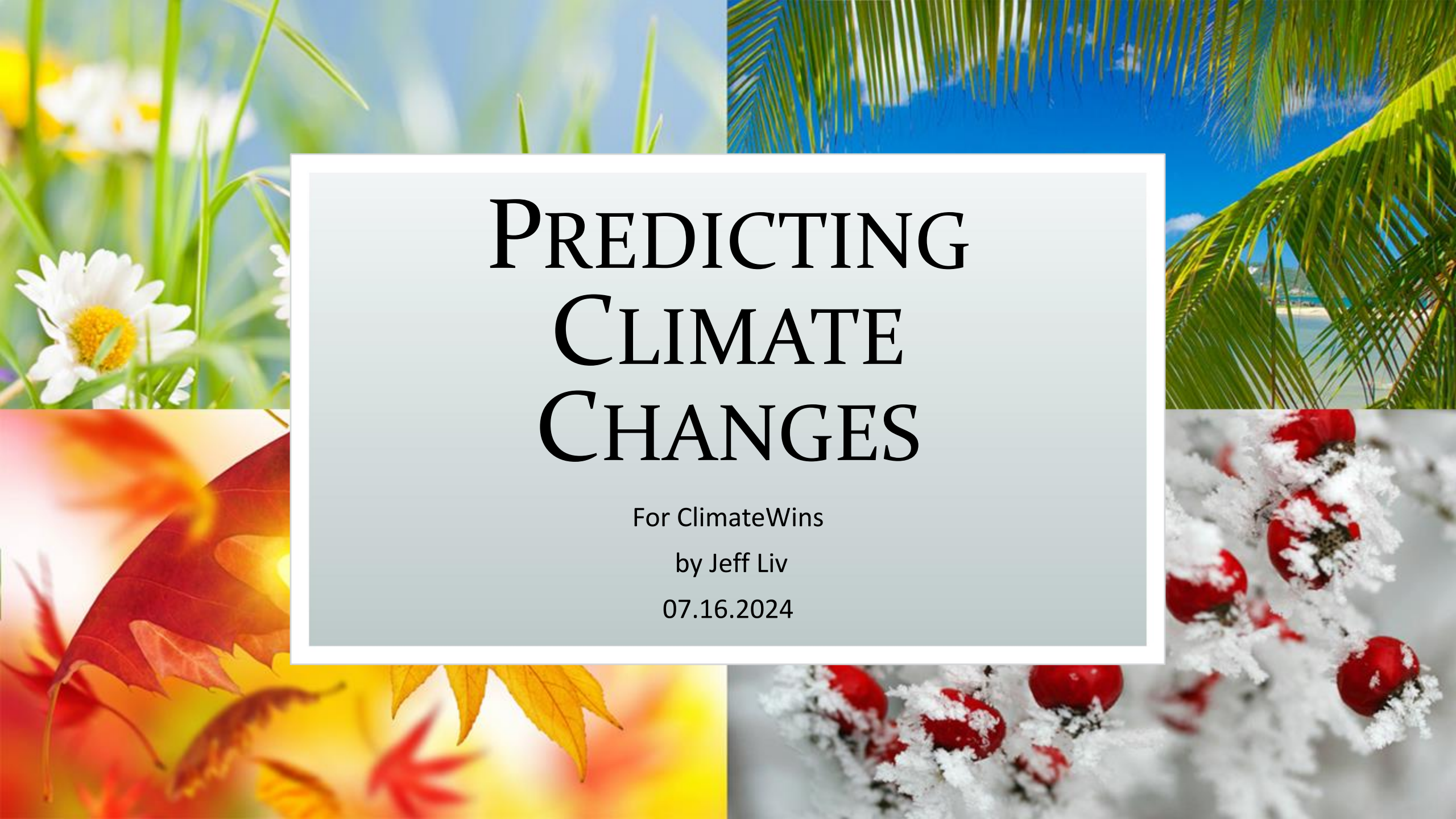


Predicting Climate Changes

For ClimateWins

by Jeff Liv

07.16.2024

Contents

Project Overview

ClimateWins is interested in using Machine Learning to help predict the consequences of climate change around Europe and potentially the world.

