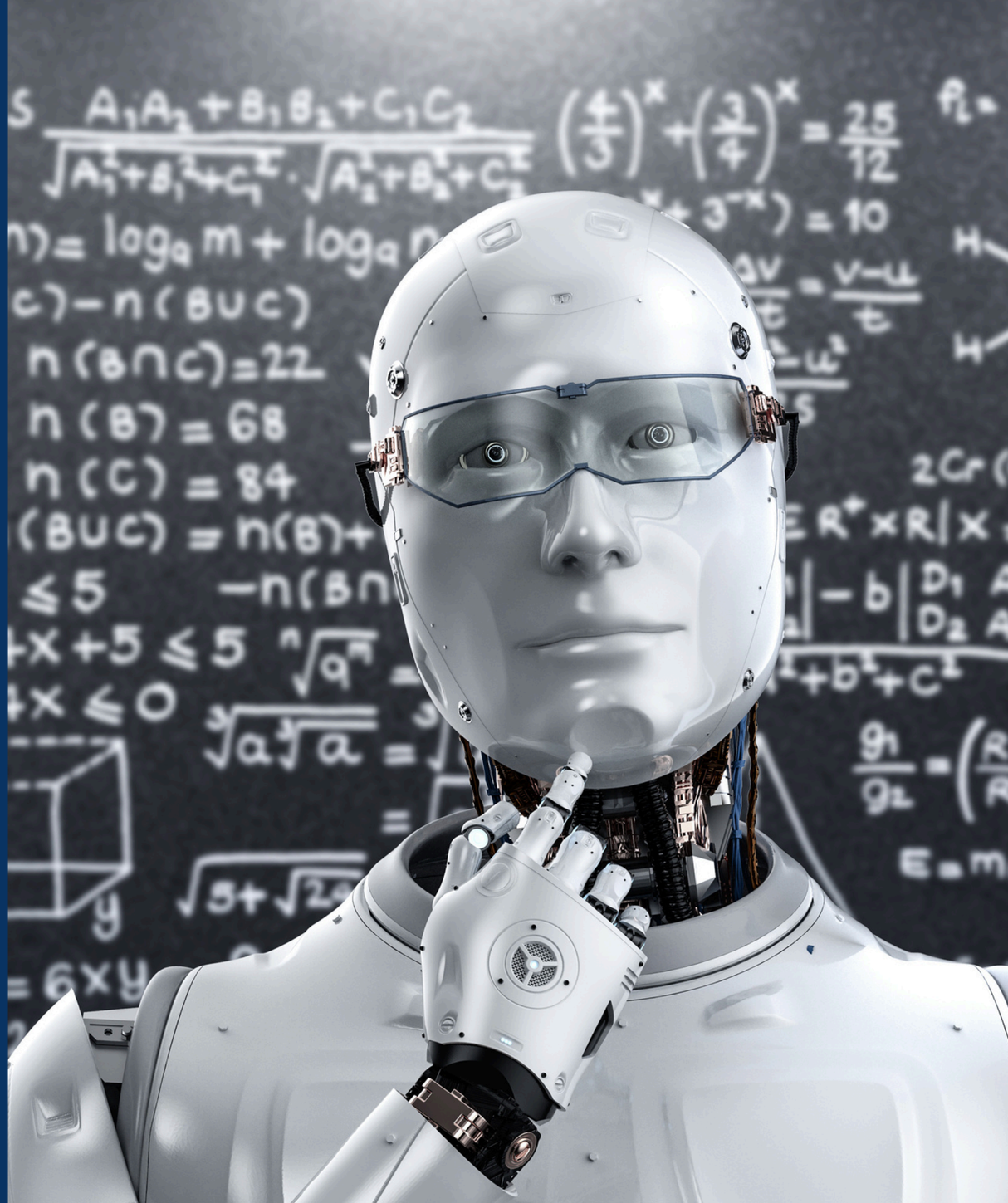
CLIMATEWINS

PREDICTING WEATHER PATTERNS

DAWN CHISM
APRIL 23, 2024

OBJECTIVE

Use machine learning to help predict the consequences of climate change.



HYPOTHESES

1

The greater number of days with a high temp_mean, the greater the number of days with pleasant weather.

2

The greater number of days with precipitation, the greater the number of days with unpleasant weather.

3

The greater number of days with high humidity, the greater number of days with unpleasant weather.

DATA SET

- The data set includes weather observations from 18 different weather stations across Europe for the years of 1960 - 2022 such as:
  - temperature
  - wind speed
  - snow
  - global radiation
  - and more

- The data was collected by the European Climate Assessment & Data Set (ECA&D).

- ECA&D was created in 1998 and has received financial support from the EUMETNET (a grouping of 33 European National Meteorological & Hydrological Services) and the European Commission, therefore it is deemed reliable and accurate.

