Prediction of the severity of an accident using a machine learning algorithm

Prediction of the severity of an accident

- Each time there is an accident, local authorities note a los of details coming form the accident, such as the road conditions, weather people involved, etc

- It could be very interesting to predict the severity of an accident given those conditions in order to warn the local authorities and the driver in order to be proud

- All this can be done using machine learning algorithms

Data acquisition and cleaning

- Data has been obtained from this website: https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv

- Data has been cleaned in order to remove missing values and remove all data which was not interesting for this Project

- At the beggining, the dimension of the dataset was (194673, 38) and after cleaning data it is (189522, 38)

Variables used

- After studying the influence of all variables on the severity of the accidents, the following variables were used (note that the variables have been codified):

|	COLLISIONTYPE	WEATHER	ROADCOND	VEHCOUNT	PERSONCOUNT
0	2	3	2	2	2
1	5	2	2	2	2
2	1	3	1	3	4
3	4	1	1	3	3
4	2	2	2	2	2

Machine Learning algorithm

- As the variables to predict are categorical, a classiffication method has been used, in this case, the Logistic Regression has been selected

- Data has been splited into a test set and a training set in order to train the model and testits accuracy with the predictions

- 20% of the data set was used for testing and 80% of the dataset for training

Results

- After predicting the values of the test set and comparing them with the values of the data set, it has been obtained a value of the Jaccar index=0.71

- The confusión matrix i is shown below:
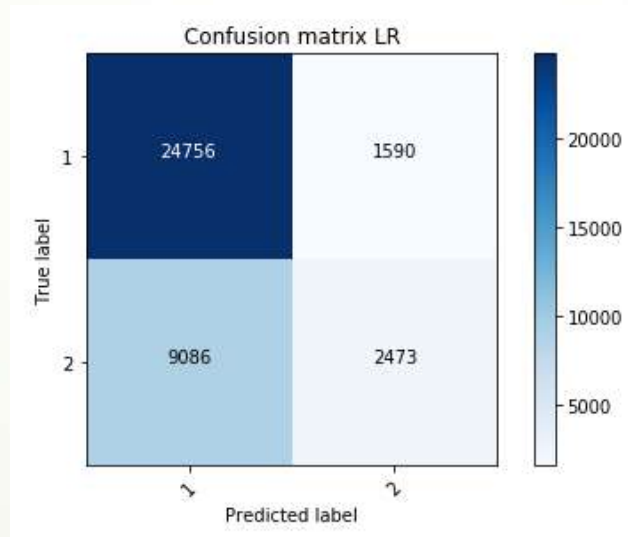
Conclusion

- There is room of improvement for the machine learning algorithm, specially for the variable injury

- We could say, that, according to the jaccard index, the overall precision of the model is 71%

- The model could be trained taking intoaccount other independent variables, such as the light condiions

- The model could be tested with other classification methods such as the KNN or the SVM