TEAM OUSE PROJECT REPORT

Objective –

The main aim of the project is to assess and analyse the flood risk in the UK.

Models -

The objective has been achieved by creating the following three models –

1. A regression model to predict the flood probability scale based on a given location:

The data has been pre-processed to find the nearest station to the input coordinates, and the model thereby makes a prediction based on the feature values for those nearest stations.

The KNN regression (neighbors=10) algorithm has been applied to predict the flood class probability. The mean squared error and the coefficient of determination (R2) have also been applied to evaluate the metrics of the model.

However, the rainfall and tidal data has only two days of data, and therefore, it does not affect the overall prediction of our model. Moreover, the model scores and the predictions remained similar with or without the data. The preprocessed rainfall & tide data has been provided in the Extra folder in the GitHub repository.

2. A regression model to estimate the median household price in England:

The data set utilized represents five influencing variables easting, northing, altitude, households, and sector. For this model, KNN regression (neighbors=5, weight='distance'), Lasso (alpha=2), and Ridge (alpha=1) algorithm were applied to estimate the median price. Comparative analysis between these models revealed that KNN outperformed lasso and ridge with 74% of R2.

However, the KNN algorithm may fail in estimating values beyond the range of values inputs used in training the model.

3. A classifier taking in an arbitrary location and predicting the Local Authority.

KNN, Decision Tree, and Random Forest algorithms have been evaluated to estimate the model. The KNN algorithm (k=5) has the highest prediction accuracy, with a score of 0.97.

However, only easting and northing have been selected for the model prediction as the input is an arbitrary coordinate (outside the unlabelled data).

GUI Design –

PyQt5 has been used to create GUI. This package requires external packages to run. However, the PyQtWebEngine package would not work properly with the ARM64 processor. Therefore, it is recommended to use MacBook with Intel CPU or Windows.

Visualisation –

Folium is the main module used for all visualisations. Offline rainfalls, stage or tide level visuals can be generated using (rain=True), (stage=True), or (tide=True), based on available data. Online API data visualisation can be done using (live_rain=True), (live_stage=True), or (live_tide=True). Different measures can be visualised together (live_rain=True, tide=True) to identify the areas that are at risk.

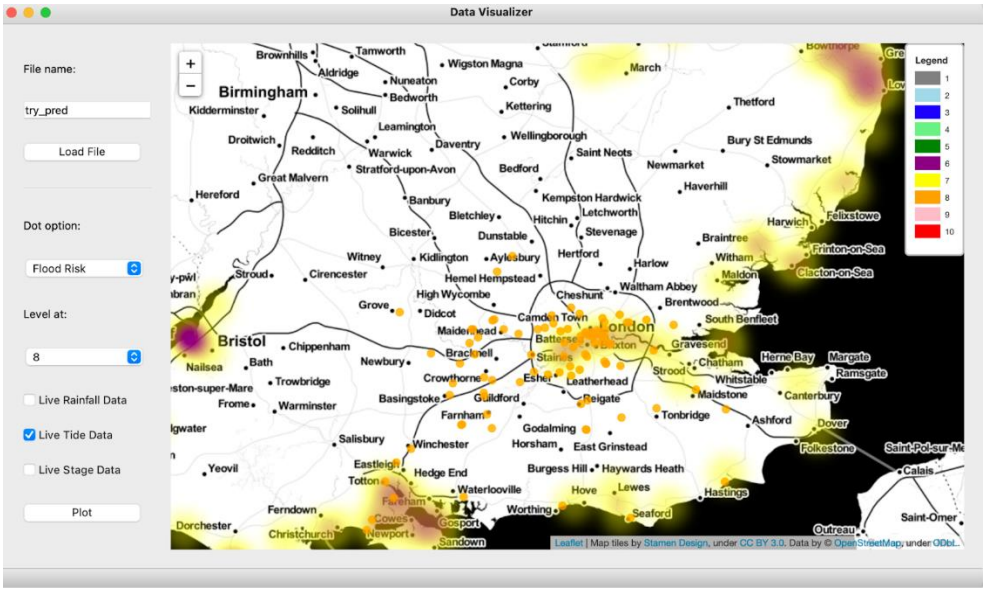

Figure 1 – GUI for predicting risk label and tide levels.

