# Classification Evaluation Metrics Explained (with Real-life Example)

## 🎯 Purpose of These Metrics

These metrics help evaluate classification models like predicting whether a customer will default on a loan or not. They tell us how well our model is performing.

## 🧠 Real-Life Example: Loan Default Prediction

Suppose you're a bank predicting if a customer will default on a loan:

- Default = Positive Class

- Not Default = Negative Class

## 📊 1. Confusion Matrix

A table that summarizes the model's performance:

|  |  |  |
| --- | --- | --- |
|  | Predicted Default | Predicted Not Default |
| Actually Default | True Positive (TP) | False Negative (FN) |
| Actually Not Default | False Positive (FP) | True Negative (TN) |

Example:

- TP = 40 (Correctly predicted default)

- FP = 10 (Wrongly predicted default)

- FN = 5 (Missed actual defaulters)

- TN = 45 (Correctly predicted non-default)

## 📌 2. Precision

Precision = Of all predicted defaults, how many were actually default?

Precision = TP / (TP + FP) = 40 / (40 + 10) = 0.80

✅ Meaning: 80% of predicted defaulters were truly defaulters. High Precision = Few false positives (low false alarms)

## 📌 3. Recall (Sensitivity / True Positive Rate)

Recall = Of all actual defaults, how many did the model detect?

Recall = TP / (TP + FN) = 40 / (40 + 5) = 0.89

✅ Meaning: Model caught 89% of true defaulters. High Recall = Few false negatives (didn't miss defaulters)

## 📌 4. F1-Score

F1-score balances precision and recall.

F1-score = 2 \* (Precision \* Recall) / (Precision + Recall) = 2 \* (0.80 \* 0.89) / (0.80 + 0.89) ≈ 0.84

✅ Meaning: Balanced score is 84%. Use this when both FP and FN matter.

## 📌 5. Accuracy

Accuracy = Overall correct predictions out of all predictions

Accuracy = (TP + TN) / (TP + TN + FP + FN) = (40 + 45) / 100 = 0.85

✅ Meaning: 85% of predictions were correct. BUT: Can be misleading in imbalanced datasets.

## 📌 6. AUC-ROC (Area Under ROC Curve)

AUC-ROC shows the model's ability to separate the classes.

- AUC = 1 → Perfect

- AUC = 0.5 → Random

- AUC = 0 → Worst

Example: AUC = 0.92 → 92% chance of ranking a defaulter higher than a non-defaulter.

## 🧪 Summary Table

Metric | Formula | What it Tells  
-------------|----------------------------------------|-----------------------------  
Accuracy | (TP + TN) / (TP + TN + FP + FN) | Overall correctness  
Precision | TP / (TP + FP) | Correctness of positive predictions  
Recall | TP / (TP + FN) | Coverage of actual positives  
F1-Score | 2 \* (Precision \* Recall) / (P + R) | Balance of precision & recall  
Confusion Matrix | TP, FP, FN, TN counts | Raw outcome counts  
AUC-ROC | Area under ROC Curve | Model's class separation ability

## ✅ Which Metric to Use When?

Scenario | Focus On  
----------------------------------|------------------  
Avoid false alarms (loan rejection) | Precision  
Catch all real defaulters | Recall  
Want balance | F1-score  
Class imbalance (e.g. few defaults) | AUC-ROC