

Introduction

ClimateWins aims to utilize temperature data in Europe to predict daily weather conditions conducive to outdoor activities. The motivation behind this research question lies in the desire to leverage machine learning techniques to accurately forecast "pleasant weather," which can have implications for various sectors such as tourism, recreation, and urban planning.

Problem

Can machine learning models accurately predict "pleasant weather" based on temperature data from European weather stations, and if so, which model performs best for this task?

Solution

I embarked on a series of analyses using different machine learning models to predict "pleasant weather" based on temperature data. Initially, had to exclude 3 of the stations since they were not part of the "pleasant weather" data, those excluded were Gdansk, Roma and Tour. I then implemented the K-nearest neighbors (KNN) algorithm on the weather data, with "pleasant weather" labels provided as the target variable. By modifying the KNN model to accommodate a multilabel answer set and utilizing a multilabel confusion matrix, we assessed the accuracy of the predictions.

Findings

- The KNN algorithm showed varying degrees of accuracy in predicting "pleasant weather" across different weather stations. The confusion Matrix and Classification Report can be found in the appendix and serves as crucial tools for evaluating the efficacy of the KNN algorithm in predicting "pleasant weather" and corroborating the findings.
- Reduction of the k-range parameter from 60 to 4 or 5 improved computational efficiency without significantly compromising prediction accuracy.
- Certain weather stations exhibited higher accuracy in predicting "pleasant weather" compared to others, suggesting the influence of geographic or climatic factors.
- Sonnblick data is all labeled as 0's - unpleasant days. Therefore, the confusion matrix produces an error.
- Overfitting was observed in some instances, indicating the need for regularization techniques or model evaluation strategies to mitigate this issue.
- Features such as temperature trends, geographic location, and proximity to bodies of water may contribute to the overall accuracy or inaccuracy of the predictions.
- Considering the strengths and weaknesses identified in the model's performance, it's recommended to explore techniques like data augmentation or hyperparameter tuning to potentially improve the model's ability to accurately classify 'pleasant weather' across diverse weather conditions.

