Car accident analysis in the UK

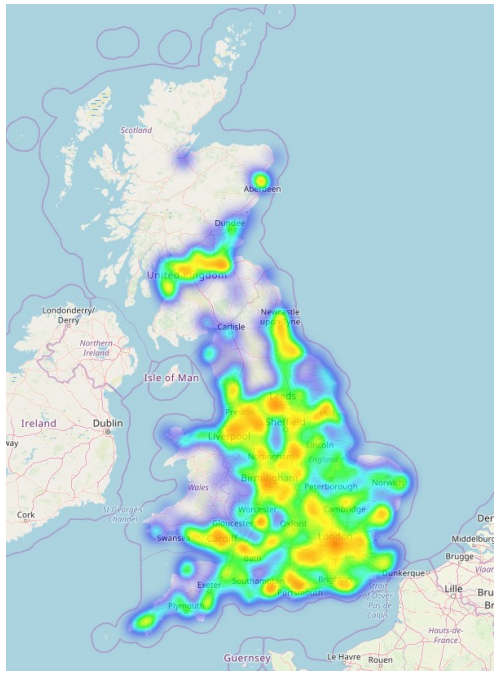
Haftamu Hailu Tefera

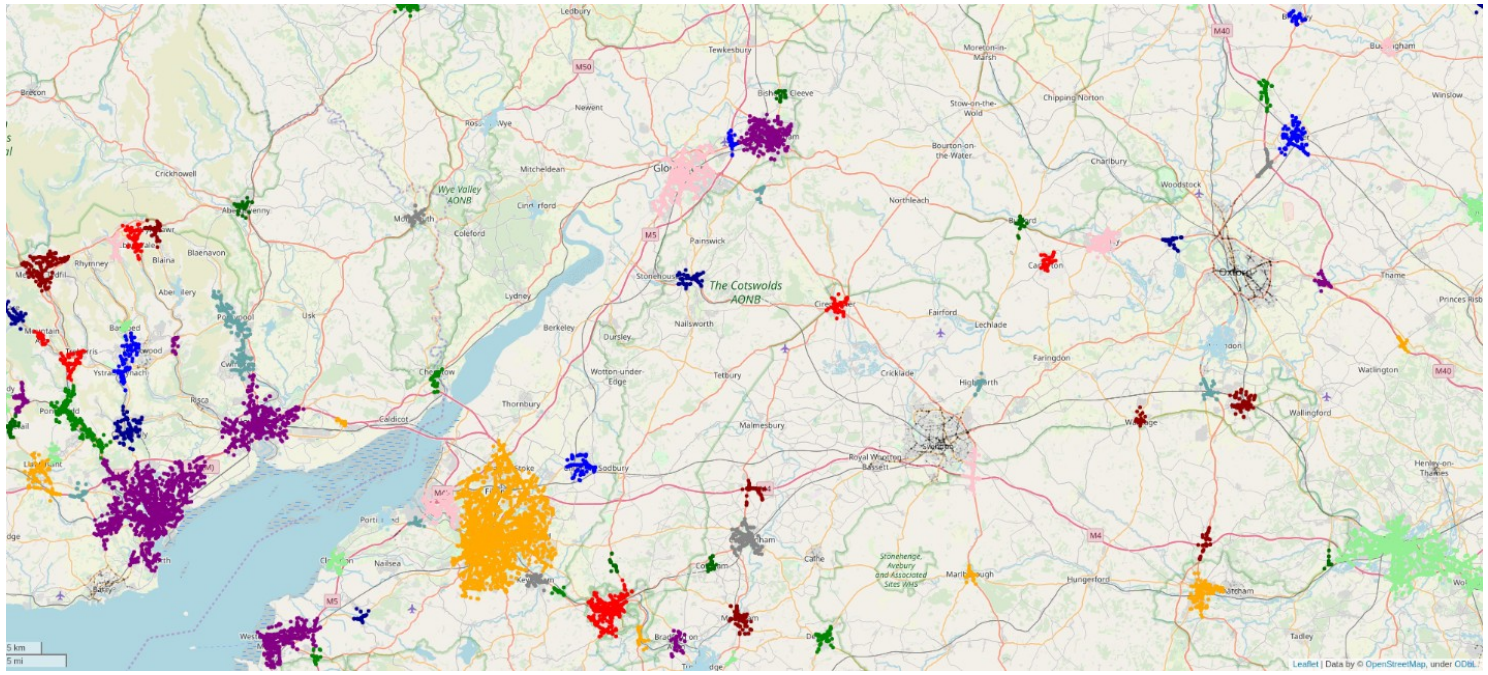
Prof:

Mahmoud Sakr

Identifying hotspots

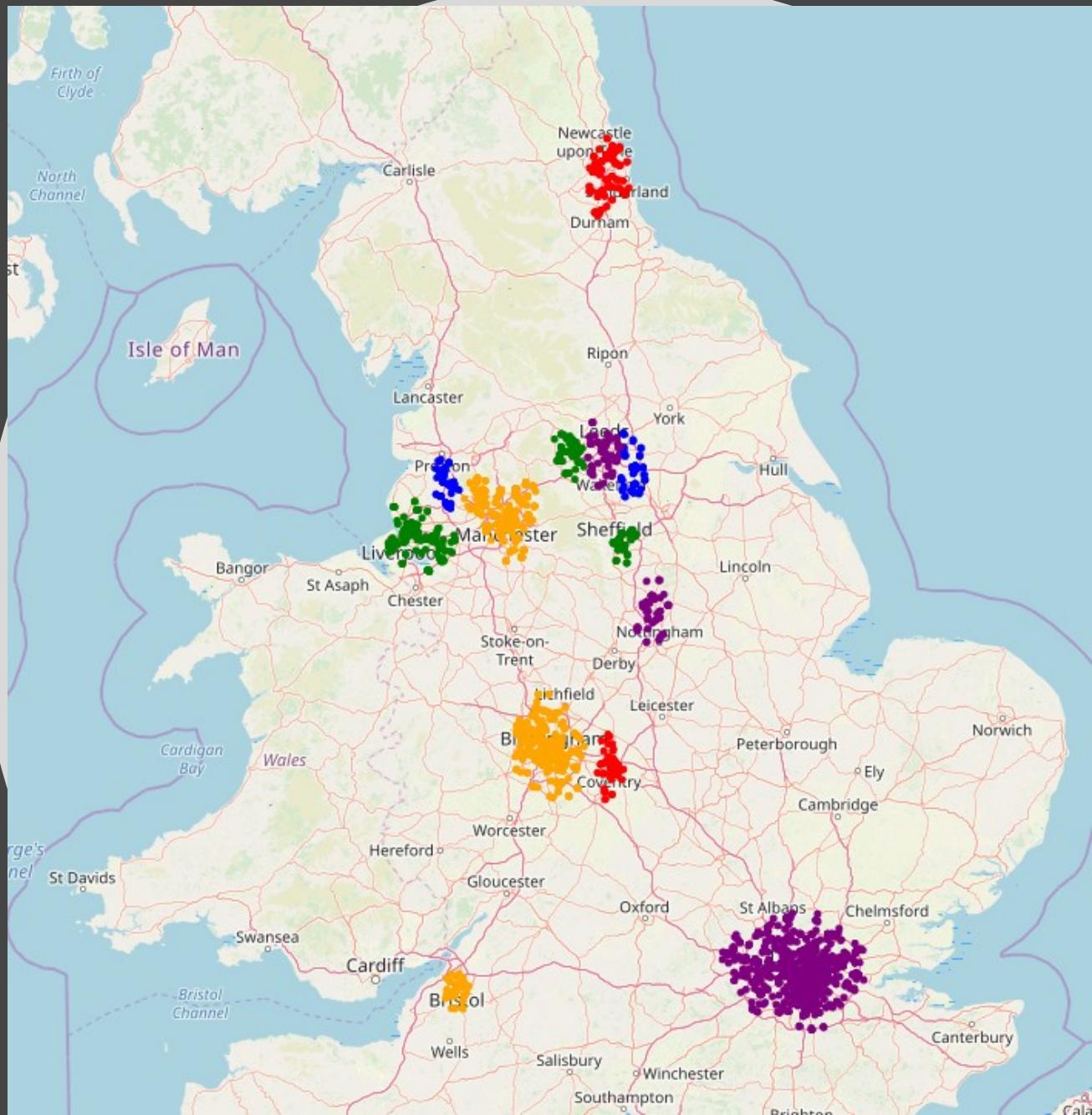
Using the whole dataset





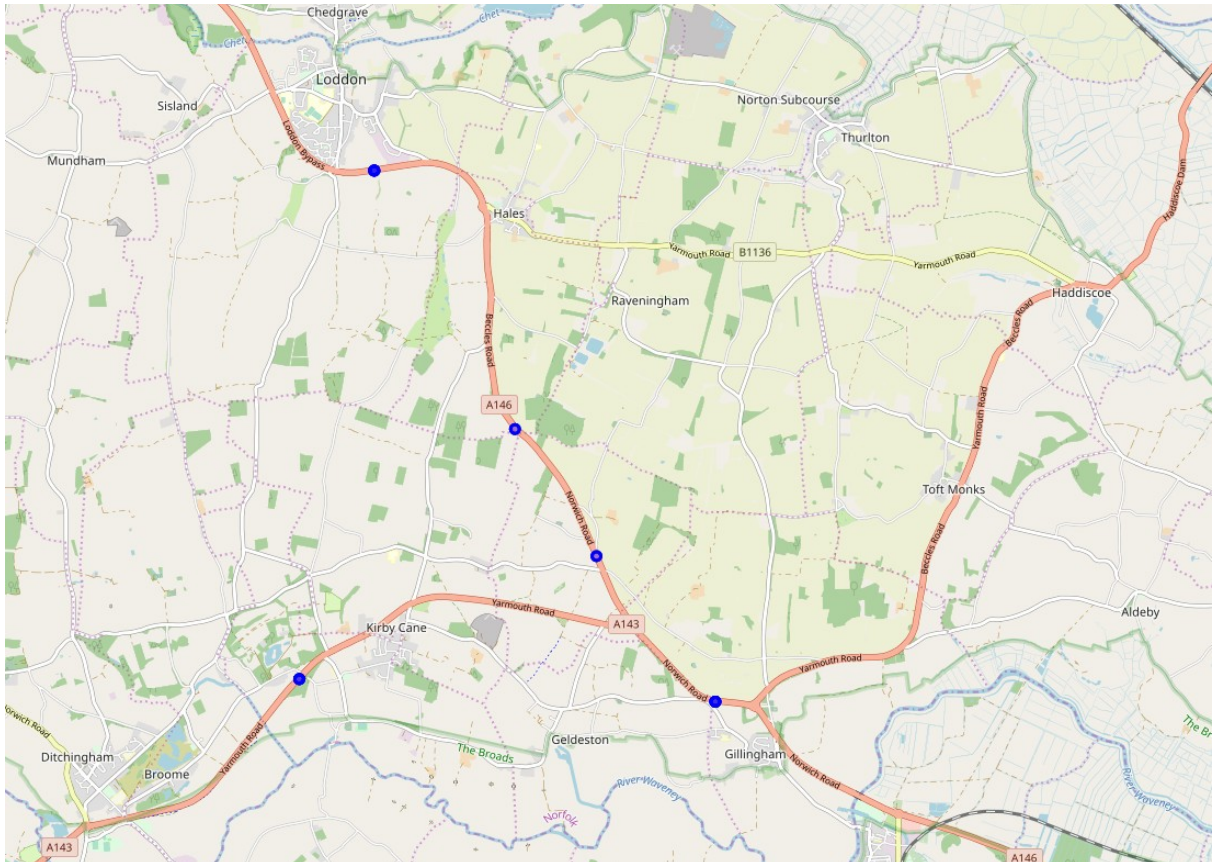
Identifying hotspots

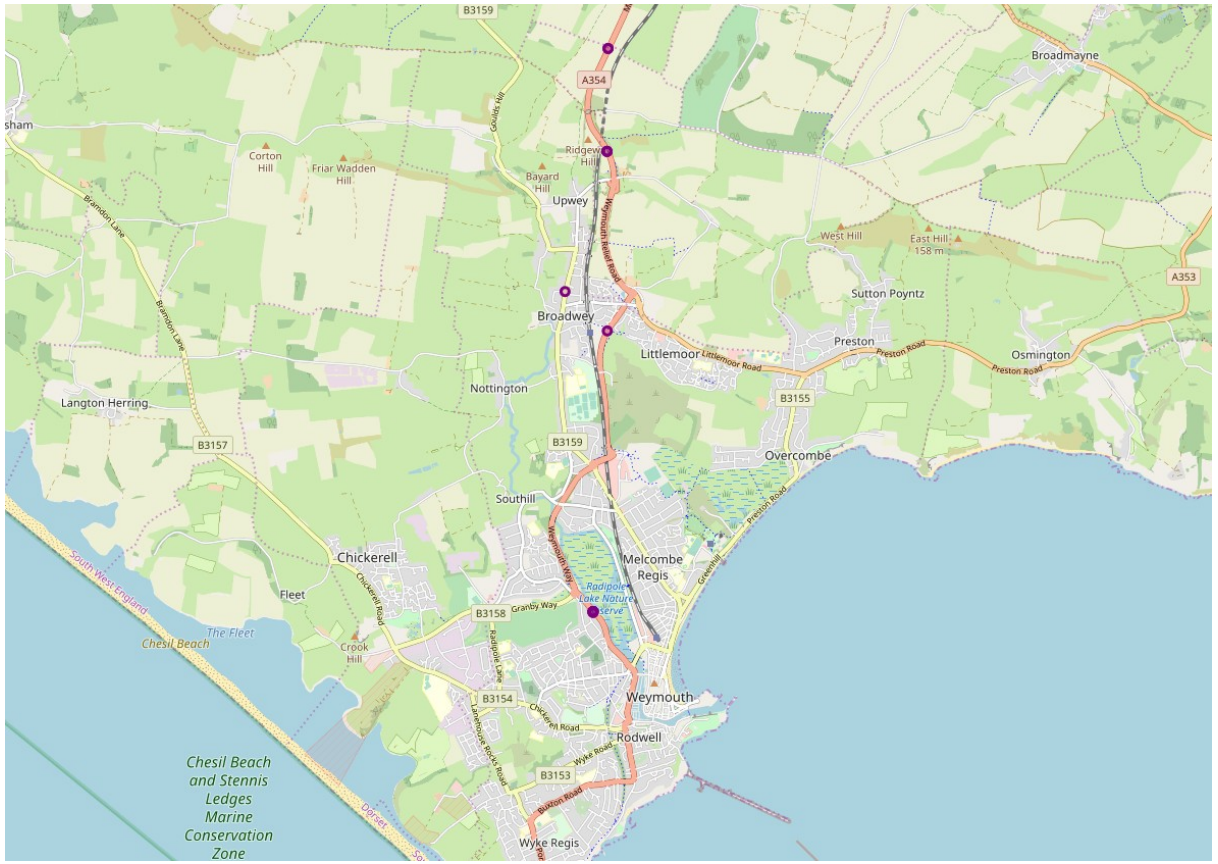
Focusing on the Fatal accidents



Identifying hotspots

Specific hotspots



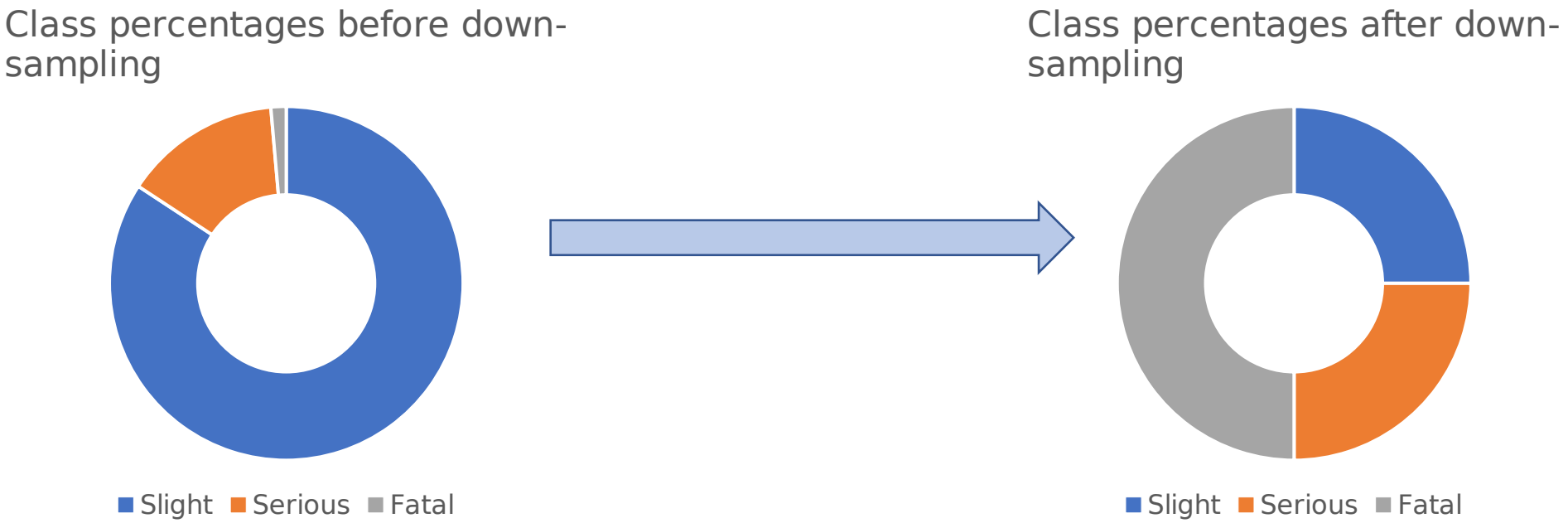


Data preprocessing

- **Dimensionality reduction**
- **Feature extraction**
- **Feature Conversion**
- **Feature Transformation – One-Hot-Encoding**

Detecting patterns for Fatal accidents

