MODEL INTERPRETABILITY REPORT

By Blessing Ilesanmi

3Signet Data Science Internship Week 5

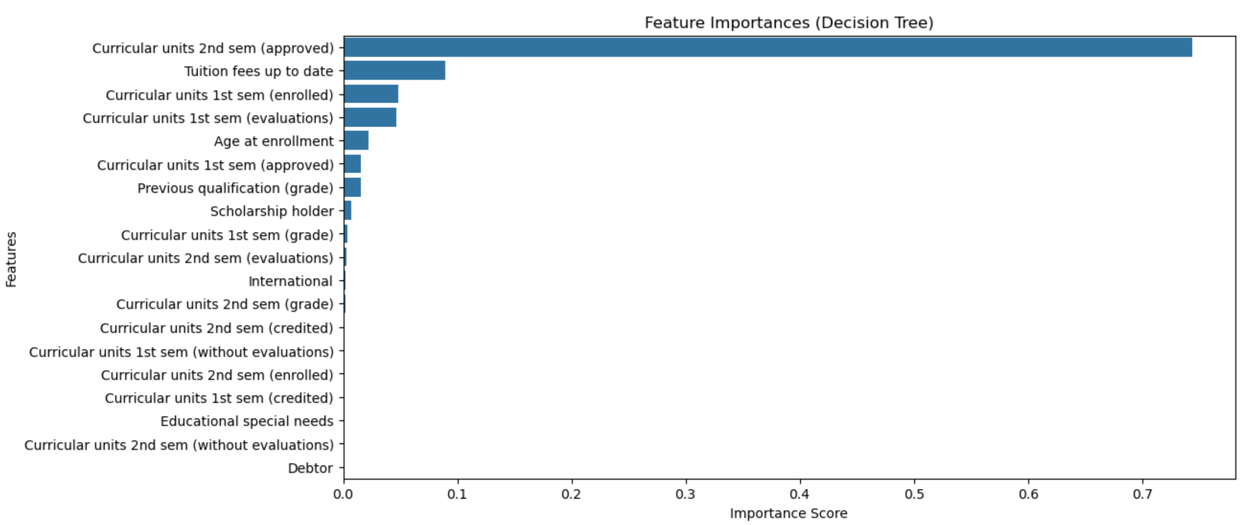
11th October, 2024

INTRODUCTION

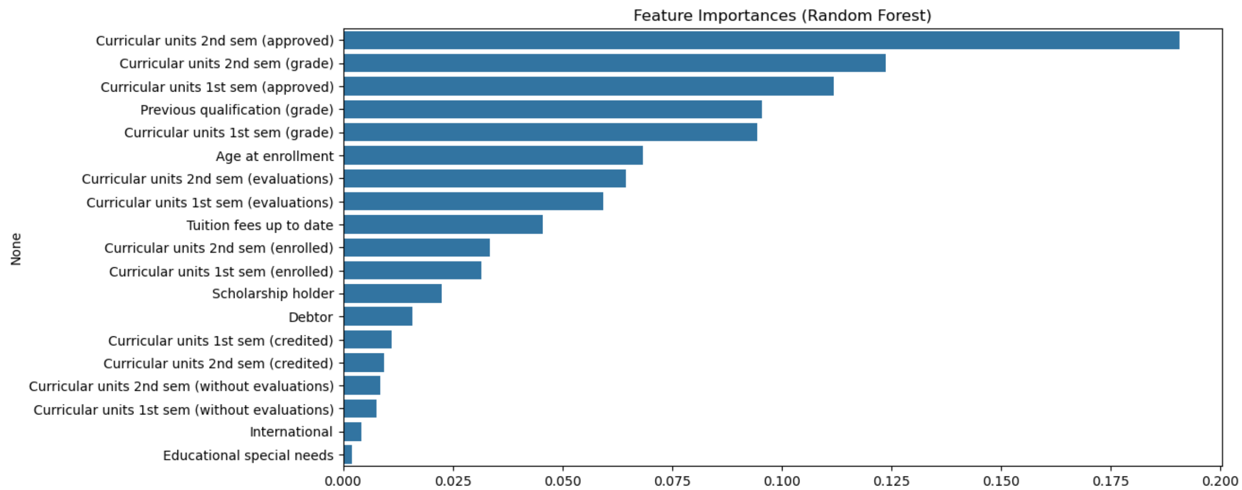
Understanding the key features that influence model predictions is critical in any machine learning analysis. This report focuses on analyzing feature importance and the relationships between features using various machine learning models, including Decision Trees, Random Forests, XGBoost, and Support Vector Machines (SVM). Through feature importance plots, SHAP (SHapley Additive exPlanations) summary and dependency plots, and partial dependence plots, we aim to identify the most significant predictors of the target variable and explore their interactions. By evaluating these metrics, we can gain insights into the underlying dynamics of the data, enabling more informed decision-making in future applications.

