QARTA: An ML-based System for Accurate Map Services

Mashaal Musleh[1], **Sofiane Abbar**[2], Rade Stanojevic[2], and Mohamed Mokbel[2]

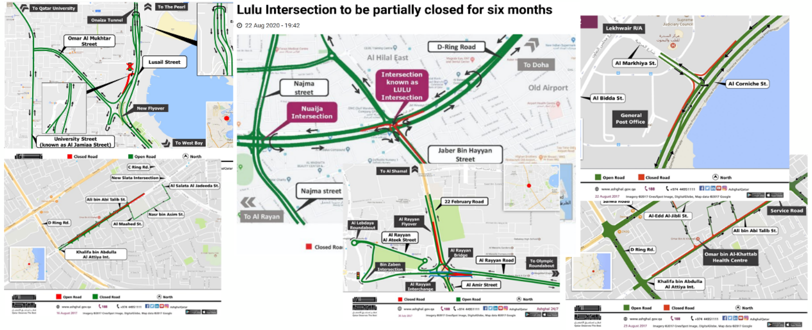
[1]Department of Computer Science and Engineering, University of Minnesota, USA
[2]**Qatar Computing Research Institute**, HBKU, Doha, Qatar

August, 2021


The Peninsula
Local focus, Global vision
QATAR'S DAILY NEWSPAPER
3rd Best News Website in the Middle East in 2017

Qatar road network increased three times between 2013-18: Ashghal

24 Apr 2018 - 11:58



QARTA leverages **big spatio-temporal data** and **machine learning**, to build a **map engine** that understands routes, **traffic**, and **drivers**.

+4%
Accuracy

-70%
Pricing

Fleet-aware
Car | buses | motorcycles

QARTA in production

All 4k taxis in Qatar

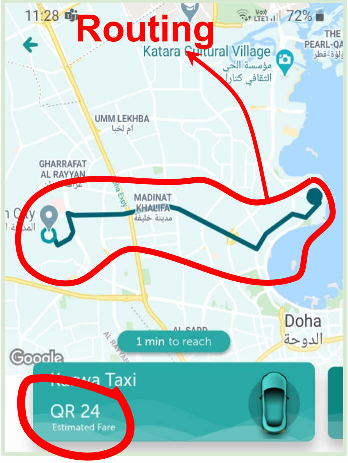
3k delivery motorbikes

Services: https://qarta.io

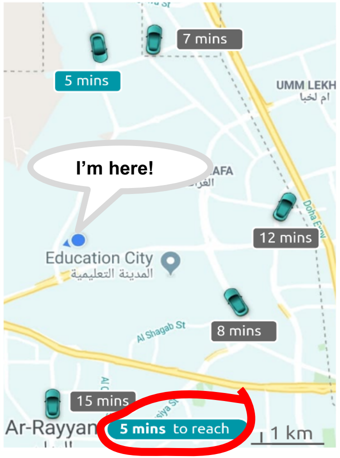
- In-traffic navigation
- Travel time estimation
- Complex route planning
- OD matrices

