ClimateWins

MACHINE LEARNING FOR CLIMATE CHANGE PREDICTION AT CLIMATEWINS

MARIIA ZAREMBA

June 2024

TABLE OF CONTENT

ClimateWins

OBJECTIVE

In this project, I'll use machine learning to help predict the consequences of climate change while working as a data analyst at a European nonprofit organization.

HYPOTHESES

1. Significant changes in maximum and minimum temperatures are observable in historical data, serving as indicators of climate change.
2. Machine learning models can accurately predict whether weather conditions will be favorable on a given day based on historical data.
3. Machine learning models can also be used to predict potential weather dangers (e.g., storms, extreme temperatures), enhancing preparedness and response strategies.

DATA SOURCE, BIASES, AND ACCURACY

DATA SOURCE AND ACCURACY

The data comes from the European Climate Assessment & Data Set project, comprising weather observations from 18 different weather stations across Europe, with records ranging from the late 1800s to 2022.

The overall accuracy of the data is high, given the long history and extensive coverage.

POTENTIAL BIASES:

1. The dataset covers weather observations from 18 different stations across Europe. While this provides a rich source of European climate data, it could introduce regional bias if used to make global climate predictions.

2. The data ranges from the late 1800s to 2022, which could include biases due to changes in data recording technologies and methodologies over time. Early data might be less accurate or incomplete, which can impact the training of machine learning models.

3. Data collection practices and priorities might have varied significantly between locations and times. These differences can introduce inconsistencies and biases into the dataset.

OPTIMIZATION

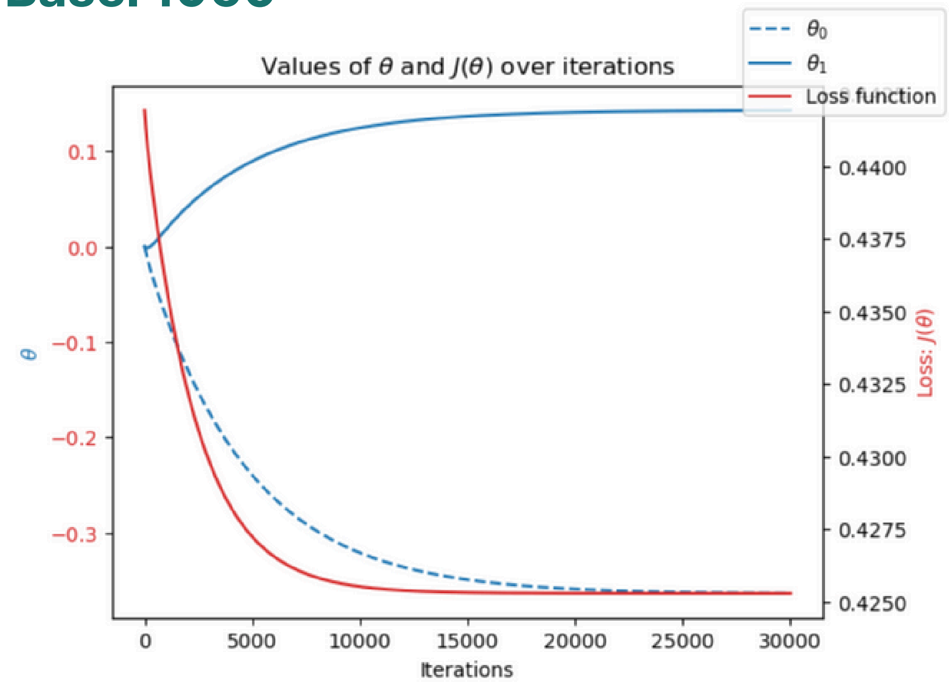
OPTIMIZATION TECHNIQUES:

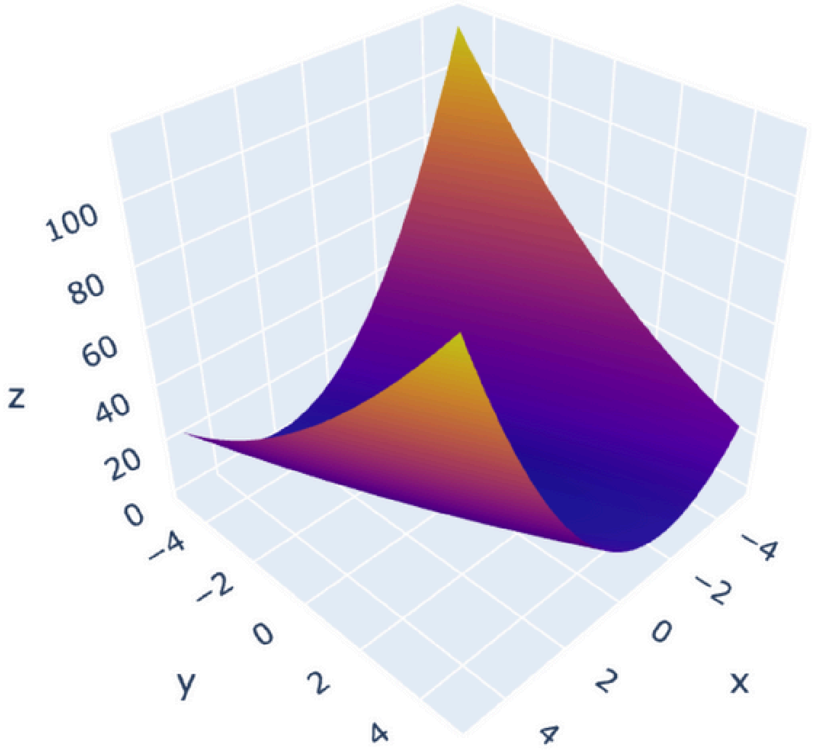
1. I used gradient descent optimization to fine-tune the model parameters (theta0, theta1) for better prediction accuracy.
2. Feature scaling and selection to enhance model performance.
3. Hyperparameter tuning helped find the best model configuration.

Weather Station	Year	Theta0 start	Theta1 start	Theta0 end	Theta1 end	Iterations	Step size
BASEL	1960	0	0	-0.4	0.14	30000	0.001
BASEL	1990	0	0	-0.07	0.07	30000	0.001
BASEL	2020	0	0	0.5	0.99	30000	0.0001
MADRID	1960	0	0	-0.03	0.095	30000	0.001
MADRID	1990	0	0	-0.14	0.11	20000	0.001
MADRID	2020	0	0	-0.07	0.12	2000	0.01

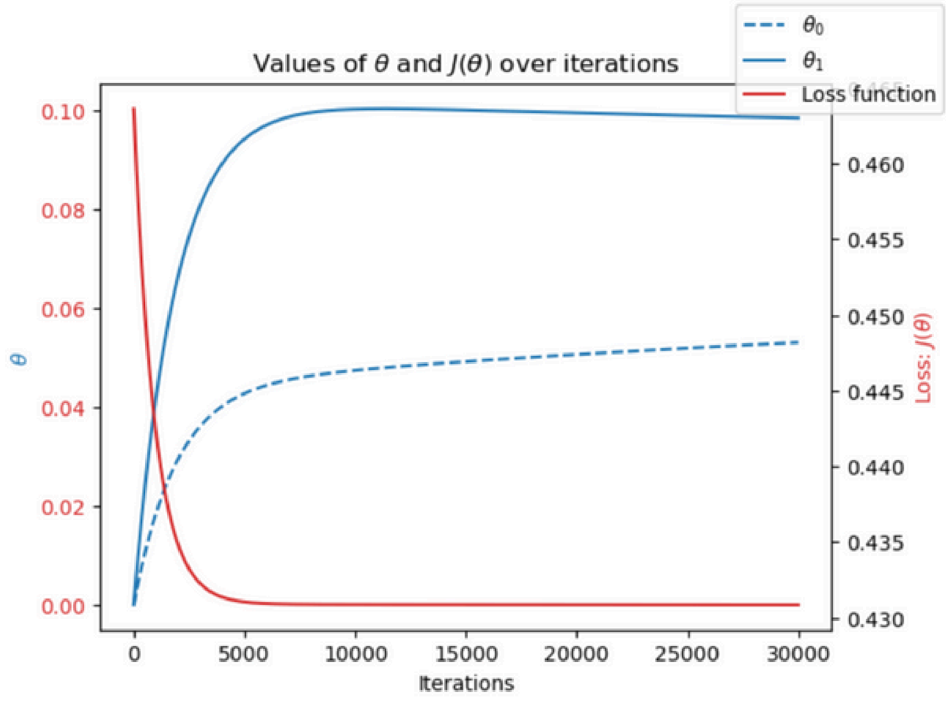
OBSERVATIONS ON MEAN TEMPERATURES OVER 60 YEARS

