Weather Conditions & Climate Change with ClimateWins

29.08.2024

GIADA GRISO

AGENDA

01

PROJECT BACKGROUND

ClimateWins is a nonprofit organization interested in addressing the challenges posed by climate change, particularly the increase in extreme weather events across mainland Europe in the last 10 to 20 years.

ClimateWins believes that advanced tools, such as **machine learning**, could play a pivotal role in predicting and preparing for such weather extremes.

By leveraging weather data from the past century, ClimateWins hopes to **develop a predictive model** that could offer insights into future weather patterns.

Key questions:

- How is machine learning used in weather forecasting?
- What ethical concerns surround the use of ML and AI?
- What are the historical extremes in temperature?
- Can machine learning be employed to predict weather conditions on specific days?

Objective

Using machine learning to help predict the consequences of climate change around Europe and, potentially, the world.

Data Set

The data set is based on weather observations from 18 different weather stations across Europe, which contain data ranging from 1960 to 2022.

Recordings exist for almost every day with values such as temperature, wind speed, snow, global radiation, and more. This data is collected by the European Climate Assessment & Data Set project.

Hypotheses

1. There is a statistically significant increase in average annual temperatures across Europe in the last two decades.
2. Machine learning models trained on historical data can accurately predict future extreme weather events.
3. Supervised learning models are particularly effective for forecasting whether pleasant weather conditions will occur on a given day.

02

DATA ACCURACY

Potential Issues

- **Sampling bias:** data be collected from 18 specific stations out of 26321 stations all over Europe, potentially misrepresenting the full diversity of weather conditions across the continent.
- **Measurement bias:** changes in instrumentation, measurement standards, or station location over time can make the data less accurate.
- **Historical bias:** older data might be less accurate due to manual recording and less reliable instruments.
- **Confirmation bias:** favouring models or analyses that confirm preconceived beliefs about climate change or weather patterns.

03

OPTIMIZATION METHOD

GOAL

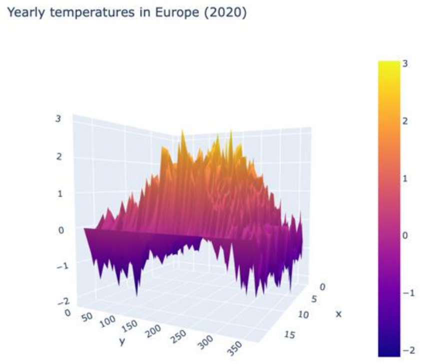
Applying **gradient descent** as a method to optimize the parameters of a predictive model that aimed to minimize the error between predicted and actual temperature values.

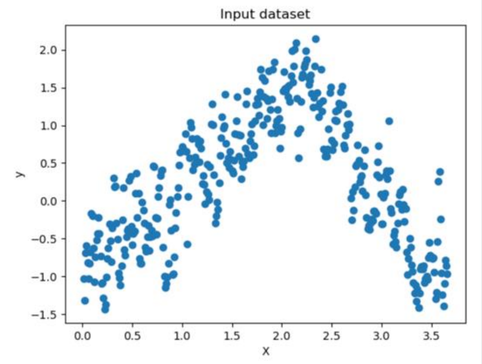
DATA

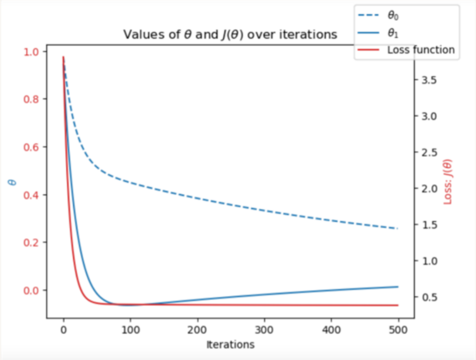
Weather data from three different years (2020, 1990, and 1960) for three different cities (Munich, Stockholm, and Rome) was used for this experiment.

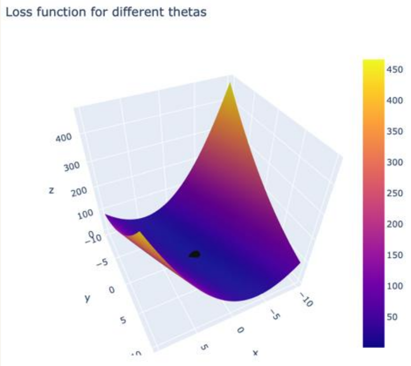
PROCEDURE

- Created 3D visualizations and scatterplots to observe temperature trends over time.
- Created a **loss function** to measure prediction errors.
- Experimented with different configurations to achieve faster and more accurate convergence (local minimum).









