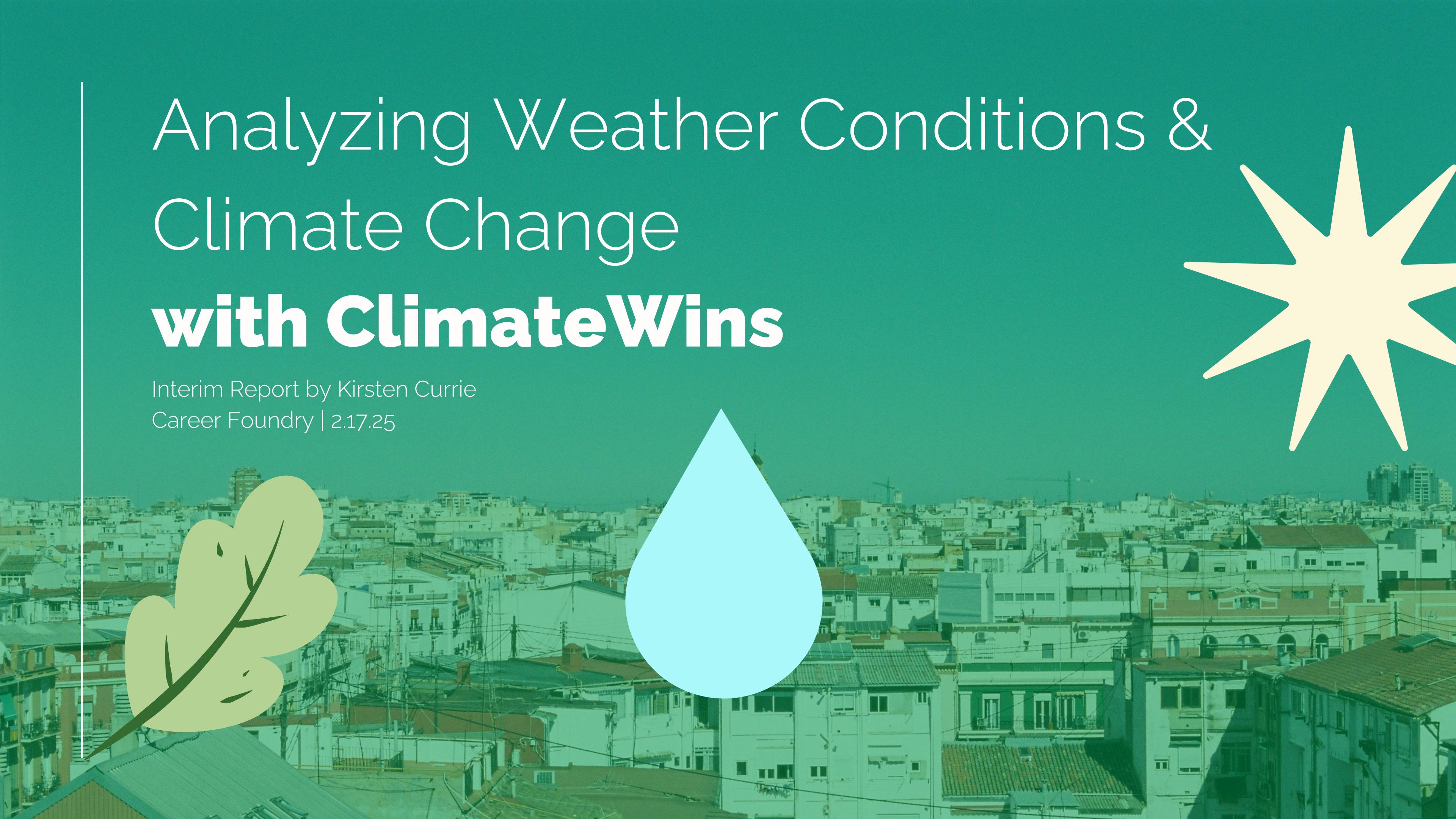


Analyzing Weather Conditions & Climate Change **with ClimateWins**



Interim Report by Kirsten Currie
Career Foundry | 2.17.25

