W210, Spring 2020

Thanh Le

Keane Johnson

Laura Chutny

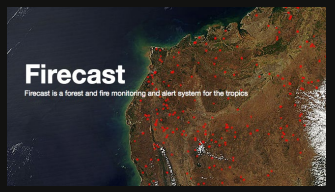
Problem Statement

- Climate change and population growth have combined to put people, animals and infrastructure at high risk of wildfire
- General public lack an easy-to-use reliable tool to predict the daily spread of wildfire
- This potentially leads to untimely evacuation plan and decisions, and potential loss of life

Current Prediction Models / Tools

- Empirical/physics based models
  - FARSITE (USA)
  - Prometheus (Canada)

- Mathematical growth prediction models
  - Wildfire Management Tool (World Wind Earth)
  - US Forest Service
- Physics-based models
  - National Center for Atmospheric Research (NCAR) Coupled Atmosphere-Wildland Fire-Environment (CAWFE) model
  - WRF-Fire
- Video/Image Monitoring
  - Alert Wildfire
- Machine Learning
  - Firecast - F-score of 6.4% (vs random of 1%)



HOTzone

- Existing approaches use exclusive data, are computationally expensive, and/or are not for general public use
  - HOTzone solves all these problems
- Uses publicly available fire data and weather data, a Convolutional Neural Network (CNN), and delivers predictions via a web application
- Users submit an address
  - A map displays the potential spread of an active fire near that area in the next 24 hours

<<DEMO>>

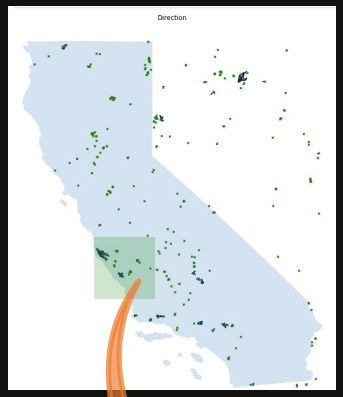
HOTzone Web App

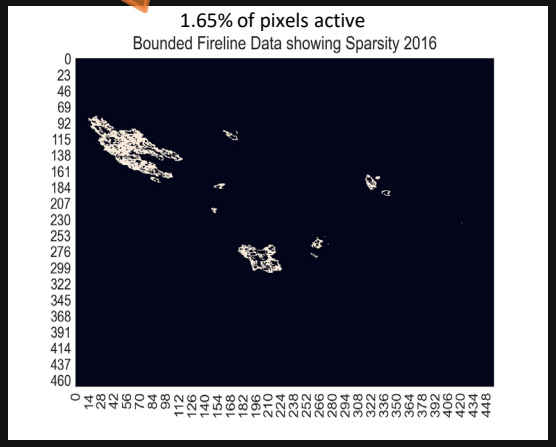
Project Infrastructure

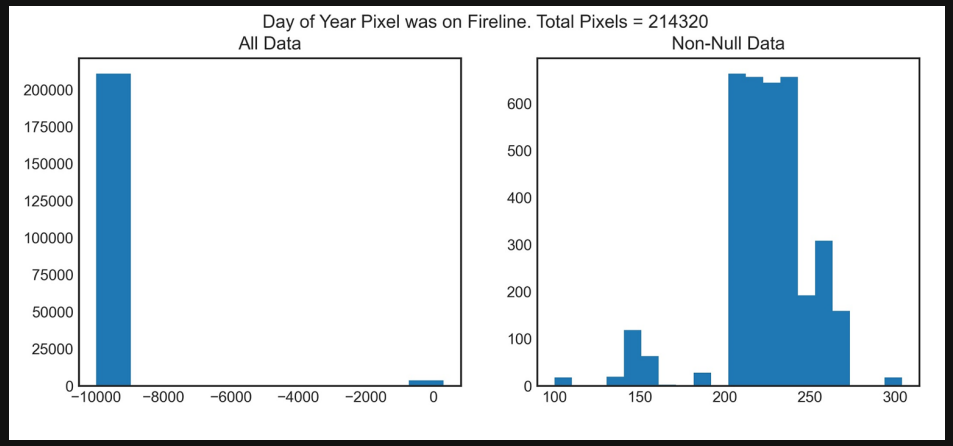
- Data Wrangling and Cleaning
  - Numpy / Pandas
  - GeoPandas, Rasterio, Earthpy (GIS)
  - Storage on shared AWS S3 bucket
- Modeling
  - Keras, Tensorflow backend
  - AWS - GPU
- Website / Webapp
  - HTML / Javascript / Python
  - Mapbox / Mapboxgl
  - Flask/Gunicorn/Nginx on AWS EC2 ML server

