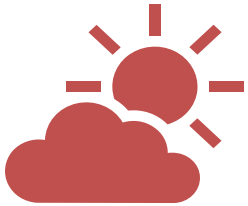
ClimateWins- Supervised Machine Learning in Python

William Zakrzewski

1/5/2026

Objective & Research Hypotheses



Objective:

Predict the consequences of climate change across European cities using Machine Learning.

Hypotheses:

Hypothesis 1: K-Nearest Neighbors (KNN): will outperform Decision Trees and Artificial Neural Network (ANN) in predicting pleasant weather.

Hypothesis 2: Increasing K in KNN improves generalization and reduces overfitting.

Hypothesis 3: Temperature-only models introduce regional bias across climates.

Weather Station Data

Data Source:

- European Climate Assessment & Dataset Project (ECA&D) collected by the European Environment Agency (EEA)
- Climatological dataset from the late 1800s to 2022 across 18 European Cities
- This analysis will primarily examine data between 1960 and 2020 and excludes 3 cities due to data limitations.

Key variables:

- Temperature, Wind speed, Precipitation, and Global Radiation

Data Bias and Accuracy

Historical data & Instrumentation Bias:

- Older measurement practices and instruments used across timespan of study.

Regional Bias:

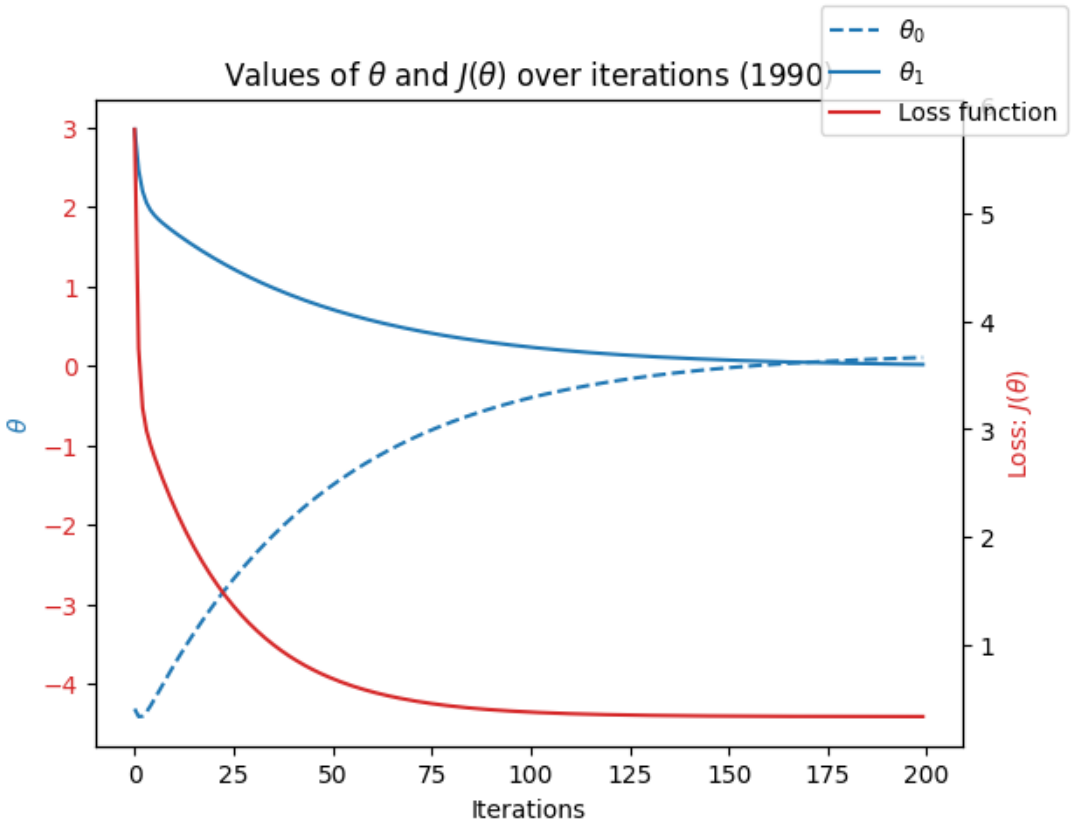
- Regional climate differences and elevation effects for those locations.
- Findings are limited to the European Region.

Data Accuracy:

- Variables are clearly defined by may be impacted by instrumentation bias.
- Cities of Gdansk, Roma, and Tours are missing data relevant for the pleasant vs. unpleasant weather modeling.

