Christopher Botts
DSC680
Draft of White Paper 3

Classification of Evapotranspiration-based Regions in Southern California

Business Problem

Climatic conditions determine the suitability of a region for the establishment of plant life, including crops, native species, and managed landscapes (Beteri et al 2024). As global warming changes regional climatic conditions, agricultural producers anticipate a yield decline in staple crops such as corn, wheat, and rice (Pugh et al 2016). In this way, ensuring the climatic suitability of plant life is essential to meet global agricultural needs and maintain economic viability.

One approach to address this concern is the use of climate analogues. Utilizing climate analogues involves classifying regional climates to identify new areas that will match a crop's ideal climate after global warming (Pugh et al 2016). As the climate changes, production can be shifted to the newly identified regions to maintain crop yield.

Background/History

The current standard for plant suitability used by landscape professionals is the USDA agricultural zones, which recommends plant selections based on a region's lowest annual temperature. This evaluation method lacks depth.

One alternative is an evapotranspiration-based classification, which utilizes a robust analysis of weather factors. The California Irrigation Management Information System has identified 18 unique regions which are numbered from 1 to 18. This system could provide plant recommendations for area in California from climatically similar regions. Evapotranspiration based classification is a good candidate for mapping agricultural regions to their climate analogues.

Data Explanation

The data used in this project was made available from the California Irrigation Management Information System, a program run by the California Department of Water Resources. This program manages 145 weather stations that collect climate data across the state. Each instance in the dataset represents one day of data from one weather station. This project contains instances from 16 different stations in the eight counties that represent the Southern California area: Los Angeles, Orange, Ventura, Santa Barbara, Riverside, San Bernadino, San Diego, and Imperial Counties.

The following data dictionary describes the features included in data downloaded from the CIMIS website.

Data Dictionary	
Jul	The day of the year
ETo	Reference evapotranspiration using a standardized surface of cool season turf, measured in inches
Precip	The number of inches of rain received in a day
Sol Rad	The amount of energy delivered by the sun measured in units of Langly per day.
Avg Vap Pres (mBars)	A component of atmospheric pressure that is derived from molecular concentration of water in the air
Max Air Temp (F)	The temperature high for the day measured in degrees Fahrenheit
Min Air Temp (F)	The temperature low for the day measured in degrees Fahrenheit
Avg Air Temp (F)	The temperature average high for the day measured in degrees Fahrenheit
Max Rel Hum (%)	The maximum amount of water vapor in the air for the day provided as a percentage
Min Rel Hum (%)	The minimum amount of water vapor in the air for the day provided as a percentage
Avg Rel Hum (%)	The average amount of water vapor in the air for the day provided as a percentage
Dew Point (F)	The temperature to which the air must be cooled to be saturated with water vapor.
Avg Wind Speed (mph)	The average speed of wind for the day, given in miles per hour
Wind Run (miles)	The amount of wind that passed in a day, given in miles
Avg Soil Temp (F)	A measurement of the warmth of the soil, given in Fahrenheit.
CIMIS Zone	Zone number designated by the CIMIS system

The CIMIS zone number is included in this dataset because of feature creation. This CIMIS zone number is provided in the station details section of the CIMIS website. For stations that did not already have an assigned CIMIS Zone number, the zone was estimated from the CIMIS map. CIMIS zones were mapped to the appropriate stations using a dictionary.

CIMIS Zone is the target feature used in this project. There are eight CIMIS Zones represented in this data set: 1, 2, 6, 9, 10, 14, 17, 18.

Methods

Multiclass classification models were built to predict the CIMIS zone based on climate data. After all data was combined into one dataframe, the class sizes were uneven. Downsampling was used to create even classes of 3173 instances.

A correlation matrix was generated from the balanced dataset using Pearson's correlation coefficient. This matrix showed high correlation between several features, suggesting multicollinearity.

