DEEP LEARNING FOR FRAUD DETECTION IN CREDIT CARD TRANSACTIONS

TEAM MEMBERS:

SOPHIA WILLIAMS

BRIAN BLANCATO

TYLER HINNENDAEL

MOTIVATION

- In 2023, US consumers reported over $10 billion in losses due to fraudulent activity which was a 14% increase from 2022 [1].
- Fraudsters target credit cards among other avenues to exploit individuals, potentially causing significant financial harm.
- Due to deep learning and artificial intelligence ability to process and find patterns in large volumes of data, we will be using these models to detect and mitigate instances of consumer exploitation.

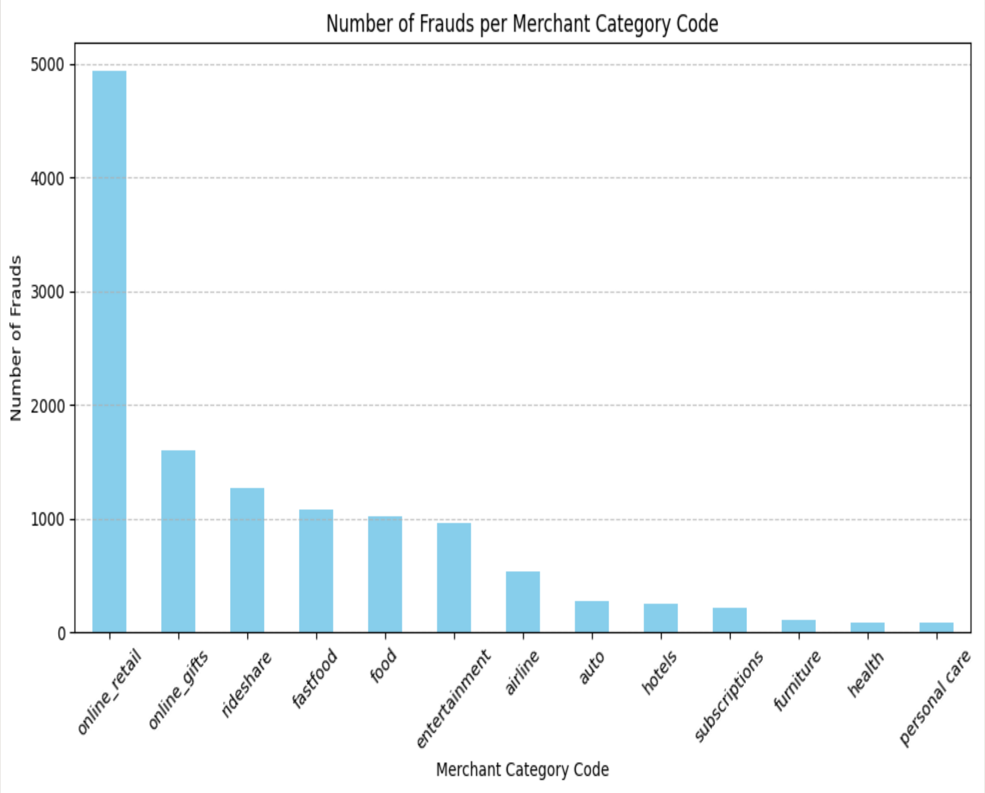
RELATED WORK

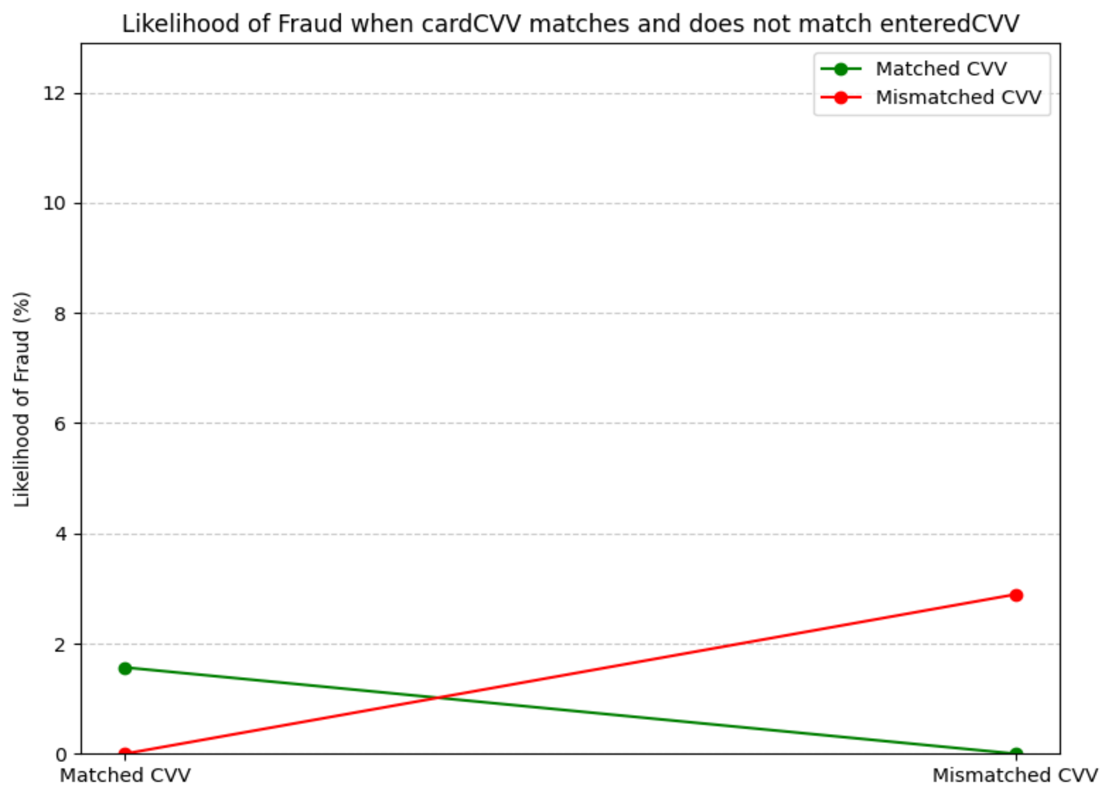
