

A decorative graphic on the left side of the slide, consisting of a network of light blue lines and small circles, resembling a circuit board or a neural network diagram.

CREDIT CARD FRAUD DETECTION

PROBLEM STATEMENT

There were over 390,000 reports of credit card fraud in 2020

In 2020, credit cards were most frequently identified as the payment method in fraud reports



DATA

- Dataset from Kaggle.com
- 1852394 observations
- 1842743 legit transactions
- 9651 fraud transaction
- Target: true class – fraudulent transaction
false class – legit transaction



METHODOLOGY

- Data acquisition

[kaggle.com](https://www.kaggle.com)



METHODOLOGY

- Data acquisition

[kaggle.com](https://www.kaggle.com)

- EDA

Dealing with missing values

Categorical features conversion

Handling class imbalance

METHODOLOGY

- Data acquisition

kaggle.com

- EDA

Dealing with missing values

Categorical features conversion

Handling class imbalance

- Model Selection and Training

Logistic Regression

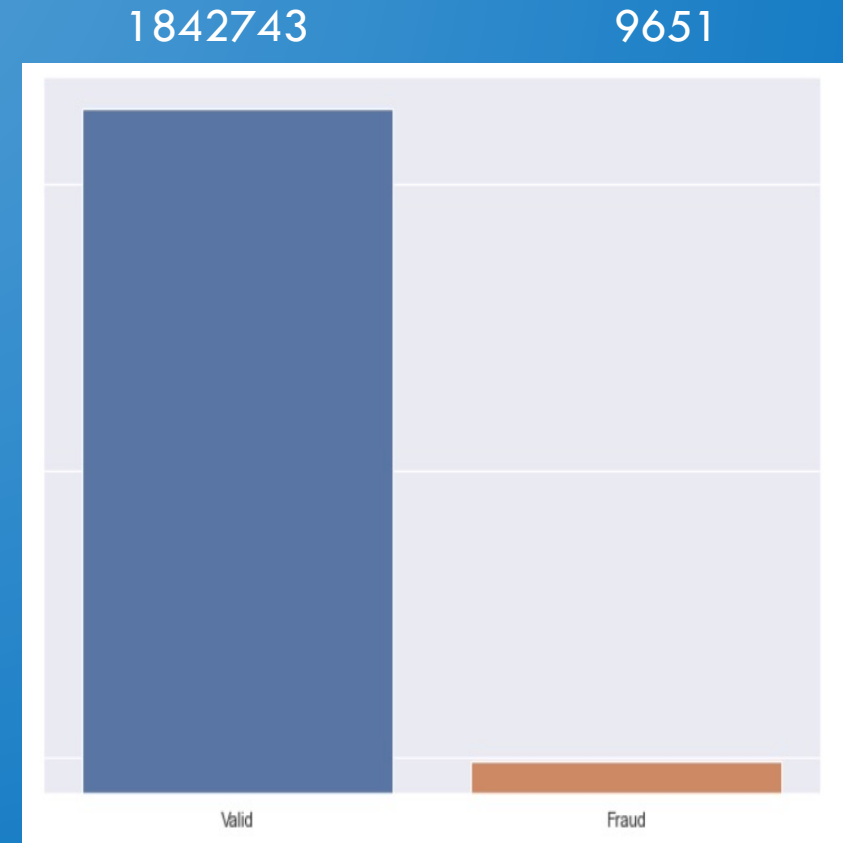
Random Forest

XGBoost

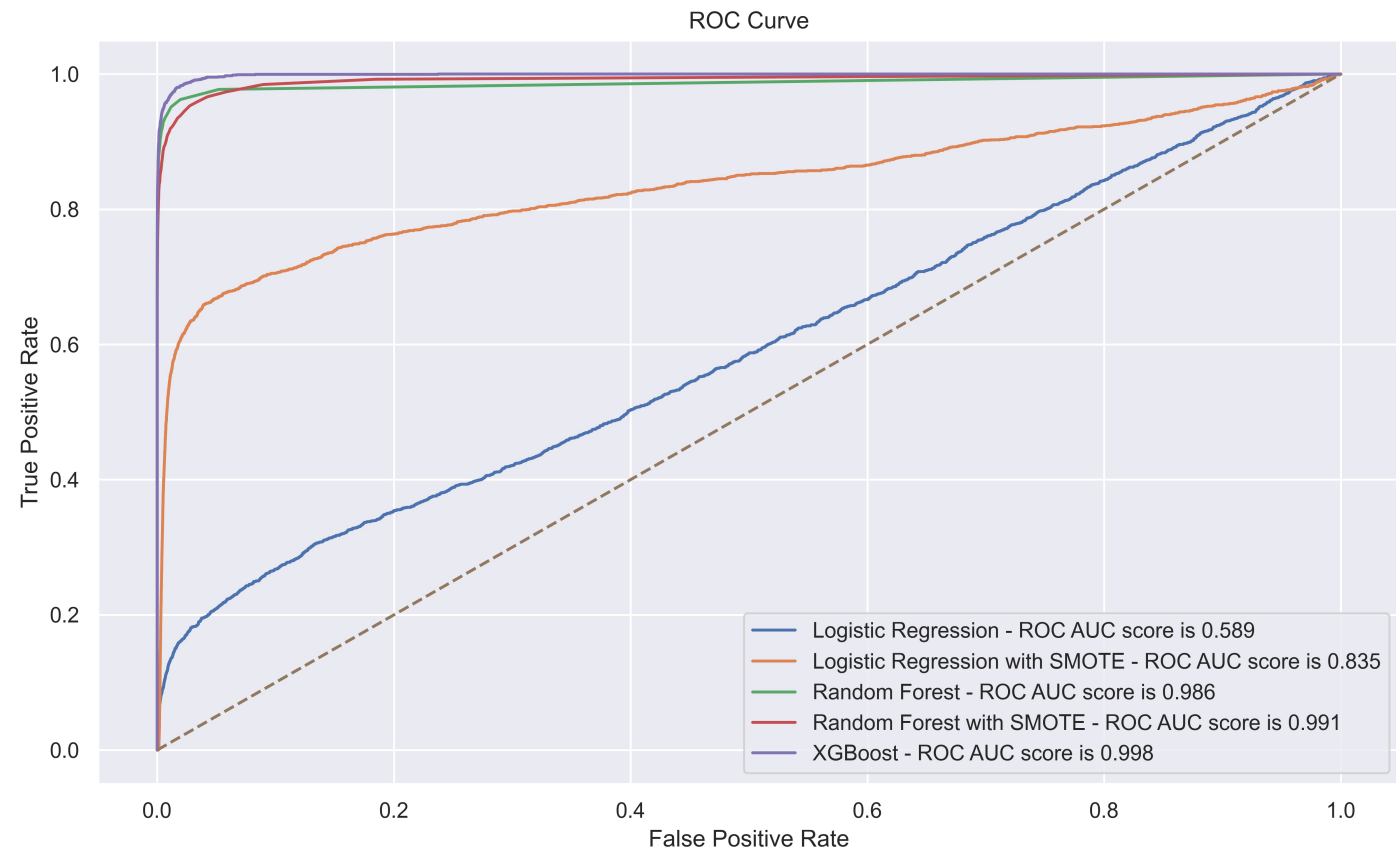
HANDLING CLASS IMBALANCE

Stratified Train – Test Split

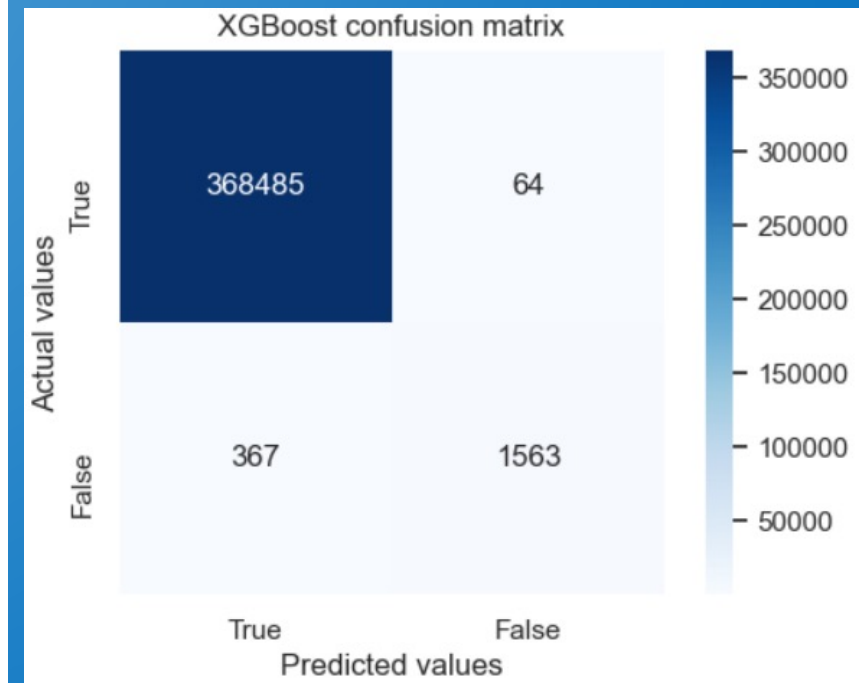
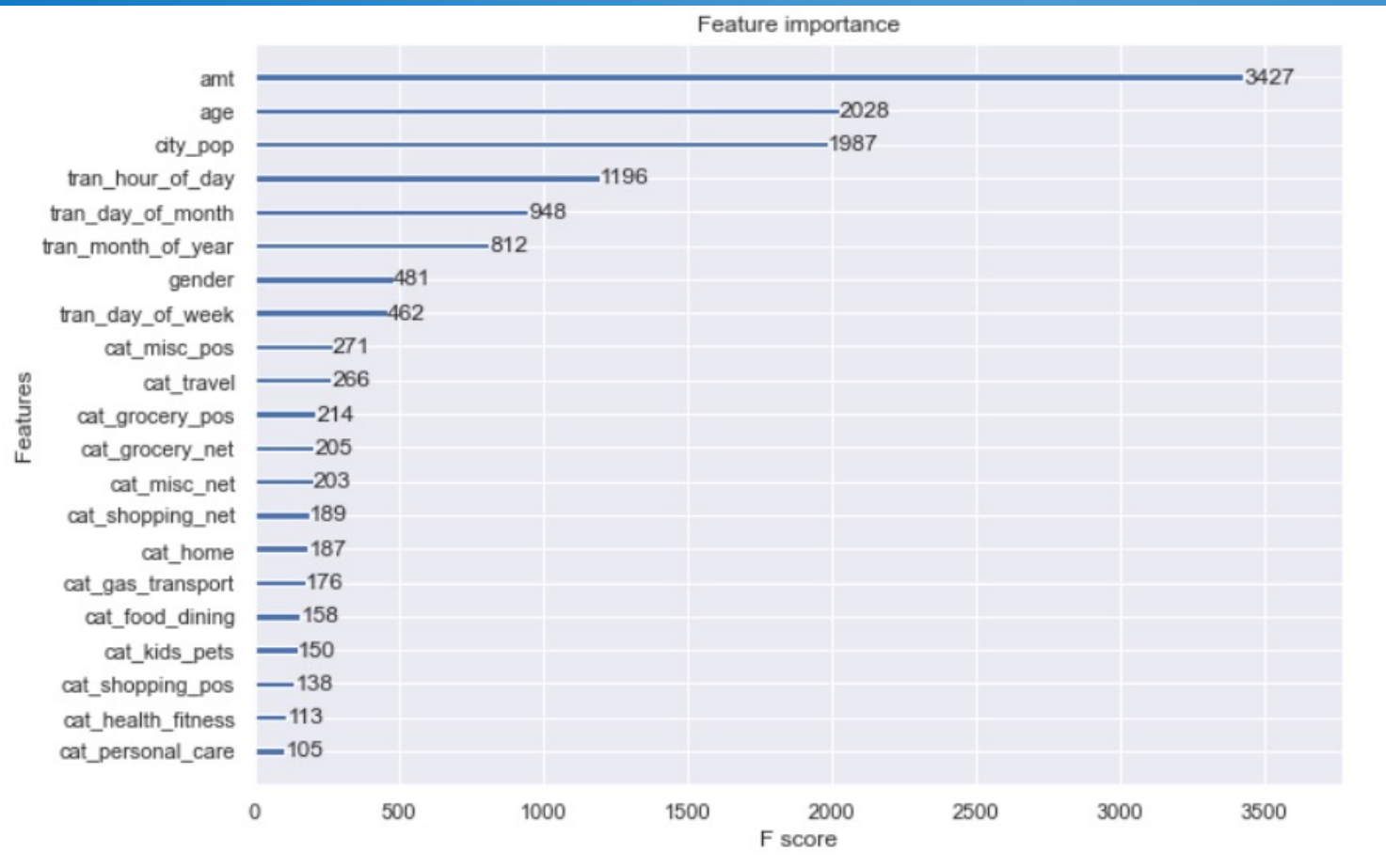
SMOTE to correct class imbalance



