1. **Week 01: Hadoop**
2. **Week 02: HDFS, MapReduce**
3. **Week 03: PySpark RDDs**
4. **Week 04: Pyspark Dataframe**
5. **Week 05: Frequent Itemset Mining**
6. A-Priori Algorithm
   1. Related Problem

Finds sets of items that frequently appear together in transactions and creates rules about item relationships.

* 1. Input/Output
* Input:
* Transaction database
* Support threshold (minimum frequency requirement)
* Confidence threshold (minimum strength for rules)
* Output:
* All frequent itemsets meeting the support threshold
* Association rules meeting the confidence threshold
  1. Pseudocode

A-Priori first identifies frequent single items, then builds larger sets step-by-step, using the knowledge that any subset of a frequent itemset must also be frequent.

* 1. Advantages & Disadvantages
* Advantages:
* Reduces candidate itemsets significantly
* Guarantees completeness of results
* Simple and straightforward implementation
* Disadvantages:
* Requires multiple database scans
* High memory usage with many frequent items
* Generates numerous candidate itemsets

1. PCY Algorithm
2. Related Problem

Improves upon A-Priori for finding frequent itemsets and association rules with better memory usage.

1. Input/Output

* Input:
* Transaction database
* Support threshold (minimum frequency requirement)
* Confidence threshold (minimum strength for rules)
* Output:
* All frequent itemsets meeting the support threshold
* Association rules meeting the confidence threshold

1. Pseudocode

PCY utilizes unused memory during the first pass by hashing potential pairs into buckets, then uses a bitmap to filter candidates in the second pass.

1. Advantages & Disadvantages

* Advantages:
* Makes better use of available memory
* Further reduces candidate pairs
* Maintains same pass count as A-Priori
* Disadvantages:
* Hash collisions reduce filtering effectiveness
* Still requires multiple database scans
* Increased algorithm complexity

1. SON Algorithm
2. Related Problem

Addresses frequent itemset mining in very large databases by processing manageable chunks.

1. Input/Output

* Input:
* Transaction database
* Support threshold (minimum frequency requirement)
* Confidence threshold (minimum strength for rules)
* Output:
* All frequent itemsets meeting the support threshold
* Association rules meeting the confidence threshold

1. Pseudocode

SON divides the database into smaller portions, finds local frequent itemsets in each portion, combines these as candidates, then verifies frequency across the entire database.

1. Advantages & Disadvantages

* Advantages:
* Only requires two database passes
* Naturally supports parallel processing
* Scalable for very large databases
* Disadvantages:
* May generate excess candidates
* Second pass memory depends on candidate count
* Partition size selection affects performance

1. **Week 06: Finding Similar Items**
2. MinHash
3. Related Problem
4. Input/Output

* Input:
* Output:

1. Pseudocode
2. Advantages & Disadvantages

* Advantages:
* Disadvantages:

1. LSH: Locality Sensitive Hashing
2. Related Problem
3. Input/Output

* Input:
* Output:

1. Pseudocode
2. Advantages & Disadvantages

* Advantages:
* Disadvantages:

1. **Week 07: Clustering**
2. K-means Clustering
3. Related Problem
4. Input/Output

* Input:
* Output:

1. Pseudocode
2. Advantages & Disadvantages

* Advantages:
* Disadvantages:

1. The BFR Algorithm
2. Related Problem
3. Input/Output

* Input:
* Output:

1. Pseudocode
2. Advantages & Disadvantages

* Advantages:
* Disadvantages:

1. The CURE Algorithm
2. Related Problem
3. Input/Output

* Input:
* Output:

1. Pseudocode
2. Advantages & Disadvantages

* Advantages:
* Disadvantages:

1. **Week 09: Dimensionality Reduction**
2. SVD: Singular Value Decomposition
3. Related Problem
4. Input/Output

* Input:
* Output:

1. Pseudocode
2. Advantages & Disadvantages

* Advantages:
* Disadvantages:

1. CUR Decomposition
2. Related Problem
3. Input/Output

* Input:
* Output:

1. Pseudocode
2. Advantages & Disadvantages

* Advantages:
* Disadvantages:

1. **Week 10: Recommender Systems**
2. Content-based Systems
3. Related Problem
4. Input/Output

* Input:
* Output:

1. Pseudocode
2. Advantages & Disadvantages

* Advantages:
* Disadvantages:

1. Collaborative Filtering
2. Related Problem
3. Input/Output

* Input:
* Output:

1. Pseudocode
2. Advantages & Disadvantages

* Advantages:
* Disadvantages:

1. **Week 13: PageRank**
2. Random Surfer Model
3. Related Problem
4. Input/Output

* Input:
* Output:

1. Pseudocode
2. Advantages & Disadvantages
3. Google Algorithm
4. Related Problem
5. Input/Output

* Input:
* Output:

1. Pseudocode
2. Advantages & Disadvantages

* Advantages:
* Disadvantages: