CLIMATE WINS

PREDICTING WEATHER CONDITIONS

WITH MACHINE LEARNING

20.10.2025, SUSANNE MAJCUG

BACKGROUND AND OBJECTIVES

Due to the increase in extreme weather events, ClimateWins wants to assess the tools availble to categorize and predict the weather in Europe. Based on previous analysis, Machine Learning has proven to be a suitable tool to use. they now want to create a model that can predict the weather in the future.

- Identify weather patterns outside the regional norm in Europe.

- Determine if unusual weather patterns are increasing.

- Generate possibilities for future weather conditions over the next 25 to 50 years based on current trends.

- Determine the safest places for people to live in Europe over the next 25 to 50 years.

PROPOSED THOUGHT EXPERIMENTS

What if we could detect anomalies in weather patterns based on historical and actual data, and use that to predict future weather conditions?

What if we could use random forests, support vector machines and unsupervised deep learning to make predictions of the climate in different locations? This could be used in areas as infrastructure planning and determine where to direct investments for future urban areas.

What if we could use historical and recent data to predict climate catastrophies, in order to evacuate people on time and have a better disaster prevention?

AVAILABLE MACHINE LEARNING OPTIONS

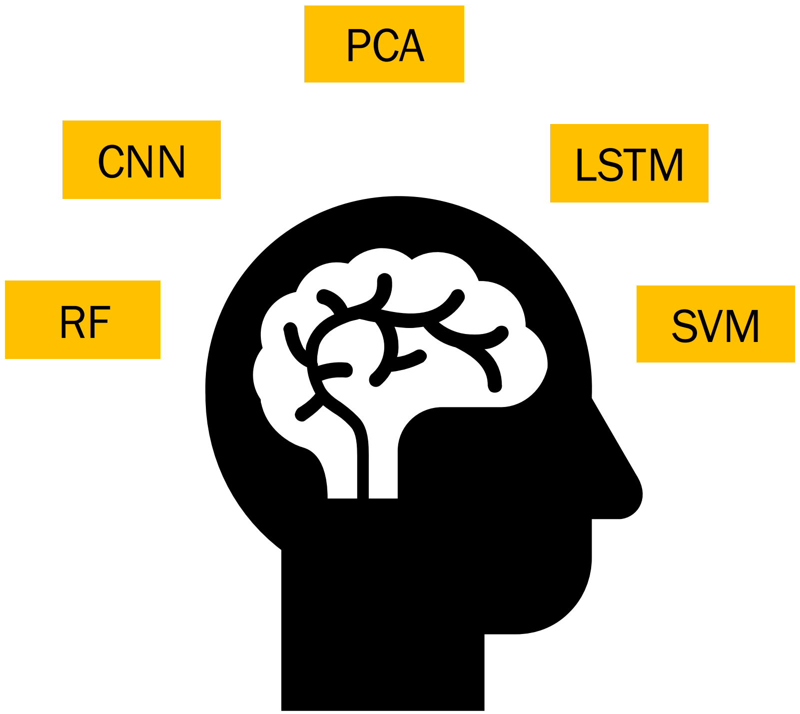
Clustering groups simliar data points together and aims to discover structure and patterns within the data. The results are displayed in a dendogram. There are different methods to perform clustering (Ward method, Complete method, Average method and Single method).

CNN (Convolutional Neural Network) is a deep learning model that is used to determine patterns in spatial data, such as images, videos or grids of numbers. It can detect features, reduce data size without information loss and combine the learned features to make predictions.

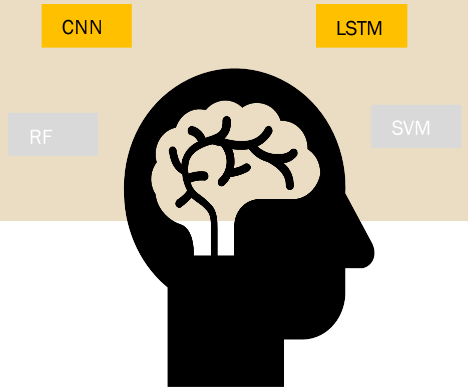
LSTM (Long-Short Term Memory) is a RNN (Recurrent Neural Network) in which order and timing matters. It remembers contexts from earlier inputs to influence later ones, and therefore suitable for sequential and time-dependent data like temperature, humidity and pressure for example. Can be used to model microclimate.

