Name : Priya Subramanian

Group: F

Project Name: Predicting Severity of Road Traffic Accidents in France

Lessons learned and Future Work:

A systematic approach was followed in this project in determining the data cleaning , data exploratory and architecture of Machine Learning techniques.

Massive amount of time was invested, only, to determine the dataset for project work. Although, the motivation behind this process was to accquire knowledge on Real-world government data and also gain experience handling large dataset, in future, I would want to take precautions to cut-down this time. Furthermore during this hunt for dataset, I learned that qualifying the Buisness value or ROI from the data is a valuable task.

Although, the wish of working with large dataset (132977 rows × 51 columns) was fulfilled, below limitations were experienced:

- You might not need all your Big Data.
- Formulating specific Hypothesis before the research is very important, and I should have followed it in work. Moreover, staying goal-focused with large variables is crucial here.
- Large number of columns raises the problem that not all possible features have uniform probability. This creates the presentation bias in models.
- Detecting outliers was challenging, and we detected outliers for two columns after processing of the models.
- Cleaning lables from categorical data is tedious when the specifications or business meaning of such labels are unclear.
- I learnt that adding more complex features to the ML models may not show an improvement in scores. It was an important reminder about feature engineering to me.
- I could practically learn that correlation does not imply causation through our ML models.

Future Work:

- It will be interesting to identify features and find scores to answer acute Business questions.
- Another interesting aspect would be to compare our models with recent year i.e. 2021 and perform research on findings.
- Performing unsupervised classification techniques for our research in classfying the features might add some interesting information.