Fire Data - Global Fire Atlas

- Daily fire progression, 2003-2016
- Rasterized GeoTIFFs with ~500m x 500m pixels:
  - Day of year on fire
  - Speed of fire
  - Direction of fire movement
- Shapefiles with: Fire Ignition Point, Final Fire Size, Final Duration
- Localized to California/Nevada and then further to sub-section of CA
- Sparse Data





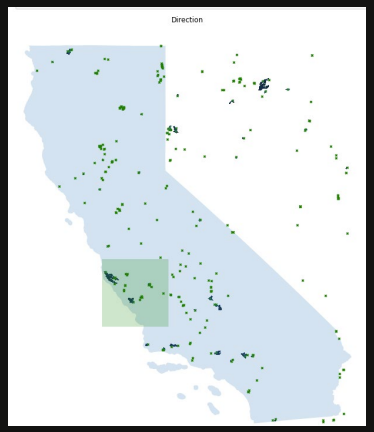


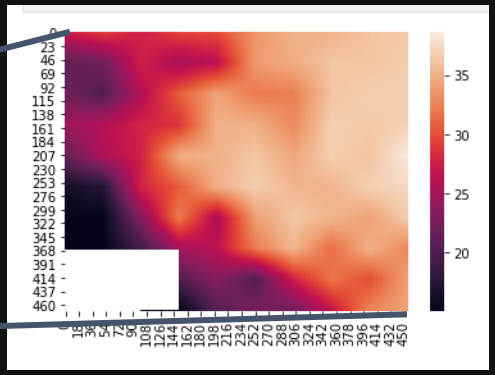
1 https://daac.ornl.gov/CMS/guides/CMS_Global_Fire_Atlas.html

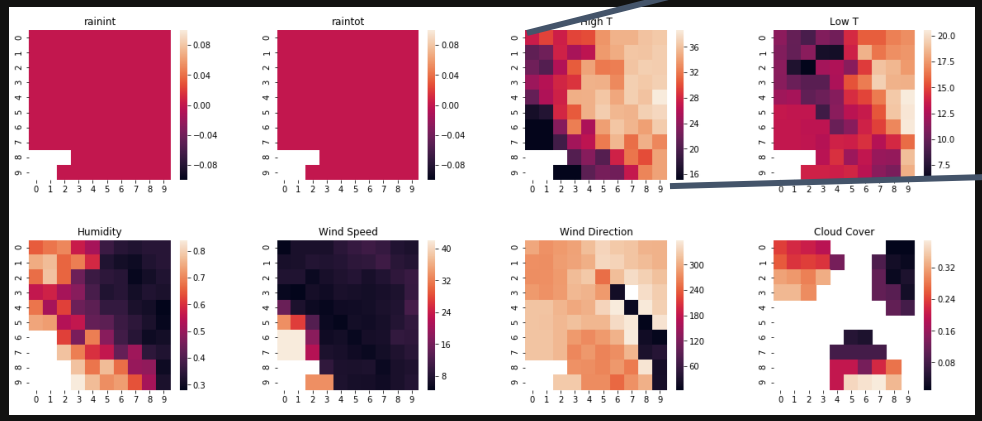
Weather Data - DarkSky

Weather Data

- Request by Lat/Long/Time
- Historical-training ⇒ Forecast-prediction
- Daily Averages
  - Rain Intensity
  - High/Low Temps
  - Wind Speed/Direction
  - Relative Humidity
- Interpolate (upsample) to match fire data grid







