Team 52: Blaire Kim, Lily Li, Linkun Ji, Vaibhav Bagaria, Zijian Xie

Binary Classification Random Forest Predictive Model for Credit Card Fraud Detection

I. Business Understanding

Nearly 2 billion credit card transactions occur globally each day, increasing the risk of fraud. In 2023, the FTC reported over 114,000 cases of credit card fraud in the U.S., leading to nearly $12 billion in losses annually and $30 billion worldwide. Detecting fraudulent transactions in bank credit card transaction data is crucial for bank reputation as fraudulent transactions lead to financial losses for both businesses and customers within the Financial Services, FinTech, and Risk Management industries. Customers who face credit fraud are faced with dissatisfaction, while for the business, fraud increases operational costs due to chargebacks. Accuracy within fraud detection systems is crucial for preventing financial losses, protecting customers, and maintaining trust in the system. With the high numbers of fraud, financial institutions and banks are more than ever seeking advanced technological solutions with data analytics to prevent fraud and protect their customers. While customers often self-report instances of fraud, machine learning and AI analytics have emerged as the new standard for efficiently processing large volumes of transactions. By analyzing consumer behavior, such as spending patterns and credit usage, these technologies can quickly identify and flag unusual activity, alerting cardholders to potential anomalies. Examples of unusual activity can include sudden increases in spending, purchases that take place in another country or location that is very far from typical locations, unusual times, and multiple transactions from the same retailer.

II. Data Understanding and Preparation

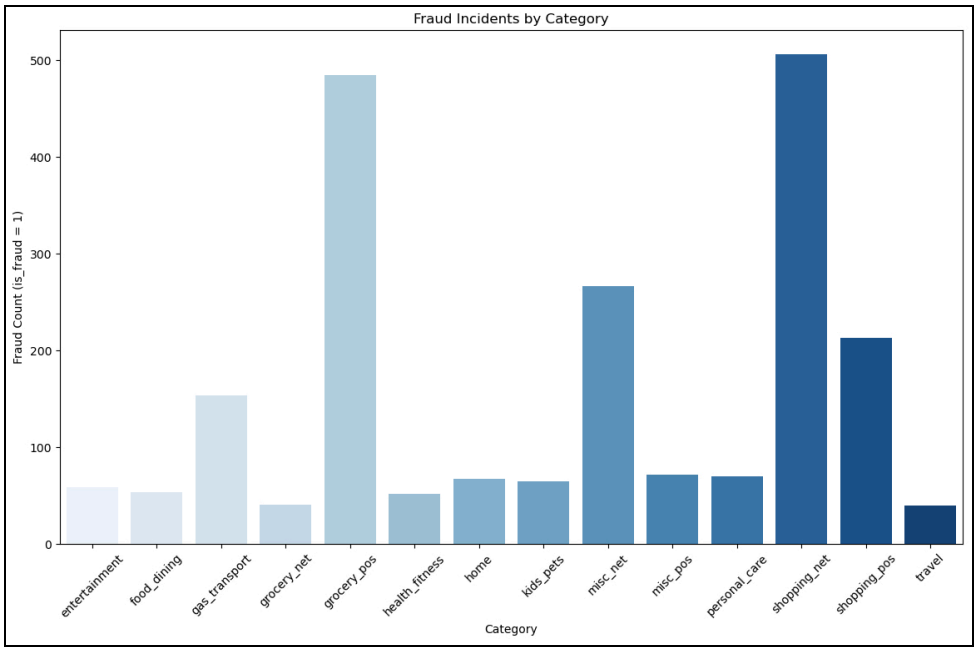
Financial institutions store historical data, through customer reports and transactional investigations, on POS systems and relational databases; they will utilize machine learning

algorithms to conduct real-time fraud detection and analysis. For the purpose of our predictive modeling project, we wanted to simulate this real-time prediction by obtaining a simulated transaction dataset from Kaggle that was randomly generated on Python for the time period of January 1, 2019 to December 31, 2020. It covers transactions and credit cards of 1000 customers' transactions within a pool of 800 merchants. The dataset came in the form of a table with various feature vectors for vectors in a CSV file with one row corresponding to a unique identifier for each transaction. Outside of the unique ID, there are 21 other variables that correspond to attributes that can be used for fraud prediction. Below is additional information on the features included in each transaction.

