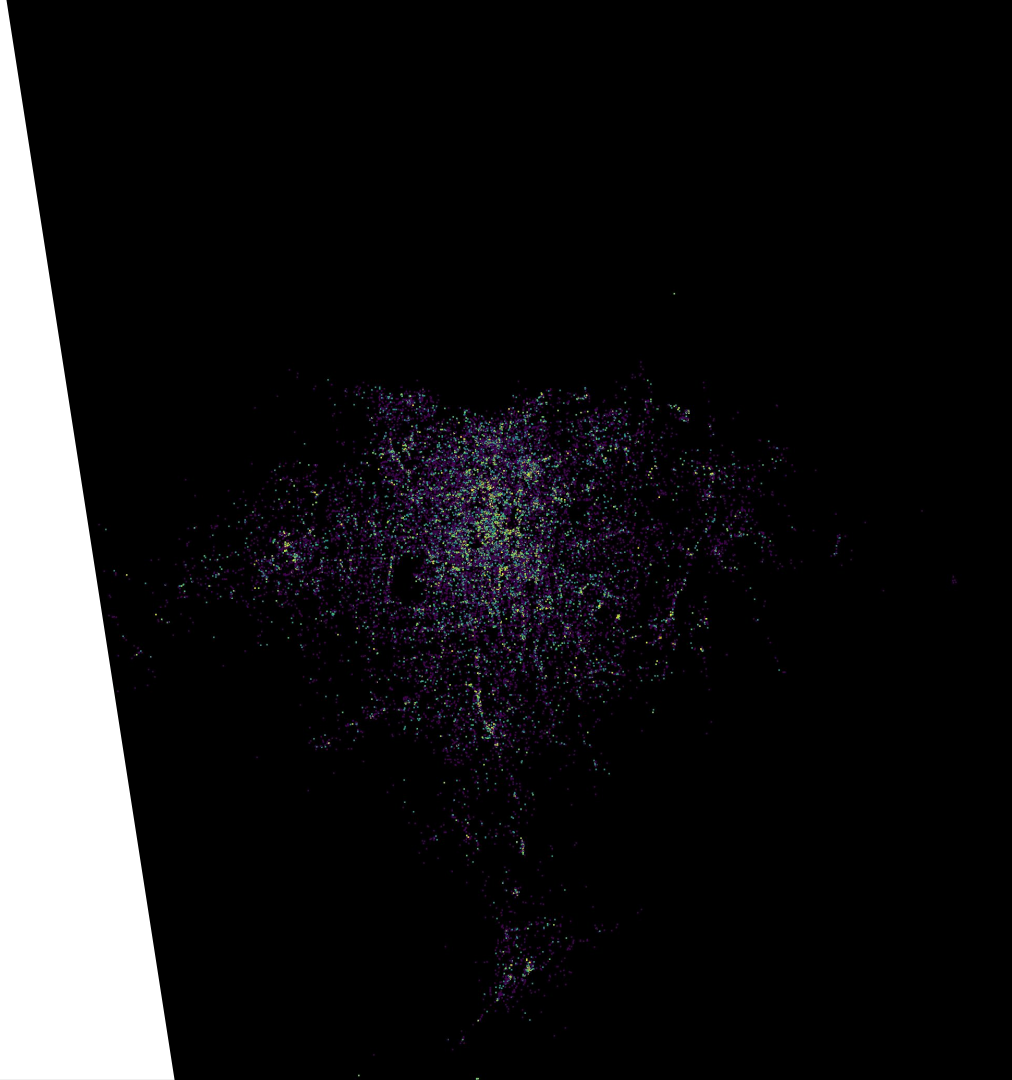
Machine Learning @ GO-JEK

Data Science Weekend
April 28, 2018

Kevin Aluwi

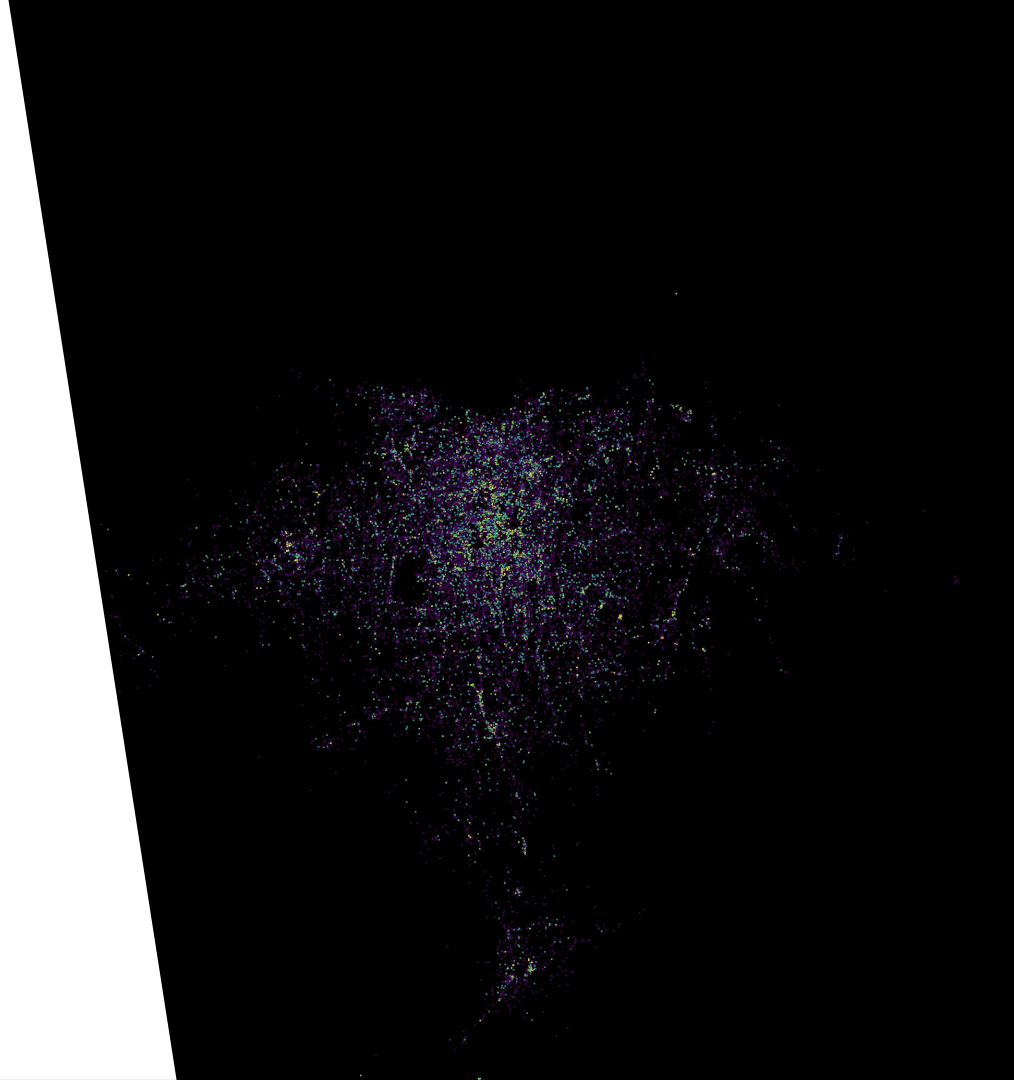
- More than 77 million downloads since the app launch in 2015
- A fleet that covers over **8 million KMs** per day, the equivalent of **nine** round-trips to the moon
- GO-PAY is the largest digital wallet in the country
- GO-FOOD alone is the **largest food delivery company in Asia** outside of China
- GO-FOOD delivers **30,000 TONS** of food monthly
- GO-JEK carries around **3 million** people daily

GO-JEK



Operating principles for data science at GO-JEK:

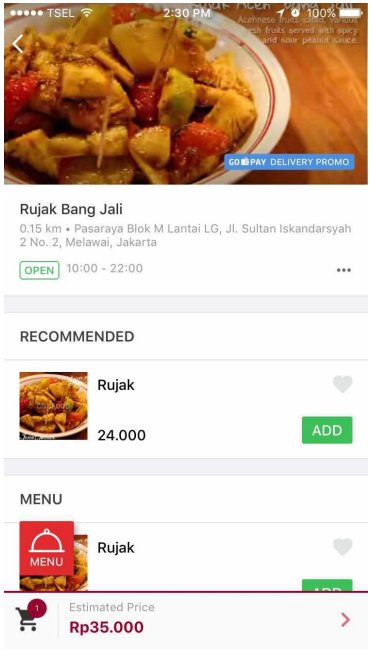
- Applied, **not** academic!
- Integrated with product engineering
- Scalable

