# WEATHER CONDITIONS AND CLIMATE CHANGE WITH CLIMATEWINS

AARON GILBERT, APRIL 2025

**Introduction:**

CimateWins is interested in using machine learning to help predict the consequences of climate change around Europe and, potentially, the world. It's been sorting through hurricane predictions from The National Oceanic and Atmospheric Administration (NOAA) in the U.S., typhoon data from The Japan Meteorological Agency (JMA) in Japan, world temperatures, and a great deal of other data.

**Objective:**

In response to the sharp rise in extreme weather events over the past two decades, ClimateWins is interested in using machine learning to help predict the consequences of climate change around Europe and, potentially, the world.

# HYPOTHESES

1. MACHINE LEARNING ALGORITHMS WILL BE ABLE TO ACCURATELY PREDICT FUTURE WEATHER CONDITIONS.

2. PREDICTION ACCURACY WILL VARY BASED ON GEOGRAPHIC LOCATION AND VARIOUS CLIMATE CONDITIONS WITHIN THE REGION.

3. MACHINE LEARNING CAN IDENTIFY SIGNS OF CLIMATE CHANGE AND ITS ADVERSE EFFECTS.

# DATA SET INFO

- Weather observations from 18 weather stations across Europe.
- Daily observations from 1800s through 2022.
- This data is collected by the Home European Climate Assessment & Dataset

# DATA BIAS & ACCURACY

**Collection Bias**

• Data is from 18 weather stations across Europe, while over 26,321 stations exist. This limited coverage may not represent Europe's diverse climates.

• Changes in instrumentation or station locations over time can introduce inconsistencies.

**Location Bias**

• Data focuses on Europe and the Mediterranean. Predictions may not generalize to regions like Brazil or Canada due to different climatic systems.

**Temporal Bias**

• Data spans from the late 1800s to 2022. Older records may no longer represent current conditions, potentially misleading machine learning models.

**Sampling Bias**

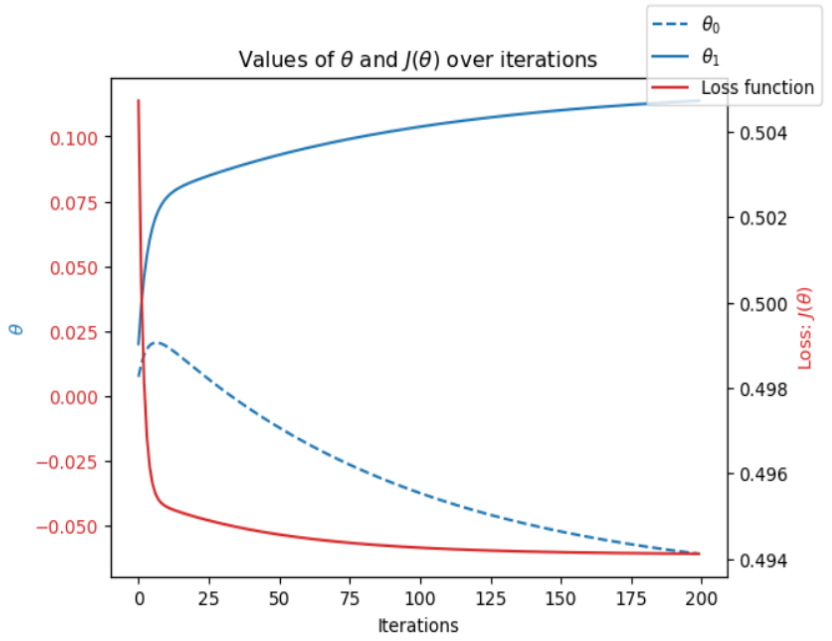
• Specific selection of stations can skew results, missing broader climate patterns across Europe.

**Data Accuracy**

- The Data Accuracy depends on the Machine Learning algorithm applied
- Artificial Neural Network (ANN) algorithm, had an overall of 46% to 49% accuracy in the test set, while presented an accuracy that ranged from 45 to 51% accuracy in the training set.
- The KNN model was the best fit, showing an accuracy of 88% on the test set.