04

SUPERVISED MACHINE LEARNING

GOAL

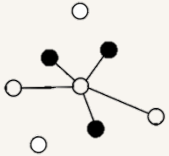
ClimateWins aims to use temperature data from Europe to assess whether it can accurately predict daily weather conditions that are favourable for outdoor activities and everyday life.

APPROACH

We conducted analyses using various ML models to identify the most effective approach for making these predictions. Additionally, we evaluated whether the concept of **"pleasant weather"** can be reliably forecasted based on available weather data. We employed a supplementary dataset to validate our findings.

04

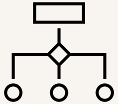
SUPERVISED MACHINE LEARNING



K–NEAREST NEIGHBOR

KNN works by comparing a new data point to nearby data points and assigning it the same label as the majority of its closest neighbors.

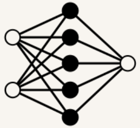
KNN Train Accuracy Score: 50.4%
KNN Test Accuracy Score: 45.4%
Individual stations accuracy: 86.7% – 96.2%



DECISION TREE

A decision tree works by narrowing down a final decision by asking a series of YES/NO questions about the data.

Decision Tree Train Accuracy Score: 61.6%
Decision Tree Test Accuracy Score: 63.0%
Individual stations accuracy: 91.0% – 98.6%



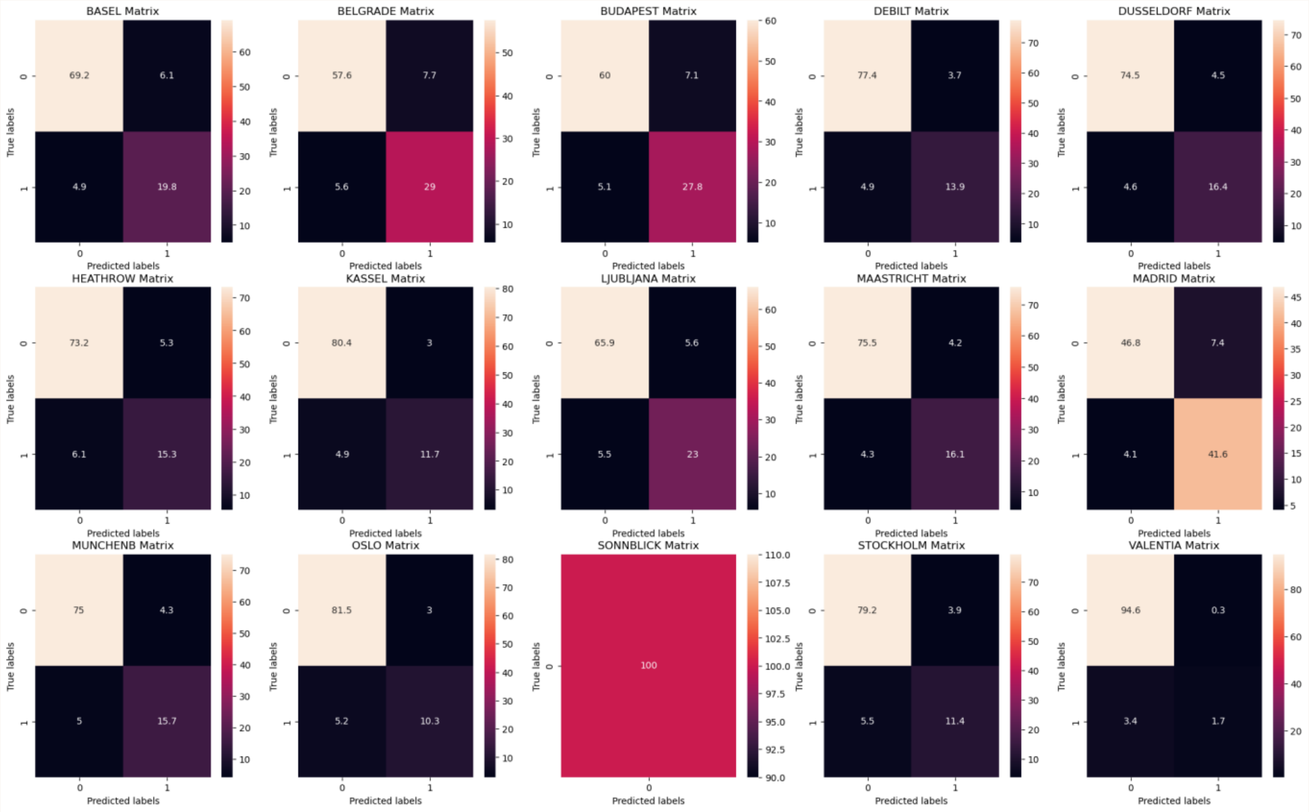
ARTFICIAL NEURAL NETWORK

ANN mimics how the brain works by connecting many small units that learn to recognize patterns in the data by adjusting themselves based on feedback.