2M requests/week

1M GPS point/day

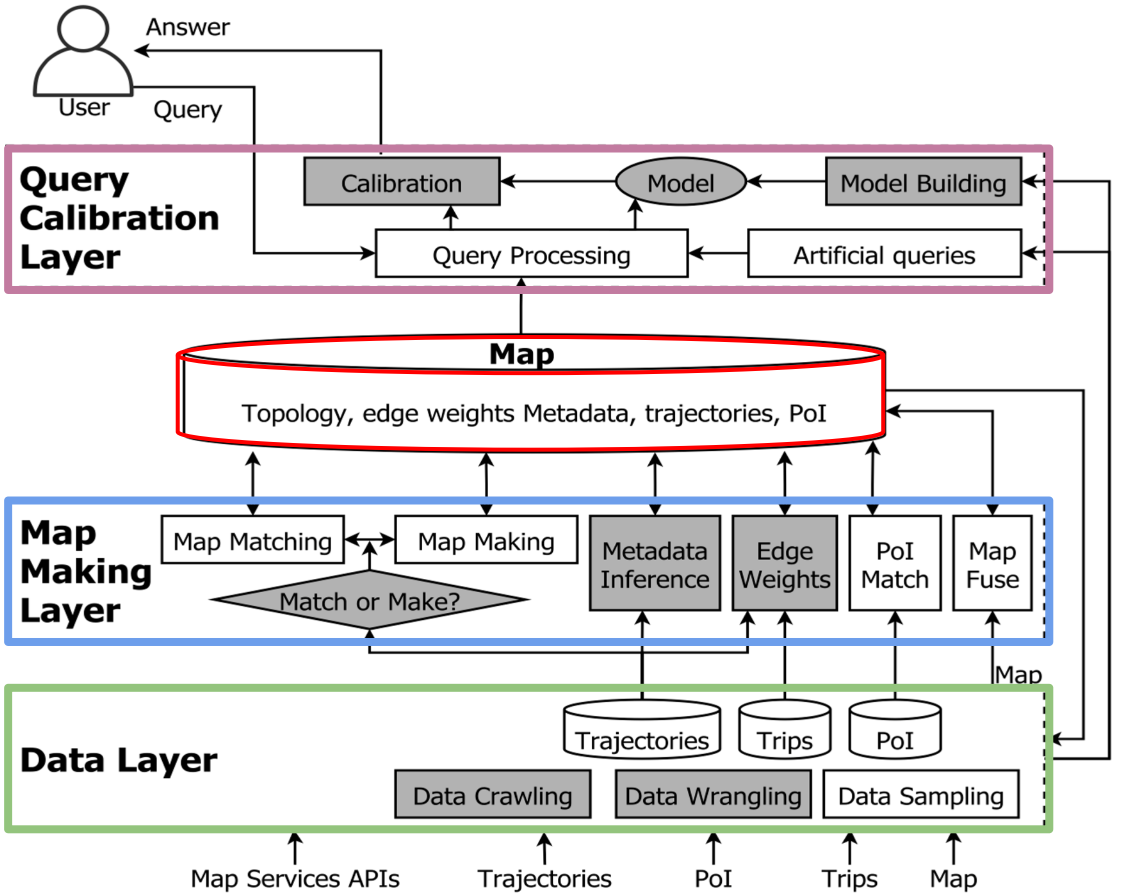


Fare estimation



Taxi Dispatching

General Architecture



DL (I): Rule-based Data Cleaning

Existing efforts for data cleaning and wrangling do not support spatial and spatio-temporal data

"After analyzing all your data, I think we can safely say that none of it is useful."

- Deployed Rules in QARTA
  - Trajectories with a stop
    - Split the trajectory
  - Unrealistic points
    - Remove the point
  - Missing points
    - Split the trajectory