# DATA OPTIMIZATION

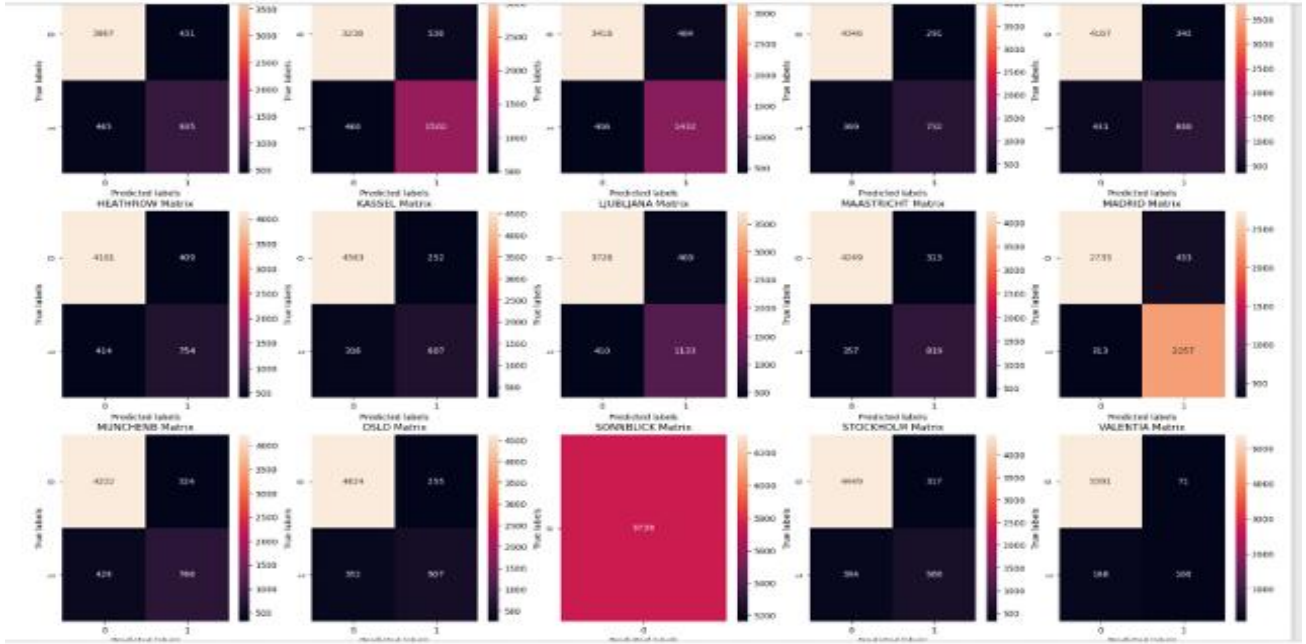
- This data set was optimized through Gradient Descent.
- The application of gradient descent, is one of the simplest ways to find a local minimum (or valley) and can be used in linear and nonlinear cases.
- In this case, we applied the gradient descent to find the minimum error, through number of iterations, as well as the number of steps (alpha), which varied as the case.



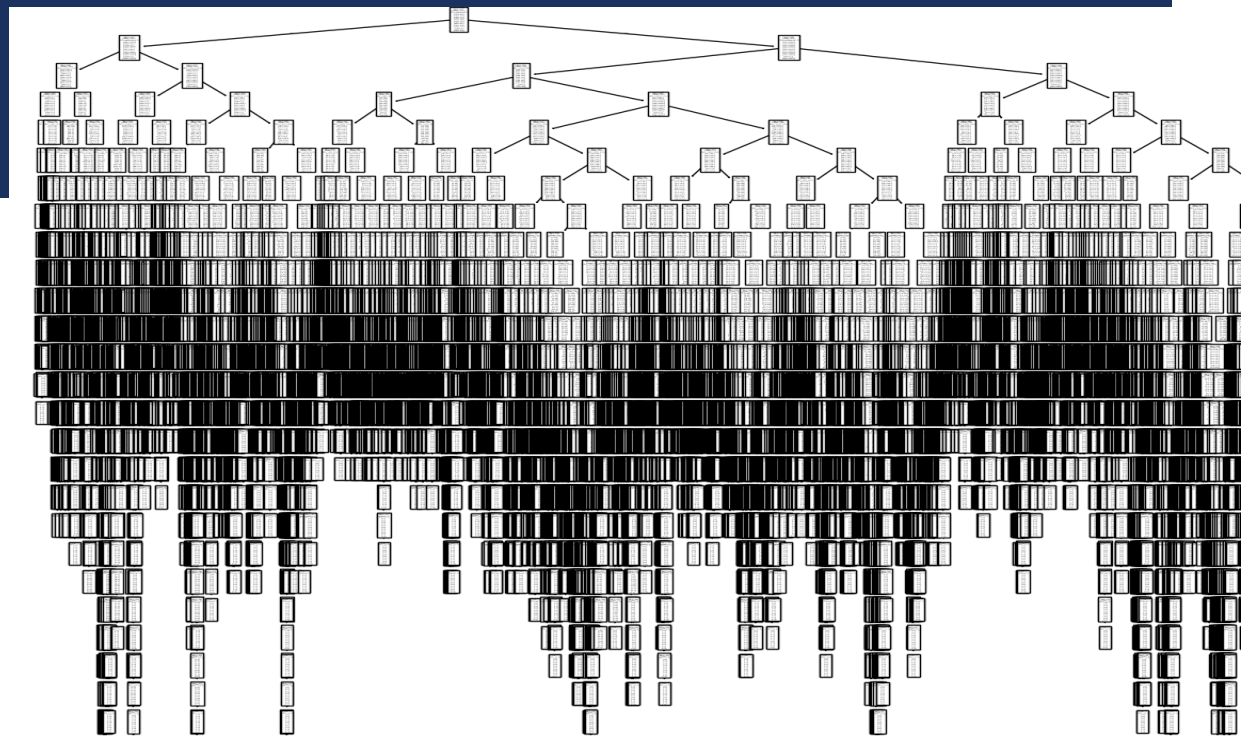
# K-NEAREST NEIGHBORS (KNN)

