

When working with “Big Data,” many different machine learning algorithms can be used to analyse the data and get useful information from it.

Here are some well-known machine learning algorithms and tools that are often used with Big Data processing systems like Hadoop, Spark, Hive, NoSQL databases, MapReduce, and Storm:

1. Linear Regression: A basic regression algorithm used to model the relationship between a dependent variable and one or more independent variables. It can be used for prediction and forecasting tasks.
2. Logistic Regression: Used for binary classification problems, where the goal is to predict a binary outcome (e.g., yes/no, true/false).
3. Decision Trees: A versatile algorithm used for classification and regression tasks. It can handle both categorical and numerical data.
4. Random Forest: An ensemble learning method that builds multiple decision trees and combines their predictions to improve accuracy and reduce overfitting.
5. Gradient Boosting Machines (GBM): Another ensemble technique that builds multiple weak learners (typically decision trees) sequentially, where each new tree corrects the errors of the previous one.
6. K-Nearest Neighbors (KNN): A simple algorithm that classifies data points based on the majority class of their K-nearest neighbors.
7. Support Vector Machines (SVM): Suitable for both classification and regression tasks, SVM aims to find the optimal hyperplane that best separates data points of different classes.
8. Naive Bayes: A probabilistic classification algorithm based on Bayes’ theorem, often used for text classification and spam filtering.
9. Clustering Algorithms (e.g., K-Means): Used to group similar data points into clusters,

helpful for segmentation and pattern recognition tasks.

10. Neural Networks and Deep Learning: Artificial neural networks, including deep learning architectures like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are utilized for complex tasks such as image recognition, natural language processing, and speech recognition.

Regarding Big Data Processing Frameworks And Technologies:

1. Hadoop: An open-source framework for distributed storage and processing of large datasets across clusters of computers. It uses the Hadoop Distributed File System (HDFS) and MapReduce for data processing.
2. Apache Spark: Another popular distributed computing framework that supports in-memory processing and offers various APIs for batch processing, real-time streaming, machine learning (Spark MLlib), and graph processing.
3. Apache Hive: A data warehousing and SQL-like query language built on top of Hadoop, allowing users to query and analyze data using SQL commands.
4. NoSQL Databases: Various NoSQL databases, such as MongoDB, Cassandra, HBase, and Couchbase, are utilized to handle large volumes of unstructured and semi-structured data.
5. MapReduce: A programming model for distributed computing, commonly used in Hadoop, where large datasets are processed in parallel across a cluster.
6. Apache Storm: A real-time stream processing system, suitable for applications that require low-latency data processing and analysis.

Related posts:

1. What is Machine Learning ?
2. Types of Machine Learning ?
3. Applications of Machine Learning
4. Data Preprocessing
5. Data Cleaning
6. Handling Missing Data
7. Feature Scaling
8. Labeled data in Machine learning
9. Difference between Supervised vs Unsupervised vs Reinforcement learning
10. Difference between Supervised vs Unsupervised vs Reinforcement learning
11. What is training data in Machine learning
12. What is Ordinary Least Squares (OLS) estimation
13. Scalar in Machine Learning
14. Scalars in Loss Functions | Machine Learning
15. Linear Algebra for Machine Learning Practitioners
16. Supervised Learning
17. Top Interview Questions and Answers for Supervised Learning
18. Define machine learning and explain its importance in real-world applications.
19. Differences Between Machine Learning and Artificial Intelligence
20. Machine Learning works on which type of data ?
21. What is target variable and independent variable in machine learning
22. Machine Learning Scope and Limitations
23. What is Regression in Machine learning
24. Statistics and linear algebra for machine learning
25. Finding Machine Learning Datasets

26. What is hypothesis function and testing
27. Explain computer vision with an appropriate example
28. Explain Reinforcement learning with an appropriate exaple
29. Reinforcement Learning Framework
30. Data augmentation
31. Normalizing Data Sets in Machine Learning
32. Machine learning models
33. Unsupervised machine learning
34. Neural Network in Machine Learning
35. Recurrent neural network
36. Support Vector Machines
37. Long short-term memory (LSTM) networks
38. Convolutional neural network
39. How to implement Convolutional neural network in Python
40. What is MNIST ?
41. What does it mean to train a model on a dataset ?
42. Can a textual dataset be used with an openCV?
43. Name some popular machine learning libraries.
44. Introduction to Machine Learning
45. Some real time examples of machine learning
46. Like machine learning, what are other approaches in AI ?
47. Statistics and Linear Algebra for Machine Learning ?
48. What is convex optimization in simple terms ?
49. What is data visualization in simple terms ?
50. What is data preprocessing in machine learning ?
51. What are data distributions, and why are they important ?
52. What is data augmentation in machine learning ?

- 53. What is labelled and unlabelled data set in Machine Learning ?
- 54. What is neural networks in Machine Learning ?
- 55. How are convolutional neural networks related to supervised learning ?
- 56. Fundamentals of Neural Networks
- 57. Linearity vs non-linearity in Machine Learning ?
- 58. Machine Learning Short Exam Notes
- 59. Machine Learning Short Exam Notes - Quick and Easy Revision Guide