POTENTIAL BIAS

COLLECTION BIAS

Changes in measurement instruments over time as well as instrument errors, calibration issues, and inconsistencies in measurement methodologies can cause collection bias.

SAMPLING BIAS

If collection sources are unevenly distributed, they may not provide an accurate representation of climate conditions. Potentially leading to underrepresented areas.

INTERPRETATION BIAS

Researchers, policymakers, and the media may interpret climate data in ways that align with their preconceived beliefs, values, or political agendas.

PUBLICATION BIAS

Studies with statistically significant or "exciting" results are more likely to be published than studies with null or inconclusive findings.

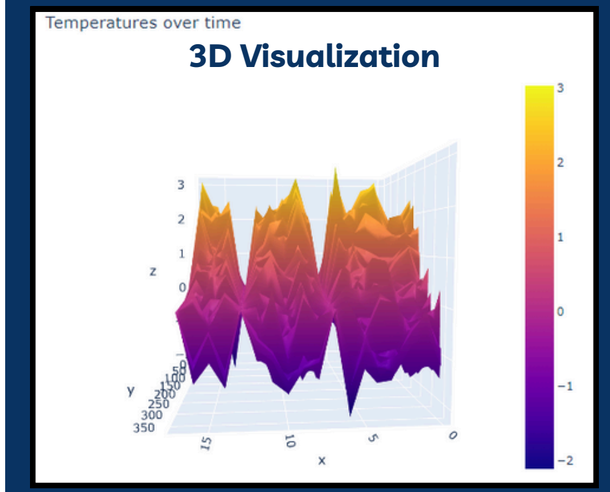
SUPERVISED MACHINE LEARNING

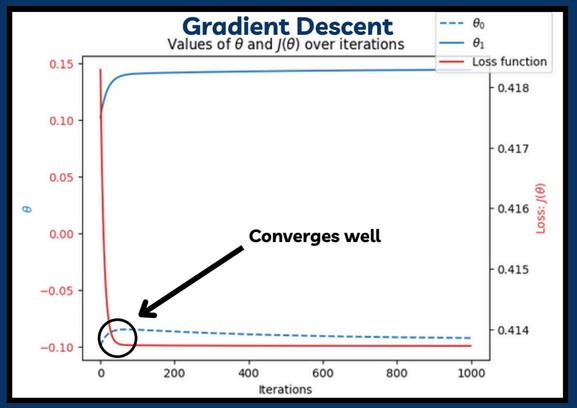
WHAT IS DATA OPTIMIZATION?

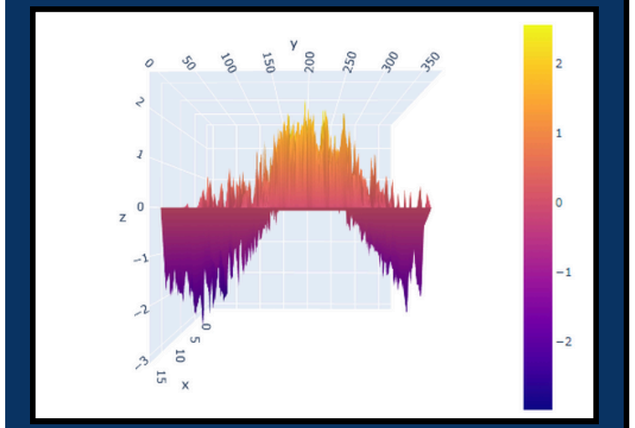
Data optimization improves the efficiency, reliability, accessibility, and overall usefulness of data through various techniques and strategies aimed at enhancing different aspects of data management to derive maximum value from the available data resources.

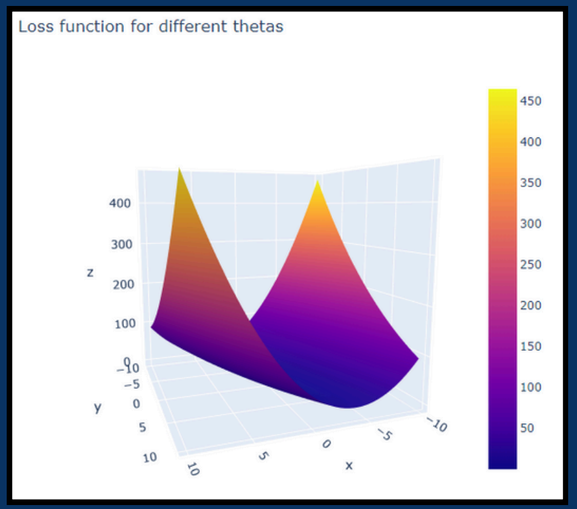
WHAT TECHNIQUES WERE USED?

