PREDICTING DROUGHTS & THEIR INTENSITIES IN CALIFORNIA

Under the guidance of Dr. Huthaifa Ashqar
DATA 601 – Intro to Data Science

University of Maryland, Baltimore County

Droughts

General Definition

Deficiency of precipitation

water shortage

Creeping Phenomenon

Slowly impacts economy

Operates on different timescales

No single definition

Difficult to define

150 published definitions in 1980s

Drought Classification

Based on both physical and socioeconomic factors:

METEOROLOGICAL DROUGHT

- precipitation deficit over a prolonged period of time
- When dry weather patterns dominate an area

AGRICULTURAL DROUGHT

- deficit in soil moisture
- affects plant production and crop yield

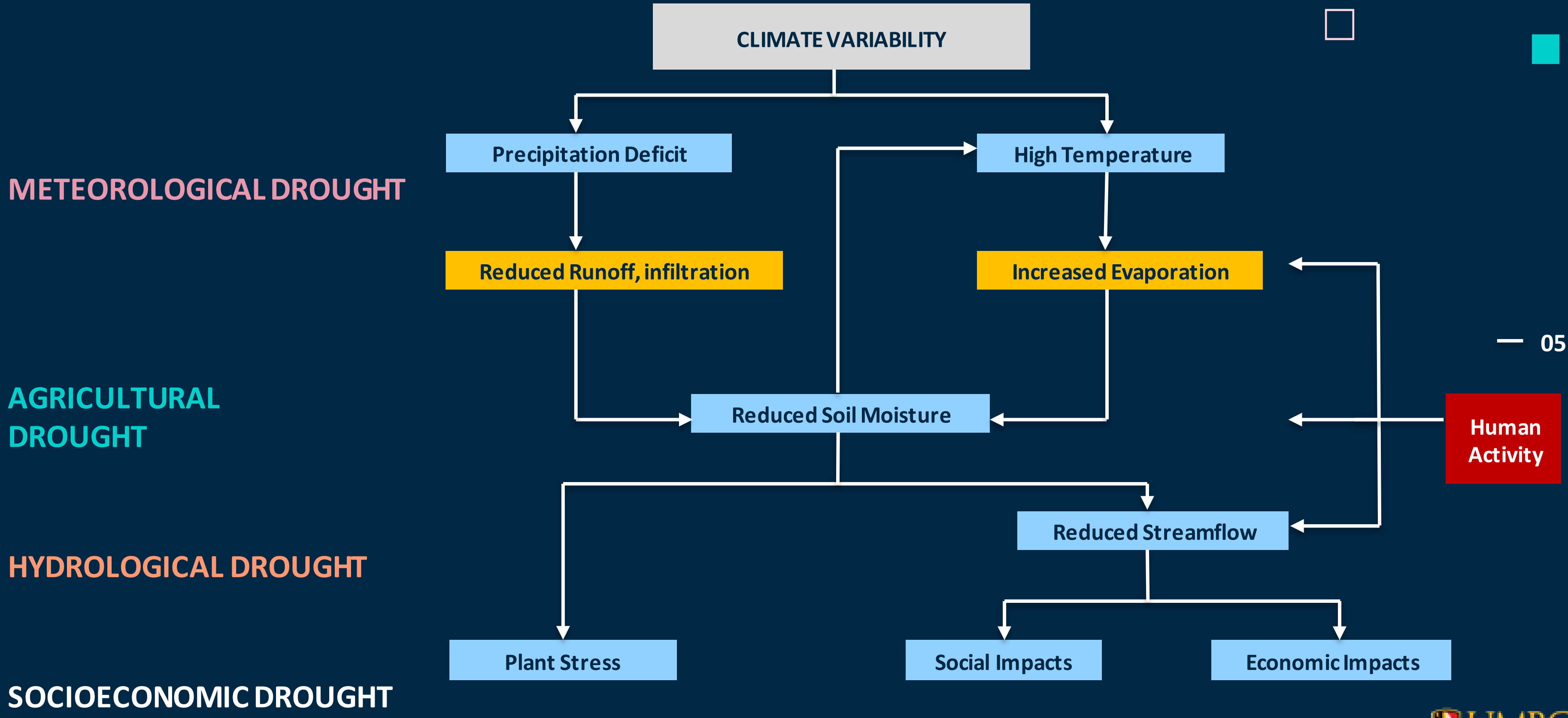
HYDROLOGICAL DROUGHT

- deficit of surface runoff, streamflow, reservoir, or groundwater level

SOCIOECONOMIC DROUGHT

- When the supply and demand of various commodities is affected by drought

Drought Classification



Meteorological Droughts in California

- Highest Agricultural Contribution : 12 % approx.
- $50 billion in cash receipts in 2019
- 20-year average : 59.54% during the winter and 63.40% in the fall
- D0 - Abnormally Dry : 100.0 % of CA

D1 - Moderate Drought : 97.5 % of CA

D2 - Severe Drought : 92.9 % of CA

D3 - Extreme Drought : 73.3 % of CA

D4 - Exceptional Drought : 5.4 % of CA

Problem Statement

What has been the trend of drought intensities in various California counties from 2014-2020?

How strong is the correlation present between droughts and meteorological data?

How effectively can drought be predicted using meteorological indicators?

Which of these indicators are significant in drought predictions?

