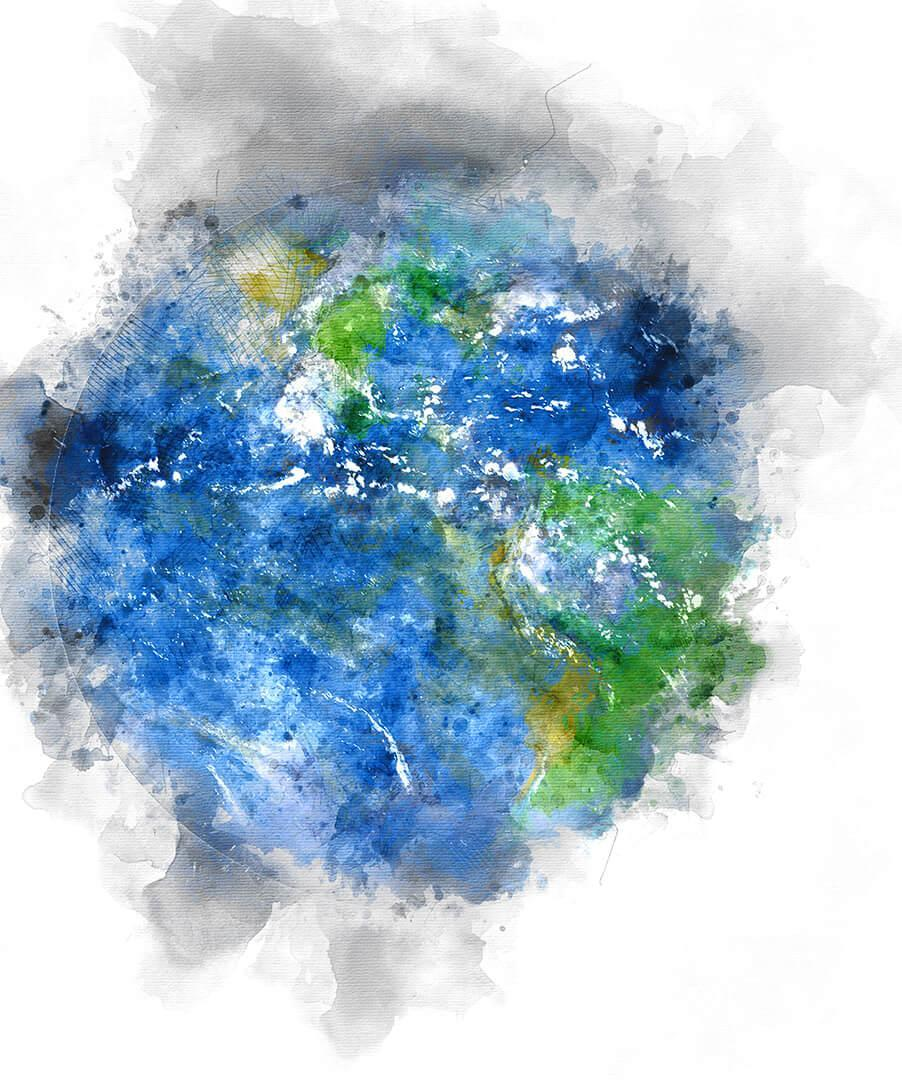


Weather Conditions & Climate Change with ClimateWins

Sydney Storer

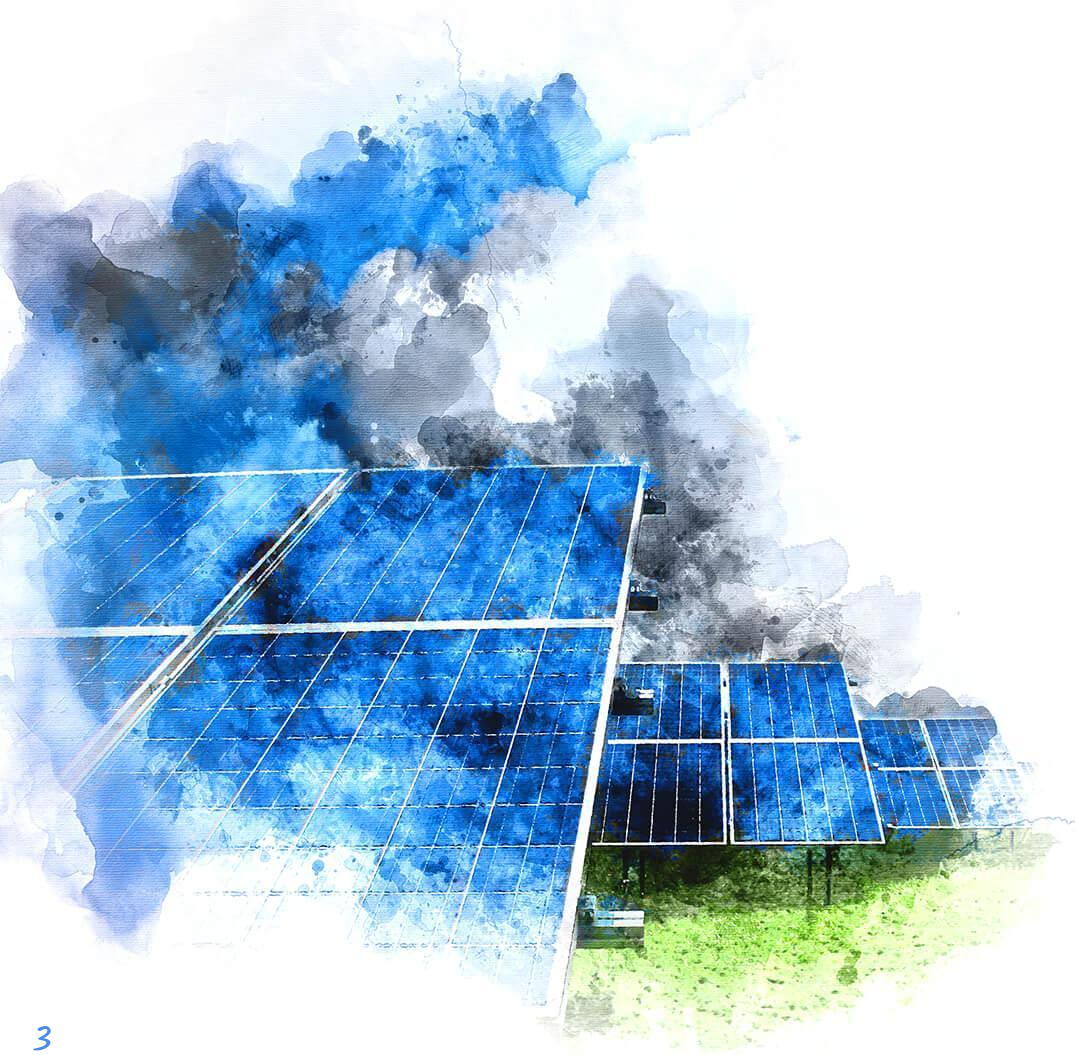
Objective & Hypotheses

Objective:

Use machine learning to help predict the consequences of climate change.

Hypotheses / Key Questions:

- How is machine learning used? Is it applicable to weather data?
- Historically, what have the maximums and minimums in temperature been?
- Can machine learning be used to predict whether weather conditions will be favorable on a certain day? (If so, it could also be possible to predict danger.)
- Can machine learning be used to predict natural disasters such as hurricanes?



