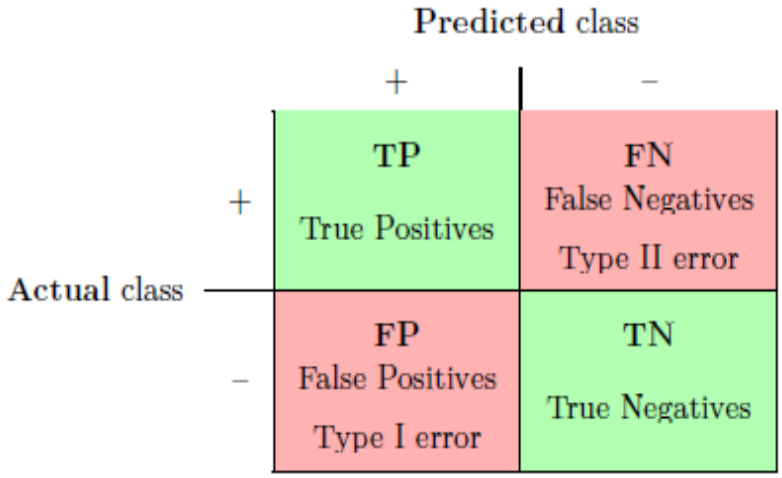
Classification:



Metric	Formula	Interpretation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Overall performance of model
Precision	$\frac{TP}{TP+FP}$	How accurate the positive predictions are
Recall Sensitivity	$\frac{TP}{TP+FN}$	Coverage of actual positive sample
Specificity	$\frac{TN}{TN+FP}$	Coverage of actual negative sample
F1 score	$\frac{2TP}{2TP+FP+FN}$	Hybrid metric useful for unbalanced classes

Random Forest Classification – Evaluation Summary

What does the confusion matrix show for the Random Forest model?

- **True Negatives (TN)**: 78
- **False Positives (FP)**: 7
- **False Negatives (FN)**: 6
- **True Positives (TP)**: 43

Classification Report - RF

Class	Precision	Recall	F1-Score	Support
0	0.93	0.92	0.92	85
1	0.86	0.88	0.87	49
Accuracy	-	-	**0.90**	134
Macro Avg	0.89	0.90	0.90	134
Weighted Avg	0.90	0.90	0.90	134

Random Forest – Technical Q&A Based on Confusion Matrix

1. What is a Confusion Matrix in classification?

Answer: It's a 2x2 table used to evaluate a classification model's performance. It shows True/False Positives and Negatives.

2. How many correct predictions were made by the Random Forest model?

Answer: TN + TP = 78 + 43 = 121, correct predictions out of 134.

3a. What is Precision for class 1? Or What is the percentage of the classification to sum of the correctly classified and wrongly classified in the test set?

Answer: Precision = TP / (TP + FP) = 43 / (43 + 7) = **0.86**

3b. What is Precision for class 0? Or What is the percentage of the classification to sum of the correctly classified and wrongly classified in the test set?

Answer: Precision = TP / (TP + FP) = 78 / (78 + 6) = **0.93**

4. What does Recall mean, and what is the value for class 1?

Or What is the percentage of correct classification of True Positives to the total input purchases of test set?

Answer: Recall = TP / (TP + FN) = 43 / (43 + 6) = **0.88**
Recall measures how well the model identifies actual positives.

5. What is F1-Score and its use?

Answer: F1-score = 2 × (Precision × Recall) / (Precision + Recall). It's the harmonic mean of precision and recall, especially useful for imbalanced datasets.

6. What is the accuracy of this Random Forest model?

Answer: TP+TN/TN+TP+FN+FP=>78+43/78+7+6+43=>121/134

= 0.90 or **90%** accuracy.

7. Is the model better at identifying class 0 or class 1?

Answer: Slightly better at identifying class 0 due to higher precision (0.93 vs. 0.86).

8. What does a False Positive mean in this context?

Answer: The model incorrectly predicted a user **would purchase** when they actually didn’t (7 such cases).

