2022年10月24日 11:48

Update since Sep 29th

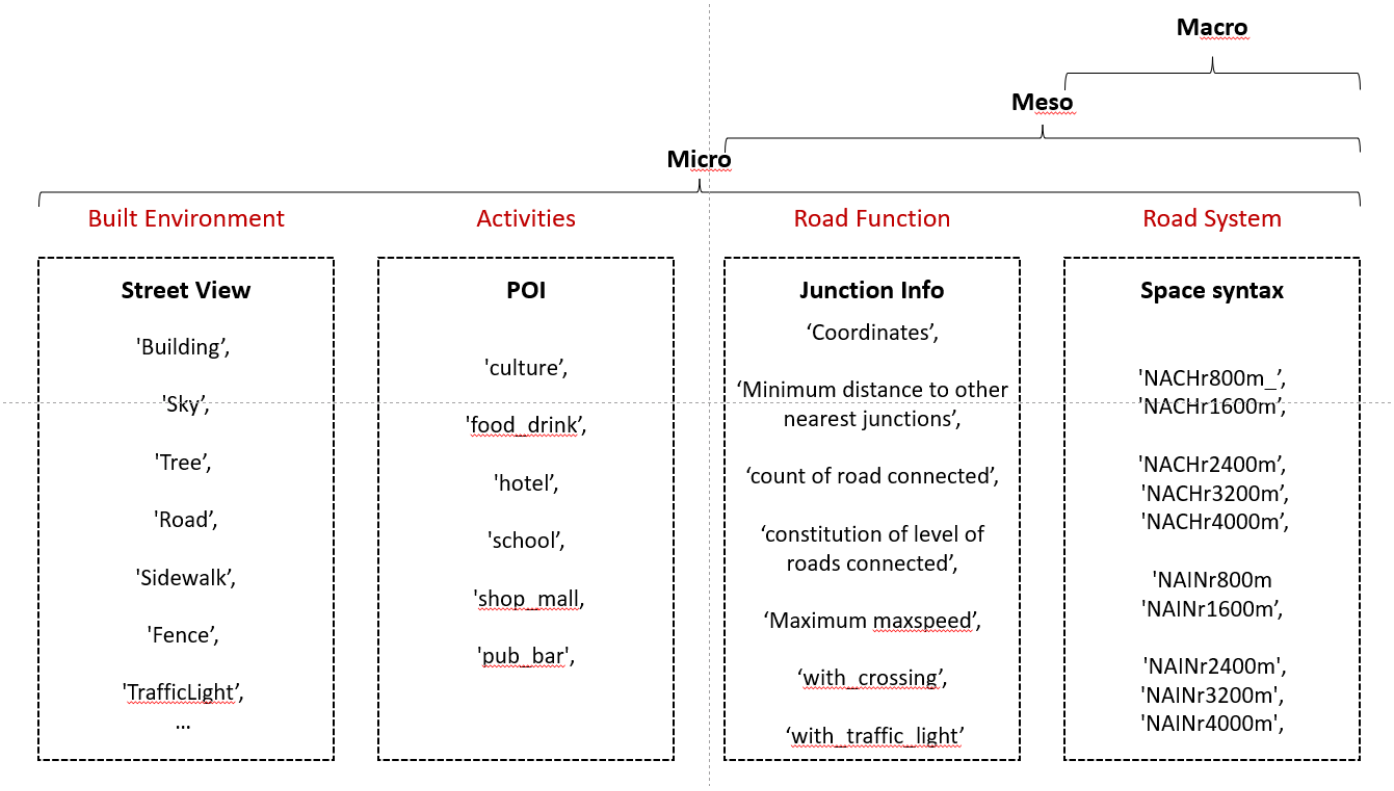
1. Considering that there is always **unbalance in junctions with accidents and junctions without accident**s, two analysis are further carried out to reduce the quantity of junctions without accidents.

   a. **Reset the searching radius of each junction for road accidents from 20m to 30m**, to have more junctions with accidents. I think this is acceptable, because in the real driving scenario, drivers have to make response to junctions earlier before they arrive junctions. Some accidents can also happen in this progress.
   b. **Delete dead ends from the current junction set,**
   These dead ends are ignored in previous analysis. In fact there are few accidents happen in dead ends and they don't intersect/ connect with other roads . Now the total count of junctions are **1877(** 2303 for previous version).

