Intelligent Techniques for Predictive Data Analytics

Edited by
Neha Singh • Shilpi Birla
Mohd Dilshad Ansari • Neeraj Kumar Shukla

IEEEPress
WILEY

Intelligent Techniques for Predictive Data Analytics

IEEE Press
445 Hoes Lane
Piscataway, NJ 08854

Intelligent Techniques for Predictive Data Analytics

Edited by

Neha Singh
Manipal University Jaipur
Jaipur, Rajasthan
India

Shilpi Birla
Manipal University Jaipur
Jaipur, Rajasthan
India

Mohd Dilshad Ansari
SRM University
Delhi-NCR, Haryana
India

Neeraj Kumar Shukla
King Khalid University
Abha
Saudi Arabia

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data
Names: Singh, Neha (Electronic engineering professor), editor. | Birla, Shilpi, editor. | Ansari, Mohd Dilshad, editor. | Neeraj Kumar Shukla, editor. | John Wiley & Sons, publisher.
Title: Intelligent techniques for predictive data analytics / Neha Singh, Shilpi Birla, Mohd Dilshad Ansari, Neeraj Kumar Shukla.
Description: Hoboken, New Jersey : Wiley, [2024] | Includes index.
Identifiers: LCCN 2024005729 (print) | LCCN 2024005730 (ebook) | ISBN 9781394227969 (hardback) | ISBN 9781394227983 (adobe pdf) | ISBN 9781394227976 (epub)
Subjects: LCSH: Engineering–Data processing. | Predictive analytics–Industrial applications. | Predictive analytics–Scientific applications.
Classification: LCC TA340 .I56 2023 (print) | LCC TA340 (ebook) | DDC 620.00285–dc23/eng/20240401
LC record available at https://lccn.loc.gov/2024005729
LC ebook record available at https://lccn.loc.gov/2024005730

Cover Design: Wiley
Cover Image: © Blue Planet Studio/Shutterstock

Set in 9.5/12.5pt STIXTwoText by Straive, Pondicherry, India

Contents

About the Editors

Dr. Neha Singh is a well-rounded and dynamic teaching professional with 19+ years of in-depth experience in educating undergraduate and postgraduate engineering students in electronics and communication engineering. She is currently working as Assistant Professor (Selection Grade) in the Department of Electronics and Communication Engineering at Manipal University Jaipur, Rajasthan, India. She did her PhD in the year 2020. Her areas of research interest encompass image processing, machine learning, data analysis, and VLSI design. She holds one Indian patent and has several papers published in journals and conferences of repute as well as many book chapters in indexed books. She has co-authored engineering textbooks and serves as an editor for indexed books. She has served as a reviewer for various international and peer-reviewed journals and conferences. She has also convened various conferences, workshops, and other events, chaired sessions, and organized summer internships in diode fabrication. She has guided several M Tech dissertations and B Tech projects and is guiding PhD scholars as well. She is a senior member of the IEEE.

Dr. Shilpi Birla is currently serving as Associate Professor (Senior Scale) in the Electronics and Communication Department at Manipal University Jaipur with over 16 years of experience in both academia and industry. She did her PhD in low power VLSI design. With a specialization in low power VLSI design, memory circuits, digital VLSI circuits, nanodevices, and image processing, she has made significant contributions to her field. Her research endeavors are reflected in over 60 research papers published in journals of repute and at various international conferences. She has played a pivotal role in organizing workshops covering HSPICE, TCAD, and

XILINX, as well as overseeing summer internships in diode fabrication and faculty development programs. Her commitment to fostering academic growth extends to guiding M Tech students and mentoring PhD candidates. Beyond her academic pursuits, she has actively participated in the academic community, serving as a session chair, conference steering committee member, editorial board member, and reviewer for international and national IEEE journals and conferences. She is also a senior member of the IEEE.

Dr. Mohd Dilshad Ansari is currently working as an associate professor in the Department of Computer Science and Engineering at SRM University Delhi-NCR, Sonepat, Haryana, India. He obtained his PhD and M Tech in Computer Science and Engineering from Jaypee University of Information Technology, Waknaghat, Solan, HP, India in 2018 and 2011, respectively. He received B Tech in Information Technology from Uttar Pradesh Technical University, Lucknow, Uttar Pradesh, in 2009. He has more than 12 years of academic/research experience; He has published more than 80 papers in international journals (SCIE/Scopus) and conferences (IEEE/Springer). He is the member of various technical/professional societies such as IEEE, UACEE, and IACSIT. He has been appointed as Editorial/Reviewer Board and Technical Programme Committee member in numerous reputed journals/conferences. He is also serving as an associate, academic, and guest editor in reputed journals and organized special sessions in IEEE/Springer conferences. His research interests include digital and fuzzy image processing, machine learning, IoT, and cloud computing.

