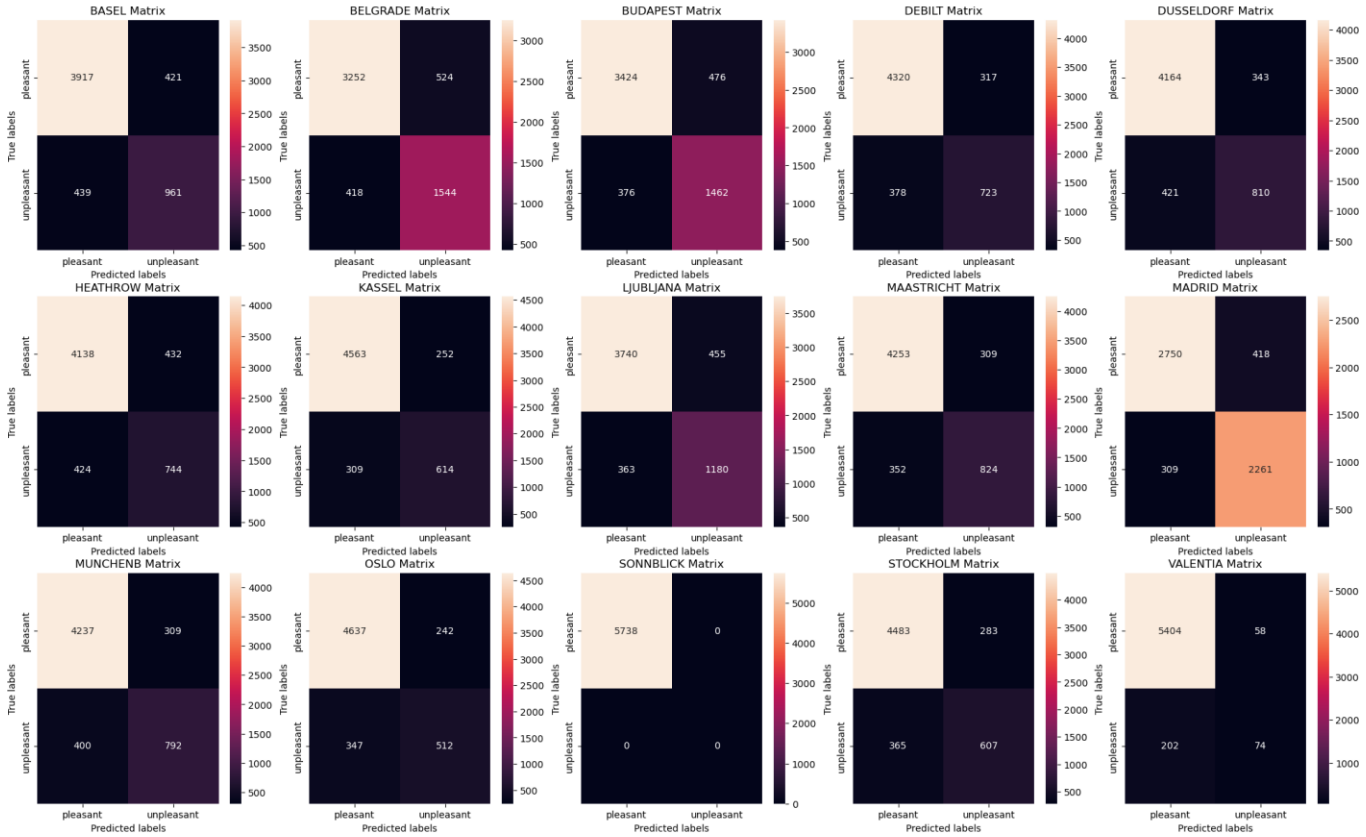
SHRAVANI
11/7/24

Task-1.4: Supervised Learning Algorithms Part 1



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%
Dusseldorf	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%

Model Performance and Accuracy

The weather prediction model achieves an average accuracy rate of 88% across the stations, indicating it performs reasonably well overall. However, the individual station accuracies reveal a more complex picture, with some locations performing significantly better than others. For instance, stations like Kassel and Oslo achieve high accuracy rates around 90%, while others, such as Belgrade and Madrid, have slightly lower accuracy rates around 84-87%, suggesting room for improvement depending on the location.

Perfect Accuracy at Sonnblick

Sonnblick stands out with a 100% accuracy rate, where the model correctly classifies all instances of unpleasant weather. While impressive, this may actually indicate overfitting rather than robust performance. The model's perfection in predicting Sonnblick's data could be due to a narrow training pattern that doesn't fully capture the variety of conditions present in other locations. This implies that the model may be overly tailored to Sonnblick's specific data, rather than being flexible enough to generalize well across different patterns.

Accuracy Variations and Misclassification Rates

Some stations experience higher rates of false positives or false negatives, which contribute to their lower accuracy. For example, Madrid has a particularly high false positive count (2261), meaning that pleasant weather is often incorrectly predicted as unpleasant. This discrepancy suggests that the model may struggle with stations that have more variable or less predictable weather conditions. These differences highlight the need to account for local climatic variability in model training.

Potential Overfitting and Training Limitations

The perfect score at Sonnblick raises concerns about overfitting, as it suggests the model might have memorized Sonnblick's weather patterns rather than learning more adaptable features. The overfitting concern is compounded by the fact that some stations with complex weather patterns, such as Madrid, show lower performance, indicating that the model might not handle diverse weather conditions well.

Influence of Local Climate and Geography on Accuracy

The model's varying accuracy across stations suggests that geographical and climatic differences impact its performance. Locations like Sonnblick, with potentially consistent weather patterns, seem to suit the model better, while stations in regions with more variable weather, such as Belgrade and Madrid, exhibit lower accuracy. This points to the need for more diverse and representative training data, particularly for stations in areas with complex weather patterns.

Improving Generalizability

To make the model more universally accurate, the training process should incorporate a broader and more varied dataset from each station, covering both pleasant and unpleasant weather conditions more evenly. Adding more balanced data for locations with lower accuracy rates could help the model learn a wider range of weather conditions, thus reducing overfitting and improving overall generalization.

Practical Implications

While the average accuracy of 88% may give an impression of reliability, the disparity across stations indicates that certain predictions might be less trustworthy depending on the location. In practical applications, such a model would benefit from further tuning to ensure reliable performance across different weather stations, reducing the risk of misleading results for end-users in stations with lower accuracy rates.

Conclusion

In conclusion, the weather prediction model performs adequately on average, but its accuracy is highly dependent on location-specific factors. Addressing overfitting, especially at stations like Sonnblick, and enhancing data diversity across stations with lower accuracy rates could improve the model's robustness and make it better suited for real-world forecasting.