## The process followed for the Safety Algorithm

1.The datasets are been split up in to multiple part file along with a label file in csv format

2. I first select the part0 file along with label file to do initially pre-processing for the features and

In due course can imply it on the remaining data ( partX files).

3. I checked on the data types as well as the size of the records and their distribution values.

4 As a part of pre-processing step we remove the Null values and most of the stages we do

Work without the BookingID field ,since it doesn’t involve and significance in the processing

Or classification

5I used RFE ( Recursive feature Elimination ) process to select the fields which do really have

Impact on our Dependent variable Label. Here I have used 5 top attributes.

6. With the reduced fields we go for standard scaling so that variation in the feature values are

Avoided and all lie on the common value range.

7. I combined all the remaining data sets and go for the full scaling process and also the train test split

8. Once the train test split done we go for the models . I used logistic regression and SVM initially

9. I also used Cross validation (SVM) to avoid variance in the training test process**. CV value of 10**

10 I used Grid search process also to identify the Best model

11. Used accuracy and Confusion matrix to check the value of both the models validation.

12. As a final step I also tried XG Boost algorithm process for the labels.

**Note**: For feature elimination we could use Backward Elimination process too. But complexity of **RFE** is bit less compared to other methods.

PCA could also be used to reduce and then be fed to the algorithm. But we won’t be able to identify the fields.