Prediction Collisions Severity

Hubert Aliaga

August 27, 2020

Business Understanding

Problem

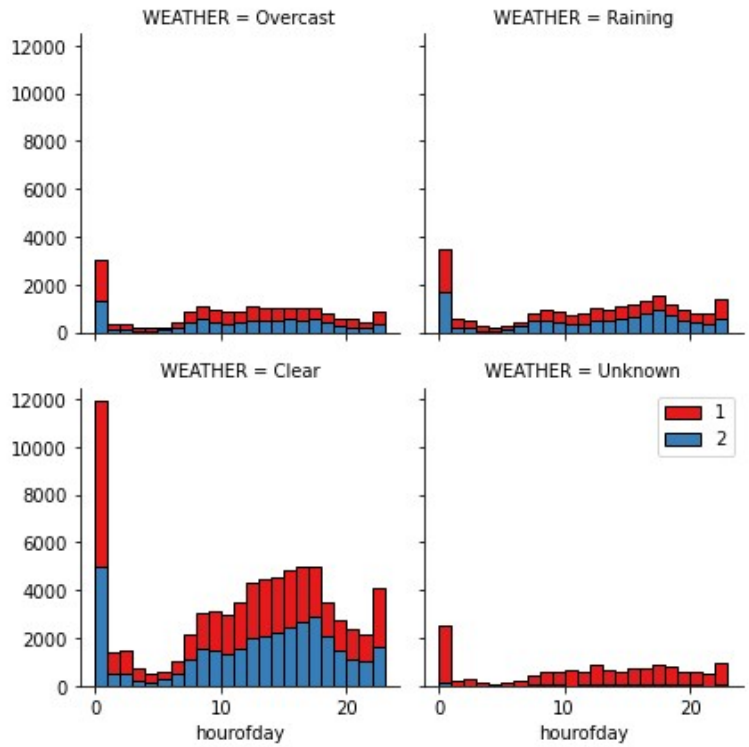
Data that might contribute to determining Collisions and his severity might include Collision Address type,Location, Collision type, Road Condition, wheather and number of objects or people involved in the collision. This project aims to predict severity type of any accidents.

Interest

Clearly, those who work long distances or frequent certain places are interested in finding routes where they are not delayed or even postponed due to sudden accidents.

Data

We will use the comparison between climates, as road conditions will show familiar behavior.



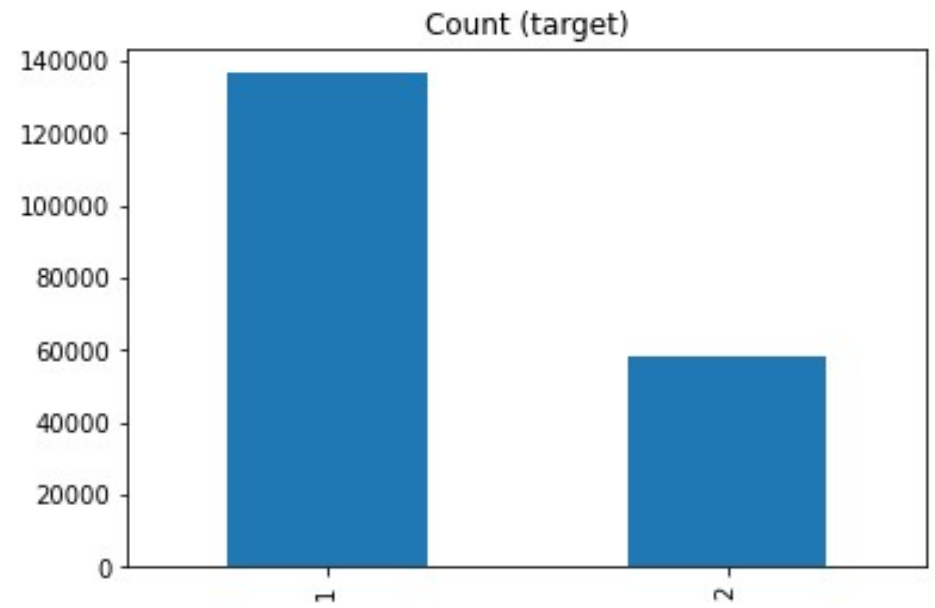
As we clearly see how it is divided into groups where more of one type of collision is concentrated than another. One visible case is collisions that occurred at midnight.

Data

Balance

You will notice that the shared data has unbalanced labels (approximately 70 % Class 1 and 30 % Class 2).
So, balance data with Random Over-Sampling (might be Random Under-Sampling).

```
Class 1: 136485
Class 2: 58188
Proportion: 2.35 : 1
```



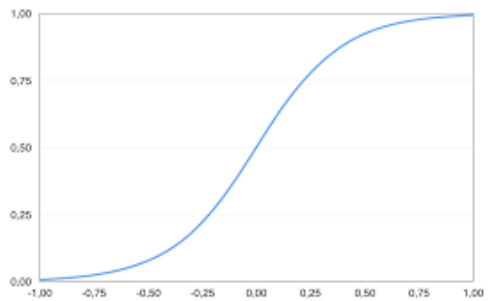
Modeling

Logistic Regression

To make this model we will do it in several traditional steps using Logistic Regression:

Steps

1. Verify that both the characteristics and the target variable have the same dimensions.
2. Normalize the data
3. Separate training and validation set
4. Train the model with the training data



Evaluation and Conclusion

Report

To evaluate the Logistic Regression model, we will focus on three specific measures:

	Algorithm	Jaccard	F1 Score	Logloss
0	LogisticRegression	0.495973	0.699479	0.541875

Conclusion

We can say that when they provide us with information about a weather, the type of street or knowing the location of the accident. We are 70 % confident in saying what kind of collision it is and thus taking other alternatives if necessary. Remember that an important variable is what time this collision occurs, as we saw that it influences a lot.