

CSCI HW-3: Machine Learning on Tornadoes Dataset

- The goal of this assignment is to predict the magnitude of Tornadoes on the F(Fujita) Scale.
- Data of Tornadoes between the years of 1949-2015.
- All Tornadoes are mapped to scale F0-F5. Tornadoes in the recent time(from 2013) have been categorized based on the Enhanced Fujita Scale(EF Scale).
- The data repository has the following distribution:

Magnitude	Number of Tornadoes	Percentage of total(%)
F0	28,086	45.4
F1	20,498	32.99
F2	9,189	14.79
F3	2,647	6.2
F4	709	2.3
F5	88	0.14

Submitted by:

Deepak Singh

Pre Processing of Data

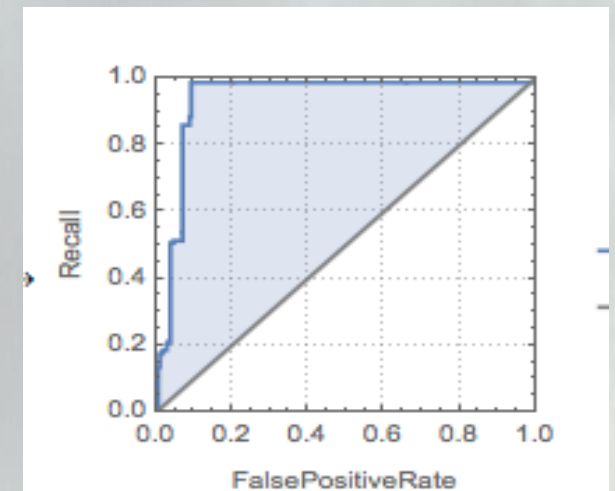
- Mapped tornadoes on EF scale to F scale.
- Computed the Average Cost of the Estimated Property Loss range in the dataset.
- Replaced the missing data in the Estimated Property Loss by the average property loss value of the class(F0-F5).
- Removed the entries with False in MagnitudeEstimatedQ as they don't conform to the methodology of the original scale.
- For Tornadoes with the end coordinates missing, replaced them with zero.
- Calculated the distance between the start and end coordinates using GeoDistance function.
- Training and Testing data selected randomly in the proportion of 4:1.
- Mapped the distance, Estimated Property Loss and the Injuries to the magnitude.



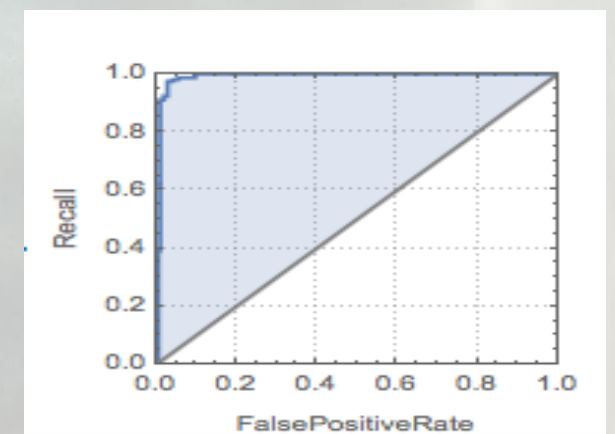
Results

Performance/ Algorithm	Accuracy (%)	Precision (%)	Recall (%)
Decision Tree	98.927	99.1	95.2
Neural Network	76.180	77.2	74.7
Nearest Neighbors	95.712	85.4	89.3

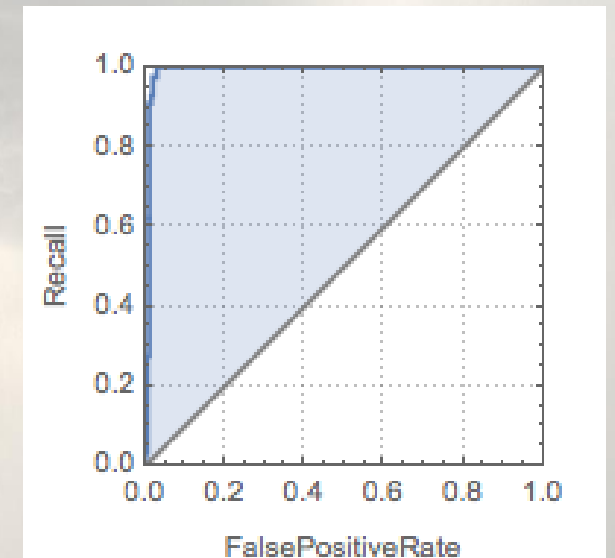
- The major factors affecting the magnitude of tornadoes are Estimated Property Loss, Injuries and Distance travelled by it.
- From the results, it can be concluded that the Decision Tree algorithm performs the best



Neural Networks



Decision Tree



Nearest Neighbors