The Data

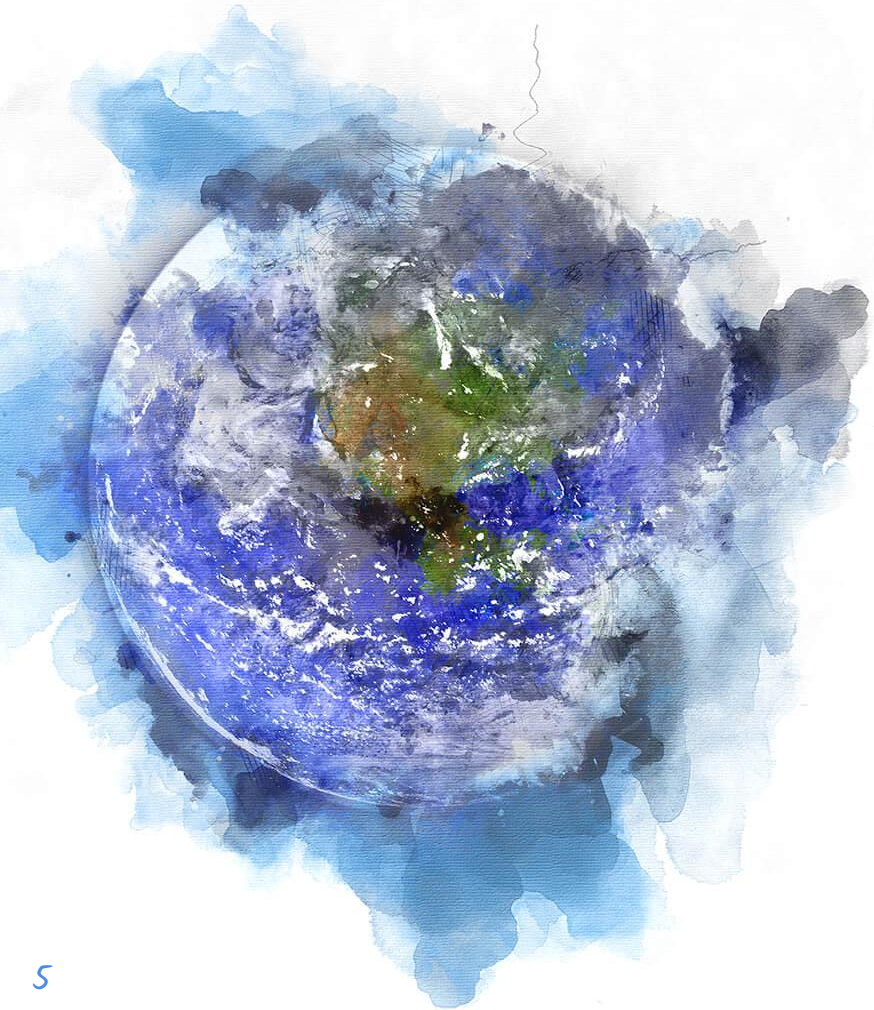
This data set is based on weather observations from 18 different weather stations across Europe, which contain data ranging from the late 1800s to 2022. Recordings exist for almost every day with values such as temperature, wind speed, snow, global radiation, and more. This data is collected by the European Climate Assessment & Data Set project.

Potential Bias

The dataset includes weather data from just 18 weather stations across Europe. It's possible that this is not a large enough sample to make assumptions about Europe's diverse climate.

Since the data goes all the way back to the 1800s, this large of a temporal sample could skew machine learning algorithms that should possibly be focusing more on recent data.

Ethical concerns: Climate change is a polarizing topic, so it is important to be impartial when analyzing the data.



Optimization

- Gradient descent was used to optimize the temperature data in the dataset
- Gradient descent is is one of the simplest ways to find a local minimum (or valley)
- Applied gradient descent to minimize loss using different values for theta0 and theta1 (values in the equation of the line), number of iterations, and step size

Supervised Learning Algorithms Used

K Nearest Neighbors (KNN)

- Classifies or predicts a data point based on the majority vote or average of its K nearest neighbors in the feature space.

Decision Tree

- Splits data into branches based on feature values, creating a tree-like structure where each path leads to a decision or prediction.

Artificial Neural Network (ANN)

- Mimics the brain's structure to process data through layers of interconnected nodes, learning patterns by adjusting weights via backpropagation.

Data Accuracy

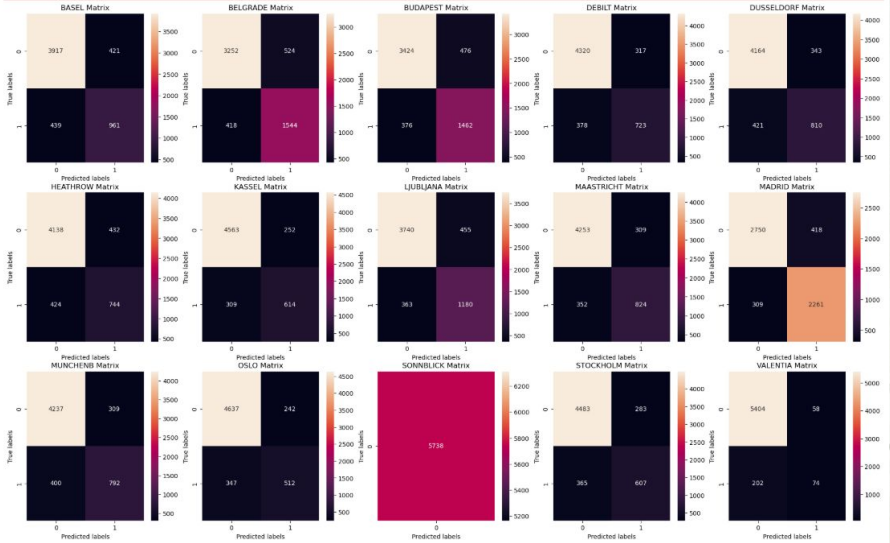
- Evaluated the accuracy of each of 3 supervised machine learning algorithms for classifying whether an algorithm can predict if weather conditions would be favorable on a certain day
- Also ran classification reports for each of the 3 algorithms to further evaluate model performance using precision, recall, and f1 score

KNN

The actual accuracy is much higher than the accuracy score of 45% might lead us to believe.