- We analyzed two papers which each identified models to classify a transaction as one of two classes - fraud or non-fraud
- In one paper [2], an ANN model architecture used 4 hidden layers and achieved a precision of 79 percent and recall of 81 percent
- A second paper [3] used an ANN model with 3 hidden layers with a stated accuracy ”approximately equal to 100 percent”
- The research can be expanded on by testing a Recurrent Neural Network.

DATASET

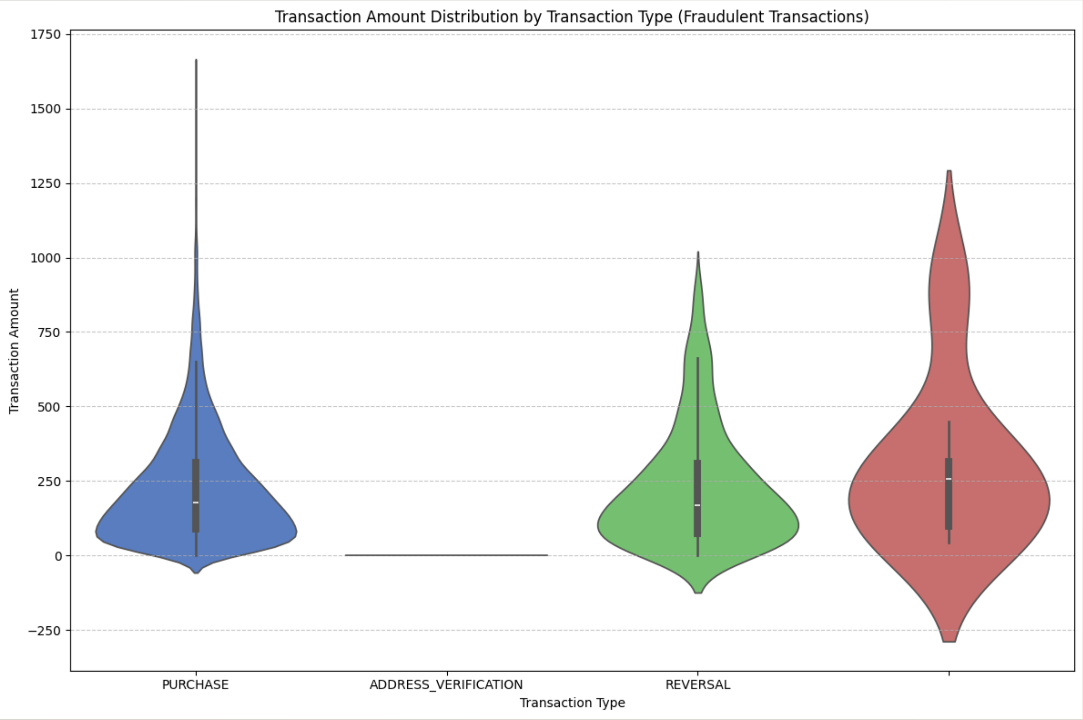
- We plan to use a synthetic dataset provided by Capital One, which includes 29 features for 786,363 instances.
- Among these instances, 12,417 are flagged as fraudulent transactions.
- 5,000 unique accounts with an average of 157 transactions each
- See reference [4] to find the dataset.

