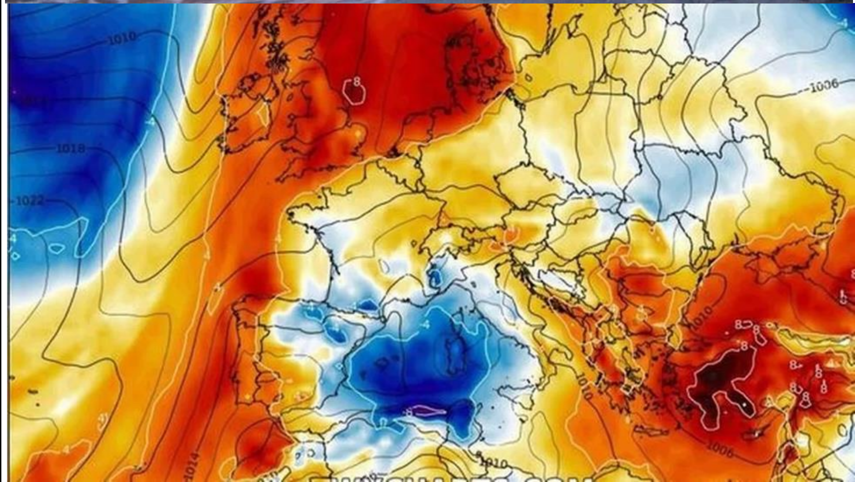
ClimateWins Climate Changes Prediction



By Andrew Fearney

22nd August 2024

Introduction

ClimateWins, a European nonprofit, is exploring the power of **machine learning (ML)** to help predict the consequences of climate change around Europe.

The project aims to predict future climate patterns using machine learning models, and analyse the potential outcomes and risks associated with climate change

Hypothesis

1. The impact of machine learning on the accuracy of predicting a significant rise in daily temperatures in Europe over the next few years

2. The influence of supervised learning on the accuracy of predicting weather conditions compared to unsupervised learning

3. The role of machine learning in predicting extreme weather conditions, such as droughts or extreme winter spells

Questions to Answer

- How is machine learning used? Is it applicable to weather data?
- ClimateWins has heard of ethical concerns surrounding machine learning and AI. Are there any concerns specific to this project?
- Historically, what have the maximums and minimums in temperature been?
- Can machine learning be used to predict whether weather conditions will be favorable on a certain day? (If so, it could also be possible to predict danger.)

Data Sets

- The dataset contains information ranging from the late 1800s to 2022
- The records were taken almost daily, capturing values such as wind speed, temperature, and global radiation, among others
- This data was collected by the **European Climate Assessment & Data Set Project**.
- You can access the dataset here

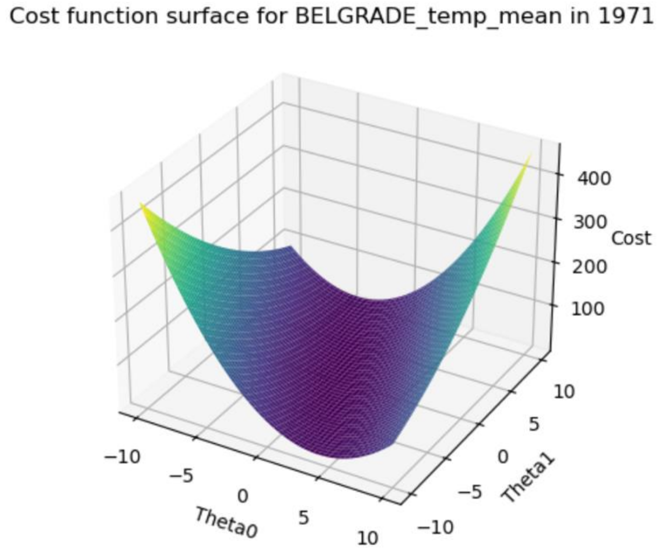
Bias

- **Collection Bias**: Limited to European stations and may not represent global patterns
- **Sampling Bias**: The 18 stations represent a small subset of the 26,321 stations across Europe, possibly leading to skewed results
- **Confirmation Bias**: In this project, analysts might unknowingly favor models or predictions that align with expected outcomes, such as extreme weather increasing due to climate change, without giving equal weight to predictions that suggest otherwise
- **Observer Bias**: This could lead researchers to interpret ambiguous weather patterns in a way that fits their hypothesis, or to overestimate the accuracy of the machine learning models due to expectations about climate change trends

