

WEATHER CONDITIONS AND CLIMATE CHANGE WITH CLIMATEWINS

BY MANISHA TOMAR

INTRODUCTION

ClimateWins is exploring the use of machine learning to predict the impacts of climate change across Europe and potentially worldwide. This presentation draws on data from hurricane forecasts by the U.S. National Oceanic and Atmospheric Administration (NOAA), typhoon data from the Japan Meteorological Agency (JMA), global temperature records, and a broad range of additional data sources.

HYPOTHESIS

- Can machine learning predict the weather?
- ClimateWins has heard of ethical concerns surrounding machine learning and AI. Are there any concerns specific to this project?
- Does Supervised or Unsupervised learning provide a better accuracy in terms of good weather.
- Can machine learning be used to predict whether weather conditions will be favorable on a certain day? (If so, it could also be possible to predict danger.)

DATA SETS & DATA BIAS

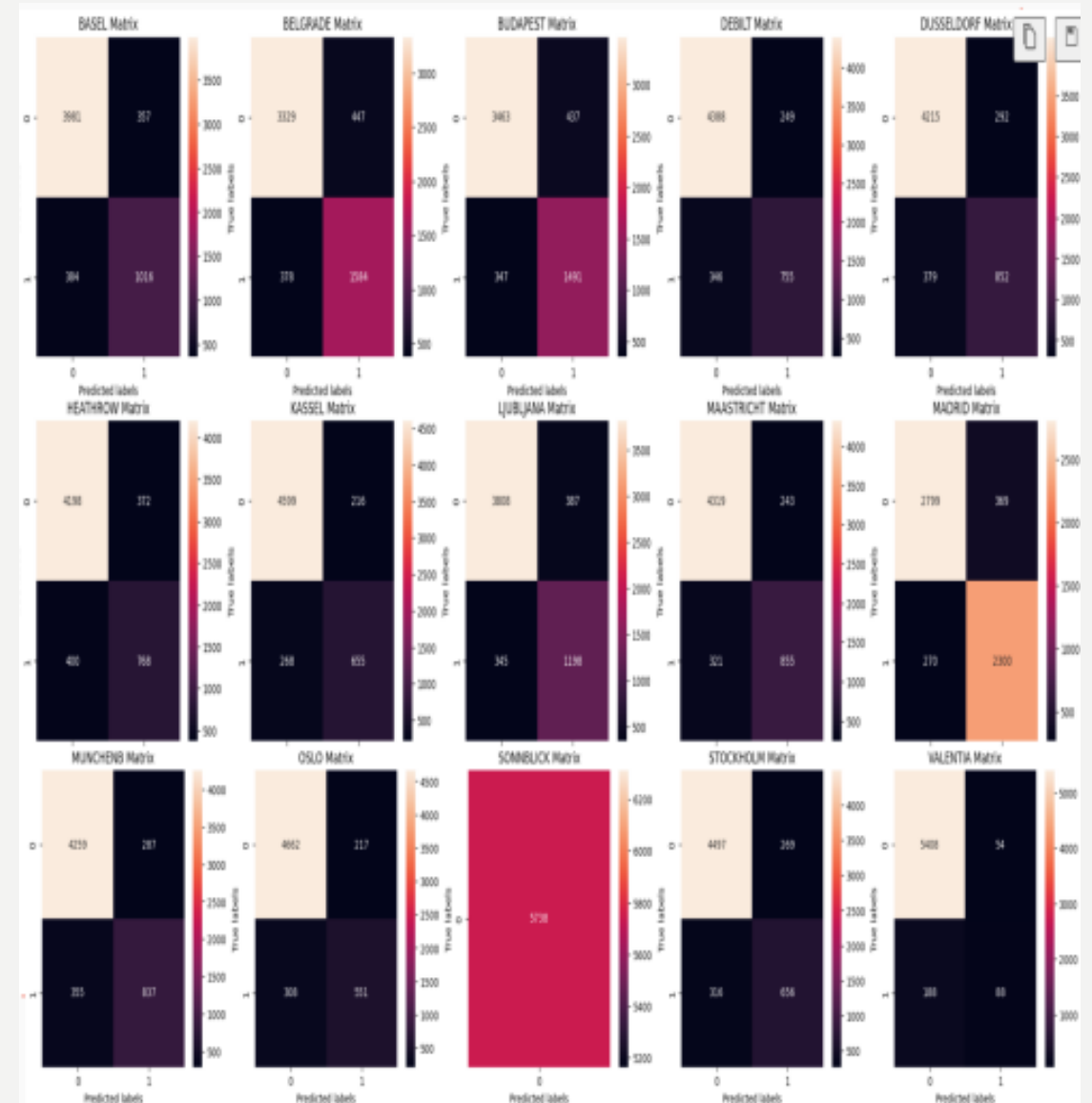
- The Dataset we use comes from European Climate Assessment & Data Set project
- The data set based on weather observations from 18 different weather stations across Europe, which contain data ranging from the late 1800s to 2022.
- Recordings exist for almost every day with values such as temperature, wind speed, snow, global radiation, and more.
- Collection Bias: The data was collected around 18 different weather stations across Europe, while, according to the European Climate Assessment and Dataset there are a total of 23755 weather stations across Europe.
- Location Bias: Considering that the data set was collected by European Assessment & Dataset and contains information about Europe and the Mediterranean, this data may not be able to predict correctly the weather from places such as Brazil or Canada, since it's different from the sample origin.
- Temporal Bias: Since the range is quite large, from late 1800s to 2022, it is possible that some data may not represent the reality anymore and could mislead the machine learning algorithm.

