CLIMATE WINS

WEATHER CONDITIONS AND CLIMATE CHANGE.

27.9.2025, SUSANNE MAJCUG

BACKGROUND

Due to the increase in extreme weather events, ClimateWins wants to assess the tools availble to categorize and predict the weather in Europe. They want to know if Machine Learning is a suitable tool to use for this. The objective is to create a model that can predict the weather in the future and that can answer a set of key questions.

- How is Machine Learning used and is it applicable to weather data?
- What have max and min temperatures been historically?
- Are there any ethical concerns specific to usage Machine Learning with the suggested data?
- Can Machine Learning be used to predict weather a certain day and potential dangers arising from extreme weather conditions?

HYPOTHESIS

MACHINE LEARNING CAN BE USED TO PREDICT FUTURE WEATHER CONDITIONS.

THE GEOGRAPHICAL LOCATION AND ITS MICROCLIMATE MATTERS WHEN INTERPRETING MACHINE LEARNING RESULTS.

MACHINE LEARNING IS A SUITABLE TOOL FOR PREDICTING WEATHER CONDITIONS.

DATA

Data set: Temperature data set (.csv).

Microsoft Excel
ma Separated Vali

Source: European Climate Assessment & Data Set project.

Description: The data set includes weather observations from 18 different weather stations across Europe. The data is from late 1800s to 2022. Recordings are available for almost every day. Data regarding temperature, wind speed, snow, global radiation amongst others were recorded.

Potential bias:

Relatively small data set.

Different locations, different microclimate.

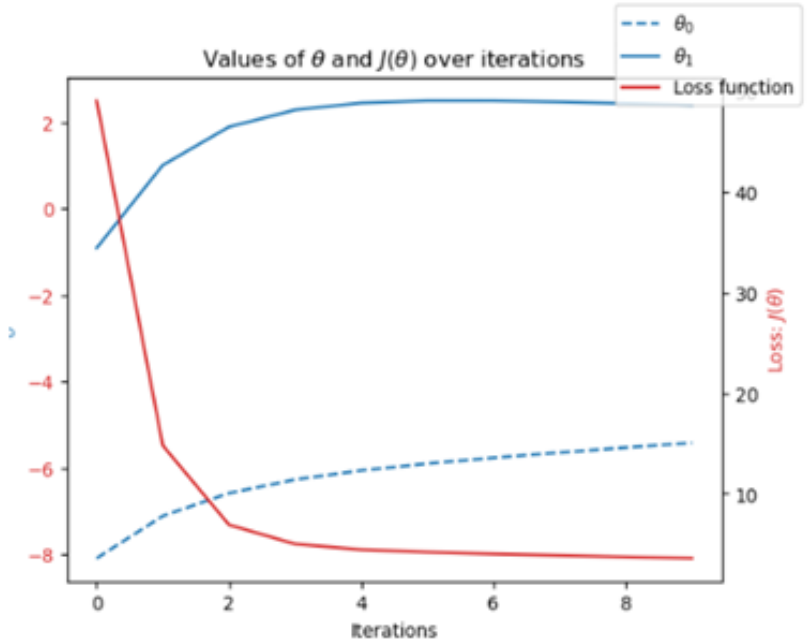
Uncertainity in early recordings (old instruments, procedures, when and how, collection method, collection source).

METHOD: OPTIMIZATION OF DATA

1. OPTIMIZATION OF DATA
2. IDENTIFY SUITABLE ALGORITHM
3. ANALYZE RESULTS
4. CONCLUSIONS

The gradient descent was used to optimize the data. The result show that most losses were eliminated in an early stage when running the function, but after a certain time (see k=2 below) the data losses decreased significantly.

Min and max temperatures are possible to determine, but difficult to interpret when using scaled data. The data set is small enough to use unscaled data.



METHOD: SUITABLE ALGORITHM

1. OPTIMIZATION OF DATA
2. IDENTIFY SUITABLE ALGORITHM
3. ANALYZE RESULTS
4. CONCLUSIONS

Three algorithms were used when analyzing the data and predicting the weather conditions.

1. KNN – K-nearest neighbor
2. Decision Tree
3. ANN – Artificial Neural Network

KNN classifies new data considering where the closest ”k” neighbors are located considering spacial data limits. Works well on small to medium sized numerical data sets with not many outliers.

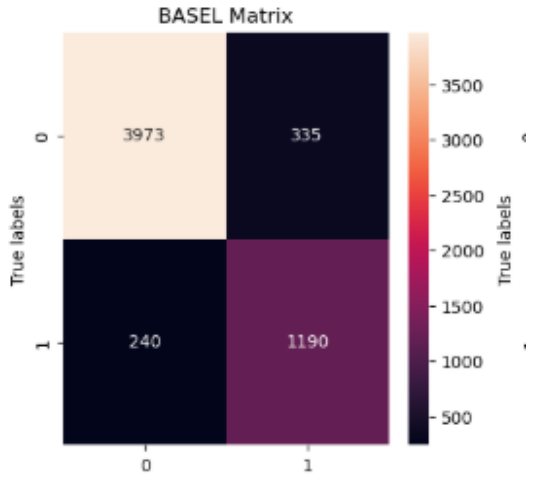
The **decision tree** splits the data into branches based on certain conditions. Works well on non-linear functions and mixed data types. The data size can be small to medium for efficient handling.

ANN is a data network that learns based on the data that it has as an input, and then it adjusts to that. It is the one that mostly mimics human behaviour of the ones mentioned. It works on large data sets and different patterns and images. The disadvantages is that it can be slow.

