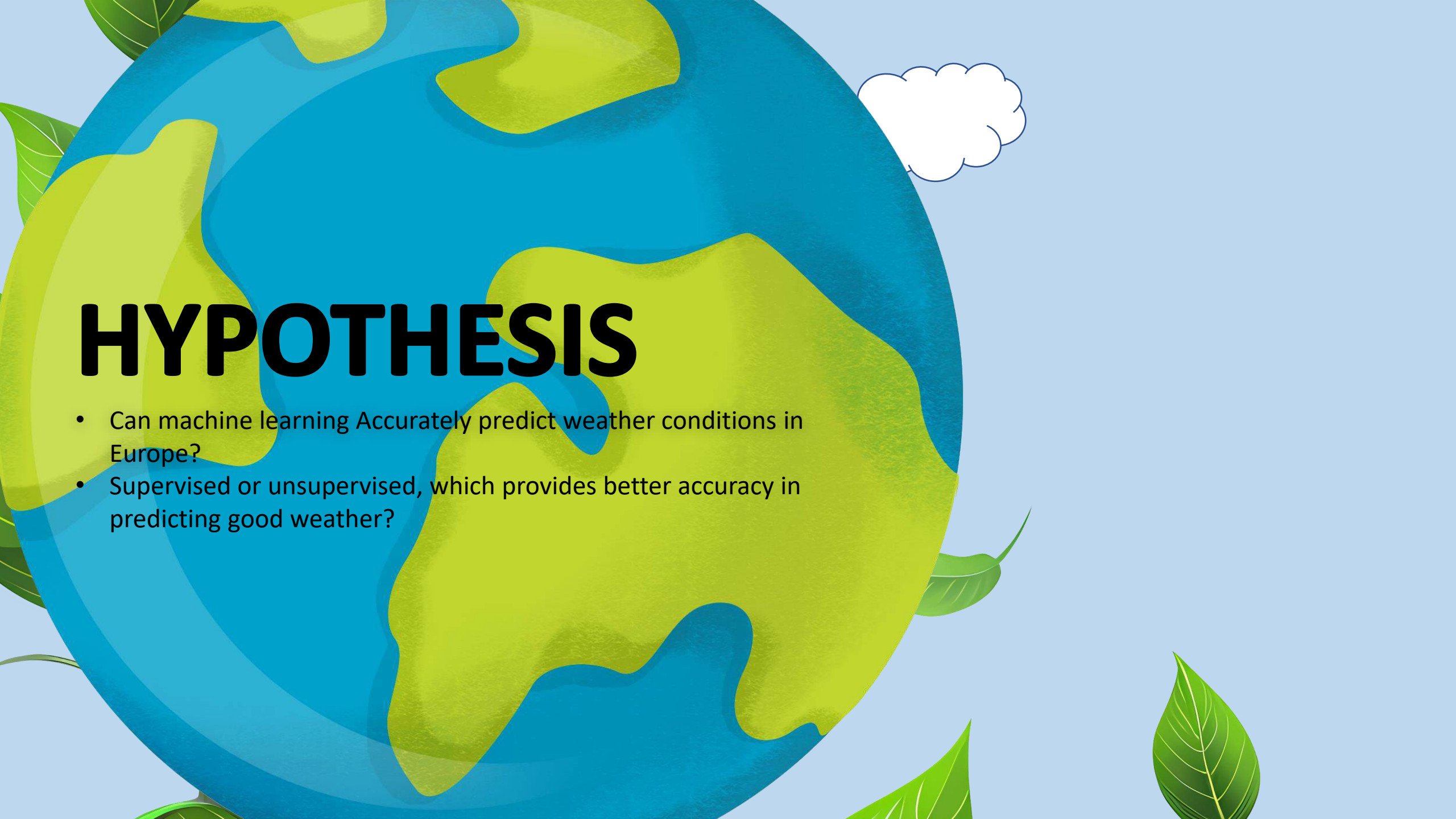
Data Specialization

Climate Conditions in Europe with ClimateWins

Shaquille Obomeghie

Introduction

Climate 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.



HYPOTHESIS

- Can machine learning Accurately predict weather conditions in Europe?
- Supervised or unsupervised, which provides better accuracy in predicting good weather?

Data

- You'll be using a data set based on weather observations from 18 different weather stations across Europe.
- The data set timeframe ranges from late 1800s to 2022.
- Recordings exist for almost every day with values such as temperature, wind speed, snow, global radiation, and more.
- The data was collected through the European Climate Assessment & Data Set project.