FEATURE IMPORTANCE PLOTS

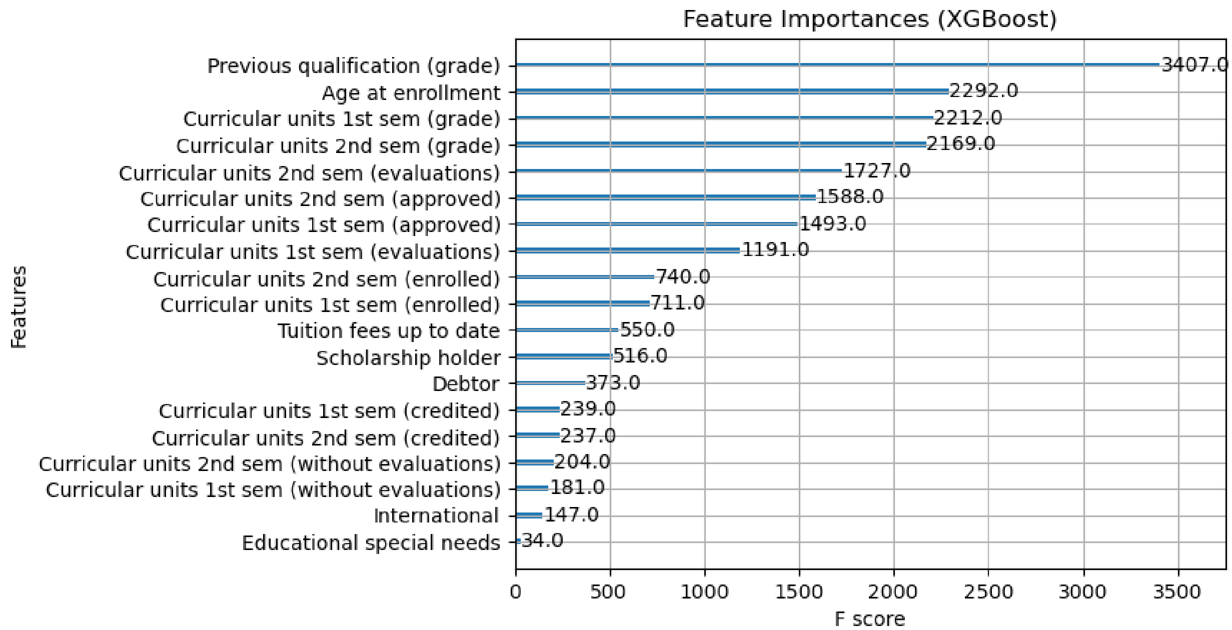
Decision Tree: The top-ranked features, such as "Curricular units 2nd sem (approved)" and "Tuition fees up to date," have the most significant influence on predicting the target variable. Less important features: Features like "Educational special needs" and "Debtor" show relatively smaller or no impacts on the model's predictions due to their lower rankings. Feature relationships: The proximity of feature rankings may indicate potential correlations or similar effects on the target variable, suggesting that closely ranked features might be related in how they influence outcomes.



Random Forest: The top-ranked features, such as "Curricular units 2nd sem (approved)" and "Curricular units 1st sem (grade)," are likely the most influential in predicting the target variable. Less important features: Features like "Educational special needs" and "International" have a smaller impact on the model's predictions, as indicated by their shorter bars. Feature relationships: The closeness of feature rankings suggests potential correlations or similar effects on the target variable, as features that are ranked similarly may have comparable influences on the outcome.



XGBoost: "Previous qualification (grade)" and "Age at enrollment" stand out as the most significant factors influencing the target variable. Less significant features: Features such as "Educational special needs" and "International" have a relatively minor impact on the model’s predictions, as shown by their lower rankings. Feature connections: Features ranked closely together may share similar effects on the target variable, hinting at possible correlations between them.



