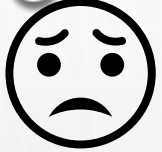


My Cool Capston Project:

# Road collisions in Montreal – Canada

By G.Fontecha

## Problem description:



**Predicting road accidents** in Montreal – Canada is quite uncertain because the relationship between the multiple factors causing an accident is undetermined.



If we could predict traffic accidents, **corrective actions** could be taken to avoid them.



The good news is that we have **data** describing the characteristics of more than 190000 accidents since 2012.



The difficulty is that influencing factors are just **too many and complex** to analyze them by simple observation.



The proposed solution is to build a prototype machine learning model to **predict road accidents and its severity, given at least weather and road conditions.**

My Cool Capston Project:

## Road collisions in Montreal - Canada

### Data description

The data-set has is downloaded from the web site of the Government of Canada.  
The data-set is composed by 190552 entries, each with 68 attributes

One of the attributes is the severity of the collision, it has the following values,

Value ID	Label	Entries
1	Minor material damages	78300
2	Major material damages	71670
3	Minor injured persons	38831
4	Major injured persons	1543
5	At least one fatality within the next 30 days	208

The number of entries seems to be enough to use the severity attribute to train and test the machine learning model, for further prediction

## As suggested, the road condition attribute is analyzed first

The road condition attribute has the following entries

Condition ID	Description
11	Dry
12	Humid
13	Water accumulation
14	Sand over the road
15	Melt snow
16	Snow
17	Hard snow
18	Icy
19	Muddy
20	Oily
99	Other

## Data assessment:

The correlation indexes to severity are:

Severity	Correlation
1	-0,18
2	-0,18
3	-0,2
4	-0,18
5	-0,85

Meaning that the linear relationship between severity and the road condition is not good.

## Then, the weather condition attribute

The weather condition attribute has the following entries

Condition ID	Description
11	Clear
12	Partially cloudy
13	mist
14	Light Rain
15	Heavy rain
16	Heavy wind

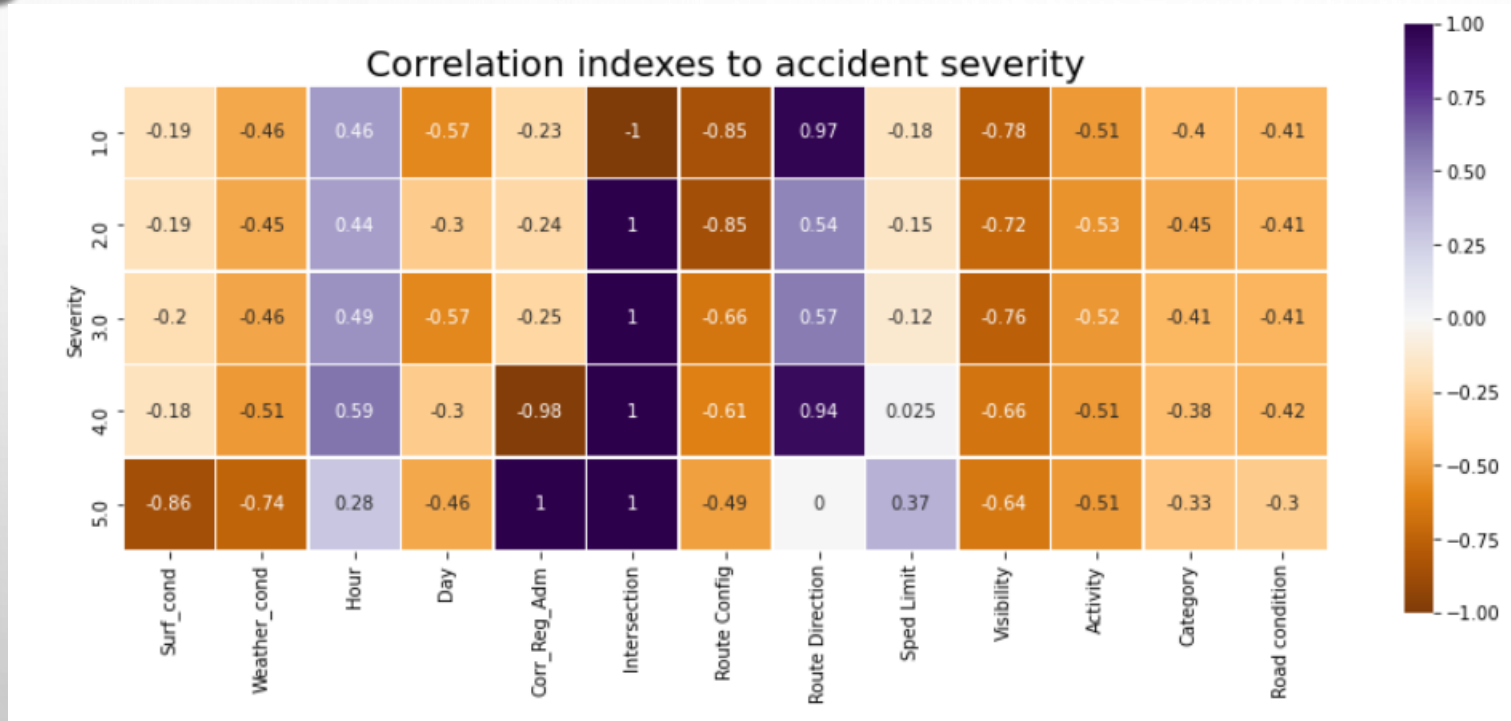
## Data assessment:

The correlation indexes to severity are:

Severity	Correlation
1	-0,46
2	-0,45
3	-0,46
4	-0,50
5	-0,73

Meaning that the linear relationship between severity and the weather condition is fair.

## Finding attributes with better correlation indexes to accident severity



The correlation heat map shows that there are attributes with better linear relationship, these are shown at the center with darker color.

It is worth to make a richer predictive model.

## Data preparation

Accident Severity	Code
Accident involving minor material damages	1
Accident involving major material damages	2
Accident involving persons with minor injures	3
Accident involving persons with major injures	4
Accident involving fatalities	5

Surface Condition	Original Code	Replaced by
Dry surface	11	11
Humid surface	12	12
Water accumulation	13	13
Sand	14	14
Melted ice	15	15
Light snow	16	16
Hardened snow	17	17
Ice	18	18
Mud	19	19
Oil	20	20
Other	99	10
Not registered	NaN	Ommited

Weather condition	Original Code	Replaced by
Clear sky	11	11
Cloudy	12	12
Mist	13	13
Rain	14	14
Heavy rain	15	15
Windy	16	16
Snow	17	17
Storm	18	18
Ice	19	19
Other	20	10
Not registered	NaN	Ommited

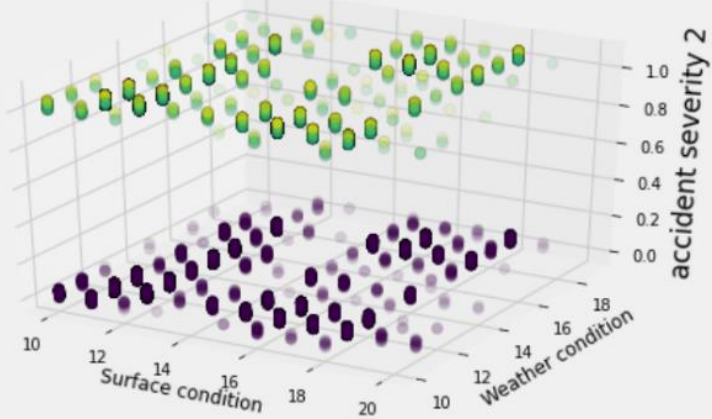


## Other attributes also prepared but not used

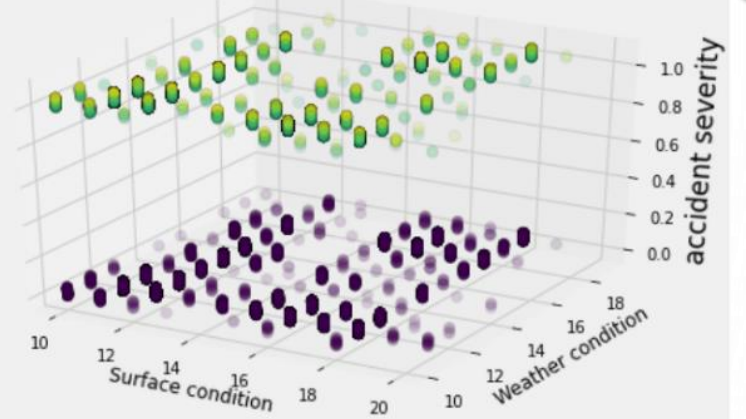
Other attributes	Description
HEURE_ACCDN	Hour of the accident
JR_SEMN_ACCDN	Day of the week
REG_ADM	Region code
TP_REPRR_ACCDN	Proximity to an intersection
CD_CONFIG_ROUTE	Configuration of the road
CD_PNT_CDRNL_ROUTE	Direction of the road
VITESSE_AUTOR	Speed limit
CD_ECLRM	Visibility
CD_ENVRN_ACCDN	Activity of the environment
CD_CATEG_ROUTE	Category of the road
CD_ETAT_CHASS	Road condition

## Scatter plots by category of accident severity

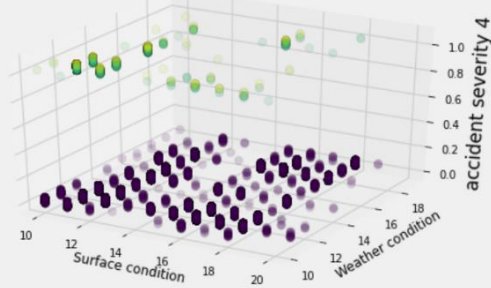
Surface condition and weather condition vs. Accident severity 2



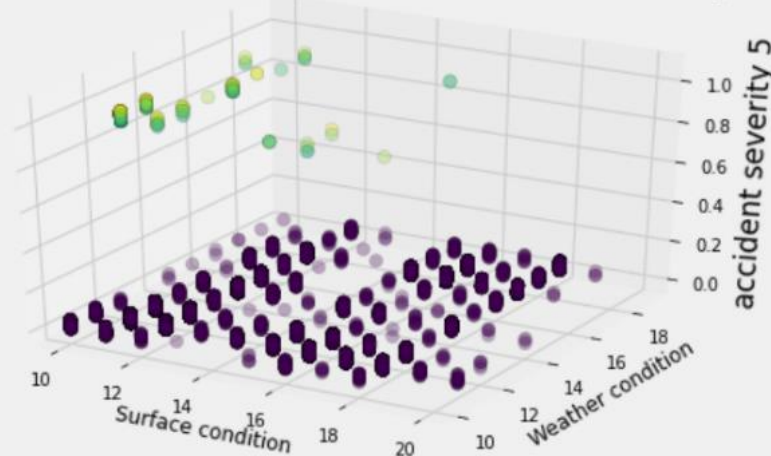
Surface condition and weather condition vs. Accident severity 1



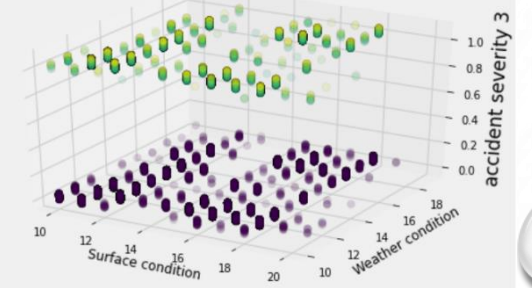
Surface condition and weather condition vs. Accident severity 4



Surface condition and weather condition vs. Accident severity 5



Surface condition and weather condition vs. Accident severity 3





## Confusion matrixes obtained from a Support Vector Machine Model

