Wildfire Modeling and Data Analysis

STAT 683

Instructor: Dr. Sutanoy Dasgupta

Project Guides: Dr. Jian Tao, Rebekah Zehnder (TFS), Don Hannenann (TFS)

Ishank Gupta | Ridham Patoliya

Objectives

- Classification model to predict fire class based on initial weather parameters
- Time Series Analysis of fire impact & fire frequency
- Optimization Model to estimate locations of new fire-stations based on fire response time

Data

Data	**# Data points**	**Important High-level Features**
Fire incidents (1992-2015) (https://www.fs.usda.gov/rds/archive/products/RDS-2013-0009.4/_metadata_RDS-2013-0009.4.html)	1.4 Million	Location, Area, Cause, Duration, Discovery Date & Time, Contained Date & Time
Weather Data (1985-2015) (https://mesonet.agron.iastate.edu/request/download.phtml?network=TX_ASOS)	2.8 Million (yearly)	Hourly Data - Temperature, Windspeed, Humidity
Fire Stations Operations Data (Current) (Received from TFS)	1,840	Location, Resources (Personals - Volunteers & Paid)
Fire Weather Data (2022) (https://data-nifc.opendata.arcgis.com/datasets/nifc::public-view-interagency-remote-automatic-weather-stations-raws/explore?location=0.000000%2C0.000000%2C1.90&showTable=true)	4,500	Temperature, Windspeed, Humidity etc, (Testing Purpose –for IWDA)

Estimation of weather parameters at fire locations

1. Inverse Distance Weighted Interpolation (IWDA)

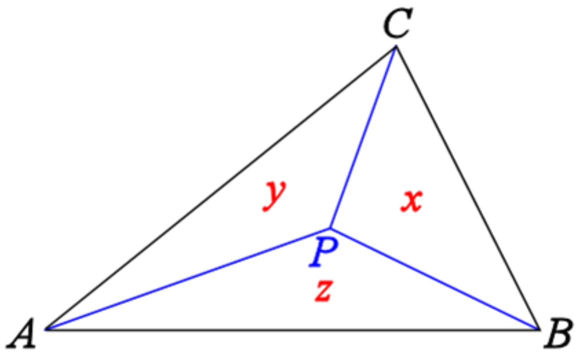
2. Baycentric Coordinate System Method

3. Kriging

$$w(s_i) = \frac{1}{d(s_0, s_i)^\wedge \mathrm{p}}$$

$$\hat{Z}(S_0) = \frac{\sum_{i=1}^{N} w(s_i) * Z(s_i)}{\sum_{i}^{N} w(s_i)}$$

IWDA weights and interpolation Formula



BayCentric Method

Classification model to predict fire class based on weather parameters

Why Classification?

- Role of randomness
- Application:
  - It is advisable for fire departments to deploy resources according to class of the fire and not as per the exact predicted value of fire impact.
- Ordinality of classes:
  - Classification gives us flexibility in merging two or more adjacent ordinal classes which gives more conservative approach to predict a disaster.
- Previous research:
  - A lot of research on wildfire prediction is based on classification.

Data Prep and Feature Engineering

Data

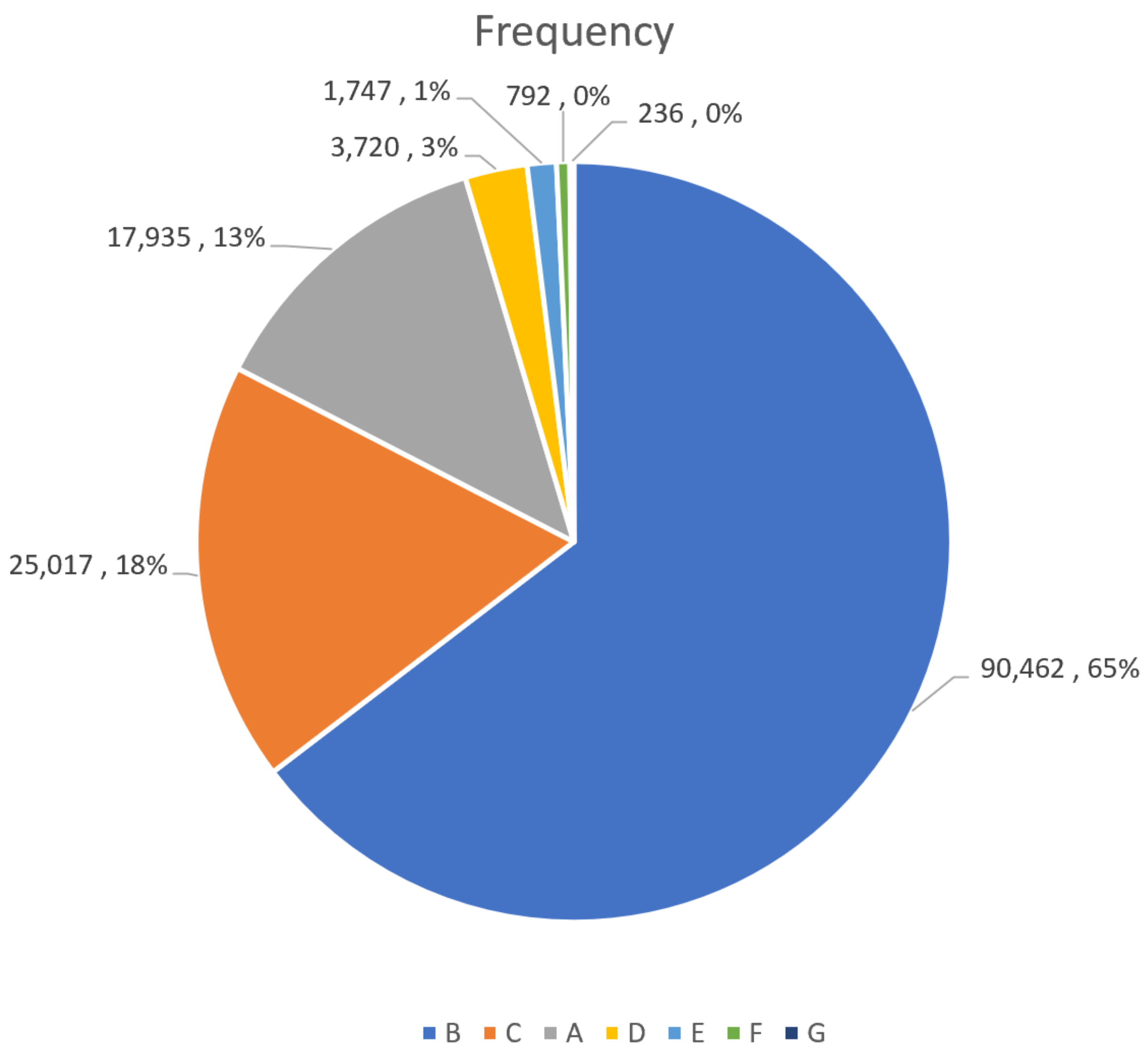
- Month of fire
- Temperature at Ignition Time
- Humidity at Ignition Time
- Wind speed at Ignition Time
- Dew point temperature at Ignition Time