Optimization Techniques

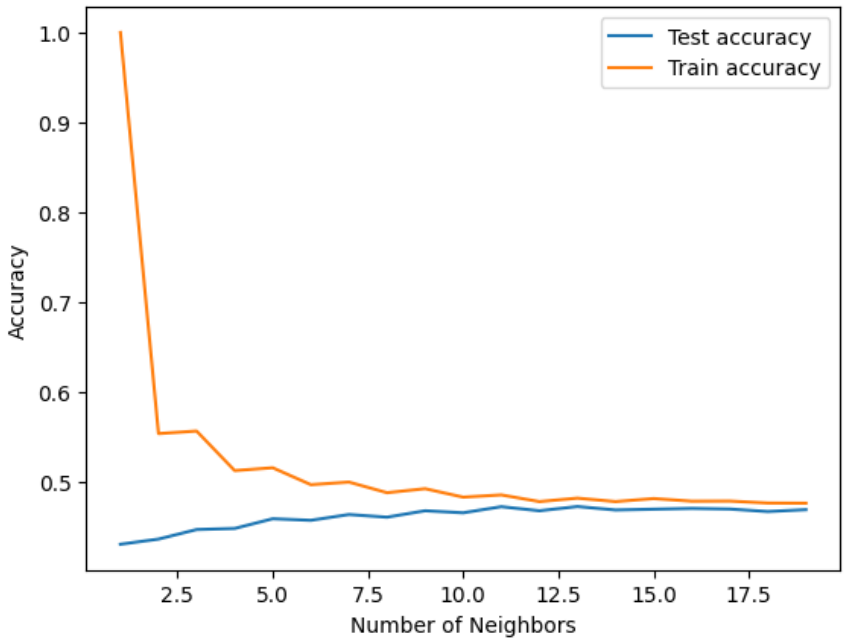
- Data optimization through Gradient Descent.
- This method helps reduce the amount of predicted error in the model.
- Iterations began at 1,1 and oftentimes had around 100 iterations.
- Data converged quickly for all weather stations.



Example: Oslo 1990

Algorithm 1: K-Nearest Neighbors

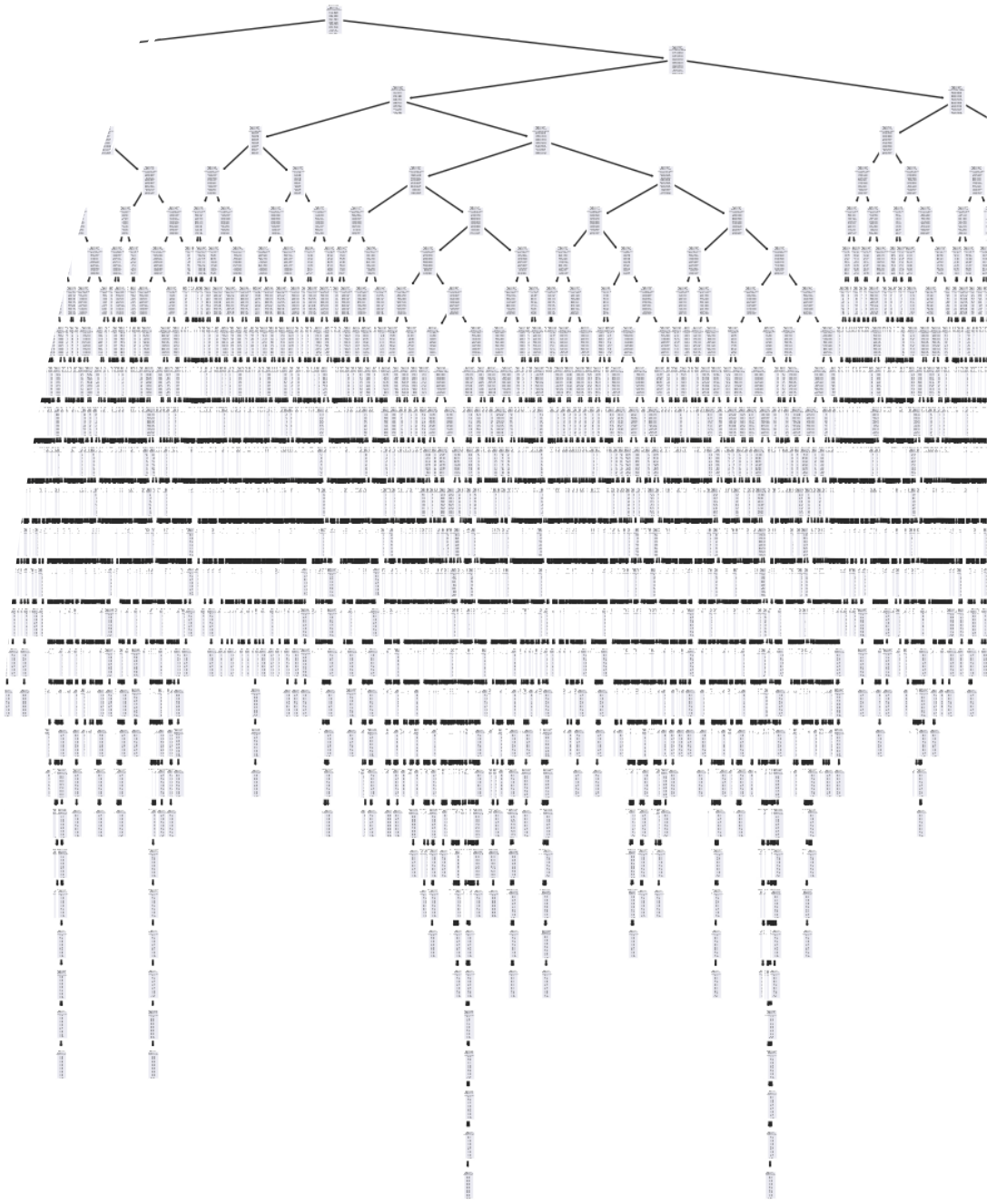
- Measures accuracy for both a train and test set of data.
- Model predicts pleasant weather with ~90% accuracy.
- Model performed well in most cities.
- KNN Accuracy:
  - Tested values 1-19
  - Final Model uses K = 19
  - Smaller K led to lower accuracy.



