CLIMATEWINS MACHINE LEARNING INTERIM REPORT

Presented by Josh Wattay
August 19, 2024

AHEAD OF THE CURVE

Machine learning makes an impact in our climate data.

CLIMATE WINS VISION

Cimate Wins 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.

EXPLORATORY QUESTIONS

How is machine learning used?

Machine Learning employs algorithms to automate tasks and foster interactive experiences between technology and users by analyzing vast troves of data.

Is it applicable to weather data?

Machine learning has been applied to weather data for over 25 years, with uses in weather and climate modeling. By identifying patterns in historical data, ML holds promise to enhance weather forecasting with greater precision.

ClimateWins has heard of ethical concerns surrounding machine learning and AI. Are there any concerns specific to this project?

The accuracy and reliability of AI algorithms are directly tied to the quality and representativeness of the training data. If the data used to train these algorithms is biased, it can lead to biased decision-making. This is a significant concern, as AI algorithms are only as unbiased as the data they are trained on.

EXPLORATORY QUESTIONS (CONTINUED)

Historically, what have the maximums and minimums in temperature been?

The maximum recorded temperature with this data was 43.6° C on July 24, 2007.

The minimum recorded temperature was -34.3° C on January 13, 1968.

Can machine learning be used to predict whether weather conditions will be favorable on a certain day? (If so, could it be possible to predict danger.)

Yes, ML can be used to predict weather conditions as favorable or dangerous on a certain day. As will be demonstrated later in this presentation, our K Nearest Neighbors model predicts pleasant weather with an 88% rate of accuracy. While not perfect, with more iterations and training of this model, it is reasonable to assess that weather conditions can be predicted by ML models.

AGENDA

Objective & Hypotheses

Addressing Data Source & Bias

Optimization Techniques

Supervised Machine Learning

Recommendation

OBJECTIVE

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

HYPOTHESES

Machine Learning aided the evaluation of these hypotheses:

1. The average temperature has changed over time since the weather stations started collecting data and we can accurately forecast that change for the future.
2. It is possible to predict how many pleasant days and unpleasant days of weather a location will experience.
3. Patterns between the variables can be investigated further in order to discover unknown or hidden correlations and relationships.

DATA SOURCE AND BIAS

The data used for this report was provided by the European Climate Assessment & Dataset. The data sets can be downloaded by using this link.

There are regional and cultural biases in climate change that can be exacerbated by the use of machine learning. If the model is not trained properly and predicts a rapid increase in temperature in the next 12-24 months, then there is a lot at stake for both those who believe strongly in the impacts of climate change (confirmation bias) and those who deny it (falsification bias).

The data in this set is also subject to temporal and location bias, since it was only gathered in Europe and for a period of time that may no longer be representative of the present.

