

# Machine Learning with Python: Weather Conditions and Climate Change with ClimateWins

Gautam Radhakrishnan Ajit  
16th December 2024



## Project Objective & Data Sources

- As a data analyst working at ClimateWins, a European organization, the aim of the project is to explore whether machine learning methods would be able to predict weather patterns across mainland Europe.
- The data was recorded at 18 different weather stations across Europe
- Different features such as wind speed, temperature, humidity, precipitation, snow depth etc. are recorded
- The values span across the 1800s to 2022.

Data source : [European Climate Assessment & Data Set project](#)

Data set : [Temperature data set](#)



# Hypothesis

- How is machine learning used? Is it applicable to weather data?
- Climate Wins has heard of ethical concerns surrounding machine learning and AI. Are there any concerns specific to this project?
- Can machine learning be used to predict whether weather conditions will be favorable on a certain day? (If so, it could also be possible to predict danger.)



## Project Objective & Data Sources

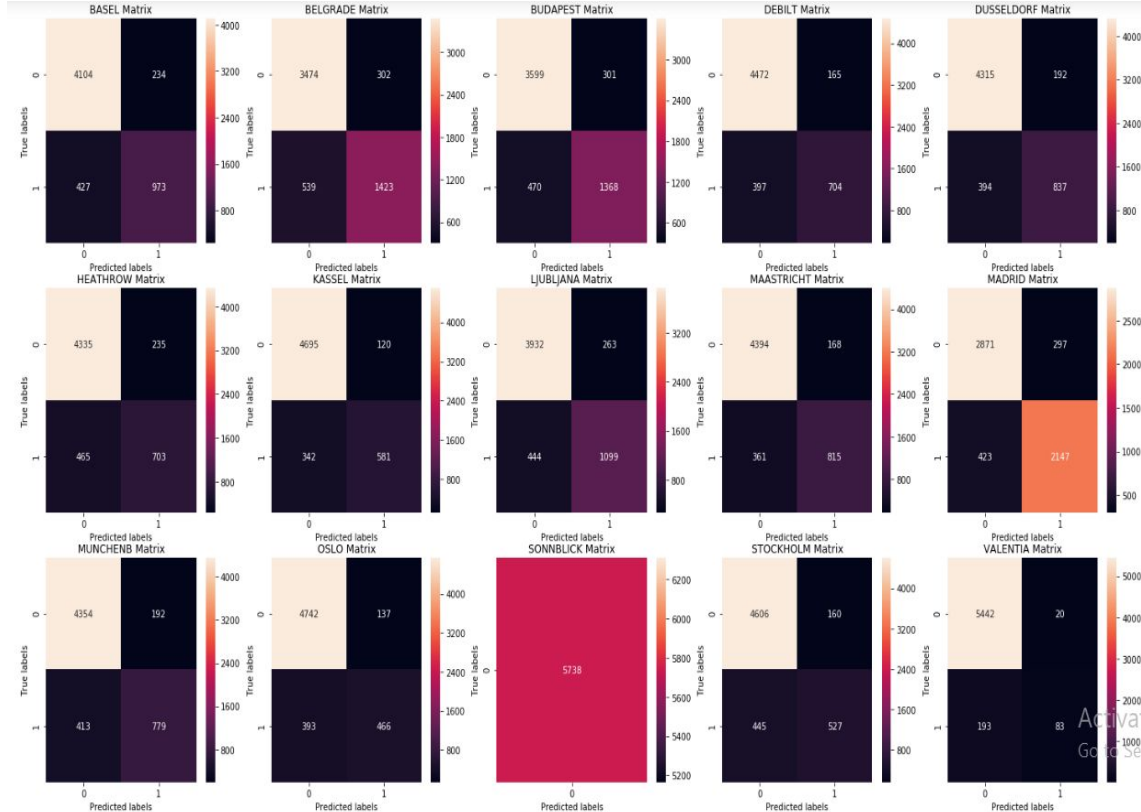
- As a data analyst working at ClimateWins, a European organization, the aim of the project is to explore whether machine learning methods would be able to predict weather patterns across mainland Europe.
- The data was recorded at 18 different weather stations across Europe
- Different features such as wind speed, temperature, humidity, precipitation, snow depth etc. are recorded
- The values span across the 1800s to 2022.

