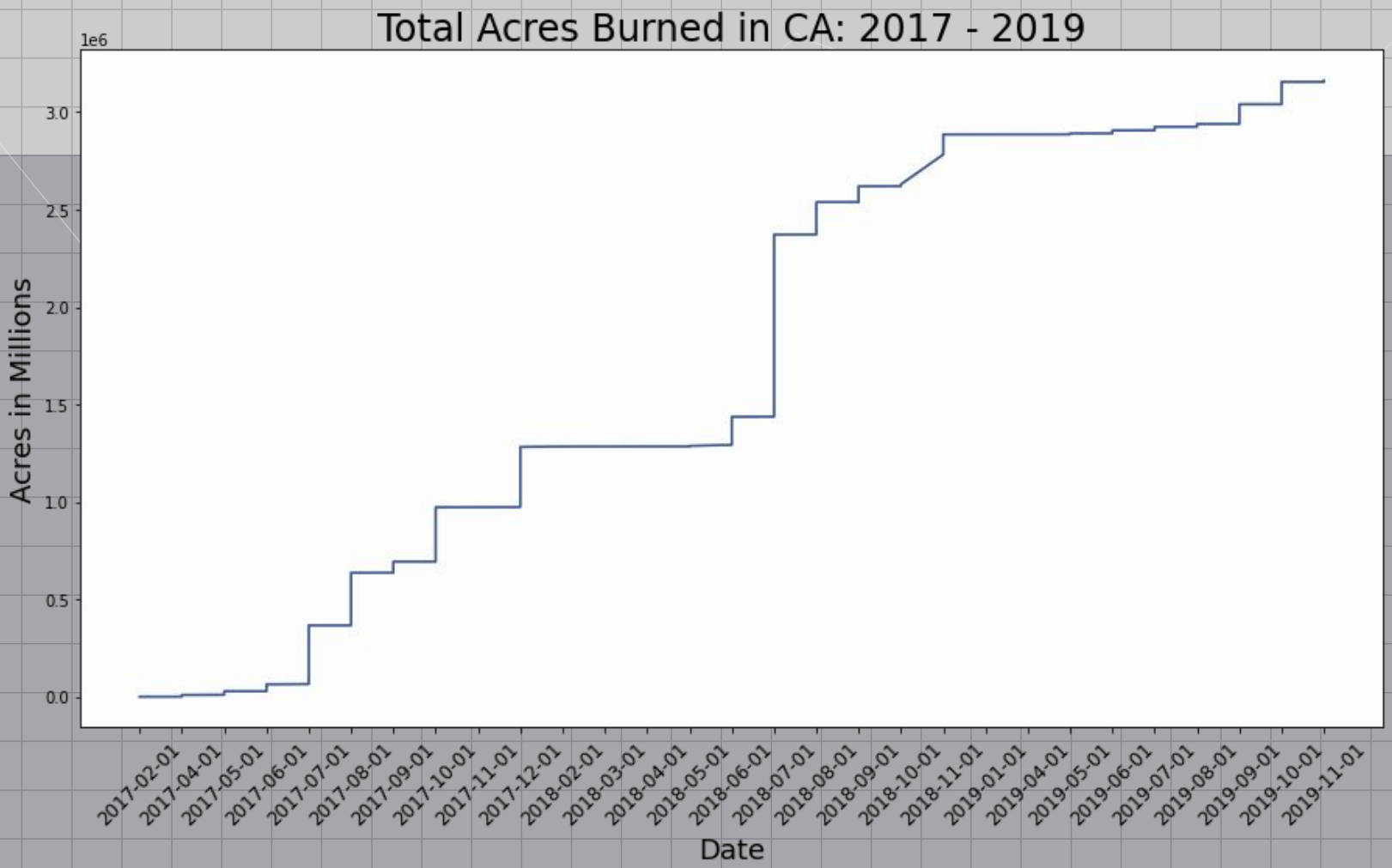
# California Wildfires

Carly Sharma, Jake Parker, Adam Zucker

# Problem Statement

We've set out to predict the potential severity of future fires based on past fire data, weather patterns and atmospheric pressure within California.


