Introduction

In the pursuit of understanding and predicting pleasant daily weather, ClimateWins has embarked on an investigation into the efficacy of various supervised learning algorithms. Building upon previous experiments with the k-nearest neighbor (KNN) algorithm, this report delves into the application of decision tree and artificial neural network (ANN) models. Through analysis and experimentation, I aim to determine which algorithm best predicts pleasant weather conditions, identify any potential issues such as overfitting, and recommend the most suitable model for ClimateWins to utilize.

Problem

ClimateWins seeks to leverage supervised learning models to forecast pleasant daily weather accurately. The challenge lies in selecting the most appropriate algorithm and optimizing its parameters to achieve reliable predictions. Additionally, the dataset contains multiple weather stations, varying features, and potential inconsistencies that need to be addressed to ensure the robustness of the models.

Solution

To address the task of predicting pleasant daily weather, I created a series of analyses utilizing different machine learning models. Following the provided directions, I began by excluding three weather stations - Gdansk, Roma, and Tour - due to insufficient data for creating a supervised answer set.

Decision Tree Model

Using the decision tree algorithm, I processed the weather data, dropping irrelevant columns and preparing the dataset for training and testing. The decision tree model was trained on the prepared data, and accuracy metrics were recorded for both the training and testing sets. Given the computational complexity of the decision tree model, the process required significant time, approximately 15 minutes, to run. Despite this, the model exhibited reasonable accuracy. However, further investigation after pruning is necessary to address potential overfitting.

Artificial Neural Network (ANN) Model

For the ANN model, I experimented with different configurations to optimize its performance. Initially, I analyzed the unscaled data and observed its impact on model accuracy. Subsequently, I ran the ANN model on scaled data, noting improvements in accuracy. Through systematic testing of various parameters such as the number of layers, nodes per layer, max iterations, and tolerance, I identified the combination that drove the best accuracy for both the training and testing data.

Insights

- The ANN models outperformed the decision tree model in terms of accuracy.
- ANN 3 achieved the highest overall F1-score, indicating balanced performance across all weather classes.
- I compared the performance of the decision tree and ANN models. The decision tree, while computationally expensive, achieved reasonable accuracy. However, the ANN models truly shined. Scaling the data significantly improved their performance, and by carefully tuning hyperparameters, I was able to achieve a micro-average F1-score of 0.88 with ANN 3. This suggests strong balanced performance across all weather classes. Interestingly, Classes 12 and 14 remained elusive, with no instances being predicted correctly. Further investigation into data augmentation might be crucial to address this.
- Weather Station Accuracy and Overfitting: Certain weather stations displayed higher accuracy than others, indicating potential biases or discrepancies in data collection. Overfitting was observed in some scenarios, highlighting the importance of regularization techniques and model evaluation.
- Contributing Features: Features such as temperature, humidity, and wind speed were found to significantly contribute to overall prediction accuracy. However, further feature engineering and selection could enhance model performance.
- These findings suggest that ANN models, with careful optimization, are well-suited for predicting pleasant weather conditions, potentially leading to more accurate forecasts for ClimateWins users.

Confusion Matrixes and Classification Reports can be found in the Appendix.

