

Capstone Project - Car accident severity

GradientBoostingClassifier, KNeighborsClassifier, and SVC

Too many people die in traffic accidents

People die in traffic accidents because of risks that cannot be eliminated.

But fatal accidents also happen due to avoidable risks.

These avoidable risks are identified for the city of Seattle (WA) to reduce the risk for road users – to save human lives.



Data acquisition

- The Seattle collisions data set contains traffic accidents in Seattle from 2004 to 2020.
- The data set can be found here: <https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv>

- SEVERITYCODE (severity of collision) is used as the label.
- The features include ADDRTYPE (collision address type), ROADCOND(ition), and LIGHTCOND(ition).

	SEVERITYCODE	ADDRTYPE	ROADCOND	LIGHTCOND
count	187630.000000	187630	187630	187630
unique	NaN	3	9	9
top	NaN	Block	Dry	Daylight
freq	NaN	123400	123808	115448

Encoding categorical data

	SEVERITYCODE	LIGHTCOND	Alley	Block	Intersection	Dry	Ice	Oil	Other	Sand/Mud/Dirt	Snow/Slush	Standing Water
0	2	Daylight	0	0	1	0	0	0	0	0	0	0
1	1	Dark - Street Lights On	0	1	0	0	0	0	0	0	0	0
2	1	Daylight	0	1	0	1	0	0	0	0	0	0
3	1	Daylight	0	1	0	1	0	0	0	0	0	0
4	2	Daylight	0	0	1	0	0	0	0	0	0	0
...
194668	2	Daylight	0	1	0	1	0	0	0	0	0	0
194669	1	Daylight	0	1	0	0	0	0	0	0	0	0
194670	2	Daylight	0	0	1	1	0	0	0	0	0	0
194671	2	Dusk	0	0	1	1	0	0	0	0	0	0
194672	1	Daylight	0	1	0	0	0	0	0	0	0	0

GradientBoostingClassifier

● Accuracy of 0.6723
~ 67%

True Neg 22,968	False Pos 6
False Neg 11,189	True Pos 2

KNeighborsClassifier

● Accuracy of 0.6599
~ 66%

With the selected value k-nearest neighbors '3' the best forecast accuracy was achieved.

True Neg 21,373	False Pos 1601
False Neg 10,019	True Pos 1,172

SVC

● Accuracy of 0.6723
~ 67%

With the selected values SVC kernel 'poly' and SVC degree '8' the best forecast accuracy was achieved.

True Neg 22,966	False Pos 8
False Neg 11,187	True Pos 4

Conclusion and remarks

- The predictability of GradientBoostingClassifier, KNeighborsClassifier, and SVC strongly depend on the choice of the values for the hyperparameters.
- ADDRTYPE (alley, block, intersection), ROADCOND(ition), and LIGHTCOND(ition) together can explain to some extent the label SEVERITYCODE.
- Maximum speed limits must be dynamically adapted to the road conditions (weather) and lighting conditions, and poorly lit areas of the street need to be better lit to decrease the severity of accidents.