Total Acres Burned in CA: 2017 - 2019
1e6
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Acres in Millions
2017-02-01
2017-04-01
2017-05-01
2017-06-01
2017-07-01
2017-08-01
2017-09-01
2017-10-01
2017-11-01
2017-12-01
2018-02-01
2018-03-01
2018-04-01
2018-05-01
2018-06-01
2018-07-01
2018-08-01
2018-09-01
2018-10-01
2018-11-01
2019-01-01
2019-04-01
2019-05-01
2019-06-01
2019-07-01
2019-08-01
2019-09-01
2019-10-01
2019-11-01
Date

# OUR PROCESS

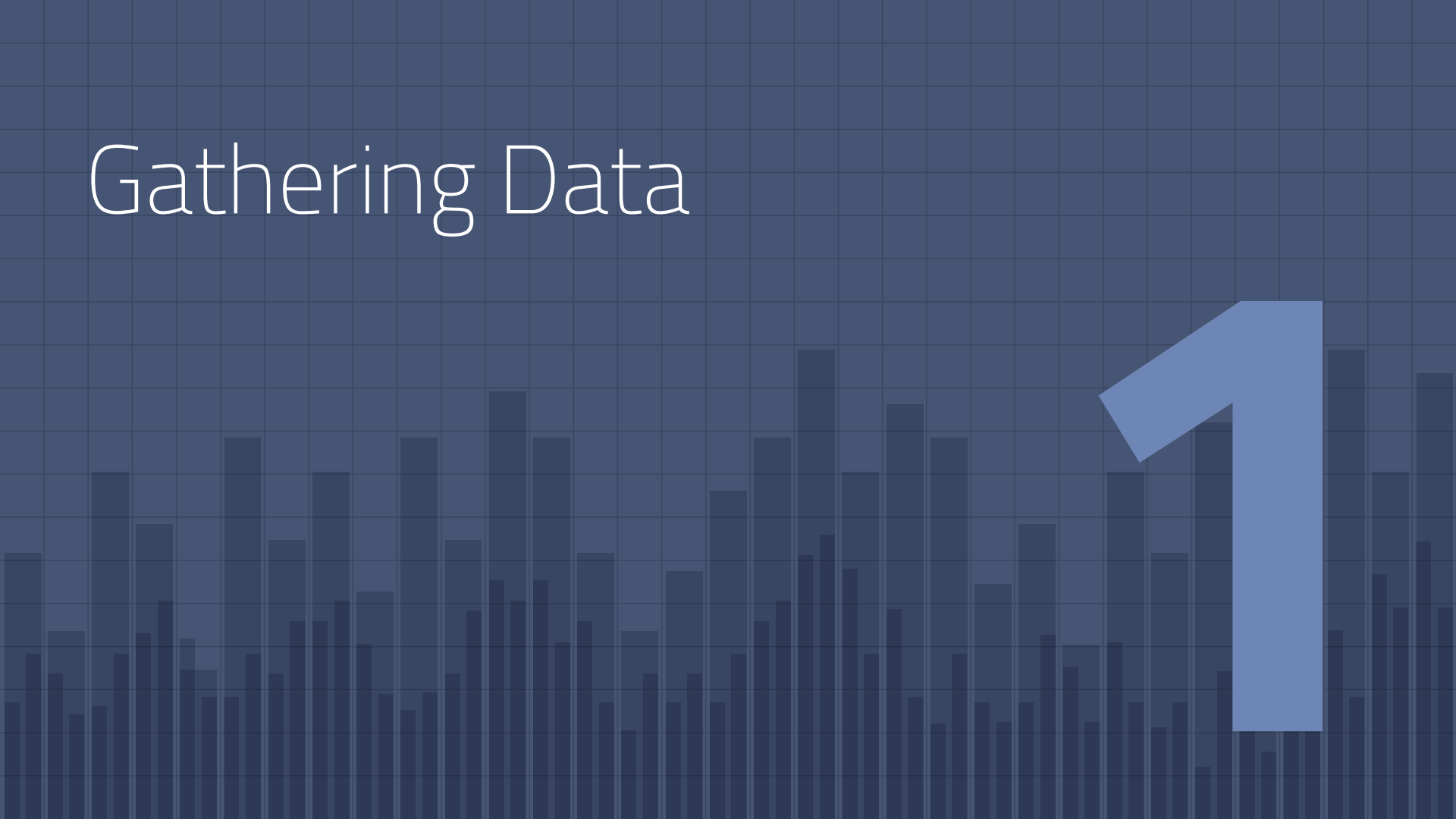


## STEP 1
Gathering Data

## STEP 2
Analysing Data

## STEP 3
Creating a Model

# Gathering Data



# DATA GATHERING: 3 TYPES

MASTER DATAFRAME

CA WildFire Data

NOAA Weather Data
(12 unique CSVs)