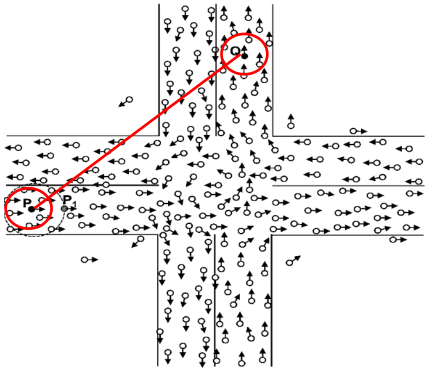
DL(II): Trajectory Imputation

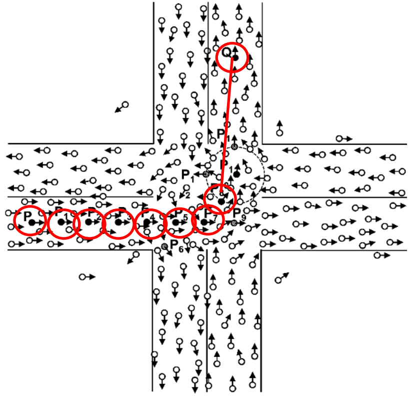
Low sampling rates

- ~400m
- Save energy & bandwidth

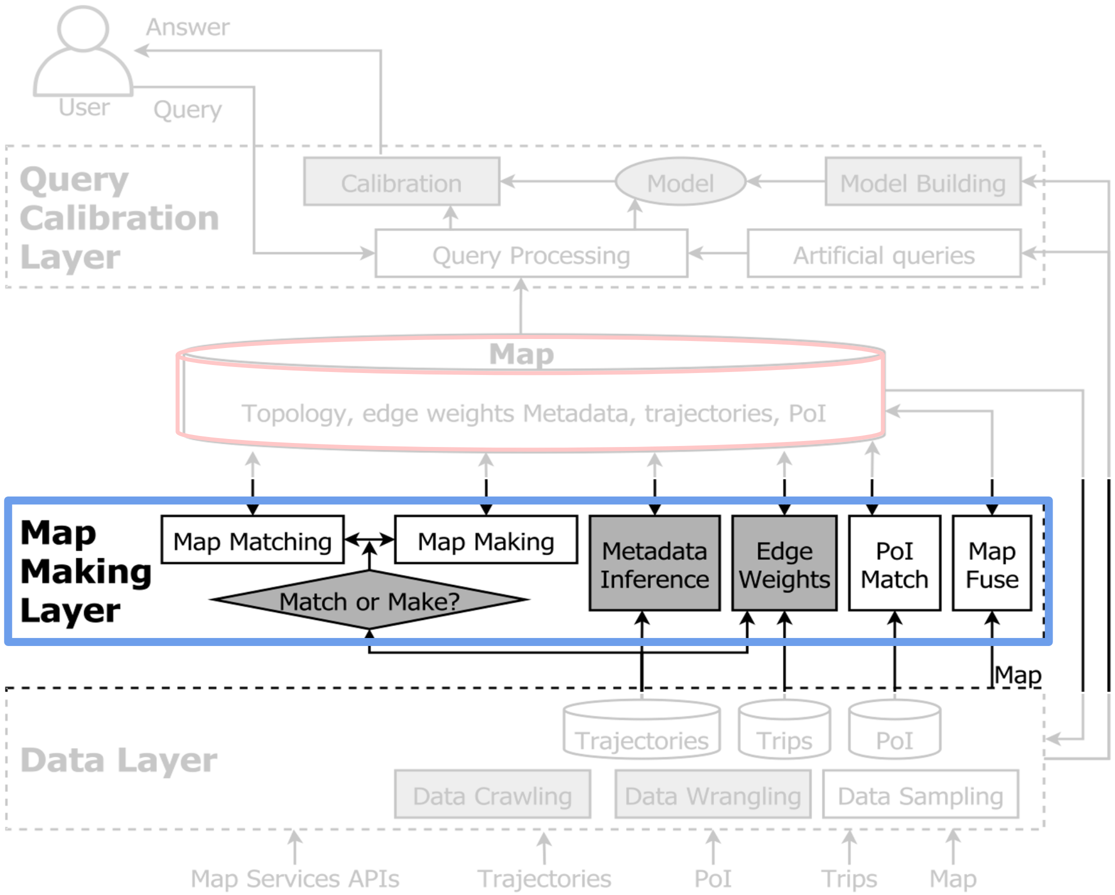
Need densification

- Use the wisdom of the crowd to impute each trajectory

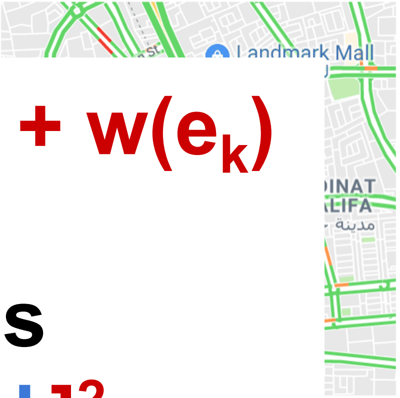




General Architecture: MapMaking Layer



MM Layer (I): Edge weights



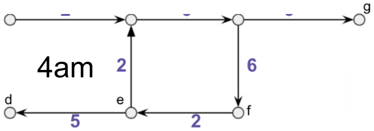
Each trip: $d_i \approx w(e_1) + w(e_2) + \ldots + w(e_k)$

Solve system of equations

$$\Sigma_i[w(e_1) + w(e_2) + \ldots + (e_k) - d_i]^2$$

Introduce constraints with Ridge Reg.

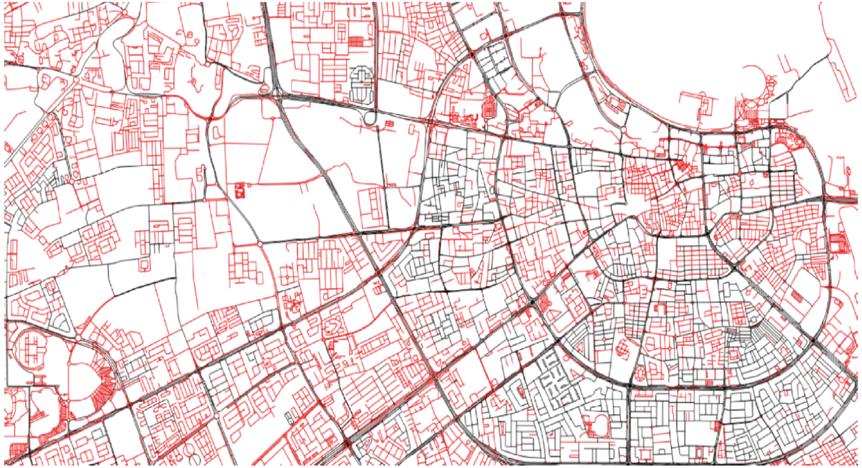
9-1-19 4:05	51.61527	25.26012	51.40927	25.24814	30.8	31
9-1-19 4:06	51.61497	25.2602	51.53813	25.32512	20.4	20



