9th MALAYSIA STATISTICS CONFERENCE 2022
Dealing with Uncertainties: Unearthing Measures for Recovery
Institut Latihan Statistik Malaysia
5th October 2022

Application of seamless hybrid geocoding solution for business location using KAWASANKU API

***Ahmad Najmi ARIFFIN** [1,2], Mohamad Hamizan ABDULLAH [1]

[1] Core Team Big Data Analytics, Department of Statistics, Malaysia
[2] Faculty of Dentistry, Universiti Kebangsaan Malaysia (UKM);

Abstract:

Addresses can be geocoded by transposing them into the corresponding longitudes and latitudes. The coordinates provide a means of pinpointing a map's location. This article provides a comprehensive methodology for online geocoding services that are widely utilised. Our study aimed to validate the distributions of geocode locations to take into account when analysing geocoded address data, and to create strategies for enriching demographic databases by utilising the centralised public data sources repository KAWASANKU API on the Github platform. Open Data is data that is accessible, usable, and shareable by the public. The private sector has been very hesitant to adopt Open Data. If businesses utilise Open Data strategically, it can be a key factor in generating a variety of uncertain business opportunities, such as enhancing new products and services. The "geopy" Python module enables the mapping of global coordinates for addresses, cities, countries, and landmarks. Risk-assessment using Fuzzywuzzy (Python library) returns the similarity percentage [1-100] between two sequences of addresses strings to match. The q-ratio score threshold is over 65. These addresses will be re-geocoded and evaluated for completeness. KAWASANKU API can query socio-demographic features and Malaysia's geospatial boundaries down to the DUN level, including national, state, district, parliament, and state legislative assembly (Malay: Dewan Undangan Negeri, DUN). We get 2,427 geojson raw lines for each property feature. This framework for a seamless, less-dependent workflow to reduce risk and benefits on data enrichment. Using the provided script, this framework allows SMD to self-perform processes. Researchers must be aware of certain peculiarities to effectively use the data, which is a research opportunity.

Keywords:

Open data; Geospatial; Python; Github API