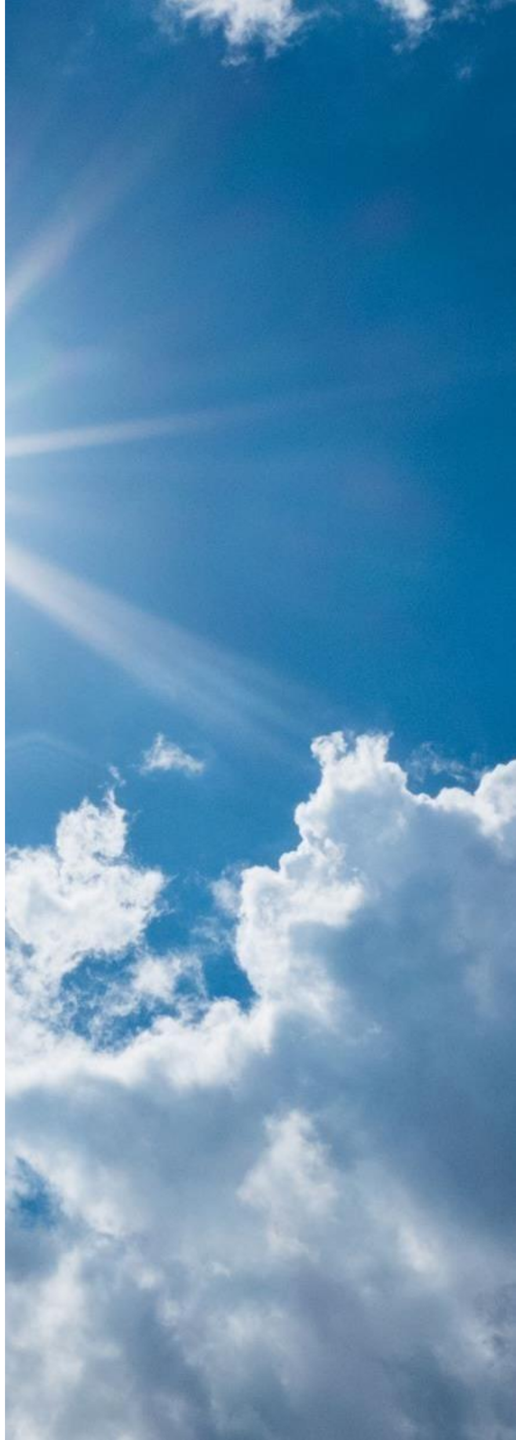
Analyzing Weather Conditions and Climate Change with ClimateWins

Shaival Mehta



Introduction

- ClimateWins leverages machine learning **to forecast the impacts of climate change across Europe**, with the potential to scale its insights globally.

- It has been analyzing **hurricane forecasts** from the U.S. National Oceanic and Atmospheric Administration (NOAA), **typhoon data** from the Japan Meteorological Agency (JMA), **global temperature trends**, and a wide range of other climate-related datasets.

Hypothesis

Machine learning algorithms can accurately predict future weather conditions.

Prediction accuracy may vary depending on geographic location and regional climate factors.

Machine learning can detect early indicators of climate change and its potential adverse effects.

Data Set Info

Collected by European Climate Assessment C Data Set Project

Based on weather observations from 18 different weather stations across Europe

Date range from late 1800s to 2022

Data points include temperature, wind speed, snow, global radiation, etc

Data Bias and Accuracy

- **Data Collection Challenges:** Historical datasets may contain human errors due to less accurate collection methods.
- **Data Diversification Bias:** Machine learning models may disproportionately favor larger regions, leading to underrepresentation of smaller or less-populated areas.
- **Temporal Bias:** Analyzing over 200 years of data can introduce discrepancies, as older records may not accurately reflect present-day climate patterns.
- **Data Uniformity:** Consistency in data format and structure is essential for machine learning algorithms to perform accurate and meaningful comparisons.

Data Optimization

- This dataset was optimized using the **gradient descent algorithm**.
- Gradient descent is a straightforward and effective method for **identifying a local minimum in both linear and nonlinear scenarios**.
- In our analysis, it was applied to **minimize prediction error** by iteratively adjusting parameters such as the learning rate (alpha) and the number of iterations.
- By fine-tuning $theta_0$ and $theta_1$, along with optimizing alpha and iteration count, we achieved a result approaching **zero error**.