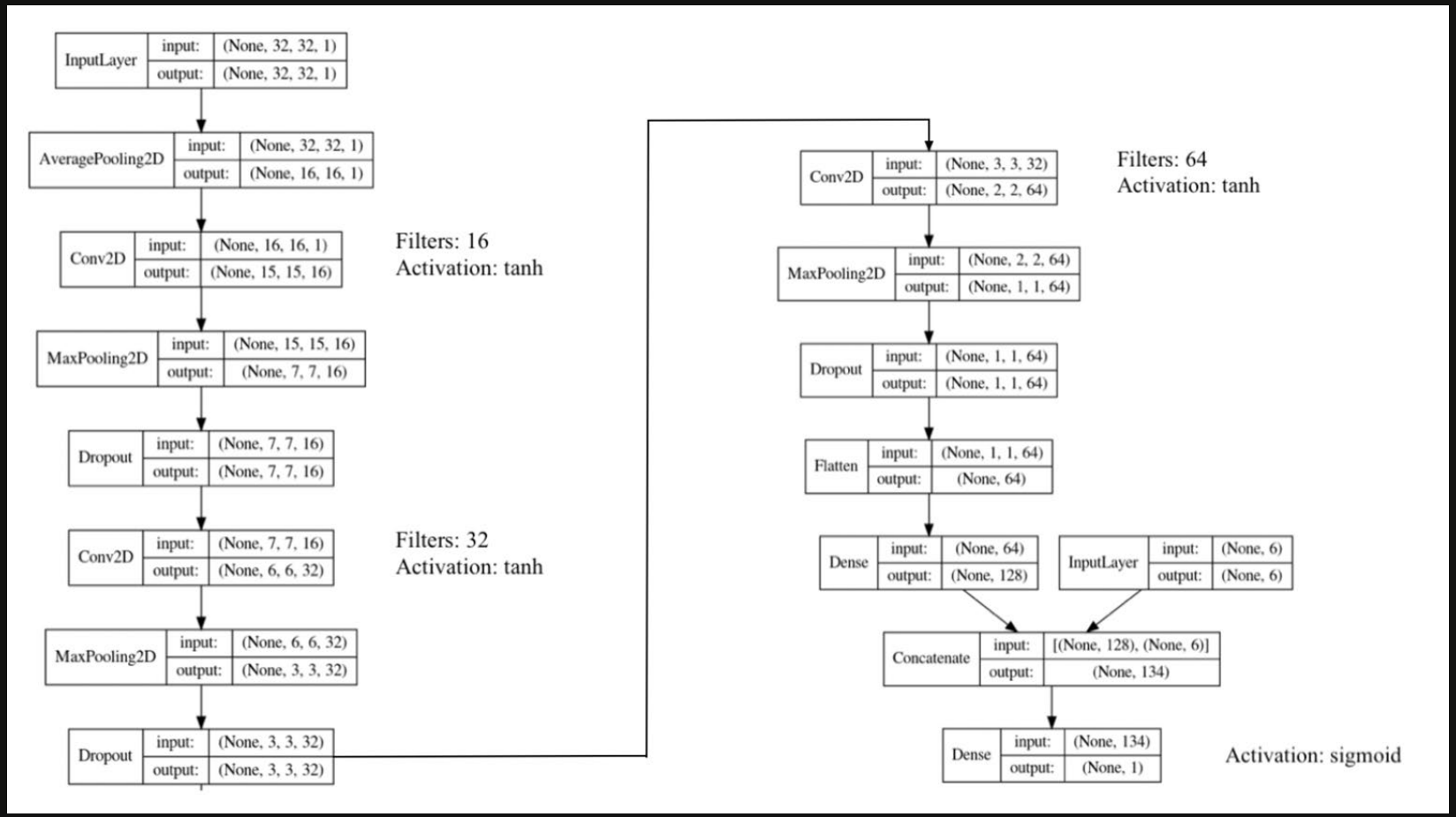
Weather Data Example - approx 19 km grid - July 21 2016; Bilinear Upsampling

Powered by Dark Sky

HOTzone Model: CNN

- Why a CNN
  - Identify hierarchical patterns in data
  - Assemble more complex patterns using smaller and simpler patterns
  - Learn dependencies between different pixels
  - Learn whether a pixel will be on fire based on whether pixels around it are on fire
- Data Used
  - 32x32 matrix of pixels around a pixel of interest
  - Vector of normalized weather data
  - Downsampled “no fire” to 80/20 “no fire”/”fire” split

