical hyperparameter to tune in LASSO in analyzing student performance data.

Figure 4 illustrates LASSO as a technique in feature selection that shrinks less significant coefficients to zero. On the x-axis, these features are standard to educational data, such as Attendance, Quizzes, Forum posts, Assignments, Video Time, Device Type, and Internet Speed. The y-axis symbolizes the normalized importance of these features before and after applying LASSO. Before LASSO, all the features were necessary at some level, with attendance, quizzes, and forum posts being the few most influential predictors of student performance. Then, after applying LASSO regularization, the importance of some features, that is, video time, device type, and internet speed, came to zero. This means that they were declared unimportant for predicting the target variable, for instance, student grades or success, in this EDM scenario. The core behaviors related to academics—like Attendance, Quizzes, Forum Posts, and Assignments—retained their importance, which confirmed their predictive capability. Thus, this visualization shows LASSO's ability to automatically filter out irrelevant or redundant features, allowing more straightforward, more interpretable, and focused predictive models for educational stakeholders.

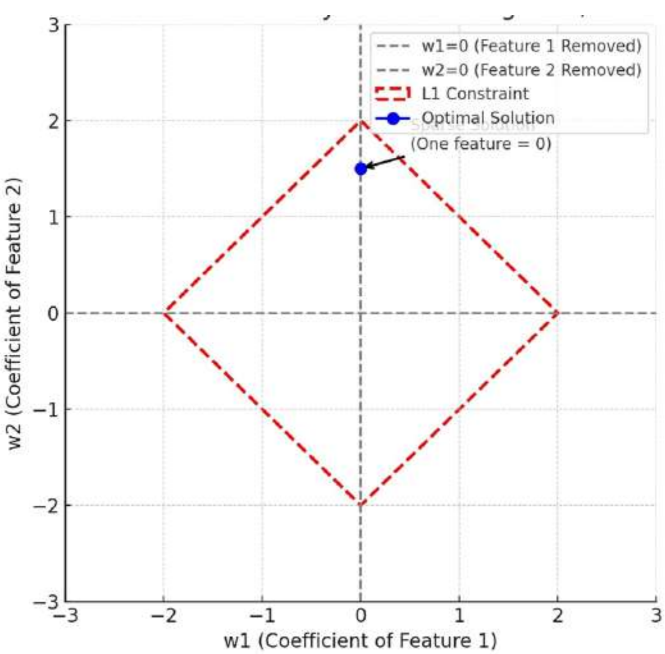

Figure 5. LASSO Decision Boundary Illustrating Feature Elimination

One can visually inspect in Figure 5 that LASSO induces sparsity into model coefficients when predicting binary outcomes such as a student's passing or failing status in educational data mining. The plot contains two axes showing two hypothetical features participating in the prediction task: average assignment score (w1) and forum participation frequency (w2). The red diamond signifies the L1 constraint enforced by LASSO, which restricts the total absolute sum of the coefficients to reside within those bounds. The dashed gray lines signify cases where either w1 or w2 is set to zero, thus excluding its corresponding feature from the model. The blue dot depicts the optimal solution, which lies precisely on the w1 = 0 axis; therefore, the model has successively removed Feature 1 from the decision boundary by assigning it a coefficient of zero. This geometrical behavior transmits the very idea behind LASSO's feature selection mechanism. In real educational applications, this will translate to keeping only one behavioral variable (perhaps forum participation) as a significant predictor of a student's pass/fail status. In contrast, other variables (assignment score) might be considered non-significant with different features. Such feature selection behavior makes LASSO a powerful machine for producing interpretable models in EDM.

6. Conclusion and Future Work in Educational Data Mining

The paper elaborated on the effective utilization of LASSO as a superior feature selection in EDM. Now that we are implementing data-driven decision-making, developing systems for deriving meaningful knowledge from high-dimensional educational datasets becomes critical. Camofy-theoretical perspective, geometry intuition, and practical implementation prove how LASSO will benefit predictive accuracy and interpretability for educators and other stakeholders. In this way, the system will automatically eliminate irrelevant features using the LASSO approach and its intra-coefficient shrinkage to zero. This action will help work with complex and heterogeneous educational datasets such as LMS logs, demographic data, and behavioral metrics.

Synthetic and real students' performance datasets show that LASSO emphasizes better performance on the model while shrinking complexity. By only selecting the most interesting features of study time, assignment behavior, or class attendance, LASSO focuses on intervention where the need is greatest. This makes it a powerful tool for modeling and developing actionable insights in teaching and learning environments. The practicalities that would less often engage the attention of a typical educational research paper include technical issues such as data preprocessing, parameter tuning, and model evaluation techniques relevant to real-world educational scenarios. Indeed, LASSO's data processing and analysis capabilities via tools such as scikit-learn also open up its promise for use in institutional analytics pipelines.