Data Source

- The data set provided is based on weather observations from 18 different weather stations across Europe from the past century.
- Data contains values such as temperature, wind speed, snow, global radiation and more.
- Data was collected by the European Climate Assessment & Data Set Project.

Data Bias

This data is susceptible to data biases that include:

- **Instrumental Biases –** Issues with Calibration of weather stations, Instrument degradation, etc.
- **Observer Biases –** If measurement observations were done manually.
- **Technological Biases –** Technological advances such as sensor upgrades, changes from manual to automatic observations.
- **Location Biases/Coverage Biases/Environmental Changes –** Things such as Urbanization, relocation of weather stations, geographic coverage, altitudes, climate and land changes.

Hypotheses

Using historical data provided by the European Climate Assessment, Machine Learning can predict weather forecasts.

Machine Learning will be able to differentiate and predict Unpleasant days and Pleasant days.

A Machine Learning technique has the potential to be the most effective method for predicting weather.

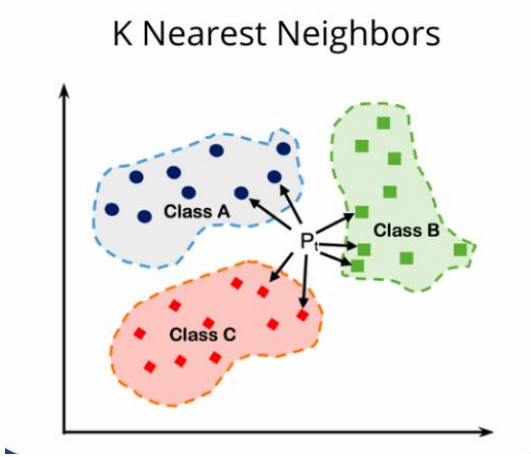
Ethical Concerns

Although there are no concerns regarding Privacy, it is possible that the design and implementation of Machine Learning algorithms can introduce some ethical concerns. This could include:

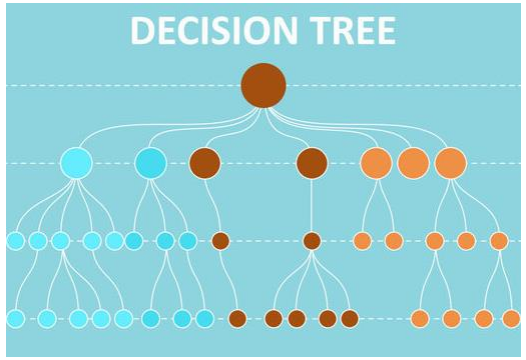
- **Bias and Fairness** – This can include historical biases and algorithmic bias.
- **Transparency and Explainability** – "Black Box" models and explainability.
- **Accountability** – Regulations and Oversight or responsibility for predictions.
- **Impact on Society** – Public trust or unintended consequences.
- **Access and Equity** – Digital divide and resource allocation.
- **Environmental Impact** – Energy consumption.

Different Machine Learning techniques

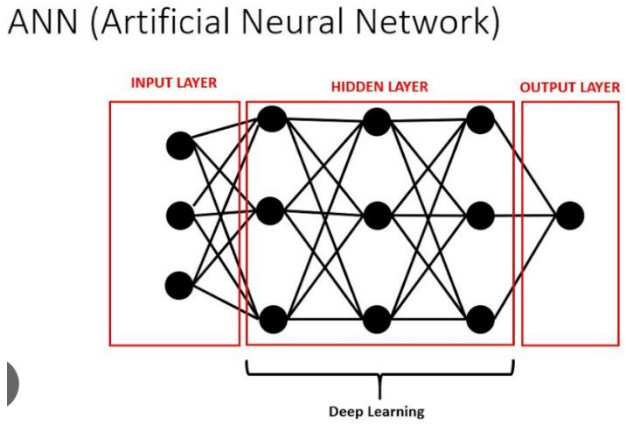
One of the first Machine Learning techniques we attempted was the **K-Nearest Neighbor** (KNN) algorithm. It belongs under the Supervised Learning domain and finds intense pattern recognition, data mining and intrusion detection.



