

Chapter 1

Optimizing

eCommerce Data:

Effective Approaches for

Data Collection, Cleansing,

and Preprocessing

Chirag Sharma

Chitkara Business School, Chitkara University, Punjab, India


Amanpreet Kaur

Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

Priyanka Datta

Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

Yonis Gulzar

 <https://orcid.org/0000-0002-6515-1569>

College of Business Administration, King Faisal University, Al Ahsa, Saudi Arabia

ABSTRACT

The process of gathering and analyzing data, on variables in a manner to address research inquiries, test hypotheses and evaluate results is referred to as data collection. After obtaining the required data the next important step is pre-processing it. Data pre-processing involves converting data into a dataset before running algorithms. It prepares the dataset by handling missing values, noisy data and inconsistencies. Sometimes additional adjustments such as handling outliers and

DOI: 10.4018/979-8-3693-5718-7.ch001

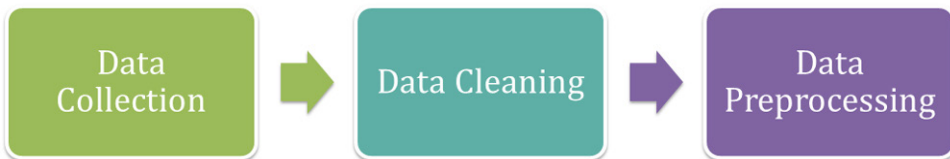
Copyright © 2025, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

creating subcategories are needed for generating outcomes. This process can be seen as a step in organizing data effectively. The methods mentioned above demonstrate that research entails steps typically carried out manually or using applications by researchers. This article will explore an instances where Artificial Intelligence (AI) and Machine Learning (ML) are utilized for cleaning and pre-processing data. By examining the data AI and ML can detect patterns efficiently. Visual representations play a role, in understanding customer satisfaction and happiness levels within a shopping platform

1. INTRODUCTION

In today's world driven by data, e commerce companies thrive on the information obtained from customer behaviour, product trends and market dynamics. However, the data collected from sources is often messy, incomplete and inconsistent. To fully utilize this data's potential businesses, need to undergo a process. Data preparation. This process involves three steps; gathering data cleaning it up and pre-processing it as illustrated in Figure 1. Each step is essential, in converting data into insights that can guide decision making and boost e commerce success.

Figure 1. Phases of converting raw data into actionable insights



1.1 Data Collection

Data collection forms the first step in the data preparation process (Taherdoost, 2021). Gathering information, from sources is important for understanding the goals of a research project, as shown in Figure 2. Taking the time to plan and decide on the type of data needed is crucial as it shapes the direction of the research. This initial

28 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/optimizing-ecommerce-data/356669

Related Content

Prediction Method of Electric Energy Metering Device Based on Software-Defined Networking

Jintao Chen, Binruo Zhu, Fang Zhao and Ruili Huang (2022). *International Journal of Information Security and Privacy* (pp. 1-20).

www.irma-international.org/article/prediction-method-of-electric-energy-metering-device-based-on-software-defined-networking/308316

OTT Platforms and Their Distributorship Agreement With Content Makers: A Study From the Perspective of Competition Law and Policy in India and Other Jurisdictions

Swati Bajaj Seth (2022). *Handbook of Research on Cyber Law, Data Protection, and Privacy* (pp. 292-302).

www.irma-international.org/chapter/ott-platforms-and-their-distributorship-agreement-with-content-makers/300917

Ontology-Based Authorization Model for XML Data in Distributed Systems

Amit Jain and Csilla Farkas (2010). *Web Services Security Development and Architecture: Theoretical and Practical Issues* (pp. 57-82).

www.irma-international.org/chapter/ontology-based-authorization-model-xml/40586

The Role and Impact of Federal Learning in Digital Healthcare: A Useful Survey

Rajasree R. S., Gopika G. S., Sree Krishna M. and Carlos Andrés Tavera Romero (2022). *Handbook of Research on Technical, Privacy, and Security Challenges in a Modern World* (pp. 127-147).

www.irma-international.org/chapter/the-role-and-impact-of-federal-learning-in-digital-healthcare/312419

Behavioral Advertising Ethics

Aaron K. Masey and Annie I. Antón (2011). *Information Assurance and Security Ethics in Complex Systems: Interdisciplinary Perspectives* (pp. 162-185).

www.irma-international.org/chapter/behavioral-advertising-ethics/46345