- Association rules using FP Growth algorithm
- Applying the data preprocessing slightly changed
  - No One-Hot-Encoding
  - Values renaming
- Class imbalance
  - Downsampling
- 50 iterations of the whole process



Detecting patterns for Fatal accidents

- ('Fine without high winds', 'local_auth_5') **0.98**
- ('Fine without high winds', 'No physical crossing within 50 meters', 'local_auth_9') **0.99**
- ('Darkness: No street lighting', 'No physical crossing within 50 meters', 'Rural') **0.76**
- ('1st_road_class_3', 'Fine without high winds', 'No physical crossing within 50 meters', 'Rural', 'Single carriageway', 'speed_60') **0.72**

Creating a model that predicts Fatal accidents

Resampling technique

- **Downsampling**

Models used

- **RandomForest**
- **SGDClassifier**

Tuning the classifiers

- **GridSearchCV**

Creating a model that predicts Fatal accidents

Fatal metrics considering 3 different types of accident:
Slightly, Serious and Fatal

RESULTS FOR DATASET 1

METRIC	Stochastic SVM	Randomforest
PRECISION	47.49%	57.48%
RECALL	80.26%	67.59%
F-MEASURE	59.63%	62.13%
ACCURACY	48.98%	53.17%

Creating a model that predicts Fatal accidents

Fatal metrics considering 2 types of accidents:
slightly or more serious one

RESULTS FOR DATASET 2

METRIC	Stochastic SVM	Randomforest
PRECISION	60.57%	56.46%
RECALL	66.53%	65.52%
F-MEASURE	63.07%	60.65%
ACCURACY	61.36%	52.82%

Creating a model that predicts Fatal accidents (additional experiments)

Upsampling technique

- RandomOverSampler
- Synthetic Minority Over-sampling Technique (SMO

Class_weighted parameter

Outlier Mining