PRESENTED BY:

Anirvan Roy

GSTN Hackathon 2024

DATE

11 Oct 2024

Flow of the Project

01.

Imputation of NaNs

After careful study and EDA, various approaches have been undertaken

02.

Training the Model

Multiple models were trained to check for best performing model

03.

Feature Engineering

Based in Feature Importance derived from the model, some new features were introduced to enhance performance

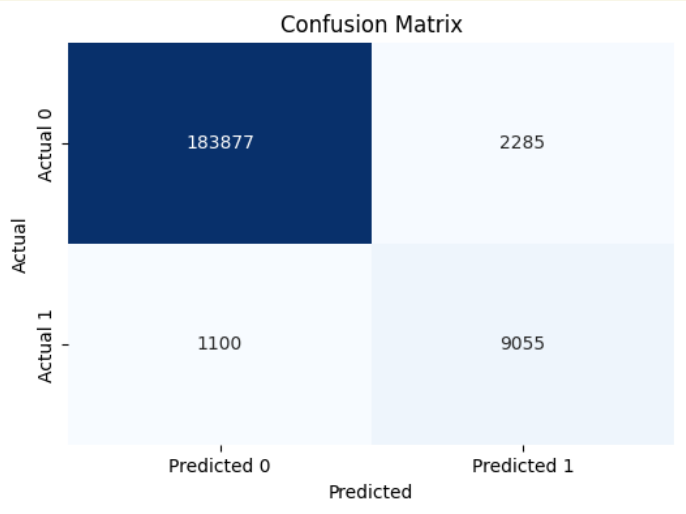
Key Metrics for LightBGM on Test Data

Accuracy
98.2 %

AUC-ROC
99.5 %

Precision
79 %

Confusion Matrix



Recall
89 %

F1 Score
0.84

References

Di Russo, J. (2021, December 12). Navigating The Hell of NaNs in Python - Towards Data Science. Medium. https://towardsdatascience.com/navigating-the-hell-of-nans-in-python-71b12558895b

Innovative Ways to Enhance ML Models with Feature Engineering. (n.d.). https://www.markovml.com/blog/feature-engineering-best-practices

Loukas, S., PhD. (2024, February 23). How To Perform Feature Selection for Regression Problems. Medium. https://towardsdatascience.com/how-to-perform-feature-selection-for-regression-problems-c928e527bbfa

Parsons, N. (2023, November 30). Feature engineering for fraud detection. Fennel Blog. https://fennel.ai/blog/feature-engineering-for-fraud-detection/

Peeters, J. (2021, December 14). A framework for feature engineering and machine learning pipelines. Medium. https://medium.com/manomano-tech/a-framework-for-feature-engineering-and-machine-learning-pipelines-ddb53867a420

Thank You