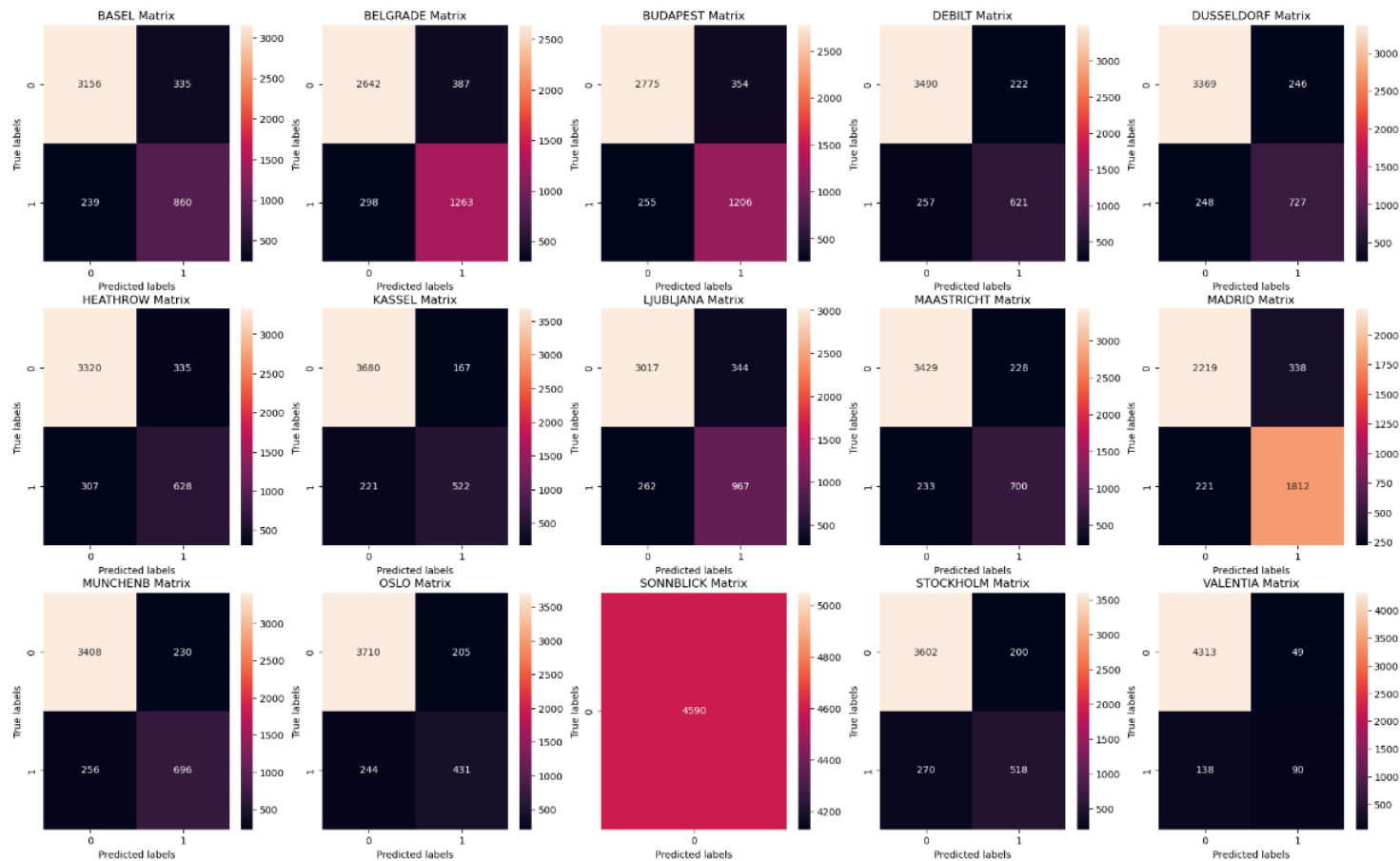
Appendix

Classification Report

	<i>precision</i>	<i>Recall</i>	<i>f1-score</i>	<i>support</i>
0	0.68	0.69	0.68	1099
1	0.63	0.80	0.71	1561
2	0.65	0.85	0.74	1461
3	0.74	0.73	0.73	878
4	0.72	0.72	0.72	975
5	0.63	0.61	0.62	935
6	0.64	0.55	0.59	743
7	0.62	0.81	0.70	1229
8	0.73	0.74	0.74	933
9	0.74	0.74	0.79	2033
10	0.70	0.61	0.65	952
11	0.64	0.48	0.55	675
12	0.00	0.00	0.00	0
13	0.63	0.46	0.54	788
14	0.00	0.00	0.00	228
<i>micro avg</i>	0.68	0.71	0.69	14490
<i>macro avg</i>	0.58	0.59	0.58	14490
<i>weighted avg</i>	0.67	0.71	0.68	14490
<i>samples avg</i>	0.32	0.33	0.31	14490

The classification report unveils a precision value of 0.68 for class 0, indicating that 68% of the model's predictions for this class were accurate. This precision is supported by a recall value of 0.69, suggesting that the model effectively identified 69% of actual instances of class 0 across different data points. However, the F1-score of 0.68 for class 0 indicates that there might be some room for improvement in balancing precision and recall for this class.

Confusion Matrix



This Confusion Matrix visually illustrates the model's predictions compared to the ground truth. By depicting the true positives, false positives, true negatives, and false negatives, it offers a comprehensive overview of the model's performance and highlights any patterns or discrepancies in the predictions. Interpretation of the results can be found in the following page.

Class 0 -> Unpleasant Days

Class 1 -> Pleasant Days

Basel Matrix

- Class 0 (True Negatives): There are 3156 instances correctly classified as class 0 (True Negatives). This suggests that the model is performing well in identifying instances belonging to class 0.
- Class 1 (True Positives): There are 860 instances correctly classified as class 1 (True Positives). This indicates that the model is also performing reasonably well in identifying instances belonging to class 1.
- False Positives: There are 335 instances incorrectly classified as class 1 (False Positives). This suggests that there is some misclassification of class 0 instances as class 1.
- False Negatives: There are 239 instances incorrectly classified as class 0 (False Negatives). This indicates that there is some misclassification of class 1 instances as class 0.

Insights

1. The model seems to have a relatively higher accuracy in classifying unpleasant days instances compared to pleasant days instances.
2. There is a notable number of false positives and false negatives, indicating that there is room for improvement in the model's ability to distinguish between pleasant and unpleasant days.

Belgrade Matrix

- Class 0 (True Negatives): There are 2642 instances correctly classified as class 0 (True Negatives). This suggests that the model is performing well in identifying instances belonging to class 0.
- Class 1 (True Positives): There are 1263 instances correctly classified as class 1 (True Positives). This indicates that the model is also performing reasonably well in identifying instances belonging to class 1.
- False Positives: There are 387 instances incorrectly classified as class 1 (False Positives). This suggests that there is some misclassification of class 0 instances as class 1.
- False Negatives: There are 298 instances incorrectly classified as class 0 (False Negatives). This indicates that there is some misclassification of class 1 instances as class 0.

Insights

1. The model seems to have a relatively higher accuracy in classifying unpleasant days instances compared to pleasant days instances.
2. There is a significant number of false positives and false negatives, indicating that there is room for improvement in the model's ability to distinguish between pleasant and unpleasant days.

Budapest Matrix

- Class 0 (True Negatives): There are 2775 instances correctly classified as class 0 (True Negatives). This suggests that the model is performing well in identifying instances belonging to class 0.
- Class 1 (True Positives): There are 1206 instances correctly classified as class 1 (True Positives). This indicates that the model is also performing reasonably well in identifying instances belonging to class 1.
- False Positives: There are 354 instances incorrectly classified as class 1 (False Positives). This suggests that there is some misclassification of class 0 instances as class 1.
- False Negatives: There are 255 instances incorrectly classified as class 0 (False Negatives). This indicates that there is some misclassification of class 1 instances as class 0.

Insights:

1. The model seems to have a relatively higher accuracy in classifying unpleasant days instances compared to pleasant days instances.
2. There is a notable number of false positives and false negatives, indicating that there is room for improvement in the model's ability to distinguish between pleasant and unpleasant days.

Debilt Matrix

- Class 0 (True Negatives): There are 3490 instances correctly classified as class 0 (True Negatives). This suggests that the model is performing well in identifying instances belonging to class 0.
- Class 1 (True Positives): There are 621 instances correctly classified as class 1 (True Positives). This indicates that the model is also performing reasonably well in identifying instances belonging to class 1.
- False Positives: There are 222 instances incorrectly classified as class 1 (False Positives). This suggests that there is some misclassification of class 0 instances as class 1.
- False Negatives: There are 257 instances incorrectly classified as class 0 (False Negatives). This indicates that there is some misclassification of class 1 instances as class 0.

Insights:

1. The model seems to have a higher accuracy in classifying unpleasant days instances compared to pleasant days instances.
2. There is a notable number of false positives and false negatives, indicating that there is room for improvement in the model's ability to distinguish between pleasant and unpleasant days.

Dusseldorf Matrix

- Class 0 (True Negatives): There are 3369 instances correctly classified as class 0 (True Negatives). This suggests that the model is performing well in identifying instances belonging to class 0.
- Class 1 (True Positives): There are 727 instances correctly classified as class 1 (True Positives). This indicates that the model is also performing reasonably well in identifying instances belonging to class 1.
- False Positives: There are 246 instances incorrectly classified as class 1 (False Positives). This suggests that there is some misclassification of class 0 instances as class 1.
- False Negatives: There are 248 instances incorrectly classified as class 0 (False Negatives). This indicates that there is some misclassification of class 1 instances as class 0.

Insights:

1. The model seems to have a relatively higher accuracy in classifying unpleasant days instances compared to pleasant days instances.
2. There is a balanced number of false positives and false negatives, indicating that the model's performance in distinguishing between pleasant and unpleasant days instances is relatively balanced.

Heathrow Matrix

- Class 0 (True Negatives): There are 3320 instances correctly classified as class 0 (True Negatives). This indicates that the model is performing well in identifying instances belonging to class 0.
- Class 1 (True Positives): There are 628 instances correctly classified as class 1 (True Positives). This suggests that the model is also performing reasonably well in identifying instances belonging to class 1.
- False Positives: There are 335 instances incorrectly classified as class 1 (False Positives). This suggests that there is some misclassification of class 0 instances as class 1.
- False Negatives: There are 307 instances incorrectly classified as class 0 (False Negatives). This indicates that there is some misclassification of class 1 instances as class 0.

Insights:

1. The model seems to have a relatively higher accuracy in classifying unpleasant days instances compared to pleasant days instances.
2. There is a balanced number of false positives and false negatives, indicating that the model's performance in distinguishing between pleasant and unpleasant days instances is relatively balanced.

Kassel Matrix

- Class 0 (True Negatives): There are 3680 instances correctly classified as class 0 (True Negatives). This indicates that the model is performing well in identifying instances belonging to class 0.
- Class 1 (True Positives): There are 522 instances correctly classified as class 1 (True Positives). This suggests that the model is also performing reasonably well in identifying instances belonging to class 1.
- False Positives: There are 167 instances incorrectly classified as class 1 (False Positives). This suggests that there is some misclassification of class 0 instances as class 1.
- False Negatives: There are 221 instances incorrectly classified as class 0 (False Negatives). This indicates that there is some misclassification of class 1 instances as class 0.

Insights:

1. The model seems to have a relatively higher accuracy in classifying unpleasant days compared to pleasant days.
2. There is a balanced number of false positives and false negatives, indicating that the model's performance in distinguishing between pleasant and unpleasant days instances is relatively balanced.

Ljubljana Matrix

- Class 0 (True Negatives): There are 3017 instances correctly classified as class 0 (True Negatives). This indicates that the model is performing well in identifying instances belonging to class 0.
- Class 1 (True Positives): There are 967 instances correctly classified as class 1 (True Positives). This suggests that the model is also performing reasonably well in identifying instances belonging to class 1.
- False Positives: There are 344 instances incorrectly classified as class 1 (False Positives). This suggests that there is some misclassification of class 0 instances as class 1.
- False Negatives: There are 262 instances incorrectly classified as class 0 (False Negatives). This indicates that there is some misclassification of class 1 instances as class 0.

Insights:

1. The model seems to have a relatively higher accuracy in classifying unpleasant days compared to pleasant days.
2. There are some false positives and false negatives, indicating that there is room for improvement in the model's ability to distinguish between pleasant and unpleasant days instances.

Maastricht Matrix

- Class 0 (True Negatives): There are 3429 instances correctly classified as class 0 (True Negatives). This indicates that the model is performing well in identifying instances belonging to class 0.
- Class 1 (True Positives): There are 700 instances correctly classified as class 1 (True Positives). This suggests that the model is also performing reasonably well in identifying instances belonging to class 1.
- False Positives: There are 228 instances incorrectly classified as class 1 (False Positives). This suggests that there is some misclassification of class 0 instances as class 1.
- False Negatives: There are 233 instances incorrectly classified as class 0 (False Negatives). This indicates that there is some misclassification of class 1 instances as class 0.

Insights:

1. The model seems to have a relatively higher accuracy in classifying unpleasant days compared to pleasant days instances.
2. There are some false positives and false negatives, indicating that there is room for improvement in the model's ability to distinguish between pleasant and unpleasant days.

Madrid Matrix

- Class 0 (True Negatives): There are 2219 instances correctly classified as class 0 (True Negatives). This indicates that the model is performing well in identifying instances belonging to class 0.
- Class 1 (True Positives): There are 1812 instances correctly classified as class 1 (True Positives). This suggests that the model is also performing reasonably well in identifying instances belonging to class 1.
- False Positives: There are 338 instances incorrectly classified as class 1 (False Positives). This suggests that there is some misclassification of class 0 instances as class 1.
- False Negatives: There are 221 instances incorrectly classified as class 0 (False Negatives). This indicates that there is some misclassification of class 1 instances as class 0.

Insights:

1. The model seems to have a relatively higher accuracy in classifying both pleasant and unpleasant days.
2. The number of false positives and false negatives is relatively low, indicating that the model's ability to distinguish between pleasant and unpleasant days is relatively balanced.