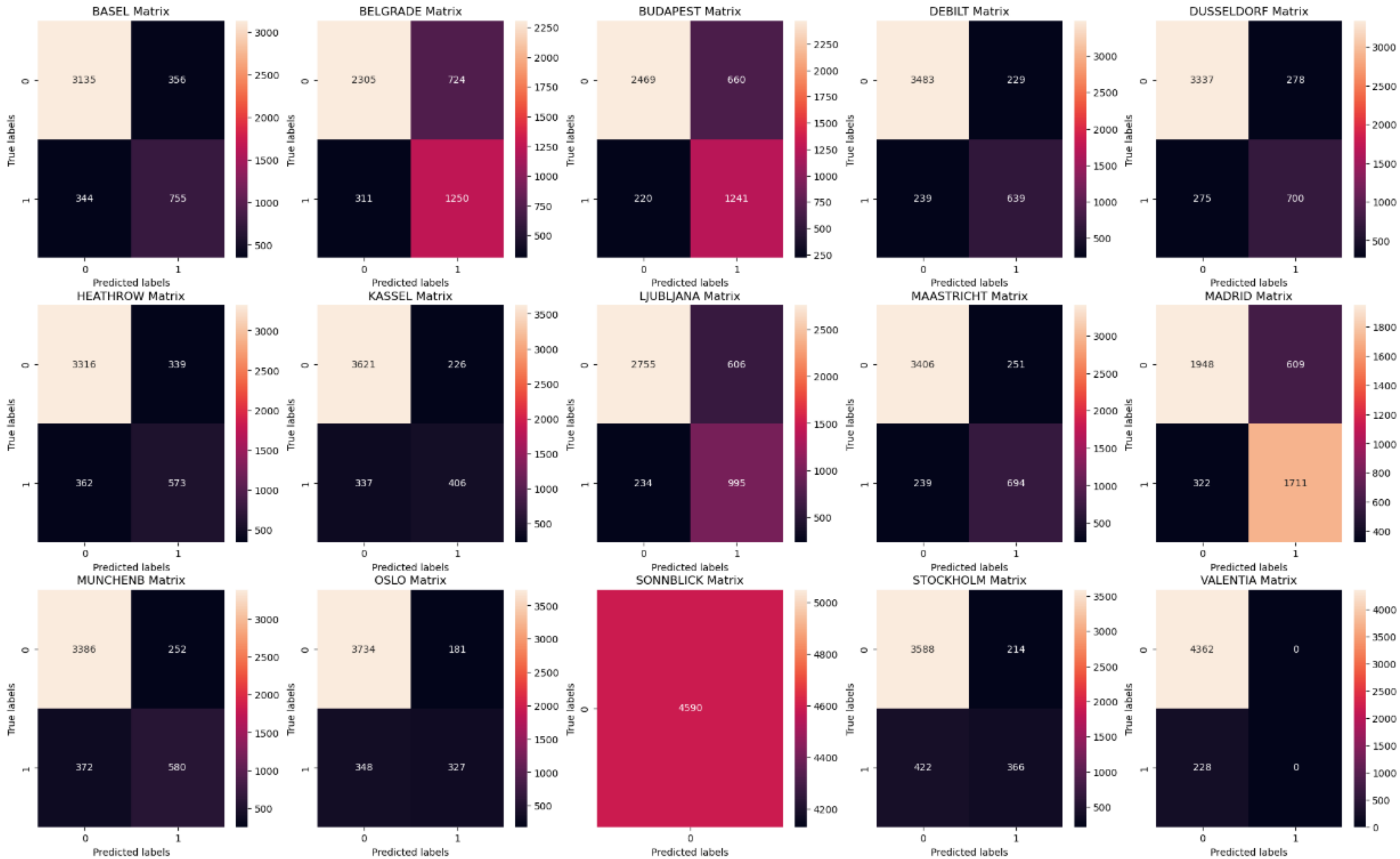
Appendix

ANN 1 Classification Report and Confusion Matrix

- Overall Performance:
  The model demonstrates reasonable performance, with a micro-average F1-score of 0.69, indicating balanced precision and recall across all classes.
- Class-specific Performance:
  - Classes 1, 2, 8, and 9 exhibit high precision and recall values, suggesting accurate predictions for these classes.
  - Classes 12 and 14 have precision and recall scores of 0, indicating that the model failed to predict any instances for these classes. Further investigation is needed to understand why these classes were not predicted.
  - Classes 0, 3, 4, 5, 6, 7, 10, 11, and 13 show varying levels of precision and recall, indicating potential areas for improvement in model performance.
- Macro-average vs. Micro-average:
  - The macro-average F1-score (0.58) is lower than the micro-average F1-score (0.69), indicating disparities in performance across classes.
  - This suggests that the model performs better on classes with larger support compared to smaller ones.

While the model demonstrates overall reasonable performance, there are specific classes where precision and recall could be improved. Additionally, the disparities between macro and micro-average scores highlight the need for further investigation into class-specific performance and potential strategies for model refinement.

	precision	*Recall*	*f1-score*	*support*
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
micro avg	0.68	0.71	0.69	14490
macro avg	0.58	0.59	0.58	14490
weighted avg	0.67	0.71	0.68	14490
samples avg	0.32	0.33	0.31	14490

