Wildfires in California

Chaimae M'SAAD

OVERVIEW

- Increased intensity due to human-induced climate change
- Drastic rise in environmental, economic, and social costs due to wildfires

4 MILLION+ ACRES

in 2020

Heat Map of CA WILDFIRES

Can we predict wildfire size based on weather elements 14 days prior and after the start of a wildfire?

Removed NA's

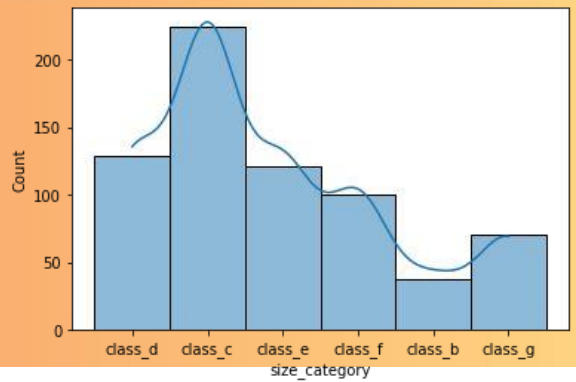
Dropped fires that contained missing data

Aggregated Data

Averaged hourly data over a 24 hr period for each feature

NAIVE BAYES

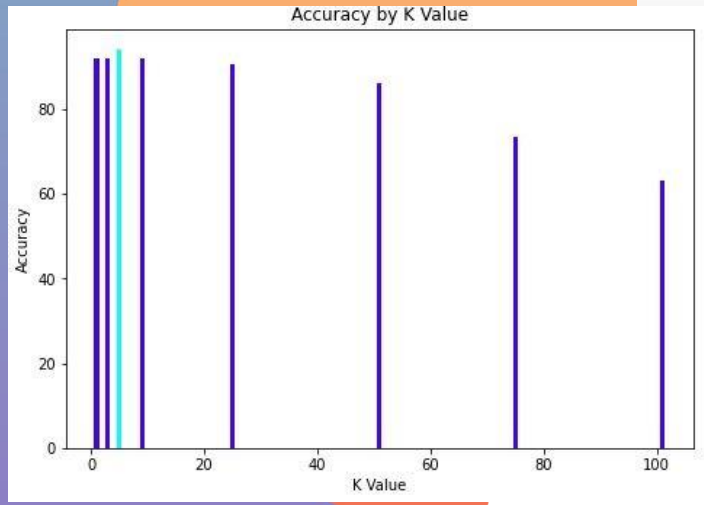
Models	F1 Score	Fits
GaussianNB	0.55	Normally Distributed Classes❌
CategoricalNB	0.94	Discrete Features✅



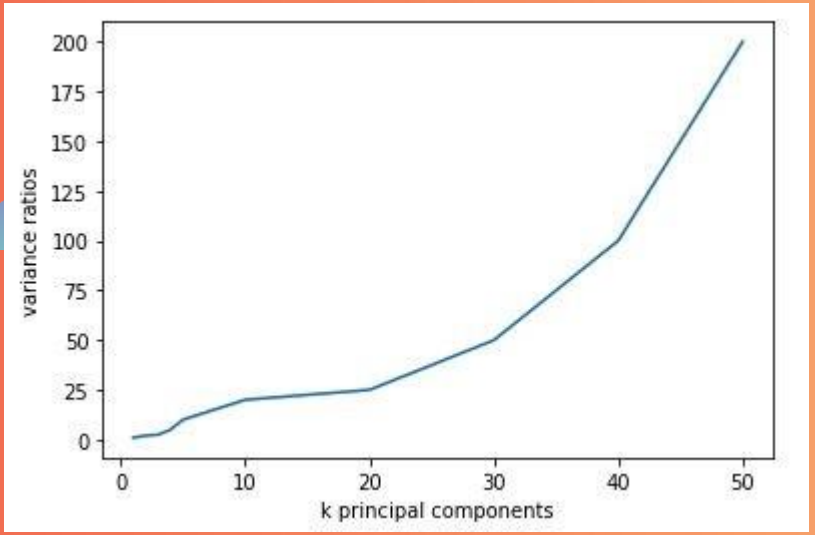
K NEAREST NEIGHBORS

- The K Nearest Neighbors model is non-linear and works well with classes
- The training data contains about 400 points, so it is understandable that K Values of 75 and 100 would produce relatively low accuracy scores, since they are classifying by using a large fraction of the data
- Best: 5 KNN with 94.11% accuracy