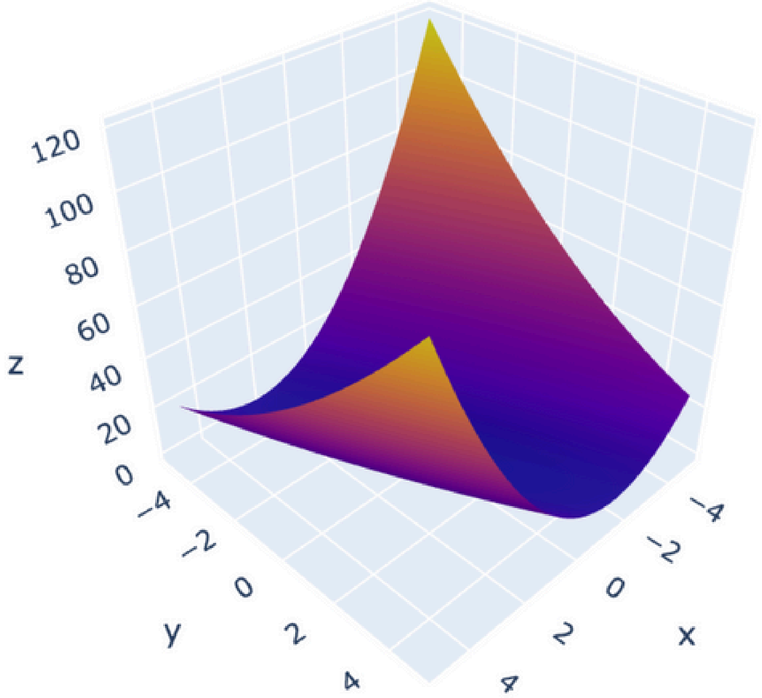
Basel 1960





Basel 2020





01 General Warming Trends:

Across all weather stations, there is a clear general warming trend observed over the past 60 years, indicating a global rise in mean temperatures.

02 Madrid's Moderate Increase:

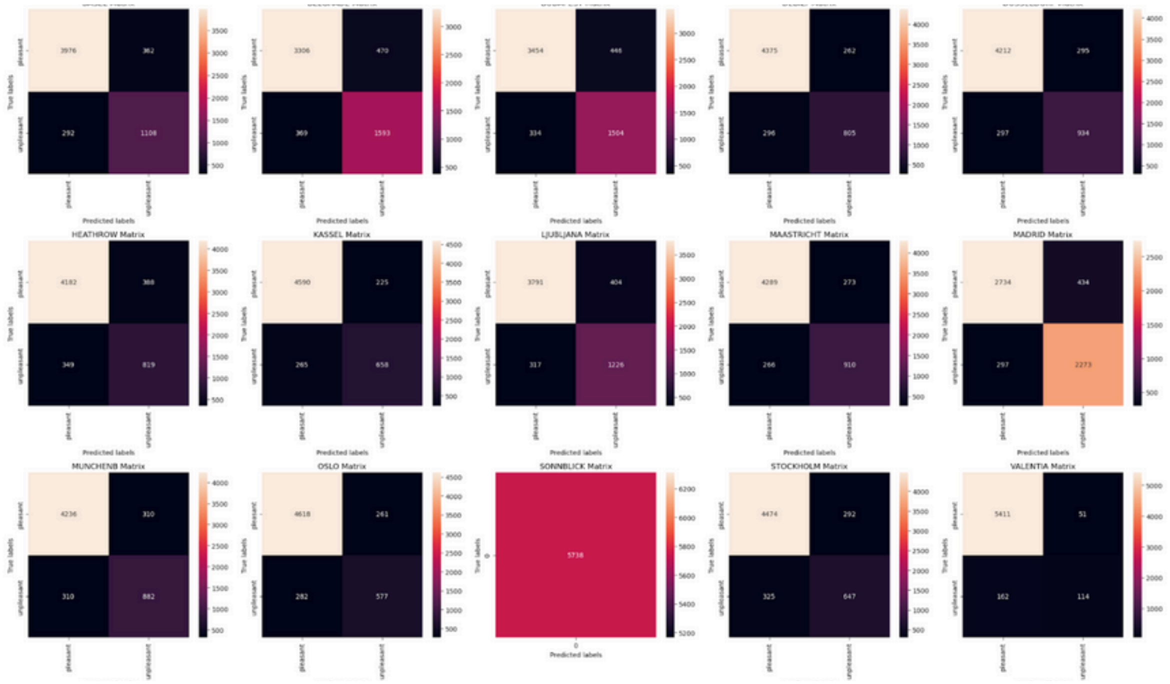
In Madrid, the increase in mean temperatures has been steady but moderate over the 60-year period, reflecting a gradual warming trend.