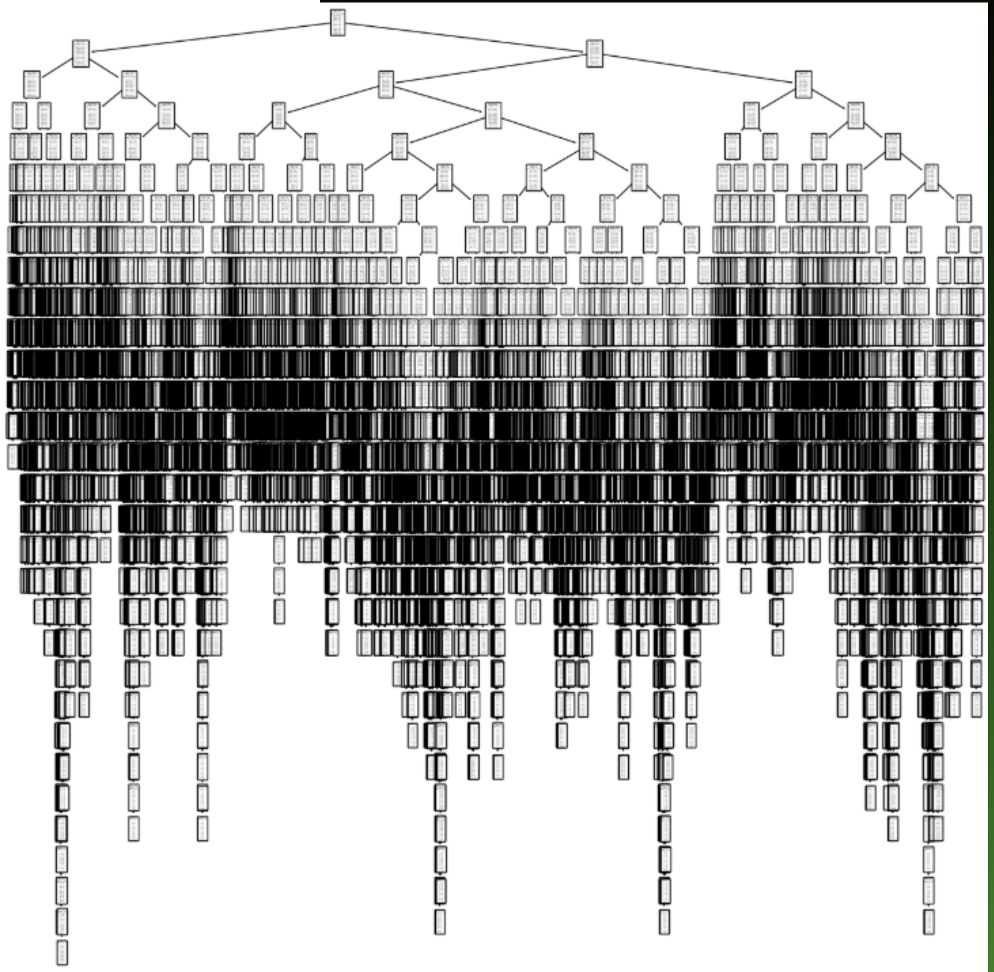
Method 1: K-Nearest Neighbor

- The KNN algorithm predicts outcomes by calculating the distance between a new data point and all data points in the training set. It then classifies the new point based on the majority class among its nearest neighbors.
- **Average accuracy:** 88%
- To enhance model accuracy and reduce disparities across weather stations, the training dataset should be diversified to include a broader range of weather conditions and geographic regions.

Weather Station	Accurate Predictions		False Positive	False Negative	Accuracy Rate
Basel	3917	961	421	439	85%
Belgrade	3252	1544	524	418	84%
Budapest	3424	1462	476	376	85%
Debilt	4320	723	317	378	88%
Desseldorf	4164	810	343	421	87%
Heathrow	4138	744	432	424	85%
Kassel	4563	614	252	309	90%
Ljubljana	3740	1180	455	363	86%
Maastricht	4253	824	309	352	88%
Madrid	2750	2261	418	309	87%
Munchenb	4237	792	309	400	88%
Oslo	4637	512	242	347	90%
Sonnblick	5738	0	0	0	100%
Stockholm	4483	607	283	365	89%
Valentia	5404	74	50	202	96%
				Average	88%

Method 2: Decision Tree

- The decision tree algorithm predicts outcomes by asking a series of feature-based questions, guiding new data points through a network of nodes and branches to reach a classification or prediction.
- **Accuracy:** 46%
- The current model is overly complex, with too many branches, making interpretation challenging and reducing practical usability.
- Additionally, the large size of the tree results in longer training times and potential overfitting.

