E6893 Big Data Analytics

Wildfire Exploratory Visualization and Cause Prediction

Project ID: 201912-2
Team Members (with UNI): Sidharth Bambah (sb4283) || Vedant Dave (vad2134)
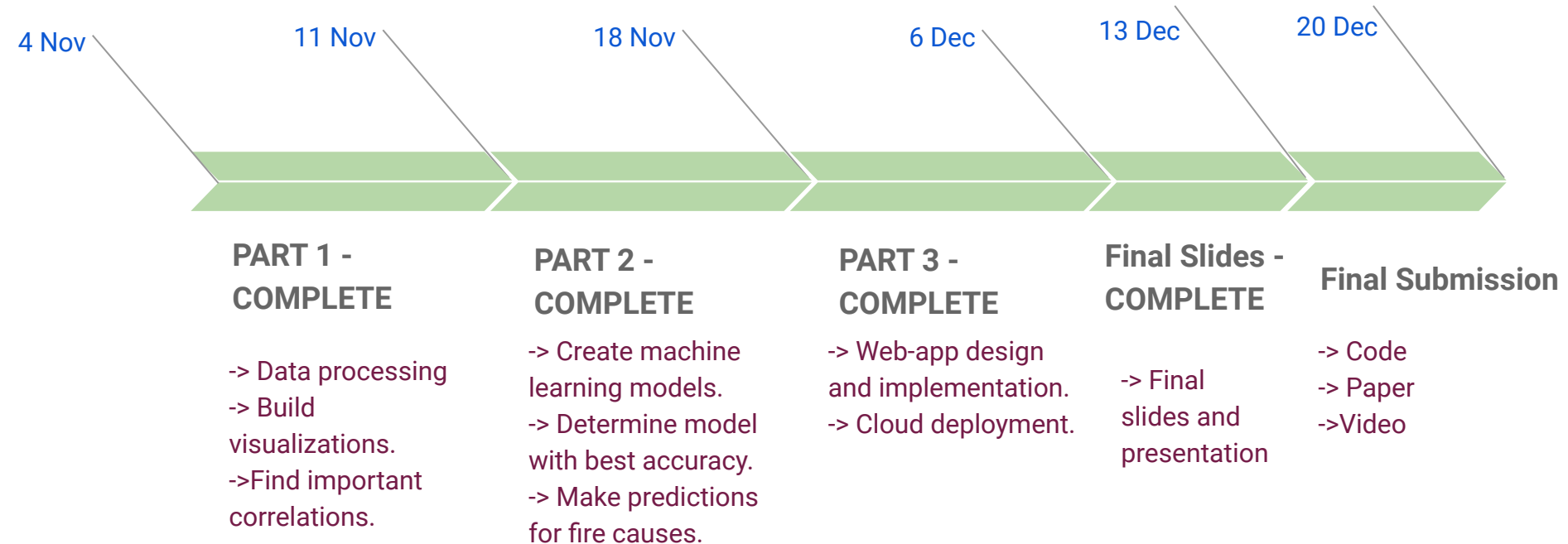
Introduction

- Predict causes of fires for prevention and government spending allocation

- Exploratory Analysis

- Machine Learning

Milestones

4 Nov

11 Nov

18 Nov

6 Dec

13 Dec

20 Dec

PART 1 - COMPLETE

-> Data processing
-> Build visualizations.
->Find important correlations.

PART 2 - COMPLETE

-> Create machine learning models.
-> Determine model with best accuracy.
-> Make predictions for fire causes.

PART 3 - COMPLETE

-> Web-app design and implementation.
-> Cloud deployment.

Final Slides - COMPLETE

-> Final slides and presentation

Final Submission

-> Code
-> Paper
->Video

E6893 Big Data Analytics – Final Project Proposal

© 2018 CY Lin, Columbia University

Dataset

- 1.88 Million US Wildfires (24 years of geo-referenced wildfire records)

