Evaluation metrics for classification

Abdul Haseeb

BS(AI)-III

Outline

- Why Model Evaluation?
- Confusion Matrix
- Accuracy
- Precision
- Recall
- F1-Score

Some Terminologies

- Features/Independent Variables/Input (X)
- Binary Classification/Multiclass Classification
- Dependent Variable/Target Variable/Output(Y)
- Training Data vs Testing Data

Why Model Evaluation?

It allows us to determine strength and weakness of a model

Evaluation on separate training and validation data, helps us to identify overfitting and underfitting

We can train different models on same data and compare which one performs best

Performance Metrics

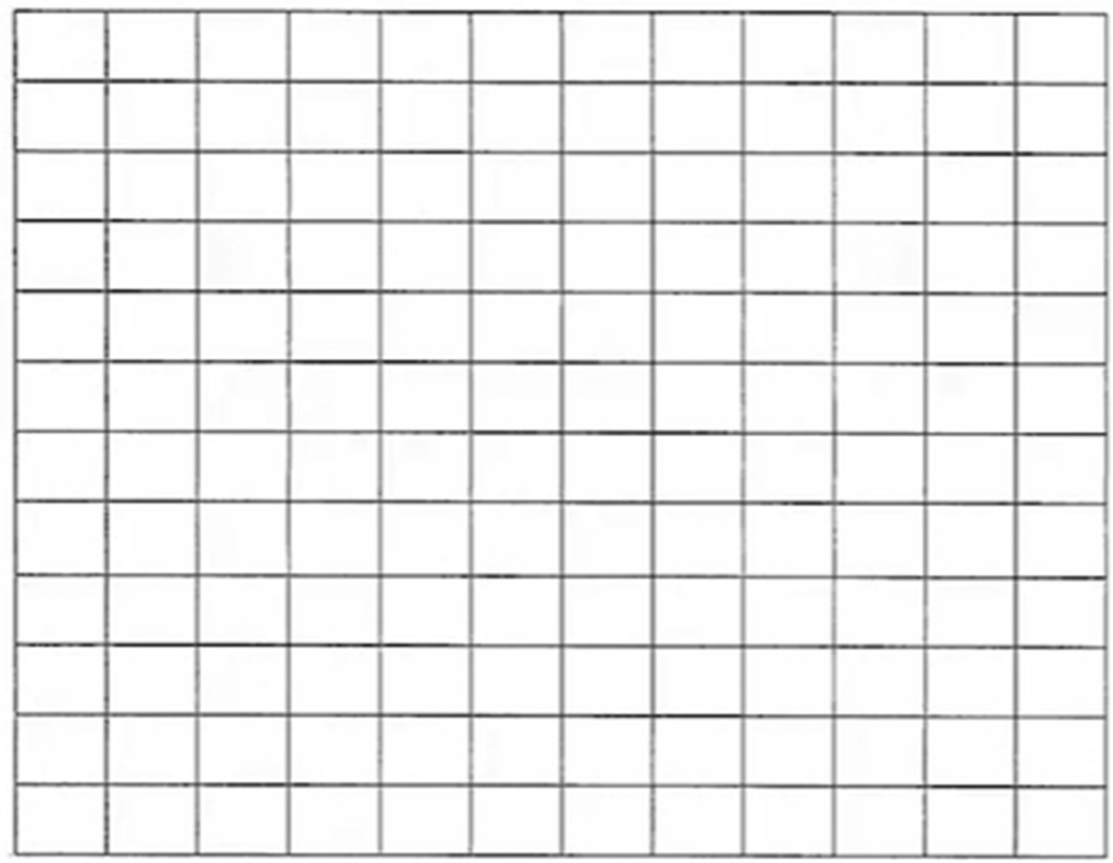
We can evaluate our Model based on different performance metrics like:

- Accuracy
- Precision
- Recall
- F1-Score

These performance metrics are calculated from confusion Matrix

Confusion Matrix

- N*N Matrix, where N represents the number of classes in Dependent variable

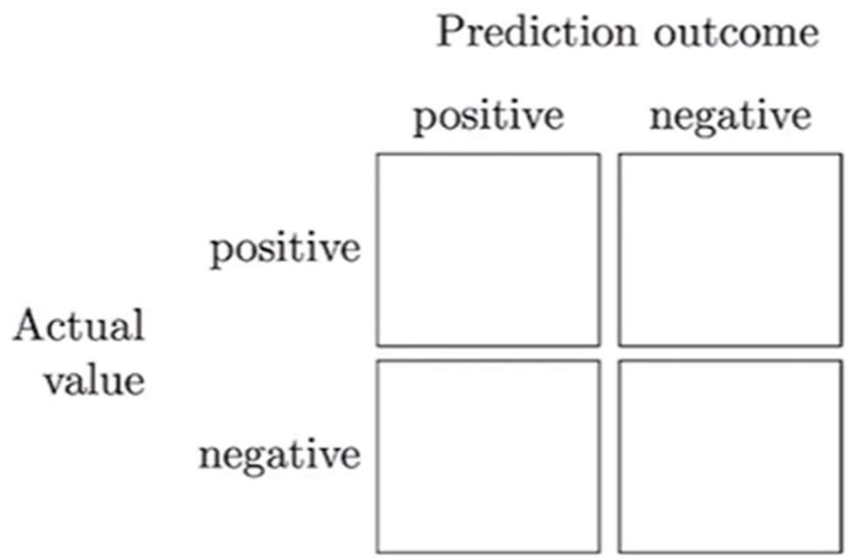


Titanic Dataset (Survived/Not Survived)

PassengerId	Survived	Pclass
1	0	3
2	1	1
3	1	3
4	1	1
5	0	3

Confusion Matrix

- Hence the output class has only two possible outcomes: 1 or 0, so we will have a 2*2 confusion matrix



Confusion Matrix

		Prediction outcome			
		positive	negative		
Actual value	positive	TP	FN	$TP + FN$	Total Actual positive
	negative	FP	TN	$FP + TN$	Total Actual negative

Confusion Matrix

- True Positive: Actual values were positive in dataset, and model also predicted those values as positive
  - Exp: In a medical test for a disease, a true positive would be a patient who has the disease and tests positive.

- True Negative: Actual values were negative in the dataset, and model also predicted those values as negative
  - Exp: A true negative would be a healthy patient who tests negative for the disease.

Confusion Matrix

- False Positive: Actual values were negative in dataset, and model predicted those values as positive
  - Exp: In a medical test for a disease, a false positive would be a healthy patient who tests positive for the disease.

- False Negative: Actual values were positive in the dataset, and model predicted those values as negative
  - Exp: A false negative would be a patient who has the disease but tests negative.

Titanic Dataset

ID	Actual Survived?	Predicted Survived?	Notation
ID1	1	0	
ID2	1	1	
ID3	1	0	
ID4	0	0	
ID5	1	1	
ID6	1	1	
ID7	0	1	
ID8	0	0	

Titanic Dataset

ID	Actual Survived?	Predicted Survived?	Notation
ID1	1	0	FN
ID2	1	1	TP
ID3	1	0	FN
ID4	0	0	TN
ID5	1	1	TP
ID6	1	1	TP
ID7	0	1	FP
ID8	0	0	TN

Build a Confusion Matrix Now

Confusion Matrix

- TP=3
- FN=2
- FP=1
- TN=2

Accuracy

- Ratio of correct predicted values over the total predicted values

		POSITIVE	NEGATIVE
ACTUAL VALUES	POSITIVE	TP	FN
	NEGATIVE	FP	TN

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

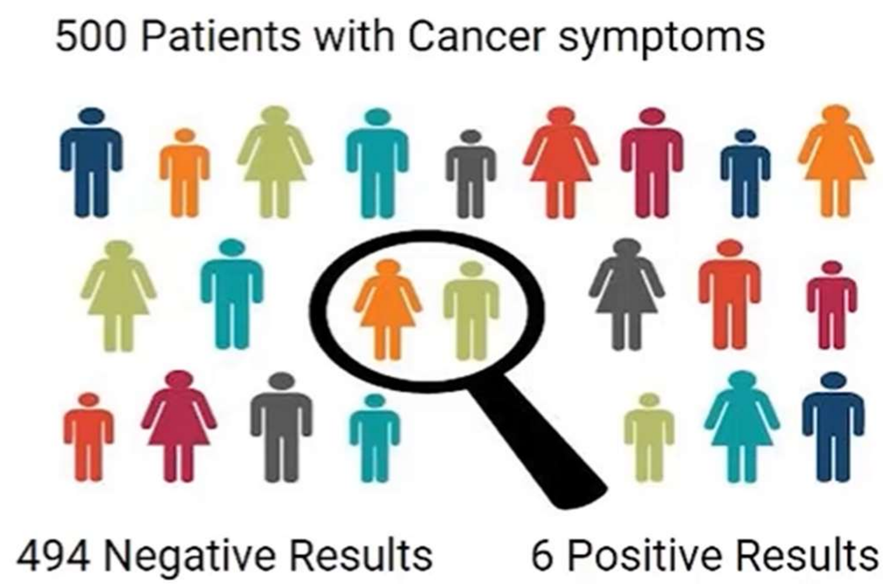
$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Accuracy

- We Train a Model to detect cancer
- This model will take inputs like: age,gender, previous medical history etc
- Confusion Matrix was like this suppose:

		Prediction outcome	
		positive	negative
Actual value	positive	4	2
	negative	8	486



Practice Problem

- Calculate accuracy from the confusion matrix on previous slide
- Do you get 98% accuracy?

