

## Model Performance Report

### 1.Evaluation of the Model Using the Provided Metrics

The CatBoost classifier, selected for its superior performance over other models like Logistic Regression, Random Forest, and Gradient Boosting, demonstrated excellent results across all evaluation metrics. Below is a breakdown of the model's performance:

#### 1. Accuracy: 97.8%

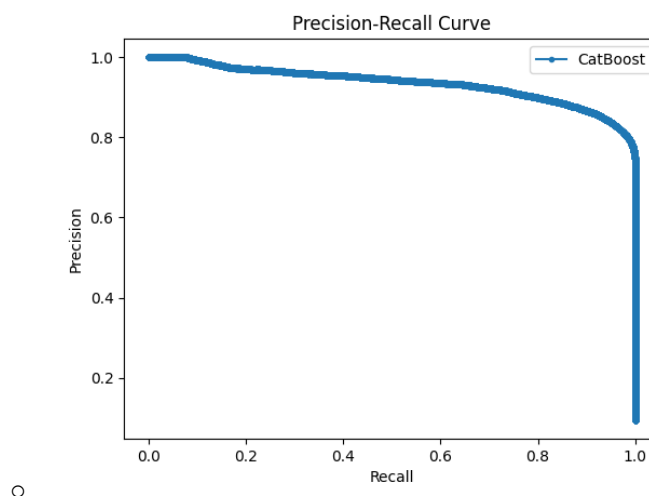
- **Explanation:** The accuracy indicates that the model correctly predicted the class labels for 98% of the test set. This high score shows that the model generalizes well to unseen data and successfully distinguishes between the target classes.

#### 2. Precision: 97.94%

- **Explanation:** Precision measures the proportion of true positive predictions out of all positive predictions made by the model. In this case, the model achieved 97.94%, meaning that nearly 98% of the time, when the model predicted a positive class (1), it was correct.

#### 3. Recall: 97.80%

- **Explanation:** Recall measures the proportion of actual positives that were correctly identified by the model. With a recall of 97.80%, the model correctly identified 98% of the true positives in the dataset.

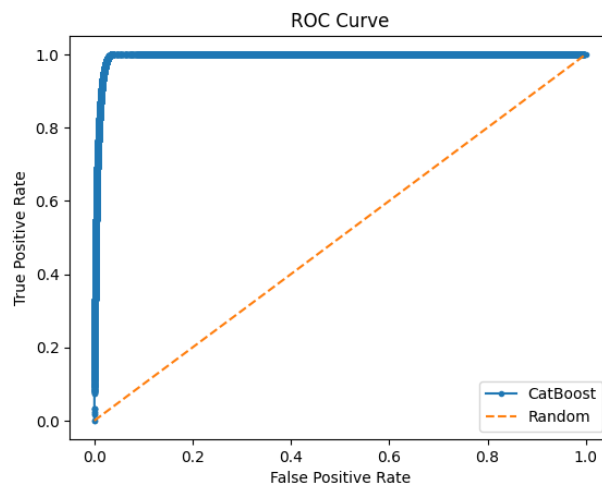


#### 4. F1-Score: 97.8%

- **Explanation:** The F1-score is the harmonic mean of precision and recall. It provides a balanced evaluation of both, ensuring that the model is not favoring either metric disproportionately.