- License - CCO: Public Domain

- Size : 759 MB

- Spatial database of wildfires in the United States from 1992 to 2015

- Originally generated to support the national Fire Program Analysis (FPA) system
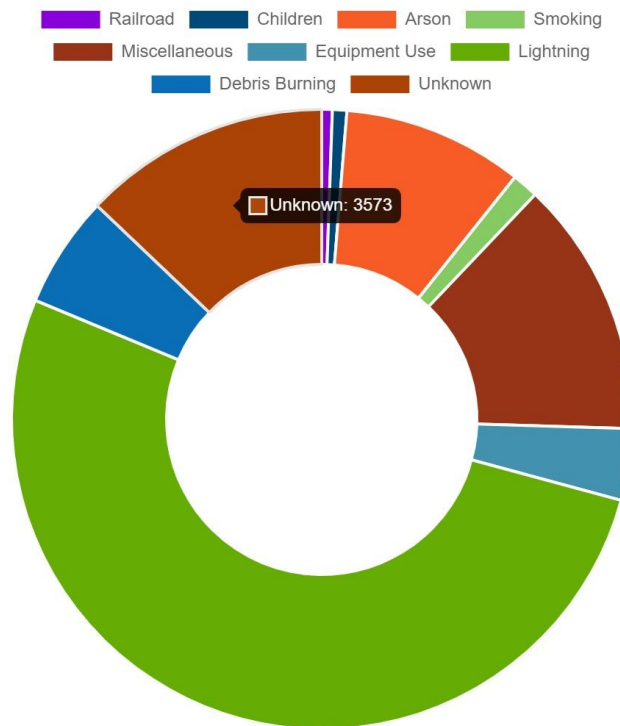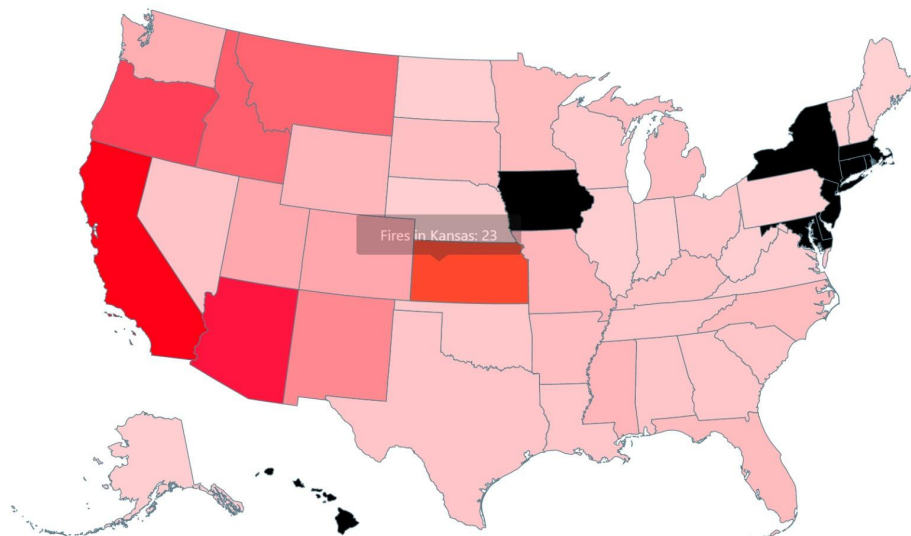
Data Processing