- Target Variable = **"FIRE SIZE CLASS"**

Feature Engineering

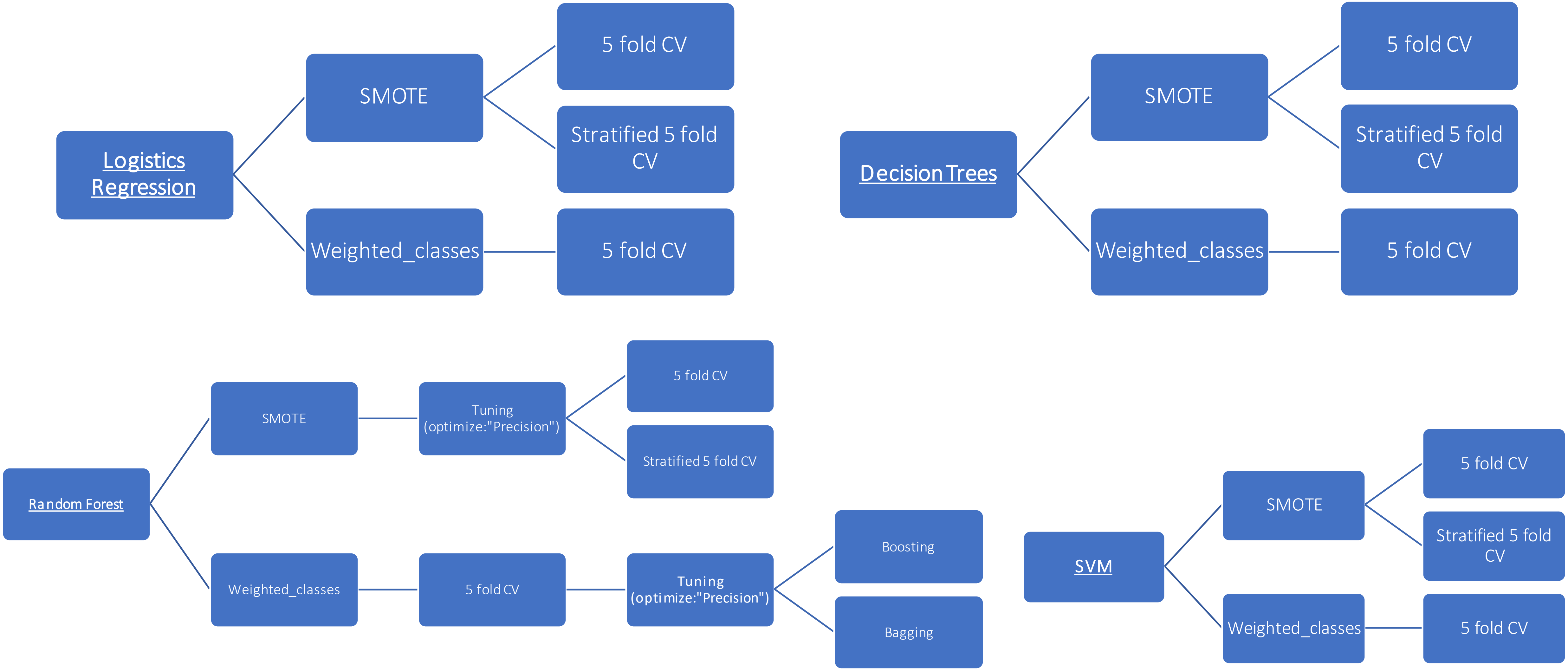
- Dropped irrelevant columns/rows
- Label-Encoding & One-Hot-Encoding
- Normalization
- Binning of numeric variable

Imbalance Dataset Problem and Handling Techniques



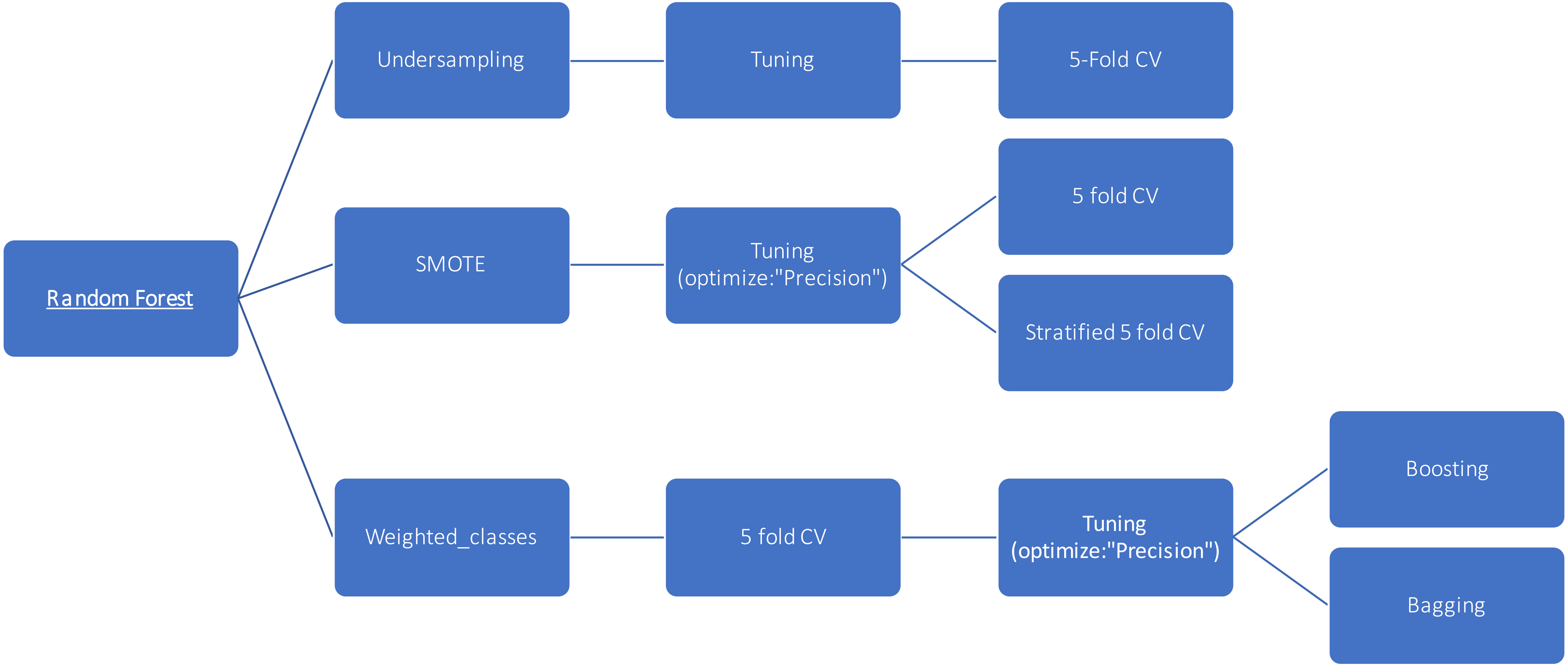
- SMOTE
- Radom over sampling
- Random under sampling
- Class weights

Model Selection



Accuracy Paradox!!!