HOTzone Architecture



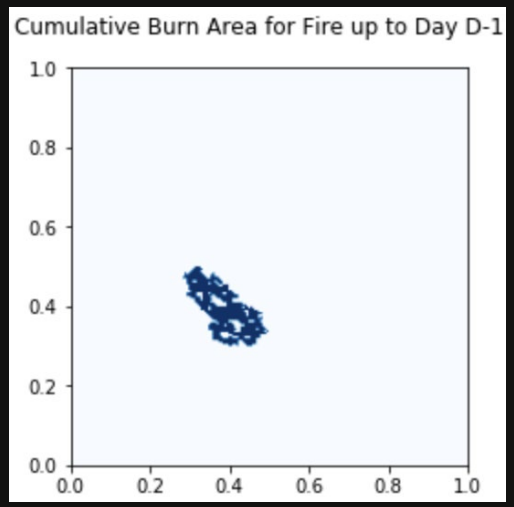
HOTzone Performance

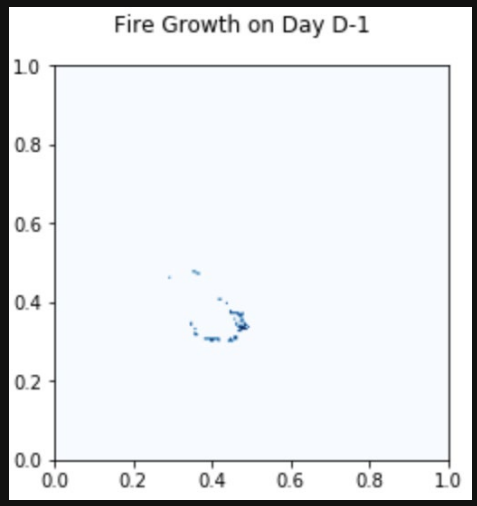
Model	Accuracy	Recall	F1 Score	AUC
Previous Work				
Farsite *	0.678 (Wet Fuel) 0.636 (Dry Fuel)	0.748 (Wet Fuel) 0.811 (Dry Fuel)	0.965 (Wet Fuel) 0.971 (Dry Fuel)	-
FireCast **	0.877	0.911	-	-
HOTzone				
Random Classifier	0.680	0.202	0.203	0.502
Naive Bayes	0.936	0.680	0.809	0.840
Logistic Regression	0.953	0.766	0.867	0.882
CNN	0.996	0.984	0.989	0.996

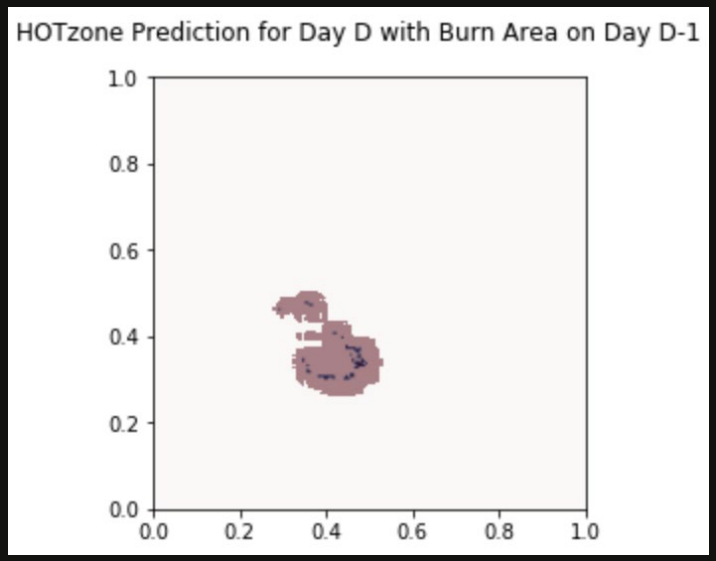
* Farsite is a physics-based fire growth model used by the US Forest Service and National Park Service

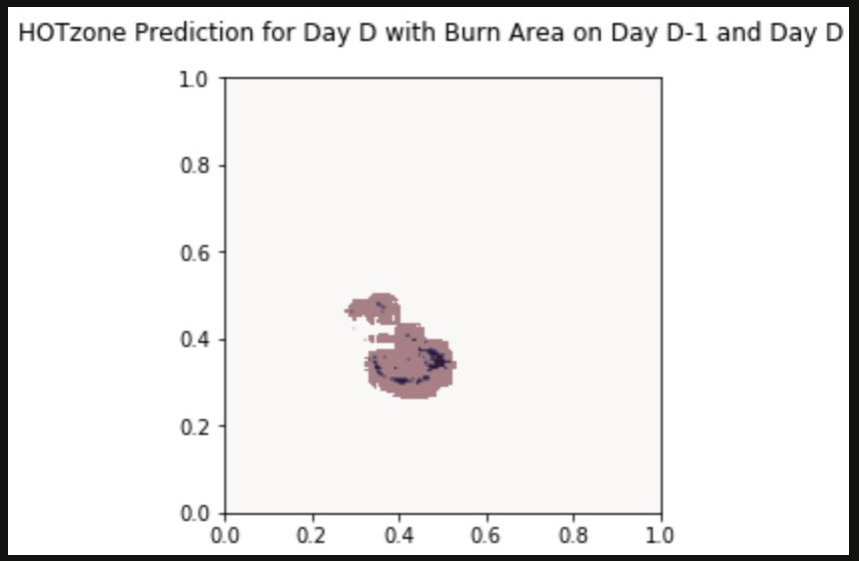
** Wildfire prediction CNN created by a joint team from the University of Waterloo and Colorado College