- Read SQLite database into CSV

- Convert Julian to Gregorian date

- Add day of week, month, and year

- Uploaded subset to MongoDB cluster
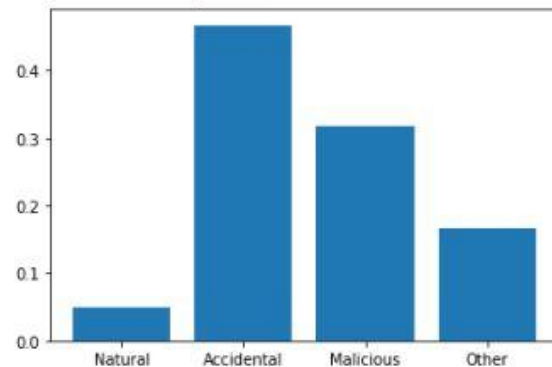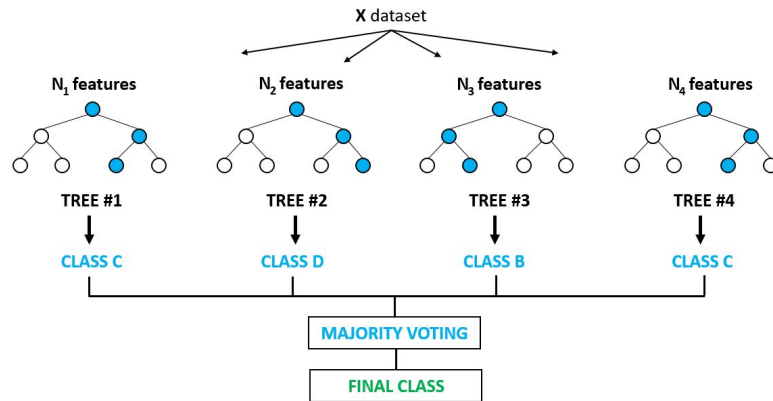
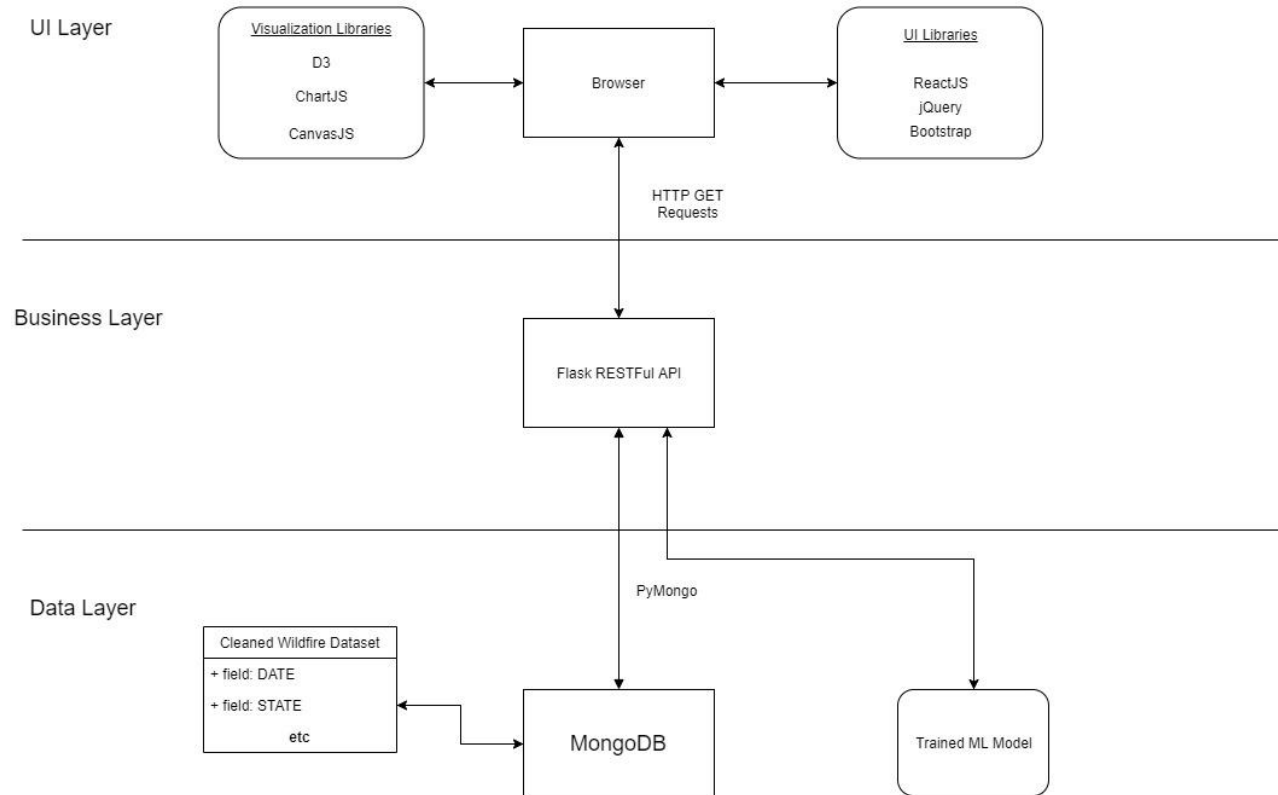# Exploratory Visualization
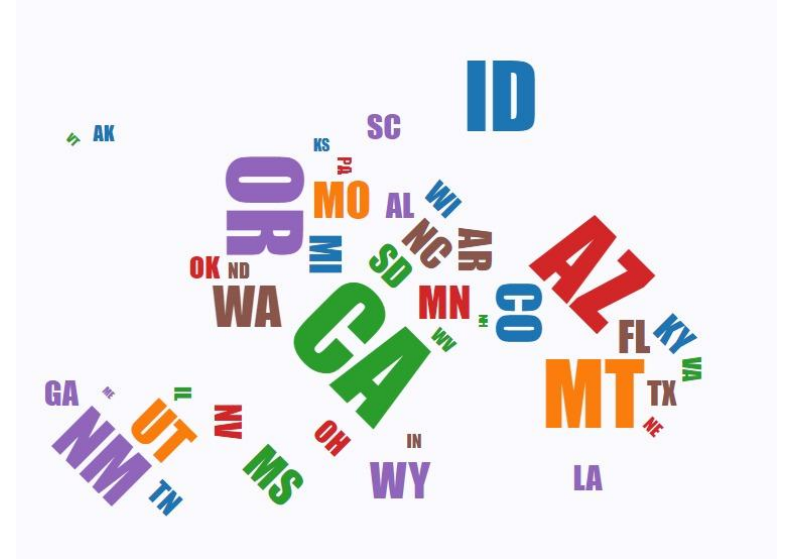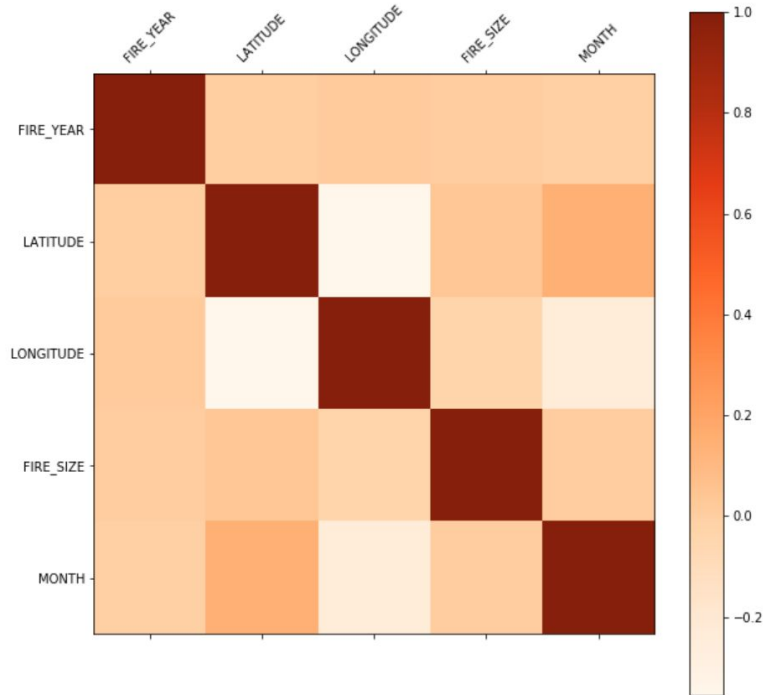
Exploratory Visualization

