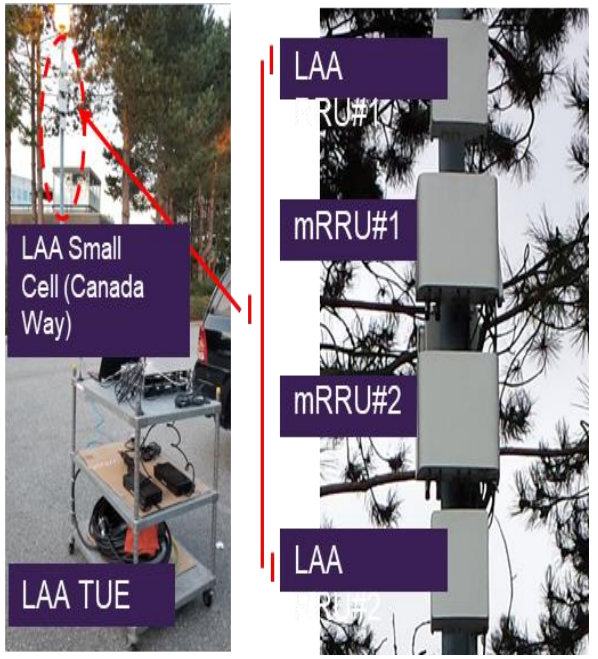
Coop-
Pole cleaning ML
Project Presentation

Charles-Zhang
5G-small cell team
Project from 2020 Jan - Current

Project problem definition

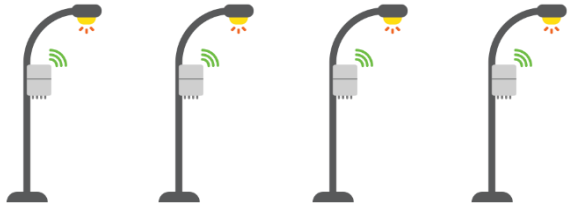
- ►**4 million poles(data entries)** that is used, or unused cross Canada stored in MSSQL server
- ► **57 different sources'** data entries joined without duplication elimination
- ► More than **100 different types** of poles that need to be distinguished for **same Latitude and longitude**



What I learned and used for my project?

Split problem into different layers

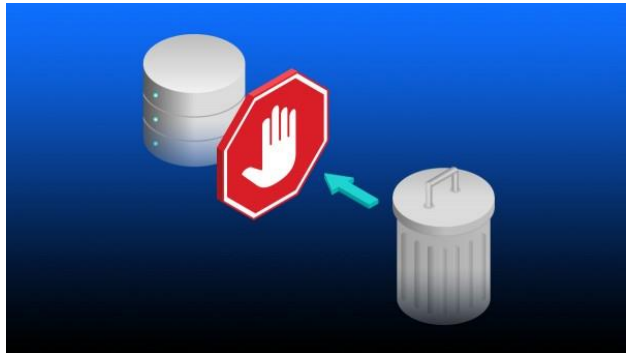
1



Eliminate Duplication poles with same latitude and longitude

Use python and sql queries to eliminate the duplication of poles

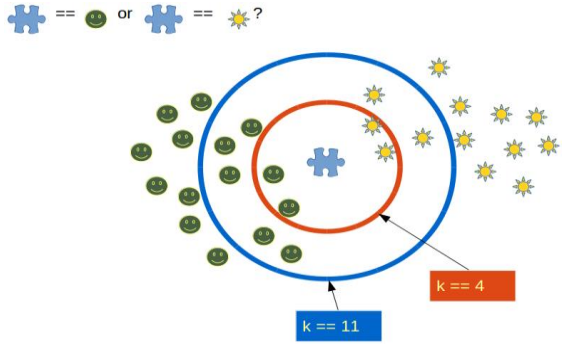
2



Get rid of the garbage data with wrong datas

Clean the data with wrong data informations which will be distract and messed up machine training later.

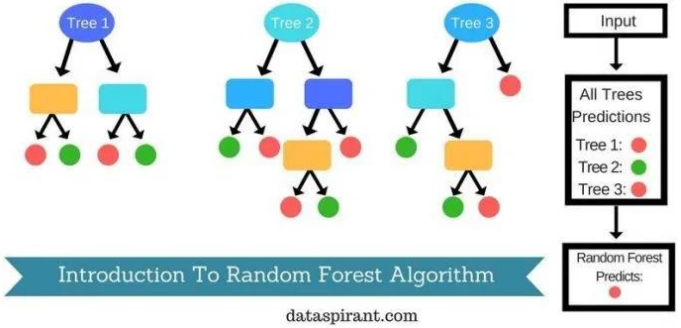
3



K Nearest Neighbor-Two layers

Use this ML to join on all poles latitude and longitude. Get closet poles around one pole within 10 meters

4



Random Forest Classification - Four layers

Use this classification to build the model and figure out if the pole is “Ghost” or “Exist”

5

Update new data back to SQL

How this project affect my team ?

For now

- Pro: The current data is cleaned and accuracy is more than 92%
- Pro: The current used pole are updated as only pole with no more duplication poles with same location from other sources
- Pro: Garbage datas are no longer confused inside sql server

For future

- Pro: The future planning use of any poles won’t get any wrong datas.
- Pro:Through machine learning we can find out the which source is more reliable
- Pro: The planning pole won’t be the Ghost pole or many duplications need to be verified

Thank you!

Questions and comments?

Email me at Charles.Zhang@telus.com