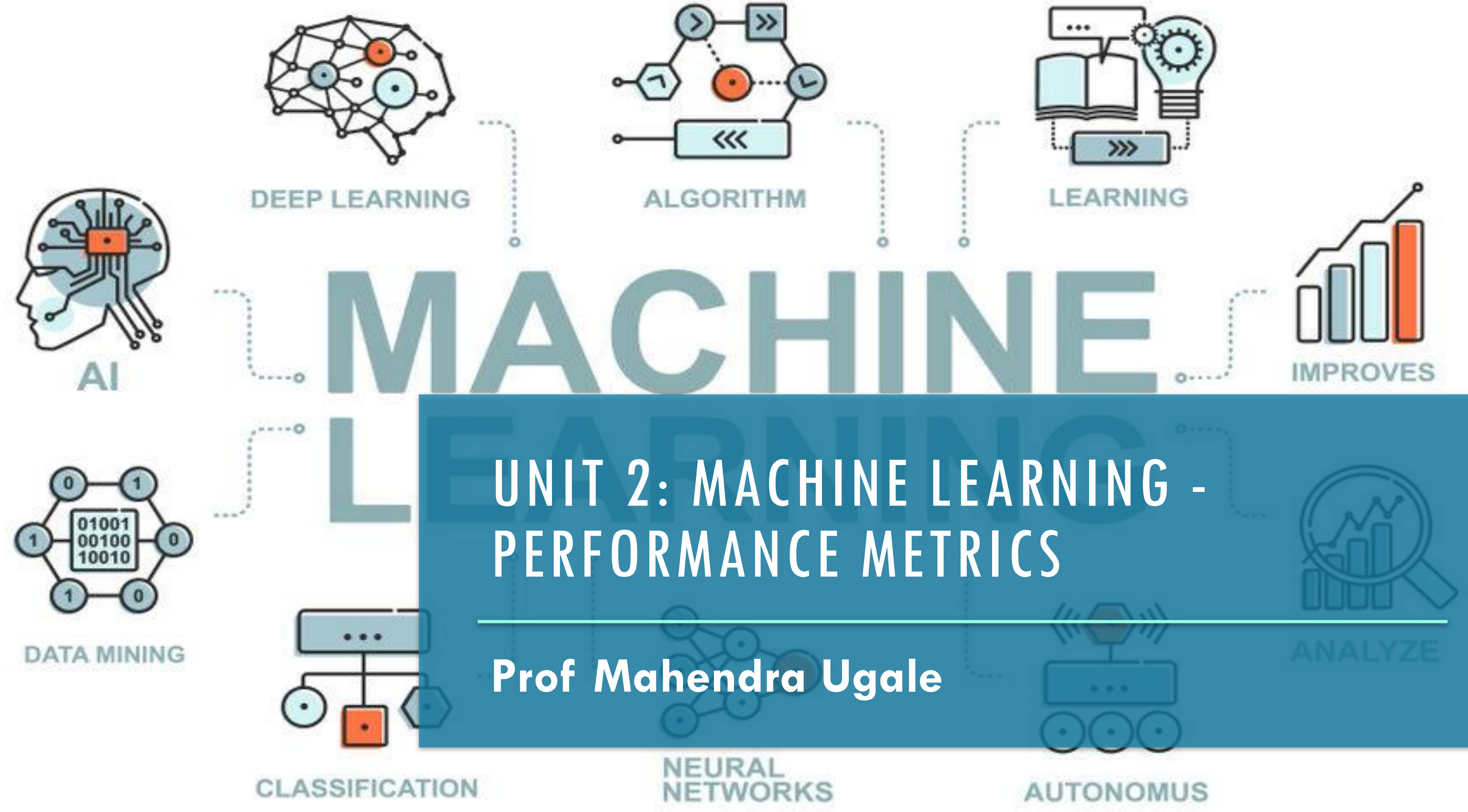


UNIT 2: MACHINE LEARNING - PERFORMANCE METRICS

Prof Mahendra Ugale

INDEX

Performance Metrics for Classification Problems-

- Confusion Matrix,
  - Classification Accuracy,
  - Classification Report- Precision, Recall or Sensitivity, F1 Score,
  - AUC (Area Under ROC curve)

Performance Metrics for Regression Problems-

- Mean Absolute Error (MAE),
- Mean Square Error (MSE),
- R Squared (R2)

INTRODUCTION

- Evaluation metrics are tied to machine learning tasks. There are different metrics for the tasks of classification and regression.
- Some metrics, like precision-recall, are useful for multiple tasks.
- Classification and regression are examples of supervised learning, which constitutes a majority of machine learning applications.
- Using different metrics for performance evaluation, we should be able to improve our model's overall predictive power before we roll it out for production on unseen data.
- Without doing a proper evaluation of the Machine Learning model by using different evaluation metrics, and only depending on accuracy, can lead to a problem when the respective model is deployed on unseen data and may end in poor predictions.

1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

- Confusion Matrix,
- Classification Accuracy,
- Classification Report- Precision, Recall or Sensitivity,

  Support,
- F1 Score,
- AUC (Area Under ROC curve)

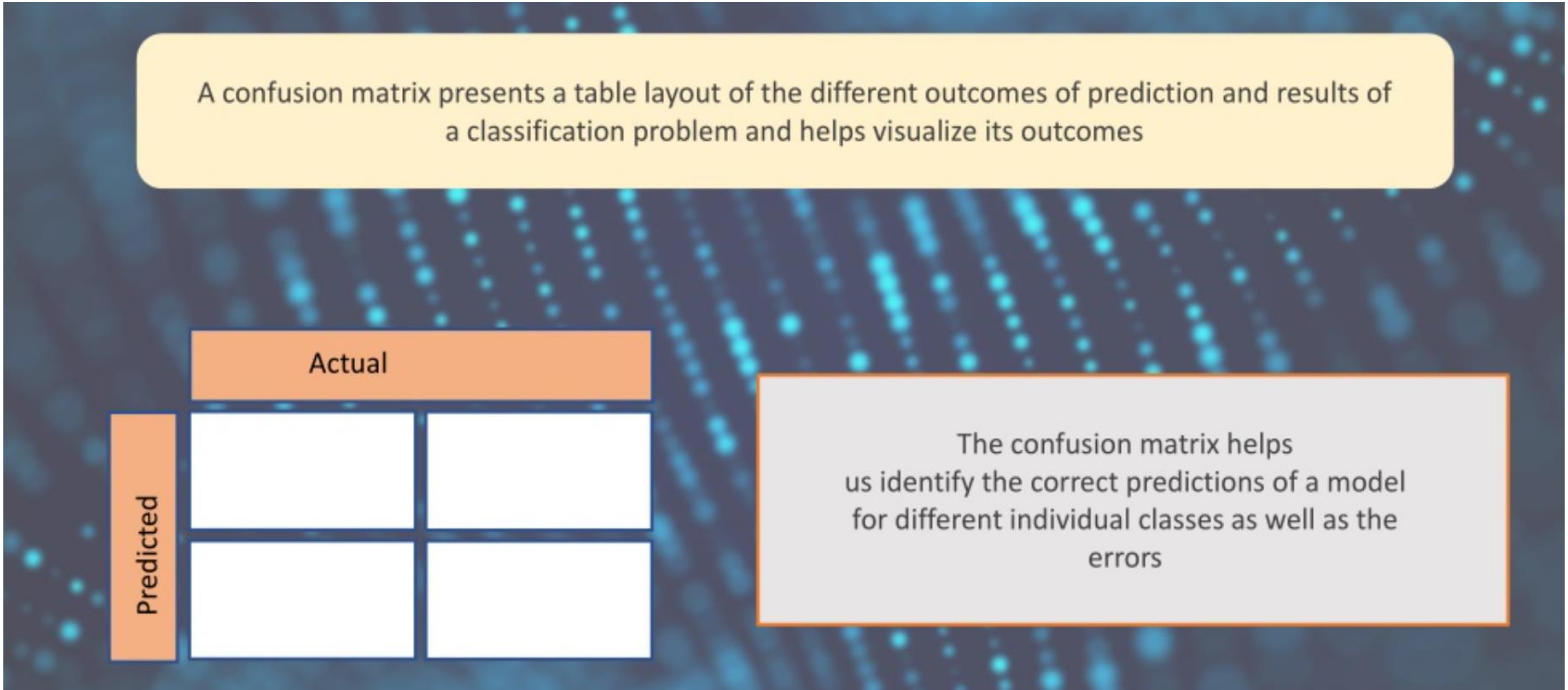
1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

▪ **Confusion Matrix-**

- The confusion matrix is a matrix used to determine the performance of the classification models for a given set of test data.
- It can only be determined if the true values for test data are known.
- The matrix is divided into two dimensions, that are **predicted values** and **actual values** along with the total number of predictions.
- Predicted values are those values, which are predicted by the model, and actual values are the true values for the given observations.
- It looks like the below table:

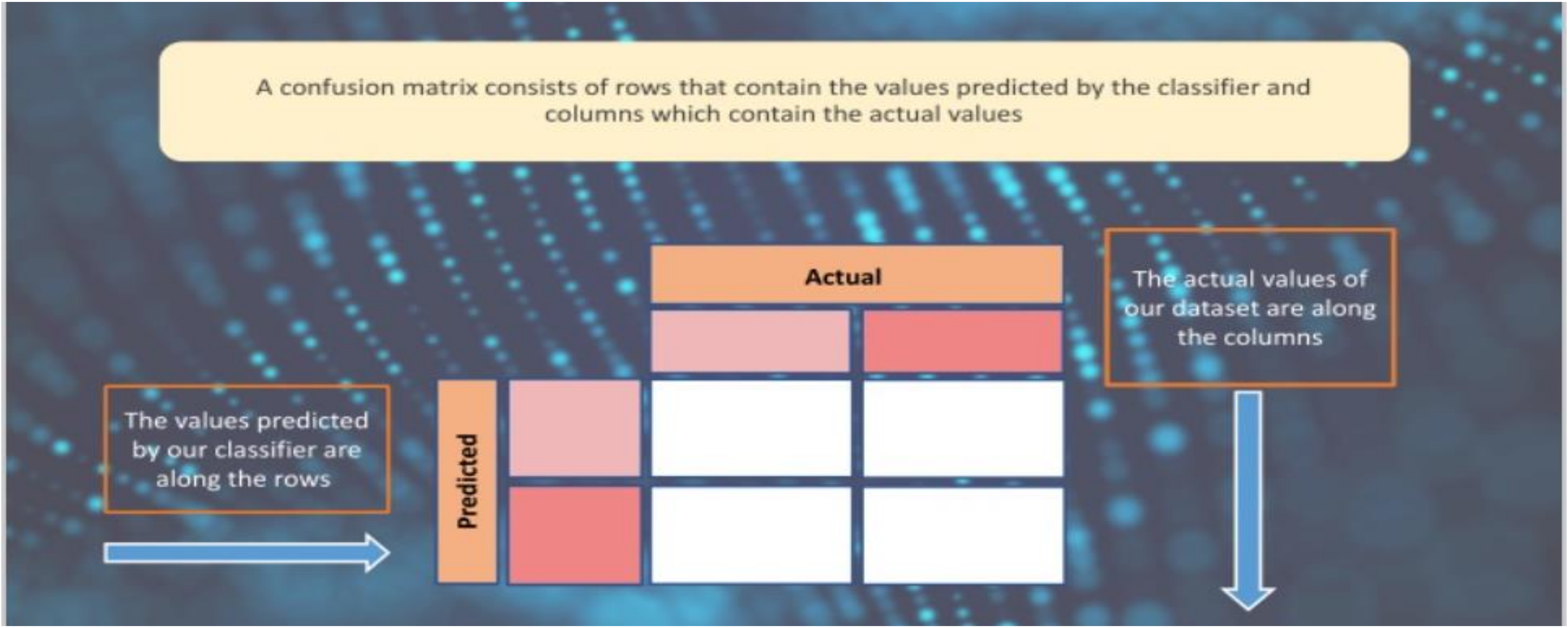
1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

- **Confusion Matrix-**The following table has the following cases:



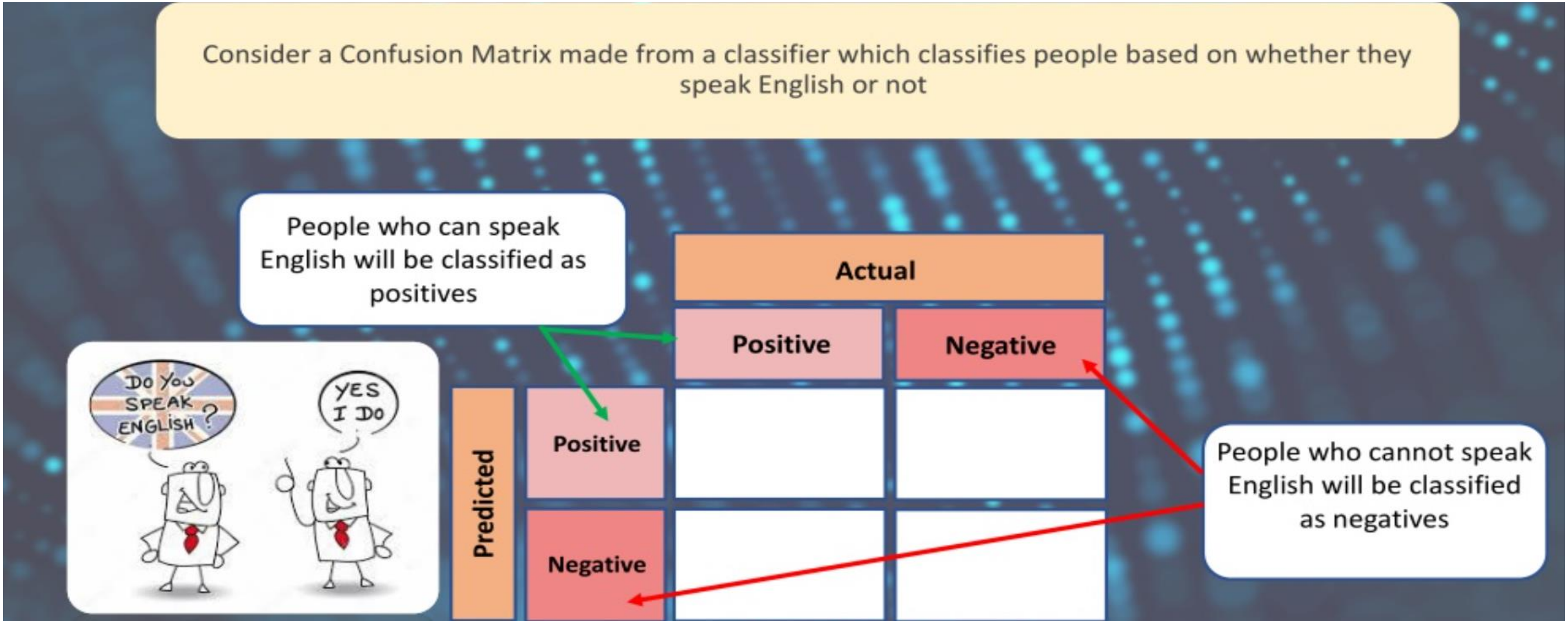
1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

- **Confusion Matrix**-The following table has the following cases:



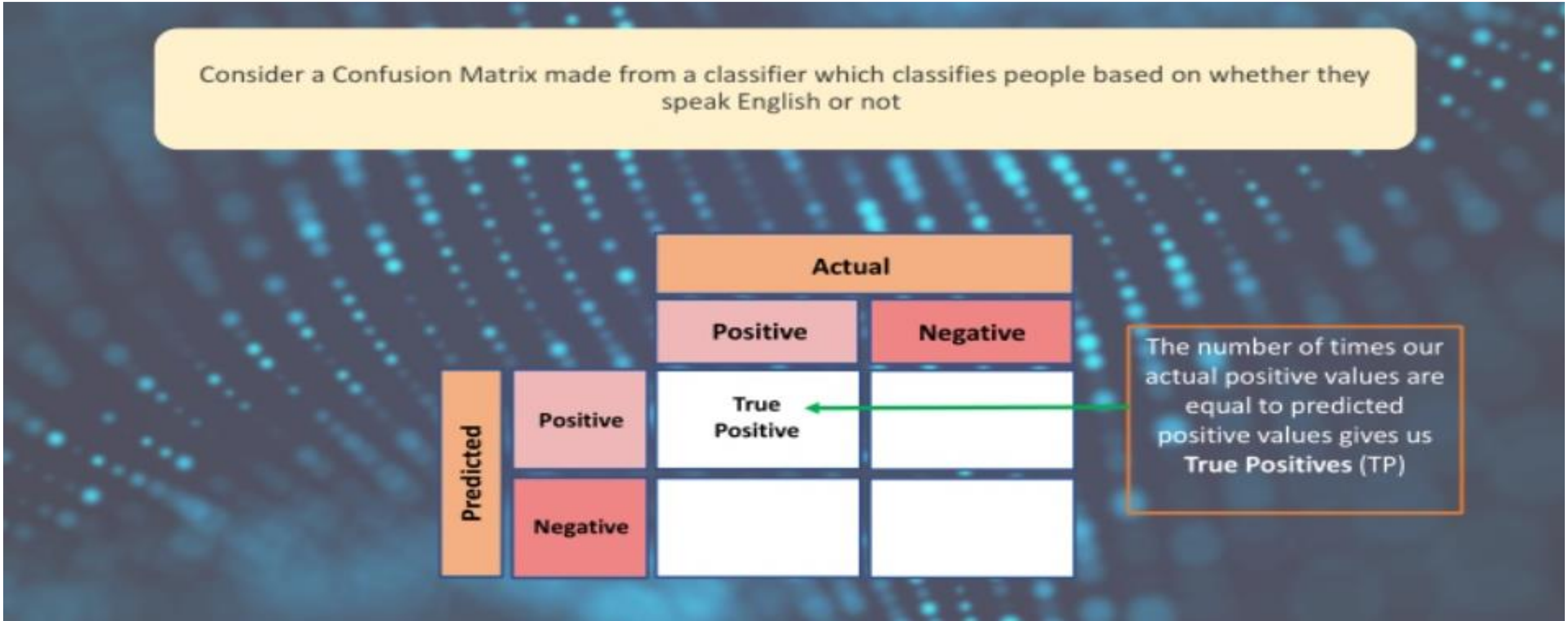
1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

