Forest Fire Area Prediction

Durga Prasad (20172139)
Ritabrata Podder (2019900031)
Akshay Kharbanda (20171037)

01

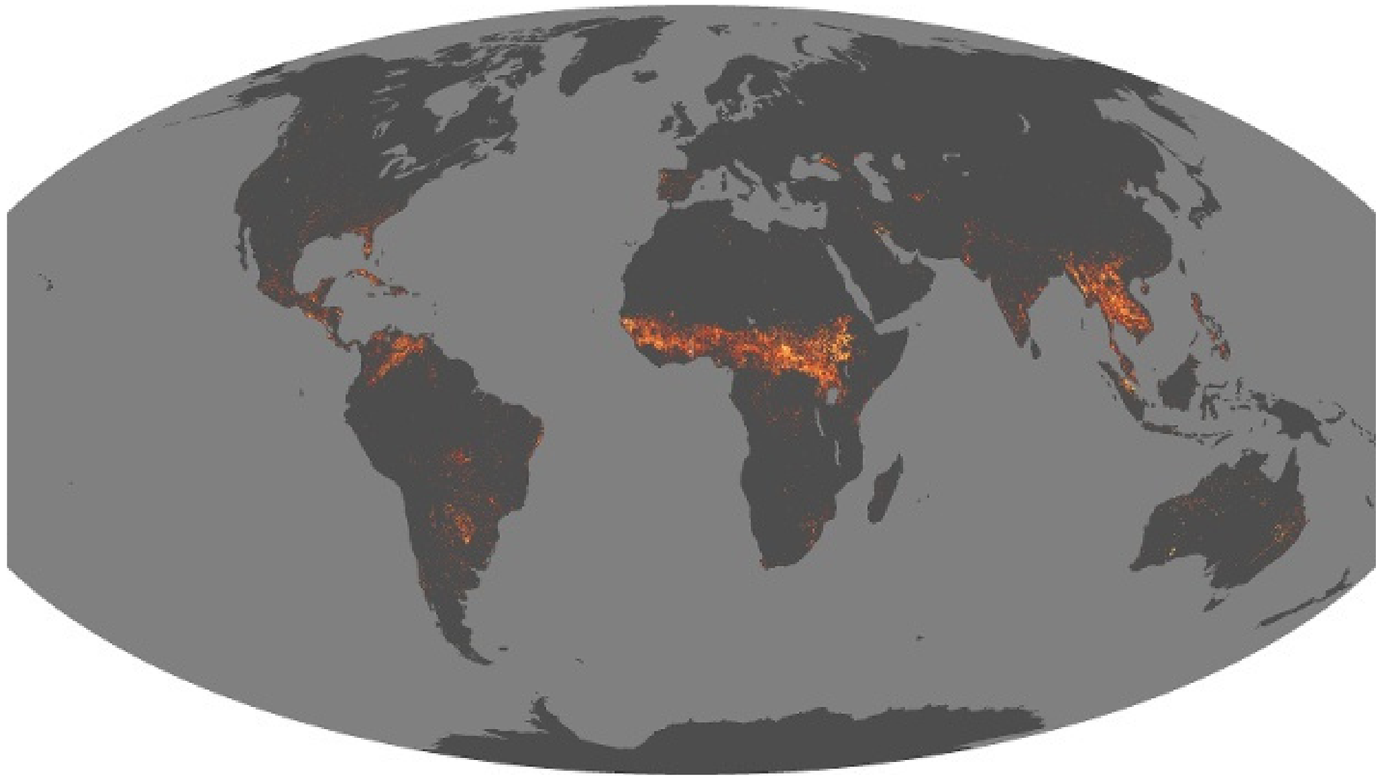
Introduction

Forest fire is the burning of tropical, temperate and boreal forest either by natural fire or man-made fire and is related to land clearing and deforestation. Natural forest fire includes an unplanned burning of forest due to lightning, while human-induced forest fire results from the unauthorized burning practice of forests for attaining farmland.

In recent years, wildfires have destroyed entire communities and damaged billions of dollars’ worth of property.

For example, the recent Australian wildfire caused:

- 46 million acres were scorched
- Nearly 3 billion animals were affected
- Destroyed over 5,900 buildings
- Killed at least 34 people



The fire maps show the locations of actively burning fires around the world on a monthly basis, based on observations from the MODIS on NASA's Terra satellite.