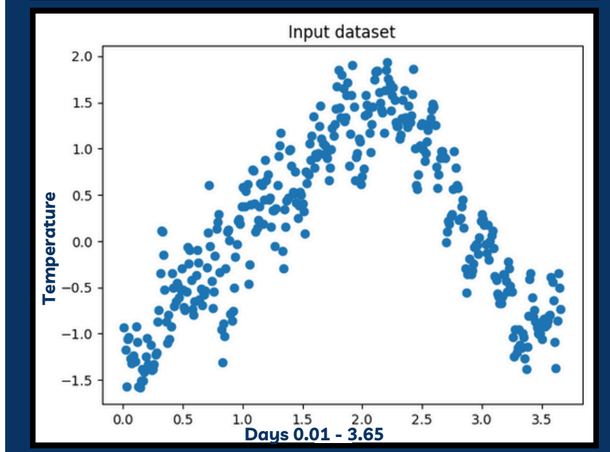
- 3D visualizations were used to plot all weather data for all stations for a year.
- Scatterplots were used to look at one year of temperature data over time.
- Gradient descent was used to find the local minimum.
  - Gradient descent uses weights, loss, backpropagation (iterations), and learning rate/step size to get to the bottom of a curve as quickly as possible, the goal is to reduce the loss to zero.
- Contour plots were used to show the convergence in the X/Y plane.
- The gradient descent converges well, therefore, we can move forward with machine learning to predict weather patterns.

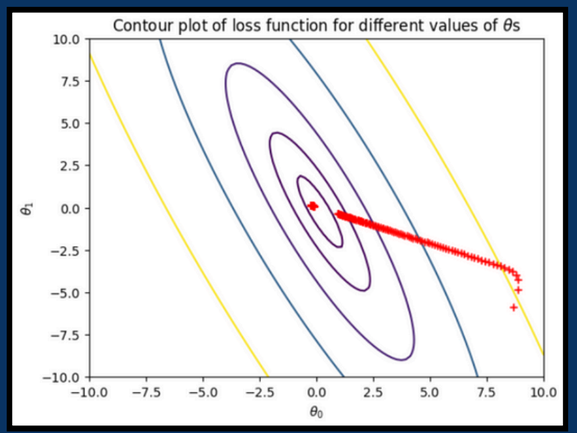








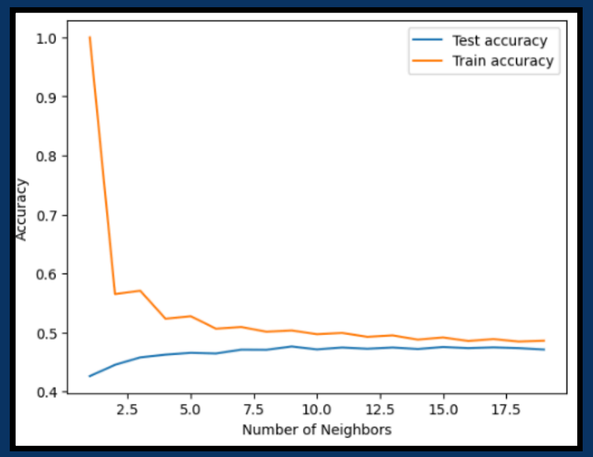




WHAT SUPERVISED LEARNING MODELS WERE USED?

- K-Nearest Neighbor was used to decide which group each data point is in based on how many of its neighbors belong to each category.
- Confusion Matrix's were used by visualizing whether the algorithm is mislabeling (or confusing) one class with another.
- Classification Report's were used to assess how well the model performed across different weather stations in the dataset.
- Decision tree's were used to narrow down a solution by asking for more and more specific information.
- Artificial Neural Network's were used to compute an answer based on a linear combination of all inputs multiplied by weights.