Although the application of LASSO has made headway in EDM, several different avenues of research have not yet been fully explored. First, such techniques as Elastic Net and Group LASSO provide more stable and structured feature selection, important in selecting educational variables with a hierarchical structure or correlation, for example, nested activity types within a course module. Second, it is clear that some temporal and sequential modeling is needed because student behavior is changing. New extensions, such as time-aware LASSO and dynamic models, offer more effective ways of modeling the learning process and predicting outcome correctness based on learning progression.

Third, advanced regularization methods would be needed for multimodal data fusion, where textual data (e.g., forum posts), interaction logs, and sensor data (e.g., eye-tracking) are combined with numerical features to create an enormous research area. Besides, an equity and fairness-aware LASSO could ensure that the models stay unbiased and fair for the diverse groups of students, addressing a pressing concern of fairness in education. Finally, LASSO can also become an essential ingredient in explainable AI or XAI systems as they pertain to education, allowing educators to predict student outcomes while being able to understand and act on insights derived from interpretable models.

With that, LASSO stands out as a robust, understandable, and scalable technique for feature selection within EDM. Such systems will continue digitizing and personalizing education systems, and trade like LASSO will keep turning raw student data into real intelligence.

Author Contributions

Mustafa Ahmed Khan: Conceptualization, methodology design, formal analysis, and supervision. **Khalid Mahboob**: Data collection, writing the original draft, and proofreading. **Urooj Yousuf**: Data cleaning, feature engineering, and preparation of datasets for LASSO modeling. **Muhammad Ramzan**: Implementation of LASSO, algorithm testing, and result interpretation. **Muhammad Taha Shaikh**: Visualization, writing of the initial draft, and literature review. **Salman Akber**: Critical review, validation, and editing of the manuscript.

Compliance with Ethical Standards

The authors declare no conflict of interest. This research study does not involve collecting data from human participants.

Funding Information

Funding information has not been received from an institution for this research.

References

[1] S. M. F. D. M. S. Mustapha, "Predictive analysis of students' learning performance using data mining techniques: A comparative study of feature selection methods," *Applied System Innovation*, vol. 5, no. 6, 2023.

[2] K. Roy and D. M. Farid, "An adaptive feature selection algorithm for student performance prediction," *IEEE Access*, vol. 12, pp. 75 577–75 598, 2024.

[3] H. C. Oz, Güven, and G. Nápoles, "School dropout prediction and feature importance exploration in malawi using household panel data: machine learning approach," *Journal of Computational Social Science*, vol. 6, pp. 245–287, December 2023.

[4] C. Jalota and R. Agrawal, "Feature selection algorithms and student academic performance: A study," in *Inter-*

national Conference on Innovative Computing and Communications, Singapore, 2021.

[5] R. L. Silva Filho, K. Brito, and P. J. L. Adeodatato, "A data mining framework for reporting trends in the predictive contribution of factors related to educational achievement," *Expert Systems with Applications*, vol. 221, 2023.

[6] M. Freo and A. Luati, "Lasso-based variable selection methods in text regression: the case of short texts," *AStA Adv Stat Anal*, pp. 69–99, 2024.

[7] J. Gui, Z. Sun, S. Ji, D. Tao, and T. Tan, "Feature selection based on structured sparsity: A comprehensive study," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 7, pp. 1490–1507, July 2017.

[8] C. Wang, T. Li, Z. Lu, Z. Wang, T. Alballa, S. A. Alhabeeb, M. S. Albely, and H. A. E.-W. Khalifa, "Application of artificial intelligence for feature engineering in education sector and learning science," *Alexandria Engineering Journal*, vol. 110, pp. 108–115, 2025.

[9] J. C. Immekus, T.-s. Jeong, and J. E. Yoo, "Machine learning procedures for predictor variable selection for schoolwork-related anxiety: evidence from pisa 2015 mathematics, reading, and science assessments," *Large-scale Assess Educ*, vol. 10, 2022.

[10] L. Cohausz, A. Tschalzev, C. Bartelt, and H. Stuckenschmidt, "Investigating the importance of demographic features for edm-predictions," in *Proceedings of the 16th International Conference on Educational Data Mining*, Bengaluru, India, 2023.

[11] Z. Y, Y. Y, A. R, C. J, D. H, and S. X, "Educational data mining techniques for student performance prediction: Method review and comparison analysis," *Front. Psychol.*, vol. 12, 2021.

