IDS\_GROUP059

2019\_Cluster-DSE-IDS\_A1\_PS1  
KODURI S CHANDRAPRABHU

SAGNIK GHOSH  
M NAVEEN KUMAR

KANDERAO LAXMI NARASIMHA RAO

Answers:

1. Without building any model, what are the contributing factors for traffic slowness?

As part of EDA using different visualizations (Stacked bar and Scatter plots) we can clearly see that the below features are impacting the target variable "Slowness\_in\_traffic\_percent" the most.   
  
a)Hour\_Evening  
b)Immobilized\_bus  
c)Broken\_Truck  
d)Accident\_victim  
e)Semaphore\_off  
f)Point\_of\_flooding

Similarly the below features are having min impact on the target

a)Vehicle\_excess  
b)Fire\_vehicles  
c)Fire  
d)Incident\_involving\_dangerous\_freight  
e)Occurrence\_involving\_freight

1. Are you able to confirm the above findings using any two modelling techniques? Give appropriate explanation for the same.

Yes almost matching.

After training and running a model these are the results observed.

Model 1: Linear Regression Lasso

From Lasso regression’s feature importance method we can see that the features that are contributing for slowness in traffic almost matches our inference from visualization.

These are the features considered by our model to predict the target variable and as we see the ranking they are,

a)Hour\_Evening   
b)Point\_of\_flooding  
c)Immobilized\_bus  
d)Semaphore\_off  
e)Broken\_truck

Model 2: Random Forest Regressor

From RandomForest regressor's feature importance method we infer that the features that are contributing for slowness in traffic matches our inference from visualization.

They are,

a)Hour\_Evening   
b)Broken\_truck  
c)Accident\_victim  
d)Immobilized\_bus  
e)Point\_of\_flooding  
f)Semaphore\_off

These are the features that are mostly deciding the output of the predicted variable as observed by the model stats.

And this matches our understanding form the exploratory data analysis.

1. Using Recursive Feature Elimination technique, what are the the contributing factors for traffic slowness?

Listed below are the features that are mainly contributing for Slowness in Traffic as selected by the model post recursive elimination of least significant features

Immobilized\_bus

Broken\_Truck

Point\_of\_flooding

Semaphore\_off

Hour\_Afternoon

Hour\_Evening

Hour\_Morning

By selecting these limited features we can achieve reasonable performance (minimum error) even for a huge set of test data with optimal usage of space and time on the system at the same time not compromising much on the accuracy.