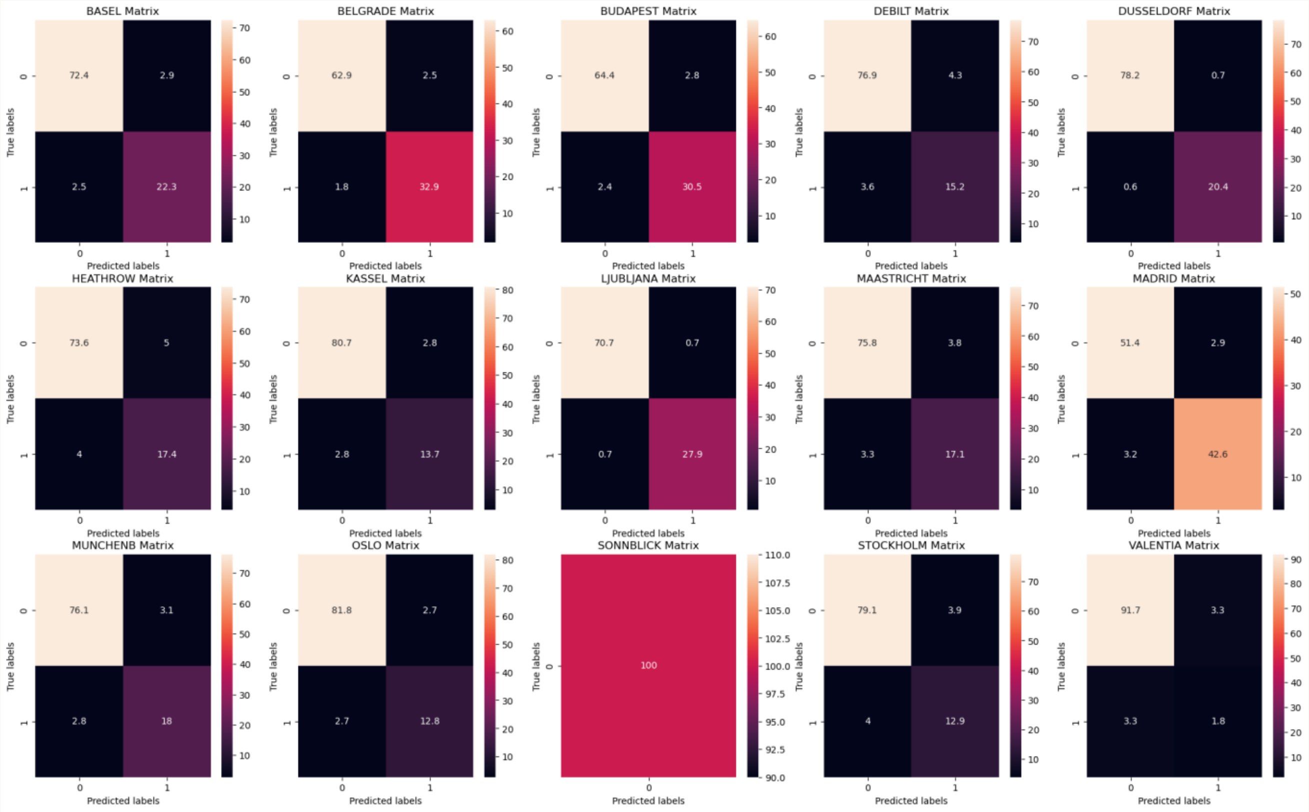
ANN Train Accuracy Score: 86.0%
ANN Test Accuracy Score: 58.5%
Individual stations accuracy: 93.3% – 97.0%

