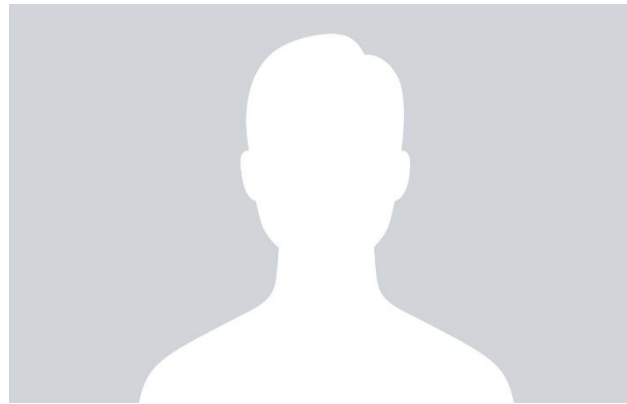


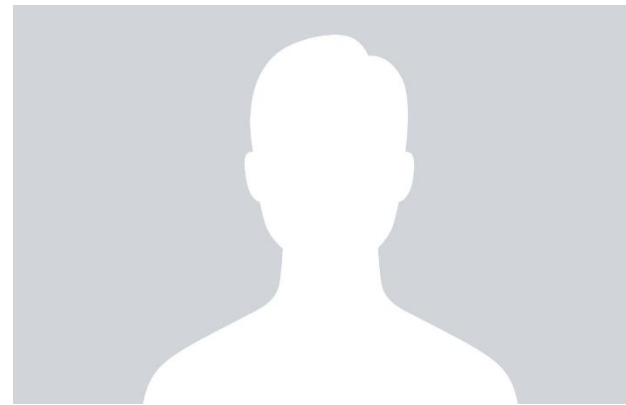


Fraud Detection

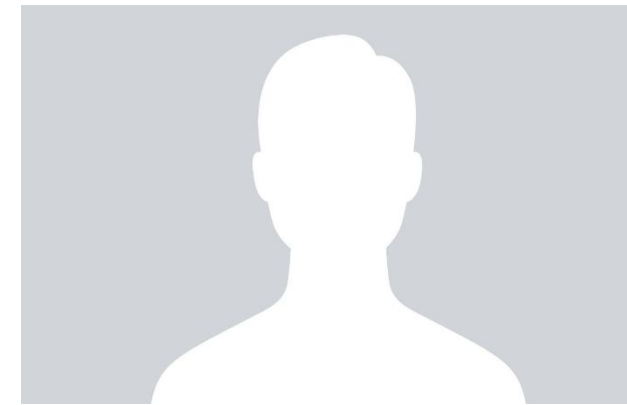
Meet the Team



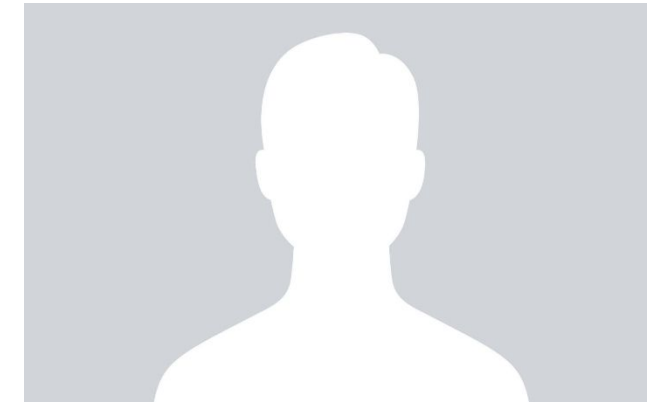
Ahmad Lutfi



Bagus Prasetyo



Achmad Muchlis

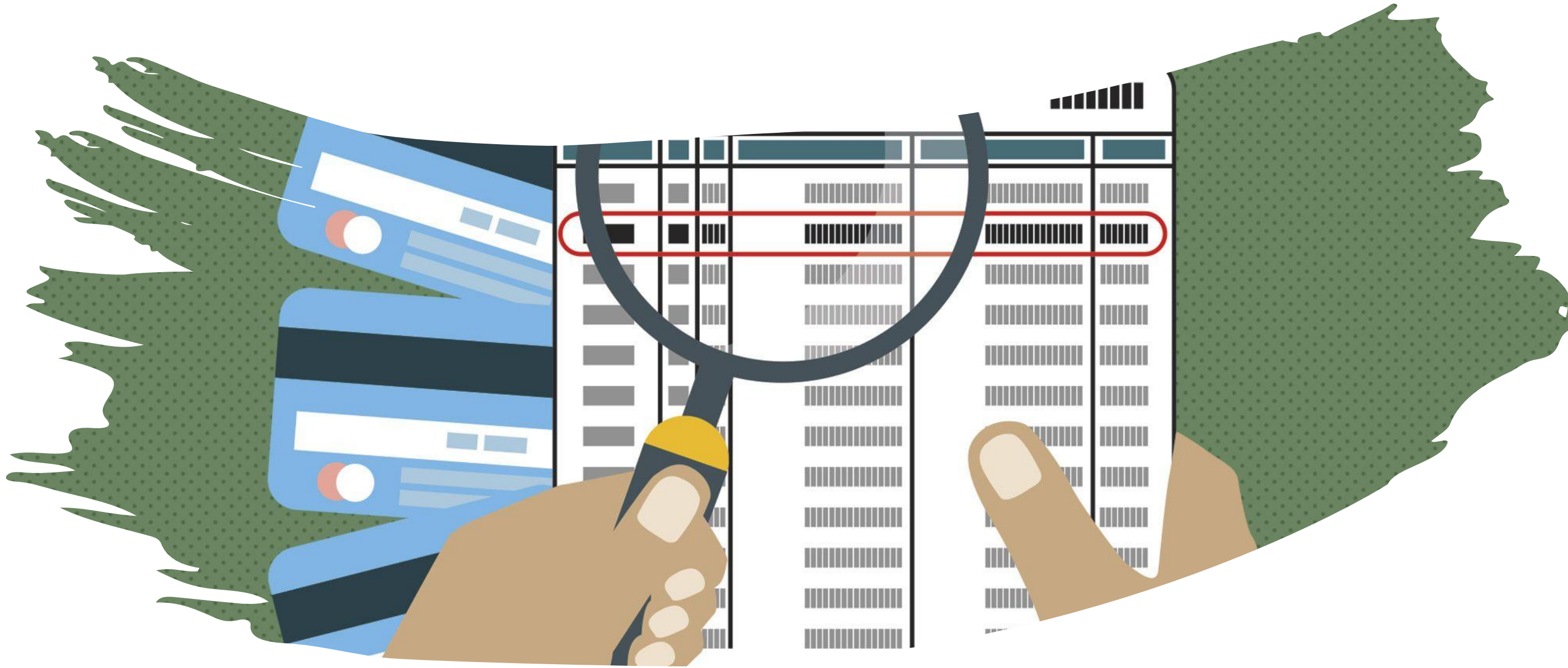


Gerry William

Background

- Menurut studi Report to The Nations yang dilakukan pada tahun 2020 oleh Association of Certified Fraud Examiners, dari 16 negara Asia Pasifik yang diteliti, Indonesia menjadi penyumbang terbanyak fraud/penipuan
- Kasus Lembaga Pembiayaan Ekspor Indonesia (LPEI). Terdapat empat debitur yang melakukan fraud dengan nilai mencapai Rp 2,5 triliun. (18/03/2024)

