# MSc Information Studies
# Data Systems Project 2022-23

**Project Title**

*Fair fraud detection*

**Project stakeholder (name and/or organization)**

**Adyen** is the payment platform of choice for the world's leading companies, delivering frictionless payments across online, mobile and in-store channels. It is the only provider of a modern end-to-end infrastructure, connecting directly to Visa, Mastercard, and consumers' globally preferred payment methods. With offices around the world, Adyen serves nine of the 10 largest U.S. internet companies and many worldwide retailers. Customers include Facebook, Uber, Dafiti, 99taxis, Rappi and Spotify.

**Stakeholder contact details**

Name of organization: Adyen
Website: https://www.adyen.com
Address: Rokin 21-49, Amsterdam, NL

Name of contact: Andreu Mora
Email: andreu.mora@adyen.com

**Brief project description:**

The goal of this challenge is to develop a system that can be used to detect and block fraudulent payments in real time. While the primary purpose of the system is to minimize the number of fraudulent transactions approved (i.e. true positive rate) while maximizing revenue (i.e. low false positive rate), the solution also needs to consider the possible harmful impact on people and society, which needs to be defined, monitored and mitigated.

**Key challenge/problem:**

***Fraud detection*** Helping merchants detect and block fraudulent payments is an important part of our business, as allowing these payments results in costs due to fines, forced refunds and loss of goods. Here is a basic overview of fraud types in the payments space. If our risk system does not block a fraudulent payment, the issuing bank of the corresponding cardholder informs us of the fraudulent payment by sending a fraudulent dispute (i.e. a chargeback). If our risk system blocks a payment, we won't know whether the payment was actually fraudulent, as in that case the issuing bank does not send a fraudulent dispute. We can use fraudulent disputes to label historical payments for training and evaluating ML models. An important datapoint to consider is that the fraudulent dispute (the label) will only be known in an undetermined time window spanning from 1 to 90 days after the transaction was

approved. Schemes (VISA, MasterCard) mandate that in order to process within the network the chargeback rate is kept always at  a maximum of 3%. Not meeting this criteria would mean that a merchant cannot process through those schemes most likely implying bankruptcy.

***Useful metrics***

- *Fraud rate* (%): the amount of payments out of the total approved payments that will result in a chargeback or a notification of fraud.
- *Block rate* (%): the amount of payments out of the total received payments that the system will block for suspected fraud reasons.
- *Conversion rate* (%): amount of approved payments out of the total received payments. Note that some of the approved payments may convert into fraudulent disputes.
- *Total revenue* (EUR): total amount of approved volume processed by the merchant minus the chargeback costs
- *Chargeback costs* (EUR): the chargeback cost of the transaction is equal to the value of the transaction plus a standard fee of 15 EUR.

***Inferential scoring*** Adyen designs and deploys a system to combat fraud on-line and is exploring two possibilities: a business rule engine and a machine learning engine. Both systems have the same goal of reducing both the block rate and the fraud rate simultaneously leaning on the features presented in the example dataset provided.

***Real time*** The system needs to be deployed in real time for inference with a Service Level Agreement (SLA) of 200 ms. This includes the network latency as well as processing time.

***Data characteristics*** As an example, consider that Adyen had 500 TPS (Transaction Per Second) in 2021 and has been growing 50% YOY (Year-On-Year) for the last 3 years. Consider that the Big Data Platform stores information for the last 3 years. The provided CSV file is a random synthetic sample of the feature matrix (table) that we have stored in our Big Data Platform.

***Human interaction*** The system is used by expert risk analysts to both configure settings and to analyze results. Risk analysts value data informed reasoning and as such it is mandatory for the system to offer an analytics overview of the performance as well as being to explain why certain transactions were blocked or accepted and the reason behind it.

***Fairness*** When deploying such models, there is a real impact for people. A false positive means that someone who legitimately wants to pay is not allowed to do so. A bias in the model can unjustly cause us to disadvantage groups in society. Adyen holds high ethical standards for its products, and it is important Fairness is considered in the fraud checks.

# MSc Information Studies
# Data Systems Project 2022-23

Solving for Fairness is not a well defined problem. Features which provide discriminatory power might be considered fair in one context but unfair in another. A first step here could be to define what type of discrimination we want to avoid within the context of a payment fraud model. Additionally, not all sensitive attributes of a shopper will be observed during a payment, while they might propagate effects into attributes which we do observe and act on. This requires a strategy to deal with this in practice.

.

**Exemplar projects or additional reference materials (books, papers, products, URLS etc):**

- synthetic dataset representative for the task available at: http://grotius.uvalight.net/DSP2223/datasets/adjen-dataset.csv

Several framework and overview papers exists in the literature, as well as tutorials and libraries for fraud/anomaly detection, for explainability, fairness and bias detection.

**Goals or key criteria qualifying a relevant *exploration*:**

The students identify and develop one or more of these dimensions:

- compare two or more competing solutions for anomaly detection (possibly adaptive, based on active learning, etc.) for the given task
- compare two or more competing solutions for XAI/fairness/bias analysis
- test architectures enabling the governance of data processing in order to support fair processing
- study the relationships between alternative control and visualization components on the interface with the actual system use (performance and impact)

**Suggested key requirements or success factors for an *implementation*:**

Real-time efficient calculation. User experience validation.

**Challenges or constraints envisioned (if any):**

**Any specific technical or content requirements:**

**Any additional comments or thoughts:**