EVALUATION

OBJECTIVE QUESTIONS (SET 01)

1. _____ is the process of understanding the reliability of any AI model, based on outputs by feeding test dataset into the model and comparing with actual answers.
a. Evaluation
b. Problem Scoping
c. Data acquisition
d. Data Exploration
Ans: a. Evaluation

2. If model will simply remember the whole training set, and will therefore always predict the correct label for any point in the training set. This is known as _____.
a. Overfitting
b. Overriding
c. Over remembering
d. None of the above
Ans: a. Overfitting

3. The result of comparison between the prediction and reality can be recorded in what we call the _____.
a. Overfitting
b. Problem Scoping
c. Confusion Matrix
d. Data acquisition
Ans: c. Confusion Matrix

4. The _____ allows us to understand the prediction results.
a. Overfitting
b. Problem Scoping
c. Confusion Matrix
d. Data acquisition
Ans: c. Confusion Matrix

5. _____ is defined as the percentage of correct predictions out of all the observations.
a. Overfitting
b. Accuracy
c. Confusion Matrix
d. Data acquisition
Ans: b. Accuracy

6. _____ is defined as the percentage of true positive cases versus all the cases where the prediction is true.
a. Overfitting
b. Accuracy
c. Precision
d. Data acquisition
Ans: c. Precision

7. _____ can be defined as the fraction of positive cases that are correctly identified.
a. Recall
b. Accuracy
c. Precision

d. Data acquisition
Ans: a. Recall

8. _____ can be defined as the measure of balance between precision and recall.
a. Recall
b. Accuracy
c. Precision
d. F1 Score
Ans: d. F1 Score

QUESTIONS AND ANSWERS (SET 01) - 1 mark

1. Define Evaluation.
Moving towards deploying the model in the real world, we test it in as many ways aspossible. The stage of testing the models is known as EVALUATION.

OR

Evaluation is a process of understanding the reliability of any AI model, based on outputs by feeding the test dataset into the model and comparing it with actual answers.

OR

Evaluation is a process that critically examines a program. It involves collecting and analyzing information about a program's activities, characteristics, and outcomes. Its purpose is to make judgments about a program, to improve its effectiveness, and/or to inform programming decisions.

2. Which two parameters are considered for Evaluation of a model?
Prediction and Reality are the two parameters considered for Evaluation of a model.
The "Prediction" is the output which is given by the machine and the "Reality"is the real scenario, when the prediction has been made?

3. What is True Positive?
- The predicted value matches the actual value
- The actual value was positive and the model predicted a positive value

4. What is True Negative?
- The predicted value matches the actual value
- The actual value was negative and the model predicted a negative value

5. What is False Positive?
- The predicted value was falsely predicted
- The actual value was negative but the model predicted a positive value
- Also known as the Type 1 error

6. What is False Negative?
- The predicted value was falsely predicted
- The actual value was positive but the model predicted a negative value
- Also known as the Type 2 error

QUESTIONS AND ANSWERS (SET 01) - 2 marks

1. What is meant by Overfitting of Data?
Overfitting is "the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably".

(OR)

An Overfitted Model is a statistical model that contains more parameters than can be justified by the data. Here, to evaluate the AI model it is not necessary to use the data that is used to build the model.

Because AI Model remembers the whole training data set, therefore it always predicts the correct label for any point in the training dataset. This is known as Overfitting

(OR)

Models that use the training dataset during testing, will always results in correct output. This is known as overfitting.

2. What is Accuracy? Mention its formula.

Accuracy is defined as the percentage of correct predictions out of all theobservations.
A prediction is said to be correct if it matches reality. Here we have twoconditions in which the Prediction matches with the Reality, i.e., True Positive and True Negative. Therefore, Formula for Accuracy is

$$Accuracy = \frac{Correct\ prediction}{Total\ cases} * 100\%$$

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

Where *TP* = True Positives, *TN* = True Negatives, *FP* = False Positives, and *FN* = False Negatives.

