Development phase

Detecting Credit-Card Fraud in the Banking Sector using Machine Learning Techniques

Author: Ajay Chandra A S

ajay.chandra-a-s@iu-study.org

Matriculation number: 321149868

Module: Project Data Science Use Case (DLMDSPDSUC01)

Date: 17-05-2023

Place: Bengaluru, Karnataka

Table of Contents

Use-Case Analysis

The goal of this project is to enhance Acme Commerce Bank's credit card fraud detection capabilities by using advanced machine learning techniques. The existing rule-based system employed by the bank is inadequate in identifying fraud due to its limited capability in detecting intricate fraud patterns. The aim is to develop a digital product that accurately identifies and prevents credit card fraud in real-time and ensure a secure banking experience for customers and in turn provide various intangible benefits to the customers and also the bank.

Value Proposition

The foundation of any system that employs machine learning is the value proposition, which comprises the What, Why, and Who of the project. This central block outlines what the system aims to accomplish, why it is significant, and who the system will serve or affect. In essence, the value proposition sets the tone for the project and guides all subsequent decision-making and implementation (Dorard, 2021a).

The following figure is the Value Proposition Canvas by Strategyzer and it consists of two sections: the customer profile and the value map. The customer profile outlines the customer's jobs to be done, their pains, and their gains. The value map outlines the products and services that the company offers, as well as the benefits they provide to the customer (Strategyzer, 2017).

In our project, we have used the Value Proposition Canvas to explain how our ML model will provide value to both Acme Commerce Bank and its customers. In the customer profile section, we identify the specific jobs we are trying to accomplish, the pains the customers experience in the current credit card fraud detection system, and the gains they hope to achieve. The value map outlines how our ML model will provide benefits to the customer and the bank.

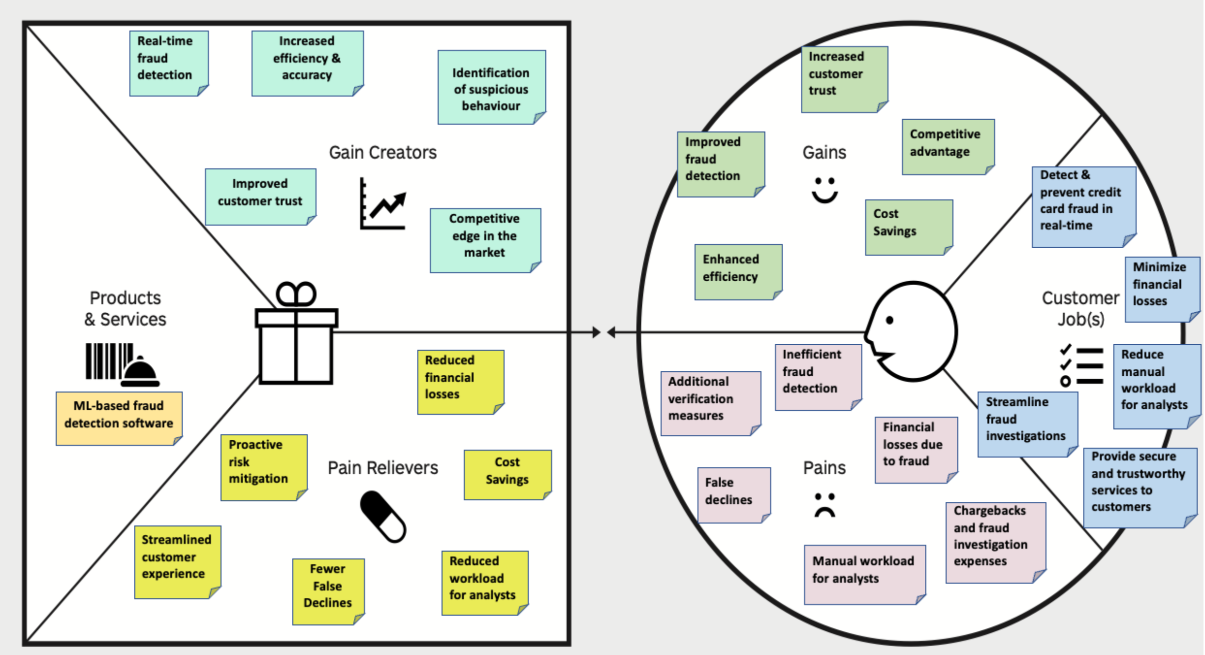
Real-time fraud detection
Increased efficiency & accuracy
Identification of suspicious behaviour
Gain Creators
Improved customer trust
Competitive edge in the market
Products & Services
ML-based fraud detection software
Reduced financial losses
Proactive risk mitigation
Pain Relievers
Cost Savings
Streamlined customer experience
Fewer False Declines
Reduced workload for analysts
Increased customer trust
Improved fraud detection
Gains
Competitive advantage
Cost Savings
Enhanced efficiency
Detect & prevent credit card fraud in real-time
Minimize financial losses
Customer Job(s)
Reduce manual workload for analysts
Inefficient fraud detection
Additional verification measures
Streamline fraud investigations
Financial losses due to fraud
Provide secure and trustworthy services to customers
Pains
False declines
Chargebacks and fraud investigation expenses
Manual workload for analysts

Figure: Value Proposition Canvas (AG, n.d.)

The implementation of the machine learning model for fraud detection in Acme Commerce Bank can offer the following benefits:

- **Minimizing Financial Losses**: The model will provide an accurate and efficient way to detect credit card fraud in real-time, reducing financial losses for the bank and protecting customers from fraudulent activities.
- **Enhancing Customer Trust**: By effectively detecting and preventing fraudulent transactions, customers will feel more secure and confident in using the bank's services, improving their trust and loyalty towards the company.
- **Cost Savings**: The bank can save money that would have otherwise been lost due expenses associated with fraud investigations, resulting in significant cost savings.
- **Competitive Advantage**: With an effective machine learning model for fraud detection, the bank can gain a competitive edge in the banking sector by providing more secure and trustworthy services to their customers, improving their reputation in the market.
- **Reduced Manual Workload**: The ML-driven system reduces the manual work required for additional verification, resulting in increased efficiency and reduced workload for analysts ("Credit Card Fraud Detection," n.d.).
- **Improved Accuracy**: The enhanced accuracy of the system reduces the burden on analysts, allowing them to focus on flagged transactions with greater confidence in their legitimacy ("Credit Card Fraud Detection," n.d.).
- **Fewer False Declines**: The number of false declines can be significantly reduced, improving the overall customer experience ("Credit Card Fraud Detection," n.d.).
- **Reduced Verification Measures**: With the ML model accurately flagging suspicious transactions, there will be fewer legitimate transactions that require additional verification measures, resulting in a smoother and more streamlined customer experience, as well as improved efficiency for the bank ("Credit Card Fraud Detection," n.d.).
- **Identification of Suspicious Consumer Behaviour**: After the model has detected or flagged a fraudulent transaction, expert analysts can analyze the transaction features to identify abnormal consumer behavior, such as atypical spending habits or transactions that deviate from typical patterns, which may have contributed to the occurrence of fraud.
- **Evaluation of Customer Risk**: After the model identifies a transaction as fraudulent, the experienced analysts can examine the customer's past and current transactions to determine if they are a high-risk customer. This analysis can help identify potential risks, allowing the bank to take preemptive measures to mitigate those risks and prevent fraudulent activities ("Risk Analysis (Business)," 2022).

Learn

Our approach to developing a machine learning model for detecting credit card fraud at Acme Commerce Bank involves an iterative supervised learning process ("Supervised Learning," 2023). We will collect data and continuously update and refine the model to enhance its accuracy and efficiency. Through this learning process, we will identify patterns and trends in credit card fraud to build a more effective fraud detection system.

Data Sources

To develop an effective credit card fraud detection model for Acme Commerce Bank, data must be collected from various sources. It is important to keep collecting data regularly, both in the short term and long term (Dorard, 2021b).

Credit card transaction data: This includes information on all credit card transactions in real-time made by the bank's customers, such as transaction amounts, merchant information, location information, transaction timestamps etc.,

Customer data: This includes information about the bank's customers, such as their demographic information (e.g. age, gender, income), credit history, past transaction data etc.,

Fraudulent transaction data: This includes information about previous fraudulent credit card transactions that have occurred at the bank, including details such as the transaction amounts, merchant information, and transaction timestamps.

Collaboration with other banks: Acme Commerce Bank can collaborate with other banks in the region to share information about credit card fraud incidents. This can help identify emerging trends and new fraud types that the model needs to be trained on.

Social media: The bank can monitor it's social media platforms for mentions of credit card fraud and use this information to update the model to recognize new types of fraud.

User feedback: The bank can collect feedback from customers regarding any fraudulent activities they have experienced with their credit cards. This information can help the bank identify new fraud patterns that the model needs to be trained on.

Data Collection

To build an effective credit card fraud detection model, Acme Commerce Bank can draw on various data sources including it's Customer Relationship Mangement system(CRM), Enterprise Resource Planning system(ERP), customer support systems, data warehouses, social media, other banks and static files. These sources contain valuable data on customer behavior and transaction history, among other things, that can help identify fraudulent activity. By analyzing the available data, experts can select the most relevant data for creating an accurate model (Dorard, 2021b).

