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

The key hyperparameters in the **K-Nearest Neighbors (KNN)** algorithm are:

1. **n\_neighbors (K)** – The number of nearest neighbors to consider when making a classification decision. A small value (e.g., 1) can lead to high variance, while a large value can lead to high bias.
2. **metric (Distance Metric)** – Determines how the distance between data points is measured. Common metrics include:
   * "euclidean" (default) – Straight-line distance.
   * "manhattan" – Absolute distance between points.
   * "minkowski" – Generalized distance metric that allows different distance calculations.
3. **weights** – Defines how neighbors contribute to the classification:
   * "uniform" (default) – All neighbors contribute equally.
   * "distance" – Closer neighbors have a greater influence.
4. **algorithm** – The method used to compute nearest neighbors:
   * "auto" (default) – Chooses the best approach.
   * "ball\_tree" – Good for low-dimensional data.
   * "kd\_tree" – Efficient for medium-dimensional data.
   * "brute" – Uses a brute-force search (best for small datasets).
5. **p (Power Parameter for Minkowski Distance)** – Determines the distance calculation:
   * p = 1 → Manhattan distance.
   * p = 2 → Euclidean distance.
6. **What distance metrics can be used in KNN?**

**Distance Metrics in KNN:**

KNN can use various distance metrics to determine the similarity between data points. Here are the most commonly used ones:

1. **Euclidean Distance ('euclidean')**
   * Formula: d(p,q)=∑(pi−qi)2d(p, q) = \sqrt{\sum (p\_i - q\_i)^2}d(p,q)=∑(pi​−qi​)2​
   * Measures straight-line distance between two points.
   * Default and most commonly used.
2. **Manhattan Distance ('manhattan')**
   * Formula: d(p,q)=∑∣pi−qi∣d(p, q) = \sum |p\_i - q\_i|d(p,q)=∑∣pi​−qi​∣
   * Measures distance in a grid-based manner (useful for city block-like data).
3. **Minkowski Distance ('minkowski')**
   * Generalization of Euclidean and Manhattan distances.
   * Defined as: d(p,q)=(∑∣pi−qi∣p)1/pd(p, q) = \left( \sum |p\_i - q\_i|^p \right)^{1/p}d(p,q)=(∑∣pi​−qi​∣p)1/p
   * When p=1, it behaves like Manhattan; when p=2, it behaves like Euclidean.
4. **Chebyshev Distance ('chebyshev')**
   * Formula: d(p,q)=max⁡∣pi−qi∣d(p, q) = \max |p\_i - q\_i|d(p,q)=max∣pi​−qi​∣
   * Measures the greatest single-dimensional difference.
5. **Hamming Distance ('hamming')**
   * Measures the number of positions at which corresponding elements are different.
   * Useful for categorical data.
6. **Mahalanobis Distance ('mahalanobis')**
   * Accounts for correlations between features.
   * Requires computation of the covariance matrix.

**GridSearchCV** is currently optimizing over 'euclidean', 'manhattan', and 'minkowski' in “**KNN\_2**” file.