1)What is anomaly detection and what is its purpose?

Ans- 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 is the ability to identify rare items or observations that don't conform to normal or common patterns found in data. These outliers are important within financial data because they can indicate potential risks, control failures, or business opportunities.

2) What are the key challenges in anomaly detection?

Ans- Null data or incomplete datasets.

Inconsistent data formats.

Duplicate data.

Different scales of measurement.

Human error.

3) How does unsupervised anomaly detection differ from supervised anomaly detection?

Ans- The main difference between supervised and unsupervised learning: Labeled data. The main distinction between the two approaches is the use of labeled datasets. To put it simply, supervised learning uses labeled input and output data, while an unsupervised learning algorithm does not.

4) What are the main categories of anomaly detection algorithms?

Ans- There are three main classes of anomaly detection techniques: unsupervised, semi-supervised, and supervised. Essentially, the correct anomaly detection method depends on the available labels in the dataset.

5) What are the main assumptions made by distance-based anomaly detection methods?

Ans- The basic assumptions for anomaly detection are that the anomalies, or outliers, occur rarely in the data, and they are significantly different from the expected pattern in the context being considered.

6) How does the LOF algorithm compute anomaly scores?

Ans- 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.

7) What are the key parameters of the Isolation Forest algorithm?

Ans- Isolation Forest is an algorithm for data anomaly detection initially developed by Fei Tony Liu and Zhi-Hua Zhou in 2008. Isolation Forest detects anomalies using binary trees. The algorithm has a linear time complexity and a low memory requirement, which works well with high-volume data.

A global interpretability method, called Depth-based Isolation Forest Feature Importance (DIFFI), to provide Global Feature Importances (GFIs) which represents a condensed measure describing the macro behaviour of the IF model on training data