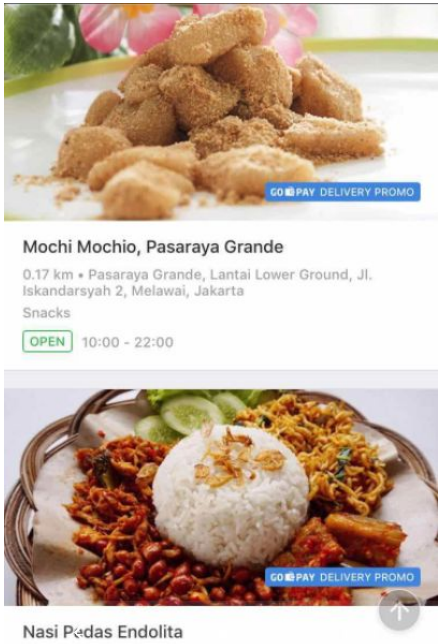


GO-JEK


Food Tagging with Deep Learning

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





The Goal:

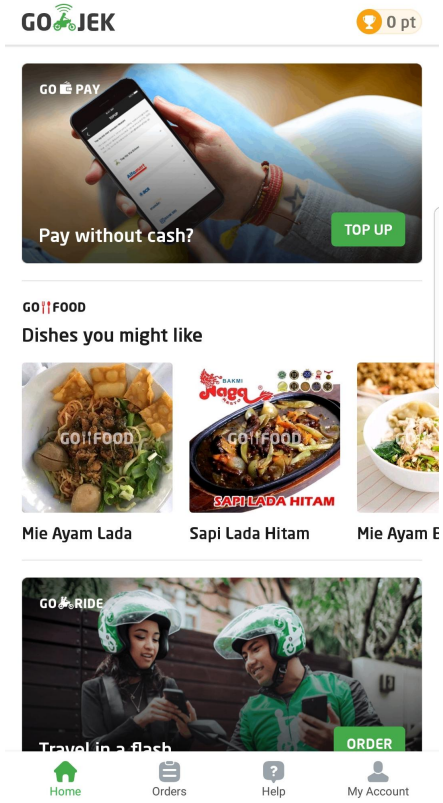
- Tag each dish
- Item level analysis
- Dish level recommendation feature

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
menu_hemat
cafe
ala
fastfood
dessert
topping
burger
aneka_ayam_bebek
geprek
penyet
minuman
praktis
padang
snacks_jajanan
crispy
sahur
bakmie
side
tea
manis
european
bakso
super
nasi
milk
pizza
paket
telur
fries
bakery
mie
coklat
biasa
bubur
hot dog
special
bento_ramadan_spesial
makanan
lauk
western
sate
sapi
sayur
shop
es
dada
besar
bebek
pisang
promo_hokben
tahu
beef
japanese
martabak
aneka
paha
rice
putih
indonesian
regular
seafood
pcs
24_hours
roti
ikan

What dishes would customers like to eat?



The Goal:

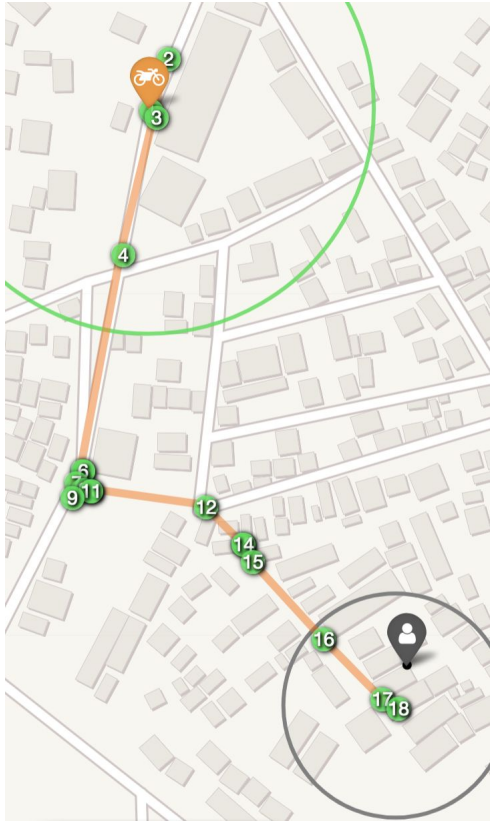
- Personalized dish level recommendation to users
- Increase discovery of new dishes based on past preferences

