

# Big Data Mining Algorithms

**M. Govindarajan**

*Annamalai University, India*

## INTRODUCTION

Data mining is widely used in fields such as science, engineering, medicine, and business. With this technique, previously hidden insights have been unearthed from large amounts of data to benefit the business community (D. Che et al., 2013). Since the establishment of organizations in the modern era, data mining has been applied in data recording. For flexible data analysis, Begoli and Horey (2012) proposed three principles: first, architecture should support many analysis methods, such as statistical analysis, machine learning, data mining, and visual analysis. Second, different storage mechanisms should be used because all of the data cannot fit in a single type of storage area. Additionally, the data should be processed differently at various stages. Third, data should be accessed efficiently. To analyze Big Data, data mining algorithms that are computer intensive are utilized. Such algorithms demand high-performance processors. Many algorithms were defined earlier in the analysis of large data set. The purpose of this chapter is to provide an analysis of different machine learning algorithms available for performing big data analytics.

## BACKGROUND

C. Yadav et al., (2013) present a review of some old algorithms that can handle large data set as Nearest Neighbor Search, Decision Tree and Neural Network. A. Fahad et al., (2014) present a survey of existing clustering algorithms of different categories (Partitioning-based, Hierarchical-based, Density-based, grid-based and model based). Tsai et al., (2015) begins with a brief introduction to data analytics followed by the discussions of big data analytics. Some important open issues and further research directions will also be presented for the next step of big data analytics. D.Rajkumar and S.Usha (2016) review on various big data mining algorithms and the methods employed to handle such a vast data is also discussed. J. Sangeetha and V. Sinthu Janita Prakash (2017) surveys about the big data mining techniques, data slicing techniques and clustering techniques and also discusses about its advantages and drawbacks. The performance and quality measurement of the big data mining algorithms, mining platforms, data slicing techniques and clustering techniques are discussed. Ripon Patgiri (2018) presents the taxonomy of Big Data and also present in-depth insight on the Big Data paradigm. Mounica Vennapusa et al., (2019) give the survey on applications and challenges of Machine Learning techniques, advanced learning methods towards Bigdata.

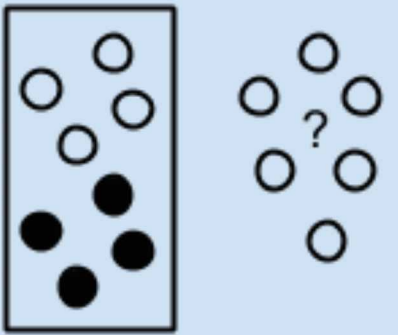
**FOCUS OF THE ARTICLE**

Machine learning is a field of computer science that try to enable computers with the ability to learn without explicitly programming them. The purpose of this chapter is to provide an analysis of different Machine learning algorithms available for performing big data analytics. The ML algorithms are categorized in three key categories, namely, supervised, unsupervised and semi-supervised ML algorithm. The Big Data Analytics or Big Data Security analytics fully depends on Data Mining, where the Machine Learning techniques are subset of Data Mining (Ripon Patgiri, 2018).

**Supervised Learning Algorithms**

The supervised learning algorithms are trained with a complete set of data and thus, the supervised learning algorithms are used to predict/forecast.

*Figure 1. Supervised Learning Algorithms*

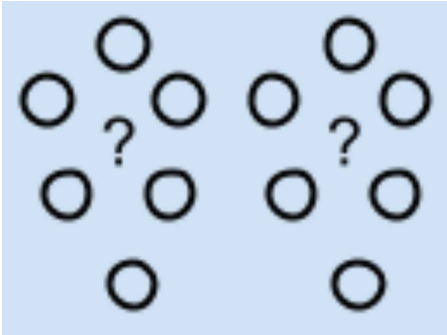


Example problems are classification and regression. Example algorithms include: Logistic Regression and the Back Propagation Neural Network.

**Unsupervised Learning Algorithms**

The unsupervised learning algorithms starts learning from scratch, and therefore, the unsupervised learning algorithms are used for clustering.

*Figure 2. Unsupervised Learning Algorithms*



8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/big-data-mining-algorithms/260227](http://www.igi-global.com/chapter/big-data-mining-algorithms/260227)

## Related Content

---

### Improving Quality of Business in Next Generation Telecom Networks

Vesna Radonji ogatovi (2018). *Encyclopedia of Information Science and Technology, Fourth Edition* (pp. 6546-6555).

[www.irma-international.org/chapter/improving-quality-of-business-in-next-generation-telecom-networks/184350](http://www.irma-international.org/chapter/improving-quality-of-business-in-next-generation-telecom-networks/184350)

### Environmental Informatics for Sustainable Development

Carlos Granell and Sven Schade (2015). *Encyclopedia of Information Science and Technology, Third Edition* (pp. 2942-2954).

[www.irma-international.org/chapter/environmental-informatics-for-sustainable-development/112717](http://www.irma-international.org/chapter/environmental-informatics-for-sustainable-development/112717)

### Deploying a Software Process Lifecycle Standard in Very Small Companies

Rory V. O'Connor (2015). *Encyclopedia of Information Science and Technology, Third Edition* (pp. 762-772).

[www.irma-international.org/chapter/deploying-a-software-process-lifecycle-standard-in-very-small-companies/112391](http://www.irma-international.org/chapter/deploying-a-software-process-lifecycle-standard-in-very-small-companies/112391)

### FLANN + BHO: A Novel Approach for Handling Nonlinearity in System Identification

Bighnaraj Naik, Janmenjoy Nayak and H.S. Behera (2018). *International Journal of Rough Sets and Data Analysis* (pp. 13-33).

[www.irma-international.org/article/flann--bho/190888](http://www.irma-international.org/article/flann--bho/190888)

### Modeling Rumors in Twitter: An Overview

Rhythm Walia and M.P.S. Bhatia (2016). *International Journal of Rough Sets and Data Analysis* (pp. 46-67).

[www.irma-international.org/article/modeling-rumors-in-twitter/163103](http://www.irma-international.org/article/modeling-rumors-in-twitter/163103)