School of Computational Sciences and Artificial Intelligence
DSAI 305: Interpretability and Explainability in AI
XGBoost results

Mid-progress Report:

XGBoost

By

Ahmed Mohamed 202201763

Model

The model employed (XGBoost) was fitted with the following parameters, that was collected from the paper:

```
model = xgb.XGBClassifier(
    n_estimators=5000,
    learning_rate=0.01,
    subsample=0.8,
    max_depth=15,
    colsample_bytree=0.8,
    scale_pos_weight=596,
    eval_metric=['logloss', 'auc', 'aucpr'],
)
```

These parameters ensured that the model did not overfit or underfit while balancing the weights of both classes.

Results

The model showcased incredible results, especially for predicting class (0). However, it had some problems correctly predicting class (1), which makes sense because class (1) observations in the data are greatly less than class (0). However, it still performed well in predicting it overall. These observations can be concluded from the following model output statistics:

```
Accuracy: 0.9995
Precision: 0.9995
Recall: 0.9995
F1 Score: 0.9995

Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     56651
           1       0.90      0.79      0.84        95

    accuracy                           1.00     56746
   macro avg       0.95      0.89      0.92     56746
weighted avg       1.00      1.00      1.00     56746


Confusion Matrix:
[[56643     8]
 [   20    75]]
```

Preprocessing, EDA, and Feature Selection

Starting with preprocessing, the methods used are simple and most common ones like checking for and dropping duplicates and null values. However, when checking for outliers, we found that the every fraud case (y==1) is an outlier. This makes sense since all fraud activities are some sort of anomaly or outlier. That is why we did not substitute or remove outliers because that would make the data unrealistic and non-representative. Finally, oversampling was used in hopes of achieving better results for the extremely undersampled class (1). However, it did not help in achieving better results at all, instead, it only increased the training time while maintaining the same results.

Going forward with the EDA, we applied univariate and multivariate analysis to understand the features and their interactions with each other. And finally, to understand the global importance of each feature to the target variable.

The overall results concluded from the univariate and multivariate analyses are that the features some features are heavily correlated with each other while most other features are not correlated with anything. Additionally, there were a few duplicates and null values.

For example, we employed Spearman correlation (Fig. 1) and ANOVA (Fig. 2) to calculate the importance of each feature to the target.

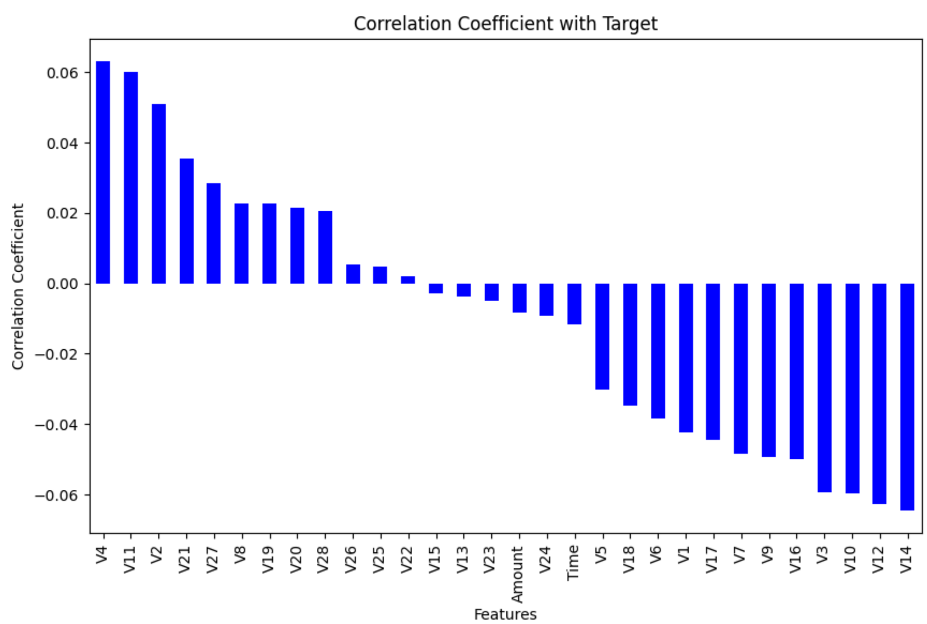

Figure 1: Spearman correlation analysis showing the relationship between each feature and the target variable. Higher absolute correlation values indicate greater importance to the target.