Dr. Neeraj Kumar Shukla is currently working as an Associate Professor in the Department of Electrical Engineering at King Khalid University, Abha, Kingdom of Saudi Arabia. He has academic, research, and industry experience of more than 20 years. He did his Ph.D. in Low Power VLSI Design. He has authored more than 110 research publications in Journals of repute and International Conferences, apart from this he has 2 Patents, 1 PhD completed under him. He guided 20 M. Tech Dissertations, 50 B. Tech Projects and conducted 50 Short terms skills development training Programs and 30 expert lectures delivered. He is reviewer in several international conferences and Journal. His research areas are Low Power VLSI Design, Digital Design, Machine Learning.

List of Contributors

Karan Aggarwal
Electronics and Communication Engineering Department
Maharishi Markandeshwar (Deemed to be University)
Ambala, India

Poojan Agrawal
Information and Communication Technology, Pandit Deendayal Energy University, Gandhinagar
Gujarat, India

Sayed Q.Y. Al Khalidi
Department of Information Technology, King Khalid University
Abha, Kingdom of Saudi Arabia

Mohd Dilshad Ansari
Department of Computer Science and Engineering
SRM University Delhi-NCR
Sonepat, India

SMD Azash
Department of Law
School of Law Presidency University
Bengaluru, Karnataka, India

Indu Bala
School of Electrical and Electronics Engineering, Lovely Professional University, Punjab, India

Paolo Barsocchi
Wireless Networks (WN)
Institute of Information Science and Technologies, National Research Council, Pisa, Italy

Pankaj Bhambri
Department of Information Technology, Guru Nanak Dev Engineering College, Ludhiana
Punjab, India

Akash Kumar Bhoi
Wireless Networks (WN)
Institute of Information Science and Technologies, National Research Council, Pisa, Italy; Directorate of Research, Sikkim Manipal University
Gangtok, Sikkim, India

Ruchi Doshi
Department of Computer Science and Engineering, Universidad Azteca
Chalco, Mexico

Nafees Akhter Farooqui
Department of CSE, Koneru Lakshmaiah Education Foundation
Vaddeswaram, Andhra Pradesh, India

Vijay Gali
Department of Electrical & Computer Engineering, Nazarbayev University
Astana, Kazakhstan

Deepika Ghai
School of Electronics and Electrical Engineering, Lovely Professional University, Phagwara, India

Umashankar Ghugar
Department of CSE
OP Jindal University
Raigarh, India

Murat Gök
Department of Computer Engineering
Yalova University, Yalova, Turkey

Nand Kishor Gupta
Department of Electical & Electronic Engineering, Poornima University
Jaipur, India

Mohammed Sayeeduddin Habeeb
Department of Electrical Engineering
College of Engineering, King Khalid University, Abha, KSA

Mohd. Haleem
Department of Computer Science
Era University, Lucknow
Uttar Pradesh, India

Kamal Kant Hiran
School of Computer Science and IT, Symbiosis University of Applied Sciences, Indore, India

Mohammad Ishrat
Department of CSE
Koneru Lakshmaiah Education Foundation, Vaddeswaram
Andhra Pradesh, India

Sardar M.N. Islam
Institute for Sustainable Industries and Liveable Cities, Victoria University
Melbourne, Australia

Wasim Khan
Department of CSE
Koneru Lakshmaiah Education Foundation, Vaddeswaram
Andhra Pradesh, India

Bidita Khandelwal
Department of General Medicine
Sikkim Manipal Institute of Medical Sciences, Sikkim Manipal University
Gangtok, Sikkim, India

Alex Khang
Department of Artificial Intelligence
Global Research Institute of Technology and Engineering
Fort Raleigh City, NC, USA

P. Pavan Kumar
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad
Telangana, India

Prakash Kuppuswamy
Department of Computer Science and Engineering, SRM University
Delhi-NCR, Sonepat, India

Maad M. Mijwil
Computer Techniques Engineering Department, Baghdad College of Economic Sciences University
Baghdad, Iraq

M. Mohan
Department of Computer Science and Engineering, SRM University Delhi-NCR, Sonepat, India

Sunpreet Kaur Nanda
Electronics and Telecommunication Engineering Department
P.R. Pote Patil College of Engineering and Management
Sant Gadge Baba Amravati University
Amravati, India

