

Unveiling the Crucial Role of Statistics in Big Data Analytics

Ms. Rinal Dineshbhai Suthar¹ and Asst. Prof. Ashwini Kale²

Department of Computer Science¹

Department of Computer Science (Statistics)²

Sarhad College of Arts, Commerce and Science, Pune

rinalsuthar23@gmail.com and kaleap30@gmail.com

Abstract: *This paper focuses on the role of statistics in big data analytics to build analytical accuracies and precise decision-making insights. Statistics and data mining techniques that are useful for big data analytics include: significance testing, classification, regression/prediction, cluster analysis, association rule learning, hypothesis testing, anomaly detection, and visualization. Statistical analysis provides a scientific justification to move from data to knowledge and turn it into actionable insights, emphasizing their role in data preprocessing, pattern recognition, and predictive modeling. The study navigates through prominent statistical methods such as regression analysis, hypothesis testing, and machine learning algorithms, illustrating their efficacy in extracting actionable knowledge from vast and complex datasets. Moreover, it sheds light on the symbiotic relationship between statistics and technological advancements, showcasing their collaborative potential in driving innovation within the Big Data domain. By synthesizing current research and practical applications, this paper aims to underscore the indispensable nature of statistical methodologies in harnessing the power of Big Data for informed decision-making. In the era of Big Data, where information proliferates at an unprecedented scale, the role of statistics has become increasingly vital in extracting meaningful insights from the vast and complex datasets that characterize this landscape. This paper seeks to unravel the intricate interplay between statistics and Big Data analytics, elucidating the significance of statistical methodologies in navigating the challenges posed by the sheer volume, velocity, and variety of data.*

Keywords: Statistics, Big Data, Analytics, Machine learning, Statistical techniques

REFERENCES

- [1]. Acharya, S., & Chellappan, S. (2019). Big Data and Analytics (2nd ed.).
- [2]. Ahmad, S. J. (Year). Big Data Manipulation- A new concern to the ICT world (A massive Survey/statistics along with the necessity). Department of Computer Science & Engineering, University of South Asia.
- [3]. Barcaroli, G. (2015). Use of Big Data in Official Statistics.
- [4]. Chen, J. J. (2015). Statistics in Big Data. Journal of the Chinese Statistical Association.
- [5]. Florescu, D., Karlberg, M., Reis, F., Rey Del Castillo, P., Skaliotis, M., & Wirthmann, A. (2014). Will 'big data' transform official statistics? ESTAT.
- [6]. Petrakos, M., Santourian, A., Farmakis, G., Stavropoulos, P., Oikonomopoulou, G., Ntakou, E., Koumaki, M., & Trampeli, A. (2014). Analysis of the potential of selected big data repositories as data sources for official statistics. Agilis SA Statistics & Informatics.
- [7]. Rahman, A. (2019). Statistics-Based Data Preprocessing Methods and Machine Learning Algorithms for Big Data Analysis.
- [8]. Jansen, R. (2015). Benefits and Challenges of using Big Data for Official Statistics. In Proceedings of the 60th ISI World Statistics Congress, 26-31 July 2015, Rio de Janeiro, Brazil.