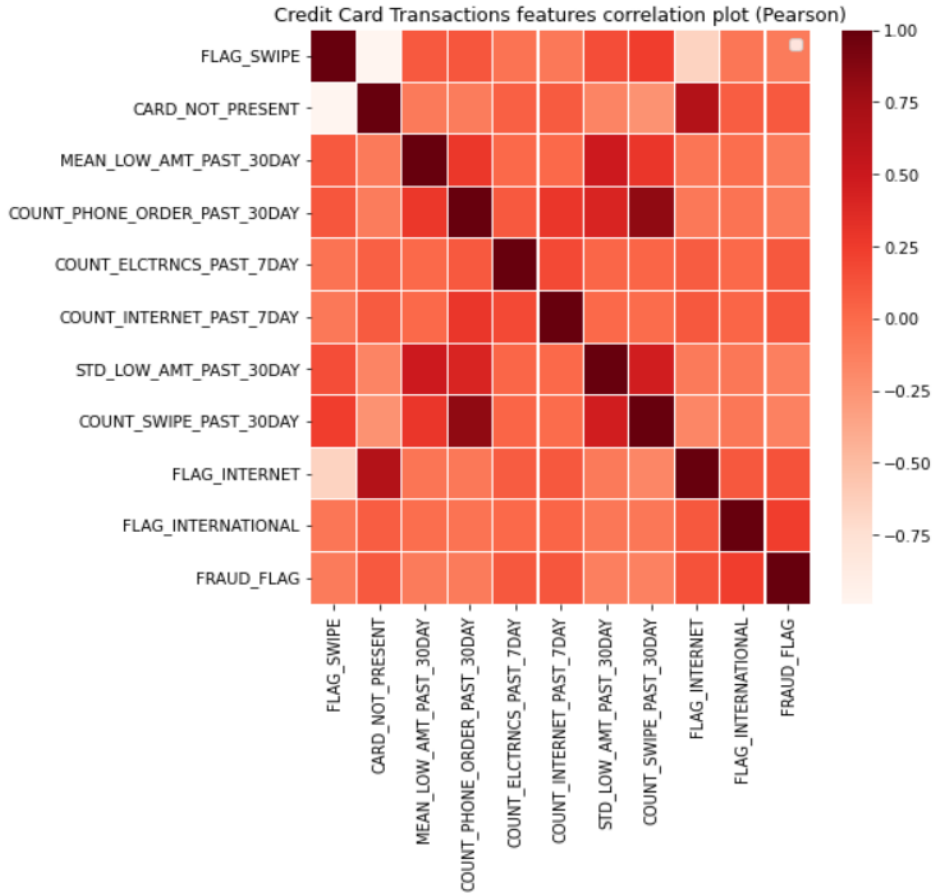
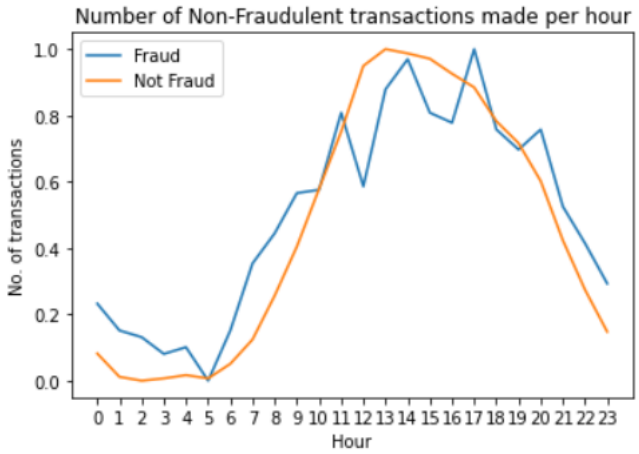
0To∞

Alex, Sarah, Joe, Rithika, Austing

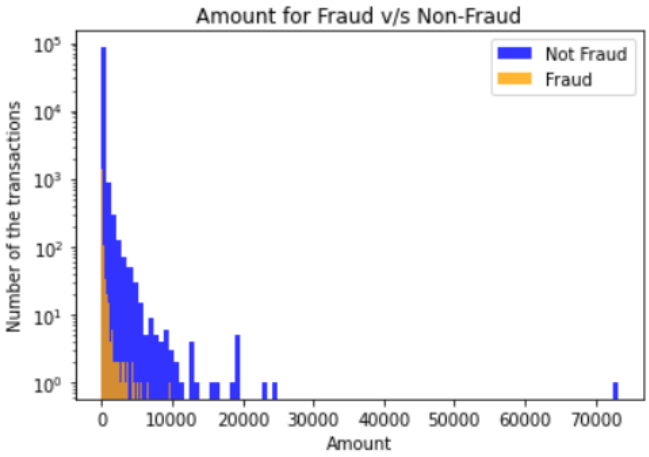
The Data



Correlation between the ten features with highest correlation magnitude fraud. Highest are FLAG_INTERNATIONAL and FLAG_INTERNET.