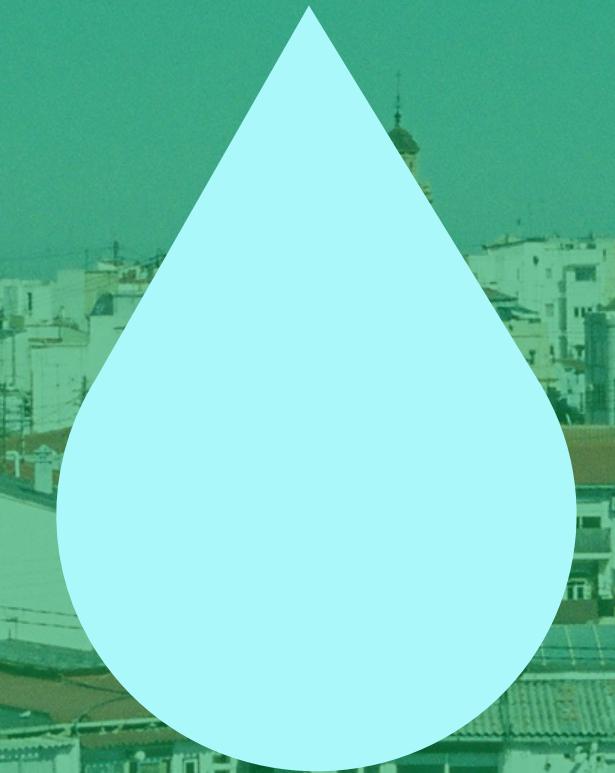
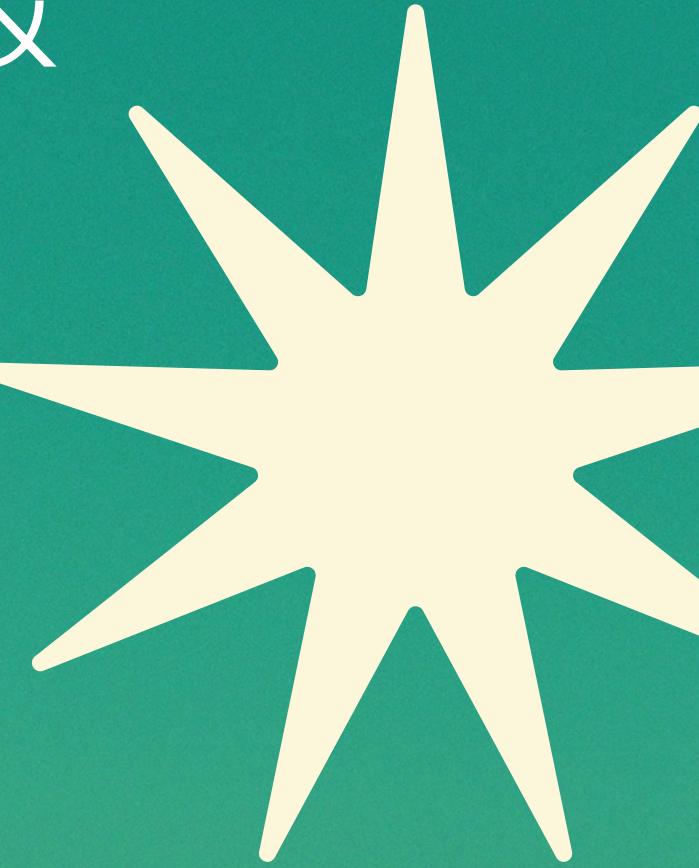




Table Of Contents

01.

INTRODUCTION

The problem at hand & data
behind it

02.