Accuracy

- Now let’s build a dumb model to detect cancer:
  - **Negative Report for every patient or No Patient has cancer**

You can observe all values in negative column in the confusion matrix!

		Prediction outcome	
		positive	negative
Actual value	positive	0	6
	negative	0	494

Accuracy

- In Practical Scenario Accuracy doesn’t works when we have **class imbalance problems.**

$$TPR = \frac{TP}{TP+FN} = 0$$

$$FNR = \frac{FN}{TP+FN} = 1$$

$$TNR = \frac{TN}{FP+TN} = 1$$

$$FPR = \frac{FP}{FP+TN} = 0$$

98.8% Accuracy for dumb model!

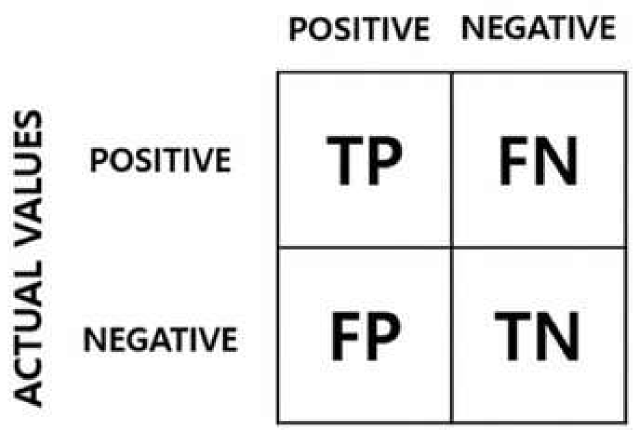
Look at FNR Rate, Means all those patients who had cancer were classified as not having the cancer

What to do for class imbalance problem

- Precision
- Recall

Precision

- Out of all positive predictions, how many are positive



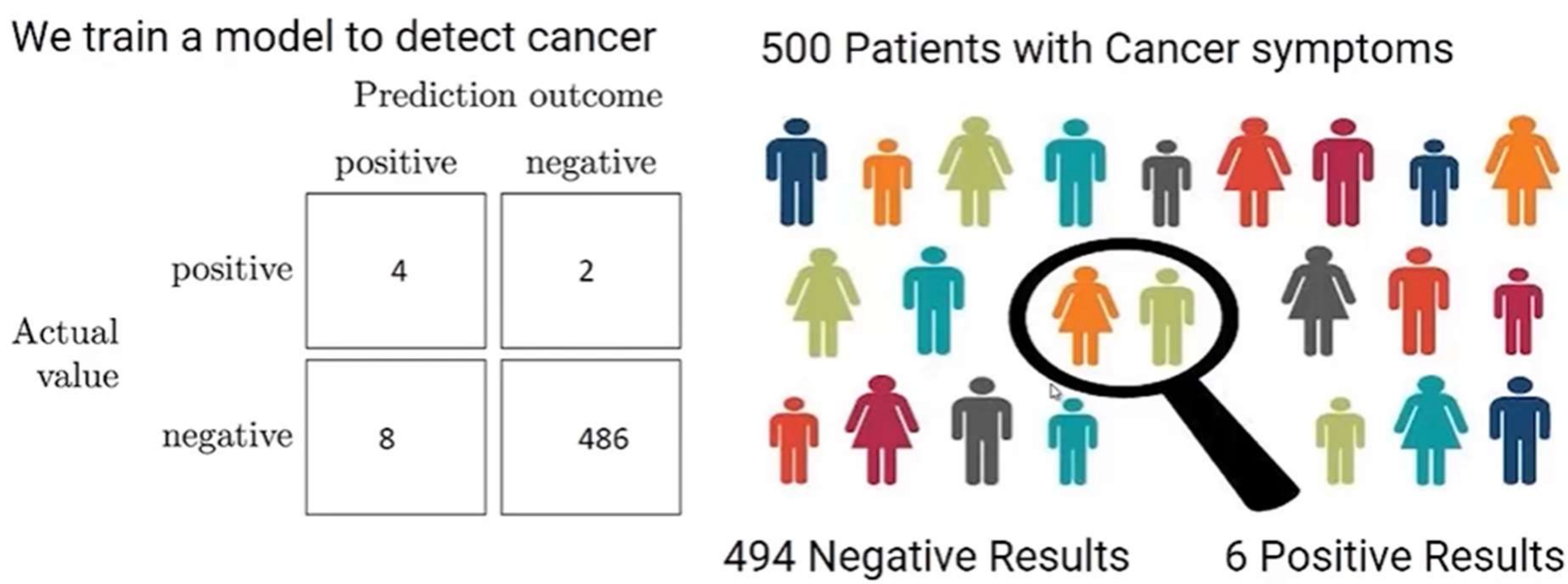
$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Precision