- **Confusion Matrix**-The following table has the following cases:



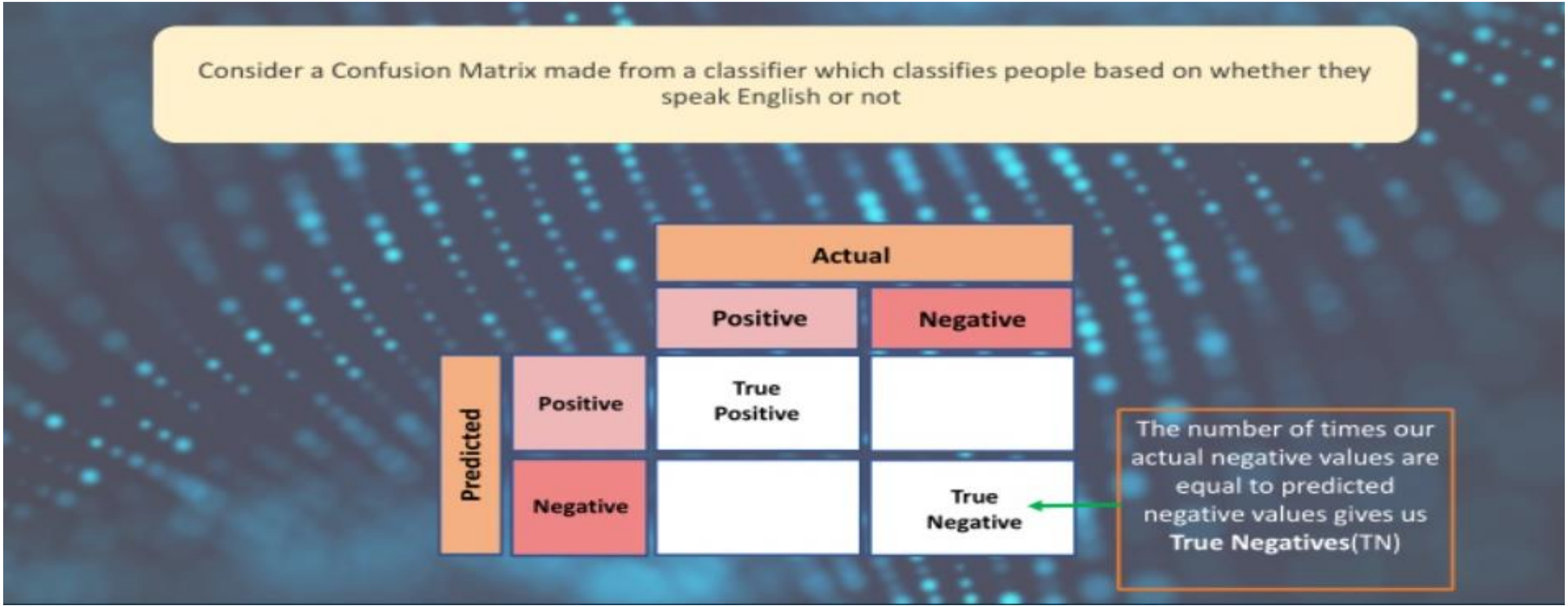
1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

- **Confusion Matrix-**The following table has the following cases:



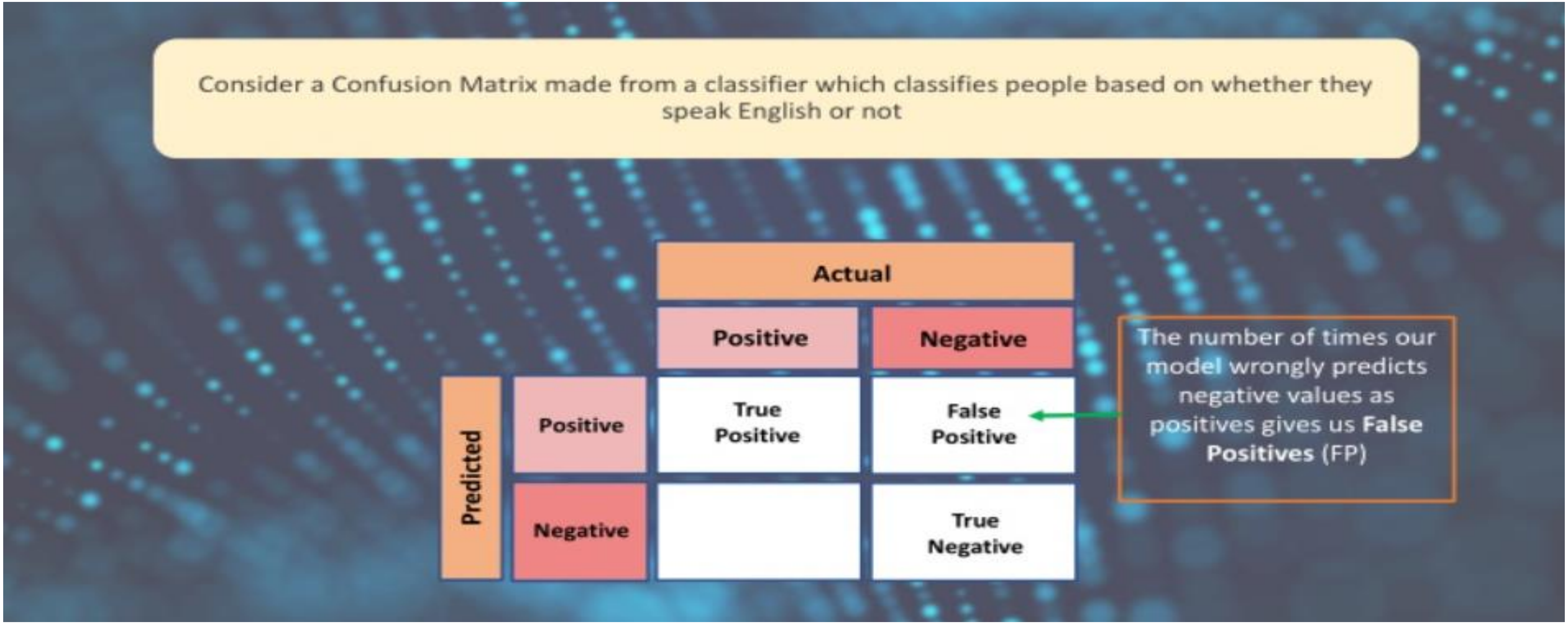
1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

- **Confusion Matrix-**The following table has the following cases:



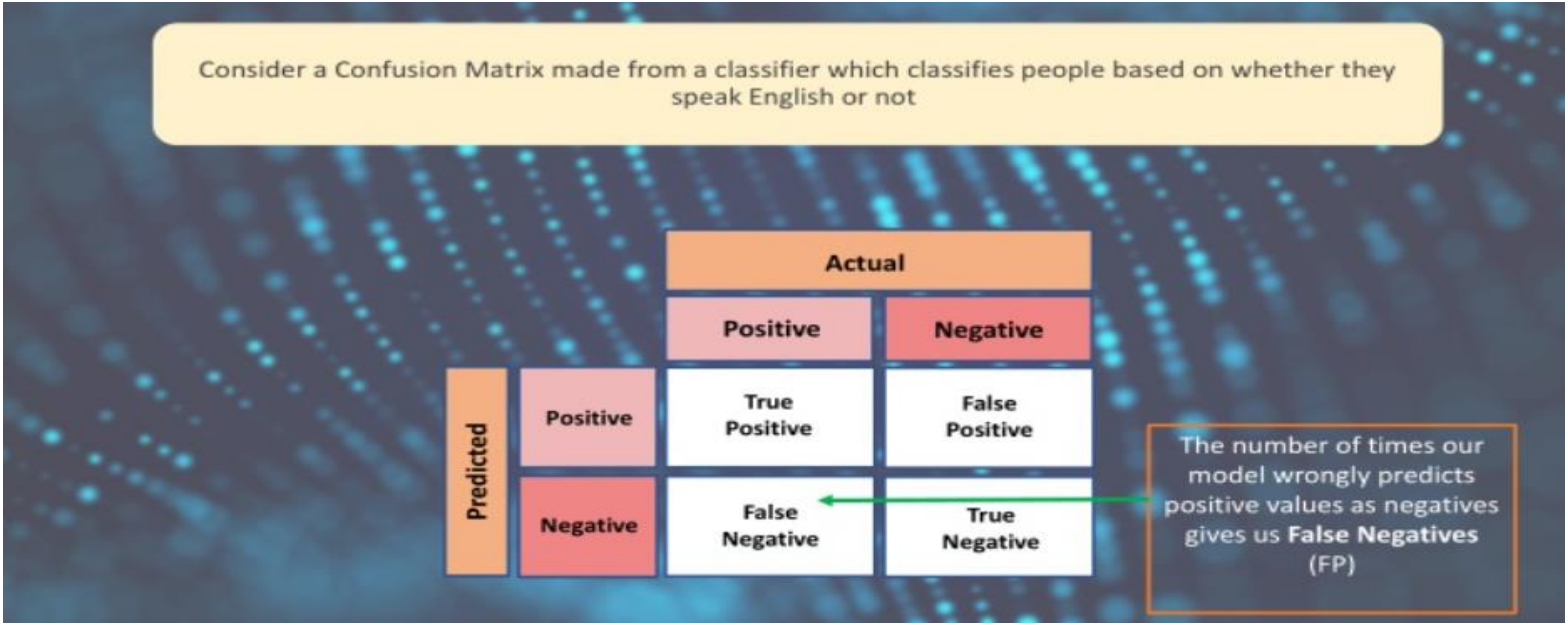
1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

- **Confusion Matrix-**The following table has the following cases:

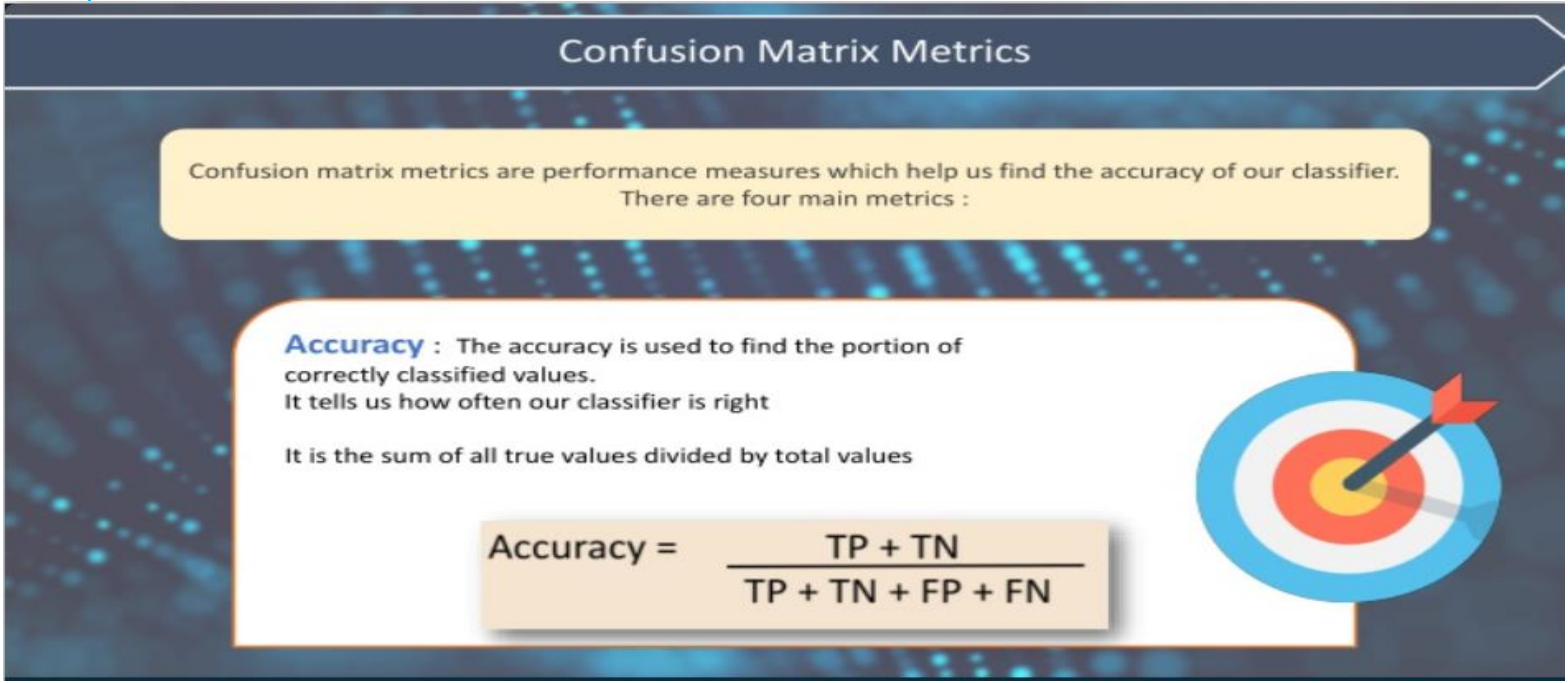


1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

- **Confusion Matrix-**The following table has the following cases:

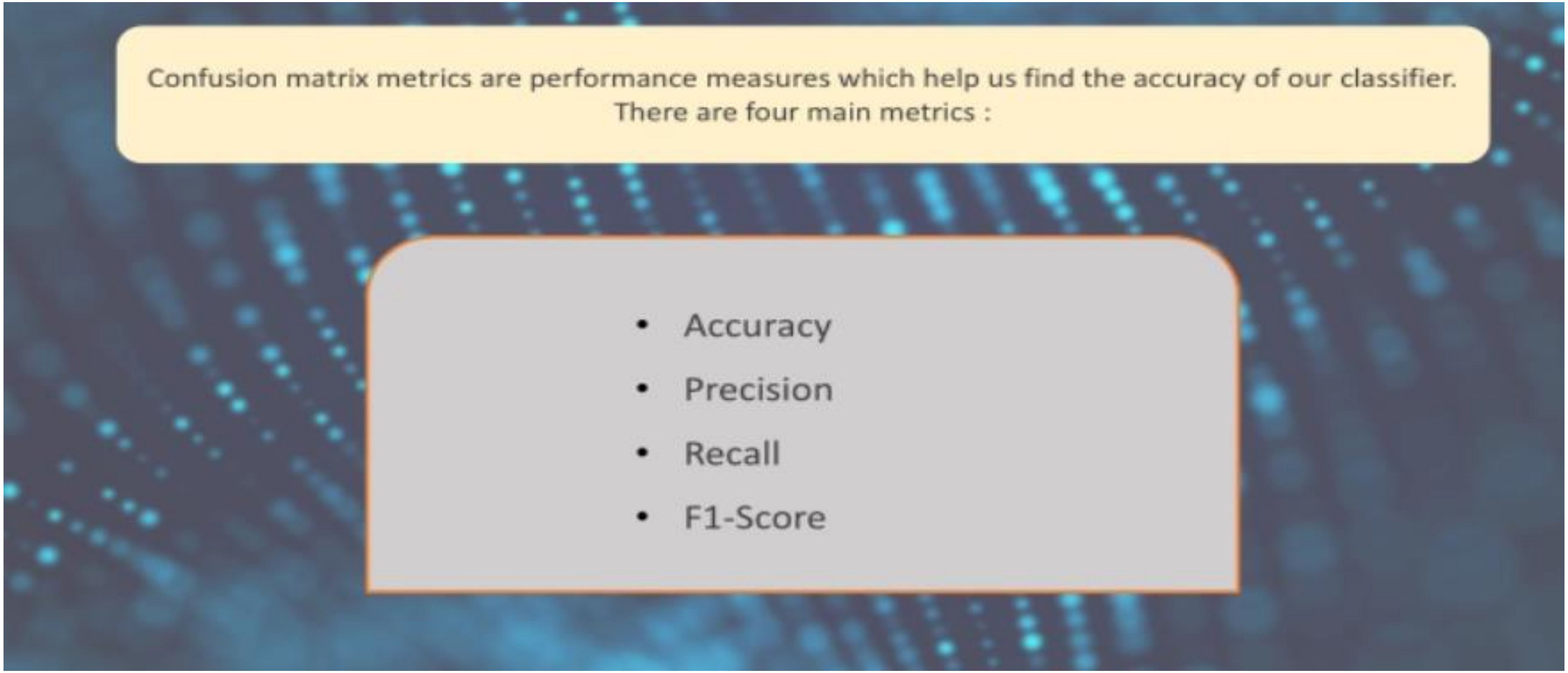


1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-



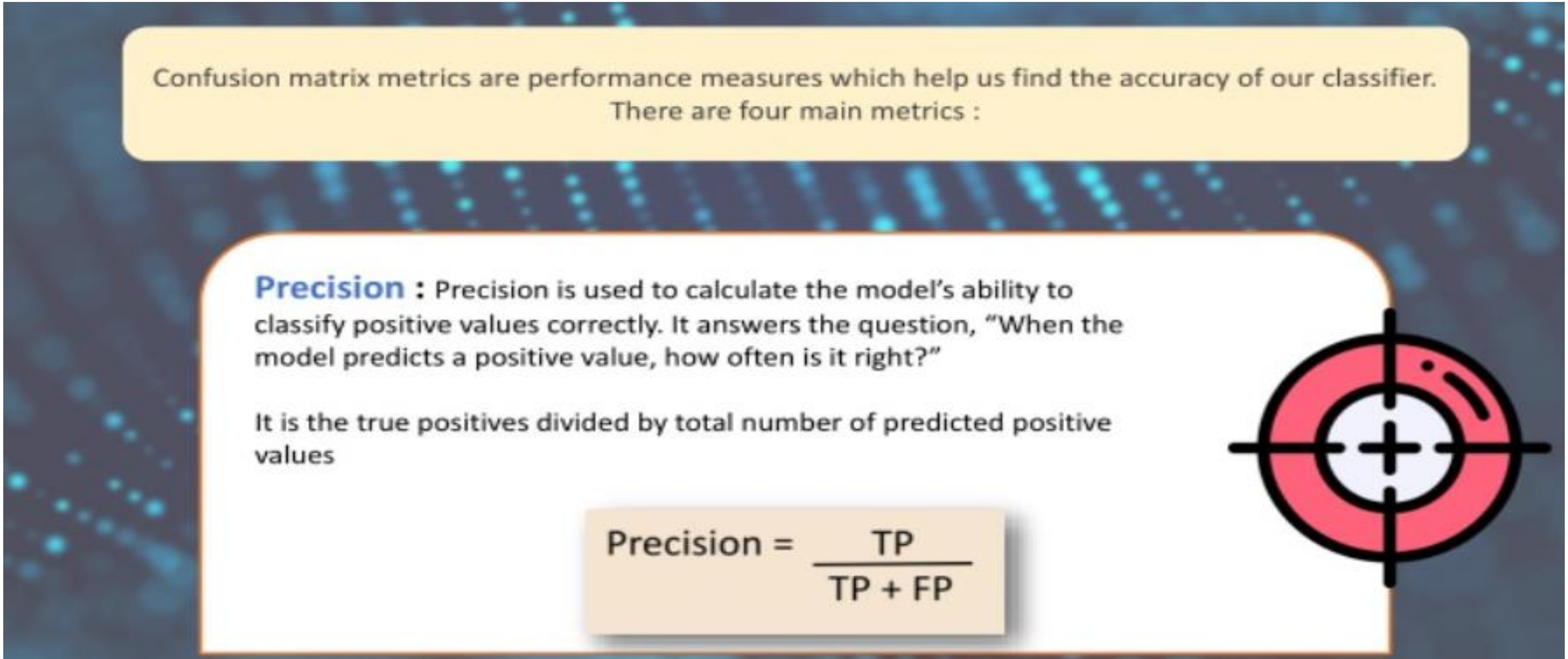
1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

- **Confusion Matrix-**



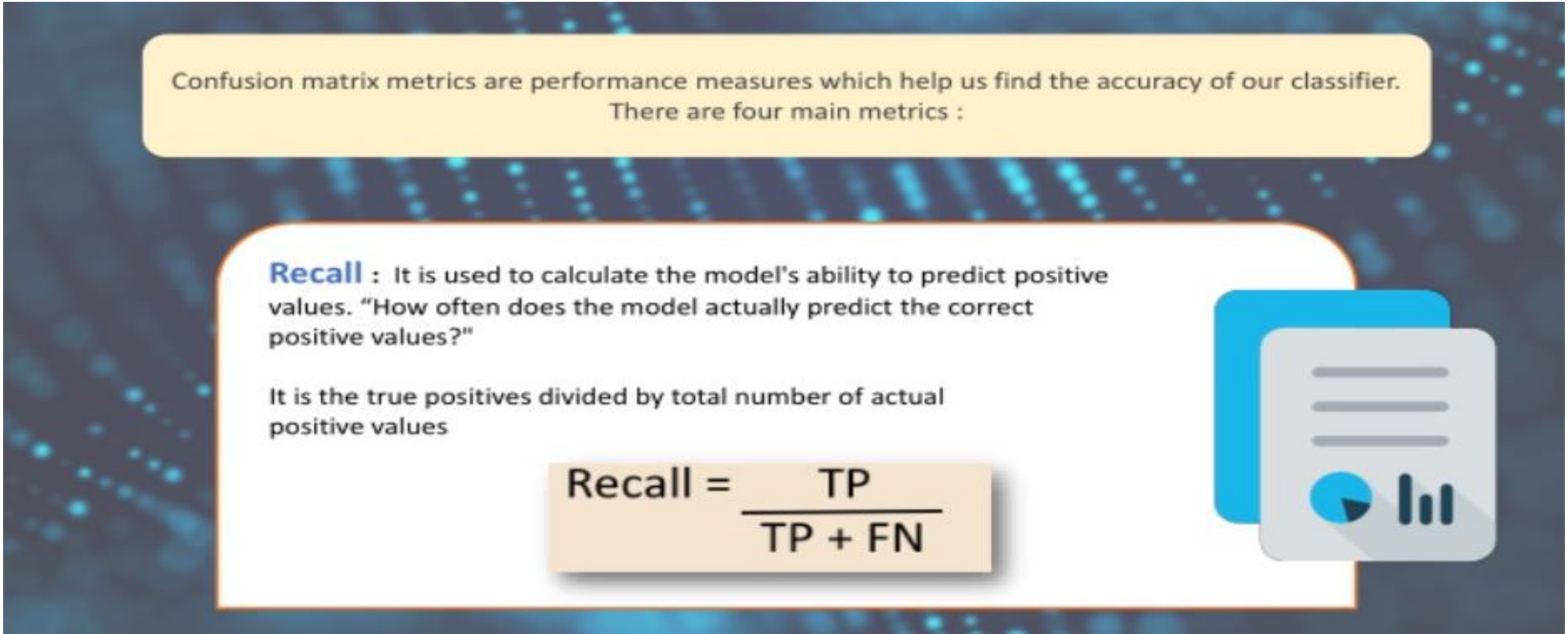
1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

- **Confusion Matrix-**



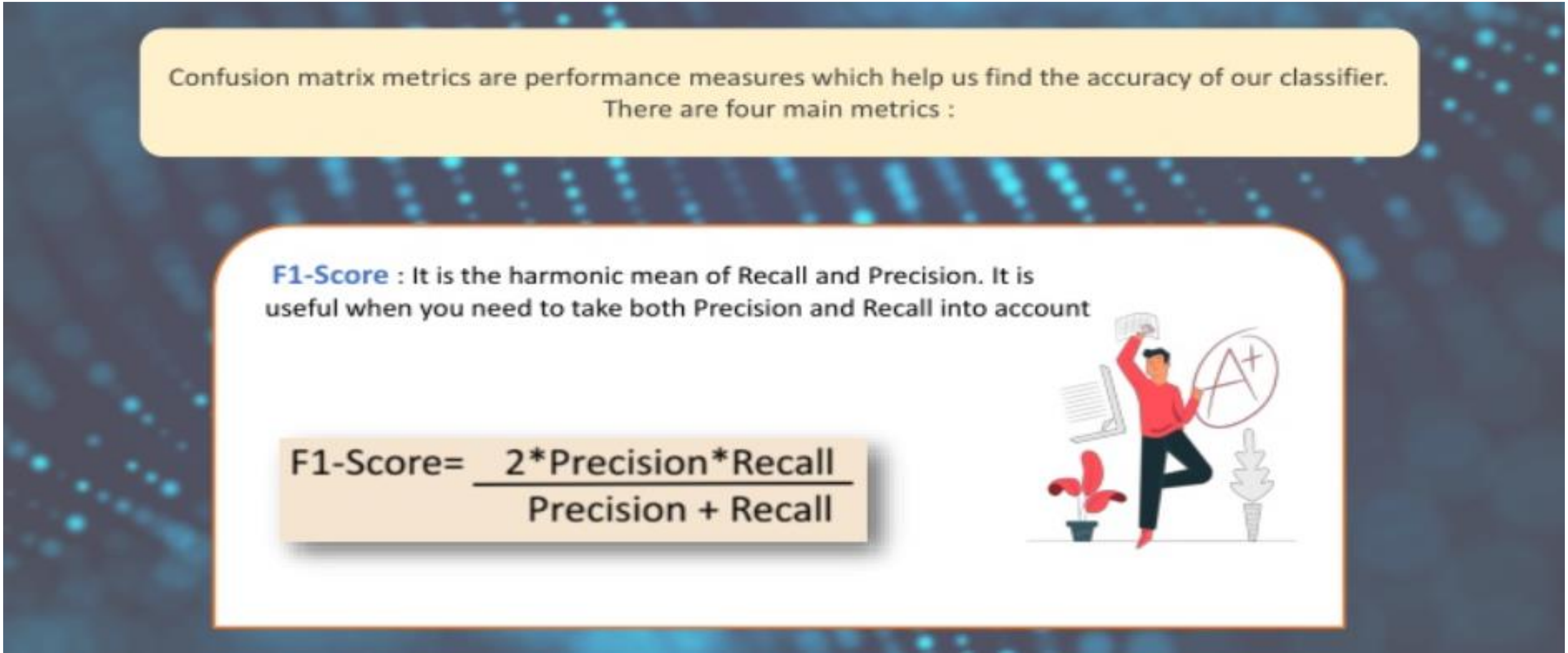
1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

