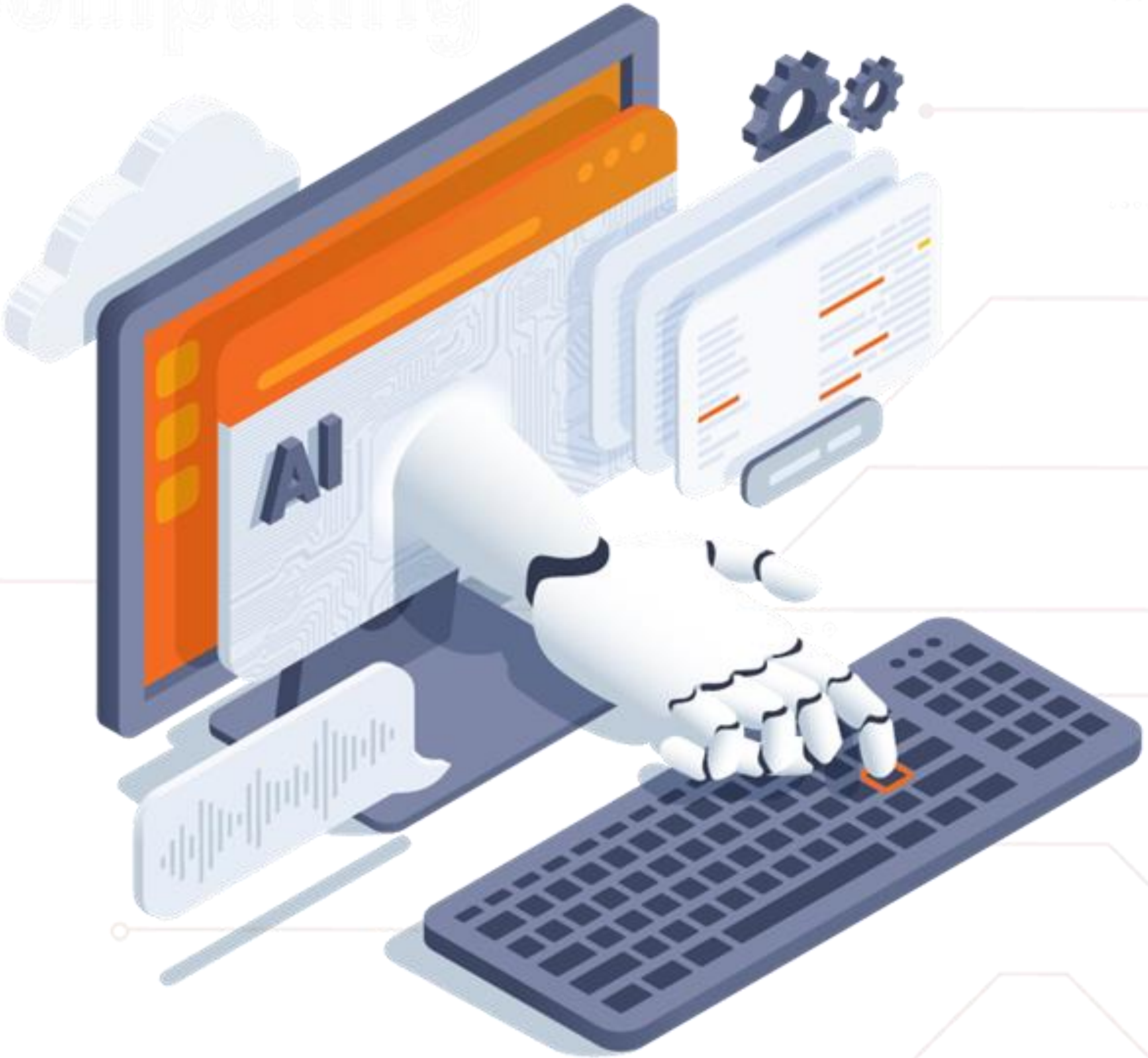
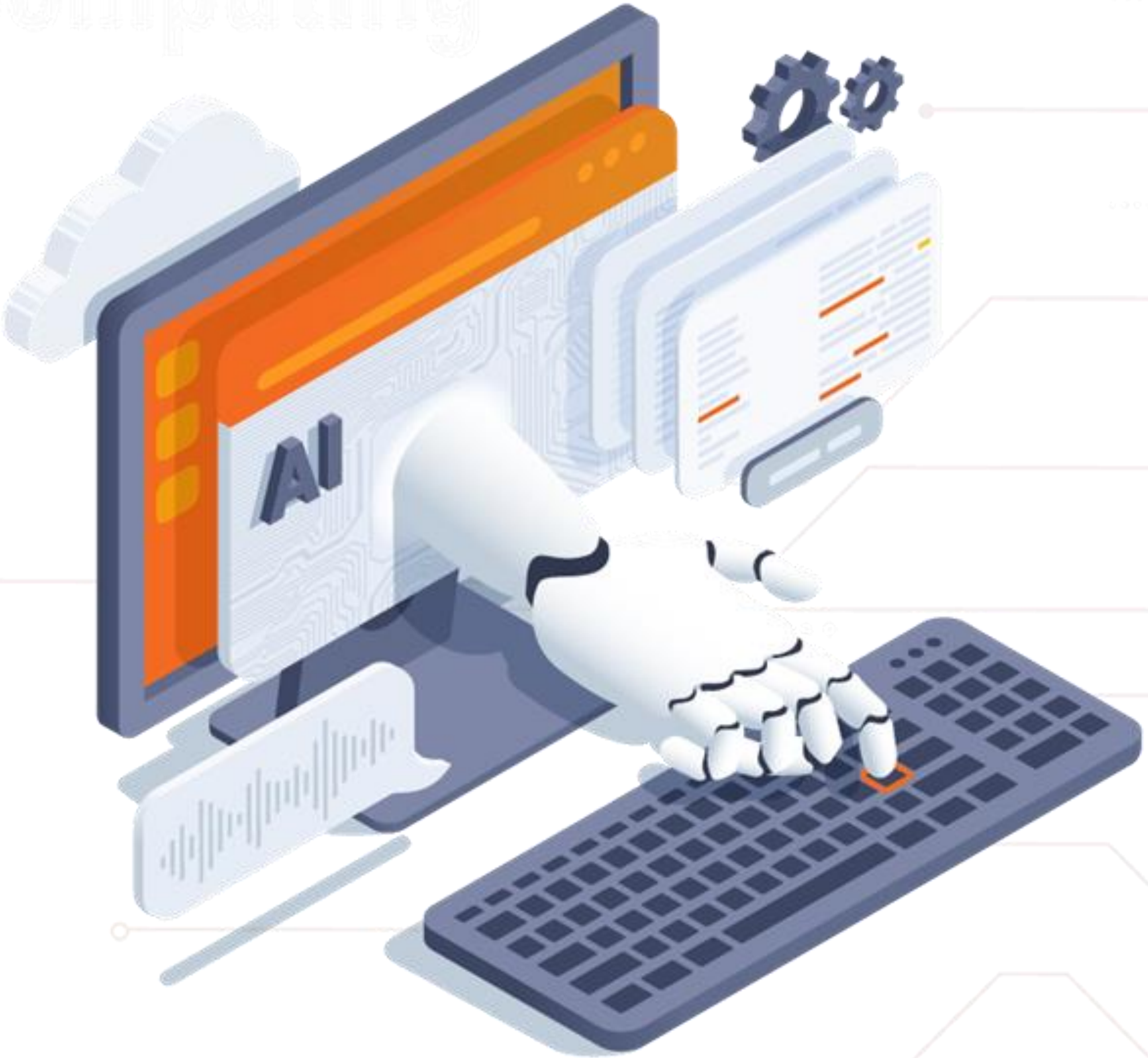