Atmospheric
Carbon Data

# GENERAL DATA CLEANING

- Drop high percentage of nulls
- Drop categories that were irrelevant(ex. all fires 100% contained)
- Average weather station data by county
- Rounded fires to counties and month

# MERGE 15 DATAFRAMES: Weather Stations

| STATION | NAME | LAT | LONG |
|---|---|---|---|
| USW00023190 | SANTA BARBARA MUNICIPAL AIRPORT, CA US | 34.4258 | -119.8425 |
| USW00023232 | SACRAMENTO EXECUTIVE AIRPORT, CA US | 38.5069 | -121.495 |


| COUNTY NAME |
|---|
| Santa Barbara |
| Sacramento |

Image from https://geopy.readthedocs.io/en/stable/

# Analysing Data

2

# FEATURE ENGINEERING

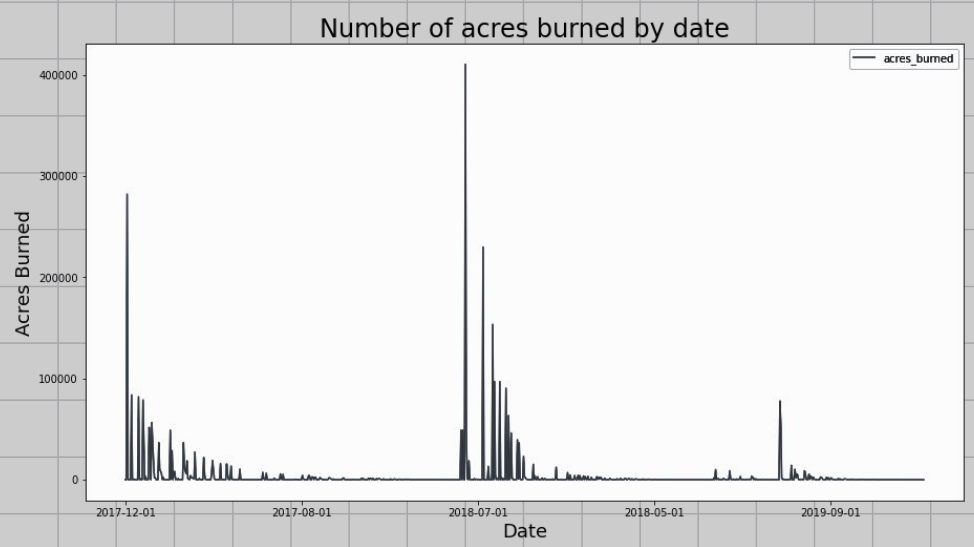
COLUMNS CREATED:

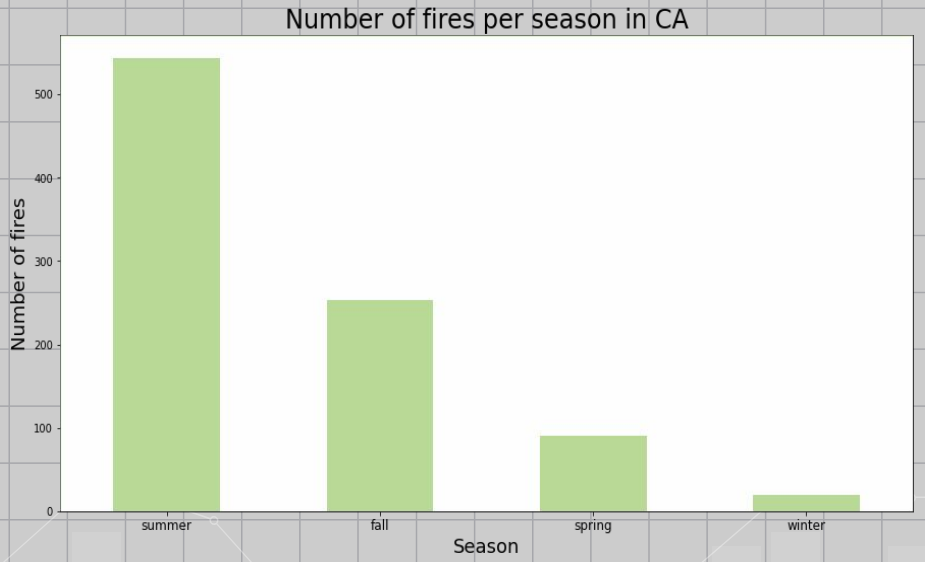
- Month
- Season
- Duration (of fire)
- Bins (of acres burned)

