

Aeroplane Severity Classifier

Name: Gaurav Kumar Daharia

Roll No.: 11

Registration No. 11711574

Section: KM032

# ABOUT DATASET

This dataset is consist 12 features and each feature is containing different values like **cabin temperature, turbulence rate, Max elevation** etc. this particular dataset gives certain information regarding Aeroplane behavior when it is at a certain elevation. It also shows certain parameter through which we can classify whether a particular flight is how much severe to meet with an accident in future if certain parameter is not taken seriously.

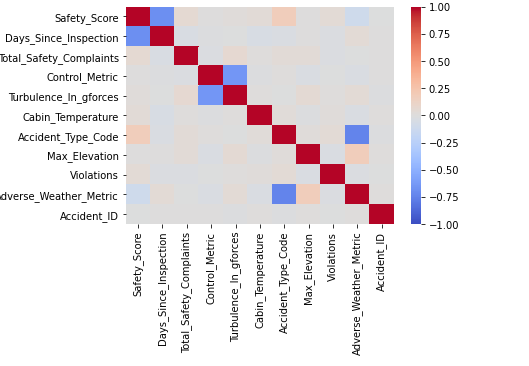
**Classes of these datasets are:**

1. Significant damage and serious injuries.
2. Significant damage and fatalities.
3. Minor damage and injuries.
4. Highly fatal and damaging.

**Features of dataset are as follows:**

1. Safety Score
2. Day Since Inspection
3. Total safety complaints
4. Control metric
5. Turbulence in GeForce
6. Cabin temperature
7. Accident type code
8. Max elevation
9. Violations
10. Adverse whether metrics
11. Accident Id
12. Severity

Dataset Correlation Analysis

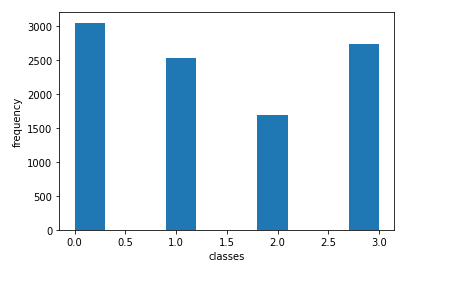


Above correlation matrix shows each feature correlation with itself as well as with other features of dataset. Its value of correlation is in between 1 to -1 and different colors of matrix shows that how much it is correlated.

As per pre-processing purpose this is a very important step to be performed because it shows whether a feature is correlated with other feature or not, if features are correlated then it important to remove them or remove either of correlated feature.

Correlation between features should be minimum because it ultimately effect the training model and later on maybe it will lead to overfitting.

Classes Analysis



Above figure shows the analysis of class frequency and as we can see that our dataset have four classes labelled [0, 1, 2, 3].

|  |  |  |
| --- | --- | --- |
| **CLASS NAME** | **LABEL ENCODEC** | **FREQUENCY** |
| Significant damage and serious injuries | 0 | 3000 |
| Significant damage and fatalities | 1 | 2500 |
| Minor damage and injuries | 2 | 1700 |
| Highly fatal and damaging | 3 | 2800 |

Algorithm and Techniques Used and Accuracy Achieved

**Random Forest Classifier:** Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees**.**

**Support Vector Machine:** Support-vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.

**Extra Tree Classifier: It** is a type of ensemble learning technique which aggregates the results of multiple de-correlated decision trees collected in a “forest” to output its classification result. In concept, it is very similar to a Random Forest Classifier and only differs from it in the manner of construction of the decision trees in the forest.

**Technique Used for Training Model:**

**Ensemble technique:** It is a type training technique in which we combine more than two algorithm and train our machine learning model on it.

**Accuracy achieved after training model:**

**Accuracy on training data – 100%**

**Accuracy on testing data – 88%**