Distribution over time of day for transactions.



Transaction amount histogram for transactions.

Our Model

We used XGBoost–an implementation of gradient-boosted-decision trees. The data is imbalanced - (2.4% of transactions are fraudulent), causing baseline XGBoost F1 to suffer due to low recall.

We try two approaches to improve recall. First, SMOTE applies data augmentation to the minority class, balancing the number of examples in each class. Next, we use Thresholding to search for the optimal prediction threshold to maximize F1 score on a validation set.

We found that thresholding works better and combining them does not help.

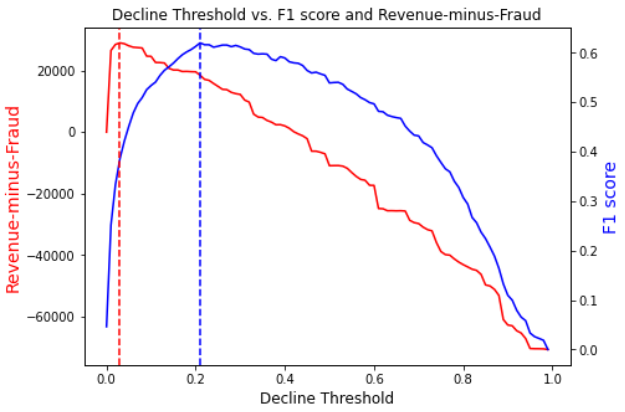
Further tuning improves F1.

Model	F1	P	R
XGBoost Baseline	0.43	**0.75**	0.30
SMOTE (1:1 ratio)	0.49	0.58	0.43
Optimal Threshold	0.58	0.55	0.60
+ Tuning	**0.62**	0.56	**0.69**

What transactions should we decline?

We consider a simplified model of revenue, approving a legitimate translation (+2% of transaction cash value) and a fraud transaction (equaling -100% of transaction cash value). We then evaluate models by revenue, considering the threshold that maximizes F1. Upweighting loss of error on fraud examples during training improves revenue.

Model	Revenue ($)	F1
F1-OPT Threshold	18,507	**0.62**
Rev-OPT Threshold	29,047	0.39
+ Upweight Pos. Training	**30,992**	0.34



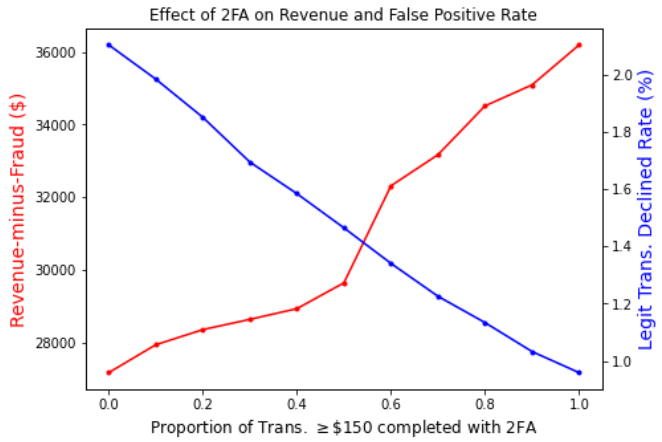
How to minimize the negative user experience while preventing fraud?

When F1 is fully optimized, we pay $9.08 for each false positive reduction. Being conservative only for transactions ≥ $150 (Balanced) buys 974 FP reductions @ $1.92 each. Note: our revenue model doesn't account for second order effects of declining legitimate transactions (customer churn).

Model	Revenue ($)	# of Legit Transactions Declined	Decline Rate of Legit Transactions
F1-OPT	18,507	**282**	**1.29%**
Revenue-OPT	**29,047**	1432	6.57%
Balanced	27,178	458	2.10%

Two-factor authentication: We should let customers submit their own preferences for personal fraud tolerance. Increased adoption of 2FA increases revenue and reduces false positives.

Privacy concerns for fraud alerts: Location, dates, and times are sensitive information to be given–we should offer general warnings instead.