Another Machine Learning technique that belongs under the Supervised Learning domain is the **Decision Tree** algorithm. This algorithm is used for classification and regression modeling. Regression is a method used for predictive modeling, so these trees are used to either classify data or predict what will come next.



Our last Machine Learning technique that belongs to the Supervised Learning domain is the **Artificial Neural Network** (ANN) algorithm. It uses interconnected nodes or neurons in a layered structure that resembles the human brain.

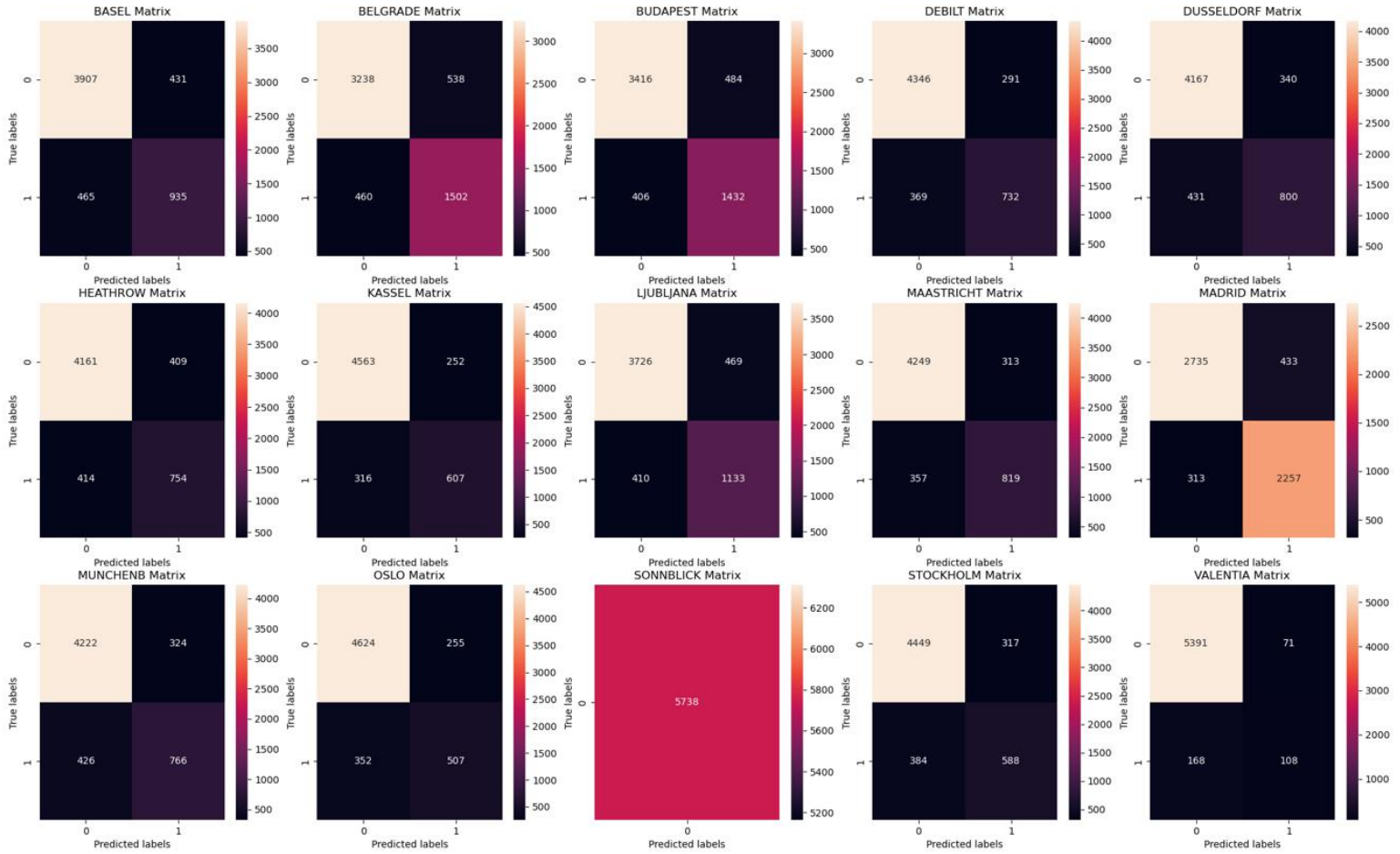


K-Nearest Neighbor (KNN) Findings

By taking our True Negative and adding our True Positive, then dividing by the overall number, we receive our accuracy percentage.

It should be noted here that the SONNBLICK weather station shows an accuracy rate of 100%, however, the data provided from the weather station showed all negative temperatures.

Overall Accuracy for Individual stations: 88.12%.