RESULTS

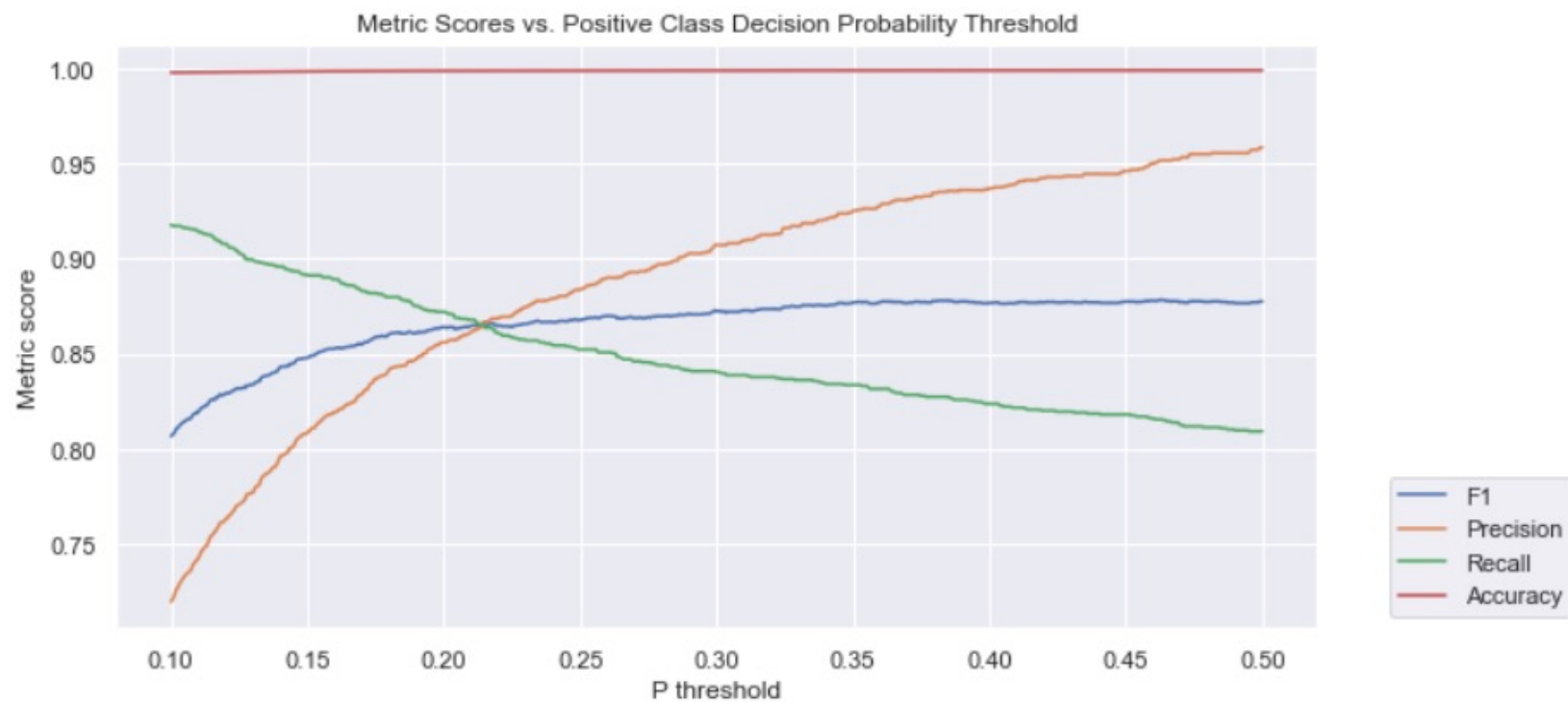


FINAL MODEL - XGBOOST



MODEL TUNING

XGBoost best F1 score 0.878 at prob decision threshold ≥ 0.462



CONCLUSIONS AND FUTURE WORK

- Classification models are good with identifying credit card fraud
- Spend more time on model tuning
- Explore more data, add more features to the analysis

The background is a solid blue color. In the four corners, there are decorative white line art elements that resemble electronic circuit traces. These lines connect to small white circles, some of which are larger than others. The lines are thin and sharp, creating a modern, tech-oriented aesthetic.

Thank you