## Project: Data Science Use Case

(DLMDSPDSUC01)

A Use Case on

# Chargeback Fraud Detection with Machine Learning at Airbnb

Author: Kaushik Puttaswamy
Matriculation No: 321150196

# Contents



# 1.Introduction

Airbnb is an online platform that connects people who rent out their homes with those who need a place to stay. Therefore, during payment via online mode, *Airbnb must detect chargeback fraud since roughly two million guests stay in Airbnb-listed houses in 191 countries at any given time.*

# 2. Problem Definition

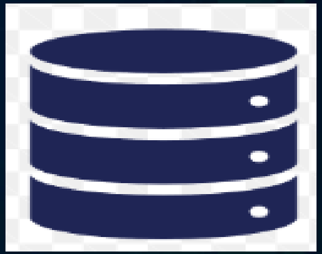
Chargebacks are transactions that are made by unauthorized users using credit cards that have been stolen, and they are common in online businesses. Airbnb takes on the full cost of chargebacks in order to avoid shifting the financial risk to the hosts. Both *revenue and customer trust* will be lost as a result of this.

# 3. Proposed Solution(Value Proposition)

It is important to develop a *machine learning canvas* as a necessary step to *fight against financial fraud prior to the final development of the ML model.* Using these machine learning techniques with the main goal of reducing *Airbnb's own exposure to chargeback fraud.*

# 4. Key highlights of ML canvas headers

**4.1 ML task:** The task of the *classification ML model* is to determine if the transaction is real or fraudulent by taking inputs such as the payment platform, currency, payment method, payment country, payment amount, and so on.

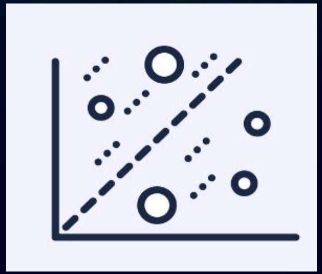


**4.2 Data source:** In this unbalance classification problem, it is important use both *internal and external* data sources.

- Internal data source: Transaction data, User data, Chargeback history.
- External data source:

1) https://www.neuraldesigner.com/files/datasets/creditcard-fraud.csv - this dataset includes legitimate transactions and contains 11 features about 3075 payments.

2) https://www.kaggle.com/datasets/dmirandaalves/predict-chargeback-frauds-payment - indicates whether the transaction was detected as a chargeback.

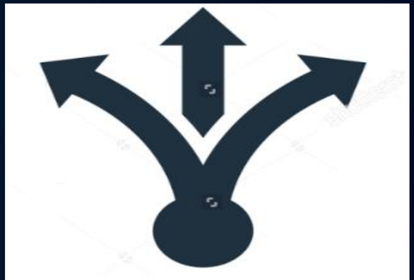


**4.3 Collecting Data:** Collecting new transactional data via the customer's transaction history, which is usually recorded through the platform that the company (Airbnb) uses to manage their website. Furthermore, redefining the data on new transactions and online bookings that should *contain a payment, features, and a representative number of chargebacks* is requested.

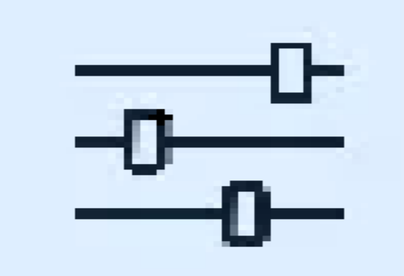
Cont'd

**4.4 Decision:** Predicting on a weekly basis from the time gap since the previous prediction gives an advantage to *taking immediate action to avoid or reduce the chargeback and to allocating resources* more efficiently. Based on prediction, it is essential to make decisions that bring value to Airbnb, including:

- *Filter out customers who are not predicted to fraudulent, and anomalous customers*
- *Sort customers by descending chargeback probability times monthly revenue loss*
- *Target the first K customers in the list*



**4.5 Making Prediction:** Models must make predictions on new input datasets after being deployed. These input datasets must be obtained from *available data sources (databases, web scraping, etc.)*. Additionally, it is imperative to predict fraud on a *weekly basis* because this enables Airbnb to *analyze* massive volumes of online transaction data *to find potential events and opportunities* before they occur so that *better decisions* can be made. Additionally, feature development *doesn't take too long because new online bookings and transactions are all taking place simultaneously.*