index	Unique Identifier for each row	state	State of Credit Card Holder
trans_date_trans_time	Transaction DateTime	zip	Zip of Credit Card Holder
cc_num	Credit Card Number of Customer	lat	Latitude Location of Credit Card Holder
merchant	Merchant Name	long	Longitude Location of Credit Card Holder
category	Category of Merchant	city_pop	Credit Card Holder's City Population
amt	Amount of Transaction	job	Job of Credit Card Holder
first	First Name of Credit Card Holder	dob	Date of Birth of Credit Card Holder
last	Last Name of Credit Card Holder	trans_num	Transaction Number
gender	Gender of Credit Card Holder	unix_time	UNIX Time of transaction
street	Street Address of Credit Card Holder	merch_lat	Latitude Location of Merchant
city	City of Credit Card Holder	merch_long	Longitude Location of Merchant
		is_fraud	Fraud Flag <--- Target Class

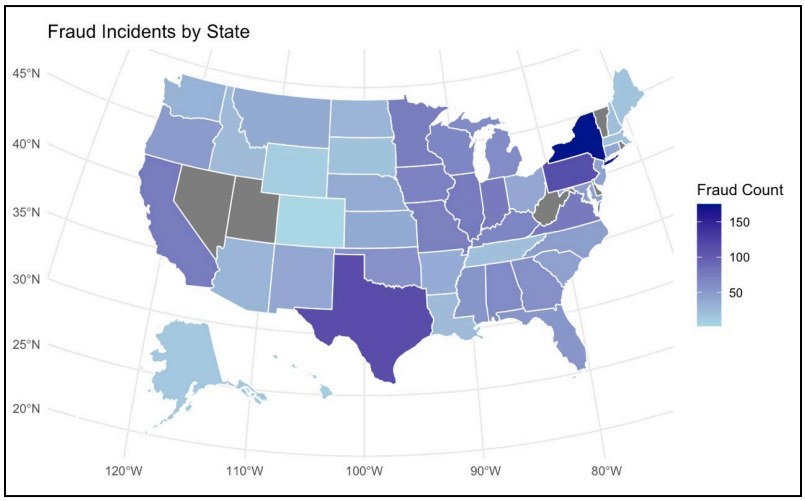
Under this supervised training and evaluation, the target variable of interest is "is_fraud" - a binary variable where 1 indicates a fraudulent transaction and 0 indicates a legitimate transaction. There are many physical costs associated with fraud detection as it involves labor hours from investigation, financial losses, and customer satisfaction. Machine learning mathematical models are used on historical data to predict future instances of fraud; it is crucial to balance costs associated with misclassifying a transaction against the cost of allowing a fraudulent transaction to go undetected. These costs are crucial for our model's performance metrics. Particularly, we will pay attention to False Positives - where there will be costs affiliated with customer dissatisfaction and wasted labor and investigation hours - and False Negatives - where there will be costs affiliated with financial losses from undetected instances of fraud.

To clean the data, we first checked to see if our Kaggle data set had any NAs - there were none. With our dataframe having 23 variables, we first did a check to see if they were all relevant in predicting fraud. Out of all the variables, we dropped columns for first name and last name, street address, transaction coordinates, credit card number, transaction number, merchant location coordinates, and unix time as they do not appear to be relevant variables in predicting credit card fraud. A new column for *Age* was added. Categorical variables were made into factors. Out of our sample size of 555,719 observations, only 2,145 observations are labeled as Fraudulent Activity, which is only 0.386% of our data set. The high proportion of non-fraudulent observations as opposed to fraudulent observations could introduce bias in our model by being very good at predicting non-fraudulent activity as opposed to fraudulent activity. To do the best we can to address this, we conducted a random downsample of the non-fraudulent transactions to ensure our model is able to gain accuracy and precision when undergoing the training process. After downsampling our dataset, we can see that most of the fraudulent transactions within the dataset happened from grocery point of sales and online shopping transactions.