Random Forest is a machine learning algorithm that consists of many decision trees. Each tree makes its own prediction. It is considered stable and accurate due to the fact that a lot of small trees together make the prediction, each one independently for a joint result.

SVM (Support Vector Machine) is a supervised machine learning algorithm that is used for mainly classification tasks (not excluding regression tasks).The goal is to find the best boundary that separates data points of different classes clearly. It can be trained on historical weather data by learning the relationships between parameters (wind, temperature, pressure, precipitation), and in this way, predict future climate.



CLUSTERING



Analysis summary:

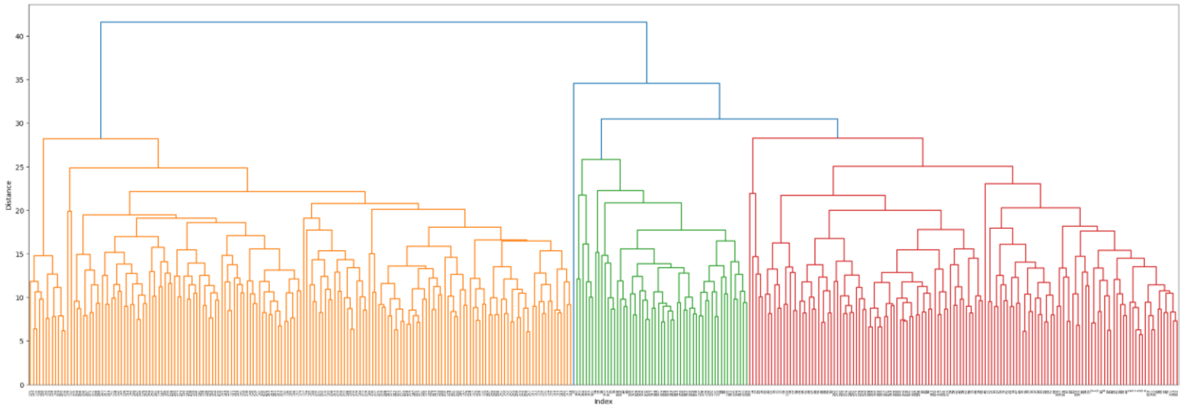
Clustering was used to find patterns or „natural groupings” in the weather data across time and locations. Clustering allows patterns and similarities in the climate to be discovered efficiently.

- Group days/weeks/months with similar weather conditions.
- Identify climate zones.
- Detect unusual weather periods.

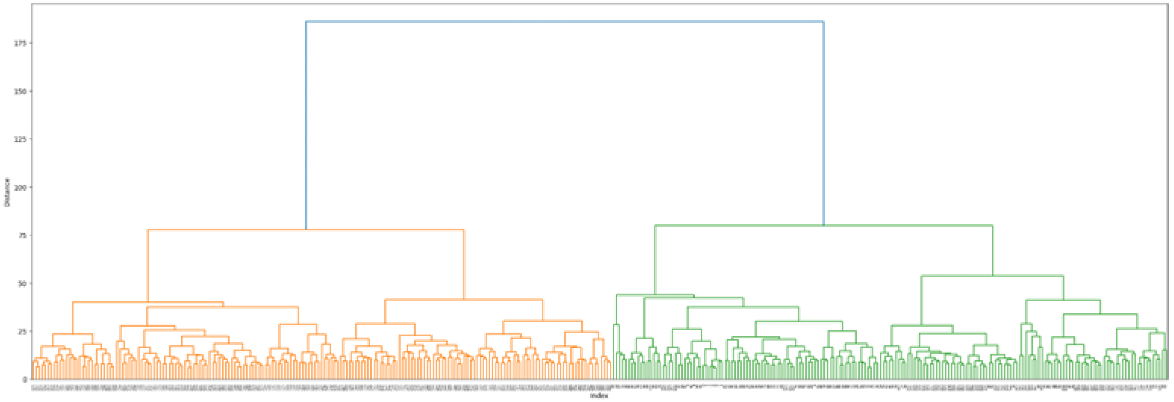
The dendogram produced by the ward method proved to give well separated and compact clusters. The small variance that was achieved gives clusters that are easier to interpret.