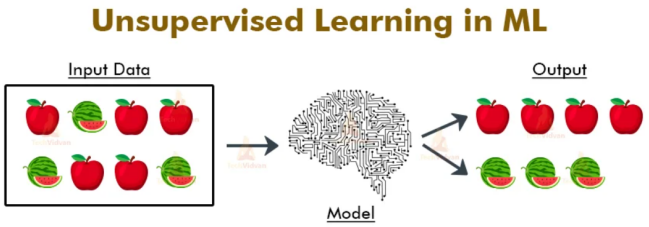


How to prevent fraud more effectively without producing operation overhead?

AI & ML: Unsupervised Machine Learning (UML) can detect unknown fraud patterns at the time of account registration, around 30 days earlier than expected.

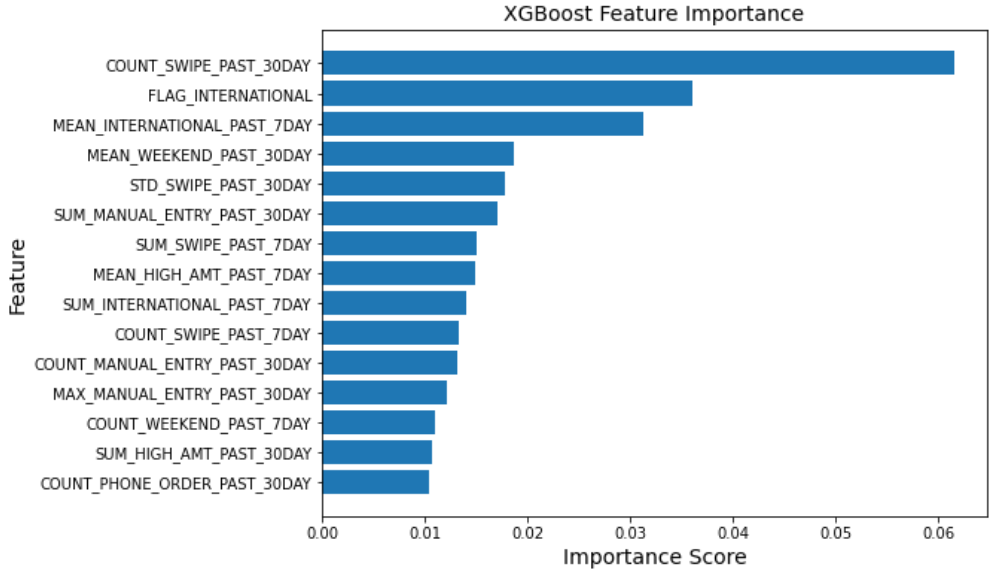
Accelerated analysis: Automation & bulk decisioning reduce operation overhead by 40% by uncovering clusters of fraudulent activity within the same fraud ring.

(Source: SDC Executive https://www.sdcexec.com/transportation/article/21196540/4-ways-to-fight-shipping-fraud-while-reducing-operational-overhead).



(Image source: TechVidvan https://techvidvan.com/tutorials/unsupervised-learning/)

What are some key attributes that help to make the decline decision?



Our model is skeptical of international transactions, and places emphasis on the typical means of using the card (history of swipe/phone order/manual entry transactions).

Amount	Decline Rate (%)
< $150	13.9%
≥ $150	2.3%

Transaction amount is a key attribute for revenue-optimal decision making.

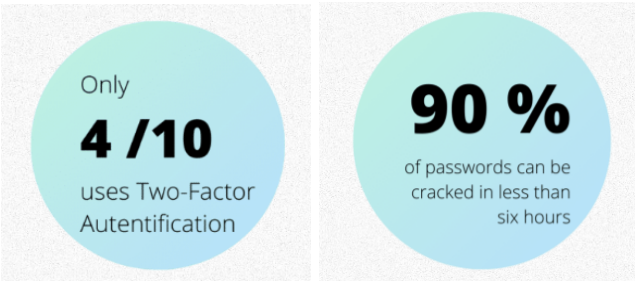
Can you make any long-term suggestions?

Conservative models with worse F1 are better at maximizing revenue.

2FA is promising and can help revenue without hurting customer experience.

International transactions are suspicious. Look into more preventative action here.

The F1 score of 0.62 is achieved when the train-test split is random. When we trained the model on data from consecutive months and tested on a holdout set of the next month, the F1 score dropped to 0.5. The model must be continuously tuned to account for distribution shifts.



(Image source: AdminControl https://blog.admincontrol.com/en/why-is-two-factor-authentication-2fa-so-important)