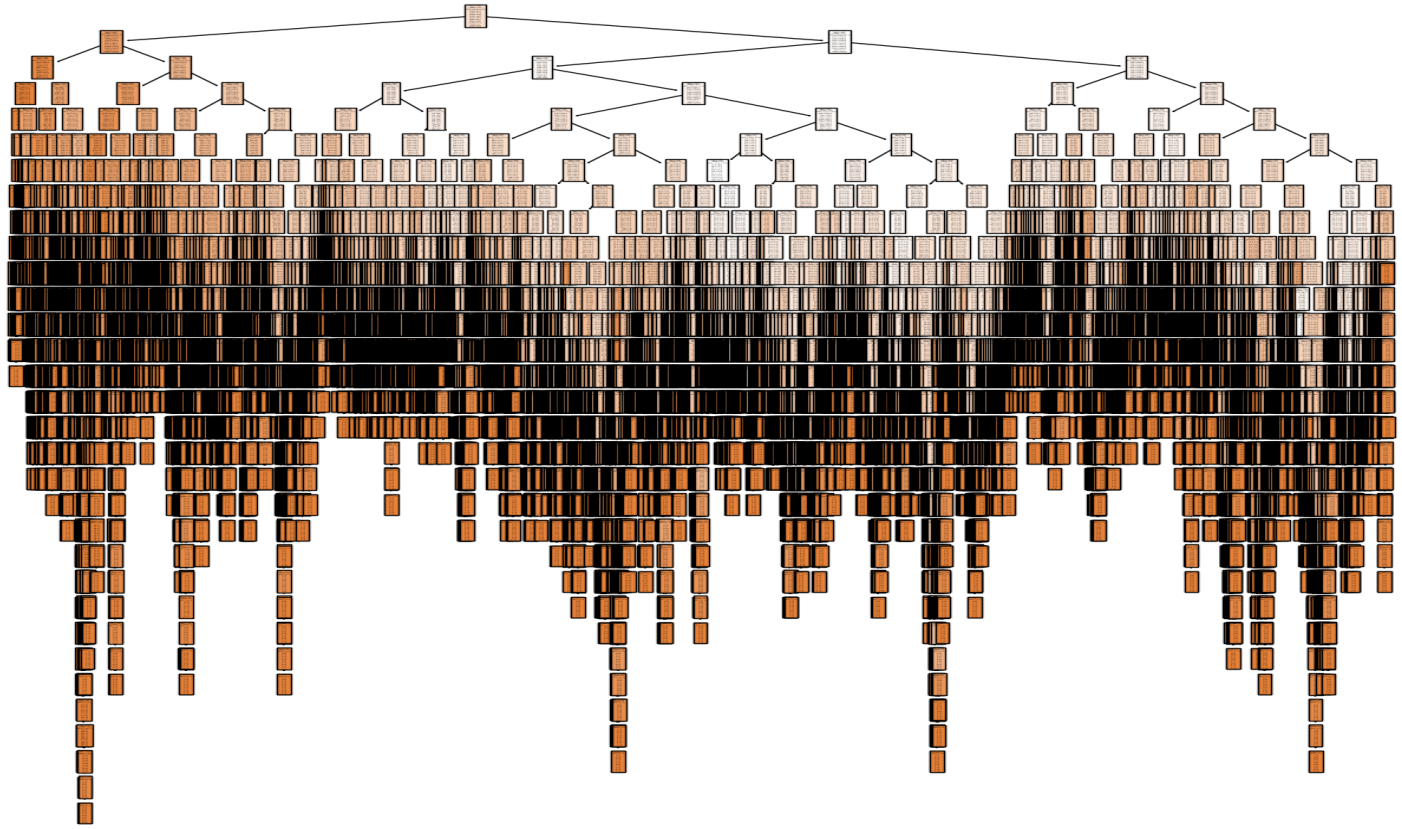


True Negative	False Positive
False Negative	True Positive

Decision Tree Findings

The decision tree algorithm we ran produced a visualization