PROJECT PRSENTATION

Presented by GSTN_1011

ABSTRACT

This project focuses on analyzing machine learning models for classification tasks, evaluating their performance, and identifying areas for improvement. The analysis leverages various metrics such as ROC-AUC and feature importance to assess model efficacy. The goal is to identify key factors influencing prediction accuracy and recommend strategies to improve model performance through balancing class predictions and tuning hyperparameters.

APPROACH

Data Preprocessing

1. **Iterative Imputer**:
    - Fills missing values by modeling them based on existing features, enhancing dataset quality.
2. **Interpolation**:
    - Estimates missing data points to ensure continuity and preserve data integrity.
3. **Scaling Transformation**:
    - Standardizes features for optimal model performance, preventing range dominance.

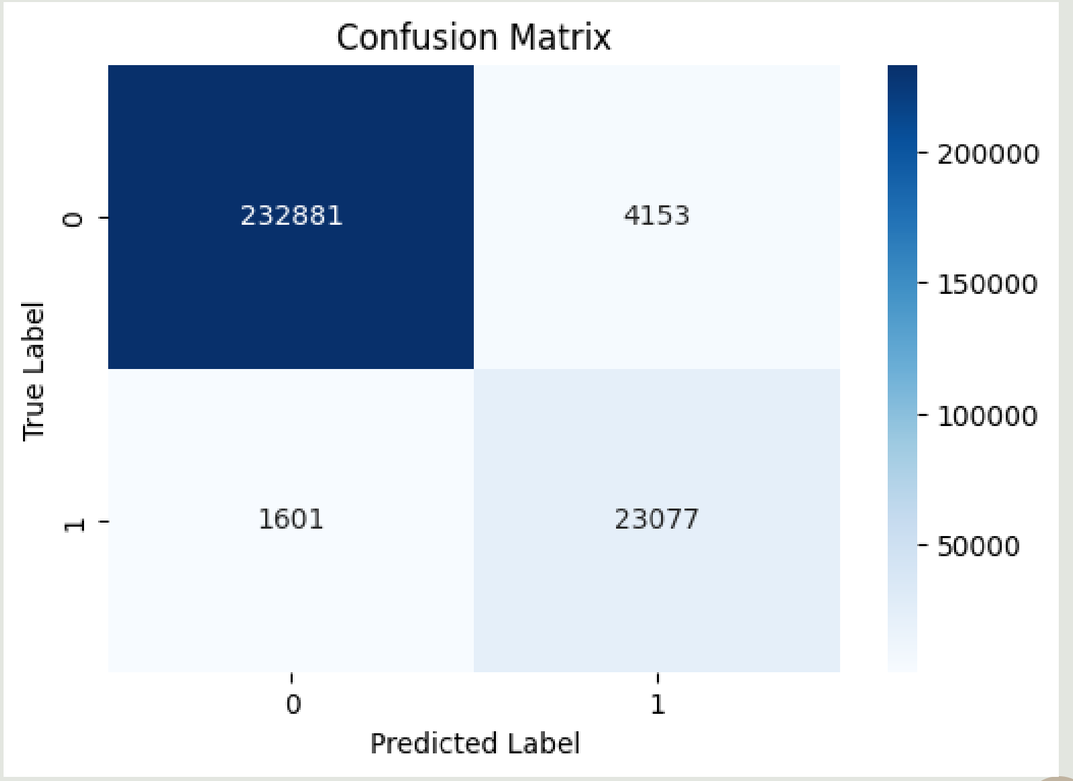
CatBoost Model Training

1. **CatBoost Selection**:
    - Selected for its robust handling of numerical and categorical data, achieving high accuracy while minimizing overfitting.