Objective

Developing a reliable model predicting forest fire burn size. Our study extracts GIS data and enhances the popular Kaggle forest fire Machine Learning dataset. We use the existing ML models but use an enhanced dataset to get a better MAE metric.

Datasets examined for Machine Learning models

Kaggle Wildfire Dataset

Contains a spatial database of wildfires that occurred in the United States from 1992 - 2015. The following core data elements were included in this data: **discovery date**, **final fire size**, and a **point location**.

Basic error-checking was performed,and redundant records were identified and removed, to the degree possible. The resulting product includes 1.88 million geo-referenced wildfire records, representing a total of 140 million acres burned during the 24-year period.

Adam and Ben (2019) used few features that are relevant for developing a regression model: **'DISCOVERY_DATE'**, **'DISCOVERY_DOY'**, **'STAT_CAUSE_CODE'**, **'FIRE_SIZE'**,**'LATITUDE'**, **'LONGITUDE'**.

UCI Dataset

UCI machine learning repository is a collection of 512fires from a large national park in northern Portugal.

Unlike the Kaggle dataset, this one provides weather features associated with each fire. We used most of the provided features, which included an **x and y location** within the national park, **month**, **day of the week**, **Fine Fuel Moisture Code (FFMC)**, **Duff Moisture Code (DMC)**, **Drought Code (DC)**, **Initial Spread Index (ISI)**, **temperature (Celsius)**, **relative humidity**, **wind**, and **rain**, to predict the burn area.

Dilemma with the Datasets

1 Kaggle Data

Lots of data points, BUT, very few features (6)

2 UCI Data

Rich in features (13), BUT, meagre data points

Our Solution

Enriching the features of the Kaggle Dataset by augmenting the dataset with more features.

03

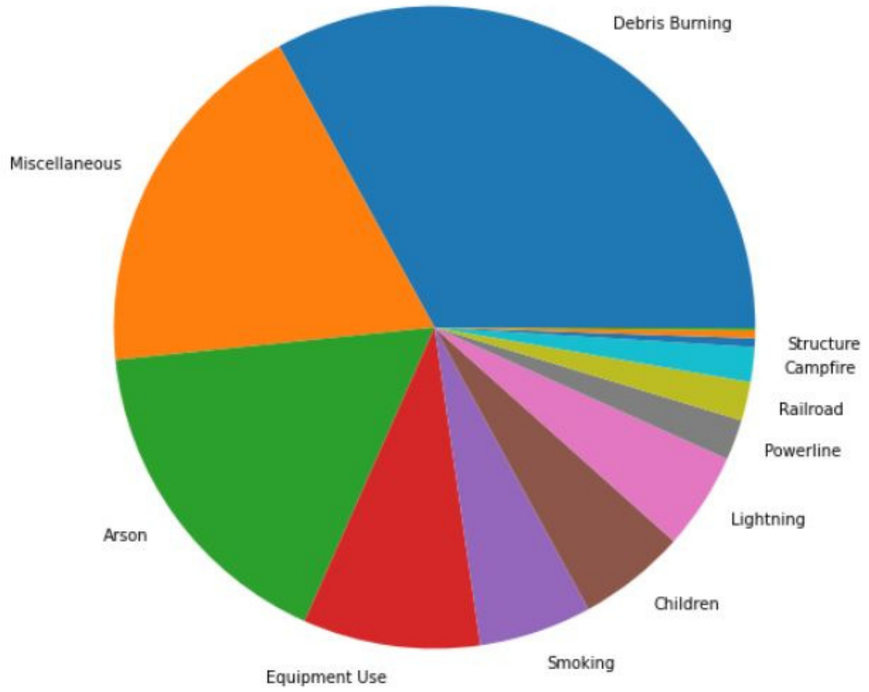
Methodology

- Augmenting the Kaggle Dataset
- Application of ML Algorithms and comparison with results from original plain Kaggle dataset.