HOTzone Performance









MVP - Where to go Next?

- Data and Feature Engineering
  - Vast variation in Volume/Type of data
  - Improving speed & ease of reprojecting coordinate systems
  - Fire data resolution: 500m x 500m pixels = 60 acres each
  - Weather data resolution: $ vs hyper-locality
  - Other types of data: Vegetation, Fuel Type, Terrain, Altitude

- Characteristics of Data
  - Fires may not burn for a long enough - insufficient measurements
  - Extremely high pixel volume (630k pixels in each image)
  - Data is highly imbalanced - 1-2% of samples are “fire”

- Modeling
  - Expand to all of California/all of United States - Tiling/Regional
  - Opportunities for CNN refinement
  - Expand to different kinds of neural nets (LSTMs) that can remember more of a fire’s history

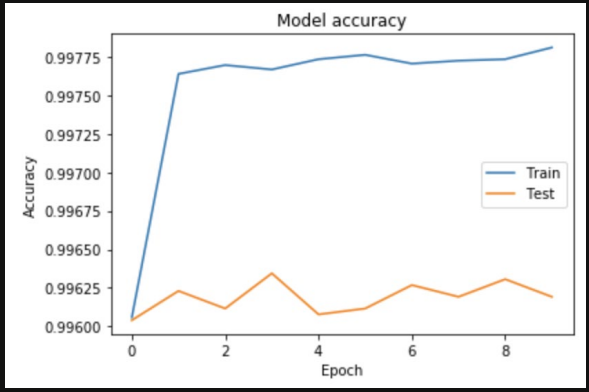
- Web Application
  - Productionize and speed up prediction.

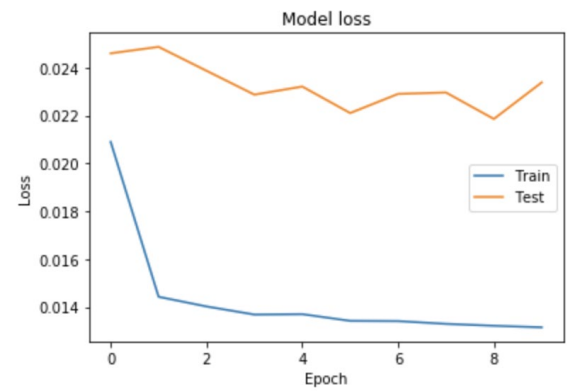
Thank you.

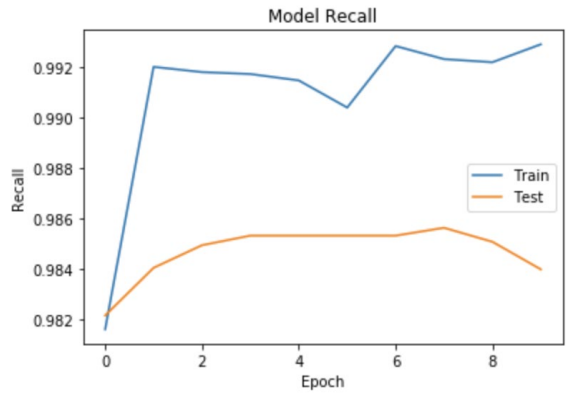
Questions?