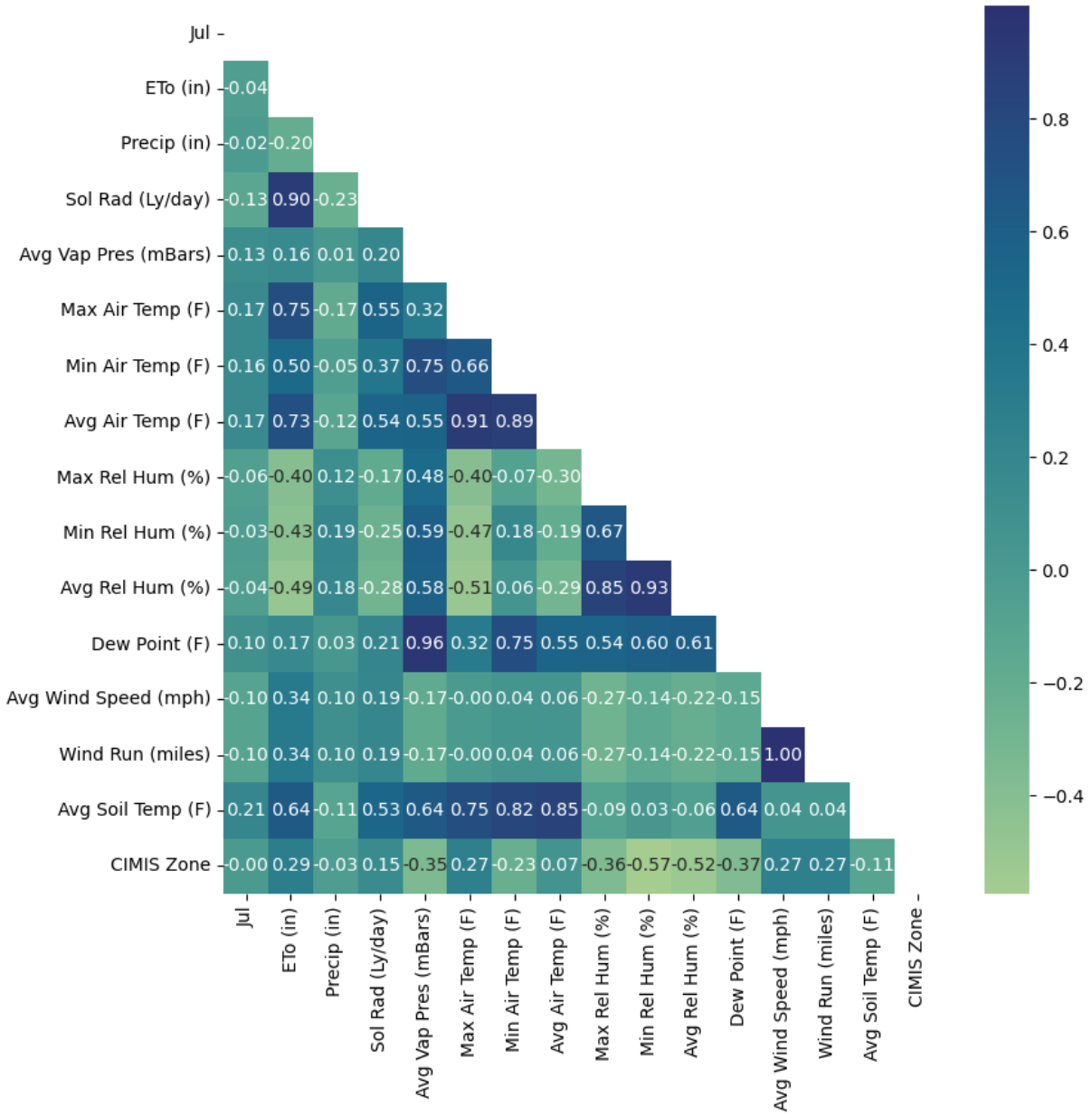


Figure 1. Correlation Matrix of CIMIS features

To remedy this, recursive feature elimination was used. Sklearn's RFECV recommended the elimination of four features:

```
'ETo (in)', 'Precip (in)', 'Avg Vap Pres (mBars)', 'Avg Wind Speed (mph)'
```

These four features were dropped from the dataset before model creation.

First, a baseline model was created to compare the results of three classification models. The accuracy of the baseline model is presented below:

```
Accuracy: 0.13016073116924046
```

Next, three models were built using the following methodology. Optimal hyperparameters were chosen using GridSearch with five fold cross validation. The dataset was split into a training and test set of 75% and 25% percent respectively. Values predicted by the model were run against the test set to calculate the model's accuracy. Lastly, a confusion matrix, f1 scores, and feature importance were generated for each model.

Model 1. Decision Tree using the following parameters:

max_depth=None,
criterion='log_loss',
max_features= None,
splitter='best',
min_samples_split= 4,
min_samples_leaf= 4

Model 2. Random Forest Classifier using the following parameters:

min_samples_split=2,
min_samples_leaf=1,
max_depth=None,
criterion='entropy',
max_features=0.4

Model 3. XGB Classifier with objective= 'multi: softmax' and the following parameters:

learning_rate=.5,
max_depth=5,
min_child_weight=2,
n_estimators=220,
subsample= 0.9,
gamma=0

For compatibility with XGBClassifier, the target feature of model 3 was label encoded and then inversed transformed to run an accuracy test.

Analysis

The accuracy scores of the three models are as follows:

Model	Accuracy
Decision Tree	0.674
Random Forest Classifier	0.786
XGB Classifier	0.820

Table 1. Accuracy scores for three CIMIS models.

Gradient Boosting provided the highest accuracy at 82% after tuned hyperparameters. The confusion matrix and f1scores generated by the XGB model are presented below.

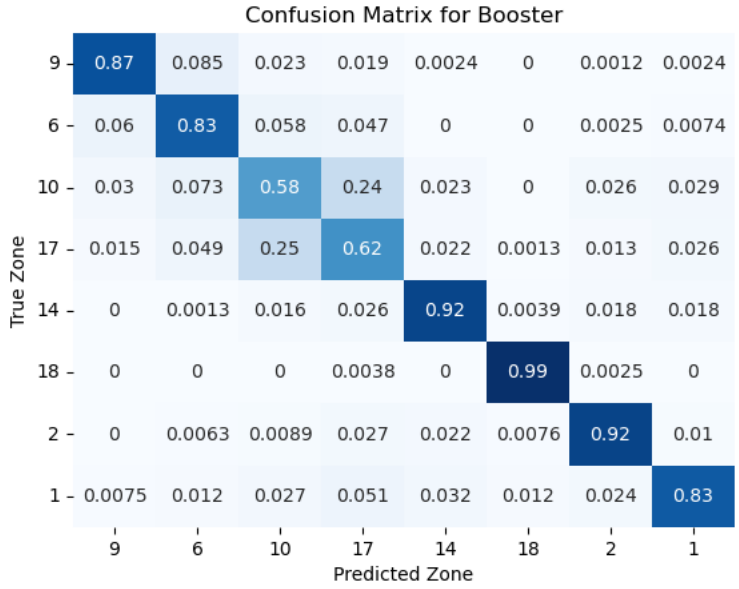


Figure 2. Confusion matrix for XGB model predictions.

	precision	recall	f1-score	support
1	0.89	0.87	0.88	826
2	0.79	0.83	0.81	815
6	0.61	0.58	0.59	793
9	0.59	0.62	0.60	758
10	0.90	0.92	0.91	762
14	0.98	0.99	0.98	799
17	0.91	0.92	0.92	788
18	0.90	0.83	0.87	805
accuracy			0.82	6346
macro avg	0.82	0.82	0.82	6346
weighted avg	0.82	0.82	0.82	6346

Table 2. Precision, recall, and F1 Score for the XGB model.

Most prediction errors for this model occurred in Zones 6, 9, 10 and 17.

24% of the time, Zone 10 was incorrectly predicted to be zone 17 and 25% of the time, zone 17 was incorrectly predicted to be zone 10. However, based on the high f1 score for both zones, false positives were rare. It is likely that too many instances were predicted to be zone 10 or 17.

On the other hand, zone 6 and 9 scored low on both precision and recall indicating that true positives were missed and predicted positives were often inaccurate. However, these two zones overall had a high accuracies with correct classification rates of 83% and 87%. Zone 9 was most often misclassified as zone 6 and zone 6 was most often misclassified as zone 9, 10 or 17.

