Case Study

Credit Card Fraud Detection

Industry – BFSI (Banking and Financial Institution)

Learning Outcomes:

After studying this case, the students will be able to:

1. **Understand how the overall Credit Card System works.**
2. **Identify challenges in fraud detection while dealing with real-world data.**
3. **Apply algorithms to analyze information from various data sources.**
4. **Create multiple machine learning models and choose the best fit as per the need.**

Understanding Credit Card Fraud

What is Credit Card Fraud?

According to an article written by Debra E. Ross,

> "**Credit card fraud**, act committed by any person who, with intent to defraud, uses a credit card that has been revoked, cancelled, reported lost, or stolen to obtain anything of value."

Fraud can be defined as an unlawful action which is intentional in nature for personal gain. A fraud done by someone without the knowledge of the card holder is termed as 'third-party fraud' whereas a fraud done by the card holder himself is termed as 'first-party fraud'. A common belief is that a fraud is mostly carried out by third-party and not first party; but this is not true. As per Experian fraud statistics the ratio of first- and third-party frauds are nearly equal.

Overall World Statistics

Organizations across the globe have adopted e-transactions in order to increase their productivity, expand their business and provide 24x7 services because of which the last two decades have witnessed an exponential rise in the use of card payment and e-tail. Along with the increase in the number of transactions there has been a tremendous rise in the fraud rate. Even if the fraud to genuine transaction ratio has been 3:100, the monetary loss is reported to be in billions of dollars. In this decade itself, financial institution and other government organizations have lost more than 50 billion dollars in card related frauds. These losses not only affect the day to day functioning of these financial institutions but also the day to day life of their customers.

History

To get into a little history; until 1980, frauds were manually checked in papers by domain experts. But in the 1990's, financial institutions started investing millions of dollars in building rule-based engines or expert systems. We will discuss about expert systems in this case study.

Note: Before we delve into the actual case scenario, we need to first understand the fundamentals of how a credit card system works and then study how a fraud detection system helps in curtailing credit card frauds. We would also look into the problems faced in designing a solution from a *developer* point of view.

Introduction to the Credit Card System

Who are the actors in a credit and debit card transactions?

- A **cardholder** gets a credit or debit card from an issuing bank and uses the account to make payments for goods and services.
- A **merchant** can be any business that accepts card payments in exchange for goods or services.
- A **merchant bank** creates and maintains **merchant accounts**. Merchant banks permit merchants to accept payments from the cardholders' credit or debit cards.
- **Payment processors** or Payment Gateways like Visa, MasterCard etc. are responsible for transmission of data from acquirer to issuer or vice versa.
- **Issuing banks** are the banks and other financial institutions that issue debit and credit cards to cardholders through the card associations.

A Brief Overview of the Transaction Cycles

For every single cashless digital transaction, an interaction is needed with each of these independent bodies - like acquirer bank, the payment networks, the issuer bank or the customer bank.

Credit card processing works in three distinct cycles: - the Authorization cycle, the Clearing cycle and the Settlement cycle.

From a financial institution point of view, a transaction has to go through all the three cycles in order to be termed as 'complete'.

First, let's look into the authorization cycle:

The Authorization Cycle

1. The Authorization Cycle is the first cycle in the process. It starts when the customer goes to a shop and provides his/her details to the shopkeeper. In this case the merchant is the shopkeeper.
2. The merchant sends the customer card information to its acquirer bank using Point of Sale device (POS). In this case not only the customer information but the merchant information is also sent to the merchant bank.
3. The acquirer bank receives the request and validates it. After validating the request, the request is enriched with data and sent to the payment network (VISA, MasterCard etc.)
4. The payment schemes validate the data and then send the request to the customer or issuer bank.
5. The issuer bank receives the request and checks for the customer information in their database. After checking for all the information, the issuer bank sends a code along with other customer details back to the payment scheme.
6. The payment network takes the request from the issuer bank and forwards it to the acquirer bank, who then sends the information back to the merchant stating the transaction was accepted or rejected.

