IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Integrating SHAP Explainability With Traditional Feature Selection Methods For Student Performance Prediction

[1]S.Suvetha, [2]Dr.C.Immaculate Mary,

[1]Research Scholar, [2]Head and Associate Professor,

[1]PG and Research Department of Computer Science,

[1] Sri Sarada College for Women(A), Salem, India

Abstract: Student performance prediction is a crucial research domain which analyses the student data aiming to anticipate academic outcomes, analyze student academic performance and to identify student at risk. In this research paper one of the efficient approaches called SHAP (SHapley Additive explanations) which is mathematically grounded with game theory is used to make feature selection and final predictions. SHAP provides both global explanations (which feature is important overall) and local explanations (why a prediction was made for a particular student). In this study, we employ different machine learning algorithm Decision Tree, Random Forest, Lasso Regression, and XGBoost with SHAP based analysis for feature selection. The work is carried out with k-fold validation technique, feature overlaps and residual plots. Paired t-test is also used to verify results statistically. Results showed that SHAP integrated with Random Forest importance based selection identified a more stable and interpretable feature subset, outperforming other combinations. The study concludes that integrating SHAP with traditional feature selection improves both predictive reliability and explainability.

Index Terms - Student performance, Explainable AI, SHAP, Random Forest, Educational Data Mining.

I. INTRODUCTION

In the area of Educational Data Mining (EDM), student performance prediction has become vital research areas. It helps educational institutions to enhance learning outcomes, reduce dropout and failure rates, identify student at risk, provide timely interventions and to draw efficient decision making for overall teaching and learning quality. The process of analyzing academic progress, behavior and socio-economic background of student empower educational institutions to make data-driven policies and strategies to improve student academic improvements. In this regard, machine learning holds an upper hand in place of contributing valid prediction accuracy. The main challenge in machine learning is feature selection. It refers to identify most relevant features that significant influencing the output and discarding the other features. Using all the features

that exist in dataset will lead the model to make invalid predictions, increases model complexity and reduces accuracy. However, predictive accuracy alone is not sufficient for practical implementations, they need to understand why such predictions have been made in order to accept the predictions given by the models. Traditional machine learning models such as Decision **Trees, Random Forests, XGBoost, and Lasso Regression** will produce strong predictions but they lacks behind to provide transparent explanations. Without making valid interpretations, educators might hesitate to use that prediction for decision making thinking that might affect student academic futures.

Explainable AI (XAI) aims to solve this problem by providing a frame work which will explain the predictions in a human understandable manner. Among different XAI method, SHAP (SHapley Additive explanations) has emerged as a leading approach and used in this research work. The SHAP is completely based on shapely values cooperative game theory which will issue credit score among features in a fair manner. By this way it ensures the consistency and accuracy are valid.

Based on important scores or coefficient made form the traditional models features are ranked and selected accordingly. While these methods lack in transparency in explaining why such features are considered for analysis. To overcome this short fall **SHapley Additive exPlanations (SHAP)** is incorporated. In this research work, we proposed an **integrated framework that combines traditional feature selection methods with SHAP explainability** for predicting the student performance. the dataset is sourced from UCI repository which includes demographic, socio-economic, behavioral, and academic results with their final grade (G3).

This main objective of the study,

1. To deploy automatic feature selecting algorithms such as Decision Tree, Random Forest, XGBoost, and Lasso regression based its importance or coefficient.
2. To apply SHAP based feature selection and compare with traditional feature selection methods.
3. To implement hybrid approach of combing model feature selection with SHAP based selection and evaluate the model with various metrics such as cross-validation, error metrics and finding feature overlay.
4. To analyse **local SHAP explanations for individual students**, highlighting how factors such as study time, failures, and absences impact specific predictions.

The personalized insights derived from models not only helps to improve the interpretations it also helps to make actionable recommendations for educators. This research works highlights the importance of incorporating the explainability based feature selection in field of education data mining. This proposed work bridges gap between machine learning models and practical decision makings.

II. LITERATURE REVIEW

In [1], The author Yan (2021) compared XGBoost against Random Forest, Lasso, Elastic Net, SVM, and Decision Tree for student performance prediction dataset taken from UCI repository. The result shows that XGBoost consistently outperformed other models, improving R^2 by 6.3% to 12.1%, and included feature engineering.

In [2], a decision tree study ("Students' performance prediction employing Decision Tree") showed that even without academic attributes, traditional models can achieve ~93% accuracy. Using full dataset, SVM scored ~95%. The Decision Tree revelas important and influencing features such as past results, family education.

In [3], The author Harif & Kassimi (2024) used a proposed technique named RFECV-RF (Random Forest + recursive feature elimination with cross-validation) to select optimal features, then tested classifiers and found SVM (with subset of 8 features) reached ~87% accuracy.

In [4], Malik et al. (2025) introduced DE-FS, a dynamic ensemble feature selection combining correlation analysis, information gain, Chi-square, and adaptive thresholding, showing outstanding performance on educational datasets.

In [5], Imran et al. (2017) applied preprocessing and classifiers (J48 / decision tree, NNge, MLP) techniques on student data taken from UCI repository. The results showed that J48 achieved 95.78% accuracy and validates the importance of both data preparation and model tuning.

In [6], Guleria (2024) the author implemented a blended proposed method MRMR, ReliefF, Chi-square, and ANOVA with SHAP values for educational dataset. Their results suggest hybrid SHAP selection improves both generalization and interpretability.

In [7], Wang et al. (2024) compared SHAP-based selection against model importance in fraud and tabular datasets; they found trade-offs between the importance methods and sometimes SHAP out performs in raw performance.