Rakesh Nayak
Department of CSE
OP Jindal University
Raigarh, India

O. Obulesu
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad
Telangana, India

Ranjit Panigrahi
Department of Computer Applications, Sikkim Manipal Institute of Technology
Sikkim Manipal University, India

Ameya Patil
Faculty of School of Business
MIT World Peace University
Pune, India

Balla Adi Narayana Raju
School of Electronics and Electrical Engineering, Lovely Professional University
Phagwara, India

Ch. Rajaramesh
Department of CSE, Vignan Institute of Technology and Science
Telangana, India

Sita Rani
Department of Computer Science and Engineering, Guru Nanak Dev Engineering College, Ludhiana
Punjab, India

Shantanu Saha
Faculty of Ramcharan School of Leadership, MIT World Peace University, Pune, India

Santosh Kumar Satapathy
Information and Communication Technology, Pandit Deendayal Energy University, Gandhinagar
Gujarat, India

Rajeev Sengupta
Faculty of School of Business
Dr. Vishwanath Karad MIT World Peace University, Pune, India

Namra Shah
Information and Communication Technology, Pandit Deendayal Energy University, Gandhinagar
Gujarat, India

Potala Venkata Subbaiah
Department of Marketing
GITAM School of Business, GITAM
(Deemed to be University)
Visakhapatnam
Andhra Pradesh, India

Suman Lata Tripathi
School of Electronics and Electrical
Engineering, Lovely Professional
University, Phagwara, India

Vivek Upadhyaya
Department of Electrical & Electronics
Engineering, Poornima University
Jaipur, India

N. Venkateswarulu
Department of Computer Science and
Engineering, Koneru Lakshmaiah
Education Foundation, Hyderabad
Telangana, India; Department of
Computer Science and Engineering
G. Narayanamma Institute of
Technology and Science, Hyderabad
Telangana, India

Lucia Vilcekova
Faculty of Management, Comenius
University Bratislava, Bratislava
Slovakia

Preface

In the dynamic landscape of contemporary data analytics, the intelligent techniques have been imperatively integrated for unlocking the full potential of predictive insights. This edited volume, titled *Intelligent Techniques for Predictive Data Analytics*, emerges as a comprehensive compendium that navigates the intricacies of data mining and predictive data analysis across diverse domains, including healthcare, agriculture, and other application areas.

The exponential growth of data in today's digital age has spurred an unprecedented demand for advanced analytical methodologies. Within this context, the chapters in this book collectively serve as a guide through the vast terrain of intelligent techniques, offering both foundational principles and innovative applications. The collaborative efforts of experts from various fields have culminated in a rich tapestry of knowledge aimed at advancing the state of the art in predictive analytics. The book extends its reach into various application areas, showcasing the versatility of intelligent techniques; from healthcare to agriculture, from financial forecasting to customer relationship management, from fraud detection to energy consumption prediction, the chapters provide a panoramic view of the impact of predictive analytics across industries.

The book is structured into 12 chapters, each dedicated to a specific facet of intelligent techniques in predictive data analytics.

The journey begins with Chapter 1 focused on various data mining techniques implemented for the intelligent predictive techniques in various application areas as covered by different chapters of this work and beyond, following which an exploration of the challenges in building predictive models and potential solutions to address them is presented in Chapter 2.

Chapter 3 explores the realm of AI-driven Digital Twins and their indispensable role within the Industry 4.0 ecosystem.

Chapter 4 presents the use of predictive data techniques in the field of healthcare, elucidating how intelligent techniques can revolutionize diagnosis, treatment planning, and patient outcomes. Following which, Chapter 5 focuses

on accurate interpretation of physiological signals associated with sleep disorders based on the signal waveform for automated sleep stage scoring systems to diagnose sleep disorders.

Chapter 6 presents the use of predictive analytics for marketing and sales of products using smart trolley with automated billing system in shopping malls.

Chapter 7 explores the predictive analytics, paired with relevant financial data, technical indicators, and advanced statistical approaches, as a potent tool for improving stock market forecasting.

Chapter 8 discusses the escalating challenges in cybersecurity due to the surge in data, communication systems, and cyber threats. The goal is to enhance cybersecurity by leveraging intelligent AI systems for dynamic decision-making and automated security updates in the face of evolving cyber threats.