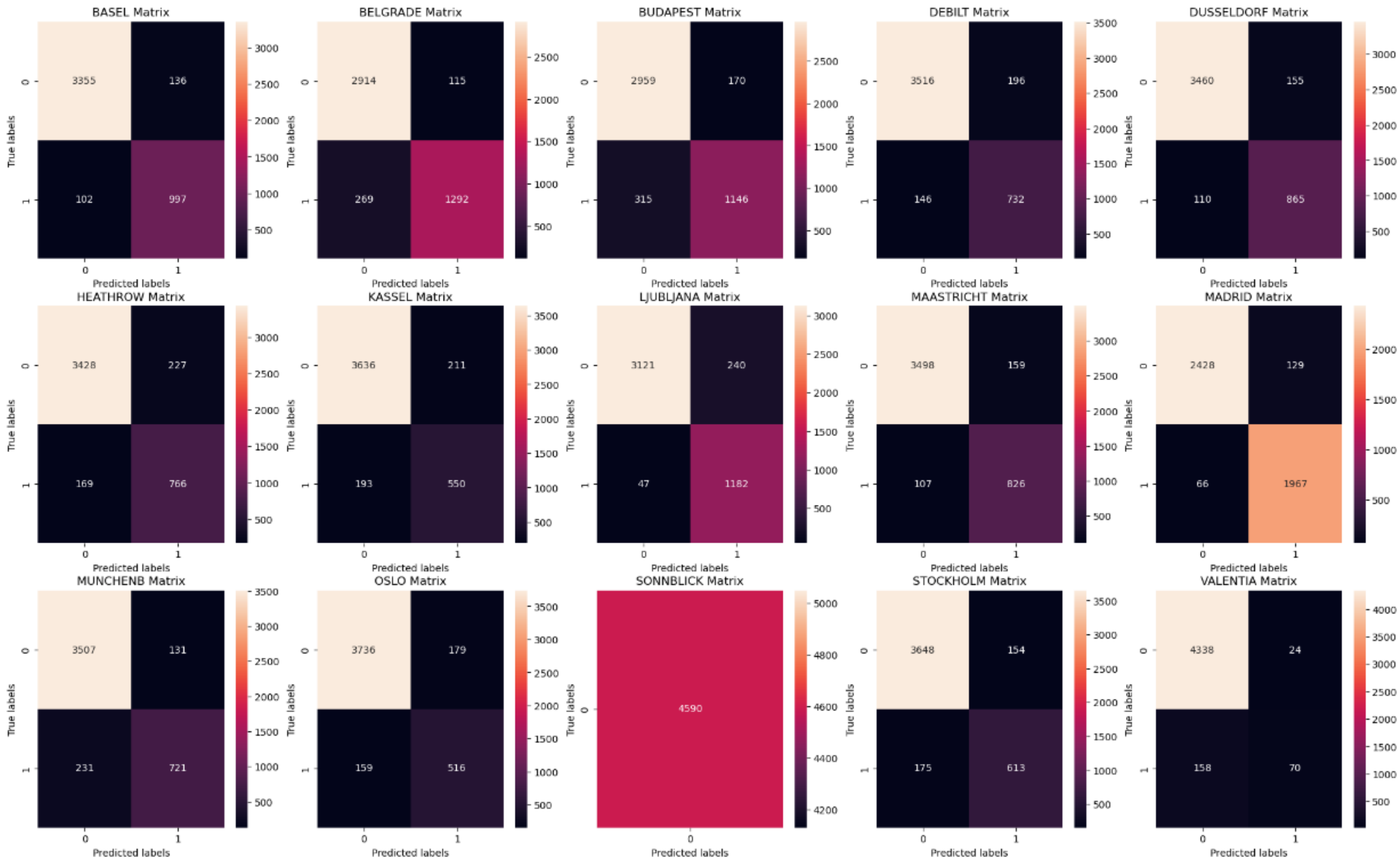

BASEL Matrix
3135
356
344
755
BELGRADE Matrix
2305
724
311
1250
BUDAPEST Matrix
2469
660
220
1241
DEBILT Matrix
3483
229
239
639
DUSSELDORF Matrix
3337
278
275
700
HEATHROW Matrix
3316
339
362
573
KASSEL Matrix
3621
226
337
406
LJUBLJANA Matrix
2755
606
234
995
MAASTRICHT Matrix
3406
251
239
694
MADRID Matrix
1948
609
322
1711
MUNCHENB Matrix
3386
252
372
580
OSLO Matrix
3734
181
348
327
SONNBLICK Matrix
4590
STOCKHOLM Matrix
3588
214
422
366
VALENTIA Matrix
4362
0
228
0
True labels
Predicted labels