However, as much as our data is very robust and contains a lot of features, there are still potential biases that are not fully accounted for, especially looking at fraudulent transactions

from each state. In particular, our data sample may include some discrepancies in quantity of data when segmenting by state. Given that states with bigger cities are naturally bound to have more instances of fraud, the model may be well trained on identifying fraud with more samples from bigger states but may underperform when analyzing underrepresented states within our dataset with fewer transactions.



III. Modeling

When it comes to the business problem, the data solution will involve looking at two outcomes - either a transaction is fraudulent or the transaction is not fraudulent, which makes our problem a binary classification prediction problem. Since fraudulent transactions are much less common than legitimate ones for any bank or financial institution, our dataset presented imbalance classes, where one outcome is far more represented than the other. This imbalance can lead to poor performance in predicting the minority class (fraudulent cases) , as the model may disproportionately favor the majority class (non-fraudulent cases). We dealt with this imbalance by using random undersampling of the non-fraudulent cases. We realized it is important to be cautious of overfitting. Given the large volume of daily card transaction data a bank would have to process, it is crucial to develop a model that makes precise and accurate decisions. As a credit card company, instances of fraud can lead to severe financial losses and lost consumer trust. Our model primarily focuses on the four metrics of Accuracy, Precision, Recall (Sensitivity), and the

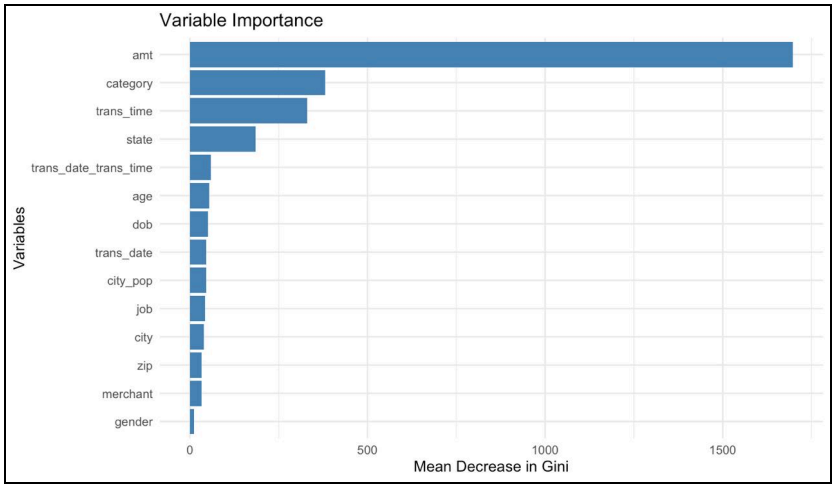
F-Measure score (balance between precision and recall). This helped improve the model's overall accuracy in identifying both fraud and non-fraud transactions. Without this adjustment, the model might have shown high accuracy simply because it was good at predicting legitimate transactions, even if it struggled to correctly identify fraud. When it came to developing our fraud detection model, we chose the Random Forest classification model as it has a strong ability to handle large datasets like ours for instance with the ability of balancing out accuracy, precision, and recall. The random forest model is also able to identify complex patterns from variables necessary to predict if fraud is true or not due to its nature of building multiple decision trees to uncover patterns in the data. While we also considered Logistic Regression, which is easier to interpret, it didn't perform as well with the more complex relationships between the variables that are crucial for predicting fraud. Ultimately, we found that the Random Forest classification model gave us the best combination of interpretability, accuracy, and computational efficiency to address detection of credit card fraud. After cleaning the data, we trained the model using 80% of the dataset and tested it on the remaining 20% to see how well it predicted fraudulent transactions. Since the dataset was downsampled to assist the model with predicting the minority class of is_fraud = 1, our final cleaned dataset included 23,595 transactions. Of these, 18,876 transactions were used to train the random forest model, while the remaining 4,719 transactions make up the test data, which was not seen during training and used for evaluation.