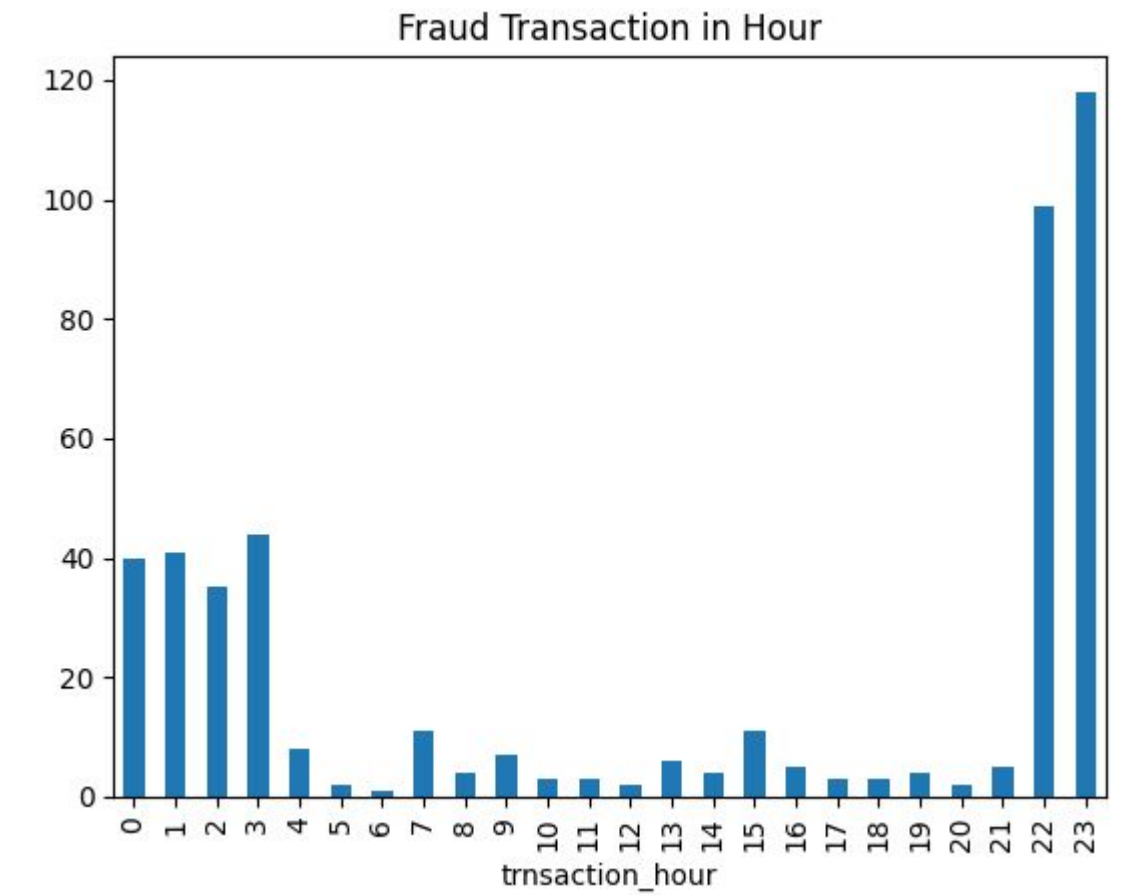
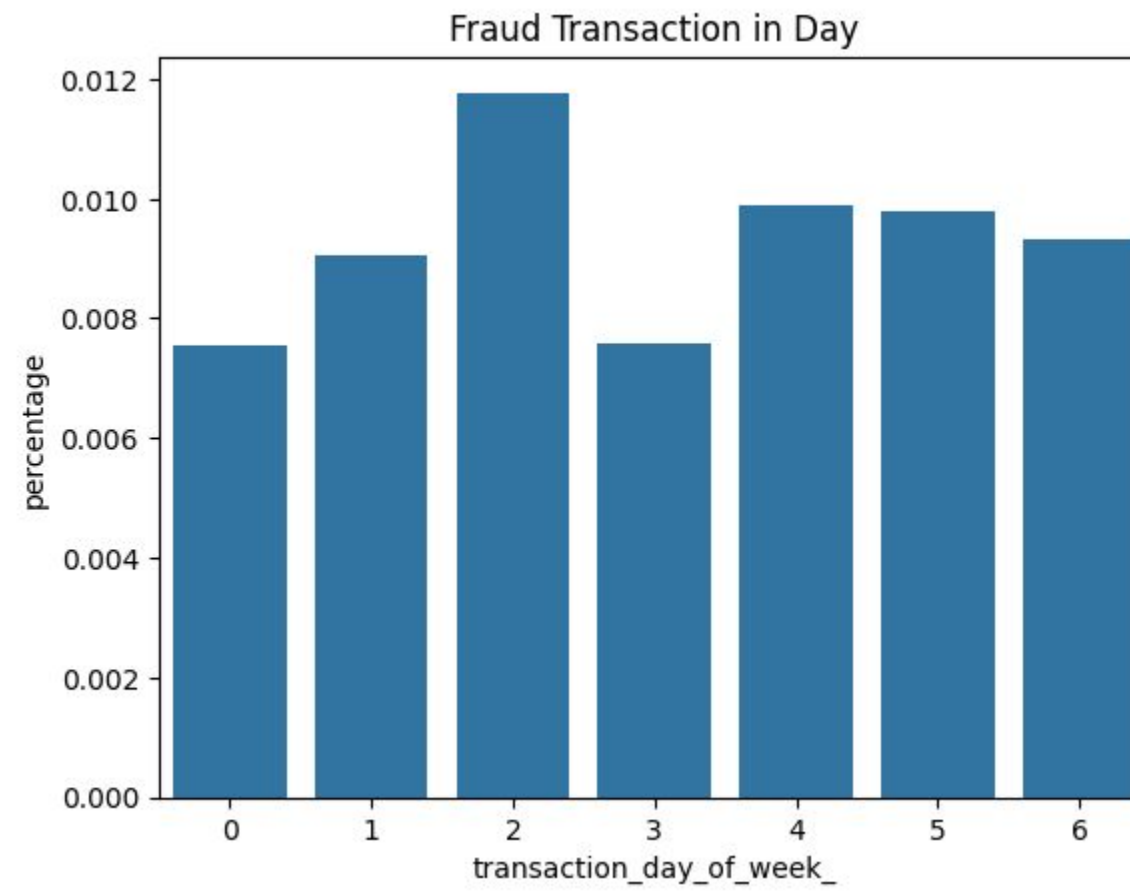
