

FraudBusters Presents..



Vikram Borra, Austin Cappetta, Gabriel Jean-jumeau, Justin Sansbury, Larry Pavlik

*** Predicting Probability of Fraud ***

Overview:

Facts

Approach

Questions

Demo

Conclusion



Facts on Fraud

- 1 in 26 of the US adult population fall victim to identity theft every year. (source: Chargebacks911.com)
- The FBI reported receiving over 880,400 complaints in 2023, with potential losses exceeding \$12.5 billion. That represents nearly a 10% increase in complaints and a 22% increase in losses compared to 2022. (source: FBI.gov)
- Actual losses resulting from fraud came to roughly \$127.5 billion and on average 51% of Americans have experienced fraud more than once. (source: LexisNexis Risk Solutions)



LARRY... IS THIS YOU?



Questions to Answer:

- Can we predict the chance of fraud?
- What is the most effective way of calculating the probability of fraud using M/L



Approach:

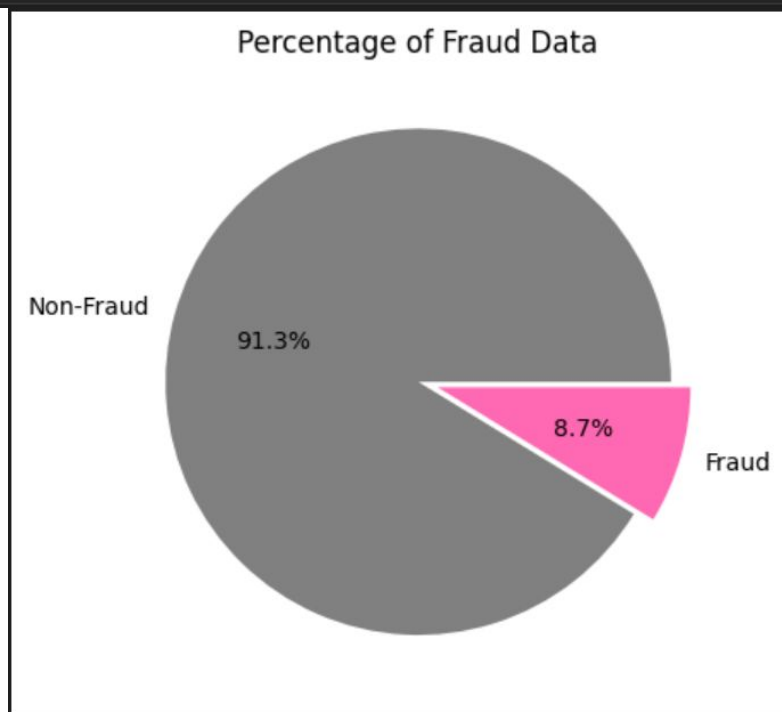
- KNN
- Decision Tree Model
- Random Forest Classifier
- Logistic Regression
- Optimization



The data is imbalance based on the distribution of the above count.

```
# Calculate the percentage of fraud in the dataset
labeled_data = ["Non-Fraud", "Fraud"]
plt.pie(fraud_tx_data_df["fraud"].value_counts(), labels=labeled_data, autopct="%1.1f%%", colors=["gray", "hotpink"], explode=(0, 0.08))
plt.title("Percentage of Fraud Data")
plt.show()
```