Below are the feature important scores for the XGB model.

```
Feature: 0, Score: 0.09219 Jul
Feature: 1, Score: 0.05304 Sol Rad (Ly/day)
Feature: 2, Score: 0.07518 Max Air Temp (F)
Feature: 3, Score: 0.10113 Min Air Temp (F)
Feature: 4, Score: 0.08357 Avg Air Temp (F)
Feature: 5, Score: 0.07600 Max Rel Hum (%)
Feature: 6, Score: 0.14629 Min Rel Hum (%)
Feature: 7, Score: 0.06131 Avg Rel Hum (%)
Feature: 8, Score: 0.07041 Dew Point (F)
Feature: 9, Score: 0.09346 Wind Run (miles)
Feature: 10, Score: 0.14742 Avg Soil Temp (F)
```

Table 3. Feature importance scores for the XGB model.

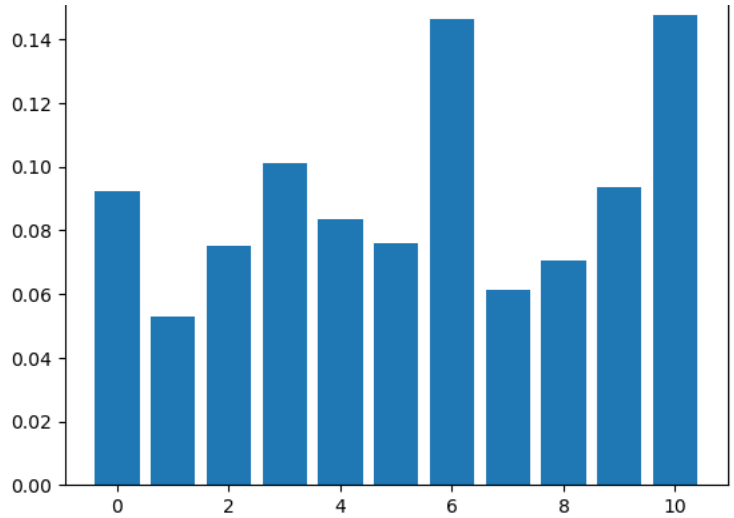


Figure 3. Bar chart displaying feature importance for XGB model.

The two most important features were average soil temperature and minimum relative humidity. Average soil temperature may be reflective of tree coverage in an area. Regions with more tree coverage are likely to see lower overall soil temperatures as well as reduced periods of high soil temperature. On the other hand, regions such as the deserts of Imperial County have little tree coverage and will consistently have high soil temperatures.

Relative humidity may be used to distinguish areas that have coastal influence. Areas with significant coastal influence observe a consistently positive minimum relative humidity, while the relative humidity in most other regions will drop to zero during the day. Exploratory data analysis shows that relative humidity did not reach zero in zones 1, 10 and 17.

The next three features in order of importance are minimum air temperature, wind run and day of the year. It is notable that the primary distinguishing feature for USDA zones is one of the significant predictors in this model. We can expect there to be overlap in the zones designated by CIMIS and the USDA.

Conclusion

Classification of regions by evapotranspiration is a complex system requiring several layers of information including average soil temperature, minimum relative humidity, minimum air temperature, wind run, and the day of the year on which data was collected.

Individually, all factors displayed a weak correlation with CIMIS Zone as observed in the bar graph below.

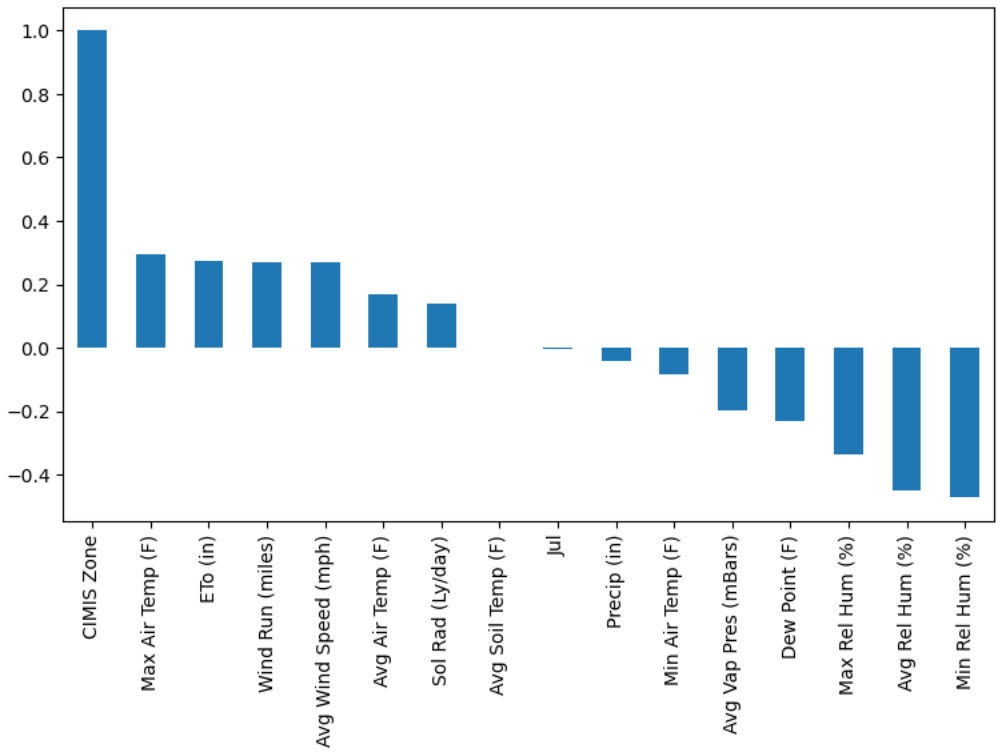


Figure 4. Bar chart of the correlation between CIMIS zone and other features.

However, when these features are combined, they have strong predictive power for determining evapotranspiration-based zones. This indicates that a combination of climatic measurements better classifies distinct regions than calculated evapotranspiration rates on their own.

It is important to note that some regions were predicted highly accurately, such as region 18 which had a 99% correct classification rate, while other zones, such as Zone 10, displayed an accuracy of only 58%. Regions that are difficult for the model to differentiate, such as zone 10 and 17, may be more similar to each other than the CIMIS classifications imply. These zones should be reevaluated on their climatic similarity.