3. What is Precision? Mention its formula.

Precision is defined as the percentage of true positive cases versus all the cases where the prediction is true.
That is, it takes into account the True Positives and False Positives.

$$Precision = \frac{True\ Positive}{All\ Predicted\ Positives} * 100\%$$

$$Precision = \frac{TP}{TP + FP} * 100\%$$

4. What is Recall? Mention its formula.

Recall is defined as the fraction of positive cases that are correctly Identified.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

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

5. Why is evaluation important? Explain.

Importance of Evaluation
Evaluation is a process that critically examines a program. It involves collecting and analyzing information about a program's activities, characteristics, and outcomes. Its purpose is to make judgments about a program, to improve its effectiveness, and/or to inform programming decisions.

- Evaluation is important to ensure that the model is operating correctly and optimally.
- Evaluation is an initiative to understand how well it achieves its goals.
- Evaluations help to determine what works well and what could be improved in a program

6. How do you suggest which evaluation metric is more important for any case?

F1 Evaluation metric is more important in any case. F1 score sort maintains a balance between the precision and recall for the classifier. If the precision is low, the F1 is low and if the recall is low again F1 score is low.
The F1 score is a number between 0 and 1 and is the harmonic mean of precision and recall

When we have a value of 1 (that is 100%) for both Precision and Recall. The F1 score would also be an ideal 1 (100%). It is known as the perfect value for F1 Score. As the values of both Precision and Recall ranges from 0 to 1, the F1 score also ranges from 0 to 1.

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

7. Which evaluation metric would be crucial in the following cases? Justify your answer.
a. Mail Spamming
b. Gold Mining
c. Viral Outbreak
Here, Mail Spamming and Gold Mining are related to FALSE POSITIVE cases which are expensive at cost. But Viral Outbreak is a FALSE NEGATIVE case which infects a lot of people on health and leads to expenditure of money too for checkups.
So, False Negative case (VIRAL OUTBREAK) are more crucial and dangerous when compared to FALSE POSITIVE cases.
(OR)

a. If the model always predicts that the mail is spam, people would not look at it and eventually might lose important information. False Positive condition would have a high cost. (predicting the mail as spam while the mail is not spam)

b. A model saying that there exists treasure at a point and you keep on digging there but it turns out that it is a false alarm. False Positive case is very costly.
(predicting there is a treasure but there is no treasure)
c. A deadly virus has started spreading and the model which is supposed to predict a viral outbreak does not detect it. The virus might spread widely and infect a lot of people. Hence, False Negative can be dangerous

8. What are the possible reasons for an AI model not being efficient? Explain.
Reasons of an AI model not being efficient:
a. Lack of Training Data: If the data is not sufficient for developing an AI Model, or if the data is missed while training the model, it will not be efficient.
b. Unauthenticated Data / Wrong Data: If the data is not authenticated and correct, then the model will not give good results.

c. Inefficient coding / Wrong Algorithms: If the written algorithms are not correct and relevant, Model will not give desired output. Not Tested: If the model is not tested properly, then it will not be efficient.

d. Not Easy: If it is not easy to be implemented in production or scalable.

e. Less Accuracy: A model is not efficient if it gives less accuracy scores in production or test data or if it is not able to generalize well on unseen data.
(Any three of the above can be selected)

9. Answer the following:
- **Give an example where High Accuracy is not usable.**

SCENARIO: An expensive robotic chicken crosses a very busy road a thousand times per day. An ML model evaluates traffic patterns and predicts when this chicken can safely cross the street with an accuracy of 99.99%.
Explanation: A 99.99% accuracy value on a very busy road strongly suggests that the ML model is far better than chance. In some settings, however, the cost of making even a small number of mistakes is still too high. 99.99% accuracy means that the expensive chicken will need to be replaced, on average, every 10 days. (The chicken might also cause extensive damage to cars that it hits.)

- **Give an example where High Precision is not usable.**

Example: “Predicting a mail as Spam or Not Spam”

False Positive: Mail is predicted as “spam” but it is “not spam”.
False Negative: Mail is predicted as “not spam” but it is “spam”.
Of course, too many False Negatives will make the spam filter ineffective but False Positives may cause important mails to be missed and hence Precision is not usable.