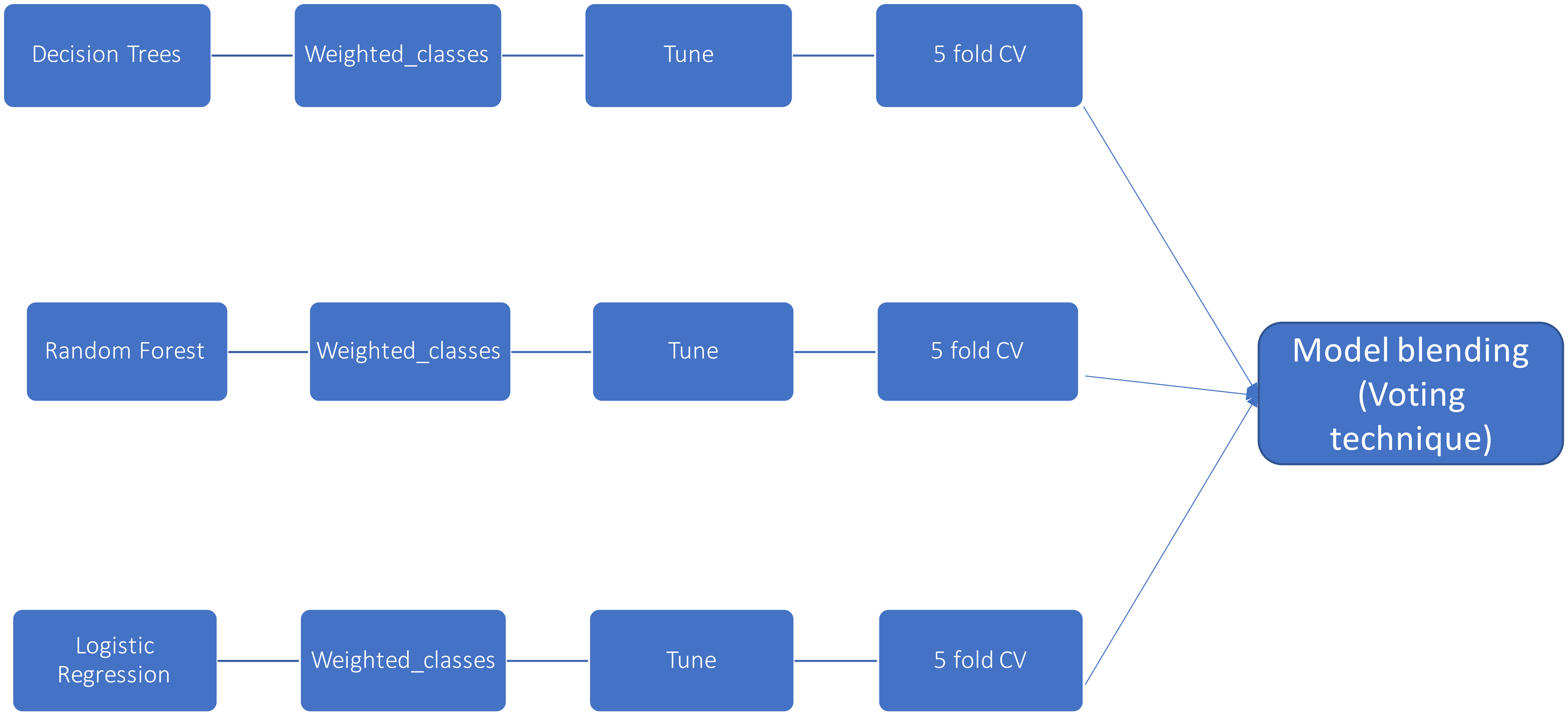
Random Forest – Option 1



Accuracy	Prec.
0.2276	0.4940

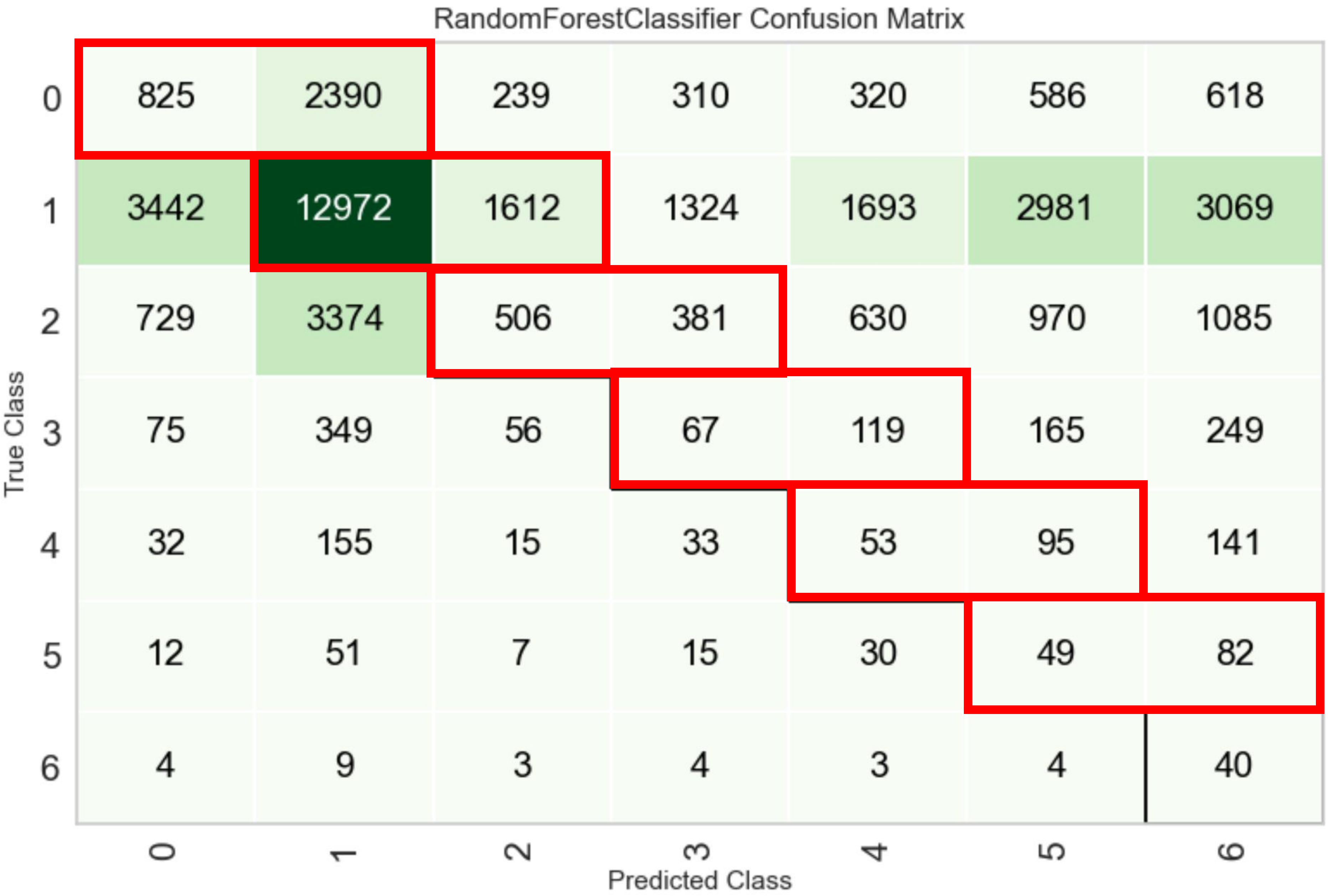
Accuracy	Prec.
0.6467	0.4182

Blending (Voting) – Option 2 (Final Model)

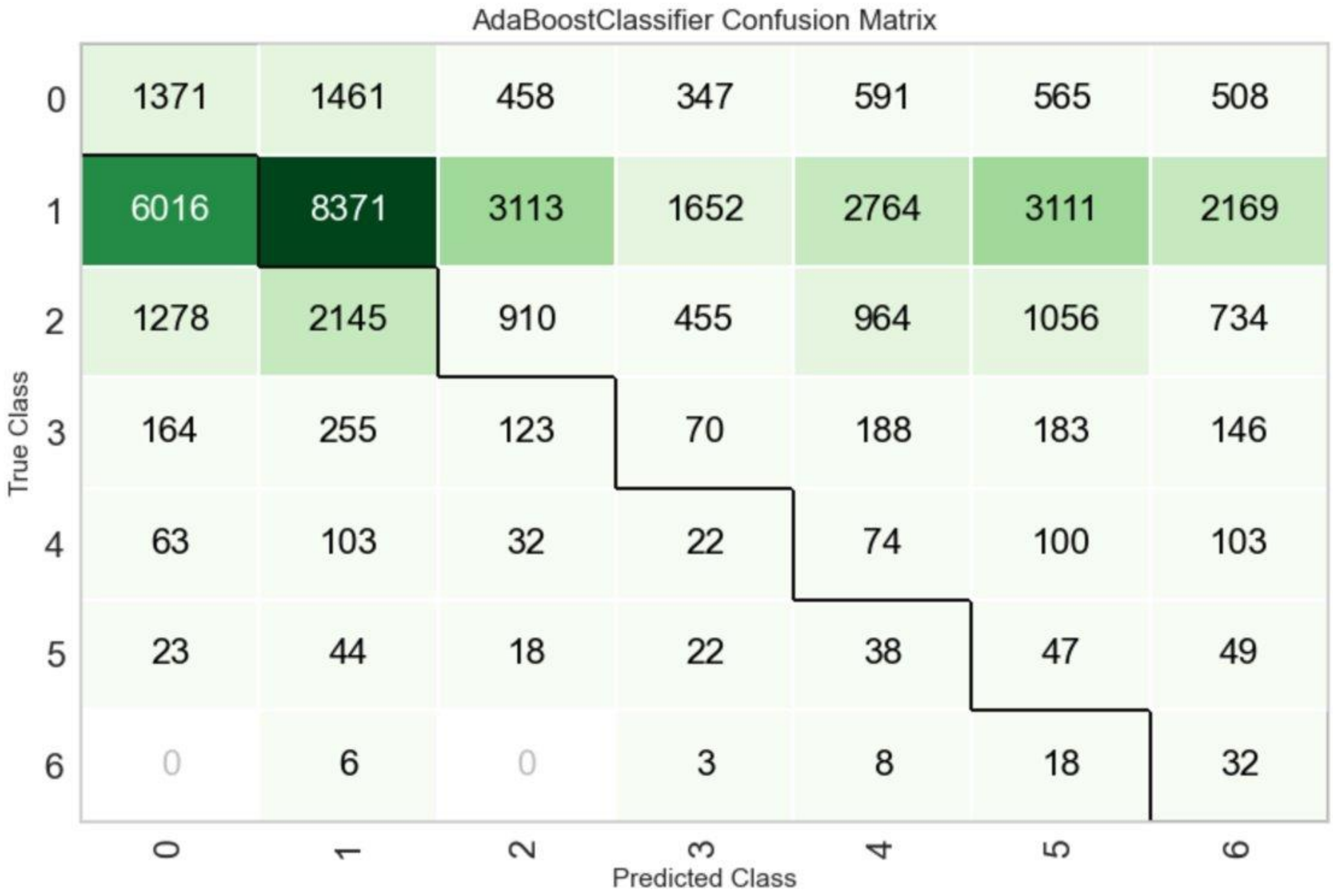


Accuracy	Prec.
0.2954	0.4978

Accuracy of more conservative classification (Example)