Data Bias

- **Sampling Bias:** Weather stations are often placed in specific locations, such as urban and developed areas, which may not accurately represent conditions across the entire region

- **Collection data:** Data was only collected for 18 weather stations while there is a total of 23755 weather stations in Europe.

- **Historical Bias**: Older data may have been collected using different methods or instruments from those used now, leading to discrepancies when comparing historical data to current measurements.

- **Temporal Bias:** There may have been inconsistency with the times when data was collected (e.g. hourly in some places and daily in some places.

Data Optimization

- The data optimization technique adopted in this project is the Gradient Descent.

- Through 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.

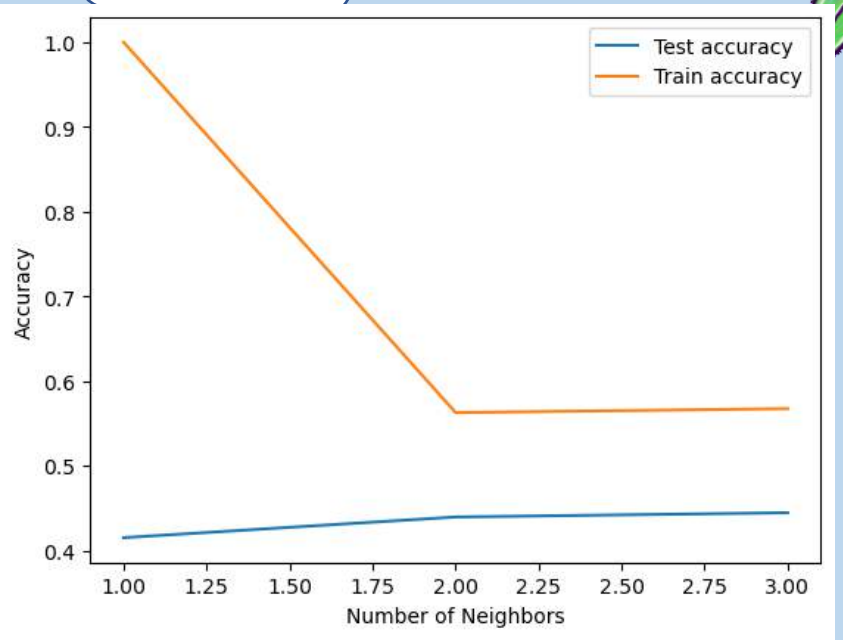
SUPERVISED LEARNING

K-NEAREST MEAN

The first Supervised Learning technique used was the K-nearest Mean. This is a classification-based algorithm. It finds the ‘K’ Closest training data point (neighbours) to a given input.

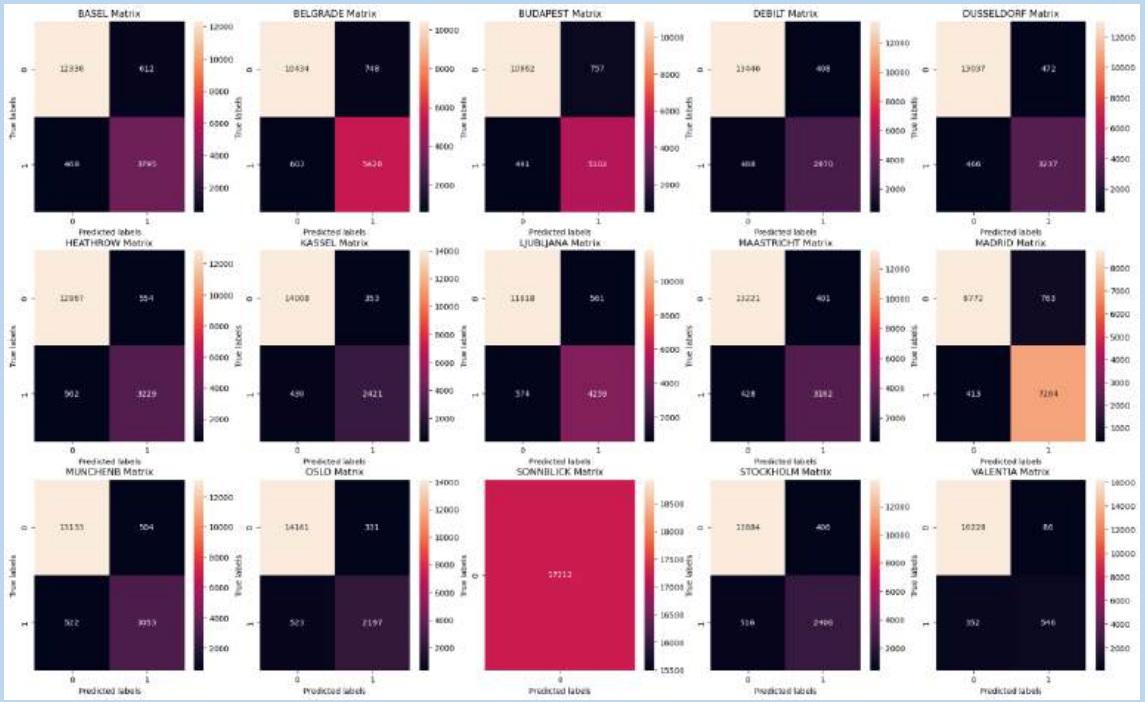
To test the accuracy, we used a confusion matrix to visualize the training and the testing

