Subject: **Online Payment Fraud Detection using Machine Learning (ML)**

Agenda:

1. Dataset insights
2. Machine learning model performance (supervised classification)

Data: Financial dataset from Kaggle

Target audience: Financial Institutions and Banks

Conclusion: ML model is capable to identify fraudulent transactions with 99% recall.

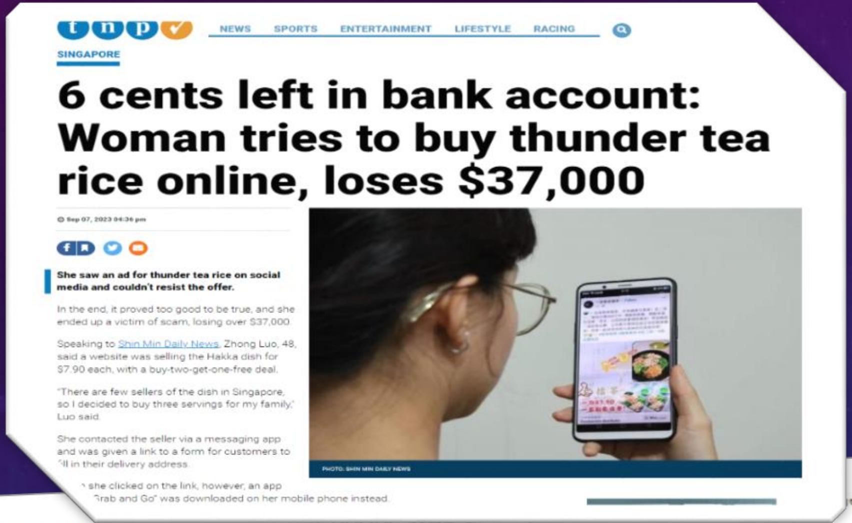
Future work:

1. Overfitting of training data – revisiting feature engineering; e.g. feature selection & obtain domain-specific knowledge.
2. Improve data generalization – collect more latest data and retrain the model regularly to capture latest fraud pattern
3. Anomaly detection using clustering (DBSCAN, K-means) - extract relevant features during data pre-processing & identify previously unknown or evolving fraud patterns.

Prepared by Emma T. (12 Sep 2023)

FRAUD DETECTION USING MACHINE LEARNING

- While incidences of major crimes in Singapore had been decreasing in the last ten years, commercial crimes such as frauds and scams have been on the rise. The increase in such cases has pushed the crime rate in Singapore to a ten-year high. Source: Statista(2022)



NEWS SPORTS ENTERTAINMENT LIFESTYLE RACING

SINGAPORE

6 cents left in bank account: Woman tries to buy thunder tea rice online, loses $37,000

Sep 07, 2023 04:36 pm

She saw an ad for thunder tea rice on social media and couldn't resist the offer.

In the end, it proved too good to be true, and she ended up a victim of scam, losing over $37,000.

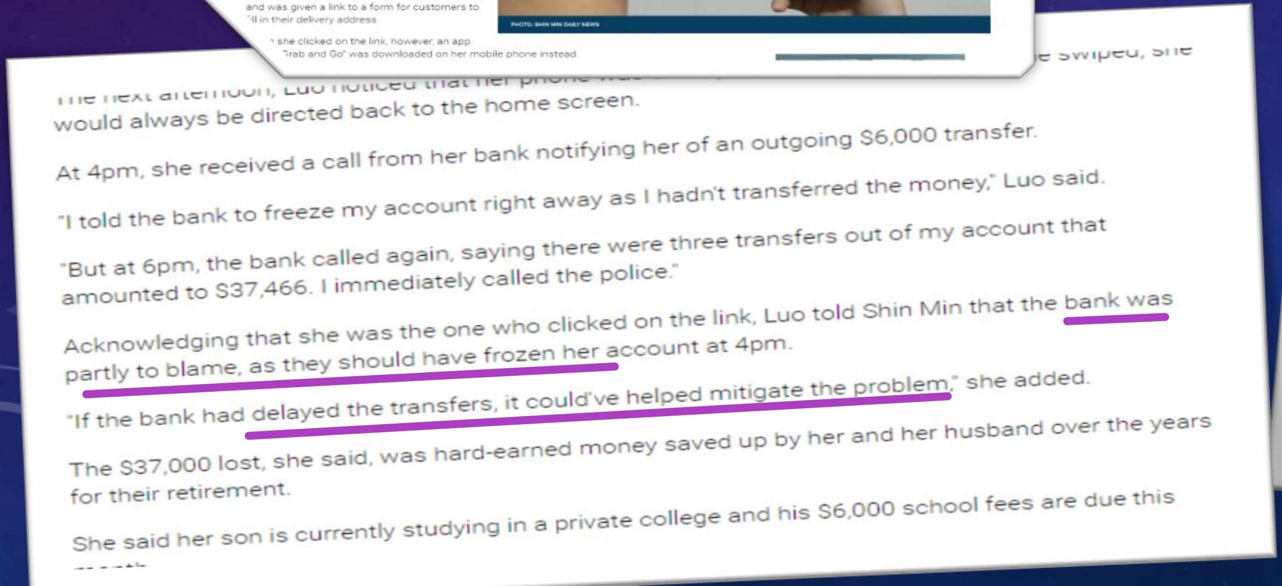
Speaking to Shin Min Daily News, Zhong Luo, 48, said a website was selling the Hakka dish for $7.90 each, with a buy-two-get-one-free deal.

"There are few sellers of the dish in Singapore, so I decided to buy three servings for my family," Luo said.

She contacted the seller via a messaging app and was given a link to a form for customers to fill in their delivery address.

she clicked on the link, however, an app "Grab and Go" was downloaded on her mobile phone instead.

PHOTO: SHIN MIN DAILY NEWS



would always be directed back to the home screen.

At 4pm, she received a call from her bank notifying her of an outgoing $6,000 transfer.

"I told the bank to freeze my account right away as I hadn't transferred the money," Luo said.

"But at 6pm, the bank called again, saying there were three transfers out of my account that amounted to $37,466. I immediately called the police."

Acknowledging that she was the one who clicked on the link, Luo told Shin Min that the bank was partly to blame, as they should have frozen her account at 4pm.

"If the bank had delayed the transfers, it could've helped mitigate the problem," she added.

The $37,000 lost, she said, was hard-earned money saved up by her and her husband over the years for their retirement.

She said her son is currently studying in a private college and his $6,000 school fees are due this

channelnewsasia.com/singapore/mooncake-scams-android-malware-facebook-instagram-3747371

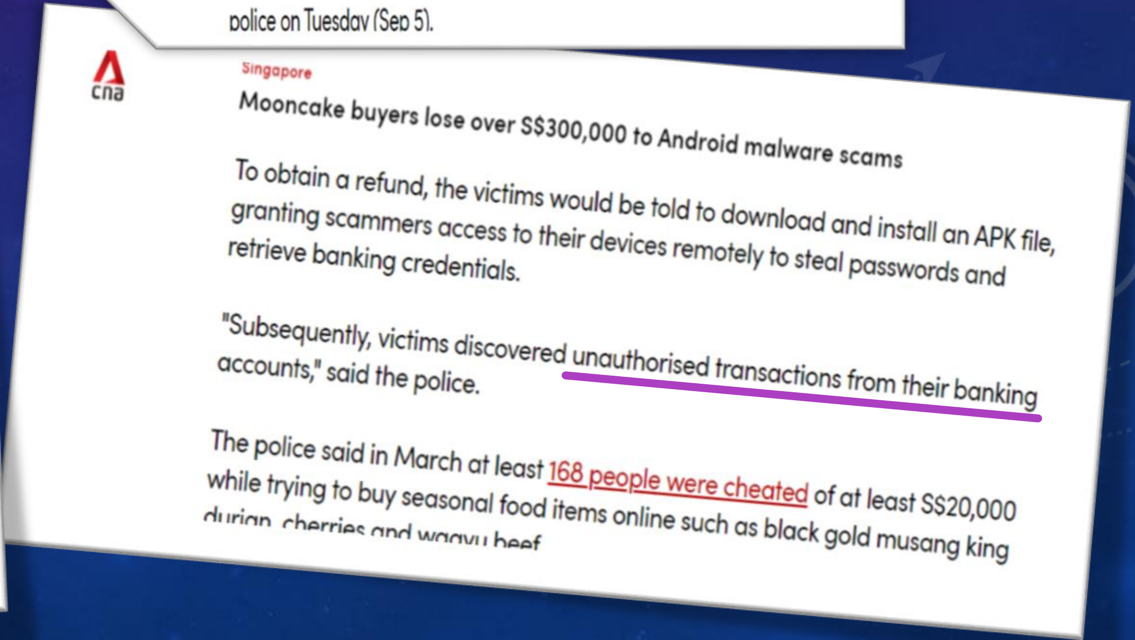
cna

Singapore

Mooncake buyers lose over S$300,000 to Android malware scams

SINGAPORE: At least 27 victims have fallen to mooncake scams perpetrated on social media platforms and involving a malicious Android mobile application.

Total losses in August have amounted to at least S$325,000 (US$238,740), said the police on Tuesday (Sep 5).



cna

Singapore

Mooncake buyers lose over S$300,000 to Android malware scams

To obtain a refund, the victims would be told to download and install an APK file, granting scammers access to their devices remotely to steal passwords and retrieve banking credentials.

"Subsequently, victims discovered unauthorised transactions from their banking accounts," said the police.

The police said in March at least 168 people were cheated of at least S$20,000 while trying to buy seasonal food items online such as black gold musang king durian, cherries and wagyu beef

FRAUD DETECTION USING MACHINE LEARNING– CONT.

- Crucial for banking institution to enhance its fraud prevention strategy, protect customers, and minimize financial losses.
- Need for more proactive and adaptive detection methods in response to the increasing sophistication of fraudsters
- Enabling banks to identify and prevent fraudulent activities with greater accuracy and speed.

cna

Singapore

S$661 million lost to scams in 2022, with young adults most likely to fall victim: SPF

The total number of scam and cybercrime cases rose by more than a quarter to 33,669 in 2022, compared to 26,886 the year before. Scams accounted for 94.2 pe cent of these cases.

The top five scam types were phishing scams, job scams, e-commerce scams, investment scams and fake friend call scams. They made up more than 80 per cer of the top 10 scam types in Singapore.

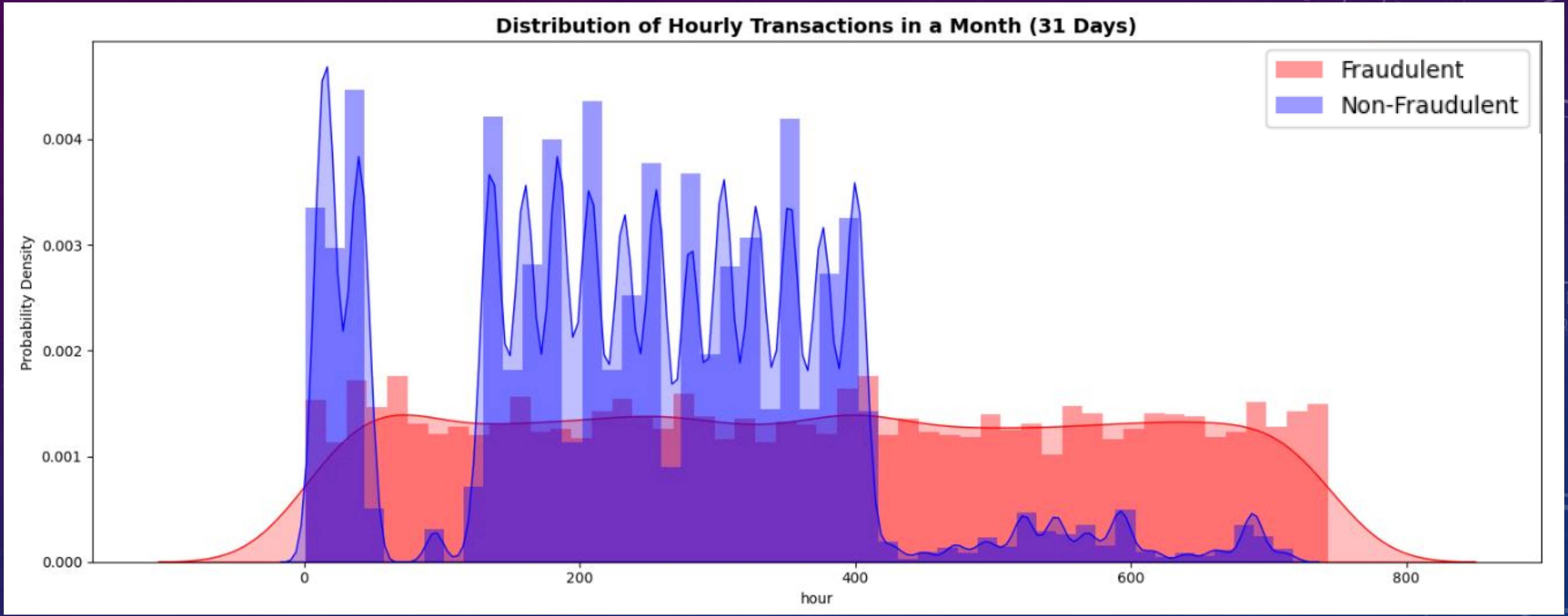
The number of cases of each of these scam types rose across the board.

Number of cases of top five scam types

	2021	2022	Percentage change
Phishing scams	5,023	7,097	+41.3%
Job scams	4,550	6,492	+42.7%
E-commerce scams	2,729	4,762	+74.5%

Prepared by Emma T. (12 Sep 2023)