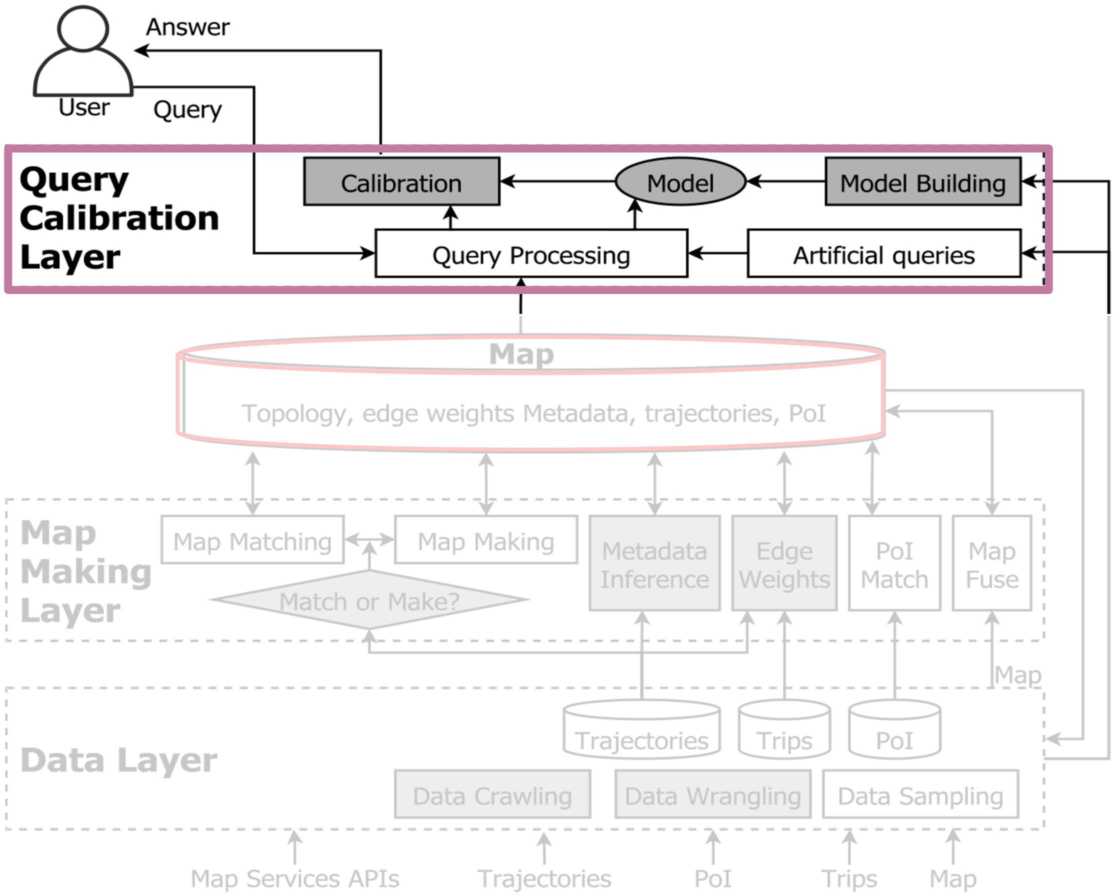
MM Layer (II): Metadata imputation

- Need rich metadata (annotation) for road networks
  - Speed limit
  - Number of lanes
  - Road type
  - ...
- Metadata inference in QARTA is framed as a **supervised learning problem**
  - **Step 1**: Find the best models that would map road features to certain metadata
  - **Step 2**: Use these models to predict the missing metadata values