Caltech | Center for Technology & Management Education

Introduction to Artificial Intelligence

Powerd by simplilearn

Cloud Computing



Caltech | Center for Technology & Management Education

Performance Metrics

Learning Objectives

By the end of this lesson, you will be able to:

- Discuss the need for performance metrics
- List and analyze the key methods of performance metrics

Need for Performance Metrics

Need for Performance Metrics

- How do you rank machine learning algorithms?
- How can you pick one algorithm over the other?
- How do you measure and compare these algorithms?

- Performance metric is the answer to these questions.
- It helps measure and compare algorithms.

Performance Metrics

> "Numbers have an important story to tell. They rely on you to give them a voice."
>
> —Stephen Few

- Performance metrics help you in assessing the machine learning algorithms.
- Machine learning models are evaluated against the performance measures you choose.
- Performance metrics help in evaluating the efficiency and accuracy of the machine learning models.

Key Methods of Performance Metrics

Key Methods of Performance Metrics

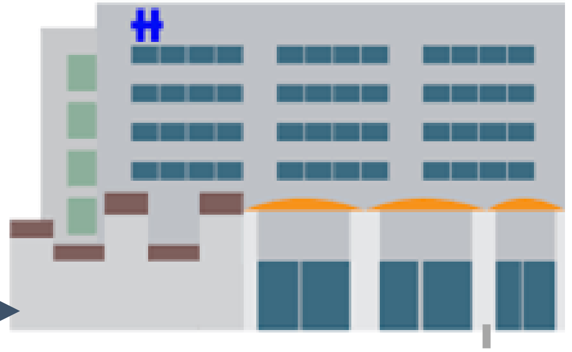
Confusion Matrix	Accuracy	Precision
Recall	Specificity	F1 Score

Meaning of Confusion Matrix

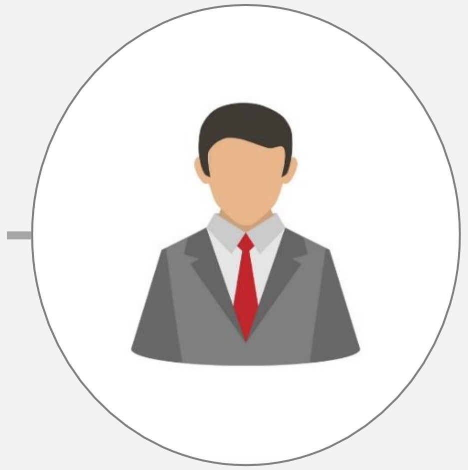
- The confusion matrix is one of the most intuitive and easiest metrics used for finding the correctness and accuracy of the model.
- It is used for classification problem where the output can be of two or more types of classes.

Confusion Metrics: Example

Cancer Prediction System



There are different approaches that can help the centre to predict cancer for a group of people.



okay

Let me first introduce you to one of the easiest matrices that can help you predict whether a person is has cancer: confusion matrix.