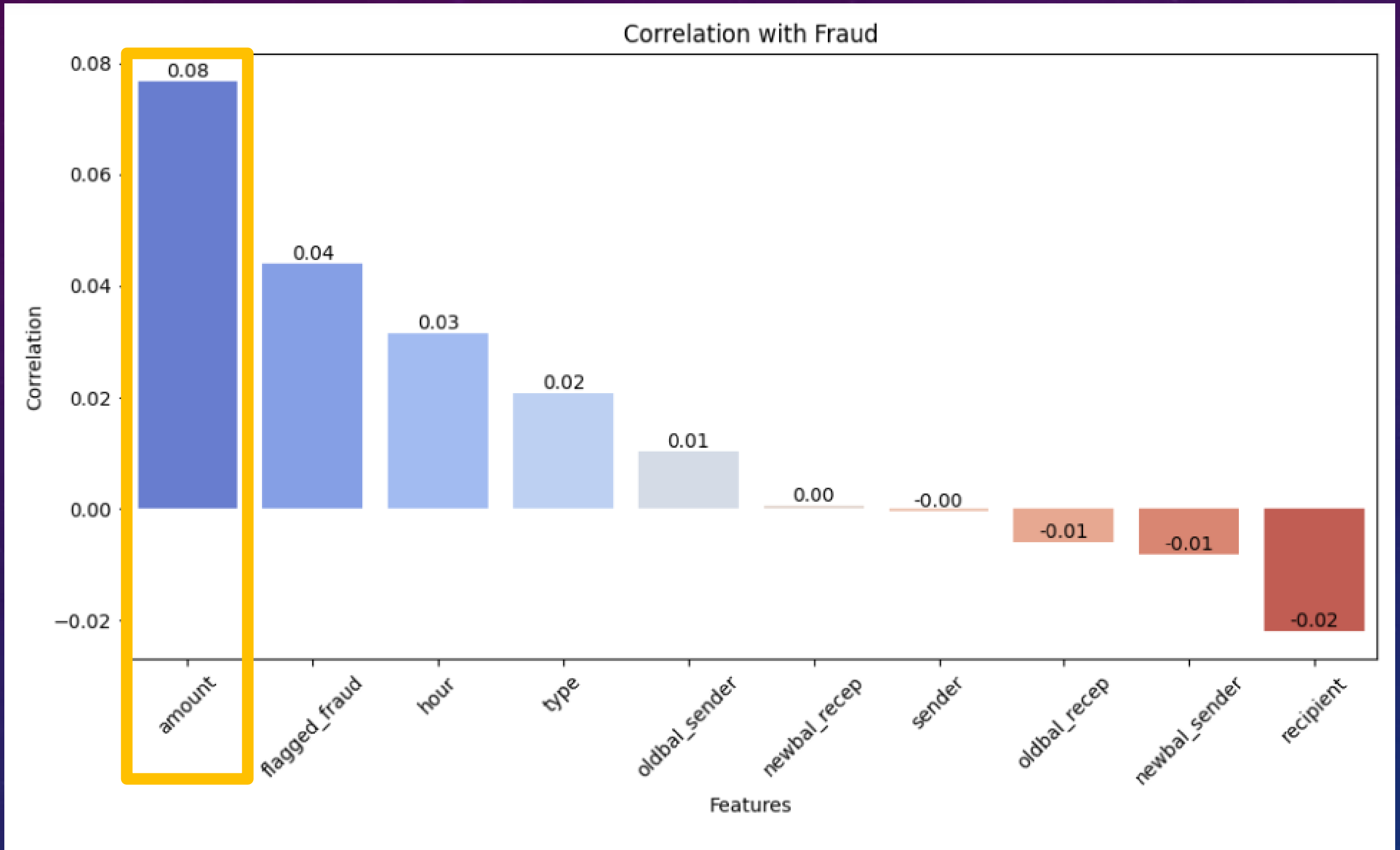
Insights:



- Fraud transaction occur evenly throughout the month.
- Genuine transaction peaks at start and mid of the month.

Prepared by Emma T. (12 Sep 2023)

Insights:

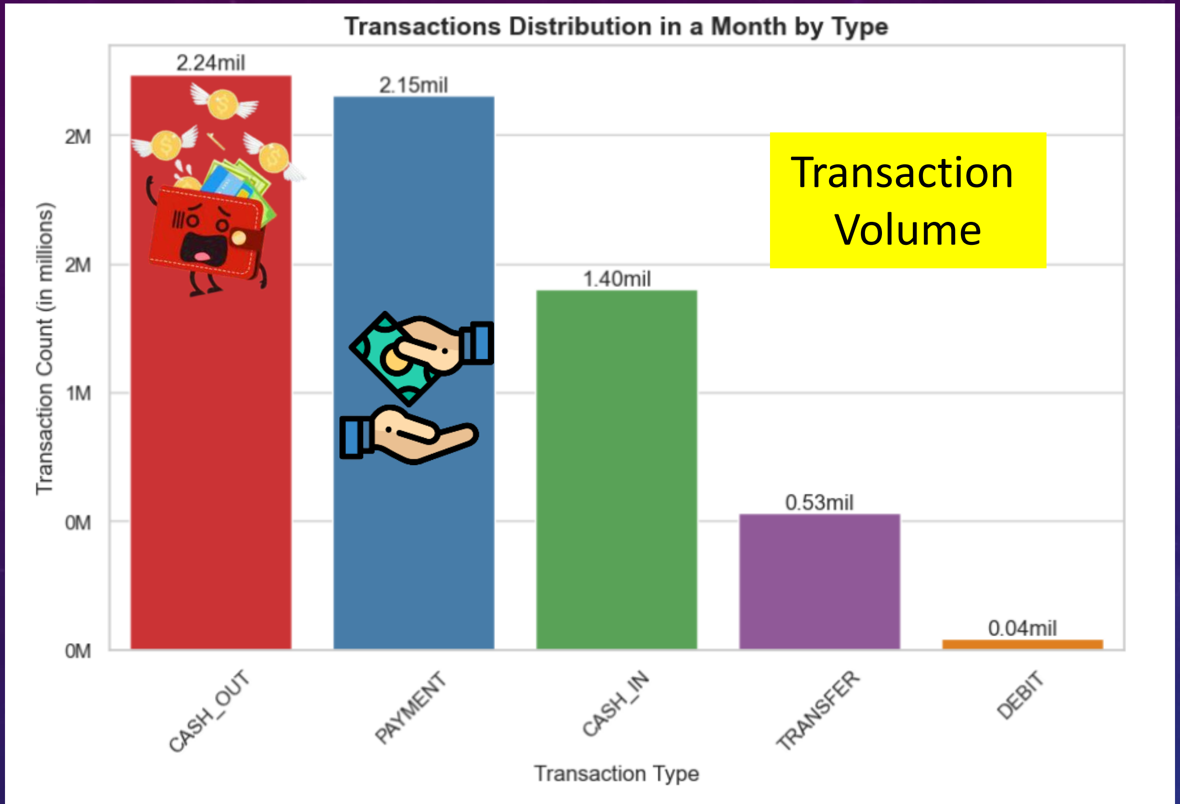


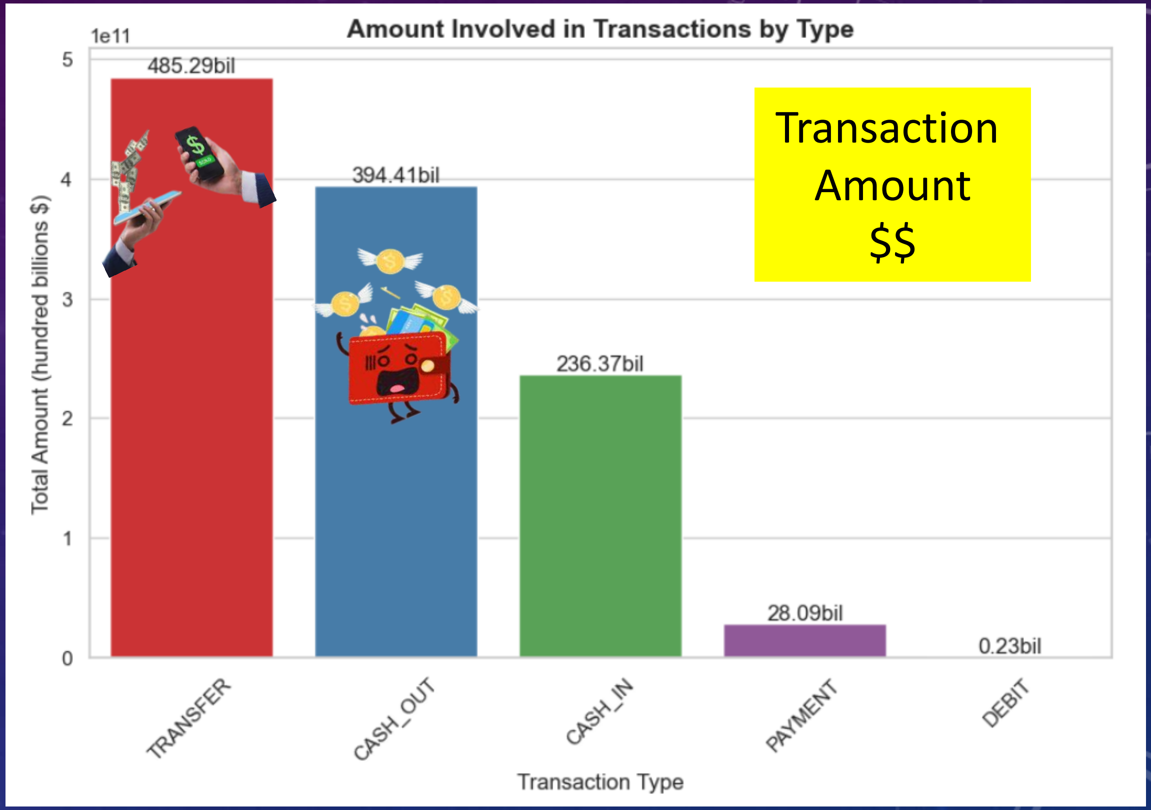
- Overall, data on hand shows weak correlation to fraudulent transaction.
- Transaction amount which have the highest correlation score is only 8% correlated statistically.

Prepared by Emma T. (12 Sep 2023)

FRAUD DETECTION USING MACHINE LEARNING– CONT.

Insights:



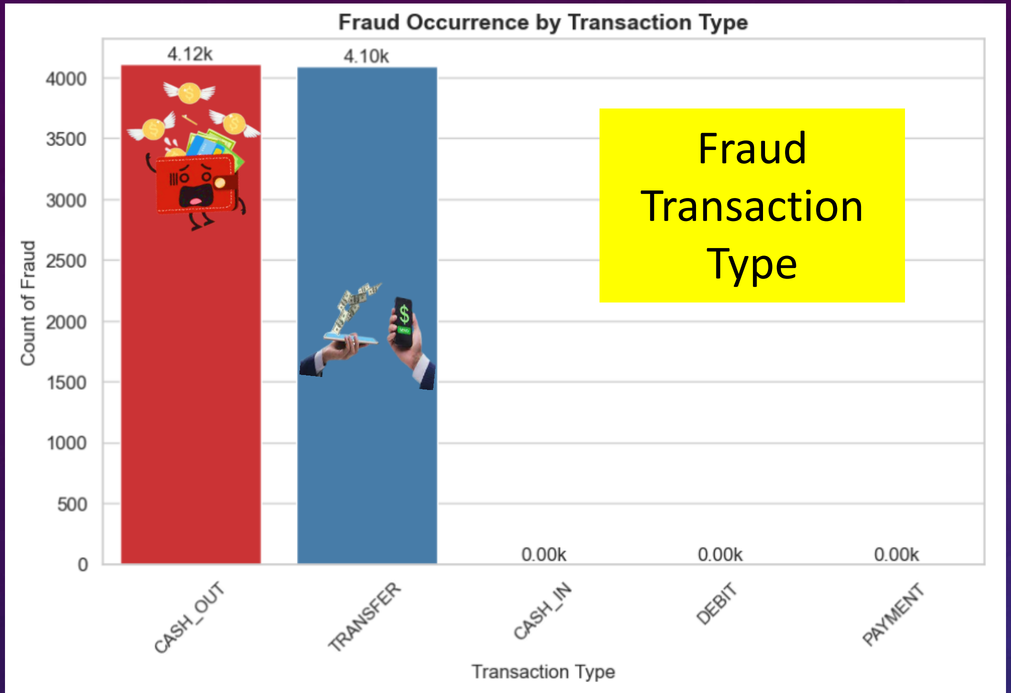


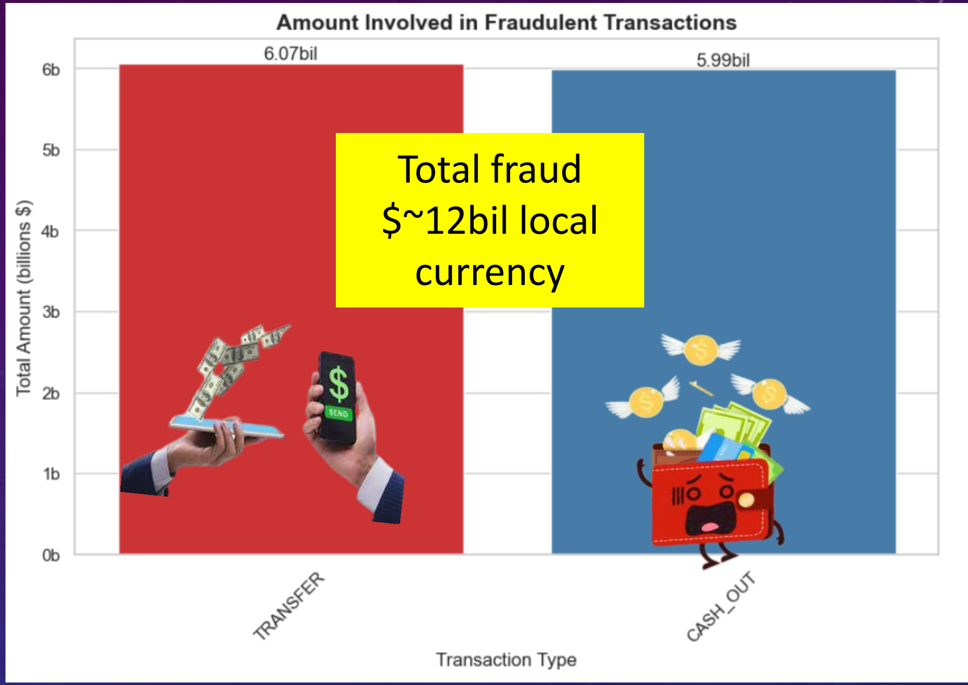
- Top transaction volume:
  1. Cash Out (~2.2 mill counts = 35%)
  2. Payment

- Top $$ involved (~$880 bil = 77% of $1.1tril):
  1. Transfer
  2. Cash out

Prepared by Emma T. (12 Sep 2023)

FRAUD DETECTION USING MACHINE LEARNING– CONT.





Highest fraud transaction type (total 8000 counts):

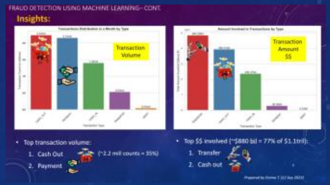
1. Cash Out
2. Transfer

High fraud amount involved (total $12 billions) :