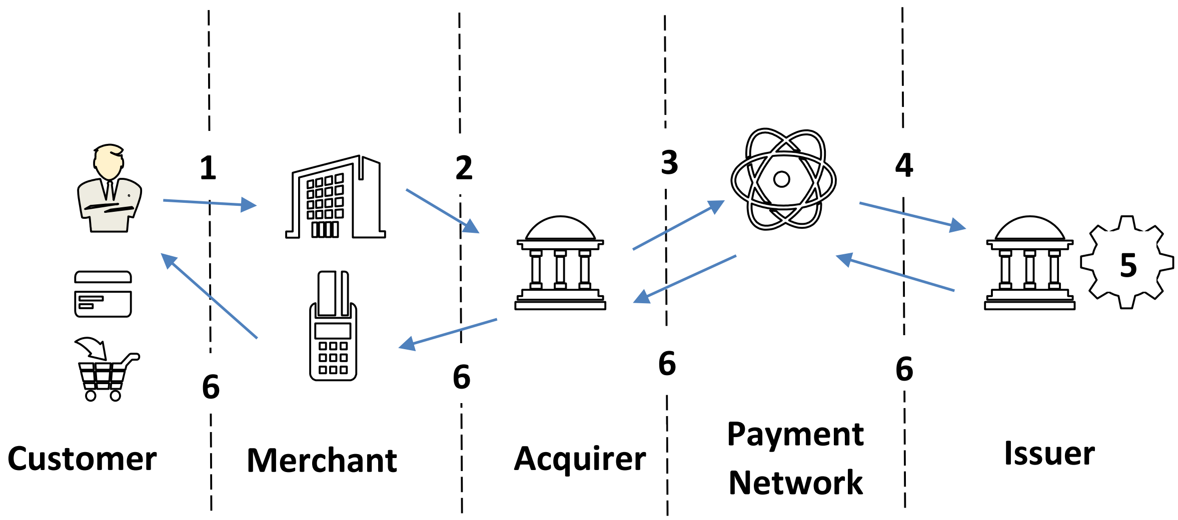

Figure 1.1: Authorization Cycle

The Clearing Cycle

In this cycle, a batch of transaction related information is sent from acquirer to issuer through the payment gateway. In the end, the payment network sends a reconciliation message to both the financial institutions. This cycle may repeat up to 3 times a day based on the participating financial instruction size.

The Settlement Cycle

The actual money movement happens in the settlement cycle where the issuer sends the amount to the acquirer and each of the independent bodies' - the acquirer bank, payment network and the issuer - with a processing fee which is also paid in this cycle.

For more details about the working of these cycles you can refer to:
https://www.mastercard.com/switching-services/our-solutions/switching.html

What is the source of data?

Financial transactions can be generated from:

- ATM
- Point of service device (POS) and
- Card not present (CNP) channels

Each of these transactions have unique codes which distinguishes them from each other. POS transaction occurs when a customer makes a purchase on the merchant's service device. Both, POS transaction channels and ATM transaction channels are card-present scenarios. Whereas a card-not-present scenario transaction includes e-commerce transactions and mail-telephone transactions where the customer and the card are not physically present but the card details and the customer details are given to the merchant. There exist different proportions of fraud in each of these transaction channels.

What is a Fraud Protection System?

In order to detect frauds, financial institutions develop **Fraud Protection Systems**. A fraud protection system consists of two layers:

The prevention layer makes sure the fraud doesn't happen or alerts in advance

When the prevention layer is by-passed, then fraud detection system comes into picture.

A fraud detection system comes into functioning when the fraud prevention system fails or is counterfeit by the fraudsters.

- Fraud prevention is the first layer in the fraud protection system. There exist a few basic techniques which are followed like firewall protection, external network detection, mapping of physical address with device id and data encryption through the networks.

- Fraud Detection system comprises of a set of predefined rules designed by the ISO 8583 standards. One of the main concerns in the finance industry is that for every falsification, the organization has to bear a cost for it.

The credit card fraud detection techniques are classified in two general categories: Misuse detection (fraud analysis) and anomaly detection (user behavior analysis).

Misuse Detection System

A Misuse system is similar to an expert system where the use cases provided by the domain expert or SME's is coded. These system can be as simple as a rule based system comprising of if-else conditions or market standard techniques like inventory optimization.

A rule-based system for fraud detection consists of the following modules:

Detection Rule Engine	Weight Configuration Engine	Velocity Rule Engine

- **Detection Rule Engine** were amongst the few most trusted detections system for e-tail. These engines are still used at present for detection.
- **Weight Configuration Systems** are typical expert based systems where each attribute is assigned a weight and a specific set of simple calculations are done. The domain expert has full control of the weights and importance placed on each attribute.
- **Velocity Rule Engine** relies on the concept of dynamic sliding window where data is assigned buckets as they are captured by the system. These buckets can comprise of either a fixed time frame length or fixed number of data stream bytes.

