PREVENTING FRAUD WITH DATA SCIENCE

CATEGORIZING FRAUDULENT MASTERCARD TRANSACTIONS

ANALYSIS BY:
JOE COWELL

The worldwide cost of credit card fraud…

$24 BILLION

…in 2018[1]

kaggle

vesta

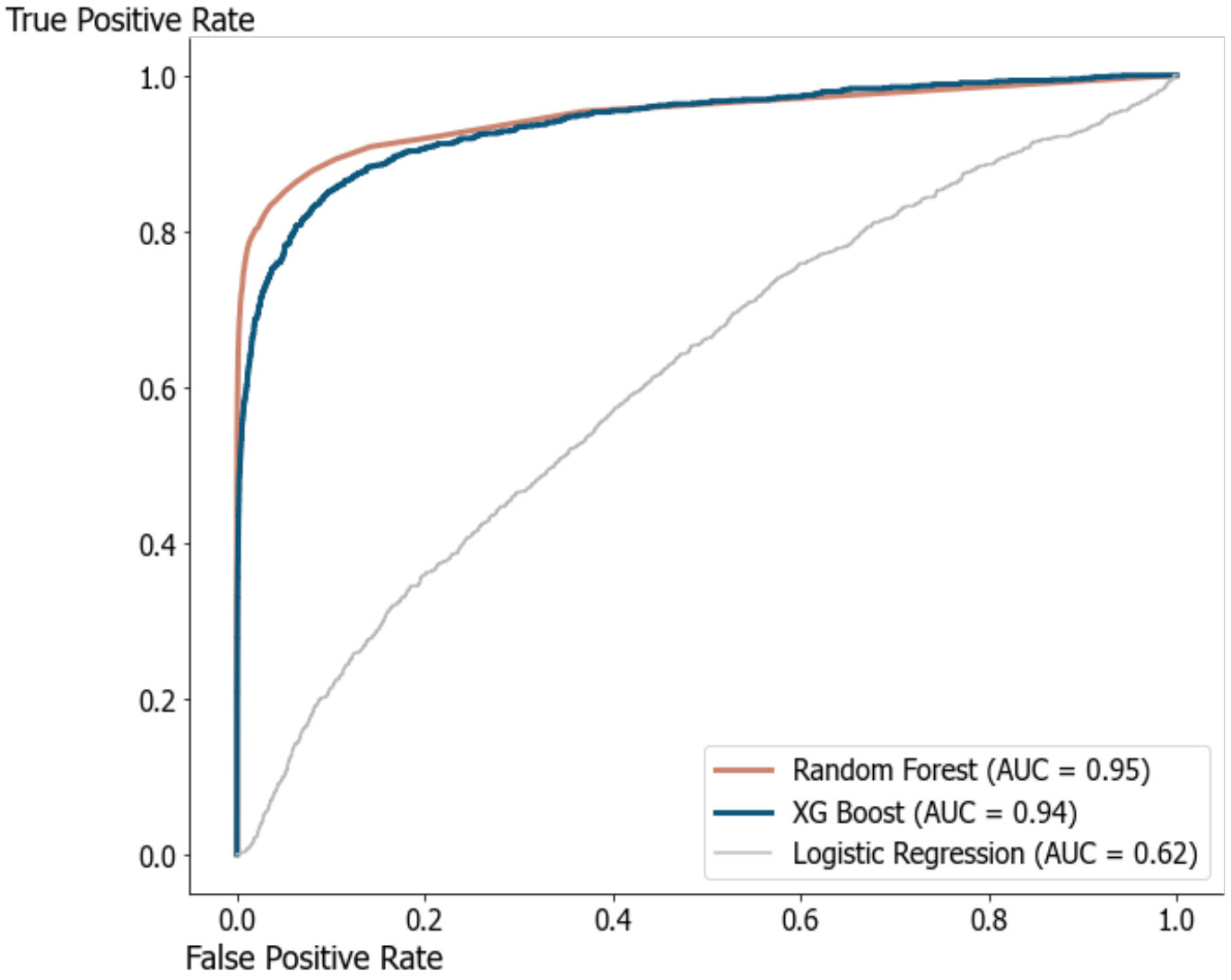