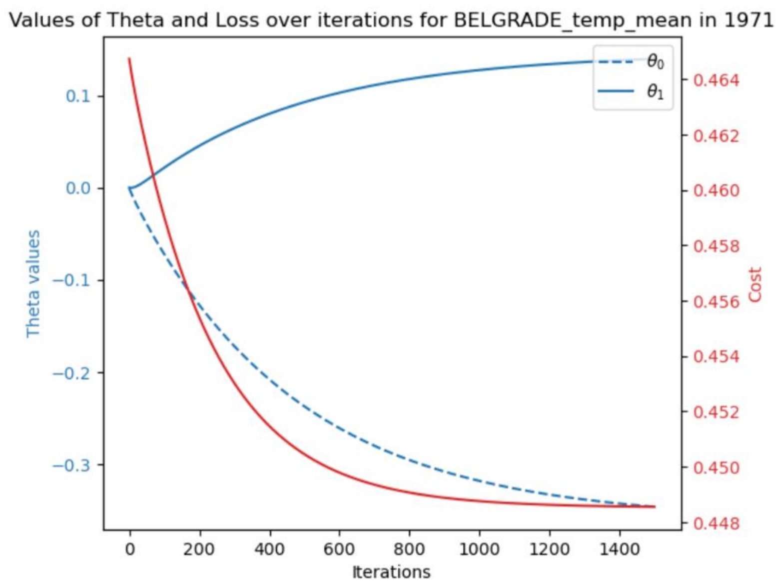
Data Optimisation

- Data optimisation techniques, such as gradient descent, were employed to ensure efficient management and analysis of large climate data sets
- Gradient descent was applied to optimise the accuracy of predictions by minimizing error

Which optimisation techniques were used?