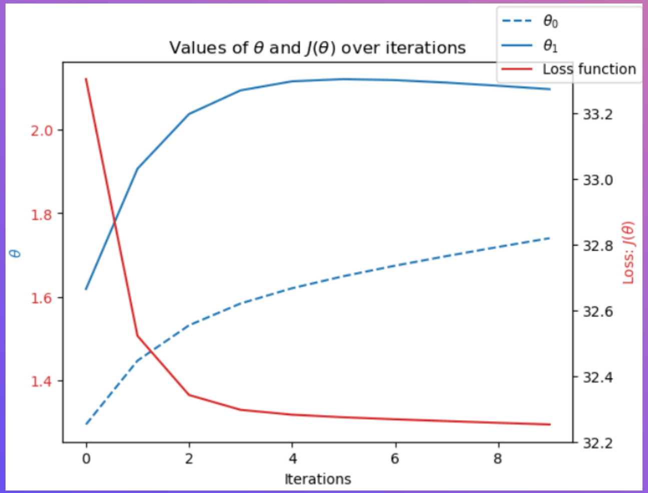
OPTIMIZATION TECHNIQUES

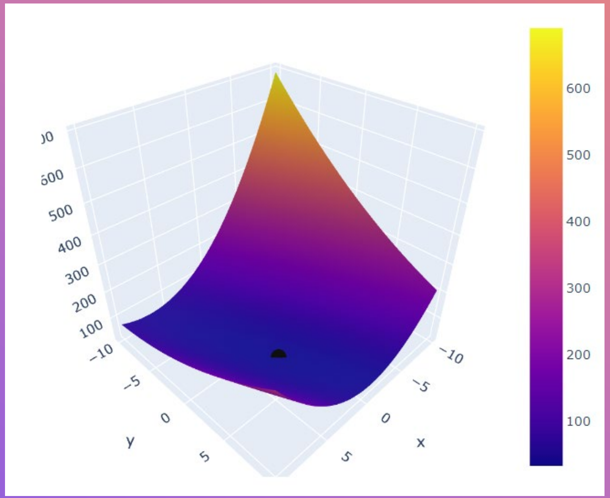
Gradient Descent Method

The gradient descent algorithm was used to optimize this data set. Gradient descent is a simple technique for finding a local minimum (or valley) that can be applied to both linear and nonlinear cases.

We leveraged gradient descent to minimize the error by adjusting the number of iterations and the step size (alpha), which were varied as needed. By tuning the theta0 and theta1 parameters, as well as the number of iterations and step size, we were able to achieve a result very close to 0.

