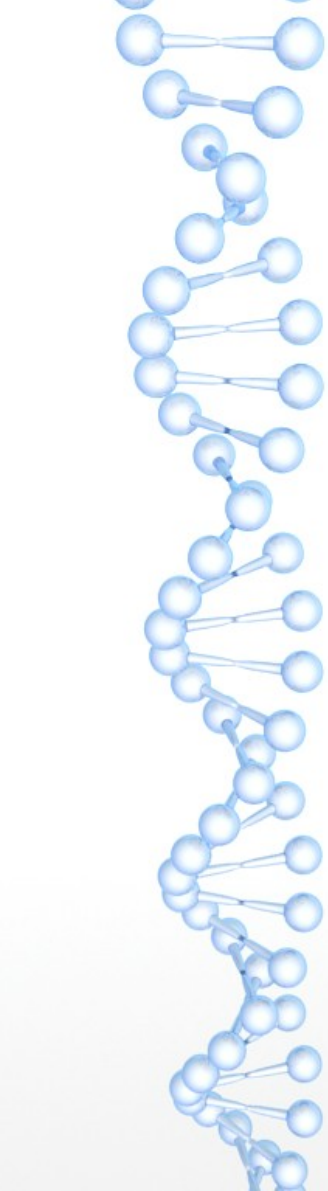


PREDICTION OF THE SEVERITY OF A TRAFFIC INCIDENT.

Problem origin

- Traffic accidents are tragedies that cost money, resources and even human lives.
-
- Have negative consequencies beyond the accident itself.
  - Traffic congestion.
  - Other accidents because of traffic diversion.
- Many factors involved, not all of them avoidable.

Possible target & stakeholder

- Governments need to find a solution to this problem.
- They must maintain the citizens safe.
- Saving resources as medical ones or money is essential.

Data source

- The original database is related to traffic collisions of the city of Seattle, provided by the state.

- The actual database is a simplified version of the original one.

Data understanding

- The database consist of 194673 cases of different kind of accidents.
- Each accident (or row) has 38 features.
- The main feature is the severity code, the aim for prediciton of this project.

Choosing model

- In this problem the aim is to predict the severity index of an accident given certain information of it.
- This is a classification kind of problem.
- Due to the non-binary target label (severity index) not all classification algorithms are valid

Choosing model

- Two algorithms are used.
  - K-Nearest Neighbors (KNN)
  - Support Vector Machine (SVM)

- Both models are used for comparison.

Select features

- The selected features for the model can be divided in two groups, focusing on the driver.
  - Out of control: type of collision, the weather, road condition, light condition, etc.
  - In control: being under the influence of alcohol or drugs, speeding, distractions, etc.

Data preparation

- Many features are categorical ones, as they represent a situations or a status.

SEVERITY CODE	ADDRTYPE	COLLISIONTYPE	PERSONCOUNT	VEHCOUNT	PEDROWNOTGRNT	UNDERINFL	INATTENTIONIND	WEATHER	ROADCOND	LIGHTCOND	SPEEDING
2	Intersection	Angles	2	2	NaN	N	NaN	Overcast	Wet	Daylight	NaN
1	Block	Sideswipe	2	2	NaN	0	NaN	Raining	Wet	Dark-Street lights on	NaN
1	Block	Parked car	4	3	NaN	0	NaN	Overcast	Dry	Daylight	NaN
1	Block	Other	3	3	NaN	N	NaN	Clear	Dry	Daylight	NaN
2	Intersection	Angles	2	2	NaN	0	NaN	Raining	Wet	Daylight	NaN

Data preparation

- They must be turned into numerical features, labelling their values with numbers.

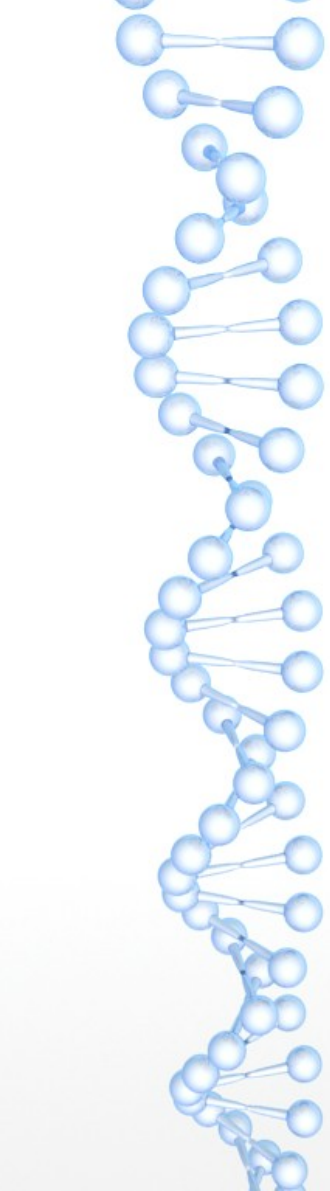
SEVERITY CODE	ADDRTYPE	COLLISIONTYPE	PERSONCOUNT	VEHCOUNT	PEDROWNOTGRNT	UNDERINFL	INATTENTIONIND	WEATHER	ROADCOND	LIGHTCOND	SPEEDING
2	0	0	2	2	0	1	0	0	0	0	0
1	1	1	2	2	0	0	0	1	0	1	0
1	1	2	4	3	0	0	0	0	1	0	0
1	1	3	3	3	0	1	0	2	1	0	0
2	0	0	2	2	0	0	0	1	0	0	0

KNN - model

- After training the KNN model, with k=44, the results are tested.
  - Mean accuracy of ≈ 0,753.
  - Jaccard index of ≈ 0,753.
  - F1-score of ≈ 0,843.

SVM - Model

- After training the SVM model, the results are tested.
  - Mean accuracy of ≈ 0,755.
  - Jaccard index of ≈ 0,755.
  - F1-score of ≈ 0,847.



Discussion & Conclusion

- Mean prediction success for both models of around the 75%.
- Acceptable but not ideal, need to test other techniques or features to be sure of the limits of this approach.