1. Transfer
2. Cash Out

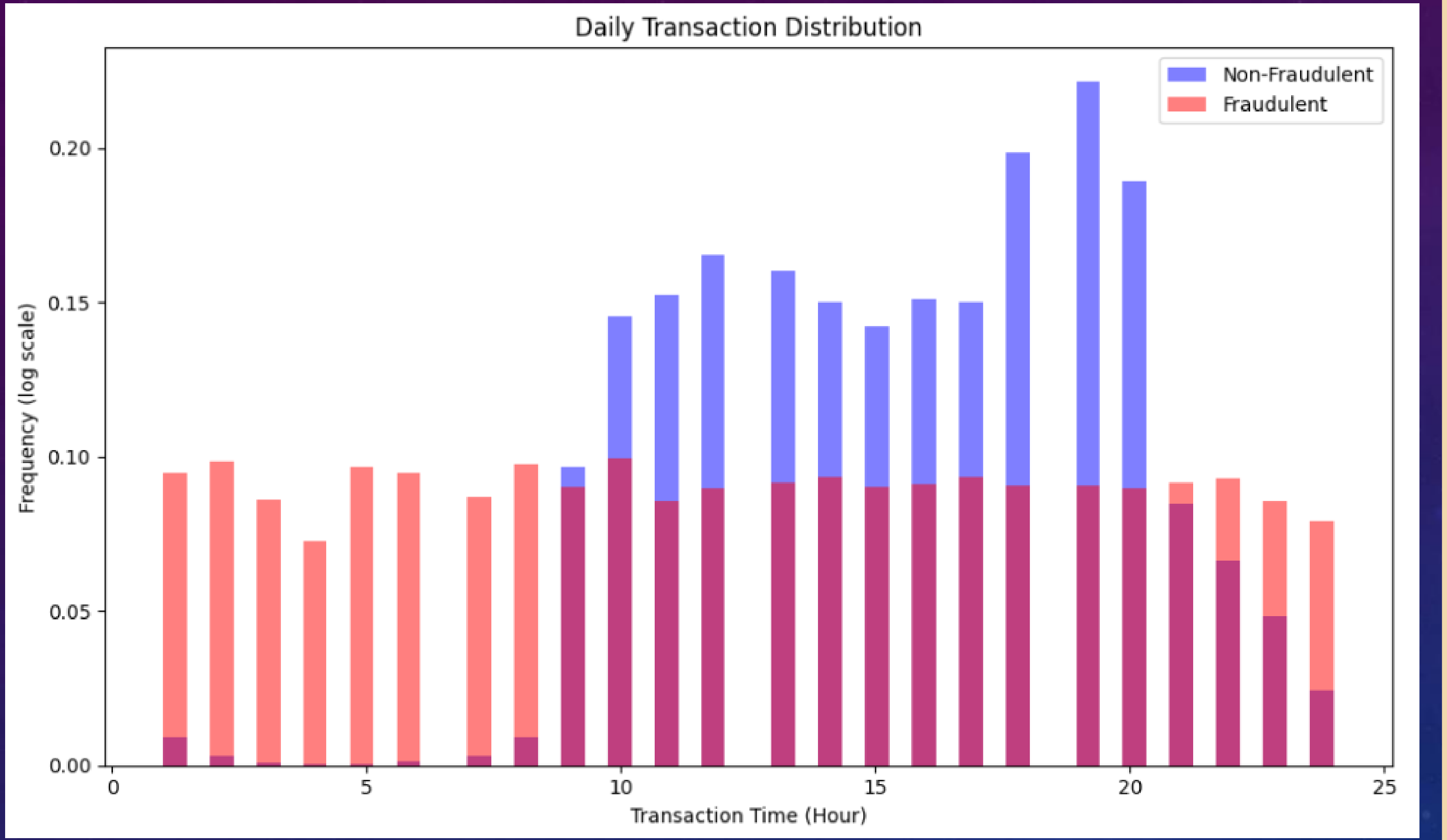
Actionable Insights:

Fraudsters tend to favour cash out and transfer.
Crucial to control fraud (high volume & high amount transacted).
To mitigate risk, we can enhance our verification process for customers for customer using cash out and transfer.



Remember?
Cash Out + Transfer = 2.7mill counts (43%)
Cash Out + Transfer = $880 bil (~77%)

Prepared by Emma T. (12 Sep 2023)



Assumption: 1st hour in dataset starts at 0000hrs.

Actionable Insights:

The probability of genuine transaction is higher during office hour while the probability of fraudulent transaction is higher during non-office hour, peak around pre-dawn.

Implementation of real-time monitoring between 11PM to 8AM is highly recommended and potentially restrict high-value transactions during this time frame.

Prepared by Emma T. (12 Sep 2023)

CURRENT SYSTEM

Transaction Monitoring:
Rule-based systems that flag transactions based on predefined rules (e.g. exceed limits on transaction amounts, frequency, or locations).

+

Manual Review:
Suspicious transactions flagged. Human analysts investigate the flagged transactions to determine if they are fraudulent.

+

Customer Verification:
Contacting the customer directly to verify the legitimacy of a transaction, especially for high-risk or unusual transactions.

"ADVANCE SYSTEM"

Adoption of Machine Learning and AI:

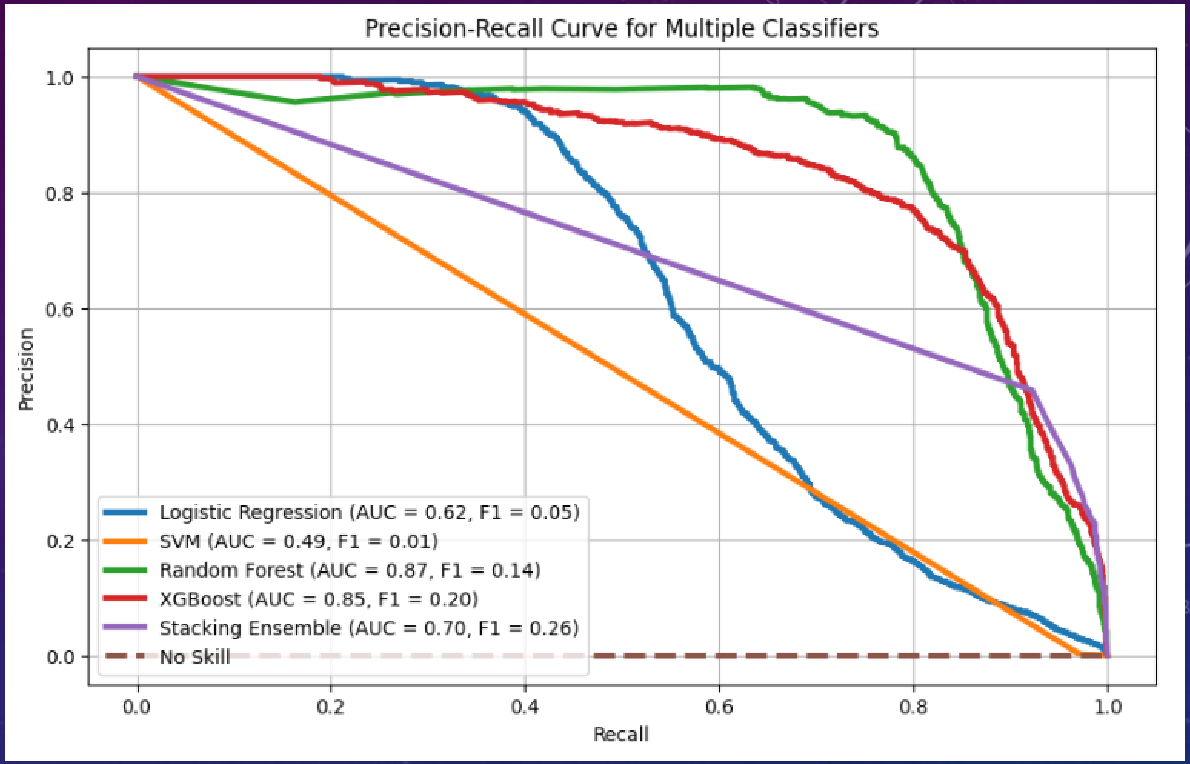
- Leverage ML and AI algorithms
- Analyze vast amounts of data and identify complex patterns & anomalies indicative of fraud in real-time.
- These systems can adapt and improve accuracy over time.

Prepared by Emma T. (12 Sep 2023)

MACHINE LEARNING RESULT

- Dataset is **severely imbalance**.
  - 6mill unique entry. **Fraudulent 1%**
  - Challenge: All models trained have high False Positive (actually genuine but predicted fraudulent) = low Precision score for fraudulent

- Model is trained using **4 algorithms (classifiers)**.
  1. Logistic Regression
  2. Support Vector Machine (SVM)
  3. Random Forest
  4. XGBoost (Gradient Boosting).
  5. Stacking Ensemble (combining prediction from 4 base models - often improve overall performance as it leverages the strengths of different models)



Precision-Recall Curve Interpretation Guideline:
1. **Ideal Precision-Recall curve** is one that starts at (0, 1) and goes to (1, 1), meaning perfect precision and recall, a curve that is as close to the top-right corner as possible.
2. **Precision:** "Out of all the cases the model predicted as positive, how many were truly positive?“
3. **Recall (Sensitivity):** "Out of all the actual positive cases, how many did the model correctly identify?“ (actual positive)

- Model performance visualisation using Precision-Recall Plot.
  - More informative and give an accurate prediction of future classification performance.
  - The plot evaluate the fraction of true positives (fraudulent) among positive predictions (predicted fraudulent).

Prepared by Emma T. (12 Sep 2023)

WHICH MODEL TO USE?

Classifier	Precision	Recall	F1-Score	AUC
Logistic Regression	0.0086	0.9937	0.0488	0.6199
SVM	0.2521	0.7355	0.0072	0.4874
Random Forest	0.0577	0.9924	0.1403	0.8693
XGBoosting	0.0127	0.9981	0.198	0.8468
Stacking Ensemble	0.0775	0.9955	0.2559	0.6968

1. If **minimizing false positives** (precision) is crucial due to the cost of investigating non-fraudulent transactions, the "**Random Forest**" model may be preferred. **Random Forest (F1 = 0.1403, AUC = 0.8693)**

2. If **maximizing the detection of fraud cases** is paramount, even at the **risk** of **some false alarms**, the "**XGBoosting**" model stands out. **XGBoosting (F1 = 0.198, AUC = 0.8468)**

3. For a **balanced approach** where both precision and recall matter, the "**Stacking Ensemble**" model offers a competitive F1-Score. **Stacking Ensemble (F1 = 0.2559, AUC = 0.6968)**

Ultimately, the choice of the best model depends on the specific goals and constraints of the fraud detection task, including the tolerance for false alarms and the consequences of missing actual fraud cases.

-Thank you! End of Presentation -

Prepared by Emma T. (12 Sep 2023)