Data source : [European Climate Assessment & Data Set project](#)

Data set : [Temperature data set](#)

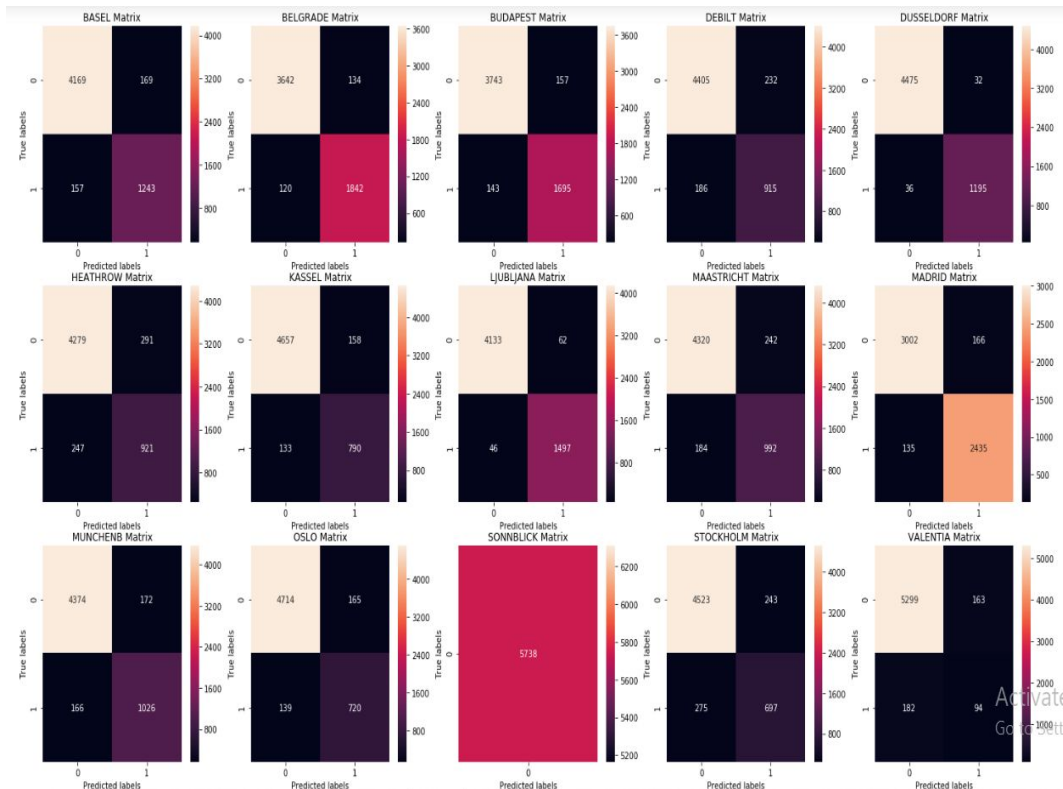
# Supervised Learning : kNN algorithm

- Training accuracy was 100% when the number of neighbours was 1.
- Model was overfitting, and the results were less impressive when the number of neighbours was increased
- Sonnblick recorded 100% test accuracy. Another station, Valentia, recorded 96%
- Average accuracy was measured to be 88%.