Augmenting the Kaggle Dataset

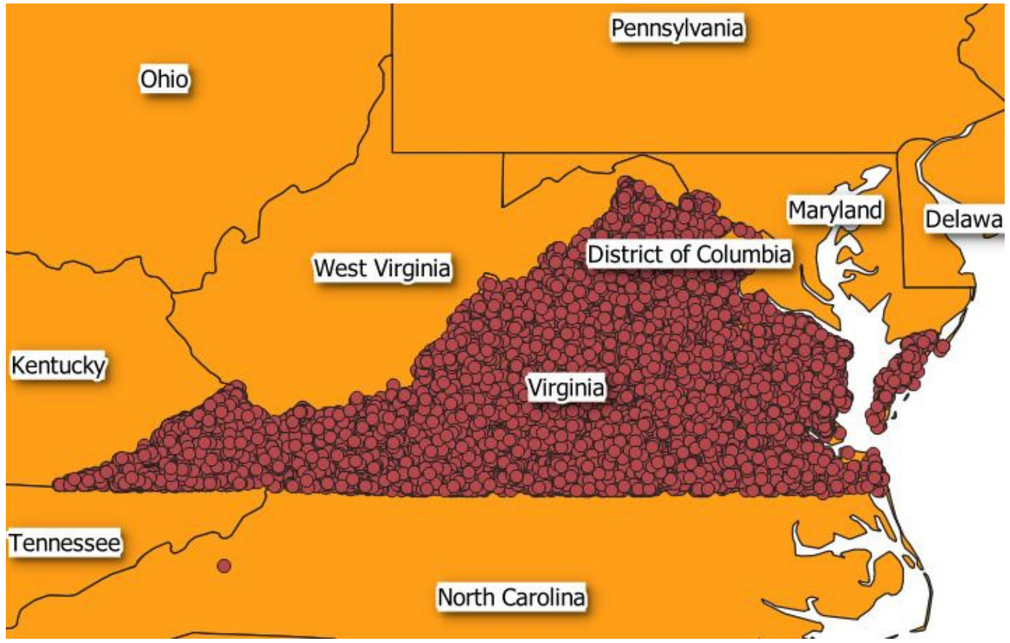
STEP 1

Preprocessing: Dropping of columns that do not contribute to the model learning. Analysis on this cleaned dataset shows that **Debris burning** is the main cause for forest fires in Virginia



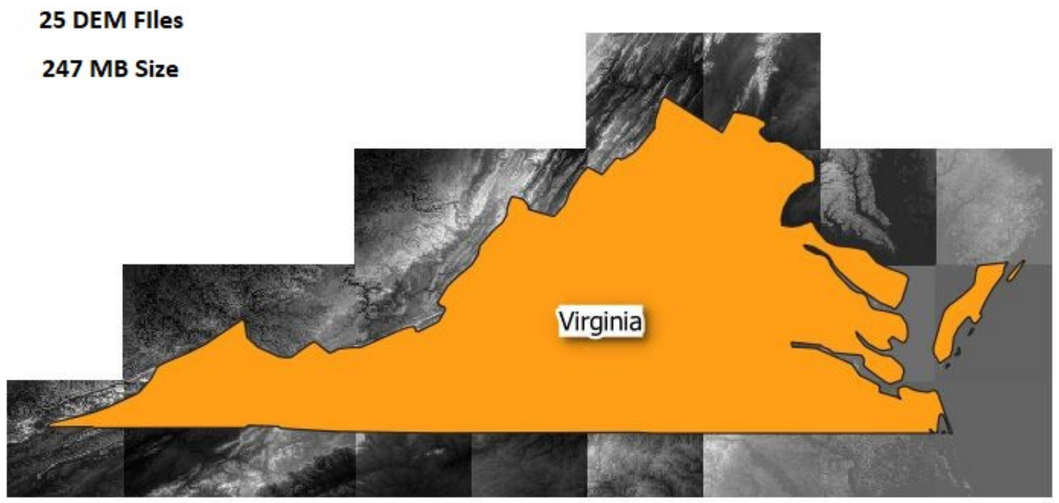
STEP 2

We used QGIS to apply spatial filter on whole dataset to select fires within Virginia state of the US.