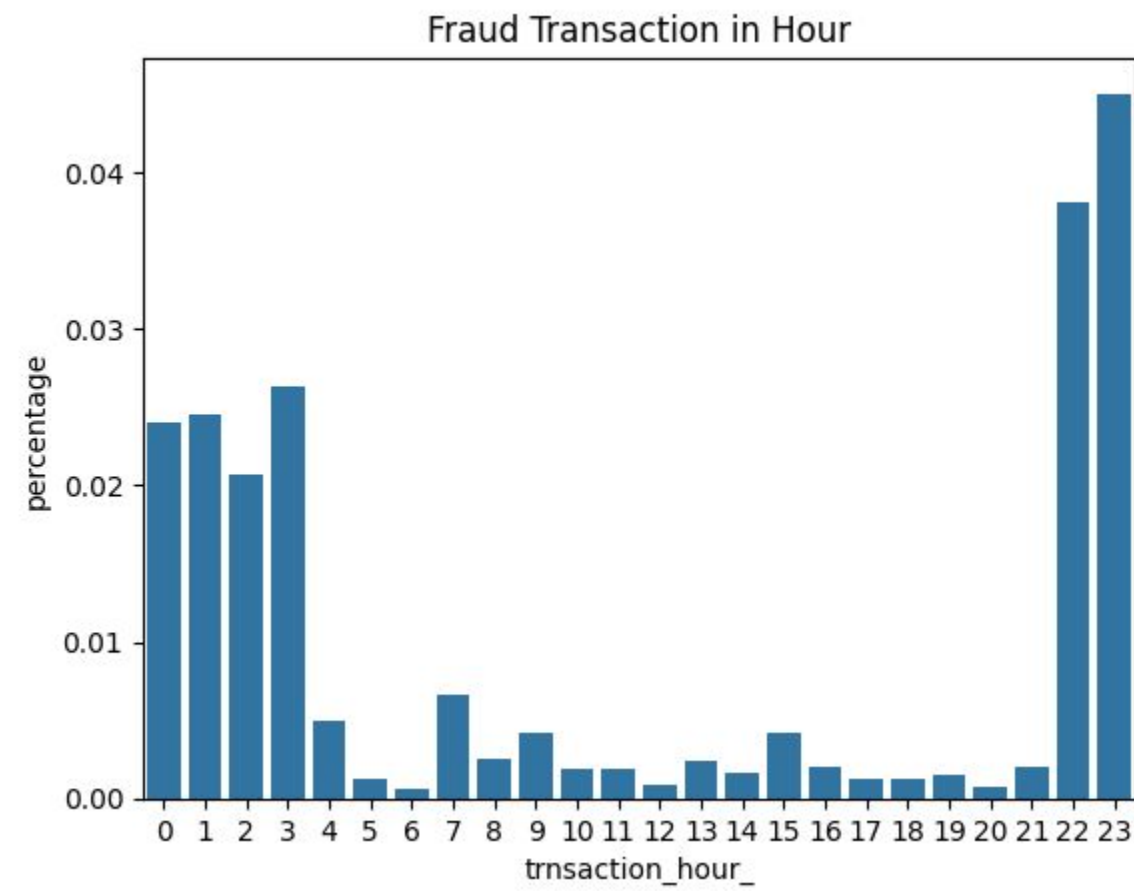