## 5. ROC-AUC Score: 0.9946

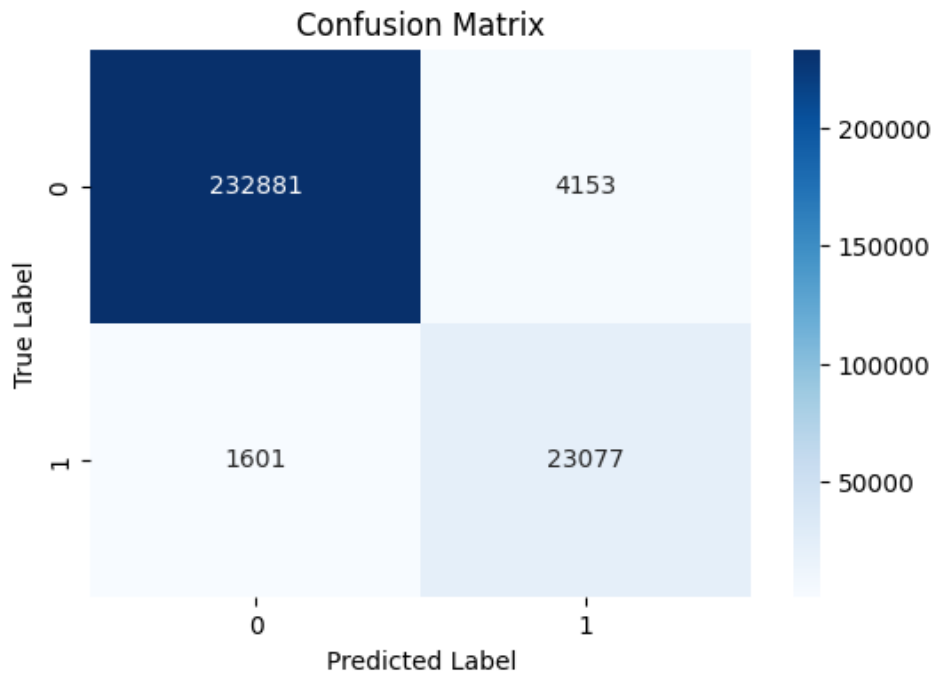
- **Explanation:** The ROC-AUC score measures the model's ability to distinguish between positive and negative classes. A score of 0.9946 is near perfect, indicating that the model is highly effective at ranking instances and predicting the correct class with a very high degree of confidence.
- **Significance:** This metric is especially useful when the dataset is imbalanced or when false negatives and false positives need to be weighed differently.



○

## 6. Confusion Matrix:

The confusion matrix reflects the model's performance in identifying correct and incorrect predictions. The high values of TP and TN, combined with the low FP and FN, indicate excellent model accuracy and generalizability.



## 2. Insights and Analysis Derived from the Model's Predictions

### 1. CatBoost Outperformance:

- CatBoost outperformed other algorithms like Logistic Regression, Random Forest, and Gradient Boosting in terms of handling both numerical and categorical variables. Its ability to efficiently process these features without the need for extensive preprocessing made it the ideal choice for this problem. Additionally, the algorithm's speed and capability of working well with relatively unbalanced datasets contributed to its success.

### 2. Low Preprocessing Overhead:

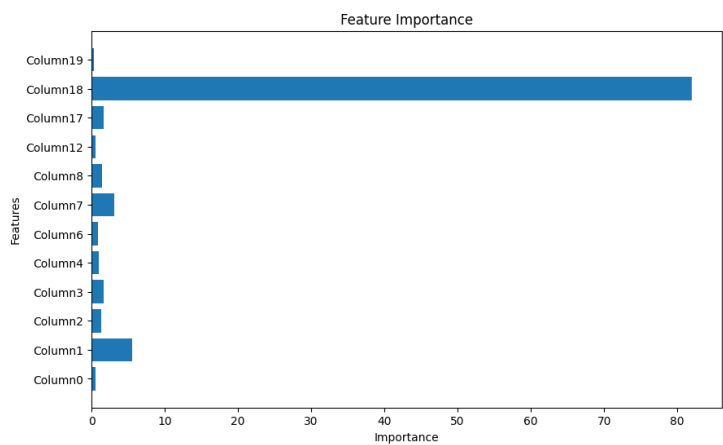
- One of the standout features of CatBoost is its low preprocessing requirements. By handling numerical and categorical variables natively, the need for complex feature engineering is reduced, significantly lowering the time and effort required for model development. This made CatBoost highly efficient not only during training but also in deployment.

### 3. Superior Classification Performance:

- The ROC-AUC score of 0.9946 highlights the model's exceptional ability to distinguish between the positive and negative classes. This high score

indicates that the model ranks predictions very well, ensuring that the true positives and true negatives are correctly separated.

4.Feature importance



5.Distribution of true and predicted label:

