Predicting the car severity accident of Seattle

Daniel Souza Lima

Predicting car accident severity is valuable for society

- Accidents classification ranges from property damage to serous injury/death
- Heavy economic burden of $ 870 billion dollars
- Life preserving, public policies and planning

Data acquisition and cleaning

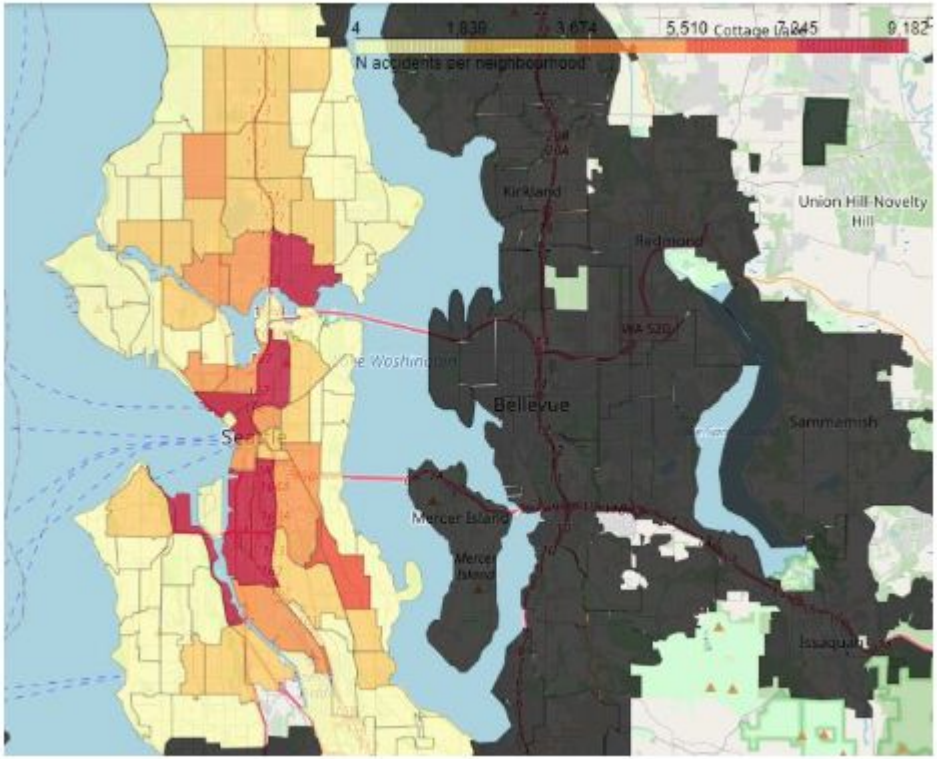
- Data collected in the Seattle City Council Area by the Seattle Department of Transport (SDOT),
- the accident dataset was download from the **Seattle GeoData**,
- time period of 2004-2017,
- in total, 222,389 rows/accidents,
- neighrbourhoods got with a geojson map extracted from this GitHub repository, and
- cleaned data contains 48 columns.

To bear in mind: severity code, the target outcome

- 0, 1: unknown, property damage
- 2: minor injure
- 2b: serious injury
- 3: fatality

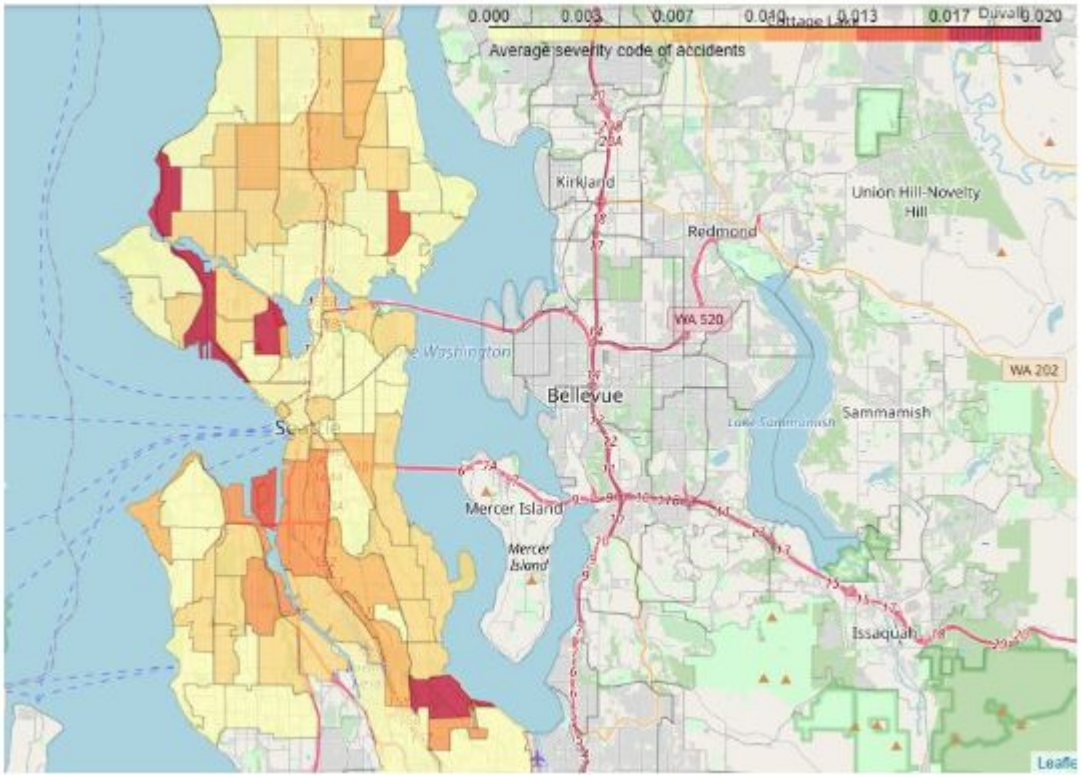
Data visualization

- Belltown has the highest count, and

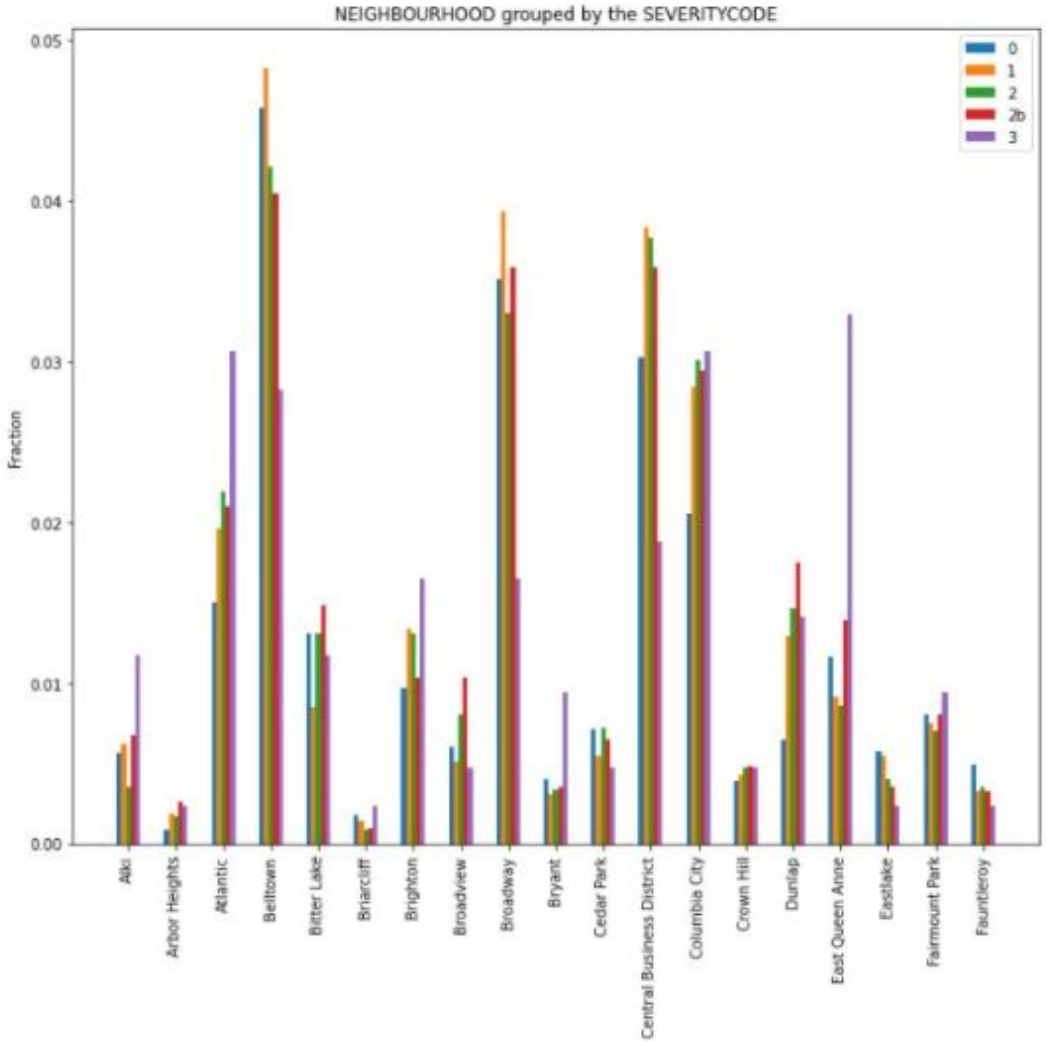


Data Visualization

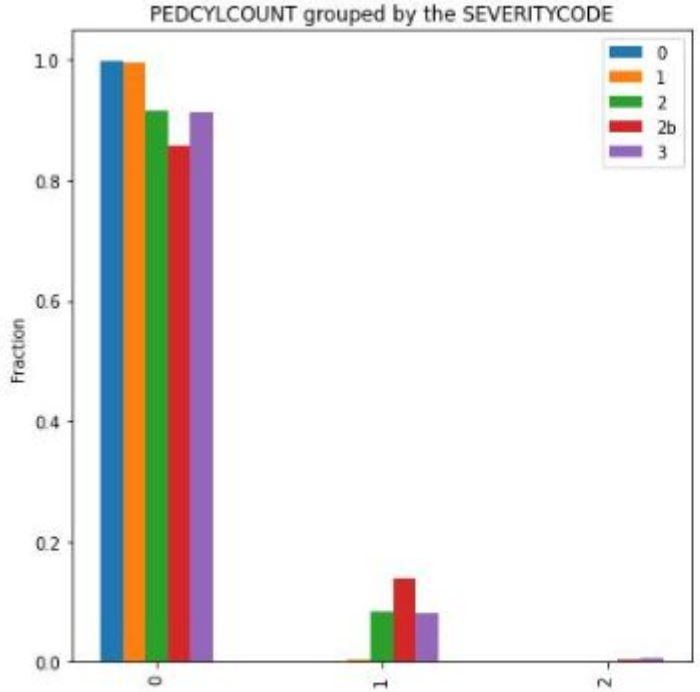
- East Queen Anne has the highest severity.



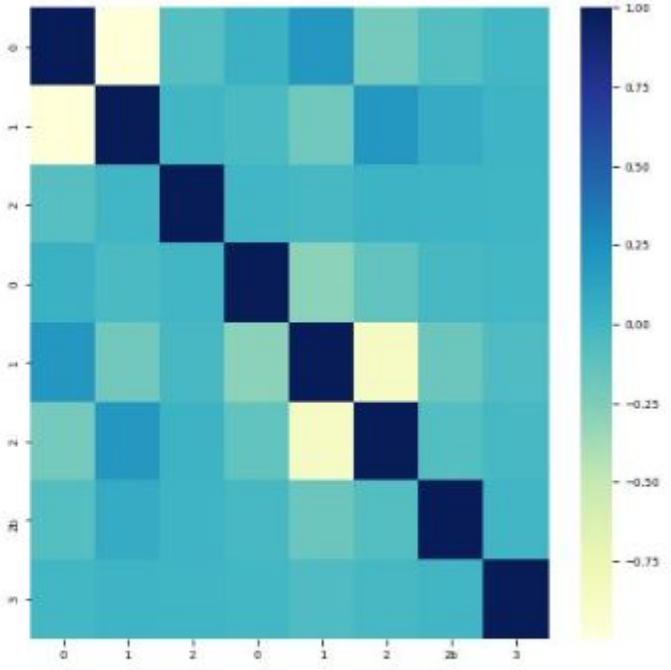
The accidents are distributed unevenly



Cyclists deaths represent 8.88% of the deaths

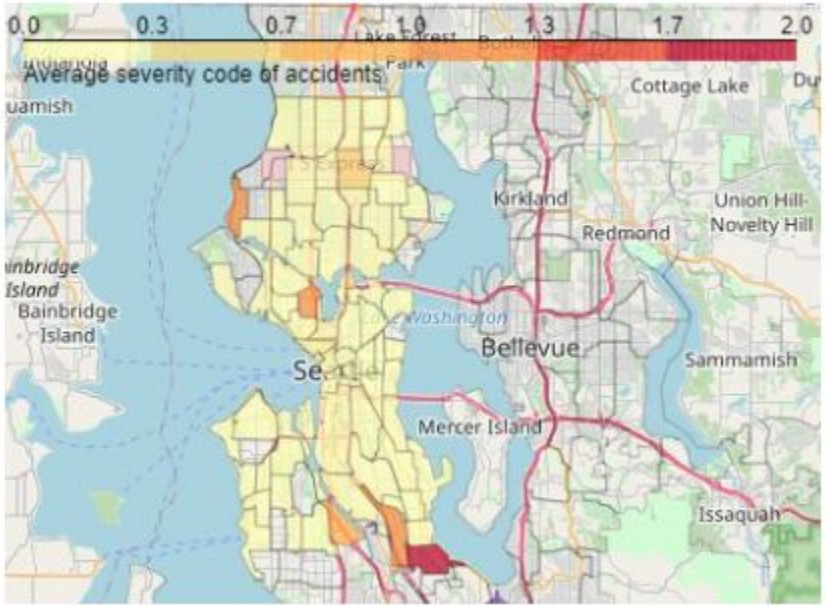


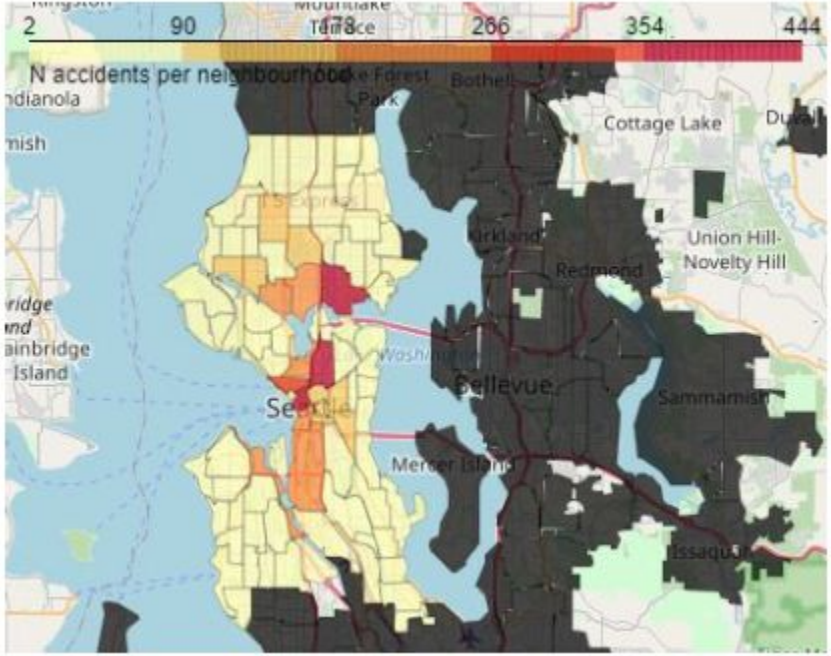
(a) Histogram of PEDCYLCOUNT gouped by the severity codes



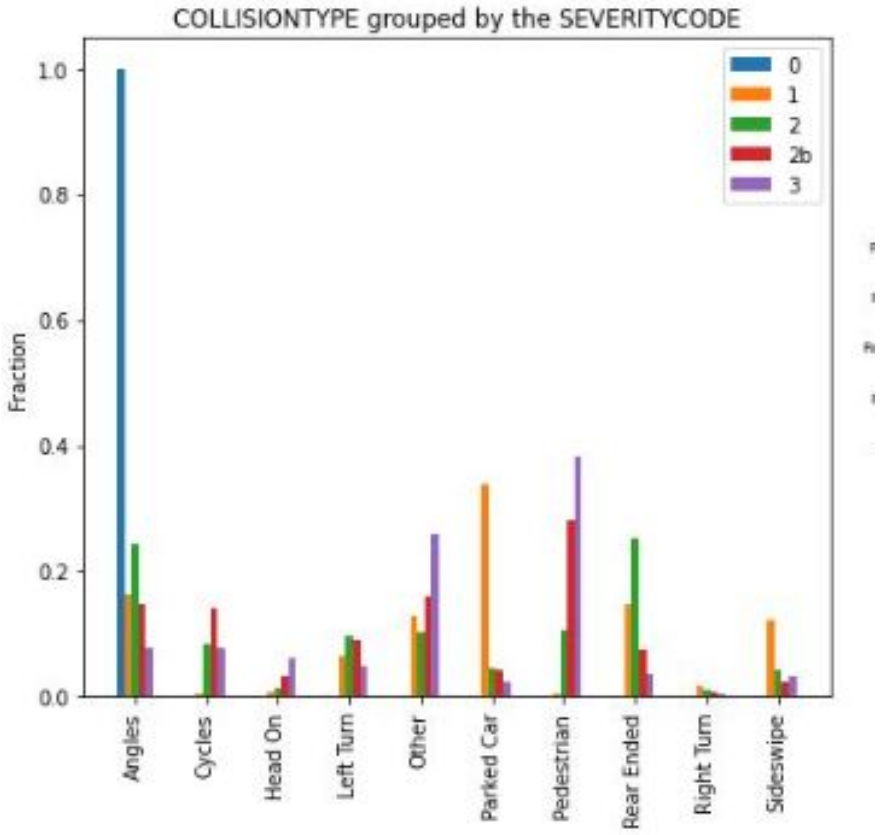
(b) The heatmap of the correlation matrix

University District is cyclist accident spot

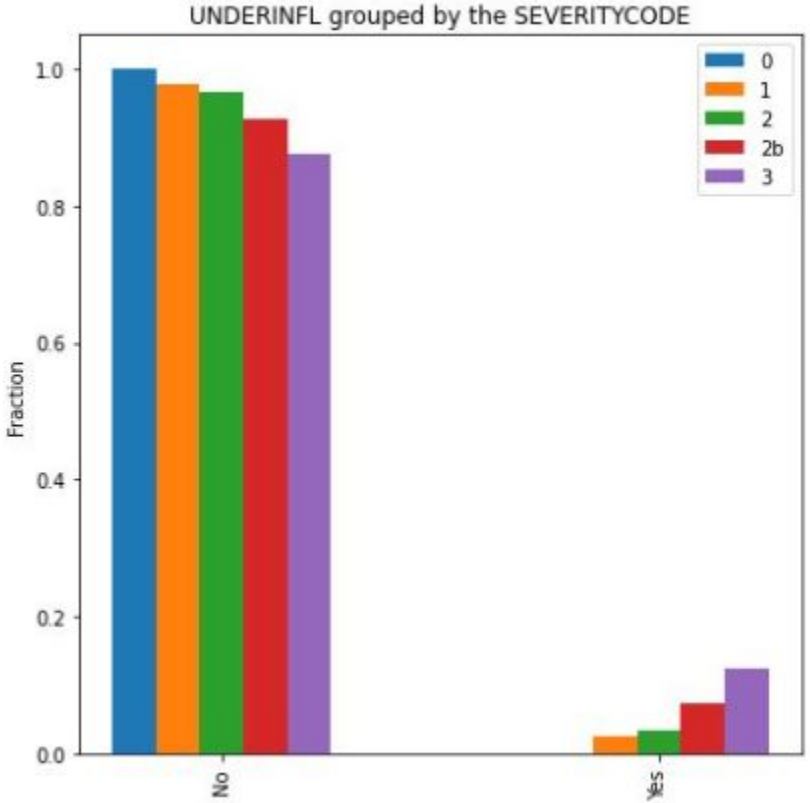




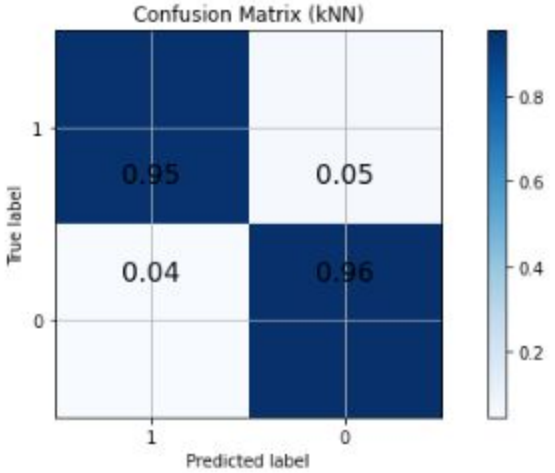
Collisions with pedestrians and cyclists result mostly in deaths



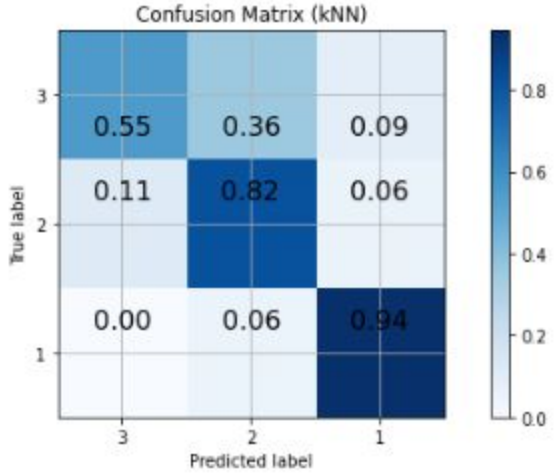
No minor accident with intoxicated driver



Classification models: kNN

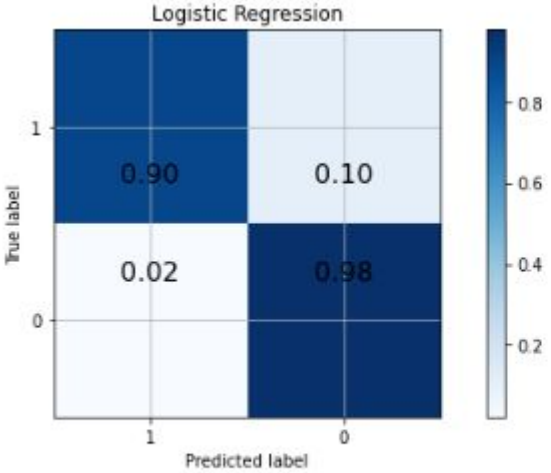


(a) Binary classification with kNN. The best k is 3.

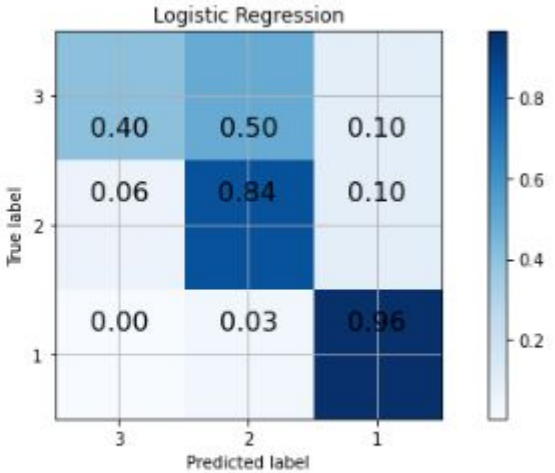


(b) Multi classification with the kNN. The best k is 13.

Classification models: Logistic Regression

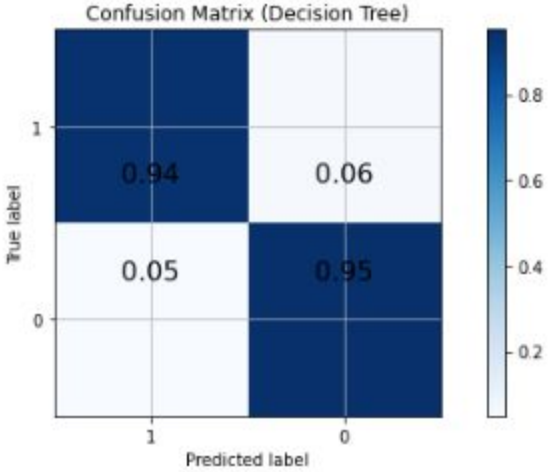


(a) Binary classification.

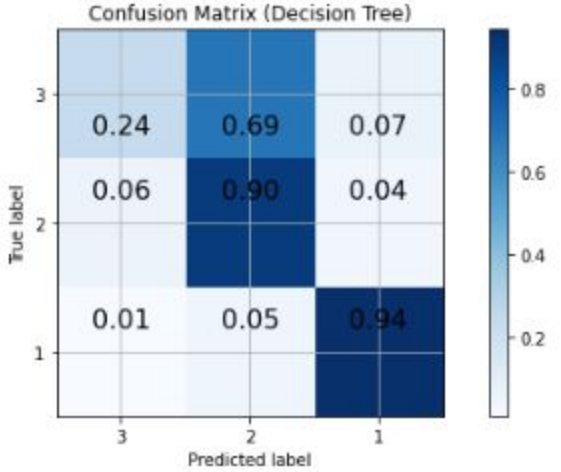


(b) Multi classification.

Classification models: Decision Tree

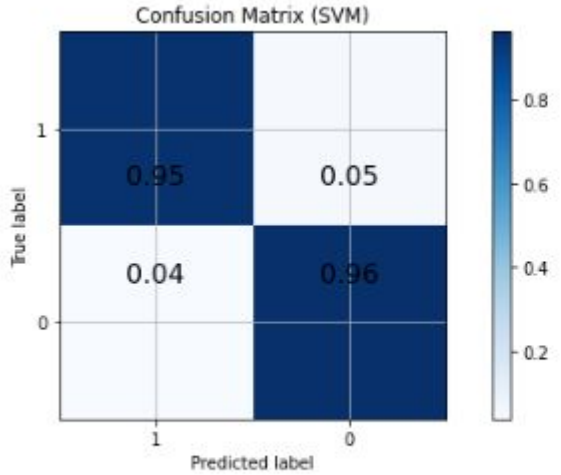


(a) Binary classification.

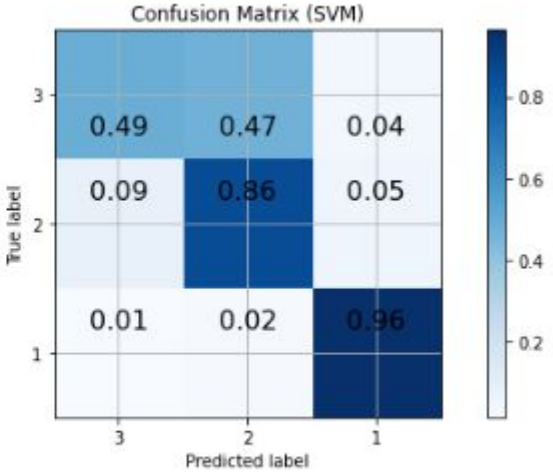


(b) Multi classification.

Classification models: SVM

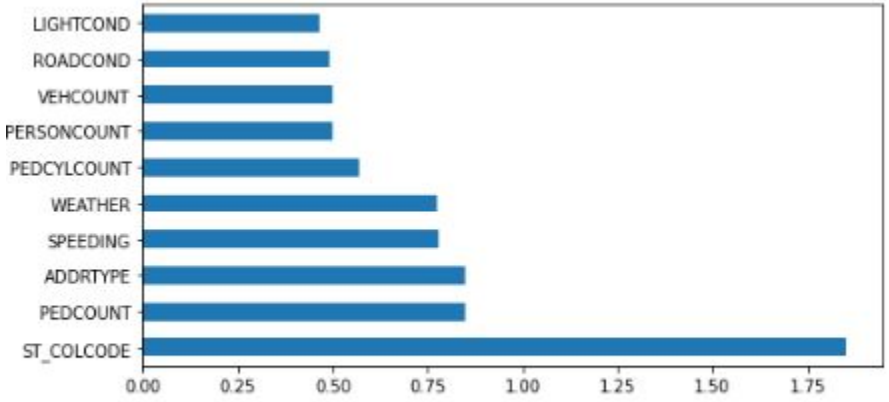


(a) Binary classification.

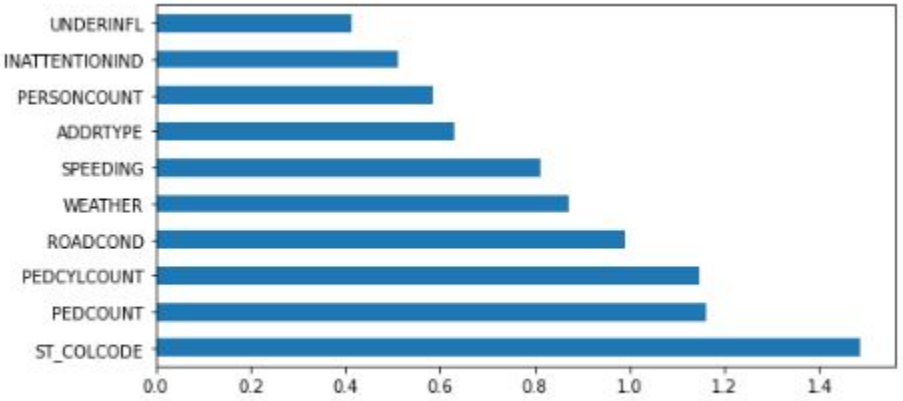


(b) Multi classification.

SVM result: ***mainly, the type of the collision and number of pedestrians determine the final outcome.***



(a) Binary classification.



(b) Multi classification.

Conclusion and future directions

- Built useful models to predict whether and how worse an accident outcome could
- Accuracy is sufficiently high
- Better discrimmination of the factors could be attained:

  - the type of the collision

- Separate the dataset in time intervals: the accident frequency may not be homogeneous in time