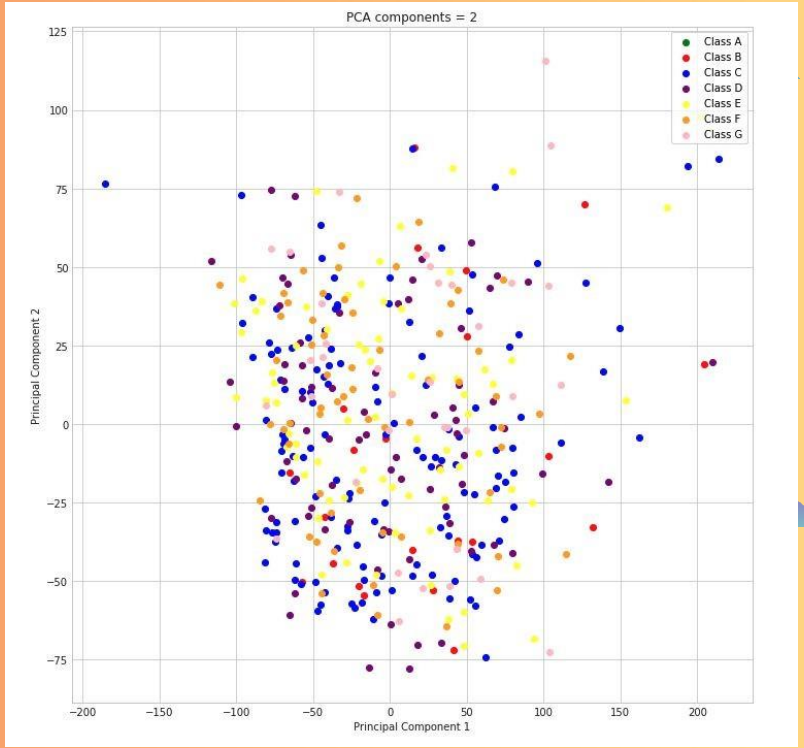
K Values	Scores
5	94.117647
1	91.911765
3	91.911765
9	91.911765
25	90.441176
51	86.029412
75	73.529412
101	63.235294



PCA



k components = [1, 2, 3, 4, 5, 10, 20, 30, 40, 50]

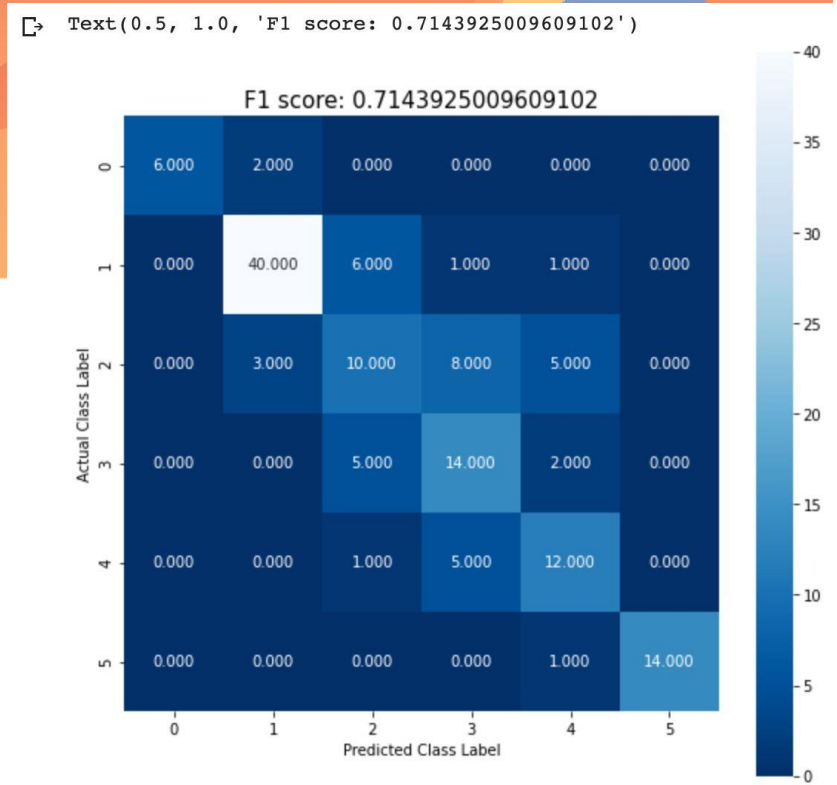


PCA with 2 principal components

LOGISTIC REGRESSION

- Now we'll use Multinomial Logistic Regression to model the probability of a certain fire size class.