Confusion Metrics

THE CLASSIFICATION PROBLEM

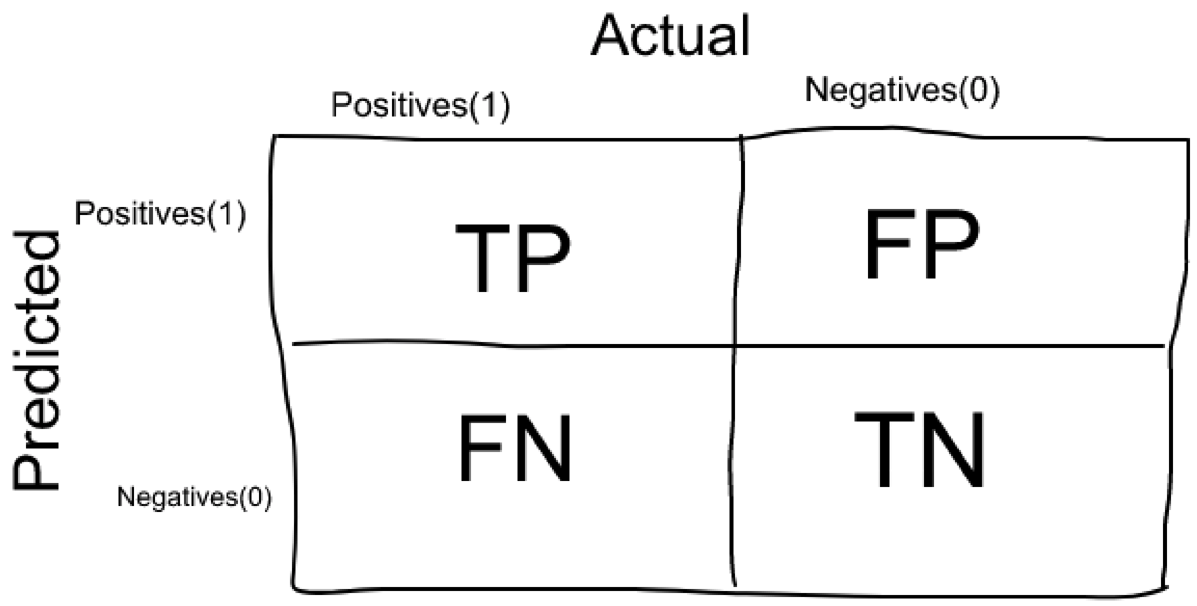
How to predict if a person has cancer?

Give a label/class to the target variables:

1	When a person is diagnosed with cancer
0	When a person does not have cancer

Confusion Metrics

THE CLASSIFICATION PROBLEM



- The confusion matrix is a 2D table with actual and predicted sections. The sets of classes are given in both dimensions.
- Actual classifications are added to columns and predicted ones are added to rows.
- The confusion matrix in itself is not a performance measure.
- However, almost all the performance matrices are based on confusion matrix and the numbers inside it.

Terms of Confusion Matrix

- TP — True Positive
- TN — True Negative
- FP — False Positive
- FN — False Negative

True Positive

TP True Positive

TN

FP

FN

- True positives are the cases where the actual class of the data point is 1 (true) and the predicted is also 1 (true).

The case where a person has cancer and the model classifies the case as cancer positive comes under true positive.

True Negative

TP

TN True Negative

FP

FN

True negatives are the cases when the actual class of the data point is 0 (false) and the predicted is also 0 (false).

The case where a person does not have cancer and the model classifies the case as cancer negative comes under true negative.

False Positive

TP

TN

FP — False Positive

FN

False positives are the cases when the actual class of the data point is 0 (false) and the predicted is 1 (true).

The case where a person does not have cancer and the model classifies the case as cancer positive comes under false positive.

False Negative

TP

TN

FP

FN False Negative

The case where person has cancer and the model classifies the case as cancer negative comes under false negative.

- False negatives are the cases when the actual class of the data point is 1 (true) and the predicted is 0 (false).
- It is false because the model has predicted incorrectly.
- It is negative because the class predicted was a negative one.

Minimize False Cases

- A model is best identified by its accuracy.
- In an analysis, missing a cancer patient is more dangerous than detecting a non-cancerous patient as cancer positive.
- No rules are defined to identify the false cases that need to be minimized.
- Two things that determine false negatives and positives are business needs and the context of the problem.