Refer to https://docs.cloudera.com/HDPDocuments/HDP2/HDP-2.6.5/bk_storm-component-guide/content/storm-windowing-concepts.html

Anomaly Detection System

Anomaly Detection system comprises of Machine Learning Algorithms or Data Mining techniques. Many white papers published provide a clear view of how data mining techniques like association, clustering and classification can be used for behavior analysis of customers and customer profiling. Machine learning algorithms of different nature are used for solving different types of scenarios in fraud detection system. The usage of the algorithm is based on the financial institution use cases and costing. Every card transaction coming into a live anomaly system is to be processed within a few hundred-mile seconds or else a penalty cost is to be paid by the serving bank.

What is the process of raising a fraud flag – Issuer?

In the first step, the customer details are verified from the customer database and run time metrics. This process is sometimes referred to as **Screening**. Screening process can be considered as the first step of fraud detection. Here the existence of the customer, OTB, transaction country and currencies are checked. After screening; the request is enriched with other customer details. The request is then sent to the Fraud Detection System (FDS) to check for fraud.

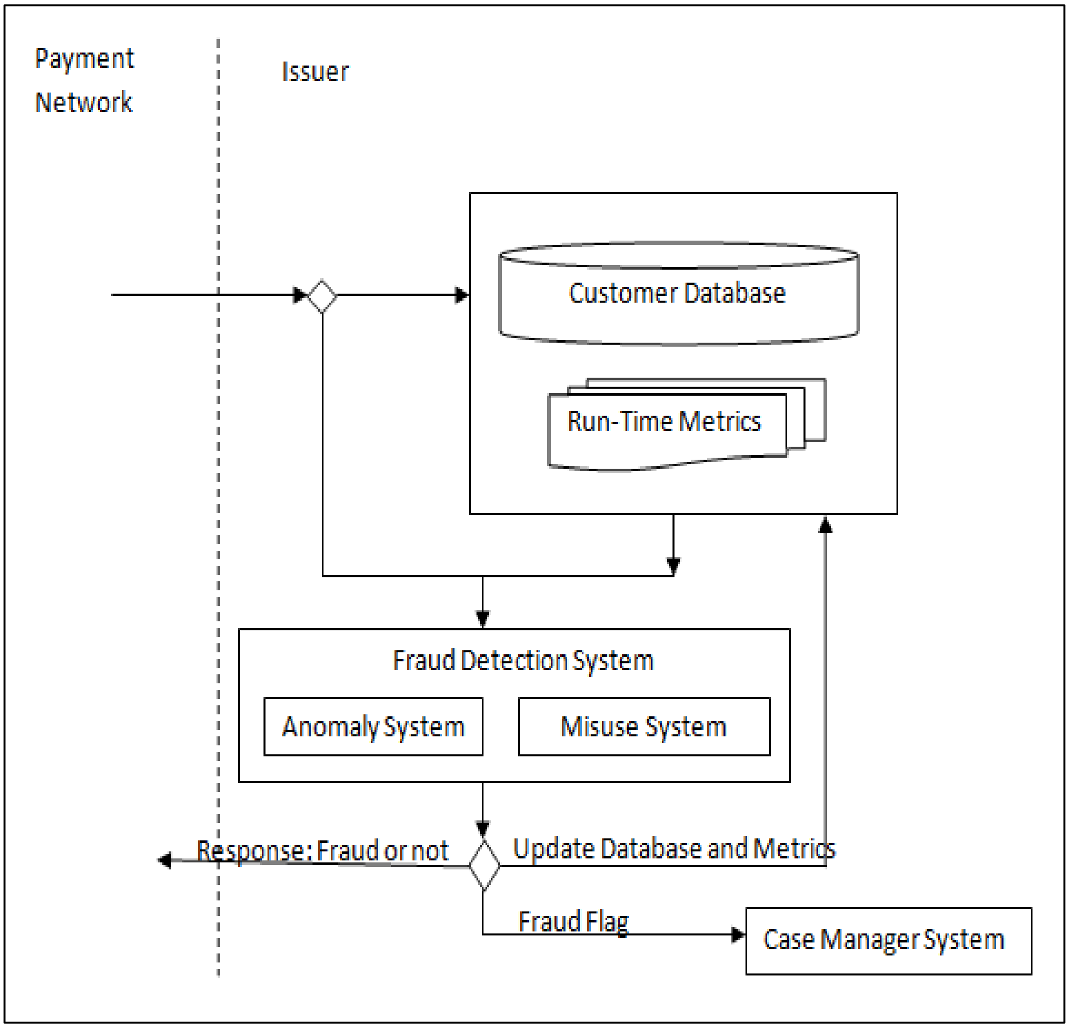

Figure 1.2: Process of raising a fraud flag

The output of the Anomaly system and the Misuse system is then sent to a heuristic system which decides whether to raise a fraud flag or not based on some pre-defined rules. The fraud flag request is then sent to a request distributer, which then sends the request back to the payment scheme stating the action performed on the initial transaction request. For every transaction request with a raised fraud flag; a case is created in the case manager system, the issuer bank then alerts the customer by either sending a message, e-mail to the customer or in some cases calling the customer for the transaction details.

Note: This is not a standard approach and may slightly differ upon institution to institution.

The Challenge of the CASE

Scenario of the Case

In this particular case, KPMG in India had to design an **anomaly detection system** for the authorization cycle which is capable of handling and processing approximately >100 transactions per second, considering the time limit on each transaction process is 300 milli seconds. This product had to be deployed in the FDS of the issuer bank and had to run in parallel with the misuse or rule base system.

Note: Machine learning cannot stand alone for such a sensitive use-case. For this we required fast data processing and data security as there were heavy penalties if the results were not delivered within the set time frame or if there is some data leak during the process.

Earlier all the electronic transactions coming from any of the medium irrespective of POS, ATM or E-com were in the xml or Json format (these formats are predefined market standards). The data had to be extracted in a structured format such that feature extraction & feature engineering can be done, and the data is sent to the ML Model for the result. This process of ingesting the data, processing the data and providing the result to the issuer needs to be completed within 300 milli seconds. In order to design such a powerful and fast system we had to bring in big data resources and work with them to achieve the goal. In the industry tools like Amazon Kinesis, Apache Kafka, Apache Storm etc. can be used for ingestion of the transaction data.

Roles of people involved in the Project Design and Development Phase

Every project ranging from a small proof of concept (POC) to deployment at production requires a team with different skill sets, expertise and experience. Every project follows a cycle, which starts from:

POC -> Prototype ->Pilot -> Production

A fully functional POC can sometimes be referred to as a prototype.

KPMG in India designed a fully functional POC with:

- An anomaly engine
- Misuse engine
- Case manager module and
- A UI

In a data science project of this kind, a normal team of 7-10 people are sufficient for building a fully functional POC. This team includes a Manager, Lead, Developers and Testers. You might come across situations where an individual might be asked to undertake multiple roles in order to increase the efficiency of the team.

1. **Technical Manager**
   A technical manager is required for the development of a project who has past development experience and knowledge of Machine Learning techniques. The technical manager is responsible for dividing the project into modules, planning tasks for each module, tracking timelines, meeting clients, understanding their requirements and gathering domain knowledge. For any data science project the role of a technical manager is very critical as he decides the proper flow of the project.

2. **Task Manager/ Project Lead/ Scrum Master**
   They are responsible for assigning tasks to the team members, checking the progress and challenges, coordinating with the SME for resolving issues and understanding the client's vision. Generally, hiring an SME become a challenging task. If an SME is not found, the company would look for consultants as an alternative. It is advisable to hire an SME if the project goes on for 2 to 3 months. The Manager, Scrum Master and SME all come together and brainstorm the requirements, design the architecture and decide on the key performance indicators (KPIs). These team members generally interact with the client, carry on research and perform activities like documentation, requirement gathering etc.

3. **Developers**

   Developers are the building blocks of the project. For the development of the misuse system – a rule building is done along with the SME and later the test are carried out to check the proper working of the scenarios. For the anomaly system, generally an experienced ML/DS developer is required who has 5-10 years of total experience in development and an analyst with around 2 years of experience and some knowledge of Machine Learning. An analyst also takes care of the database activities like storing the data, extraction of the data etc. We always need an analyst for a DS Project as they tend to act as backup support for a ML developer.

> **Note:** Remember the thumb rule, never let any individual take up an entire project module. Always give them a support developer and keep a backup resource ready.

For this case, we required a big data developer who will set up the big data ingestion, fast querying for misuse system and feature extraction/ feature engineering for anomaly system. And same goes for the development of the graphical interface. An UI/UX developer can be hired or instead a full stack developer can be hired based on the project budget. For testing such a prototype system or a POC, we usually do not require manual testers. Hiring people just for manual testing purposes has become outdated. In the POC stage, testing can be done by the SME and project manager as they have wide knowledge of the domain.

This is what makes a DS project special and different than the others; different skills, different expertise come together to build a final product. A data science project is a product of collaborative efforts of all the team members involved in the project.

Project Plan

Any successful data analysis project requires a strong project plan. This plan illustrates many basic requirements of the project, outlines the structure of the data, declares the objectives, describes the data sources and identifies the procedure which will be used to carry out the study.

The Project Plan document becomes a vital part of the project, because it shows the methods and purpose of the study to supervisors, grant writers and experts in the field.

Working on an amateur data science task or implementing an article is not the same as working in a professional working environment. Machine Learning is still a new technology for many, and that can make it hard to manage. Project Planner or Project Manager often do not know what questions should be asked and get everyone on the same page without overlooking something essential. Data revenue analyzed nearly 30 machine learning project and designed a checklist to avoid missing of important points during project planning and development.

- ✓ Project Motivation
  - What is the problem we are trying to solve?
  - What strategic goal is it linked to?
- ✓ Problem Definition
  - What output is expected?
  - What is the description of input data?
    - Number of fields
    - Number of entities
    - Are there any geo spatial data?
    - How is the data being captured?
  - What might be the relevant factors which might affect the output?
- ✓ Performance Measurement
  - On the basis of what KPI's will the performance be checked
  - Is there any documentation required?
  - What will be the KPI's?
  - What will the output do?
  - Are there reference solution (research article)?
- ✓ Timeline
  - What is the overall deadline of the entire project?
  - What is the timeline for each module?

  - By when will the first solution be provided?

- ✓ Contacts
  - Who is responsible for the project/ project manager?
  - Who is to be contacted for getting data/server rights?
  - Who will help provide the background knowledge?
  - Who is the SME for this project?
- ✓ Collaboration
  - Set weekly/daily updates between business people and developers?
  - Who is to involve? What should they learn?
  - Define where the codes are stored and how to access them.

To learn more about this checklist or how to go about it please refer:
https://towardsdatascience.com/the-machine-learning-project-checklist-d9ee6e33a2b2

Data Collection and Unavailability

The biggest challenge we faced were at the initial stages i.e. data availability. Financial data is very sensitive, and a real dataset is very hard to find. Without a real dataset designing a system becomes very difficult as real fraud scenarios are hard to get and understand. Most of the credit card datasets present online don't provide proper explanation or description of the nature of the data like which transaction cycle does the data belong to or at which institution level will the dataset be used. We were not handed the data entirely; just some use cases therefore we had to look for domain expert and gather understanding of what data is, how it looks like, what are the use cases etc.

Let us provide some simple use cases:
• More than 2 cash withdrawal from an ATM using a credit card within 15 minutes is considered a suspicious transaction in India.
• A small amount of transaction and then a sudden large amount of transaction raises suspicion in most parts of the globe.
• A transaction coming in between 12:00 – 4:00 am from a high-risk country leads to suspicion.

After gathering the use cases, doing aggressive research on use cases for that specific country and getting the idea of data, we created scripts for generating the training and test data.

For model development and testing we created two sample datasets. The datasets were created such that the transactions with high risk are the same in both the samples, but the number of transactions or data records differ in size. The sample 1 was the smaller sample dataset with around 70K records whereas sample 2 dataset had more than 50Lakhs records. The idea behind 2 sample datasets is that until the servers are fully ready for large scale model development, the data scientist/engineers will work on the smaller dataset.

Data Understanding and Preparation Concepts for the Case

Credit Card data is highly imbalanced or skewed i.e. the total number of fraud scenario are very less as compared to the legitimate transactions or non-fraud transactions. Machine learning aims to improve the accuracy without the knowledge of the class distribution. Using machine learning algorithms without balancing the class distribution will result in high error rates and unsatisfactory accuracy. This unbalanced distribution can be handled using data re-sampling techniques. The data re-sampling techniques can be broadly classified as oversampling and under sampling techniques.

Under-sampling aims at reducing the number of major class distribution. In our scenario, under sampling will eliminate the total number of non-fraud transactions in order to balance the fraudulent and legate transactions. Whereas over sampling aims at balancing the class distribution by replicating the instances of the minority class distribution.

Under-sampling is not widely used data re-sampling technique as it reduces the training data size, causing it to lose important scenarios or use cases which were initially required and therefore reducing the accuracy/performance. Oversampling technique can be sub categorized as random oversampling, cluster based oversampling and informed oversampling.

Random Oversampling	Cluster based Oversampling	Informed Oversampling

- In Random over sampling the instances of the minority class distribution are selected at random and replicated.
- Cluster based oversampling aims at creating equal size clusters for both the majority and minority class distributions. This re-sampling technique is based on K-means algorithm. It is based on the idea that both the majority class and minority class distributions comprise of multiple smaller clusters. Each of these clusters then oversample to a predetermined number n. By replicating the instances of each sub cluster results in the problem of over fitting to the training data. This problem can be solved using informed over sampling.
- In informed oversampling like synthetic minority over sampling approach or SMOTE approach is the most widely used informed sampling technique, the subset of the minority class distribution is taken using KNN. This technique helps tackle the imbalance class distribution without over fitting the model by creating smaller synthetic instances in the neighborhood of the minority class. Researchers use different data re-sampling techniques for balancing the data for their experiments.

Data Reduction

Card transactions for VISA/Mastercard/AMEX follow ISO-8583 format [3], which means there are more than are 127 fields. With the sensitive customer details and other client specific details – encrypted in alpha-numeric codes might lose their significance. Therefore it is considered a good practice to reduce the high dimensionality data such that the number of fields for the processing gets reduced but still retain its accuracy. The Categorization shows data reduction strategies which can be adopted to reduce the data for fast processing. Data reduction can be defined as reducing the data into a smaller subset without losing value. Data cube aggregation, attribute sub-selection, dimensionality reduction, numerosity reduction, discretization and concept hierarchy generation methods are among the few data reduction approaches which will be discussed in this section.

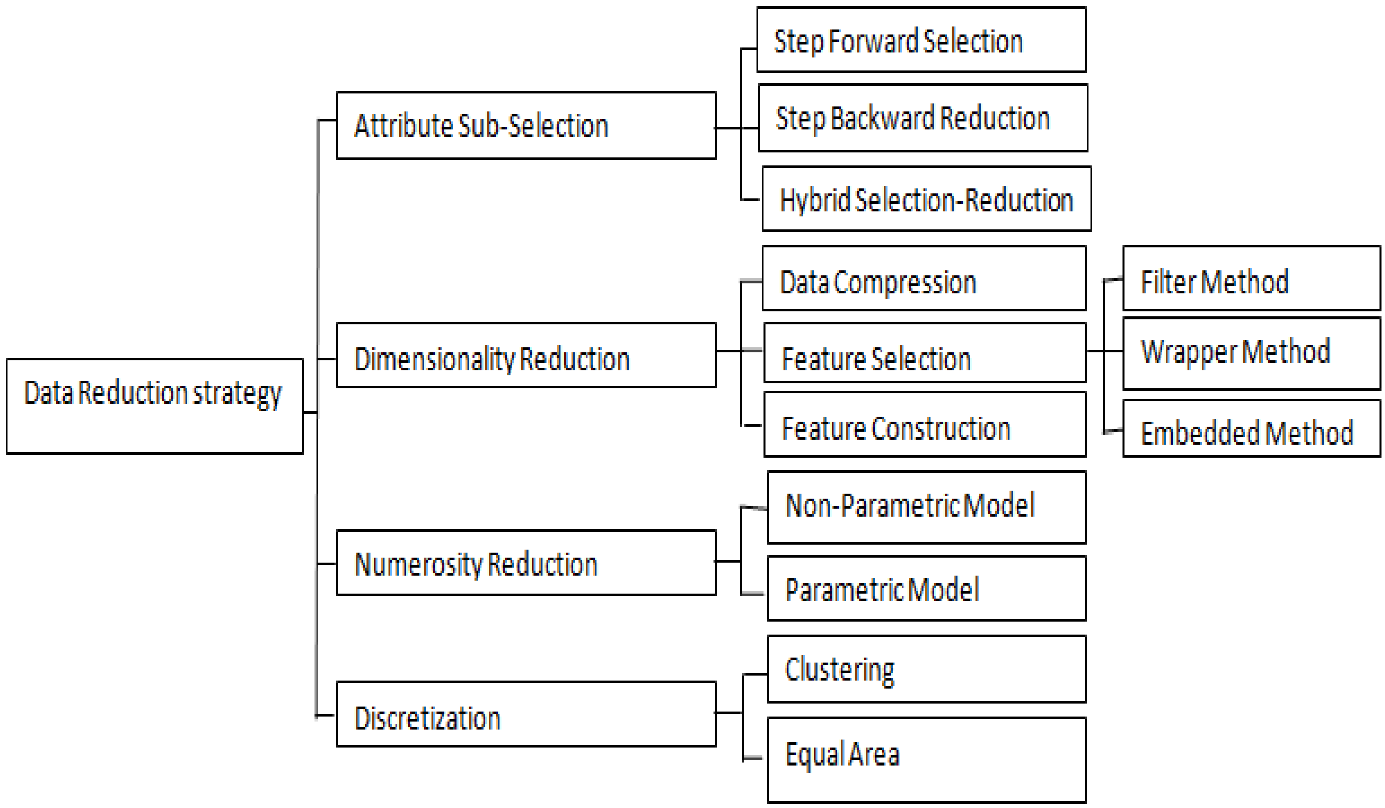

Figure 1.3: Types of Data Reduction Strategy

Attribute subset selection method consists of techniques used for selection of a smaller subset of variables to train the model. Attribute subset selection method has been subcategorized as step forward selection, step backward elimination, hybrid selection-elimination and decision tree induction.

- In step forward selection, the best-chosen attributes are added to the reduced subset.
- Backward elimination considers all the attributes of the initial dataset and starts eliminating the least important attribute in each iteration.
- Hybrid selection-elimination takes a two-way approach where the best attribute is selected, and the least important attribute is eliminated in each iteration.

Attribute subset selection method is aggressively used in our use case in order to save time during pre-processing of data and raising a flag.

Eliminating variables is very risky, not only for our use case but for any real-world problem. Therefore, don't just go around trying to avoid variables based on variable importance, it is always advisable to consult the SME's before taking such a step.

To know more please refer to learn about types of data reduction strategies: https://medium.com/@m96mridula/dimensionality-reduction-667438932a07,

https://www.electronicsmedia.info/2017/12/26/data-reduction-strategies-in-data-mining/

The Solution

ML Model Development

KPMG in India used both black box and white box ML models in the quest for finding the best fit. For those who don't know the difference between white box and black box machine learning models- Black box model are machine learning techniques or algorithms in which the result reasoning cannot be backtracked whereas in the white box machine learning model there are methods/ algorithms in which results can be backtracked.

To explain this further - in this case if we used Artificial neural network the backward propagation would have in general reduced the error providing more accuracy and reliability but if we needed to justify or find out the reason of a transaction being fraud we couldn't back track the result. In the case of a decision tree we can find the reason for which a transaction was termed fraud but with stacked decision tree structure like random forest it becomes very difficult. Therefore, we call decision tree white box and random forest black box techniques.

Now getting back to model selection and evaluation -We built multiple ML models like regression models, decision trees, ANN, super stacked model structure, random forest, XGboost etc. and evaluated them.

Regression models have been known to be widely used in the finance industry for solving most of the problems, this was not for our case as the model did provide expected accuracy as compared to other linear models.

Model tuning and result observations were done multiple times by different samples for testing. We also went on following a totally different approach by creating a new data set with some lakhs of records and tested these already built models.

During the evaluation, Random Forest and ANN model provided a higher MCC value and a Mean square error, making them more suitable for our scenario. Usually ML model's performance is evaluated using accuracy, precision etc. In some research articles model evaluation for this use case will be based on MCC or Matthews Correlation Coefficient Score. It is considered a good practice when you have an unbalanced binary classification problem, where the accuracy metric can't estimate well if the predictor is accurate. It is a contingency matrix method of calculating the Pearson product-moment

		Predicted	
		Non-Fraud	Fraud
Actual	Non-Fraud	**TN**	**FP**
	Fraud	**FN**	**TP**

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Similarly Log-loss or Logarithmic Loss can also be used in order to understand cross entropy for a binary classifier or the divergence of our model from being the perfect model. You need to evaluate your models based on complex mathematical functions and not just accuracy, precision or recall values.

ML model	MSE	Log-Loss	MCC
ANN	0.31	0.61	0.74
Random forest	0.53	0.2	0.92
Super Stacked Model	0.2	0.3	0.77
Linear regression	0.6	3.12	0.66

**Please note these figures are Illustrative and not actual output.*

As mentioned earlier, real-world machine learning is different, even with the evaluation of model. Companies will buy your product believing that it will save cost.