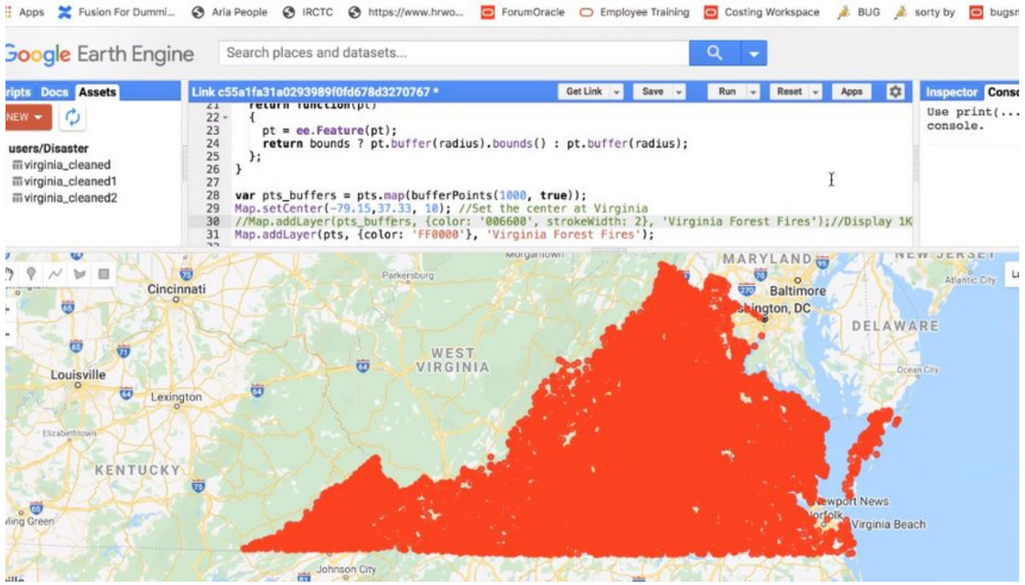
STEP 3

We used the USGS DEM data to derive E**levation** and **Slope** for each of the spatial location of fire. We either used tools in QGIS manually or wrote python scripts to derive this data. For Virginia region, DEMs were **247 MB** in size and **25** in number.



STEP 4

We customized our Google Earth Engine (GEE) scripts to derive **min**, **max** and **average elevation** and **mean slope** for a 1x1 km square boundary around each spatial location of fire. All these four features are used to augment the Kaggle data for Virginia.



Models Applied to the Enriched Dataset

The following machine learning models are experimented with plain Kaggle data and the augmented dataset :

- Deep Learning
- SVM
- MLP Regressor
- KNN Regressor
- DecisionTree Regressor
- Stacking Regressor