Train accuracy – 0.57 or 57%
Test accuracy – 0.44 or 44%

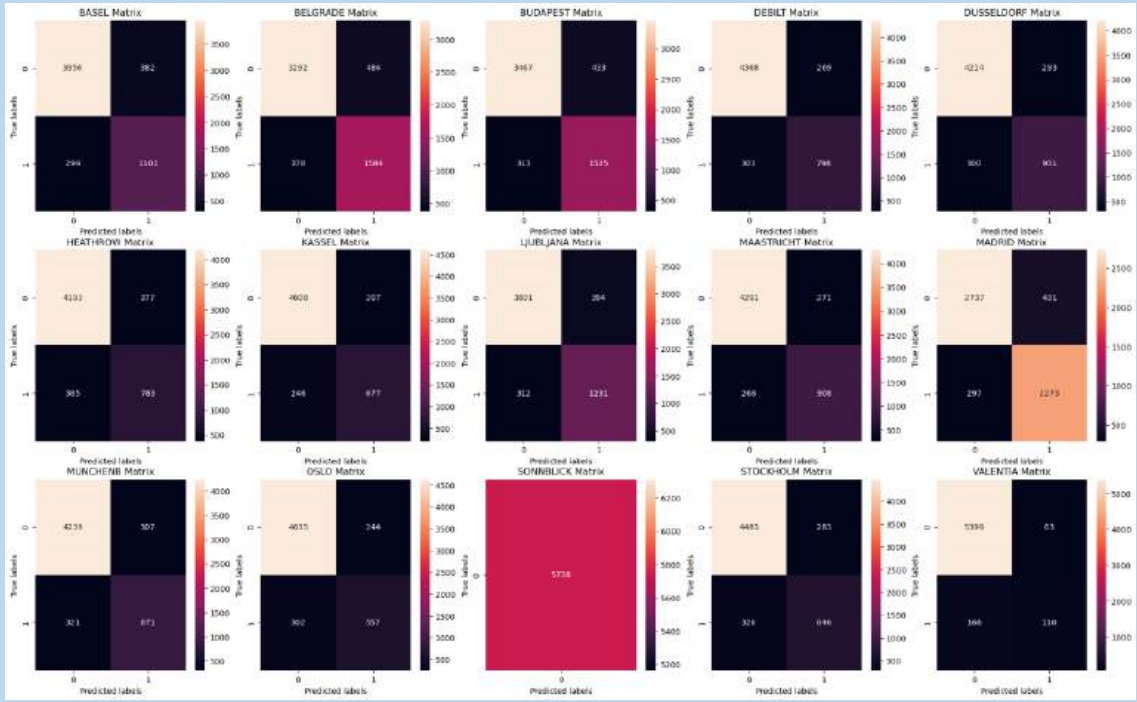


✦ This plot shows the relationship between the number of neighbors and the accuracy.