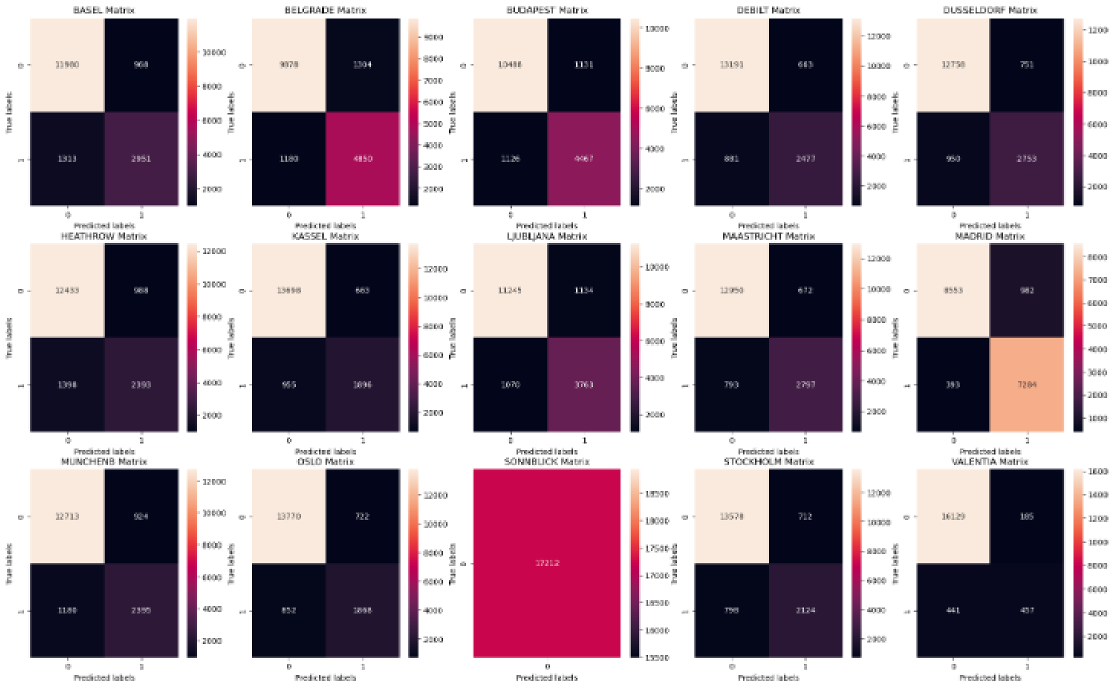


Method 3: Artificial Neural Network

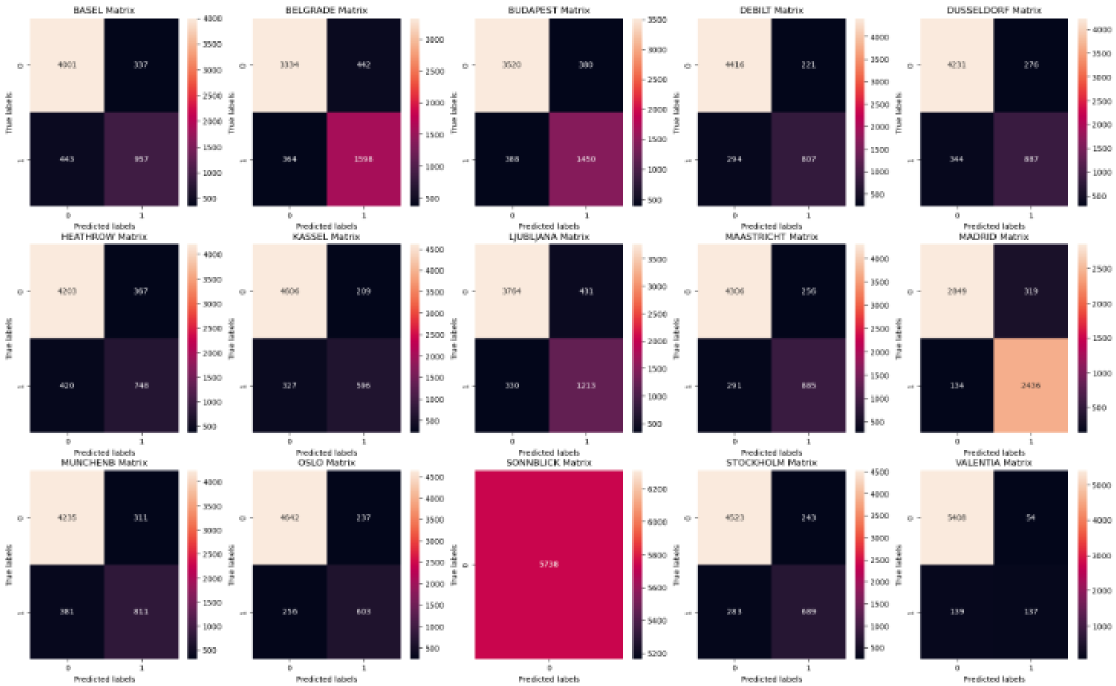
- Artificial Neural Network (ANN) is a computer model that mimics the way the human brain processes information.
- It consists of layers of interconnected "neurons" that learn patterns from data through training.
- By adjusting connections between neurons, it improves its ability to make predictions or classifications.
- Adjustments to layer size and max iteration
- **Accuracy:** 48% to 58%