1. What is the accuracy of the Random Forest model?

- **Answer:** The accuracy of the Random Forest model is **90%**, as shown in the classification report. This means the model correctly predicted 90% of the test instances.
- TP+TN/TN+TP+FN+FP

2. What does the precision of 0.93 for class 0 indicate?

- **Answer:** A precision of 0.93 for class 0 means that when the model predicts a user did not purchase (class 0), it is correct **93%** of the time. In other words, out of all predicted non-purchases, 93% were actual non-purchases.

3. What does the recall of 0.92 for class 0 represent?

- **Answer:** A recall of 0.92 for class 0 means the model correctly identified **92%** of the actual non-purchases in the test set. It missed 8% of the true non-purchases.

4. How is the F1-score calculated, and what does it represent?

- **Answer:** The F1-score is the harmonic mean of precision and recall. For class 0, it is calculated as:
  $F1=2\times Precision\times RecallPrecision+Recall=2\times 0.93\times 0.92$
  $0.93+0.92=0.92F1=2\times \frac{Precision+Recall}{Precision\times Recall}$
  $=2\times \frac{0.93+0.92}{0.93\times 0.92}=0.92.$
  It represents a balance between precision and recall, with 0.92 indicating strong performance for class 0.

5. What does the confusion matrix show for the Random Forest model?

- **Answer:** The confusion matrix is:
  - **True Negatives (TN):** 78 (correct non-purchases).
  - **False Positives (FP):** 7 (incorrectly predicted as purchases).
  - **False Negatives (FN):** 6 (incorrectly predicted as non-purchases).
  - **True Positives (TP):** 43 (correct purchases).

6. How would you interpret the macro avg and weighted avg in the report?

- **Answer:**
  - **Macro Avg:** Averages the metrics (precision/recall/F1) across both classes without considering class imbalance. Here, it is **0.89** for precision and **0.90** for recall.
  - **Weighted Avg:** Averages metrics by weighting them based on class support (number of instances). Here, it is **0.90** for precision and recall, reflecting the model's performance accounting for class distribution or PROPORTION RATE.

Decision Tree Classification – Evaluation Summary

- **True Negatives (TN)**: 78
- **False Positives (FP)**: 7
- **False Negatives (FN)**: 6
- **True Positives (TP)**: 43

Classification Report - DT

Class	Precision	Recall	F1-Score	Support
0	0.90	0.89	0.90	85
1	0.82	0.84	0.83	49
Accuracy	-	-	**0.87**	134
Macro Avg	0.86	0.87	0.86	134
Weighted Avg	0.87	0.87	0.87	134

Decision Tree – Technical Q&A Based on Confusion Matrix

1. How many correct predictions were made by the Decision Tree model?

Answer: TN+TP=78+43=121,

TN + TP = 78 + 43 = 121

correct predictions out of 134.

2a. What is Precision for class 1? Or What is the percentage of the classification to sum of the correctly classified and wrongly classified in the test set?

Answer: Precision = TP / (TP + FP) = 43 / (43 + 7) = **0.82**

2b. What is Precision for class 0? Or What is the percentage of the classification to sum of the correctly classified and wrongly classified in the test set?

Answer: Precision = TN / (TN + FN) = 78 / (78 + 7) = **0.90**

3a. What does Recall mean, and what is the value for class 1?

Answer: Recall = TP / (TP + FN) = 43 / (43 + 6) = **0.84**
Recall measures how well the model identifies actual positives.

4. What is F1-Score and its use?

Answer: F1-score = 2 × (Precision × Recall) / (Precision + Recall). It's the harmonic mean of precision and recall, especially useful to find out **overall performance** each imbalanced datasets

5. What is the accuracy of this DT model?

Answer: 0.87

6. Is the model better at identifying class 0 or class 1?

Answer: Slightly better at identifying class 0 due to higher precision (0.90 vs. 0.82).

7. What does a False Positive mean in this context?

Answer: The model incorrectly predicted a user **would purchase** when they actually didn't (7 such cases).

8. What is the accuracy of the Decision Tree model?

- **Answer:** The accuracy of the Decision Tree model is **87%**, as shown in the classification report. This is slightly lower than the Random Forest model.

9. What does the precision of 0.90 for class 0 indicate?

- **Answer:** A precision of 0.90 for class 0 means that when the model predicts a user did not purchase (class 0), it is correct **90%** of the time. This is slightly lower than the Random Forest's precision for the same class.