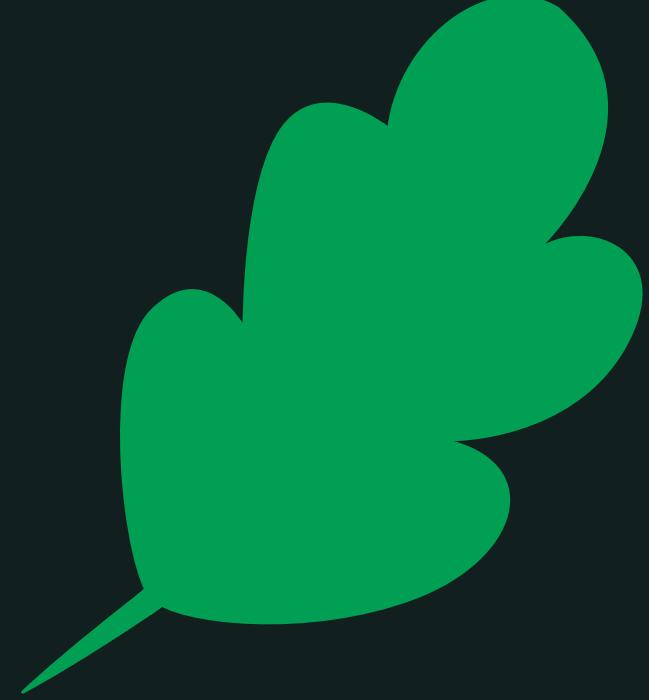
THE METHODS

Supervised model testing &
evaluation

03.

SUMMARY

Conclude best supervised
machine learning model for
ClimateWins



Introduction

THE PROBLEM AT HAND & DATA BEHIND IT

Why It Matters

The past 10-20 years have brought increasingly concerning weather extremes to our planet.

ClimateWins, a European Nonprofit Organization, is pushing to investigate weather patterns over the past century in order to address climate conditions not just for Europe, but hopefully the entire globe.





The Resources

- The data represents weather data collected from **18 different weather stations** across Europe
- Years between **1800 - 2022** collected
- [Downloadable Dataset Here](#)
- Data was collected by the [European Climate Assessment & Dataset Project](#)

Proposed Hypotheses

- Machine learning models will vary in their accuracy for model prediction.
- Temperature trends in European weather stations show a significant increase over the past century.
- Weather patterns will vary across station and may not always correctly be used to predict weather patterns globally.





Possible Biases

HISTORICAL

- The data collection methods as well as quantity of data may not be consistent over the last century, leading to potential data misrepresentation.

HUMAN OPINION

- Some of the following analyses will be looking at the relationships between actual temperature averages compared to ratings of weather "pleasantness". The way in which people perceive weather being pleasant or not is inherently biased and can present issues with accurately predicting "pleasant" weather days.

LOCATION

- Weather stations will experience different variations of weather conditions and could cause the algorithms to learn from the data incorrectly.



The Methods

SUPERVISED MODEL TESTING & EVALUATION

Machine Learning Supervised Models

CLIMATEWINS

OPTIMIZATION

Using gradient descent for temperature predictions

K NEAREST NEIGHBORS

Will data neighbors be an accurate way to classify weather?