Assumptions

One assumption in this model is that USDA zones, which are based on an area's minimum temperature alone, is inadequate for distinguishing climatic regions. To challenge this assumption, models were constructed using a variety of climate data.

A second major assumption involves the original designation of the CIMIS zones. CIMIS zones use cool season turf-grass as the base material for evapotranspiration calculations. Cool-season grass may not be a good baseline plant material for the diverse climates of Southern California.

Limitations

One limitation to this project is that the CIMIS zones are not clearly defined. The CIMIS zone map provides average Eto per zone but does not provide quantitative climatic definitions for zoning. This makes classification of different zones, and cities that border two zones challenging.

Another limitation with the data is that soil temperature is only provided as an average. The minimum value tends to hold the most predictive power for features that include a minimum, maximum and average value. This raises the question if minimum soil temperature could improve predictive accuracy if provided instead of average soil temperature.

Lastly, this model accounts for many of the above ground climatic factors that a plant experiences but does not include data about the below ground environment in which these plants are growing. Including information from soil analysis could generate a more holistic prediction of plant success.

Challenges

One challenge with this dataset is the distribution of weather reporting stations. Without a uniform distribution, it is difficult to ensure representation of all areas in California.

Another challenge is the inconsistency in station activation dates; stations began recording data across a range of dates, so that data does not line up across all stations. Some stations have also been decommissioned, which further restricts the dates of the data. This limits the ability to do a time series analysis on the data.

Future Uses and Additional Applications

As climate change modifies the environment, this model could serve additional functions outside of agriculture. Regional classification can be used to discover cities with climates similar to a post-climate change environment. Plants from these areas could planted in urban forests ahead of time, so that the city greenery can grow in the future climate.

Furthermore, this model offers the opportunity to analyze restoration techniques across different climates zones. Efforts that have been successful in similarly classified regions may also be successful in analogous zones in California.

Recommendations

Depending on the exact inquiry, this dataset can easily grow quite sizable. If it is desired to utilize more stations (this inquiry used 16 of 145) or a longer time frame (this inquiry utilized 10 years of data from 2013-2023), then this system should modified so that it is compatible with big data tools.

Implementation Plan

Weather data should be collected from stations in the United States and worldwide. Special attention should be paid to regions that are climatically similar to California, such as Texas, Chile, Australia, the Mediterranean, and Mexico. The model can then classify the expected climate conditions after global warming to identify the climate analogues. These regions should be evaluated for the economic viability of adopting crop production in response to climate change.

Ethical Assessment

The regions used for this project are predefined by the California Irrigation Management Information System. This may mean the regions are biased towards irrigated landscapes systems, such as turf grass. The assumption that the CIMIS regions are broadly applicable may be problematic.

There are also ethical implications regarding the application of this model. The model can be used to identify plant species from other areas that could succeed in specific regions in California. However, if new introductions perform too well, they may escape cultivation and become invasive or outcompete native species. For this reason, the model should be used to identify potential candidates, but no species should be introduced without further research into the environmental impacts of introduction.

References

Beteri, J., Lyimo, J.G. & Msinde, J.V. (2024). The influence of climatic and environmental variables on sunflower planting season suitability in Tanzania. *Sci Rep* **14**, 3906 (2024). https://doi.org/10.1038/s41598-023-49581-5

Pugh, T. A. M. *et al*. Climate analogues suggest limited potential for intensification of production on current croplands under climate change. *Nat. Commun.* 7:12608 doi: 10.1038/ncomms12608 (2016).

State of California (2024). California Irrigation Management Information System. *California Department of Water Resources*. Retrieved from: https://cimis.water.ca.gov/WSNReportCriteria.aspx

U.S. Department of Agriculture (2023). 2023 USDA Plant Hardiness Zone Map. Retrieved from: https://planthardiness.ars.usda.gov/

Appendix.

Comparison of CIMIS Zones and USDA Plant hardiness Zones in California

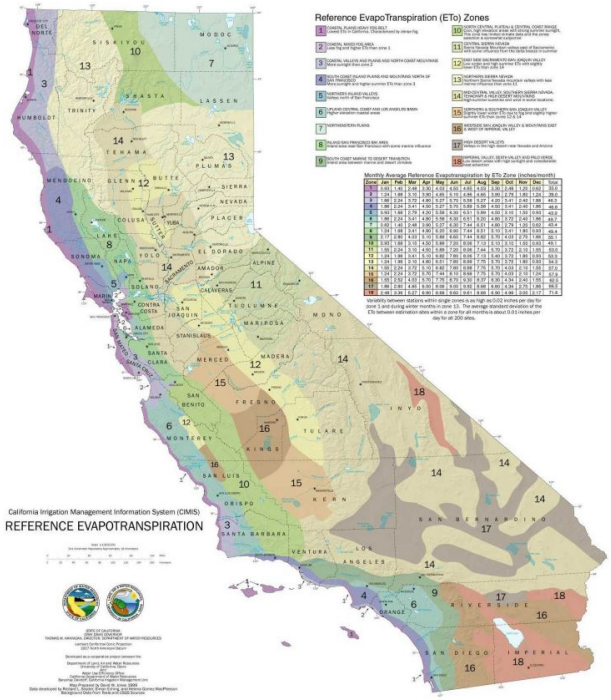


Figure 5. Map of CIMIS Zones. Image from https://cimis.water.ca.gov/App_Themes/images/etozonemap.jpg

