

## LEVERAGING DATA SCIENCE AND BIG DATA IN IoT APPLICATIONS: A COMPREHENSIVE REVIEW OF TECHNOLOGIES, ARCHITECTURES AND CHALLENGES

Arindam Ghosh<sup>1</sup>, Subhadip Das<sup>2</sup>, Subrata Kumar Majumdar<sup>3</sup>, Pritoma Saha<sup>4</sup>, Sudipta Bhanja<sup>5</sup>, Deblina Roy<sup>3</sup>, and Kousik Roy<sup>2\*</sup>

<sup>1</sup>Department of Computer science and Engineering, Academy of Technology (AOT), Adisaptagram- 712502

<sup>2</sup>Department of Computer Science and Engineering, Bengal College of Engineering and Technology, Durgapur-713212, West Bengal

<sup>3</sup>Department of Mechanical Engineering, Sanaka Educational Trust, Durgapur-713212, West Bengal

<sup>4</sup>Department of Computer Application, Techno India university, Salt Lake City, Sector V, Kolkata, West Bengal 700091

<sup>5</sup>Kalyani Government Engineering College Kalyani, Nadia, West Bengal, India. Pin-741235

\*Corresponding Author: kousikroy002@gmail.com

### ABSTRACT

The proper fusion of IoT devices data with the utilization of Data Science and Big Data is nothing but the demand of the hour. This thorough review paper unveils all the technologies, architectures, and challenges at this crucial intersection. The review not only discusses these topics but also emphasizes the main points of data ingestion, distributed storage, real time stream processing and application oriented machine learning models being developed for smart cities and industrial automation. Furthermore, the paper examines the competition between cloud and edge computing technologies aimed at building fast and efficient analytics. To illustrate the rivalry between these two strategies the paper points out the main problems of data heterogeneity, security, privacy and power limitations. The review keeps itself well acquainted with the latest trends of federated learning and digital twins that are not only very much beneficial for researchers but also for practitioners in the field of smart decision making with the IoT technology.

**Keywords:** *Data Science, Big Data, Internet of Things, Machine Learning, Edge Computing, Real Time Analytics, Federated Learning, Digital Twins.*

### 1. INTRODUCTION

Dependence on Internet of Things (IoT) devices is one of the main reasons why there is a continuously increasing amount of data that requires advanced analytics [1]. By combining Data Science and Big Data, the application of machine learning has already expanded from mere monitoring to predictive analytics and even further [2]. One promising solution to the problem is the implementation of the cloud and edge computing infrastructure which is scalable and by that providing very fast and real-time distribution of data storage and processing thus enabling the almost instant processing of huge amounts of data for decision making [3]. Still, the generation of data remains an issue along with the other main challenges like heterogeneous data, security of the systems and availability of power [4]. Along with federated learning and blockchain these are some of the solutions that will not just help in overcoming challenges but also in boosting the reliability that underpins the technology's growth. The analysis of technologies, architectures and challenges associated with Data Science and Big Data in IoT paves the way for the new maturity this technology has reached.