Session 3

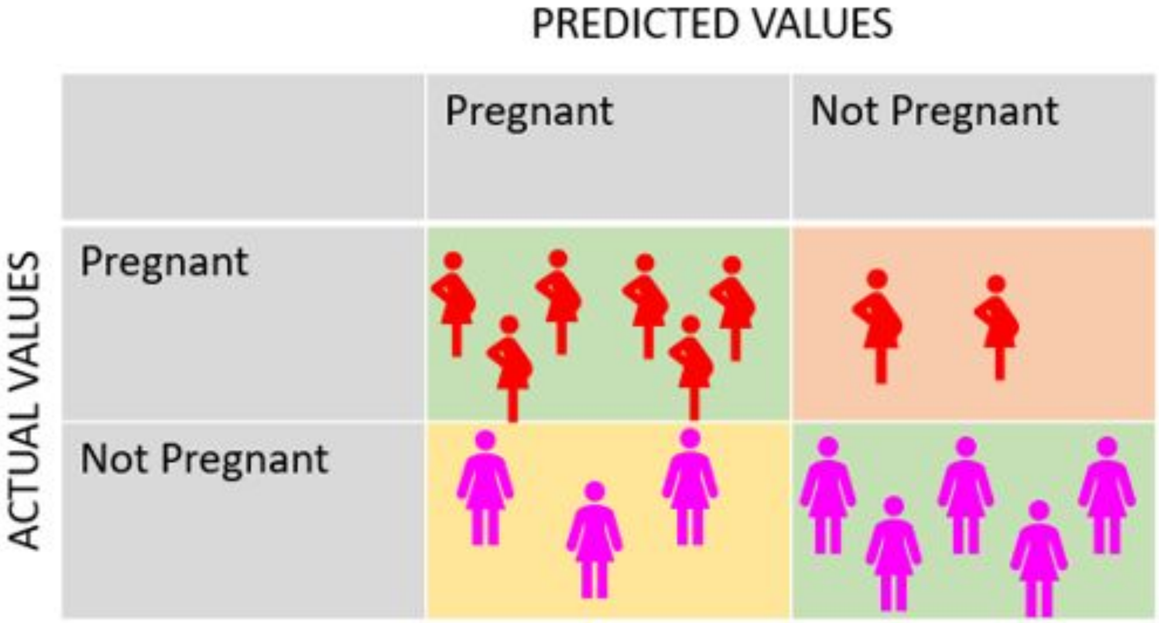
Y	Y^
0	0
1	1
1	0
0	1

Classification Metrics

What is Confusion Matrix?

		Predicted Class	
		Positive	**Negative**
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)

What is Confusion Matrix?