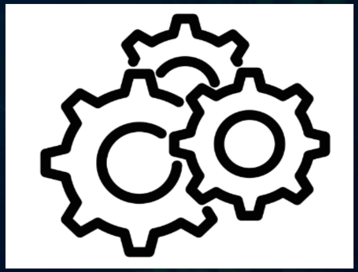


**4.6 Building model:** In a variety of cases, including *new fraud patterns, changes in the distribution of the data, the addition of new data sources, and performance degradation*, the ML model needs to be updated with new training data. Additionally, there are some strategies to keep a machine learning model up-to-date, such as the *collection and analysis of new data, monitoring model performance, using user feedback, staying consistent with industry trends, using multiple detection methods, and continuously improving and refining the model.*

# Cont'd

**4.6 Features:** Input representations extracted from available raw data sources includes *merchant_id, avg_amount_day, transaction_amount, is_declined, number_declines_day, foreign_transaction, high_risk_country, daily_chbk_avg_amt, 6m_avg_chbk_amt, 6m_chbk_freq, is_fradulent*. In addition, it is important to consider some of the other categorical features such as *payment platform, currency, payment method, product ID, product group, most called country, users last bought service, product subcategory, area for use of product, product country, response type, IP connection type, user's country, payment country*, and numerical features such as *payment amount, days since last payment, IP fraud score*, etc.

**4.7 Offline evaluation:** To evaluate the performance of our classification model *Confusion matrix, precision, recall, F1 score, and AUC-ROC* are all ML model performance metrics that address the imbalanced data classification issues. Because missed fraudulent transactions can cause significant losses for Airbnb and chargeback fraud can have detrimental financial and reputational effects, it was recommended that recall optimization be used to catch all fraud, regardless of whether there are false alarms.



**4.9 Live evaluation and monitoring:** KPIs like *Precision and Recall, False Positive Rate, False Negative Rate, Average Time to Detect, Reductions in Chargeback and Detection Rates* and others are frequently used to track DSUC performance after ML model deployment.

# 5. Filled-in version of the Machine Learning Canvas

### Decisions

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

Computing predictions for all payment transaction data that had occurred recently (a couple of weeks ago) and that are used to make decisions:

•Filter out customers who are not predicted to fraudulent, and anomalous customers.

• Sort customers by descending chargeback probability times monthly revenue loss.

• Target the first K customers in the list.

### ML Task

Input, output to predict, type of problem

Input: Transaction details such as Payment platform, Currency, Payment method, Payment country, Payment amount, and so on.

Output: Classification task- Predict whether a fraudulent online booking or not.

### Value Propositions

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

Developing a machine learning canvas using various machine learning techniques with the goal of reducing Airbnb's own exposure to chargeback fraud; moreover, the main aim of this task is to provide a reliable and effective ML canvas in which the final outcome of the predictive system should enable end-users to detect and prevent fraudulent chargeback requests, minimize revenue losses, and maintain a positive business reputation.

Objective:

• Determining the data sources (internal and external)

• Explaining the amount of new data that needs to be collected

• Describing a set of decisions to provide desired value to the end user

• Stating an approach to regularly analyze and retrain a model

• Representation of different input data fields (features)

• Explaining the evaluation metrics and methods for measuring the proposed model's performance

### Data Sources

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

• Internal: Transaction data, User data, Chargeback history.

• External:

(https://www.neuraldesigner.com/files/datasets/creditcard-fraud.csv) This dataset includes many legitimate transactions and contains 11 features about 3075 payments.

(https://www.kaggle.com/datasets/dmirandaalves/predict-chargeback-frauds-payment) This data source containing one month of raw credit card transactions.

### Collecting Data

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

Collecting new transactional data via the customer's transaction history, which is usually recorded automatically through the point-of-sale system or the platform that company (Airbnb) uses to manage their website. Furthermore, redefining the data on new transactions and online bookings that should contain a payment, features, and a representative number of chargebacks is requested.

### Making Predictions

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

Making predictions on new inputs when the trained model is deployed in a production environment or used for inference on a dataset not previously seen during training requires the new input dataset to be obtained from available data sources (databases, web scraping, etc.).

Featurizing is instantaneous since feature development doesn't take too long in our use case because new online bookings and transactions are occurring simultaneously. Finally, making predictions for forthcoming events can also be relatively quick, because the dataset size is relatively small since we are collecting a certain week of transaction data and it is comparatively less complex.

### Offline Evaluation

Methods and metrics to evaluate the system before deployment?

Metrics to measure classification performance offline are:

• Confusion matrix
• Precision
• Recall
• F1 score
• AUC-ROC

### Features

Input representations extracted from rawdata sources.

Features that are significantly associated with results among all features, and these selected features allow a model to perform better.