# FEATURE ENGINEERING: Bins

- **Bin 1:** 50 acres or below
- **Bin 2:** 100 acres or below, and greater than 50 acres
- **Bin 3:** 250 acres or below, and greater than 100 acres
- **Bin 4:** 500 acres or below, and greater than 250 acres
- **Bin 5:** 1000 acres or below, and greater than 500 acres
- **Bin 6:** Greater than 1000 acres

# FEATURE ENGINEERING: Seasons



# FEATURE ENGINEERING: Duration

- Duration = Fire Extinguished - Fire Started
- Duration ratio = Acres Burned / Duration

# CORRELATION FINDINGS: Weather Data

Avoid Multicollinearity:

Drop weather data that is over 80% correlated to one another

# Modeling

# TARGET & BASELINE ACCURACY

- Multiclass Classification Problem
- Our target is to classify fires into bins representing severity
  - An engineered feature based on acres burned
  - Fires are binned smallest to largest, 1 to 6
- Baseline Accuracy: 39%
  - Bin 1: 50 acres or below

# MODELS

**Logistic Regression**

- A simple model to predict categorical outcomes
- Similar to a linear model in structure

**KNN**

- Supervised Machine Learning
- Assumes if points are near each other they are related

# MODELS

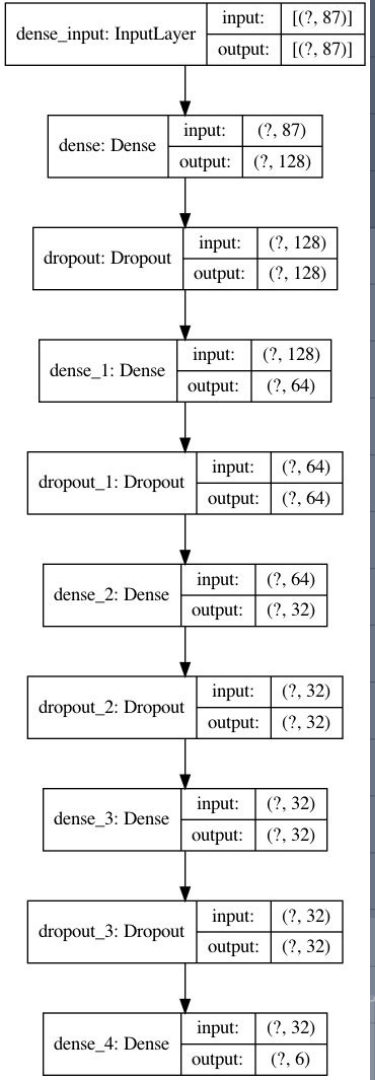
**Random Forest**

- An industry standard model for classification
- Improves decoupling of our decision trees
- Every tree in the forest gets a “vote”