By using the PCA method, the data was reduced into small principal components which simplified the data without changing the main direction of the complete data set.

Dendrogram Complete Method for all stations



Dendrogram Ward Method for all stations



CNN

LSTM

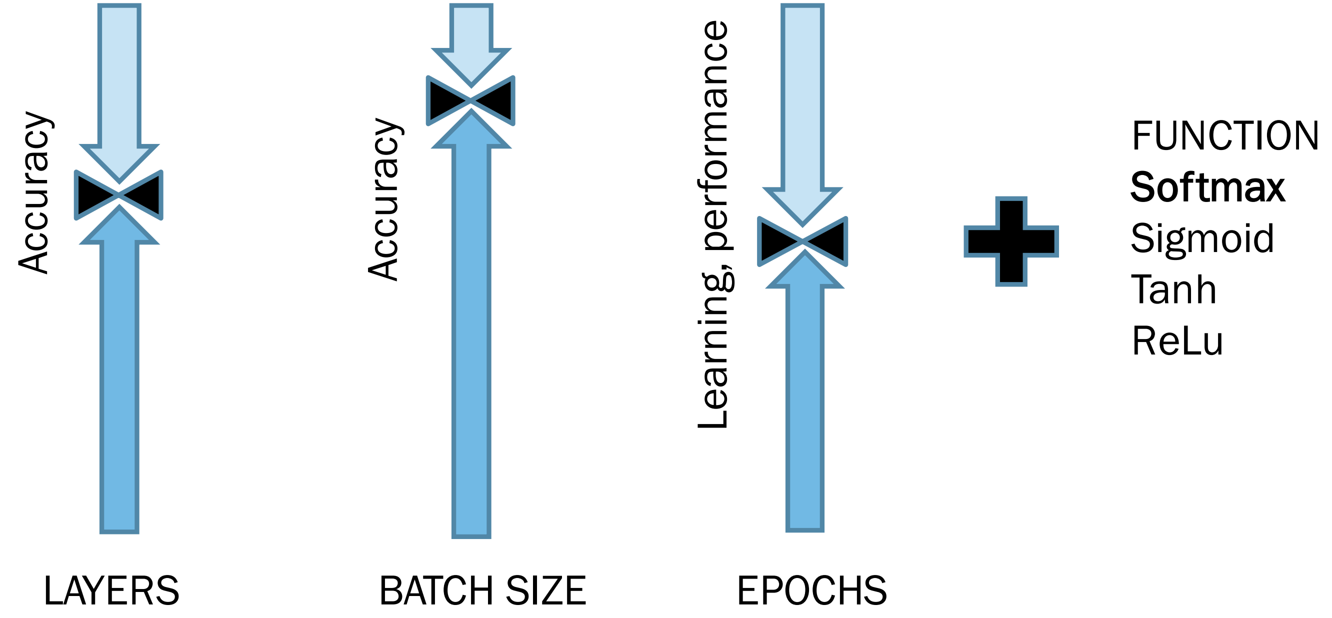
RF

SVM

CNN AND LSTM

Analysis summary:
The confusion matrix using the CNN model resulted in too many misclassifications. Instead, the LSTM/RNN model contributed to a more accurate analysis, even if the result was not good enough to draw any conclusions.

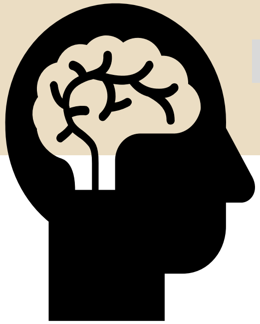
By balancing number of layers, batch size and number of epochs, the optimal balance can be found.