mastercard

LOOKING INTO THE DATA:$_2$

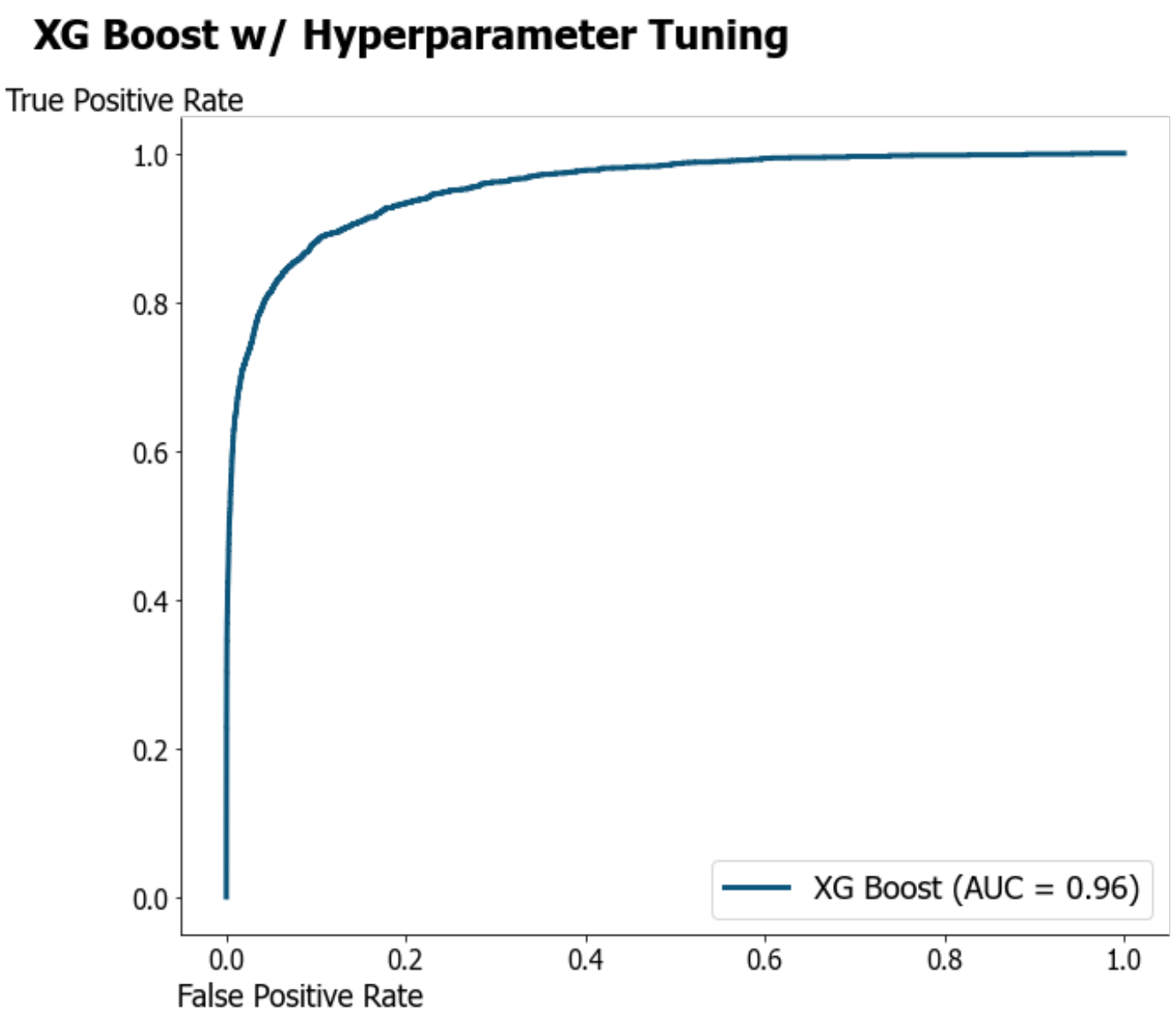
Transactions
189,217

Features		
Matched information	Timedelta	Transaction Amount
Debit vs. Credit Card	Product Code	Card Information