- **Confusion Matrix-**



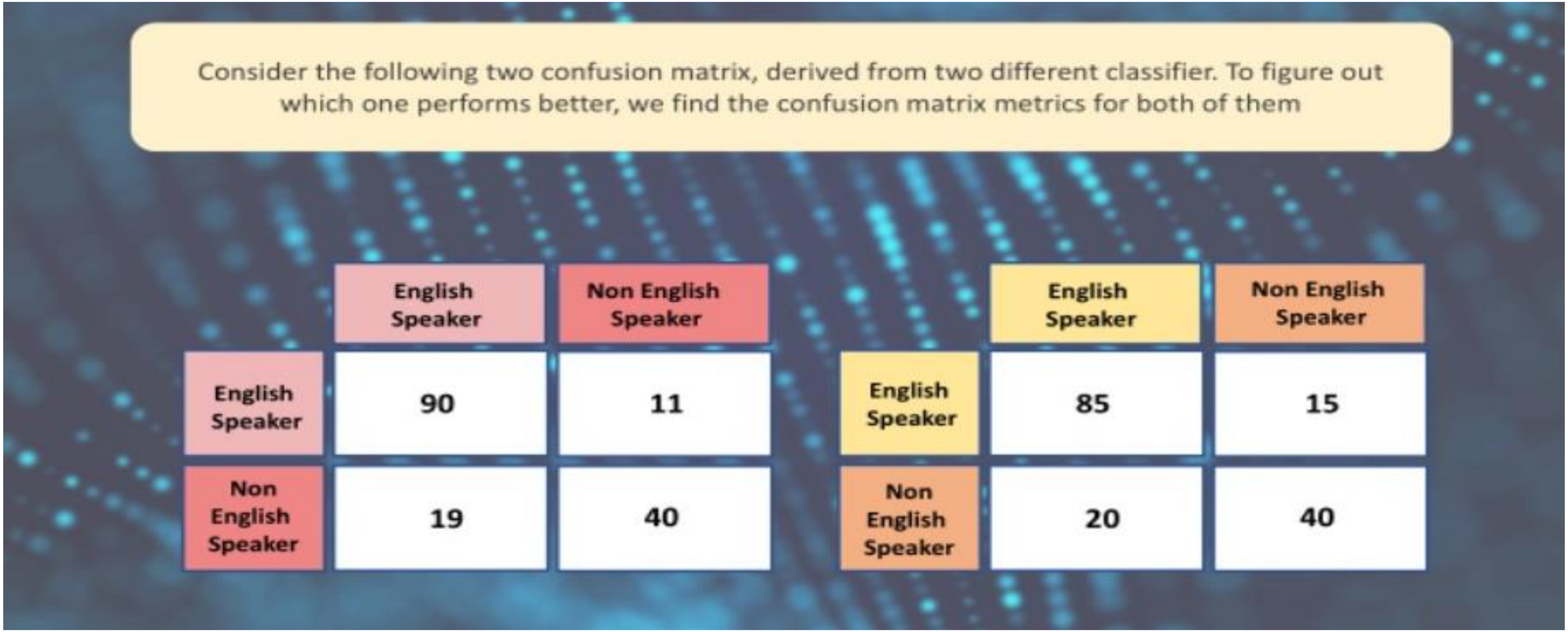
1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

- **Confusion Matrix-**



1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

- **Confusion Matrix-**



Consider the following two confusion matrix, derived from two different classifier. To figure out which one performs better, we find the confusion matrix metrics for both of them

	English Speaker	Non English Speaker
English Speaker	90	11
Non English Speaker	19	40

	English Speaker	Non English Speaker
English Speaker	85	15
Non English Speaker	20	40

1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

- **Confusion Matrix-**

	English Speaker	Non English Speaker
English Speaker	90	11
Non English Speaker	19	40

- Accuracy = (TP + TN) / (TF+TN+FP+FN)
  = (90+40)/ (90+40+19+11) = 0.8125
- Precision = TP/(TP+FP)
  = 90/(90+ 11) = 0.891
- Recall = TP/(TP+FN)
  = 90/ (90 + 19) = 0.8256
- F1-Score = 2* Precision*Recall / (Precision + Recall)
  = 2*0.891*0.8256/ (0.8256+ 0.891) = 0.857

	English Speaker	Non English Speaker
English Speaker	85	15
Non English Speaker	20	40

- Accuracy = (TP + TN) / (TF+TN+FP+FN)
  = (85+40)/ (85+40+15+20) = 0.781
- Precision = TP/(TP+FP)
  = 85/(85+ 15) = 0.85
- Recall = TP/(TP+FN)
  = 85/ (85 + 20) = 0.809
- F1-Score = 2* Precision*Recall / (Precision + Recall)
  = 2*0.85*0.809/ (0.85+ 0.809) = 0.828

1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

- **Confusion Matrix-**

  Need for Confusion Matrix in Machine Learning-

- It evaluates the performance of the classification models, when they make predictions on test data, and tells how good our classification model is.
- It not only tells the error made by the classifiers but also the type of errors such as it is either type-I or type-II error.
- With the help of the confusion matrix, we can calculate the different parameters for the model, such as accuracy, precision, etc.

1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

AUC-ROC Curve in Machine Learning

- In Machine Learning, performance measurement is an essential task. So when it comes to a classification problem, we can count on an AUC - ROC Curve.

- When we need to check or visualize the performance of the multi-class classification problem, we use the AUC (**Area Under The Curve**) ROC (**Receiver Operating Characteristics**) curve.

- It is one of the most important evaluation metrics for checking any classification model's performance.
- It is also written as AUROC (**Area Under the Receiver Operating Characteristics**)

1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

AUC-ROC Curve in Machine Learning

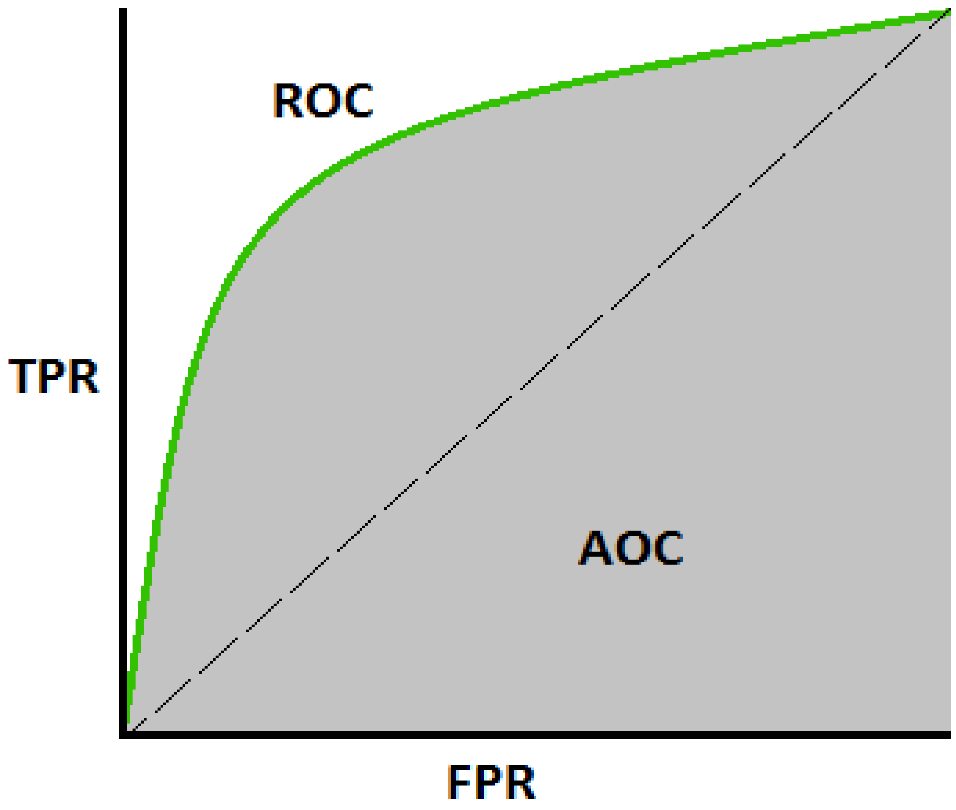
What is the AUC - ROC Curve?