| Weather Station | Accurate | Predictions | False positive | False Negative | Accuracy rate |
|---|---|---|---|---|---|
| Basel | 3907 | 935 | 431 | 465 | 84% |
| Belgrade | 3238 | 1502 | 538 | 460 | 83% |
| Budapest | 3416 | 1432 | 484 | 406 | 84% |
| Debilt | 4346 | 732 | 291 | 369 | 88% |
| Dusseldorf | 4167 | 800 | 340 | 431 | 87% |
| Heathrow | 4161 | 754 | 409 | 414 | 86% |
| Kassel | 4563 | 607 | 252 | 316 | 90% |
| Ljubljana | 3726 | 1133 | 469 | 410 | 85% |
| Maastricht | 4249 | 819 | 313 | 357 | 88% |
| Madrid | 2735 | 2257 | 433 | 313 | 87% |
| Munchenb | 4222 | 766 | 324 | 426 | 87% |
| Oslo | 4624 | 507 | 255 | 352 | 89% |
| Sonnblick | 5738 | 0 | 0 | 0 | 100% |
| Stockholm | 4449 | 588 | 317 | 384 | 88% |
| Valentia | 5391 | 108 | 71 | 168 | 96% |
| | | | | Average | 88% |

- The k-nearest neighbors (KNN) algorithm makes predictions by measuring the distance between new data points and nearby data points, or neighbors, then groups the new data with nearby data points based on the number of neighbors in each group.
- KNN was used to predict pleasant weather days in 15 of the European weather stations.
- Using a k-range value of 3, the model was able to predict pleasant weather days with an average accuracy of 88%. Overfitting is likely occurring in this model.



# DECISION TREE



- The decision tree algorithm classifies data by starting with a root, or all the data, and creating branches based on features of the data.

- Predictions can be made based on new data features by following the flow of the tree starting with the root, flowing into branches that best fit the new data's features, and finding an end point, or leaf, for the data.

- The model performed with training data accuracy of 46.2% and testing data accuracy of 47.3%

- The decision tree needs to be pruned back.

```
In [12]: #What is the training accuracy score? Using the cross validation method
         y_pred_train = weather_dt.predict(X_train)
         print('Train accuracy score: ',cross_val_score(weather_dt, X_train, y_train, cv = 3, scoring='accuracy').mean())

Train accuracy score:  0.46177089441145663
```

