# Classification Report Interpretation *(applies to all four models)*

A classification report provides a summary of key performance metrics for each class:

* Precision: The proportion of correct positive predictions out of all positive predictions made by the model. It reflects how accurate the positive predictions are.
* Recall: The proportion of actual positive instances that were correctly predicted. It indicates how well the model captures true positives.
* F1-Score: The harmonic mean of precision and recall, calculated as 2 \* (Precision \* Recall) / (Precision + Recall)  
  A higher F1-score (closer to 1) indicates better overall performance, especially in the presence of class imbalance.
* Support: The number of actual instances for each class in the test set.

In addition to per-class metrics, the report typically includes:

* Accuracy: The overall percentage of correctly classified instances across the entire test set.
* Macro Average: The unweighted mean of precision, recall, and F1-score across all classes. This treats each class equally, regardless of its support.
* Weighted Average: The average of each metric (excluding support), weighted by the number of true instances per class. This reflects the model's performance more accurately on imbalanced datasets.

# Confusion matrix interpretation (the same for all models)

A **confusion matrix** is an *n × n* table that visualizes the model’s performance by comparing predicted labels with true labels, where *n* is the number of classes.

Each cell in the matrix shows how many instances of a given true class were predicted as each class, allowing detailed insight into misclassifications.

The key components include:

**True Positive (TP)**: The model correctly predicts an instance as belonging to a specific class.

**True Negative (TN)**: The model correctly predicts that an instance does *not* belong to a specific class.

**False Positive (FP)**: The model incorrectly predicts an instance as belonging to a class it does not actually belong to.

**False Negative (FN)**: The model fails to predict an instance as belonging to its actual class.

The confusion matrix helps identify specific patterns of misclassification, highlighting which classes the model confuses most often, and offering deeper insight into its decision-making behavior.

# 40-60 Insights

The model achieves an overall accuracy of 80%, with strong results on certain classes but clear areas for improvement on others.

Classes 1, 3, and 4 show consistently high performance with class 3 standing out with perfect precision (1.00) and recall (0.91), indicating clear feature separation. Classes 1 and 4 also perform well, with F1-scores of 0.85 and 0.86, respectively, and only a handful of confusion with other classes.

On the other hand, Class 6 suffers from low recall rate (0.50) despite perfect precision, indicating many actual Class 6 instances are missed—mostly misclassified as Class 2 as observable from the confusion matrix. In addition, Class 5’s modest performance with 0.64 at F1-score indicates probable overlapping features, which results in misclassifications into Classes 1, 2, and 4.

Last but not least, Class 2 proves to be a troublemaker when being prone to both receiving and causing misclassifications. While it achieves high recall (0.89), its precision is low (0.56), meaning the model often predicts Class 2 incorrectly for other classes, especially Classes 1, 5, and 6 based on the confusion matrix.

In short, the model performs reliably on well-defined classes but struggles with minority and feature-overlapping classes, particularly Class 2 as a frequent source of confusion.

# 60-40 Insights

The model achieves an accuracy of 81% on the 60/40 split, showing reliable overall performance.

We have excellent performance coming from Class 1, 3 and 6. Class 1 performs exceptionally well with a precision of 0.98 and recall of 0.96, indicating the model almost always gets this class right. Class 3 also possesses reasonably well statistics with an F1-score of 0.82 and a precision score of 0.85, though the confusion matrix reveals a few misclassifications into Classes 2 and 4. Especially, Class 6 stands out with perfect precision and recall despite having only 8 instances.

However, we have more modest figures with the other classes. Class 2 has moderate performance, with decent recall (0.75) but lower precision (0.67), suggesting the model often predicts the instances to be Class 2 even when they are not . Class 4 and Class 5 continue to show more modest performance, with F1-scores of 0.67 and 0.65, respectively. Both classes experience scattered misclassifications among neighboring classes, including each other, indicating possible feature overlap.

Overall, the model remains solid on dominant or well-separated classes, but confusion among Classes 2, 4, and 5 remains a notable weakness.

# 80-20 Insights

With an accuracy of 92%, the model performs very well on the 80/20 split, showing strong generalization even with a smaller test set.

Class 1, which has the highest support, maintains high precision and recall (both 0.96), indicating consistent reliability. Surprisingly Class 6 also achieves perfect scores across all metrics, though with the least test and train sizes.

Most classes exhibit balanced precision and recall, namely Class 4 and Class 5, both scoring F1-scores of 0.90 and 0.95, respectively, with minimal confusion. Class 2 performs flawlessly in terms of recall (1.00), although precision is slightly lower at 0.75 due to some misclassified instances from Class 3.

Class 3 has a perfect precision of 1.00 but a slightly lower recall of 0.80, as three of its instances are misclassified as Class 2, which reflects some feature similarity between the two.

In summary, the model demonstrates excellent predictive power in this split, particularly for clearly distinguishable or well-represented classes. The confusion between Classes 2 and 3 is the only notable weakness, but even this has limited effect on overall performance.

# 90-10 Insights

With an accuracy of 89%, the model proves to have no difficulty learning the patterns and the correlation.

Class 1 is predicted very well, achieving perfect precision and high recall (0.92), with just one instance mistaken as Class 5. Additionally, Class 3 and Class 6 both show perfect scores across all metrics, indicating clear feature representation and separability for those categories.

Class 2 performs consistently, with both precision and recall at 0.83, and only one instance confused with Class 5. Similarly, Class 4 also has a balanced performance with an F1-score of 0.80 with just one misclassification into Class 2.

However, Class 5 is the weakest among the group, with an F1-score of 0.73. It suffers a considerable drop in precision, likely due to drawing in misclassified instances from Classes 1 and 4.

In general, most classes are classified accurately, with minimal confusion thanks to the large test size. The primary confusion occurs between some classes suggesting mild feature overlap but overall performance remains robust across the board.

# 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)