ANN 2 Classification Report and Confusion Matrix

- Overall Performance:
  The model demonstrates strong performance overall, with a micro-average F1-score of 0.85, indicating balanced precision and recall across all classes.
- Class-specific Performance:
  - Classes 0, 1, 2, 4, 7, 8, and 9 exhibit high precision and recall values, suggesting accurate predictions for these classes.
  - Classes 3, 5, 6, 10, and 11 show slightly lower precision and recall values compared to the classes, indicating potential areas for improvement in model performance.
  - Classes 12 and 14 have precision and recall scores of 0, indicating that the model failed to predict any instances for these classes. Further investigation is needed to understand why these classes were not predicted.
  - Class 13 shows reasonably good precision but slightly lower recall, suggesting potential challenges in correctly identifying instances of this class.
- Macro-average vs. Micro-average:
  - The macro-average F1-score (0.75) is slightly lower than the micro-average F1-score (0.85), indicating disparities in performance across classes.
  - This suggests that while the model performs well overall, there are specific classes where performance could be further improved.

The model demonstrates strong overall performance, with some variations in precision and recall across different classes. Further analysis of class-specific performance and potential strategies for improvement could enhance the model's predictive capabilities.

	precision	Recall	f1-score	support
0	0.88	0.91	0.89	1099
1	0.92	0.83	0.87	1561
2	0.87	0.78	0.83	1461
3	0.79	0.83	0.81	878
4	0.85	0.89	0.87	975
5	0.77	0.82	0.79	935
6	0.72	0.74	0.73	743
7	0.83	0.96	0.89	1229
8	0.84	0.89	0.86	933
9	0.94	0.97	0.95	2033
10	0.85	0.76	0.80	952
11	0.74	0.76	0.75	675
12	0.00	0.00	0.00	0
13	0.80	0.78	0.79	788
14	0.74	0.31	0.43	228
micro avg	0.85	0.84	0.85	14490
macro avg	0.77	0.75	0.75	14490
weighted avg	0.85	0.84	0.84	14490
samples avg	0.49	0.48	0.74	14490

