

Intelligent Techniques for Predictive Data Analytics



Edited by

Neha Singh • Shilpi Birla

Mohd Dilshad Ansari • Neeraj Kumar Shukla

 IEEE Press

WILEY

**Intelligent Techniques for Predictive
Data Analytics**

IEEE Press

445 Hoes Lane
Piscataway, NJ 08854

IEEE Press Editorial Board

Sarah Spurgeon, *Editor-in-Chief*

Moeness Amin
Jón Atli Benediktsson
Adam Drobot
James Duncan

Ekram Hossain
Brian Johnson
Hai Li
James Lyke
Joydeep Mitra

Desineni Subbaram Naidu
Tony Q. S. Quek
Behzad Razavi
Thomas Robertazzi
Diomidis Spinellis

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*

 **IEEE Press**

WILEY

Copyright © 2024 by The Institute of Electrical and Electronics Engineers, Inc.
All rights reserved.

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

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

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://lcn.loc.gov/2024005729>

LC ebook record available at <https://lcn.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 *xiii*

List of Contributors *xv*

Preface *xix*

Acknowledgments *xxi*

1	Data Mining for Predictive Analytics	1
	<i>Prakash Kuppuswamy, Mohd Dilshad Ansari, M. Mohan, and Sayed Q.Y. Al Khalidi</i>	
1.1	Introduction	1
1.1.1	Data Mining Process	2
1.2	Background Study	3
1.3	Applications of Data Mining	4
1.3.1	Healthcare Industry	4
1.3.2	Finance and Banking	5
1.3.3	Retail and E-commerce	5
1.3.4	Manufacturing and Supply Chain	6
1.3.5	Recommender Systems	6
1.3.6	Social Media Analysis	6
1.4	Challenges of Data Analytics in Data Mining	7
1.5	Significance of Data Analytics Tools for Data Mining	7
1.6	Life Cycle of Data Analytics	8
1.6.1	Defining the Problem	8
1.6.2	Data Collection and Preparation	8
1.6.3	Data Exploration and Visualization	9
1.6.4	Model Development	10
1.6.5	Model Training and Validation	10
1.6.6	Model Evaluation and Selection	10
1.6.7	Model Deployment and Integration	10
1.6.8	Model Monitoring and Maintenance	10

1.6.9	Interpretation and Reporting	11
1.7	Predictive Analytics Model	11
1.7.1	Regression Analysis	11
1.7.2	Time Series Analysis	12
1.7.3	Classification	12
1.7.4	Clustering	12
1.7.5	Neural Networks	12
1.7.6	Decision Trees	12
1.7.7	Ensemble Methods	13
1.7.8	Text Analytics	13
1.7.9	Anomaly Detection	14
1.7.10	Collaborative Filtering	14
1.8	Data Analytics Tools	14
1.8.1	RapidMiner	15
1.8.2	Konstanz Information Miner (KNIME)	15
1.8.3	Statistical Analysis System (SAS)	16
1.8.4	Python with Scikit-learn	16
1.8.5	Programming R	17
1.8.6	Waikato Environment for Knowledge Analysis (Weka)	17
1.8.7	Tableau	17
1.8.8	Microsoft Power BI	17
1.9	Benefits of Predictive Analytics Techniques	18
1.10	Applications of Predictive Analytics Model	18
1.10.1	Business Intelligence	19
1.10.2	Healthcare	19
1.10.3	Finance	19
1.10.4	Supply Chain Management	19
1.10.5	Manufacturing	20
1.10.6	Marketing and Advertising	20
1.10.7	Smart Cities	20
1.11	Conclusion	20
	References	21

2 Challenges in Building Predictive Models 25

Rakesh Nayak, Ch. Rajaramesh, and Umashankar Ghugur

2.1	Introduction	25
2.1.1	Different Predictive Models	26
2.1.2	Challenges Associated with Data Acquisition	27
2.2	Literature Survey	30
2.2.1	Challenges in Feature Selection	31
2.2.2	Challenges in Model Selection	32

2.2.3	Challenges in Training and Evaluating the Model	33
2.2.4	Challenges in Deploying and Maintaining the Model Parameter Tuning	34
2.2.5	Challenges in Interpretability, Explain Ability and Scalability	37
2.2.6	Challenges for Interpretability	38
2.2.6.1	Challenges for Explainability	39
2.2.6.2	Challenges for Scalability	40
2.2.6.3	Challenges in Ethical and Regulatory Challenges	41
2.3	Few Suggestions to Overcome the Above Challenges	42
2.4	Conclusion and Future Directions	44
	References	44

