PREDICTING WEATHER PATTERNS USING MACHINE LEARNING

AN ANALYTICAL APPROACH TO CLIMATE CHANGE

BY : URVI PATEL

OBJECTIVE

Utilize machine learning techniques to forecast the impacts of climate change.

HYPOTHESES

- Increased average temperatures correlate with a higher number of pleasant weather days.
- Higher humidity levels are associated with more unpleasant weather days.
- Greater precipitation levels correspond with a higher number of unpleasant weather days

DATA SET DESCRIPTION

Source: European Climate Assessment & Dataset (ECA&D)

Data Coverage: Weather observations from 1960 to 2022 across 18 weather stations in Europe.

Variables Included: Temperature, wind speed, snow, global radiation, and others.

Data Integrity: Financially backed by EUMETNET and the European Commission, affirming the data's accuracy and reliability.

Relevant Links: EUMETNET, ECA&D Involvement Page

ADDRESSING POTENTIAL BIASES

Collection Bias: Evolving tools and methods over time may affect data consistency.

Sampling Bias: Data may not fully represent climate conditions due to uneven collection distribution.

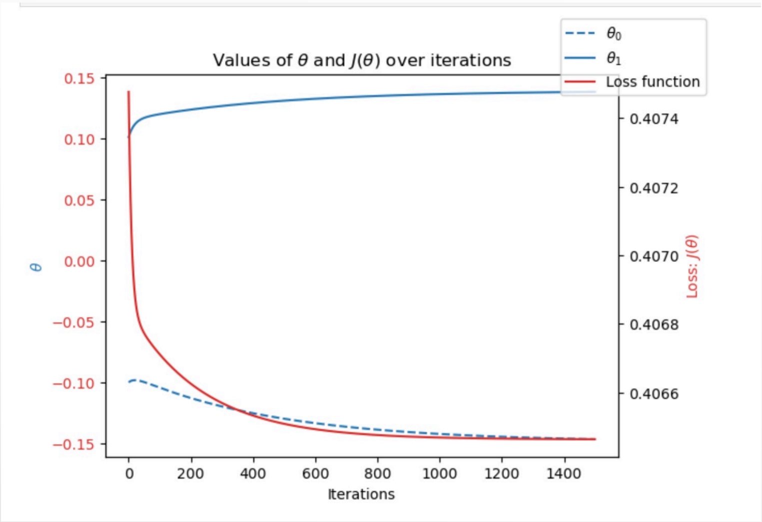
Interpretation Bias: Climate data could be skewed by the subjective views of interpreters.

Publication Bias: Significant results are published more often, potentially overshadowing less dramatic findings.

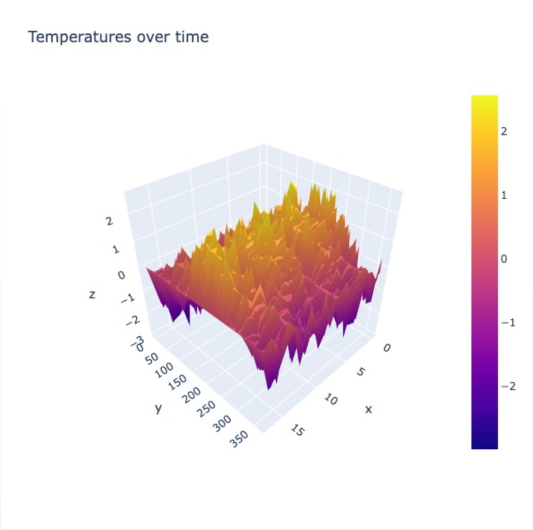
DATA OPTIMIZATION STRATEGY

The optimization of data involved refining data handling and processing techniques to enhance the accessibility, reliability, and overall utility of the dataset. These improvements are crucial for ensuring that our predictive models operate on the most accurate and relevant data.

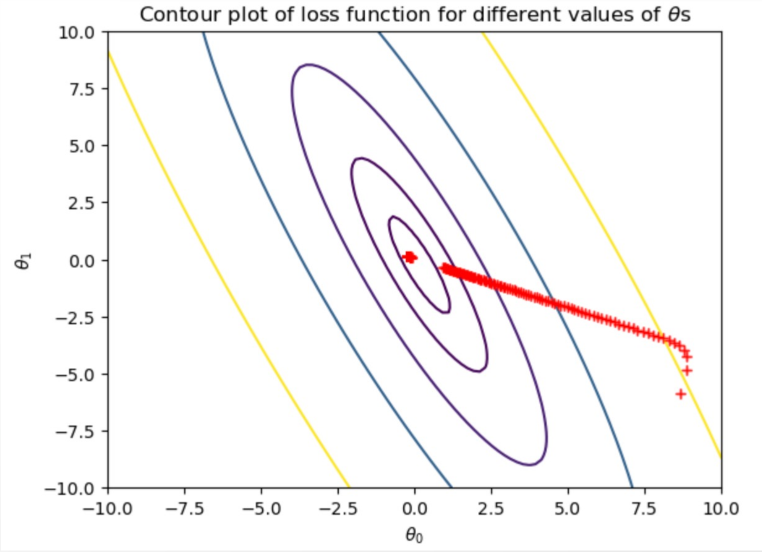
SUPERVISED LEARNING TECHNIQUES APPLIED



Gradient Descent Optimization: Employed as a core technique to fine-tune our predictive models by iteratively reducing the prediction error.



