CAPSTONE PROJECT: BATTLE OF THE NEIGHBORHOODS, JAKARTA

Jakarta Food Lover Recommendation

I. PURPOSE

This document provides the details of work of Capstone Project IBM Data Science Certified program.

II. INTRODUCTION

Jakarta is capital on Indonesia with a population of 10,075,310 as of 2014. A population with mix of cultures – Arab, Chinese, Javanese, Malay, Indian and European – has influenced its architecture, language and cuisine.

Jakarta also home of many oil, gas and natural resources company operate in Indonesia. There is 4231 international liaison office in Jakarta based on 2014 dataset from Jakarta open data. There are 4165 men and 1018 women work as a professional. Most professional come because their company transferred them there under contract as an expat. People struggle to find neighborhood with adequate living experience without sacrificing the quality. People need apartment, house rental, restaurant, breakfast spot, or coffee place with high quality rating reviews. They also want neighborhood close to their work without waste too much time in traffic each day.

Sometimes expat use a travel agency to locate their best neighborhood, however many of them attempt to push the transaction based on commission, not the right fit of customer. Other problems involve a language barrier. Many expats express their experience communicate with local. Some of them ask questions before he or she realizes his Indonesian colleague actually doesn't know what he's talking about.

The need for data-driven recommendation system based on rating or review would be significantly helpful to guide people to the right neighborhood with confident.

II. DATA ACQUISITION

To help solve the problem, some activity involving data acquisition of the Jakarta area need to be done.

- Jakarta Neighborhood Data

The need of dataset containing a zip code of Jakarta neighborhood can be found on Jakarta open data website. Some of them with location provided but most of them are not. Data location is: http://data.jakarta.go.id/dataset?tags=kelurahan

- Geo location acquisition

The use of Geo-location API use to retrieved Neighborhood with missing coordinate. This step providing information about Latitude and Longitude of neighborhood.

- Real estate data

Jakarta real estate dataset from Jakarta open data can be provide on Open data Jakarta website. Link. "http://data.jakarta.go.id/dataset q=perumahan&sort=score+desc%2C+metadata_modified+desc". Dataset provide basic information about living situation in the area.

- Neighborhood Nearby Venue

Information about Jakarta neighborhood venue provide by Foursquare API. Coordinate location of neighborhood is request to Foursquare web service and Foursquare reply with a set of venue nearby in JSON format. The data then clean and process as a data-frame of neighborhood with vunues.

- Neighborhood Food related rating

Foursquare also provide limited request for rating of the restaurant base of the venue ID. First, dataset from previous step will be group, filter to contain data about food related only, using blacklist keyword technique.

New data-frame produce from step above contain food related venue only. Columns on dataset also contain neighborhood name, zip code, location and venue ID. By using foursquare limited call api, customer rating for venue can be extract.

This activity would take couple of day to complete. Foursquare allow 500 call perday only for this kind of call.

Recommendation system develop using data set above can be utilized to segmented neighborhood by living cost, high rating restaurant, location to the work, and other perk like park on the neighborhood and so on.