Looking at the actual accuracy scores for each station as well as the classification report, we can see that this model is performing relatively well to predict pleasant vs unpleasant weather days.

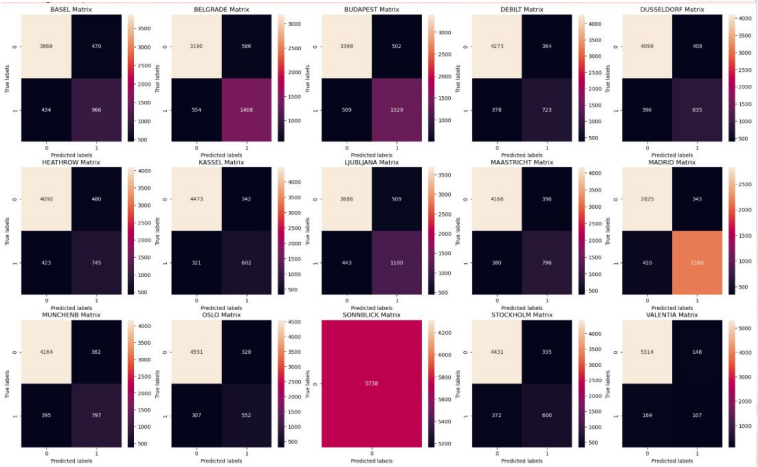
	precision	recall	f1-score	support
0	0.70	0.69	0.69	1400
1	0.75	0.79	0.77	1962
2	0.75	0.80	0.77	1838
3	0.70	0.66	0.68	1101
4	0.70	0.66	0.68	1231
5	0.63	0.64	0.63	1168
6	0.71	0.67	0.69	923
7	0.72	0.76	0.74	1543
8	0.73	0.70	0.71	1176
9	0.84	0.88	0.86	2570
10	0.72	0.66	0.69	1192
11	0.68	0.60	0.63	859
12	0.00	0.00	0.00	0
13	0.68	0.62	0.65	972
14	0.56	0.27	0.36	276
micro avg	0.73	0.72	0.73	18211
macro avg	0.66	0.63	0.64	18211
weighted avg	0.73	0.72	0.72	18211
samples avg	0.41	0.40	0.39	18211

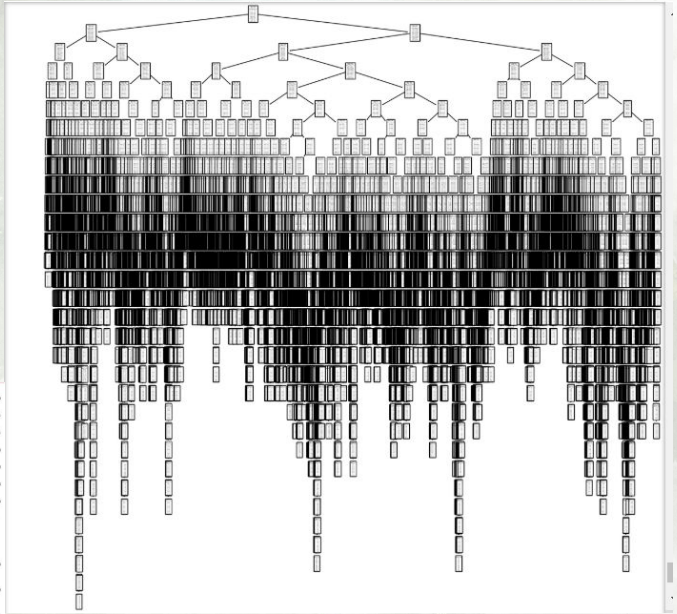


Weather Station	Accurate - Unpleasant	Accurate - Pleasant	False Positive	False Negative	Accuracy Rate
BASEL	3917	961	421	439	82%
BELGRADE	3252	1544	524	418	80%
BUDAPEST	3424	1462	476	376	83%
DEBILT	4320	723	317	378	86%
DUSSELDORF	4164	810	343	421	85%
HEATHROW	4138	744	432	424	82%
KASSEL	4563	614	252	309	89%
LJUBLJANA	3740	1180	455	363	83%
MAASTRICHT	4253	824	309	352	87%
MADRID	2750	2261	418	309	85%
MUNCHENB	4237	792	309	400	86%
OSLO	4637	512	242	347	89%
SONNBLICK	5738	0	0	0	100%
STOCKHOLM	4483	607	283	365	87%
VALENTIA	5404	74	58	202	95%