Minimize False Negatives: Example

Out of
100 people

Actual cancer patient
= 5

Bad Model

Predicts everyone as
non-cancerous

Accuracy = 95%

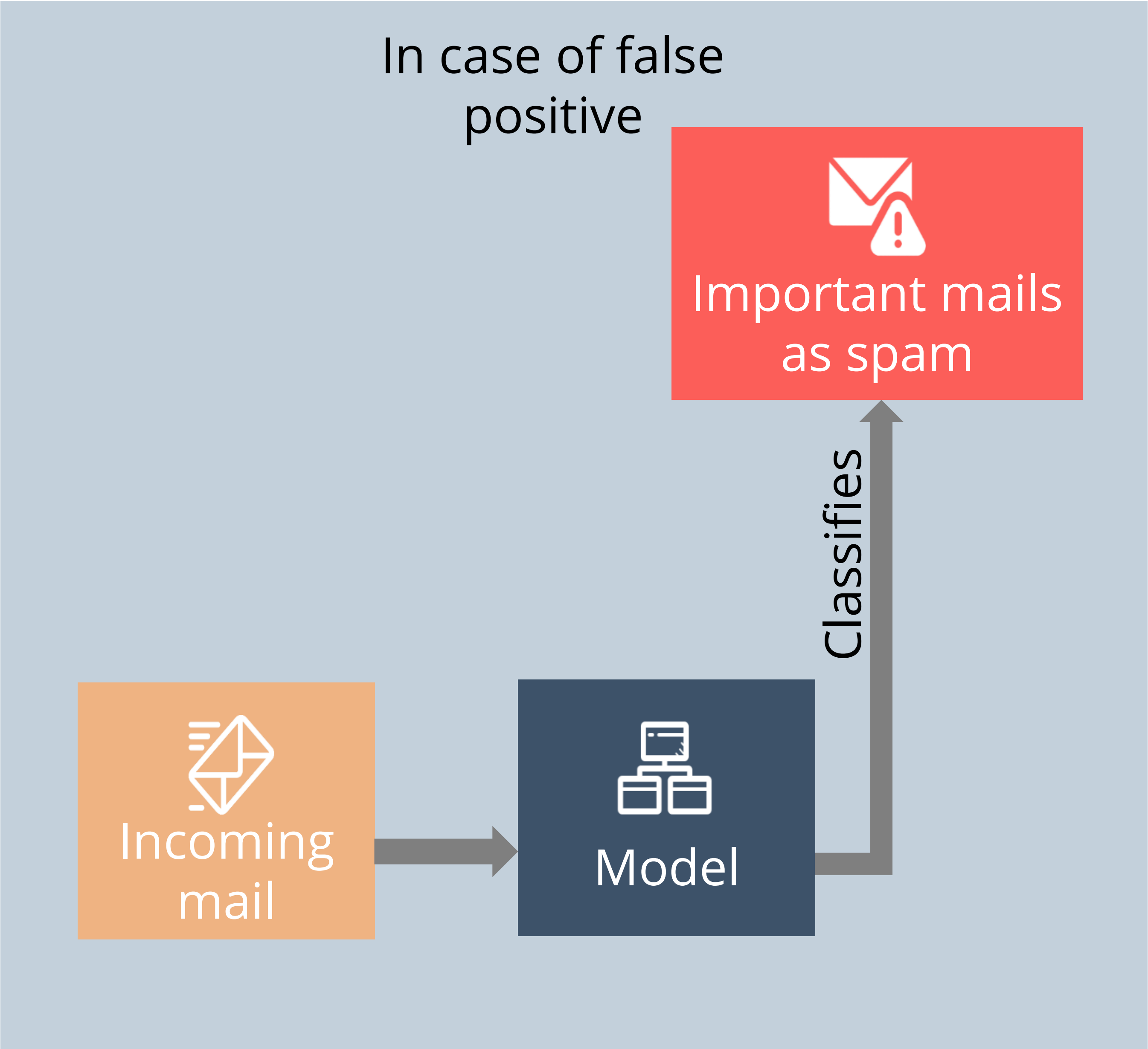
Minimize False Positives: Example

The model needs to classify an email as spam or ham (term used for genuine email).

Assign a label/class to the target variables:

1	Email is spam
0	Email is not spam

Minimize False Positives: Example (contd.)



- If a model classifies an important email as spam, it is a case of false positive.
- Business stands a chance to miss an important communication.
- An important email marked as spam is more business critical than diverting a spam email to the inbox.
- Therefore, in case of spam email classification, minimizing false positives is more important than false negatives.

Accuracy

In classification problems, accuracy is defined by the total number of correct predictions made out of all the predictions.

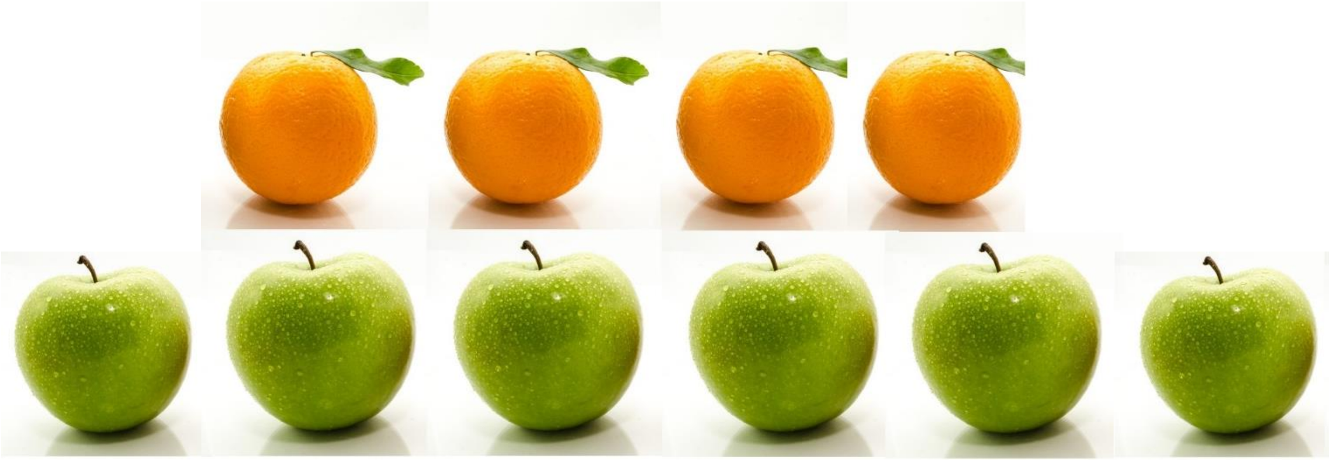
Accuracy: Calculation

		Actual	
		Positives (1)	Negatives (0)
Predicates	Positives (1)	TP	FP
	Negatives (0)	FN	TN

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

Accuracy: Example

- Accuracy is a good measure when the target variable classes in the data are nearly balanced.
- In this image, 60% classes are apple and 40% are oranges.
- With this type of data, the machine learning model will have approximately 97% accuracy in any new predictions.