DECISION TREES

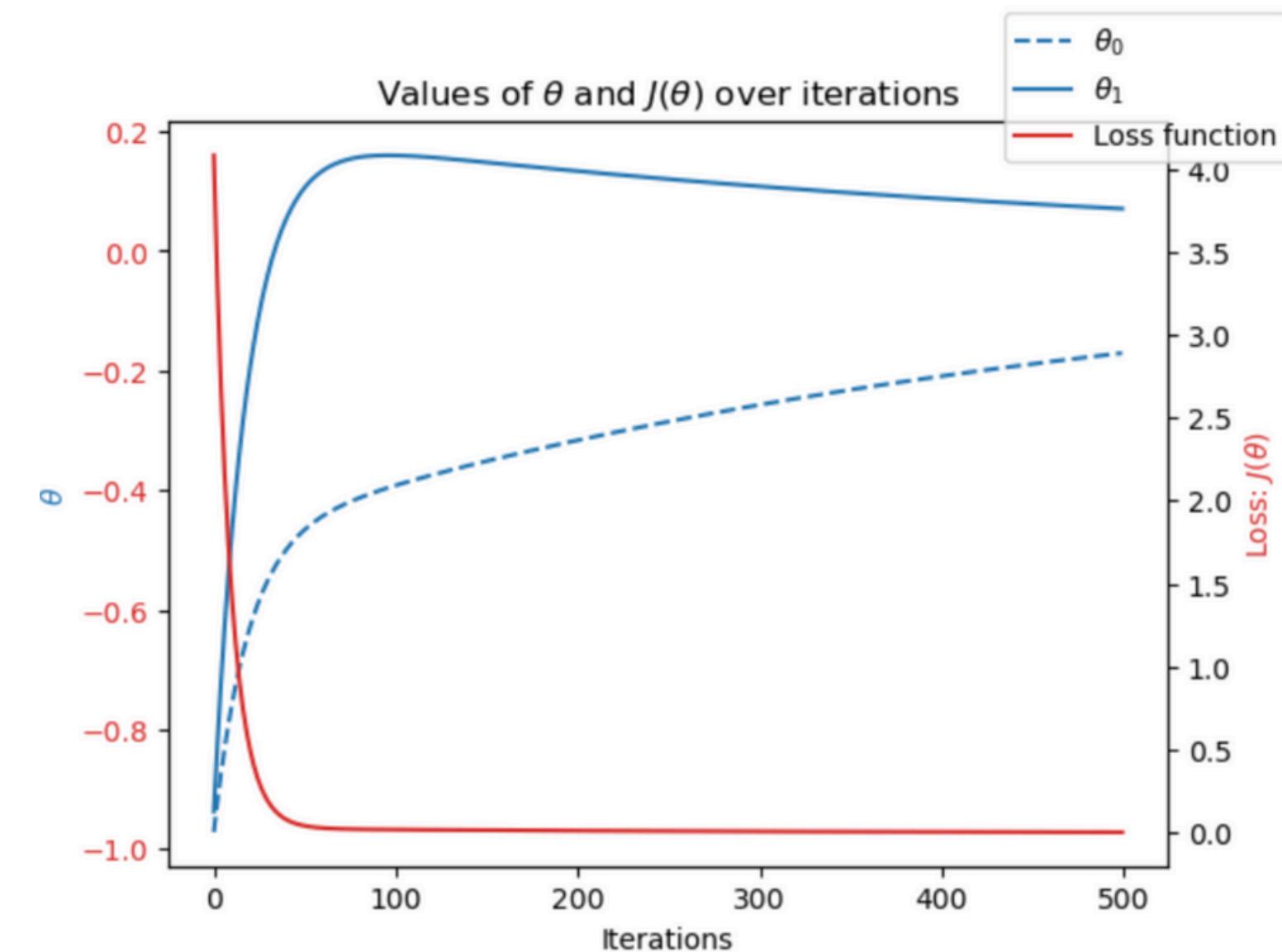
A flowchart-like structure to make predictions or classifications based on input data

ARTIFICIAL NEURAL NETWORK

Models inspired by the structure of the human brain.

Optimization

- Gradient descent was applied to optimize temperature prediction models for three European cities.
- The loss function helped measure the accuracy of temperature predictions against actual historical data.
- Iterative adjustments were made to model parameters (Theta values) to minimize prediction errors.
- For all three cities, min and max temperature means tend to increase in extremes over time.