- AUC - ROC curve is a performance measurement for the classification problems at various threshold settings.
- ROC is a probability curve and AUC represents the degree or measure of separability.
- It tells how much the model is capable of distinguishing between classes.
- Higher the AUC, the better the model is at predicting 0 classes as 0 and 1 classes as 1.
- By analogy, the Higher the AUC, the better the model is at distinguishing between patients with the disease and no disease.

1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

AUC-ROC Curve in Machine Learning

- The ROC curve is plotted with TPR against the FPR where TPR is on the y-axis and FPR is on the x-axis.



1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

AUC-ROC Curve in Machine Learning

Defining terms used in AUC and ROC Curve.

1. TPR (True Positive Rate) / Recall /Sensitivity

$$\text{TPR /Recall / Sensitivity} = \frac{TP}{TP + FN}$$

1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

AUC-ROC Curve in Machine Learning

Defining terms used in AUC and ROC Curve.

2. Specificity

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

1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

AUC-ROC Curve in Machine Learning

Defining terms used in AUC and ROC Curve.

3. FPR

$$\text{FPR} = 1 - \text{Specificity}$$

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

1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

AUC-ROC Curve in Machine Learning

How to speculate about the performance of the model?

- An excellent model has AUC near to the 1 which means it has a good measure of separability.

- A poor model has an AUC near 0 which means it has the worst measure of separability. In fact, it means it is reciprocating the result. It is predicting 0s as 1s and 1s as 0s. And

- When AUC is 0.5, it means the model has no class separation capacity whatsoever.

1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

AUC-ROC Curve in Machine Learning

How to speculate about the performance of the model?

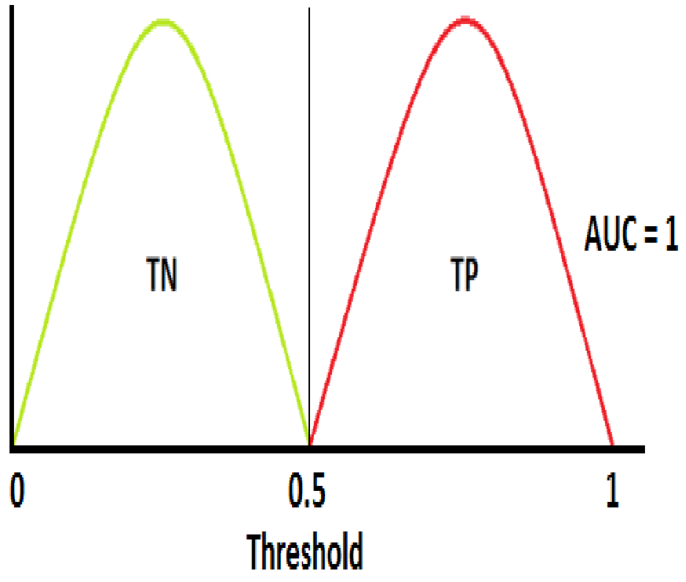
- As we know, ROC is a curve of probability.
- So let's plot the distributions of those probabilities:

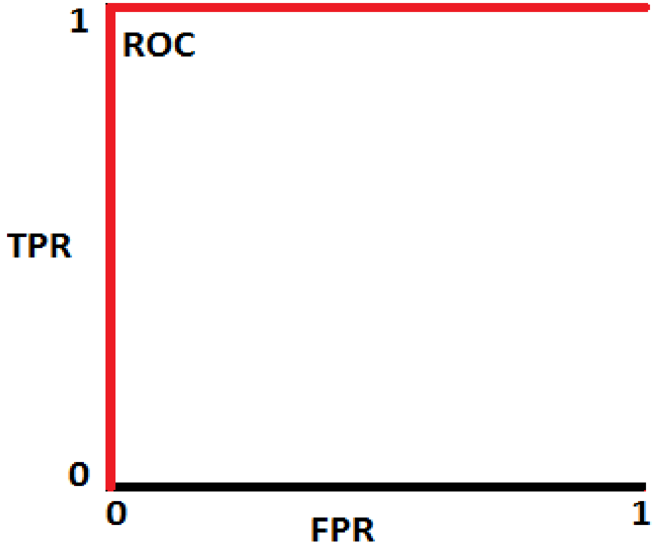
1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

AUC-ROC Curve in Machine Learning

How to speculate about the performance of the model?

- Note: Red distribution curve is of the positive class (patients with disease) and the green distribution curve is of the negative class(patients with no disease).



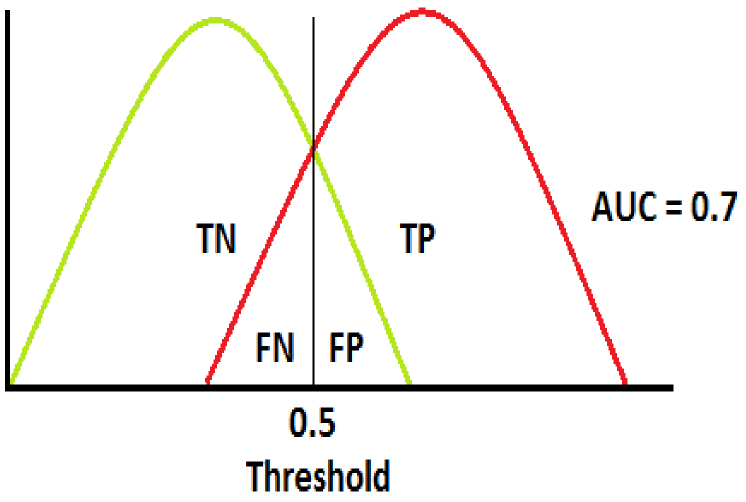


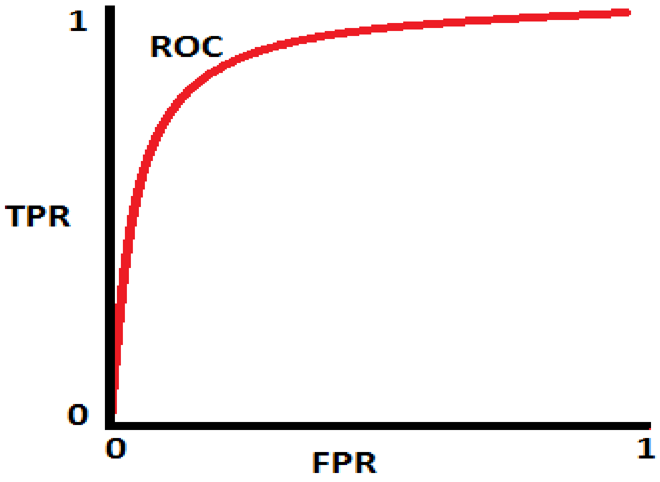
- This is an ideal situation. When two curves don't overlap at all means model has an ideal measure of separability.
- It is perfectly able to distinguish between positive class and negative class.

1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

AUC-ROC Curve in Machine Learning

How to speculate about the performance of the model?