Accuracy as a Measure

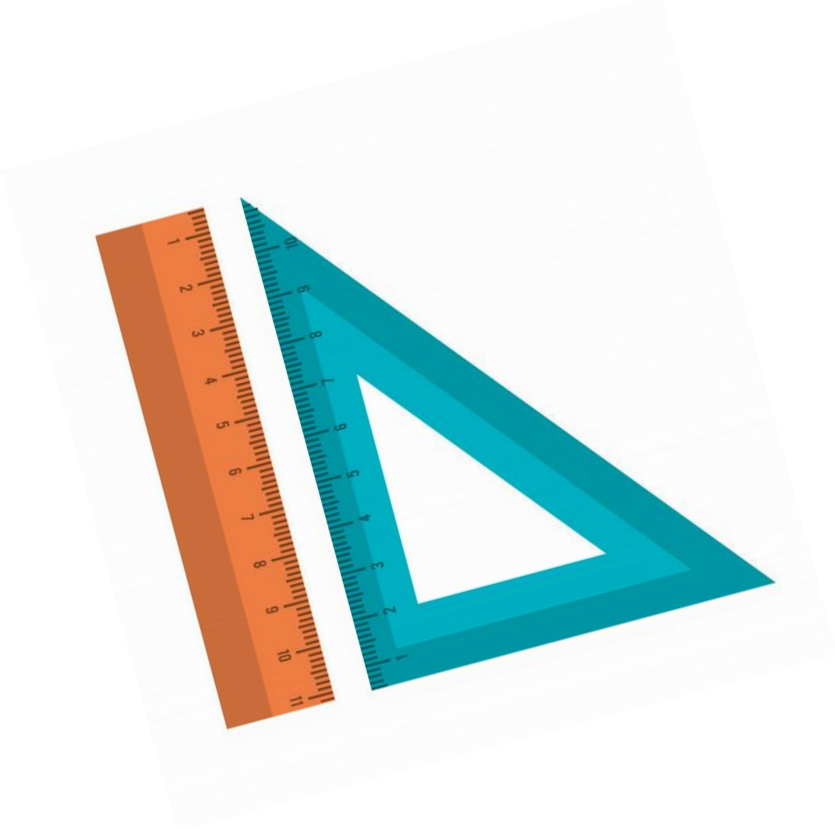
- Consider the previously discussed cancer detection example where only 5 out of 100 people have cancer.
- Suppose, it is a bad model and predicts every case as non-cancerous.
- In doing so, it classifies 95 non-cancerous patients correctly and 5 cancerous patients as non-cancerous.
- Now, even though the model did not accurately predict the cancer patients, the accuracy of the model is 95%.

Note:

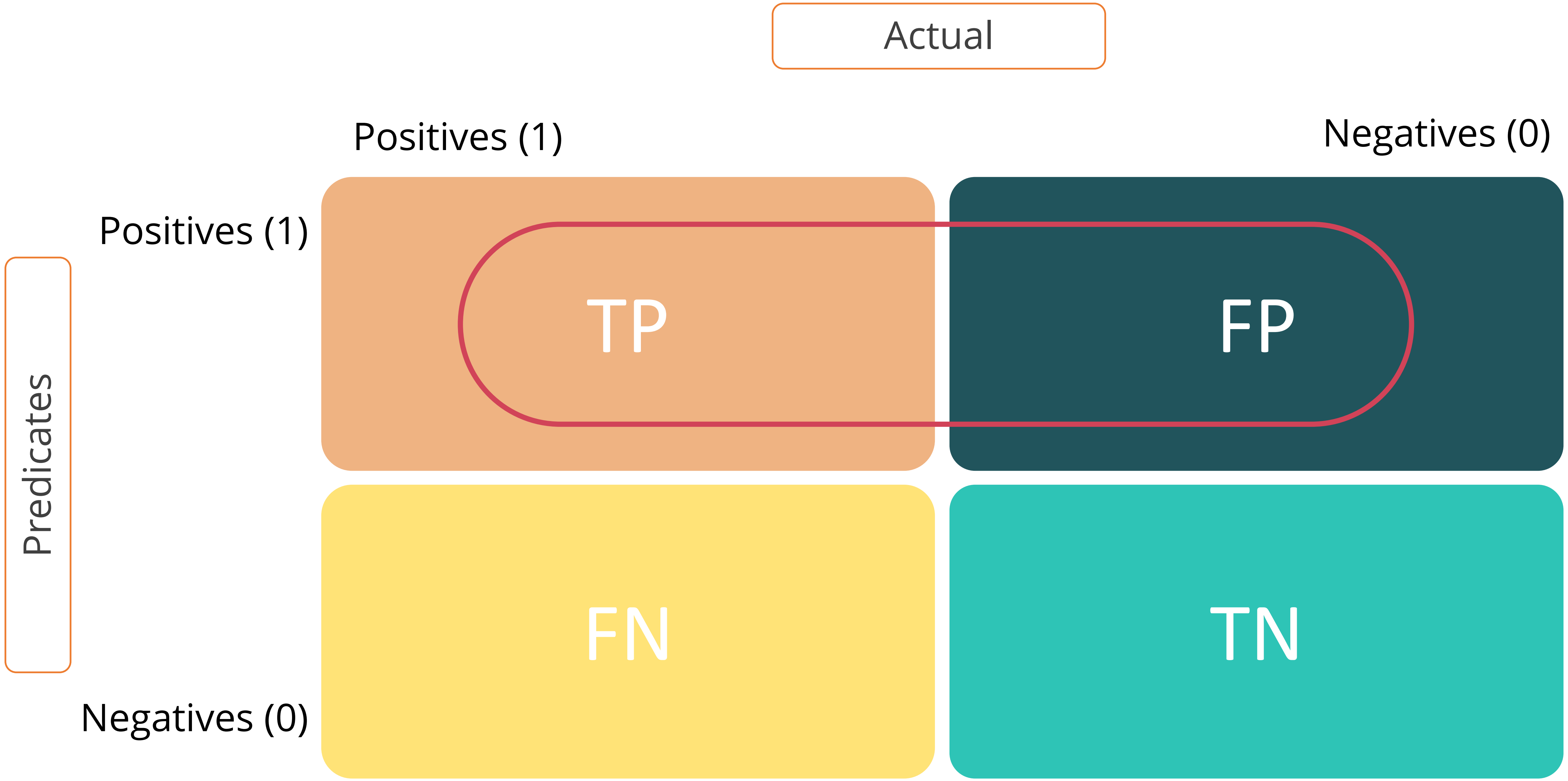
When the majority of the target variable classes in data belong to a single class, accuracy should not be used as a measure.