180/180 ━━━━━━━━━━━━━━━━━━━━ 0s 1ms/step

Pred / True	BELGRADE	DEBILT	DUSSELDORF	HEATHROW	KASSEL	LJUBLJANA \
BASEL	60	2360	239	494	493	14
BELGRADE	6	824	45	102	74	19
BUDAPEST	2	156	13	20	17	0
DEBILT	0	78	1	0	2	1
DUSSELDORF	0	26	2	1	0	0
HEATHROW	0	49	6	12	13	1
KASSEL	0	11	0	0	0	0
LJUBLJANA	0	28	2	16	15	0
MAASTRICHT	0	6	1	1	1	0
MADRID	0	165	26	101	145	6
MUNCHENB	0	2	0	4	1	1
OSLO	0	4	0	1	0	0
STOCKHOLM	0	3	0	0	0	1
VALENTIA	0	1	0	0	0	0

Pred / True	MAASTRICHT	MADRID
BASEL	7	15
BELGRADE	2	20
BUDAPEST	1	5
DEBILT	0	0
DUSSELDORF	0	0
HEATHROW	1	0
KASSEL	0	0
LJUBLJANA	0	0
MAASTRICHT	0	0
MADRID	2	13
MUNCHENB	0	0
OSLO	0	0
STOCKHOLM	0	0
VALENTIA	0	0