In [8], Sebastián et al. (2024) proposed Boruta-Shap, Shapicant, and other SHAP-based selection extensions, combining SHAP values with permutation and shadow features to yield more robust feature subsets.

In [9], Kraev et al. (2024) introduced *Shap-Select*, regressing the target on SHAP values to select features whose SHAP coefficients are significant and benchmarked favourably against RFE, Boruta.

In [10], Hancock et al. (2025) presented a general SHAP-based selection and reduction framework for various label settings. The result indicates that the models maintain performance even as there is decrease in future counts.

III. METHODOLOGY

The proposed methodology for student performance prediction is structured into six main phases: dataset preparation, preprocessing, model implementation, feature selection, evaluation, and explainability analysis.

3.1 Dataset Description

The dataset used in this study is taken from UCI repository. It contains academic, demographic and behavioral details of the students. The academic data such as their grade 1, grade 2 and grade 3 marks. Behavioral data such as study time, interest, activities, school up, travel time, guardian details, father mother education and occupation. The demographic details like age, higher (intends to higher education), internet are included. Some features have numerical features and some features have categorical values. In total, there are 649 student records with 33 attributes.

3.2 Data Preprocessing:

Data preprocessing on one of the curious steps in data analytics because if it is not handled properly it might lead to inconsistent and invalid results to make our prediction more accurate the preprocessing is essential. Raw data will have many null, missing values and duplicate values.

- Categorical attributes were encoded using one-hot encoding (low cardinality) and Label Encoding (high cardinality).
- Missing values in numeric attributes were imputed using the statistical method median of the respective feature.
- Non-numeric residual columns were transformed into numeric form using Label Encoding.
- A train-test split (70:30) was applied with a random seed of 42 for reproducibility.

3.3 Regression Models Implemented

Linear Regression: By simulating a straight-line relationship between input features and the target variable, linear regression provides a baseline for continuous outcome prediction. To estimate coefficients, the sum of squared residuals is minimized. It is easy to understand, but it makes the assumption that the predictors are independent and linear. When relationships are extremely non-linear or there is multicollinearity in the data, its efficacy may be restricted.

Ridge Regression: By including L2 regularization, which penalizes large coefficients, Ridge Regression builds upon Linear Regression. This reduces variance and helps control overfitting, especially when predictors are highly correlated. All features remain in the model, but their impact is reduced proportionately. When there are a lot of correlated variables in the dataset, it is especially helpful.

Decision Tree Regression: Decision Tree Regression is useful for identifying non-linear patterns because it divides data into regions according to feature thresholds. It can easily handle both categorical and numerical features. Single trees, however, frequently overfit, resulting in erratic predictions with a high variance. They are nevertheless interpretable and serve as the foundation for numerous ensemble approaches.

The Random Forest Regressor: Several Decision Trees constructed using bootstrapped samples and random feature subsets are combined by Random Forest. This ensemble preserves accuracy while lowering variance and overfitting. It can successfully capture intricate feature interactions and is resilient to noisy data. The model's excellent generalization performance across a variety of datasets has earned it widespread recognition.

Gradient Boosting : Gradient Boosting builds trees one after the other, fixing the mistakes of the previous tree. High predictive accuracy and robustness are produced by this iterative learning process. To prevent overfitting, it necessitates meticulous adjustment of parameters such as learning rate and depth. For structured data, gradient boosting is effective and frequently outperforms more straightforward models.

XGBoost Regressor: A refined version of gradient boosting, XGBoost adds sophisticated regularization and computational efficiency. It is well-liked in competitive machine learning because of its accuracy, scalability, and speed design. It efficiently manages big datasets thanks to parallelization and integrated cross-validation. It is a cutting-edge boosting framework because of its versatility and excellent performance.

Lasso Regression: The regression technique that enhances the model's performance by incorporating L1 regularization is called least absolute shrinkage and selection operator (Lasso regression), which places a limit on the absolute magnitude of the regression coefficients. By reducing a few feature coefficients to precisely zero, this restriction helps to simplify the model.

$$\min_{\beta} \sum_{k=1}^{m} (y_j - \hat{y}_j)^2 + \lambda \sum_{d=1}^{q} |\beta_j|$$

where:

- y_j represents the actual values, and $\hat{y}_j$ represents the predicted values,
- $\beta_j$are the regression coefficients,
- λ is the regularization parameter that controls the amount of shrinkage.

3.3 Evaluation Metrics:

Multiple error and accuracy metrics were used to assess the regression models' performance. These metrics measure each model's efficiency as well as the difference between expected and actual student grades.

Mean Squared Error (MSE):

The average squared difference between expected and actual values is measured by the Mean Squared Error, or MSE. Larger deviations are penalized more severely because errors are squared, which makes it susceptible to outliers.

Root Mean Squared Error (RMSE):

The square root of MSE, expressed in the same units as the target variable, is known as the root mean squared error, or RMSE. It is frequently used to compare model accuracy and offers a more comprehensible indicator of prediction error magnitude.

Mean Absolute Error (MAE):

The average absolute difference between expected and actual values is calculated by the Mean Absolute Error, or MAE. It provides a reliable indicator of model performance since it handles all errors equally and is less inclined to outliers than RMSE.

Coefficient of Determination (R2):

R2 is the percentage of the target variable's variance that the model can account for. Stronger predictive fit is indicated by values near 1, while poor explanatory power is indicated by values close to 0.

Cross- Validation:

K-fold cross-validation was used to ensure robustness. In particular, tests were carried out using various fold values (k = 3, 5, 7, 10, 15), and the mean RMSE and R2 values for each fold were reported. This method mitigates bias, lessens reliance on a single train-test split, and offers a more accurate assessment of model generalization.

