WEATHER CONDITIONS AND CLIMATE CHANGE WITH CLIMATEWINS

AMRIT GILL

APRIL 2025

INTRODUCTION:

ClimateWins is interested in using machine learning to help predict the consequences of climate change around Europe and, potentially, the world. The data was collected 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.

HYPOTHESIS

Machine Learning will enable highly accurate weather predictions in the future

The accuracy of weather predictions will differ from place to place because each region has unique climate characteristics.

Machine Learning can detect patterns that indicate climate change and help reveal its negative impacts.

DATA SETS

A data set based on weather observations from 18 different weather stations across Europe, which contain data ranging from the late 1800s to 2022. The records were made almost every day with values such as temperature, wind speed, snow, global radiation, and more.

The data was collected by the European Climate Assessment & Data Set project.

Weather Data Set Link

Pleasant Data Set Link

DATA BIAS & ACCURACY

❖ BIAS

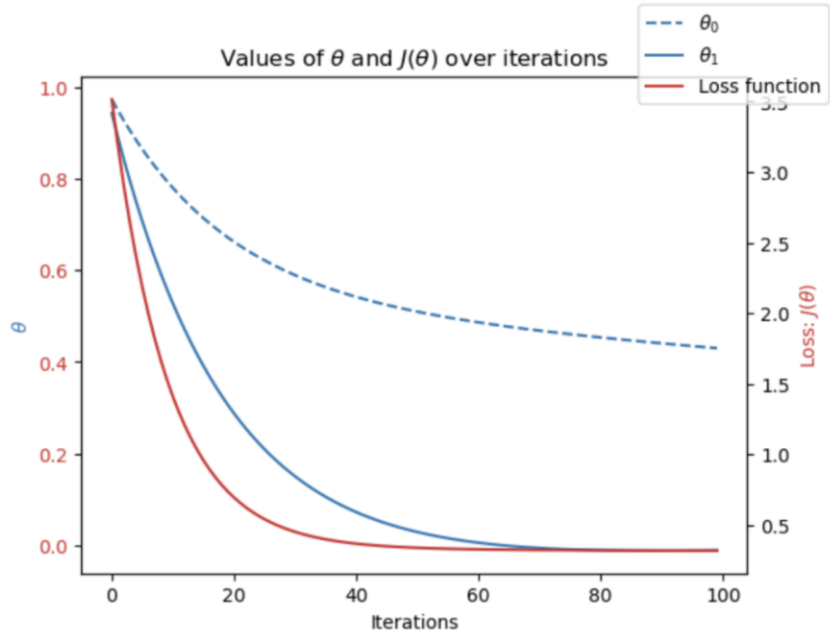
- **Data Collection:** The historical weather data comes from 18 different stations across Europe, but it may not equally represent all regions.
- **Location Bias:** Since the data is centered on Europe and the Mediterranean, the predictions may not be applicable to other regions—such as Brazil or Canada—because they have different climate systems.
- **Human Bias**: Older measurement instruments were less precise, making historical data less reliable.
- **Sampling Bias**: Specific selection of stations can skew results, missing broader climate patterns across Europe.

❖ ACCURACY

- The accuracy of machine learning models depends on the algorithms and optimization technique applied.

DATA OPTIMIZATION

- The data was optimized to find the minimum error, through number of iterations, as well as the number of steps (alpha), which varied as the case through Gradient Descent.
- It can be used in both linear and non-linear models.
- Iterative optimization was able to minimize the loss function to nearly zero, after adjusting the theta0 and theta1, as well as the number of iterations and alpha.