QUESTIONS AND ANSWERS (SET 01) - 3/4 marks

1. Deduce the formula of F1 Score? What is the need of its formulation?

The F1 Score, also called the F score or F measure, is a measure of a test’s accuracy.
It is calculated from the precision and recall of the test, where the precision is the number of correctly identified positive results divided by the number of all positive results, including those not identified correctly, and the recall is the number of correctly identified positive results divided by the number of all samples that should have been identified as positive.
The F1 score is defined as the weighted harmonic mean of the test’s precision and recall. This score is calculated according to the formula.
Formula:

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

Necessary:
F-Measure provides a single score that balances both the concerns of precision and recall in one number.
A good F1 score means that you have low false positives and low false negatives, so you’re correctly identifying real threats, and you are not disturbed by false alarms.
An F1 score is considered perfect when it’s 1, while the model is a total failure when it’s 0.
F1 Score is a better metric to evaluate our model on real-life classification problems and when imbalanced class distribution exists.

2. What is a confusion matrix? Explain in detail with the help of an example.

Confusion Matrix:
A Confusion Matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

(OR)

A 2x2 matrix denoting the right and wrong predictions might help us analyse the rate of success. This matrix is termed the Confusion Matrix.
Evaluation of the performance of a classification model is based on the counts of test records correctly and incorrectly predicted by the model.
Therefore, Confusion Matrix provides a more insightful picture which is not only the performance of a predictive model, but also which classes are being predicted correctly and incorrectly, and what type of errors are being made.
The confusion matrix is useful for measuring Recall (also known as Sensitivity), Precision, Accuracy and F1 Score.
The following confusion matrix table illustrates how the 4-classification metrics are calculated (TP, FP, FN, TN), and how our predicted value compared to the actual value in a confusion matrix

The Confusion Matrix		Reality	
		Yes	No
Prediction	Yes	True Positive (TP)	False Positive (FP)
	No	False Negative (FN)	True Negative (TN)

Let’s decipher the matrix:
The target variable has two values: Positive or Negative
The columns represent the actual values of the target variable
The rows represent the predicted values of the target variable
True Positive, True Negative, False Positive and False Negative in a Confusion Matrix
True Positive (TP)
The predicted value matches the actual value
The actual value was positive and the model predicted a positive value
True Negative (TN)
The predicted value matches the actual value
The actual value was negative and the model predicted a negative value
False Positive (FP) – Type 1 error
The predicted value was falsely predicted
The actual value was negative but the model predicted a positive value ● Also known as the Type 1 error
False Negative (FN) – Type 2 error
The predicted value was falsely predicted
The actual value was positive but the model predicted a negative value also known as the Type 2 error
Example:
Case: Loan (Good loan & Bad loan)

The result of TP will be that bad loans are correctly predicted as bad loans.
The result of TN will be that good loans are correctly predicted as good loans.
The result of FP will be that (actual) good loans are incorrectly predicted as bad loans.
The result of FN will be that (actual) bad loans are incorrectly predicted as good loans.
The banks would lose a bunch of money if the actual bad loans are predicted as good loans due to loans not being repaid. On the other hand, banks won't be able to make more revenue if the actual good loans are predicted as bad loans. Therefore, the cost of False Negatives is much higher than the cost of False Positives.

3. Calculate Accuracy, Precision, Recall and F1 Score for the following Confusion Matrix on Heart Attack Risk. Also suggest which metric would not be a good evaluation parameter here and why?

The Confusion Matrix	Reality: 1	Reality: 0
Prediction: 1	50	20
Prediction: 0	10	20

The Confusion Matrix	Reality: 1	Reality: 0	
Prediction: 1	50	20	70
Prediction: 0	10	20	30
	60	40	100

Calculation:
Accuracy:
Accuracy is defined as the percentage of correct predictions out of all the observations

$$Accuracy = \frac{Correct\ prediction}{Total\ cases} * 100\%$$

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

Where True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).
Accuracy = (50+20) / (50+20+20+10)
= (70/100)
= 0.7
Precision:
Precision is defined as the percentage of true positive cases versus all the cases where the prediction is true.

$$Precision = \frac{True\ Positive}{All\ Predicted\ Positives} * 100\%$$

$$Precision = \frac{TP}{TP + FP} * 100\%$$

= (50 / (50 + 20))
= (50/70)
= 0.714
Recall: It is defined as the fraction of positive cases that are correctly identified.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

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

= 50 / (50 + 60)
= 50 / 110
= 0.5

F1 Score:
F1 score is defined as the measure of balance between precision and recall.