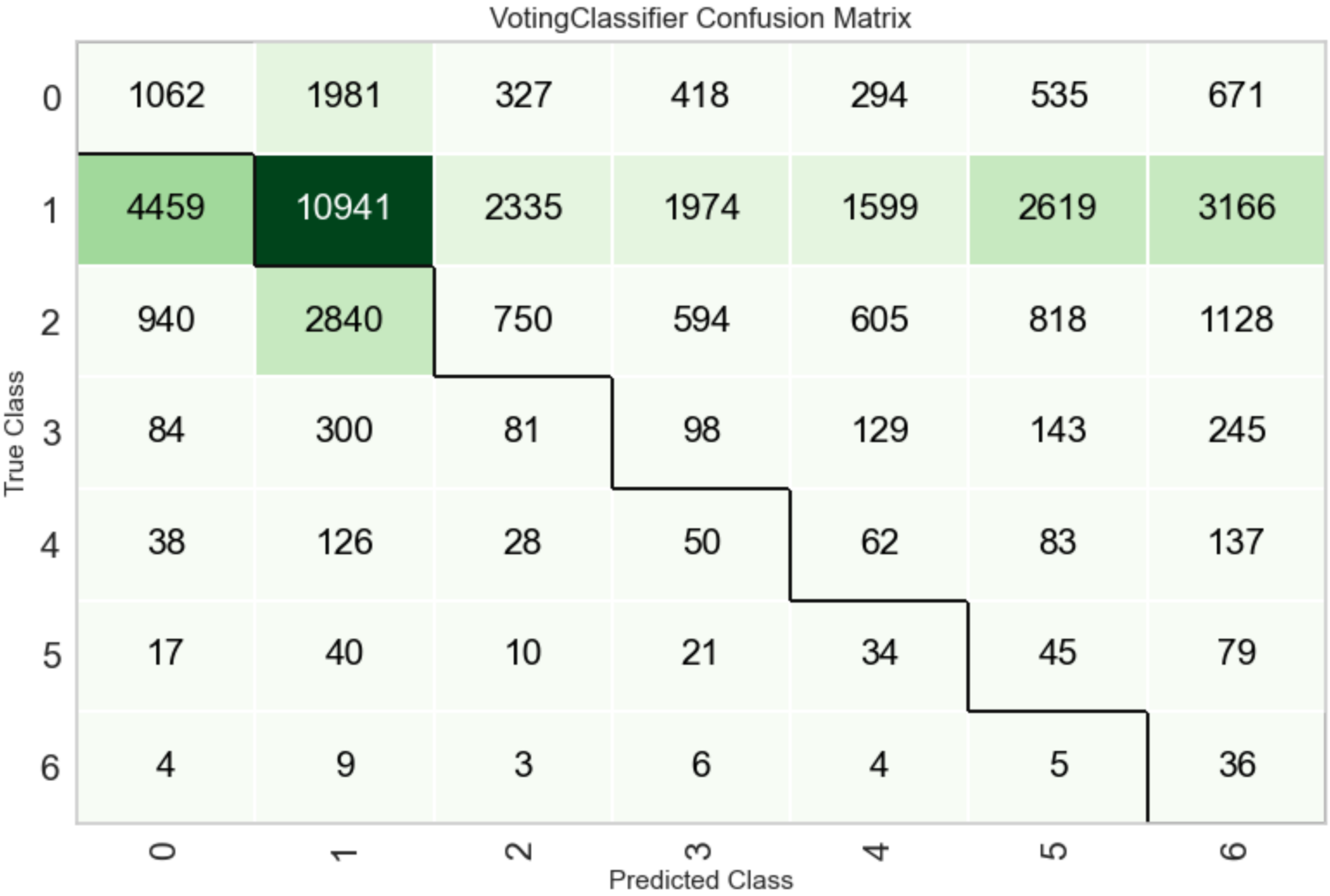


Confusion Matrix – Random Forest & Adjusted accuracy



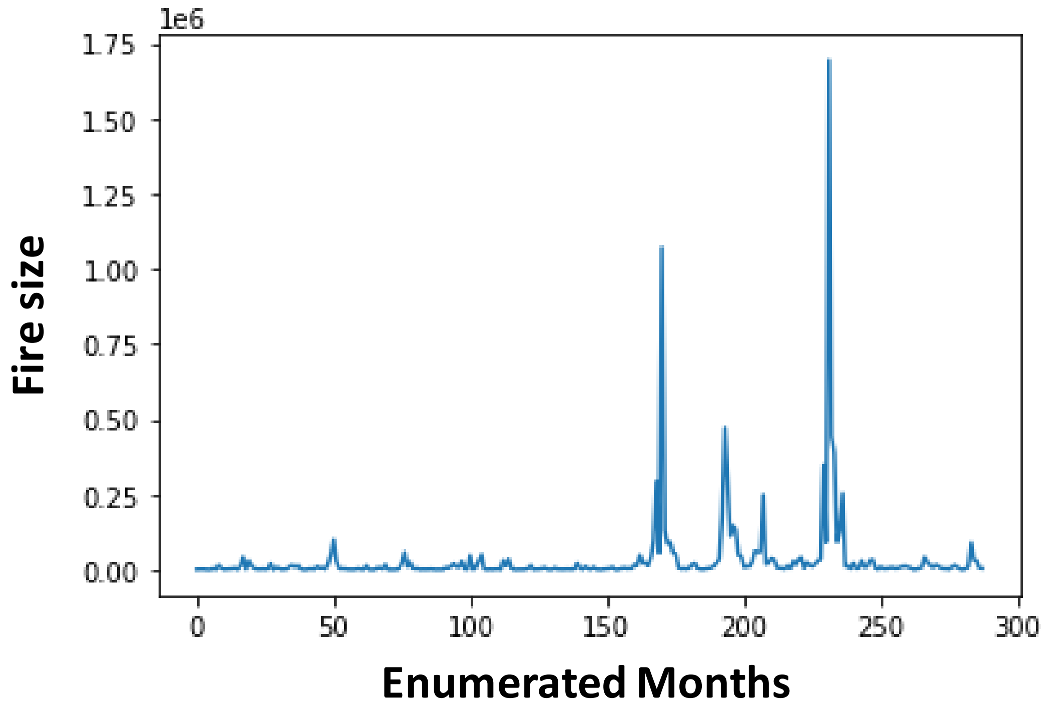
	Random Forest Tuned Without Conservative Approach	**Random Forest Tuned With Conservative Approach**
A	13.62	60.79
B	49.86	55.81
C	6.59	11.43
D	6.01	17.03
E	10.11	28.05
F	20.32	53.65
G	59.701	59.701

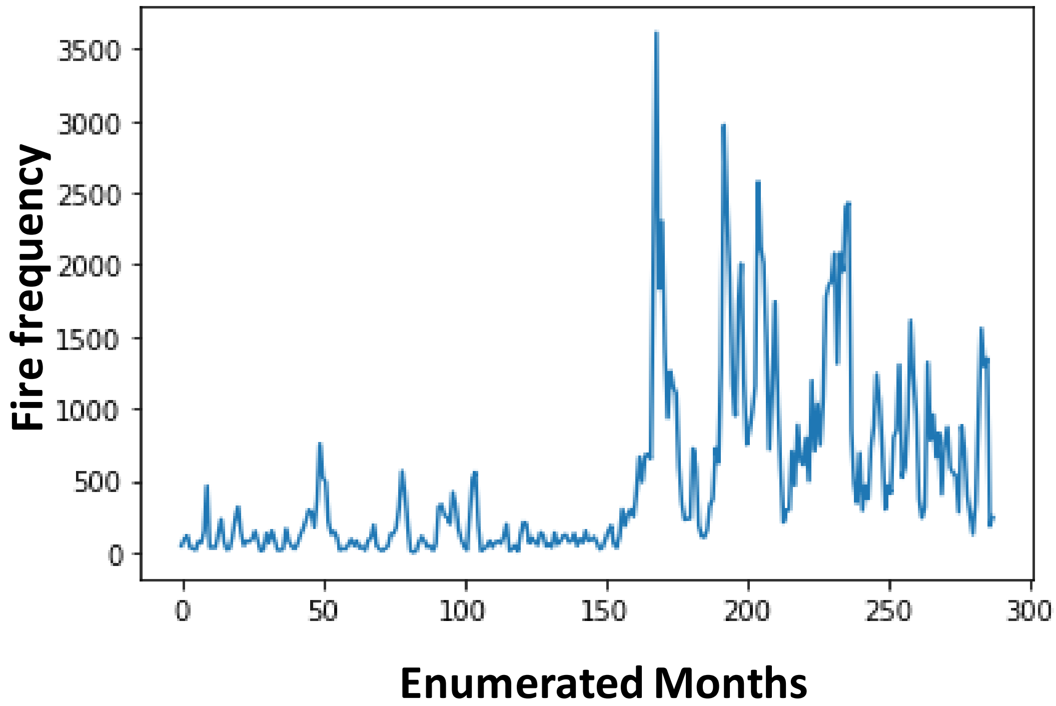
Confusion Matrix – Blended (Voting) Model & Adjusted accuracy