**SVC**

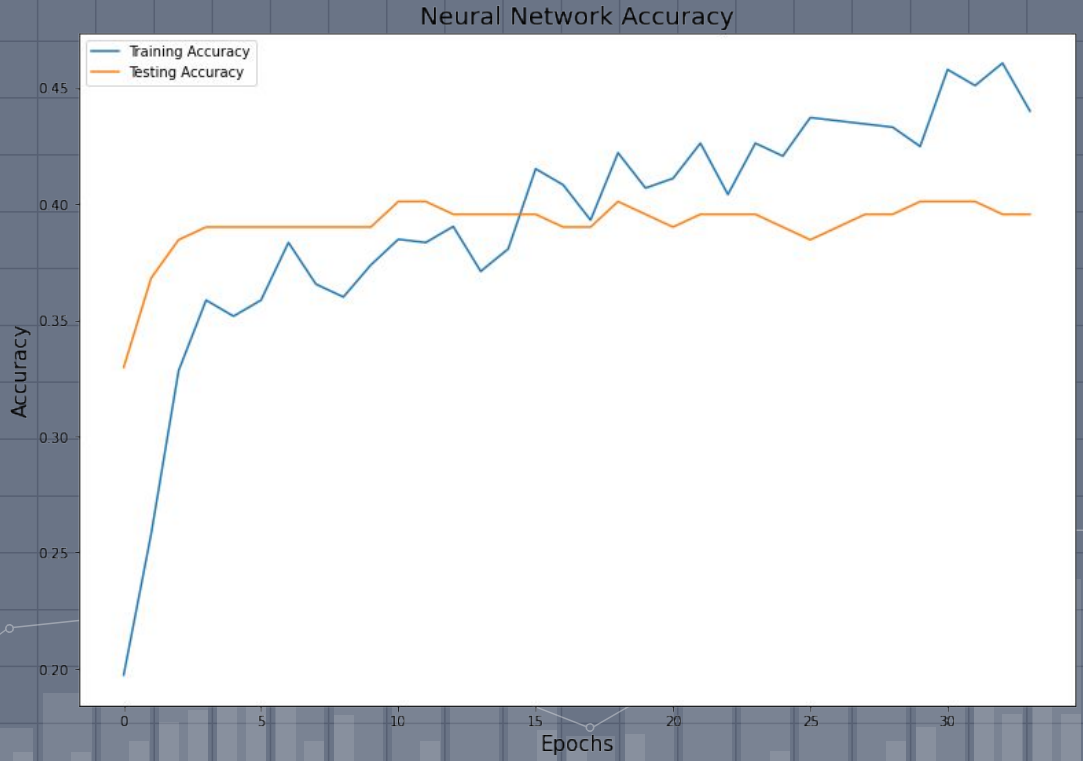
- Kernel Support Vector Classifier
- Our model displays moderate accuracy, but lacks potentially crucial interpretability

# NEURAL NETWORK




**Left:** Network structure. For fun!

This model gives us the probability of a fire falling into one of the 6 given bins

