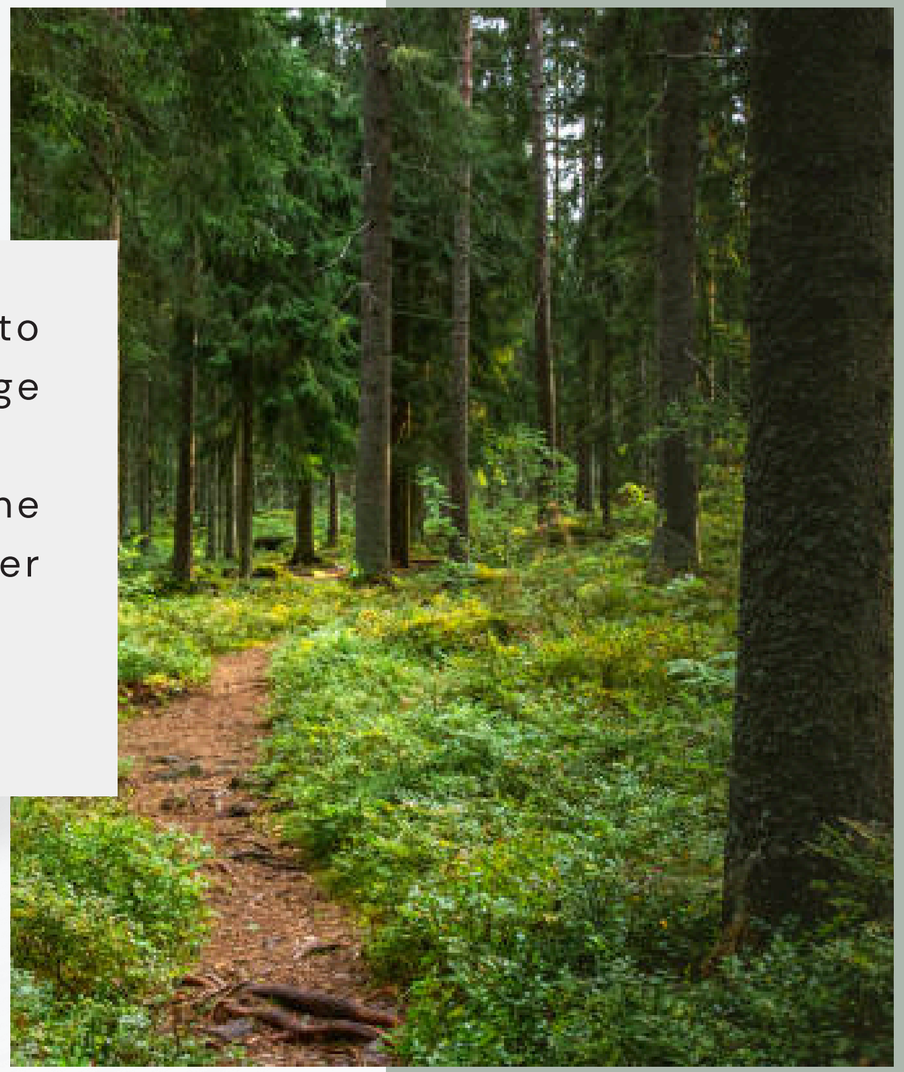
Nuria Miquel - September 2024

CLIMATE WINS

Weather Conditions and Climate Change

INTRODUCTION

Cimate Wins is interested in using machine learning to help predict the consequences of climate change around Europe and, potentially, the world.
For this project, ClimateWins aims to use Machine Learning to analyze data collected from Weather Stations across Europe for decades.



HYPOTHESIS

- Some areas of Europe are more prone to catastrophic climate change consequences than others. Through the use of Machine learning, we can get ahead and identify them.
- Countries in Southern Europe will have more records of pleasant weather.
- Supervised ML will help predict what type of atmospheric conditions can strike harder in each region.

DATASET

The data used for this project consists of two different datasets based on weather observations from 18 different weather stations from 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.

Temperature Data Set

Pleasant Data Set

DATA BIAS

Collection bias

Not all weather stations have been equiped with up-to-date systems for the whole of the years that data has been collected. There is a big probability that some regions developed their technology greater than others, which can ultimately lead to bias in the data collection process.

Not all weather stations collect the same information. Some atmospheric conditions are not measured by all the stations, which can lead to skewed analyses.

Sample Bias

Only 18 of the 23775 existing weather stations across Europe were featured in this dataset. This decision might lead to biases and it is important to assess why these were the ones selected.