DATA ACCURACY & DATA OPTIMIZATION

- Data accuracy varies with the machine learning algorithm used. The following slides show significant differences in accuracy between each model on both training and test sets. The decision tree's accuracy couldn't be evaluated without pruning. The Artificial Neural Network (ANN) achieved 40% accuracy on the test set, ranging from 40-70% on the training set. The KNN model performed best, with 88% accuracy on the test set and 92% on the training set.
- This dataset was optimized using gradient descent, a simple method for finding a local minimum that works in both linear and nonlinear cases. Here, gradient descent was applied to minimize error by adjusting parameters such as the number of iterations and step size (alpha), which varied as needed. As shown in the next slide, we achieved an error close to zero after tuning θ_0 , θ_1 , iterations, and alpha.

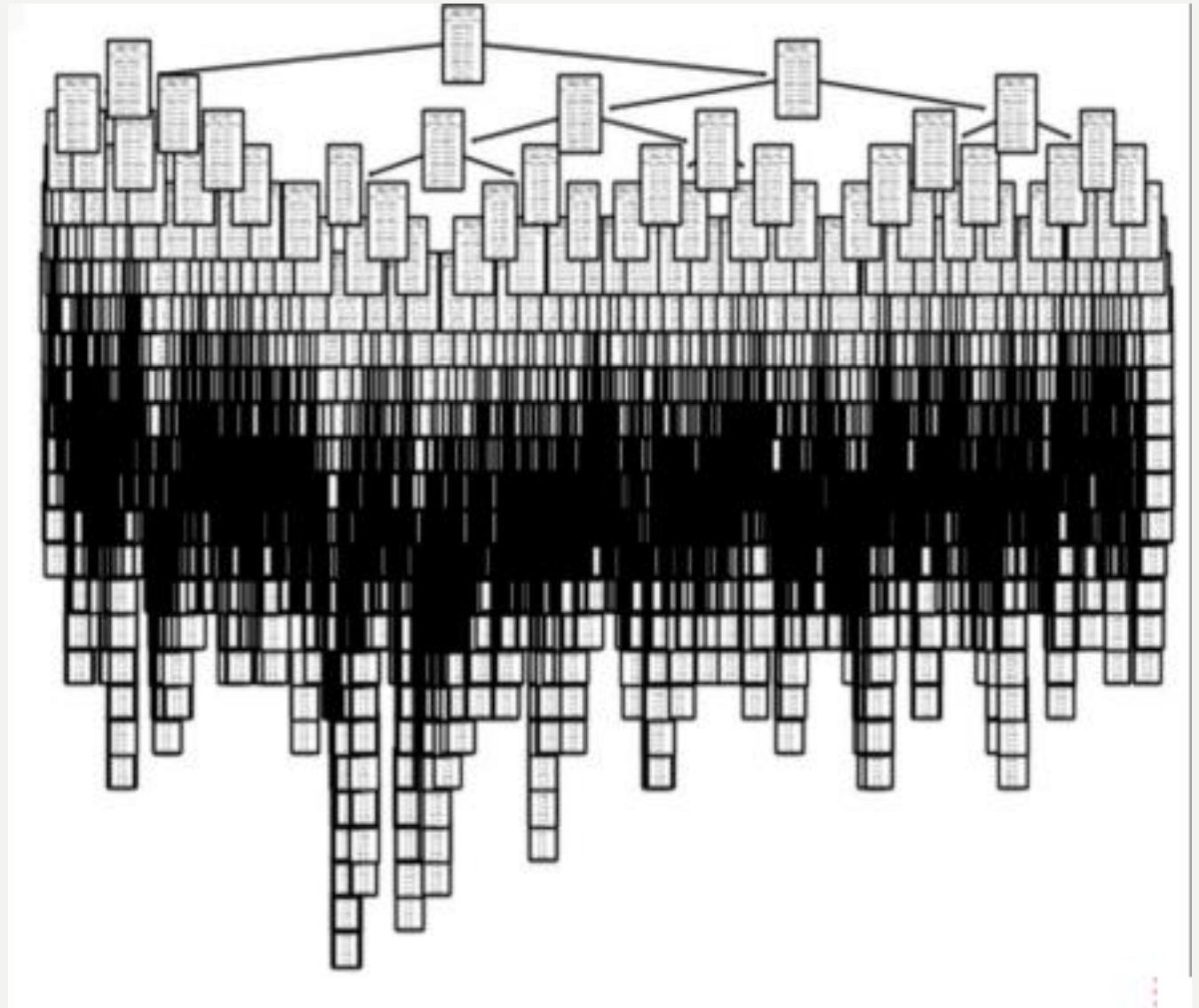
SUPERVISED LEARNING VISUALIZATION

- Using Complex Matrix, the weather stations that shows promising good weather is Basel, Belgrade, Budapest, Ljubljana and Madrid.
- The reason for that is accuracy predictions are higher than the false positives. Sonblick Overall accuracy for Accuracy Rate is 90%.
- For Sonblick the accurate rate is 100%, for unpleasant weather, which is inflating the overall average of 90%. This is implying possible overfitting or learning the model too well. This may be due to having one category of data (e.g. unpleasant weather) or noise in the model.
- Valentia gave us the most believable accuracy rate of 96% which is above the 90% mean. Belgrad and Budapest had the lowest accuracy, 86% below the 90% mean.



SUPERVISED LEARNING VISUALIZATION

- Created a Decision Tree
- The decision tree shows how to recognize patterns and structures within data sets.
- The decision tree needs to be pruned as it is too deep and complex.



