

A photograph of a modern, open-plan office space. In the foreground, a man is seated at a desk, working on a laptop. The desk is equipped with a green ergonomic chair. In the background, other employees are visible at their desks, and the office is decorated with various items like a red floral arrangement and a gold balloon. The office has a high ceiling with exposed ductwork and modern lighting fixtures.

# DS-Bootcamp(Day-1)

## DS Case Studies



- To develop valuable insights, building confidence for decisions.
- The work may influence ***decisions worth billions or billions of small decisions***
- Data Science projects require teams, not sole players.

***Paco Nathan , Concurrent Inc. 2012***



Few decisions worth Billions?  
**SERIOUSLY?**



# WAKTU TEMPUH

2X PINDAH ANGKUTAN **VS** LANGSUNG



Dari **Kemanggis** ke **Gedung WTC**

source: googlemaps



**54**  
MENIT

**VS**



**15**  
MENIT

**&**

ORANG MEMPUYAI

**WAKTU LEBIH BANYAK**

UNTUK DIHABISKAN BERSAMA DENGAN KELUARGA & SAHABAT



**PENDAPATAN MITRA GO-FOOD**  
dari pesanan di sekitar **Jl. Jendral Sudirman**



## Keterangan Warna

Kurang dari Rp 1 jt	Rp 50 jt hingga Rp 100 jt
Rp 1 jt hingga Rp 10 jt	Rp 100 jt hingga Rp 200 jt
Rp 10 jt hingga Rp 50 jt	Lebih dari Rp 200 jt



# Dampak *bagi* mitra GO-JEK

Pengemudi kehilangan lebih dari

90.000

perjalanan/hari



PENDAPATAN **5.000** MITRA

**GO**  **FOOD** 



terancam  
**HILANG**  
sebanyak

**10** **MILYAR**  
**/BULAN**

# What about billions decisions?



# What about billions decisions?







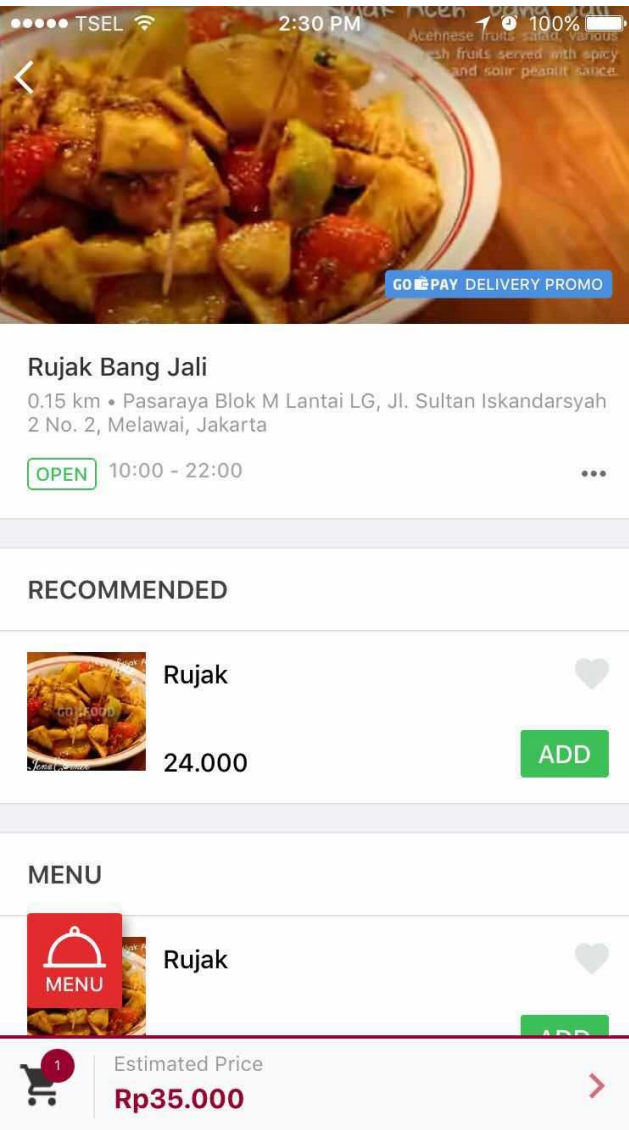




# Food Tagging with Deep Learning



# GO-FOOD has over **8M** dishes without tags



## Mochi Mochio, Pasaraya Grande

0.17 km • Pasaraya Grande, Lantai Lower Ground, Jl. Iskandarsyah 2, Melawai, Jakarta

