Company Name - EY
Role - AI/ML engineer

Q. Kernels in SVM?

A. Kernel Function is a method used to take data as input and transform into the required form of processing data. “Kernel” is used due to set of mathematical functions used in Support Vector Machine provides the window to manipulate the data.

Q. Overfitting and Underfitting? How do you handle them?

A. Your model is underfitting the training data when the model performs poorly on the training data. Your model is overfitting your training data when you see that the model performs well on the training data but does not perform well on the evaluation data.
To avoid overfitting: Keep your model simple, Use regularization technique, Use cross-validation. To avoid underfitting: Decrease regularization, increase the training duration.

Q. P value? Why threshold is 0.05 or less?

A. A p-value is a measure of the probability that an observed difference could have occurred just by random chance. The lower the p-value, the greater the statistical significance of the observed difference. P-value can be used as an alternative to or in addition to pre-selected confidence levels for hypothesis testing. A p-value less than 0.05 is statistically significant. It indicates strong evidence against the null hypothesis, as there is less than a 5% probability the null is correct (and the results are random). Therefore, we reject the null hypothesis, and accept the alternative hypothesis.

Q. Regularization and its use?

A. Regularization is a technique used for tuning the function by adding an additional penalty term in the error function. Regularization, significantly reduces the variance of the model, without substantial increase in its bias.

Q. Type 1 and type 2 error in confusion matrix?

A. A type I error is the mistaken rejection of the null hypothesis (also known as a "false positive" finding or conclusion; example: "an innocent person is convicted"), while a type II error is the mistaken acceptance of the null hypothesis (also known as a "false negative" finding or conclusion; example: "a guilty person is not convicted").

Q. How root node is selected in decision tree.

A. While building the decision tree, we would prefer choosing the attribute/feature with the least Gini index as the root node.

Q. Std deviation vs variance?

A. Variance is the average squared deviations from the mean, while standard deviation is the square root of this number. Both measures reflect variability in a distribution, but their units differ:

- Standard deviation is expressed in the same units as the original values (e.g., minutes or meters).
- Variance is expressed in much larger units (e.g., meters squared).

Variance helps to find the distribution of data in a population from a mean, and standard deviation also helps to know the distribution of data in population, but standard deviation gives more clarity about the deviation of data from a mean.

Company Name - EY
Role - Data Scientist

Q. What is Yolo?

A. YOLO - You Only Look Once is an algorithm proposed by by Redmond et. al in a research article published at the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) as a conference paper, winning OpenCV People's Choice Award.

Compared to the approach taken by object detection algorithms before YOLO, which repurpose classifiers to perform detection, YOLO proposes the use of an end-to-end neural network that makes predictions of bounding boxes and class probabilities all at once.

Q. Select perfect k for k means?

A. There is a popular method known as elbow method which is used to determine the optimal value of K to perform the K-Means Clustering Algorithm. The basic idea behind this method is that it plots the various values of cost with changing k. As the value of K increases, there will be fewer elements in the cluster.

Q. Model metrics when you have outliers?

A. The bigger the MAE, the more critical the error is. It is robust to outliers. Therefore, by taking the absolute values, MAE can deal with the outliers.

Q. Face recognition?

A. Facial recognition is a way of identifying or confirming an individual's identity using their face.

Q. Features for food delivery data to give discount to selected customers?

A. The features for same can be: Total number of orders, Frequency of ordering per week, Amount paid per order, Distance travelled by delivery man etc.

Q. What is RMSE and MSE?

A. MSE (Mean Squared Error) represents the difference between the original and predicted values which are extracted by squaring the average difference over the data set. It is a measure of how close a fitted line is to actual data points. The lesser the Mean Squared Error, the closer the fit is to the data set. The MSE has the units squared of whatever is plotted on the vertical axis. RMSE (Root Mean Squared Error) is the error rate by the square root of MSE. RMSE is the most easily interpreted statistic, as it has the same units as the quantity plotted on the vertical axis or Y-axis. RMSE can be directly interpreted in terms of measurement units, and hence it is a better measure of fit than a correlation coefficient.

Q. What is Specificity?

A. Specificity (SP) is calculated as the number of correct negative predictions divided by the total number of negatives.