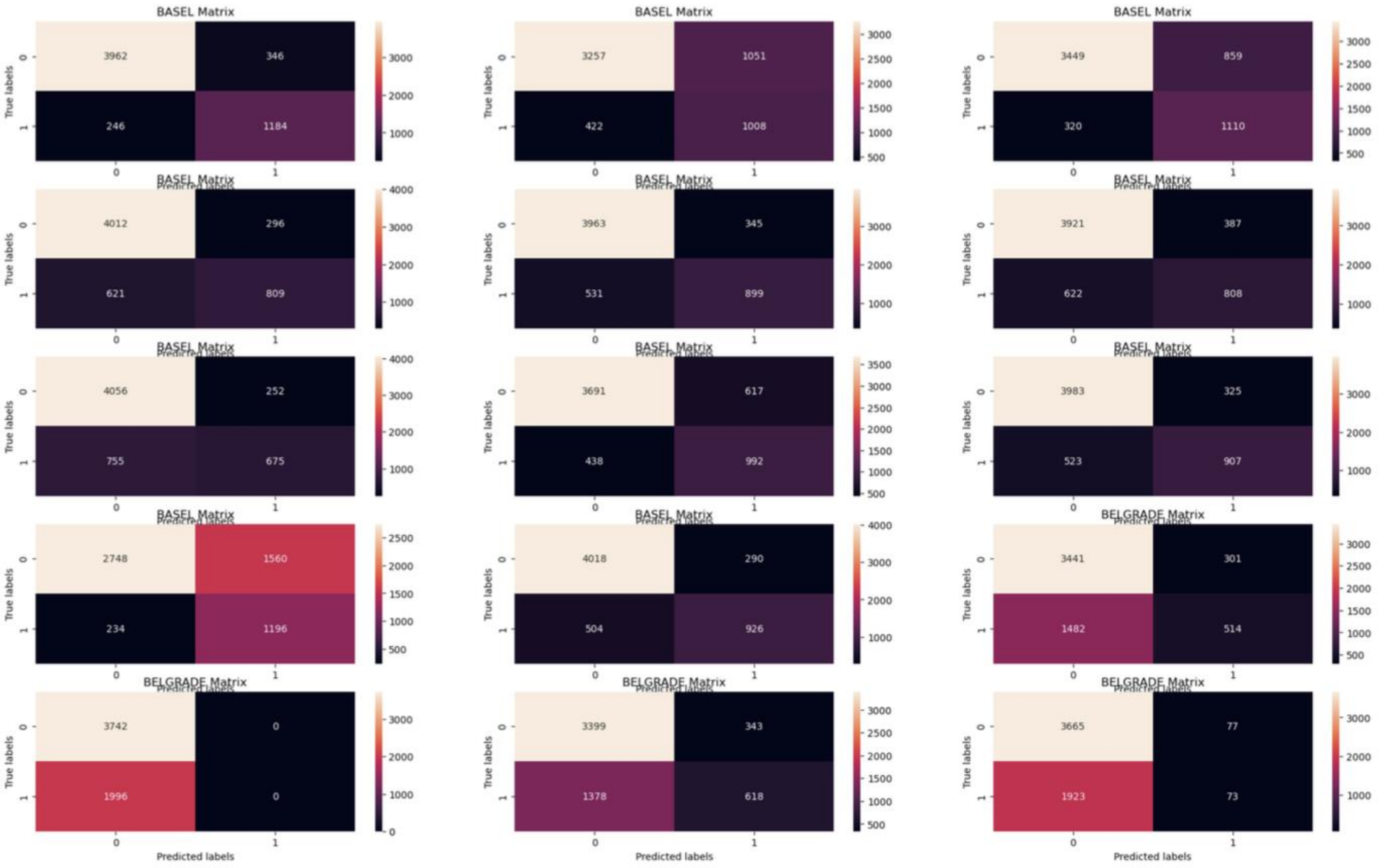


- **Gradient Descent**:
    - An optimization algorithm used to minimize the cost function.
    - The **3D Cost Surface** visualizes the cost changes with respect to $\theta 0\theta 0$ and $\theta 1\theta 1$.
    - The **Loss Function Plot** shows how the cost reduces over iterations, indicating the model learning process.
- **Cost Function**:
    - Measures the error between predicted and actual values.
    - The goal is to adjust $\theta\theta$ values to minimize this error.
- **Theta Updates**:
    - The **Theta Convergence Graph** displays the changing values of $\theta 0\theta 0$ and $\theta 1\theta 1$ as they adjust over time to achieve better predictions.