Public maps have very poor metadata coverage



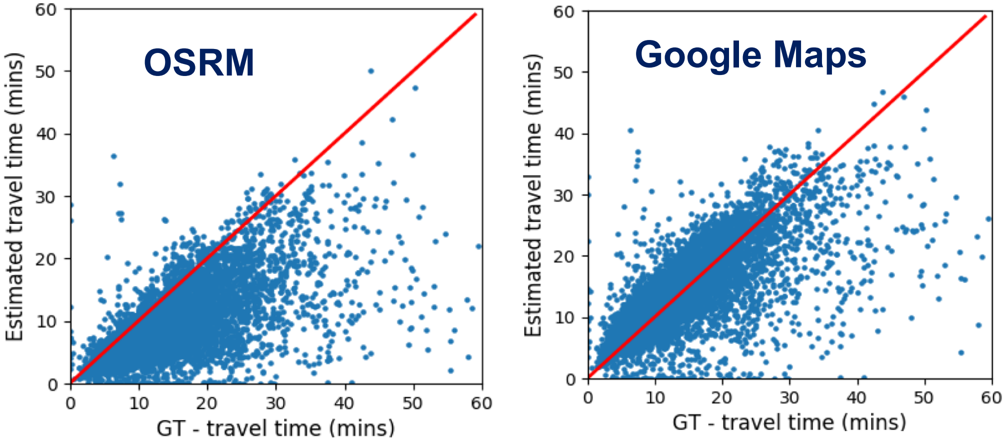
General Architecture: Calibration



Cal. Layer (I) : Estimated time of arrival

Important for:

- Route planning
- Logistics | Deliveries
- Fare estimation



Depends on:

- Time | Location
- Route
- Vehicle type
- Weather

Can we learn the per-trip **ETA offset** (error) **distribution** for each map service, and use that to calibrate our queries?

Cal. Layer (I): Supervised Learning of ETA offsets

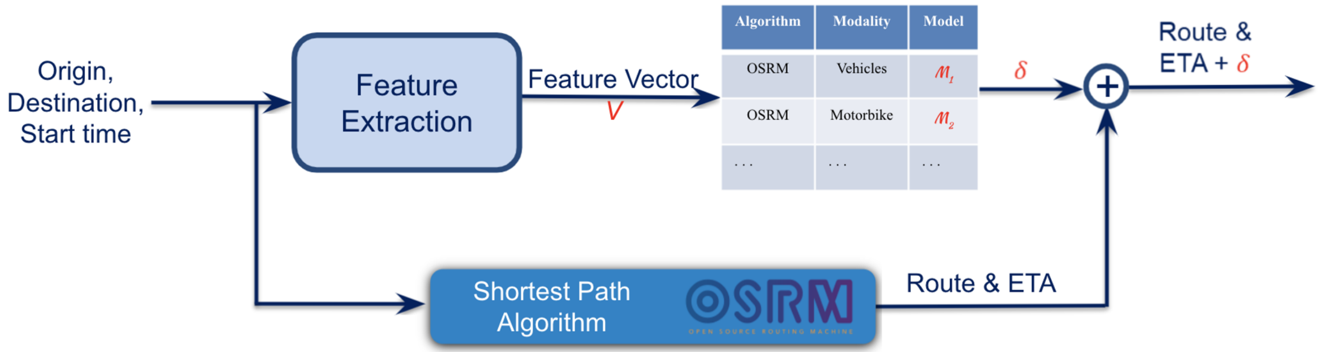
free flow features | Temporal features | Spatial features

ETA_OFFSET(s,d,t) ~ f(osrm_tt(s,d), distance, hour_of_day, day_of_week, hour_of_week, zone(s), zone(d))

Trips

+5	17	12.5	8	2	32	'14'	'53'
+2	15	9.8	22	6	142	'03'	'12'
-7	32	11.2	7	1	7	'51'	'63'
...	...	...	...	...	...	...	...