Chapters 9 and 10 together explore the integration of cutting-edge technologies with conventional farming methods for precision agriculture with fine-grid crop management. Large amount of data produced with smart sensors is used intelligently for plant disease prediction and providing controlled, data-driven insights to farmers to manage decisions that optimize resource utilization, increase yields, and promote sustainable farming methods.

Chapter 11 sheds light on the use of intelligent techniques for predictive maintenance and load forecasting to optimize operations, enhance reliability, and save maintenance costs by proactively scheduling maintenance tasks in various industries.

Forecasting the likelihood of a future event occurring in several facets of a company's operations, including customer churn, loan default, staff performance, and so on, can save another big cost for an industry. Lastly, Chapter 12 targets to study the use of predictive analytics tools for predicting employee turnover and to classify the employee group into stayer and leaver so that effective planning of human resource can be made.

As editors, our goal has been to curate a collection that not only reflects the current landscape of intelligent techniques but also anticipates future trends and challenges.

We believe that this book will be an invaluable resource for researchers, professionals, and students seeking to deepen their understanding of predictive data analytics and harness the power of intelligent techniques. The interdisciplinary nature of the content ensures that readers from diverse backgrounds will find relevant insights, fostering a holistic appreciation for the transformative potential of predictive analytics in our data-driven world.

Happy reading!
Editors

Acknowledgments

As editors, we extend our thanks with profound gratitude to the multitude of individuals who have contributed to the creation of this edited volume.

First and foremost, we express our deepest appreciation to the esteemed authors whose expertise, dedication, and willingness to share their knowledge have been pivotal in shaping this collection into a comprehensive resource.

Editors are indebted to their respective organizations, colleagues, and mentors whose unwavering support and invaluable guidance have been instrumental throughout every stage of this project. Your insightful feedback, encouragement, and scholarly counsel have undoubtedly elevated the quality of this work.

We extend our sincere gratitude to the team at Wiley for their professionalism, expertise, and tireless efforts in bringing this project to fruition. From initial concept to final publication, your meticulous attention to detail and commitment to excellence have been evident at every turn.

Special recognition is also due to the reviewers whose constructive feedback and thoughtful critiques have helped refine and strengthen the content of this volume. Your expertise and dedication to advancing scholarly discourse are deeply appreciated.

To our families and loved ones, we owe a debt of gratitude for your unwavering patience, understanding, and encouragement throughout the countless hours spent immersed in this project. Your steadfast support has been a source of strength and inspiration.

Lastly, we dedicate this book to the readers—the students, scholars, practitioners, and enthusiasts—who will engage with its contents. May the book serve as a beacon of knowledge, sparking curiosity, fostering critical thinking, and inspiring further inquiry in the years to come.

Editors

1

Data Mining for Predictive Analytics

Prakash Kuppuswamy[1], Mohd Dilshad Ansari[1], M. Mohan[1], and Sayed Q.Y. Al Khalidi[2]

[1] *Department of Computer Science and Engineering, SRM University Delhi-NCR, Sonepat, India*
[2] *Department of Information Technology, King Khalid University, Abha, Kingdom of Saudi Arabia*

1.1 Introduction

Data mining is the process of discovering meaningful patterns, correlations, and relationships in large datasets using computational algorithms [1–3]. It involves various steps, including data preprocessing, pattern discovery, and result interpretation, extracting knowledge, and making informed decisions from large and complex datasets [4–6]. It combines techniques from various domains such as statistics, machine learning, artificial intelligence, and database systems to explore and analyze structured and unstructured data to gain insights and generate actionable information [7–9]. This chapter aims to delve into the concept of data mining, its methods, applications, challenges, and future research scope.

The primary goals of data mining are prediction, classification, association, clustering, and anomaly detection [10, 11]. These techniques help businesses and researchers derive valuable insights, support decision-making processes, and uncover hidden patterns within data [12, 13]. In Figure 1.1, the process of data mining typically includes data collection, data preprocessing, feature selection, algorithm selection, pattern discovery, interpretation, and evaluation. Each step plays a crucial role in extracting meaningful information from the data [14–16].

Intelligent Techniques for Predictive Data Analytics, First Edition.
Edited by Neha Singh, Shilpi Birla, Mohd Dilshad Ansari, and Neeraj Kumar Shukla.

Published 2024 by John Wiley & Sons, Inc.