GRADIENT DESCENT RESULTS





Example of Gradient Descent Results : Oslo 1960

Weather Station	Year	Theta0	Theta1	Iterations	Step Size
Roma	1960	5	5	10	0.1
Roma	1985	5	5	10	0.1
Roma	2000	5	5	10	0.1
Budapest	1960	4	3	10	0.1
Budapest	1985	4	3	10	0.1
Budapest	2000	4	4	10	0.1
Oslo	1960	2.1	1.7	10	0.1
Oslo	1985	1.7	1.1	10	0.1
Oslo	2000	2	2	10	0.1

ACCURACY

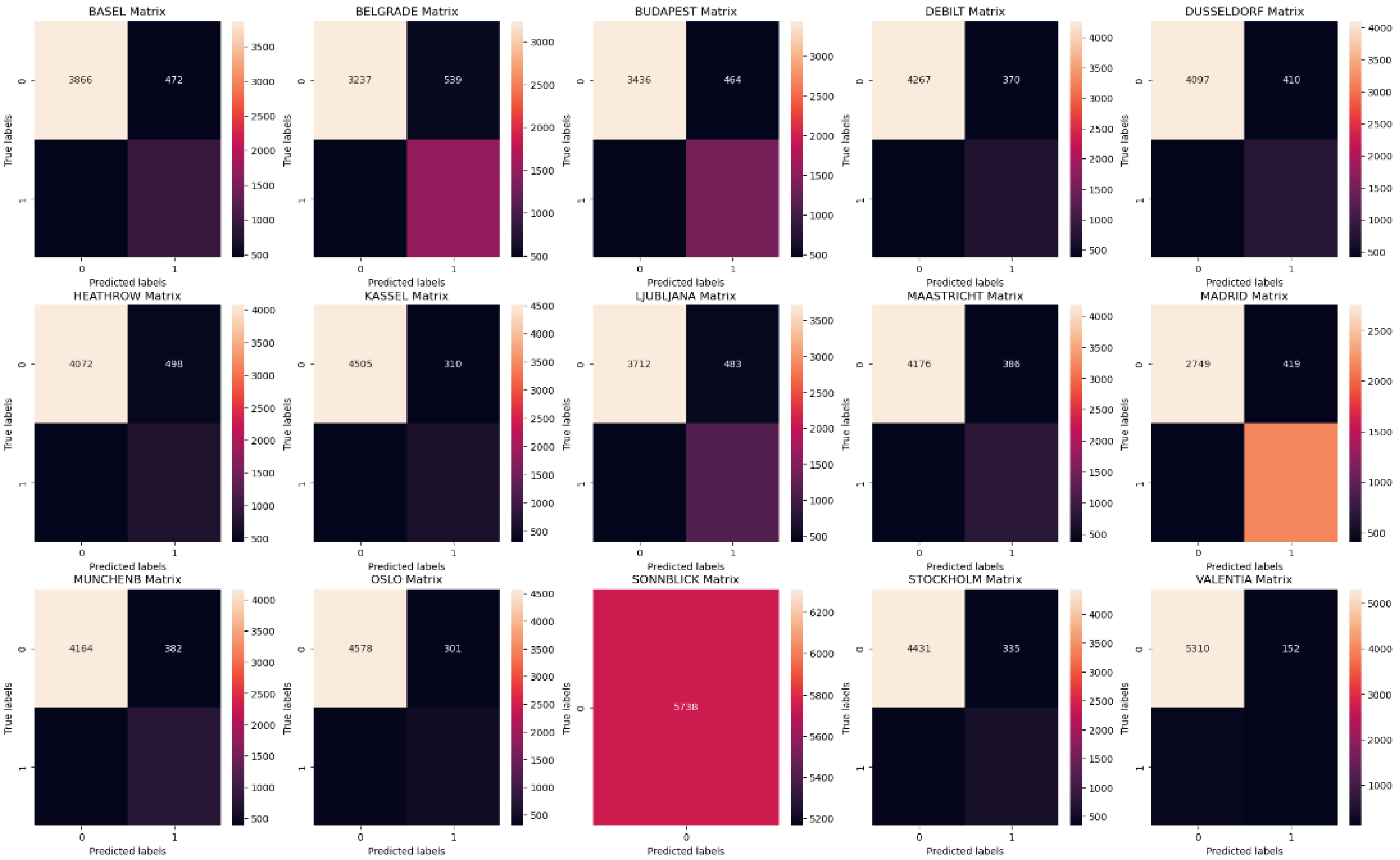
Three supervised machine learning models (K Nearest Neighbor (KNN), Decision Tree (DT), and Artificial Neural Networks (ANN) were introduced in order to determine the most accurate model for the data.