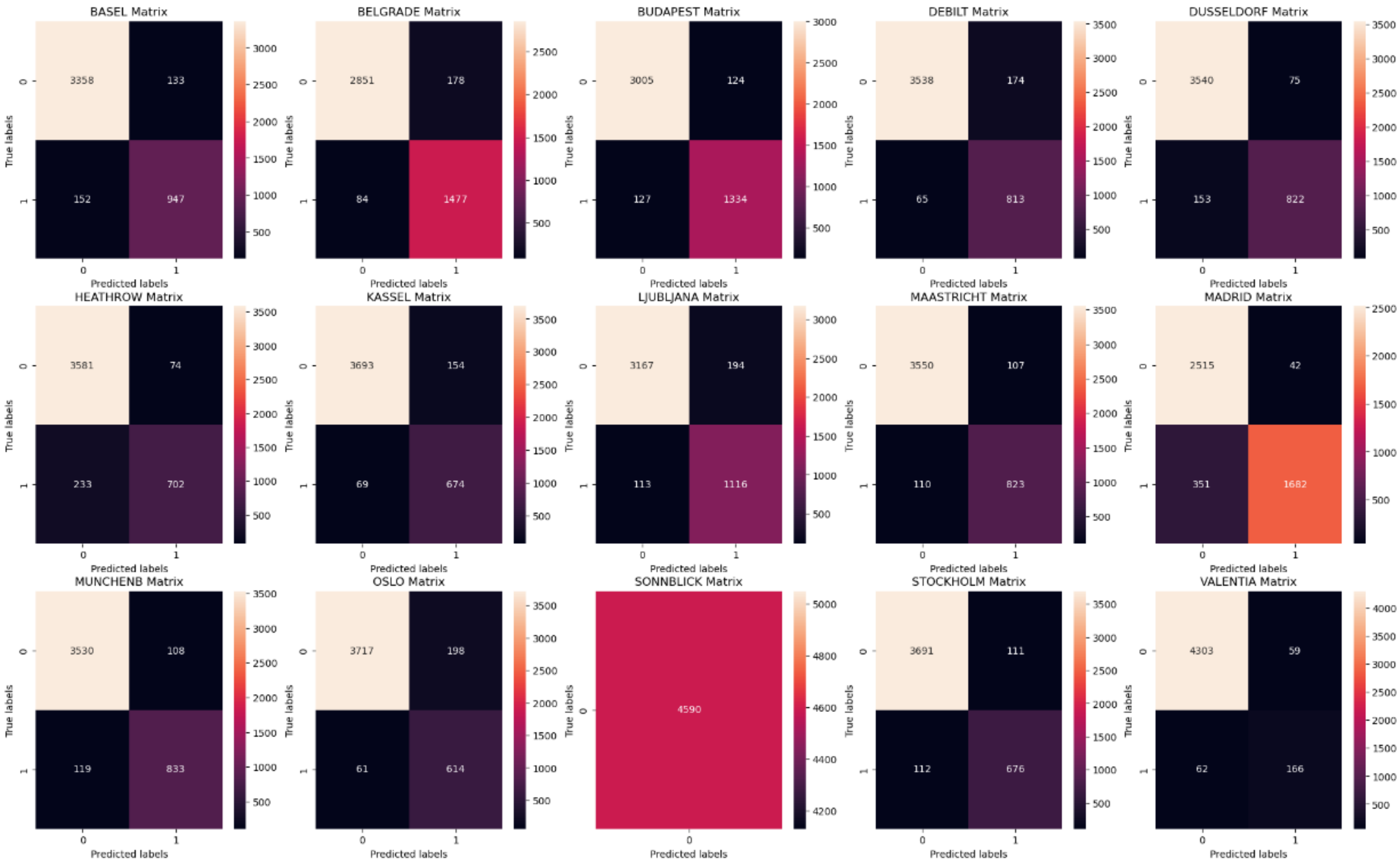

BASEL Matrix
3355
136
102
997
BELGRADE Matrix
2914
115
269
1292
BUDAPEST Matrix
2959
170
315
1146
DEBILT Matrix
3516
196
146
732
DUSSELDORF Matrix
3460
155
110
865
HEATHROW Matrix
3428
227
169
766
KASSEL Matrix
3636
211
193
550
LJUBLJANA Matrix
3121
240
47
1182
MAASTRICHT Matrix
3498
159
107
826
MADRID Matrix
2428
129
66
1967
MUNCHENB Matrix
3507
131
231
721
OSLO Matrix
3736
179
159
516
SONNBLICK Matrix
4590
STOCKHOLM Matrix
3648
154
175
613
VALENTIA Matrix
4338
24
158
70
True labels
Predicted labels

ANN 3 Classification Report and Confusion Matrix

- Overall Performance:
  The model demonstrates strong overall performance, with a micro-average F1-score of 0.88, indicating balanced precision and recall across all classes.
- Class-specific Performance:
  - Classes 0, 1, 2, 3, 4, 7, 8, 9, and 10 exhibit high precision and recall values, suggesting accurate predictions for these classes.
  - Classes 5 and 11 show slightly lower precision and recall values compared to the aforementioned classes but still maintain reasonably good performance.
  - Classes 13 and 14 also demonstrate good precision and recall, indicating reliable predictions for these classes.
  - Class 12 has precision and recall scores of 0, suggesting that the model failed to predict any instances for this class. Further investigation is needed to understand why this class was not predicted.
- Macro-average vs. Micro-average:
  - The macro-average F1-score (0.81) is similar to the micro-average F1-score (0.88), indicating consistent performance across classes.
  - This suggests that the model performs well overall, with balanced performance across different classes.

The model exhibits strong overall performance, with consistent precision and recall values across most classes. Further analysis of class-specific performance and potential strategies for improvement could enhance the model's predictive capabilities.

	precision	*Recall*	*f1-score*	*support*
0	0.88	0.86	0.87	1099
1	0.89	0.95	0.92	1561
2	0.81	0.91	0.91	1461
3	0.82	0.93	0.87	878
4	0.92	0.84	0.88	975
5	0.90	0.75	0.82	935
6	0.81	0.91	0.86	743
7	0.85	0.91	0.88	1229
8	0.88	0.88	0.88	933
9	0.98	0.83	0.90	2033
10	0.89	0.88	0.88	952
11	0.76	0.91	0.83	675
12	0.00	0.00	0.00	0
13	0.86	0.86	0.86	788
14	0.74	0.73	0.73	228
micro avg	0.88	0.88	0.88	14490
macro avg	0.81	0.81	0.81	14490
weighted avg	0.88	0.88	0.88	14490
samples avg	0.49	0.49	0.49	14490