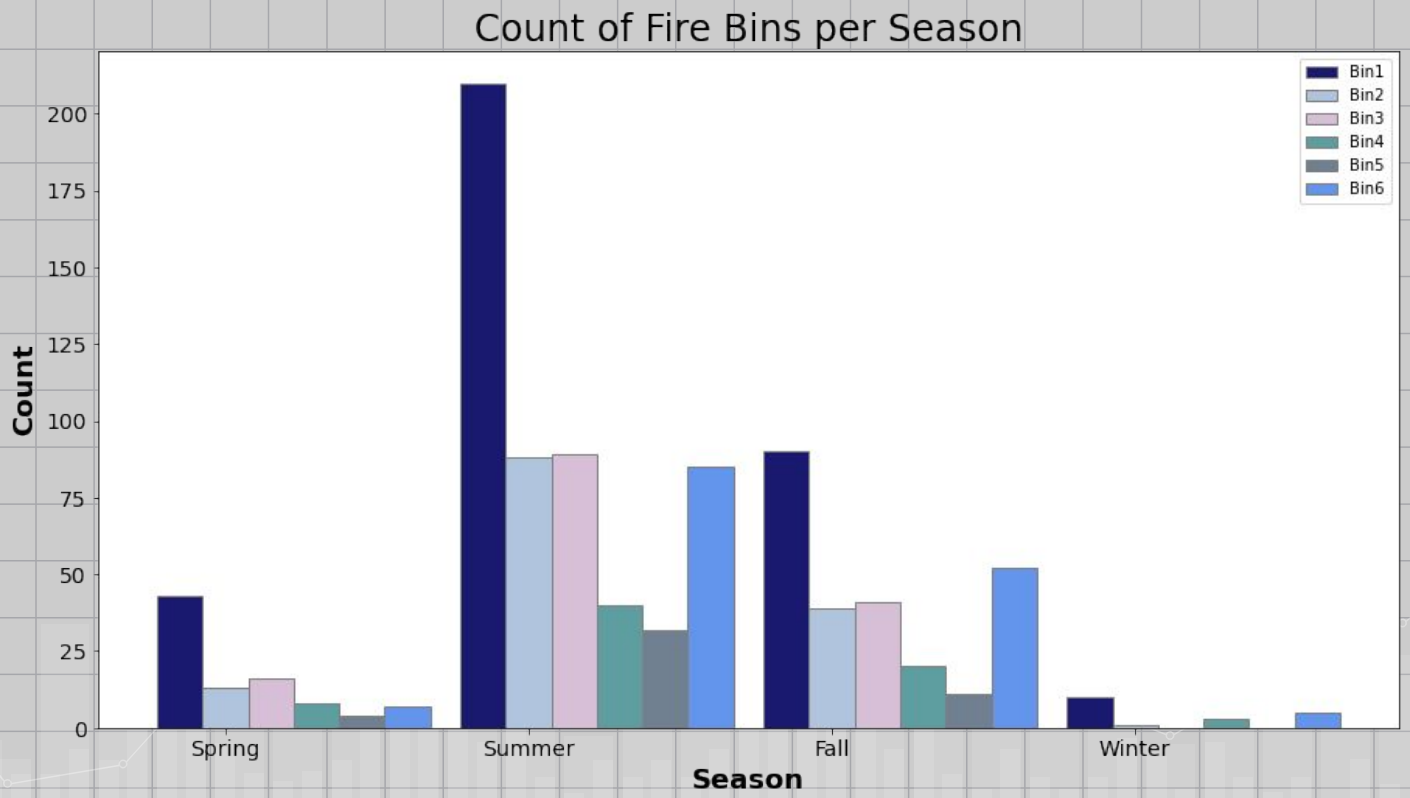



**Above:** Ultimately, our NN topped out at 42% accuracy, with significant loss

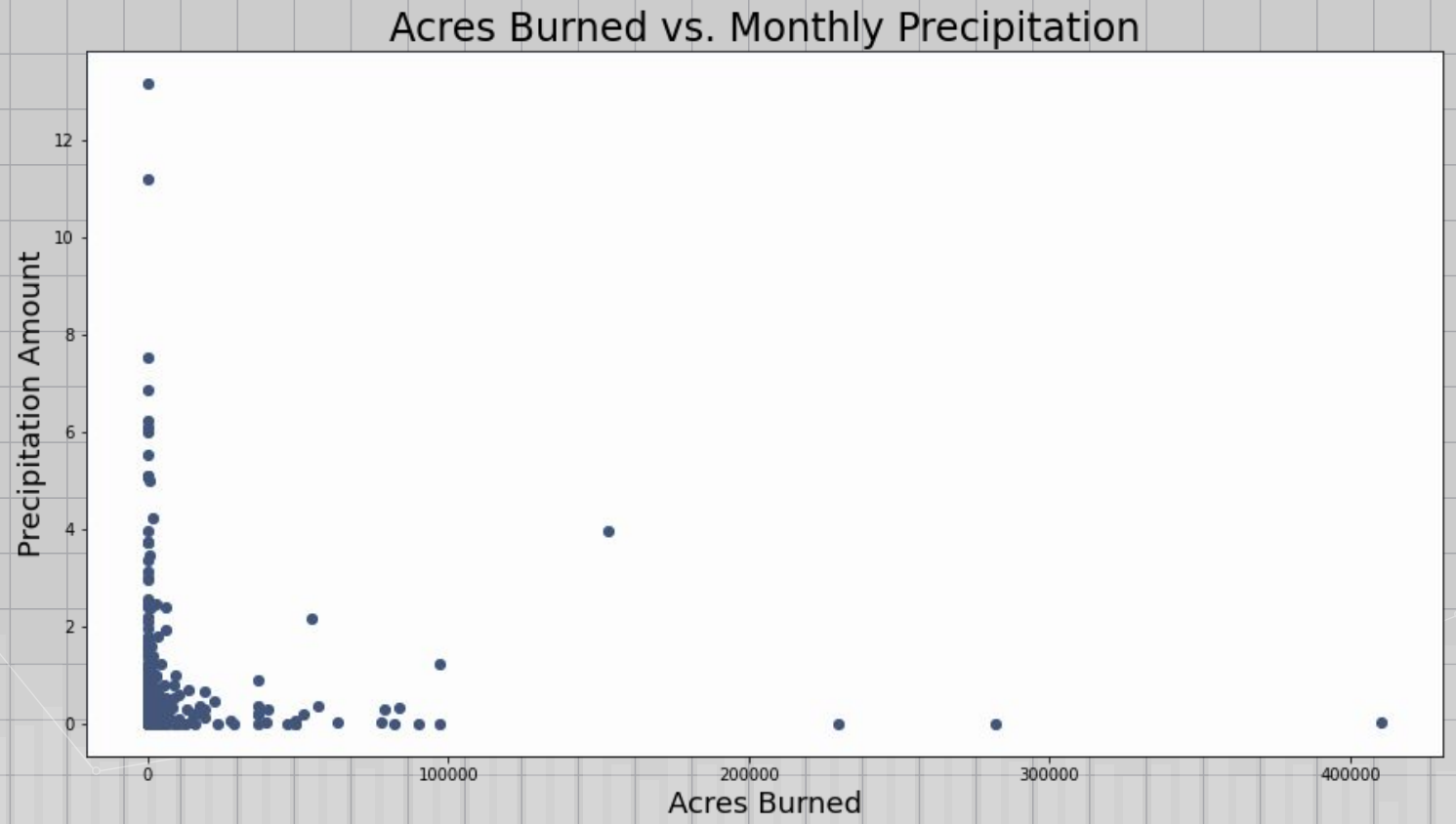
# Results

| | Training Accuracy | Testing Accuracy |
|---|---|---|