Decision Tree

	precision	recall	f1-score	support
0	0.67	0.68	0.68	1400
1	0.71	0.72	0.72	1962
2	0.73	0.73	0.73	1838
3	0.66	0.65	0.65	1101
4	0.67	0.66	0.67	1231
5	0.60	0.62	0.61	1168
6	0.65	0.66	0.65	923
7	0.67	0.70	0.68	1543
8	0.66	0.68	0.67	1176
9	0.86	0.84	0.85	2570
10	0.67	0.67	0.67	1192
11	0.62	0.62	0.62	859
12	0.00	0.00	0.00	0
13	0.63	0.59	0.61	972
14	0.39	0.37	0.38	276
micro avg	0.69	0.69	0.69	18211
macro avg	0.61	0.61	0.61	18211
weighted avg	0.69	0.69	0.69	18211
samples avg	0.42	0.42	0.40	18211

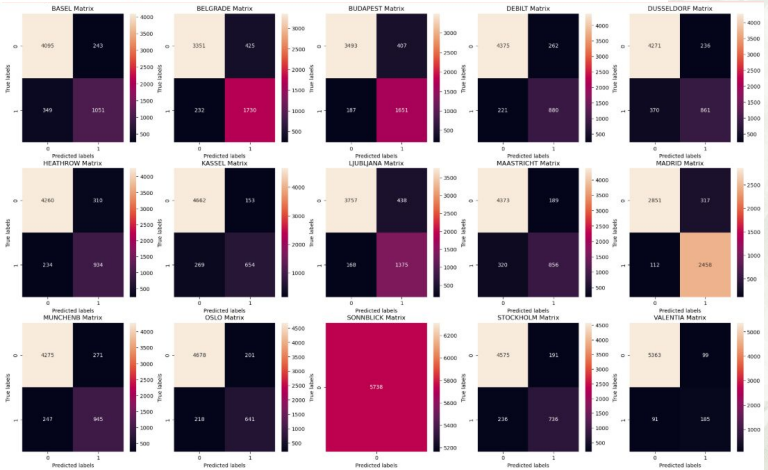




This decision tree could be pruned because of how complex it is with so many branches.

The accuracy score of the testing data was 47%, which we can infer to be actually higher than this based on the classification report.

	precision	recall	f1-score	support
0	0.76	0.84	0.80	1400
1	0.83	0.84	0.83	1962
2	0.81	0.88	0.84	1838
3	0.76	0.81	0.78	1101
4	0.78	0.72	0.75	1231
5	0.78	0.74	0.76	1168
6	0.79	0.73	0.76	923
7	0.77	0.85	0.81	1543
8	0.81	0.75	0.78	1176
9	0.87	0.96	0.92	2570
10	0.75	0.83	0.79	1192
11	0.75	0.75	0.75	859
12	0.00	0.00	0.00	0
13	0.79	0.74	0.76	972
14	0.68	0.68	0.68	276
micro avg	0.80	0.82	0.81	18211
macro avg	0.73	0.74	0.73	18211
weighted avg	0.80	0.82	0.81	18211
samples avg	0.48	0.49	0.47	18211



ANN

Based on the accuracy score of the testing data (52%) and the classification report having better numbers than the other algorithms, I recommend ClimateWins use the ANN model to predict favorable weather conditions.

After 3 scenarios tried, the best accuracy received was using 3 layers with 100, 50, and 25 nodes each, with a max iteration of 1000 and tolerance of 0.0001.

Again, we can infer that the actual accuracy rate of the model would be higher than 52% if calculated based on the confusion matrix.

Conclusion

We have determined that machine learning can be useful for analyzing weather data.

The ANN model best predicts favorable weather conditions on a certain day of the year.

Next Steps:

Prune the decision tree to hopefully achieve better accuracy of the model

Explore unsupervised learning methods to further analyze climate data

Combine supervised and unsupervised learning to create a comprehensive analysis of whether machine learning can predict the weather.

Thanks!

For More Information Visit:
GitHub