

## Current Big Data Issues and Their Solutions via Deep Learning: An Overview

Asif Ali Banka\*, Roohie Naaz Mir

\*Department of Computer Science and Engineering, National Institute of Technology, Srinagar,  
Jammu and Kashmir 190006 India

**Abstract:** The advancements in modern day computing and architectures focus on harnessing parallelism and achieve high performance computing resulting in generation of massive amounts of data. The information produced needs to be represented and analyzed to address various challenges in technology and business domains. Radical expansion and integration of digital devices, networking, data storage and computation systems are generating more data than ever. Data sets are massive and complex, hence traditional learning methods fail to rescue the researchers and have in turn resulted in adoption of machine learning techniques to provide possible solutions to mine the information hidden in unseen data. Interestingly, deep learning finds its place in big data applications. One of major advantages of deep learning is that it is not human engineered. In this paper, we look at various machine learning algorithms that have already been applied to big data related problems and have shown promising results. We also look at deep learning as a rescue and solution to big data issues that are not efficiently addressed using traditional methods. Deep learning is finding its place in most applications where we come across critical and dominating 5Vs of big data and is expected to perform better.

**Index Terms:** Big data, machine learning, deep learning

### I. INTRODUCTION

The advancements in modern day computing and architecture focuses on harnessing parallelism and achieve high performance computing resulting in generation of massive amounts of data. The interconnected devices have practically resulted into data intensive computing where data is now essential part of human life. Sharing data, professional or non-professional due to high availability of anywhere-computing and usage of internet oriented application like social media, internet of things, GPS, Cloud computing etc. have resulted in generation of massive amounts of data.

Every minute Facebook records more than 3.2 million likes, stores more than 3.4 million posts and generates around 4 GB of data. In March 2013, Facebook launched a graph search feature that enables its users to search the social graph for users with similar locations or hobbies. Also in 1 min, Google answers about 300K searches, 126 h uploaded to YouTube and more than 140K video views, about 700 users created in Twitter and more than 350K tweets generated, and more than 11K searches on LinkedIn performed [1]. The massive data is being produced, consumed and transferred and is expected to increase with more people gaining access to technology.

Computing technology alone is not the