Precision

- Precision refers to the closeness of two or more measurements with each other.
- It aims at deriving the correct proportion of positive identifications.



Precision: Calculation



$$\text{Precision} = \frac{TP}{TP + FP}$$

Precision: Example

- Consider the previously discussed cancer detection example where only 5 out of 100 people have cancer.
- Precision will help you identify the proportion of cancer diagnosed patients who actually have cancer.
- The predicted positives are the people who are predicted to have cancer and include true positives and false positives.
- The actual positives are the people actually having cancer and include true positives.

Precision: Example (contd.)

Consider that the model is bad and predicts every case as cancer positive. In such scenario:

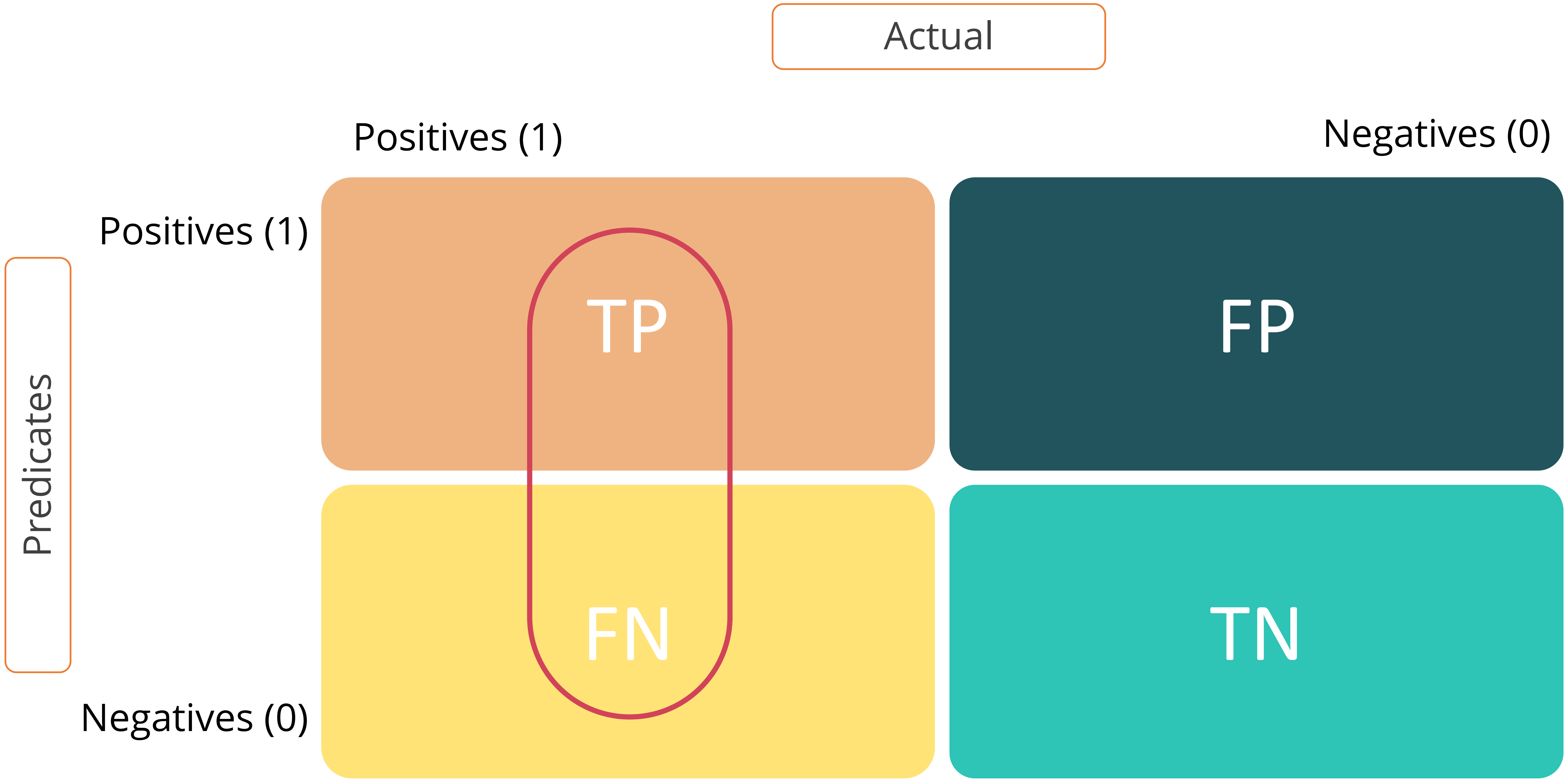
$$\text{Precision} = \frac{TP}{TP + FP}$$

- The model is predicting that everyone is cancerous.
- The denominator (true positives and false positives) is 100.
- The numerator is 5 and denotes a person having cancer. The model predicts the case as cancer positive.
- In this case, the precision of this model is 5%.

Recall or Sensitivity

Recall or sensitivity measures the proportion of actual positives that are correctly identified.

Recall or Sensitivity: Calculation



$$\text{Recall} = \frac{TP}{TP + FN}$$

Recall or Sensitivity: Example

- Consider the previously discussed cancer detection example where only 5 out of 100 people have cancer.
- Recall helps you identify the actual proportion of cancerous patients that are diagnosed by the algorithm.

$$\text{Recall} = \frac{TP}{TP + FN}$$

- The denominator (true positives and false negatives) is 5.
- The numerator is 5 and denotes that the person has cancer.
- The model predicting the case as cancer is also 5 (Since five cancer cases are predicted correctly).
- Recall of such model is 100% and the precision is 5%.

