Exercise 2.4: Hyperparameter Optimization

All Weather Stations Pre-Optimization (1990s)

Random Forest (96% accuracy)

A diagram of a company

AI-generated content may be incorrect.

Feature Importances

A graph of a number of different weather stations

AI-generated content may be incorrect.

All Weather Stations Post-Optimization (1990s)

Random Forest (97% accuracy)

A diagram of a company

AI-generated content may be incorrect.

Feature Importances

A graph of different weather stations

AI-generated content may be incorrect.

Belgrade after Optimization

Random Forest (Accuracy 100%)

A computer screen shot of a network

AI-generated content may be incorrect.

Feature Importance

A graph with blue bars

AI-generated content may be incorrect.

Observations

1. Overall Accuracy increased by 1%
2. Belgrade was the highest weighted weather station after Optimization whilst previously it was Madrid
3. Decision Tree complexity reduced after Optimization less crowded and easier to interpret.
4. Previously we saw that the most important features where temp\_mean, temp\_max and precipitation whilst after optimization for Belgrade we see that temp\_mean has less weight than sunshine

# Deep Learning

**CNN Pre-Optimization:**

Accuracy = 12%

Loss = 60413308.0000

**CNN Post-Optimization:**

Accuracy = 91%

Loss = 0.2606

|  |  |  |
| --- | --- | --- |
| **Hyperparameters** | **Pre-Optimization** | **Post-Optimization** |
| Neurons | 32 | 61 |
| Epochs | 32 | 47 |
| Batch\_size | 16 | 460 |
| Learning\_rate | - | 0.763 |
| Kernel | 2 | 2 |
| Activation | softmax | softsign |
| Optimizer | Adam | Adadelta |
| Layers 1 | - | 1 |
| Layers 2 | - | 2 |
| Normalization | - | 0.771 |
| Dropout | - | 0.730 |
| Dropout\_rate | - | 0.191 |

Confusion Matrix

A table of numbers and names

AI-generated content may be incorrect.

• Optimizer: Switched from Adam to Adadelta.

• Learning Rate: Fine-tuned to 0.763.

• Batch Size: Increased from 32 to 460.

• Hidden Layer Neurons: Increased from 32 to 61

. • Activation Function: Changed from "softmax" to "softsign".

These Changes Increased our overall Accuracy by 79% from 12% to 91% whilst reducing the loss exponentially, it has however only predicted 11 weather stations from the total 15 this could be due to the 1990s not having enough data I would rerun the test with data from the 2010s and to see if the accuracy still increases or it can predict all weather stations.

# Iterations

To improve weather predictions for the Air Ambulance company, I suggest breaking the data into smaller parts based on location, time, or weather features. For location, we can use data from individual weather stations or group similar ones, like continental and coastal stations, to focus on local weather. Dividing data by time (like seasons or months) can help identify trends, and organizing by weather factors like temperature and precipitation can show how they affect flight safety.

The choice of model depends on what is needed. An optimized random forest model is easy to understand and manage, as it shows which factors are important and reduces overfitting. However, it may not handle complex data well and can be slow with large datasets. On the other hand, an optimized CNN is good at finding complex patterns and works well with large datasets, but it is harder to interpret and requires careful tuning.

I recommend starting with the optimized random forest for analysing single weather stations, where accuracy can be very high. Later, we can use the optimized CNN to explore more complex patterns. The most important factors for weather predictions in the random forest were maximum temperature and precipitation, so we should focus on these in our weather models to improve flight safety for Air Ambulance operations.