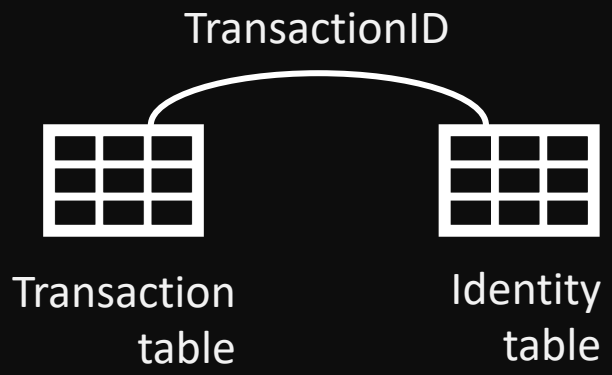
Online Transaction Fraud Detection

Jung-a Kim

Online Payment Fraud

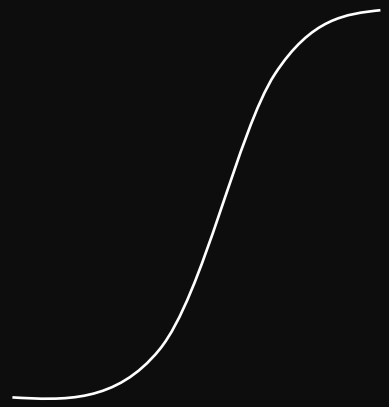
- Online sales make up 17% of all global sales.
- For every $1 of fraud from chargebacks, e-commerce businesses lose an extra $2.94.
- Fraud also impacts brand and customer loyalty.



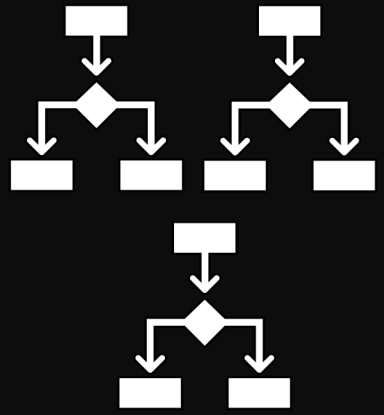
Dataset

- 590,540 transactions with labels
- 383 numeric features
  - e.g) Vesta- engineered rich features , time(D), payment amount, count(C), identity information
- 50 categorical features
  - e.g) Product/service code, address, card type, device type, identity information
- Class imbalance
  - 3.5% fraudulent transactions