	Voting Without Conservative Classification	Voting With Conservative Approach
A	20.08	57.55
B	40.383	49.075
C	9.77	17.51
D	9.07	21.01
E	11.83	27.67
F	18.29	50.4
G	53.73	53.73

Time Series Analysis of fire impact

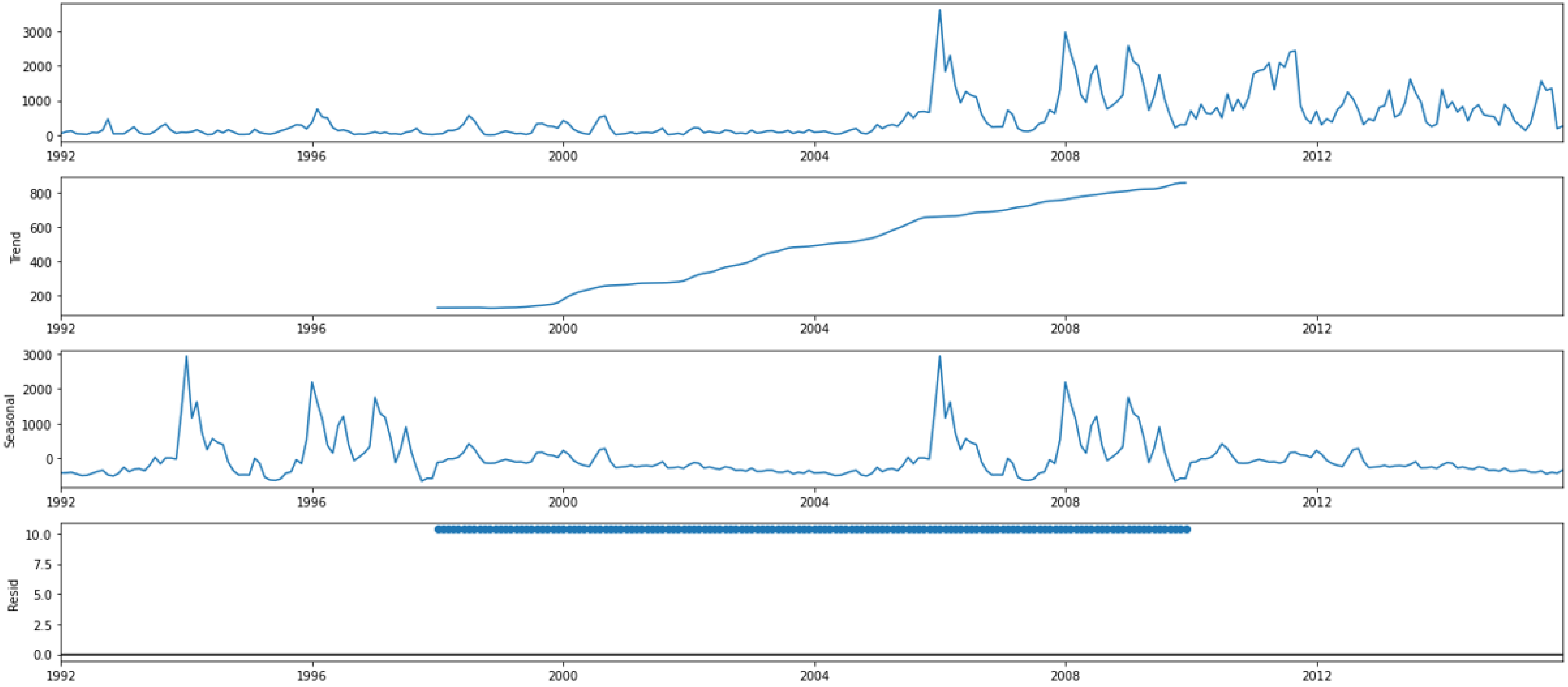




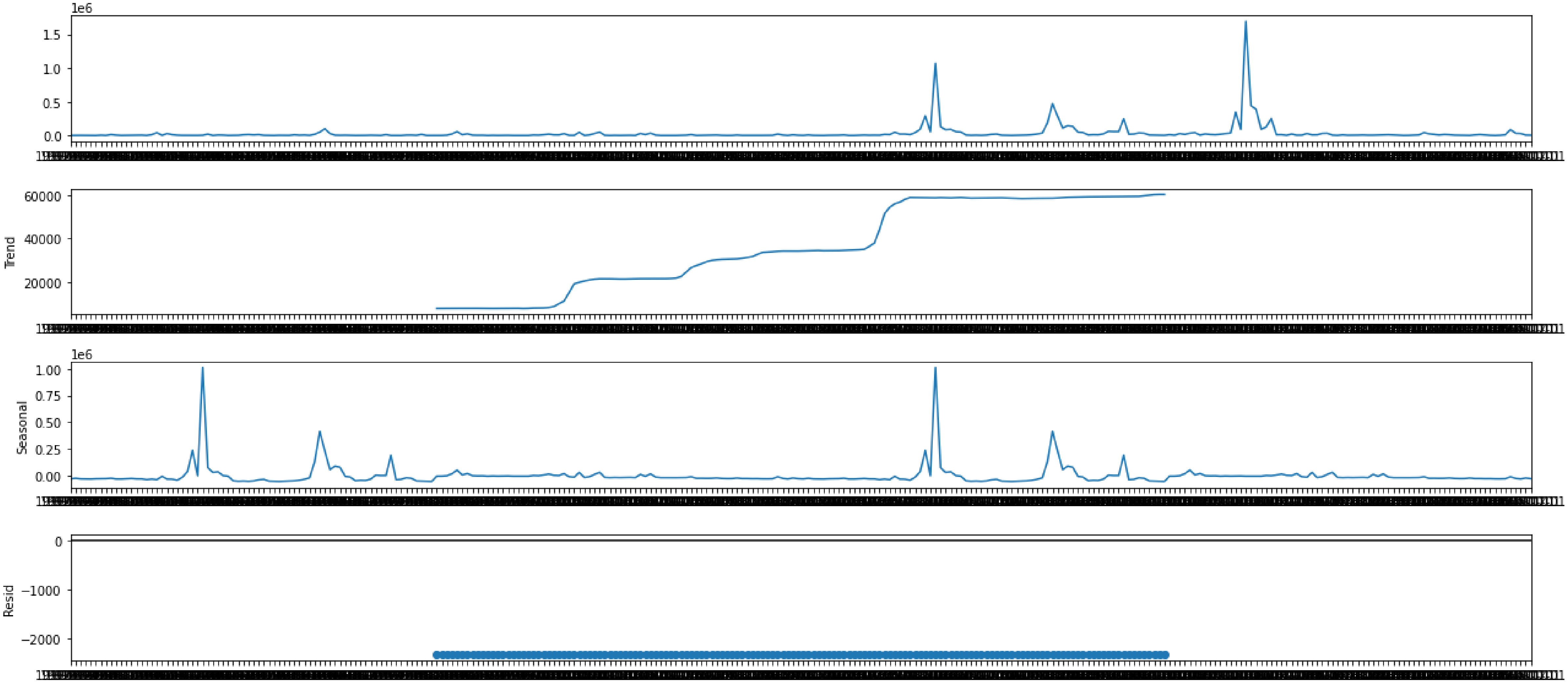
<u>Time-series</u>

- Aggregation:
  - Monthly fire size impact of historical fires
  - Monthly frequency of fires
- Modelling (Methods Explored):
  - ARIMA - autoregressive integrated moving average
  - SARIMA - Seasonal Autoregressive Integrated Moving Average,
  - Exponential Smoothening