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

= 2 * (0.714 *0.5) / (0.714 + 0.5)
= 2 * (0.357 / 1.214)
= 2* (0.29406)
= 0.58
Therefore,
Accuracy= 0.7 Precision=0.714 Recall=0.5
F1 Score=0.588
Here within the test there is a tradeoff. But Recall is not a good Evaluation metric. Recall metric needs to improve more.
Because,
False Positive (impacts Precision): A person is predicted as high risk but does not have heart attack.
False Negative (impacts Recall): A person is predicted as low risk but has heart attack.
Therefore, False Negatives miss actual heart patients, hence recall metric need more improvement.
False Negatives are more dangerous than False Positives.

4. Calculate Accuracy, Precision, Recall and F1 Score for the following Confusion Matrix on Water Shortage in Schools: Also suggest which metric would not be a good evaluation parameter here and why?

The Confusion Matrix (Water Shortage inSchool)	Reality: 1	Reality: 0
Prediction: 1	75	5
Prediction: 0	5	15

	Reality: 1	Reality: 0	
Prediction: 1	75	5	80
Prediction: 0	5	15	20
	80	20	100

Calculation:

Accuracy
Accuracy is defined as the percentage of correct predictions out of all the observations

$$Accuracy = \frac{Correct\ prediction}{Total\ cases} * 100\%$$

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

Where True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).
= (75+15) / (75+15+5+5)
= (90 / 100)
=0.9
Precision:
Precision is defined as the percentage of true positive cases versus all the cases where the prediction is true.

$$Precision = \frac{True\ Positive}{All\ Predicted\ Positives} * 100\%$$

$$Precision = \frac{TP}{TP + FP} * 100\%$$

= 75 / (75+5)
= 75 /80
= 0.9375
Recall:
It is defined as the fraction of positive cases that are correctly identified.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

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

= 75 / (75+5)
= 75 /80
= 0.9375
F1 Score:
F1 score is defined as the measure of balance between precision and recall.

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

= 2 * ((0.9375 *0.9375) / (0.9375+0.9375)
Therefore,
= 2 * (0.8789 / 1.875)
= 2 * 0.46875
= 0.9375
Accuracy= 0.9% Precision=0.9375% Recall=0.9375%
F1 Score=0.
Here precision, recall, accuracy, F1 score all are same

5. Calculate Accuracy, Precision, Recall and F1 Score for the following Confusion Matrix on SPAM FILTERING: Also suggest which metric would not be a good evaluation parameter here and why?

Confusion Matrix on SPAM FILTERING:	Reality: 1	Reality: 0
Prediction: 1	10	55
Prediction: 0	10	25

Confusion Matrix on SPAM FILTERING:	Reality: 1	Reality: 0	
Prediction:1	10	55	65
Prediction: 0	10	25	35
	20	80	100

Accuracy is defined as the percentage of correct predictions out of all the observations

$$Accuracy = \frac{Correct\ prediction}{Total\ cases} * 100\%$$

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

Where True Positive (TP), True Negative (TN), False Positive (FP) and False Negative(FN).
= (10 + 25) / (10+25+55+10)
= 35 / 100
= 0.35
Precision:
Precision is defined as the percentage of true positive cases versus all the cases wherethe prediction is true.

$$Precision = \frac{True\ Positive}{All\ Predicted\ Positives} * 100\%$$

$$Precision = \frac{TP}{TP + FP} * 100\%$$

= 10 / (10 +55)
= 10 /65
= 0.15
Recall:
It is defined as the fraction of positive cases that are correctly identified.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

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

= 10 / (10 + 10)
= 10 / 20
= 0.5

F1 Score:
F1 score is defined as the measure of balance between precision and recall.

