Predicting the severity of an accident

Ing. Fernando Villa

CONTENT

- INTRODUCTION
- DATA
- MODELLING
- RESULTS
- CONCLUSIONS
- FUTURE DECISIONS

1. INTRODUCTION

- This project is talk about how act more quickly if an accident occur, for that reason is necessary to predict the severity of the accident.

2. DATA

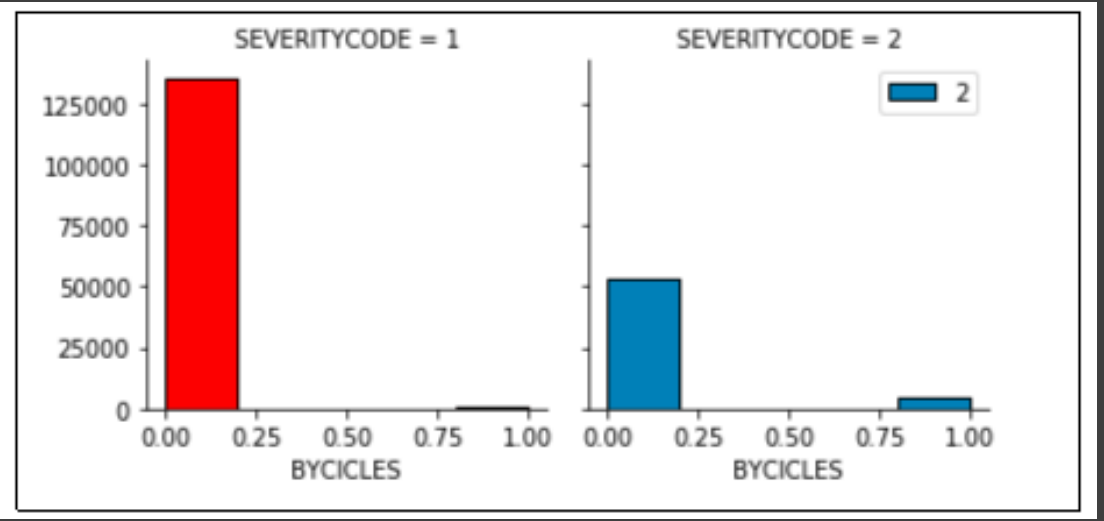
Data Users:

- PERSONCOUNT
- VEHCOUNT
- Bicycle (Calculated)

Other Features:

- TYPEWEATHER (Calculated)
- SPEEDING: (change)
- INCDATE
- INCDTTM
- WEEKDAYTYPE : (Calculated)
- LIGHTTYPE: (Calculated)
- CROSS: (Calculated)

Severity Code (Target)



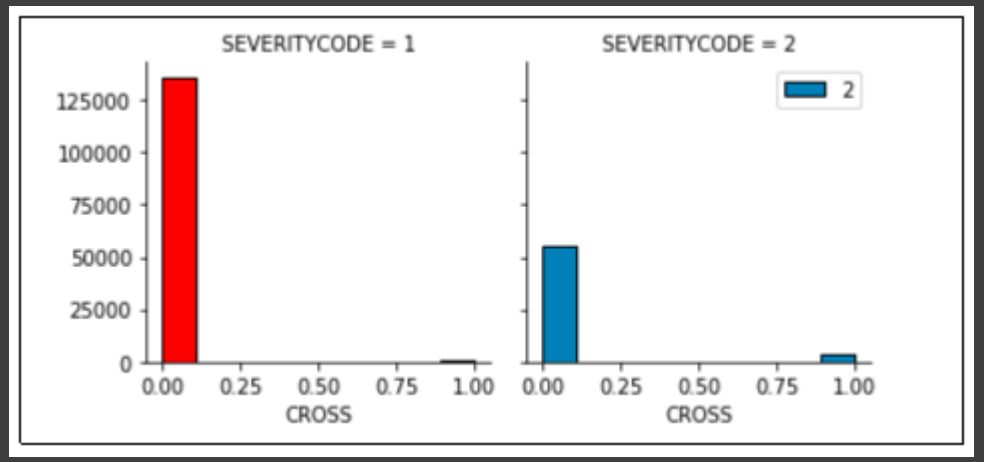
2. DATA

CORRELATION

SEVERITYCODE	1.000000
BYCICLES	0.214702
PEDCYLCOUNT	0.214218
CROSS	0.182314
CROSSWALKKEY	0.175093
PERSONCOUNT	0.130949
SPEEDING	0.038938
dayofweek	-0.015246
WEEKDAYTYPE	-0.017153
VEHCOUNT	-0.054686
TYPEWEATHER	-0.104996
LIGHTTYPE	-0.119548

Only 5 features is choose to explain the severity of accident: BYCICLES, PEDCYLCOUNT,CROSS,PERSONCOUNT and SPEEDING.

CROSS explain better than CROSSWALK KEY and is calculate from that feature. For that reason CROSSWALK KEY is not use.



3. MODELLING

I prefer use a Classification Models and Logistic Regression because the target is binary. Before training the models, I use a 20% to the data to test the value, and 80% to train the model.

The different models that I train are these:

- KNN NEIGHBOARD: with 6 clusters
- DECISION TREE
- SVC
- LOGISTICS REGRESSION

4. RESULTS

Using the method of F1-Score and Jaccard (Accuracy) I have these results:

Algorithm	Jaccard	F1-score	LogLoss
KNN	0.73	0.68	NA
Decision Tree	0.74	0.66	NA
SVM	0.74	0.67	NA
LogisticRegression	0.74	0.67	0.56

About the results, I choose the KNN algorithm because have a better F1-score even if the Jaccard result is a little less than the others. F1- Score is a better to evaluate than the Jaccard method.

5. CONCLUSION

In summary, it was found that the KNN predictive model is more suitable for predicting the severity of an accident. Additionally, the model considers only 5 variables, so it makes prediction more efficient (since too much data should not be collected) and would allow the authorities to act more quickly.

6. FUTURE DIRECTIONS

In the case study, an accuracy of 73% was obtained, which for a predictive model is a good indicator, however there are still 27% of precision to be explained, which would merit the evaluation of new variables or the use of more sophisticated models such as networks. neuronal. This would be a better version for future work