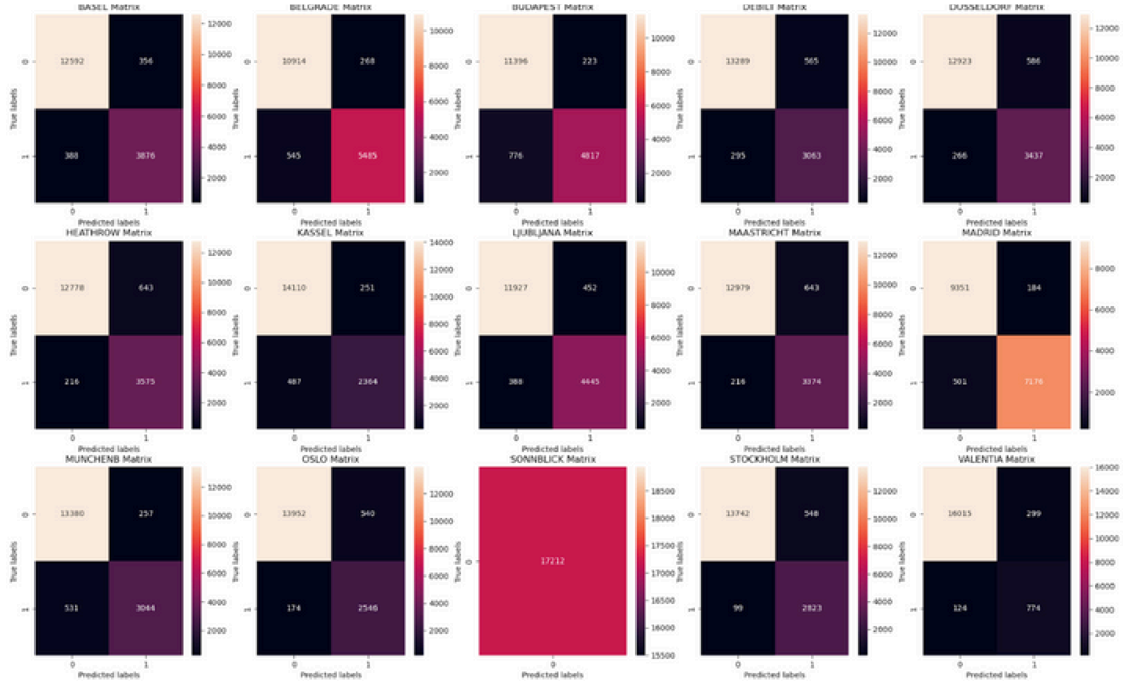
03 Basel's Steeper Positive Trend:

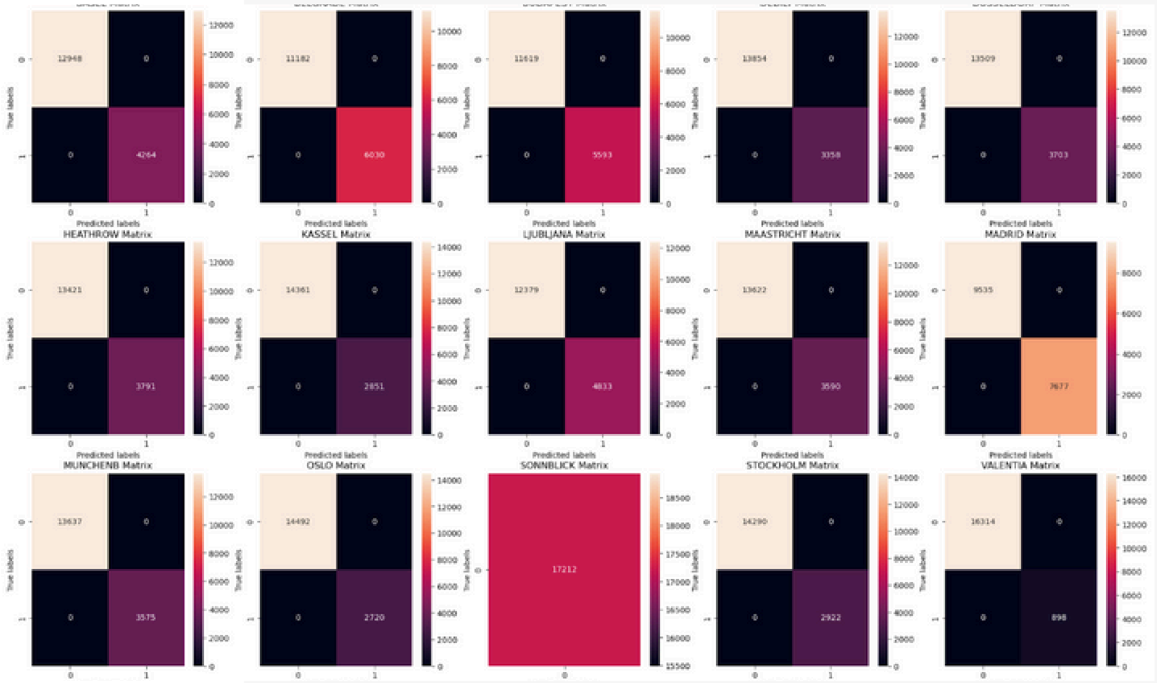
Basel shows a significantly increased temperature over the years, indicating a steeper positive trend in mean temperatures. This suggests that the rate of warming in Basel is higher compared to other stations.

SUPERVISED LEARNING MODELS

ClimateWins







K-Nearest Neighbors (KNN):

Accuracy rates ranging from around 85% to 96% for most weather stations.

The model achieves high accuracy, particularly in cities like Sonnblick (100%) and Valentia (96%)

Decision Tree:

Training data accuracy: 60%

Test data accuracy: 63%

Artificial Neural Network (ANN):

3 layers with 70, 60, 60 nodes each Max iteration: 1000

Tolerance: 0.0001

Accuracy for training data: 84%

Accuracy for testing data: 62%

SUMMARY AND NEXT STEPS

Summary

- Based on the analysis, the K-nearest neighbors (KNN) algorithm appears to best predict the current data, demonstrating accuracy rates between 85% and 96% for most weather stations.
- However, no weather stations are fully accurate, as indicated by the variation in prediction errors across different stations.
- Notably, Sonnblick's perfect accuracy suggests potential overfitting, where the model may be memorizing the training data rather than generalizing well to new, unseen data.
- The disparity in accuracy rates across stations, such as lower accuracy in Belgrade (85%) and Budapest (86%), implies that certain features or noise levels in the data significantly impact overall accuracy.

Next Steps

- Further optimize the ANN model to reduce overfitting.
- Explore ensemble methods that combine KNN and ANN models.
- Conduct localized adjustments for specific weather stations to improve prediction accuracy.

Future Analysis

- Incorporate additional data features such as wind speed and atmospheric pressure.
- Implement cross-validation techniques to ensure model robustness and generalization.

ClimateWins

ClimateWins

THANK YOU

FOR YOUR NICE ATTENTION

June 2024