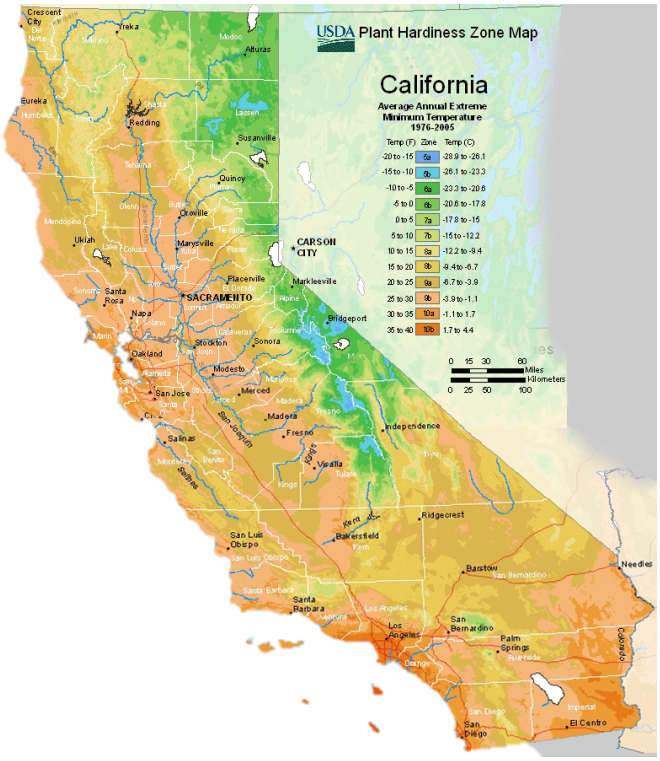


Figure 6. Map of USDA Hardiness Zones. Image from https://planthardiness.ars.usda.gov/

Feature Distribution

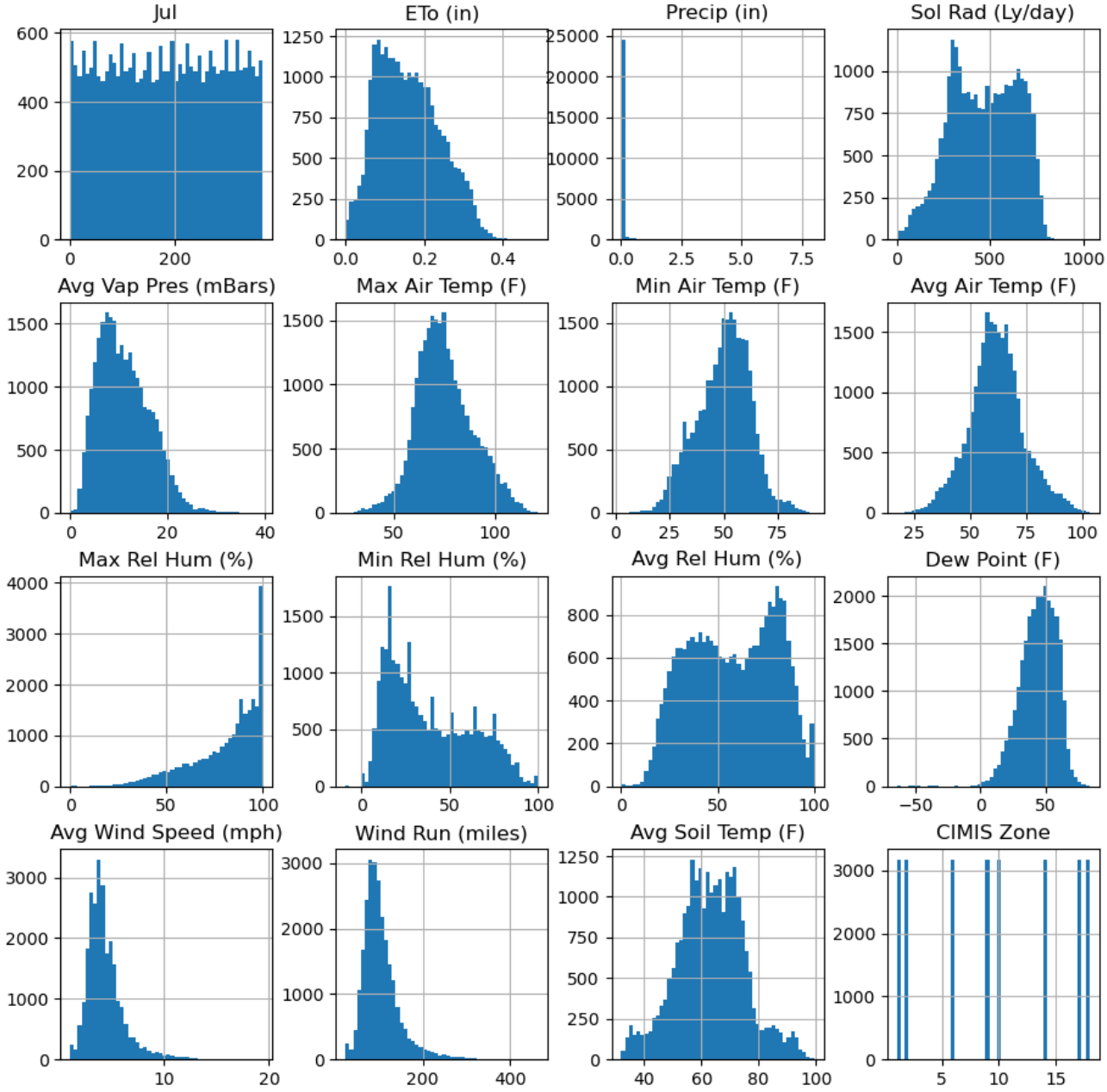


Figure 7. Histogram of each feature in CIMIS data.

