CLIMATEWINS
CLIMATE CHANGES PREDICTION

QUINN HA

26.06.2024

AGENDA

OBJECTIVES & HYPOTHESES

DATA SETS

OPTIMIZATION METHODS

SUPERVISED MACHINE LEARNING

SUMMARY & RECOMMENDATION

OBJECTIVES

ClimateWins is interested in using machine learning to help predict the consequences of climate change around Europe and, potentially, the world.

HYPOTHESES

1. Using historical weather data, machine learning can accurately predict the increasing frequency of extreme weather.
2. Supervised learning models are particularly effective for forecasting specific weather conditions
3. Warmer temperatures correlate positively with the occurrence of pleasant weather days.

DATA SETS

DATA SOURCE

The European Climate Assessment & Data Set project provided the primary dataset, which includes comprehensive weather observations from 18 weather stations across Europe, meticulously recorded from 1960 to 2022.

DATA BIAS

- **Collection Bias**: Changes in instrumentation, measurement methods, or station location over time can introduce bias.
- **Sampling Bias:** Data be collected from 18 specific stations out of 26321 stations all over Europe, potentially skewing the representation of climate patterns.

DATA OPTIMIZATION

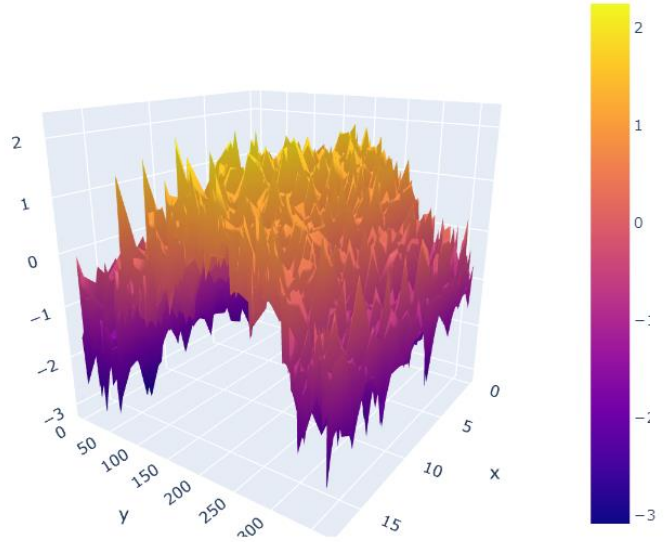
Data Optimization employs various techniques and strategies to enhance data management across multiple aspects, thereby boosting efficiency, reliability, accessibility, and overall utility, aiming to extract maximum value from existing resources.

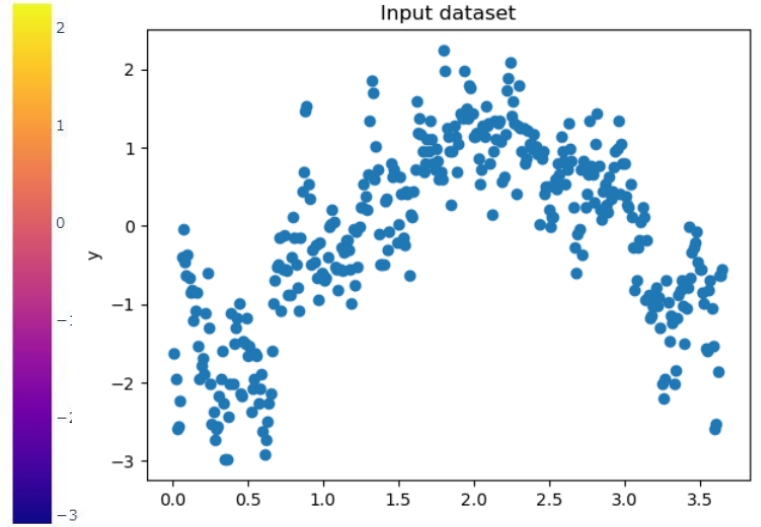
Gradient Descent:

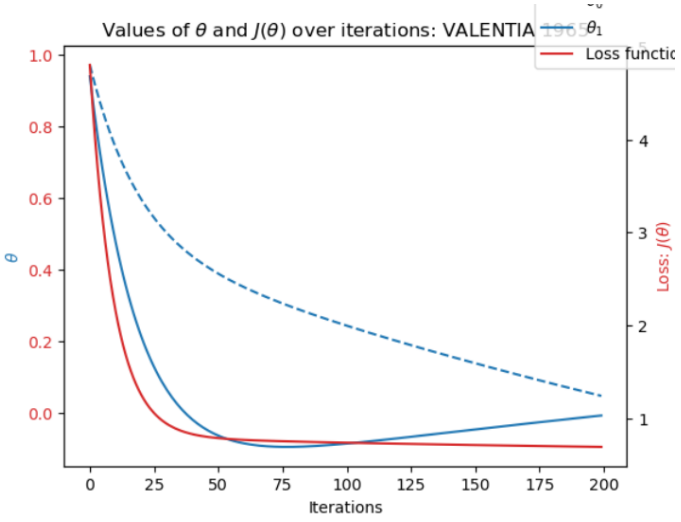
Gradient descent is a method used to find the best-fit line or curve (the optimal parameters) for data points by minimizing the error (or cost).

