

EAS Group#40

Proposal

The Fire Risk Assessment project was developed by the National Park Service's (NPS) Fire and Aviation Management program to respond to the devastating 2011 wildfire season and hold data from 1970-2020. The current dataset consists of the data collected during the assessment process, identifying, planning, prioritizing, and investigating fire hazard treatments at different units and developing incident response plans for facilities and communities within NPS units. Each unit has been examined has been rated and scored accordingly. More information on the dataset has been provided in <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).

The overall dataset consists of 68 attributes with 44398 observations. The dataset holds different fire hazard information collected from several National Parks; based on that, the units have been rated or scored on how prone they are against fire hazards. In the current study, we would be building a predictive model that would estimate for each of the units against fire hazards. The following steps would be taken to build the suitable ML model in predicting "Score/Rating" from the dataset to identify fire hazards in different facilities.

1. The data frame is very messy, and several values are missing from the record. Suitable data imputation and data cleaning have been performed on the dataset. Different visualization techniques (via folium, plotly, bokeh) and EDA (correlation plot, heatmap, histogram, scatter, violin plots) have also explored the dataset to get familiar with the data's several attributes.
2. After performing data imputation, data conversion, and transformation, shrinkage methods have been committed to identifying the dataset's essential features to predict "Score/Rating". Feature selection has been achieved via Random Forest, Lasso, and Ridge Regression (shrinkage).
3. The predictive model has been build by different ensemble methods (like Random Forest, Bagging, XGBoost, Decision Tree).