2. Add a new **junction dimension**(and a bunch of related variables) in classifiers.
   Details can be found in the graph below.

   Maxspeed of road, crossing, and traffic light data are collected from osm;
   Road level data are collected from Ordnance Survey road centre line

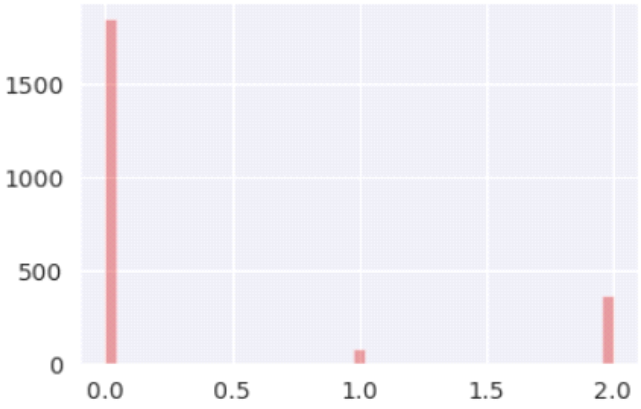
   Just think a junction dimension can be necessary for a study on junction classification?

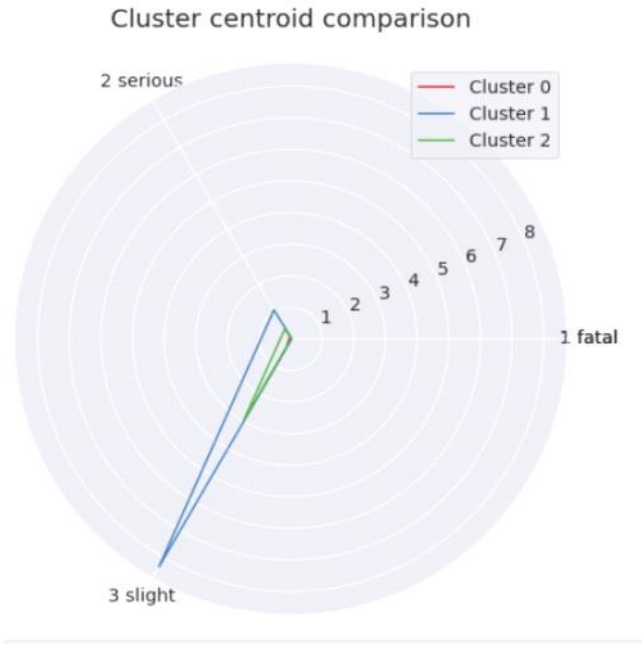


3. **Test the k-means clustering method**

Have tested the k-means clustering method and the count of slight/serious/fatal accidents are used as three dimensions.

a. The main difference between clusters is the total count of accidents
b. Significant difference in sample count included in each clusters
c. Randomforest classifier receive unbalanced result. Poor performance on severe/fatal accident featured clusters.





