Data Science Use Case (DLMDSPDSUC01)

Finalization Phase

Portfolio: Chargeback fraud detection with machine learning at Airbnb

Abstract:

Chargeback fraud is becoming a more serious problem for many companies, particularly those that carry out business online. Traditional rule-based fraud detection systems frequently fall short of keeping up with fraudsters' constantly evolving strategies. We used a real scenario of chargeback fraud at Airbnb for this use case analysis. Airbnb is an online marketplace that connects people who rent out their homes with those who need a place to stay. When a company bears the full cost of chargebacks rather than shifting the financial risk to their hosts, it is caused by unauthorised users using stolen credit cards to make online payments. Therefore, the credit card company requests a refund from the merchant (Airbnb). As a result, creating a machine-learning canvas serves as a necessary step in the fight against financial fraud. In view of this, we begin by giving an overview of the ML canvas, a machine learning project framework that can help ensure successful results. We then go into detail about how we used the canvas for a chargeback fraud detection use case by illustrating various subject areas. The value proposition is the first step, and its primary goal is to offer end users a trustworthy and efficient ML canvas that, in turn, will enable them to identify and stop fraudulent chargeback requests, minimise revenue losses, and maintain a positive business reputation. Inputs like payment platform, currency, payment method, payment country, payment amount, and so on are further defined as part of the ML model's task. Thereby producing a classification result indicating whether the transaction is genuine or fraudulent.

The use of internal data sources, such as transaction data, user data, and chargeback history, is an essential part of any ML model development. Additionally, we are referring to two openly accessible external data sources that deal with credit card fraud. The csv dataset file contains 11 features regarding 3075 payments, and the second dataset contains the most recent trends and methods in fraud. Additionally, new information must be gathered for the model to be kept updated from the resources that are available, like the database of transaction data for Airbnb. Additionally, to enable our model to recognise current chargeback patterns, the data must include a representative number of chargebacks.

As predictions are merely information and have no independent usefulness, they must be transformed into value before they can be delivered to the end user. Additionally, to make things more concrete in terms of what to do with prediction, consider when decisions should be made, it makes sense to predict chargebacks on a weekly or monthly basis because Airbnb can take immediate action to avoid or reduce the chargeback, more efficiently allocate resources that are likely to result in a chargeback in the near future, and we can potentially achieve higher accuracy and reduce false positives and negatives by focusing on these areas. In addition to predicting chargebacks in the coming weeks of new datasets, each incoming transaction should be rated for its potential for chargebacks. These facts can offer a more thorough picture of the transaction risk and assist in guiding longer-term plans and choices. Additionally, based on the ML model's predictions, we suggest making certain decisions, like eliminating customers who are not expected to be fraudulent and anomalous customers, sorting customers by decreasing chargeback probability times monthly revenue loss, and focusing on the first K customers in the list. The goal of developing intelligent systems may ultimately be to fully automate decision-making. However, it is advised to avoid total automation when we are just getting started with ML. Instead, we could make a list of all potential decisions, let the machine sort them according to predictions, and then let a human review and make the final choices.

A new input dataset must be obtained from readily available data sources (databases, web scraping, etc.) when the trained ML model is deployed in a production environment in order to make predictions on new data that has not yet been trained. Variables (also known as features) in a dataset will have values in the form of numbers, categories, and/or text. Features should also be carefully chosen so that they can be used to define output-value propositions. The term "featurization" generally refers to this process. It is now necessary to input this data into the trained ML model, which has to have been repeatedly run until it reaches a relatively high level of accuracy in comparison to the testing set. Additionally, it is essential to predict fraud on a weekly basis because doing so enables Airbnb to analyse huge volumes of online transaction data in order to identify potential opportunities and events in advance and make smarter decisions.

The performance of the ML may deteriorate over time if it is not trained on new data that reflects current trends in fraud. As a result, the model needs to be updated in a number of situations, including the emergence of new fraud patterns, modifications to the data distribution, the creation of new data sources, and performance decline. Additionally, here are some strategies to think about when maintaining a chargeback fraud detection ML model. Monitor model performance, gather and analyse new data, utilize user suggestions, follow industry trends, employ a variety of detection techniques, and continuously enhance and improve the model.

The goal of feature selection is to find features that are significantly associated with results among all features, and these selected features allow a model to perform better. Furthermore, input representations extracted from selected external raw data sources include, 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 also numerical features such as Payment amount, Days since last payment, IP fraud score etc.

In the case of ML model evaluation, we considered metrics such as Confusion matrix, Precision, Recall, F1 score, and AUC-ROC. A Confusion matrix is a two-dimensional table that displays actual and predicted values, whereas an F1 score is a more sophisticated statistic that assists us in achieving more accurate results in unbalanced classification situations. Furthermore, when addressing the trade-off between precision and recall, it is important to consider Recall optimisation rather than Precision optimisation because in our chargeback fraud detection ML model, it is more important to catch all fraud, regardless of whether there are false alarms.

After successfully evaluating the prediction model prior to deployment, it is now essential to monitor the actual raw performance of the ML model in real-world scenarios after deployment. In our case, effective KPIs frequently used to monitor DSUC performance post-ML model deployment include Precision and Recall, False positive rate, False negative rate, Average time to detect, Chargeback rate reduction, and Detection rate. These KPIs aid in tracking the effectiveness of our ML model so that adjustments can be made as needed to effectively identify and stop chargeback fraud.