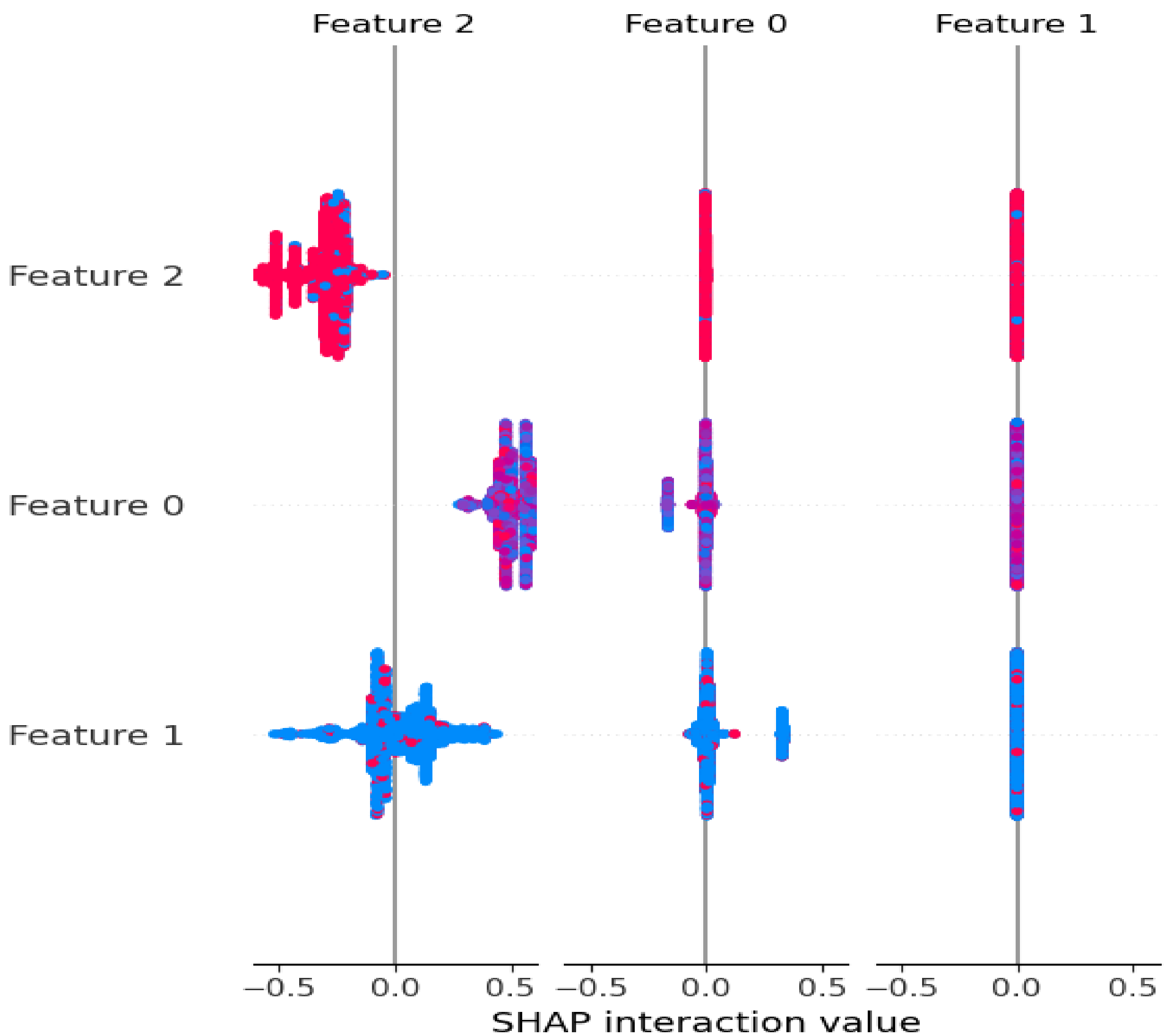
SHAP SUMMARY AND DEPENDENCY PLOTS

Decision Tree

Strong interactions: There are notable strong interactions between Feature 0 and Feature 1, as well as between Feature 2 and Feature 1, indicated by the clusters of red and blue dots along these axes.

Weak interactions: The interaction between Feature 0 and Feature 2 appears to be weaker, characterized by fewer and more scattered dots in that area.

Feature importance: The density of dots along each feature's axis suggests their overall significance in the model. Here, Feature 1 stands out as particularly important, as evidenced by the high density of dots along its axis.