[12] U. Stańczyk, *Feature Evaluation by Filter, Wrapper, and Embedded Approaches*. Berlin, Heidelberg: Springer, 2014, vol. 584.

[13] S. Bashir, I. U. Khattak, A. Khan, F. H. Khan, A. Gani, and M. Shiraz, "A novel feature selection method for classification of medical data using filters, wrappers, and embedded approaches," *Complexity*, vol. 2022, 2022.

[14] H. Deng and G. Runger, "Feature selection via regularized trees," in *International Joint Conference on Neural Networks (IJCNN)*, Brisbane, QLD, Australia, 2012.

[15] S. V, M. Y, K. M, and D. S, "Harnessing machine learning models and explainable ai to understand mooc continuance intention," *Information Discovery and Delivery*, 2025.

[16] F. Forouhideh and H. Aliakbarimajid, "From description to prediction: unveiling student performance in online learning through data-driven analysis and machine learning," 2023–24.

[17] J. C. Obi and I. C. Jecinta, "A review of techniques for regularization," *International Journal of Research in Engineering and Science*, vol. 11, no. 1, pp. 360–367, 2023.

[18] M. E. Anbari and A. Mkhadri, "Penalized regression combining the l 1 norm and a correlation based penalty," *Sankhya B*, vol. 76, pp. 82–102, 2014.

[19] L. Liu, J. Gao, G. Beasley, and S. H. Jung, "Lasso and elastic net tend to over-select features," *Mathematics*, vol. 11, no. 17, pp. 4531–4543, 2023.

[20] M. El Jihaoui, O. El Kheir Abra, and K. Mansouri, "Factors affecting student academic performance: A combined factor analysis of mixed data and multiple linear regression analysis," *IEEE Access*, vol. 13, pp. 15 946–15 964, 2025.

[21] M. Iloska, P. M. Djurić, and M. F. Bugallo, "Fast sparse learning from streaming data with lasso," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing*, Hyderabad, India, 2025, pp. 1–5.

[22] A. Dutta, K. Lakshmanan, R. Karthik, S. S. Priya, and A. Ramamoorthy, "An ellipsoid restrictive region-based regularization for regression analysis," *International Journal of Information Technology (Singapore)*, vol. 17, no. 3, pp. 1865–1871, 2024.

[23] S. Christina, S. Sowjanya, C. Lakshmhyma, L. Prathiba, and M. S. Amzad Basha, "Data-driven insights into student performance: Benchmarking machine learning models for grade prediction using regression and classification approaches," in *2025 International Conference on Intelligent Systems and Computational Networks (ICISCN)*, Bidar, India, 2025, pp. 1–6.

[24] Y. Luo and Z. Wang, "Feature mining algorithm for student academic prediction based on interpretable deep neural network," pp. 1–5, 2024.

[25] B. Bouihi, A. Bousselham, E. Aoula, F. Ennibras, and A. Deraoui, "Prediction of higher education student dropout based on regularized regression models," *Engineering, Technology and Applied Science Research*, vol. 14, no. 6, pp. 17 811–17 815, 2024.

[26] J. C. J. Luza and C. Rodriguez, "Predictive attributes in machine learning for university academic performance: A feature engineering approach," in *2024 IEEE 16th International Conference on Computational Intelligence and Communication Networks (CICN)*, Indore, India, 2024, pp. 443–456.

[27] S. Gupta, B. Kishan, and P. Gulia, "Comparative analysis of predictive algorithms for performance measurement," *IEEE Access*, vol. 12, pp. 33 949–33 958, 2024.

[28] I. Papadogiannis, M. Wallace, and G. Karountzou, "Educational data mining: A foundational overview," *Encyclopedia*, vol. 4, no. 4, pp. 1644–1664, 2024.

[29] S. Pirenne and G. Claeskens, "Parametric programming-based approximate selective inference for adaptive lasso, adaptive elastic net and group lasso," *Journal of Statistical Computation and Simulation*, vol. 94, no. 11, pp. 2412–2435, 2024.

[30] A. Abdelhadi, S. Zainudin, and N. S. Sani, "A regression model to predict key performance indicators in higher education enrollments," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, pp. 454–460, 2022.

[31] P. Cortez, "Student performance," https://doi.org/10.24432/C5TG7T, 2008.

[32] S. Ansari and A. B. Nassif, "A comprehensive study of regression analysis and the existing techniques," in *2022 Advances in Science and Engineering Technology International Conferences (ASET)*, Dubai, United Arab Emirates, 2022, pp. 1–10.