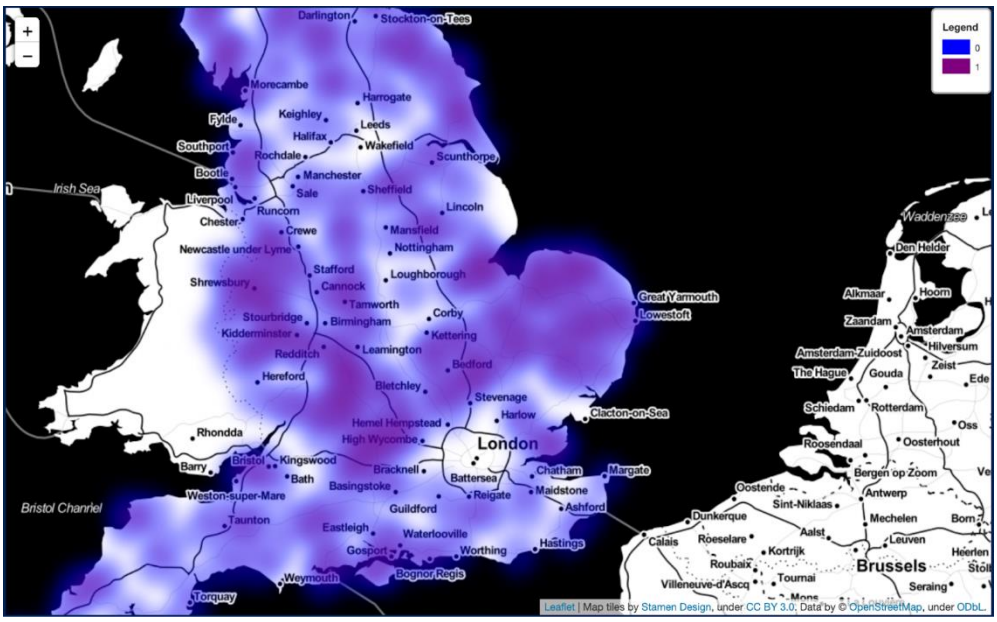

Figure2 - GUI for predicting online rain.

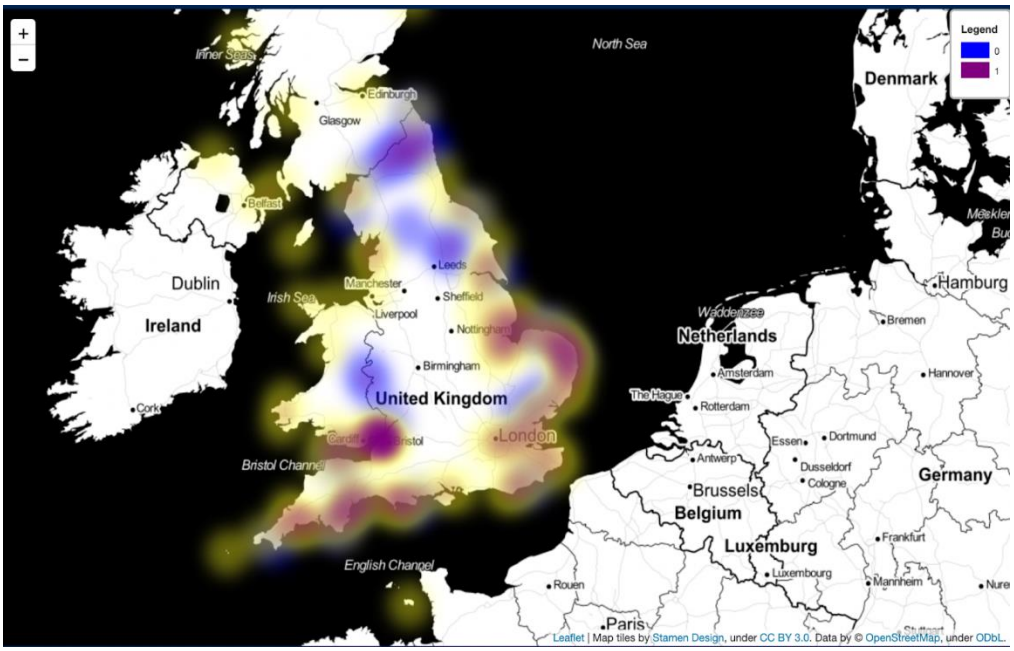

Figure 3 – GUI for predicting offline rain and live tide.