Fraud Detection in Ride Hailing Apps - *A Product Manager’s Perspective*

You have been appointed to be the Data Science Product Manager of a ride-hailing app in South East Asia. You were tasked by your manager to help the regional operations team to reduce the number of frauds in the platform. You are assigned with a team of Data Scientists to figure out the best way to develop a model for fraud detection. After spending a month of gathering information, you are asked to share your plan to the regional operations management team.

**Presentation:**

Your task:

1. **Product Requirement Documentation (PRD) in Microsoft Word:** all your product design, ML model, features, planning, evaluation method, etc.

2. P**resentation Slides (ppt / pdf) :** During your presentation, usually the PRD will be the reference at most of the time since it’s long and detailed, not very efficient for presentation, thus please prepare some slides for the presentation.

**Stakeholder**:

* Regional Operations Management Team, a bunch of technical + non-technical people
* What is the relationship / flow of work btwn Regional Ops team and Developers?
  + Dev build the application
  + Ops team take the product to the front-end consumers

**Objective:** Devise a plan to **reduce number of Frauds on App**

**Goals in Objective:**

* **Develop a** **plan** to build a **working fraud detection ML model** for the regional ops mgmt team
* **Write & Communicate the plan to the stakeholder(s)** with simple-to-understand technical details

**Business Requirements for Product Requirement Documentation**

1. Product Design: Fraud Detection System
   1. **Workflow** behind the product
   2. **Architecture** (backend (sql server) / languages / front-end)
2. Define Fraud: Self-defined + State assumptions on why it is most common
3. Explain the objective of ML Model, and what kind of approach you will take to develop this model.
   1. High level explanation of what the BASELINE model and what the user will expect.
   2. What is the BASELINE performance for the BASELINE model that needs to be achieved?
   3. Iterative Process. There will be new features, new preprocessing steps and new models tested over time if the performance needs improvement
4. State what are the **data features** you will use **in the model**
5. **What kind of resources** you need **from the business team**.
6. Using a **Product Mgmt Approach,** explain the **short-term and long-term roadmap** for this feature and how you will **prioritize** the tasks
   1. **Timeline**
7. Define how you will measure (evaluate) and track the performance of the ML models.
   1. Define the metrics for evaluation
      1. Speed of query (complete in 10 seconds)
      2. Accuracy performance
   2. How can the performance metrics be improved?

* **As a Product Manager, I plan to deliver by A months**
* **Dataset is from B to C month..**
* **Plan for an MVP for release 1.0**
* **Communicate to the stakeholders that for 1.0, we need to release it on deadline X , and must start delivering value Y - Define deadline X and value Y**
* **Version Control the Data & Model**
* **Framework**

# Why is it important to us to detect fraud?

The key stakeholders in our businesses (drivers and passengers) are both being impacted greatly by fraud

# Is fraud common

1 out of 5 rides in Indonesia are fraudulent.

Up To 20% Of Rides In Southeast Asia Impacted

Internal research by Grab indicates that fraud in the ride-hailing industry in Southeast Asia, if left unchecked, can represent up to 20% of all rides. In other words, without stern commitments from ride-hailing companies, for every dollar invested in ride-hailing, as much as 20 cents could be misspent on fraudulent activities.

A stolen ride-hailing driver profile today is [worth up to US$30](http://fortune.com/2018/03/22/data-sheet-uber-dark-web-fraud/) on the black market, even more than stolen credit card information.

# **1: Definition of Fraud we are addressing today + Assumptions on why it is most common**

Reference: <https://www.todayonline.com/singapore/grab-outlines-ways-its-drivers-try-defraud-ride-sharing-firm>

The more **common cases of fraud** faced by Grab **stem from drivers**, rather than passengers, Mr Foo Wui Ngiap, head of trust, identity, safety and infosecurity at Grab, told TODAY.

More money is at stake for the drivers, whereas customers looking to cheat the system would tend to focus on getting yet another promotional code — which is not scalable, he added.

*Reference: https://www.todayonline.com/singapore/grab-outlines-ways-its-drivers-try-defraud-ride-sharing-firm*

**GPS SPOOFING:**

Use spoof GPS tools to pretend they are in a particular location, such as a shopping mall, so they can secure a ride ahead of drivers who are actually at the location.

Some drivers may use fake GPS tools and “rooted” mobile phones, which feature unconventional software allowing them to change their position arbitrarily without actually moving, to simulate driving behaviour and complete rides in order to game the system.

In some cases, drivers may use such apps to plant their location in the middle of a shopping mall when they are actually far away.

Passengers are then allocated to fraudulent drivers given their apparent proximity and experience longer waiting times, while honest drivers physically waiting in the area lose out.

“This has a cascading impact on how they mess up the ecosystem for everybody else. The customers will be angry and say that Grab does not know how to allocate rides. The honest driver is not earning money. And the fraudulent driver is collecting money and getting away with it,” Mr Foo said.

**FAKE ACCOUNTS TO ABUSE PERFORMANCE INCENTIVE SYSTEM**

Drivers setting up multiple phones and multiple accounts to artificially reach benchmarks to earn performance incentives

Bonus payouts from hitting incentive benchmarks, pegged to completing a certain number of rides, are also a prime target for fraudsters.

This involves the setting up of multiple phones and multiple accounts, and pretending to complete many rides with payment made in cash.

“At some point, even money flows are finite. If the company is giving out all their money to bad drivers, there will eventually be a trickle down — where other (honest) drivers will earn less,” Mr Foo added.

**DRIVERS USE FAKE APPS TO MAKE UNLIMITED CANCELLATIONS WITHOUT PENALTY**

Dodgy drivers getting hold of fake Grab apps to enable them to make unlimited cancellations of rides they do not want, without penalty

The company has also seen its app being reverse-engineered and tampered with. These non-sanctioned apps are obtained outside the official Google play store or Apple app store.

Driver use these fake apps to utilise features such as unlimited cancellations of passengers that they have been allocated but prefer not to fetch, without any penalties.

# 2: Explain the objective of ML Model, and what kind of approach you will take to develop this model.

[**https://medium.com/grab/why-anti-fraud-technology-is-the-secret-sauce-to-winning-ride-hailing-platforms-2ae60a12ab75**](https://medium.com/grab/why-anti-fraud-technology-is-the-secret-sauce-to-winning-ride-hailing-platforms-2ae60a12ab75)

[**https://eng.uber.com/advanced-technologies-detecting-preventing-fraud-uber/**](https://eng.uber.com/advanced-technologies-detecting-preventing-fraud-uber/)

[**https://www.youtube.com/watch?v=pIfDlQt6HPo**](https://www.youtube.com/watch?v=pIfDlQt6HPo)

[**https://www.youtube.com/watch?v=yjy90pdBLfo**](https://www.youtube.com/watch?v=yjy90pdBLfo)

[**https://www.youtube.com/watch?v=cOzLO37mkwk**](https://www.youtube.com/watch?v=cOzLO37mkwk)

# 3: State what are the **data features** you will use **in the model**, and **what kind of resources** you need **from the business team**.

# 4: Using a **Product Mgmt Approach,** explain the **short-term and long-term roadmap** for this feature and how you will **prioritize** the tasks

# 5: Define how you will measure and track the performance of the ML models.