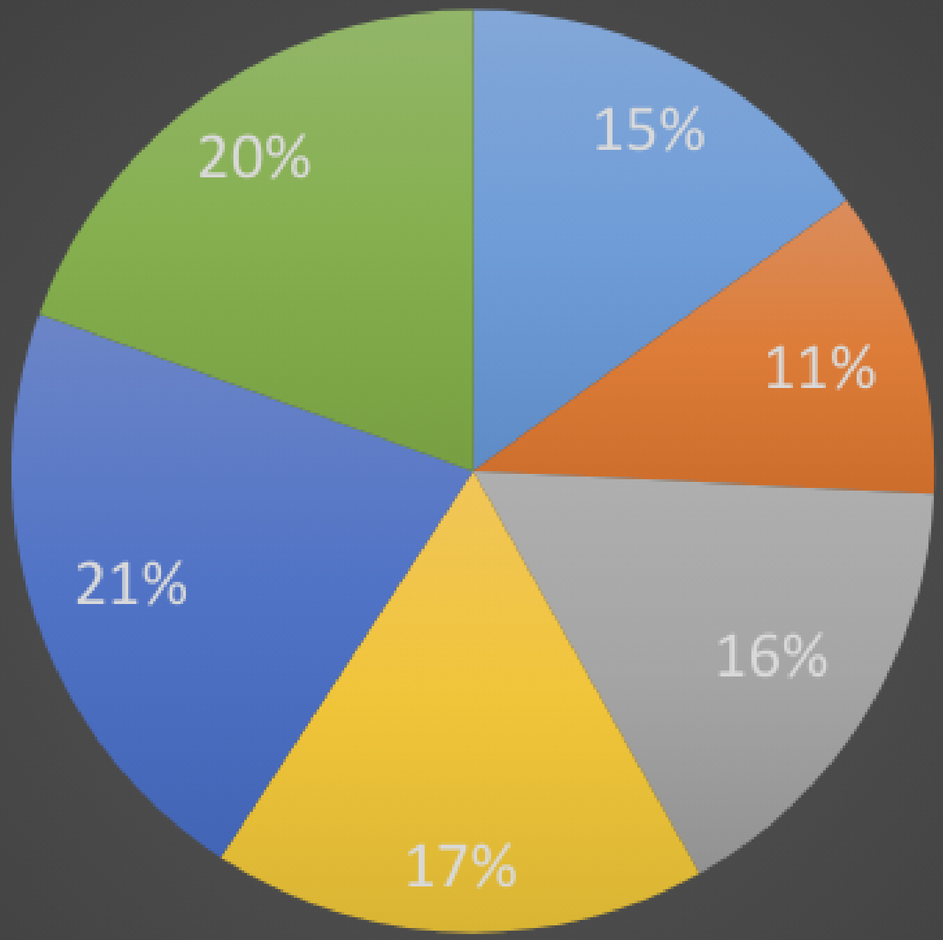
04

Results and Discussion

MAE with kaggle data

01

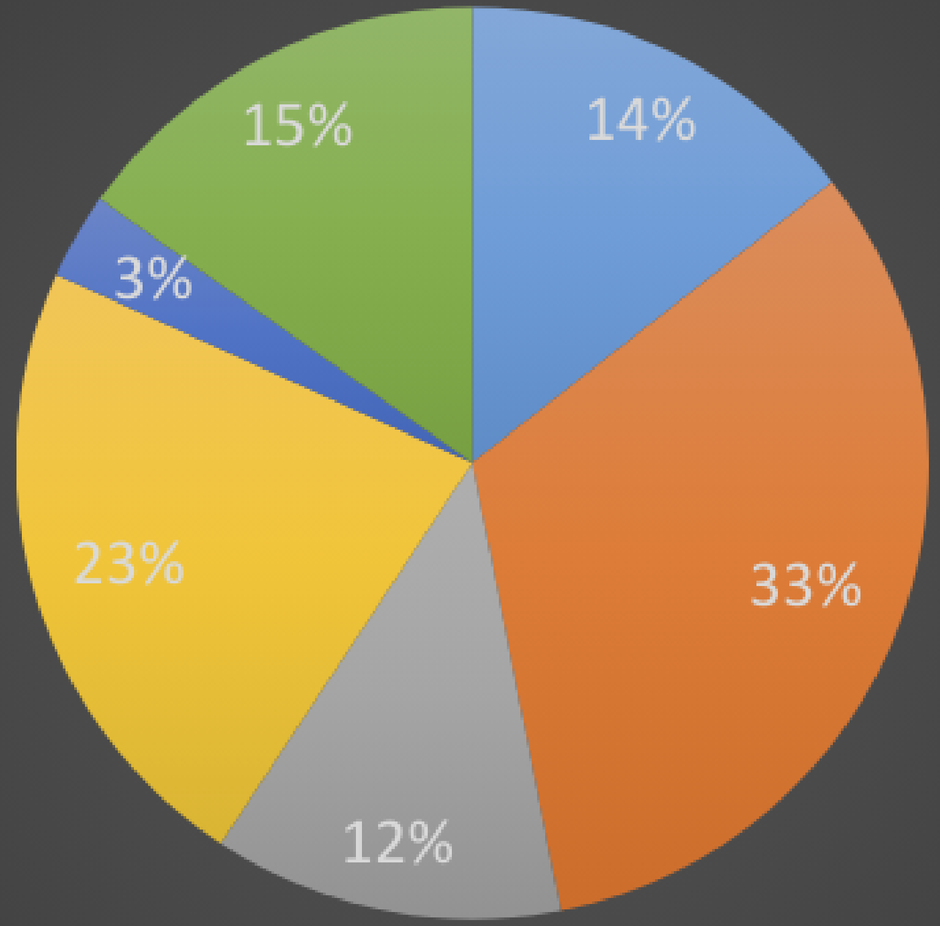
SVM model performed better among all models in the original paper.