Dataset

SOURCE

NASA Earth Science/Applied Science Program → NASA Langley Research Center (LaRC) POWER Project → **Kaggle**

FEATURES

- PRECTOT
- PS
- QV2M
- TS
- WS50M
- WS50M_MIN
- WS50M_MAX
- WS50M_RANGE
- WS10M
- WS10M_MIN
- WS10M_MAX
- WS10M_RANGE
- T2M
- T2M_MIN
- T2M_MAX
- T2M_RANGE
- T2MDEW
- T2MWET

LABELS

0 – No Drought

1 - Abnormally Dry (D0)

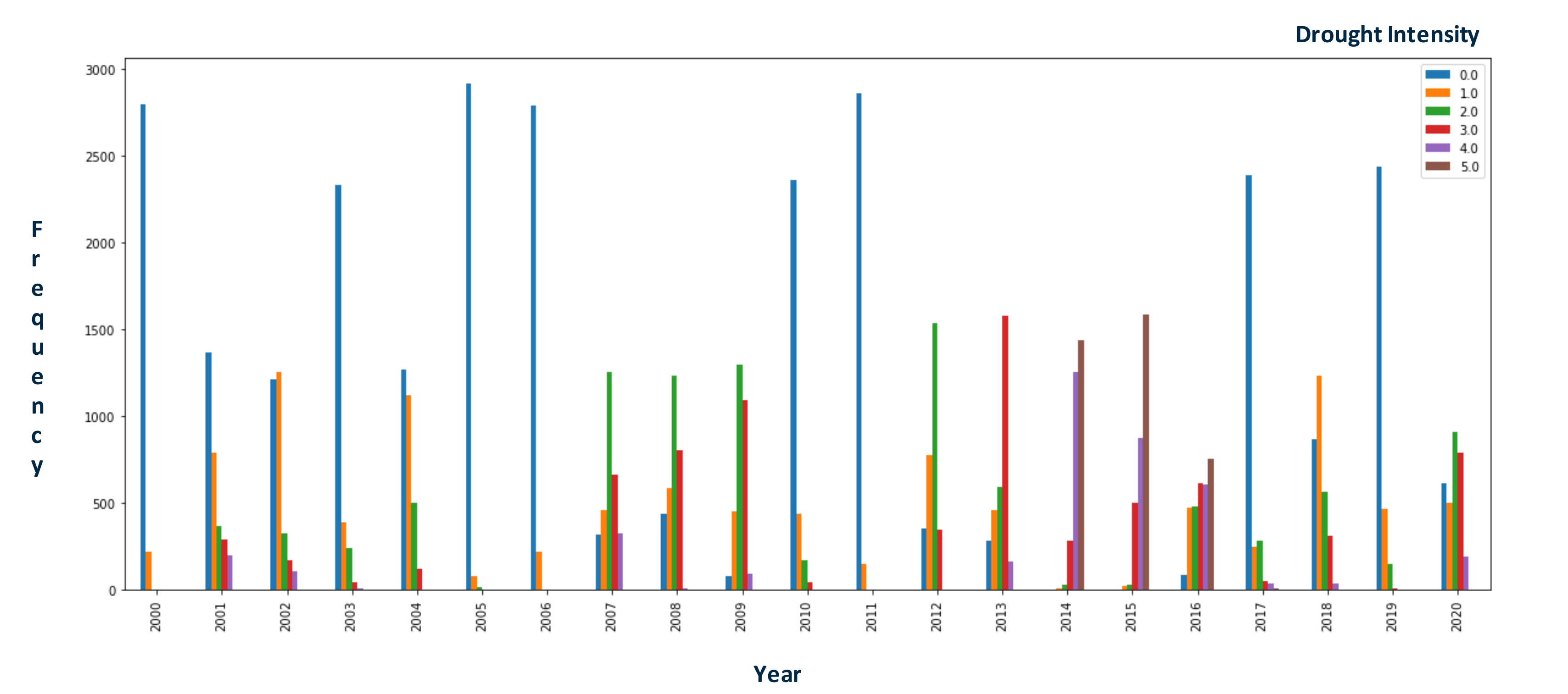
2 - Moderate Drought (D1)

3 – Severe Drought (D2)

4 - Extreme Drought (D3)

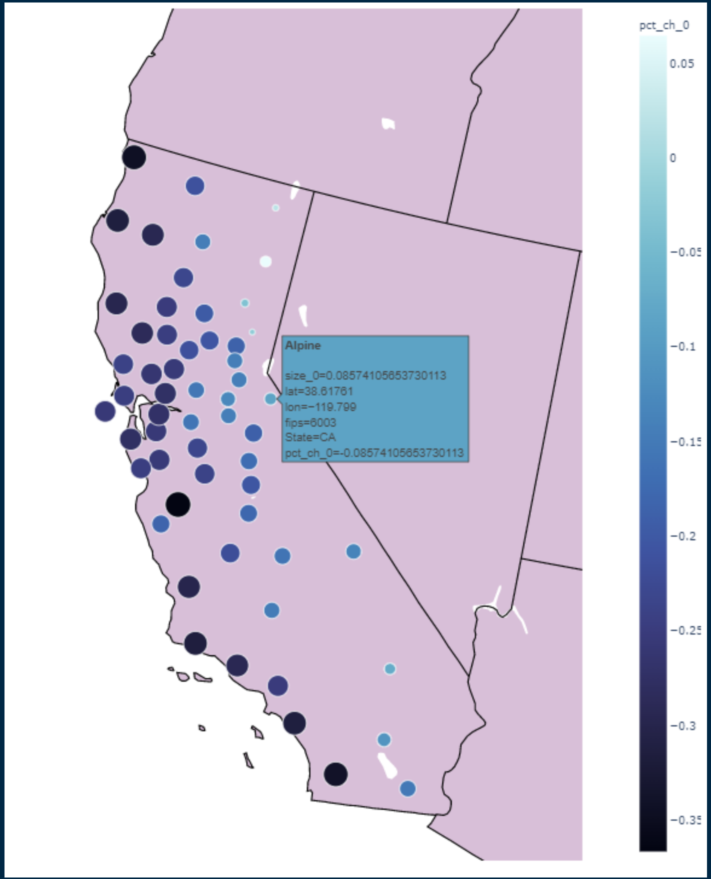
5 - Exceptional Drought (D4)

