APPLIED DATA SCIENCE CAPSTONE PROJECT: BATTLE OF THE NEIGHBORHOODS

Restaurant Recommendation System Model

INTRODUCTION

Special Capital Region of Jakarta is the capital of Indonesia with a population of 10.77 million as of 2020. As the capital of Indonesia, Jakarta is the melting point of cultures of all ethnic groups of the country. The culture of the city represents many languages and ethnic groups, support differences in regard to religion, traditions and linguistics, rather than any single and dominant culture.

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 neighbourhood 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 neighbourhood close to their work without waste too much time in traffic each day.

Sometimes expat use a travel agency to locate their best neighbourhood, 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.

For my capstone project, I wanted to build something that would be meaningful for these expats in their everyday lives. This ultimately led me to build a recommendation system model that could recommend people restaurants near their location based on restaurant ratings and reviews from other people.

I will be using Foursquare API to find neighbourhood nearby restaurants and group them together based on geographical location is by using the K-Means Clustering Algorithm.

DATA ACQUISITION

For the purpose of this capstone project, the dataset that I built my models on was taken from:

1. **Jakarta Neighbourhood Data**
   The need of dataset containing a zip code of Jakarta neighbourhood can be found on Ministry of Home Affairs (Kementrian Dalam Negeri) website (https://ditjenbinaadwil.kemendagri.go.id/wp-content/uploads/2019/01/ DAERAH-KHUSUS-IBUKOTA.pdf) . I already convert the pdf into dataset before.

2. **Geo Location Acquisition**
   The use of Geo-location API use to retrieved neighbourhood with missing coordinate. This step providing information about Latitude and Longitude of neighbourhood.

3. **Neighbourhood Nearby Venue**
   Information about Jakarta neighbourhood venue provide by Foursquare API. Coordinate location of neighbourhood 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 neighbourhood with venues.

4. **Neighbourhood 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 neighbourhood name, zip code, location and venue ID. By using foursquare limited call API, customer rating for venue can be extract.

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