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

= 2 * ((0.15 * 0.5) / (0.15 + 0.5))
= 2 * (0.075 / 0.65)
= 2 * 0.115
= 0.23
Therefore,
Accuracy= 0.35
Precision= 0.15
Recall= 0.5
F1 Score= 0.23

Here within the test there is a tradeoff. But Precision is not a good Evaluation metric.Precision metric needs to improve more.

Because, False Positive (impacts Precision): Mail is predicted as "spam" but it is not.False Negative (impacts Recall): Mail is predicted as "not spam" but spam Of course, too many False Negatives will make the Spam Filter ineffective. But False Positives may cause important mails to be missed. Hence, Precision is more importantto improve

QUESTIONS AND ANSWERS (SET 02)

1. What is evaluation?
Answer – By feeding test datasets into AI models and comparing the outputs to real-world results, evaluation is the process of determining the dependability of any AI model is known as evaluation. Various evaluation methods may be used, depending on the kind and function of the model. Keep in mind that it is not advised to utilize the model's construction data for its evaluation.

2. What is overfitting?
Answer – Overfitting happens when a statistical model matches its training data exactly. When this occurs, the algorithm's goal is lost because it is unable to accurately execute against unseen data.

3. What is Confusion Matrix?
Answer – The result of comparison between the prediction and reality can be recorded in what we call the confusion matrix. The confusion matrix allows us to understand the prediction results. Note that it is not an evaluation metric but a record which can help in evaluation.

4. What is the purpose of Accuracy in AI and give the example of equation?
Answer – The percentage of accurate predictions among all the observations is what is meant by the term accuracy. A prediction is deemed accurate if it agrees with reality. There are two circumstances in this case where the Prediction and Reality match: True Positive and True Negative. The equation for accuracy is –

$$Accuracy = \frac{Correct\ prediction}{Total\ cases} * 100\%$$

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

$$Precision = \frac{True\ Positive}{All\ Predicted\ Positives} * 100\%$$

$$Precision = \frac{TP}{TP + FP} * 100\%$$

5. What is Precision and give the example of equation?
Answer – Precision is defined as the percentage of true positive cases versus all the cases where the prediction is true. That is, it takes into account the True Positives and False Positives.

6. What is Recall in AI and give the example of equation?
Answer – The percentage of pertinent documents that are successfully retrieved is known as recall. Recall, for instance, is the ratio of the number of accurate results to the number of results that should have been returned for a text search on a set of documents.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

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

7. What is F1 Score and give the example of equation?
Answer – F1 score can be defined as the measure of balance between precision and recall. By

calculating the harmonic mean of a classifier's precision and recall, the F1-score integrates both into a single metric. It mainly used to compare the effectiveness of two classifiers.

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

8. Write the stages of Al project cycle? (1 Mark)
Answer : Al Project Cycle stages are - (i) Problem Scoping (ii) Data Acquisition (iii) Data Exploration (iv) Modelling (v) Evaluation

9. What is Evaluation Stage? (1 Mark)
Answer : The evaluation stage follows modeling stage. In Evaluation stage, the desired model is trained with the training data set and ready for testing with the help of testing data set.

10. What is Testing data set? (1 Mark)
Answer : Testing data set is use to test the model before the complete deployment. Testing data set is prepared by the trained professionals after exploring and cleaning the raw big data. Testing data set is a completely different, separate and new entity for the model to be tested.

11. What is Scenario? (1 Mark)
Answer : The problem area for which a model has been developed is called scenario. Scenario is the reality in which the real problem exists and the model has to be developed.

12. Define Scenario with reference to evaluation of Al model? (2 Marks)
Answer : Scenario is the source of real data which is fed into the model for processing either at regular intervals (hourly, daily, weekly, etc.) or in real time. Scenario is the reality in which the real problem exists and the model has to be developed. The model has to deal with the scenario the way it has been trained.

13. How many types of Scenario? (2 Mark)
Answer : There are two types of Scenario. (i) A regular-interval Scenario, and (ii) Real-time Scenario

14. What is Regular-interval Scenario? Give example. (2 Mark)
Answer : A regular-interval Scenario can be a less critical problem area in which emergent threat is not there. For example, pollution monitoring in a region needs weekly or monthly data, studying a cancer patient for research, monitoring the diet and fitness of sports persons, monitoring performance of students and their study habits, etc.