3 AI-driven Digital Twin and Resource Optimization in Industry 4.0 Ecosystem 47

Pankaj Bhambri, Sita Rani, and Alex Khang

3.1	Introduction	47
3.1.1	Overview of Industry 4.0	48
3.1.2	Role of AI-driven Digital Twin	49
3.2	Digital Twin Technology	50
3.2.1	Definition and Concepts	51
3.2.2	Components of a Digital Twin	51
3.2.3	Real-Time Data Integration	52
3.3	Industry 4.0 Ecosystem	53
3.3.1	Components and Technologies	53
3.3.2	IoT Integration	54
3.3.3	Cyber-Physical Systems	55
3.4	AI in Digital Twins	56
3.4.1	Machine Learning and AI Algorithms	56
3.4.2	Data Analytics for Predictive Maintenance	56
3.5	Resource Optimization	57
3.5.1	Importance in Industry 4.0	57
3.5.2	Cost Reduction Strategies	58
3.5.3	Energy Efficiency	58
3.6	AI-driven Resource Allocation	59
3.6.1	Resource Monitoring	59
3.6.2	Roles and Types of Optimization Algorithms	60
3.6.3	Optimization Algorithms for AI-Driven Resource Allocation	61
3.7	Challenges and Consideration	62
3.8	Future Trends	62
3.9	Conclusion	63
	References	63

- 4 Predictive Analytics in Healthcare 71**
N. Venkateswarulu, P. Pavan Kumar, and O. Obulesu
- 4.1 Predictive Analytics 71
- 4.2 Predictive Analysis in Medical Imaging 73
- 4.3 Predictive Analytics in the Pharmaceutical Industry 75
- 4.4 Predictive Analytics in Clinical Research 78
- 4.5 AI for Disease Prediction 81
- 4.6 Medical Image Classification for Disease Prediction 83
- References 86
- 5 A Review of Automated Sleep Stage Scoring Using Machine Learning Techniques Based on Physiological Signals 89**
Santosh Kumar Satapathy, Poojan Agrawal, Namra Shah, Ranjit Panigrahi, Bidita Khandelwal, Paolo Barsocchi, and Akash Kumar Bhoi
- 5.1 Introduction 89
- 5.2 Review of Related Works 91
- 5.2.1 Electroencephalogram (EEG) 91
- 5.2.2 Electrocardiogram (ECG) 94
- 5.2.3 Respiratory Effort and Electrooculography 94
- 5.2.4 Multiple Signals Integration (PSG) 96
- 5.3 Methodology 98
- 5.3.1 Preprocessing 100
- 5.3.2 Feature Extraction 100
- 5.3.3 Feature Selection 101
- 5.3.4 ML and DL for Sleep Disorder Detection 101
- 5.3.4.1 Limitations of Traditional Methods 103
- 5.3.4.2 Need for Automated Approaches (ML/DL) 103
- 5.4 Conclusion 105
- 5.5 Future Work 105
- References 106
- 6 Predictive Analytics for Marketing and Sales of Products Using Smart Trolley with Automated Billing System in Shopping Malls Using LBPH and Faster R-CNN 111**
Balla Adi Narayana Raju, Deepika Ghai, Suman Lata Tripathi, Sunpreet Kaur Nanda, and Sardar M.N. Islam
- 6.1 Introduction 111
- 6.2 Major Contributions 112
- 6.3 Related Work 113
- 6.4 Proposed Methodology 119

- 6.4.1 Module 1 – Face Recognition 119
- 6.4.2 Module 2 – Trolley Movement 123
- 6.4.3 Module 3 – Items Detection 123
- 6.4.4 Module 4 – Object and Item Classification 125
- 6.5 Experimental Results and Discussions 126
 - 6.5.1 Experimental Results 126
 - 6.5.2 Discussions on the Security Aspect of IoT 129
- 6.6 Conclusion 130
- References 130

- 7 Enhancing Stock Market Predictions Through Predictive Analytics 135**
Ameya Patil, Shantanu Saha, and Rajeev Sengupta
 - 7.1 Introduction 135
 - 7.2 Factors Influencing Stock Prices 137
 - 7.3 Can Markets Be Predicted? 138
 - 7.3.1 Different Stock Prediction Techniques 138
 - 7.4 Using Predictive Analytics for Stock Prediction 140
 - 7.4.1 Stock Prediction Using ML and AI 141
 - 7.5 Neural Networks 141
 - 7.5.1 Artificial Neural Network 142
 - 7.5.1.1 Recurrent Neural Network (RNN) 143
 - 7.5.1.2 Deep Learning Models 145
 - 7.6 Conclusion 146
 - References 147