Statistical Testing:

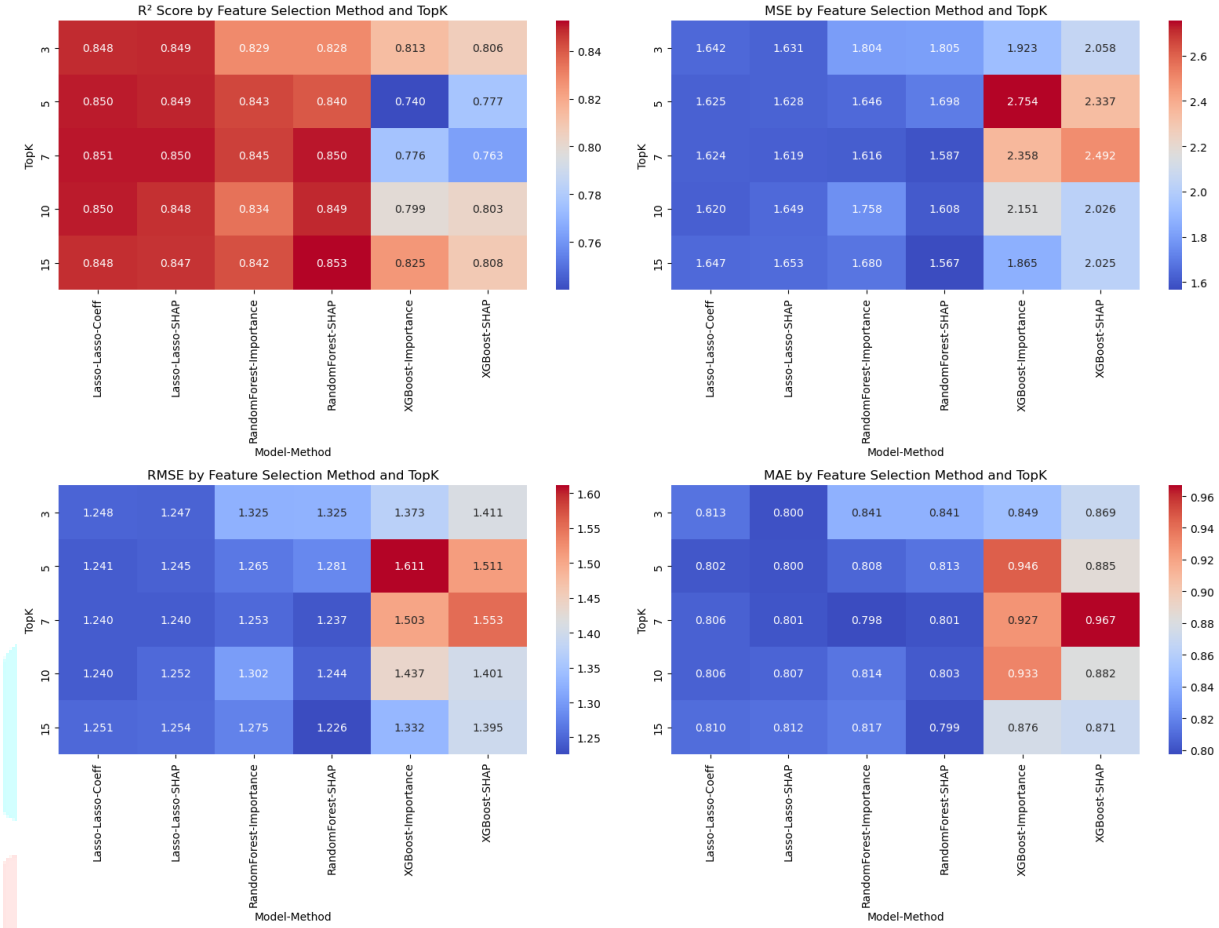
The t-statistic is a test which is primarily used to check whether the difference between two methods is statistically significant or it is just due to some random changes. In this research work this test is conducted between SHAP-based feature selection and traditional importance-based methods. If **|t| is large** (positive or negative), it means there is a strong difference between the two methods. If **|t| is small**, it means the difference is weak and likely due to chance. The **p-value** associated with t tells us whether the difference is significant ($p < 0.05$) or not.

3.5. Feature Selection Approaches:

Feature Selection is process of selecting relevant features and ignoring irrelevant features to make a solid prediction from the machine learning models. It helps to reduce the error and invalid predictions. In this research work feature selection is done in two different ways. First method is automatic model-based selection, here the machine learning algorithms *Decision Tree, Random Forest, XGBoost* are deployed. The impurity will be reduced and based on important scores is allotted for each feature by the model. In Lasso regression instead of scores coefficient after L1-regularization is used to rank the features. The second method is Explainability model-based selection (SHAP). Here the SHapley Additive exPlanations is applied to compute both global and local feature contributions. Based on the SHAP values the features are ranked and overlapping features between both the methods are analysed.

IV. RESULTS

Figure :1 Heat Map of Performance Metrics



The above fig:1 depicts the correlation of heatmap between numbers of Top-K features (3, 5, 7, 10, 15) and its performance metrics such as R^2, MSE, RMSE, and MAE for various machine learning algorithms such as Random Forest, XGBoost, and Lasso regression under different strategies of feature selection. From the heatmap, it is clearly seem that Lasso regression achieved highest accuracy across all feature subsets where R^2 values ranging between 0.847 and 0.862. Random Forest also yields competitive results (≈0.84), while XGBoost showed relatively lower performance (≈0.76 – 0.80). Notably, SHAP- and importance-based selections provided nearly identical R^2 values, indicating that SHAP did not compromise predictive power while offering interpretability.