Smote method + without tuning

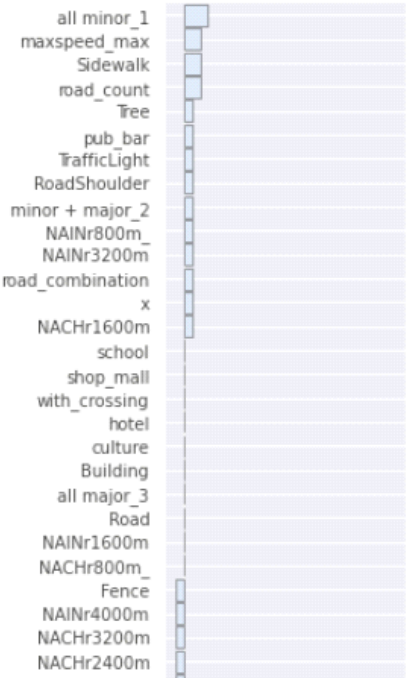
	precision	recall	f1-score	support
0	0.87	0.94	0.90	140
1	0.25	0.09	0.13	11
2	0.58	0.51	0.54	37
accuracy			0.80	188
macro avg	0.56	0.51	0.53	188
weighted avg	0.77	0.80	0.79	188

SMOTE + Without tuning

	precision	recall	f1-score	support
0	0.91	0.91	0.91	140
1	0.45	0.45	0.45	11
2	0.56	0.54	0.55	37
accuracy			0.81	188
macro avg	0.64	0.64	0.64	188
weighted avg	0.81	0.81	0.81	188

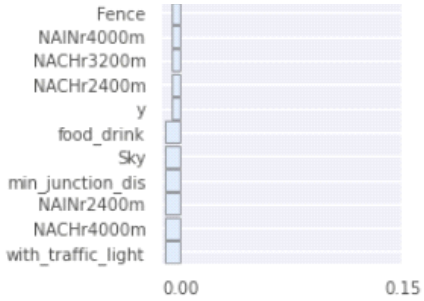
SMOTE + Tuning

SMOTE + Tuning
Feature Importance



SMOTE + Tuning

	precision	recall	f1-score	support
0	0.91	0.89	0.90	140
1	0.45	0.45	0.45	11
2	0.54	0.57	0.55	37
accuracy			0.80	188
macro avg	0.63	0.64	0.64	188
weighted avg	0.81	0.80	0.81	188



4. **Classification method**

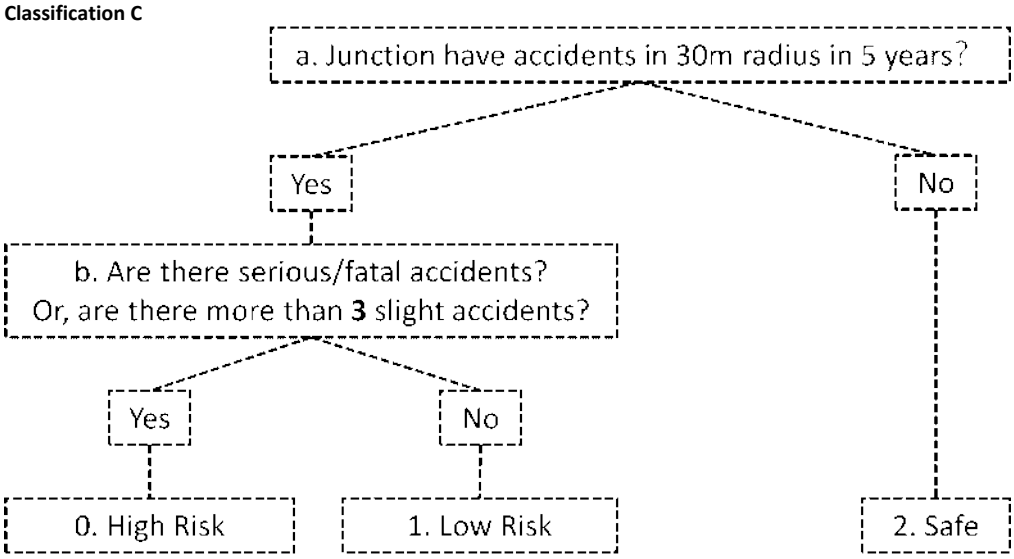
Have tested the **classification C** with new variables and with more balanced dataset

a. The classification precision and recall for high risk junctions have increased compared to previous version.
b. For safe and low risk junctions, seems there is a lack of supporting evidence.
c. Poor performance in classifying low-risk junctions

Have raised another classification **method E**.
The **method E** also consider accidents happened on road segments(beyond the junction) and provide more rules to define low risk and safe junctions.

a. Good performance in classifying high-risk and low risk juncitons and relatively low performance in safe juncitons
b. Relatively more balanced performance for the overall model.