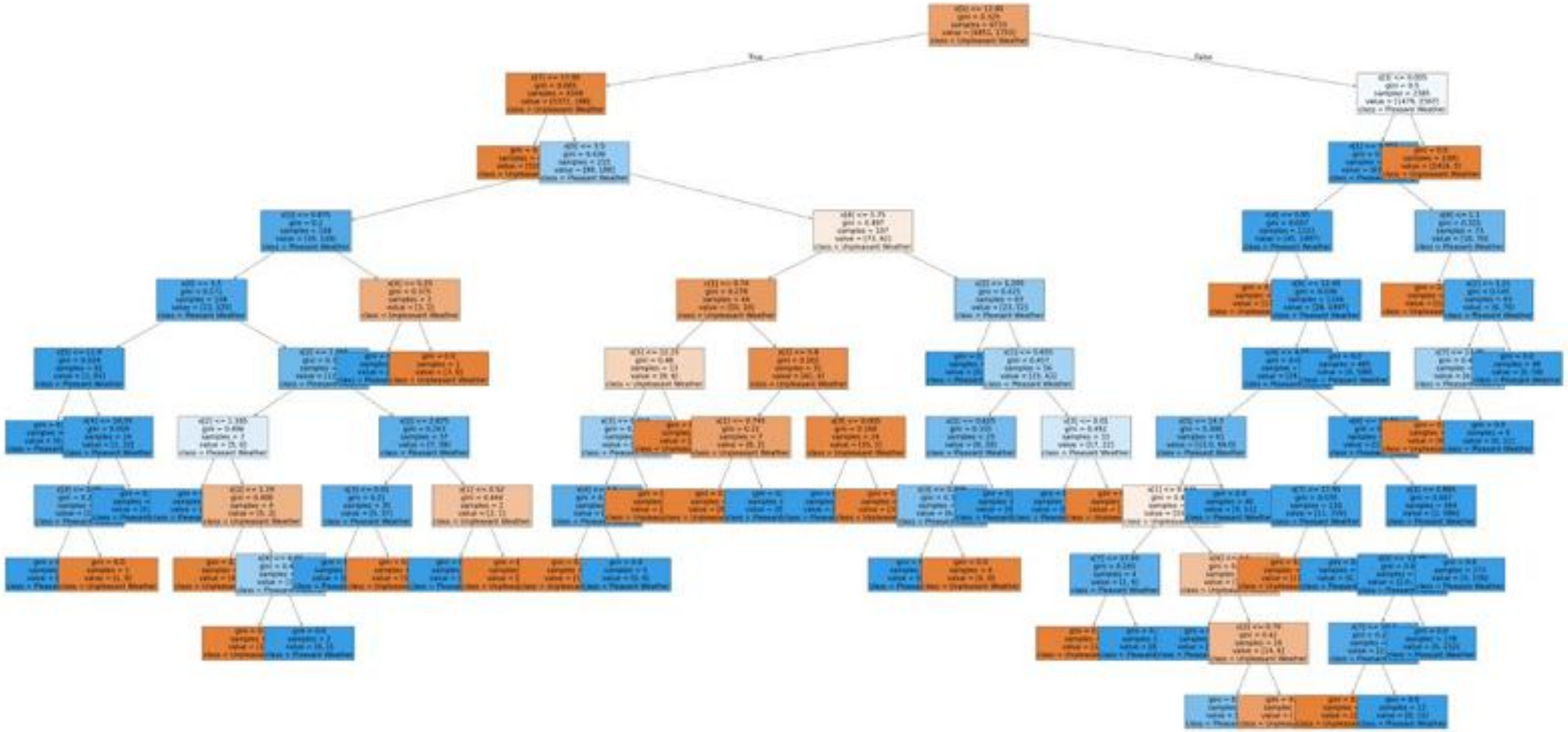


RANDOM FOREST

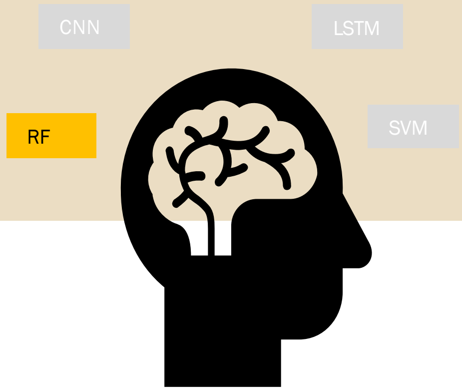
Analysis summary:

The Random Forest achieved a high accuracy score (100 %) when predicting the microclimate in specific locations. However, it achieved a low accuracy score (58.6 %) when measuring across all weather stations.

The high accuracy score can indicate overfitting and that the model memorizes data, which is a risk. The model has to be trained more to become reliable.