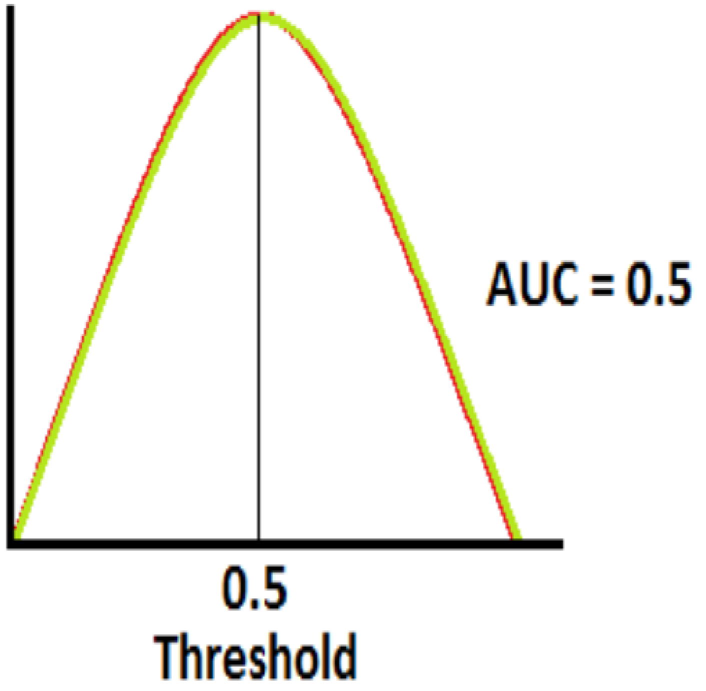


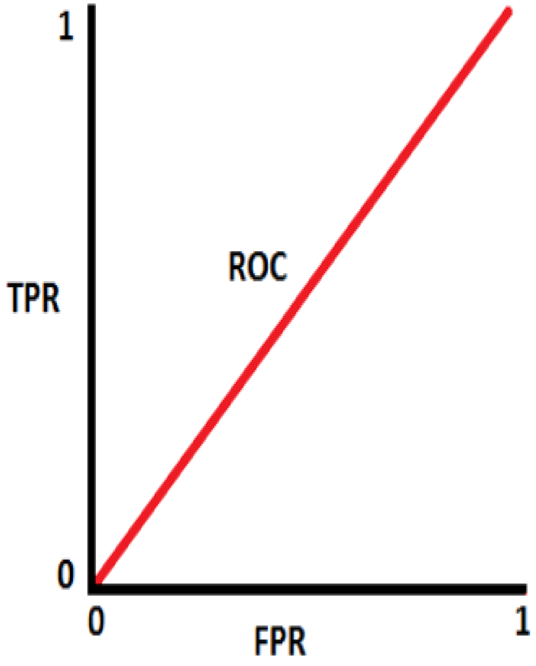
- When two distributions overlap, we introduce type 1 and type 2 errors. Depending upon the threshold, we can minimize or maximize them.
- When AUC is 0.7, it means there is a 70% chance that the model will be able to distinguish between positive class and negative class.

1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

AUC-ROC Curve in Machine Learning

How to speculate about the performance of the model?



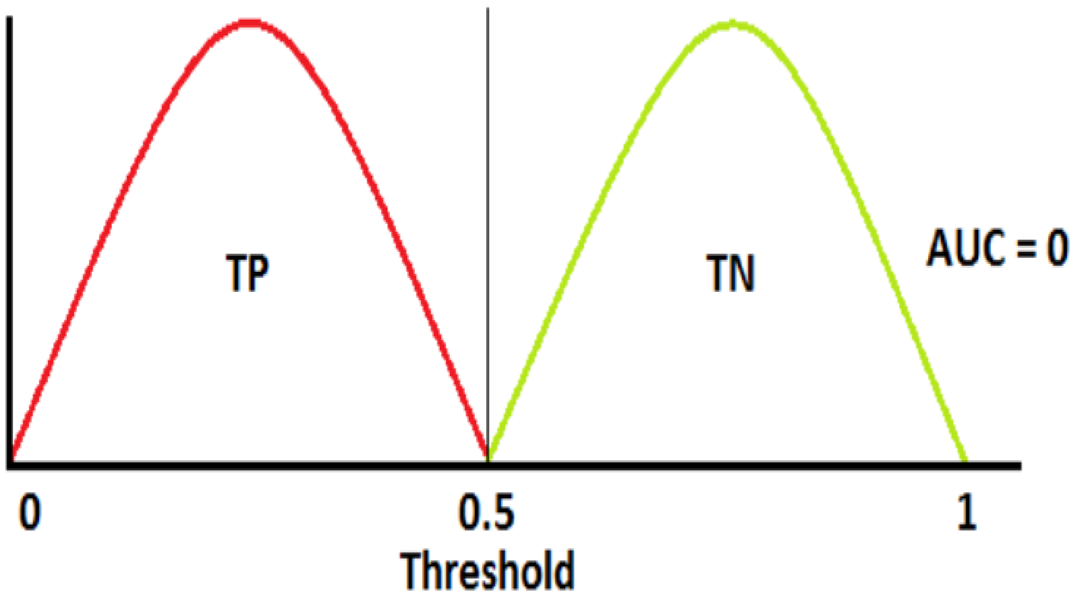


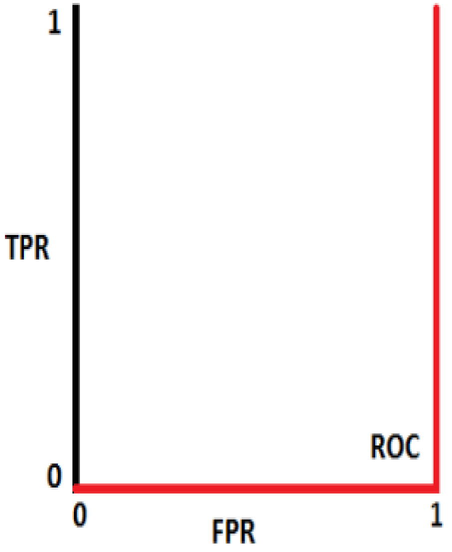
- This is the worst situation.
- When AUC is approximately 0.5, the model has no discrimination capacity to distinguish between positive class and negative class.

1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

AUC-ROC Curve in Machine Learning

How to speculate about the performance of the model?





- When AUC is approximately 0, the model is actually reciprocating the classes.
- It means the model is predicting a negative class as a positive class and vice versa.

1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

AUC-ROC Curve in Machine Learning

How to speculate about the performance of the model?

The relation between Sensitivity, Specificity, FPR, and Threshold.

- Sensitivity and Specificity are inversely proportional to each other. So when we increase Sensitivity, Specificity decreases, and vice versa.

Sensitivity⬆️, Specificity⬇️ and Sensitivity⬇️, Specificity⬆️

- When we decrease the threshold, we get more positive values thus it increases the sensitivity and decreasing the specificity.

1. PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS-

AUC-ROC Curve in Machine Learning

How to speculate about the performance of the model?

The relation between Sensitivity, Specificity, FPR, and Threshold.

- Similarly, when we increase the threshold, we get more negative values thus we get higher specificity and lower sensitivity.
- As we know FPR is 1 - specificity. So when we increase TPR, FPR also increases and vice versa.

TPR ⬆, FPR ⬆ and TPR ⬇, FPR ⬇

2. PERFORMANCE METRICS FOR REGRESSION PROBLEMS-

Following are the various performance metrics that can be used to evaluate predictions for regression problems.

- **Mean Absolute Error (MAE)**
- **Mean Square Error (MSE)**
- **R Squared (R2)**

2. PERFORMANCE METRICS FOR REGRESSION PROBLEMS-

1.**Mean Absolute Error (MAE)**

- It is the simplest error metric used in regression problems.
- It is basically the sum of average of the absolute difference between the predicted and actual values.
- In simple words, with MAE, we can get an idea of how wrong the predictions were.
- MAE does not indicate the direction of the model i.e. no indication about underperformance or overperformance of the model.
- $$MAE = \frac{1}{n}\sum|Y - \hat{Y}|$$

Here, Y=Actual Output Values

And $\hat{Y}$ = Predicted Output Values.

We can use mean_absolute_error function of sklearn.metrics to compute MAE.

2. PERFORMANCE METRICS FOR REGRESSION PROBLEMS-

2.Mean Square Error (MSE)

- MSE is like the MAE, but the only difference is that the it squares the difference of actual and predicted output values before summing them all instead of using the absolute value.

- The difference can be noticed in the following equation –

$$MSE = \frac{1}{n}\sum(Y - \hat{Y})$$

Here, Y=Actual Output Values

And $\hat{Y}$ = Predicted Output Values.

We can use mean_squared_error function of sklearn.metrics to compute MSE.

2. PERFORMANCE METRICS FOR REGRESSION PROBLEMS-

3 . R Squared (R2)

- R Squared metric is generally used for explanatory purpose and provides an indication of the goodness or fit of a set of predicted output values to the actual output values.
- The following formula will help us understanding it –

$$R^2 = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y_i})^2}{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y_i})^2}$$

- In the above equation, numerator is MSE and the denominator is the variance in Y values.
- We can use r2_score function of sklearn.metrics to compute R squared value.

2. PERFORMANCE METRICS FOR REGRESSION PROBLEMS-

The following python code will give us an insight about how we can use the above explained performance metrics on regression model −

```
from sklearn.metrics import r2_score
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
X_actual = [5, -1, 2, 10]
Y_predic = [3.5, -0.9, 2, 9.9]
print ('R Squared =',r2_score(X_actual, Y_predic))
print ('MAE =',mean_absolute_error(X_actual, Y_predic))
print ('MSE =',mean_squared_error(X_actual, Y_predic))
```

Output

```
R Squared = 0.9656060606060606
MAE = 0.42499999999999993
MSE = 0.5674999999999999
```

QUESTIONS?

mahendra_ugale@csmssengg.org

DISCLAIMER

The material for the presentation has been compiled from various sources such as books, tutorials (offline and online), lecture notes, several resources available on Internet. The information contained in this lecture/presentation is for general information and education purpose only. While we endeavor to keep the information up to date and correct, we make no representation of any kind about the completeness and accuracy of the material. The information shared through this presentation material should be used for educational purpose only.