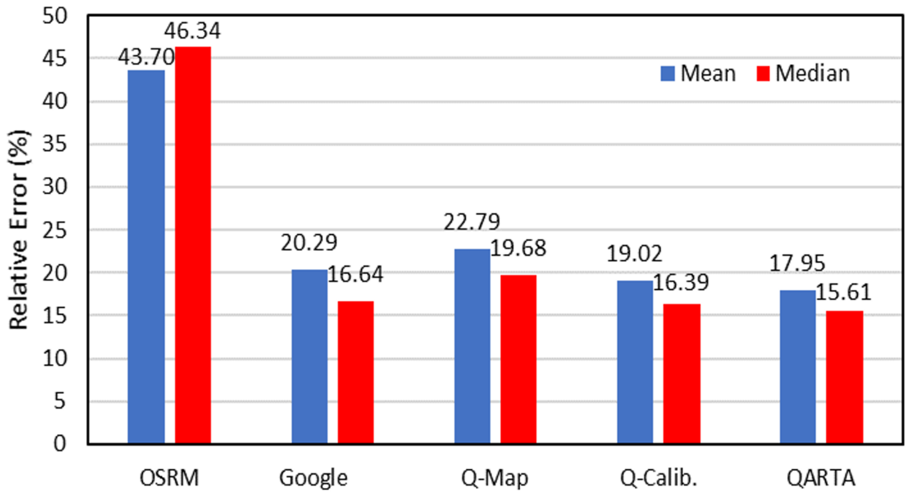
Shortest path queries



Experimental evaluation

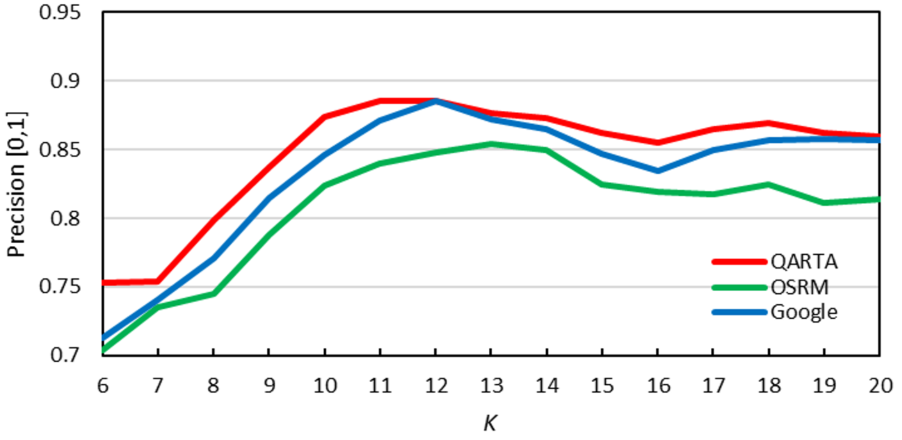
- **Data**
  - Qatar taxi fleet
  - 250k trips
- **ML models**
  - Trained on 200k trips
  - Tested on 50k trips
- **Underlying algorithms**
  - OSRM for shortest path
  - OSM map
  - Off-the-shelf kNN&Range queries

Shortest Path

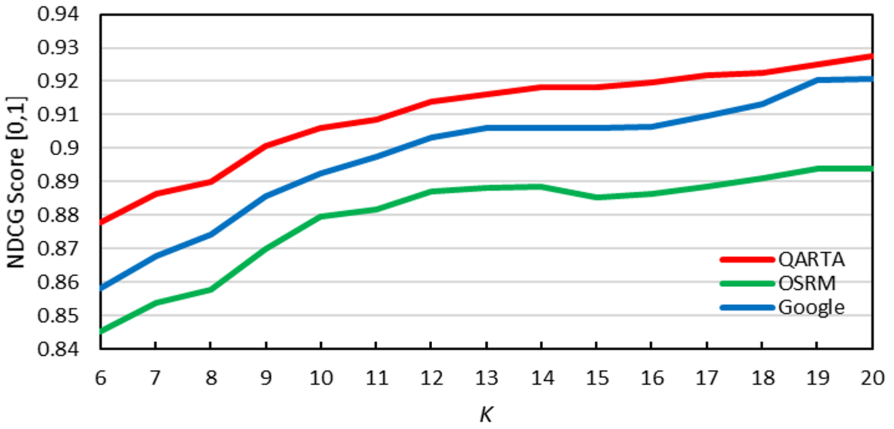


k-NN Queries

- *Precision*: Number of items in *K*NN list that overlap with ground truth
  - All very similar performance



- *NDCG*: A ranking quality measure that takes into account the order if items in the list



Thank You!

Do not forget to attend our demo

7 **A Demonstration of QARTA: An ML-based System for Accurate Map Services**

Sofiane Abbar (Qatar Computing Research Institute), Rade Stanojevic (Qatar Computing Research Institute), Mashaal Musleh (University of Minnesota), Mohamed M Elshrif (Qatar Computing Research Institute), Mohamed Mokbel (University of Minnesota - Twin Cities)*