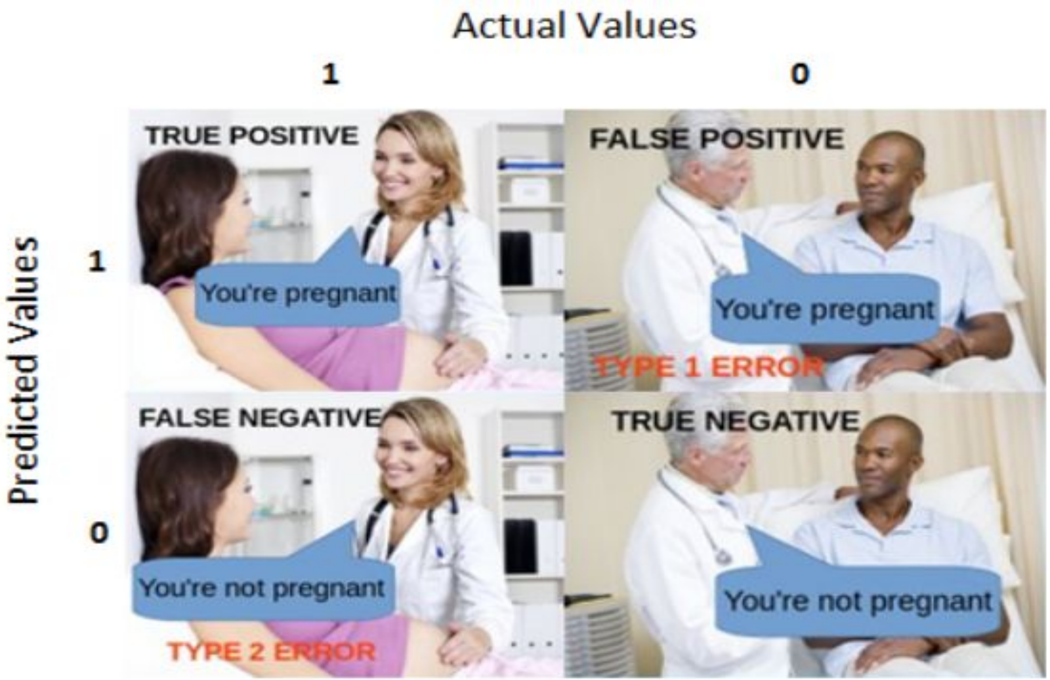


What is Confusion Matrix?

		Predicted Class	
		Positive	**Negative**
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)

- **True Positive (TP) —** When the model says that the patient has cancer and the patient actually has it
- **False Positive (FP) —** When the model says that the patient has cancer but the patient doesn't have it
- **True Negative (TN)** — When the model says that the patient does not have cancer and the patient actually doesn't have it
- **False Negative (FN)** — When the model says that the patient doesn't have cancer but the patient actually has it. *We don't want this, do we?*

What is Confusion Matrix?



Exercise

- We select 100 people which includes pregnant women, not pregnant women and men with fat belly. Let us assume out of this 100 people 40 are pregnant and the remaining 60 people include not pregnant women and men with fat belly. We now use a machine learning algorithm to predict the outcome.

- Out of 40 pregnant women 30 pregnant women are classified correctly and the remaining 10 pregnant women are classified as not pregnant by the machine learning algorithm.
- On the other hand, out of 60 people in the not pregnant category, 55 are classified as not pregnant and the remaining 5 are classified as pregnant.

Performance evaluation Measures

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error Missed Alarm
	Negative	False Positive (FP) Type I Error False Alarm	True Negative (TN)

ما شفت انو في مشكلة،
ولكن كانت موجودة

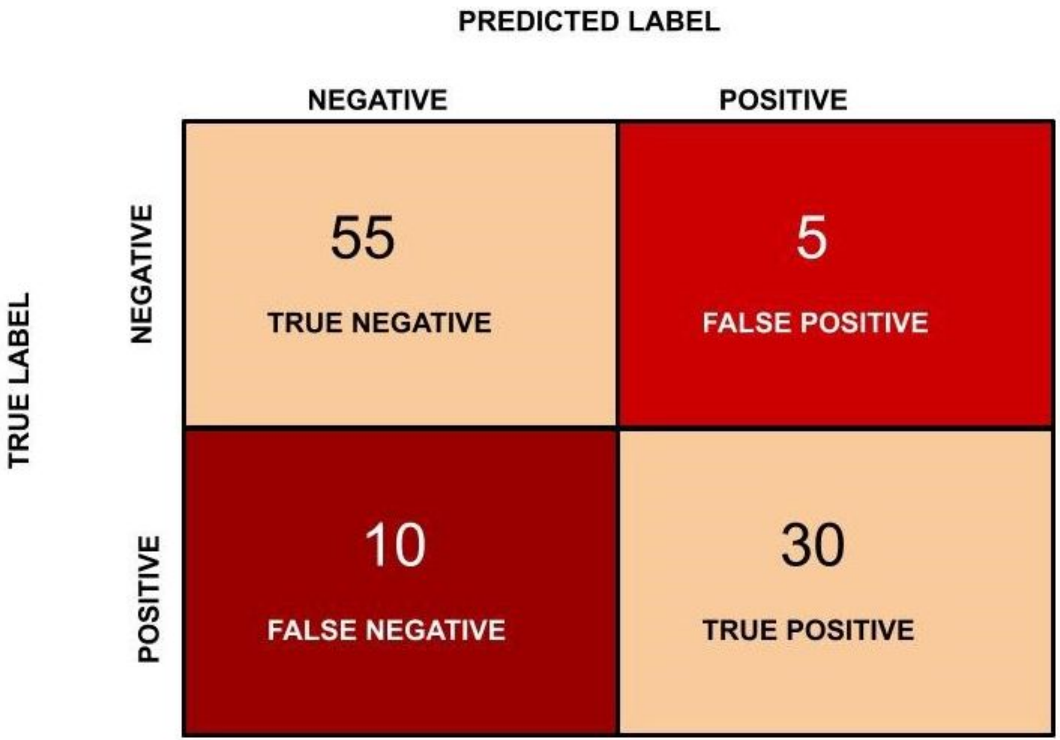
"سمعت صوت غريب
بالليل، بس طنّشت وقلت
يمكن الريح، وفعليًا كان في
لص! "