| Logistic Regression | 55.58% | 42.85% |
| Support Vector | 48.97% | 51.65% |
| K Nearest Neighbor | 43.72% | 38.46% |
| Random Forest | 99.86% | 100% |

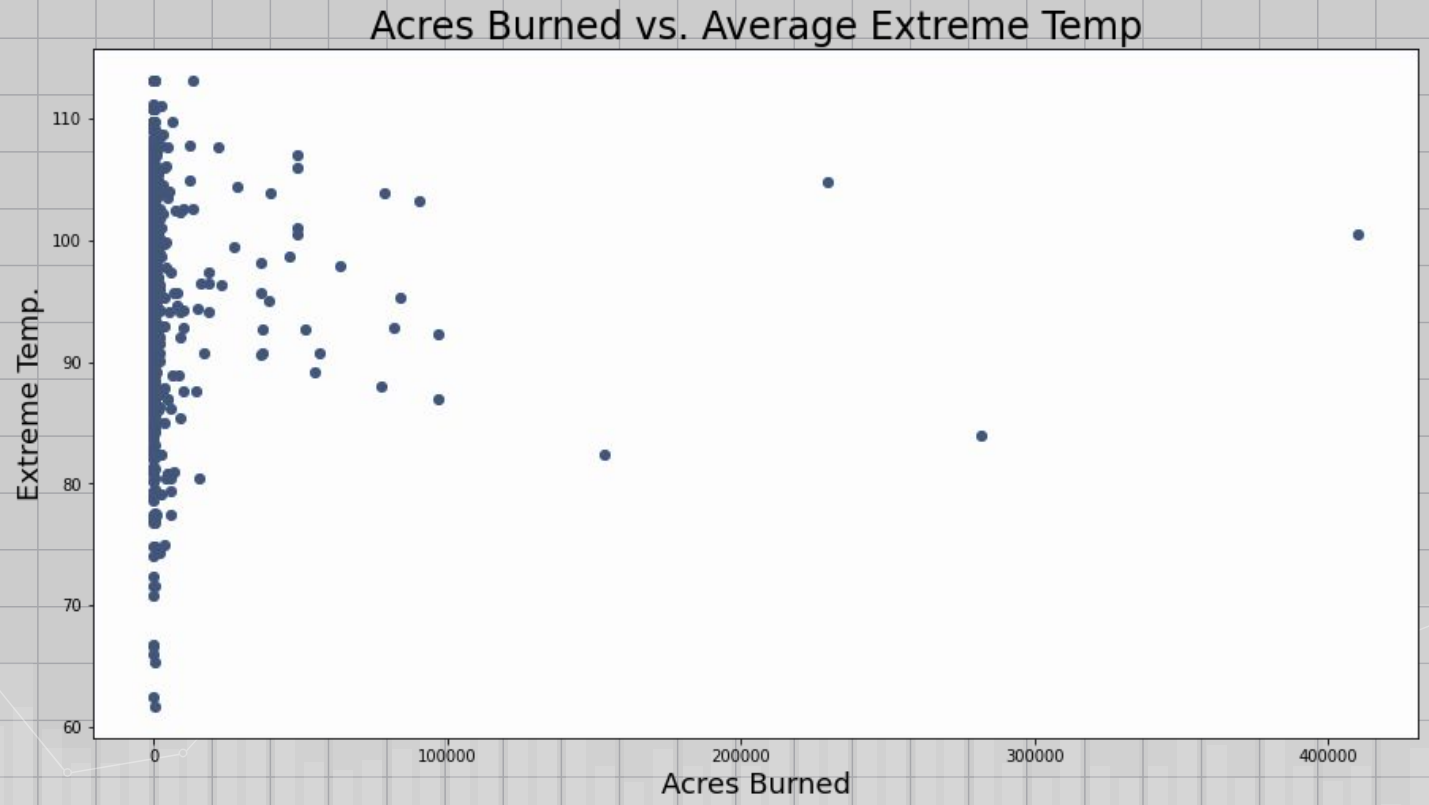
# RESULTS: Seasonality and Severity of Fires