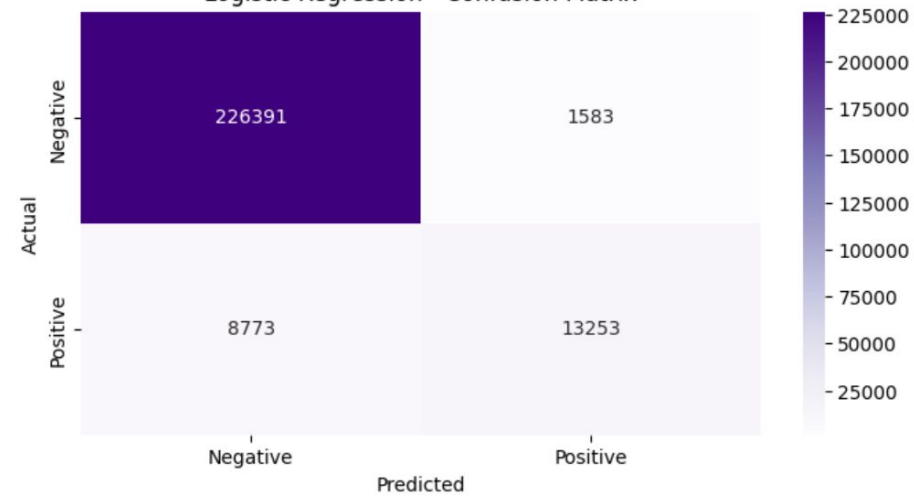
✓ 0.0s



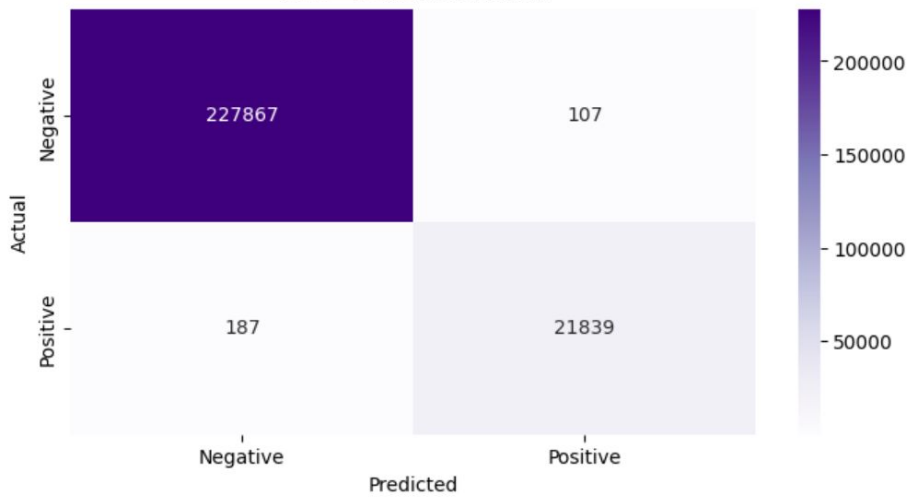
Random Forest Classifier - Confusion Matrix



Logistic Regression - Confusion Matrix



KNN - Confusion Matrix



Decision Tree - Confusion Matrix



Accuracy and R2 Comparsion across 4 models

	Logistic Regresssion	Random Forest Classifier	Decision Tree	K Neighbors Classifier (K=5)
Accuracy Score	0.95858	0.99999	0.99999	0.99882
Score	0.95858	0.99999	0.99999	0.99882
R2	0.48440	0.99990	0.99990	0.98536

RandomizedSearchCV

	precision	recall	f1-score	support
negative	0.92	1.00	0.96	227974
positive	0.78	0.07	0.13	22026
accuracy			0.92	250000
macro avg	0.85	0.53	0.54	250000
weighted avg	0.91	0.92	0.88	250000

```
{'weights': 'distance', 'n_neighbors': 5, 'leaf_size': 493}
```

Random Undersampling

KNN (k=5) - Classification report before under sampling				
	precision	recall	f1-score	support
negative	1.00	1.00	1.00	227974
positive	1.00	0.99	0.99	22026
accuracy			1.00	250000
macro avg	1.00	1.00	1.00	250000
weighted avg	1.00	1.00	1.00	250000

KNN (k=5) - Classification report After under sampling				
	precision	recall	f1-score	support
negative	1.00	0.99	1.00	227974
positive	0.94	1.00	0.97	22026
accuracy			0.99	250000
macro avg	0.97	1.00	0.98	250000
weighted avg	0.99	0.99	0.99	250000

DEMO TIME

Probability Findings:

We can now predict fraud:

- in the store
- Online
- Accuracy of 99.99%



Next Steps

1. Monitor your accounts
2. Avoid using public wifi
3. Avoid unsecured websites
4. Use mobile wallet
5. If all else fails, call your card issuer



Appendix:

<https://github.com/VR-BORRA/credit-card-fraud-analysis>

<https://chargebacks911.com/>

<https://www.fbi.gov/>

<https://risk.lexisnexis.com/>



WE ACTUALLY FOUND AND CHASED A FRAUDSTER AND CAUGHT THEM, HERE IS THEIR IDENTITY





QUESTIONS?