Machine Learning

- Cleaned the data by eliminating un-important Feature labels.

- Applied Random Forest Classifier Machine Learning algorithm after careful considera-tion to predict the most probable cause of the wildfire.

- Grouped all the different causes of wildfires into 4 main categories:
  1. Natural 2. Accidental 3. Malicious 4. Other

- User inputs for the prediction of causes of wildfires:
  1. Latitude 2. Longitude 3. Month 4. Day of Week

System Structure

Software Architecture for Wildfire Visualization and ML Application

UI Layer

Visualization Libraries
D3
ChartJS
CanvasJS

Browser

UI Libraries
ReactJS
jQuery
Bootstrap

HTTP GET Requests

Business Layer

Flask RESTFul API

Data Layer

PyMongo

Cleaned Wildfire Dataset
+ field: DATE
+ field: STATE
etc

MongoDB

Trained ML Model

S3 Bucket

Elastic Beanstalk

MongoDB

BigDataAnalytics

Overview   Real Time   Metrics   **Collections**   Profiler   Performance Advisor   Command Line Tools

DATABASES: 1 COLLECTIONS: 1

⟳ REFRESH

+ Create Database

🔍 NAMESPACES

▼ **Fires**

   | **fires**

Fires.fires

COLLECTION SIZE: 435.55MB   TOTAL DOCUMENTS: 1880465   INDEXES TOTAL SIZE: 17.94MB

Find   Indexes   Aggregation

INSERT DOCUMENT

FILTER {"filter":"example"}   **Find**   Reset

Thank You!

Demo

Website:
http://big-data-analytics-columbia.s3-website.us-east-2.amazonaws.com/

YouTube Video: https://youtu.be/aAZ1EnyyqEk