Correlation Matrix of df features

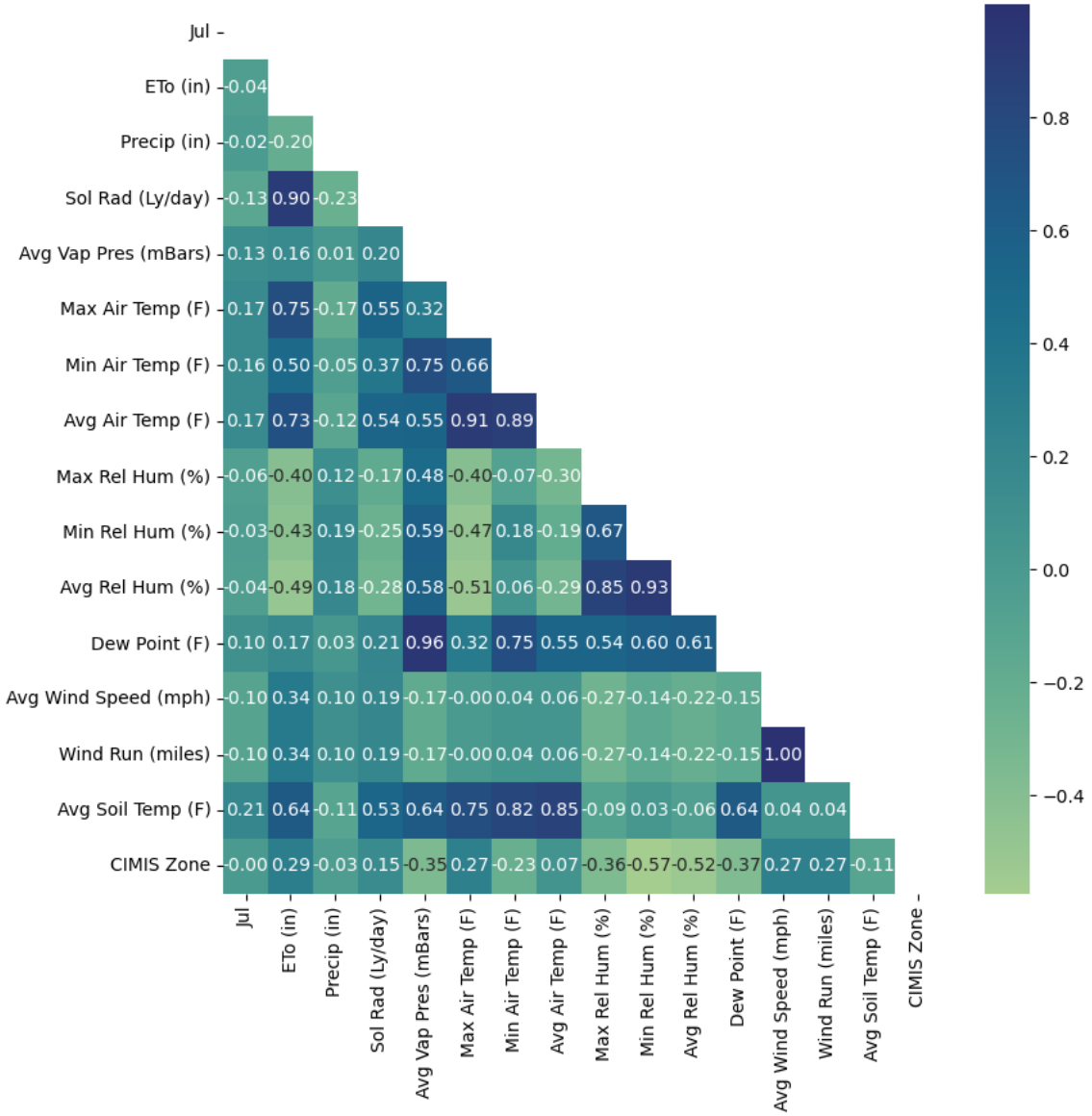


Figure 8. Correalation matrix of CIMIS features after balancing.

Statistical summary of features grouped by CIMIS Zone

ETo (in)

CIMIS Zone	count	mean	std	min	25%	50%	75%	max
1	3173.0	0.118235	0.051054	0.0	0.08	0.12	0.16	0.35
2	3173.0	0.138733	0.056138	0.0	0.10	0.14	0.18	0.29
6	3173.0	0.147548	0.068169	0.0	0.09	0.15	0.20	0.37
9	3173.0	0.143029	0.065062	0.0	0.09	0.14	0.20	0.31
10	3173.0	0.181179	0.089904	0.0	0.10	0.18	0.26	0.37
14	3173.0	0.151109	0.079837	0.0	0.08	0.15	0.23	0.30
17	3173.0	0.192969	0.095988	0.0	0.10	0.19	0.28	0.42

18	3173.0	0.198894	0.093093	0.0	0.12	0.20	0.28	0.49

Precip (in)

CIMIS Zone	count	mean	std	min	25%	50%	75%	max
1	3173.0	0.032077	0.208481	0.0	0.0	0.0	0.0	8.01
2	3173.0	0.026013	0.155242	0.0	0.0	0.0	0.0	2.96
6	3173.0	0.028802	0.148411	0.0	0.0	0.0	0.0	1.94
9	3173.0	0.029474	0.140111	0.0	0.0	0.0	0.0	2.21
10	3173.0	0.007488	0.049580	0.0	0.0	0.0	0.0	0.85
14	3173.0	0.047996	0.242590	0.0	0.0	0.0	0.0	4.11
17	3173.0	0.009036	0.067289	0.0	0.0	0.0	0.0	1.35
18	3173.0	0.008686	0.082371	0.0	0.0	0.0	0.0	2.45

Sol Rad (Ly/day)

CIMIS Zone	count	mean	std	min	25%	50%	75%	max
1	3173.0	415.147810	155.861764	0.0	296.0	410.0	537.0	784.0
2	3173.0	451.428932	163.447038	15.0	317.0	461.0	591.0	919.0
6	3173.0	435.165459	174.680922	0.0	304.0	439.0	581.0	851.0
9	3173.0	446.079735	171.325843	0.0	309.0	447.0	592.0	800.0
10	3173.0	486.466120	187.117300	17.0	323.0	497.0	658.0	1030.0
14	3173.0	484.438071	198.260000	1.0	321.0	488.0	665.0	958.0
17	3173.0	505.743460	178.926548	31.0	348.0	517.0	669.0	970.0
18	3173.0	491.962811	168.801328	0.0	351.0	506.0	639.0	984.0

Avg Vap Pres (mBars)

CIMIS Zone	count	mean	std	min	25%	50%	75%	max
1	3173.0	14.958304	4.354019	2.7	12.0	14.9	18.2	28.7
2	3173.0	13.537662	4.780446	1.8	10.2	13.8	17.4	24.2
6	3173.0	12.264482	4.685728	0.0	8.7	12.4	15.7	26.4
9	3173.0	12.917365	4.504093	2.3	9.6	12.9	16.5	24.9
10	3173.0	8.166625	2.645407	1.8	6.3	8.1	10.0	20.7
14	3173.0	7.233659	3.381463	0.1	4.7	6.8	9.3	19.8
17	3173.0	7.170596	3.057015	1.2	4.9	6.8	9.0	19.7
18	3173.0	12.810495	6.630720	0.0	8.0	11.1	15.8	39.6

Max Air Temp (F)