The K Nearest Neighbor method has been the most accurate model for predicting pleasant weather conditions at this point in the testing.

The Sonnblick weather station has been overfitted with 100% accuracy. I believe this is due to the location of the weather station being on the top of the Austrian Alps. This is the only station with 100% accuracy at the moment.

The absence of Precipitation, Cloud Cover, and Wind Speed may have an impact on the accuracy of the model. For example, if the temperature is within what the model considers to be a pleasant range, but on that same date there was 3” of rainfall, then the model is not accounting for that variable and predicting that this day is pleasant, even though it was not due to the rain.

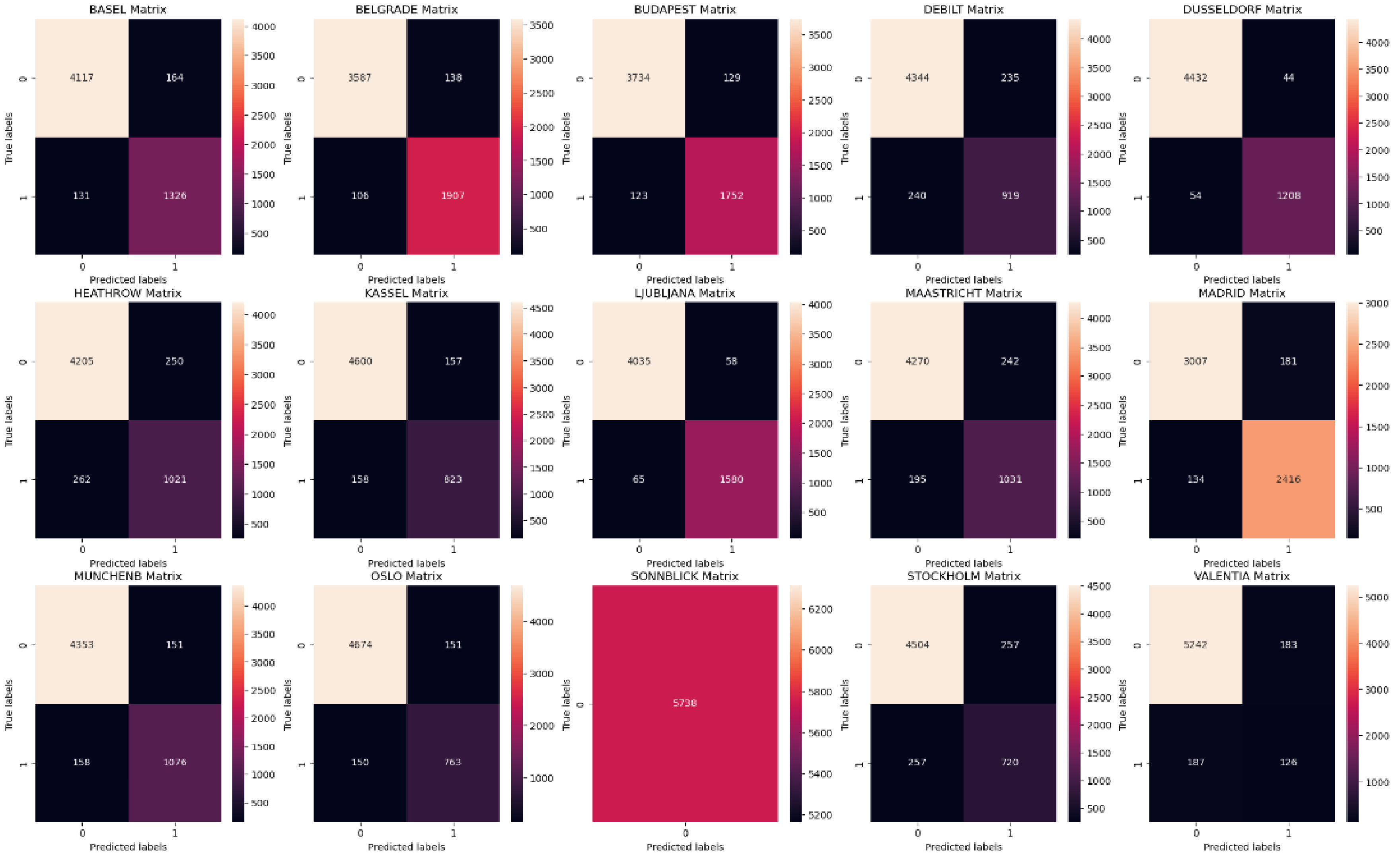
K NEAREST NEIGHBOR (KNN)



K NEAREST NEIGHBOR (KNN)