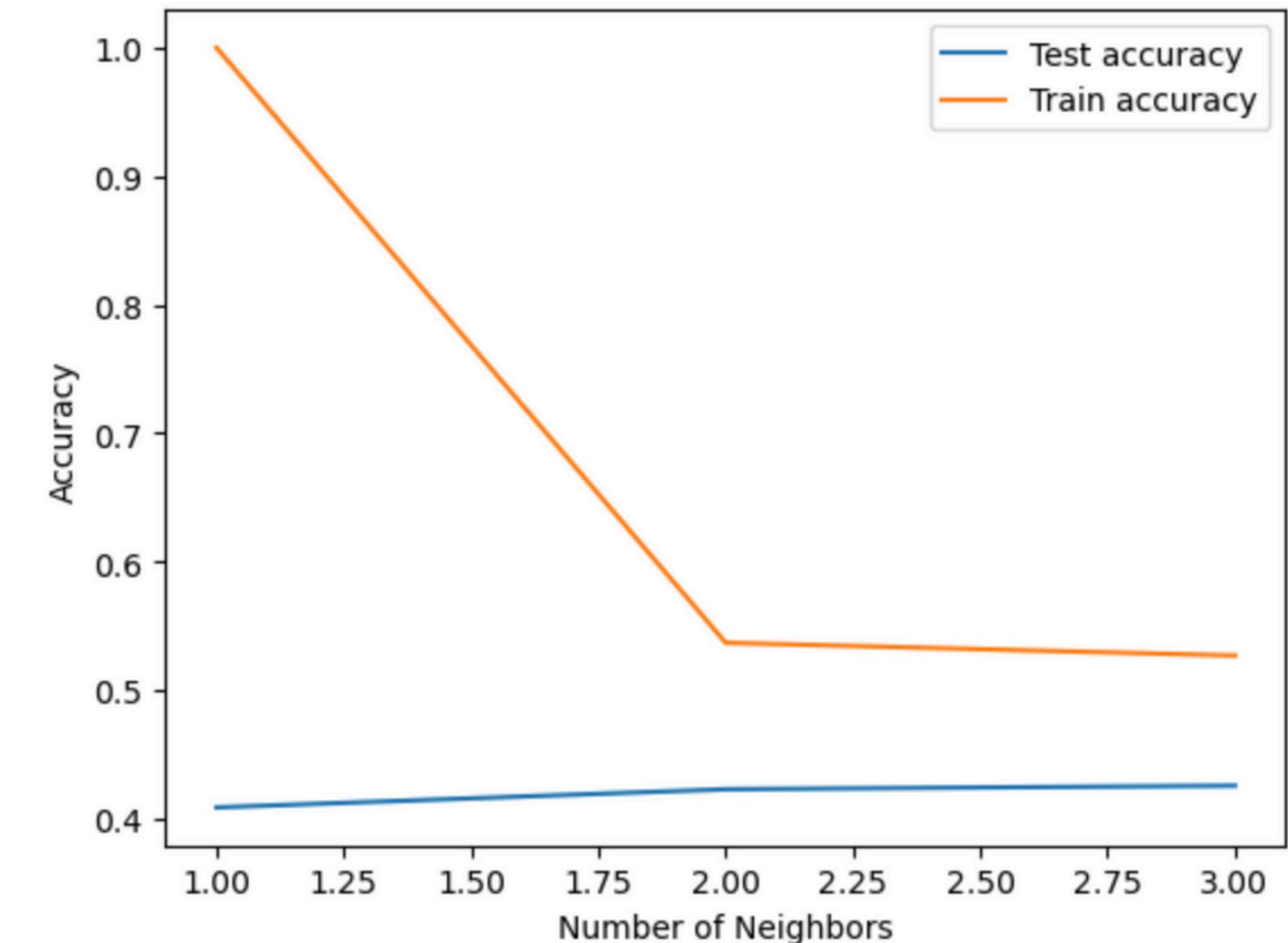


This chart represents the average temperature of the city of Kassel during the year 2019. The red line shows a rapid descent towards near 0, which is what the model is optimizing for depending on number of iterations, steps, and step sizes.

K Nearest Neighbors

- KNN is an algorithm used for classification tasks. It works on the principle that similar data points are typically close to each other.
- Using scaled data (to prevent possible weighting towards larger values), average temperatures across 15 weather stations were compared to ratings of weather "pleasantness".
- Beyond 2 "neighbors", model accuracy began to dip.
- Risk of overfitting was present