OPTIMIZATION

The purpose of optimization is to **lower the risk of error** and improve the accuracy of a model.
For this project I optimized the data using the **Gradient Descent** algorithm, this helps minimize the cost function in the model.

SUPERVISED MACHINE LEARNING

KNN

K-Nearest Neighbours (KNN) was applied to determine the category of each data point by evaluating the number of neighboring points that belonged to each class.

The average accuracy was **88%**

DECISION TREE

A Decision Tree was employed to refine the solution by splitting the data into more specific subsets, leading to a decision or classification based on the most informative features.

The accuracy was **~60%**

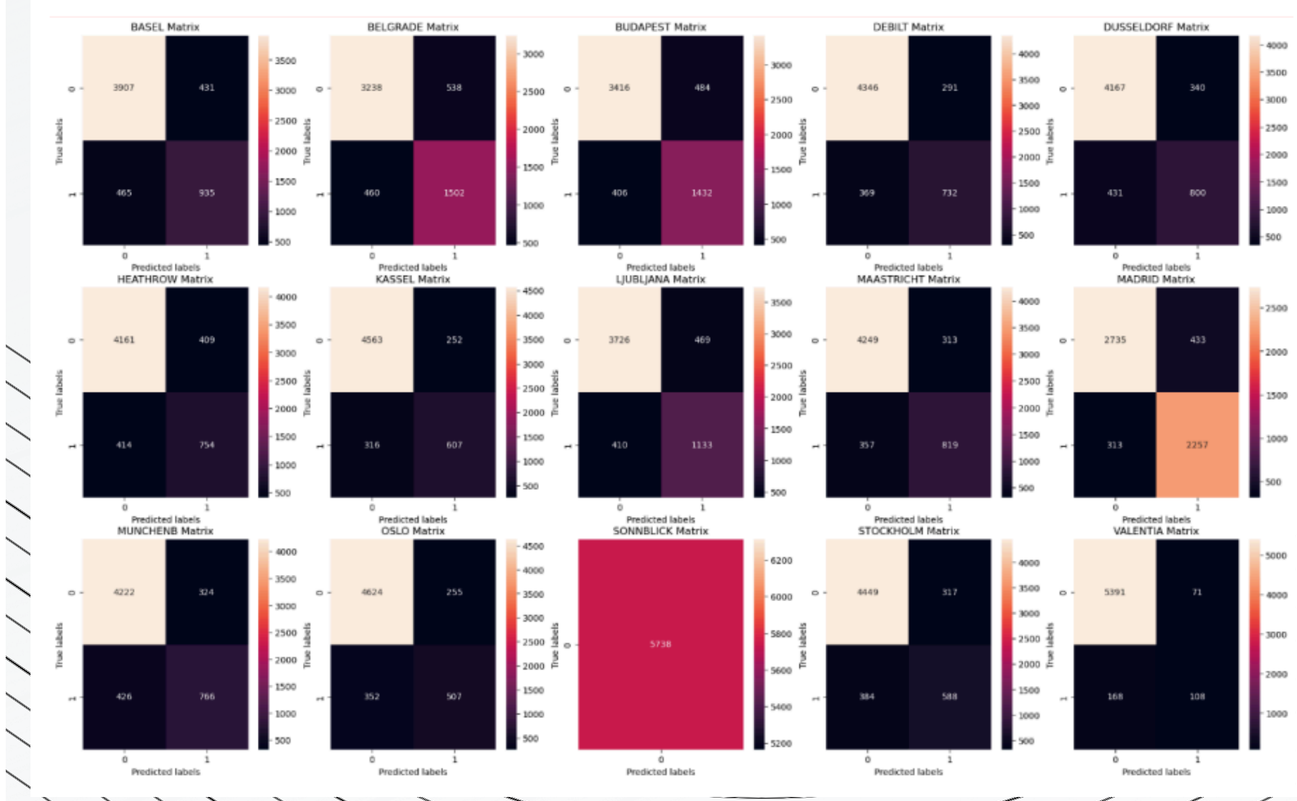
ANN

An Artificial Neural Network (ANN) processesed inputs through interconnected layers of neurons, where each input was multiplied by learned weights.