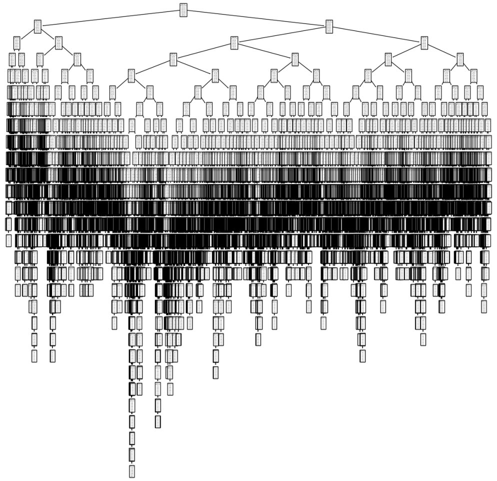
3D Data Visualization: Implemented to explore complex weather variables over time, aiding in the identification of key patterns.



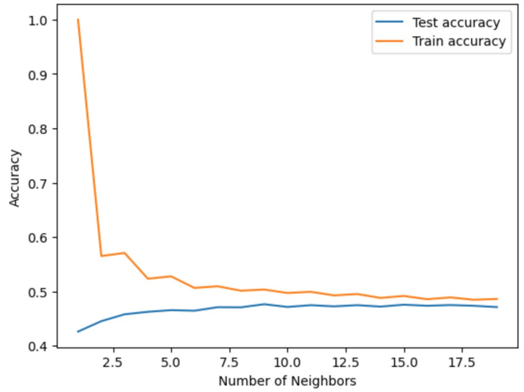
Contour Plot Analysis: Utilized to monitor the progress and convergence of our gradient descent algorithm, confirming the model's learning efficacy.

SUPERVISED LEARNING MODELS USED

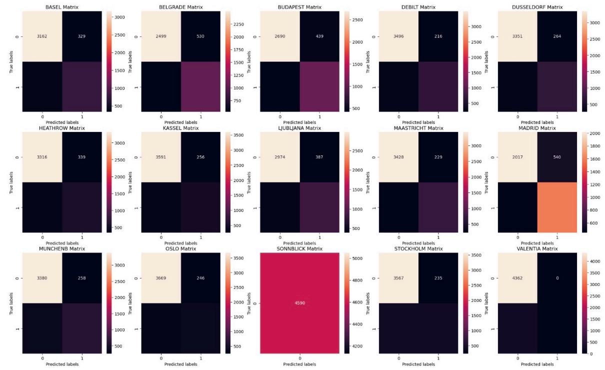
Decision tree



K- nearest neighbor



- **K-Nearest Neighbor (KNN):** Implemented to categorize data points by considering the class of their 'k' nearest neighbors.
- **Confusion Matrix:** Applied to visually inspect if the KNN model was incorrectly classifying any classes.
- **Classification Reports:** Utilized to evaluate the KNN model's performance metrics across various weather stations.
- **Decision Trees:** Deployed to iteratively divide the dataset into increasingly specific subsets based on feature values.
- **Artificial Neural Networks (ANNs):** Developed to predict outcomes through a weighted combination of all input features.



Confusion matrix

	precision	recall	f1-score	support
0	0.86	0.85	0.85	1099
1	0.87	0.90	0.88	1561
2	0.90	0.91	0.90	1461
3	0.82	0.78	0.80	878
4	0.88	0.77	0.82	975
5	0.84	0.77	0.80	935
6	0.87	0.80	0.83	743
7	0.87	0.88	0.87	1229
8	0.88	0.83	0.85	933
9	0.93	0.89	0.91	2033
10	0.86	0.86	0.86	952
11	0.79	0.84	0.81	675
12	0.00	0.00	0.00	0
13	0.84	0.76	0.79	788
14	0.67	0.64	0.65	228
micro avg	0.87	0.84	0.85	14490
macro avg	0.79	0.76	0.78	14490
weighted avg	0.87	0.84	0.85	14490
samples avg	0.50	0.48	0.48	14490

Classification report

BEST MODEL ?

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

Test accuracy score: 0.6381263616557734

Train accuracy score: 0.6039215686274509

Artificial neural network (ANN)

	precision	recall	f1-score	support
0	0.87	0.82	0.85	1099
1	0.89	0.85	0.87	1561
2	0.90	0.86	0.88	1461
3	0.81	0.82	0.81	878
4	0.85	0.78	0.82	975
5	0.84	0.76	0.80	935
6	0.89	0.71	0.79	743
7	0.83	0.89	0.86	1229
8	0.85	0.82	0.84	933
9	0.90	0.96	0.93	2033
10	0.87	0.82	0.84	952
11	0.78	0.78	0.78	675
12	0.00	0.00	0.00	0
13	0.82	0.77	0.79	788
14	0.67	0.61	0.64	228
micro avg	0.86	0.83	0.85	14490
macro avg	0.79	0.75	0.77	14490
weighted avg	0.86	0.83	0.85	14490
samples avg	0.49	0.48	0.48	14490

RECOMMENDATION & NEXT STEPS

- **Model endorsement:** I recommend the use of artificial neural networks (ANN) due to their superior accuracy in predicting weather patterns in our dataset.
- **Continued analysis:** i will further analyze the extensive data collected from 15 weather stations over the past 62 years using ann to explore machine learning's capability in projecting the future consequences of climate change.

THANK YOU

GITHUB REPOSITORY:
https://github.com/urvippatel/predicting-weather-patterns-using-machine-learning