**Interview Questions**

**K-Nearest Neighbours**

**1. What are the key hyperparameters in KNN?**

In K-Nearest Neighbors (KNN), the key hyperparameters include:

1. Number of Neighbors (k):
   * This determines how many nearest neighbors are considered when making predictions. A small kkk can make the model sensitive to noise, while a larger kkk may smooth out predictions but could ignore local patterns.
2. Distance Metric:
   * The metric used to calculate the distance between points. Common options include:
     + Euclidean Distance: The default choice for continuous variables.
     + Manhattan Distance: Useful in higher dimensions or for certain types of data.
     + Minkowski Distance: A generalization that can represent both Euclidean and Manhattan distances depending on the parameter ppp.
3. Weights:
   * Determines how neighbors influence the classification:
     + Uniform: All neighbors contribute equally.
     + Distance: Closer neighbors have more influence than farther ones.
4. Algorithm:
   * The method used for nearest neighbor searches:
     + brute: Direct computation of distances.
     + kd\_tree: Efficient for low-dimensional data.
     + ball\_tree: Effective for higher-dimensional spaces.
5. Leaf Size:
   * This affects the efficiency of tree-based algorithms (like kd\_tree and ball\_tree). Smaller leaf sizes can speed up searches but may require more memory.
6. P (for Minkowski Distance):
   * In the Minkowski distance formula, ppp defines the distance type. p=1p = 1p=1 gives Manhattan distance, while p=2p = 2p=2 gives Euclidean distance.

Understanding these hyperparameters and how they impact the KNN model can significantly improve your ability to optimize its performance.

**2. What distance metrics can be used in KNN?**

In K-Nearest Neighbors (KNN), the following distance metrics can be used:

1. Euclidean Distance: The straight-line distance between two points.
2. Manhattan Distance: The sum of the absolute differences of their coordinates.
3. Minkowski Distance: A generalization that includes both Euclidean (p=2) and Manhattan (p=1) distances.
4. Chebyshev Distance: The maximum absolute difference along any dimension.
5. Hamming Distance: Measures the number of differing positions between categorical variables.
6. Cosine Similarity: Measures the cosine of the angle between two vectors, often used for text data.