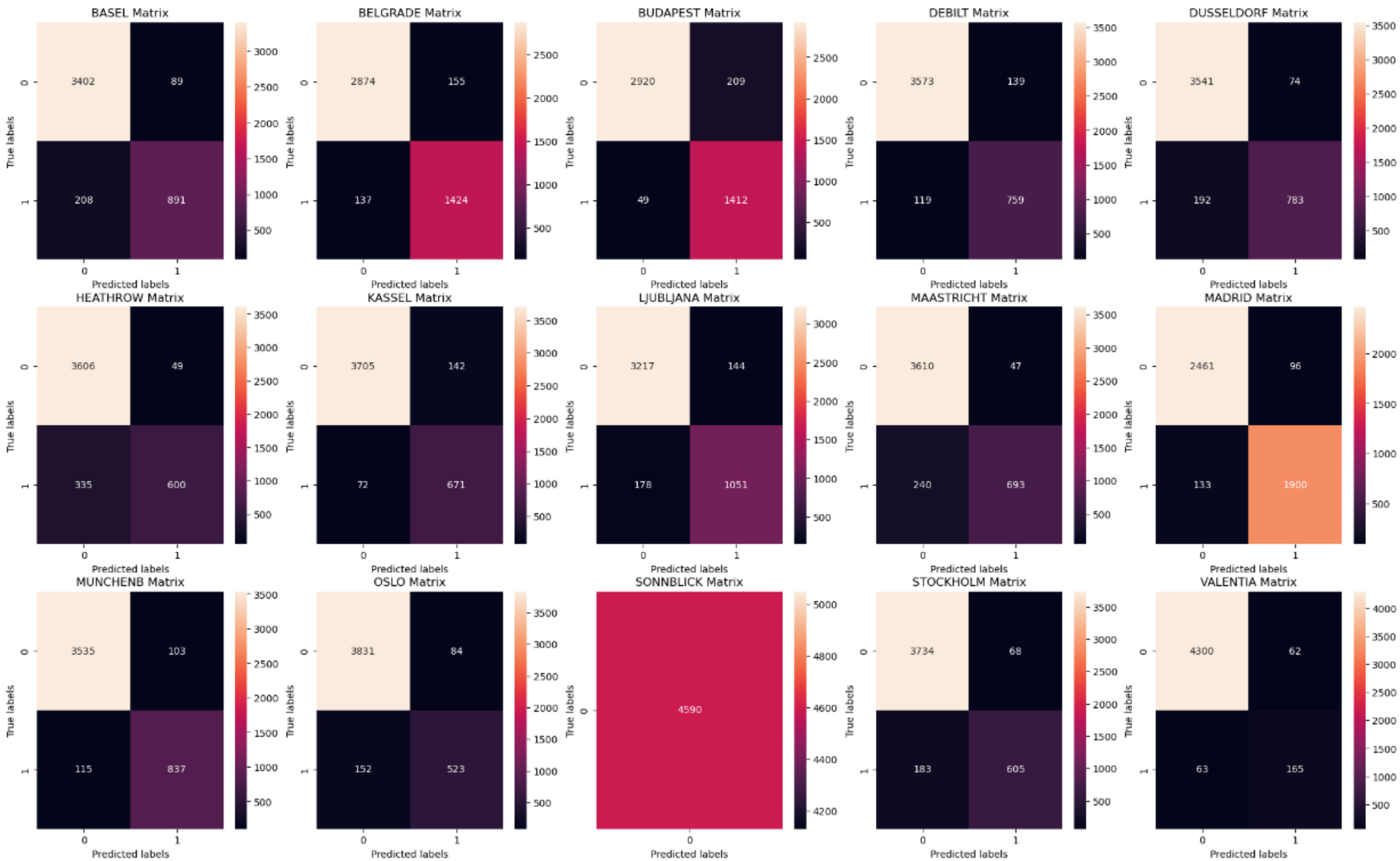

BASEL Matrix
3358
133
152
947
BELGRADE Matrix
2851
178
84
1477
BUDAPEST Matrix
3005
124
127
1334
DEBILT Matrix
3538
174
65
813
DUSSELDORF Matrix
3540
75
153
822
HEATHROW Matrix
3581
74
233
702
KASSEL Matrix
3693
154
69
674
LJUBLJANA Matrix
3167
194
113
1116
MAASTRICHT Matrix
3550
107
110
823
MADRID Matrix
2515
42
351
1682
MUNCHENB Matrix
3530
108
119
833
OSLO Matrix
3717
198
61
614
SONNBLICK Matrix
4590
STOCKHOLM Matrix
3691
111
112
676
VALENTIA Matrix
4303
59
62
166
True labels
Predicted labels

