Chapter 9
Advances in Data Processing, Machine Learning, and Data Security

N. Manjunathan
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, India

V. Sathya
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, India

K. Aravindaraj
SRM Institute of Science and Technology, India

R. Senthil
Majan College, Oman

J. Anitha
Panimalar Engineering College, India

R. Siva Subramanian
https://orcid.org/0000-0002-7509-9223
RMK College of Engineering and Technology, India

V. K. Ramya Bharathi
Sathyabama Institute of Science and Technology, India

ABSTRACT

This study examines how data processing, sophisticated machine learning (ML), and data security are crucial to data-driven decision-making. It covers accurate data collecting, cleaning, and pre-processing methods, which are the foundation for trustworthy ML models. Exploratory data analysis and feature engineering provide difficult dataset insights. Data quality depends on how missing data and outliers are handled. Predictive modelling uses ML approaches such supervised, unsupervised, semi-supervised, ensemble, and deep learning. Reducing dimensionality and selecting

DOI: 10.4018/979-8-3693-5718-7.ch009

features improve model efficiency and interpretation. The model creation, training, and assessment methods ensure performance quality. The study also emphasises data security, ML ethics, privacy, and justice. New technologies in eCommerce data security are revolutionising protection methods, and AI-based solutions are moving cybersecurity towards transparency and confidentiality. However, future study will examine ethical problems and explainability to improve data-driven applications.

1.INTRODUCTION

The world of today which is constantly changing is unique in terms of the incredible volume of digital data and the novel technologies that emerge from time to time. This has led to the convergence of powerful data processing techniques, trustworthy machine learning algorithms, and strict data security protocols, which have become necessary and fundamental(Mahesh 2020). The digital era has come, and data is the new oil, and it is the driving force of innovation, it is the one that influences choices, and it is the one that reshapes almost every industry. Organizations from any sector in the global economy, from retail to transportation, healthcare to finance, are using data-driven approaches to underline crucial features, to make their business more effective, and even to stay in competition in a connected world. The multi-faceted data science approach, comprised of the combination of ML methods and data processing algorithms, drives the data-driven change. A successful ML model is built on a strong base of effective data collection, cleansing, and pre-processing(Sharifani, K., & Amini, M 2023). Feature engineering and EDA techniques are the main choices for data scientists that help them make sense of raw data, find significant insights, and extract useful knowledge from gigabytes of various unprocessed data(Ben Jabeur, S et al 2023).

Along with these factors, the problems of managing missing data and outliers will inevitably become more and more of a topic of discussion as the scale of data keeps expanding without showing any signs of slowing down. Well-structured strategies for these problems are vital to ensure the effectiveness and robustness of machine learning models as well as to ensure the reliability and consistency of data. ML, technically, is a set of various techniques and approaches, among which are supervised, unsupervised, and semi-supervised classifying systems (Sarker, I. H 2021). Predictive modeling, problem detection, clustering, and pattern recognition are among the many problems that can be solved by the diversified machine learning algorithms ranging from the standard regression and classification to the newest NN and DL architecture(Ahsan et al 2021).

Improved Transmission of Data and Information in Intrusion Detection Environments Using the CBEDE Methodology

Reinaldo Padilha França, Yuzo Iano, Ana Carolina Borges Monteiroand Rangel Arthur (2020). *Handbook of Research on Intrusion Detection Systems (pp. 26-46).*
www.irma-international.org/chapter/improved-transmission-of-data-and-information-in-intrusion-detection-environments-using-the-cbede-methodology/251795