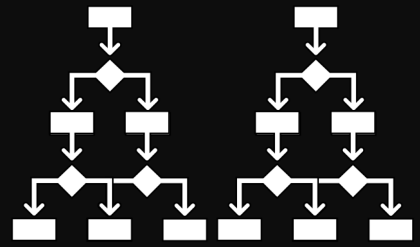
Candidate models



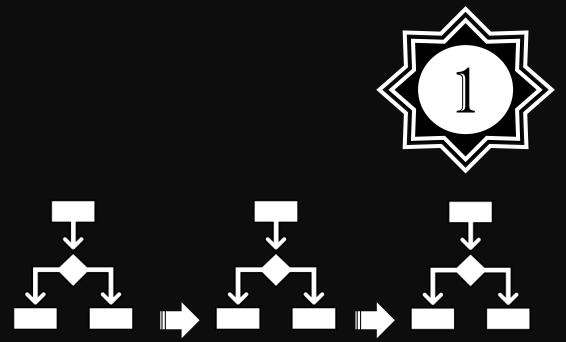
Logistic Regression



Random Forest

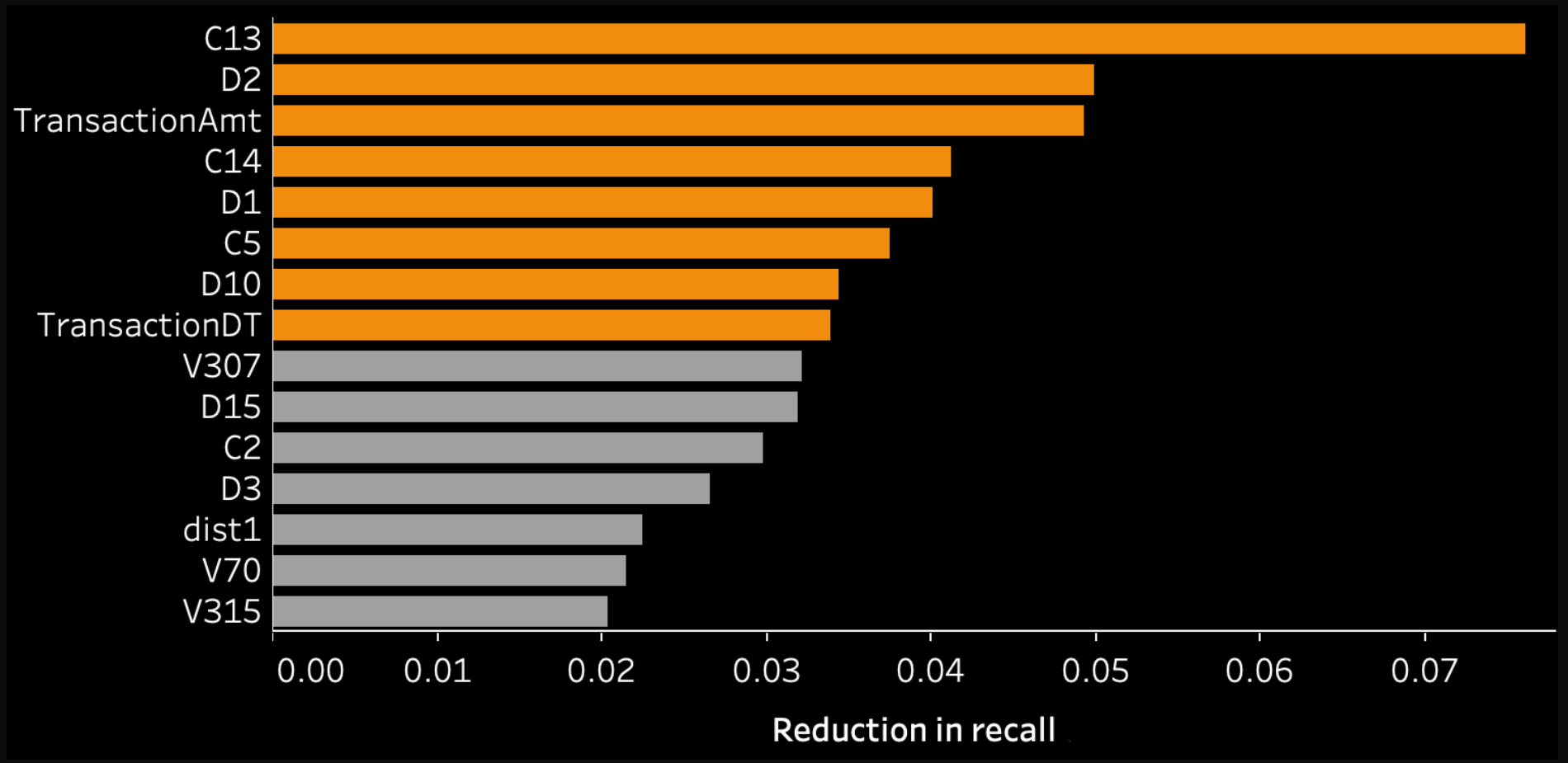


Extra Trees

