ClimateWins Analysis

Predicting Climate Change Trends With Supervised Machine Learning

Laura DeCesare

October 2024

Project Objective

To identify the machine learning approach that most accurately predicts the effects of climate change

Hypotheses

1) Using historical climate data, machine learning methods can predict weather's impact on human populations.
2) Countries and regions with widespread unpleasant weather may be at higher risk of impactful climate effects
3) Changing patterns of inclement weather may predict greater impact of climate change.

The Data

The European Climate Assessment & Data Set project (ECAD) collects data on 13 weather elements from over 23,000 weather stations across Europe. This data set is a sampling of data from 18 of those stations, recorded nearly daily from 1960 to 2022.

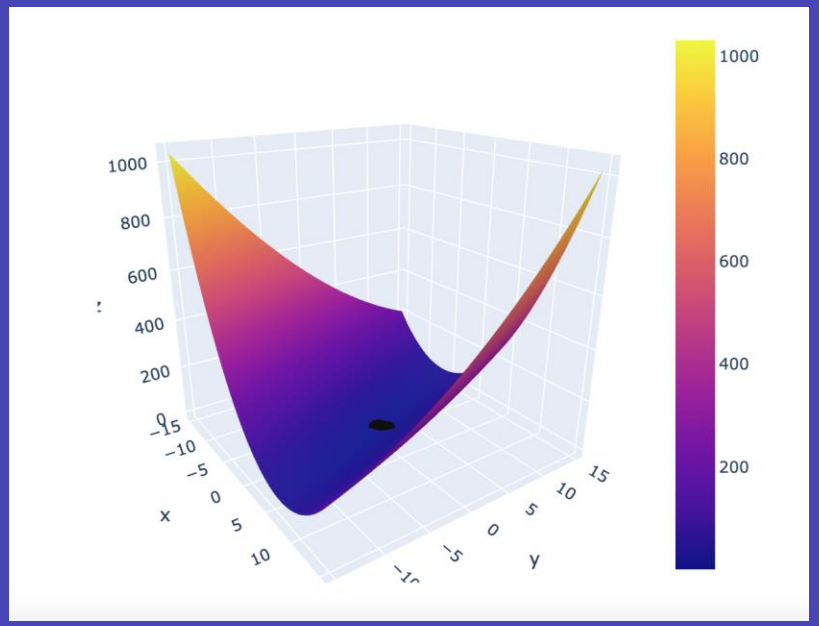
A Note on Accuracy and Bias:

Although this data has a high likelihood of accuracy, coming from moitoreid scientific recording, it is is subject to several potential sources of bias, including:

- **Overgeneralization bias:** Data is only available from European countries, which scientists believe may see fewer climate disasters than the global South. If the data records mild climate impacts, audiences may incorrectly assume the impact is less severe than it is.
- **Implicit bias:** Designations of “pleasant” and “unpleasant” weather are necessarily based on human opinion, which may not be shared.
- **Collection bias:** Improvements in reporting technology since the 1960s may lead to differences in measurements, suggesting trends that don’t exist.

To learn more about these data bias types, see TowardsDataScience.com.

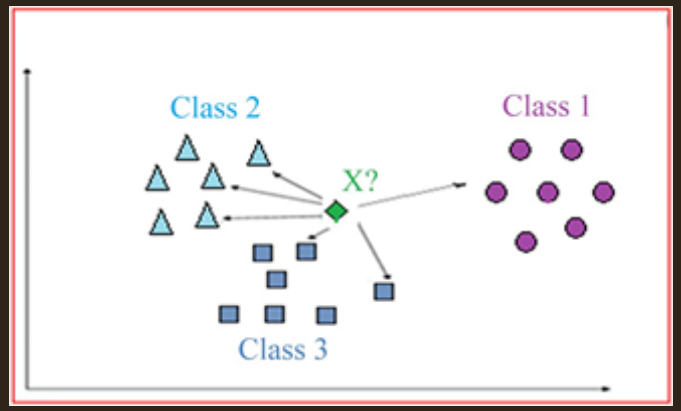
Data Optimization



Optimization is the first step in the machine learning process. In this case, it fits the model to the data to ensure we use resources efficiently.

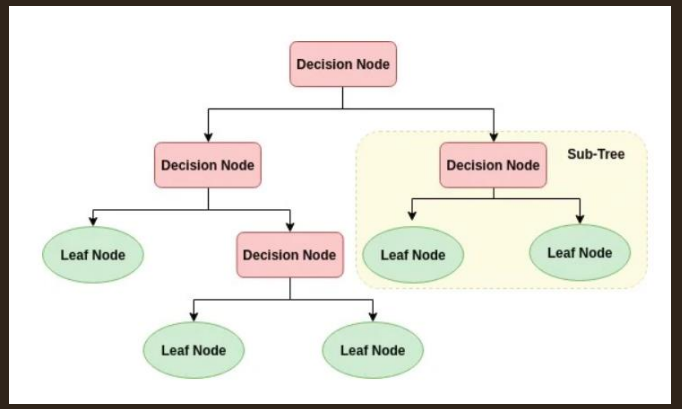
In this case, we used the **gradient descent** method to check the model's fit and identify key features such as the local minimum, pictured here.