Training



Testing

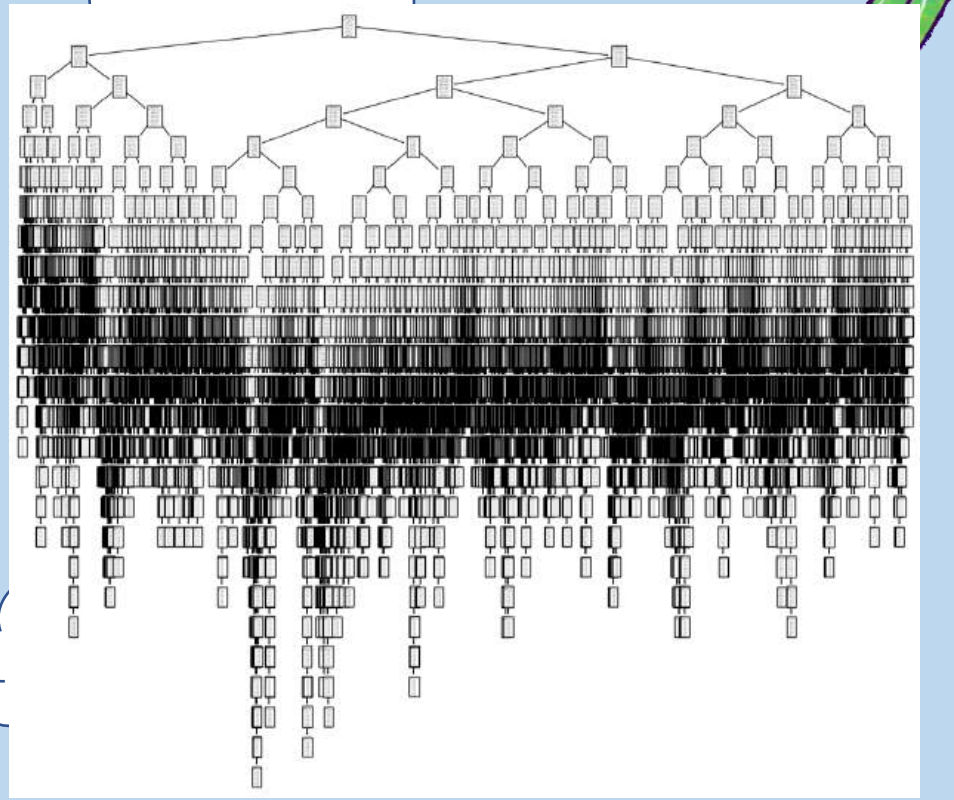


SUPERVISED LEARNING

Decision Tree

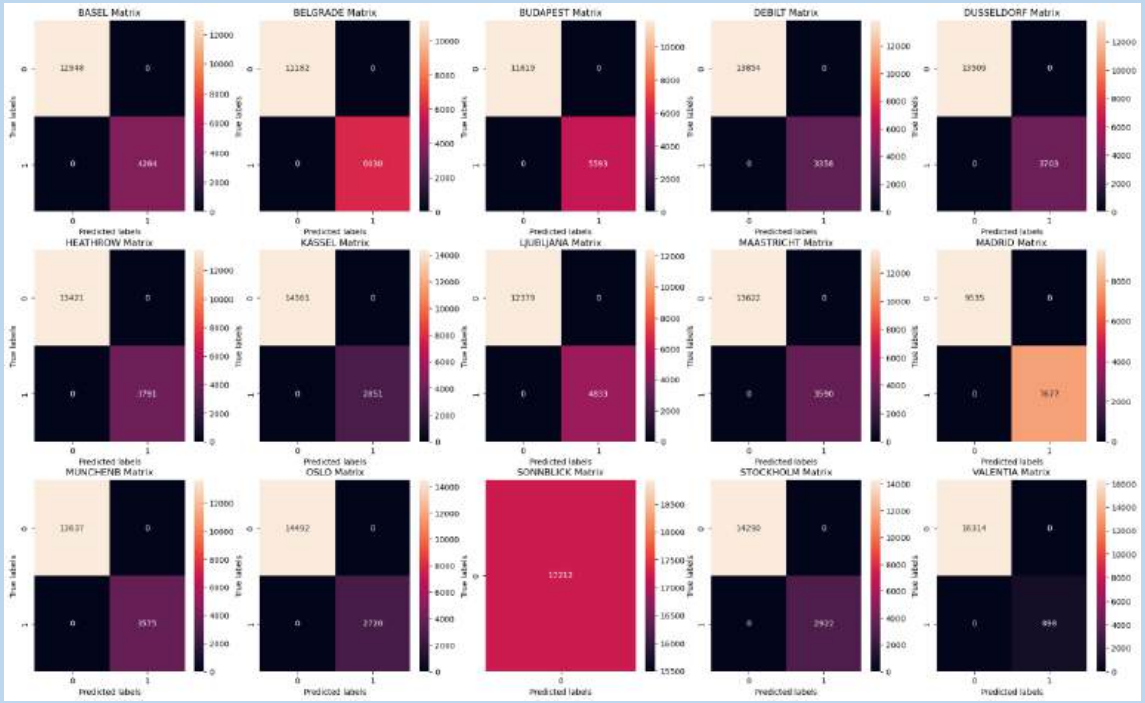
A **Decision Tree** is a type of machine learning model that makes decisions by asking a series of yes/no questions. It is used for **classification** (e.g., deciding if an email is spam or not) or **regression** (predicting a value, like house price).