METHOD: ANALYZE RESULTS

1. OPTIMIZATION OF DATA
2. IDENTIFY SUITABLE ALGORITHM
3. ANALYZE RESULTS
4. CONCLUSIONS

The results of the KNN, decision tree and ANN tests were analyzed and measured based on accuracy rate. The results were presented in a confusion matrix per city.



KNN method samples:

Weather Station	Accurate predictions		False positive	False negative	Accuracy rate
Basel	3981	1016	357	384	85%
Belgrade	3329	1504	447	378	83 %
Budapest	3463	1491	437	347	84 %
Debilt	4388	755	249	346	88 %
Düsseldorf	4215	852	297	379	87 %

Decision tree samples:

Weather Station	Accurate predictions		False positive	False negative	Accuracy rate
Basel	4138	1239	200	161	93%
Belgrade	3643	1852	133	110	96 %
Budapest	3742	1700	158	138	95 %
Debilt	4381	905	256	196	91 %
Düsseldorf	4475	1183	12	48	99 %

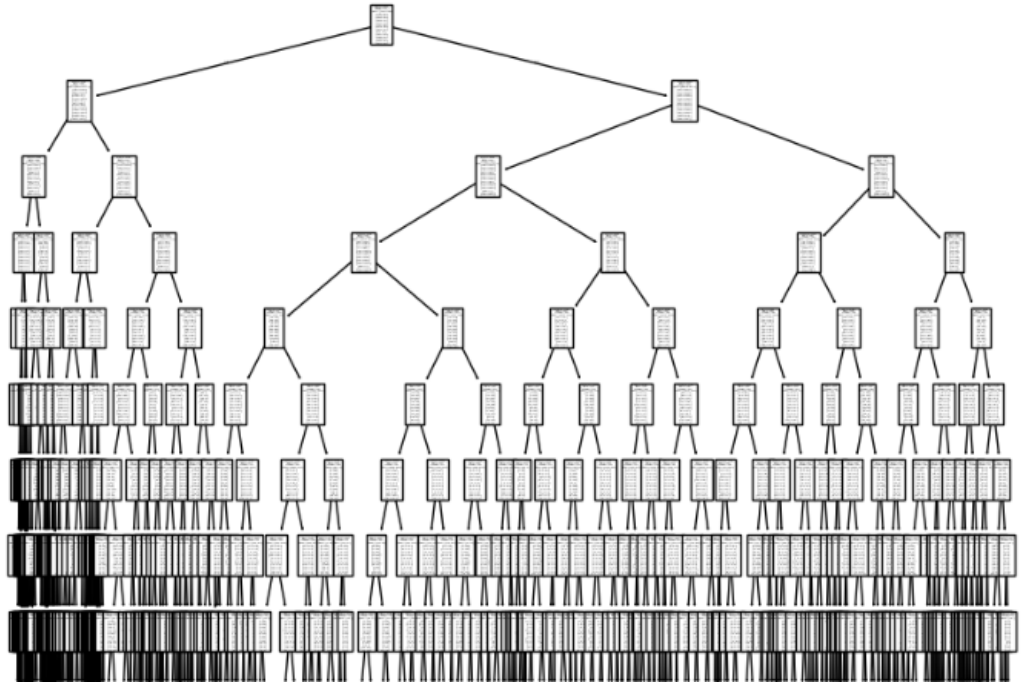
ANN method samples:

Weather Station	Accurate predictions		False positive	False negative	Accuracy rate
Basel	4267	1051	71	349	92%
Belgrade	3668	1751	108	211	94 %
Budapest	3581	1821	319	17	94 %
Debilt	4400	896	237	205	92 %
Düsseldorf	4313	1049	194	162	93 %

METHOD: ANALYZE RESULTS

1. OPTIMIZATION OF DATA
2. IDENTIFY SUITABLE ALGORITHM
3. ANALYZE RESULTS
4. CONCLUSIONS

- The decision tree resulted in the highest test accuracy score, hence should be the model to choose for further processing. The train accuracy score between the ANN and decision tree models were the same.
- The decision tree also has a lower cost of implementation, which is favourable and a factor to consider.



Model	Train accuracy score	Test accuracy score
KNN	53 %	48 %
Decision Tree	60 %	63 %
ANN	60 %	58 %

CONCLUSIONS

Yes! Machine Learning can be used to predict future weather conditions. It has different functions available to cover for needs and different kinds of data (KNN, ANN, decision tree). It is possible to determine and predict temperatures in different locations, and determine the accuracy of the trained and tested results, in order to make them as reliable as possible. In this case, the decision tree model is the best option and most cost efficient one as well!

Yes! The geographical location and its microclimate matters when interpreting machine learning results. Some locations had a higher accuracy rate, which can depend on different factors, such as microclimate and height at which the location is located (Sonnenblick) for example. We saw that different cities had different accuracy rates.

Yes! Machine learning is a suitable tool for predicting weather conditions. However, it is important to choose the correct function when analyzing data and reduce the bias. The size and type of data should be considered when choosing between KNN, decision tree and ANN. The aim is to reduce the gap between trained and tested results.

RECOMMENDATIONS AND NEXT STEPS

1. Include more cities and micro locations across Europe in the data set. A larger base of data can train the model better.

2. Due to the climate changes we can expect more ”noise” in data. Depending on the data size and type of data, it could become relevant to evaulate the model (decision tree or ANN?), but this is uncertain at this point in time.

3. Start looking at combinations of data that can lead to certain weather conditions. For example, will high wind speed and a fast drop in temperature in combination potentially lead to a dangerous weather condition? Is there a need for checking the labeling of data?

4. Further reduce existing bias. For example, measure at the same time and/or more times each day to capture fast changes in weather conditions.

THANK YOU!

Susanne Majcug

susanne.majcug@gmail.com

https://github.com/susanne-maj