# Classification report

A classification report is a summary of performance metrics *for each class*:

Precision: Percentage of correct positive predictions relative to total positive predictions

Recall: Percentage of correct positive predictions relative to total actual positives.

F1 Score: A weighted harmonic mean of precision and recall. The closer to 1, the better the model, with F1 being 2 \* (Precision \* Recall) / (Precision + Recall)

Support: the number of actual samples in the test set when evaluate the model.

It typically includes:

* Per-class metrics
* Averages: **macro** (simple mean), **weighted** (weighted by support), **micro** (overall TP/FN/FP aggregated)

This report helps diagnose individual class performance, especially in multi-class settings [geeksforgeeks.org+2linkedin.com+2en.wikipedia.org+2](https://www.linkedin.com/advice/3/what-difference-between-confusion-matrix-classification-hsehf?utm_source=chatgpt.com)

# Confusion matrix

# Overall

Due to small size, class 6 performance metric is not as stable as the other classes.

# 40-60

A screenshot of a computer screen

AI-generated content may be incorrect.

A diagram of a heatmap confusion matrix

AI-generated content may be incorrect.

Overall, the model achieves an accuracy of **87%**, indicating that it correctly classifies the vast majority of instances.

A closer look at the per-class metrics reveals outstanding performance on **classes 1, 3, and 5**:

* **Class 1**: Precision = 0.98, Recall = 0.96, F1-Score = 0.97
* **Class 3**: Precision = 0.98, Recall = 0.98, F1-Score = 0.98
* **Class 5**: Precision = 1.00, Recall = 0.97, F1-Score = 0.98

These near-perfect scores—especially the 100% precision on class 5—suggest that the decision tree has successfully learned the key feature patterns that distinguish these classes from all others.

**Class 6** has the fewest test examples, yet the model attains high precision. However, recall is lower (0.75), meaning a notable fraction of true class 6 samples were misclassified as other classes. Despite that, when the model predicts class 6, it’s almost always right.

**Classes 2 and 4**, although better represented in the dataset, show relatively poor performance — both precision (~0.68 for 2 and ~0.62 for 4) and recall (0.65 and 0.70 respectively) are low. They are often misclassified, including frequent swaps between each other as observable in the confusion matrix.

Taken together, the pattern indicates the model particularly struggles to distinguish among **classes 2 and 4**, suggesting overlaps in feature representation or inadequate separability under the default tree settings.

60-40

80-20

90-10

Comparison between 4 models

# References

<https://www.statology.org/sklearn-classification-report/> (23/6/2025)

<https://www.evidentlyai.com/classification-metrics/accuracy-precision-recall> (23/6/2025)