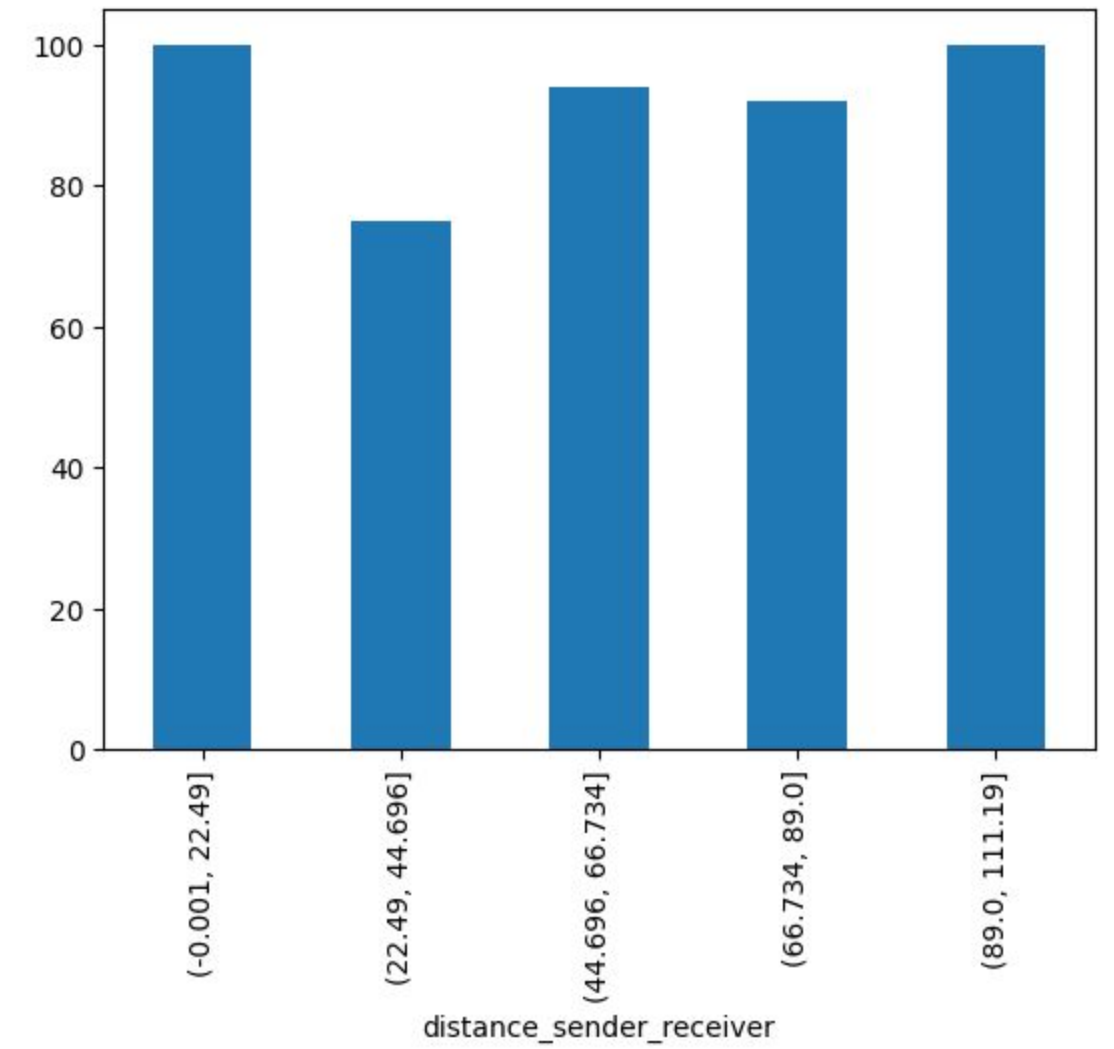
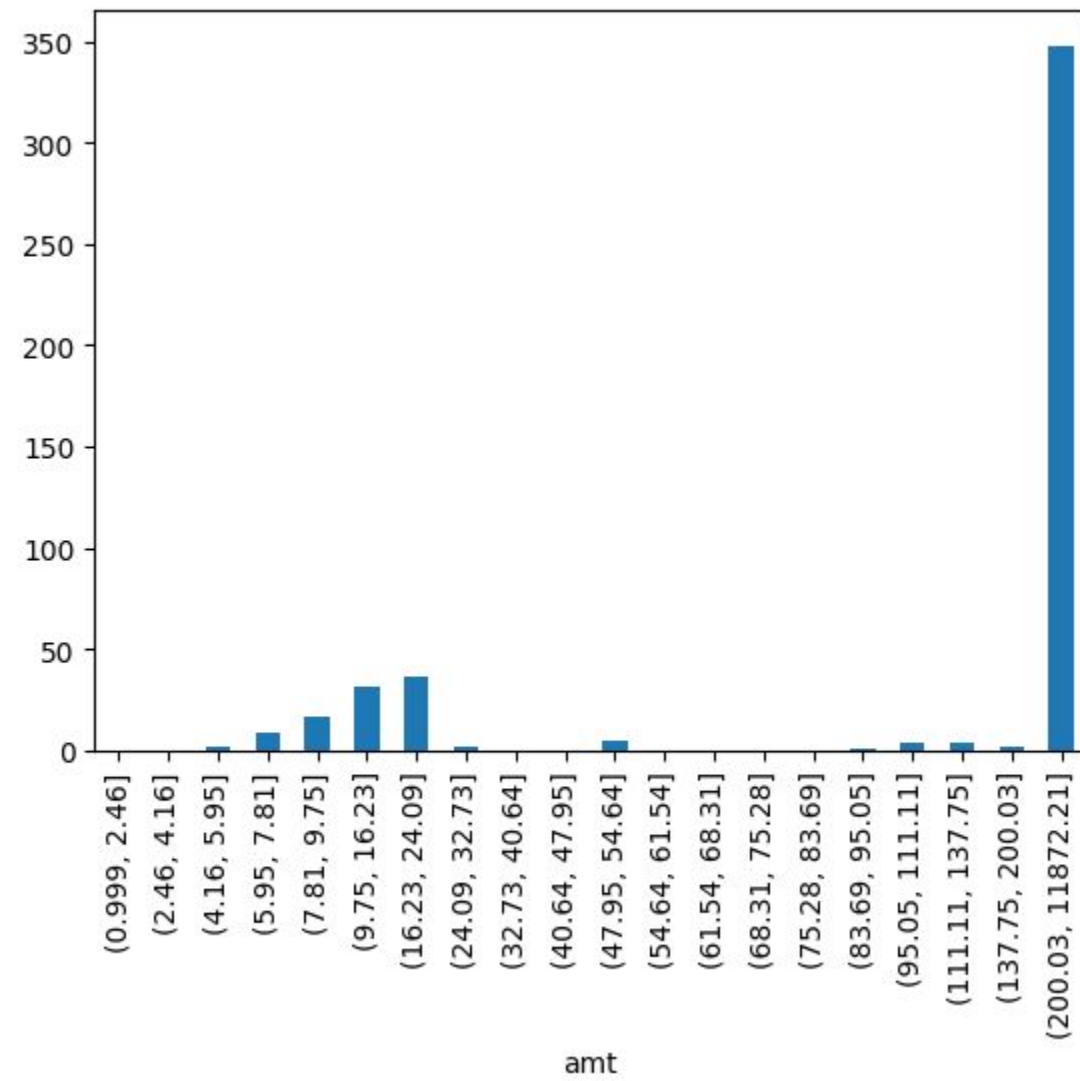