CLIMATEWINS

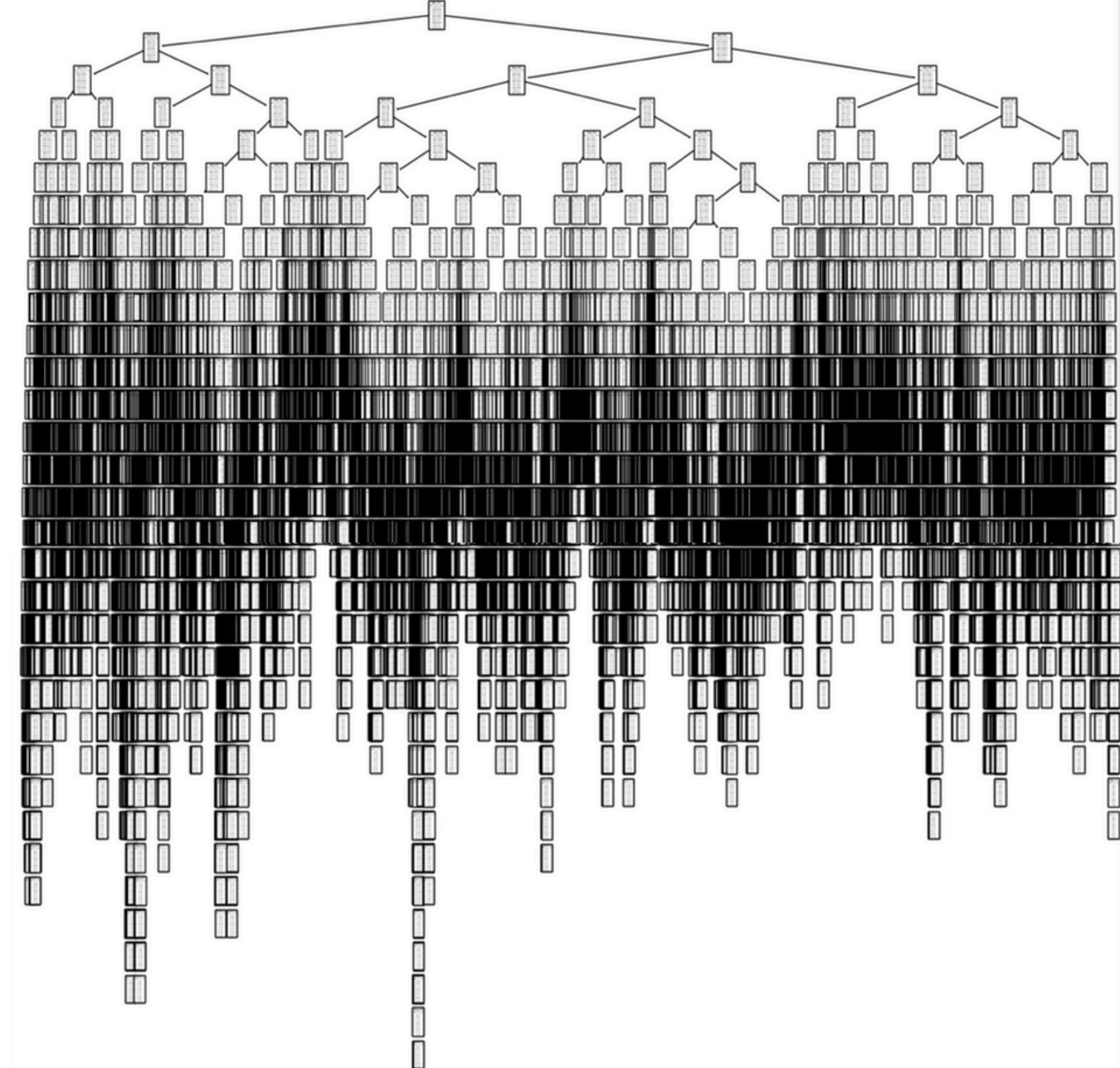


The best training accuracy prediction is about 53% and the best testing accuracy prediction was about 43% using K Nearest Neighbors algorithm.

