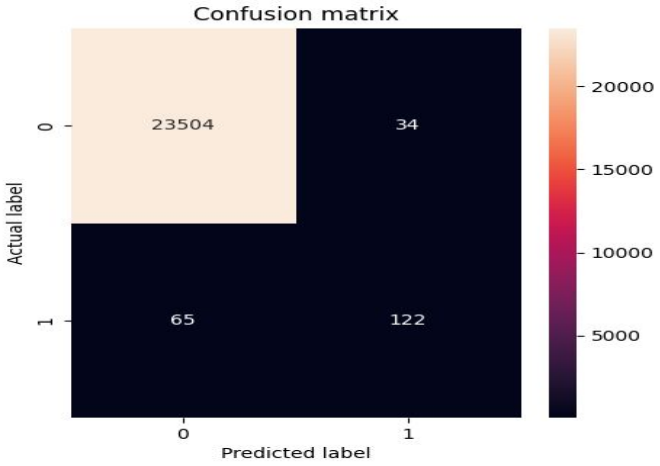
Fraud detection in bank transactions

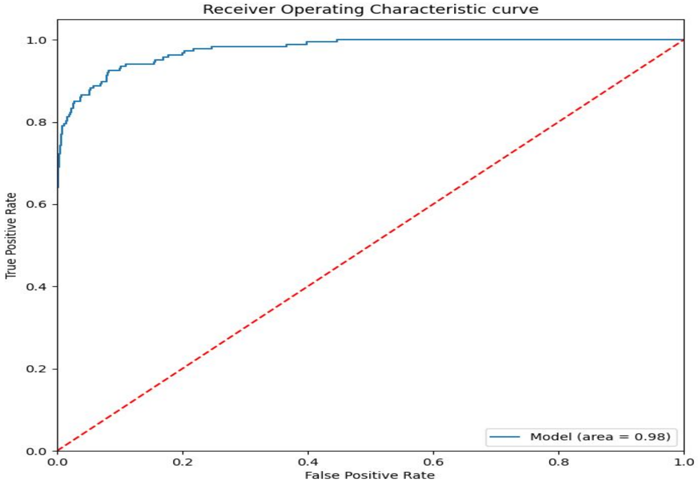
A machine learning model to detect fraud transactions

FEATURE SPACE

Model performance

- The model was trained using 94K rows and tested on 23K. Less than 1% were fraud transactions
- The model achieves an accuracy of 99.58%
- Precision (78%): This suggests that when the model predicts a transaction is fraudulent, it is correct about 78% of the time.
- Recall (65%): This suggests that the model is able to correctly identify 65% of all actual fraudulent transactions.
- FPR (0.27%): model classifies as non fraud the 0.27% of the transactions.
- Model positive rate (0.65%): percentage of fraud transactions classified by the model
- Methodology used: oversampling dataset using SMOTE and training XGBoost with best hyperparameters.



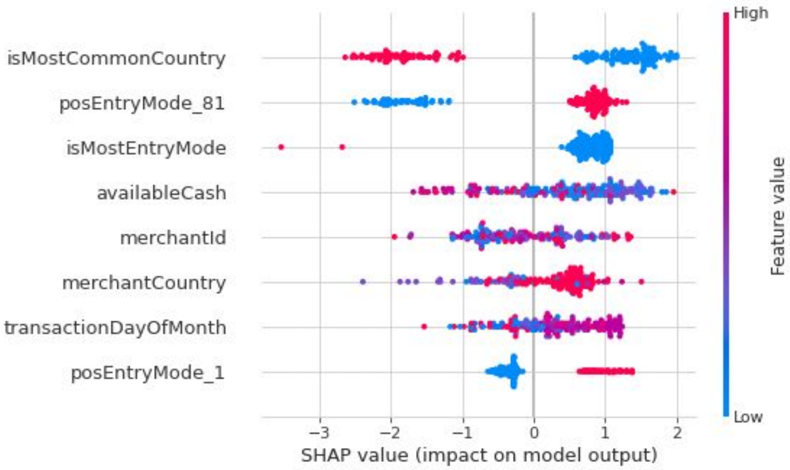


Explainability and insights

- We utilized SHAP values to provide some insight into the model's decision-making process. These SHAP values highlight the impact of each feature on the model's prediction for each individual instance.
- One crucial observation from our analysis is that transactions involving either electronic or manual entry modes are particularly risky.
- Another significant finding from our SHAP analysis is the correlation between high available cash and the likelihood of a transaction being flagged as fraudulent.
- Most notably, **<u>our analysis revealed that transactions originating from the country coded as "826" pose an exceptionally high risk</u>**. As such, any transactions associated with this country should be prioritized for thorough checking.

