

PREDICTING FOREST FIRES

A project by Marwa Allam, Lisa Weiss, Sanskriti Bajaj,
Sameeksha Mahajan and Gabriela Urquieta

OUR MOTIVATION

- Forest Fires are a major environmental issue, causing economical and ecological damage.
- Use Descriptive Statistics and Python Programming to work towards the prediction of such catastrophic events.
- Experiment with different Machine Learning Algorithms.

INITIAL QUESTIONS

Before starting the project we formulated the following questions :

- Which Machine Learning Model is the best for predicting Forest Fires ?
- Which months report the highest number of Forest Fires ?
- Which days report the highest number of Forest Fires ?

MACHINE LEARNING PROCESS

Gathering Data

We have used the Forest Fires open Data set from UCI Machine Learning Repository. The Data set enlists features that affect the occurrence of Forest Fires, like rain, wind and temperature.

Cleaning Data

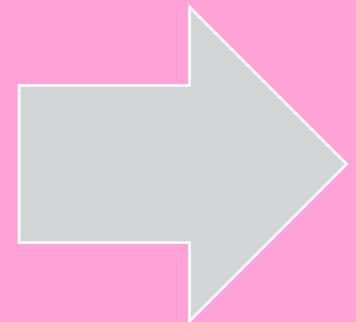
We performed Exploratory Data Analysis and found outliers as well as some unimportant features. We transformed the data by normalization and one hot encoding.

Building ML Models

Each member chose different ML models, allowing us to explore several algorithms like Regression, Neural Networks, Lasso and SVM. We used Python Libraries like Scikit Learn and Keras.

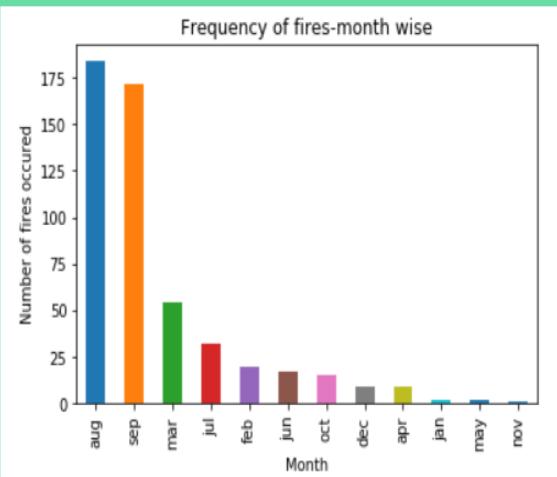
Evaluating Results

Through comparison of the mean squared error (as an evaluation criteria) and plots of predicted data vs test data, we could determine which algorithms produced better results.

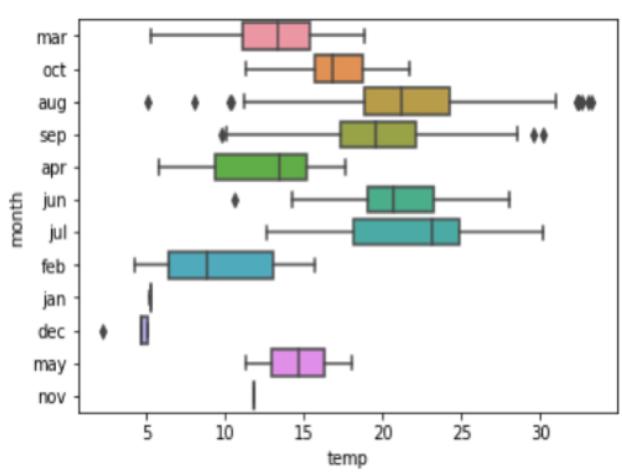


Data Visualizations
Using Python Libraries like Matplotlib and Seaborn

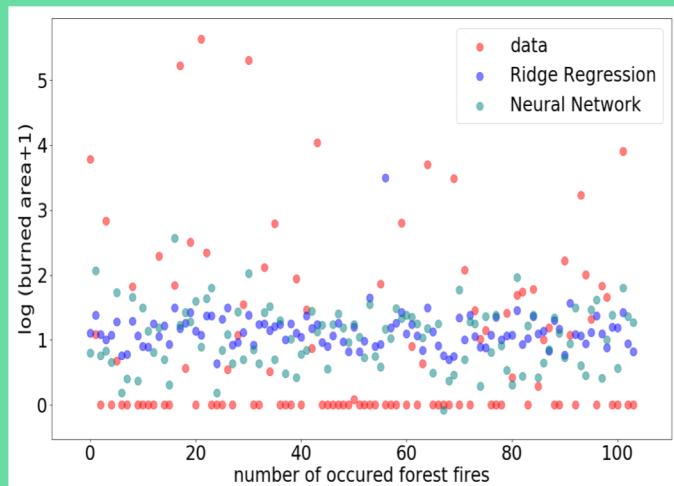
VISUALIZATION OF RESULTS



Exploring the data:
Occurrence of Forest
Fires per month



Illustrating the temperature in each
month to find the outliers in the data



Predicting the burned area
caused by forest fires using
Neural Networks and Ridge
Regression

CONCLUSIONS



DAYS WITH MOST
NUMBER OF
FOREST FIRES

FRIDAY

SUNDAY

SATURDAY

MONTHS WITH
MOST NUMBER OF
FOREST FIRES

AUGUST

SEPTEMBER

BEST ML
MODELS FOR
PREDICTION

RIDGE
REGRESSION

NEURAL
NETWORK

WORLDWIDE WOMEN PROGRAMMERS



LISA WEISS - Germany



SANSKRITI BAJAJ - India



GABRIELA URQUIETA - Bolivia



SAMEEKSHA MAHAJAN - India



MARWA ALLAM - Egypt

REFERENCES

SOURCE OF DATA :

- Cortez, Morais, "A Data Mining Approach to Predict Forest Fires using Meteorological Data"; 21/7/2018/<http://www3.dsi.uminho.pt/pcortezfires.pdf>
- <https://archive.ics.uci.edu/ml/datasets/Forest+Fires>

DOCUMENTATION OF LIBRARIES USED :

- <http://scikit-learn.org/stable/>
- <https://matplotlib.org/contents.html>
- <https://keras.io/>
- <https://seaborn.pydata.org/>
- <https://pandas.pydata.org/>
- <https://www.scipy.org/>
- <http://www.numpy.org/>

OTHER RESOURCES :

- Jason Brownie, Why One-Hot Encode Data in Machine Learning, 21/7/18, <https://machinelearningmastery.com/why-one-hot-encode-data-in-machine-learning/>
- Jason Brownie, Develop Your First Neural Network in Python With Keras, Step-By-Step, 21/7/18, <https://machinelearningmastery.com/tutorial-first-neural-network-python-keras/>
- <https://www.istockphoto.com/in/photo/hand-counting-three-fingers-gm471731759-26258534> ; 21/07/2018