The similar kind of observation is also seen in MSE and RMSE heatmaps. Higher errors are noted while comparing XGBoost with Random Forest and Lasso Regression. If features get added after the Top-7 threshold, the results began to worsen. This suggest that core feature subset of (G1, G2, failures, absences, study time) carried out more predictive information.

In terms of MAE, Lasso and Random Forest outperformed XGBoost, yielding average absolute errors around 0.80 compared to 0.93 – 0.96 for XGBoost. SHAP-based selection closely mirrored traditional methods, highlighting its robustness.

Overall, the results demonstrate that Lasso regression, whether with coefficient or SHAP-based feature selection, provides the best balance between accuracy and sparsity. Random Forest also performed competitively, while XGBoost lagged behind while comparing with these models. Mainly, the statistical validation (paired t-test) confirmed that the differences between SHAP and traditional importance were not statistically significant ($p > 0.05$), but SHAP offered the crucial advantage of explainability at both global and local levels.

Figure:2 plot between R^2 variation with different Top-K features

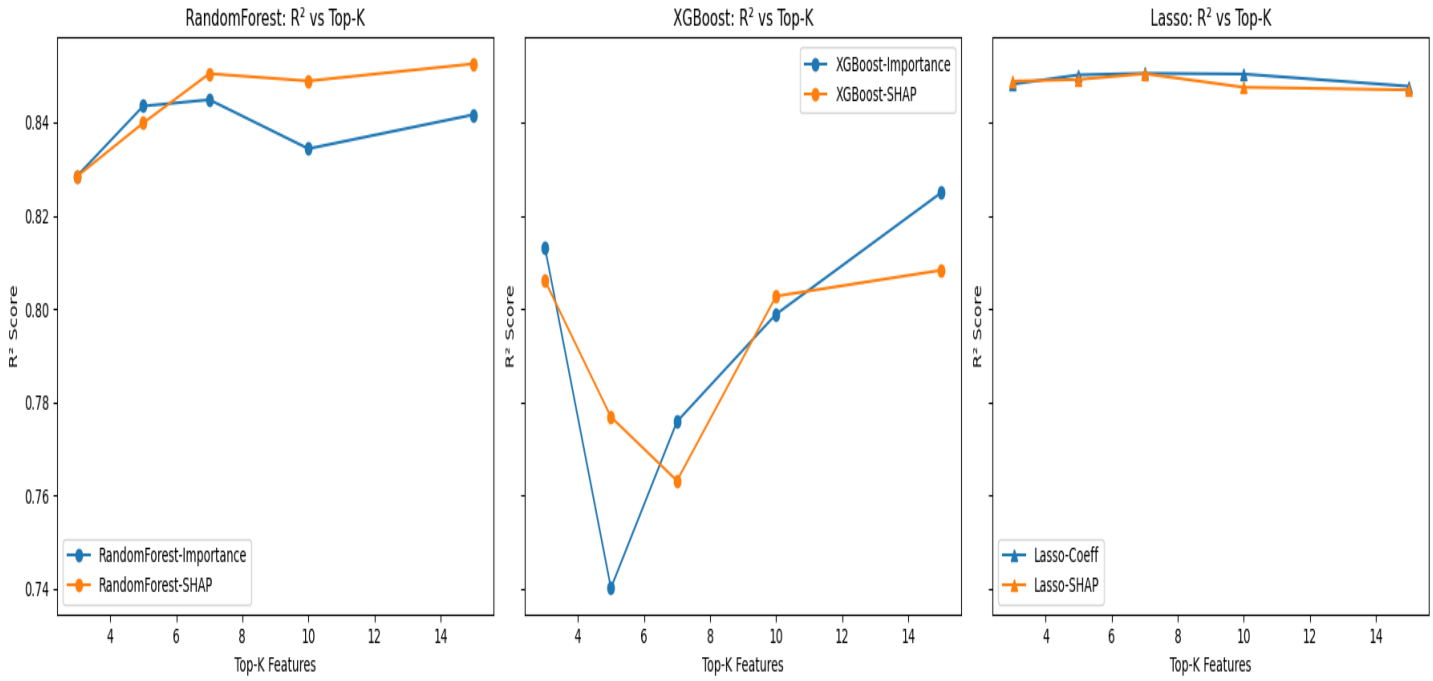


Figure 2 presents the R^2 variation with different Top-K features for different machine learning algorithms Random Forest, XGBoost, and Lasso. Stable performance has been achieved in Random Forest after Top-7 features, in means of SHAP-based selection it is showing slightly better stability than Importance based method. XGBoost exhibited major fluctuations in accuracy depending on the feature subset, indicating higher sensitivity, though SHAP and Importance results were comparable. Lasso consistently produced the highest and most stable R^2 (~0.85–0.86), with both coefficient and SHAP-based rankings leading to nearly identical outcomes. Overall, the results confirm that SHAP-based selection performs equally with traditional methods, it also holds upper hand because of offering the additional benefit of feature interpretability.