Time Series Decomposition – Frequency of fire

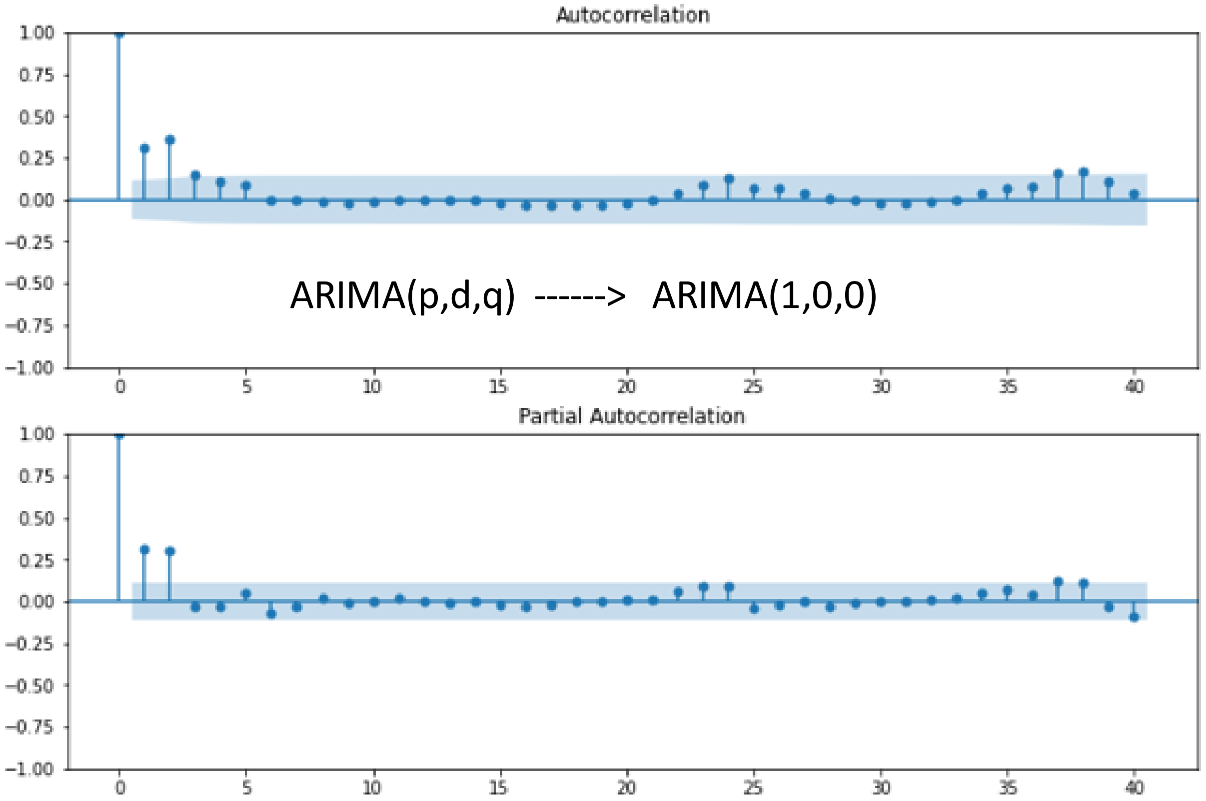


Time Series Decomposition – Fire Size

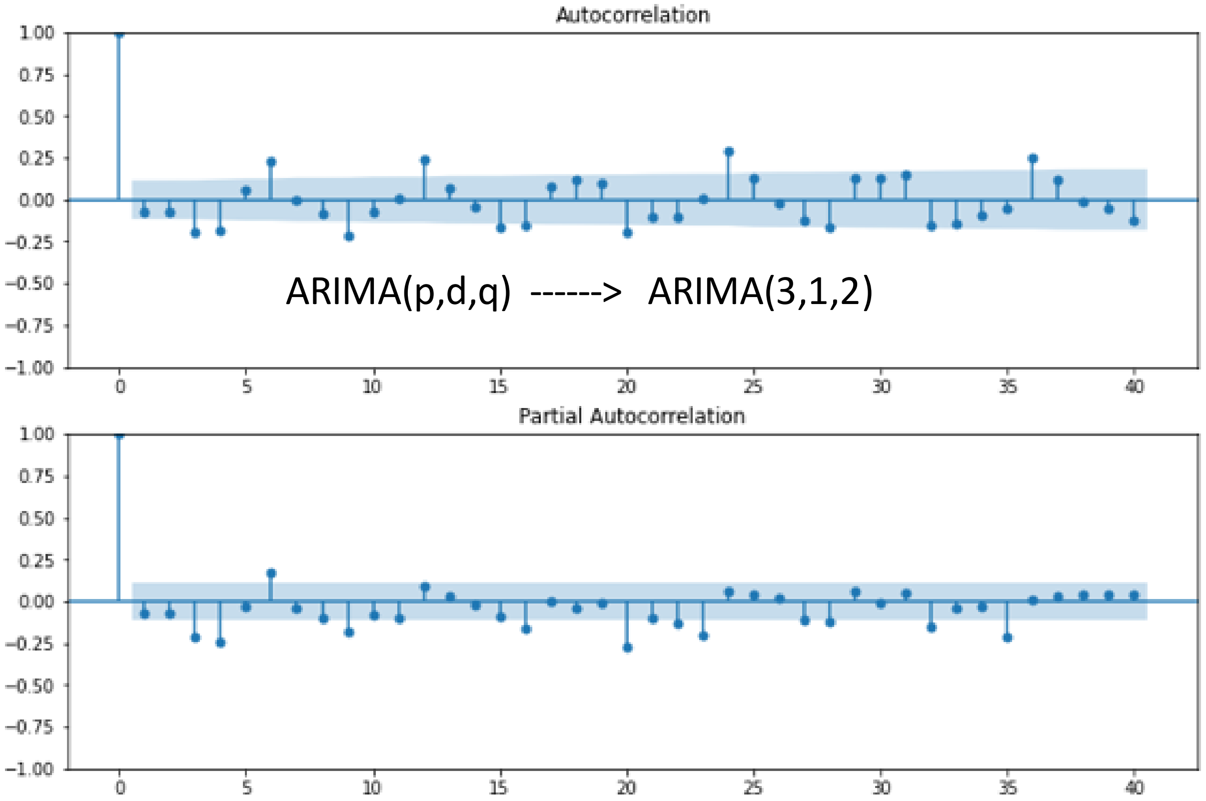


Methodology

1. Tested if Both Data are stationary or not through AD-Fuller Test .
2. For Fire-Size It was stationary as per AD-Fuller Test, but for frequency data we needed to perform 1 level Differencing.
3. Grid search was performed to calculate best parameters of p,d,q