Features

Feature selection is a crucial step in data pre-processing when it comes to detecting credit card fraud for Acme Commerce Bank using machine learning. This process involves the use of algorithms to identify the most significant features from a large pool of options, as well as techniques to select features that improve the performance of the model. To create a robust credit card fraud detection model for Acme Commerce Bank, various features extracted from different sources should be taken into account (Dorard, 2021b).

To select the initial features for our credit card fraud detection model, we referred to sources such as Research papers (Dornadula & Geetha, 2019), fraud detection dataset on Kaggle (*Fabricated Fraud Detection, n.d.*) and the examples provided on neuraldesigner.com (Credit Card Fraud Detection Using Machine Learning, n.d.) as follows:

- transaction_amount: the amount of the transaction that the customer had made.
- transaction_currency: the currency of the transaction.
- transaction_time: the time when the transaction was made.

- transaction_location: the location where the transaction was made.
- transaction_type: the type of the transaction (e.g., online, in-store, etc.).
- transaction_category: the category of the transaction (e.g., travel, shopping, etc.).
- merchant_id: id of the merchant is used to identify and track the transactions processed by the merchant.
- merchant_category: the category of the merchant (e.g., clothing, electronics, etc.).
- avg_amount_day: average amount of transaction per day.
- foreign_transaction: yes = it is a foreign transaction, no = it is not a foreign transaction.
- high_risk_country: yes = it is a high-risk country, no = it is not a high-risk country.
- credit_limit: the maximum amount that the customer can spend using their credit card.
- frequency_of_transactions: how often the customer uses their credit card.
- previous_fraud_history: whether the customer has a history of fraud or not.
- customer_demographic_information: details about the customer (e.g., age, income, etc.).
- collaborative_data_with_other_banks: data shared between banks to identify potential fraud.
- social_media_mentions_of_credit_card_fraud: monitoring social media for mentions of credit card fraud.
- nameOrig: the unique identifier of the customer who initiated the transaction.
- oldbalanceOrg: the account balance of the customer before the transaction was initiated.
- newbalanceOrig: the account balance of the customer after the transaction was completed.
- nameDest: the unique identifier of the customer who received the transaction.
- oldbalanceDest: the account balance of the recipient before the transaction was initiated.
- newbalanceDest: the account balance of the recipient after the transaction was completed.
- card_age: the age of the customer's credit card account.
- IP_address: the IP address of the device used to initiate the transaction.
- browser_info: information about the browser used to initiate the transaction.
- transaction_amount_category: the category of transaction amount, such as low, medium, or high.
- credit_score: the credit score of the customer.
- account_age: the age of the customer's account with the bank.
- average_balance: the average account balance of the customer over a specified period.
- isFlaggedFraud: a binary variable that indicates whether a transaction was flagged as suspicious by the system due to a potential fraud risk.
- isFraud: a binary variable that indicates whether the transaction is fraudulent or not.

These features will serve as input to the machine learning algorithm to train the model to accurately detect fraudulent activities (“Feature (Machine Learning),” 2023). The model's accuracy will improve as it gains more data and evolves, and additional features may be added or updated to improve its precision. Evaluating the model's performance regularly and adjusting its features accordingly is essential to ensure that it can identify fraudulent activity accurately. The goal is to create a flexible and adaptive system that can stay up-to-date with emerging trends and new fraud types to protect the bank and its customers.

Building Models

We will implement a strategy for Acme Commerce Bank to regularly update their fraud detection models with new training data (Dorard, 2021b). These models will be updated every quarter or year depending on various factors such as the rate of fraudulent activities, changes in customer behavior, and changes in the bank's policies or systems. We will also create new models or update existing ones in response to major shifts in the banking industry or changes in regulatory requirements to ensure that the fraud detection system remains effective and accurate.

To train these machine learning models, we will use supervised learning, as it is a binary classification problem ("Binary Classification," 2022). We will train the model using historical data from Acme Commerce Bank that has already been classified with the intended output feature, and we will adjust the weights to produce the desired output. We will compare the accuracy of the output to the actual results, and the training will be considered complete when the measure reaches an acceptable low level. This will allow the model to identify patterns and predict labels for unlabelled data.

To keep the models up-to-date, we will continuously analyze and retrain the model with new data (Watson, 2018). We will save new training data as it comes in and compare its accuracy to the existing models. If the accuracy of the models deteriorates over time, we will construct and deploy a new model using new data or a combination of new and old training data. Detecting credit card fraud involves an ever-evolving problem, and fraudsters are always introducing new patterns of fraudulent behavior. Therefore, we must continually update the models and incorporate new data to ensure the effectiveness and accuracy of the fraud detection system (Dorard, 2021b).

Predict

In our approach, we will employ a process of continuous learning. This will involve gathering data and refining the model to improve its accuracy and efficiency over time ("Incremental Learning," 2023). By using this approach, we aim to identify patterns and trends in credit card fraud and develop a more effective fraud detection system. We will also make predictions on new credit card transactions based on the model's output, which will be used to inform decisions and prevent fraudulent activity. Additionally, we will evaluate the model's performance using various methods and metrics before deployment to ensure its effectiveness.

Machine Learning Task

The goal of Acme Commerce Bank's credit card fraud detection with machine learning is to classify credit card transactions as either fraudulent or non-fraudulent using a set of features as input and generating a prediction as output (Dorard, 2021c). The bank's current rule-based fraud detection system can serve as a baseline for this task. It's a supervised learning problem as we have labeled data for both fraudulent and non-fraudulent transactions. The task is a binary classification problem as we are predicting whether a transaction is fraudulent or not. The dataset is unbalanced, with most transactions being non-fraudulent and only a small percentage being fraudulent (Ravaglia, 2022). The specific distribution of the dataset depends on the bank's customer base and the prevalence of credit card fraud in the area.

Decision

The task of the ML system in Acme Commerce Bank is to predict the fraudulent credit card transactions. However in the end, predictions are just information, and they do not do anything useful on their own. Therefore, we need to turn prediction into decision that deliver the intended value (Dorard, 2021c).

Decisions are often based on the model’s confidence in its predictions. The model provides a confidence value, typically ranging from 0 to 1, with each of its predictions. This confidence value can be utilized to automate decisions when it exceeds certain thresholds (Dorard, 2021c). It is recommended to seek guidance from experts in the relevant field to establish appropriate confidence thresholds. If the credit card transaction is predicted to be suspicious but not definitively fraudulent, the system should flag the transaction for review by the bank's fraud detection team. If the credit card transaction is predicted to be fraudulent with a high degree of confidence, the system should automatically block the transaction in real-time and notify the customer and the bank's fraud detection team. The bank's fraud detection team should use the machine learning model's predictions as one of several inputs to make a final decision on whether a transaction is fraudulent or not.

After every credit card transaction, predictions are made to detect any fraudulent activity. To prioritize the high-risk transactions, the first step is to filter out the non-fraudulent ones. The fraudulent transactions are then sorted in descending order of their probability of fraud, and the bank can target the list by focusing their efforts on the transactions with the highest probability of fraud. This strategy enables the bank to allocate their resources more efficiently and prioritize their efforts towards preventing fraudulent activities among the most suspicious customersc (Dorard, 2021c).

The ultimate vision when building intelligent systems can be to automate decisions completely. But it is recommended to avoid full automation in our case when we are just starting out with the project. Instead, all possible fraudulent transactions should be flagged for the expert analysts to review them and finalize the decision based on the prediction (Dorard, 2021c).

Making Prediction

Acme Commerce Bank should use prediction methods to detect potential fraudulent activities before they take place. In order to achieve this, the ML model generates predictions for each new credit card transaction based on previously learned patterns of fraudulent activity. Predictions are made for every new transaction, and the number of predictions is equal to the number of decisions. It is important that predictions are made in real-time immediately after the transaction is processed, taking into account the time required for the model to generate a prediction and the time needed to extract feature values (Dorard, 2021d).

In order to detect and prevent fraudulent activities in a timely manner, it is essential that the featurization of new inputs and the prediction of new transactions occur with speed and efficiency. The model must be able to rapidly process and analyze the transaction data, utilizing learned patterns of fraudulent activities to produce accurate predictions promptly.

When making predictions, it is crucial to ensure that all listed features are utilized. The selection of features should be done thoughtfully so that they can be utilized to accurately define output value propositions. In order to avoid potential issues, it is important to confirm that each feature's value can be efficiently extracted from the data sources during prediction time. Prior to being input into the ML model, the features for the new input dataset should be sourced from their respective data sources and subjected to pre-processing (“Data Pre-Processing,” 2023).

Offline Evaluation

To ensure that our machine learning system for detecting credit card fraud at Acme Commerce Bank is ready for deployment, we need to evaluate the model through a simulation before actual deployment. This pre-evaluation process will help us build confidence and trust in the system's performance and readiness for deployment (Dorard, 2021d).

To ensure the trustworthiness of our simulation, we should run it on test cases that are representative of what the system would encounter when deployed. Similarly, we should train our system on data that is representative of what it will be tested on. However, we should be cautious not to expose the system to too much information about what it will be tested on (Dorard, 2021d).

As our dataset has a significantly higher number of normal transactions compared to fraudulent ones, we should ensure that our model is trained and evaluated on a balanced dataset by using a sampling method that proportionally represents both normal and fraudulent transactions in the training and test sets. However, we also need to consider domain-specific requirements such as correctly distributing time-bound inputs between the training and test sets. This is important for capturing potential changes in fraud patterns over time. Splitting the data into two sets based on the time period represented, such as using the first six months of the year for the training set and the last six months for the test set, can be helpful. By considering these factors and consulting with domain experts, we can ensure that our evaluation results are accurate and meaningful (Dorard, 2021d).

In addition to a balanced dataset, the size of the test set is also important to achieve meaningful results in the domain of application. For instance, if we plan on updating our model every quarter, we should test on at least a quarter's worth of data. However, it may be beneficial to test on a longer period, such as a year, if fraudulent behavior exhibits a yearly pattern. Consulting with domain experts can help us determine the appropriate test set size for our problem and ensure that our evaluation results are reliable.

It is good practice to choose the test set in a way that makes it easy to present meaningful results, and to interpret them in the domain where the system will be used. One way to do this is to use the most recent data as test, so we can answer “how well would we have done if we had deployed this system X days/weeks/months ago?” (Dorard, 2021d).

In the context of making decisions in detecting credit card fraud, some possible metrics for offline evaluation include precision, recall, F1-score, and ROC-AUC. These metrics can help evaluate the effectiveness of the model in identifying fraudulent transactions and minimizing false positives. Additionally, other metrics like cost of fraud and cost of misclassifications can also be considered to assess the real-world impact of the model's performance.

Machine learning tasks are aligned to evaluation measures. There are several evaluation metrics for binary classification of credit-card fraud detection using Machine Learning such as confusion matrix, precision, recall, F1-score, and ROC-AUC (Reddy et al., n.d.).

Confusion Matrix: It is a useful tool to evaluate the performance of our machine learning model for detecting credit card fraud. It presents actual and predicted values in a two-dimensional table. In our case, the goal is to develop a classifier that can accurately predict fraudulent transactions. The confusion matrix includes class instances such as (“Confusion Matrix,” 2023):

- True Positive (TP): When a transaction is actually fraud and predicted as chargeback

- True Negative (TN): When a transaction is actually genuine and predicted as such

- False Positive (FP): When a transaction is actually genuine but predicted as fraud

- False Negative (FN): When a transaction is actually fraud but predicted as genuine

Precision: It is a performance metric used to assess the accuracy of the model's positive predictions in detecting fraudulent credit card transactions. It is calculated by dividing the number of correctly predicted fraudulent transactions by the total number of transactions predicted as fraudulent by the model. Precision is crucial in credit card fraud detection since it is essential to avoid false positives. High precision means that the model accurately predicts fraudulent transactions, reducing the number of false positives. However, evaluating the effectiveness of the model based only on precision may not be enough since it does not consider false negatives and does not take into account the balance between precision and recall ("Precision and Recall," 2023).

Recall: It is a performance metric used to identify all positive fraudulent transactions in the case of credit card fraud detection. It is the ratio of correctly predicted fraudulent transactions to the total number of actual fraudulent transactions. High recall indicates that the model can identify most of the fraudulent transactions and has a low false negative rate. Recall is crucial in credit card fraud detection because missing a fraudulent transaction could lead to significant financial losses for the bank and its customers. However, recall alone may not be enough to evaluate the effectiveness of the model as it ignores false positives and does not consider the trade-off between precision and recall ("Precision and Recall," 2023).

F1 Score: it is essential to strike a balance between precision and recall to optimize the overall performance of the model. A common approach is to use F1 score. F1 score is a measure of a model's accuracy that combines precision and recall into a single metric. In the context of detecting credit card fraud at Acme Commerce Bank, F1 score can be used to evaluate the effectiveness of a fraud detection model by considering both the false positives and false negatives. The F1 score is calculated as the harmonic mean of precision and recall, with a score of 1 indicating perfect precision and recall, and a score of 0 indicating the worst possible precision and recall ("F-Score," 2023).

An F1 score is a useful metric to assess the overall performance of a fraud detection model as it provides a balance between precision and recall. A high F1 score indicates that the model has a high accuracy rate, while a low F1 score indicates that the model has a lower accuracy rate. Therefore, it is important to optimize the F1 score when building a fraud detection model to ensure that both precision and recall are balanced to provide accurate and reliable predictions of fraudulent transactions.

AUC-ROC (Area Under the Receiver Operating Characteristic Curve): It is a performance metric used to evaluate the classification models. In the context of detecting credit card fraud in Acme Commerce Bank, AUC-ROC score can be used to measure how well the model can distinguish between fraudulent and genuine transactions. The ROC curve is a graph that represents the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR) for different threshold values. The AUC-ROC score is the area under this curve, which ranges from 0 to 1 ("Receiver Operating Characteristic," 2023).

A higher AUC-ROC score indicates that the model has a better ability to distinguish between fraudulent and genuine transactions. An AUC-ROC score of 0.5 means that the model is not performing better than a random guess, while a score of 1 indicates that the model has perfect discrimination ability. AUC-ROC is useful when dealing with imbalanced datasets where the number of fraudulent transactions is much smaller than genuine transactions since it is not affected by the imbalance in the data. Therefore, AUC-ROC can be used in combination with other performance metrics such as precision, recall, and F1 score to evaluate the overall performance of the credit card fraud detection model.

To measure the performance of our system, we need to compute the gain or reduction in cost of using the system compared to not using it or some other system, when running the simulation ("Loss Function," 2023). This approach provides a more practical and meaningful metric than just reporting abstract accuracy measures like F1-score. We need to make certain assumptions about the cost and gain values for the different types of errors. In the case of binary classification, we have two types of errors, False Positives and False Negatives. For instance, in our fraud detection scenario, a False Positive is a transaction that our system flags as fraudulent but is actually a legitimate transaction. A False Negative is a transaction that our system fails to flag as fraudulent, but is actually fraudulent. We need to assign appropriate cost and gain values to these errors, based on the specific context of our problem. The cost of an FP would be the cost of investigating a legitimate transaction, and the cost of an FN would be the potential loss incurred by the bank due to fraudulent activity. Similarly, we can estimate the gain for a True Positive as the cost of preventing the fraud minus the cost of investigation, and the gain for a True Negative is 0. These cost and gain values are fixed for the problem at hand. During the simulation, we can count the number of FPs, FNs, TPs, and TNs and multiply each value by its associated cost or gain value, and then sum everything to compute the overall gain or cost reduction achieved by the system (Dorard, 2021d).

We must also consider performance constraints, such as a maximum allowable error rate or a maximum number/proportion of certain types of errors that we can tolerate. For instance, we might set a constraint that the number of fraudulent transactions that go undetected (False Negatives) should be less than a certain amount every X days to maintain an acceptable level of security for our customers (Dorard, 2021d).

Another important consideration for evaluating the performance of the system is to ensure that it performs better than the current baseline, which is the existing rule-based fraud detection system. This implies that we need to first compare the baseline with offline evaluation before deploying the machine learning model based on the data (Dorard, 2021d).

Live Evaluation and Monitoring

After the successful evaluation of the fraud detection model in a simulation using relevant metrics, it can be implemented in the production environment. However, the decision-makers need to ensure that the model's deployment aligns with the organization's use case value. To determine whether the model meets the business objectives, a set of key performance indicators (KPIs) are used, including the following (“Performance Indicator,” 2023):

False Positive Rate: The rate of false positives, or the proportion of transactions wrongly labeled as fraudulent, should be minimized to prevent any inconvenience to customers with genuine transactions (“False Positive Rate,” 2023).

False Negative Rate: The percentage of fraudulent transactions that are not detected by the system. This should be kept as low as possible to ensure that fraudulent activities are detected and prevented (“False Positives and False Negatives,” 2023).

Real-Time Performance Monitoring: Continuously monitoring the performance of the model in real-time to detect any changes in its accuracy or false positives.

Key Metric Tracking: Tracking the key metrics such as precision, recall, F1 score, and ROC-AUC to evaluate the performance of the model.

Precision and Recall: These are the key metrics used to evaluate the performance of our classification model. Precision measures the accuracy of the positive predictions, while recall measures the ability of the model to identify all positive instances. Both metrics should be optimized to achieve the best performance (“Precision and Recall,” 2023).

Feedback Loop: To improve the model's performance over time, a feedback mechanism should be incorporated to allow it to learn from new data. This can be achieved by continuously monitoring the system and retraining the model. The model's algorithms should be refined based on feedback from stakeholders and retrained on new data to continuously improve its accuracy (Watson, 2018). Additionally, a feedback loop should be established to collect feedback from fraud analysts and investigators on the model's accuracy. This will help to identify any new patterns or trends in fraudulent activity that can be used to refine the model further (“Credit Card Fraud Detection,” n.d.).

Detection Impact Monitoring: The impact of the machine learning model on reducing financial losses caused by credit card fraud and the number of prevented fraudulent transactions should be monitored. The detection rate measures the proportion of fraudulent transactions correctly identified by the model, which is crucial in preventing losses due to credit card fraud. It is essential to achieve a high detection rate to minimize the impact of credit card fraud.

Cost-Benefit Analysis: The system should be evaluated based on the cost of implementing and maintaining it compared to the benefits it provides. This can include factors such as the reduction in financial losses due to fraud, increased customer satisfaction, and improved operational efficiency ("Cost–Benefit Analysis," 2023).

By tracking these KPIs, we can assess the performance of the credit card fraud detection model and determine whether it is meeting our business objectives. This information can help us make informed decisions about whether adjustments are needed to improve the model's accuracy and effectiveness in identifying and preventing fraudulent transactions. With the help of these KPIs, we can ensure that our ML model is continuously improving and responding to new patterns of fraudulent activity. Ultimately, the goal is to stop credit-card fraud as effectively as possible while minimizing the impact on legitimate transactions.

Machine Learning Canvas

The following figure depicts a visual representation of the Machine Learning Canvas that has been filled out to summarize our ideas about the use of machine learning in detecting credit-card fraud at Acme Commerce Bank. It is a graphical representation of our thought process and the various aspects we have considered for this use case.

DECISIONS

How are predictions used to make decisions that provide the proposed value to the end-user?

Automatically block and notify if a credit card transaction is predicted to be fraudulent with high confidence.

Flag suspicious transactions for review by the bank's fraud detection team, who will use the machine learning model's predictions as one of the input to make a final decision on whether a transaction is fraudulent or not.

Sort fraudulent transactions by probability of fraud to prioritize high-risk ones, allowing for more efficient allocation of resources towards preventing fraudulent activities.

ML TASK

Input, output to predict, type of problem.

Input: A set of features for each credit card transaction. These features can be grouped into 3 main categories: Transaction information, Customer information, Transaction outcome.

Output: A binary classification prediction for each transaction, indicating whether it is fraudulent or not.

The task is a supervised learning problem and the dataset is unbalanced, with most transactions being non-fraudulent and only a small percentage being fraudulent.

VALUE PROPOSITION

What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?

Enhance Acme Commerce Bank's credit card fraud detection capabilities by developing a digital product that uses advanced machine learning techniques to accurately identify and prevent fraud in real-time, providing a secure banking experience for customers and various intangible benefits for both the bank and its customers.

Targets:

- ○ Minimizing Financial Losses
- ○ Enhancing Customer Trust
- ○ Cost Savings
- ○ Competitive Advantage
- ○ Reduced Manual Workload
- ○ Improved Accuracy
- ○ Fewer False Declines
- ○ Reduced Verification Measures
- ○ Identification of Suspicious Consumer Behaviour
- ○ Evaluation of Customer Risk

DATA SOURCES

Which data sources can we use (internal and external)?

To develop an effective credit card fraud detection model for Acme Commerce Bank, data must be collected from various sources. It is important to keep collecting data regularly, both in the short term and long term.

Sources:

- ○ Credit card transaction data
- ○ Customer data
- ○ Fraudulent transaction data
- ○ Collaboration with other banks
- ○ Social media
- ○ User feedback

DATA COLLECTION

How do we get new data to learn from (inputs and outputs)?

To build an effective credit card fraud detection model, Acme Commerce Bank can collect data from various sources including:

- ○ Customer Relationship Management (CRM) System
- ○ Enterprise Resource Planning (ERP) System
- ○ Customer support systems
- ○ Data Warehouses
- ○ Social Media
- ○ Other banks
- ○ Static files

MAKING PREDICTIONS

When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?

- ○ The ML model generates predictions for each new credit card transaction based on previously learned patterns of fraudulent activity.
- ○ Predictions are made for every new transaction, and the number of predictions is equal to the number of decisions.
- ○ Predictions are generated in real-time, considering the time for prediction generation and feature value extraction after the transaction is processed.
- ○ The model must efficiently process and generate predictions for new transactions by extracting relevant features.
- ○ It is crucial to ensure that all listed features are utilized thoughtfully, efficiently extracted from the data sources, subjected to pre-processing and input into the ML model to accurately define output value propositions.

OFFLINE EVALUATION

Methods and metrics to evaluate the system before deployment?

The performance of the classification can be evaluated offline using certain metrics including:

- ○ True Positive, True Negative, False Positive, False Negative rate
- ○ Precision
- ○ Recall
- ○ F1 Score
- ○ AUC-ROC
- ○ Computing the cost savings or increase in cost incurred by using the system.

FEATURES

Input representations extracted from raw data sources.

The credit card fraud detection model will be using several features to identify potential fraudulent activity. These features in general can be clubbed into 3 parent categories including:

- ○ Transaction information
- ○ Customer information
- ○ Transaction outcome

BUILDING MODELS

When do we create/update models with new training data?

The credit card fraud detection model might require periodic updates with new training data every quarter or year due to various reasons including:

- ○ Increase in the rate of fraudulent activities
- ○ Changes in customer behavior
- ○ Performance degradation
- ○ New fraud patterns
- ○ Changes in the bank's policies or systems
- ○ Major shifts in the banking industry or changes in regulatory requirements

LIVE EVALUATION AND MONITORING

Methods and metrics to evaluate the system after deployment, and to quantify value creation.

- ○ False Positive Rate
- ○ False Negative Rate
- ○ Real-Time Performance Monitoring
- ○ Key Metric Tracking
- ○ Precision and Recall optimization
- ○ Feedback Loop
- ○ Detection Impact Monitoring
- ○ Cost-Benefit Analysis

Figure: ML Canvas for detecting credit-card fraud using Machine Learning (*Machine Learning Canvas*, n.d.)

References

AG, S. (n.d.). *Value Proposition Canvas – Download the Official Template*. Retrieved May 4, 2023, from https://www.strategyzer.com/canvas/value-proposition-canvas

Binary classification. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Binary_classification&oldid=1119691786

Confusion matrix. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Confusion_matrix&oldid=1148699071

Cost–benefit analysis. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Cost%E2%80%93benefit_analysis&oldid=1152653549

Credit Card Fraud Detection: How Machine Learning Can Protect Your Business From Scams. (n.d.). *AltexSoft*. Retrieved May 9, 2023, from https://www.altexsoft.com/blog/credit-card-fraud-detection/

Credit card fraud detection using machine learning. (n.d.). Retrieved May 8, 2023, from https://www.neuraldesigner.com/learning/examples/credit-card-fraud#DataSet

Data pre-processing. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Data_pre-processing&oldid=1138293751

Dorard, L. (2021a, March 9). From Data to AI with the Machine Learning Canvas (Part I). *Own Machine Learning*. https://medium.com/louis-dorard/from-data-to-ai-with-the-machine-learning-canvas-part-i-d171b867b047

Dorard, L. (2021b, March 9). From Data to AI with the Machine Learning Canvas (Part II). *Own Machine Learning*. https://medium.com/louis-dorard/from-data-to-ai-with-the-machine-learning-canvas-part-ii-b02c71067da8

Dorard, L. (2021c, March 9). *From Data to AI with the Machine Learning Canvas (Part III)*. Medium. https://medium.com/louis-dorard/from-data-to-ai-with-the-machine-learning-canvas-part-iii-868fe17b9be6

Dorard, L. (2021d, March 9). *From Data to AI with the Machine Learning Canvas (Part IV)*. Medium. https://medium.com/louis-dorard/from-data-to-ai-with-the-machine-learning-canvas-part-iv-2badcfa6c8fb

Dornadula, V. N., & Geetha, S. (2019). Credit Card Fraud Detection using Machine Learning Algorithms. *Procedia Computer Science*, *165*, 631–641. https://doi.org/10.1016/j.procs.2020.01.057

Fabricated Fraud Detection. (n.d.). Retrieved May 8, 2023, from https://www.kaggle.com/datasets/giladmanor/fraud-detection

False positive rate. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=False_positive_rate&oldid=1143988780

False positives and false negatives. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=False_positives_and_false_negatives&oldid=1131727577

Feature (machine learning). (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Feature_(machine_learning)&oldid=1150327846

F-score. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=F-score&oldid=1148225663

Incremental learning. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Incremental_learning&oldid=1139894476

Loss function. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Loss_function&oldid=1152207168

Machine Learning Canvas. (n.d.). OWNML. Retrieved May 9, 2023, from https://www.ownml.co/machine-learning-canvas

Performance indicator. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Performance_indicator&oldid=1152150191

Precision and recall. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Precision_and_recall&oldid=1149017180

Ravaglia, A. (2022, December 21). Imbalanced classification in Fraud Detection. *Data Reply IT | DataTech*. https://medium.com/data-reply-it-datatech/imbalanced-classification-in-fraud-detection-8f63474ff8c7

Receiver operating characteristic. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Receiver_operating_characteristic&oldid=1150954317

Reddy, D. K., Singh, D. T. C., Suresh, B., & Aparna, K. (n.d.). *Credit Card Fraud Detection Using Machine Learning Based Classification Algorithms*. *13*(05).

Risk analysis (business). (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Risk_analysis_(business)&oldid=1105639218

Strategyzer (Director). (2017, March 8). *Strategyzer's Value Proposition Canvas Explained*. https://www.youtube.com/watch?v=ReM1uqmVfP0

Supervised learning. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Supervised_learning&oldid=1144919032

Watson, M. (2018, March 8). Keeping Your Machine Learning Models Up-To-Date. *Center for Open Source Data and AI Technologies*. https://medium.com/codait/keeping-your-machine-learning-models-up-to-date-f1ead546591b