PERFORMANCE

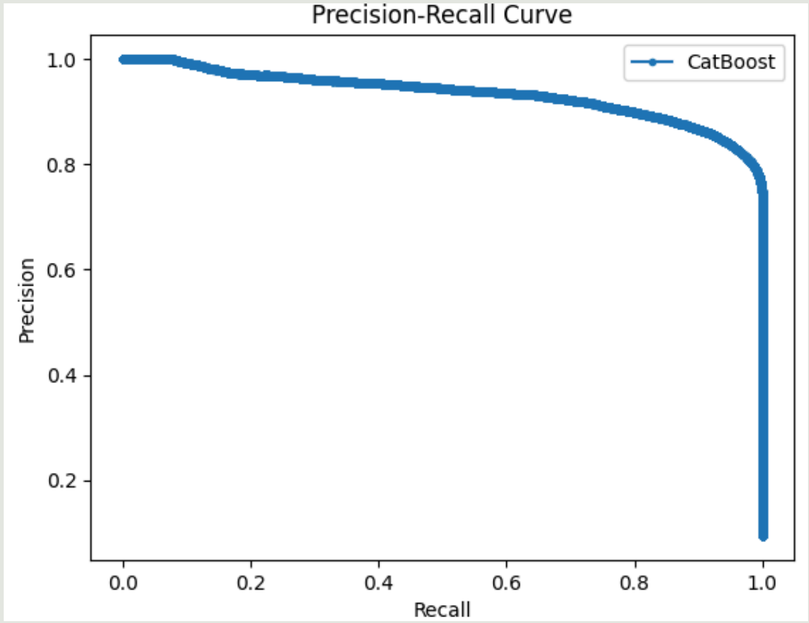
- **Accuracy: 97.8%** – The model demonstrates strong generalization across the dataset, correctly classifying the vast majority of instances.

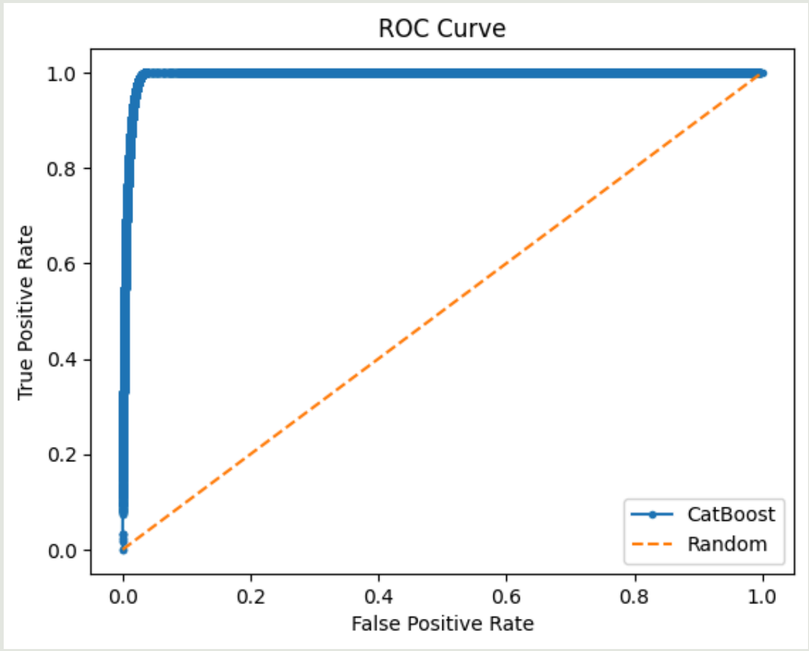
- **Precision: 97.94%** – The model is highly accurate in its positive predictions, with minimal false positives, ensuring reliability in correct classifications.



PERFORMANCE

- **Recall: 97.80%** – The model has high sensitivity, accurately identifying most actual positive cases.
- **F1-Score: 97.8%** – A strong balance between precision and recall, indicating robust overall performance.
- **ROC-AUC: 0.9946** – Excellent ability to distinguish between classes, showing near-perfect classification performance.





FINDINGS

- **CatBoost Dominance:** Outperformed models like Logistic Regression and Random Forest, with a ROC-AUC of 0.9946, excelling in class differentiation.
- High Performance Metrics:
- **Accuracy:** 97.8%, Precision: 97.94%, Recall: 97.80%, and F1-Score: 97.8% – Indicating strong generalization, low false positives, high sensitivity, and balanced performance.
- **Class Discrimination:** ROC-AUC of 0.9946 confirms excellent class ranking accuracy.
- **Balanced Trade-off:** Near-equal precision and recall ensure reliability in both sensitivity and precision.
- **Efficiency:** Minimal preprocessing, making it easy to train and deploy.

RECOMMENDATIONS

- **CatBoost Recommendation:** CatBoost is highly recommended for its exceptional performance in classification tasks, particularly in handling numerical features efficiently.
- **Monitoring and Retraining:** Regular monitoring and retraining with new data are essential to maintain the model's accuracy and adapt to evolving data patterns.
- **Hyperparameter Tuning:** Future efforts should explore hyperparameter tuning to further enhance model performance and improve classification metrics.

Thank You

For your attention