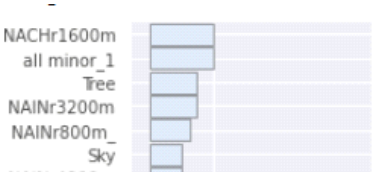
Classification C



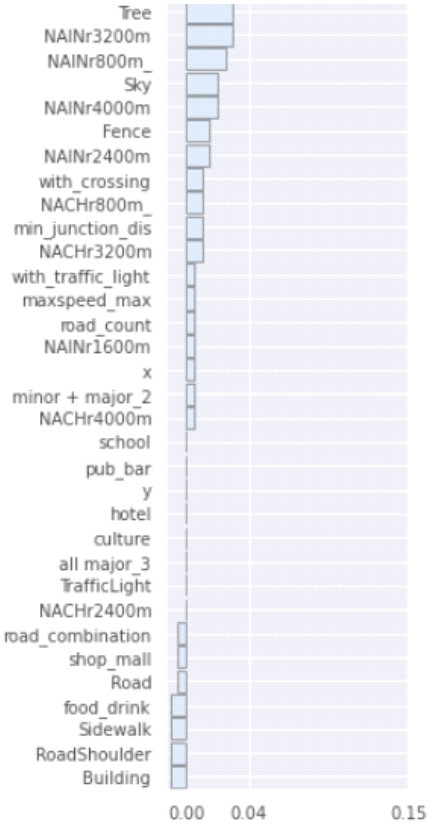
Smote method + without tuning

	precision	recall	f1-score	support
0	0.80	0.86	0.83	100
1	0.31	0.26	0.28	39
2	0.71	0.71	0.71	49
accuracy			0.70	188
macro avg	0.61	0.61	0.61	188
weighted avg	0.68	0.70	0.69	188