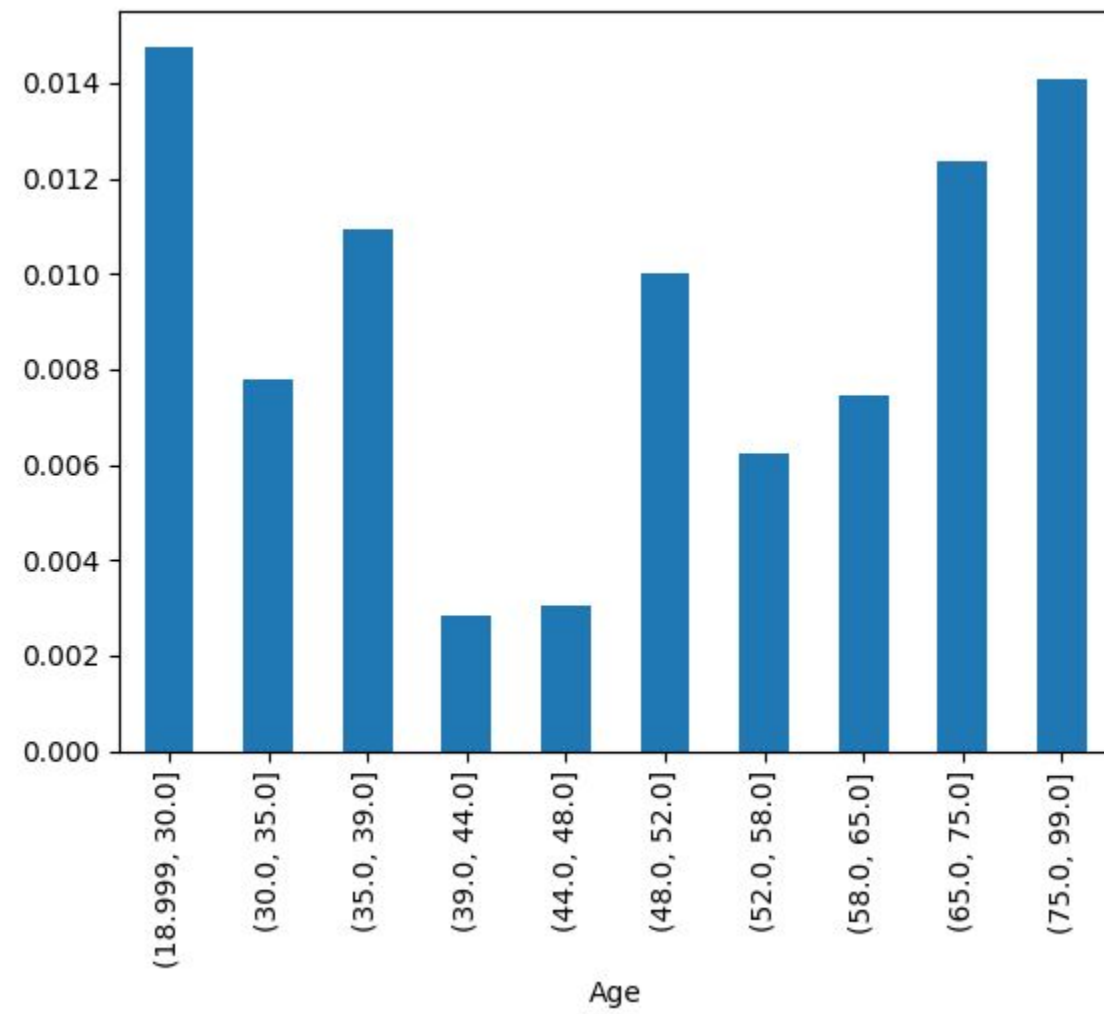
Objective

