# Classification report interpretation (the same for 4 all models)

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 evaluating the model.

For the overall information, it typically includes:

Accuracy: The rate of correct prediction the model has made.

Macro avg: The unweighted average per label for each metric (Precision, Recall, F1-Score, Support)

Weighted avg: The weighted average per label for each metric (all metrics except for Support)

# Confusion matrix interpretation (the same for all models)

A confusion matrix is a *n-by-n* grid that shows the performance of the model by comparing its predicted labels with the true labels, with n being the number of classes. It presents the number of true positives, true negatives, false positives and false negatives the model has made for each class, with ‘*positive’* and ‘*negative’* referring to the target class.

* True positive: this is when the model predicts the instance to be of the target class and it is indeed.
* True negative: this is when the model predicts the instance to not belong to the target class and it is not.
* False positive: this is when the model predicts the instance to be of the target class although it is not.
* False negative: this is when the model predicts the instance to not belong to the target class even though it actually is.

# 40-60 Insights

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** revealnear-perfect scores (above 0.96 for all metrics) —especially the 100% precision on class 5— which suggests 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 one to be 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 model performs well in differentiating class 1, 3, 5 but struggles to distinguish when it comes to class 2, 4 and 6. This suggests overlaps in feature representation or inadequate separability under the default tree settings.

# 60-40 Insights

A screenshot of a computer screen

AI-generated content may be incorrect.A graph of a heatmap

AI-generated content may be incorrect.

Overall, this model achieves an impressive 0.95 accuracy, meaning it can efficiently differentiate between instances of the classes.

Detailed inspection at the class-wise performance shows remarkably high precision, recall rate as well as F1-Score on virtually all classes. This is concrete proof that the model has learned the relation between input features to such a great extent that it can reliably classify an instance. The only class whose statistics seem a bit low compared to others is class 4, which is regularly mistaken as class 2 as revealed by the confusion matrix.

Despite having the least samples, class 6 has a flawless recall rate, meaning this model has learned the key feature that helps tell apart class 6 from other classes and thus, it doesn’t mislabel samples of this class as other classes.

In general, the classification tree performs exceptionally well, with only a slight notable difficulty in discriminating between class 2, class 3 and class 4.

# 80-20 Insights

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 has a high chance of correctly label the sample, with 0.95 accuracy.

Examining the details of the report and the matrices, we can see that:

Class 1, 3, 5, 6 have an impression precision rate of 100%, which means the model has learned the patterns for classes 1, 3, 5 thoroughly and can predict precisely. In spite of this superb result, class 6 has a considerably lower recall rate, which means out of 4 test sample, there is one that gets misclassified as another class (class 2 as pointed out by the confusion matrix). This indicates the overlapping features between the two classes that the model failed to clearly discern.

In contrast to those classes’ extraordinary performance, class 2 and class 4 prediction precision experiences a sharp drop, suggesting that instances of other classes are falsely identified as these two classes. This goes on to show that the model struggles to tell apart some of the feature correlation due to their similarities and common points, which results in mistaking the samples and significantly afflict precision score.

In short, this classification model can work out the correlation between features for class 1, 3, 5, 6 very well but lack clear classification capability when it comes to the other two classes.

# 90-10 Insights

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.

With a relatively high accuracy of 0.89, this model also possesses great capability to distinguish between samples of the classes.

Upon further observation, we see that class 1 has an outstanding performance when scoring 1.0 on all three metrics, meaning the model has truly recognize the class’s ffeature pattern.

For class 3, 5 and 6, the precision score is flawless but the other fluctuate greatly, especially when class 6’s recall rate is only 0.5. This indicates that the model confuses instances of class 6 with other classes more.

On the contrary, class 2 and class 4’s statistics plummet for the precision metrics, while still attaining relatively good recall rate and f1-score. This means the model mistakes other classes as either 2 or 4 a lot and that it struggle but will have no trouble identifying a class 2 or class 4 instance if it is indeed those two.

Overall, the model performs relatively well, being able to differentiate instances of classes 3, 5 and 6. Although for class 2 and class 4, the model exhibit a considerably lower score on precision, the high recall rate indicates that if the model predicts one sample to belong to either these two classes, it has a high chance of being correct.

# Comparison between 4 models

As for class 6’s modest performance, apart from the poor pattern recognition among the models, it can be attributed to the fact that it has a significantly smaller population, meaning less data to train the model on, i.e. – less chance to examine and study the feature correlation.

The four models perform well, especially when it’s about class 1, 3, 5, suggesting the model’s capability to comprehend and discern the feature relation concerning those classes. However, the models suffer from the overlapping features of class 2, class 4 and thus show inconsistent performance in predicting those classes, notwithstanding the ratio for train-test sets.

Despite having more data to train the model with, the 90\_10 model experience a sharp drop in performance, which is a clear sign of over-fitting and calls for further tuning.

# References

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

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

<https://www.evidentlyai.com/classification-metrics/confusion-matrix> (24/6/2025)