FEATURE IMPORTANCE CHART



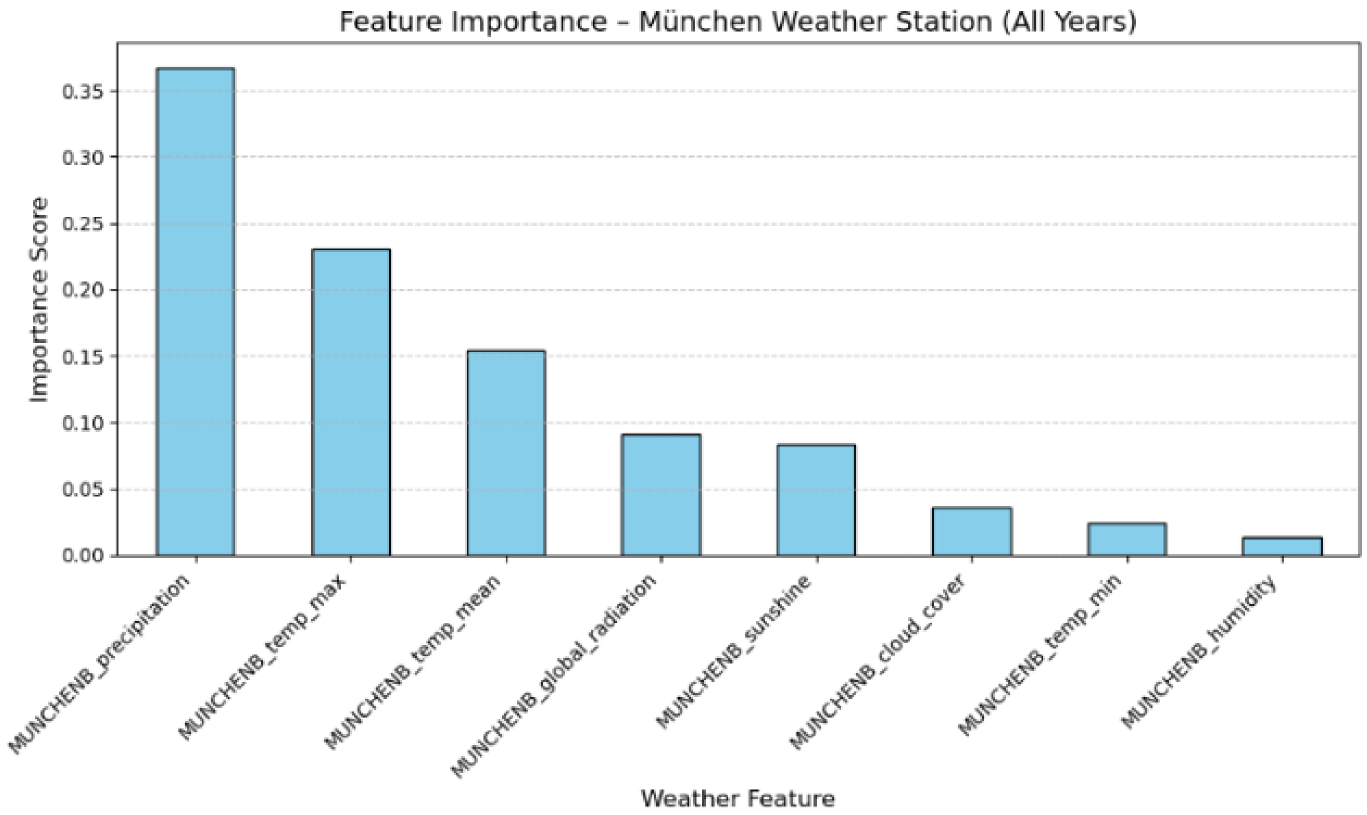
Analysis summary, cont.

The importance bar chart shows what weather that was most influential in determining the prediction. In the case of Munich, it seems like there is more rain than sunshine for example. This should change in a different location, which leads to the question if the same model can be used for different climate zones.

Possible usage of Random Forests and Feature Importance Chart:

Prediction of weather in specific locations.

It can be a powerful tool to use when determining future urban or infrastructure investments. If it rains a lot in an area, water pipe infrastructure that leads away water has to be sufficient and maybe include underground pools. Flooding could be avoided (disaster prevention).



CONCLUSIONS

THOUGHT EXPERIMENT	APPROACH
What if we could detect anomalies in weather patterns based on historical and actual data, and use that to predict future weather conditions?	• Clustering into regions / locations to detect unusual weather periods. • Split data into regions and use as input in next step.
What if we could use random forests, support vector machines and unsupervised deep learning to make predictions of the climate in different locations? This could be used in areas as infrastructure planning and determine where to direct investments for future urban areas.	• Random forests and feature importance charts to predict microclimate. • Merge with infrastructure plans and maps to detect future urban areas. • Visualize in a map.
What if we could use historical and recent data to predict climate catastrophies, in order to evacuate people on time and have a better disaster prevention?	• Use random forests to cluster data to identify dangerous weather conditions. • Continously add new data points and check clustering.

DATA

- ClimateWins weather data set.
- Data gathered from previous climate catastrophies.
- Maps of regions with indication of urban areas, nature, forests, infrastructure.
- City planning maps.

RECOMMENDATIONS

1. Use clustering and random forests to look for 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?
2. By merging the data above with weather data that occured during climate catastrophies, patterns for „dangerous weather conditions” could be identified.
3. By applying the random forest and feature importance index it is most likely to receive accurate data regarding weather conditions in local areas or specific regions. The best results are achieved if the data set is split into areas with similar climate.

THANK YOU

susanne.majcug@gmail.com

https://github.com/susanne-maj