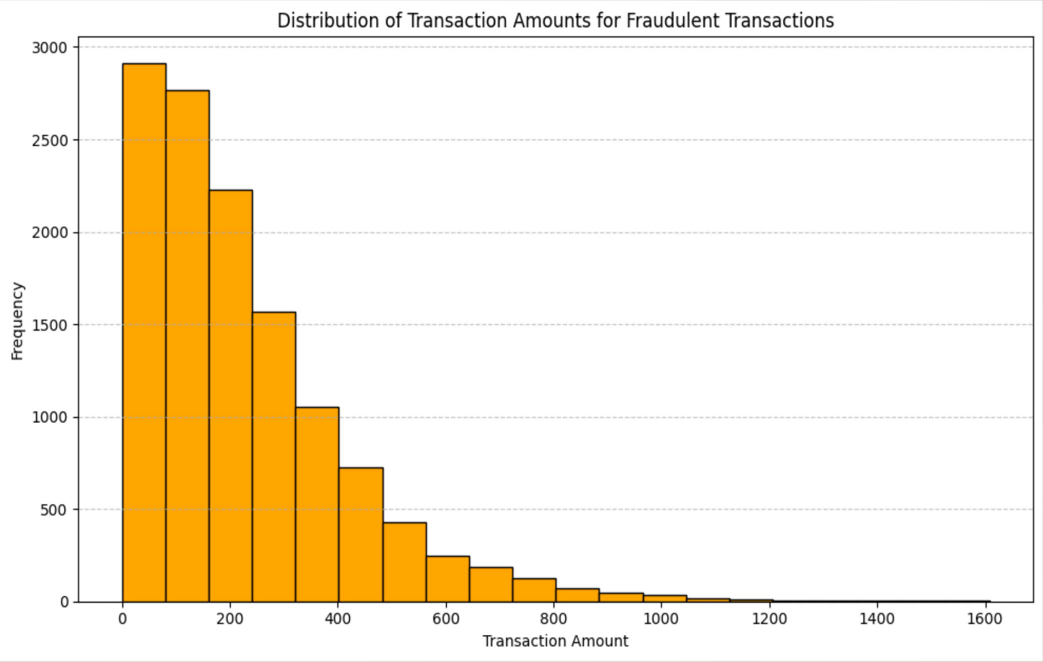
DATA ANALYSIS





DATA ANALYSIS





DATA PREPROCESSING

- Removed empty and unneeded columns along with rows containing NaN values
- Transformed datetime columns
- Encoded all categorical columns and converted Boolean columns to int
- Correlation matrix – ensured low correlation values between columns
- Implemented SMOTE on training dataset to address significant class imbalance
- Standardized data with StandardScalar
- Verified the data was ordered by account number to run the RNN models

NEURAL NETWORK

- Trained numerous ANN models experimenting different architectures and hyperparameters
  - Adjusting number of layers, activation functions, number of filters and regularization layers showed minimal improvement on validation and test sets.
  - Final Model: 4 Dense layers with bias, Leaky ReLU activation, 3 dropout layers

Layer	Output Shape	Param #
Dense	(None, 64)	3968
LeakyReLU	(None, 64)	0
Dropout	(None, 64)	0
Dense	(None, 32)	2080
LeakyReLU	(None, 32)	0
Dropout	(None, 32)	0
Dense	(None, 32)	1056
LeakyReLU	(None, 32)	0
Dropout	(None, 32)	0
Dense	(None, 1)	33