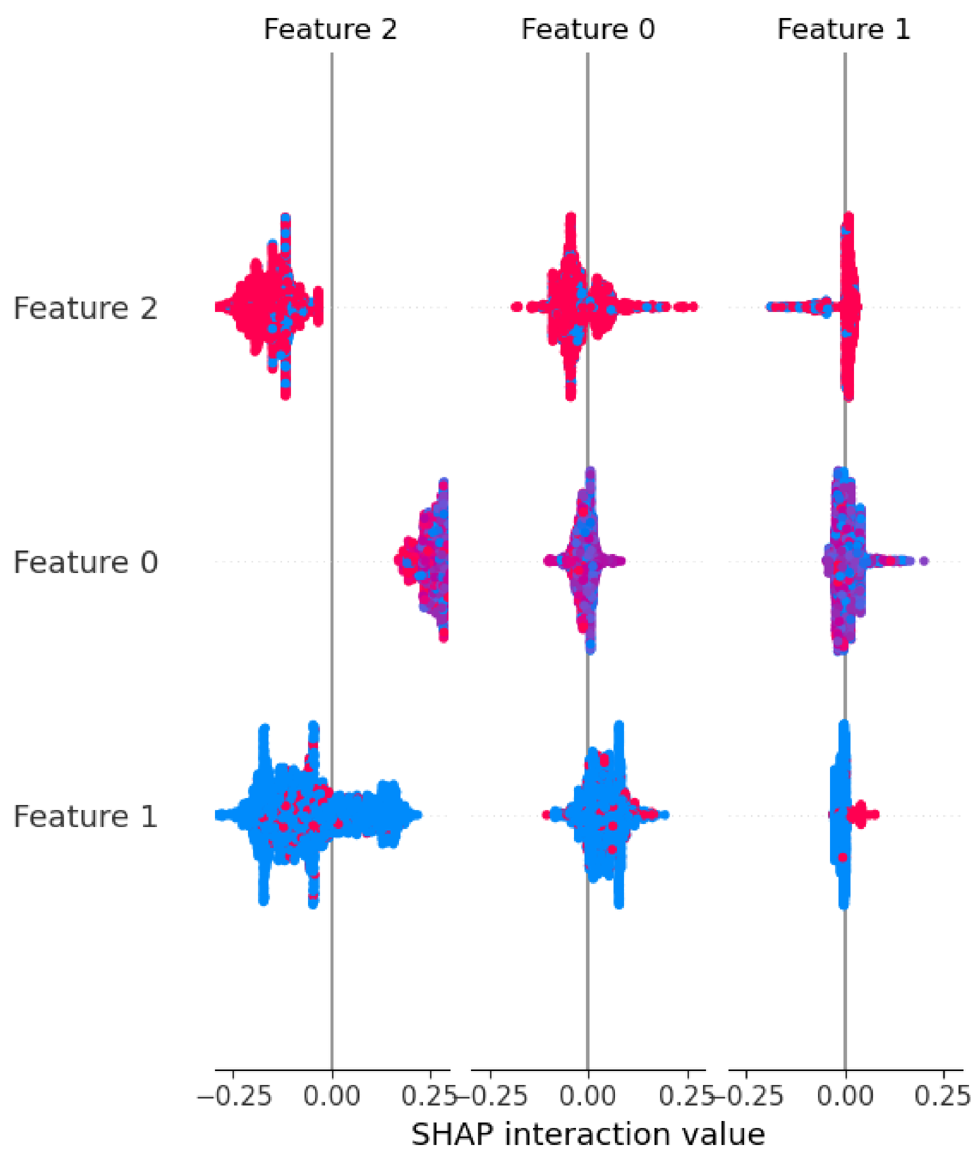


Random Forest

Strong interactions: There appear to be strong interactions between Feature 0 and Feature 1, as well as between Feature 2 and Feature 1. This is evident from the clusters of red and blue dots along these axes.

Weak interactions: The interaction between Feature 0 and Feature 2 seems to be weaker, as there are fewer and more scattered dots in this region.

Feature importance: The density of dots along a particular feature's axis can provide an indication of its overall importance in the model. In this case, Feature 1 seems to be particularly important, as it has a high density of dots along its axis.