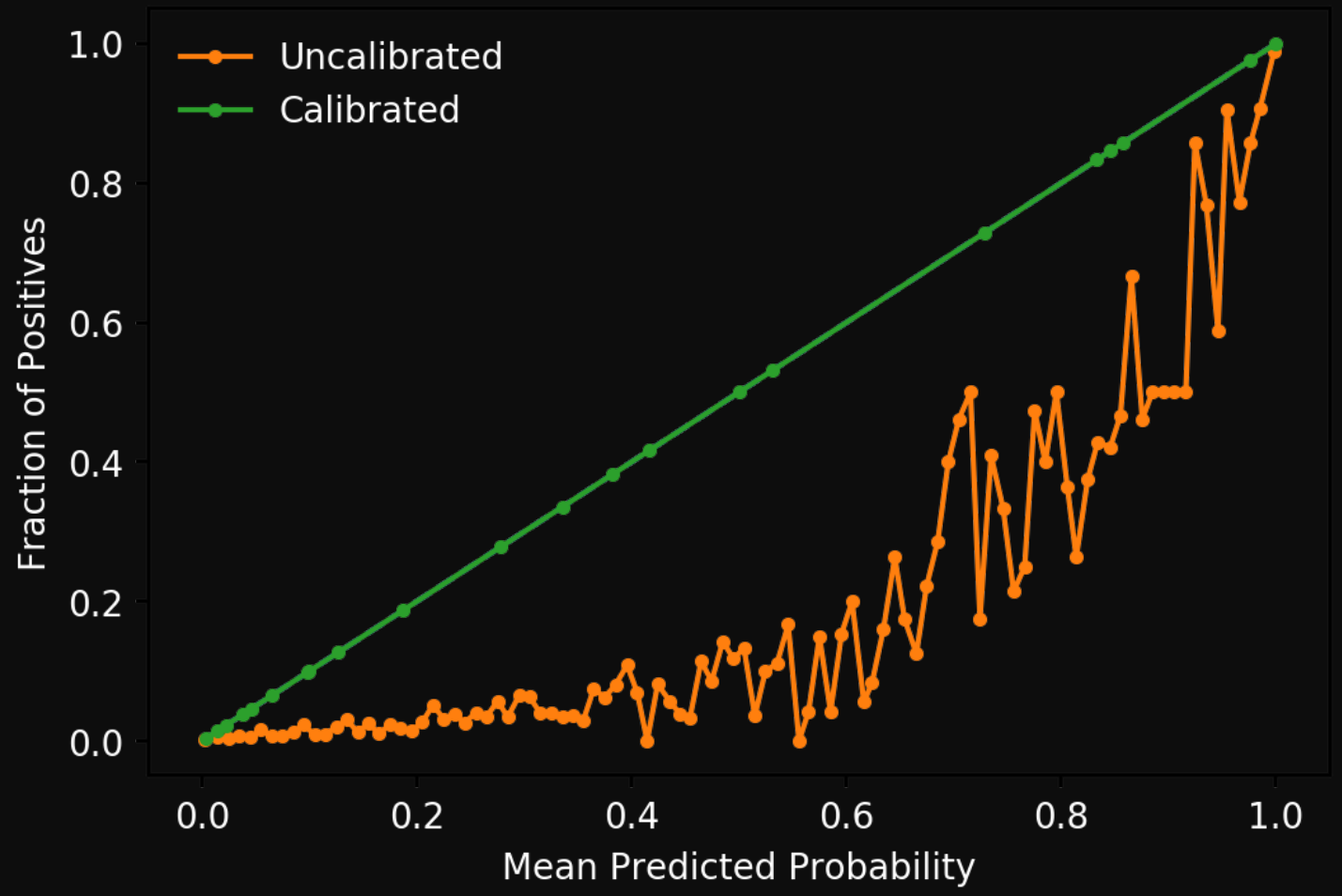


XGBoost

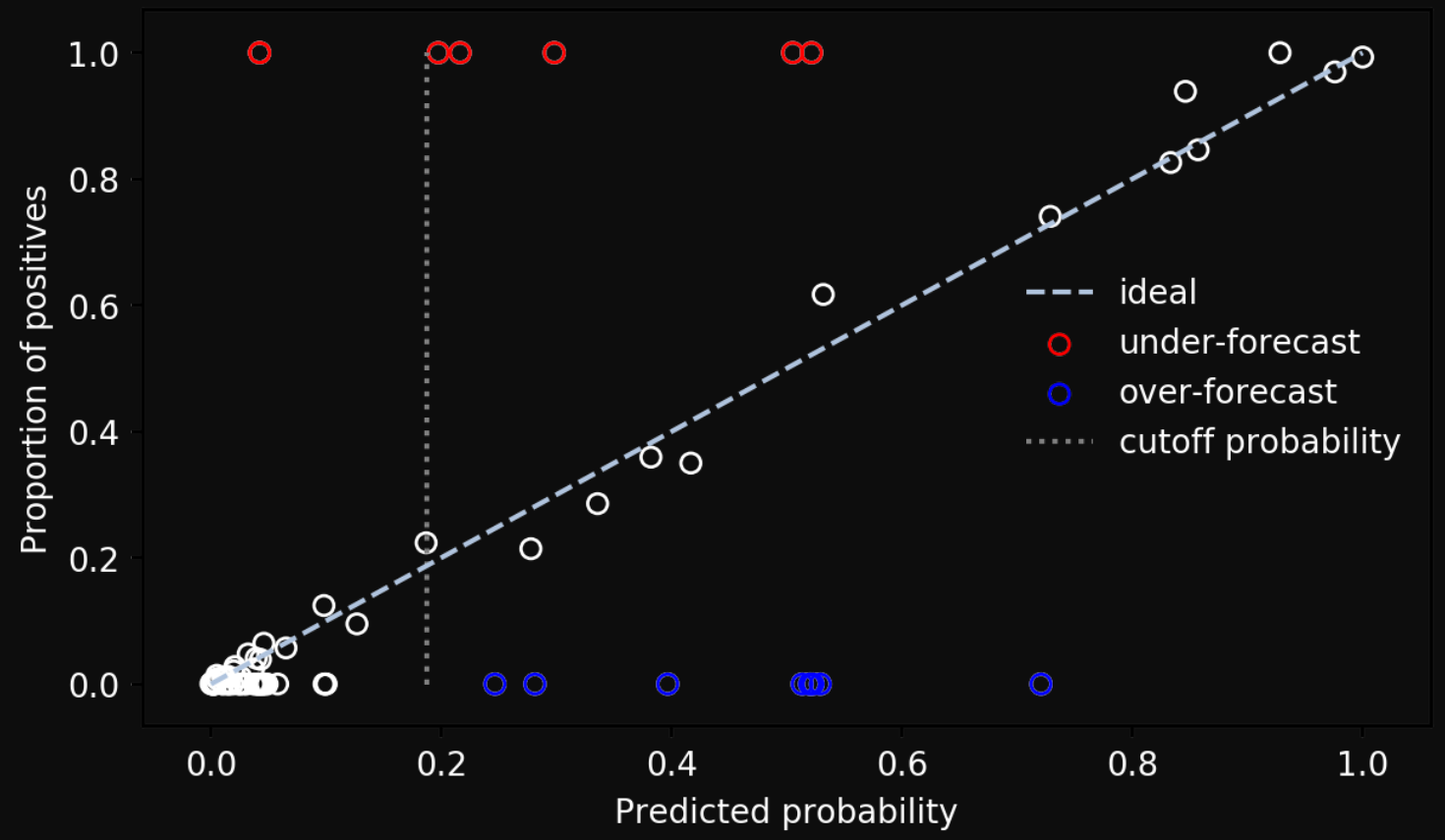
Permutation Feature Importance



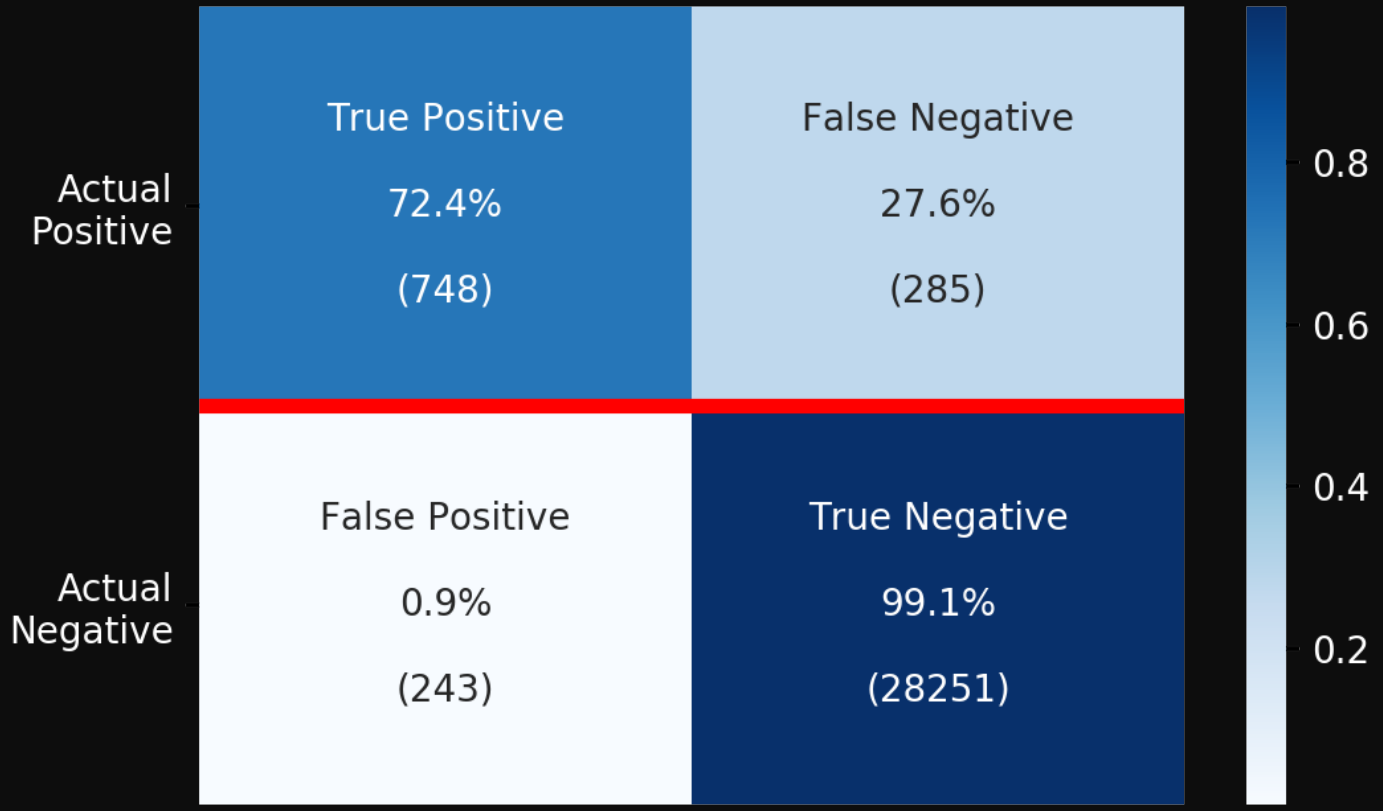
Calibration on validation set