- 8 Predictive Analytics and Cybersecurity 151**
Mohammed Sayeeduddin Habeeb
 - 8.1 Introduction 151
 - 8.2 Cybersecurity and Predictive Analysis 152
 - 8.3 Machine Learning 153
 - 8.3.1 Supervised and Unsupervised Learning 153
 - 8.3.2 ML Approach to Cybersecurity 154
 - 8.4 Proactive Cybersecurity and Real-Time Threat Detection 156
 - 8.4.1 Understanding Proactive Cybersecurity 157
 - 8.4.2 Real-Time Threat Detection 158
 - 8.4.3 Use of Predictive Analytics for Threat Detection 158
 - 8.5 Network Security Analytics 159
 - 8.5.1 Enhancing Network Security with Predictive Analytics 159
 - 8.5.2 The Role of ML in Network Security 159
 - 8.6 Cyber Risk Analytics 160
 - 8.6.1 Understanding Cyber Risk Management 161

- 8.6.2 Predictive Modeling for Cyber Risk Assessment 161
- 8.7 Impact of Predictive Analytics on the Cybersecurity Landscape 162
- 8.8 Challenges in Applying Predictive Analytics to Cybersecurity 162
- 8.9 Conclusion 164
- References 164

**9 Precision Agriculture and Predictive Analytics:
Enhancing Agricultural Efficiency and Yield 171**

*Nafees Akhter Farooqui, Mohd. Haleem, Wasim Khan,
and Mohammad Ishrat*

- 9.1 Introduction 171
- 9.2 Background 173
 - 9.2.1 IoT and Networks 173
 - 9.2.2 Predictive Analytics 174
 - 9.2.3 Applications of Precision Agriculture System 176
- 9.3 Precision Agriculture Technologies and Methods 178
 - 9.3.1 The Role of GPS in Modern Crop Management 178
 - 9.3.2 Architecture of a Smart Farming Monitoring System 179
 - 9.3.3 Agricultural Economic Management 181
- 9.4 Smart Agriculture Cultivation Recommender System 183
- 9.5 Conclusion 184
- References 185

**10 A Simple Way to Comprehend the Difference and the Significance
of Artificial Intelligence in Agriculture 189**

*Karan Aggarwal, Ruchi Doshi, Maad M. Mijwil, Kamal Kant Hiran,
Murat Gök, and Indu Bala*

- 10.1 Introduction 189
- 10.2 Machine Learning 191
- 10.3 Deep Learning 192
- 10.4 Data Science 193
- 10.5 AI in the Agriculture Industry 194
 - 10.5.1 Techniques in Agriculture 194
 - 10.5.2 Smart Agriculture 195
- 10.6 Conclusions 198
- References 199

11 An Overview of Predictive Maintenance and Load Forecasting 203

Nand Kishor Gupta, Vivek Upadhyaya, and Vijay Gali

- 11.1 Introduction 203
- 11.2 PdM: Revolutionizing Asset Management 204
 - 11.2.1 Development of Maintenance Procedures 204

11.2.1.1	Reactive Maintenance	206
11.2.1.2	Preventive Maintenance	206
11.2.1.3	Predictive Maintenance	206
11.2.2	Data-Driven Insights for Proactive Decision-Making	208
11.2.2.1	Role of IoT and Sensor Networks	208
11.2.2.2	Big Data Analytics and Machine Learning	209
11.2.2.3	Predictive Algorithms	209
11.2.2.4	Real-Time Monitoring and Alerts	209
11.2.2.5	Benefits of Data-Driven Insights	209
11.2.2.6	Limitations and Challenges	209
11.2.3	Benefits of PdM	211
11.2.4	Challenges in Implementing PdM	213
11.3	Load Forecasting: Illuminating the Path Ahead	216
11.3.1	Navigating Energy Demand Variability	216
11.3.1.1	Load Patterns and Peaks	216
11.3.1.2	Seasonal and Weather Influences	216
11.3.2	Techniques for Accurate Load Forecasting	217
11.3.2.1	Time Series Analysis	217
11.3.2.2	Machine Learning Models	218
11.3.2.3	Hybrid Approaches	218
11.3.3	Advantages of Load Forecasting	218
11.3.3.1	Grid Stability and Resource Planning	218
11.3.3.2	Integration of Renewable Energy Sources	218
11.3.3.3	Demand-side Management and Consumer Empowerment	218
11.3.4	Troubleshooting Load Forecasting	220
11.3.4.1	Data Granularity and Quality	220
11.3.4.2	Incorporation of Dynamic Factors	220
11.3.4.3	Regulatory and Market Uncertainties	220
11.3.5	Application Scenarios: From Smart Cities to Rural Electrification	220
11.3.5.1	Smart Grids: Enabling DR	221
11.3.5.2	Developing Countries: Bridging the Energy Gap	222
11.4	Synergies and Future Prospects	222
11.4.1	Convergence of PdM and Load Forecasting	223
11.4.1.1	Holistic Asset Management and Increasing Efficiency	223
11.4.1.2	Real-time Decision-Making for Energy Systems	223
11.4.2	The Role of Artificial Intelligence (AI) and Edge Computing	223
11.4.2.1	Edge Analytics for Localized Statistics	223
11.4.2.2	Automation of Maintenance and Energy Operations Controlled by AI	224
11.4.3	Sustainability and Environmental Impact	224
11.4.3.1	Reduced Carbon Footprint through Optimized Operations	224