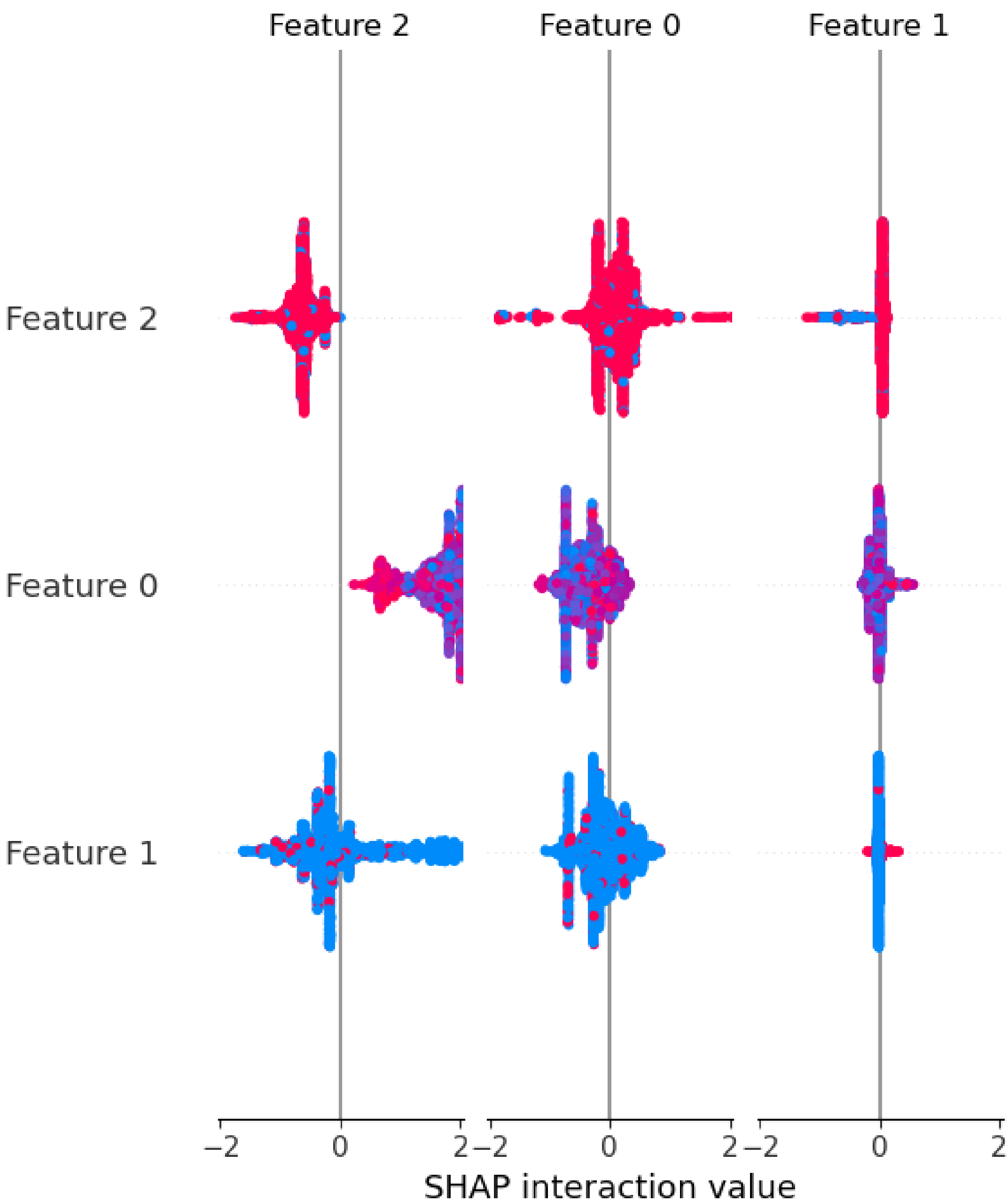


XGBoost

Strong interactions: Clusters of red or blue dots in specific regions of the plot indicate strong interactions between the corresponding features.

Weak interactions: When the dots are scattered randomly across the plot, it suggests weak or no significant interaction between those features.

Feature importance: A higher concentration of dots along a feature's axis suggests its greater influence or importance in the model.

