Q1 Anomaly detection identifies suspicious activity that falls outside of your established normal patterns of behavior. A solution protects your system in real-time from instances that could result in significant financial losses, data breaches, and other harmful events. Anomaly detection can be used to improve the quality of datasets by identifying and removing outliers. Outliers are data points that stand out from the rest of the dataset.By removing outliers, anomaly detection can improve the accuracy of machine learning models.

Q2 Data quality issues and small training samples also make anomaly detection algorithms less effective. Without a high-quality dataset to reference, the system develops unreliable anomaly detection, meaning that the model can miss glaring outliers.

Q3 Supervised anomaly detection is ideal for scenarios which contain distinct types of anomalies that can be labeled and learned from. Unsupervised anomaly detection is best for scenarios without labels or when the anomalies are unknown or ever-changing. If we have a labeled target and are using supervised learning techniques, we can just calculate the precision and recall on the validation data to evaluate model performance. If we are using unsupervised anomaly detection techniques, we will just have normal operating data with no labeled anomalies.

Q4 Main Categories of Anomolt detection:- 1.Unsupervised Clustering:-An unsupervised learning strategy should be used for data lacking prior knowledge, especially when the data points have not been pre-labeled as normal or pathological. 2.Supervised Classification:-This method requires pre-labeled data that are classified as normal or abnormal or even specific known categories of abnormal behavior. It supports both normality and abnormality modeling. 3.Semi-supervised Detection. This focuses solely on modeling normalcy, necessitating either pre-classified data that has been designated as normal, or the presumption that the training set exclusively consists of normal data.

Q5 Distance-based outlier detection method consults the neighbourhood of an object, which is defined by a given radius. An object is then considered an outlier if its neighborhood does not have enough other points. A distance the threshold that can be defined as a reasonable neighbourhood of the object.

Q6 The Local Outlier Factor (LOF) algorithm is an unsupervised anomaly detection method which computes the local density deviation of a given data point with respect to its neighbors. It considers as outliers the samples that have a substantially lower density than their neighbors.The LOF of a point is based on the ratios of the local density of the area around the point and the local densities of its neighbors. It considers relative density of data points. In simple words, LOF compares the local density of a point to local density of its k-nearest neighbors and gives a score as final output.

Q7 1.When given a dataset, a random sub-sample of the data is selected and assigned to a binary tree. 2.Branching of the tree starts by selecting a random feature (from the set of all N

features) first. And then branching is done on a random threshold (any value in the range of minimum and maximum values of the selected feature). 3.If the value of a data point is less than the selected threshold, it goes to the left branch else to the right. And thus a node is split into left and right branches. 4.This process from step 2 is continued recursively till each data point is completely isolated or till max depth(if defined) is reached. 5.The above steps are repeated to construct random binary trees.

Q8 Anomaly detection is the identification of rare events, items, or observations which are suspicious because they differ significantly from standard behaviors or patterns. Anomalies in data are also called standard deviations, outliers, noise, novelties, and exceptions. The range is computed from the standard deviation std from the n lastest points. Thus, the range of a non-anomaly is from (x_value - std) to (x_value + std). So, if the value is in the range, it is not an anomaly; but if the value is out of the range, it is an anomaly/outlier

Q9 At inference time, the anomaly score can be calculated by calculating the difference between the predicted and actual value for that time point. Values that fall outside the prediction's confidence interval can be directly classified as anomalous.Isolation forest will then provide a ranking that reflects the degree of anomaly of each data instance according to their path lengths. The ranking or scores are called the anomaly scores which is calculated as follows: H(x) : the number of steps until the data instance x is fully isolated.

The anomaly score is a value from 0 to 100, which indicates the significance of the anomaly compared to previously seen anomalies. The highly anomalous values are shown in red and the low scored values are indicated in blue. An interval with a high anomaly score is significant and requires investigation.