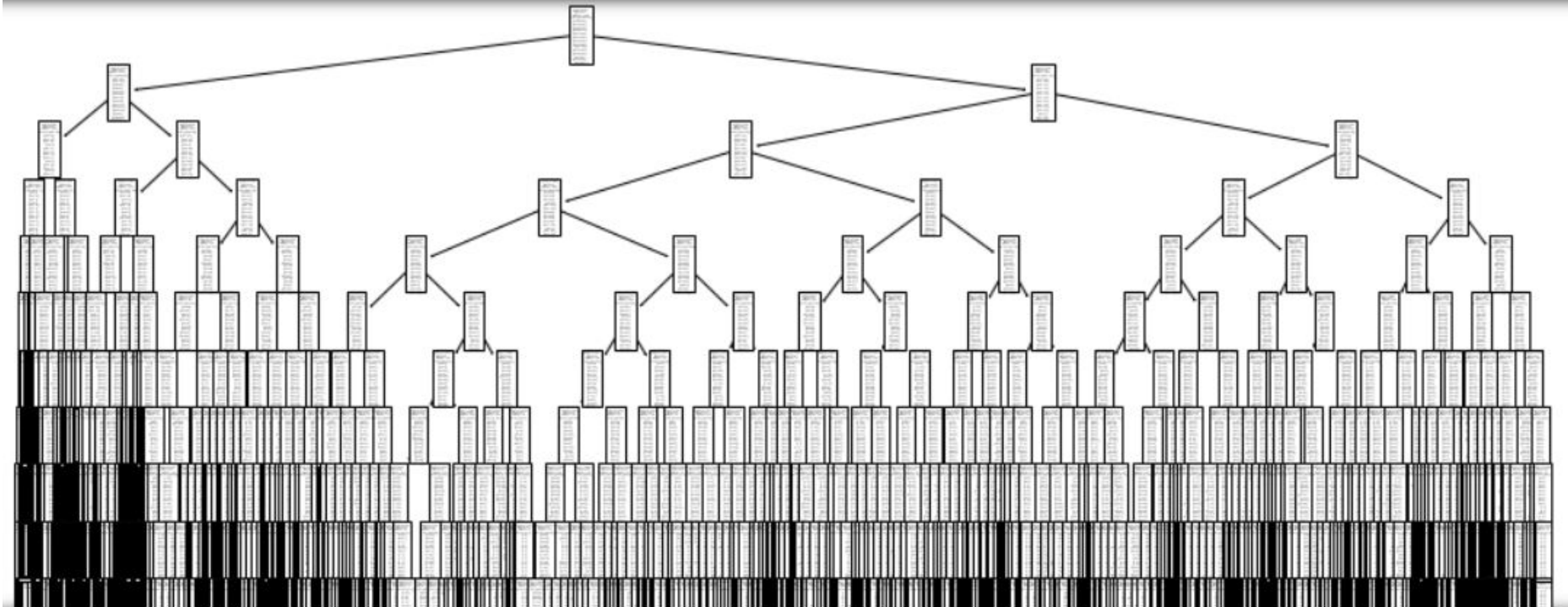


# Supervised Learning : Decision Tree - 1

- The Decision Tree achieved a test accuracy around 55%, with an approx 60% training accuracy.
- Much lower accuracy score than the kNN accuracy

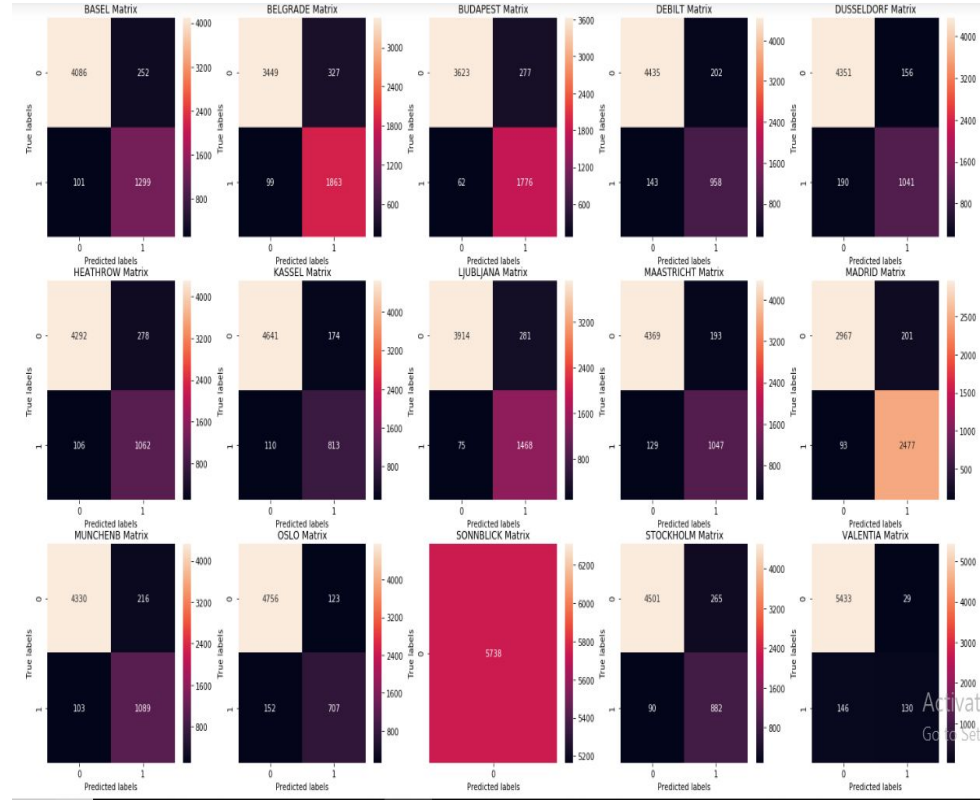


# Supervised Learning : Decision Tree - 2



# Supervised Learning : ANN model

- The first ANN model tested had two hidden layers of 5 nodes each, and was run for 500 epochs with a tolerance of 0.000001. This achieved an accuracy of 41.45% and 41.40% on the train and test data respectively.
- An additional hidden layer was added, along with changes in the number of nodes (100,50,25) and was run for 2000 epochs.
- This model achieved a much better training accuracy of 75.11 % and a test accuracy of 56%.







## Conclusion and Future Analysis

- The kNN model records the highest train accuracy with 88% on the training data. However, the test accuracy falls below 50%, which is the lowest among the 3 different models tested. This is a sign of overfitting, and hence cannot be recommended for this dataset.
- ANN is more recommended for this data, as tweaking the parameters have shown to yield better scores than the other models.
- Different hidden layers and hyper-parameters (ANN) and perhaps pruning the Decision Tree could improve the results further

# Potential Ethical Concerns



- Privacy is not an issue as Personally Identifiable Information (PII) is not collected.
- Some European regions may be less developed and data may not be accurately captured. This could hinder fair data representation.
- Data from the 1800s may be less accurate due to under-developed instrumentation and recording mechanisms.
- Not all governments would be open to sharing the data publicly, and may filter out data which could potentially show them in a bad light. This could affect the integrity of any study conducted on the data



**Thank you**