Confusion Matrix of K–Nearest Neighbor

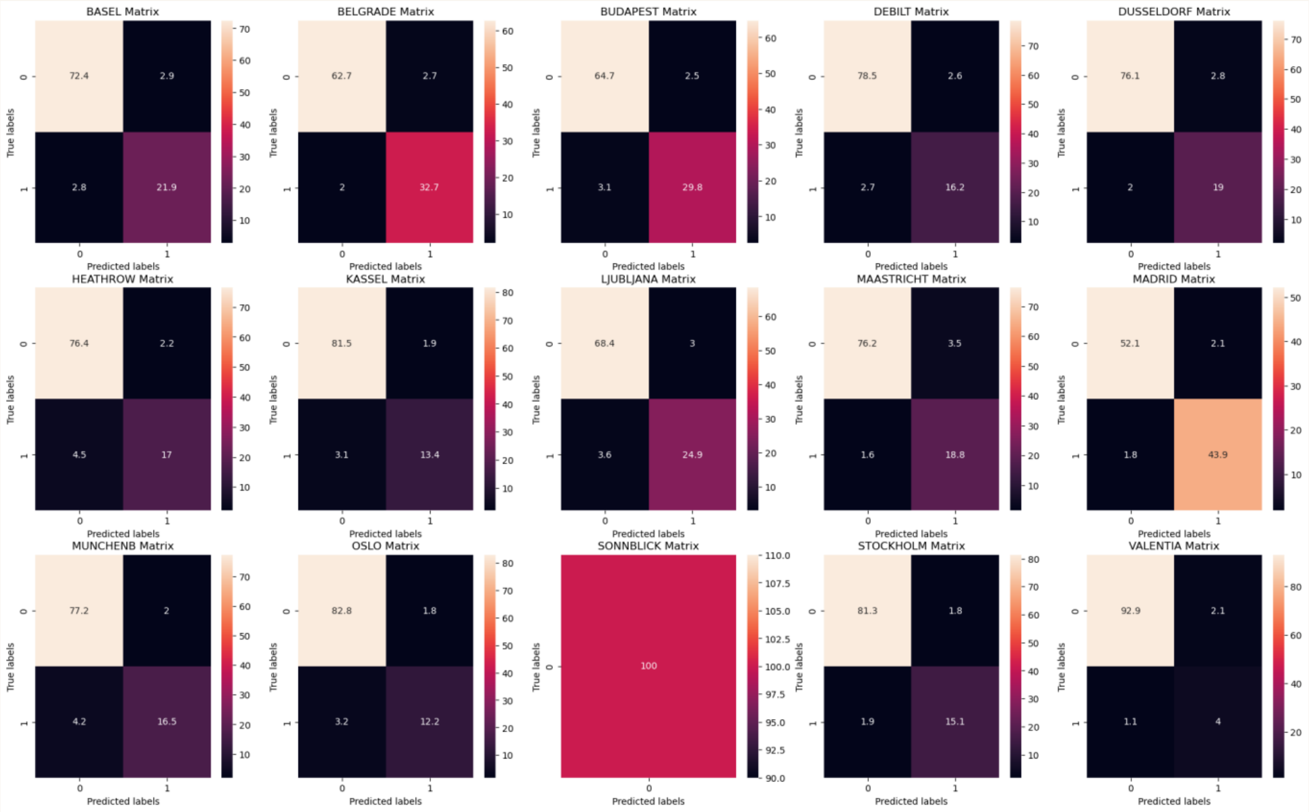
A confusion matrix shows how well a classification model performed by comparing actual and predicted classes, displaying the percentage of correct predictions and errors for each class (true positives, true negatives, false positives, and false negatives). In this case, 1 = pleasant weather, 0 = not pleasant.

