



A Study on Cleaning and Pre-Processing of Large-Scale Mobile Data

Gururaj A Nagalika, Research Scholar, Department of Computer Science, OPJS University, Churu (RAJ)

Abstract

The rapid growth of mobile devices and digital communication has resulted in the generation of massive volumes of heterogeneous data. Effective analysis of large-scale mobile data requires robust data cleaning and pre-processing techniques to ensure accuracy, consistency, and reliability. This study explores the importance of cleaning and pre-processing mobile datasets collected from smartphones, mobile applications, sensors, and communication networks. It examines common data quality issues, including missing values, duplicate records, inconsistent formats, noise, outliers, and incomplete observations, and discusses techniques for addressing these challenges. The paper also reviews preprocessing methods such as data integration, transformation, normalization, feature selection, and dimensionality reduction to improve the efficiency of data analytics and machine learning models. The findings indicate that appropriate preprocessing significantly enhances data quality, reduces computational complexity, and improves the performance of predictive and analytical systems. Furthermore, the study highlights the role of scalable big data frameworks and automated preprocessing pipelines in handling high-volume mobile datasets. It concludes that systematic data cleaning and preprocessing are fundamental steps for extracting meaningful insights and supporting reliable decision-making in mobile computing, smart applications, and data-driven research.

Keywords: Mobile Data, Data Cleaning, Data Preprocessing, Big Data Analytics, Data Quality, Missing Data.