- Our main findings were:
  - The optimal value for C is 7.0.
  - F1 score for a Logistic Regression classifier using the optimal value for C: 0.7144.



LINEAR REGRESSION

Use All Features

No manipulation was done on the feature space

Negative R^2

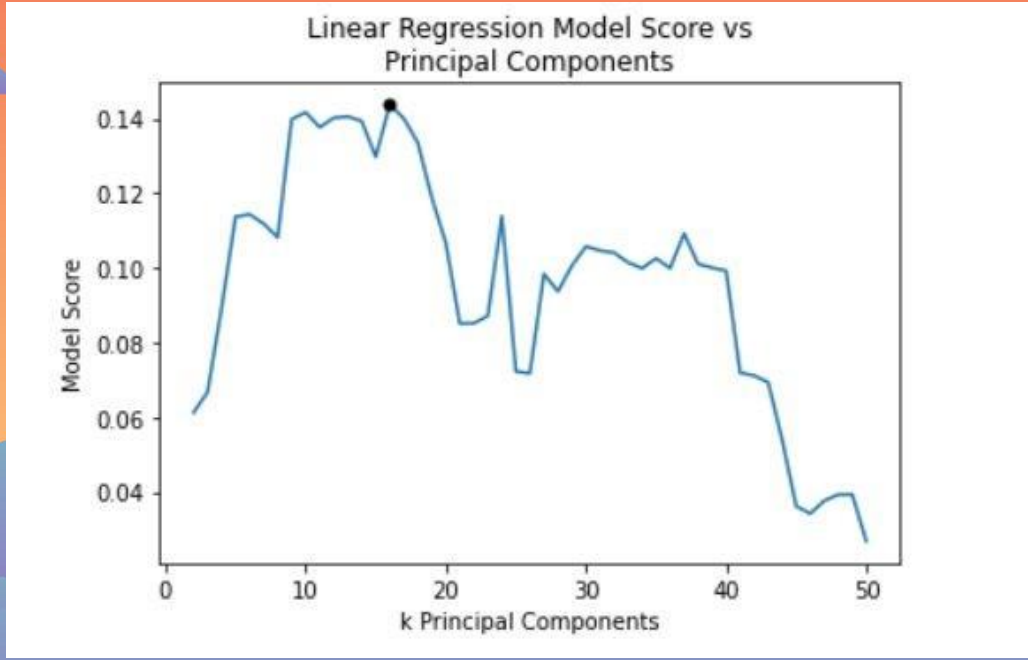
Can't capture variation in the dataset

	log_sizes	predicted
0	8.052615	3.815292
1	5.521461	5.127927
2	5.802118	6.329192
3	4.983607	3.508105
4	4.700480	0.045942
...	...	...
131	4.290459	5.277446
132	3.970292	5.545927
133	7.293018	2.325221
134	3.332205	4.941196
135	1.791759	9.273558

136 rows × 2 columns

First Model

LINEAR REGRESSION: PCA



LINEAR REGRESSION

Reduce Feature Space

Use PCA with 16 components

Improved R^2

Model captures about 14.38% of variation ($R^2 = 0.1438$)

	log_sizes	predicted
0	8.052615	4.591681
1	5.521461	5.165711
2	5.802118	8.600421
3	4.983607	3.921614
4	4.700480	4.268586
...	...	...
131	4.290459	4.482017
132	3.970292	4.442077
133	7.293018	3.823545
134	3.332205	5.873307
135	1.791759	4.584004

136 rows × 2 columns

First Model

BEST PERFORMING MODELS

Categorial Naive Bayes

Produces 94% accuracy

KNN

Nearest Neighbor Value of 5 produces 94.1% accuracy

What i Learnt :

- We are able to predict with fairly high accuracy wildfire size based on weather elements 14 days before and after the start of the fire
- This information can be highly useful in containing fires
- However, human actions are necessary to reduce the increasingly harmful effects of climate change