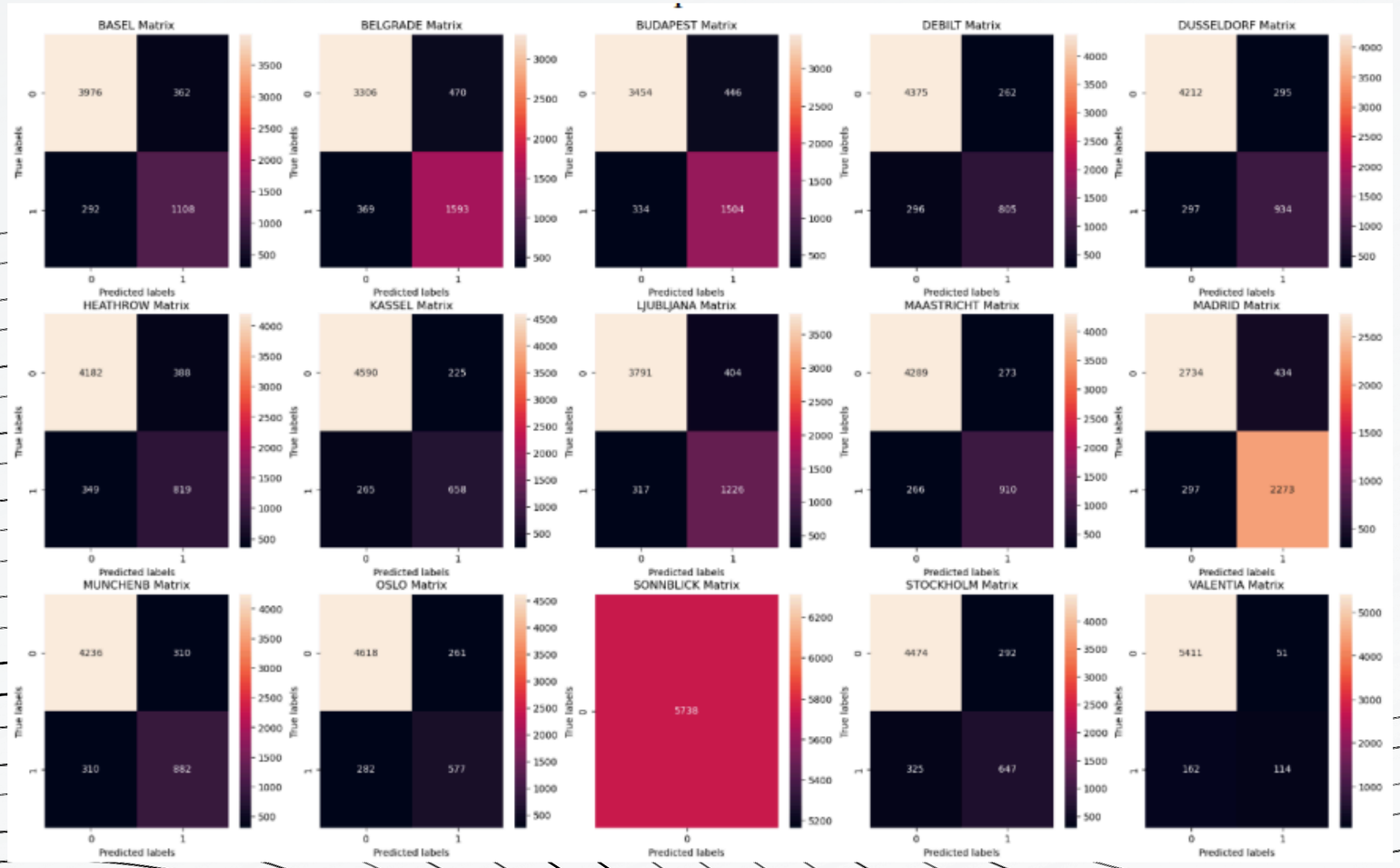
The accuracy was **~72% (Train)**
~63% (Test)

K-NEARSEST NEIGHBOURS(KNN)