XGBoost results

ANOVA:

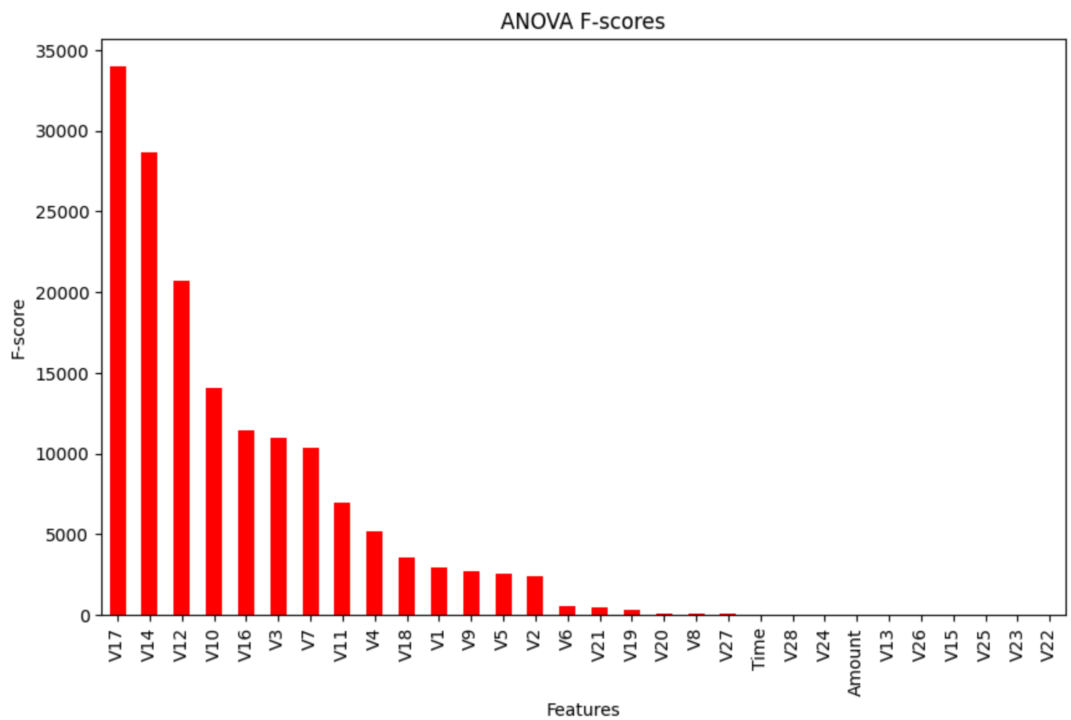

Figure 2: ANOVA results illustrating the importance of each feature to the target variable. Features with higher F-values demonstrate greater significance in explaining the target variance.

The most important 2 features across both methods are V12 and V14. Other features who appear to be important but not consistent in both methods are V4, V7, V10, V11, V16, V17, V21, and V27.
Their importance will be confirmed when applying post-hoc techniques in the next section.

XAI techniques and results

To choose the suitable features to train the model with, recursive feature elimination (RFE) was employed. The features selected from this wrapper method are ['Time', 'V4', 'V6', 'V8', 'V12', 'V13', 'V14', 'V17', 'V18', 'V19', 'V20', 'V26', 'V27', 'Amount'].