Snacks

**OPEN** 10:00 - 22:00



## Nasi Pedas Endolita

## The Goal:

- Tag each dish
- Item level analysis
- Dish level recommendation feature

## The Challenge:

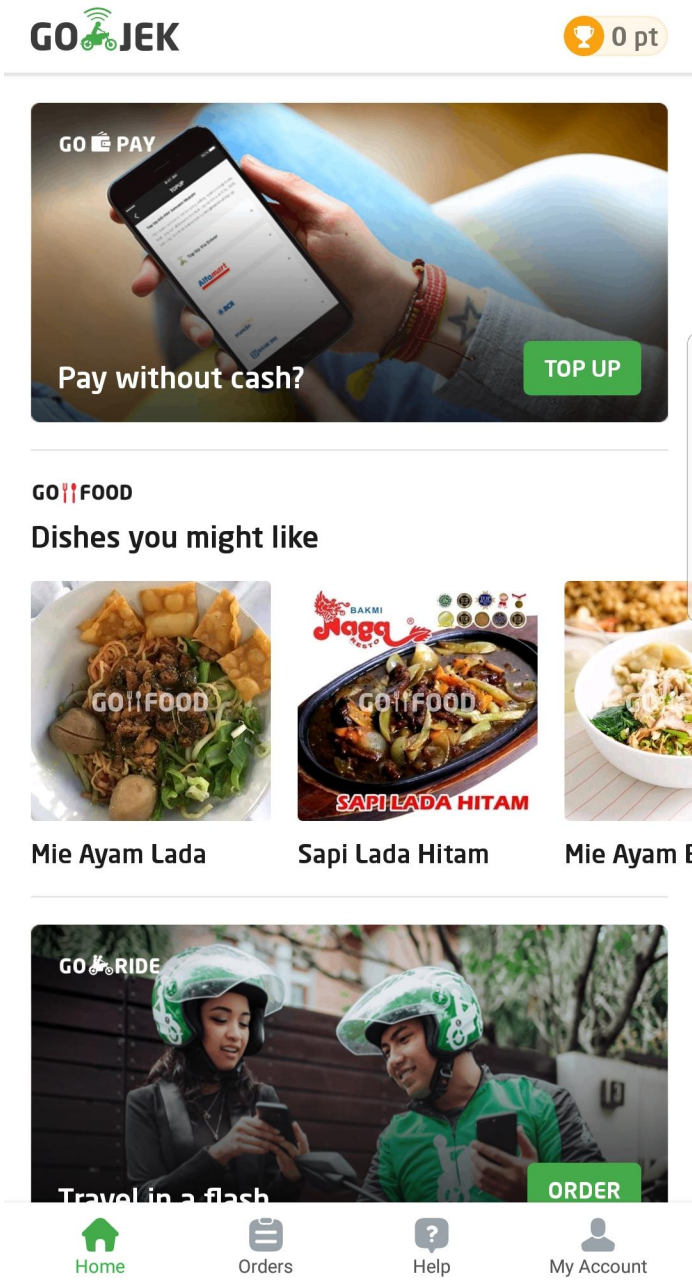
- Each item may have multiple tags:  
*e.g. Beverage, sweet, iced, coffee*
- Manual labels are messy and duplicates exist: *e.g. Dairy and Milk*
- Some tags appear very few times





# Content Based Dish Recommendation

# What dishes would customers like to eat?



## The Goal:

- Personalized dish level recommendation to users
- Increase discovery of new dishes based on past preferences

## The Challenge:

- No rich and clean text data on dishes
- Thresholding to prevent same dishes from different merchants from being recommended

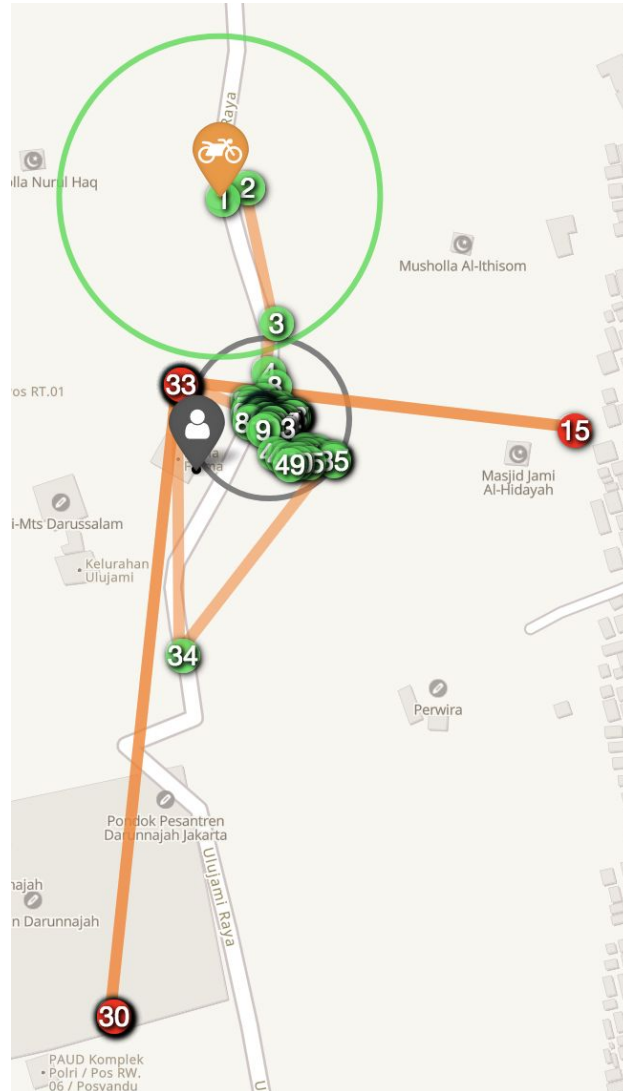
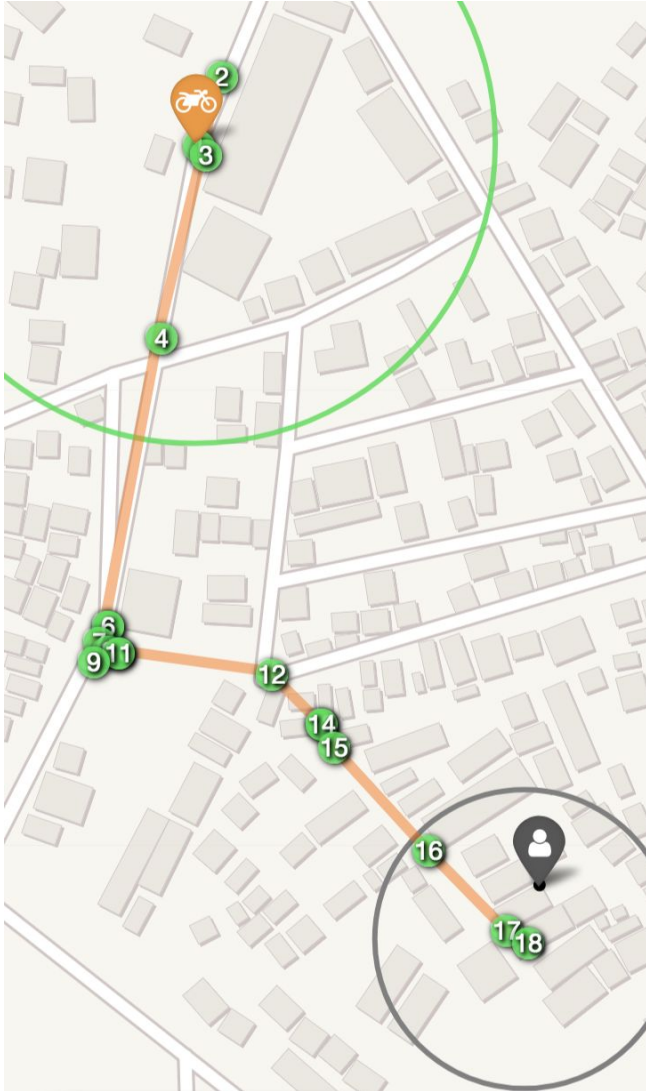