ANN First Scenario

Training Data



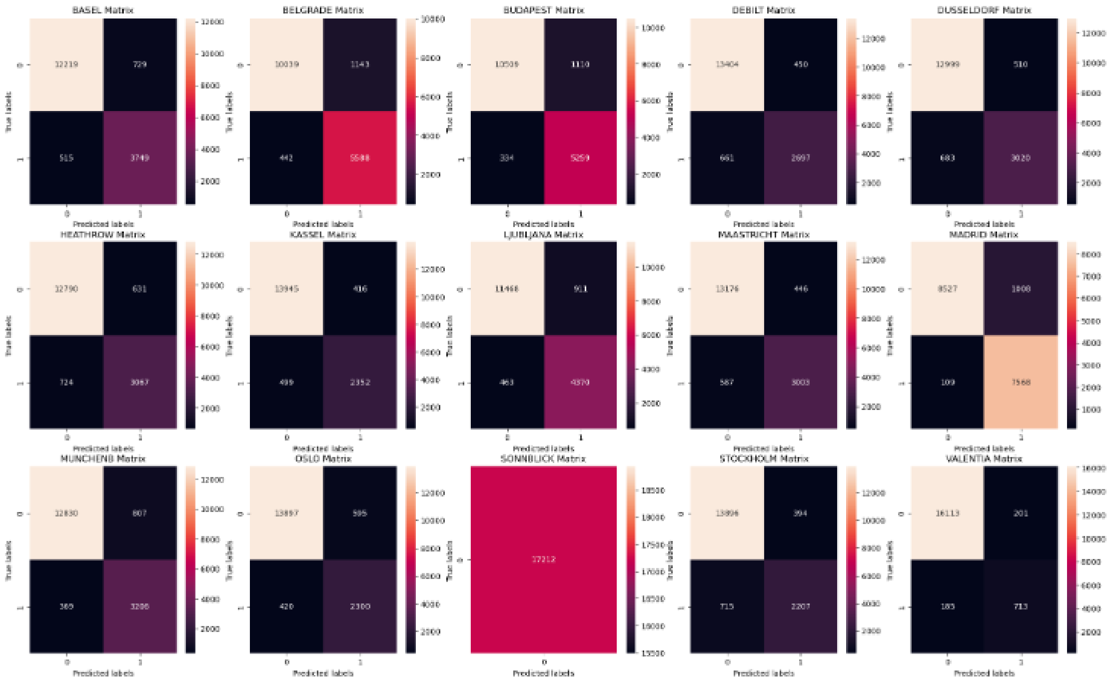
Testing Data



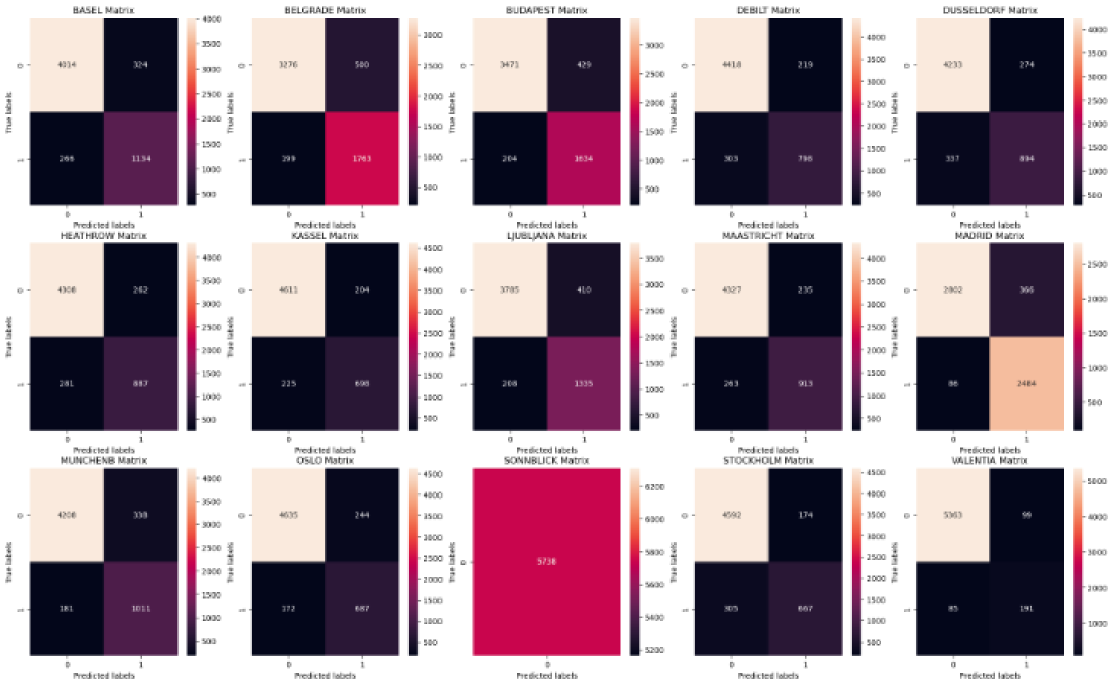
Testing Accuracy: 48%

ANN Second Scenario

Training Data



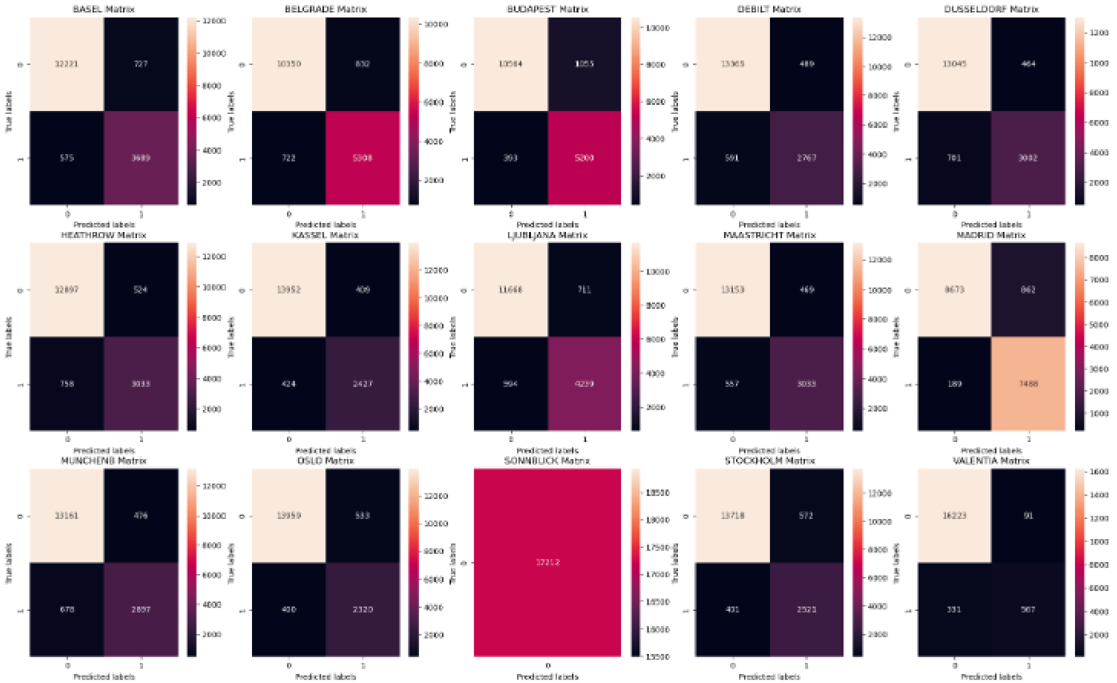
Testing Data



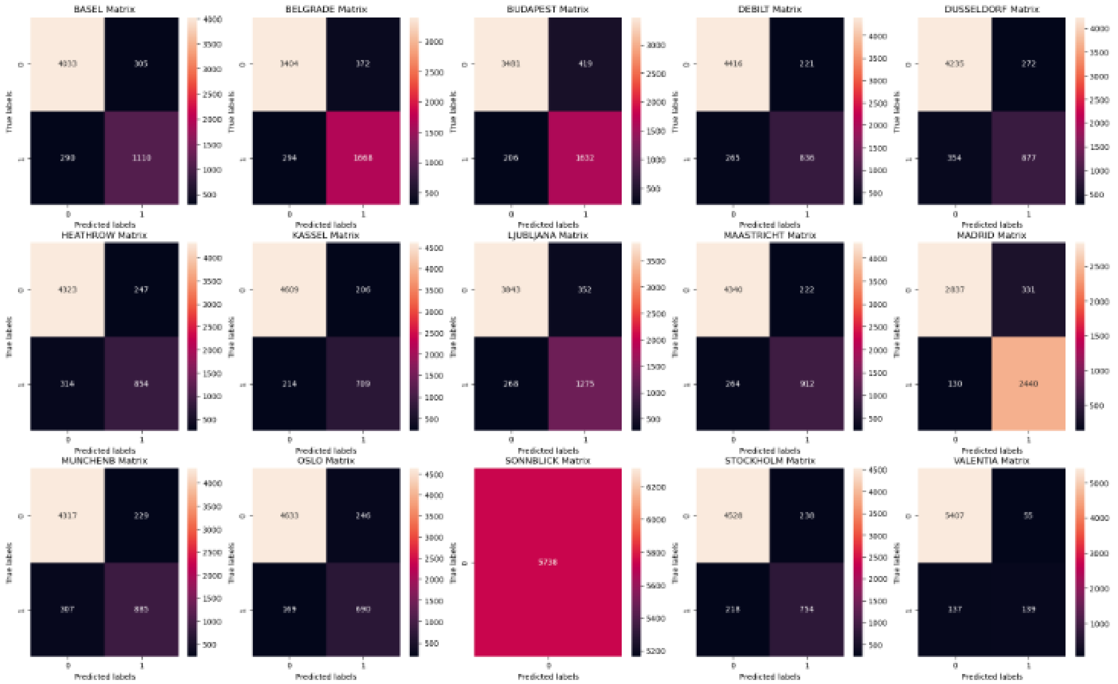
Testing Accuracy: 58%

ANN Third Scenario

Training Data



Testing Data



Testing Accuracy: 58%

Algorithm Recap

- The **decision tree** model is overly complex and difficult to interpret.
- While the **KNN model** achieves the highest average accuracy of 88% across all weather stations—compared to 55% with ANN—it is less suitable for handling complex, non-linear data.
- Despite its lower accuracy in this case, the **ANN model** is more powerful, scalable for larger datasets, and less sensitive to noise.
- Given the complexity and non-linearity of weather data, the **ANN model is recommended over KNN** for weather prediction tasks.

Summary

Machine learning algorithms can effectively predict future weather conditions, with some models achieving accuracy rates as high as 88%.

Prediction accuracy varies depending on geographic location and the specific climate conditions of the region.

Machine learning can detect indicators of climate change and its adverse effects by analyzing patterns of extreme weather conditions over time.

Next Steps

- Incorporate unsupervised machine learning algorithms
- Expand data inputs to include additional variables for analysis
- Continuously test and refine adjustments to improve model accuracy
- Monitor more weather stations to enhance machine learning performance

Thank you!

Check out my video

Check out my Github for the Python scripts and dataset used in this analysis