CIMIS Zone	count	mean	std	min	25%	50%	75%	max
1	3173.0	66.783517	5.929577	50.3	62.3	66.5	70.9	91.7
2	3173.0	71.459817	7.323757	51.7	66.2	71.3	76.0	96.9
6	3173.0	76.762433	10.912691	48.2	68.7	76.7	84.5	108.4
9	3173.0	75.420611	9.767956	48.6	68.3	75.0	82.1	105.5

10	3173.0	78.533407	15.226955	39.9	66.7	79.0	92.0	108.2
14	3173.0	64.849259	13.967408	26.0	54.8	65.8	76.4	95.2
17	3173.0	77.280744	15.314504	39.6	65.1	77.3	90.9	109.9
18	3173.0	88.822597	14.936851	48.0	76.6	88.9	102.3	121.2

Min Air Temp (F)

CIMIS Zone	count	mean	std	min	25%	50%	75%	max
1	3173.0	56.341380	6.714398	32.0	52.0	56.5	61.5	73.8
2	3173.0	55.550457	6.228869	32.0	51.2	55.5	60.4	74.4
6	3173.0	51.987614	9.308568	21.4	46.0	52.3	58.8	77.3
9	3173.0	51.770470	8.254755	28.7	45.4	51.8	58.4	73.4
10	3173.0	44.009297	10.836063	12.8	35.4	43.3	52.6	71.1
14	3173.0	35.846328	10.989324	2.5	27.4	34.3	44.2	66.1
17	3173.0	47.907312	13.031497	16.8	37.5	47.3	58.4	77.3
18	3173.0	56.043902	14.879709	23.0	44.0	54.7	68.2	91.8

Avg Air Temp (F)

CIMIS Zone	count	mean	std	min	25%	50%	75%	max
1	3173.0	61.353325	5.518853	44.2	57.3	61.3	65.5	78.2
2	3173.0	62.791049	6.180359	46.7	58.2	62.8	67.2	81.8
6	3173.0	62.957989	8.825698	37.7	56.7	62.6	69.0	92.0
9	3173.0	62.528081	7.901724	42.1	56.7	62.3	68.2	88.0
10	3173.0	60.312071	12.992953	28.3	49.5	59.0	71.9	90.0
14	3173.0	50.407595	12.437175	17.4	40.4	49.4	61.6	74.3
17	3173.0	62.934478	14.153392	29.3	51.2	61.6	75.8	91.9
18	3173.0	72.437882	14.779719	39.0	59.9	71.6	86.1	103.4

Max Rel Hum (%)

CIMIS Zone	count	mean	std	min	25%	50%	75%	max
1	3173.0	89.622439	10.818427	25.0	87.0	92.0	96.0	100.0
2	3173.0	84.630949	16.125540	10.0	84.0	89.0	95.0	100.0
6	3173.0	85.567602	18.396405	3.0	81.0	93.0	98.0	100.0
9	3173.0	88.688308	13.036286	14.0	85.0	93.0	98.0	100.0
10	3173.0	74.634415	15.970574	23.0	63.0	76.0	88.0	100.0
14	3173.0	83.838954	15.907844	1.0	75.0	87.0	98.0	100.0
17	3173.0	63.110621	17.645756	16.0	49.0	63.0	78.0	100.0
18	3173.0	73.053892	16.795876	0.0	62.0	75.0	86.0	100.0

Min Rel Hum (%)

CIMIS Zone	count	mean	std	min	25%	50%	75%	max
1	3173.0	67.618973	20.769621	1.0	60.0	75.0	82.0	100.0
2	3173.0	53.089820	21.060393	0.0	37.0	60.0	70.0	99.0
6	3173.0	39.341002	19.586963	-9.0	24.0	39.0	54.0	100.0
9	3173.0	43.892531	18.716276	0.0	29.0	47.0	59.0	100.0
10	3173.0	25.750709	16.567545	3.0	14.0	21.0	33.0	100.0
14	3173.0	32.614245	20.094448	0.0	19.0	27.0	40.0	100.0
17	3173.0	20.177750	13.784984	4.0	11.0	15.0	25.0	100.0
18	3173.0	24.438386	12.722437	0.0	15.0	22.0	31.0	90.0

Avg Rel Hum (%)

CIMIS Zone	count	mean	std	min	25%	50%	75%	max
1	3173.0	79.110621	16.186533	12.0	75.0	84.0	89.0	100.0
2	3173.0	68.378191	19.814828	7.0	59.0	76.0	82.0	100.0
6	3173.0	61.833281	20.222554	0.0	49.0	67.0	77.0	100.0
9	3173.0	65.496061	17.505369	8.0	56.0	70.0	78.0	100.0
10	3173.0	47.167665	18.567120	12.0	32.0	43.0	60.0	100.0
14	3173.0	55.698078	19.670607	0.0	42.0	53.0	67.0	100.0
17	3173.0	36.956193	16.284503	9.0	25.0	33.0	47.0	94.0
18	3173.0	44.913646	15.730473	0.0	34.0	43.0	54.0	100.0

Dew Point (F)

CIMIS Zone	count	mean	std	min	25%	50%	75%	max
1	3173.0	54.162748	8.925201	13.1	49.5	55.3	60.8	72.1
2	3173.0	50.720517	11.506946	4.0	45.1	53.2	59.6	69.0
6	3173.0	47.789474	12.252705	-64.3	40.9	50.2	56.6	71.5
9	3173.0	49.639616	10.715802	9.4	43.4	51.3	58.1	69.8
10	3173.0	37.972518	9.015938	3.7	32.9	38.9	44.5	64.4
14	3173.0	33.419950	13.393307	-53.7	25.4	34.6	42.6	63.2
17	3173.0	33.860416	11.274522	-4.3	26.4	34.8	41.8	63.1
18	3173.0	48.095241	13.733593	-64.3	38.7	47.3	56.8	83.8

Avg Wind Speed (mph)