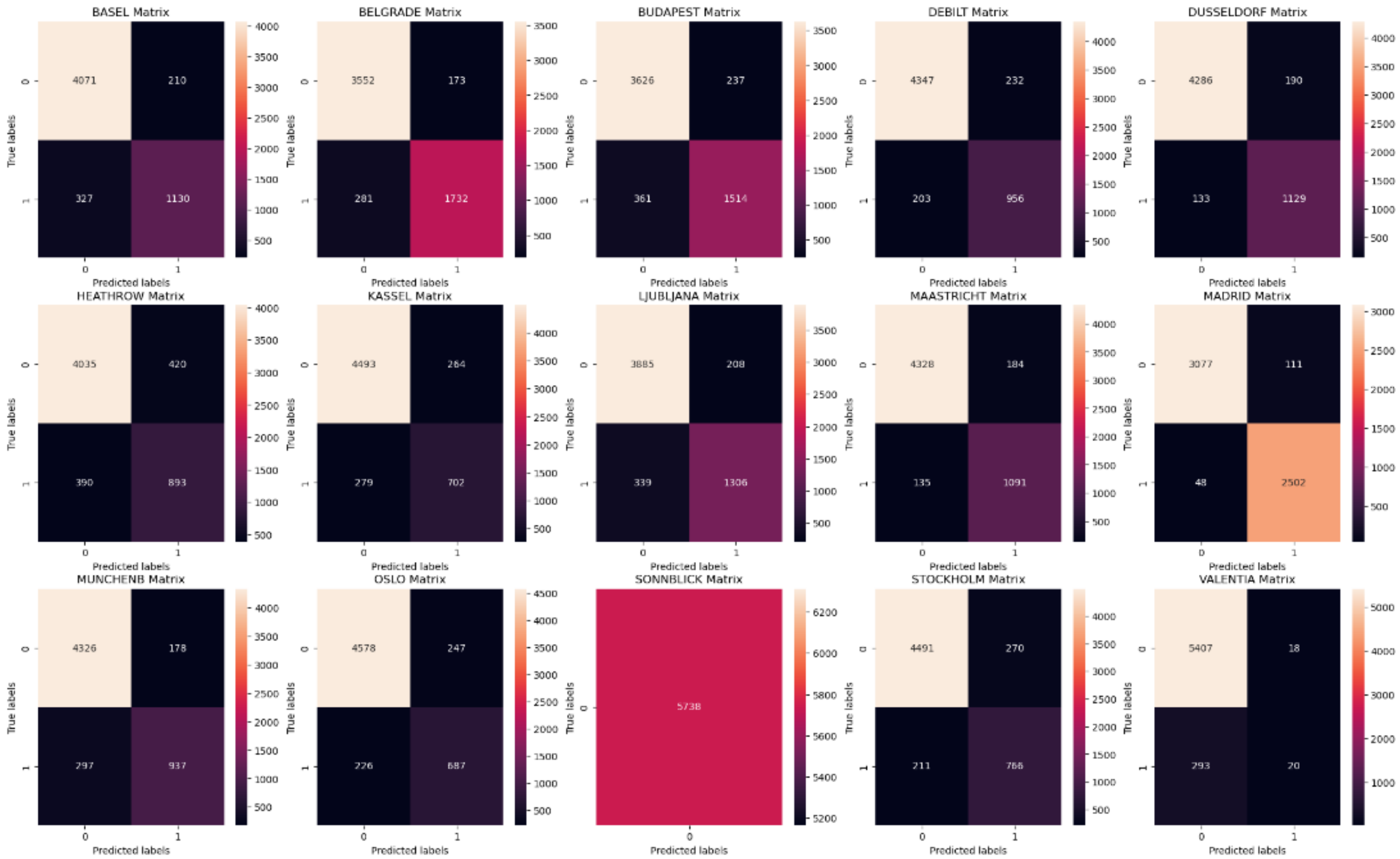
Station	Accurate Prediction		False (+)	False (-)	Accuracy
BASEL	3866	949	472	451	84%
BELGRADE	3237	1482	539	480	82%
BUDAPEST	3436	1361	464	477	84%
DEBILT	4267	716	370	385	87%
DUSSELDORF	4097	798	410	433	85%
HEATHROW	4072	728	498	440	84%
KASSEL	4505	590	310	333	89%
LJUBLJANA	3712	1122	483	421	84%
MAASTRICHT	4176	803	386	373	87%
MADRID	2749	2163	419	407	86%
MUNCHENB	4164	771	382	421	86%
OSLO	4578	500	301	359	88%
SONNBLICK	5378	N/A	N/A	N/A	100%
STOCKHOLM	3875	530	891	442	77%
VALENTIA	3588	147	1874	129	65%

DECISION TREE (DT)



