CS 225 - Spatial Computing
Predicting Forest Fires

Our Team(Group-5)

Abhishek Premnath (862323549)

Balaji Arunachalam (862321425)

Madhusudhan Tungamitta (862325125)

Shreya Godishala (862313765)

Thirumalai Vinjamoor Akhil Srinivas (862320790)

Data Cleaning

Three different datasets were compared for geo spatial data and the most appropriate dataset for our particular model from University of California, Irvine's forest fire dataset was chosen. The dataset has a lot of unnecessary features and some attributes had missing values too.

```
v=df.dropDuplicates(["State"]).select("State").collect()
```

Thus, the dimensions of the dataset had to be reduced and the null values were addressed by either removing the rows associated with it entirely or taking the average values from similar records. Also, to reduce the complexity of the dataset, certain attributes such as rain and relative humidity were combined.

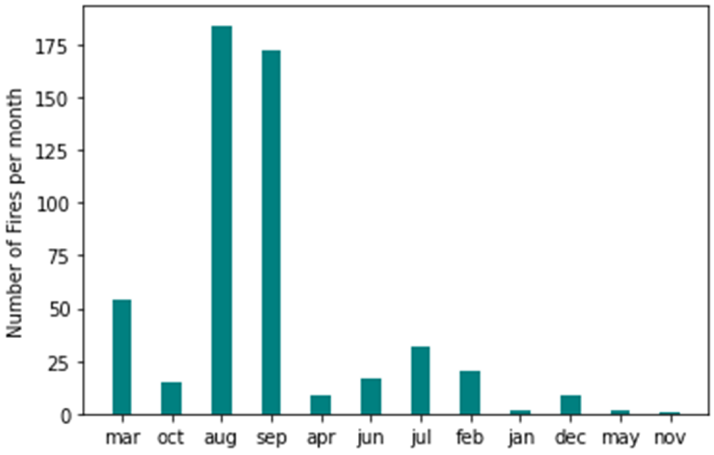
For the appropriate locations of the data collected, X and Y spatial coordinates within Montesinho park map have also been provided on a scale of 1-9. These values would also be normalized prior to training the model.

```
for l in states:
  temp=df.filter(df.State==l).select("Type","Severity","EndTime(UTC)","Precipitation(in)")
```

Data Visualization

To get a better understanding of the dataset used and to find the underlying corelation between various attributes used. Some correlated attributes are Duff Moisture Code

(DMC) with respect to humidity and the Fire Weather Index (FWI) with respect to the Fine Fuel Moisture Code (FFMC)
As a part of initial visualtion we are able to see that most of the forest fires happens in the months of August and September months



Further progress

Algorithms Planning to use
KNN, Random Forest,Support Vector Regression,Linear Regression

Taxonomy

1.	Click here	A Data Mining Approach to Predict Forest Fires using Meteorological Data
2.	Click here	Learning to predict forest fires with different data mining techniques
3.	Click Here	Predicting Forest Fire Using Remote Sensing Data And Machine Learning

Deliverables & Timelines From Midterm:

Deliverables Type	Old Deadline Time	New Deadline Adjustment
1. Literature Review	10/22/2022	10/22/2022

2. Data Cleaning	10/29/2022	10/29/2022
3. Implementing Model using different Machine learning Algorithms	11/12/2022	11/20/2022
4. Training our Model	11/19/2022	11/24/2022
5 Visualizing Data	11/26/2022	11/26/2022
6. Final Report	12/03/2022	12/03/2022

Clarifications From the Feedback Of Project Proposal:

1.The x,y coordinates themselves are not meaningful in this case?Basically x,y co ordinates are the grid representation of the Montesinho national park

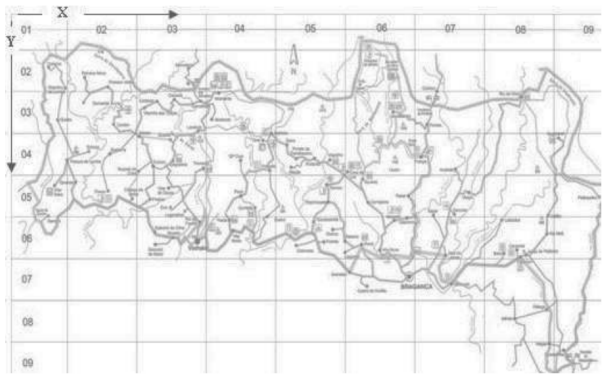


Fig. 2. The map of the Montesinho natural park

2. Can it be used to predict forest fires from other regions?

I don’t think so because the data varies and here they represented as x,y co-ordinated and it may not be same in all other datasets