Decision Tree

CLIMATEWINS

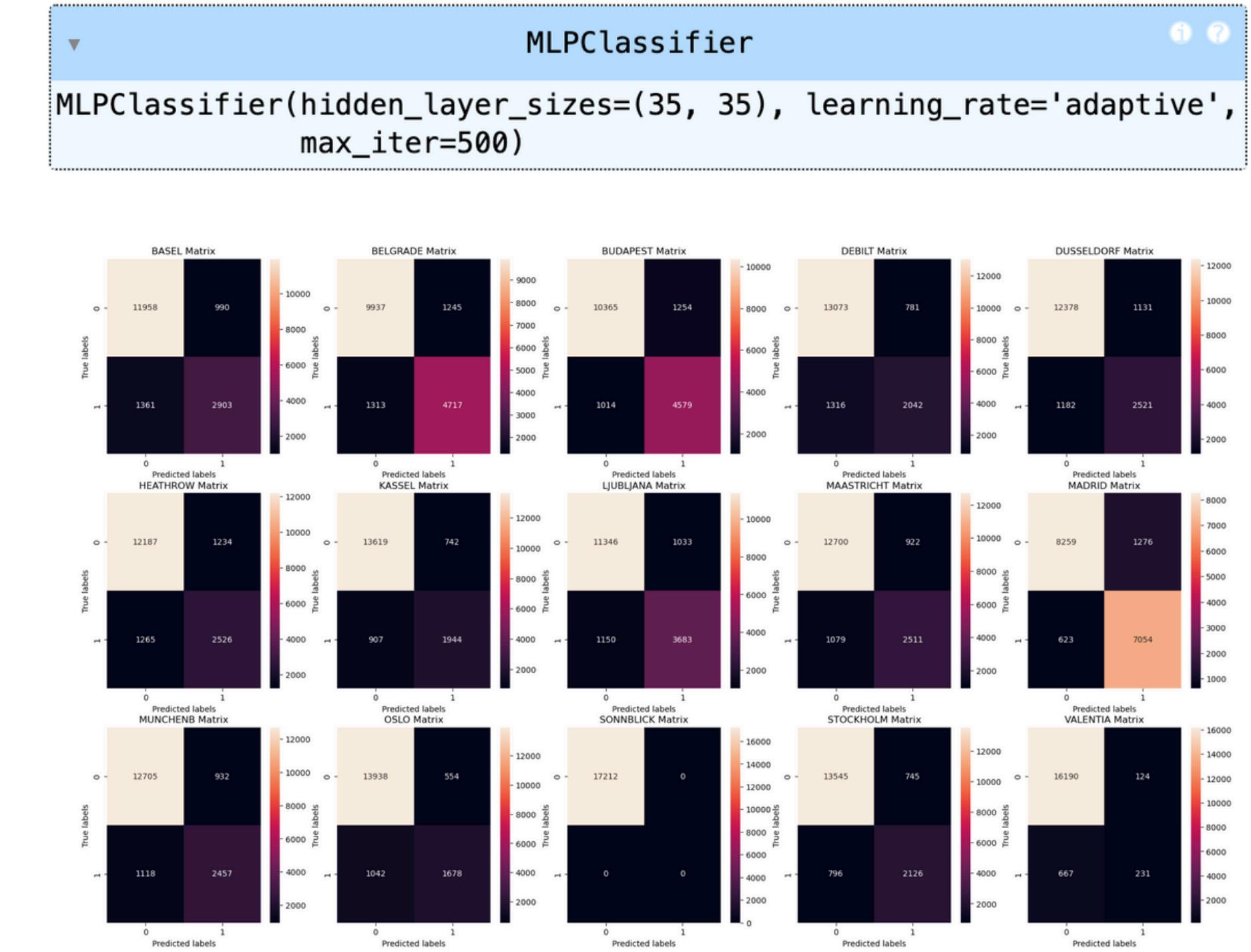
- The tree begins with root nodes and evaluates the best way to split features into sub groups. It aims to create increasingly homogeneous subgroups, making it easier to predict outcomes or classify data points.
- For this purpose of this study, it's advisable that branches are "pruned" as it's not decipherable in it's current state.
- Model testing accuracy was only about 39.8%.



Decision Tree of the average temperature across 15 European weather stations. Undecipherable and needing "pruned".

Artificial Neural Networks

- Artificial Neural Networks (ANNs) are machine learning models inspired by the structure and function of the human brain.
- Various model “hidden layers”, nodes, & iterations were tested in order to see which variation produced the most accurate model.
- The dataset was quite large, and it was challenging to determine the appropriate amount of layers and nodes, though it appeared that too many layers & nodes could result in overfitting, as the accuracy score was higher with few layers & nodes.

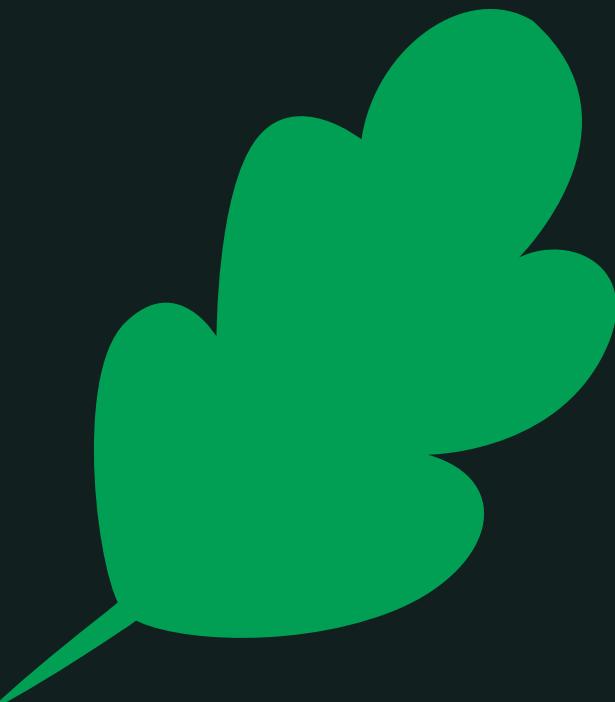


The 4th iteration of ANN on the average temperature / weather pleasantness data resulted in the best accuracy score (Train - 44%, Test - 46%) with few number of layers & nodes than the previous iterations. However, the confusion matrix here still shows several false predictions (black squares).



Summary

CONCLUDE BEST SUPERVISED MACHINE
LEARNING MODEL FOR CLIMATEWINS



Which Model is the Best?



Of the three supervised models, the K Nearest Neighbors appeared to have provided the most accurate scores of the three.

The Artificial Neural Network could potentially test closer depending if more time was spent adjusting layers and nodes, but it should be noted that there was some discrepancy between the train/test scores appearing high while the confusion matrix still showed a large amount of false predictions.

ANN & KNN Classification Report F1 scores were comparable, with Train scores being slightly better for KNN (ANN - 70% Train, 73% Test; KNN - 83% Train, 68% Test).

To summarize, KNN appears to be the better model of the three, and would be recommended for ClimateWise.

The Summary

HYPOTHESES

- Models varied in their accuracy in weather predictions.
- More extreme in min and max average temperatures were witness in European cities in later years.
- Overfitting & data bias were present as certain weather stations (such as Sonneblick) had 100% prediction accuracy and indicated that not enough feature variety was present in data from cities.

METHOD SELECTED

- It is recommended that ClimateWins proceeds with the K Nearest Neighbors based for future weather prediction testing.

NEXT STEPS

- Further refine and optimize the KNN model, as it showed the best performance
- Investigate and address potential overfitting issues, particularly for the Sonneblick station data
- Expand the analysis to include more diverse geographical locations
- Consider incorporating additional climate variables to improve prediction accuracy

Thank You

CLIMATEWINS

QUESTIONS?

Please contact me at
kirstenlynncurrie@gmail.com

GITHUB

Feel free to explore the Jupyter
Notebook behind this project on my
[GitHub](https://github.com/kirstencurrie) (github.com/kirstencurrie).

PRESENTATION

[Link](#) to video presentation



Slides Carnival

This presentation template is free for everyone to use thanks to the following:

SlidesCarnival for the presentation template

Pexels for the photos

Happy designing!