Logistic Regression is out of the Picture



ENSEMBLE METHODS DOMINATE



XG BOOST IMPROVED WITH TUNING

Predicting Fraud for Mastercard

Model: XG Boost

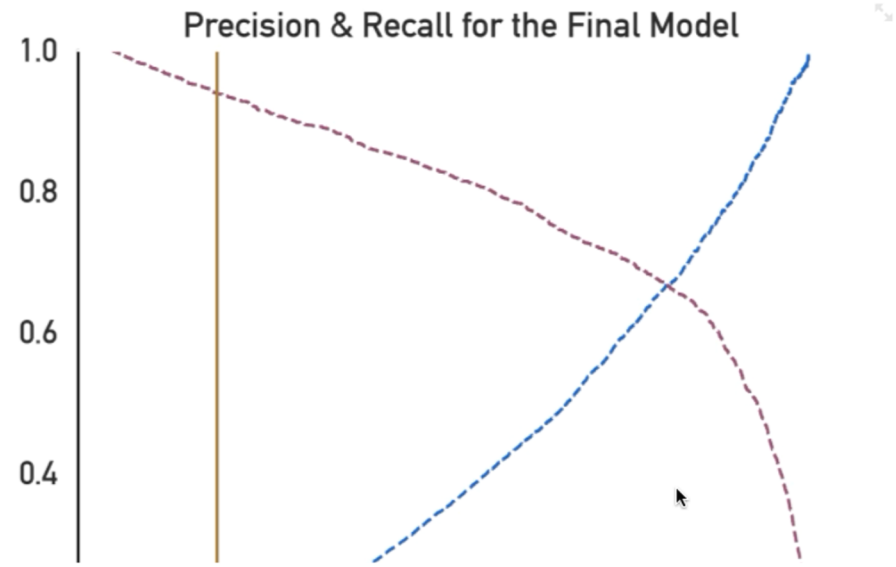
The model behind the scenes consists of Mastercard transaction data. Move the 'Probability Threshold' slider to tailor the model and see how it performs. This app will help it's user decide how the model works and find the perfect threshold for it's final purpose.

Total observed transactions: 56,766

Probability Threshold

0.15

0.00 1.00



AN IDEAL XG BOOST MODEL

(56,776 OBSERVATIONS)

Threshold

- 0.25

Recall

- 90.20%

False Positives

- 13.75%

Savings

- $227,573.27

296 MILLION MASTERCARD TRANSACTIONS PER DAY[1]

Savings of…

$1.18 BILLION

…per day

- Joe Cowell
- Github: github.com/josephpcowell/

WORKS CITED

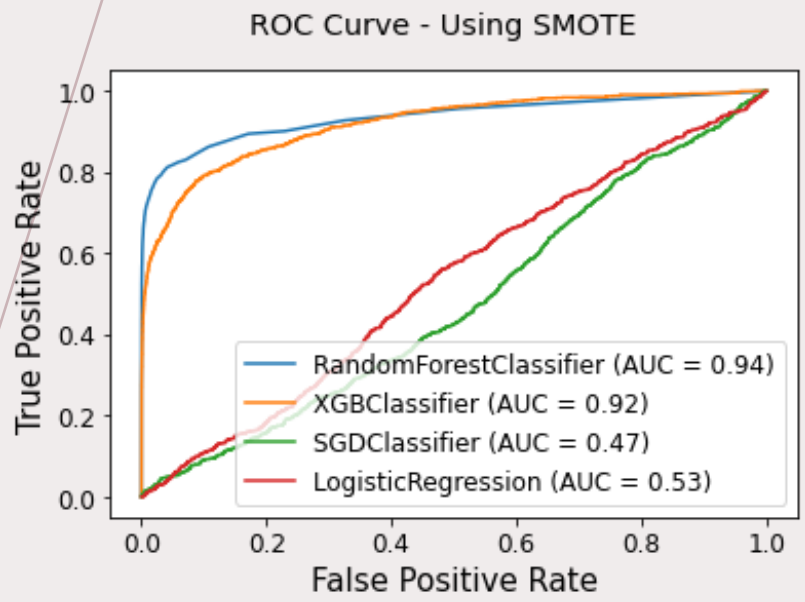
1. Letić, Jovana. “Credit card fraud statistics: What are the odds?”. December 10, 2019. *DataProt.* October 27, 2020. <https://dataprot.net/statistics/credit-card-fraud-statistics/>.
2. Vesta. “IEEE-CIS Fraud Detection”. September 24, 2019. *Kaggle.* October 27, 2020. <https://www.kaggle.com/c/ieee-fraud-detection>.
3. Cover Photo by Kay on Unsplash <https://unsplash.com/photos/PbZ79P_M4IA>.