# RESULTS: Acres Burned to Average Monthly Prec.

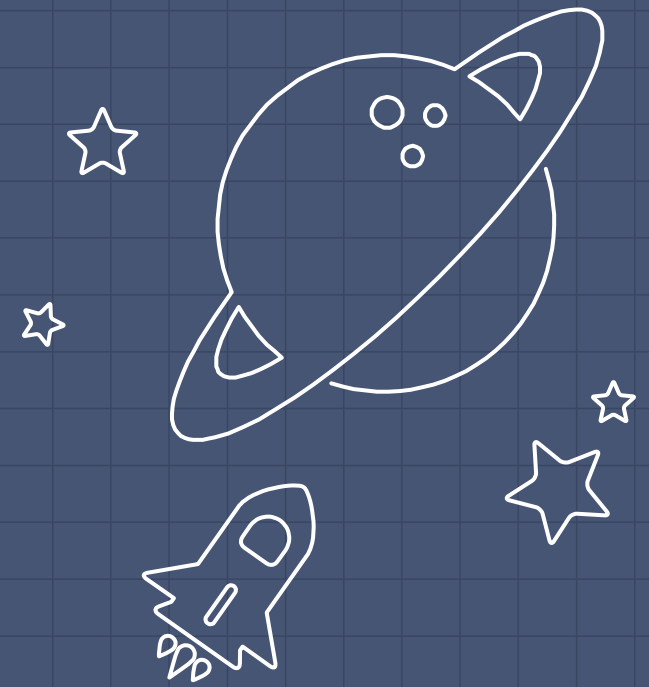


# RESULTS: Acres Burned to Average Extreme Temp



# RESULTS: Acres Burned to Average Elevation

# Recommendations & Conclusions



# RECOMMENDATIONS: Hire extra crews for

- Counties in higher elevations
- If there has been low rain amounts for the season
- If there has been an extreme monthly temp. above 90 deg.
- Summer/Fall season (specifically July you can expect a lot of small fires to occur)

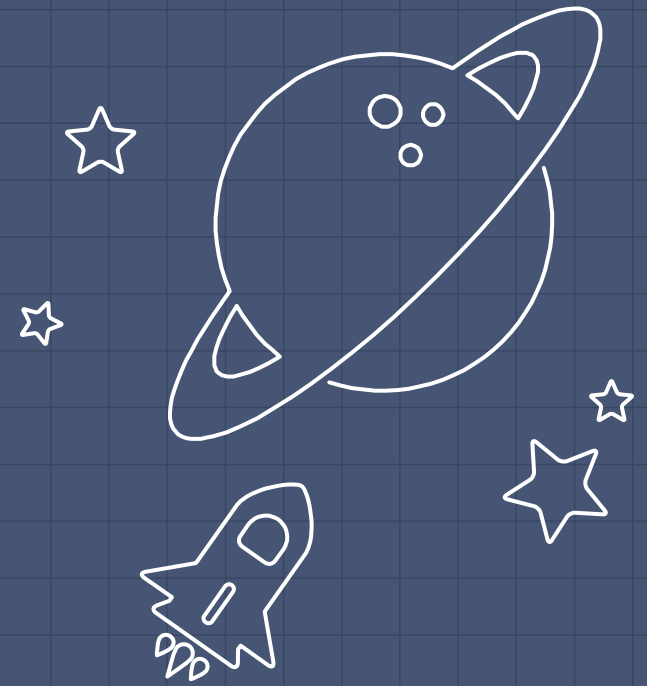
# RECOMMENDATIONS: Helping at risk areas

- Controlled burns in at risk areas
- Clear fuel for wildfires
- Reallocation of resources and firefighting units based on environmental and weather conditions

# FURTHER EXPLORATION:

- Look at wind speeds, look at daily numbers instead of monthly
- Compare rural vs. urban areas
- Look at a longer time period
- Research climate change in more depth

# Questions?



# SOURCES

- Kaggle California Wildfire Incidents Dataset
- NOAA Meteorological Datasets
- NOAA/Mauna Loa Observatory Atmospheric Carbon Dataset
- California Department of Forestry and Fire Protection