Project proposal

Fantastic FouR: Hannah, Eli, Preetha

2020-10-19

```
library(tidyverse)
library(broom)
library(patchwork)
library(knitr)
```

```
fire <- readr::read_csv('data/all_conditions.csv')
```

Section 1. Introduction

For our project, we will be looking at which factors are most dominant in predicting forest fires in the state of California. Climate change represents an existential threat to humanity, and we are already seeing its effects, particularly on America's west coast. We want to understand the specific factors that contribute to fires to more fully understand what can be done to prevent these tragedies from spreading. The research question we wish to look at is: What are the strongest environmental predictors in forest fires? We think that precipitation, humidity, Eto, and air temperature will all be strongly correlated with forest fires.

Section 2. Data description

There are 128,125 observations in the data set. Each observation represents information on the weather conditions on a given weather station on a specific date. Each observation looks at a variety of factors such as Eto, precipitation, solar radiation, average vapor, maximum air temperature, minimum air temperature, average air temperature, maximum humidity, minimum humidity, average humidity, dew points, wind speed, soil temperature. The response variable we will be investigating is Target, which corresponds to fires on the respective observation date, in the observation region. The Target variable is a binary indicator, with a value of 1 indicating there was a fire and a value of 0 indicating there was not a fire. Each of the predictors represents a different aspect of the climate in California. Eto looks at the presence of a natural flammable gas, precipitation and solar radiation look at the amount of rain, snow, or sunshine an area will receive on the given day. Average vapor, maximum, minimum and average temperature and humidity, and wind speed represents qualities of the air. Dew points and soil temperature look at factors that impact the ground. This data was scraped from CIMIS (California Irrigation Management Information System) weather stations by github user czaloumi using a selenium chromedriver. The dataset was combined with Wikipedia tables listing California fires by county and city to create the Target column, which indicates whether or not there was a fire on a particular day. Additionally, the curator of the dataset adds that this dataset was "used in conjunction to building an XGBoost Classifier to accurately predict probability for fire given environmental condition feature."

Section 3. Glimpse of data

```
glimpse(fire)
```

```
## Rows: 128,125
## Columns: 19
## $ `Stn Id`               <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,...
```

```
## $ `Stn Name`             <chr> "FivePoints", "FivePoints", "FivePoints", "F...
## $ `CIMIS Region`         <chr> "San Joaquin Valley", "San Joaquin Valley", ...
## $ Date                   <chr> "1/1/2018", "1/2/2018", "1/3/2018", "1/4/201...
## $ `ETo (in)`             <dbl> 0.06, 0.04, 0.04, 0.07, 0.07, 0.02, 0.06, 0....
## $ `Precip (in)`          <dbl> 0.00, 0.00, 0.00, 0.01, 0.00, 0.04, 0.00, 0....
## $ `Sol Rad (Ly/day)`     <dbl> 219, 127, 125, 219, 239, 101, 202, 65, 259, ...
## $ `Avg Vap Pres (mBars)` <dbl> 7.3, 7.4, 8.4, 11.6, 12.7, 13.4, 11.6, 13.8,...
## $ `Max Air Temp (F)`     <dbl> 63.4, 59.8, 61.1, 69.2, 73.8, 60.2, 65.7, 60...
## $ `Min Air Temp (F)`     <dbl> 35.3, 37.7, 37.3, 48.7, 47.5, 47.6, 50.9, 51...
## $ `Avg Air Temp (F)`     <dbl> 47.8, 47.2, 49.9, 56.8, 59.8, 55.2, 56.4, 55...
## $ `Max Rel Hum (%)`      <dbl> 82, 80, 79, 94, 94, 96, 90, 97, 94, 96, 97, ...
## $ `Min Rel Hum (%)`      <dbl> 46, 52, 49, 52, 49, 80, 58, 79, 60, 64, 72, ...
## $ `Avg Rel Hum (%)`      <dbl> 65, 67, 68, 74, 72, 90, 74, 92, 75, 80, 89, ...
## $ `Dew Point (F)`        <dbl> 36.6, 36.7, 39.9, 48.5, 50.8, 52.3, 48.4, 53...
## $ `Avg Wind Speed (mph)` <dbl> 3.3, 3.1, 4.5, 5.8, 4.2, 6.7, 3.5, 5.1, 5.7,...
## $ `Wind Run (miles)`     <dbl> 78.3, 74.5, 107.5, 140.2, 101.4, 162.0, 83.3...
## $ `Avg Soil Temp (F)`    <dbl> 51.1, 51.3, 51.3, 53.0, 54.4, 55.4, 55.5, 55...
## $ Target                 <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
```