WEATHER STATION	TRUE NEGATIVES	FALSE POSITIVES	FALSE NEGATIVES	TRUE POSITIVES	PRECISION
BASEL	3978	360	378	1022	87%
BELGRADE	3266	510	317	1645	86%
BUDAPEST	3403	497	293	1545	86%
DEBILT	4437	200	358	743	90%
DUSSELDORF	4238	269	423	808	88%
HEATHROW	4256	314	443	725	87%
KASSEL	4634	181	310	613	91%
LJUBLJANA	3767	428	308	1235	87%
MAASTRICHT	4310	252	332	844	90%
MADRID	2747	421	205	2365	89%
MUNCHENB	4297	249	375	817	89%
OSLO	4727	152	378	481	91%
SONNBLICK	5738	0	0	0	100%
STOCKHOLM	4560	206	352	620	90%
VALENTIA	5454	8	251	25	96%
				Overall Average:	90%

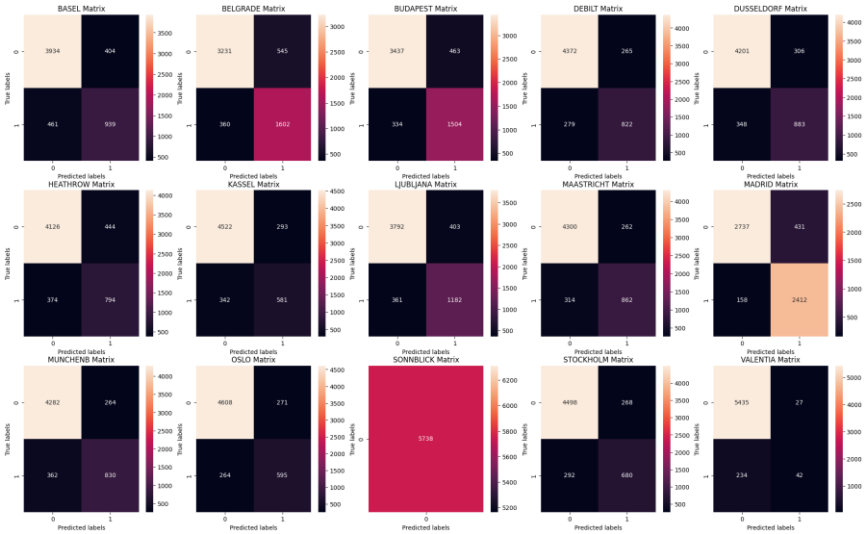
Algorithm 2: Decision Tree

- The decision tree algorithm is extremely vast and difficult to even show on this slide with all of its branches.
- Complex trees rely too heavily on overfitting the model.
- Model accuracy was low:
  - Training Data Accuracy: 46%
  - Test Data Accuracy: 47%



Algorithm 3: Artificial Neural Network

- The ANN model minimizes prediction error and looks to recognize complex patterns through neuron/node development.
- ANN accuracy:
  - Test = 47%
  - Train = 46%



	Hidden Layers	Iterations	Tolerance	Train Accuracy	Test Accuracy
Scenario 1	(5, 5)	500	.0001	.433	.457
Scenario 2	(50, 10, 5)	1000	.005	.445	.473
Scenario 3	(100, 50, 25)	1000	.05	.463	.476

Algorithm Effectiveness by Hypothesis

KNN had the highest model accuracy given the amount of data available. Average accuracy was 90%.

The Decision Tree model was useful for demonstrating the extent of overfitting.

ANN shows potential but requires feature expansion and finetuning. It would likely benefit from additional scenarios and data.

Summary & Future Analysis

- This modeling was conducted for ClimateWins on weather station data across 15 European Cities.
- The KNN supervised ML model is the most accurate and recommended in the given context.
- Creating a model dictated largely by a single parameter does limit its ability to measure across various climates.

Next Steps & Future Analysis

- **Expand available data:**
  - Increase the number of cities analyzed
  - Provide additional continents and parts of the globe
  - Include additional factors to control for regional variation
    - *May resolve inability to measure pleasant weather in SONNBLICK*
- **Review data collection methods across various weather stations**
- **Finetune ANN models after adding more data.**

Questions?

Thank You

William Zakrzewski Jr.
wjzakrzew@gmail.com