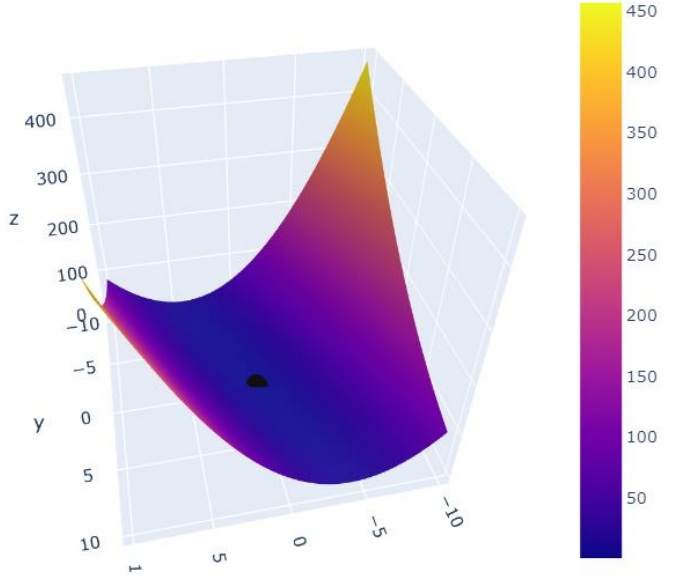
WHICH TECHNIQUES ARE USED?

- 3D Visualizations plot all weather data for all stations for a year.
- A scatterplot examines one year of temperature data over time.
- The Loss function assesses the deviation between predicted and actual weather data.
- Gradient descent is applied to pinpoint the local minimum in the dataset, optimising the model performance towards the optimal option.

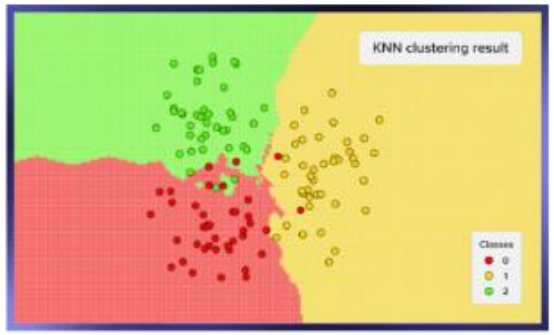








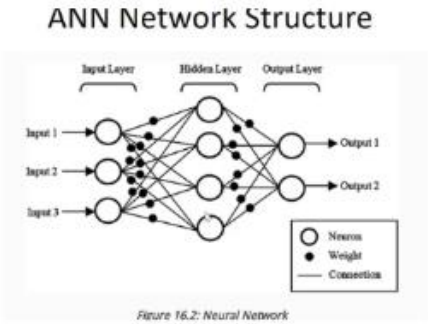
SUPERVISED MACHINE LEARNING



K-NEAREST NEIGHBOURS (KNN)

KNN was used to decide which group each data point was in based on how many of its neighbours belonged to each category.

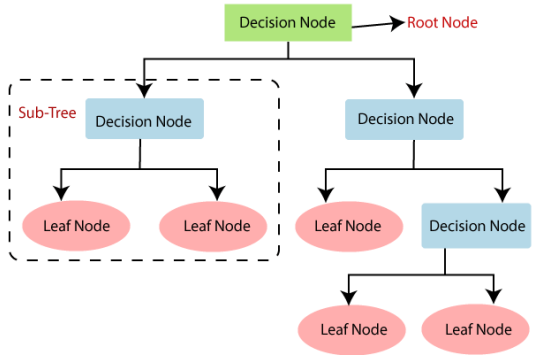
Accuracy of KNN, Training Set: 56.83%
Accuracy of KNN, Test Set: 44.72%
Individual station accuracy: 85.38% ~ 96.23%



ARTIFICIAL NEURAL NETWORK (ANN)

ANN was used to compute an answer based on a linear combination of all inputs and multiplied by weights.

Accuracy of ANN, Training Set: 99.83%
Accuracy of ANN, Test Set: 55.6%
Individual station accuracy: 93% ~ 96.9%



DECISION TREE MODEL

Decision Tree was used to narrow down a solution by asking for more and more specific information.

Accuracy of DC, Training Set: 60.36%
Accuracy of DC, Test Set: 63.56%
Individual station accuracy: 90.9% ~ 95%

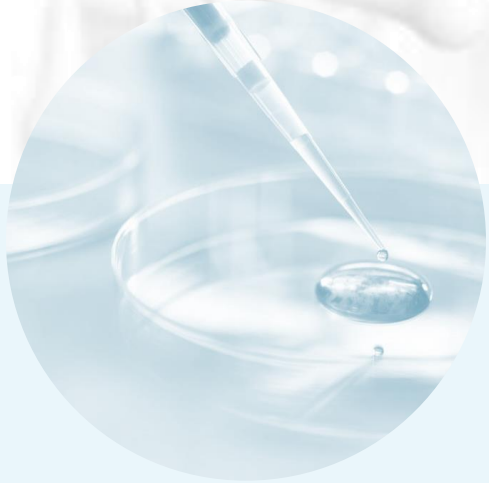
CONCLUSION & NEXT STEPS

CONCLUSION

- Our findings so far show that these models can accurately predict various weather events, affirming our hypothesis about the potential of machine learning to enhance our understanding and prediction of weather patterns influenced by climate change.
- The KNN Model seems to be underfitting (low accuracy on both training and test sets). Meanwhile, ANN is slightly overfitting (very high training accuracy but a significantly lower test accuracy)
- The decision tree model is the best among the three models because it has the highest test accuracy, and its performance is consistent between the training and test sets.

NEXT STEPS

- Further pruning the Decision Tree Model to improve the accuracy.
- Explore ensemble methods that combine the Decision Tree Model with other models
- Discover unsupervised learning machine models to identify previously uncategorized patterns and abnormalities in weather data.
- Combine both supervised and unsupervised methods to create a complete climate model that predicts both specific weather events an wider climate trens.



THANK YOU

QUINN HA
JUNE 2024