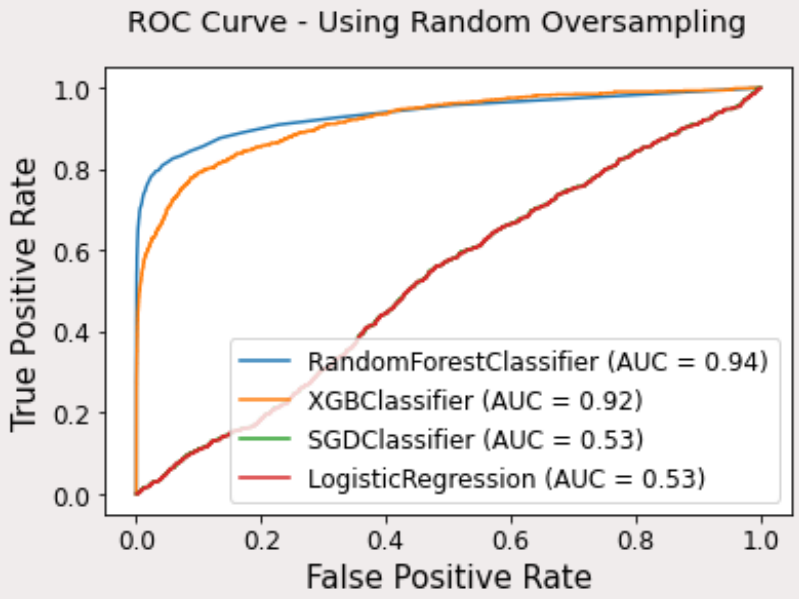
APPENDIX: CALCULATIONS

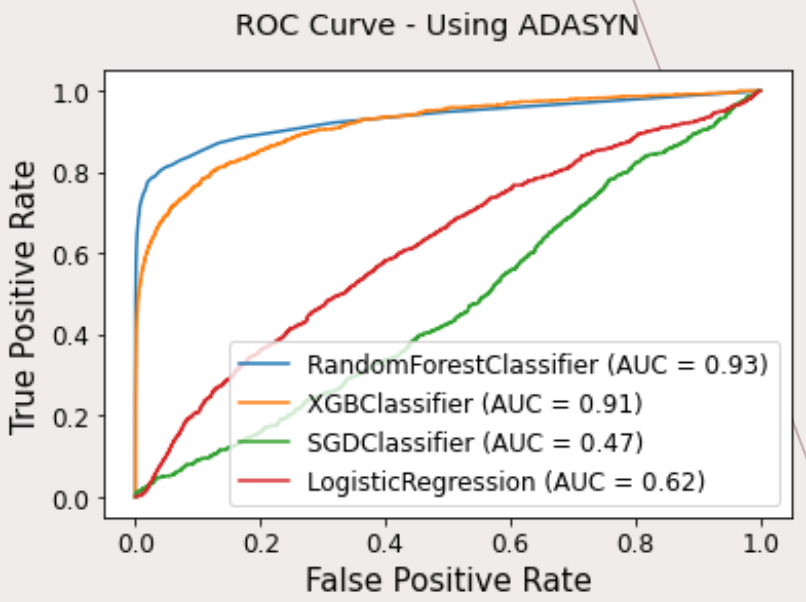
Calculation for the last slide:

$$\frac{295{,}000{,}000\ mc\ transactions}{day} \times \frac{\$227{,}573.27\ saved}{56{,}766\ transactions} = \$1{,}182{,}470{,}000\ saved/day$$

APPENDIX: OVERSAMPLING METHODS







APPENDIX: SQL QUERY

```
mastercard_full = pd.read_sql(
        """SELECT * FROM train_trans
        LEFT JOIN train_ident ON train_trans."TransactionID" = train_ident."TransactionID"
        WHERE train_trans.card4='mastercard'""",
    con=engine)
```