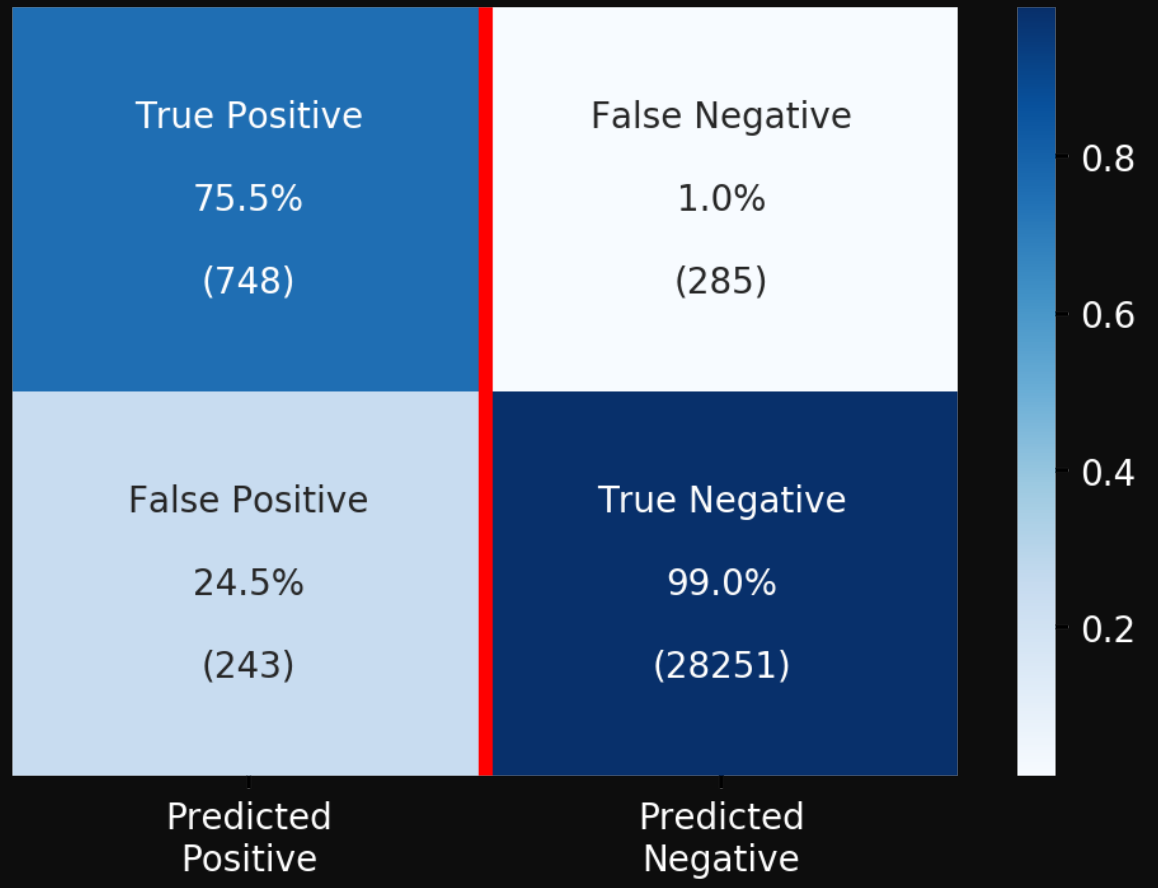
Calibration on test set



Confusion Matrix



Confusion Matrix



Conclusion

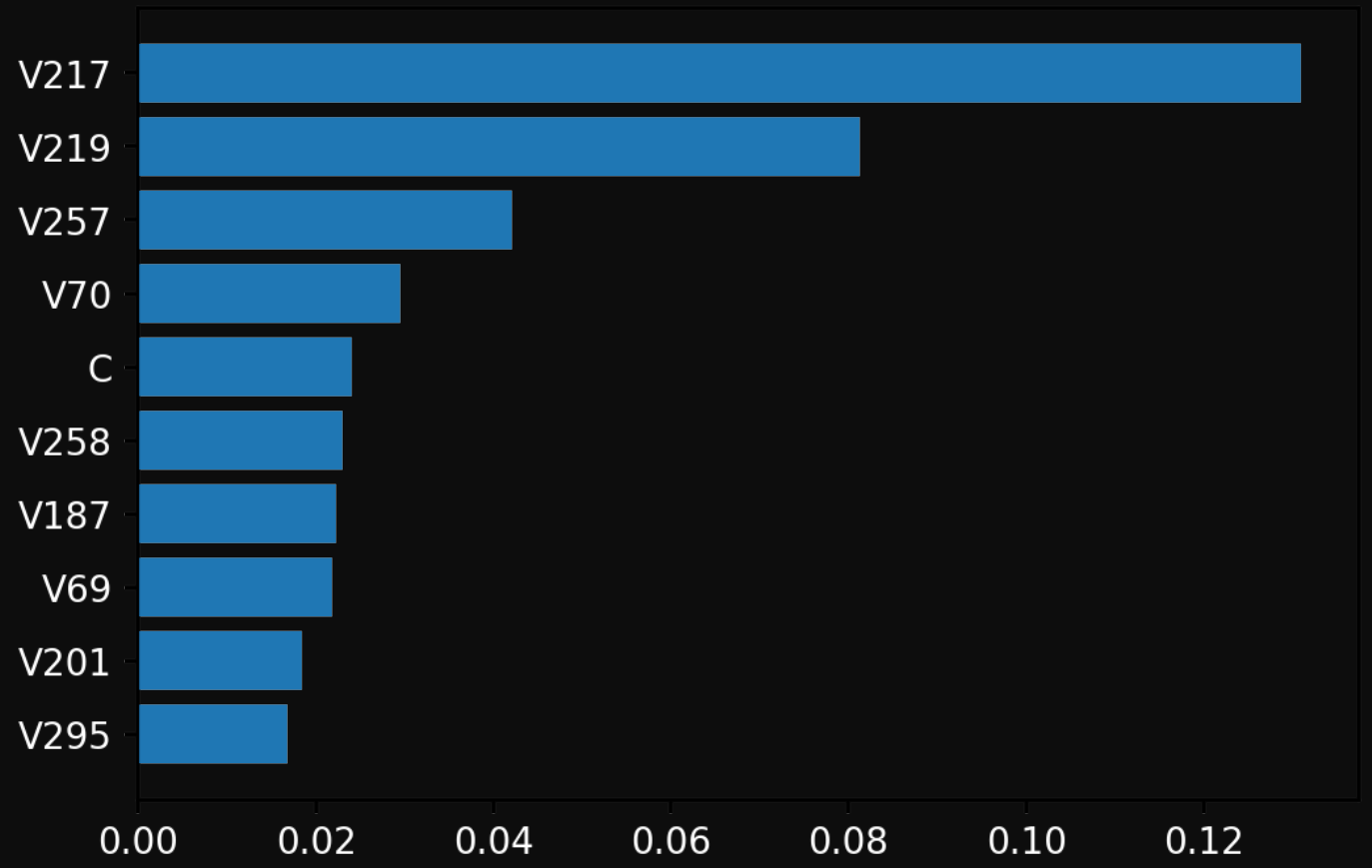
- Count, time, payment amount >>> product, address, card types, email domain
- Predicted probabilities ~ true probabilities
- Moderate recall and precision
- High specificity and negative predictive value

