



CAR ACCIDENT ANALYSIS

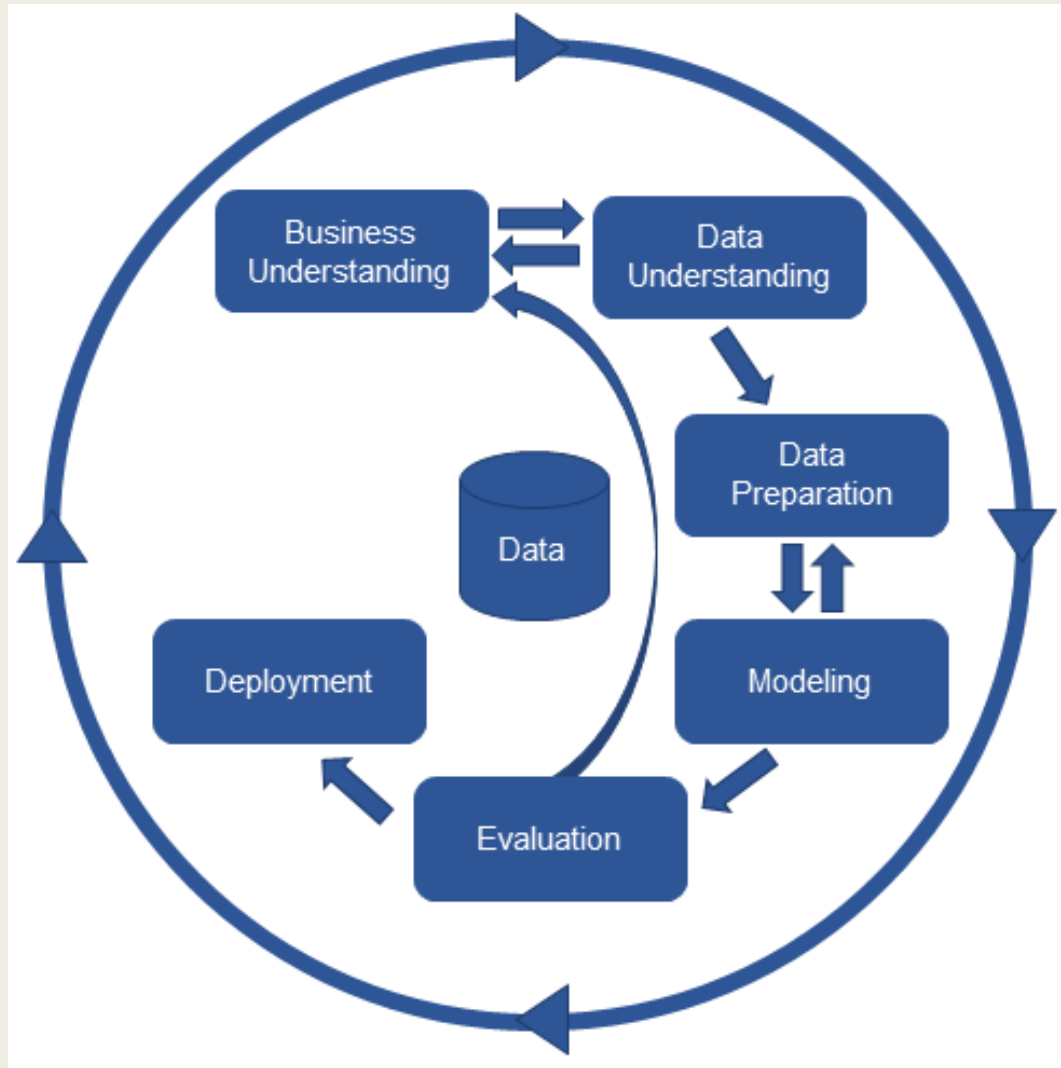


Problem Statement

- Based on the reports from ASIRT, around 38,000 people die every year in road accidents in the United States. The fatality rate is quite significant which is around 12 deaths per 100,000 inhabitants. It is pretty apparent that as the amount of vehicles increase on the roads the number of accidents are also increasing, which obviously means as the number of vehicles increase exponentially it also increases the fatality rate within accidents. In this project we aim to analyze and predict the common causes of accidents allowing us to gain more insight towards this ongoing problem within the United States. We will be using data provided by the Seattle Department of Transport which will allow us to come to a final conclusion for this ongoing problem.

Data Used

- To reach our target goal of analyze and predict the common causes of accidents we are going to use the data provided by the Seattle Department of Transport, the data set is a very extensive data set which includes around 40 columns which describe a particular accident including the road condition (dry or wet), people involved in the accident, the severity of the accident, if the accident place was a block or an intersection, the type of accident on how the accident occurred (MOTOR VEHICLE STRUCK MOTOR VEHICLE, FRONT END AT ANGLE), the weather and the light condition the accident occurred in. These types of data is very essential for us to actually reach to a final conclusion towards our analysis of the main conditions which actually cause accidents.



Methodology

- CRISP DM framework for Machine Learning

Results

- Logistic Regression
 - *Score: 0.6261580874045284*
- Decision Tree Classifier
 - *Score: 0.7552085687169522*
- Random Forest Classifier
 - *Score: 0.7250418041307001*
- BEST ALGORITHM: Decision Tree Classifier

Discussion

- Based on this analysis of the data set, it is a very unbalanced data set and with too many missing values, this resulted in a lot of data cleaning to give me a result which I was looking for in which the data can be worked in, for example according to me the "SPEEDING" factor does play a major role when it comes to accidents and the cause of accidents and had to be dropped due to multiple instances of missing values. This would also give police officers a way to add more stop signs and technology-based accident control systems which will allow them to stop people from speeding and causing accidents.

Conclusion

- To conclude, this was a good insight which was gained, there are many other variables which need to be considered within our analysis since most of the accidents are minor and avoidable, but at the same time do cause traffic disruptions thus increased travel time and also costs towards increased loss of property for the council of the city.