# GPS Quality Anomaly Detection





# GPS is a core data requirement of GO-JEK



## The Goal:

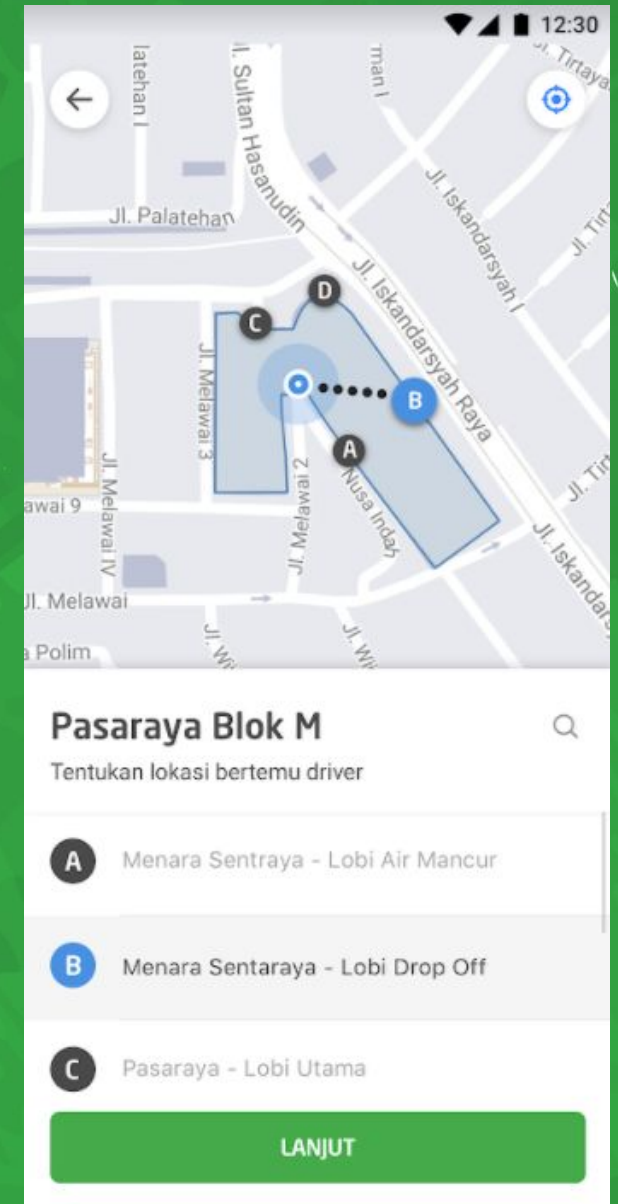
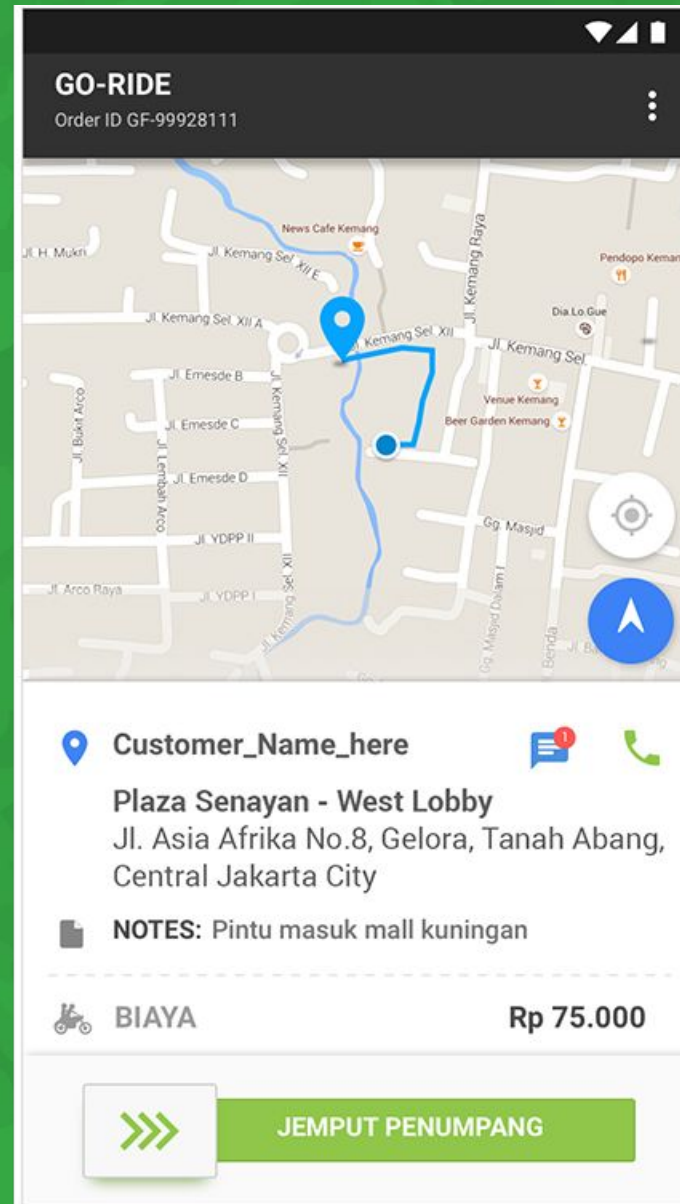
- Identify anomalous GPS points
- Correct anomalous GPS data
- Track anomalous GPS behaviour

## The Challenge:

- There are many reasons why GPS data may be inaccurate: buildings, bad devices, fraud
- Our systems require reasonably accurate GPS data to function optimally else user experience suffers
- GPS data is one of the biggest data we have so the models need to be fast and scalable



# Intelligent Places of Interest



Questions ?  
[dipta@go-jek.com](mailto:dipta@go-jek.com)