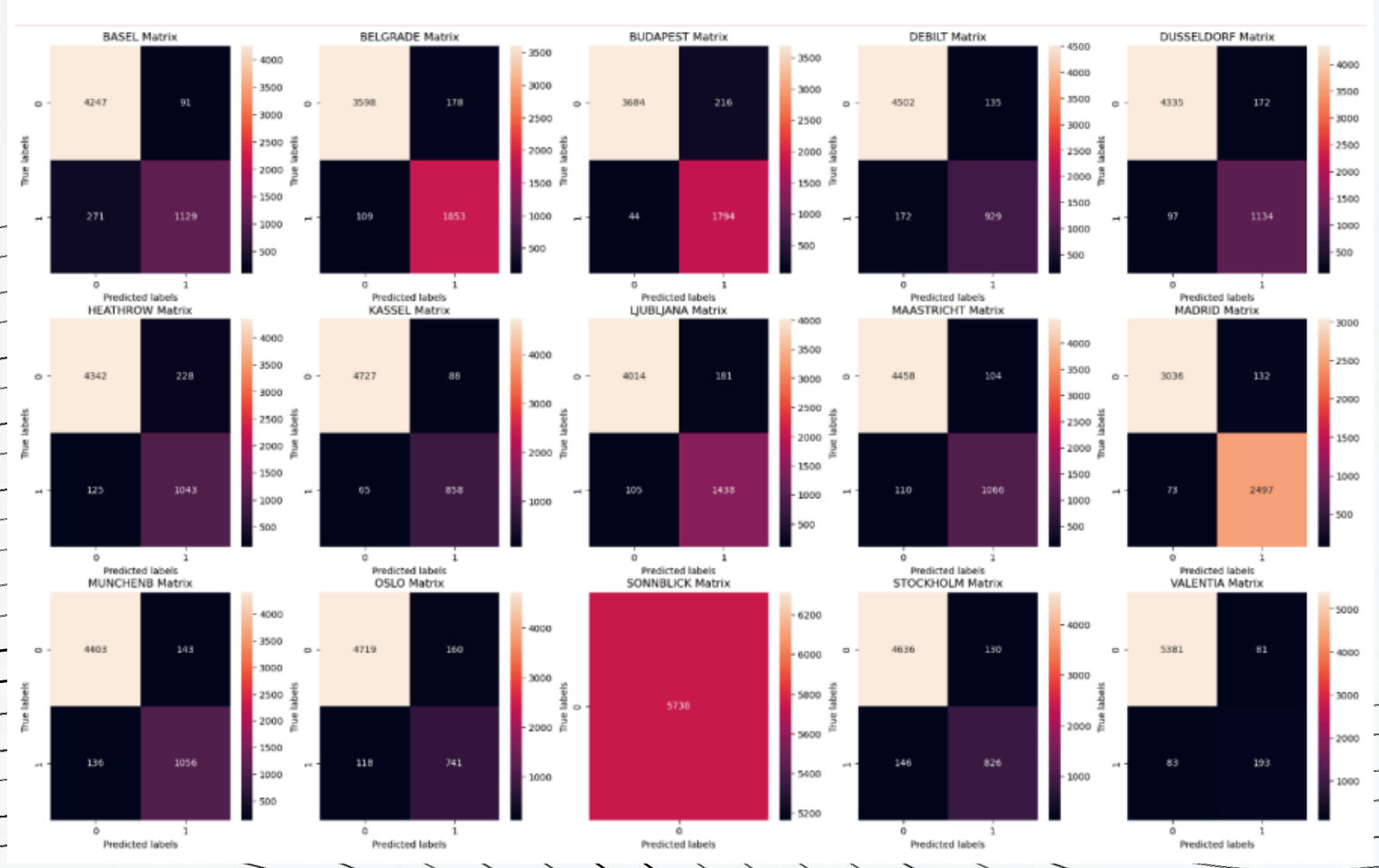
Weather Station	Accurate Predictions		False Positive	False Negative	Accuracy Rate
Basel	3907	935	465	431	84%
Belgrade	3238	1502	460	538	83%
Budapest	3416	1432	406	484	84%
Deblit	4346	732	369	291	88%
Dusseldorf	4167	800	431	340	87%
Heathrow	4161	754	414	409	86%
Kassel	4563	607	316	252	90%
Ljubljana	3726	1133	410	469	85%
Maastricht	4249	819	357	313	88%
Madrid	2735	2257	313	433	87%
Munchenb	4222	766	426	324	87%
Oslo	4624	507	352	255	89%
Sonnblick	5738		0	0	100%



DECISION TREE



ARTIFICIAL NEURAL NETWORK (ANN)



SUMMARY AND NEXT STEPS

Summary

- This project has identified KNN as the most suitable algorithm for predicting weather events. With an accuracy of 88% ClimateWins can use this model to understand weather patterns and how these are affected by Climate Change.

- If the Decision Tree model wants to be applied successfully, I would suggest prunning it for a higher accuracy.
- Try new weights for the ANN algorithm.
- Consider combining algorithms that could uncover patterns in the data.

Next Steps

THANK YOU

Any questions?

nuria@ClimateWins.com

ClimateWins.com/questions