Metrics		Threshold
Precision	0.8889	0.5
Recall	0.7754	0.5
Accuracy	0.9712	0.5
F-Measure	0.8283	0.5
OOS R^2	0.6468	
AUC	0.9487	

IV. Evaluation

Our confusion matrix shows the breakdown of our model's performance. Out of the 4,719 test data points, our model was able to correctly identify 4255 values as non-fraudulent and 328 values as fraudulent. However, there were 95 values that were predicted as non-fraudulent when they were actually fraudulent, and 41 values were predicted as fraudulent when they were legitimate. Since this model is not perfect, it is crucial to assess our binary classification problem's performance with a few key metrics that use a decision threshold of 0.5: precision, accuracy, recall, and the F-Measure score, which balances precision and recall. In terms of precision, out of all the cases that the model predicted as fraud, 89% of them were correct. Out of all the cases that were actually fraudulent, 77.54% of them were explained by the model. In terms of overall accuracy, 82.8% were accurately captured by this model, whether it is fraudulent or non-fraudulent. As shown in the feature importance chart below, a combination of the transaction amount, category of transaction, state the transaction took place in, and transaction time all contributed to being important features that allowed the random forest model to predict if a transaction was fraudulent or not.

	Reference	
Prediction	**Is_Fraud = 0**	**Is_Fraud = 1**
Is_Fraud = 0	**4255**	**95**
Is_Fraud = 1	**41**	**328**



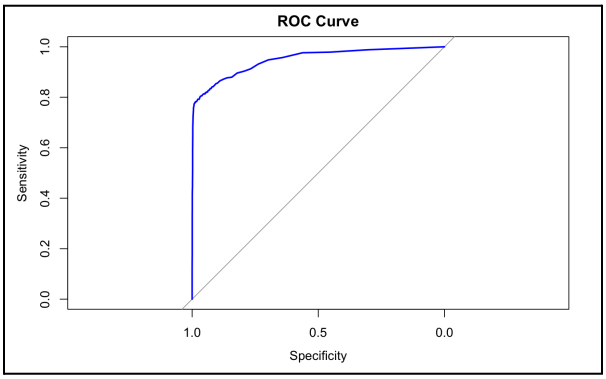
When credit card fraud occurs, the bank will have to make decisions that involve trade-offs affecting their own profitability and customers. For instance, when a bank correctly detects a fraudulent transaction, there will be losses incurred but they can take immediate action to prevent further loss; as much as they will incur costs from time spent investigating and

corrective measures, they will gain benefits from customer satisfaction and lower chance of customer churn. This scenario is seen as a true positive, where fraud is detected when it should be. Nevertheless, there can be two types of errors in their decision-making process - a false positive or a false negative. False positives occur when transactions are incorrectly flagged as fraudulent when they are legitimate. This not only frustrates customers but also imposes costs and loss of resources, including investigation time and labor costs. Similarly, a false negative happens when fraudulent transactions are mistaken for legitimate transactions. In this scenario, the bank and consumers affected will incur financial losses, and fraudulent activity goes undetected. Therefore, it is absolutely crucial for the bank to minimize any type of loss from both scenarios to maintain operational efficiency, consumer trust and loyalty, and profits. See below for the mathematical model to represent this problem.

$$max(EV) = E[TP + TN] - E[FP + FN]$$

$$max(EV) = P(Yi=1 \mid Xi) \cdot RTP + P(Yi=0 \mid Xi) \cdot 0 - P(Yi=1 \mid Xi) \cdot CFP - P(Yi=0 \mid Xi) \cdot CFN$$

To assess how well our model performs in detecting fraudulent and legitimate transactions, we looked at the ROC curve and calculated the AUC (Area Under the Curve), which came out to 0.9487. The ROC curve helps us understand the balance between identifying legitimate fraud cases while minimizing false positives at different decision thresholds. The AUC score of 0.9487 shows that our model is highly effective, with strong predictive power to differentiate between fraud and legitimate transactions. A score close to 1 means the model performs very well in reducing both false positives (flagging legitimate transactions as fraud) and false negatives (missing actual fraud).