Future work

- Fine-tuning XGBoost
  - Tune Maximum depth, minimum samples-to-split, subsample ratio, γ, λ, etc
- Simpler model
- Domain research

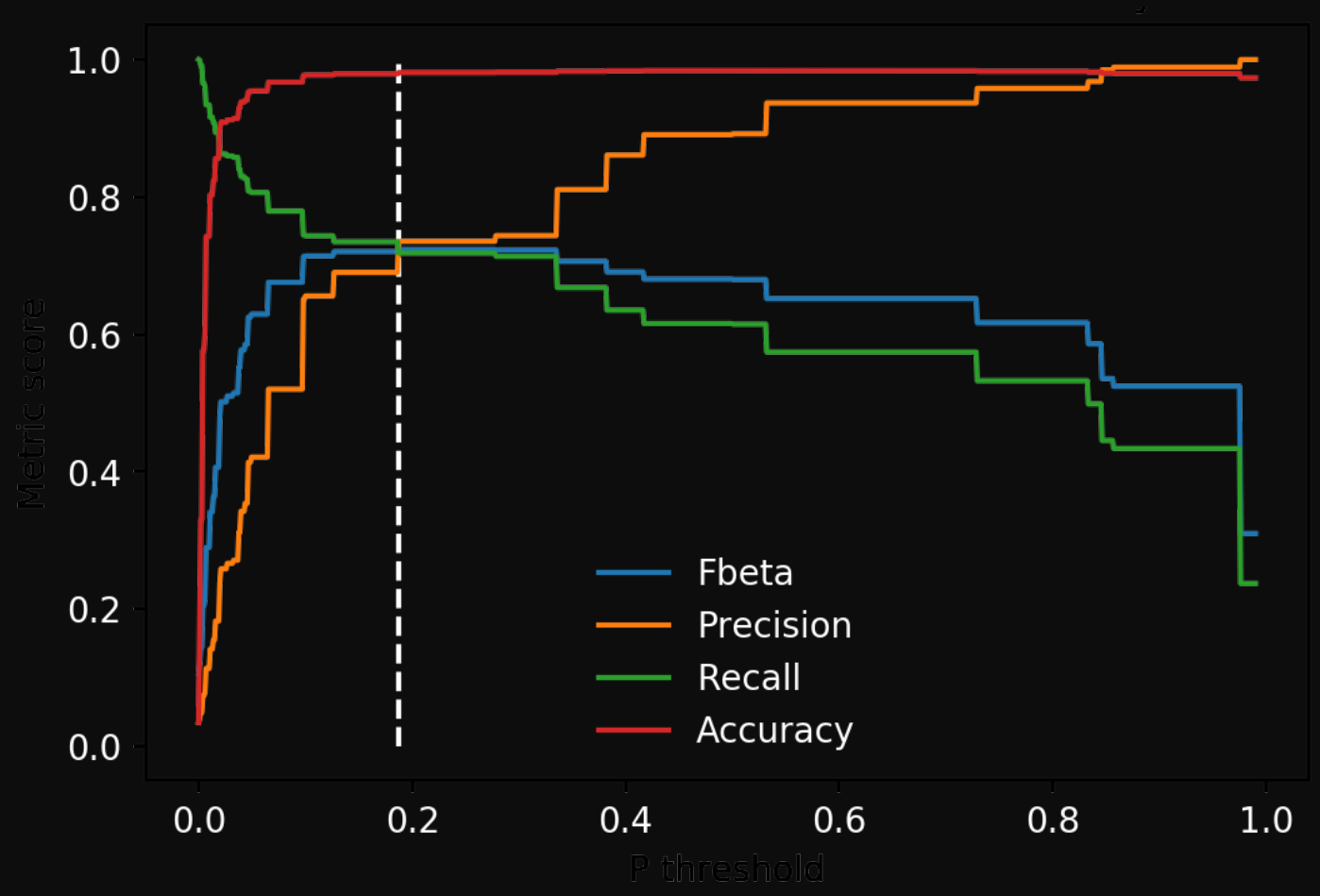