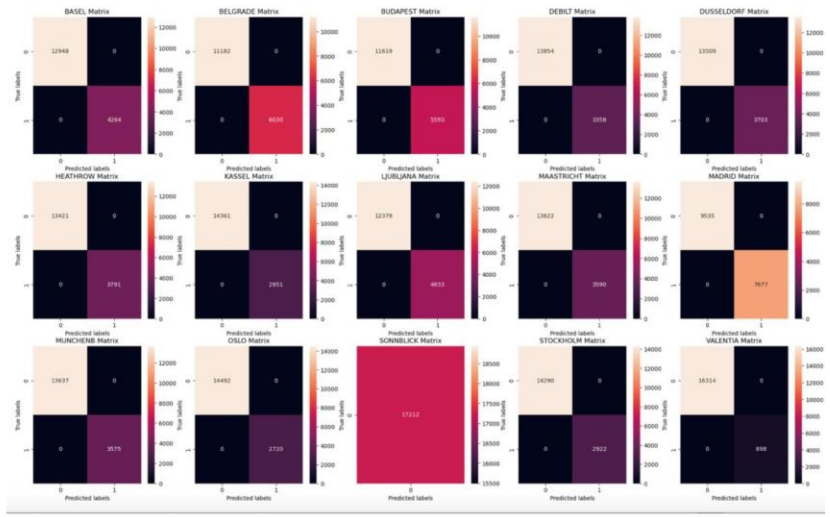
METHOD 1. K-NEAREST NEIGHBOR

Weather Station	Accurate Prediction		False Positive	False Negative	Accurate Rate
Basel	3888	933	450	467	84%
Belgrade	3265	1491	511	471	83%
Budapest	3423	1425	477	413	84%
Debilt	4338	690	299	411	88%
Dusseldorf	4158	808	349	423	86%
Heathrow	4187	752	383	416	86%
Kassel	4561	600	254	323	90%
Ljubljana	3762	1117	433	426	85%
Maastricht	4234	785	328	391	87%
Madrid	2765	2226	403	344	87%
Munchenb	4197	753	349	439	86%
Oslo	4645	524	234	335	90%
Sonnblick	5738	0	0	0	100%
Stockholm	4482	622	284	350	88%
Valentia	5382	112	80	164	95%
				Average	88%

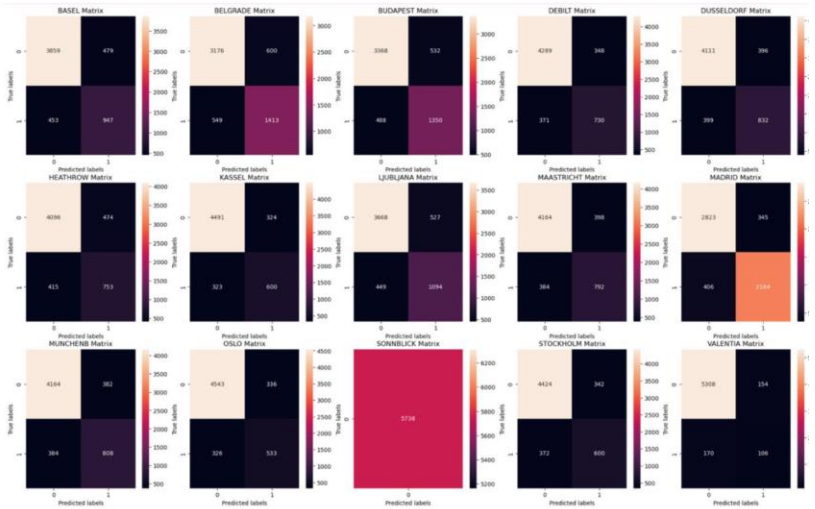
- The k-nearest neighbors (KNN) algorithm makes predictions by measuring the distance between new data points and nearby data points, or neighbors, then groups the new data with nearby data points based on the number of neighbors in each group.
- Best Accuracy: Sonnblick (100%) had perfect accuracy, with zero false positives and false negatives. Valentia (95%), Oslo (90%), and Kassel (90%) also showed very high accuracy. Belgrade (83%) had the lowest accuracy, indicating possible challenges in prediction accuracy for this station.
- Using a k-range value of 3, the model was able to predict pleasant weather days with an average accuracy of 88%.
- To improve forecasting accuracy, the model requires training on diverse datasets, better handling of varied weather patterns, and rigorous cross-validation, enhancing adaptability for broader, reliable, and practical weather predictions.