شفت انو في مشكلة،
وما كانت موجودة

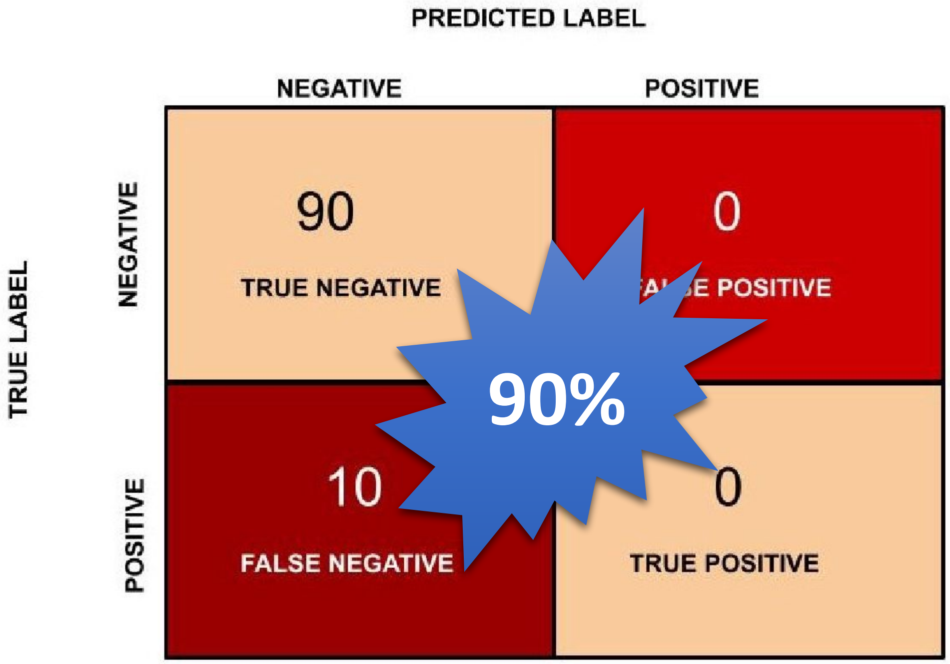
"زي إنك بلغت الشرطة إنه
في لص بالبيت، ولما إجوا
طلع كلب الجيران."

Accuracy = (TP+TN) / (TP+FP+FN+TN)

Example


PREDICTED LABEL
NEGATIVE
POSITIVE
TRUE LABEL
NEGATIVE
POSITIVE
55
TRUE NEGATIVE
5
FALSE POSITIVE
10
FALSE NEGATIVE
30
TRUE POSITIVE

Is accuracy the best measure?