Figure:3 Overlap of Features between traditional based and SHAP

	Model	TopK	Method1	Method2	Top_Features_Method1	Top_Features_Method2	Overlap
0	RandomForest	3	Importance	SHAP	[G1, absences, G2]	[G1, absences, G2]	3
1	RandomForest	5	Importance	SHAP	[absences, G1, famrel, age, G2]	[absences, G1, age, G2, failures]	4
2	RandomForest	7	Importance	SHAP	[freetime, absences, G1, famrel, age, G2, health]	[freetime, absences, G1, famrel, age, G2, fail...	6
3	RandomForest	10	Importance	SHAP	[freetime, absences, G1, famrel, age, G2, reas...	[freetime, absences, G1, famrel, age, G2, fail...	9
4	RandomForest	15	Importance	SHAP	[Mjob, freetime, failures, absences, Medu, tra...	[Mjob, freetime, Fedu, absences, school, G1, f...	12
5	XGBoost	3	Importance	SHAP	[school, absences, G2]	[G1, absences, G2]	2
6	XGBoost	5	Importance	SHAP	[failures, absences, G2, schoolsup, school]	[absences, G1, age, G2, school]	3
7	XGBoost	7	Importance	SHAP	[failures, absences, traveltime, G1, G2, schoo...	[freetime, failures, absences, G1, age, G2, sc...	5
8	XGBoost	10	Importance	SHAP	[freetime, failures, absences, traveltime, G1,...	[freetime, failures, absences, G1, age, G2, st...	7
9	XGBoost	15	Importance	SHAP	[freetime, failures, Walc, absences, traveltim...	[freetime, failures, Walc, Fedu, absences, sex...	13
10	Lasso	3	Lasso-Coeff	Lasso-SHAP	[failures, school, G2]	[school, G1, G2]	2
11	Lasso	5	Lasso-Coeff	Lasso-SHAP	[traveltime, school, G1, G2, failures]	[traveltime, G1, G2, school, reason]	4
12	Lasso	7	Lasso-Coeff	Lasso-SHAP	[traveltime, school, G1, famsup, G2, failures,...	[failures, traveltime, G1, G2, school, reason,...	5
13	Lasso	10	Lasso-Coeff	Lasso-SHAP	[traveltime, school, sex, G1, famsup, G2, high...	[failures, traveltime, absences, G1, G2, Dalc,...	6
14	Lasso	15	Lasso-Coeff	Lasso-SHAP	[traveltime, school, sex, G1, famsup, G2, stud...	[freetime, failures, traveltime, absences, sex...	13

The above fig:3 shows the overlap of top features selected by Importance/Coefficient-based methods and SHAP across different models and Top-K values. In Random Forest model both the approaches importance based and SHAP based approaches consistently identified the same set of dominant predictors with over lapping score of 12 out of 15 in Top-15 fold. Whereas in XGBoost model overlapping was smaller when folds are minimal and increased gradually when subset is larger. For Lasso model the over lap was consistently high 13 out of 15 at Top-15fold which is more efficient than Random Forest. It shows strong relation between SHAP and coefficient-based selection. The result evident that SHAP yields high accuracy while reinforcing the interpretability of feature ranking.