In a cost sensitive scenario like ours the model gets evaluated based on:

- Business loss = False Positive (Non fraud flagged suspicious) = sum of all transaction amount

- Financial loss = False Negative (Fraud flagged non-fraud) = sum of all transaction amount

- Admin charges = sum of all transactions predicted fraud (TP, FP)

- Operational charges =0.2% of all transactions predicted non-fraud

Note: There is a way of evaluating your results! Initially try to see how much loss the organization is facing without using an anomaly system. And then use your model to check, how much the loss is getting reduced by using your product (by adding the business loss, operational charges and financial loss). Always remember in cost sensitive problems it's about the loss you can avoid rather than the profit you gain.

So, what about the admin charges? We studied the process of raising the fraud flag i.e whenever a case is created the customer service representative sends a warning and sometimes calls the customer. In such cases, the customer service representative should be able to state the reason for the decline transaction which is not possible with these black-box models.

Type of loss	Before Model Deployment	After Model deployment
Business loss	297273	169532
Financial loss	452873	327498
Admin Charges	175K	179K

**Please note these figures are Illustrative and not actual output.*

So, have we made a model which can enable us to backtrack our cases? No! Linear regression, logistic regression, decision tree models if used can provide backtracking easily, but their performance is typically not what we want for our system, therefore we look into complex black box models. With Blackbox models like Adaboost, Random Forest, Artificial Neural Network we get high performance, but we need a reason for every flag raised in order to calculate the admin charges.

Finding a way to at least provide a reason becomes our next task. After brainstorming and research when we couldn't find a method to backtrack the reason, we tried looking for interpretability concepts. Interpretability concepts is similar to variable importance which is quite similar to what we do during EDA. The concept relies on using local linear approximations for our complex machine-learned models to derive explanations is also known as local interpretable model-agnostic explanations (LIME).

There exists another popular technique Leave-One-Covariate-Out (LOCO). But there were some challenges with this technique. Like LOCO, and LIME, both have shortcomings with respect to accuracy, consistency, or speed. LOCO works by analyzing individual feature at a time making the number of features slower and more expensive to run as compared to LIME. Both only work with the inputs and outputs of the model which means they have no access to look at the internal structure of the model. For our admin charge calculation and opening a case we do not require high speed therefore LIME fits our need of interpretability.

To know more you can refer to:
https://www.groundai.com/project/visualizing-the-feature-importance-for-black-box-models/1

The Outcome

This project success was a sum up of experience, expertise, teamwork and patience. This project was an eye opener especially for those team members who were fresher's in this field. This project made us realize how big the scope of data

science in the real world is and that there is no fixed approach to solve a data science problem. The approach will change as the project moves forward.
Our system was able to reduce the financial loss or "fraud transactions flagged non fraud" by 30%. Even though the admin charge was marginally affected, the manpower used for tracking the raised flag was however reduced. The FDS not only helped the organization by reducing the financial and business loss but also helped in providing near-real time data visualization and monitoring, which in turn made it easy to get a detailed distribution of the transactions coming into the system and the reasons of fraud flag raised. It enabled the management team to create new rules and policies in order to reduce the risky transactions. With the implementation of such a system there comes an added advantage - the power of transparent view, which every top level band of the organization wants. The deployment also helped in identifying new unusual behavior which was previously not encountered.

The full-fledged implementation takes time and tremendous management effort / focus for a cost sensitive scenario like this. So it's important to understand that a data science project is more than just machine learning code, it is working of different organizational components/concepts in a way to reach its goal.

General points to remember

Implementation of machine learning and data science in real world is very different as compared to what you learn in research articles and colleges. When the implementation is for a company which involves processing of real data then things become more complicated; as each wrong prediction will directly cost the company and its customers. Along with this many external factors need to be observed like fashion trends, political issues, government policies etc. There might be a case where-by the time your model gets ready for production and deployment the real-world scenarios/external factors might have aggressively changed and you have to redo some of your work.

The redoing of the entire model is rare and is considered a complete model failure which can bear a huge cost to the company if it fails during production. But this usually happens when the requirements are not discussed and noted correctly by the business analyst or if the requirements were not communicated properly to the data scientist/data engineers. Our team also faced this situation. During the development of the Proof of Concept (POC) and the project prototype at initial stages we had to redo the development of the model. It surely doubled our efforts at various phases but it is always better to get these things figured out during prototype development.