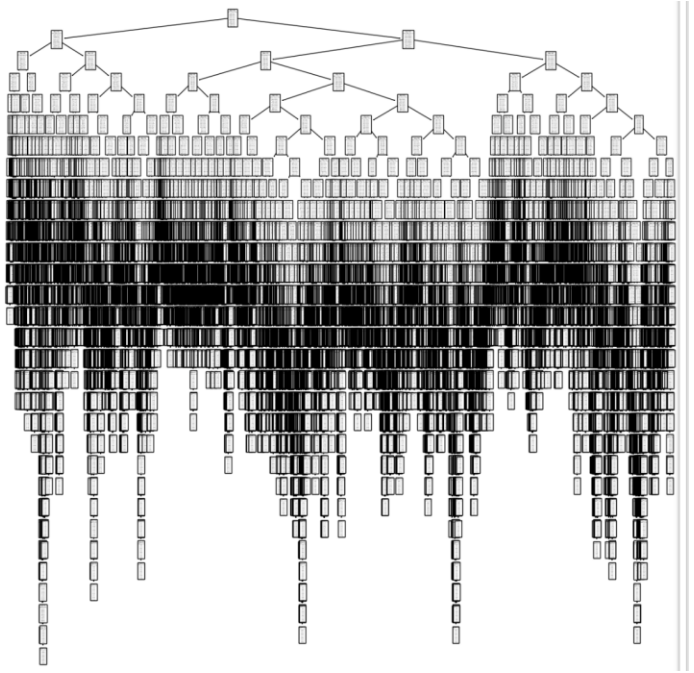
METHOD 2: DECISION TREE

Decision Tree Training Set:



Decision Tree Test Set:

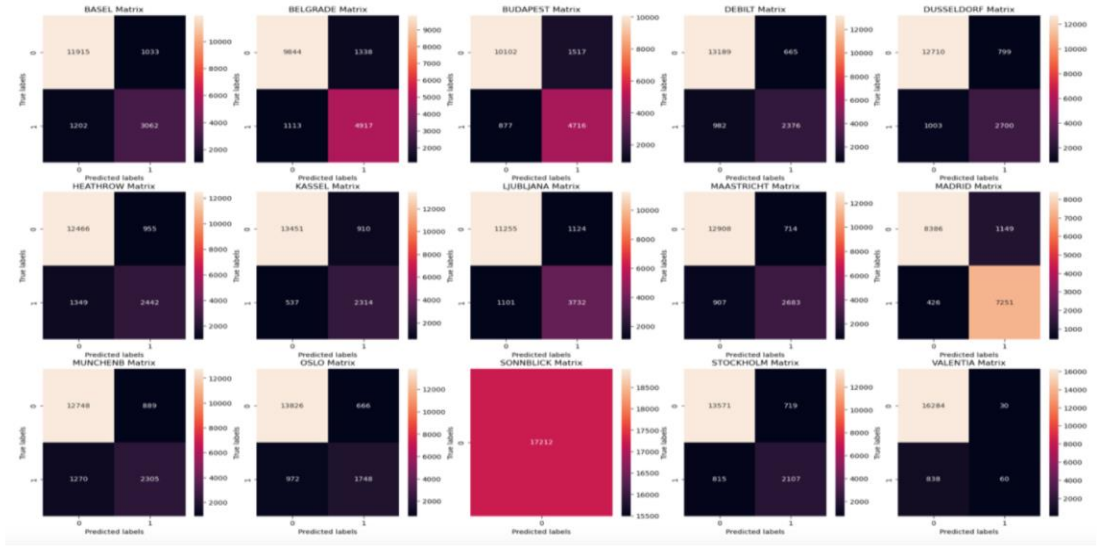




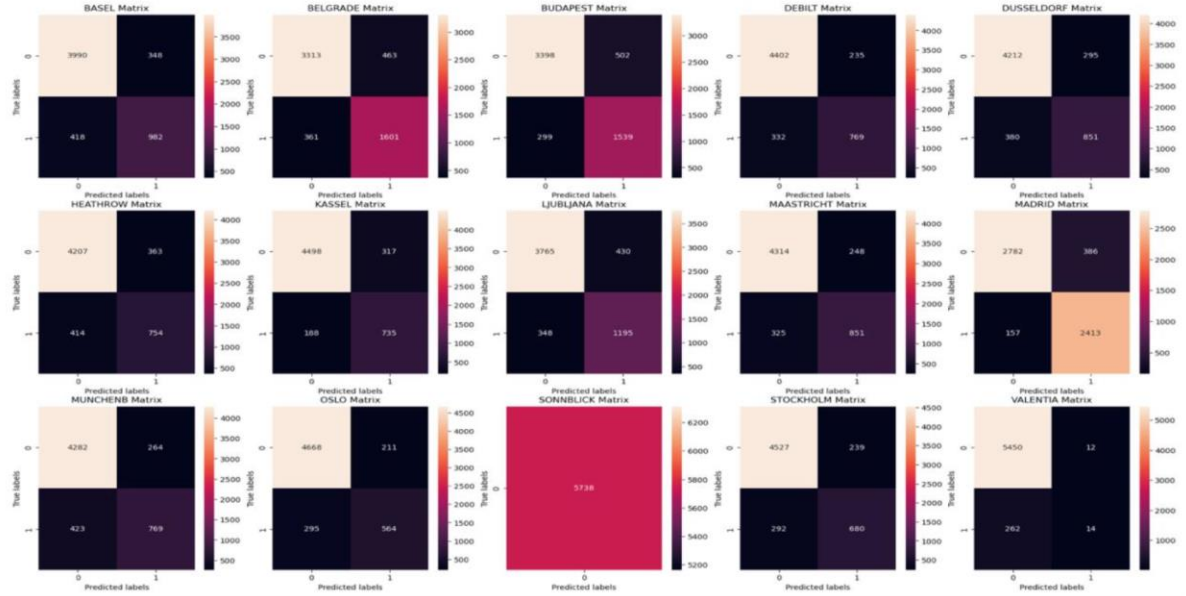
- The decision tree model currently illustrates poor accuracy, with only 46% on the training data and 47% on the test data.
- The decision tree algorithm breaks down the classification process into a series of "yes" or "no" questions based on the features of the data. Starting from the root node, the algorithm splits the data at each branch based on the most relevant feature, continuously refining the classification.
- The decision tree is very complex with too many branches which can lead to a model being too difficult to interpret. The tree is also very large, which can take a long time to train.

METHOD 3: ARTIFICAL NEURAL NETWORK (Scenario 1)

Training Data Set: Accuracy for this set is 47%.



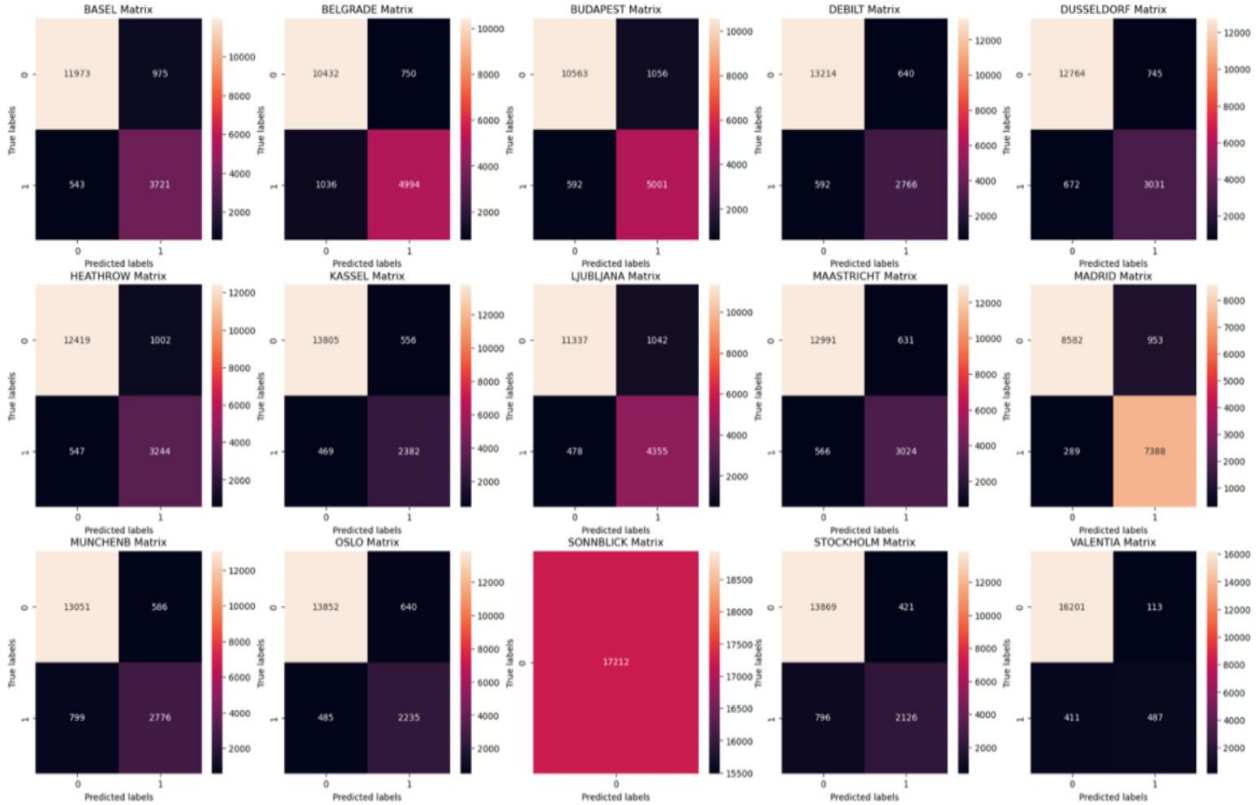
Test Data Set: Accuracy for this set is 47%.