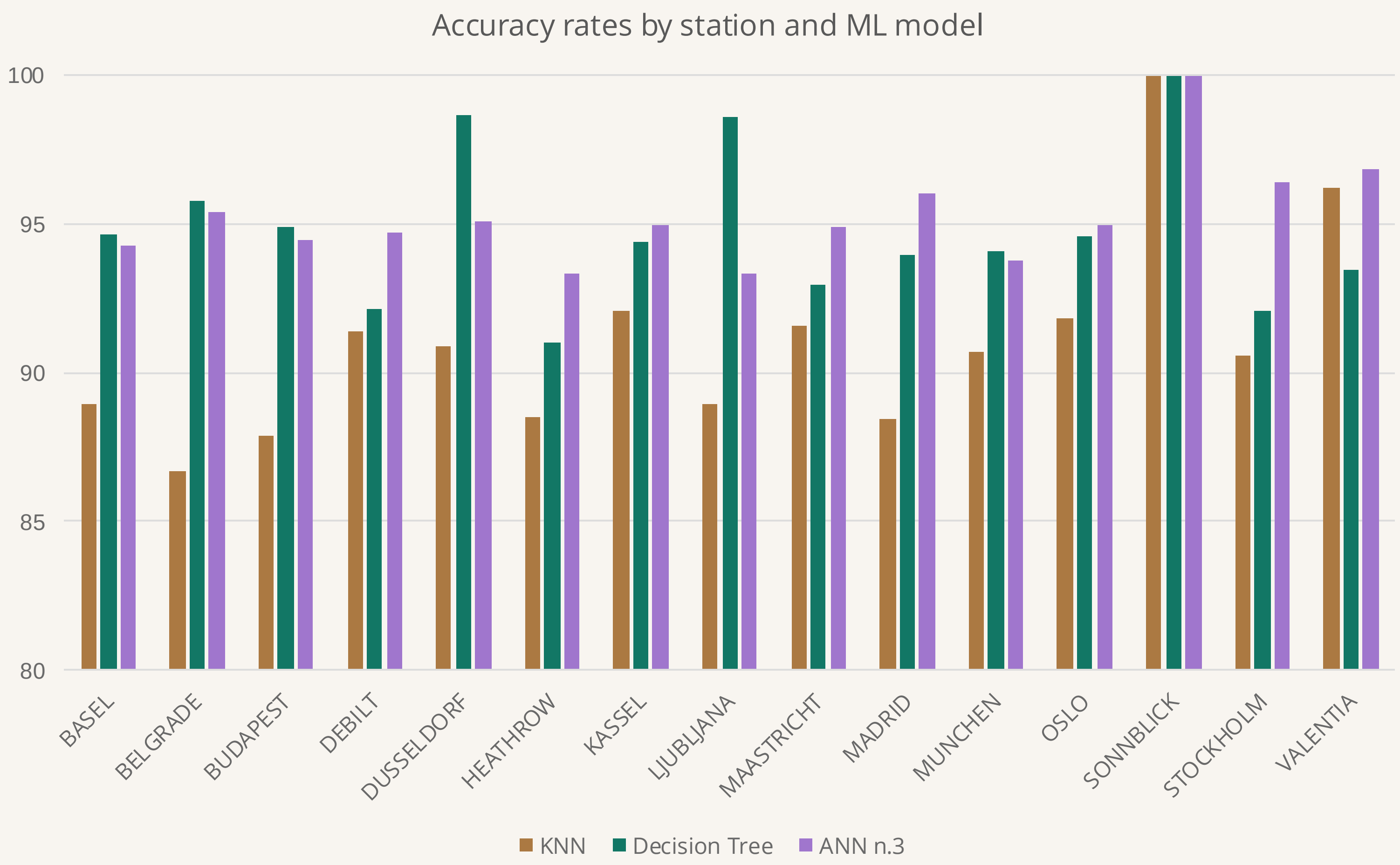


Confusion Matrix of Decision Tree



Confusion Matrix of Artificial Neural Network




Accuracy rates by station and ML model
100
95
90
85
80
BASEL
BELGRADE
BUDAPEST
DEBILT
DUSSELDORF
HEATHROW
KASSEL
LJUBLJANA
MAASTRICHT
MADRID
MUNCHEN
OSLO
SONNBLICK
STOCKHOLM
VALENTIA
KNN
Decision Tree
ANN n.3

05

RESULTS

- Our analysis indicates that supervised learning models can efficiently classify weather conditions as “pleasant” or “unpleasant”, supporting our initial hypothesis.
- The KNN model seems to be **underfitting** as it shows low accuracy on both training (50.4%) and test sets (45.4%).
- The Decision Tree model provides a reliable performance, with a reasonable balance between training accuracy (61.6%) and test accuracy (63.0%).
- The ANN model achieves the highest accuracy on the training set (86.0%) but sees a significant drop in test accuracy (58.5%), suggesting **overfitting**.
- **The Decision Tree is the optimal choice** for ClimateWins because it has the highest test accuracy, and it performs more consistently across both training and test data compared to the other models.

NEXT STEPS

- Adjust hyperparameters – such as tree depth – and pruning strategies to further improve model performance.
- Try out ensemble methods like Random Forests or Gradient Boosting to provide better accuracy and generalization.
- Delve deeper into the model to understand which parameters contribute most significantly to the classification of “pleasant” or “not pleasant” weather.
- Explore unsupervised learning models to uncover previously unrecognized patterns and anomalies in weather data.
- Integrate both supervised and unsupervised methods to develop a comprehensive climate model that predicts specific weather events as well as broader climate trends.

THANKS!

DO YOU HAVE ANY QUESTIONS?

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon** and infographics & images by **Freepik**