that was very deep and complex. It was difficult to visually interpret due to its size.

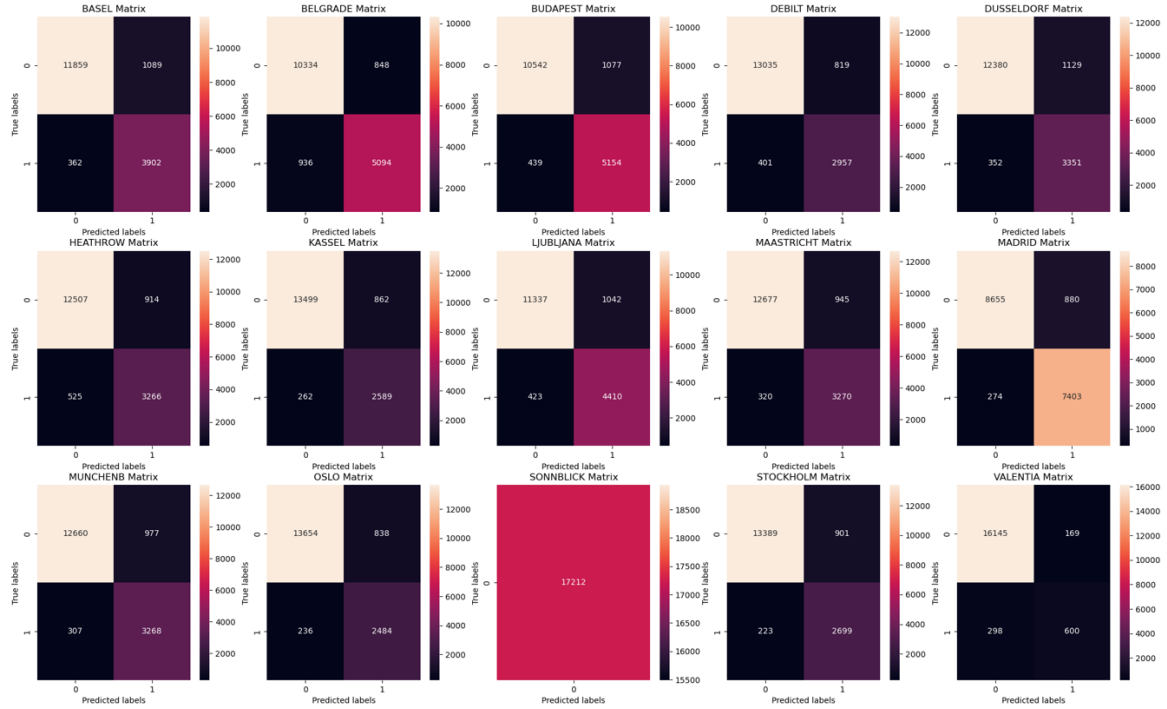
Overall Accuracy for Individual stations: 87.11%.

