Main Repository: CMPT459-Final-Group-Project Merging Guide

Note: Please feel free to ask questions about the implementation below and update/delete/insert your implementation of Step 1- 6 into the main branch where specified.

Folder: Steps-1-2-3
Status: Complete

****Used dataset files (../Steps-1-2-3/data):**

cases_2023_test.csv
cases_2021_train.csv
Location_2021.csv

****Resulting dataset files(../Steps-1-2-3/results):**

case_test_processed .csv
case_train_processed.csv

1) Downloaded Datasets from the introductory paragraph
2) Visualized data using 'data_visuals.py'
    a) This includes produced heatmaps and bar graphs as shown in (../Steps-1-2-3/plots)
        - **Include your plots here**
            - **This includes heatmap visuals and any other data visuals you could think of.**
3) Pre-processed Datasets from the intro paragraph
    a) Removed LAST_UPDATE column
    b) Replaced missing age/sex values with -1
    c) Removed unlikely ages from the dataset
    d) Removed unlikely long/lat values from the dataset
    e) Removed row if province is unknown
        - **Feel free to add/delete to accommodate for your implementation of Steps-1-2-3 of pre-processing the data**
4) Lastly joined datasets by combined key
5) Created a .ipynt file to show code

Folder: Steps-4-5
Status: Complete

****Used dataset files (../Steps-4-5/data) as provided in highlighted text in step 2:**

cases_2021_test_processed_unlabelled_2.csv

Cases_2021_train_processed_2 - cases_2021_train_processed_2.csv

****Resulting dataset files(../Steps-4-5/result):**

oversampled_processed_data.csv

1) Only performed pre_processing on training data! Not sure what to do with testing data as it didnt say anything about that.
2) Converted Sex to categorical [for mapping] -> new column name is sex_code
3) Converted Chronic_Disease_binary to categoical [for mapping] -> new column name is chronic_disease_binary_code
4) Converted province to categorical [for mapping] -> new column name is province_code
5) Converted country to categorical [for mapping] -> new column name is country_code
6) Converted **outcome_group** to categorical [for mapping] -> new column name is **outcome_group_code.**
7) Ensured that all types are numeric!
8) Removed date confirmation (unecessary?)
9) **Balanced dataset using Oversampling**

Folder: Steps-6
Status: KNN Complete

****Used dataset files (../Steps-4-5/result) for training the classifier models:**

oversampled _processed_data.csv

****Used dataset files (../Steps-6/hyperparameter_tuning_data) for hyperparamter tuning:**

oversampled_processed_data.csv

1) Pre-processed the dataset in the Introducotry paragraph as it told us to us that for hyperparamter tuning

2) **However**, that specified dataset was not fully pre-processed for hyperparamter tuning.
3) In ../Steps-6/Preprocessing/main.py I preprocessed the dataset so it could be used for hyperparamter tuning.
4) Implemented KNN model
    - **Here is where your implemented classifier models should go in the given folders**
    - **../Steps-6/Logistic_Regression for Logisitic Regression**
    - **../Steps-6/Random_Forest for Random Forest**

****For step 6, To train/validate your classification models use dataset [../Step-4-5/result/oversampled_processed_data.csv]. Make sure to drop categorical columns!**
For hyperparameter tuning use the dataset [../Steps-6/hyperparameter_tuning_data/hyperparameter_tuning_data.csv]