Figure:4 Residual plots for Different Models

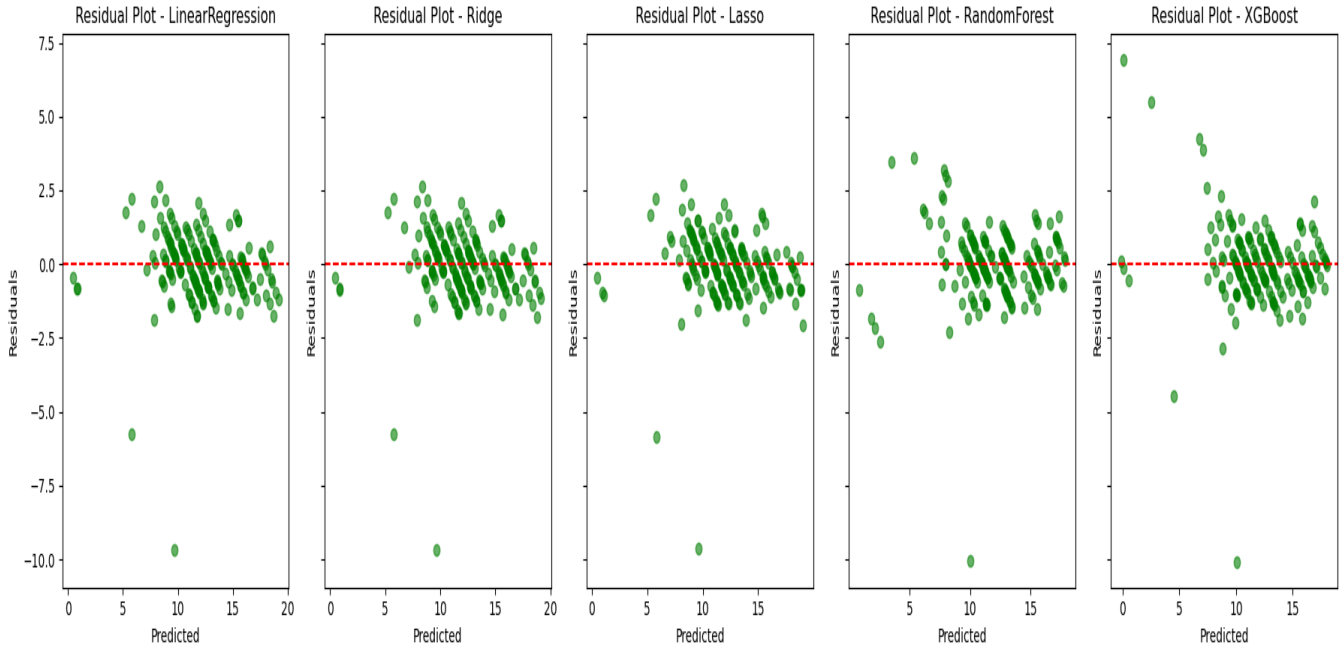
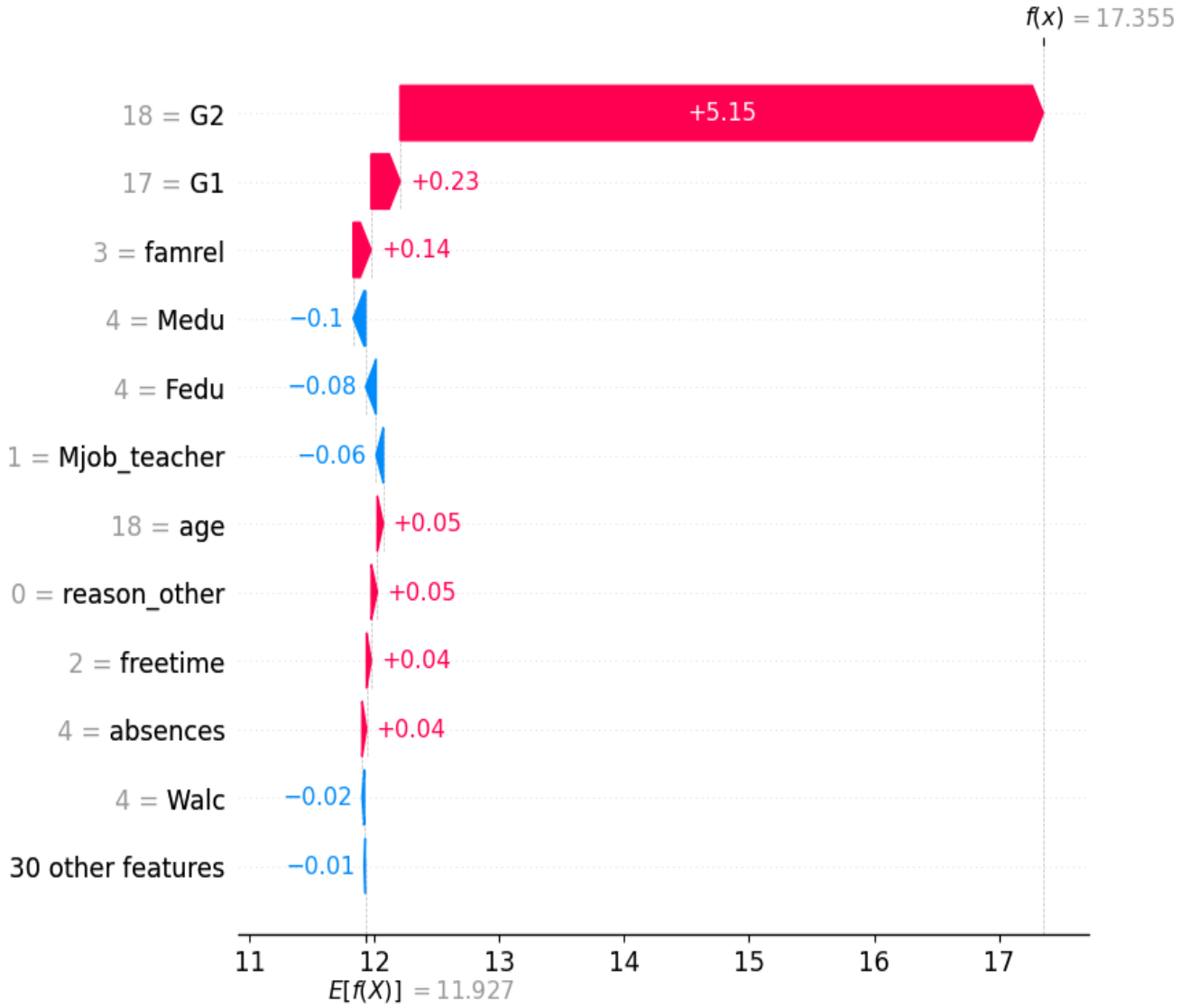


Figure 4 illustrates the residual plots for different regression models. Linear Regression, Ridge, Lasso, Random Forest and XGBoost. The linear and regularized models linear, ridge and Lasso have very minimal clustered residuals within ±2, demonstrating stable error distribution and effective generalization. In contrast, Decision Tree–based models show wider error ranges. Random Forest residuals reach ±5, while XGBoost