For financial institutions and banks, a high AUC from this binary classification predictive model

is crucial because it ensures the system can effectively protect both customers and the company from fraud-related losses about 95 times out of 100. Additionally, we calculated an out of sample R^2 of 0.6468 provides further confidence in its predictive power; this means that when the model is tested on data it has never seen before, it will capture about 64.68% of the variance in our target variable is_fraud from the input features. An R^2 value above 0.5 suggests that the model is relatively effective in distinguishing between fraudulent and non-fraudulent transactions. However, given the high stakes involved in fraud detection, there is room for improvement, particularly in reducing potential misclassifications that could lead to financial losses or customer dissatisfaction. Below is a summary of our key metrics.

Metrics		Threshold
Precision	0.8889	0.5
Recall	0.7754	0.5
Accuracy	0.9712	0.5
F-Measure	0.8283	0.5
OOS R^2	0.6468	
AUC	0.9487	

V. Deployment

Moving forward, the bank should deploy this fraud detection model by integrating it with AI into its transaction monitoring system so it could flag suspicious activity, which can then be sent off to fraud analysts. It will be essential for the bank to monitor model performance continuously and update the model with new data regularly, as fraud patterns and tactics can evolve over time. Additionally, further fine-tuning of the model could potentially enhance its performance, especially in balancing the trade-offs between false positives and false negatives. Overall, this model provides a solid foundation for identifying potential fraud, helping financial institutions minimize losses, maintain customer trust, and optimize operational efficiency.

The deployment of the credit fraud detection model will involve integrating it into the bank's existing transaction monitoring system. This integration ensures that the model can assess transactions in real-time and automatically flag any suspicious activity for further investigation.

Once deployed, the system will analyze transactions as they occur and, based on the model's predictions, either allow them to proceed or flag them for review. A crucial aspect of the deployment will be to establish a feedback loop to ensure continuous learning. Fraud patterns often evolve, so the model needs to be retrained periodically with new transaction data to maintain its effectiveness. When deploying a fraud detection model, the firm must be aware of several critical issues. First, model accuracy is essential, especially in a changing environment where fraud tactics can shift rapidly. Without regular updates to the model, its performance will deteriorate over time, leading to more false positives (incorrectly flagged legitimate transactions) and false negatives (missed fraud). Second, the cost of false positives can be significant, leading to customer dissatisfaction and wasted resources on unnecessary investigations. Conversely, false negatives can result in substantial financial losses. Balancing these two outcomes is critical for ensuring the model's effectiveness and operational efficiency. Finally, the firm must ensure strict compliance with data privacy and security regulations, given the sensitive nature of the transaction data being processed. Failure to comply could result in legal penalties and damage to the bank's reputation. There are important ethical considerations when deploying a fraud detection model, primarily related to fairness and transparency. One major concern is the potential for bias in the model. If the data used to train the model reflects historical biases, certain demographic groups could be disproportionately flagged for fraud, leading to unfair treatment. To mitigate this, the firm should regularly audit the model for bias and take steps to retrain it using more representative data. Transparency is another ethical concern. Customers may be frustrated when their legitimate transactions are flagged as fraud, leading to inconvenience, bad experience, and reputational damage. Clear communication about why a transaction was flagged and providing easy ways for customers to validate flagged transactions can help these concerns. Deploying the fraud detection model also comes with several risks that

must be addressed. One key risk is model drift, where the model's performance deteriorates over time as fraud tactics evolve. This can lead to inaccurate predictions and increased false negatives.

To mitigate this, the model should be retrained on a regular basis using updated transaction data. Another risk is the impact of false positives on customer satisfaction. If too many legitimate transactions are flagged as fraud, customers may lose trust in the system and feel inconvenienced. To manage this, the decision thresholds can be adjusted to reduce false positives, and manual review can be employed for certain flagged transactions. Another risk of this model is that it does not account for segmentation by demographics as a factor in predicting fraud.

Appendix Citations:

https://www.capitalone.com/learn-grow/privacy-security/credit-card-fraud-detection/

https://spd.tech/machine-learning/credit-card-fraud-detection/

https://www.dataversity.net/how-data-is-used-in-fraud-detection-techniques-in-fintech-business/

Team Contribution:

Business Understanding: Everyone

Data Preparation and Understanding: Everyone

Modeling: Everyone

Evaluation: Everyone

Deployment: Everyone