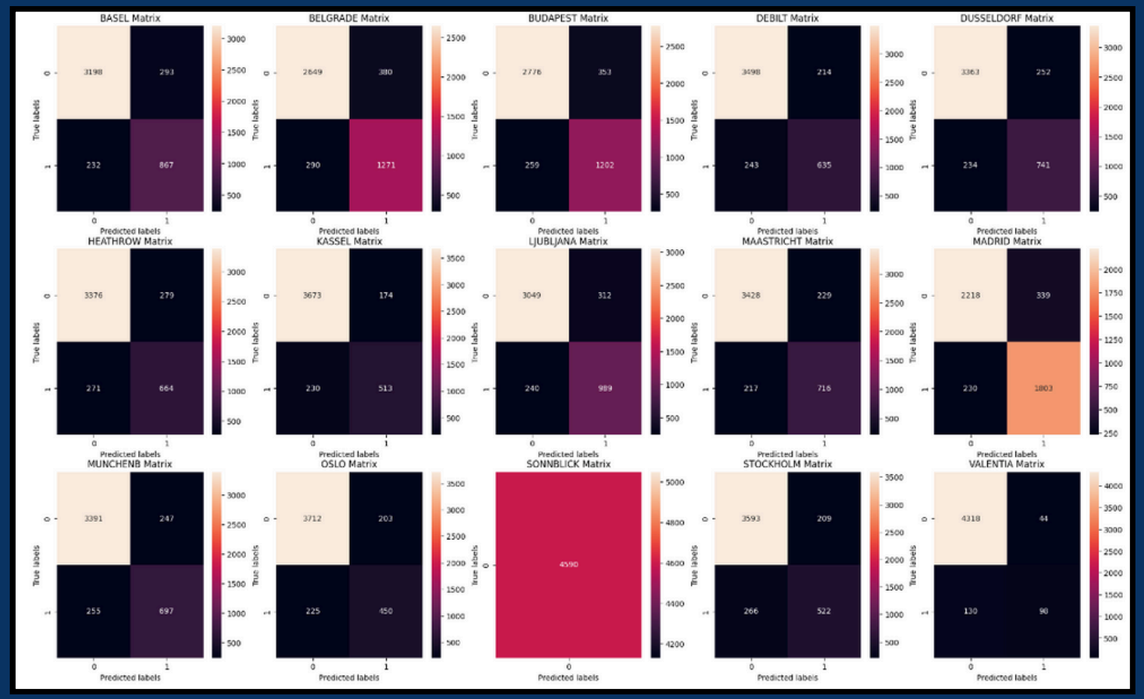
K-Nearest Neighbor



Classification Report

	precision	recall	f1-score	support
0	0.75	0.79	0.77	1099
1	0.77	0.81	0.79	1561
2	0.77	0.82	0.80	1461
3	0.75	0.72	0.74	878
4	0.75	0.76	0.75	975
5	0.70	0.71	0.71	935
6	0.75	0.69	0.72	743
7	0.76	0.80	0.78	1229
8	0.76	0.77	0.76	933
9	0.84	0.89	0.86	2033
10	0.74	0.73	0.74	952
11	0.69	0.67	0.68	675
12	0.00	0.00	0.00	0
13	0.71	0.66	0.69	788
14	0.69	0.43	0.53	228
micro avg	0.76	0.77	0.77	14490
macro avg	0.70	0.68	0.69	14490
weighted avg	0.76	0.77	0.76	14490
samples avg	0.41	0.42	0.40	14490

Confusion Matrix



WHICH MODEL IS BEST?

K-Nearest Neighbor

	precision	recall	f1-score	support
0	0.75	0.79	0.77	1099
1	0.77	0.81	0.79	1561
2	0.77	0.82	0.80	1461
3	0.75	0.72	0.74	878
4	0.75	0.76	0.75	975
5	0.70	0.71	0.71	935
6	0.75	0.69	0.72	743
7	0.76	0.80	0.78	1229
8	0.76	0.77	0.76	933
9	0.84	0.89	0.86	2033
10	0.74	0.73	0.74	952
11	0.69	0.67	0.68	675
12	0.00	0.00	0.00	0
13	0.71	0.66	0.69	788
14	0.69	0.43	0.53	228
micro avg	0.76	0.77	0.77	14490
macro avg	0.70	0.68	0.69	14490
weighted avg	0.76	0.77	0.76	14490
samples avg	0.41	0.42	0.40	14490

Decision Tree

- The train accuracy score is 0.60266884531590.
- The test accuracy score is 0.5474945533769063.

Artificial Neural Network

	precision	recall	f1-score	support
0	0.90	0.87	0.89	1099
1	0.86	0.96	0.91	1561
2	0.95	0.89	0.92	1461
3	0.74	0.97	0.84	878
4	0.86	0.90	0.88	975
5	0.72	0.96	0.82	935
6	0.86	0.88	0.87	743
7	0.93	0.78	0.85	1229
8	0.84	0.89	0.86	933
9	0.91	0.98	0.94	2033
10	0.87	0.96	0.91	952
11	0.87	0.80	0.84	675
12	0.00	0.00	0.00	0
13	0.90	0.81	0.86	788
14	0.73	0.83	0.77	228
micro avg	0.86	0.90	0.88	14490
macro avg	0.80	0.83	0.81	14490
weighted avg	0.87	0.90	0.88	14490
samples avg	0.50	0.52	0.51	14490

RECOMMENDATION AND NEXT STEPS

- Recommendation:
  - I recommend using artificial neural network (ANN) as it had the highest rate of accuracy for predicting the weather data.
- Next Steps:
  - Further evaluate the 15 weather stations over the 62 years worth of data that we have using ANN to determine if machine learning can help predict the consequences of climate change .

THANK YOU!