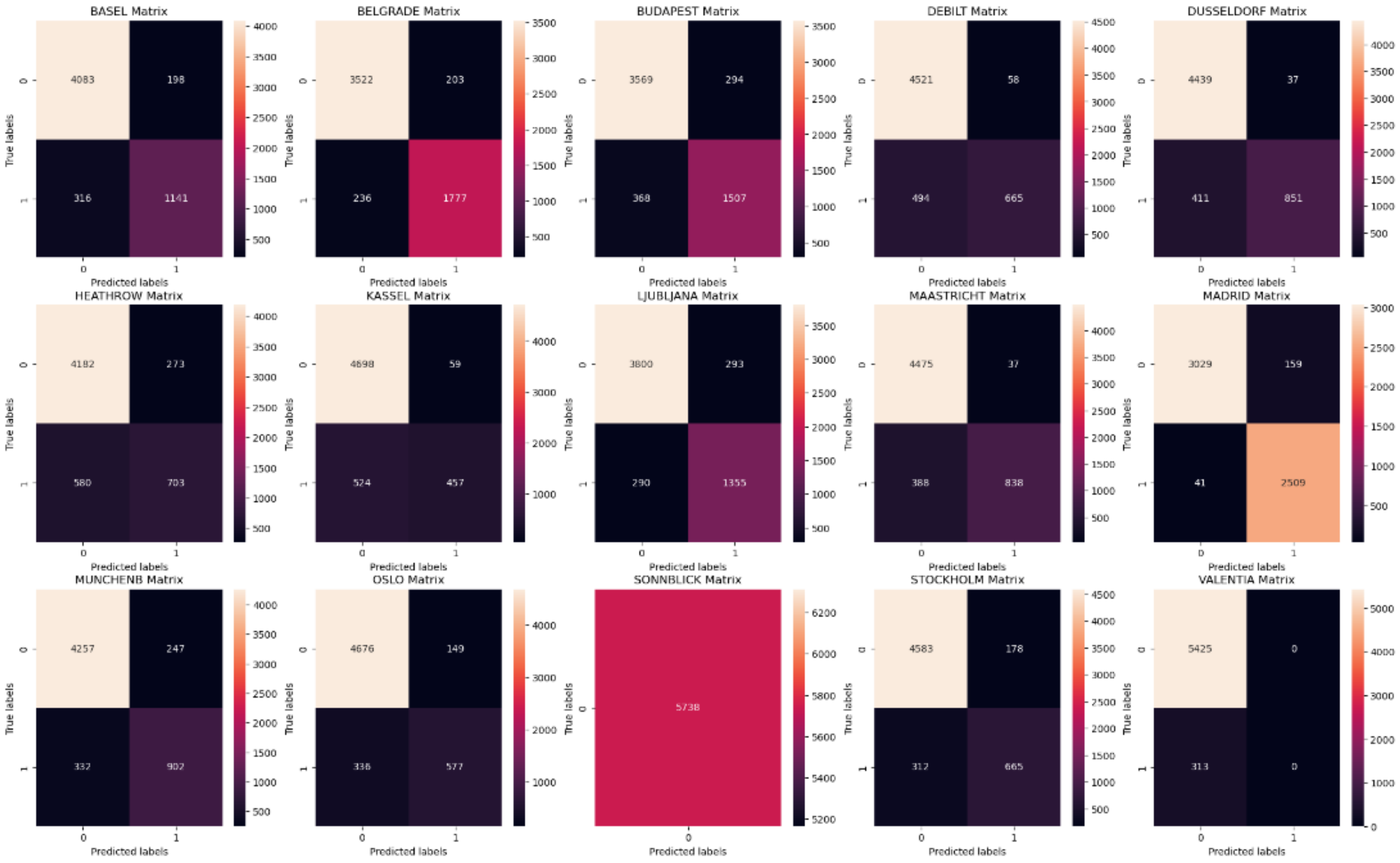
ARTIFICIAL NEURAL NETWORK (ANN)

Scenario 1 - Confusion Matrix – mlp = MLPClassifier(hidden_layer_sizes=(15, 10, 5), max_iter=1000, tol=0.00001)