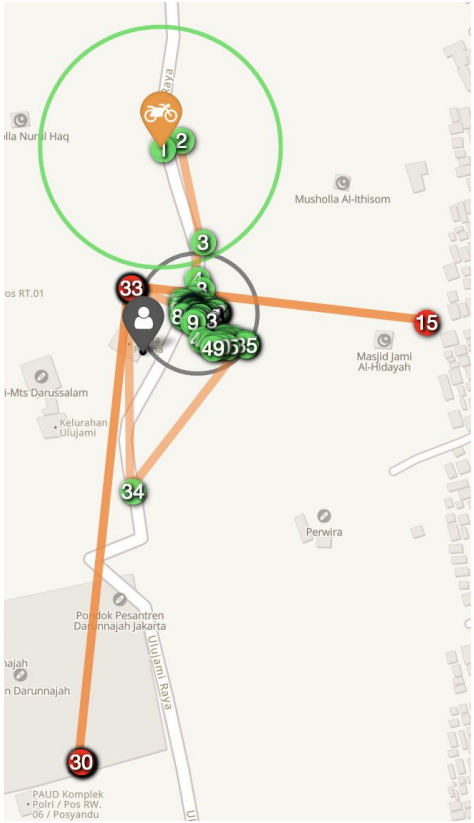
The Challenge:

- 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:

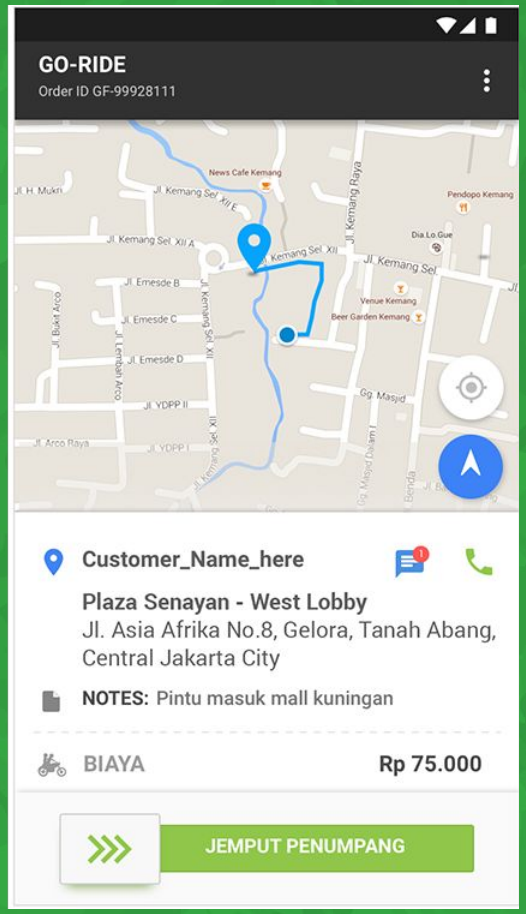
- Identify anomalous GPS points
- Correct anomalous GPS data
- Track anomalous GPS behaviour

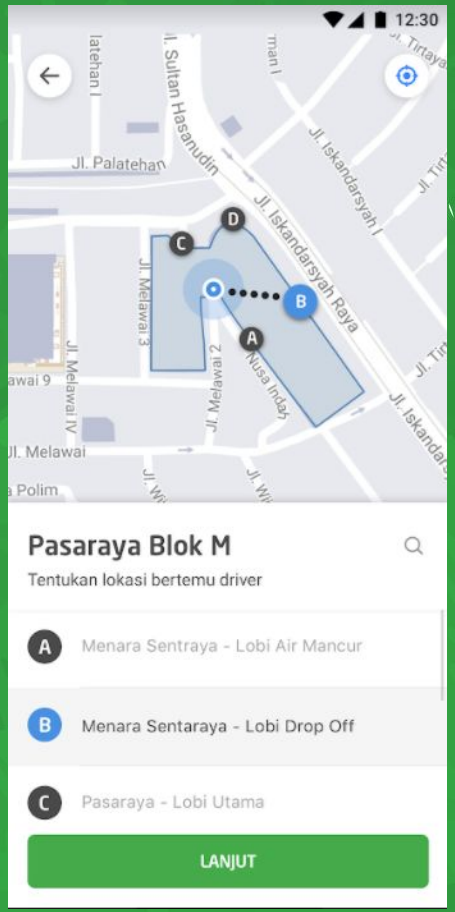
The Challenge:

- 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

GO-JEK

Intelligent Places of Interest





Q&A

Thank you!

kaluwi@go-jek.com