Analysis of Drought trends 2000-2020



UMBC

'No drought' trend



2000-2013
&
2014-2020

Out of 58 CA counties:

+ve change : 3 Counties

-ve change : 55 Counties

2007-2013
&
2014-2020

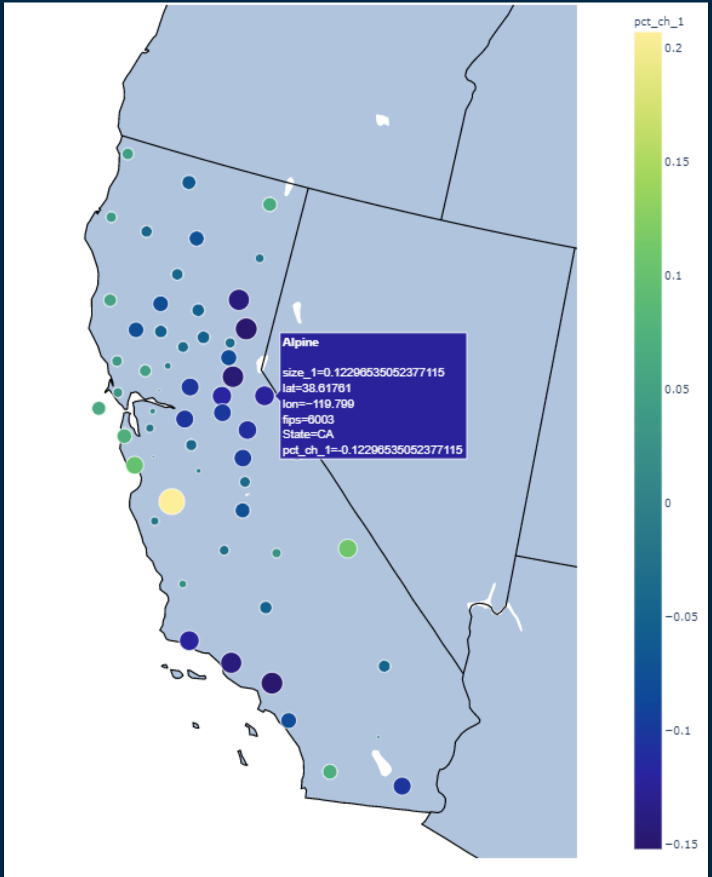
Out of 58 CA counties:

+ve change : 57 Counties

-ve change : 1 Counties

UMBC

‘Drought intensity 1’ trend



Out of 58 CA counties:

+ve change : 16 Counties

-ve change : 42 Counties

2000-2013 & 2014-2020

Out of 58 CA counties:

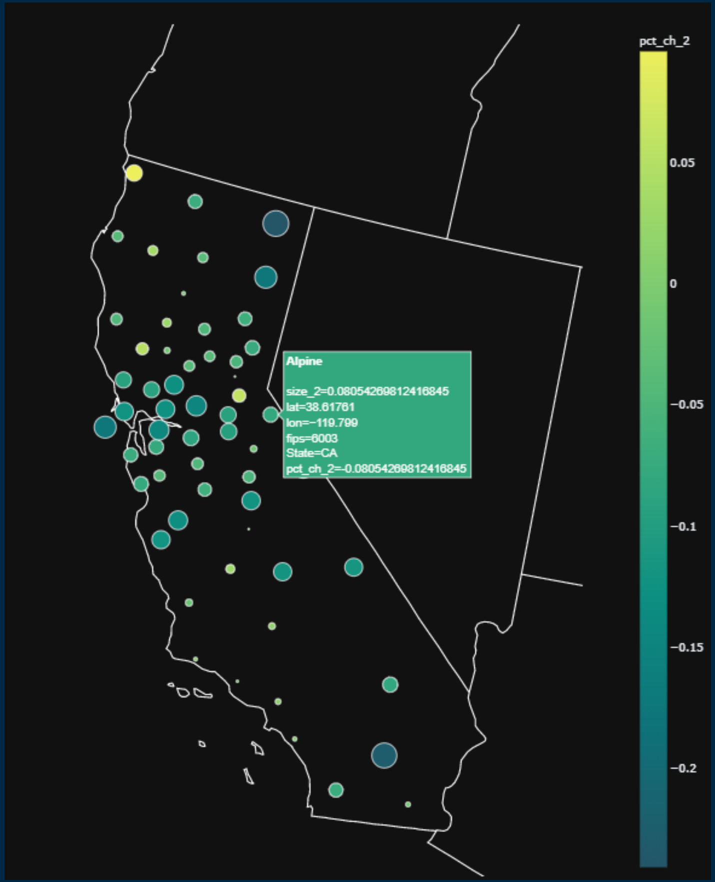
+ve change : 48 Counties

-ve change : 10 Counties

2007-2013 & 2014-2020

UMBC

‘Drought intensity 2’ trend



2000-2013
&
2014-2020

Out of 58 CA counties:

+ve change : 10 Counties

-ve change : 48 Counties

2007-2013
&
2014-2020

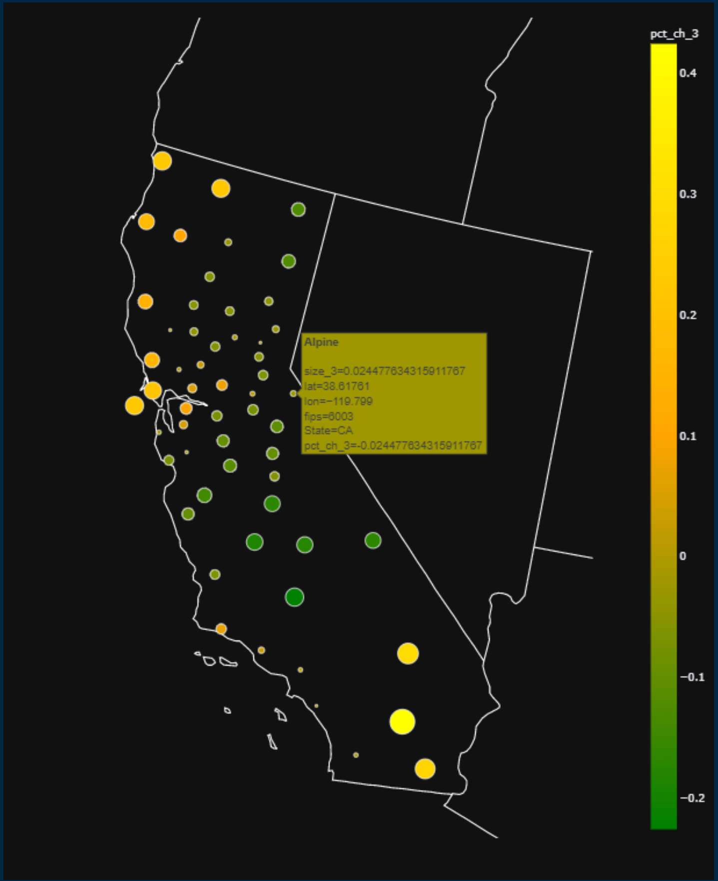
Out of 58 CA counties:

+ve change : 19 Counties

-ve change : 39 Counties

UMBC

‘Drought intensity 3’ trend



2000-2013
&
2014-2020

Out of 58 CA counties:

+ve change : 22 Counties

-ve change : 36 Counties

2007-2013
&
2014-2020

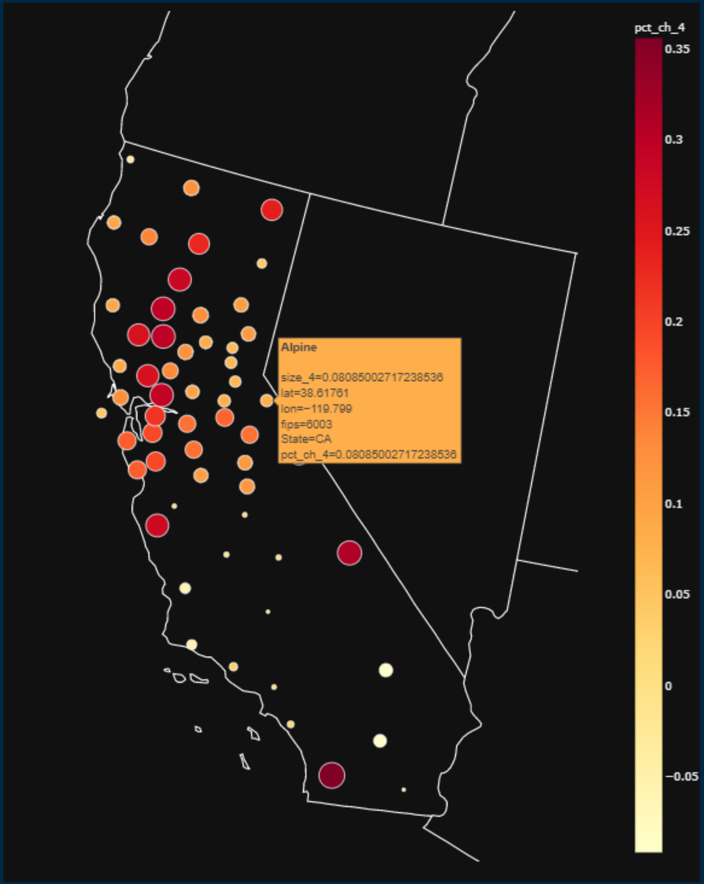
Out of 58 CA counties:

+ve change : 25 Counties

-ve change : 33 Counties

UMBC

‘Drought intensity 4’ trend



Out of 58 CA counties:

+ve change : 50 Counties

-ve change : 8 Counties

2000-2013 & 2014-2020

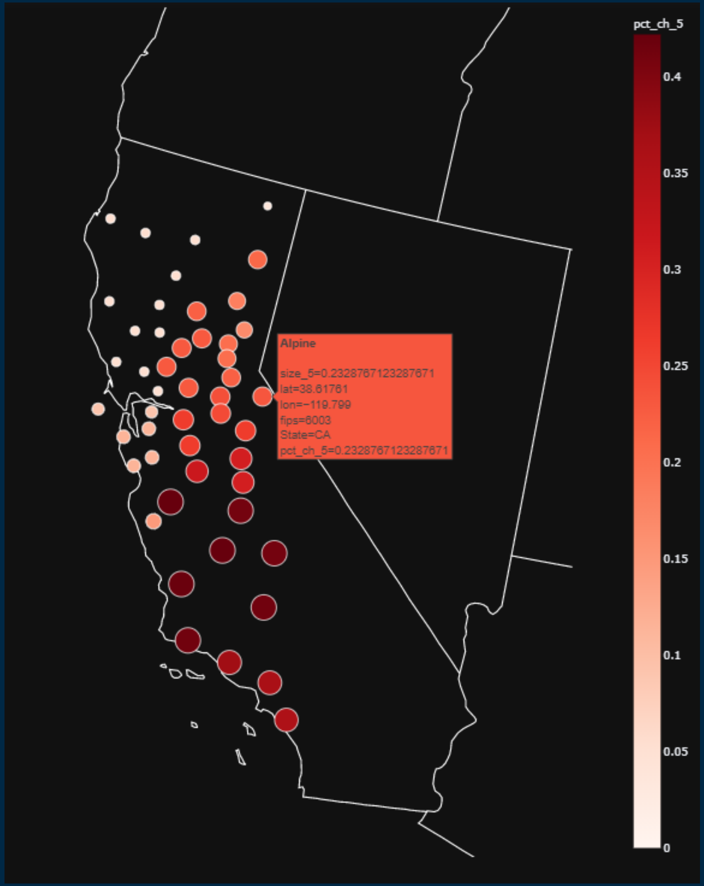
Out of 58 CA counties:

+ve change : 52 Counties

-ve change : 6 Counties

2007-2013 & 2014-2020

‘Drought intensity 5’ trend



2000-2013 & 2014-2020

Out of 58 CA counties:

+ve change : 58 Counties

-ve change : 0 Counties

2007-2013 & 2014-2020

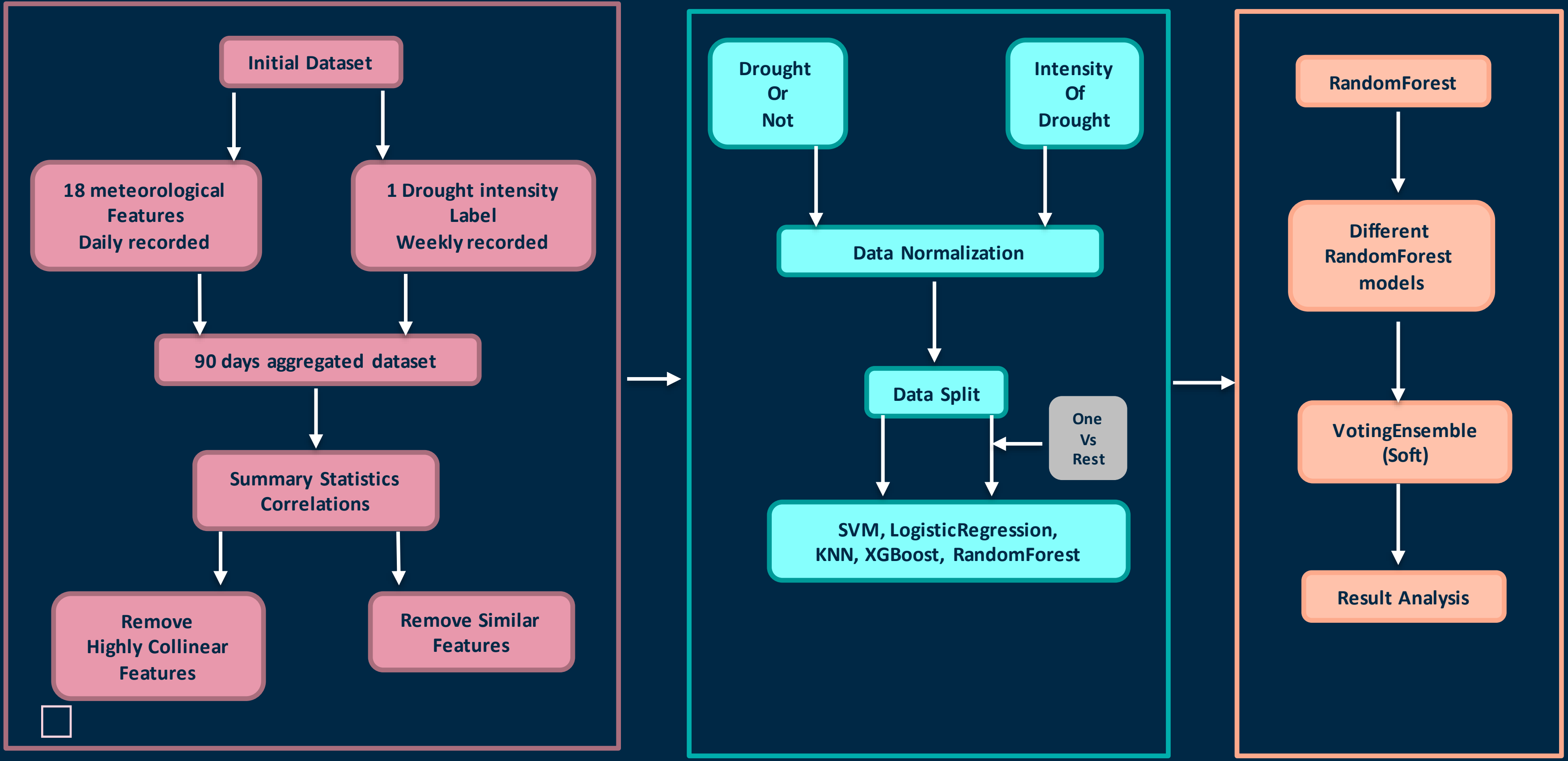
Out of 58 CA counties:

+ve change : 58 Counties

-ve change : 0 Counties

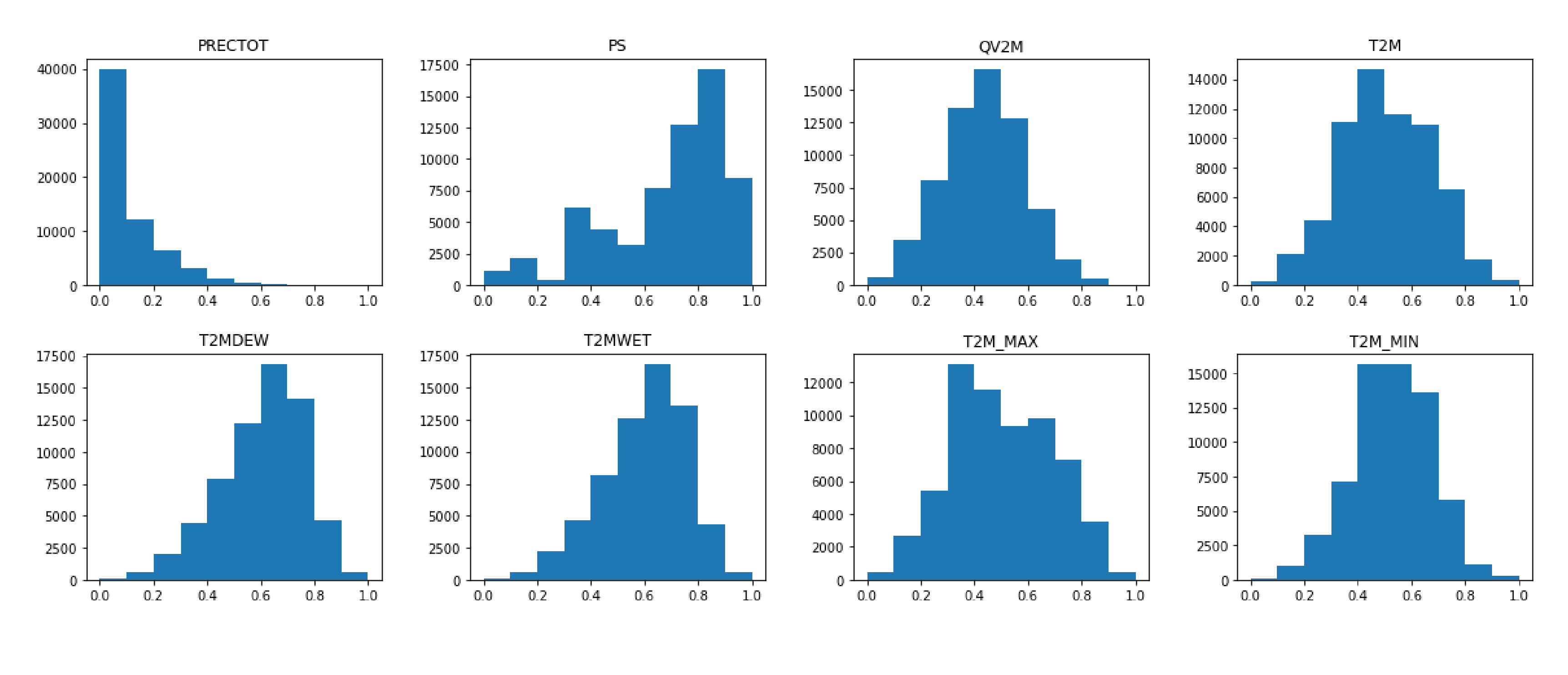
UMBC

Methodology

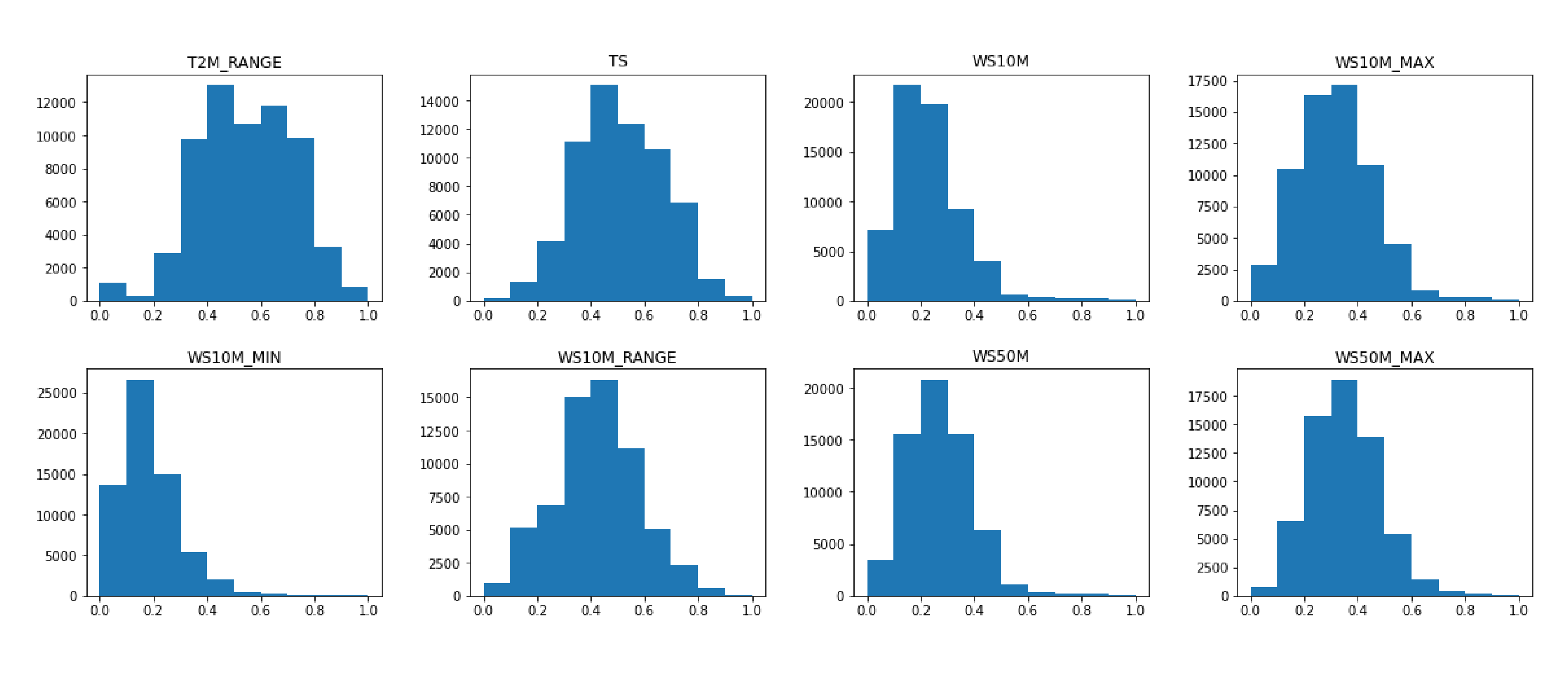


Data Filtering　　Model Selection　　Ensemble

Distribution Graphs: Features & Labels

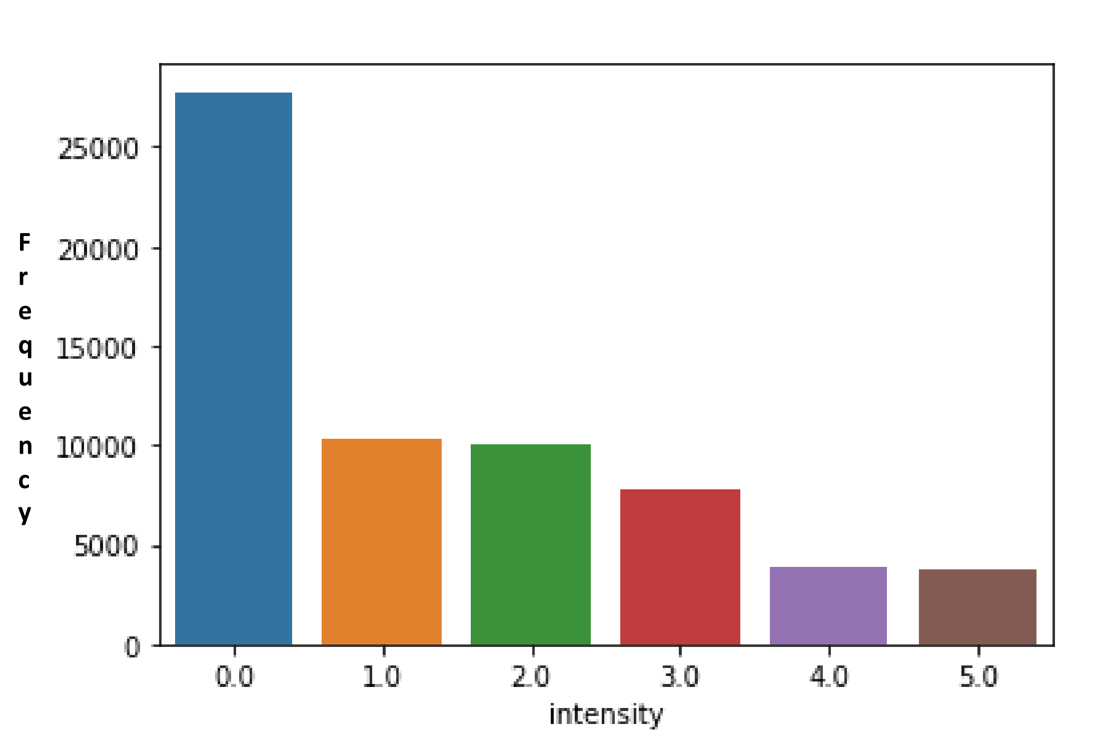


Distribution Graphs: Features & Labels



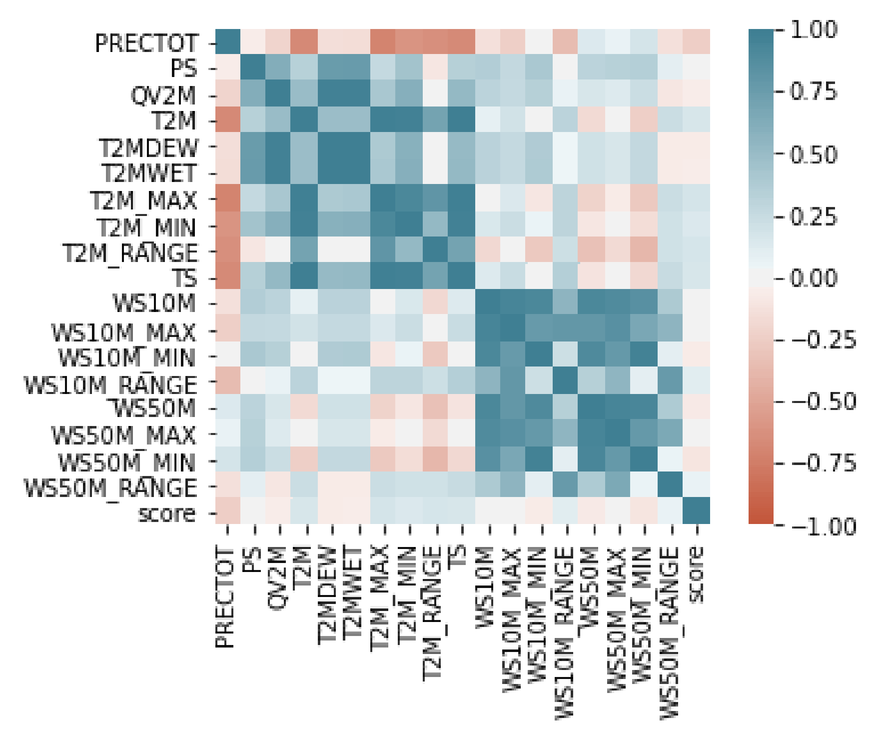
UMBC

Distribution Graphs: Features & Labels



Frequency	Intensity
27785	0.0
10371	1.0
9985	2.0
7722	3.0
3911	4.0
3794	5.0

Correlations



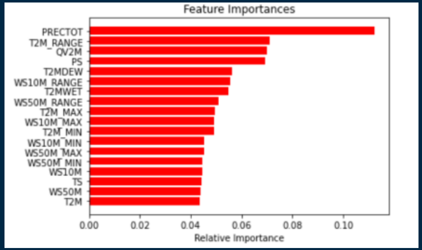
UMBC

RESULTS

Drought Or Not Prediction

accuracy_score: 0.8465733312359079
f1_score: 0.8465733312359078

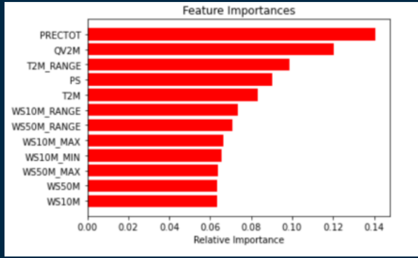
	precision	recall	f1-score	support
0	0.84	0.90	0.87	10731
1	0.86	0.78	0.82	8340
accuracy			0.85	19071
macro avg	0.85	0.84	0.84	19071
weighted avg	0.85	0.85	0.85	19071