Mencari transaksi yang mencurigakan dari transaksi banking dengan metode random forest dengan optimalisasi 'hyper tuning'

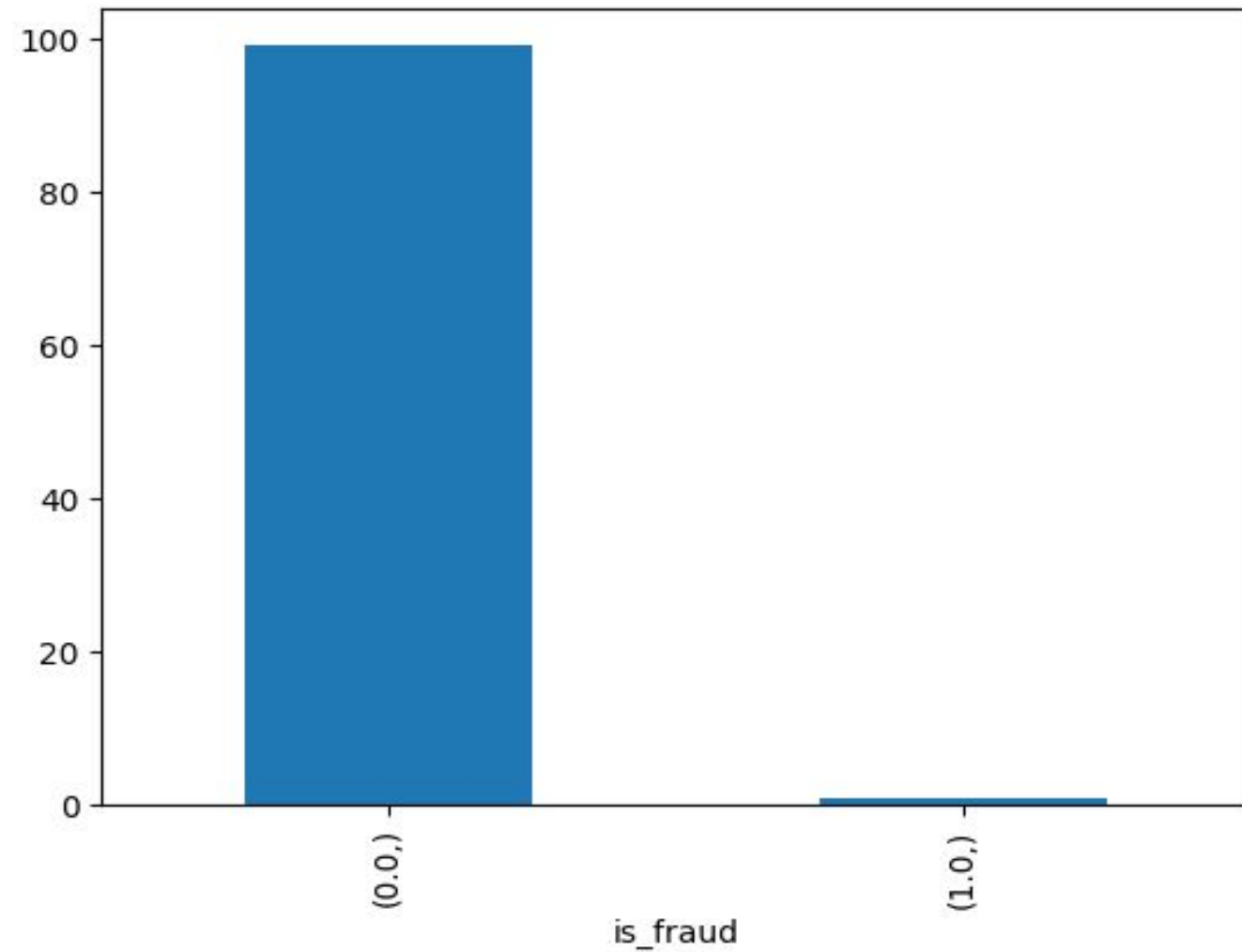
Preparation Data



Preparation Data



Proporsi Target Y



```
is_fraud
0.0      99.086975
1.0       0.911049
Name: count, dtype: float64
```