All models and code applied can be seen in these links:

- https://github.com/hdnh2006/featurespace
- https://wandb.ai/hdnh2006/tree-based-models
- https://wandb.ai/hdnh2006/autoencoder



Without the model vs With our model

The mean of the number of transactions is 9124.

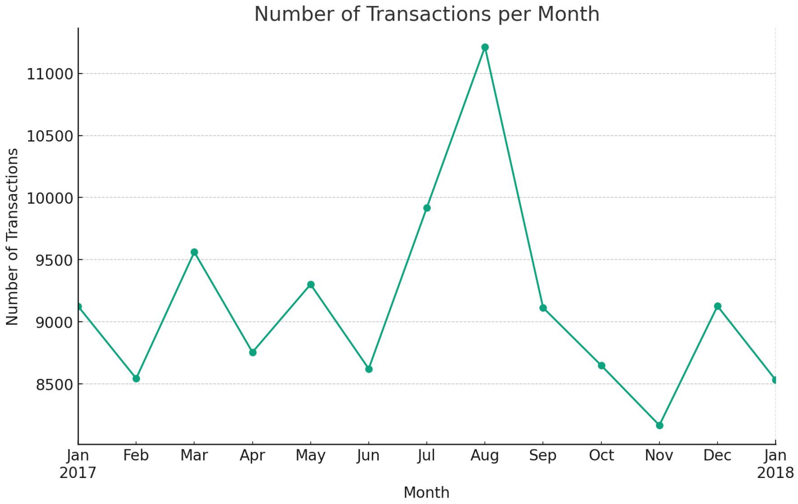
From these transactions, our model classifies the 0.65% as fraud transactions. This equates to an average of 59 transactions per month.

The team has the capacity to review 400 transactions a month, and they will only have to review about 59.

That represents a reduced amount of approximately 85.25% in the volume of transactions that the team needs to manually review each month.

Even in month like July (9919) and August (11213), your team will have the capacity to check these transactions (152 and 172 respectively)

This reduction frees up significant time and resources, allowing the team to focus on other crucial tasks, and thus improving the overall operational efficiency of the fraud detection process.



85% less transactions to check!

FEATURE SPACE

How to improve the model (next steps)

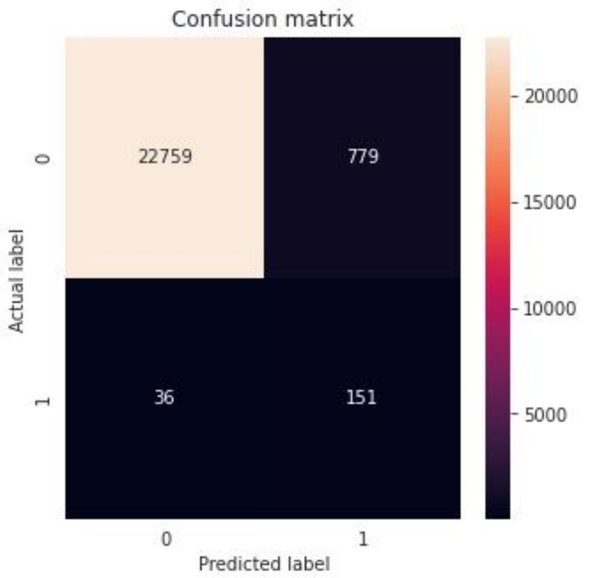
While our model has greatly reduced the amount of manual reviews necessary, there remains a small proportion of transactions, approximately 0.27%, that neither the team nor our model can currently detect as fraudulent.

However, by employing feature selection techniques using SHAP values, we have been able to further minimize this number.

With the implementation of this feature selection, your team will be required to review approximately 357 transactions/month. Despite this, there will still be around 14 transactions/month, representing 0.15% of total transactions, that will not be identified as fraudulent.

Moving forward, we will follow these steps:

- Perform a comprehensive analysis on the characteristics of these 0.15% undetected fraudulent transactions using descriptive statistics.
- Conduct an A/B test on these transactions.



FEATURE SPACE

Thanks for trusting in Featurespace

Featurespace

“making the world a safer place”

FEATURE SPACE