CIMIS Zone	count	mean	std	min	25%	50%	75%	max
1	3173.0	4.179861	1.299314	1.0	3.4	4.0	4.6	17.9
2	3173.0	3.725875	0.777831	1.0	3.3	3.6	4.0	9.1
6	3173.0	3.544122	1.305199	1.0	2.8	3.4	4.1	14.9
9	3173.0	3.438449	1.047327	1.0	2.8	3.4	4.0	10.7
10	3173.0	5.076867	0.918443	1.0	4.5	5.0	5.5	12.6
14	3173.0	4.664608	1.785834	1.0	3.3	4.4	5.6	13.0
17	3173.0	5.625370	2.753188	1.1	3.4	5.2	7.2	17.2
18	3173.0	4.681721	2.471454	1.0	3.0	4.0	5.5	19.4

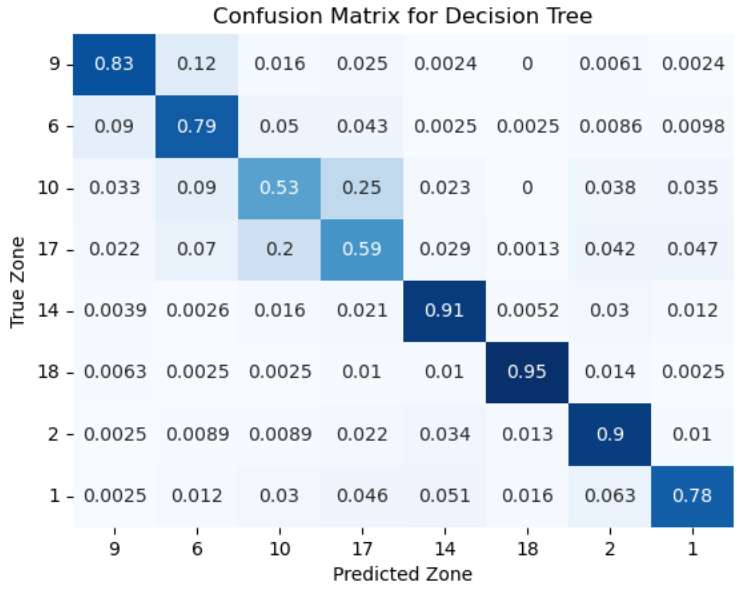
Wind Run (miles)

CIMIS Zone	count	mean	std	min	25%	50%	75%	max
1	3173.0	100.289032	31.161136	24.0	82.5	95.7	110.1	430.6
2	3173.0	89.409140	18.660471	24.0	78.4	87.5	97.1	219.4
6	3173.0	85.066183	31.320596	23.4	66.4	82.4	99.6	357.6
9	3173.0	82.514623	25.117818	24.0	66.7	81.8	95.4	256.7
10	3173.0	121.856350	22.023859	24.0	108.9	121.0	133.0	303.6
14	3173.0	111.941664	42.834575	24.0	79.7	104.7	135.2	312.3
17	3173.0	134.992657	66.088607	26.3	81.6	125.2	172.3	413.3
18	3173.0	112.366026	59.290220	24.5	73.1	96.2	133.0	464.4

Avg Soil Temp (F)

CIMIS Zone	count	mean	std	min	25%	50%	75%	max
1	3173.0	65.729720	6.842533	49.7	59.5	65.8	72.0	78.4
2	3173.0	63.751371	6.623753	47.0	58.7	64.6	69.4	79.1
6	3173.0	65.151812	8.601239	42.4	57.5	65.1	72.9	84.1
9	3173.0	64.305988	8.155090	45.6	57.4	64.5	71.5	80.5
10	3173.0	68.500536	14.892653	38.1	55.0	68.2	80.9	99.3
14	3173.0	49.634794	10.836238	31.9	39.3	49.2	59.7	69.6
17	3173.0	59.260227	9.204713	40.5	51.0	59.7	67.9	75.7
18	3173.0	69.980681	14.461967	36.6	58.1	70.8	82.8	99.1

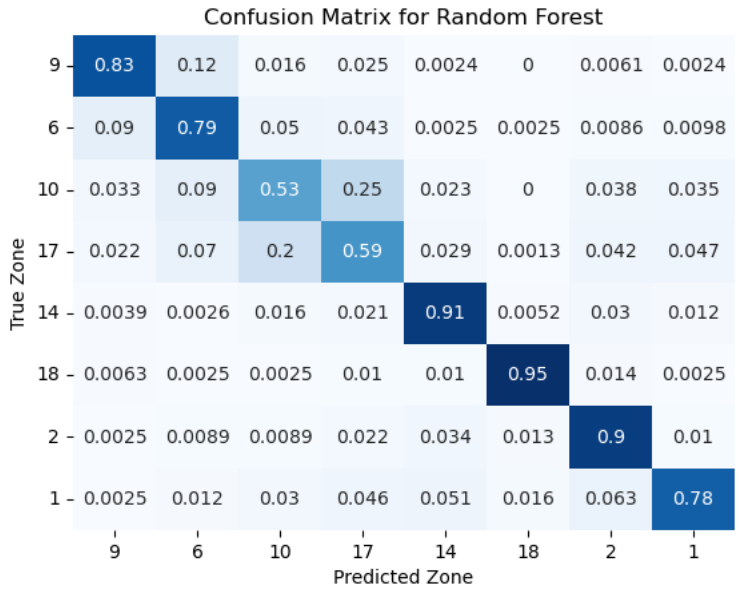
Results of Model 1. Decision Tree



	precision	recall	f1-score	support
1	0.73	0.75	0.74	826
2	0.61	0.59	0.60	815
6	0.44	0.48	0.46	793
9	0.44	0.47	0.45	758
10	0.78	0.80	0.79	762
14	0.89	0.91	0.90	799
17	0.79	0.75	0.77	788
18	0.75	0.66	0.70	805
accuracy			0.67	6346

macro avg	0.68	0.67	0.68	6346
weighted avg	0.68	0.67	0.68	6346

Results of Model 2. Random Forest



	precision	recall	f1-score	support
1	0.84	0.83	0.84	826
2	0.73	0.79	0.76	815
6	0.63	0.53	0.57	793
9	0.57	0.59	0.58	758
10	0.85	0.91	0.88	762
14	0.96	0.95	0.96	799
17	0.82	0.90	0.86	788
18	0.87	0.78	0.82	805
accuracy			0.79	6346
macro avg	0.78	0.79	0.78	6346
weighted avg	0.78	0.79	0.78	6346