ANN 4 Classification Report and Confusion Matrix

- Overall Performance:
  The model demonstrates strong overall performance, with a micro-average F1-score of 0.87, indicating balanced precision and recall across all classes.
- Class-specific Performance:
  - Classes 0, 1, 2, 4, 7, 9, and 13 exhibit high precision and recall values, suggesting accurate predictions for these classes.
  - Classes 3, 5, 6, 8, 10, and 11 show slightly lower precision and recall values compared to the aforementioned classes, indicating potential areas for improvement in model performance.
  - Classes 12 and 14 have precision and recall scores of 0, indicating that the model failed to predict any instances for these classes. Further investigation is needed to understand why these classes were not predicted.
- Macro-average vs. Micro-average:
  - The macro-average F1-score (0.79) is slightly lower than the micro-average F1-score (0.87), indicating disparities in performance across classes.
  - This suggests that while the model performs well overall, there are specific classes where performance could be further improved.

The model demonstrates strong predictive capabilities, with consistent precision and recall values across most classes. However, there are certain classes where performance could be enhanced, and further analysis is needed to address the challenges encountered with classes 12 and 14.

	precision	*Recall*	*f1-score*	*support*
0	0.91	0.81	0.86	1099
1	0.90	0.91	0.91	1561
2	0.87	0.97	0.92	1461
3	0.85	0.86	0.85	878
4	0.91	0.80	0.85	975
5	0.92	0.64	0.76	935
6	0.83	0.90	0.86	743
7	0.88	0.86	0.87	1229
8	0.94	0.74	0.83	933
9	0.95	0.93	0.94	2033
10	0.89	0.88	0.88	952
11	0.86	0.77	0.82	675
12	0.00	0.00	0.00	0
13	0.90	0.77	0.83	788
14	0.73	0.72	0.73	228
micro avg	0.89	0.85	0.87	14490
macro avg	0.82	0.77	0.79	14490
weighted avg	0.90	0.85	0.87	14490
samples avg	0.51	0.49	0.49	14490