Thank you

Feature Importance

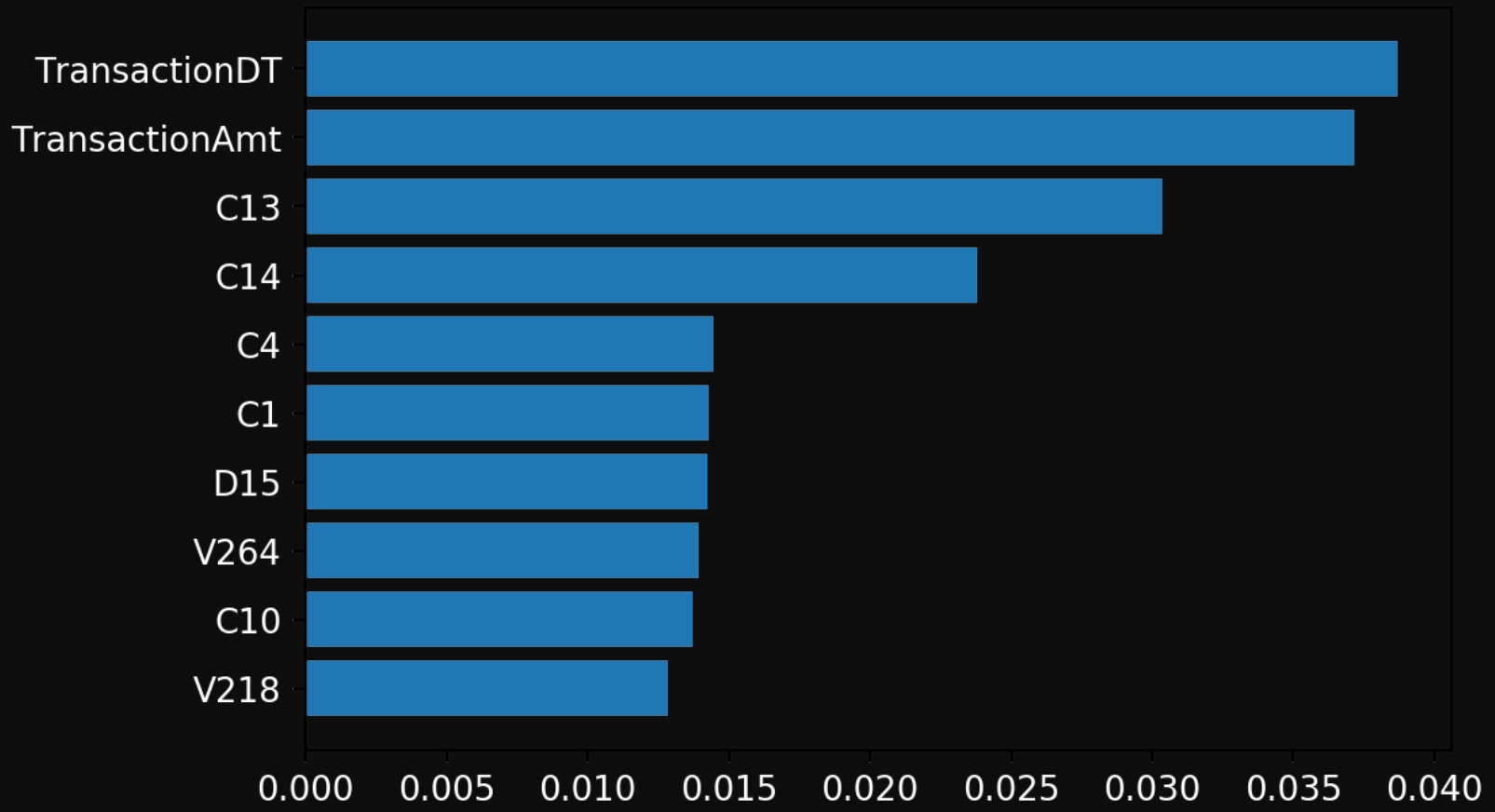


Vesta-engineered features are advantaged

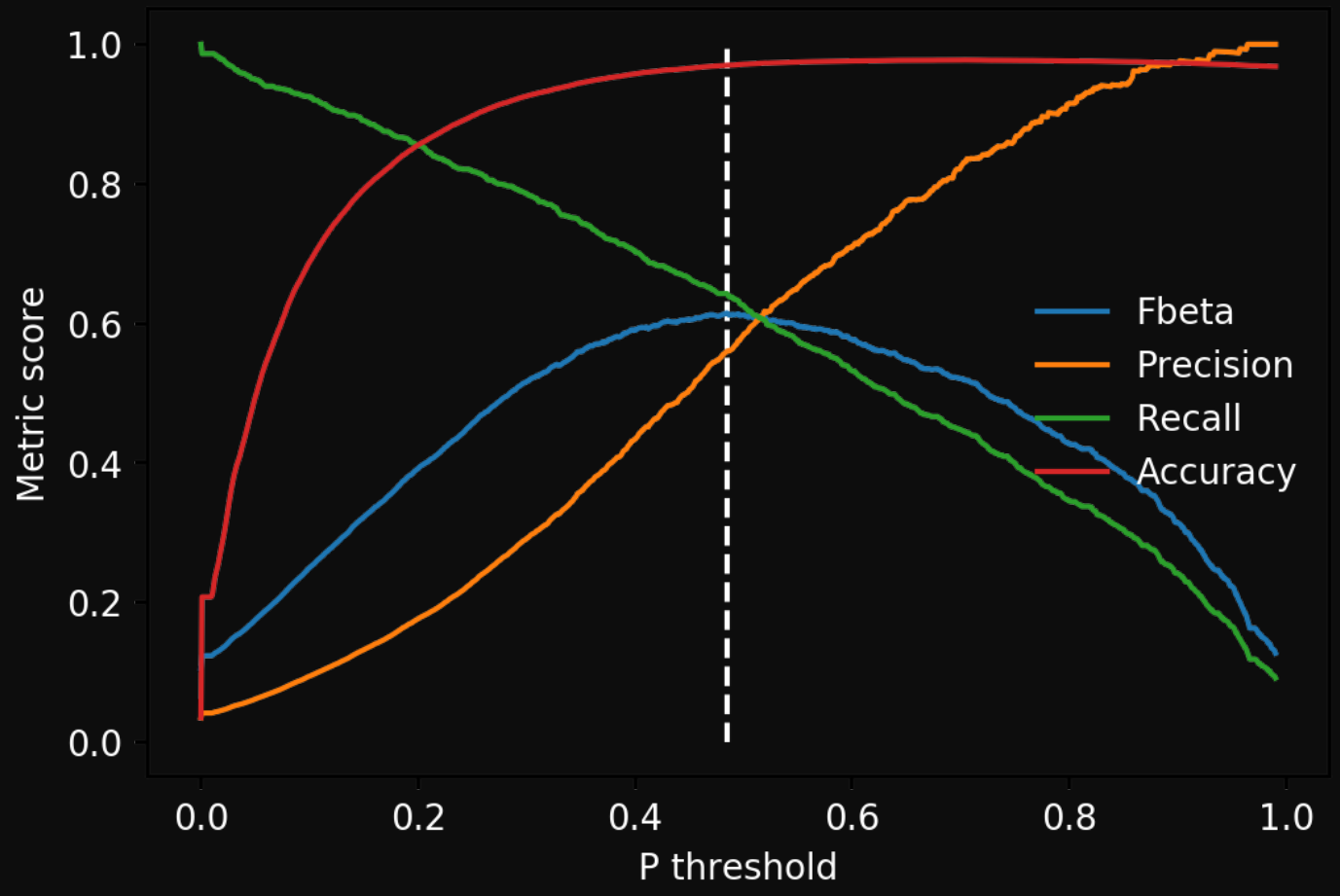
Recall-Precision



