Class -10 , AI-417 Question Answer of Evaluation

1Q – What is a confusion matrix? why is it used?

Ans – A Confusion matrix is a performance measure for classification-based AI models where output can belong to two or more classes. It is a table with 4 different combinations of predicted and actual values.

Actual value	Predicted value	
	Positive	Negative
Positive (1)	No. of True Positive (TP)	No. of False Negative (FN) Type II error
Negative (0)	No. of False Positives (FP) Type I error	No. of True Negatives (TN)

When both the predicted and actual values match, they are True Positive or True Negatives.

Prediction	Reality(actual)	True/False	Positive/Negative
Yes	Yes	True	Positive
No	No	True	Negative
Yes	No	False	Positive
No	Yes	False	Negative

- ✓ If Prediction and Reality match it becomes True.
- ✓ If Prediction is Yes then it is Positive and
- ✓ If Prediction is No then it is False

2Q – What are different metrices used to evaluate classification based AI models ?

Ans – Different evaluation metrics used for classification-based AI models are Recall, Precision, Accuracy and F1 score, etc.

3Q– What is precision? What is the formula to compare precision? What is its significance?

Ans – The Precision (also known as the positive predictive value) is defined as the proportion of positive predictions that are truly positive.

Its formula is :

Precision = $\frac{TP}{(TP+FP)}$

A model with high precision is considered trustworthy.

4Q – What is Recall? What is its formula? What is its significance?

Ans – The recall is a measure of the completeness of the results.

Its formula is :

Recall = $\frac{TP}{(TP+FN)}$

It tells ratios of correctly predicted positive outcomes (True Positives) to all actual positive outcomes (Actual Positives).

5Q – Why is accuracy not considered a complete and sufficient metric?

Ans. The accuracy metric itself is not sufficient to determine the efficiency of an Al model as it can be misleading in cases where the training data is not a balanced one. Thus, other metrics must also be taken in account to evaluate an Al model.

Accuracy is only useful when the training dataset is a balanced dataset.

6Q. Suggest a situation where the impact of false negatives would be more costly than false positives.

Ans. Suppose we have an Al-based weather forecasting device to predict rainy days. If this device has high precision, it means when it says "it is going to rain", we can trust it. But this doesn't give us any information about all the times it says "it is not going to rain"! If our business is related to drying up the crops in open, the false claims are going to be costly and would ruin all our drying crops and material.

7. Suggest a situation where the impact of false positives would be more costly than false negatives.

Ans. Military people are facing a bizarre situation where a specific disease is spreading in an unpredictable way. To ensure timely test and diagnosis, the military has put up an Al-based system that predicts which soldier may be vulnerable for that specific disease that needs timely intervention, diagnosis, and treatment. Timely detection is important as testing all soldiers is costly as the individual diagnostic tests real costly. However, if the Al model keep predicting vulnerability for many soldiers, which is actually not, this would be a really costly situation.

8. Give example of the situation when you would go for precision metric

Ans. If we have an AI-based device to predict rain and that it ALWAYS claims that "tomorrow is going to rain"! Then, surprisingly, it is not going to mis predict one single rainy day! But this does not make it a perfect device. Now Precision would be needed to evaluate the model.

9. For an AI model developed to check if a painting is authentic or not, its confusion matrix is given below (0 means False and 1 means True).

N = 192	Predicted 0	Predicted 1
Actual 0	118 TN	12 FP
Actual 1	47 FN	15 TP

As per given Confusion matrix : find Precision, Recall, Accuracy rate and F1 Score

Ans –

True Positive (TP) = 15

True Negative (TN) = 118

False Positive (FP) = 12

False Negatives (FN) = 47

Various Evaluation Matrix are calculated as :

Accuracy = $\frac{(TP+TN)}{TP+TN+FP+FN)} = \frac{(15+118)}{(15+118+112+47)} = 0.45547945205 = 45.54\%$

Precision = $\frac{TP}{TP+FP} = \frac{15}{(15+12)} = \frac{15}{27} = 0.5555555555 = 55.55\%$

Recall = $\frac{TP}{TP+FN} = \frac{15}{(15+47)} = \frac{15}{62} = 0.24193548387 = 24.1935\%$

F1 Score = $\frac{TP}{TP+(FP+FN)/2} = \frac{15}{15+(12+47)/2} = \frac{15}{15+59/2} = \frac{15}{15+29.5} = 0.33707865168 = 33.707\%$

10.Q-People of a village is dependent on the farmers for their daily food items. Farmers grow new seeds by checking the weather conditions every year. An AI model is being deployed in the village which predicts the chances of heavy rain to alert farmers which helps them in doing the farming at the right time. Which evaluation parameter out of precision, recall, and F1 Score is best to evaluate the performance of this AI model? Explain.

ANS - Let us take each of the factors into consideration at once,

- ⇨ **If precision is considered, FN cases will not be taken into account**, so it will be of great loss if the machine will **predict there will be no heavy rain**, **but if the rain occurred**, it will be a big monetary loss due to damage to crops.
- ⇨ **If only recall is considered, then FP cases will not be taken into account**. This situation will also cause a lot of loss, as all the village's people depend on farmers for food, and if the **model predicts there will be heavy rain and the farmers may not grow crops,** it will affect the basic needs of the people.
- ⇨ Hence F1 Score is the best-suited parameter to test this AI model, which is the balance between Precision and Recall.