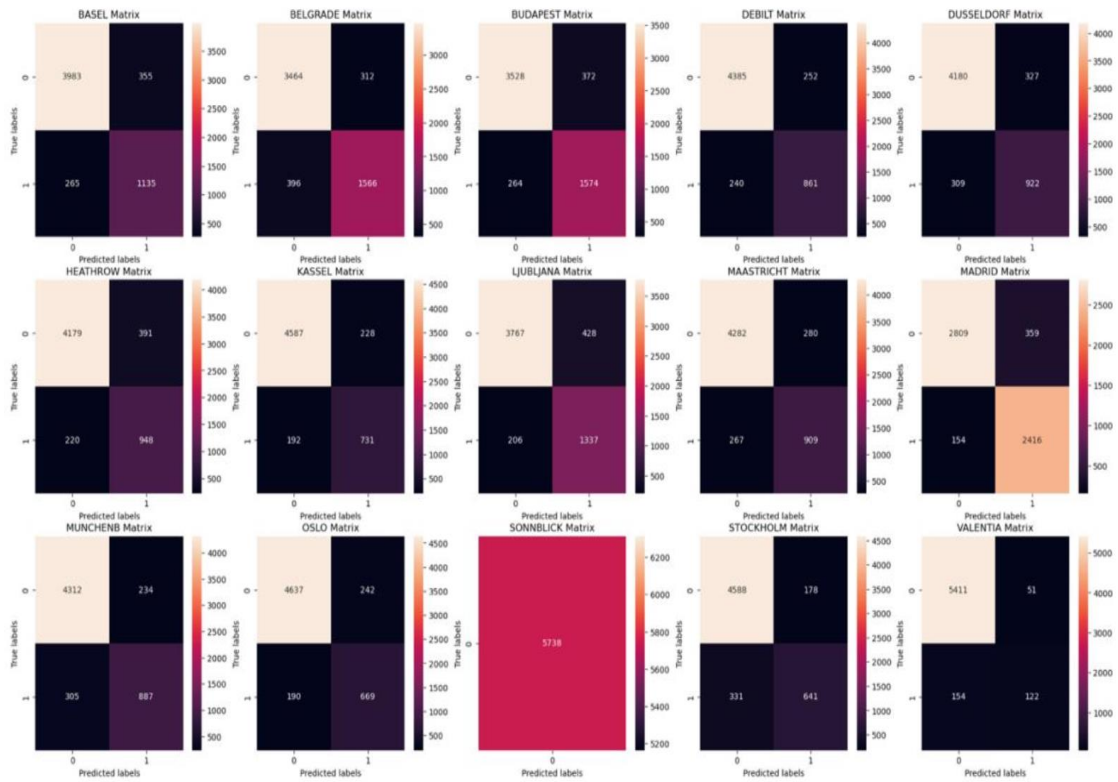
- Artificial Neural Networks (ANNs) make predictions by processing data through a set of layers.
- Upper left = accurate unpleasant, Upper right = inaccurate unpleasant.
- Bottom left = inaccurate pleasant, Bottom right = accurate pleasant.