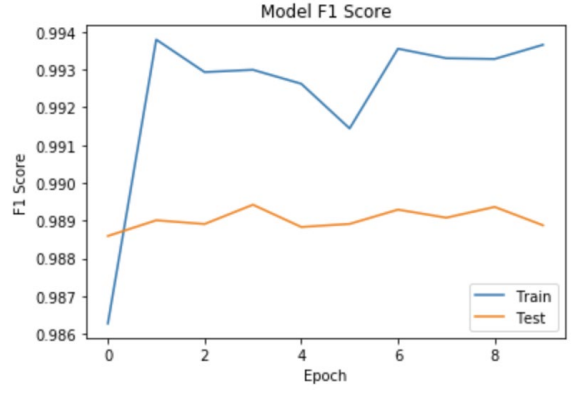
Appendix

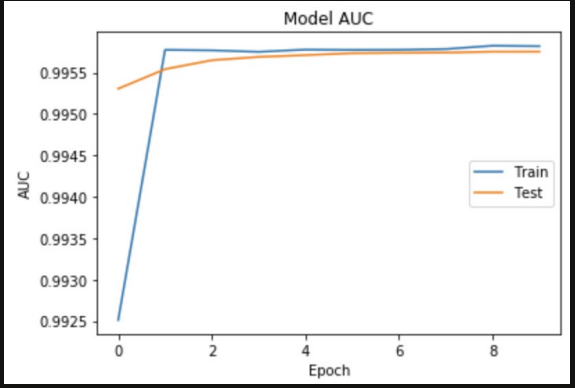
HOTzone Performance by Epoch









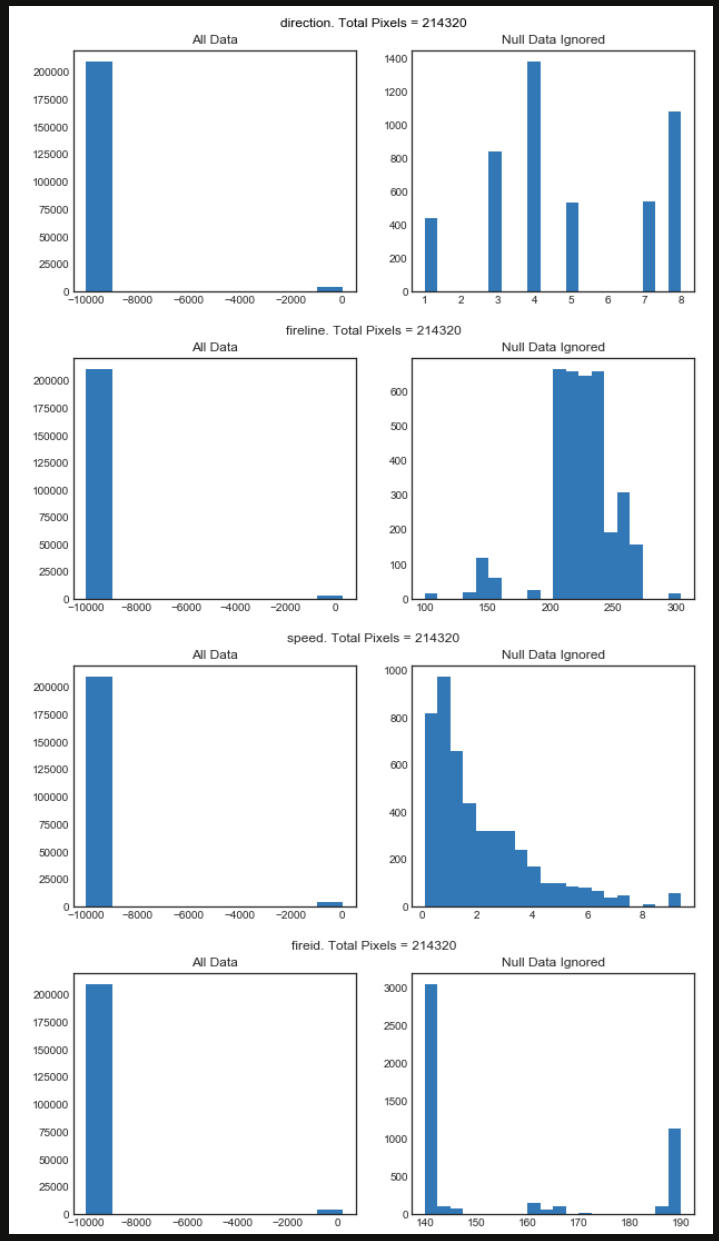


Fire Data EDA

FireData
Distributions
(for sample area)

