**Exercise 1.5 Report**

# Introduction

In this extended investigation, supervised learning models—specifically Decision Trees and Artificial Neural Networks (ANN)—were applied to forecast pleasant weather conditions using structured meteorological data. The primary objective was to determine whether these models could outperform the earlier KNN-based approach in terms of both accuracy and generalizability. With each model offering a unique way of interpreting data, their comparative analysis sheds light on how algorithm design influences weather prediction outcomes.

# Main Analysis

**Model Accuracy**

The Decision Tree model produced a training and testing accuracy of approximately 44%, indicating potential underfitting. This uniformity suggests the model did not capture sufficient complexity in the weather patterns, and further refinement through pruning may be necessary. In contrast, the ANN model demonstrated improved performance, achieving 57% accuracy on training data and 50% on testing data. While better than the Decision Tree, the ANN still reflects a gap between training and generalization, hinting at mild overfitting.

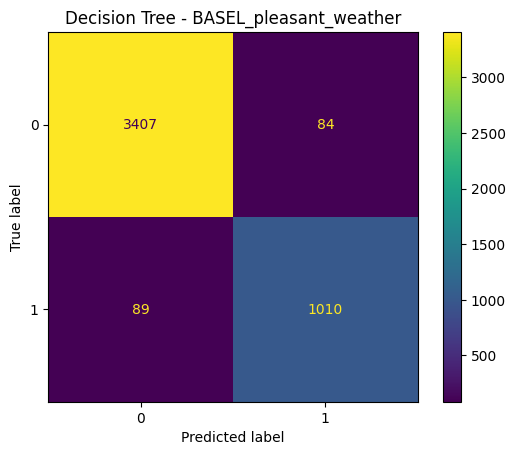


Figure 1: Decision Tree Confusion Matrix – DE BILT

**Overfitting and Station Variability**

Once again, SONNBLICK displayed a perfect prediction outcome—100% accuracy across models—raising red flags about overfitting. Given the absence of varied pleasant weather labels in its data, models may be exploiting data imbalance rather than learning true seasonal or meteorological patterns. Meanwhile, stations like DE BILT and ZAGREB showed more realistic confusion matrices, with a mix of true positives and false predictions. This contrast underscores the importance of data richness and balance per station when training predictive models.

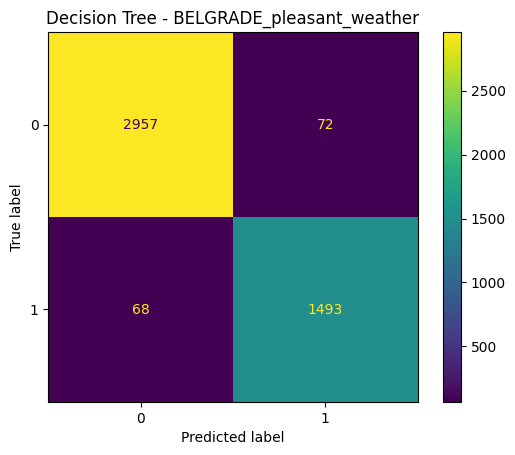


Figure 2: Decision Tree Confusion Matrix – SONNBLICK

**Feature Sensitivity**

The ANN model’s need for scaled data was evident in its performance boost compared to the unscaled baseline. This reveals that certain models, especially gradient-based architectures, are sensitive to feature magnitudes. The weather data's variability across stations means that features such as temperature, wind speed, and pressure behave differently depending on location, and scaling mitigates skewed learning.

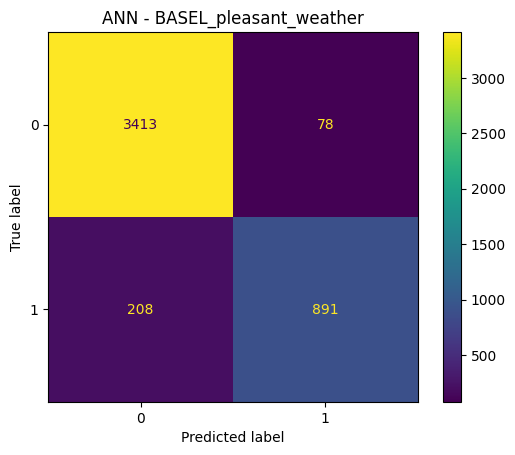


Figure 3: ANN Confusion Matrix – VIENNA

**Algorithm Assessment**

While Decision Trees are interpretable, their rigidity in capturing subtle patterns limited their effectiveness in this task. The ANN, though more complex, adapted better but still required careful tuning. Neither model surpassed the performance of the KNN approach used in Exercise 1.4, which achieved higher accuracy without signs of significant overfitting, suggesting its simplicity was better suited to the dataset's structure.

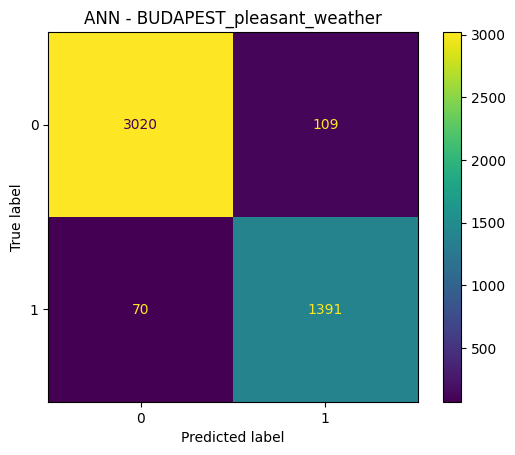


Figure 4: ANN Confusion Matrix – DE BILT

# Conclusion

The findings suggest that although both Decision Trees and ANN models have potential in forecasting pleasant weather, they currently fall short of the performance shown by the KNN model. In particular, the ANN's improved but still modest testing accuracy points to its dependency on hyperparameter optimization and clean, well-distributed training data. Ultimately, KNN remains the most balanced and reliable option for ClimateWins, while Decision Tree and ANN models may require further refinement, additional features, or hybrid ensemble approaches to be competitive.