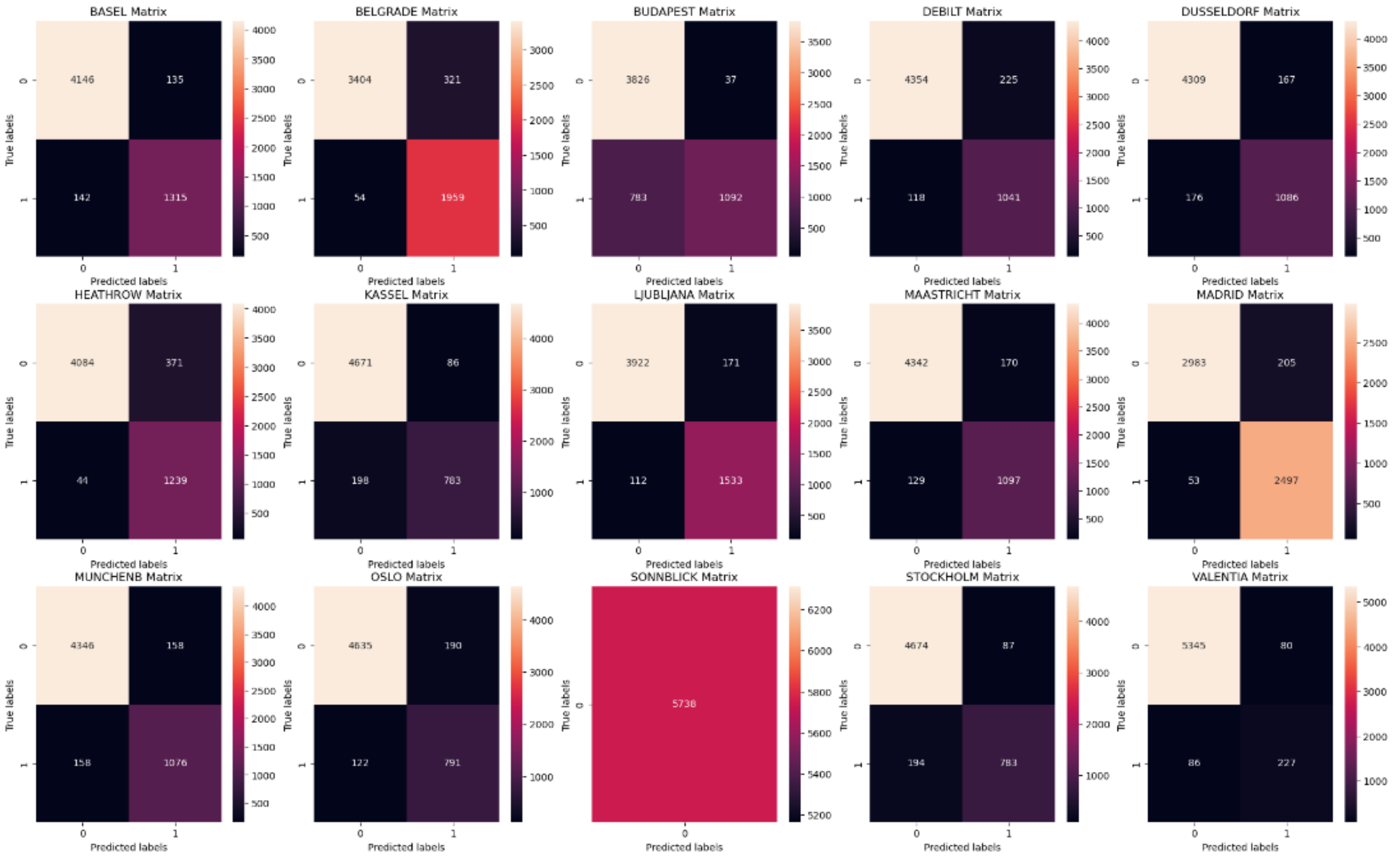
ARTIFICIAL NEURAL NETWORK (ANN)

Scenario 2 Confusion Matrix - mlp = MLPClassifier(hidden_layer_sizes=(20, 10, 5), max_iter=1500, tol=0.00001)



ARTIFICIAL NEURAL NETWORK (ANN)

Scenario 3 Confusion Matrix - mlp = MLPClassifier(hidden_layer_sizes=(100, 50, 25), max_iter=5000, tol=0.0000000001)



RECOMMENDATION

I recommend that Climate Wins use the KNN model for predicting pleasant weather. It has performed the best, is relatively efficient and inexpensive compared to the Decision Tree and Artificial Neural Network (ANN) models, and if we account for other variables moving forward, it should be able to perform above 95% accuracy.

With only a 60% train accuracy and 53% test accuracy, the Decision Tree needs more criteria to more accurately classify pleasant and unpleasant weather patterns.

With more iterations the Artificial Neural Network (ANN) model does have the ability to become quite accurate (>95%) however, this will require significantly more resources than the KNN model.

THANK YOU

Josh Wattay

jawattay@gmail.com