Accuracy = (TP+TN) / (TP+FP+FN+TN)

Performance evaluation Measures

Confusion Matrix

		Predicted Class	
		Positive	**Negative**
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)

Accuracy = (TP+TN) / (TP+FP+FN+TN)

Precision = TP / (TP+FP)

Recall = TP / TP+FN

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

فعليا عندي 10 نساء (8 حامل، 2 مش حامل)

8 فعلا حامل، وتم التنبؤ انهم حامل >>> TP = 8

2 مش حامل ولكن الموودل تنبأ انهم حامل >>> FP = 2

Percision = 8 / (8+2) = 80%

من بين كل النساء الي **الموودل** قال عنهم حامل، **فعليا** 80% منهن حامل و 20% التنبؤ غلط

=====================================

فعليا عندي 10 نساء حامل

7 تم التنبؤ انهم حامل >>> TP = 7
3 المودل تنبأ انهم مش حامل >>> FN = 3

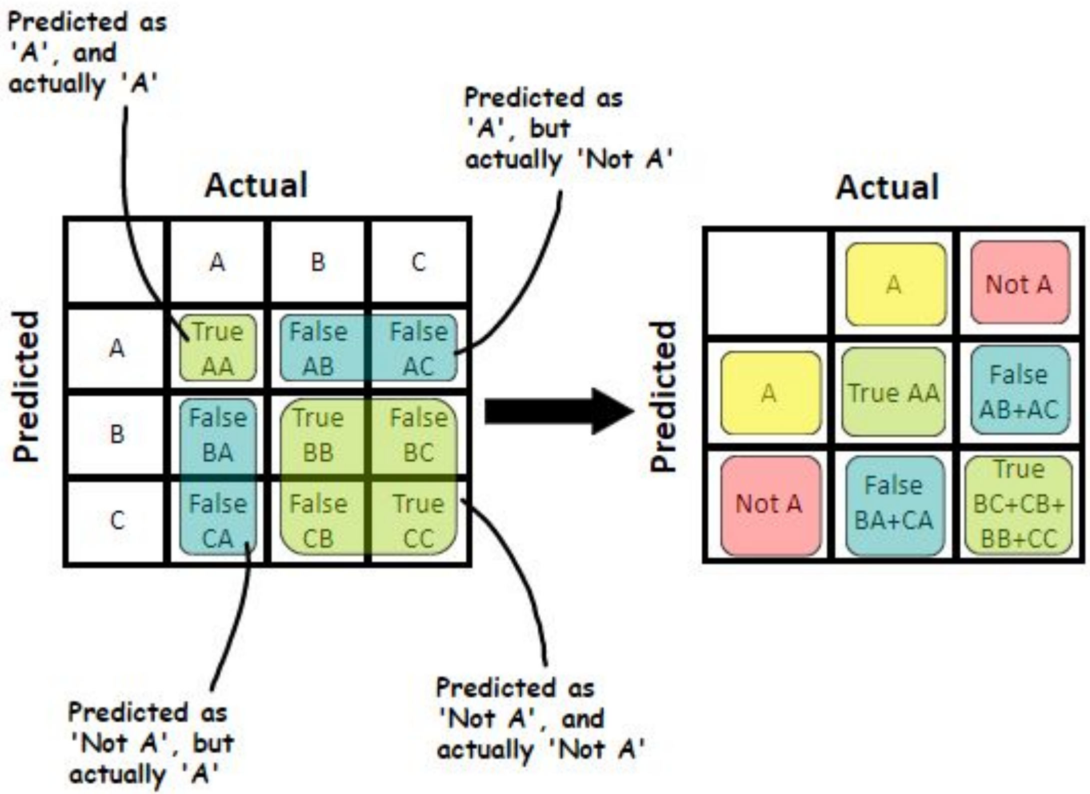
Recall = 7 / 7+3 = 70%

من بين كل الي **فعليا** حامل، المودل قدر يكتشف 70% منهم

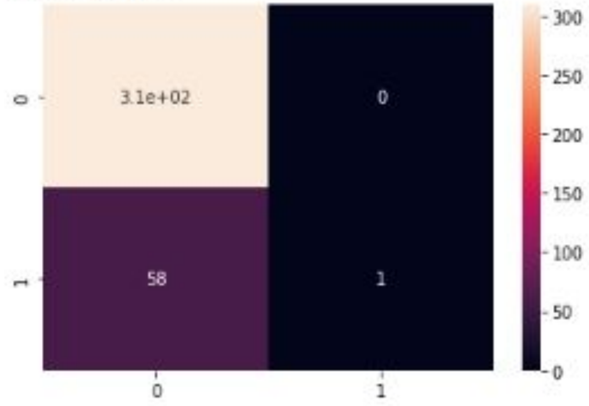
==

المقياس	التعريف	متى منيح نستخدمه؟	المشكلة	Type of Error
Accuracy =(TP+TN) / (TP + TN + FN + FP)	نسبة التوقعات الصحيحة (TP,TN) من كل العينات	لما تكون الداتا عندي Balanced	ممكن تخدعك مع الداتا اللي مش متوازنة (imbalance)	بيحسب الكل ومفيش خطأ محدد
Precision = TP /(TP+FP)	من كل اللي توقعهم المودل صح، كام وحدة منهم صح فعلاً؟	**لما يكون مهم أقلل FP:** **لما يكون الFP مُكلف:** ما بدي يحط ايميل مهم بال spam بالغلط	هيتجاهل ال FN	Type 1 error (FP)
Recall (Sensitivity) = TP /(TP+FN)	من كل ال positive الحقيقين، كام وحدة اكتشف المودل؟	**لما يكون مهم اقلل FN** **او لما يكون ال FN مُكلف:** اهم اشي اني اكتشف المريض المصاب بالسرطان وما افوت اي حال	ممكن يرفع ال FP	Type 2 error