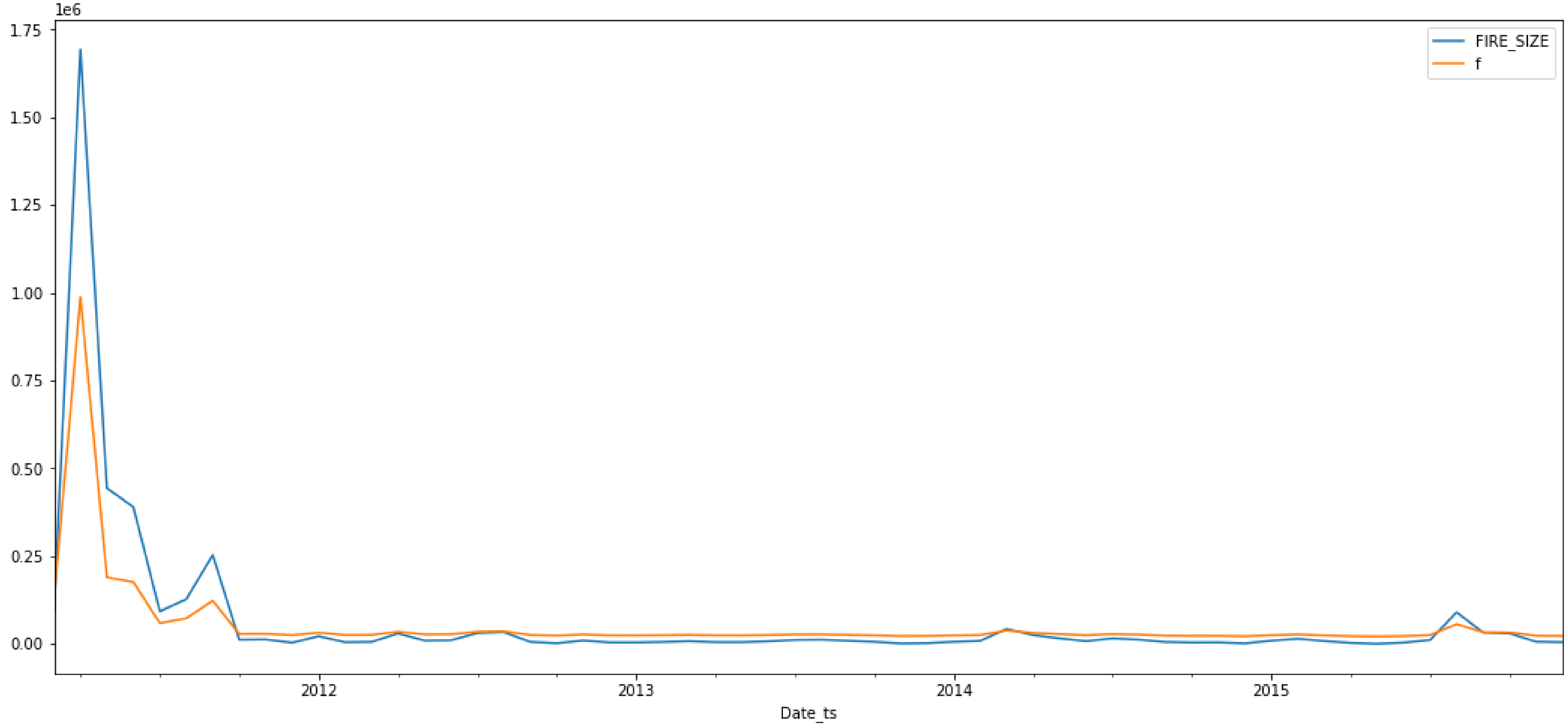


Fire size ACF v/s PACF

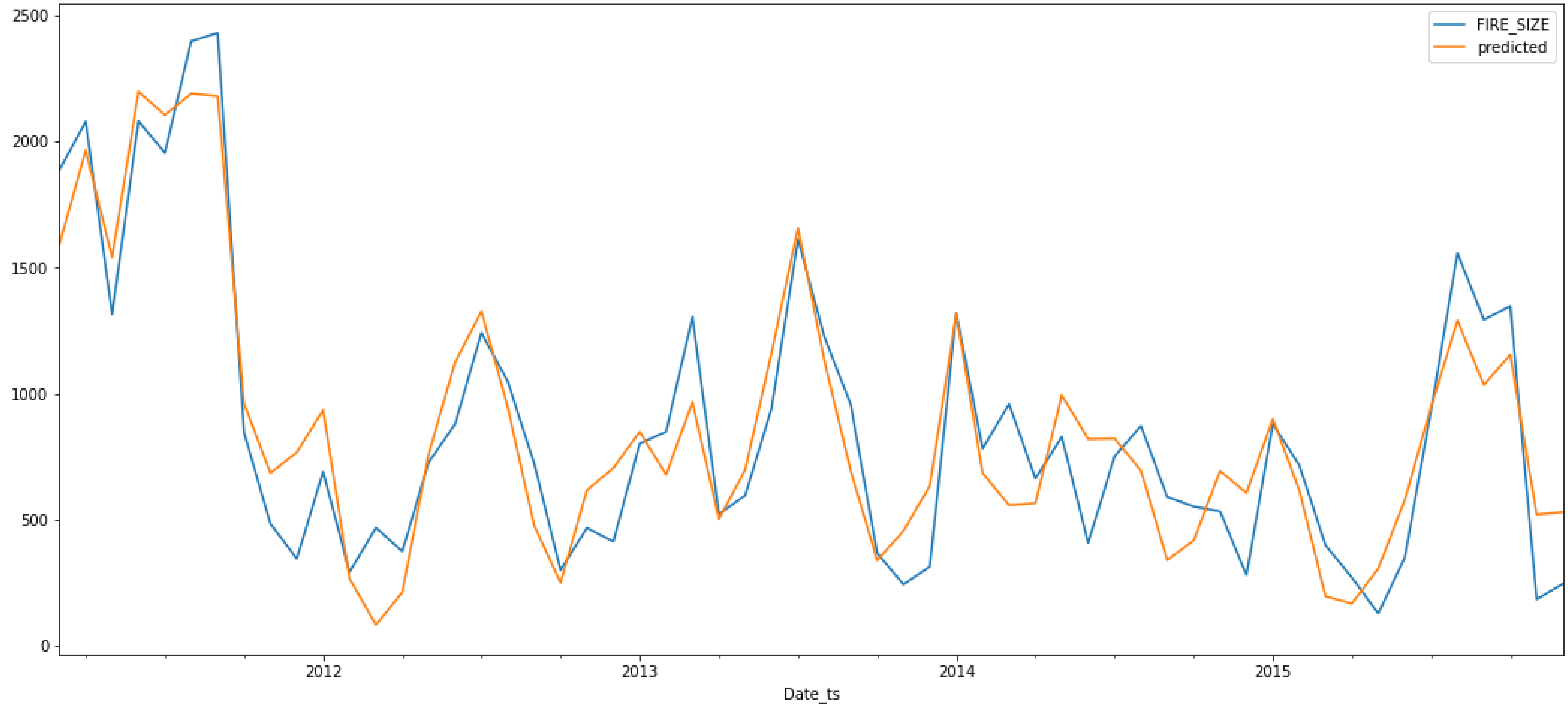


Fire Frequency ACF v/s PACF

Fire Size ARIMA

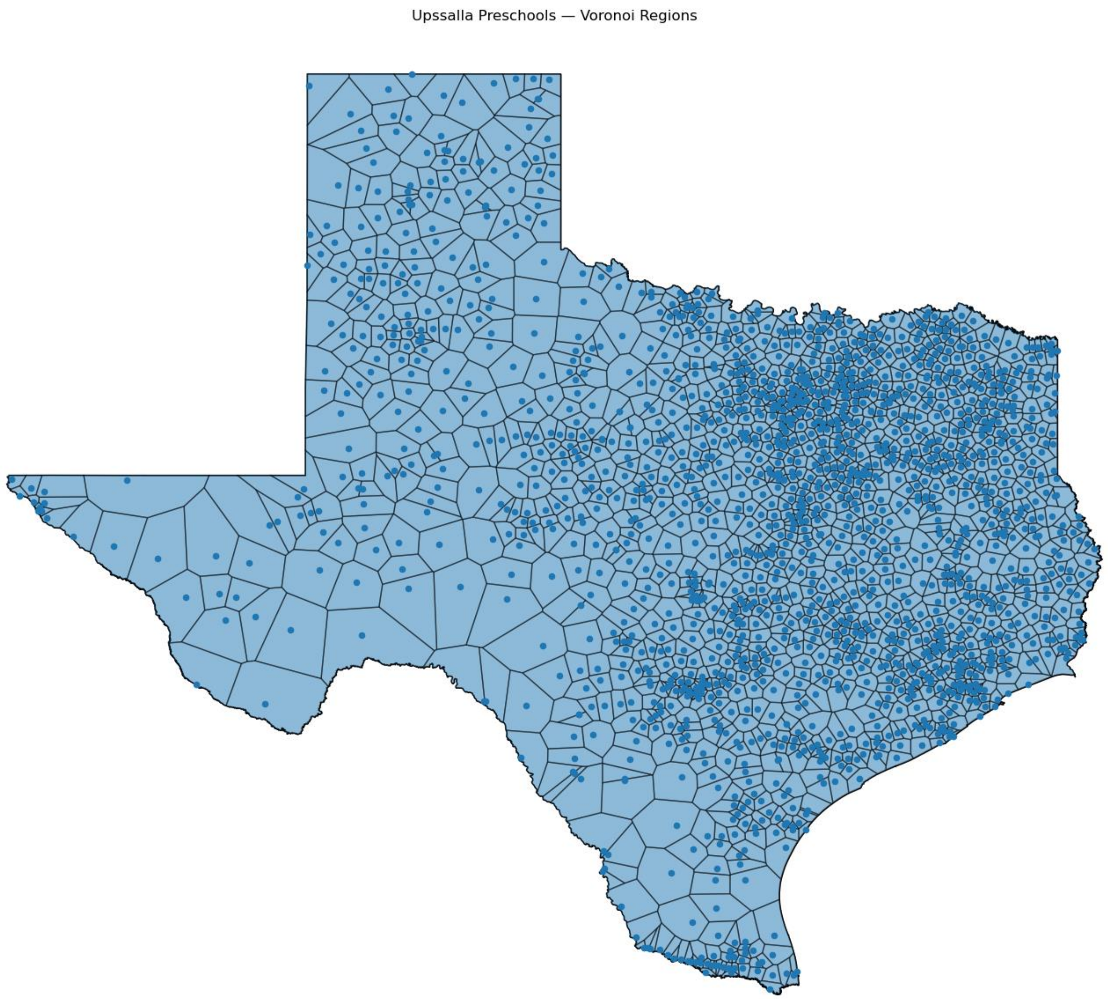


Fire Frequency ARIMA



Optimization Model

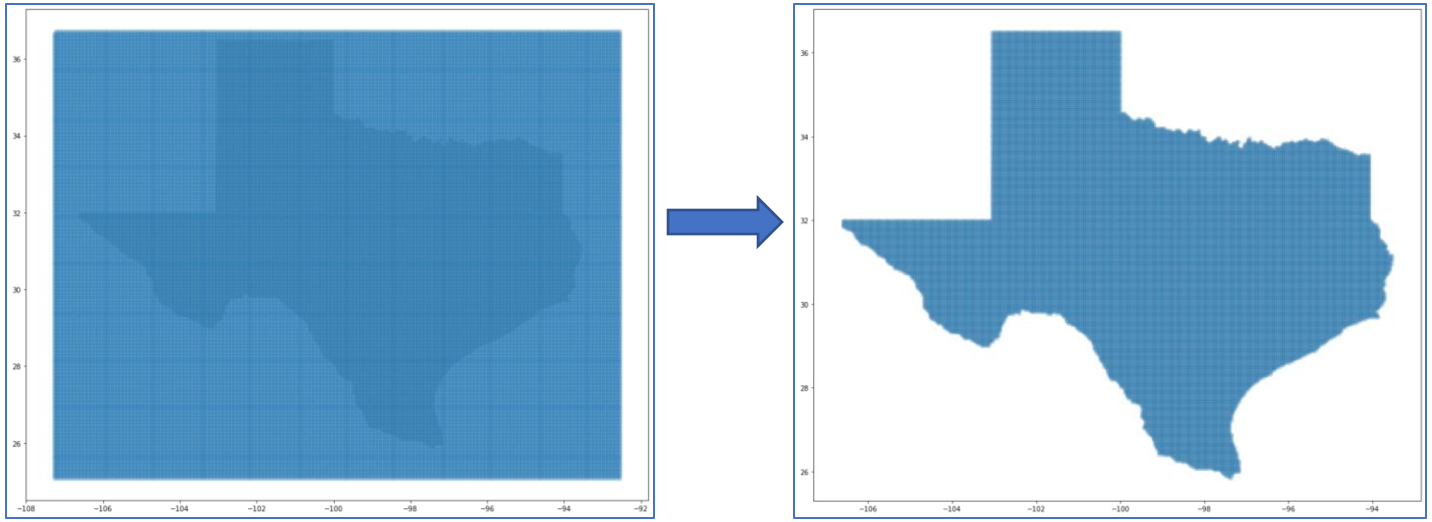
Optimization Problem Approach



1. From Voronoi Diagram Intention was to explore how spread out the Fire-stations locations are and estimate areas it needs to cover.
2. From the image can see how thinly spread on west side of Texas, also most of the fire stations are concentrated in east side because its most fire prone zone.

Optimization

1. Discretize Texas into grids of 5 km x 5 km
2. Identify fires which could not be responded within 8 mins
3. Assign grid_id to each fire
4. Calculate number of grids that can respond to the fires within a specific distance
5. Run optimization model



Discretization

$$T = (0.65 + (Distance\ km) \times K) \times 60$$

T = travel time in minutes

K = defined constant based on the average speed of a given apparatus over a 5-mile course

Optimization Model

Objective Function

$$MAX\ Z = \sum_{t\ belonging\ to\ Set\ N} X_t * F_t$$

Constraints:

$$\sum_{i\ belongs\ to\ coverage_i} X_i \leq 1$$

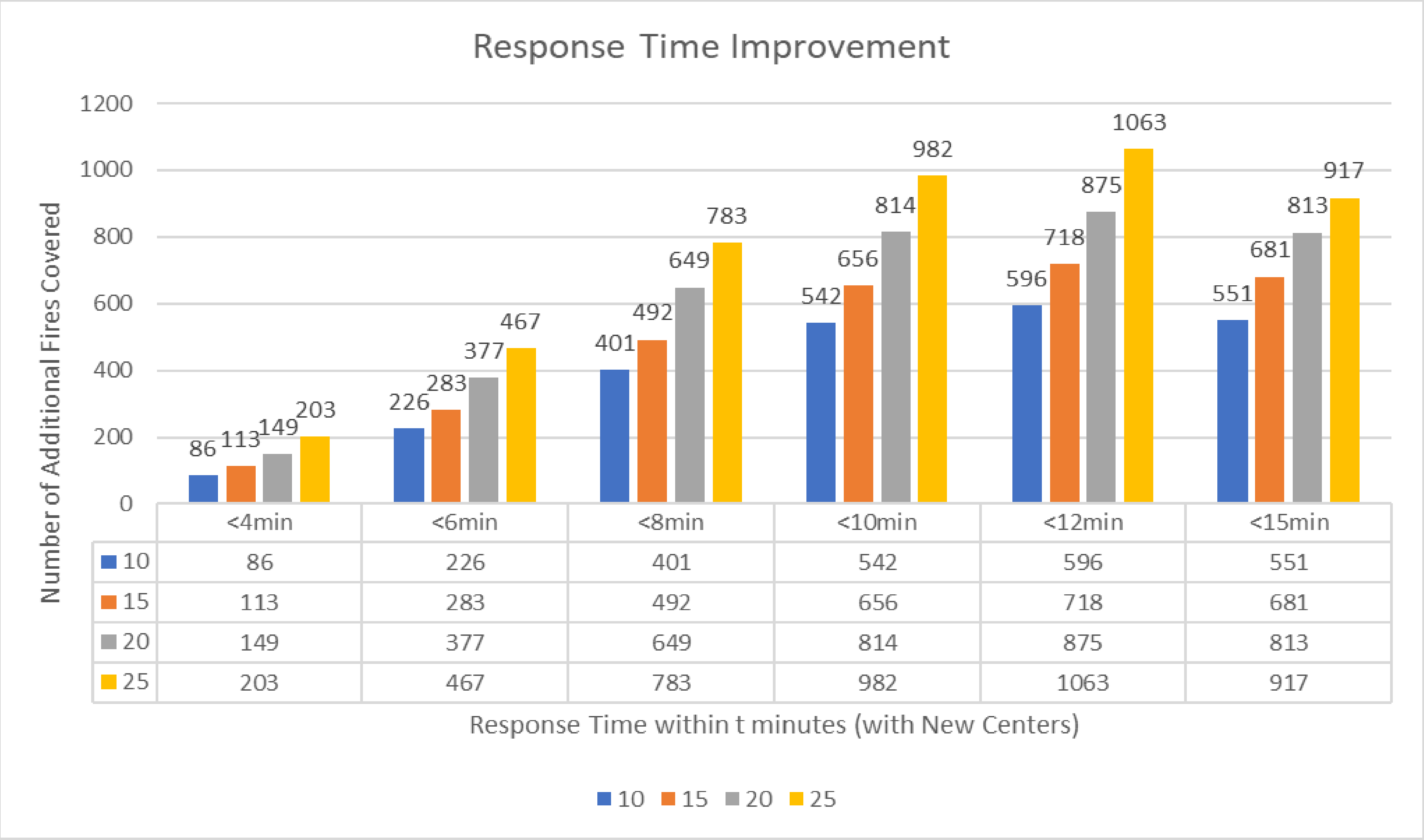
$$\sum_i X_i \leq P$$

Modification of Set Coverage Problem

N : Number of Grids in which Missing fire is Present
Coverage: Set of those grids which if have fire stations will be able to respond to Grid "I" within the specified response time
P : Upper limit on Number of Fire stations to be Built
F: Aggregated fire size (Sum) of all the grids that will become accessible if fire station is built at Grid I.

