Data

Exploring search parameters: initial search parameters listed 5 dengue-related phrases, each containing the word “dengue”. Additional phrases and combinations of words may aid in gathering better results. Further testing should include more common search terms, possibly covering other symptoms of the illness. Using vernacular and local dialects may also hint at other related search terms. This must be tempered with other illnesses with similar search fields by prioritizing combinations of symptoms related to dengue.

Additional functions for geocoding and assigning points are needed. Consider having a roster of locations (barangay/street address vs lat-long) to reference locations listed in agency dengue data. Any enhancements may be for accuracy

Installation

Current packages require the user to manually install them. The team is considering other automated or batch installation procedures. The following packages have to be manually installed as of this update: TPOT, PyTrends and GDAL.

Specific version of Pandana is necessary for osm to work (Pandana =0.4.4)

Model

The team is currently exploring other parameters and data preprocessing necessary in the Auto ML. Additional documentation is being prepared to specify them.

The feature importances graph does not display after running the perform\_classification function.

Exported models trained via TPOT cannot be read by the Pickle module.

User must be able to change number of generations for ML

PerformRegression function is using TPOTClassifier for ML which does not work for non-categorical target variable

Others

[under Presenation] It would be much better to see the relative risk of each area, not just the hotspots.

[on Location Input] Add support for string input for location instead of geojson.

[On code] Users should be able to query date on the get\_search\_trends function like the one on get\_satellite\_measures\_from\_points.

get\_satellite\_measures\_from\_points function takes a long time to get remote sensing data (NDWI,FAPAR, NDMI, etc.). Ex: Takes 6 minutes for 50 random coordinates.

[On code] The get\_OSM\_network\_data function didn’t plot the data when the parameter was set to true.

Similar to the 4th issue, Instead of having a geojson input, users should be able to opt for locating areas like with the Google Search trends method by ISO Geotag. E.g. (PH-40 for Laguna and etc.)

[Streamlit app] There should be a legend for the map markers on the prototype. E.g. Cluster 1 is Yellow, Cluster 2 is Red and etc.

Variables used in the documentation (code snapshots) was not consistent throughout. (E.g. QC\_AOI and qc\_df was used in the first part then it latter became aoi\_geojson and satellite\_df.)