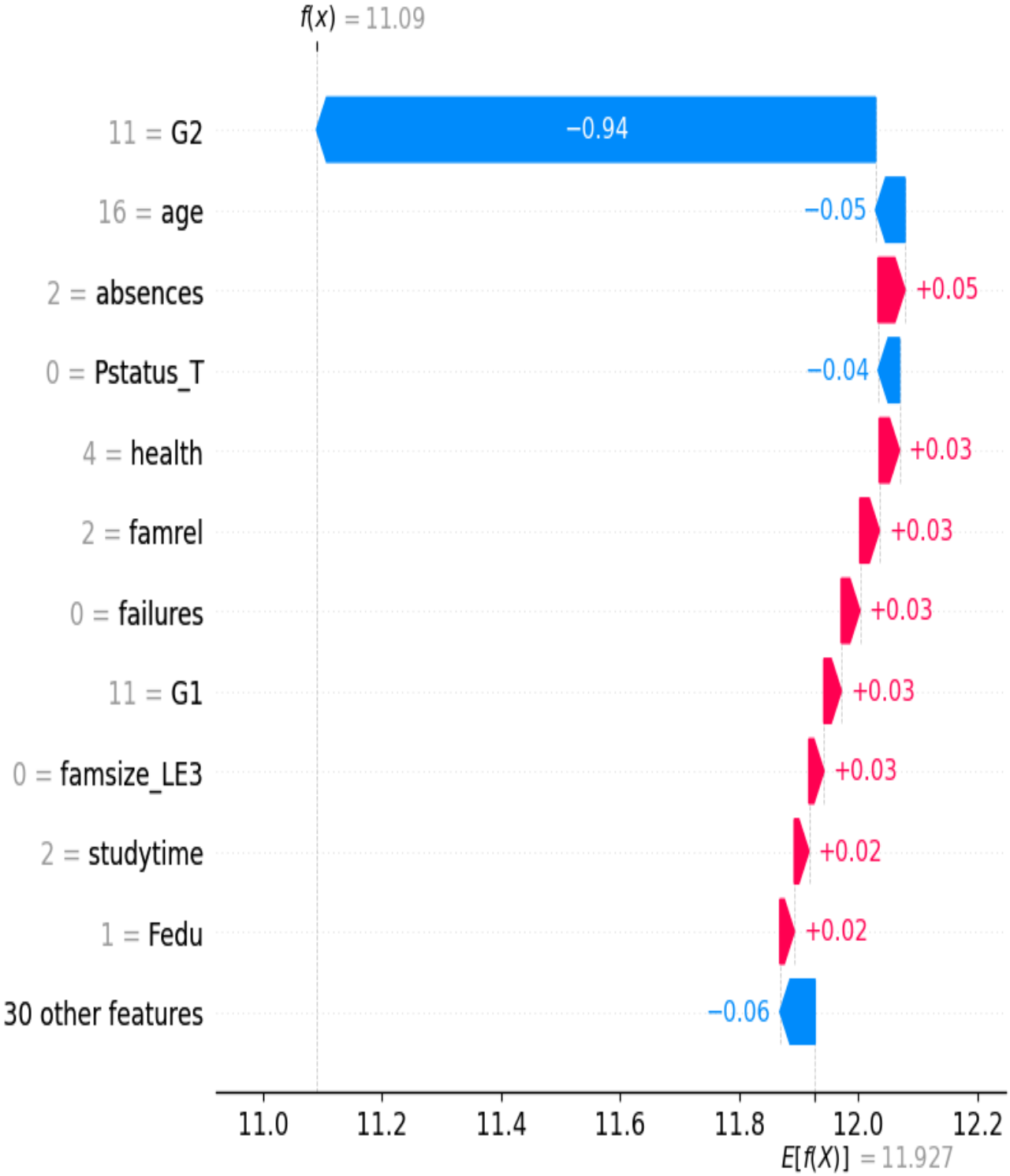
residuals exceed ±10 for some students, indicating higher variance and occasional overfitting. Overall, regularized linear models (Ridge, Lasso) provide the most consistent performance, though Random Forest balances accuracy with interpretability.

Figure:5 Shap analysis for Student 1



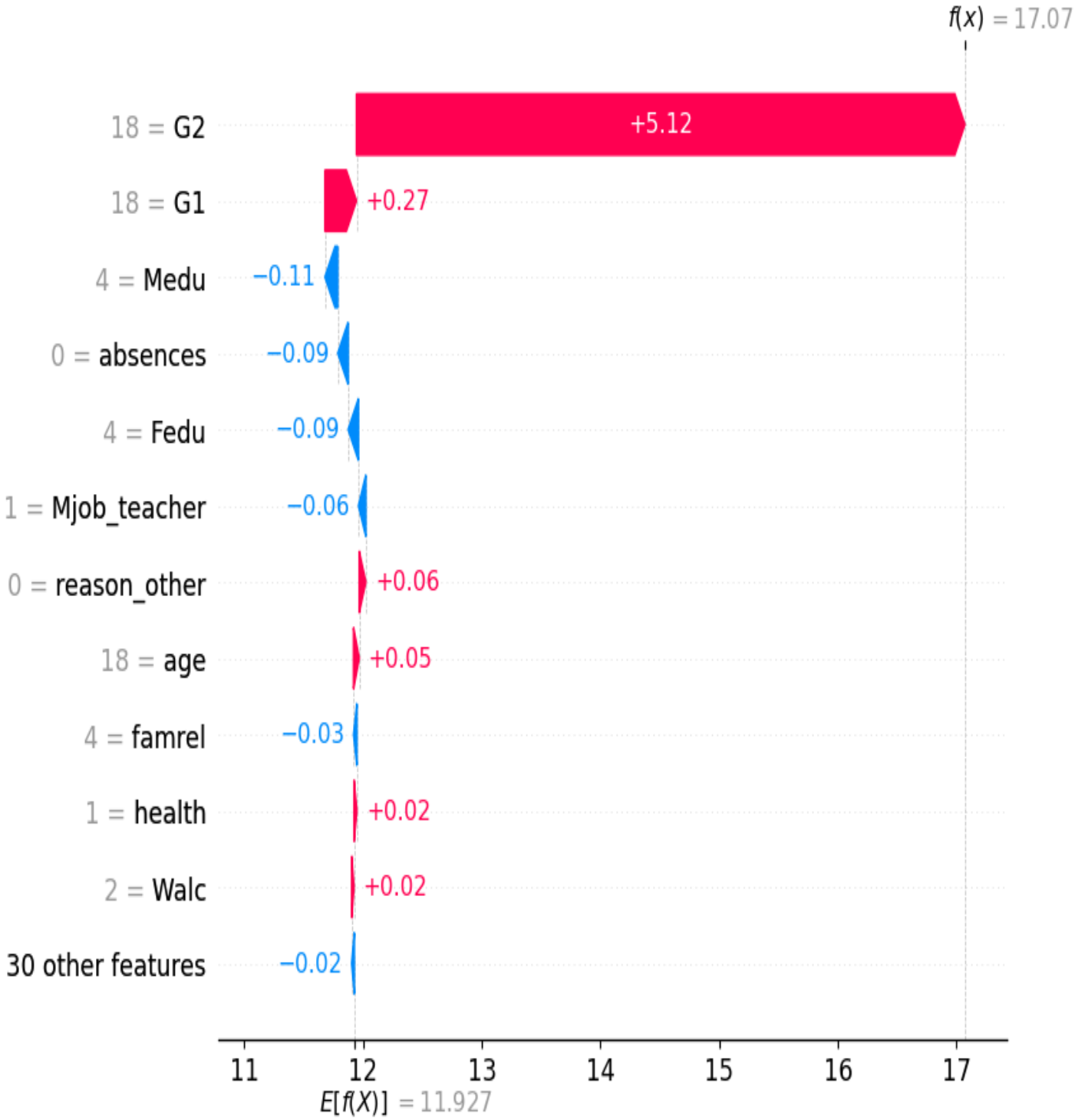
- The Grade **G2 (+5.1)** and Grade **G1 (+0.2 to +0.3)** has an strong push for the upward prediction.
- features named **Medu, Fedu, Mjob_teacher** slightly reduce the score. It shows Negative Influence for making predictions.
- **famrel, age, reason_other, absences, freetime**. Having a small level of positive influence.
- Incase of this record the Academic performance (G2, G1) dominates, with some behavioral and family background features adding marginal effects.