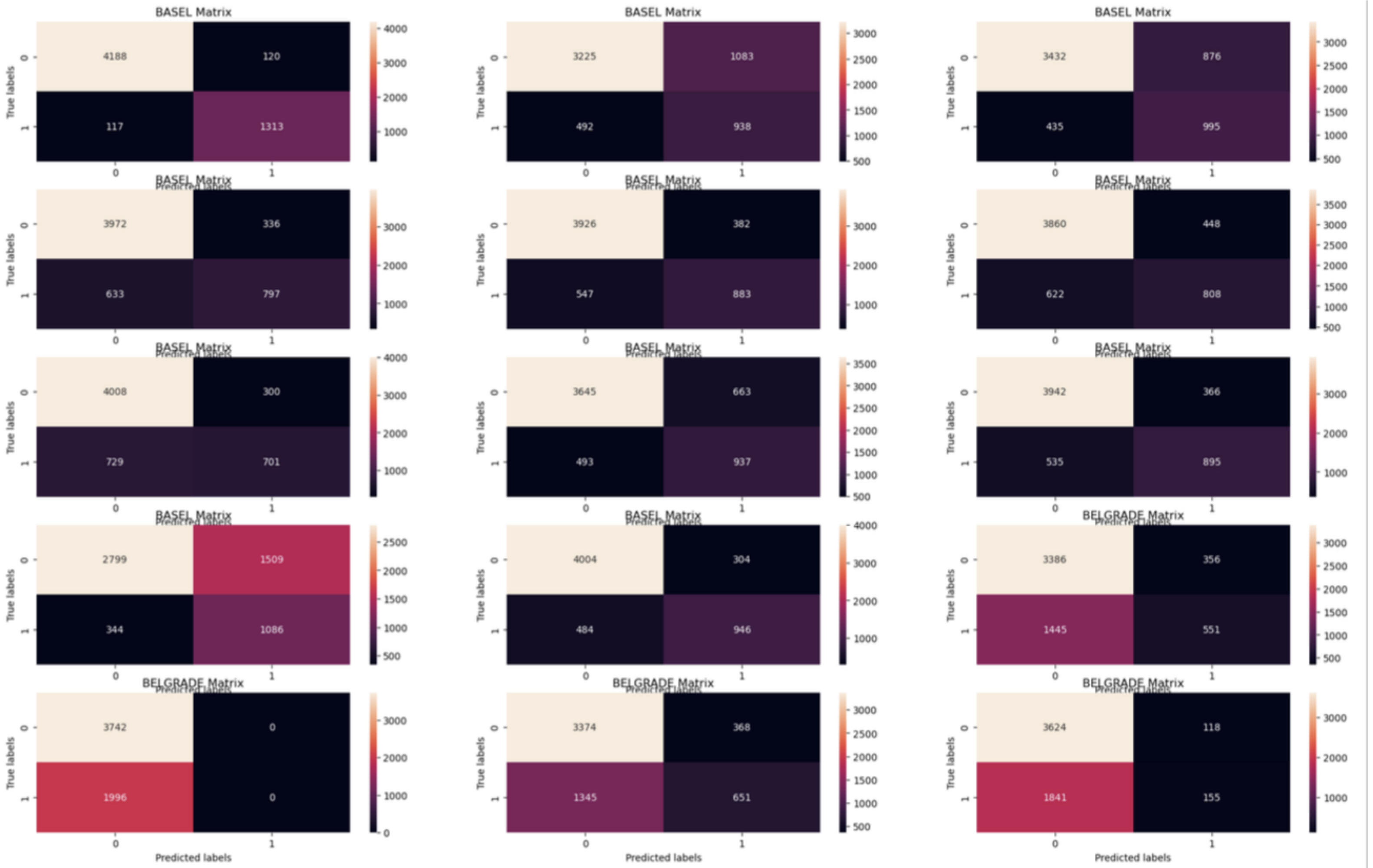
Supervised Learning

- K-Nearest Neighbors (KNN) was applied to categorise weather patterns. Certain irrelevant weather stations, such as Gdansk and Roma Tours, were excluded to improve model accuracy