- Out of all positive predictions, how many are actually positive

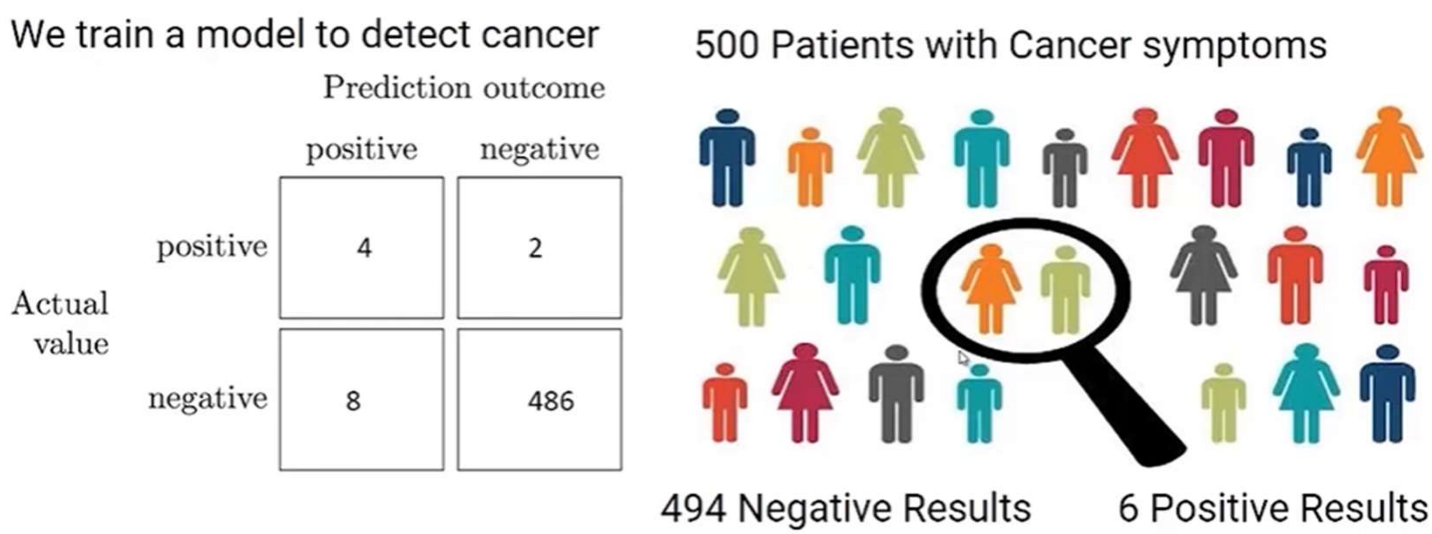


Precision

- Calculate Precision
- Is it 33.33%

Recall

- Out of all actual positives, how many are predicted positives
- Minimize False Negatives



Problem

- Calculate Recall for the confusion matrix

We train a model to detect cancer

		Prediction outcome: positive	Prediction outcome: negative
Actual value	positive	4	2
	negative	8	486

500 Patients with Cancer symptoms

494 Negative Results

6 Positive Results

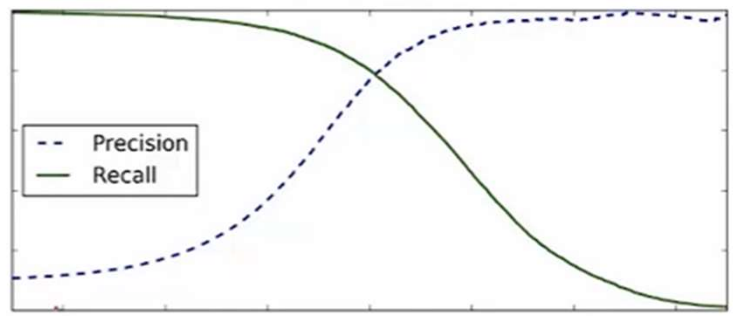
Solution

- 66.66%

- We can again infer that accuracy was misleading

Tradeoff between Precision and Recall

- High Precision, Low Recall
- High Recall, Low Precision
- Choice depends upon the use case
- Combined using F1 Score



$$F_1 = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}}$$