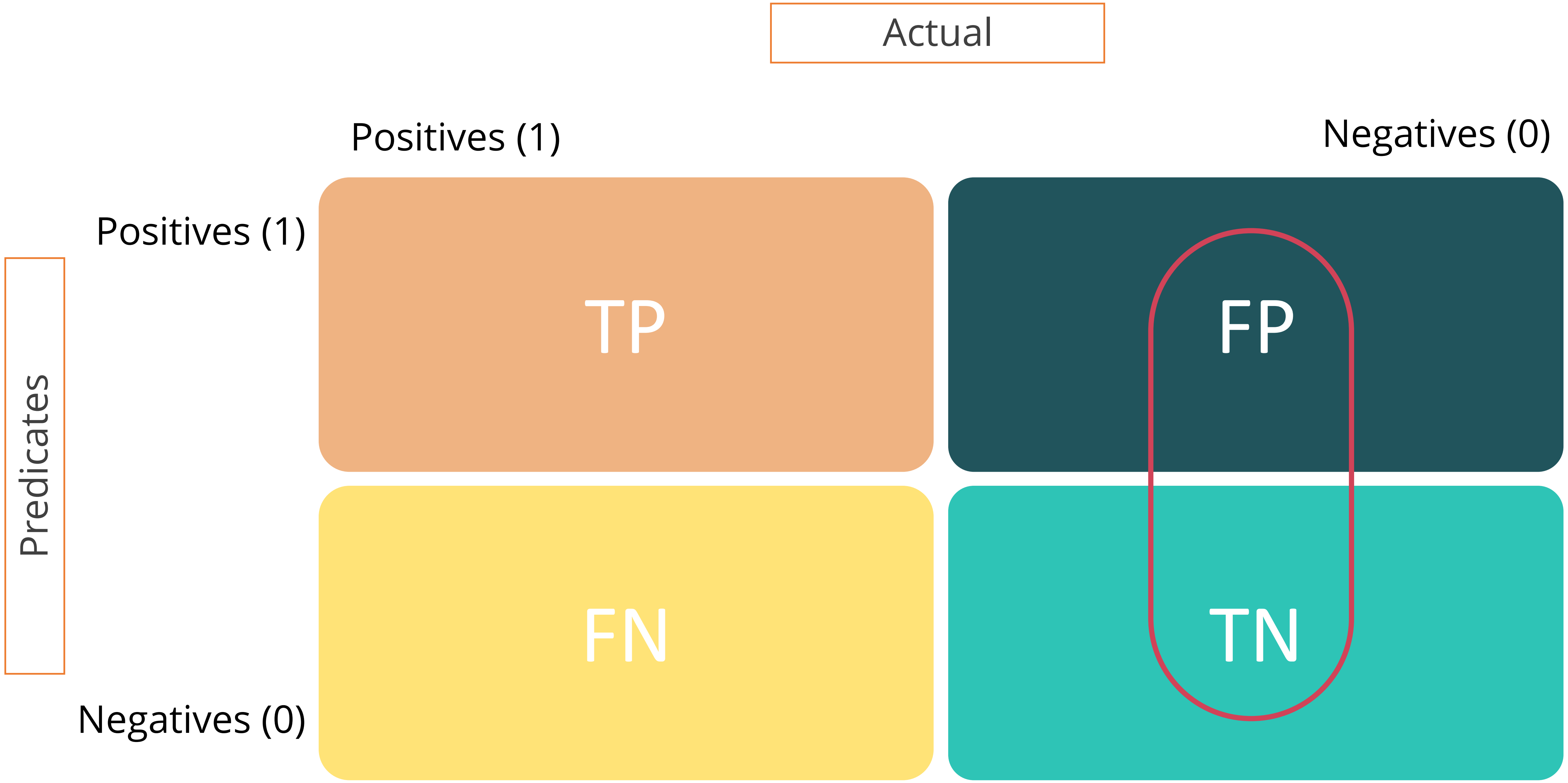
Recall as a Measure

- Precision is about being precise, whereas recall is about capturing all the cases.
- Therefore, even if the model captures one correct cancer positive case, it is 100% precise.
- If the model captures every case as cancer positive, you have 100% recall.
- If you want to focus more on minimizing false negatives, you would want 100% recall with good precision score.
- If you want to focus on minimizing false positives, then you should aim for 100% precision.

Specificity

- Specificity measures the proportion of actual negatives that are correctly identified.
- Specificity tries to identify the probability of a negative test result when input with negative example.

Specificity: Calculation



$$\text{Specifically} = \frac{TN}{TN + FP}$$

Specificity: Example

- Consider the previously discussed cancer detection example where out of 100 people, only 5 people have cancer.

- Specificity identifies the proportion of patients who were predicted as not having cancer and are non-cancerous.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

- The denominator (false positives and true negatives) is 95.

- The numerator (person not having cancer and the model predicting the case as no cancer) is 0.

- Since every predicted case is cancerous, the specificity of this model is 0%.

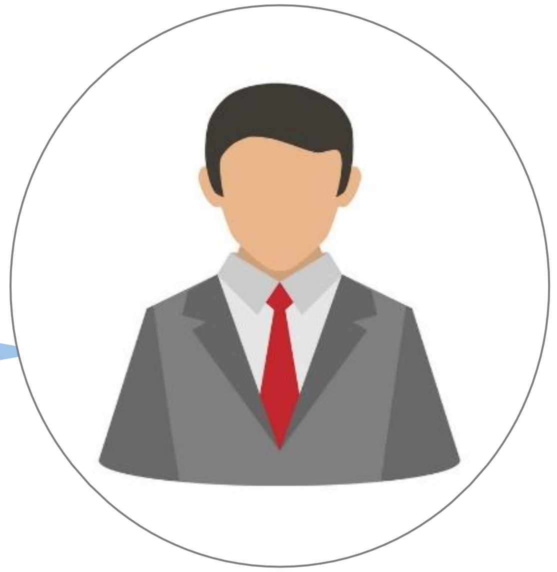
Note:

Specificity is exact opposite of recall.