Munchenb Matrix

- Class 0 (True Negatives): There are 3408 instances correctly classified as class 0 (True Negatives). This suggests that the model is performing well in identifying instances belonging to class 0.
- Class 1 (True Positives): There are 696 instances correctly classified as class 1 (True Positives). This indicates that the model is also performing reasonably well in identifying instances belonging to class 1.
- False Positives: There are 230 instances incorrectly classified as class 1 (False Positives). This suggests that there is some misclassification of class 0 instances as class 1.
- False Negatives: There are 256 instances incorrectly classified as class 0 (False Negatives). This indicates that there is some misclassification of class 1 instances as class 0.

Insights:

1. The model seems to have a relatively higher accuracy in classifying unpleasant days compared to pleasant days.
2. There are some false positives and false negatives, indicating that there is room for improvement in the model's ability to distinguish between unpleasant and pleasant days.

Oslo Matrix

- Class 0 (True Negatives): There are 3710 instances correctly classified as class 0 (True Negatives). This indicates that the model is performing well in identifying instances belonging to class 0.
- Class 1 (True Positives): There are 431 instances correctly classified as class 1 (True Positives). This suggests that the model is also performing reasonably well in identifying instances belonging to class 1.
- False Positives: There are 205 instances incorrectly classified as class 1 (False Positives). This suggests that there is some misclassification of class 0 instances as class 1.
- False Negatives: There are 244 instances incorrectly classified as class 0 (False Negatives). This indicates that there is some misclassification of class 1 instances as class 0.

Insights:

1. The model seems to have a relatively higher accuracy in classifying unpleasant days compared pleasant days.
2. There are some false positives and false negatives, indicating that there is room for improvement in the model's ability to distinguish between unpleasant and pleasant days.

Stockholm Matrix

- Class 0 (True Negatives): There are 3602 instances correctly classified as class 0 (True Negatives). This indicates that the model is performing well in identifying instances belonging to class 0.
- Class 1 (True Positives): There are 518 instances correctly classified as class 1 (True Positives). This suggests that the model is also performing reasonably well in identifying instances belonging to class 1.
- False Positives: There are 200 instances incorrectly classified as class 1 (False Positives). This suggests that there is some misclassification of class 0 instances as class 1.
- False Negatives: There are 270 instances incorrectly classified as class 0 (False Negatives). This indicates that there is some misclassification of class 1 instances as class 0.

Insights:

1. The model seems to have a relatively higher accuracy in classifying unpleasant days compared to pleasant days.
2. There are some false positives and false negatives, indicating that there is room for improvement in the model's ability to distinguish between unpleasant and pleasant days.

Valentia Matrix

- Class 0 (True Negatives): There are 4313 instances correctly classified as class 0 (True Negatives). This indicates that the model is performing well in identifying instances belonging to class 0.
- Class 1 (True Positives): There are 90 instances correctly classified as class 1 (True Positives). This suggests that the model is also performing reasonably well in identifying instances belonging to class 1.
- False Positives: There are 49 instances incorrectly classified as class 1 (False Positives). This suggests that there is some misclassification of class 0 instances as class 1.
- False Negatives: There are 138 instances incorrectly classified as class 0 (False Negatives). This indicates that there is some misclassification of class 1 instances as class 0.

Insights:

1. The model seems to have a relatively higher accuracy in classifying unpleasant days instances compared to pleasant days.
2. There are relatively few instances of pleasant days, which might indicate an imbalance in the dataset.
3. There are few false positives and false negatives, but they are still present, indicating that the model's performance could be further improved.

Sonnblick Matrix

- Class 0 (True Negatives): There are 4590 instances correctly classified as class 0.

Insights:

1. The confusion matrix indicates that there is only one class present in the dataset, and all instances are correctly classified as unpleasant days.
2. The model seems to have a perfect accuracy in classifying instances belonging to unpleasant days.
3. Since there is only one class, there are no false positives or false negatives to analyze.
4. Further evaluation metrics like precision, recall, or F1-score might not be applicable in this scenario since there is only one class.