MAE with enhanced data

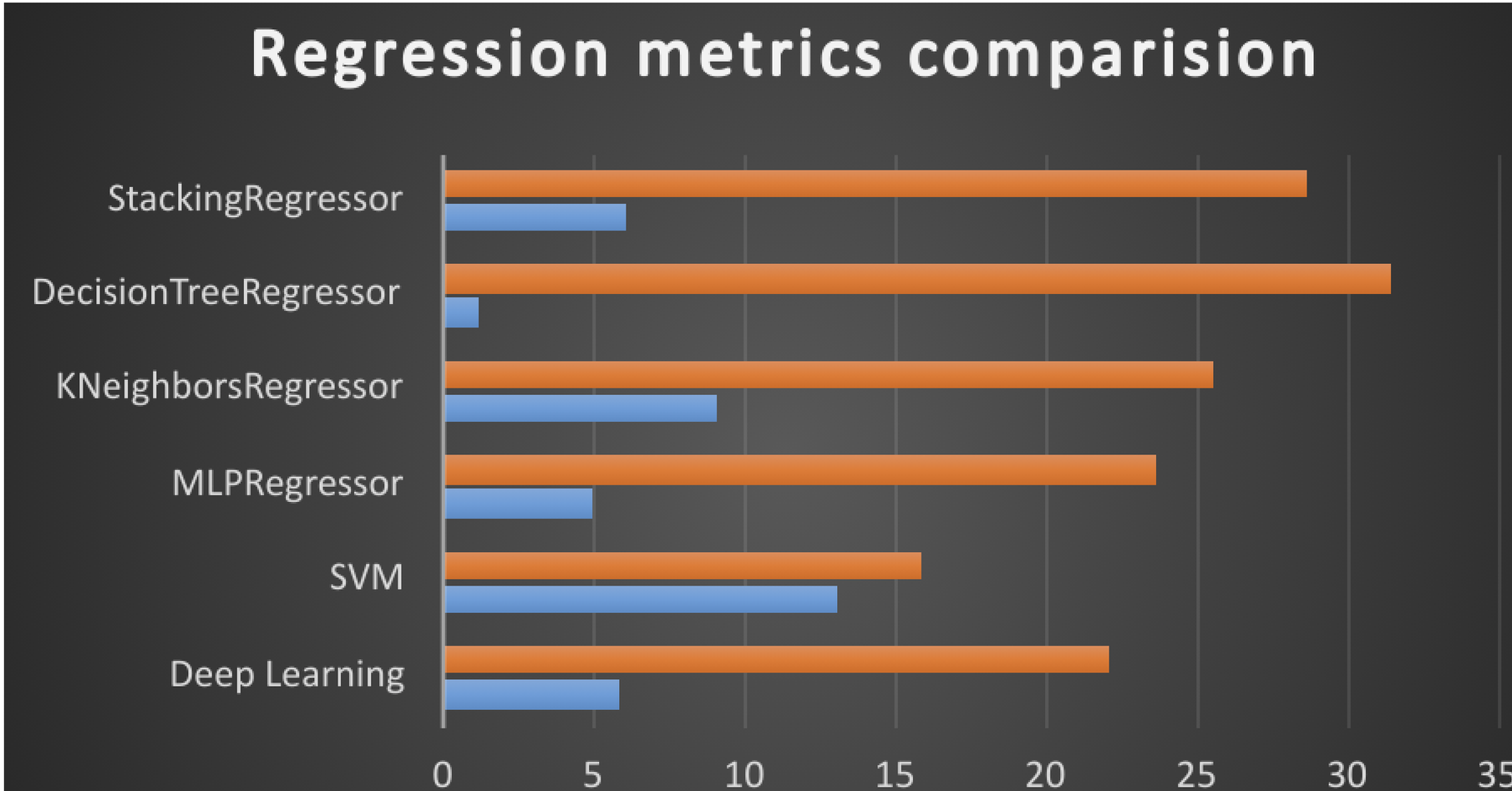
02

Models using augmented dataset performed well on Decision Tree regression with least MAE at 3%.



03

Figure below illustrates the MAE of models run on both the datasets.



05

Conclusion

The convergence of Remote Sensing, GIS and Machine Learning make the wildfire prediction models more accurate. There is a definite advantage of making use of cloud-based single-window GIS APIs lies in GEE for faster access and analysis. Cloud based GIS APIs offer more flexibility as historical data is readily available.

06

Future Work

These results encourage us to continue with our future research that intends to expand the study to include parameters beyond elevation and slope. In addition to Topography, Drought, Vegetation and Weather are recommended by several other studies. It is argued that forest fires occur based on seasons if the causes are natural. It will be interesting to study phenomenon using Recurrent Neural Networks (RNN) like LSTM or GRU.

Questions ?