11.4.3.2	Integration of Electric Vehicles (EVs) and Energy Storage	224
11.4.4	Ethical Considerations and Data Privacy	224
11.4.4.1	Balancing Data Utilization and Consumer Privacy	224
11.4.4.2	Transparency and Accountability in Automated Systems	224
11.4.5	Future Trajectories and Uncharted Territories	225
11.4.5.1	Cognitive PdM: Merging AI and Human Expertise	225
11.4.5.2	Load Forecasting in a Decentralized Energy Landscape	225
11.5	Conclusion	225
	References	226

12 Predictive Analytics: A Tool for Strategic Decision of Employee Turnover 231

SMD Azash, Potala Venkata Subbaiah, and Lucia Vilcekova

12.1	Introduction	231
12.2	Literature Review	232
12.3	Need and Importance of the Study	233
12.4	Objectives of the Study	235
12.5	Hypothesis of the Study	235
12.6	Research Method	235
12.7	Data Analysis Procedures and Discussion	236
12.8	Recommendations	240
12.9	Conclusion	241
	References	242

Index 245

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, Sonapat, 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, UACCE, 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
Sonapat, 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 Electrical & 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, Sonapat, 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, Sonapat, 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 Kuppaswamy¹, Mohd Dilshad Ansari¹,
M. Mohan¹, and Sayed Q.Y. Al Khalidi²

¹ Department of Computer Science and Engineering, SRM University Delhi-NCR, Sonapat, 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.

© 2024 The Institute of Electrical and Electronics Engineers, Inc.

Published 2024 by John Wiley & Sons, Inc.

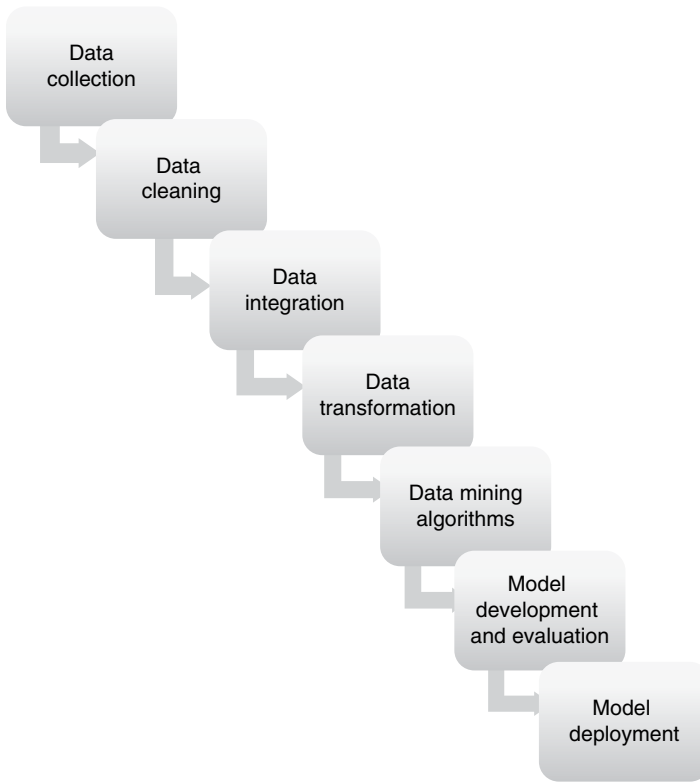


Figure 1.1 Stages of data mining process.

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