SMOTE + Tuning
Feature Importance



accuracy			0.70	188
macro avg	0.61	0.61	0.61	188
weighted avg	0.68	0.70	0.69	188



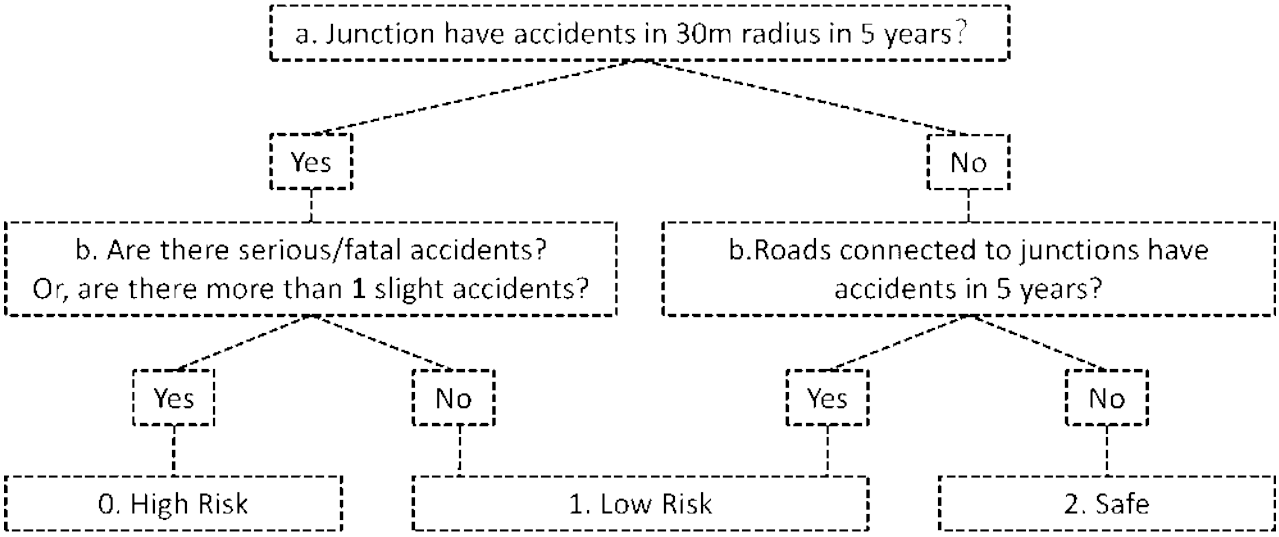
SMOTE + Without tuning

	precision	recall	f1-score	support
0	0.80	0.86	0.83	100
1	0.31	0.26	0.28	39
2	0.71	0.71	0.71	49
accuracy			0.70	188
macro avg	0.61	0.61	0.61	188
weighted avg	0.68	0.70	0.69	188

SMOTE + Tuning

	precision	recall	f1-score	support
0	0.80	0.87	0.83	100
1	0.37	0.28	0.32	39
2	0.73	0.73	0.73	49
accuracy			0.71	188
macro avg	0.63	0.63	0.63	188
weighted avg	0.69	0.71	0.70	188

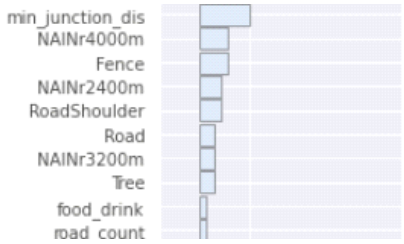
Classification E



Without SMOTE

	precision	recall	f1-score	support
0	0.67	0.31	0.43	32
1	0.69	0.86	0.76	97
2	0.81	0.71	0.76	59
accuracy			0.72	188
macro avg	0.72	0.63	0.65	188
weighted avg	0.72	0.72	0.70	188

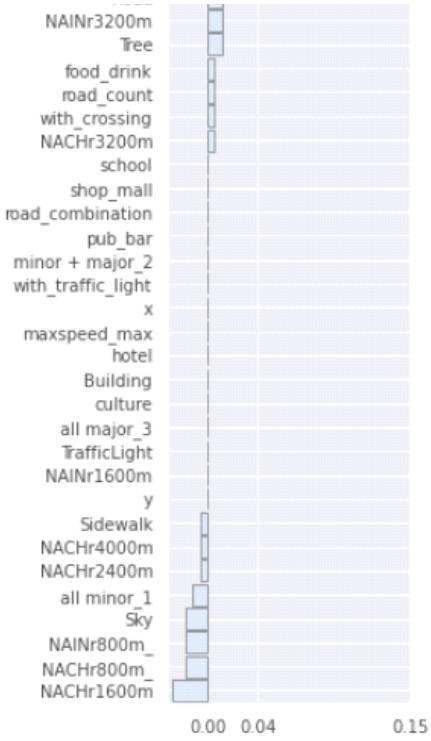
SMOTE + Tuning
Feature Importance



SMOTE + Without tuning

	precision	recall	f1-score	support
0	0.54	0.59	0.57	32

	precision	recall	f1-score	support
0	0.54	0.59	0.57	32
1	0.70	0.70	0.70	97
2	0.77	0.73	0.75	59
accuracy			0.69	188
macro avg	0.67	0.67	0.67	188
weighted avg	0.70	0.69	0.69	188



SMOTE + Tuning

	precision	recall	f1-score	support
0	0.54	0.59	0.57	32
1	0.70	0.71	0.70	97
2	0.78	0.71	0.74	59
accuracy			0.69	188
macro avg	0.67	0.67	0.67	188
weighted avg	0.70	0.69	0.69	188