METHOD 3: ARTIFICAL NEURAL NETWORK (Scenario 2)

Training Data Set: Accuracy is 54%.

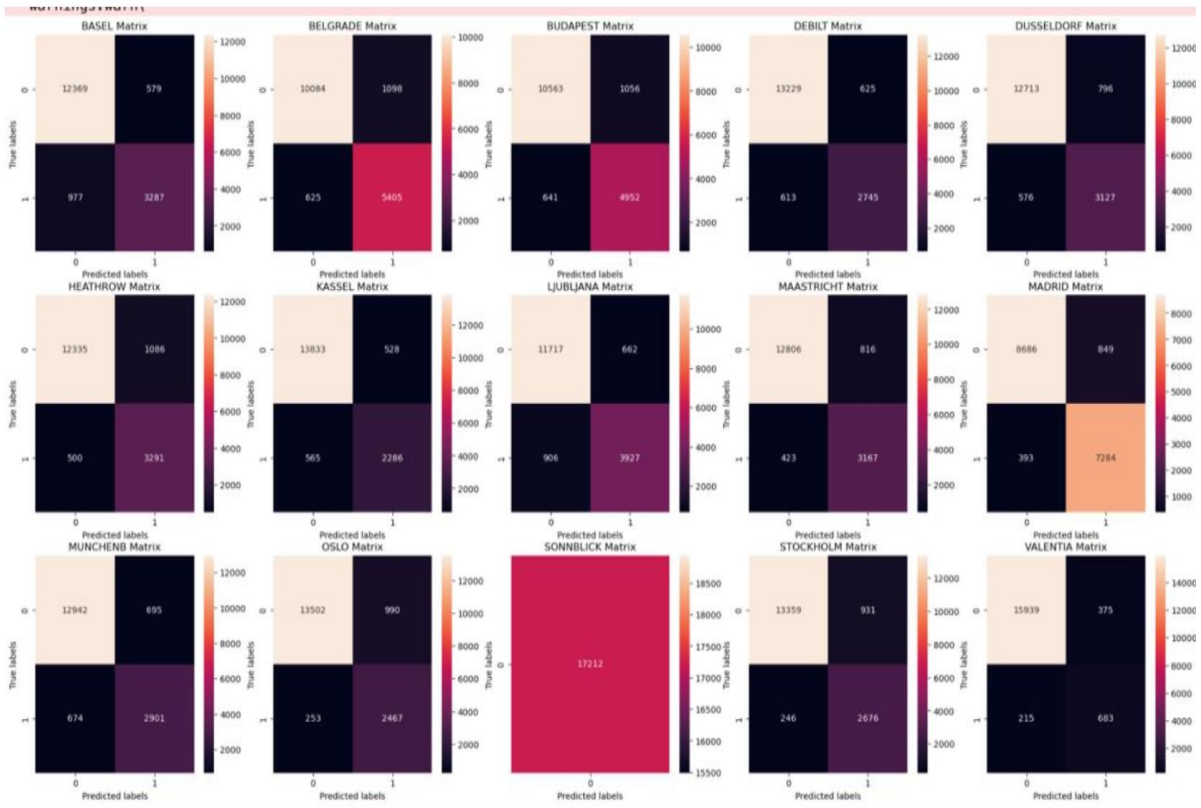


Test Data Set: Accuracy for this set is 49%.