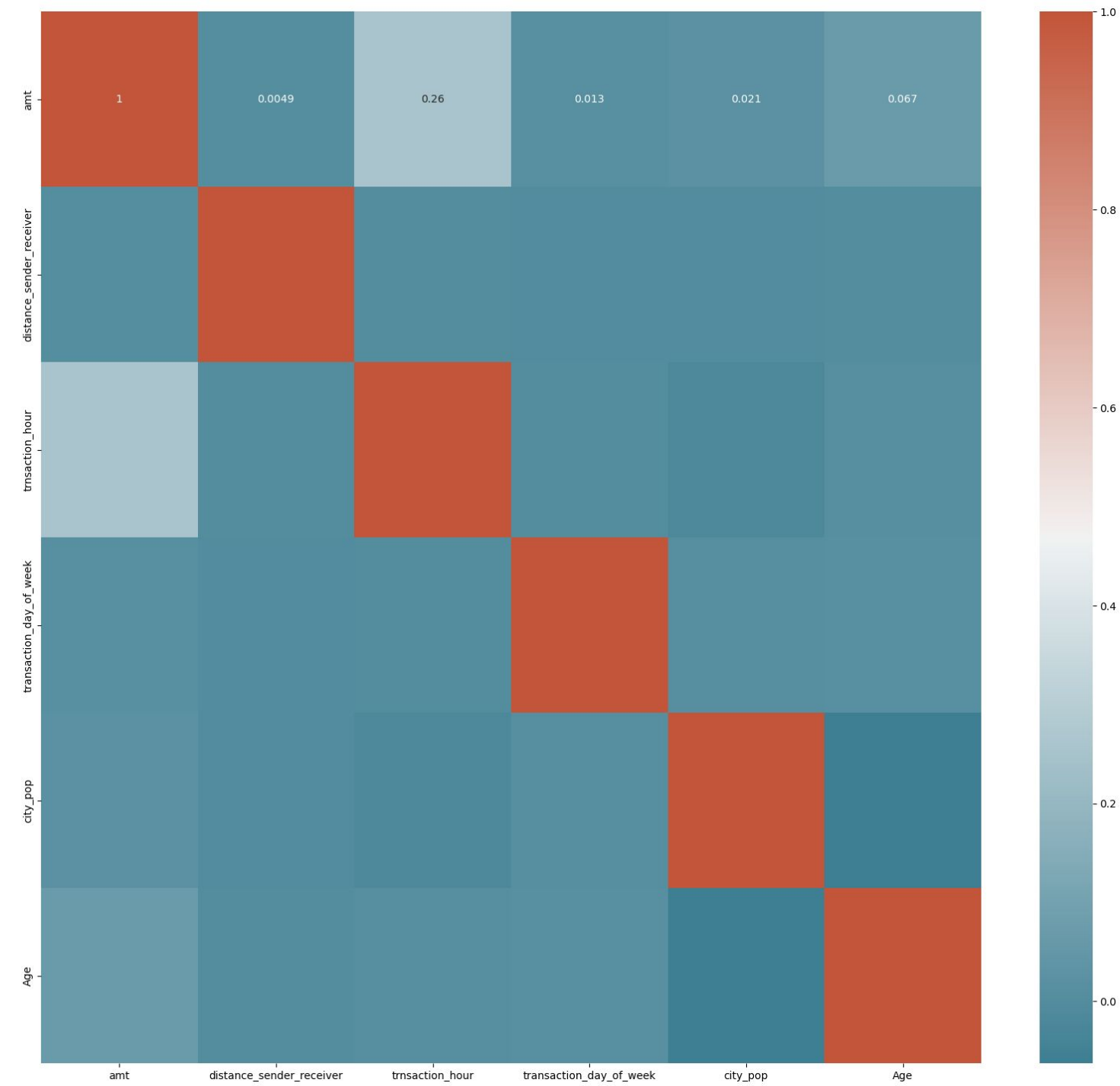
Perhatikan bahwa proporsinya sangat jauh antara kategori 1 (fraud) dengan 0 (no fraud), bisa dikatakan data imbalance extreme, Perlu adanya solusi sebelum proses modeling terjadi, terlebih modeling adaboost yang sangat terpengaruh hasilnya ketika data memiliki nilai outlier

Solusi *Imbalance* Data

```
is_fraud
0.0    0.99087
1.0    0.00911
Name: count, dtype: float64
```

Data sudah menjadi data yang balance data kategori 1 (fraud) dengan 0 (no fraud) sudah menjadi seimbang data (0) 0.5/50% dan data (1) 0.5/50%

Preprocessing



Data Modeling

Metode yang digunakan :

- Adaboost
- Logistic Regression
- Random Forest

Metrik :

- Recall (Utama)
- Akurasi
- Presisi
- F1 Score
- ROC

Pemilihan Metrik

Recall :

- Dalam kasus fraud, sangat penting untuk meminimalkan jumlah fraud yang tidak terdeteksi (false negative). Memiliki recall yang tinggi berarti model cenderung menemukan sebagian besar kasus fraud yang sebenarnya.
- Banyaknya false negatives dapat memiliki konsekuensi seperti kerugian keuangan yang signifikan atau reputasi yang rusak bagi perusahaan tersebut.

Hasil Evaluasi

	Model	Precision	Recall	F1 Score	Accuracy	ROC AUC
0	AdaBoostClassifier	0.783757	0.355684	0.489310	0.995702	0.677556
1	LogisticRegression	0.005348	0.000444	0.000820	0.993735	0.499982
2	RandomForestClassifier	0.840845	0.530195	0.650327	0.996699	0.764806

Hyperparameter

GridSearchCV

```
param_grid = {  
    'n_estimators': [1, 20, 50],  
    'max_features': ['auto', 'sqrt'],  
    'max_depth': [None, 10, 20],  
}
```

Scoring : Recall
CV : 3

Hasil Hyperparameter

Best Parameters: {'max_depth': 20, 'max_features': 'sqrt', 'n_estimators': 50}

Sebelum

Model	Precision	Recall	F1 Score	Accuracy	ROC AUC	Balanced Accuracy
RandomForestClassifier	0.840845	0.530195	0.650327	0.996699	0.764806	0.764806

Sesudah

Model	Precision	Recall	F1 Score	Accuracy	ROC AUC	Balanced Accuracy
RandomForestClassifier	0.830556	0.531083	0.647887	0.996658	0.765226	0.765226



Terima Kasih