Every model is analysed with a metric function that prints the classification report and displays a confusion matrix for further visualization of the results. This chapter will only focus on the best performing model.

The most important metric for the model performance is the accuracy as it represents the ratio of the sum of true positives and true negatives out of all predictions. A higher accuracy indicates a better performance. When there is a high difference between precision and recall it is recommended to chose the weighted average of the f1-score for model comparison, but this is not the case here. In the following table all tested models are ranked based on their accuracy score.

|  |  |  |  |
| --- | --- | --- | --- |
| Nr. | Model Description | Data Preprocessing | Accuracy |
| 9 | Decision Tree with Entropy and max Depth 15 | Drop Rows with Null Values  Drop least important Features  Encode Categorical to Numerical | 0.884 |
| 5 | Decision Tree with Entropy and max Depth 15 | Drop Rows with Null Values  Encode Categorical to Numerical | 0.881 |
| 8 | Decision Tree with Entropy and max Depth 15 | Drop Rows with Null Values  Create Dummies for Categorical | 0.879 |
| 7 | Decision Tree with Entropy and max Depth 15 | Replace Null Values with ‘-1’  Encode Categorical to Numerical | 0.878 |
| 6 | K-Nearest-Neighbour | Drop Rows with Null Values  Encode Categorical to Numerical | 0.862 |
| 3 | Decision Tree | Drop Rows with Null Values  Encode Categorical to Numerical | 0.858 |
| 4 | Weighted Decision Tree | Drop Rows with Null Values  Encode Categorical to Numerical | 0.858 |
| 1 | Unscaled Linear Regression | Drop Rows with Null Values  Encode Categorical to Numerical | 0.759 |
| 2 | Scaled Linear Regression | Drop Rows with Null Values  Encode Categorical to Numerical | 0.759 |

Model 9 has achieved the highest accuracy. Therefore the results of this model will be explained in more detail.

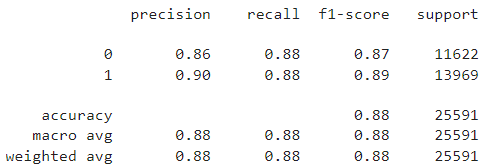
When checking the performance with the training dataset all metrics are approximately 92% which shows that the model is not overfitted to the training dataset, but still quite accurate. This is proven when reviewing the classification report for the test dataset, shown in figure x.x, as all metrics are still around 88%.

Figure x.x: Classification Report for Model 9

As the precision score shows, 90% of the predictions for a good overall experience where correct and the predictions that the customer was unsatisfied were correct in 86% of the cases.

The recall is 88% for both classes, meaning that 88% of the actual cases were caught by the model.

The f1-score is a weighted harmonic mean of precision and recall and for this model it is 87% for bad and 89% for good overall experience, meaning that the model is slightly better in detecting satisfied customers.

For each of these metrics there is also a macro average and a weighted average displayed, being 88% for all. The column support shows how many samples exist for each class.

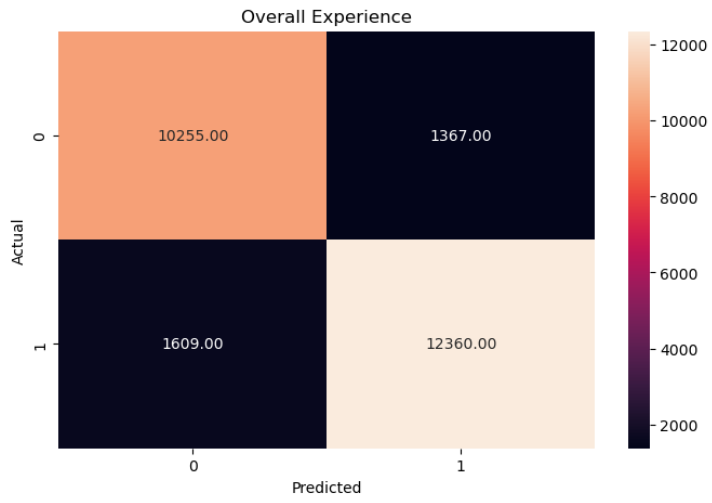
The confusion matrix, illustrated in figure x.x, visualizes the absolute number of true negatives/positives and false negatives/positives. As these are used for the calculation of the metrics it is a easy overview of the model’s performance.

Figure x.x: Confusion Matrix for Model 9

There are four ways to check if the predictions are right or wrong:

1. **True Negative (TN):** the customer was unsatisfied (0) and the model predicted a bad overall experience (0)
2. **True Positive (TP):** the customer was satisfied (1) and the model predicted a good overall experience (1)
3. **False Negative (FN):** the customer was satisfied (1), but the model predicted a bad overall experience (0)
4. **False Positive (FP):** the customer was unsatisfied (0) and the model predicted a good overall experience (1)

The total numbers of these cases are represented in the confusion matrix as shown in figure x.x



Figure x.x: Confusion Matrix

In the classification report there are 3 different metrics for each class (0/1):

1. **Precision:**