Note: Here, Xi represents binary decision variable which is 1 if fire-station is built at grid I and 0 otherwise.

Optimization Results



	<4min	<6min	<8min	<10min	<12min	<15min
10	86	226	401	542	596	551
15	113	283	492	656	718	681
20	149	377	649	814	875	813
25	203	467	783	982	1063	917

Future Prospects

- Connect the optimization model with google API, to find possible fire locations in real-time.
- Explore better methods for interpolating weather data
- Understand forest sampling Procedures and merging it with Fire and Weather Data
- Explore Bayesian Network (ADG) for classification
- ML Model for estimated duration of fire using already created features like mapping with closest station, resources available etc.
- As it is highly imbalanced classification problem can maybe convert to Anomaly Detection with anomaly behaviour as large fires of type "F" and "G"
- Streamlit app/Tableau/Live app for easy accesibility.

Limitations/Challenges

- Challenges in FIA
- High Class Imbalance
- Huge Dataset with huge computational effort in data wrangling as well as Model Training
- Understanding Actual Physics behind Fire spread

Learning/Tools Used

- Software's – Python, Jupyter Notebook, Google Collab, SQL, Tableau
- Python Libraries – Pandas, Numpy, Scikit-Learn, Plotly, Matplotlib, Geopandas, Folium, Bokeh, Vaex, Dask, Pycaret, Shapely, StatsModels, BaseMap, Descartes, etc.
- ML Algorithms :
  - Clustering – K-means, DBSCAN
  - Regression
  - Classification – Random Forest, Decision Trees, Logistic Regression, KNN, Gradient Boosting, SVM
- Additional Techniques:
  - Boosting, Bagging, Imbalance Methods(SMOTE, Class_weight, Random Under & Over sampling), Ensambles - Voting