Figure:6 Shap analysis for Student 2



- The feature **G2 (–0.94)** drops the prediction towards down.
- Small positives from **absences (+0.05), studytime (+0.02)**, **health (+0.03), famrel (+0.03), failures (+0.03)** add minor benefits.
- In this case the feature G2 severely lowers prediction, even though behavioral and family factors provide small positive compensations.

Figure:7 Shap analysis for Student 3



- Similar to Student 1, **G2 (+5.12)** dominates.
- **G1 (+0.27)** also positive.
- Small negatives relation exist between features **Medu, Fedu, absences** which reduce score slightly.
- Other features (reason_other, age, famrel, health) add minor balance.
- In this case the features G2 and G1 have high influence and family background seems to be disadvantage.

Figures 5,6,7 present SHAP local explanations for three individual students. In all cases, **G2 (second period grade)** was the most predicting factor, it either contribute positively or negatively depending on the student's performance. For high-performing students, G2 and G1 strongly increased the predicted final grade, while negative influences such as parental education, absences, or family job background had only marginal effects. On other hand, for lower-performing students, low G2 scores significantly reduced predictions, and small positive contributions from behavioural or demographic features were insufficient to balance this impact. These case studies confirm that **academic history dominates predictions at the individual level**, with contextual features providing secondary refinements. SHAP thus provides actionable insights by showing which factors institutions can target to support at-risk students.

V. CONCLUSION

The research work integrates traditional feature selection based on its importance or coefficient and SHAP Ranking based feature selection with the different machine learning algorithms such as **Decision Trees, Random Forests, XGBoost, and Lasso regression** in the student performance prediction dataset. The results revealed that SHAP with traditional feature selection holds upper hand than other methods. SHAP combined with Decision Tree feature selection yielded the most consistent and insightful feature subset among all other methods. Additionally, SHAP analysis added the explanation for both global level prediction (important features for all the students) and local level predictions (why specific student was predicted in certain way). Results are tested with Cross-fold validation and statistical testing to make valid and reliable output. The personalised insights for 3 students from test data is also derived how attributes like study time, failure and absences influence the predictions. Finally, the SHAP enhances both the model accuracy and interpretability for predicting the performance of the students.

REFERENCES

[1] K. Yan, "Student Performance Prediction Using XGBoost Method from A Macro Perspective," *Proc. CDS*, pp. 1–6, 2021, doi: 10.1109/CDS52072.2021.00084.

[2] P. Cortez and A. Silva, "Using data mining to predict secondary school student performance," *Proc. Future Business Technology Conference*, 2008.

[3] A. Harif and M. Kassimi, "Predictive Modeling of Student Performance Using RFECV-RF for Feature Selection and Machine Learning Techniques," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 15, no. 7, pp. 168–176, 2024, doi: 10.14569/IJACSA.2024.0150723.

[4] S. Malik, S. G. K. Patro, C. Mahanty, R. Hegde, et al., "Advancing educational data mining for enhanced student performance prediction: a fusion of feature selection algorithms and classification techniques with dynamic feature ensemble evolution," *Scientific Reports*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-92324-x.

[5] M. Imran, R. S. Bajaj, and M. A. Khan, "Student Academic Performance Prediction Using J48, NNge, and MLP with Preprocessing," *International Journal of Computer Applications*, vol. 168, no. 6, pp. 1–7, 2017.

[6] P. Guleria, "Blending Shapley values for feature ranking in machine learning: an analysis on educational data," *Neural Computing and Applications*, 2024. [Online]. Available: https://www.researchgate.net/publication/380346949

[7] H. Wang, et al., "Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods," *Journal of Big Data*, vol. 11, no. 5, pp. 1–15, 2024, doi: 10.1186/s40537-024-00905-w.

[8] C. Sebastián, J. M. Moguerza, and A. Muñoz, "A feature selection method based on Shapley values," *Soft Computing*, vol. 28, pp. 10563–10577, 2024, doi: 10.1007/s00521-024-09745-4.

[9] E. Kraev, B. Koseoglu, L. Traverso, and M. Topiwalla, "Shap-Select: Lightweight Feature Selection Using SHAP Values and Regression," *arXiv preprint*, arXiv:2410.06815, 2024.

[10] J. T. Hancock, T. M. Khoshgoftaar, and A. B. Dittman, "A problem-agnostic approach to feature selection and data reduction using SHAP," *Journal of Big Data*, vol. 12, no. 1, pp. 1–19, 2025, doi: 10.1186/s40537-024-01041-1.