Chapter 5
A Survey of Big Data Analytics Using Machine Learning Algorithms

Usha Moorthy
Vellore Institute of Technology, India

Usha Devi Gandhi
Vellore Institute of Technology, India

ABSTRACT

Big data is information management system through the integration of various traditional data techniques. Big data usually contains high volume of personal and authenticated information which makes privacy as a major concern. To provide security and effective processing of collected data various techniques are evolved. Machine Learning (ML) is considered as one of the data technology which handles one of the central and hidden parts of collected data. Same like ML algorithm Deep Learning (DL) algorithm learn program automatically from the data it is considered to enhance the performance and security of the collected massive data. This paper reviewed security issues in big data and evaluated the performance of ML and DL in a critical environment. At first, this paper reviewed about the ML and DL algorithm. Next, the study focuses towards issues and challenges of ML and their remedies. Following, the study continues to investigate DL concepts in big data. At last, the study figures out methods adopted in recent research trends and conclude with a future scope.

DOI: 10.4018/978-1-5225-2863-0.ch005

1. INTRODUCTION

Big data analytics is the vast level investigation and preparing of data in dynamic utilize in a few fields and, as of late, has pulled in light of a legitimate concern for the security group for its guaranteed capacity to dissect and correspond security related data effectively and at phenomenal scale (Shirudkar et al., 2015). Separating between customary data examination and enormous data investigation for security is, in any case, not clear (Imperva, 2015). All things considered, the data security group has been utilizing the investigation of system movement, framework logs, and other data sources to recognize dangers and identify noxious exercises for over 10 years, and it's not clear how these customary methodologies vary from big data (Mulanee et al., 2015). "Big Data Analytics for Security Intelligence," concentrates on big data's part insecurity (Raja et al., 2014). In advanced world, data are produced from different sources and the quick move from computerized innovations has prompted the development of enormous data (Suryawanshi et al., 2015). It gives transformative leaps forward in numerous fields with an accumulation of vast datasets. When all is said in done, it alludes to the accumulation of extensive and complex datasets which are hard to process utilizing customary database administration instruments or data handling applications (UK Data Archive, 2011). These are accessible in the organized, semi-organized, and unstructured organization in peta bytes and past (Tsai et al., 2015). Some of these extraction strategies for acquiring accommodating data were examined by Gandomi and Haider (Gandomi et al., 2015). The, however, correct definition for big data is not characterized, and there is trusted that it is issue particular. This will help us in getting upgraded basic leadership, knowledge disclosure, and advancement while being inventive and financially savvy (Kaur and Kaur, 2016). Extensive scale data sets are gathered and examined in various spaces, from designing sciences to interpersonal organizations, trade, bimolecular examination, and security (Tsai et al., 2015). Especially, advanced data produced from an assortment of computerized gadgets, and are developing at amazing rates. As per Gandomi and Haider (2015), in 2011, computerized data is grown nine times in volume in only 5years, and its sum on the planet will be reached 35 trillion gigabytes by 2020 (Lynch, 2008). In this manner, the expression "Enormous Data" was begotten to catch the significant importance of this data blast pattern (Qiu et al., 2016).

The aim of Machine Learning (ML) is to empower a framework to gain from the past or present and utilize that data to settle on expectations or choices with respect to obscure future occasions (Rajkumar et al., 2016). In the broadest terms, the work process for an administered ML errand comprises of three stages: manufacture the model, assess and tune the model, and afterward put the model into creation