F1 Score	المتوسط بين ال **Recall** و **Precision**	لما تكون الداتا عندي unBalanced وبدي مقياس عادل بين الاثنين	-	بيركز على type1 ,type2 مع بعض

MultiClass Classification



<AxesSubplot:>




```python
accuracy = metrics.accuracy_score(y_test, y_pred)
print("Accuracy: %.2f%%" % (accuracy * 100.0))
```

```
Accuracy: 84.24%
```

```python
from sklearn.metrics import classification_report
print(classification_report(y_true=y_test, y_pred=y_pred))
```

	precision	recall	f1-score	support
0	0.84	1.00	0.91	309
1	1.00	0.02	0.03	59
accuracy			0.84	368
macro avg	0.92	0.51	0.47	368
weighted avg	0.87	0.84	0.77	368

$$\text{Macro Avg} = \frac{\text{Metric (Class 0)} + \text{Metric (Class 1)} + \ldots}{\text{Number of Classes}}$$

$$\text{Weighted Avg} = \frac{\text{Metric (Class 0)} \times \text{Support 0} + \text{Metric (Class 1)} \times \text{Support 1}}{\text{Total Support}}$$

Classifcation metrics

6. ROC Curve and AUC (Area Under the Curve)

- The **Receiver Operating Characteristic (ROC) curve** is a graphical representation that shows the performance of a binary classifier as the discrimination threshold is varied.

  - True Positive Rate (TPR) = Recall = $\frac{TP}{TP+FN}$
  - False Positive Rate (FPR) = $\frac{FP}{FP+TN}$

The **ROC curve** plots the TPR against the FPR at different threshold levels. The closer the curve is to the top-left corner, the better the model.