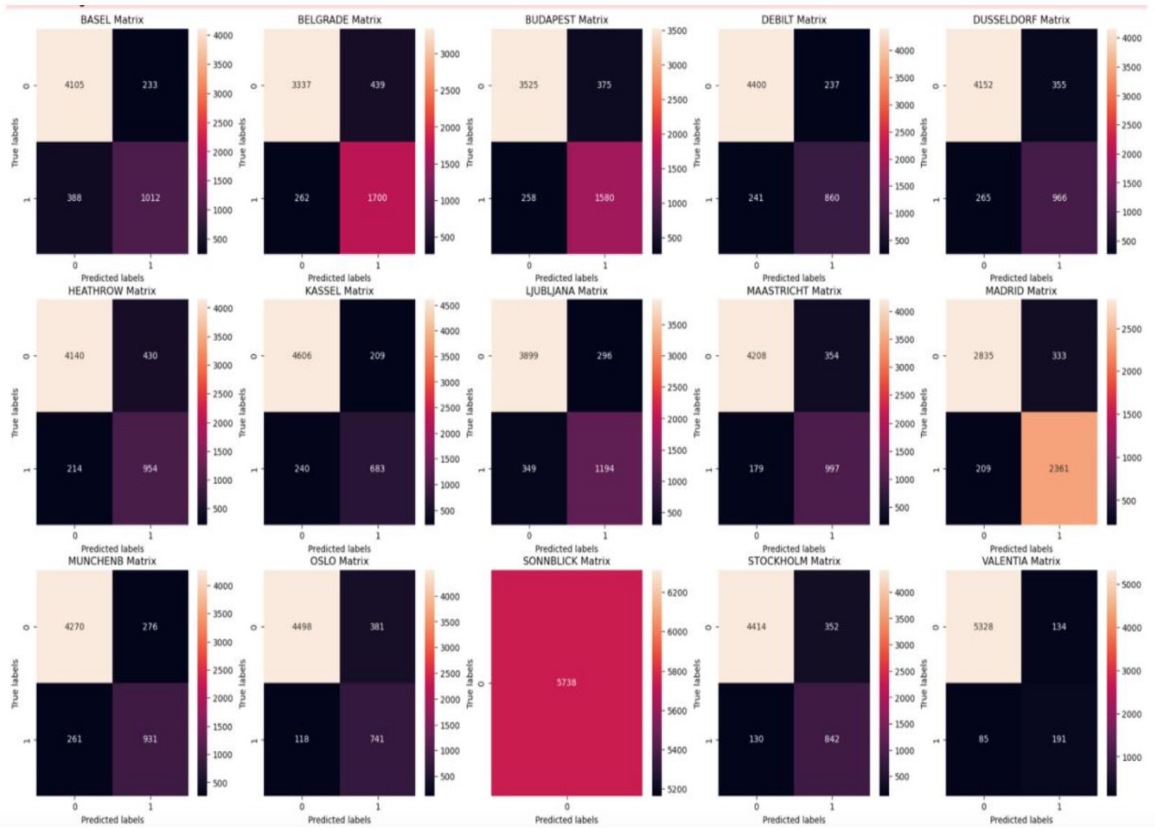


METHOD 3: ARTIFICAL NEURAL NETWORK (Scenario 3)

Training Data Set: Accuracy is 53%.



Test Data Set: Accuracy is 48%.



WHICH ALGORITHM IS BEST?

- Based on model accuracy, **K-Nearest Neighbors (KNN)** is the most effective algorithm for ClimateWins to provide accurate predictions of good picnic weather since it effectively identifies patterns and makes reliable predictions.
- On the other hand, The **decision tree** model is too complex to interpret. The **ANN model** is more powerful and better suited for complex tasks.

CONCLUSION & NEXT STEPS

- The KNN algorithm predicted the data the best, with an average accuracy of 88% on the test set.
- The decision tree still needs some adjustments (like pruning) to work better.
- The ANN (artificial neural network) was not very reliable and only got about 46% accuracy, which is much lower than KNN.
- Since the project isn't finished yet, KNN is the best choice for now, because it did the best job at predicting the climate temperature.
- Additionally, Keep testing both supervised and unsupervised algorithms after better optimizing them.
- Try to improve the decision tree even more by pruning it further to increase accuracy.
- In the end, combine supervised and unsupervised methods to build a more complete climate model — one that can predict both specific weather events and long-term climate trends.

GILLAMRIT03@OUTLOOK.COM

HTTPS://GITHUB.COM/GILLAMRIT92/CLIMATEWINS