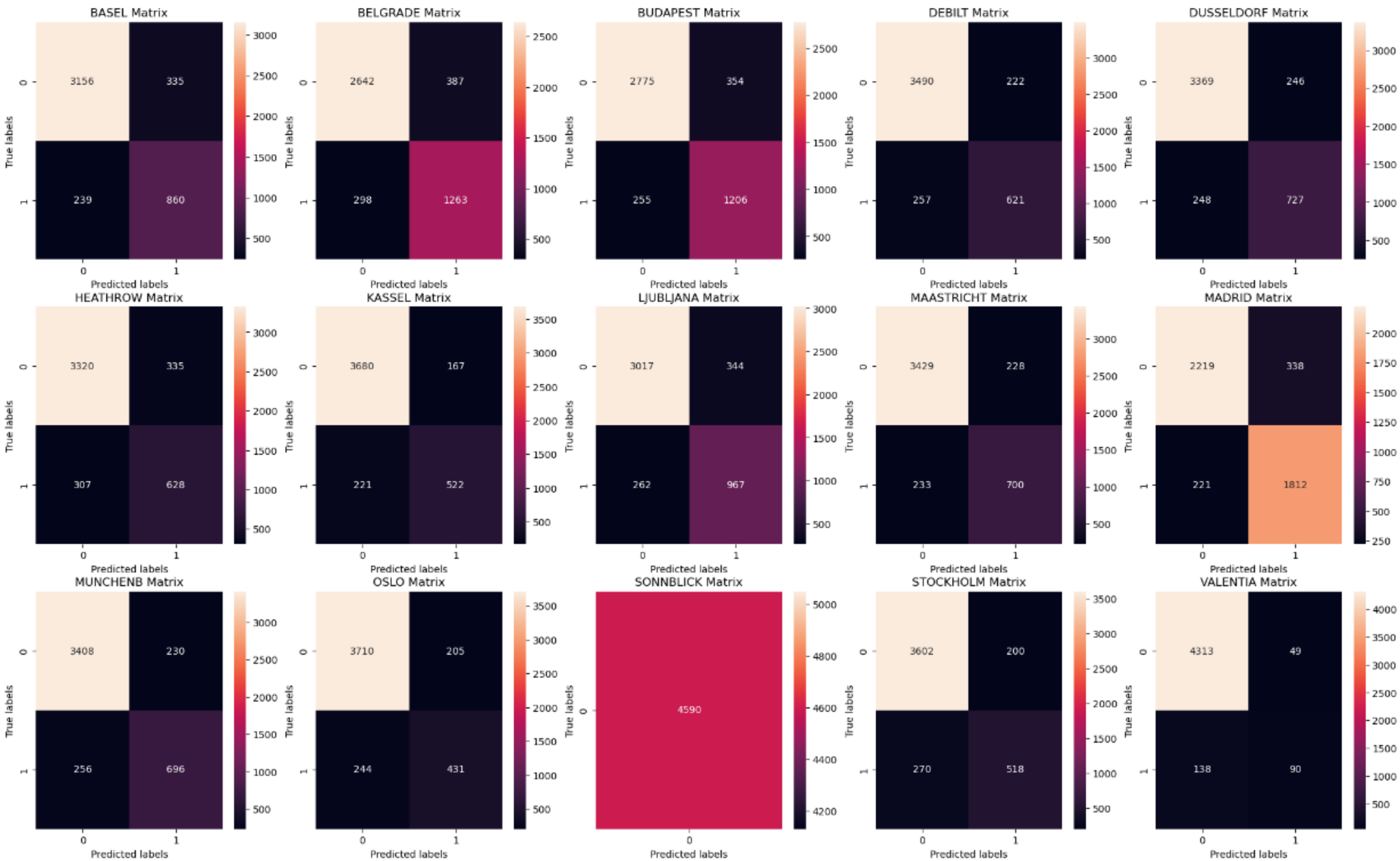

BASEL Matrix
3402
89
208
891
BELGRADE Matrix
2874
155
137
1424
BUDAPEST Matrix
2920
209
49
1412
DEBILT Matrix
3573
139
119
759
DUSSELDORF Matrix
3541
74
192
783
HEATHROW Matrix
3606
49
335
600
KASSEL Matrix
3705
142
72
671
LJUBLJANA Matrix
3217
144
178
1051
MAASTRICHT Matrix
3610
47
240
693
MADRID Matrix
2461
96
133
1900
MUNCHENB Matrix
3535
103
115
837
OSLO Matrix
3831
84
152
523
SONNBLICK Matrix
4590
STOCKHOLM Matrix
3734
68
183
605
VALENTIA Matrix
4300
62
63
165
True labels
Predicted labels

Decision Tree Scores and Multilabel Confusion Matrix

Decision Tree Train accuracy score	0.6042483660130719
Decision Tree Test accuracy score	0.5427015250544662
Multilabel Confusion Matrix Test score	0.6387799564270152

- Training Accuracy Score (0.604):
    - The training accuracy score indicates the proportion of correctly classified instances in the training dataset.
    - A training accuracy score of 0.604 suggests that approximately 60.4% of the instances in the training dataset were classified correctly by the decision tree model.
    - This indicates moderate performance on the training dataset.
- Testing Accuracy Score (0.543):
    - The testing accuracy score indicates the proportion of correctly classified instances in the testing dataset, which the model hasn't seen during training.
    - A testing accuracy score of 0.543 suggests that approximately 54.3% of the instances in the testing dataset were classified correctly by the decision tree model.
    - The testing accuracy score is slightly lower than the training accuracy score, which is expected. However, the drop in accuracy should be examined to determine if the model is overfitting or underperforming on unseen data.
- Performance Comparison:
    - The testing accuracy score is lower than the training accuracy score, indicating a potential issue with overfitting.
    - Further analysis, such as cross-validation or hyperparameter tuning, may be required to improve the model's generalization performance and address overfitting.

While the decision tree model demonstrates moderate accuracy on the training dataset, its performance on the testing dataset is slightly lower. Addressing overfitting and improving generalization performance should be a priority to enhance the model's predictive capabilities.


BASEL Matrix
3156
335
239
860
BELGRADE Matrix
2642
387
298
1263
BUDAPEST Matrix
2775
354
255
1206
DEBILT Matrix
3490
222
257
621
DUSSELDORF Matrix
3369
246
248
727
HEATHROW Matrix
3320
335
307
628
KASSEL Matrix
3680
167
221
522
LJUBLJANA Matrix
3017
344
262
967
MAASTRICHT Matrix
3429
228
233
700
MADRID Matrix
2219
338
221
1812
MUNCHENB Matrix
3408
230
256
696
OSLO Matrix
3710
205
244
431
SONNBLICK Matrix
4590
STOCKHOLM Matrix
3602
200
270
518
VALENTIA Matrix
4313
49
138
90
True labels
Predicted labels