Without Removing Any Features

accuracy_score: 0.852760736196319
f1_score: 0.852760736196319

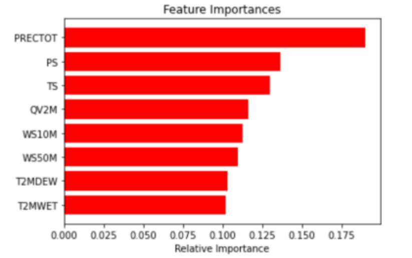
	precision	recall	f1-score	support
0	0.84	0.91	0.87	10731
1	0.87	0.79	0.82	8340
accuracy			0.85	19071
macro avg	0.85	0.85	0.85	19071
weighted avg	0.85	0.85	0.85	19071



After Removing Highly Collinear Features

accuracy_score: 0.815531435163337
f1_score: 0.815531435163337

	precision	recall	f1-score	support
0	0.81	0.87	0.84	10731
1	0.82	0.74	0.78	8340
accuracy			0.82	19071
macro avg	0.82	0.81	0.81	19071
weighted avg	0.82	0.82	0.81	19071



After Removing Similar Features

Drought Or Not Prediction

accuracy_score: 0.8465733312359079
f1_score: 0.8465733312359078

	precision	recall	f1-score	support
0	0.84	0.90	0.87	10731
1	0.86	0.78	0.82	8340
accuracy			0.85	19071
macro avg	0.85	0.84	0.84	19071
weighted avg	0.85	0.85	0.85	19071

RandomForest 1

accuracy_score: 0.8432698862146715
f1_score: 0.8432698862146715

	precision	recall	f1-score	support
0	0.84	0.90	0.87	10731
1	0.85	0.77	0.81	8340
accuracy			0.84	19071
macro avg	0.85	0.84	0.84	19071
weighted avg	0.84	0.84	0.84	19071

RandomForest 2

accuracy_score: 0.8463111530596193
f1_score: 0.8463111530596193

	precision	recall	f1-score	support
0	0.84	0.90	0.87	10731
1	0.85	0.78	0.82	8340
accuracy			0.85	19071
macro avg	0.85	0.84	0.84	19071
weighted avg	0.85	0.85	0.85	19071

RandomForest 3

accuracy_score: 0.8464160243301347
f1_score: 0.8464160243301347

	precision	recall	f1-score	support
0	0.84	0.90	0.87	10731
1	0.86	0.78	0.82	8340
accuracy			0.85	19071
macro avg	0.85	0.84	0.84	19071
weighted avg	0.85	0.85	0.85	19071

VotingEnsemble (Soft) of RandomForest1, RandomForest2 and RandomForest3.

Drought Intensity Prediction

accuracy_score: 0.7224033535165347
f1_score: 0.7224033535165348

	precision	recall	f1-score	support
1	0.72	0.77	0.75	3126
2	0.67	0.75	0.71	2970
3	0.71	0.71	0.71	2323
4	0.81	0.57	0.67	1164
5	0.84	0.71	0.77	1152
accuracy			0.72	10735
macro avg	0.75	0.70	0.72	10735
weighted avg	0.73	0.72	0.72	10735

RandomForest 1

accuracy_score: 0.7321844434094085
f1_score: 0.7321844434094085

	precision	recall	f1-score	support
1	0.73	0.78	0.75	3126
2	0.68	0.76	0.72	2970
3	0.72	0.72	0.72	2323
4	0.82	0.59	0.69	1164
5	0.86	0.71	0.78	1152
accuracy			0.73	10735
macro avg	0.76	0.71	0.73	10735
weighted avg	0.74	0.73	0.73	10735

RandomForest 2

accuracy_score: 0.7312529110386586
f1_score: 0.7312529110386586

	precision	recall	f1-score	support
1	0.73	0.78	0.75	3126
2	0.68	0.75	0.72	2970
3	0.72	0.72	0.72	2323
4	0.82	0.59	0.69	1164
5	0.86	0.71	0.78	1152
accuracy			0.73	10735
macro avg	0.76	0.71	0.73	10735
weighted avg	0.74	0.73	0.73	10735

RandomForest 3

accuracy_score: 0.7336748952026083
f1_score: 0.7336748952026082

	precision	recall	f1-score	support
1	0.73	0.78	0.76	3126
2	0.68	0.76	0.71	2970
3	0.73	0.72	0.73	2323
4	0.83	0.60	0.69	1164
5	0.86	0.70	0.77	1152
accuracy			0.73	10735
macro avg	0.77	0.71	0.73	10735
weighted avg	0.74	0.73	0.73	10735

VotingEnsemble (Soft) of RandomForest1, RandomForest2 and RandomForest3.

Conclusions

- Drought is persistent
- Meteorological data is useful for predicting droughts
- Precipitation (PRECTOT), Surface Pressure (PS) and Earth Skin Temperature (TS) are the most significant

Future Scope of Work

- Days aggregation can be varied
- Location data can be used
- Time series trends can be incorporated
- Scope can be expanded to analyse all US states

Do you have any questions?

THANK YOU

ıs