

EVALUATION

Activity: 1

Imagine that you have come up with an AI-based prediction model which has been deployed on the roads to check traffic jams. Now, the objective of the model is to predict whether there will be a traffic jam or not. Now, to understand the efficiency of this model, we need to check if the predictions which it makes are correct or not. Thus, there exist two conditions which we need to ponder upon: Prediction and Reality.

Traffic Jams have become a common part of our lives nowadays. Living in an urban area means you have to face traffic each and every time you get out on the road. Mostly, school students opt for buses to go to school. Many times, the bus gets late due to such jams and the students are not able to reach their school on time.

Considering all the possible situations make a Confusion Matrix for the above situation.

Ans:

Case 1: Is there a traffic Jam?

Prediction: Yes Reality: Yes True Positive

Case 2: Is there a traffic Jam?

Prediction: No Reality: No True Negative

Case 3: Is there a traffic Jam?

Prediction: Yes Reality: No False Positive

Case 4: Is there a traffic Jam?

Prediction: No Reality: Yes False Negative.

The confusion Matrix		Reality	
		Yes	No
Prediction	Yes	True Positive TP	False Positive FP
	No	False Negative FN	True Negative TN

Activity: 2

Shweta is learning NLP. She read about the F1 score but did not understand the need for the F1 score formulation. Support her by giving an answer.

Ans.:

The F1 score is also known as the F score or F measure of AI model's test accuracy. It is calculated from the precision and recall of the test. Here

1) The precision is the number of correctly recognized positive results divided by the number of all positive results, including those not identified correctly.

2) The recall is the number of correctly identified positive results divided by the number of all samples that should have been identified as positive.

3) The F1 score is defined as the harmonic mean of the test's precision and recall. The formula of F1 score is as below:

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

From this formula,

i) A good F1 score means that you have high true positives and high true negatives. It means that the machine is correctly identifying real threats and not disturbs the users by false alarms.

ii) An F1 score is considered perfect when it's 1 while the model is a total failure when it's 0.

iii) F1 score is a better metric to evaluate the model on real-life classification problems and when imbalanced class distribution exists.

Activity: 3

Calculate accuracy, precision, recall, and F1 score for the following Confusion Matrix. Suggest which metric would not be a good evaluation parameter and why?

The Confusion Matrix	Reality: 1	Reality: 0
Prediction: 1	50 TP	30 FP
Prediction: 0	15 FN	25 TN

Ans:

a) Accuracy: It is defined as the percentage of correct predictions out of all the observations.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+FN+TN+FP)} \times 100\%$$

$$\frac{(50+25)}{(50+15+25+30)} \times 100\% = \frac{80}{120} \times 100\% = 0.67$$

b) Precision: It is defined as the percentage of true positive cases versus all the cases where the prediction is true.

$$\text{Precision} = \frac{TP}{(FP+TP)} = \frac{50}{(30+50)} = \frac{50}{80} = 0.625$$

c) Recall: It is defined as the fraction of positive cases that are correctly identified.

$$\text{Recall} = \frac{TP}{(FN+TP)} = \frac{50}{(15+50)} = \frac{50}{65} = 0.769$$

d) F1 Score: It is identified as the measure of balance between precision and recall.

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

$$2 \times \frac{(0.625 \times 0.769)}{(0.625 + 0.769)} = 2 \times \frac{0.96125}{1.394} = 0.7$$

Accuracy = 0.67 Precision = 0.625

Recall = 0.769 F1 Score = 0.7

By this results, recall is not good evaluation metric it needs to improve more. Two conditions are very risk here:

1) False Positive: Suppose if we see the above confusion matrix for the heart attack cases then a person is predicted with heart attack but does not have heart attack in reality.

2) False Negative: A person is predicted with no heart attack but in reality person is suffering from the heart attack.

Hence False negative miss the actual patients. So recall needs more improvement.

Activity: 4

In schools, a lot of times it happens that there is no water to drink. At a few places, cases of water shortage in schools are very common and prominent. Hence, an AI model is designed to predict if there is going to be a water shortage in the school in the near future or not. The confusion matrix for the same is:

The Confusion Matrix	Reality: 1	Reality: 0
Prediction: 1	22 TP	12 FP
Prediction: 0	47 FN	118 TN

Find out Accuracy, Precision, Recall and F1 Score for the given problems.

Ans:

Calculation:

$$1. \quad \text{Accuracy} = \frac{(TP+TN)}{(TP+FN+TN+FP)} \times 100\%$$

$$\frac{(22+118)}{(22+47+118+12)} \times 100\% = \frac{140}{199} \times 100\% = 0.703$$

$$2. \quad \text{Precision} = \frac{TP}{(FP+TP)} = \frac{22}{(12+22)} = 0.647$$

$$3. \quad \text{Recall} = \frac{TP}{(FN+TP)} = \frac{22}{(47+22)} = 0.318$$

$$4. \quad \text{F1 Score} = 2X \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$$

$$2X \frac{(0.647 \times 0.318)}{(0.647 + 0.318)} = 2X \frac{0.205746}{0.965} = 2X0.213 = 0.426$$

Accuracy = 0.703

Precision = 0.647

Recall = 0.318

F1 Score = 0.426