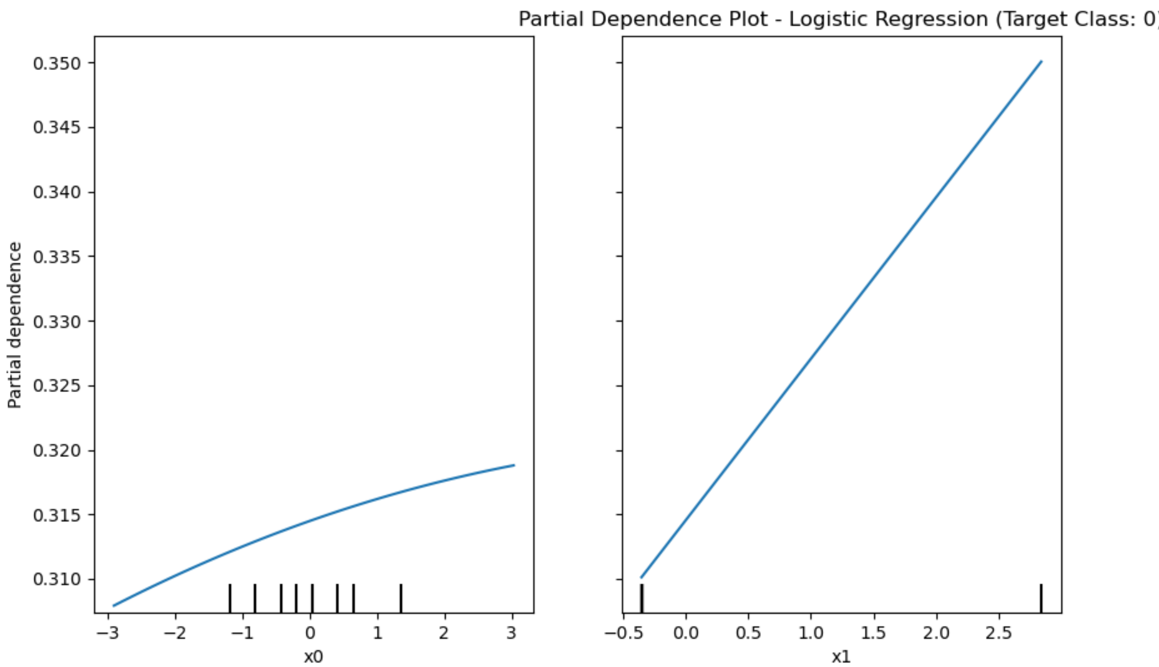


PARTIAL DEPENDENCE PLOTS FOR KEY FEATURES

Logistic Regression:

Feature x0: The left plot illustrates the partial dependence of the predicted probability on feature x0. The curve reveals that as x0 increases, the predicted probability of belonging to class 0 rises but at a decreasing rate, indicating a diminishing effect of x0 on the predicted probability.

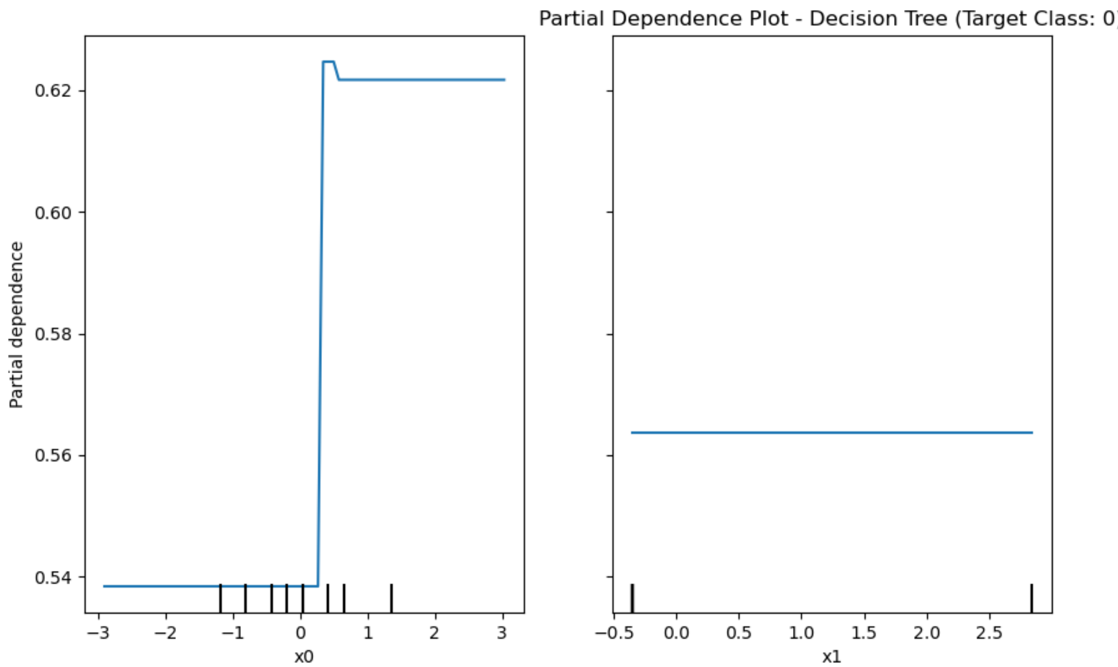
Feature x1: The right plot displays the partial dependence of the predicted probability on feature x1. The curve shows a linear relationship, where an increase in x1 corresponds to a proportional increase in the predicted probability of belonging to class 0.



Decision Tree:

Feature x0: The left plot illustrates the partial dependence of the predicted probability on feature x0, characterized by a step function. This indicates that the predicted probability changes sharply at specific values of x0, suggesting that the decision tree model has established decision boundaries based on particular thresholds of this feature.