Most of the features selected make sense because most of them are already known to be important from the EDA tests in the section above. The order of the features do not reflect their importance, they are only ordered in alphabetical order.

To confirm the contribution and importance of each feature, further post-hoc techniques were employed.

School of Computational Sciences and Artificial Intelligence
DSAI 305: Interpretability and Explainability in AI
XGBoost results

[1] Visualizing the XGBoost tree

Multiple trees were visualized to observe how the model used the features to make decisions. An example of such trees is shown below in Fig. 3.

Figure 3: Visualization of Tree #75, illustrating the decision-making process of the model. Each node represents a feature-based split, with branches showing the criteria and corresponding leaf values contributing to the prediction.

In this tree, for example, when observing all the decision nodes. We find that the tree used all the features selected from the RFE in at least 1 node to reach a decision at the end.

[2] Partial Dependence Plot (PDP)

When observing the PDP of each feature, it can be concluded that changing the value of most features does significantly change the value of the prediction, which further confirms the importance of each feature to the prediction. An example of such PDP plots is shown in Fig. 4.

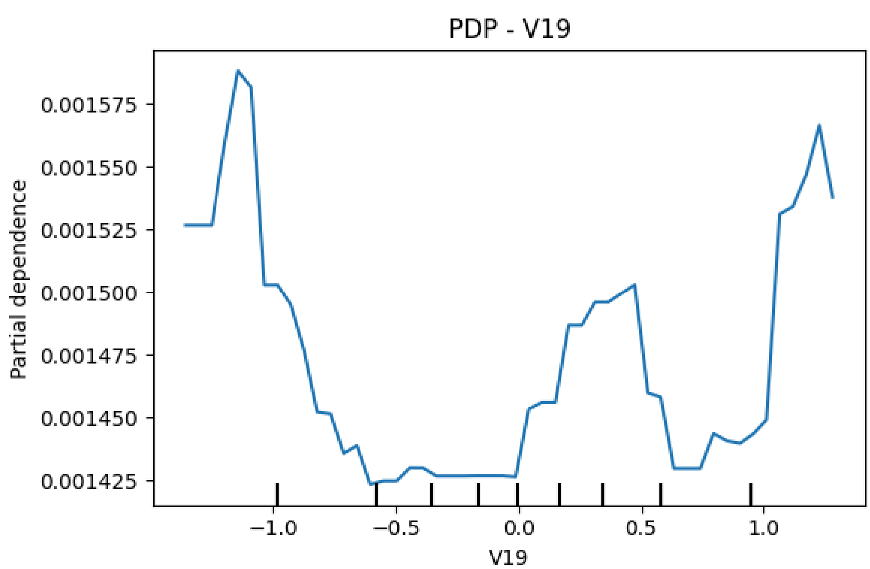

Figure 4: Partial Dependence Plot (PDP) for feature V19, demonstrating its impact on the model's predictions. The y-axis represents the partial dependence, while the x-axis shows the range of V19 values.

XGBoost results

[3] Accumulated local effects (ALE)

Due to the limitations and pitfalls of PDP that result from correlated features, extrapolation and marginalization, it is not recommended to depend entirely on interpretations gathered from PDP. That is why ALE was applied as well. By calculating the ALE for each feature again, we further confirm that when the feature value changes the prediction changes like in Fig. 5.