Train accuracy – 0.60 or 60%
Test accuracy – 0.63 or 63%%

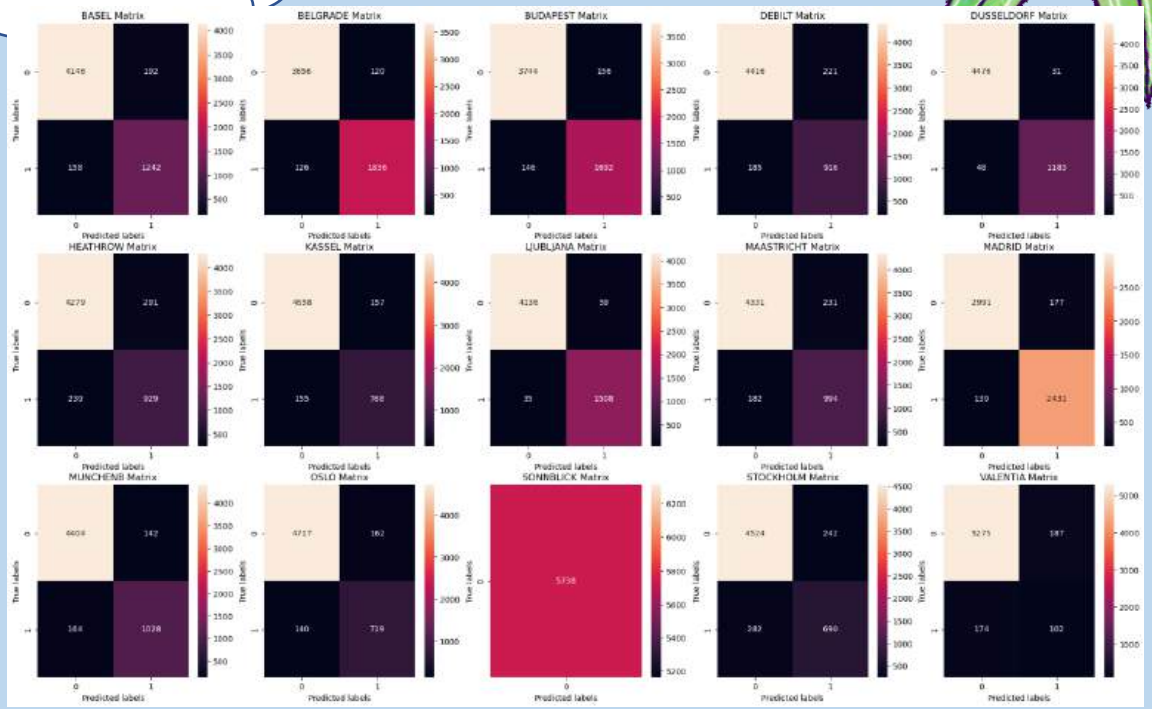


✦ The decision tree in this project was complex based on the data size. It is difficult to understand and interpret.

Training



Testing



SUPERVISED LEARNING

Artificial Neural Network Model

An ANN is a computational model made up of layers of interconnected nodes (neurons), which work together to process data. The network learns patterns from the input data by adjusting weights and biases during training, helping it make predictions.

For this Model, we manually ran three scenarios and continuously modified the hidden layers, max number of iterations and tolerance.

The most accurate (scenarios 3) gave:
Training: 0.72 or 72%
Testing: 0.65 or 65%

```
MLPClassifier
MLPClassifier(hidden_layer_sizes=(100, 100, 100), max_iter=1500)
```

```
MLPClassifier
MLPClassifier(hidden_layer_sizes=(100, 100), max_iter=1500)
```

```
MLPClassifier
MLPClassifier(hidden_layer_sizes=(30, 30), max_iter=1500)
```

- Ran three different scenarios, adjusting the elements of the MLP classifier

CONCLUSION AND NEXT STEPS

Conclusion and Findings

- The findings from this project show that all the models can accurately predict weather events, confirming the hypothesis, “Can machine learning accurately predict weather conditions in Europe?”
- The Decision Tree needs to be pruned to make it less complicated.
- When working with ANN model, there was an issue of overfitting whereby the training accuracy was substantially larger than the testing accuracy in various scenarios.

Next Steps

- Continuous pruning of the decision tree to make it less complicated and for better accuracy.
- Testing the dataset with other supervised algorithms.
- Look into unsupervised learning algorithms and apply them to the dataset

QUESTIONS??

THANK YOU

Shaquille Obomeghie