Example: Payment platform, Currency, Payment Method, Product ID, Product Group, most called country, users last bought service, Payment amount, Days since last payment etc.…

### Building Model

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

ML models for chargeback fraud detection may need to be updated with new training data in a variety of cases, including:

• New fraud patterns
• Changes in the data distribution
• New data sources

Performance degradation keeping a chargeback fraud detection ML model up-to-date is important to ensure its effectiveness in detecting and preventing fraudulent activities. Here are some strategies to consider:

• Collect and analyze new data
• Monitor model performance
• Use feedback from users
• Stay up-to-date with industry trends
• Use multiple detection methods
• Continuously improve and refine the model

### Live Evaluation and Monitoring

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

• Precision and recall

• False positive rate

• False negative rate

• Average time to detect

• Reduction in chargeback rate

# 6. Conclusion

In conclusion, chargeback fraud is a serious problem for businesses that process online payments. To reduce Airbnb's exposure to chargeback fraud in this use case, we used the ML canvas. In this regard, we followed the procedure outlined below:

- We have to choose internal and external data sources to train the ML model.
- In order to predict chargebacks based on the most recent data, we need to collect recent transaction data from the available sources.
- Decisions are made by targeting and removing those fraudulent and anomalous customers based on the predicted results.
- After ML model deployment, it is time to make predictions on new data using the week's or month's worth of new data that is currently available.
- It is important to represent some of input features that are extracted from the selected data sources.
- In a variety of situations, the model must be updated with the most recent fraud trends and new data. It is also essential to employ strategies to prevent the model's performance from degrading.
- The model has to be evaluated to measure its performance using methods and metrics that address the imbalance issue.
- After deployment the model needs to monitor and quantify the value creation by using desired metrics.

Therefore, this process concludes the overall development of ML canvas in context to our chargeback fraud detection use case to fight against revenue loss at Airbnb.

## CONCLUSIONS

# 7. References


- Agrawal, S. K. (2021, July 20). Evaluation Metrics For Classification Model | Classification Model Metrics. *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2021/07/metrics-to-evaluate-your-classification-model-to-take-the-right-decisions/
- *Credit card fraud detection using machine learning*. (n.d.). Retrieved March 19, 2023, from https://www.neuraldesigner.com/learning/examples/credit-card-fraud#DataSet
- Dima. (2022, January 20). *How to Use Machine Learning in Fraud Detection and Prevention*. Intellias. https://intellias.com/how-to-use-machine-learning-in-fraud-detection/
- Dorard, L. (2021, March 9). From Data to AI with the Machine Learning Canvas (Part III). *Own Machine Learning*. https://medium.com/louis-dorard/from-data-to-ai-with-the-machine-learning-canvas-part-iii-868fe17b9be6
- Fick, O., & Gunther, T. (n.d.). *Detecting Chargebacks in Transaction Data with Artificial Neural Networks*.
- *Fighting Financial Fraud with Machine Learning at Airbnb*. (n.d.). InfoQ. Retrieved January 14, 2023, from https://www.infoq.com/news/2018/03/financial-fraud-ml-airbnb/
- Kalirane, M. (2023, January 20). Ensemble Learning Methods: Bagging, Boosting and Stacking. *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2023/01/ensemble-learning-methods-bagging-boosting-and-stacking/
- *Machine learning for fraud detection*. (n.d.). Ravelin. Retrieved January 14, 2023, from https://www.ravelin.com/insights/machine-learning-for-fraud-detection
- Machine Learning Metrics: How to Measure the Performance of a Machine Learning Model. (n.d.). *AltexSoft*. Retrieved January 25, 2023, from https://www.altexsoft.com/blog/machine-learning-metrics/
- Pant, H., & Srivastava, D. R. (2015). *A SURVEY ON FEATURE SELECTION METHODS FOR IMBALANCED DATASETS*.
- 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
- Shmueli, B. (2022, July 21). *Multi-Class Metrics Made Simple, Part II: The F1-score*. Medium. https://towardsdatascience.com/multi-class-metrics-made-simple-part-ii-the-f1-score-ebe8b2c2ca1
- Ucar, M. (2020). Classification Performance-Based Feature Selection Algorithm for Machine Learning: P-Score. *IRBM*, *41*. https://doi.org/10.1016/j.irbm.2020.01.006
- 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
- Wei, Y.-C., Lai, Y.-X., & Wu, M.-E. (2022). An evaluation of deep learning models for chargeback Fraud detection in online games. *Cluster Computing*. https://doi.org/10.1007/s10586-022-03674-4