Classifcation metrics

6. ROC Curve and AUC (Area Under the Curve)

AUC (Area Under the Curve)

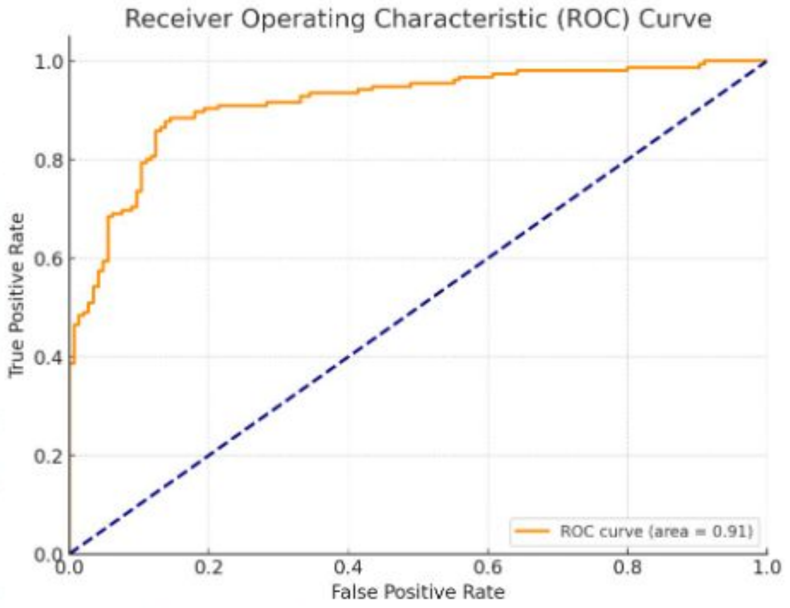
- The **AUC** value represents the degree of separability. Higher AUC means the model is better at distinguishing between positive and negative classes.

AUC ranges from 0 to 1

- 1 = Perfect classifier
- 0.5 = Random guess
- 0 = Completely wrong classification

Classifcation metrics

AUC (Area Under the Curve)



- ROC Curve is used to evaluate a classification model's performance.
- It plots True Positive Rate (TPR) against False Positive Rate (FPR) at different thresholds.
- The diagonal line represents random guessing.
- The orange line shows the model's performance with an AUC of 0.91, indicating strong performance.
- The closer the curve is to the top left corner, the better the model distinguishes between classes.

Classifcation metrics

6. ROC Curve and AUC (Area Under the Curve)

Key Insight

- **ROC-AUC** is ideal when you care about the ranking of predictions rather than the exact predicted class.
- It helps to understand how well the model distinguishes between classes across all thresholds.

Metric	Definition	When to Use?	When is it Useful?	When is it Not Useful?
Confusion Matrix	Table that shows the count of true positives, true negatives, false positives, and false negatives.	When you want detailed insight into the types of errors a model makes.	Useful for in-depth analysis of model behavior and error types.	Not useful for large datasets where analyzing the matrix becomes impractical.
Accuracy	Measures the ratio of correctly predicted instances to the total instances.	When classes are balanced and overall accuracy matters.	Useful for general tasks with balanced data.	Not useful when there's a class imbalance.

Metric	Definition	When to Use?	When is it Useful?	When is it Not Useful?
Precision	Measures the ratio of correctly predicted positive instances to all instances predicted as positive.	When false positives are costly or problematic.	Useful in tasks like medical diagnosis where false positives can cause unnecessary treatments.	Not useful when missing positives is more problematic than false positives.
Recall	Measures the ratio of correctly predicted positive instances to all actual positive instances.	When false negatives are costly or problematic.	Useful in tasks like security, where missing a positive case is critical.	Not useful when false positives are a bigger issue than false negatives.

Metric	Definition	When to Use?	When is it Useful?	When is it Not Useful?
F1 Score	Harmonic mean of Precision and Recall, balancing both.	When you need a balance between Precision and Recall.	Useful in cases where both false positives and false negatives need to be minimized.	Not useful if Precision or Recall is much more important than the other.
ROC/AUC	Measures the model's ability to distinguish between classes at various thresholds.	When you need to evaluate a probabilistic model's performance across thresholds.	Useful in cases with a need to understand model performance across thresholds, like fraud detection.	Not useful when there is a large class imbalance.