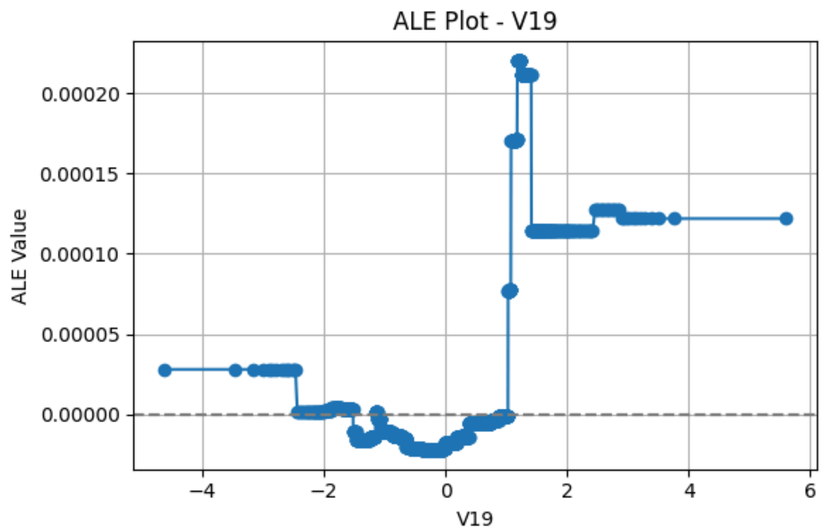

Figure 5: ALE Plot for feature V19, showing its effect on the model's predictions while accounting for feature interactions. The y-axis represents the ALE value, and the x-axis shows the range of V19 values.

However, some ALE plots of other features showed barely any change, like the ALE in Fig. 6.

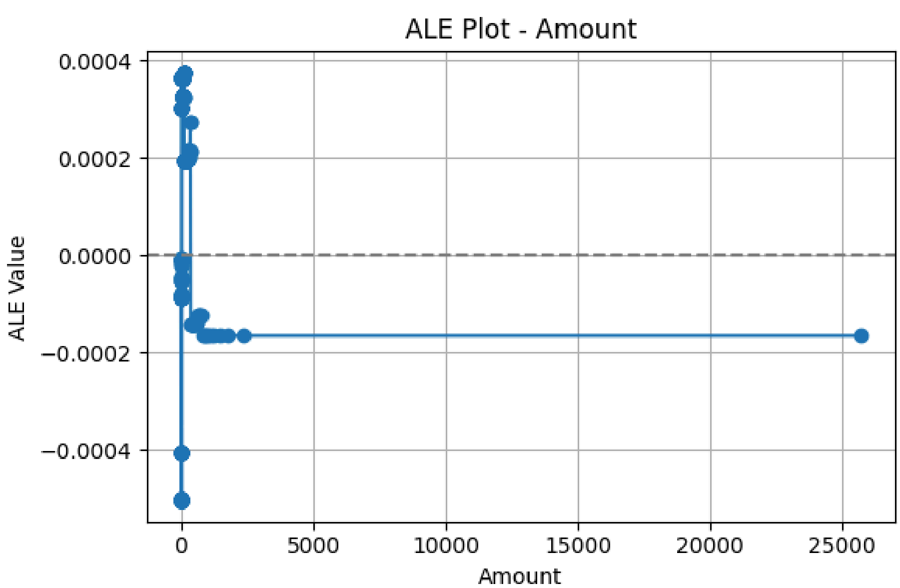

Figure 6: Accumulated Local Effects (ALE) Plot for the 'Amount' feature, illustrating its minimal impact on the model's predictions. The y-axis represents the ALE value, and the x-axis shows the range of Amount values.

This indicates that the wrapper feature selection technique (RFE) might have chosen less significant features like ‘Amount’. However, finding out that ‘Amount’ is not as significant as other features is no surprise because it did not appear as the top affecting feature in any EDA techniques employed above, like ANOVA, so it all makes sense.

Conclusion

By combining the interpretations from all the above-mentioned XAI techniques, a conclusion about why the model got high accuracy (0.9995) can be reached.

The combination of the XAI techniques enabled us to choose the most important and significant features, which were able to lessen the noise in the data and let the model focus on only the important features. Additionally, XGBoost is a very robust model that can effectively train and provide satisfactory results even with all the noise in the data.

Finally, by combining all the techniques’ interpretations. The most significant features for predicting are:

V4, V6, V8, V,10, V11, V12, V13, V,14, V17,V18, V19, V20, V26, V27