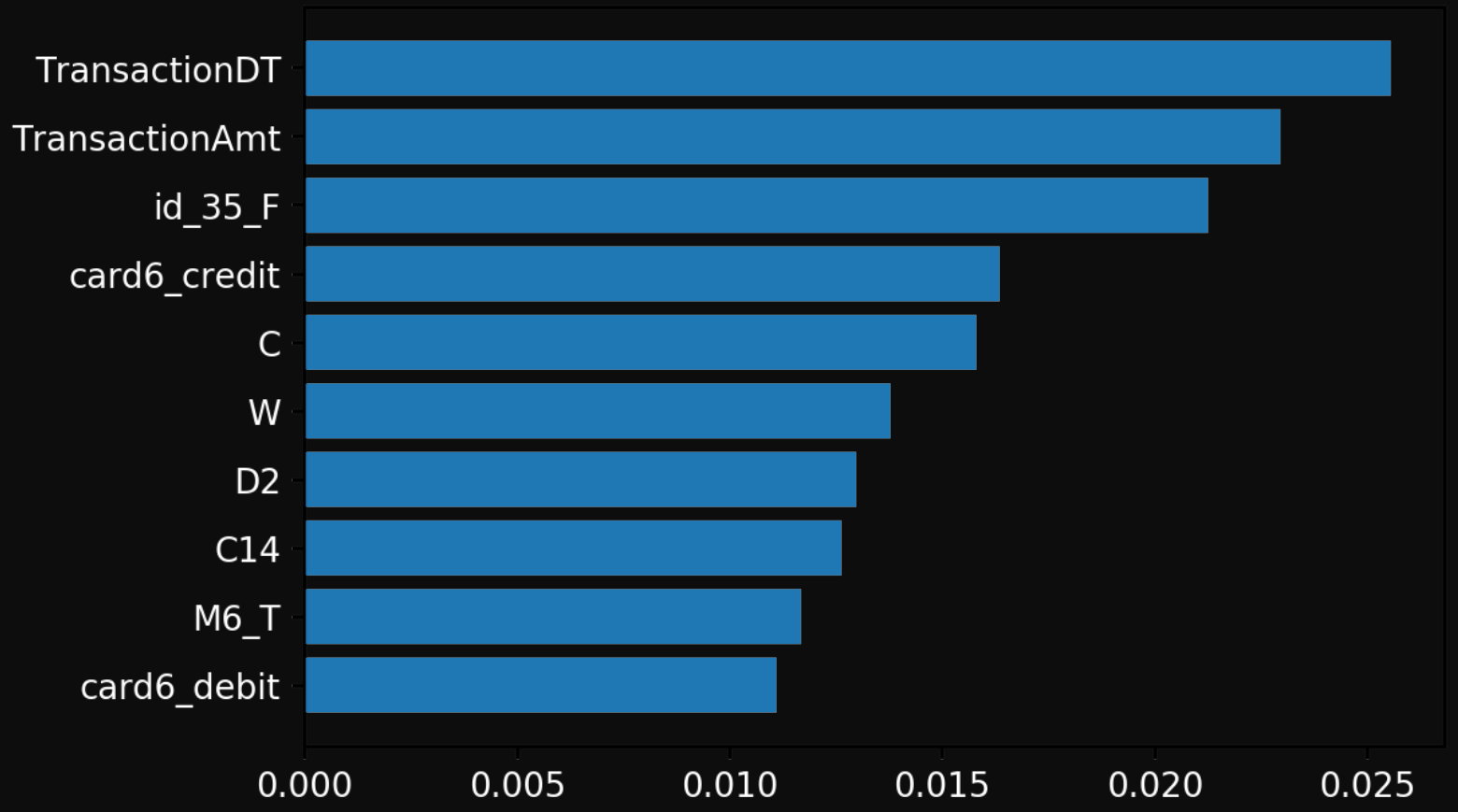
Random Forest Feature Importance



Random Forest Recall-Precision



Extra Trees
Feature Importance



Extra Trees
Recall-Precision