

Description of the Data Obtained:

Author: K Uday Karthik

1. I have chosen the open-source dataset from
<https://www.kaggle.com/sobhanmoosavi/us-accidents>
2. It is free to use.
3. The dataset is very huge. So to simplify it. I have chosen a particular state called
“New York” (NY) from all the states in the USA.
4. The records show the accident details of “New York”.
5. In the dataset, I have used the columns like Accident Severity, Date, Time, Latitude, Longitude, Temperature, Wind Speed, Day / Night and many more. Since accidents are dependent on all these factors.
6. Actually there were numerous attributes. I have cleaned the data by considering only relevant attributes that are necessary for my project.
7. Missing Values all have been checked.
8. I considered mainly 3 models for this project.
9. Namely Logistic Regression, Decision Tree Regression, KNN and Random Forest Regression.
10. I have fit my data to all these models and predicted accident severity.
11. To my knowledge, I have observed with given dataset, Random Forest was a better one.