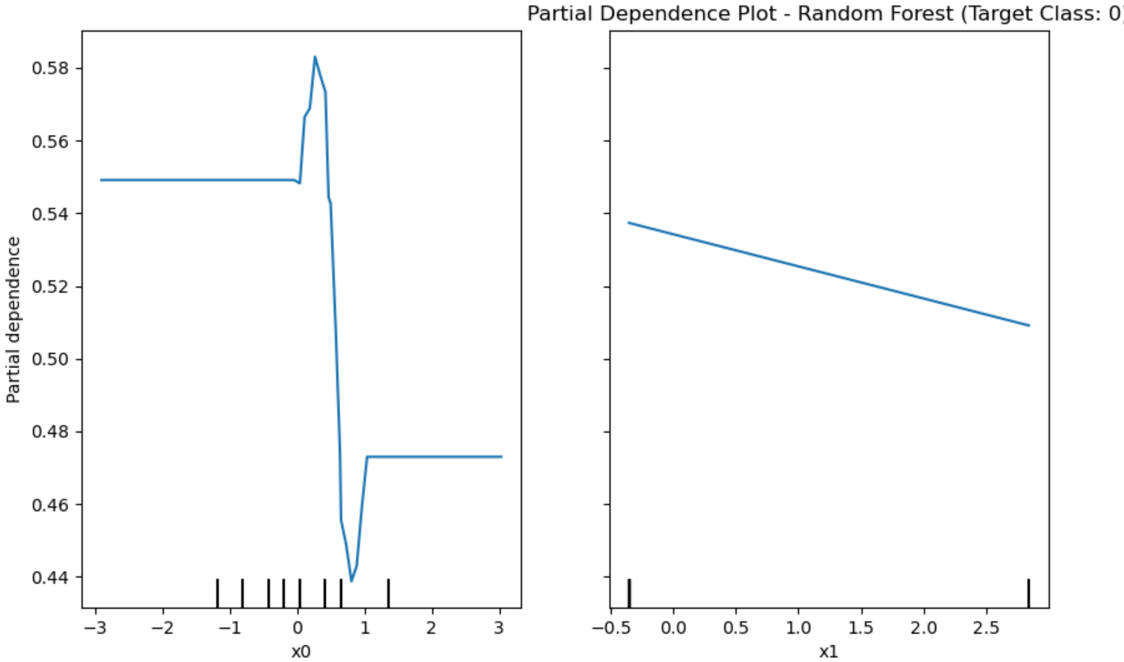
Feature x1: The right plot presents the partial dependence of the predicted probability on feature x1, represented as a flat line. This indicates that variations in x1 do not influence the predicted probability, suggesting that x1 is not a significant predictor of the target variable within the decision tree model.



Random Forests:

Feature x0: The left plot illustrates the partial dependence of the predicted probability on feature x0, depicted as a step function with multiple jumps. This indicates that the predicted probability changes sharply at specific thresholds of x0, suggesting that the random forest model has established several decision boundaries based on varying thresholds of this feature.

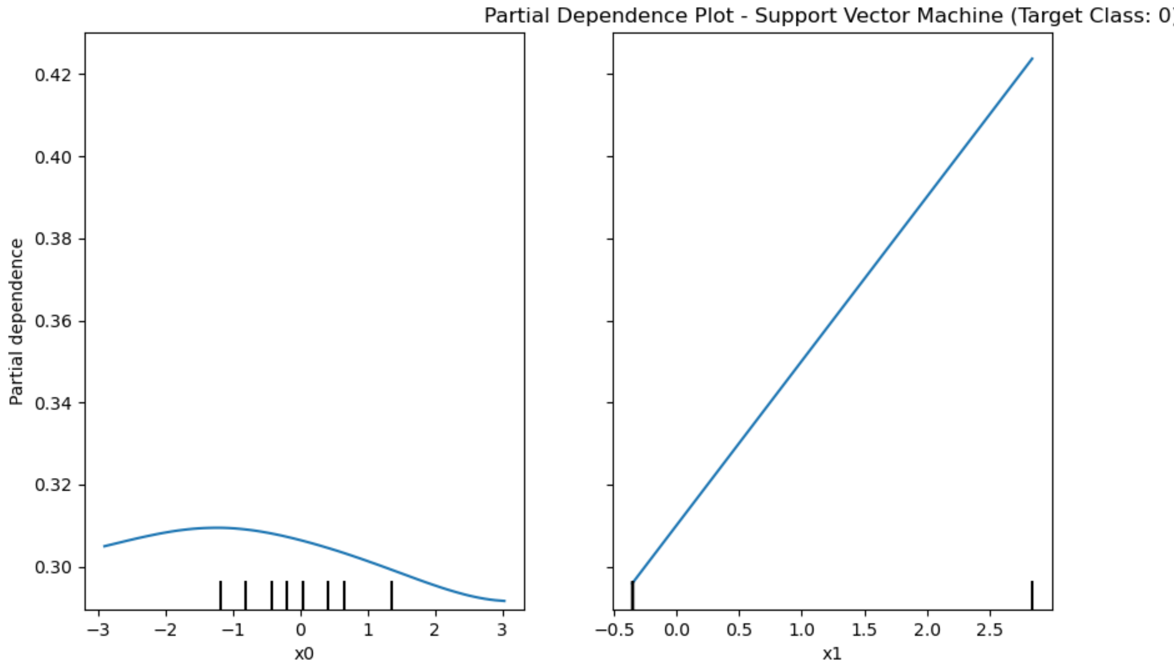
Feature x1: The right plot presents the partial dependence of the predicted probability on feature x1, shown as a slightly decreasing line. This indicates that as x1 increases, the predicted probability of belonging to class 0 decreases slightly, although the impact of x1 on the predicted probability is relatively minor.



Support Vector Machines:

Feature x0: The left plot displays the partial dependence of the predicted probability on feature x0, characterized by a non-linear function with a peak around x0 = 0. This indicates that the impact of x0 on the predicted probability is most significant near this value, decreasing as x0 deviates from it.

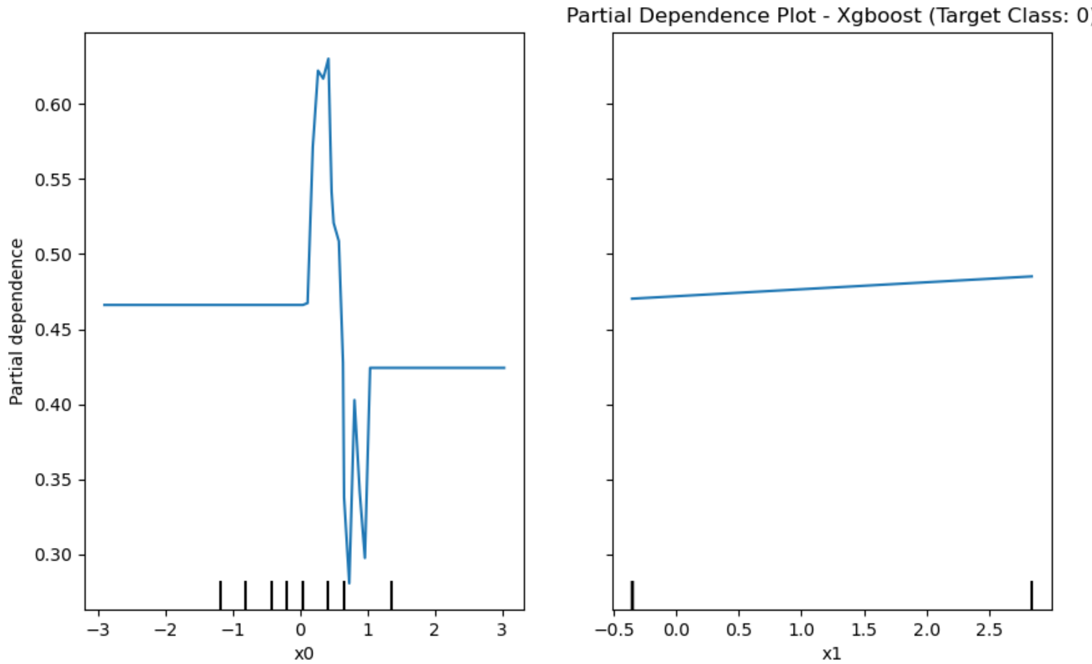
Feature x1: The right plot illustrates the partial dependence of the predicted probability on feature x1, represented as a linear function with a positive slope. This suggests that as x1 increases, the predicted probability of belonging to class 0 rises linearly.



XGBoost:

Feature x0: The left plot illustrates the partial dependence of the predicted probability on feature x0, exhibiting a complex function with multiple peaks and valleys. This indicates that the influence of x0 on the predicted probability is highly non-linear and varies across specific ranges of x0.

Feature x1: The right plot presents the partial dependence of the predicted probability on feature x1, showing a slightly increasing line. This suggests that as x1 increases, the predicted probability of belonging to class 0 rises marginally; however, its overall effect on the predicted probability is relatively minor.



CONCLUSION

In conclusion, the analysis of feature importance across multiple machine learning models reveals valuable insights into the predictive dynamics of the dataset. The Decision Tree, Random Forest, and XGBoost models highlight specific features such as "Curricular units 2nd sem (approved)" and "Previous qualification (grade)" as significant predictors, while less impactful features, including "Educational special needs," show minimal influence. The SHAP analysis further underscores the interactions between features, providing a nuanced understanding of how they collectively affect predictions. Finally, the partial dependence plots illustrate the varying relationships between features and the predicted probabilities, emphasizing the importance of considering both linear and non-linear effects. Overall, these findings enhance our understanding of the model's decision-making process and pave the way for optimizing predictive performance in future analyses.