EAS-503 ~ THE FIRE RISK ASSESSMENT PROJECT
Submitted by Group#40

INTRODUCTION and MOTIVATION:

- The Fire-Risk Assessment project is developed by the National Park Service's (NPS) Fire and Aviation Management program to respond to the devastating 2011 wildfire season and holds data from 1970-2020.
- The dataset can be found at https://data-nifc.opendata.arcgis.com/datasets/facility.
- The aim of the project is to build a predictive model that evaluates certain contributing factors such as access to the facility, the surrounding environments, construction design, and materials and resources available to protect facilities from wildland fire.
- The data set was procured from:

https://data-nifc.opendata.arcgis.com/datasets/facility?geometry=50.977%2C-89.991%2C-50.977%2C-89.336 (last accessed on 10 Oct 2020). An overview of the dataset has been given in table 1.

- We divided the dataset into nine different tables in the SQL database.

Table 1. An overview of the raw dataset

Dataset	Observations	Data Description
NPS_Facilities__Wildfire_Risk_	64 attributes and 44397 records	The dataset contains information such as facility location, facility features, owner details, access routes, surrounding environment, structure, and protection resources.

PROPOSED ANALYSIS:

- Performing EDA (.ipynb notebook)
- Data Cleaning
- Suitable Data Imputation
- Dividing the data into train and test sets
- Feature importance analysis
- ML algorithms
- Summary

CONCLUSION:

- Ensemble methods like random Forest and XGBoost are performing well in predicting 'Rating' variable against Wild Fire hazards!
- Sigmoid kernel from SVM is not performing in a desirable manner.
- Suitable ML algorithms would be depending on dataset. No Hands-on rule!!

ML Algorithm Accuracy Matrices, %		Train Set				Test Set			
		Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
Random Forest		98	100	99	99	92	92	93	93
XGBoost		100	100	100	100	93	95	95	95
Decision Tree		96	98	98	98	87	91	91	91
KNN, k=1		100	100	100	100	84	88	88	88
AdaBoost		100	100	100	100	91	95	95	95
Bagging		98	100	99	99	91	94	94	94
SVM	Linear	80	83	91	87	87	90	93	91
	Radial	93	95	97	96	91	94	94	94
	Poly.,3	92	94	95	95	90	94	94	94
	Sigmoid	29	40	34	37	26	26	42	32