

Spark 2.4.7 - Documentation Helper Bot

Prompt

which clustering algorithms are available in spark 2.4.7

what is spark



Spark is a powerful open-source processing engine that provides high-level APIs in Java, Scala, Python, and R, and an optimized engine that supports general execution graphs. It also supports a rich set of higher-level tools including Spark SQL for SQL and structured data processing, MLlib for machine learning, GraphX for graph processing, and Spark Streaming for real-time data processing.

what is mlib



MLlib is a machine learning library in Apache Spark that provides a set of machine learning algorithms for classification, regression, clustering, collaborative filtering, and dimensionality reduction. It is designed to be scalable and easy to use, making it suitable for big data processing tasks.

clustering algorithms list





Some common clustering algorithms include:

1. K-means
2. Hierarchical clustering
3. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
4. Mean Shift
5. Gaussian Mixture Models
6. Agglomerative clustering
7. Spectral clustering
8. Affinity Propagation

These algorithms are used to group similar data points together based on certain criteria or distance metrics.

which clustering algorithms are available in spark 2.4.7



In Apache Spark 2.4.7, the MLlib library provides the following clustering algorithms for use:

1. K-means
2. Gaussian Mixture Models (GMM)
3. Bisecting K-means

These algorithms can be used for clustering tasks in Spark 2.4.7.