F1 Score

Do you have to carry both precision and recall in your pockets every time you make a model for solving a classification problem?

No! To avoid taking both precision and recall, it is best to get a single score (F1 score) that can represent both precision(P) and recall(R).



F1 Score: Calculation

Predicates \ Actual	Fraud	Not Fraud
Fraud	3	97
Not Fraud	0	0

y

Arithmetic Mean
= (x + y)/2

Harmonic Mean
= (x + y)/2

x

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

F1 Score: Example

Fraud detection

- Let's consider 100 credit card transactions, out of which 97 are legit and 3 are fraud.
- Assume that a model predicts everything as fraud.
- The precision and recall for this example would be:

$$\text{Precision} = \frac{3}{100} = 3\%$$

$$\text{Recall} = \frac{3}{3} = 100\%$$

- The arithmetic mean of precision and recall would be:

$$\text{Arithmetic mean} = \frac{3+100}{2} = 51.5\%$$

Note: A model that predicts every transaction as fraud should not be given a moderate score. Instead of arithmetic mean, harmonic mean can be used.

Harmonic Mean

- Harmonic mean is an average used when x and y are equal.
- The value of the mean is smaller when x and y are different.
- With reference to the fraud detection example, F1 score can be calculated as:

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

$$= \frac{2 * 3 * 100}{100 + 3} = 5\%$$

Key Takeaways

- Confusion metrics is used for finding the correctness and accuracy of the model.
- Accuracy is the number of correct predictions made by the model over all kinds of predictions.
- Precision tries to answer the question, “what proportion of positive identifications was actually correct?”
- Recall measures the proportion of actual positives that are identified correctly.

Key Takeaways

- Specificity measures the proportion of actual negatives that are identified correctly.
- F1 Score gives a single score that represents both precision(P) and recall(R).
- The harmonic mean is used when the sample data contains extreme values (too big or too small) because it is more balanced than arithmetic mean.

Knowledge Check

Knowledge Check
1

What is precision?

A. Precision is also known as the true positive rate.

B. Precision is also known as the positive predictive value. It is a measure of the amount of accurate positives that our model claims in comparison with the number of positives it actually claims.

C. Precision is also known as the positive predictive value. Recall is the number of correct predictions made by the model over all kinds predictions made.

D. Precision is the number of correct predictions made by the model over all kinds predictions made.

Knowledge Check 1

What is precision?

A. Precision is also known as the true positive rate.

B. Precision is also known as the positive predictive value. It is a measure of the amount of accurate positives that our model claims in comparison with the number of positives it actually claims.

C. Precision is also known as the positive predictive value. Recall is the number of correct predictions made by the model over all kinds predictions made.

D. Precision is the number of correct predictions made by the model over all kinds predictions made.

The correct answer is **B**

Precision is also known as the positive predictive value, and it is a measure of the amount of accurate positives our model claims compared to the number of positives it actually claims.

Knowledge Check

2

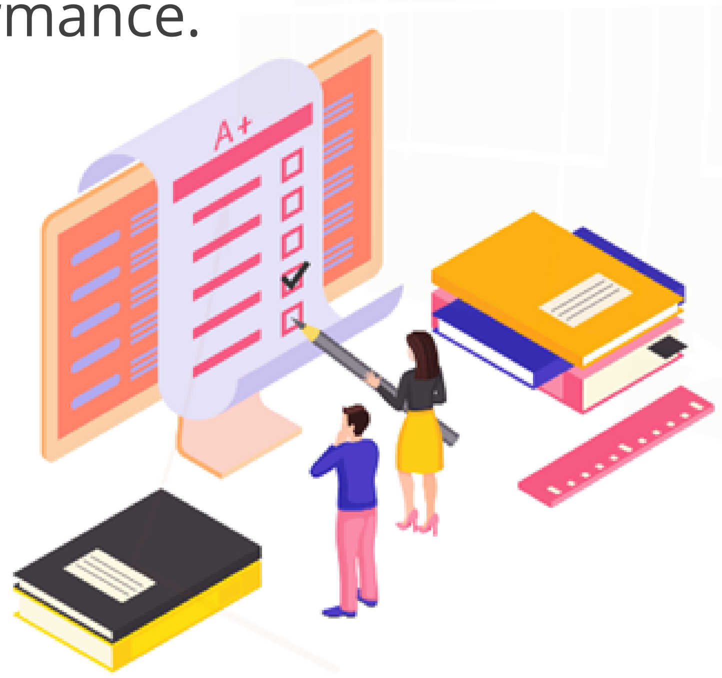
Which is more important: model accuracy or model performance?

A. Model performance is maximized when precision and recall are maximized.

B. Model accuracy is a subset of model performance and can be addressed or maximized.

C. There are models with higher accuracy that can perform worse in predictive power. it has everything to do with how model accuracy is only a subset of model performance.

D. Confusion matrix is the key for model performance and is determined by accuracy. The more the accuracy better the model.



Knowledge Check

2

Which is more important: model accuracy or model performance?

A. Model performance is maximized when precision and recall are maximized.

B. Model accuracy is a subset of model performance and can be addressed or maximized.

C. There are models with higher accuracy that can perform worse in predictive power. it has everything to do with how model accuracy is only a subset of model performance.

D. Confusion matrix is the key for model performance and is determined by accuracy. The more the accuracy better the model.

The correct answer is **C**

There are models with higher accuracy that can perform worse in predictive power. it has everything to do with how model accuracy is only a subset of model performance.