Daily Fire Data

- shows ‘daily’ fire progression
- Shapefiles give final fire data
- Rasterized GeoTIFFs with ~500m x 500 m pixels give for each pixel:
  - Day of year on fire
  - Speed of fire
  - Direction of fire movement
  - Day of burn



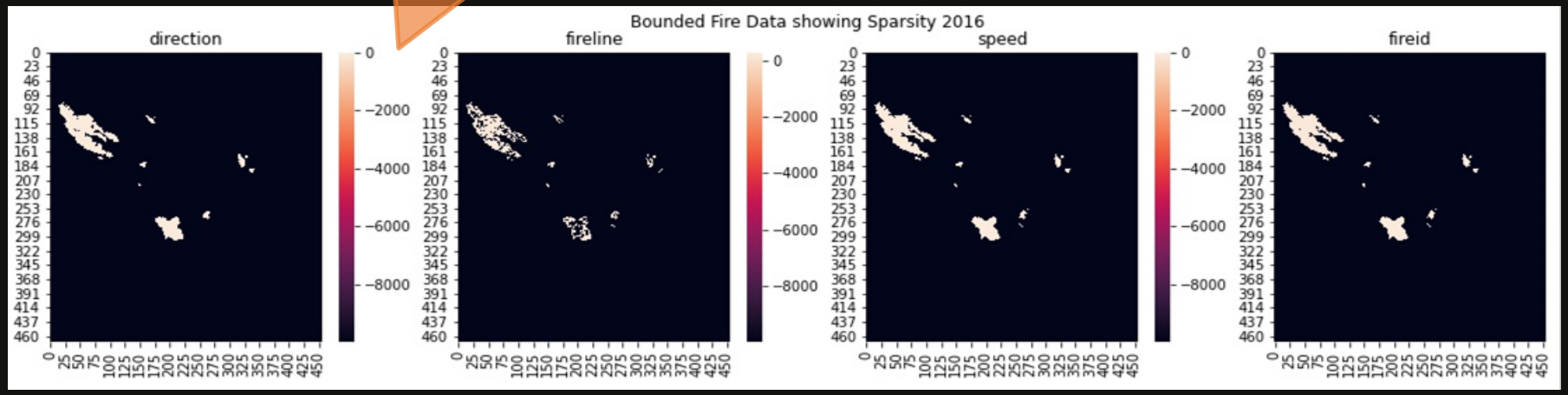
Sparse Fire Data

Fire data for 2016[1]

- localized to California/Nevada and then further to sub-section of CA

- sparse coverage
For Toy Data: Coverage ~ 1.65% of Total area

For Comparison: in 2008, the year with most burned ha in our study, 1.8% of the State's surface area was burned.



1 https://daac.ornl.gov/CMS/guides/CMS_Global_Fire_Atlas.html

Thanh's contributions ...

- Research
  - Map (google and mapbox) related research
  - Map coding for product prototype
  - Satellite image researching
  - NDVI - Python codes for NDVI calculation from satellite image
  - EDA on data source

- Data Engineering
  - Model Development (Naive Bayes model)
  - API connection to active fire database
    - Gather geolocation data for input to model, output for mapping, references for checking active fires and boundary
  - Mapbox coding
    - Reverse & forward geocoding
    - Mapping visualization of predictions

- Capstone website & web app development
  - HTML / Python / Javascript coding for website and app
  - Mapbox implementation & integration with ML model
  - Flask implementation / development

Laura's contributions ...

- Data Sourcing
  - GIS - Fire Historical Data
  - Weather - Historical Hyperlocal data
  - EDA

- Data Engineering
  - GIS - Model Inputs - (GeoTiffs and Shapefiles); changing geo references; cropping, reprojecting, resampling
  - Casting weather data into geo format with same coordinate reference system
  - Data pipelines
  - Implementing Flask App with WSGI (Gunicorn) and Nginx on AWS EC2 Ubuntu Server, built custom Geostack

- General:
  - Code review
  - Website/app updates and troubleshooting
  - Domain and Application Resource Management and Maintenance

Keane's contributions ...

- Data Engineering
  - Turn fire data tiffs into matrices that can be fed to a CNN
  - Relate weather data to individual pixels
- Model Research and Development
  - Identify and understand previous models
  - Develop a CNN in Keras
- Data Viz
  - Turning predictions into plotable lat/long coords
- AWS Implementation