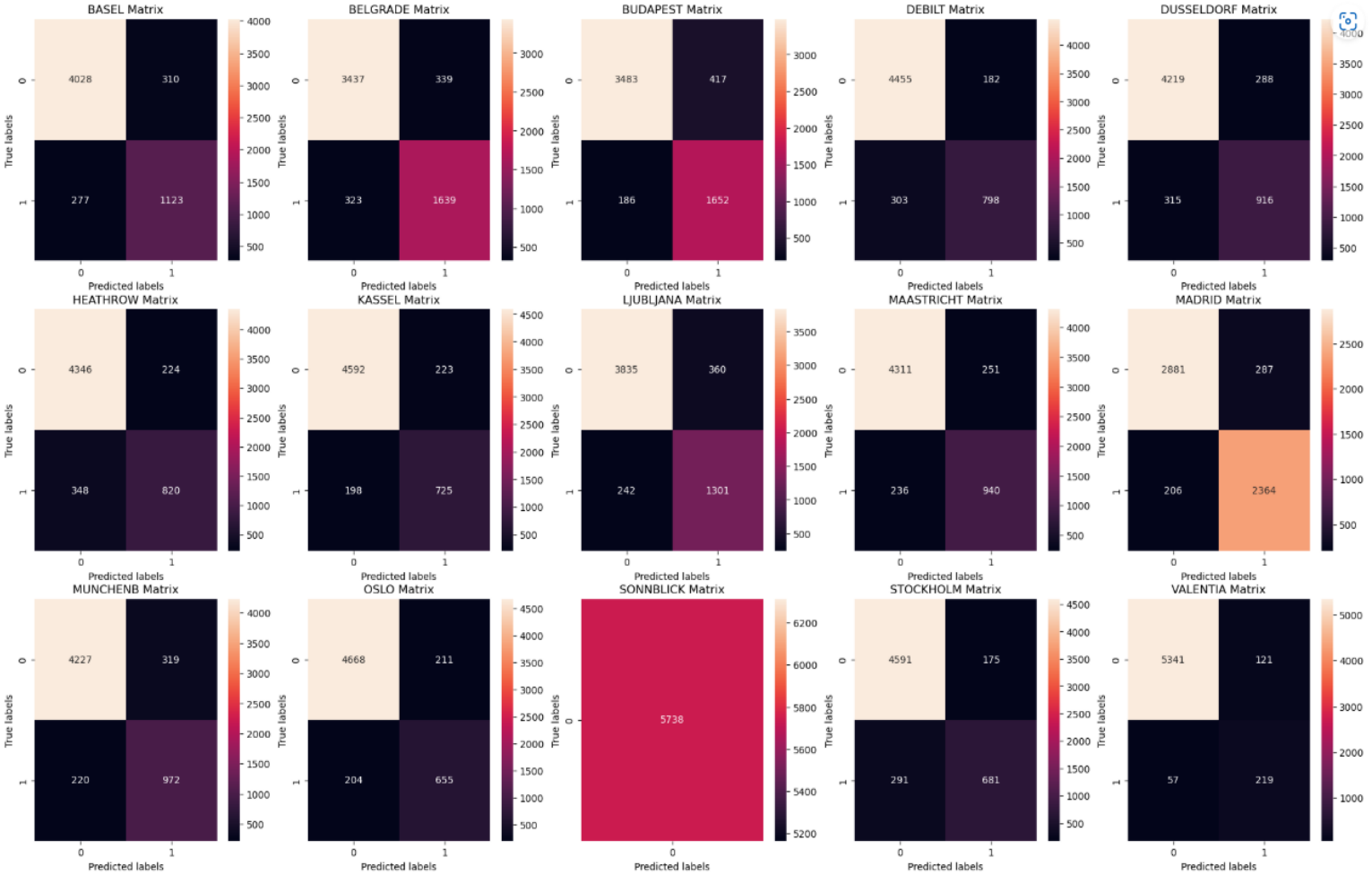
```
In [17]: #What is the testing accuracy score? Using the cross validation method
         y_pred_test = weather_dt.predict(X_test)
         print('Test accuracy score: ',accuracy_score(y_test, y_pred))
         multilabel_confusion_matrix(y_test, y_pred)

Test accuracy score:  0.4731613802718717
```

# ARTIFICIAL NEURAL NETWORK (ANN)

- Artificial Neural Network (ANN) is a computer model that mimics the way the human brain processes information.
- It consists of layers of interconnected "neurons" that learn patterns from data through training.
- By adjusting connections between neurons, it improves its ability to make predictions or classifications.
- Notice Sonnblick's (solid pink) matrix, which represents the ANN's 100% accuracy in predicting only unpleasant weather days for this location.

**Best accuracy score:** 51%




Above is a series of Confusion Matrices

Upper left quadrant = accurate unpleasant
Upper right quadrant = inaccurate unpleasant
Bottom left quadrant = inaccurate pleasant
Bottom right quadrant = accurate pleasant

# BEST FIT ALGORITHM

- The Decision tree model is too complex to interpret.
- Based on current accuracy rate,  the KNN algorithm is likely the best algorithm for predicting future weather. Thus, I would recommend Climate Wins to use the KNN model to predict if the weather is suitable for a picnic compared to the ANN model.
- The ANN does require more manual input and manipulation, but this model could be a better option if more data scenarios were available to use for pleasant and unpleasant conditions.

## SUMMARY

- Machine learning algorithms can accurately predict future weather conditions with some models with accuracy up to 88%.
- Machine learning can identify signs of climate change and its adverse effects by comparing unpleasant weather condition over time.

## NEXT STEPS

- Continue testing through supervised and unsupervised algorithm, after optimizing them.
- Further prune the decision tree for better accuracy.
- Diversify data points to include other variables for analysis.
- Observe more weather stations to increase machine learning capabilities

# THANK YOU

aarongilbert9/ClimateWins-Machine-Learning-with-Python-

For any questions, please contact me at aaron_gilbert101@yahoo.com