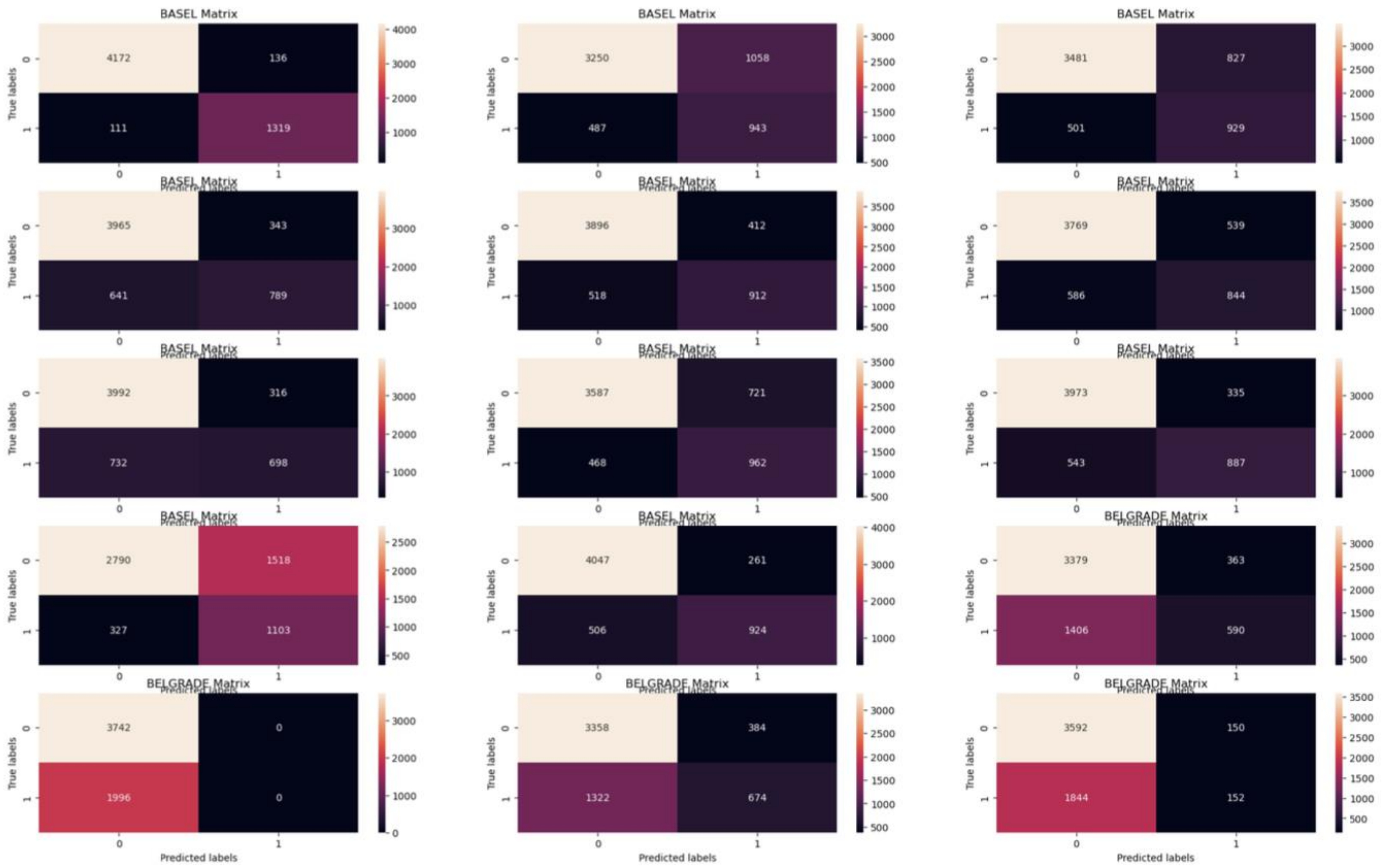


Unsupervised Learning

- A Decision Tree was used for unsupervised learning. The model identified weather patterns but required pruning due to its complexity

