<div align="center">**Machine Learning Model Performance Report**</div>

Model Overview

The following report summarises the performance of a classification model trained on a dataset to predict target data. Model was trained using the XGBoost ML algorithm.

Evaluation of the model using the suggested metrics

Accuracy :

Accuracy measures the proportion of correctly classified instances out of all instances.

Value: 0.98

Interpretation: The model accurately predicted the target variable for **98%** of the instances in the test dataset.

Precision :

Precision measures the proportion of true positive predictions out of all positive predictions.

Value:

Classifications	- Precision score	- Test dataset support
1	- 0.81	- 24678
0	- 1.00	- 237034

Interpretation: For every instance predicted as positive, the model correctly predicted **81%** of them and **100%** for **negative**.

Recall (Sensitivity or True Positive Rate) :

Recall measures the proportion of true positive predictions out of all actual positive instances.

Value:

Classifications	- Precision score	- Test dataset support
1	- 0.97	- 24678
0	- 0.98	- 237034

Interpretation: The model correctly predicted **97%** of the actual **positive** instances and **98%** of the **negative** instances in the test dataset.

F1 Score :

F1 score is the harmonic mean of precision and recall.

Value:

Classifications	- Precision score	- Test dataset support
1	- 0.88	- 24678
0	- 0.99	- 237034

Interpretation: The model achieved a balance between precision and recall, indicating a good trade-off between the two.

AUC-ROC (Area Under the Receiver Operating Characteristic Curve) **:**

AUC-ROC measures the model's ability to distinguish between positive and negative classes.

Value: 0.97

Interpretation: The model demonstrated excellent ability to distinguish between the positive and negative classes, with an AUC-ROC score of 97%.

Confusion Matrix :

Predicted Positive	Predicted Negative
231267 (TN)	**5767** (FP)
689 (FN)	**23989** (TP)

Interpretation:

* True Positives (TP): **23989** instances were correctly predicted as positive.

* True Negatives (TN): **231267** instances were correctly predicted as negative.

* False Positives (FP): **5767** instances were incorrectly predicted as positive.

* False Negatives (FN): **689** instances were incorrectly predicted as negative.

Insights and Analysis

The model is very accurate, with an overall accuracy of 98%. This means it correctly classifies most of the data. Good at identifying true positive cases (class '1'), with a recall of 97%. This means it correctly identifies most of the important cases it's supposed to catch. Also to identify true negative cases (class '0'), with a recall of 98%. This means it rarely misclassifies a negative case as positive.

The model has a good balance between precision and recall for both classes, as shown by the F1 scores. It is performing well at understanding between positive and negative classes, as shown by the AUC-ROC score of 0.97. This means it can easily tell the difference between the two classes, even with the imbalance. The model correctly identified a large number of negative instances and a good number of positive instances, as shown in the confusion matrix.

Even though the model's overall accuracy is good, the imbalance in the training and testing data may influence its performance as well. As we can see that the model is good at finding the majority class - negative(class '0') and has some confusions going around while dealing with the minority class - positive(class '1'). This can be seen in the confusion matrix itself. Here we performed SMOTE Oversampling technique to resolve the data imbalance. Bringing good balance between the classes in the datasets might give more good results.