Machine Learning Models



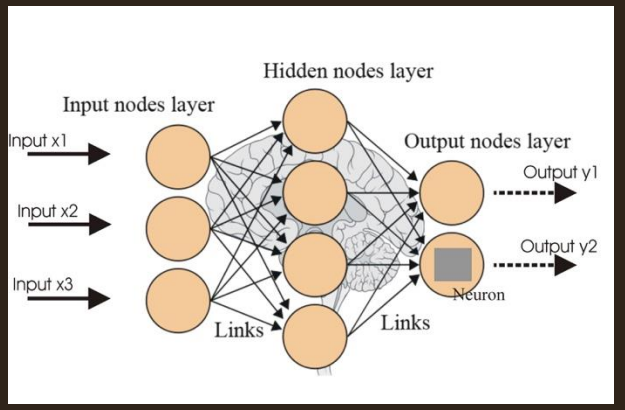
K-Nearest Neighbor:
Categorizes data points based on the classification of nearby known points

Training data accuracy: 48%
Test data accuracy: 47%



Decision Tree:
Finds patterns in the data by evaluating its qualities, as if playing a complex game of "20 Questions."

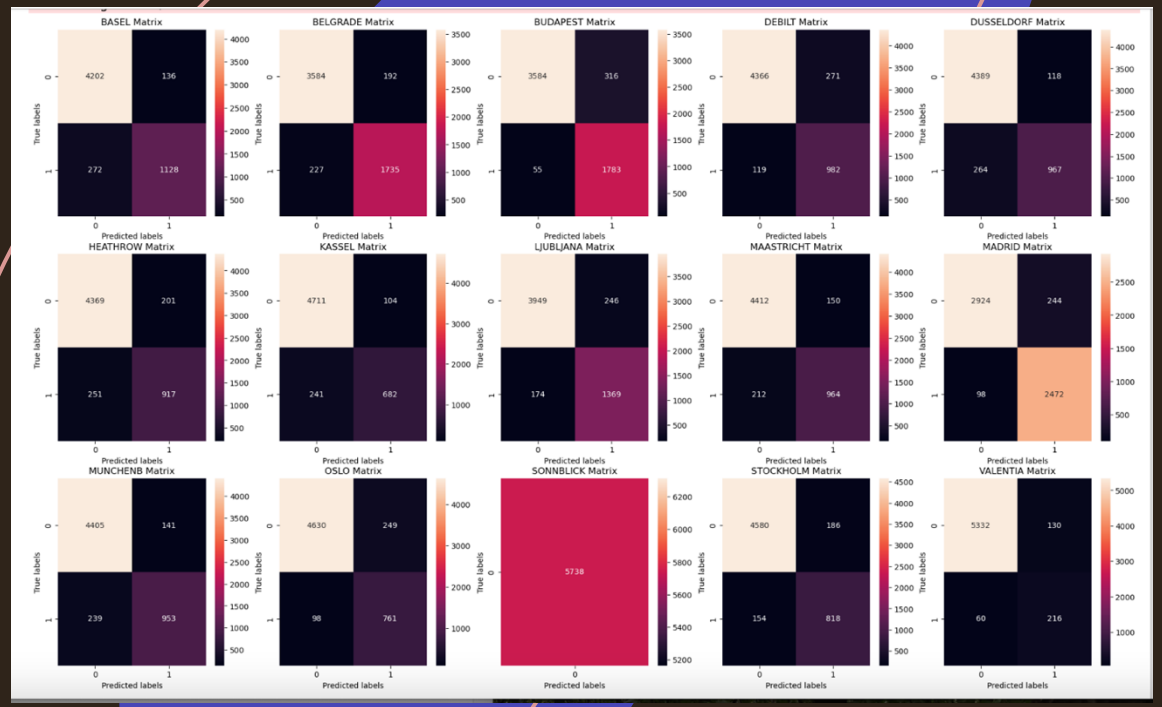
Training data accuracy: 60%
Test data accuracy: 54%



Artificial Neural Network (ANN):
Passes data points through complex decision layers, which generate a final output

Training data accuracy: 66
Test data accuracy: 54%

Conclusions and Recommendations



Primary Hypothesis: "Using historical climate data, machine learning methods can predict weather's impact on human populations."

Conclusion: Each supervised learning method shows promise, with the KNN and Decision Tree models leading for accuracy.

Recommendation:

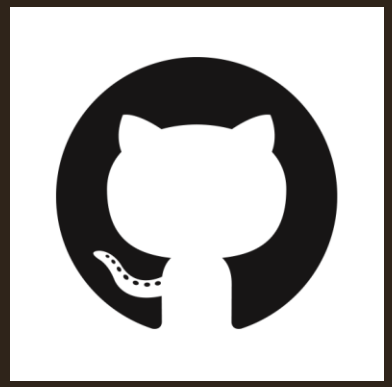
With the highest combined accuracy, the Artificial Neural Network (ANN) model holds the most promise for ClimateWins. As you can see from the matrix displayed here, it predicts weather patterns reliably for 17 of the 18 weather stations. Sonnblick is an outlier due to its mountaintop location.

Next Steps for Continuing Analysis

- Re-run Decision Tree and KNN analyses eliminating Sonnbruck, checking for any improvements in accuracy
- Prune the Decision Tree model and re-assess
- Further refine the ANN model with the assistance of an experienced data scientist, if resources allow
- Experiment with unsupervised machine learning methods and combinations of methods to find the best possible fit

Thank You!

To learn more about the process used in this analysis, see the GitHub repository:



Additional questions? Email me at
lldecesare@gmail.com