SUPERVISED LEARNING TRAINING DATA

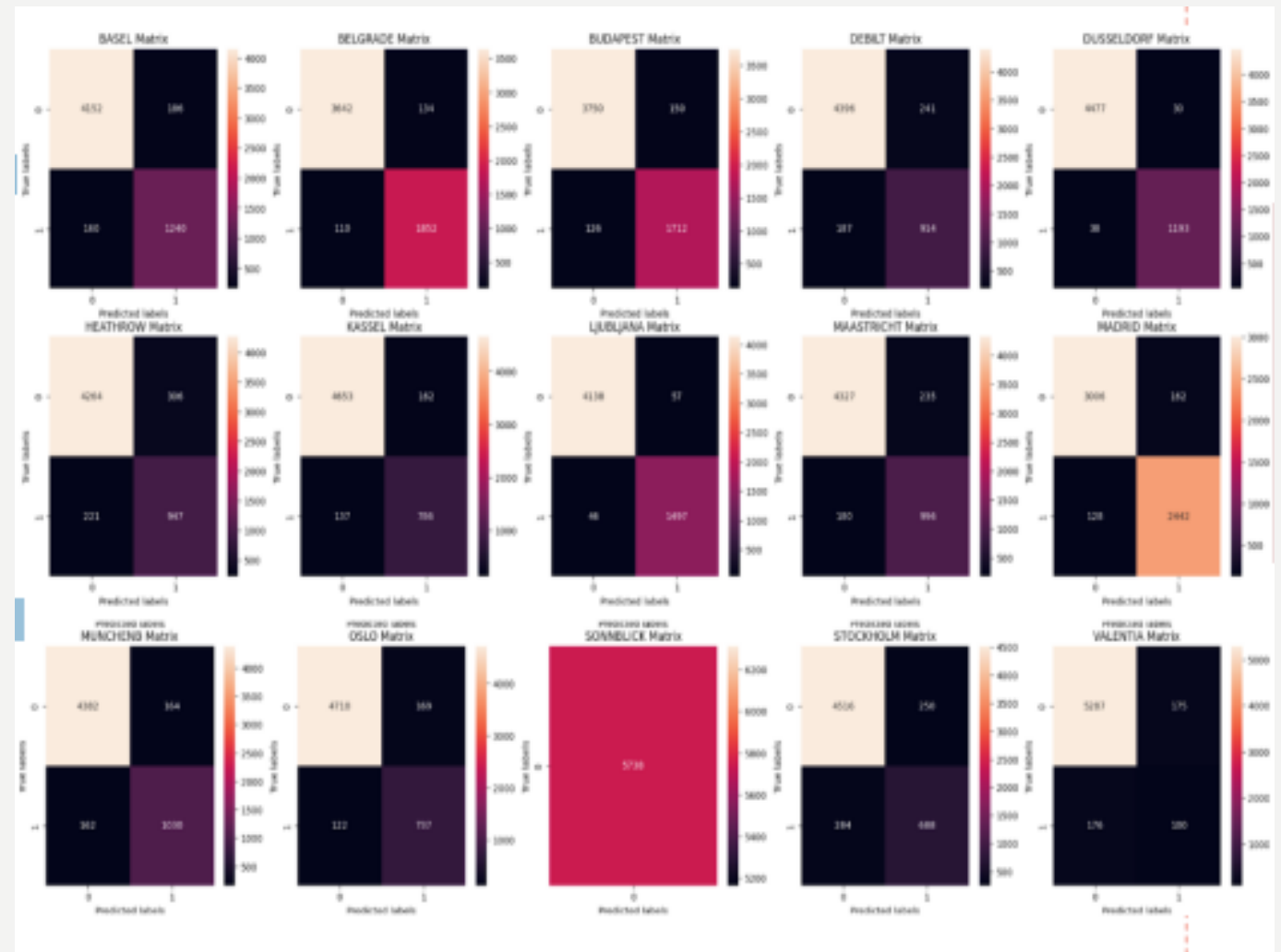
- Using Complex Matrix, we performed training and testing data.
- We used the ANN algorithm that learns from historical data however it didn't perform better than KNN for accuracy.
- Training data includes input features and continuous labels (Predicting future weather)
- Training data accuracy: 60%



SUPERVISED LEARNING TESTING DATA

- Using Complex Matrix, we performed training and testing data.
- Test data accuracy: 63%
- Testing data in unsupervised learning helps validate the patterns, clusters, or structures discovered by the model. Since there are no labels, evaluation focuses on internal validation metrics and, if possible, comparison with external data or domain knowledge.

Testing Data



HYPOTHESIS

- Can machine learning predict the weather? The KNN model shows the highest accuracy, making it the best choice for ClimateWins to predict picnic-friendly weather. However, models trained for Sonnblick may be overfitting; although they achieve 100% accuracy by consistently predicting unpleasant weather, this could be misleading due to limited data on pleasant conditions. This lack of variation suggests the model may not generalize well to different scenarios, potentially inflating its accuracy.

QUESTIONS

- Will Sonnblick be the only country that has 100% accuracy when you perform the complex matrix every time?
- Would the countries with good weather ever shift to having bad weather?
- Would the countries with bad weather shift to having good weather?

CONCLUSION AND NEXT STEPS

- Combine both supervised and unsupervised methods to create a complete climate model that predicts both specific weather events and wider climate trend
- Prune the Decision tree to get a better accuracy
- Go further to discover to identity uncategorized from Unsurprised data.
- Our findings so far show that these models can accurately predict various weather events.
- KNN 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.



THANK YOU

Manisha Tomar