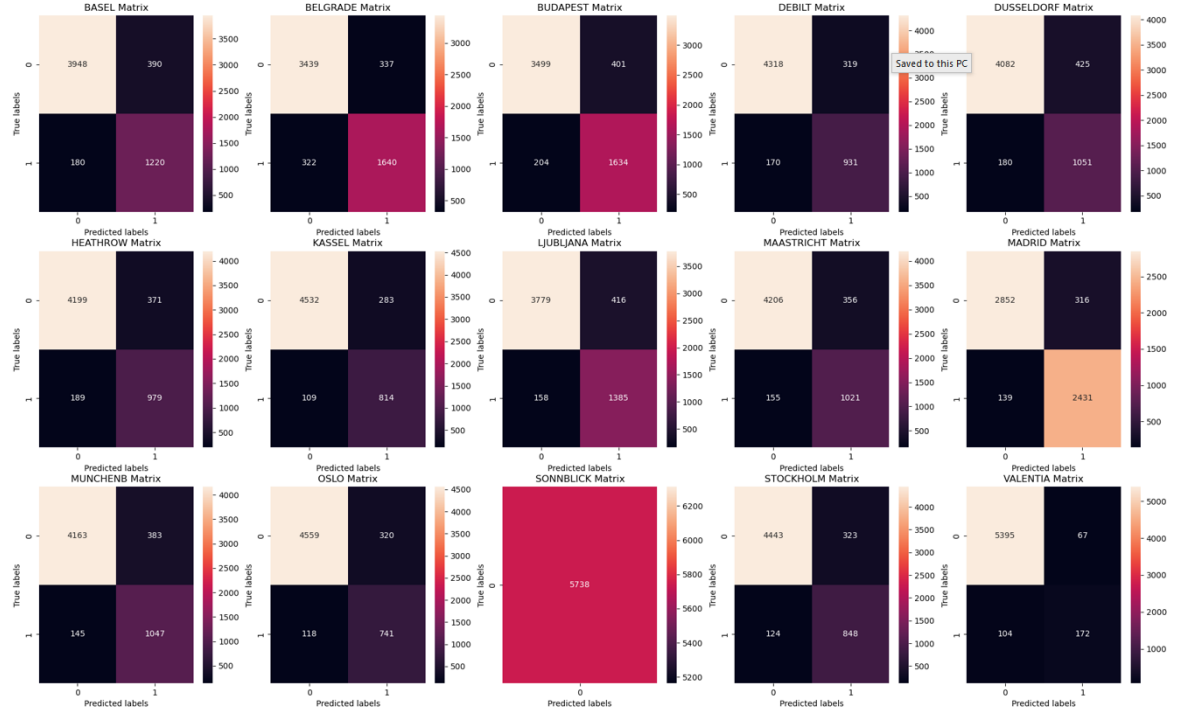


Artificial Neural-Network (ANN) Findings

Below we have our ANN algorithm results. To the left is our training results and on the right is our test results. These results are based off three nodes of 50, max iterations of 1000 and a tolerance of 0.0001.

Overall Accuracy for Individual stations after refined layers and nodes: 91.86%.





Conclusion & recommendations

- All three algorithms can predict Pleasant and Unpleasant days well.
- Of the three algorithms, ANN requires the most testing as different numbers of nodes, quantity for each node, how many iterations and tolerance can affect the outcome substantially.
- The Decision Tree visualization is very clustered, leaving it difficult to visualize.
- The K-Nearest Neighbors predicted very well for individual weather stations for the test without overall need to modify the algorithm and is our recommended approach.

Next Steps

- Implement KNN Model for predicting weather.
- Conduct further validation to ensure models reliability across different weather stations.
- Address overfitting in KNN
  - Investigate and address the overfitting issue observed in the SONNBLICK weather station.

Thank You

Jeff Liv

jchadliv@gmail.com

07.16.2024