Metric	Score
Recall	0.14
Precision	0.08
F-1 Score	.10
Accuracy	.96

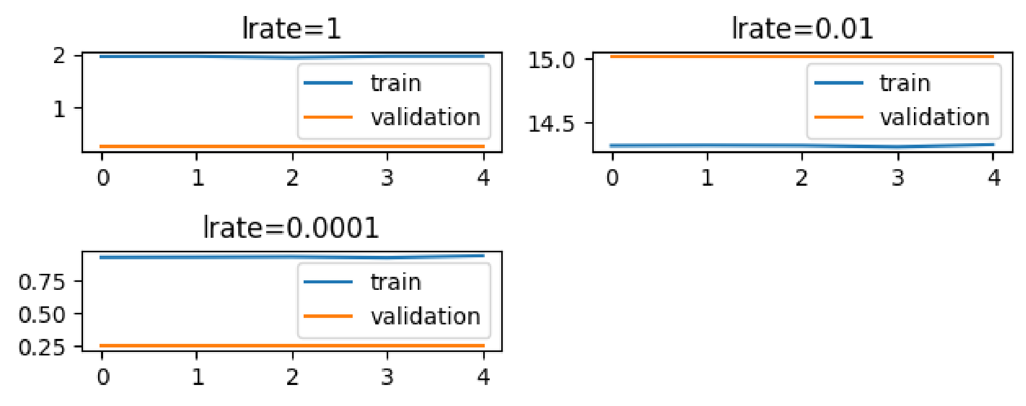
RECURRENT NEURAL NETWORK (RNN)

- Trained over 20 RNN models with multiple combinations of hyperparameters.
  - Able to achieve a high recall value (capture true positives), other metrics low
  - Architecture: embedding layer, dropout layer, pooling layer, dropout layer, dense layer
- Utilize customer transaction history for fraud detection – prior data for accounts
  - Sequential Transaction Data: RNN remembers prior data - captures patterns over multiple transactions.
- Best Model:

Recall	Precision	F-1 Score	Accuracy	Epochs	Optimizer	Learn Rate	Thresholds	Batch size
0.79	0.02	0.03	0.26	3	Adam	(1e-5)	0.4 (Recall)	32

HYPER PARAMETER TUNING

- Epochs and Early stopping
- Number of units per hidden layer
- Learning Rate
- Batch Size



LONG SHORT-TERM MEMORY (LSTM) RNN

- RNNs are well-suited for modeling sequential data because they can capture temporal dependencies in the data
  - A standard RNN can only capture short term dependencies
  - Training a RNN suffers from vanishing gradients (further we go back in time gradients get smaller and smaller)
  - As a result LSTM was used as a solution to handling vanishing gradients

Recall	Precision	F-1 Score	Accuracy	Epochs	Optimizer	Learn Rate	Thresholds	Batch size
0.0391	0.016	0.0748	0.9845	5	Adam	(0.01)	Recall score	32

RESULTS

- Able to achieve a high recall score with the RNN models, but unable to increase the other metrics – difficult to prevent overtraining on the unbalanced data
- When tuning the models, there was a trade-off between Accuracy and Recall

Model Type	Recall	Precision	F-1 Score	Accuracy
Basic NN	0.14	0.08	0.10	0.96
RNN	0.79	0.02	0.03	0.26
RNN- LSTM	0.039	0.8810	0.007	0.98

CONCLUSIONS

Fraud detection is hard!

Next Steps before deployment

- Automated Hyper Parameter Tuning option - Grid Search/Random Search (limited GPU)
- Need to keep recall high while increasing precision and F1 Score
- Implement other methods to balance data for neural networks – Mixup research
- Engage with the client to identify ideal recall threshold – dependent on the bank's resources
- Collect more data on fraudulent transactions

REFERENCES

- [1] Federal Trade Commission. ”AS Nationwide Fraud Losses Top $10 Billion in 2023, FTC Steps Up Efforts to Protect the Public”.https://www.ftc.gov/news-events/news/pressreleases/ 2024/02/nationwide-fraud-losses-top-10-billion-2023-ftc-stepsefforts-protect-public
- [2] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic, and A. Anderla. (2019). Credit Card Fraud Detection - Machine Learning methods. 1-5. 10.1109/INFOTEH.2019.8717766.
- [3] V. Agarwal. ”Identity theft detection using machine learning”. International Journal for Research in Applied Science and Engineering Technology. 2021;9(8):1943–1946. doi: 10.22214/ijraset.2021.37696
- [4] Capital One transaction dataset. https://github.com/CapitalOneRecruiting/DS
- [5] W.C. Cheng, T.H. Mai, and H.T. Lin. "From SMOTE to Mixup for Deep Imbalanced Classification." International Conference on Technologies and Applications of Artificial Intelligence. Singapore: Springer Nature Singapore, 2023.