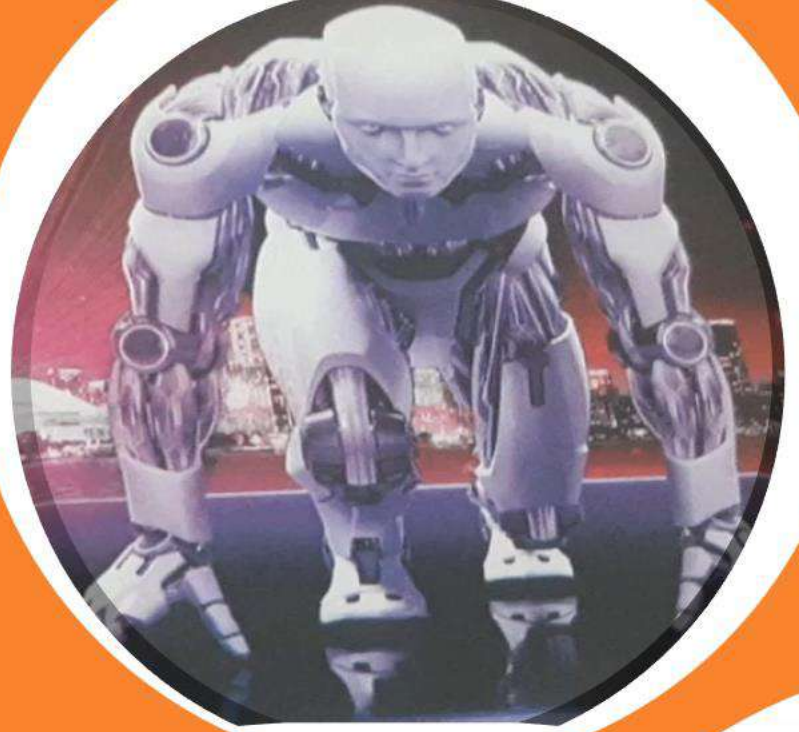
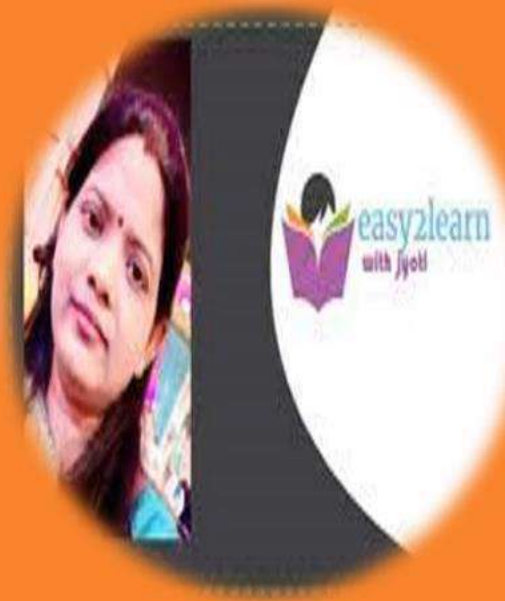
AI-417

CLASS – 10

CLASS- IX AI-417





EVALUATION

ARTIFICIAL INTELLIGENCE

EVALUATION

- ✓**Evaluation** refers to systematically checking and analyzing the merit, correctness, and reliability of an AI model based on the outputs produced by it.

- ✓Aim of an AI model to produce results for any unknown data, based on its learning.

Causes behind the Performance of an AI model

1. Overfitting –

- Overfitting refers to a situation when an AI model performs so well as the test data it got, fitted exactly against its training data, and thus AI model always produced the correct result.
- However, the AI model fails to fit the unknown data, i.e., can't predict reliable results for unknown data.
- An overfitted model will appear to have a higher accuracy when we apply it to the training data.
- The model developers may take it as a highly accurate model, whereas in reality, it will underperform in production when given new data.

2. UNDERFITTING

- Underfitting is the opposite of overfitting.

- Underfitting refers to a situation when an AI model is not complex enough to accurately capture the structure and relationships of training data and predict effective outcomes.

- An underfitted model results in a problematic or erroneous outcome on new data, or data not the same as training data, and often performs poorly even on training data.

overfitting and underfitting are the two biggest causes of the poor performance of AI models and thus, model developers employ certain protection methods to control these two.

3. GENERALIZATION -

Generalization refers to how well the concept learned by a machine learning model applies to specific examples not seen by the model when it was learning.

✓To ensure that the AI model is accurate, smart, and good at learning **the AI model is evaluated through various evaluation matrices**.

Model Evaluation Metrics

CONFUSION MATRIX

A Confusion Matrix is a technique using a chart of the table for summarizing the performance of a classification-based AI model by listing the predicted values of an AI model and the actual/correct outcome values.

PREDICTED VALUE

ACTUAL VALUE

	Positive	Negative	
True	True Positive (TP)	False Negative (FN) Type II Error	Accuracy $\frac{TP + TN}{TP + TN + FP + FN}$
False	False Positive (FP) Type I Error	True Negative (TN)	F1 Score $\frac{TP}{TP+\frac{1}{2}(FP+FN)}$
	Precision $\frac{TP}{TP+FP}$	Recall $\frac{TP}{TP + FN}$	F1 Score= $2 * \frac{PERCISION * RECALL}{PRECISION + RECALL}$

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, otherwise it is False.
- ✓ If the Prediction is Yes then it is Positive and
- ✓ If the Prediction is No then it is Negative.

A Confusion Matrix is a technique using a chart or table (N X N matrix) for summarizing the **performance of a classification-based AI** model by listing the predefined values of an AI model and the actual/correct outcome values.

The **Actual Value** represents the actual result (observed or measured)

True ——— **False**

The **Predicted Value** is the value of the outcome/result of the AI model, produced on the basis of its algorithm and learning.

Positive ——— **Negative**

Using the Confusion Matrices, we need to compute the following values to evaluate an AI model :

1. **Accuracy rate –All True cases, i.e. TP + TN) out of all the observations or tests conducted (i.e. TP + FP +TN +FN).**

The formula to determine Accuracy Rate is :

$$\text{Accuracy} = \frac{Number\ of\ correct\ predictions\ (TP + TN)}{Total\ Number\ of\ predictions\ made\ (TP + TN + FP + FN)} X\ 100\%$$

		Predicted Value	
		Positive	Negative
Actual Value	True	True Positive (TP)	False Negative (FN) Type II Error
	False	False Positive (FP) Type I Error	True Negative (TN)

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

2. Precision rate – True Positives out of all positive instances.

Thus, the formula for Precision rate is : $\frac{TP}{(TP+FP)}$

In percentage, Precision rate is : $\frac{TP}{(TP+FP)} X\ 100\%$

In precision, FN cases will not be taken into account.

		Predicted Value	
		Positive	Negative
Actual Value	True	True Positive (TP)	False Negative (FN) Type II Error
	False	False Positive (FP) Type I Error	True Negative (TN)

3. Recall – It is a rate of correct positive predictions to the overall number of positive instances in the dataset.

$$\textbf{Recall} = \frac{Predictions\ actually\ positive(TP)}{Actual\ positive\ values\ in\ the\ dataset\ (TP+FN)} = \frac{TP}{(TP+FN)}$$

In recall, FP cases will not be taken into account

		Predicted Value	
		Positive	Negative
Actual Value	True	True Positive (TP)	False Negative (FN) Type II Error
	False	False Positive (FP) Type I Error	True Negative (TN)

Consider the example given below-

- For an AI model developed to check for gas leakage for a manufacturing unit, **False Negatives** (i.e. it shows no gas leakage while in reality, it is) will be highly risky and may cost lives and money.

- For an AI model to predict the onset of calamities like flood danger etc. for a Dam management group, a **False Negative** will result in huge losses of money, resources, and human lives.

- For an AI model developed to locate the position of a lump not visible through normal diagnosis, a **False Positive** for a body part may lead to a wrong operation/surgery and the patient's life may be in danger.

- For an AI model developed to test if a customer can pay back their loan considering various parameters before approving their loan application, a **False Positive** will result in turning down the application of a valid and rightful applicant.

To minimize such cases, Precision and Recall metrics are used.

- ✓ **Precision is used as a metric when our objective is to minimize false positives**
- ✓ **Recall is used as a metric when our objective is to minimize false negatives.**

Note - Precision and Recall are inversely proportional to each other so we can't increase the value of these metrics simultaneously.

Thus, we have to decide which is more important in our situation or depending on our case/problem.

Once we decide which metric is more important, we can optimize our model performance on the selected metric.

For this purpose, another metric, F1 Score, is useful for striking a balance between Precision and Recall values.

F1 Score (F Measure) – F1 Scores refers to a metric that balances Precision and Recall and hence balances the impact of False Positives and false negatives. It is computed as per the following formula :

$$F1 = 2 \times \frac{PRECISION * RECALL}{PRECISION + RECALL} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

Note-

1. **The Precision, Recall, and F1 score range from 0 to 1.**
2. **F1 Score is the best-suited parameter to test this AI model, which is the balance between Precision and Recall.**

		Predicted Value	
		Positive	Negative
Actual Value	True	True Positive (TP)	False Negative (FN) Type II Error
	False	False Positive (FP)	True Negative (TN)

Precision and Recall impact on the F1 score as listed in the table

Precision	Recall	F1 Score
Low	Low	Low
Low	High	Low
High	Low	Low
High	High	High

Note – F1 Score is very useful when we need to combine two or more classification AI models for the same data.

We opt for the AI model whose F1 score is higher.

Practice Time

1.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.

Q. 21	Automated trade industry has developed an AI model which predicts the selling and purchasing of automobiles. During testing, the AI model came up with the following predictions.	4

Confusion Matrix		Reality	
		Yes	No
Predicted	Yes	60	25
	No	05	10

(i) How many total tests have been performed in the above scenario?

(ii) Calculate precision, recall and F1 Score.

		Actual/Reality Value	
		YES	NO
Predicted Value	YES	True Positive (TP)	False Negative (FP) Type I Error
	NO	False Positive (FN) Type II Error	True Negative (TN)

(i) TP=60, TN=10, FP=25, FN=5

TP+ TN + FP + FN = 60+25+5+10=100 total cases have been performed

(ii) (Note: For calculating Precision, Recall and F1 score, we need not multiply the formula by 100 as all these parameters need to range between 0 to 1)

Precision =TP/(TP+FP) =60/(60+25) =60/85 =0.7

Recall=TP/(TP+FN) =60/(60+5) =60/65 =0.92

F1 Score = 2*Precision*Recall/(Precision + Recall)

F1 Score = 2*0.7*0.92/(0.7+0.92) = 0.79

		Actual/Reality Value	
		True	False
Predicted Value	Positive	TP 60	FP 25
	Negative	FN 5	TN 10