Q. e.1 How can I measure the performance of my model?

Ans. e.1. The model's performance can be evaluated using various measures, which are specific to the type and purpose of a model. For regression models this is measured by mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and R-squared while for classification models it is usually done through accuracy, precision, recall, and Fuzzy F1 score. Clustering models use metrics such as the Davies–Bouldin index and silhouette scores for evaluation. Cross-validation methods as well as domain-specific metrics give more insights about the generalization and applicability of the model. The selection of appropriate metrics that correspond to the goals of data, and uniquely measure other factors affecting model efficiency allows an accurate evaluation of model effectivity. I therefore have many ways to evaluate my model depending on its type (classification or regression) and aim (e.g., picture recognition, sales forecast). Such measures include but are not limited to accuracy, precision, recall, F1 score, MSE, RMSE MAE, and R^2. Therefore, I must consider various indicators; compare my model with baselines; check how well it generalizes using a different test set.

Q. e.2 What are: Accuracy, Confusion Matrix, Precision, Recall & F1 Score, ROC & AUC, Log Loss?

Ans. e.2. The proportion of correctly classified out of the total instances is what accuracy measures. Precision, on the other hand, determines how good are positive predictions when it looks at true positives against all those predicted to be positive. Recall or sensitivity tests for this case compare correctly detected true positives among all actual positives. The F1 score goes between precision and recall, particularly helpful with imbalanced classes - it is the harmonic mean of precision and recall. True positive, true negative, false positive number, and false negative count provide a picture of model performance (confusion matrix). ROC Curve presents a binary classifier's performance across many thresholds illustrating a trade-off between true positive rate and false positive rate, where AUC summarizes this performance over all thresholds. Lastly, log loss simply measures the difference between actual outcomes as well as predicted probabilities.

Definitions:

$$\boldsymbol{Accuracy} = \frac{\textbf{Number of Correct Predictions}}{\textbf{Total Number of Predictions}}$$

$$\boldsymbol{Precision} = \frac{\textbf{True Positives }(\boldsymbol{TP})}{\textbf{True Positives }(\boldsymbol{TP}) + \textbf{False Positives }(\boldsymbol{FP})}$$

$$\boldsymbol{Recall} = \frac{\textbf{True Positives }(\boldsymbol{TP})}{\textbf{True Positives }(\boldsymbol{TP}) + \textbf{False Negatives }(\boldsymbol{FN})}$$

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

Confusion Matrix:

	Actual Positive	**Actual Negative**
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

Logistic Regression (Fertility Dataset) Analysis: The logistic regression model has a decent accuracy of 90%, indicating that it correctly predicts the class for 90% of the instances. However, looking at the confusion matrix, it predicts only one class (class 0) and completely misses the other class (class 1). This results in precision, recall, and F1 score of 0, indicating that the model fails to classify any instances of class 1 correctly. The ROC AUC of 0.5 suggests that the model's ability to distinguish between classes is no better than random guessing.