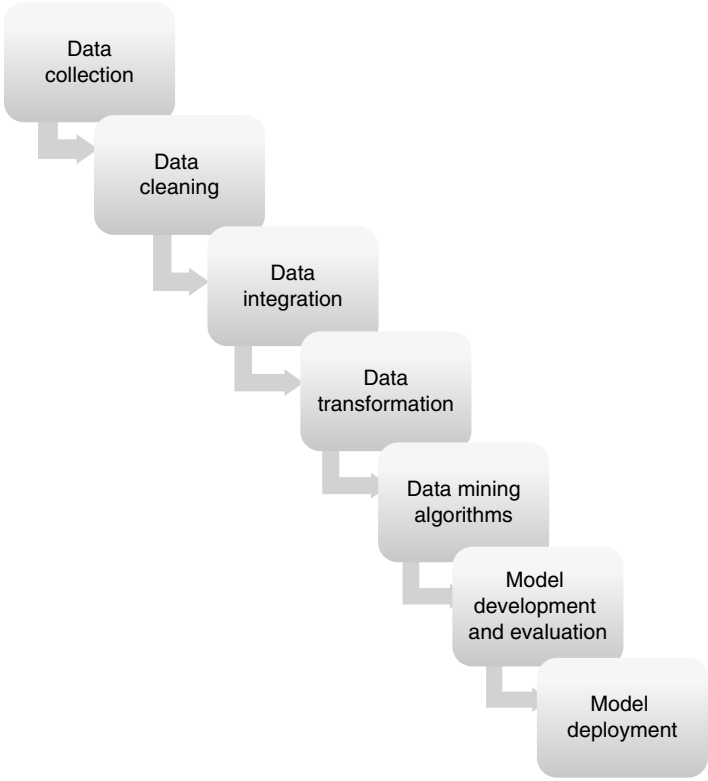

Figure 1.1 Stages of data mining process.

1.1.1 Data Mining Process

Data mining involves a systematic process to extract knowledge or patterns from large datasets. The data mining process typically consists of the following stages:

a) **Data Collection:** Collecting relevant and comprehensive datasets is crucial for accurate predictions. This data can be obtained from internal databases, third-party sources, or through data scraping techniques [17, 18].
b) **Data Cleaning:** Raw data often contains inconsistencies, missing values, and anomalies. Data cleaning involves removing or correcting these errors to ensure the accuracy and reliability of the data [19, 20].
c) **Data Integration:** In many cases, data is obtained from multiple sources and requires to be integrated into a unified dataset. This stage involves consolidating and merging data from different sources to establish a comprehensive dataset for analysis [21, 22].
d) **Data Transformation:** Data transformation involves converting raw data into a suitable format for analysis. This process includes data normalization, standardization, and feature engineering [13, 23].

e) **Data Mining Algorithms:** Selecting the appropriate algorithms based on the nature of the data and the desired outcome is crucial. Various data mining algorithms such as decision trees, neural networks, and association rules can be employed for predictive analytics [24, 25].
f) **Model Development and Evaluation:** Once the data mining algorithms have been applied, models are developed using historical data. These models are subsequently evaluated using performance metrics to assess their accuracy and effectiveness [26, 27].
g) **Model Deployment:** Once the models are considered reliable, they can be deployed in real-world scenarios to make predictions based on new data [28, 29].

Data analytics tools encompass a range of software and platforms designed to enable the extraction, analysis, and visualization of data. These tools utilize various techniques, including data mining, statistical analysis, machine learning, and visualization, to transform raw data into actionable insights. Data analytics tools come with a variety of features that enable efficient data mining, including data integration, data preprocessing, exploratory data analysis, statistical modeling, machine learning algorithms, and visualization capabilities. Using data analytics tools for data mining offers several advantages, including improved decision-making, enhanced productivity, increased efficiency, greater accuracy, and the ability to uncover hidden patterns and trends in data [12, 30, 31]. The most popular data mining techniques are as follows:

- Classification
- Clustering
- Association Rule Mining
- Regression
- Anomaly Detection

1.2 Background Study

There is a great deal of data being generated worldwide in the modern era, which is the age of data. An analysis of small, limited-dimensional data can be conducted using a basic computer. In the medical sciences, the utilization of data mining techniques has emerged as an emerging trend. However, traditional text-mining approaches are not adequate to manage the current surge in published data [32]. To find hidden features and correlations in large volumes of data, artificial intelligence-based text mining tools are being developed and used. In the field of industrial data mining, large-scale data and predictive analytics present the greatest challenges [33]. In the context of extracting and analyzing underlying knowledge from a data repository, it is necessary to use data mining and machine learning techniques due to the rapid changes in the industrial environment and the availability of huge customer data. Better administration will enable better decisions in the future to manage resources and improve efficiency [34]. With the