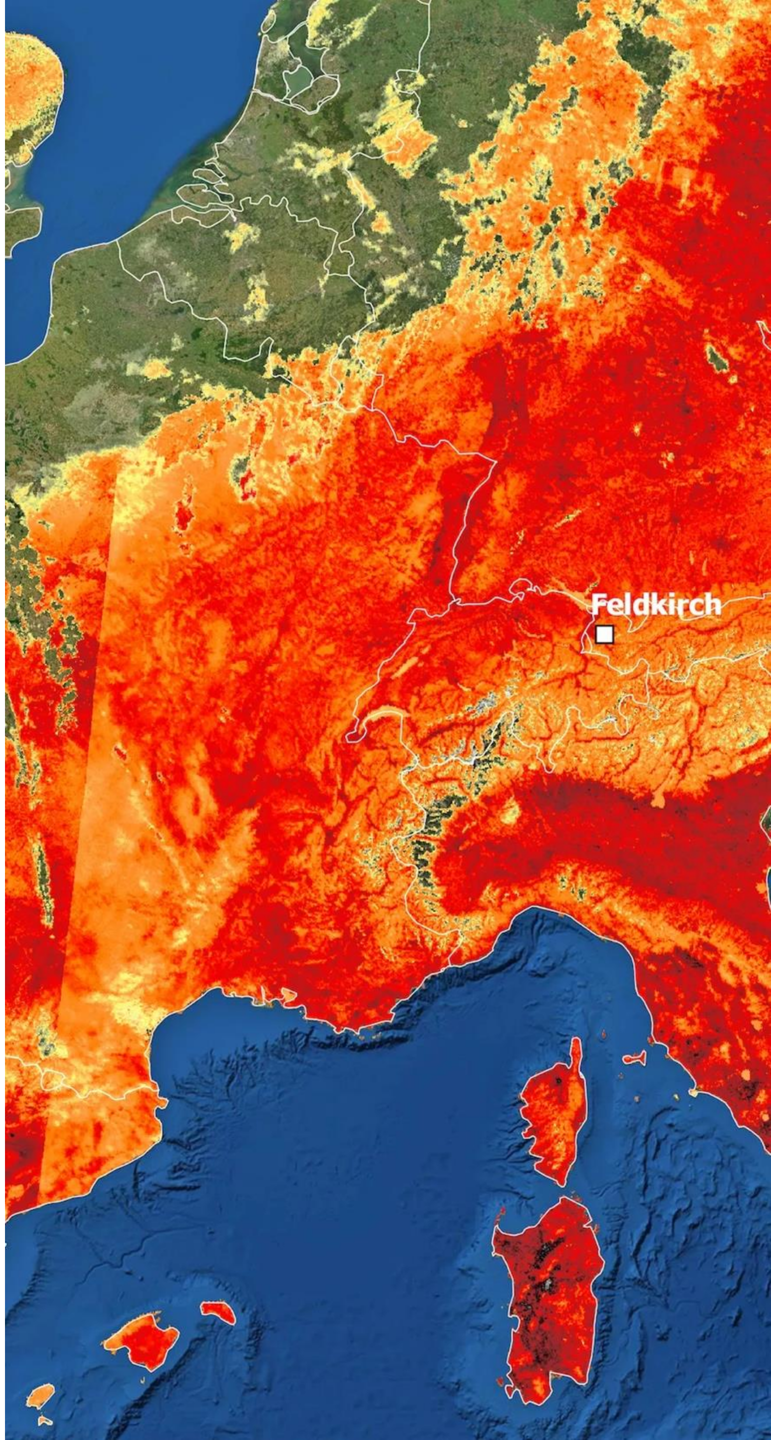


- Training data included input features such as historical weather observations
- The Artificial Neural Network (ANN) model was employed but did not outperform the KNN



Questions to be asked

1. **How reliable is the data from the weather stations in predicting future climate trends?**
    1. Could we enhance predictions by incorporating data from global weather stations?
2. **What are the limitations of using historical data to predict future weather patterns?**
    1. How do sudden shifts in weather extremes (like extreme droughts or flash floods) affect the prediction model’s accuracy?
3. **How can machine learning help identify new weather patterns or anomalies that current models are not picking up?**
4. **Could the models be expanded beyond Europe to include global weather predictions for different regions?**
    1. Would integrating data from Asia, Africa, or the Americas improve model accuracy?



Conclusion

- **KNN** was found to have the highest test accuracy, showing consistency between the training and test sets. It is the most reliable model for predicting climate change outcomes
- However, the ANN and Decision Tree models require further refinement, especially in their ability to generalise and avoid overfitting

Future Work:
•Incorporating **ensemble models** that combine KNN, Decision Trees, and ANNs could enhance overall prediction accuracy.
•Expanding the dataset to include global weather stations, oceanic data, and more diverse climate indicators would likely improve model performance and provide a more comprehensive view of global climate change


Thank You!