1 Local Outlier Factor

On a high level, the local outlier factor uses distance to estimate the density of each point, and then points with very low densities would be considered outliers.
For the distance of each point, we indicate this using reachability distance, denoted $reachability{-}distance_k(A,B) = max\{k - distance(B), d(A,B)\}$, where $k - distance(B)$ indicates the distance of the object B to the k-th nearest neighbor.
We then have the local reachability density to calculate the density around a point P, where $|N_k(P)|$ represents the number of k nearest neighbors to P: $lrd_k(P) = 1/(\frac{sum_{B \in N_k(P)} reachability-distance_k(P,B)}{|N_k(P)|})$ And then finally, we have the local reachability density: $LOF_k(P) = \frac{\sum_{B \in N_k(P)} \frac{lrd_k(B)}{lrd_k(P)}}{|N_k(P)|}$

This demonstrates the average local reachability density of neighbors divided by the points' own local reachability density, and thus if this value is greater than one that means that the density of neighbors is greater than the local point, meaning it could potentially be an outlier.

2 Isolation Forest

On a high level, the isolation forest method uses tree splitting to determine if there are anomalies because anomalies require fewer splits to be isolated, we can compare their path lengths to the average path lengths to determine if it is an anomaly. Isolation forest has two main stages to detect outliers: build isolation trees using training, and then for each point pass it through the isolation trees and assign a proper ”anomaly score”.
For the anomaly score, we have a function to represent the average of $h(x)$ given m, where m is the size of the sample set to normalize $h(x)$, a function that represents the path length for x:

$$c(m) = \begin{cases} 2H(m-1) - \frac{2(m-1)}{n}, & \text{for } m > 2 \\ 1 & \text{for } m = 2 \\ 0 & \text{otherwise} \end{cases}$$

Thus, to get the anomaly score we normalize this $c(m)$ value: $s(x,m) = 2^{\frac{-E(h(x))}{c(m)}}$. If s is close to 1, x is very likely to be an anomaly. If s is smaller than 0.5, then x is likely to be a normal value.

3 K-nearest neighbors

This method simply uses the kNN algorithm to find neighbors, and calculate the distance between them to see if the point is an anomaly. We will simply try to use statistical methods, like the average k-th distance, to compare with each point's distance to its k-th nearest neighbor to detect if it is an outlier.