

1. Record Nr.	UNINA9910954327303321
Titolo	Data science : theory, analysis, and applications / / edited by Qurban A Memon, Shakeel Ahmed Khoja
Pubbl/distr/stampa	Boca Raton : , : CRC Press, , [2020]
ISBN	0-429-55882-1 0-429-26379-1 0-429-55435-4
Edizione	[1st ed.]
Descrizione fisica	1 online resource (345 pages)
Disciplina	006.312
Soggetti	Data mining - Statistical methods
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Cover -- Half Title -- Title Page -- Copyright Page -- Dedication -- Contents -- Preface -- Editors -- Contributors -- PART I: Data Science: Theory, Concepts, and Algorithms -- Chapter 1 Framework for Visualization of GeoSpatial Query Processing by Integrating MongoDB with Spark -- 1.1 Introduction -- 1.1.1 Integration of Spark and MongoDB -- 1.2 Literature Survey -- 1.3 Proposed System -- 1.3.1 Methodology for Processing Spatial Queries -- 1.3.2 Spark Master-Slave Framework -- 1.3.3 Algorithms for Sharding -- 1.3.3.1 Algorithm for Range Sharding -- 1.3.3.2 Algorithms for Zone Sharding -- 1.3.4 Dataset and Statistics -- 1.4 Results and Performance Evaluation -- 1.5 Conclusion -- References -- Chapter 2 A Study on Metaheuristic-Based Neural Networks for Image Segmentation Purposes -- 2.1 Introduction -- 2.2 Supervised Image Segmentation -- 2.3 Literature Review -- 2.4 Artificial Neural Networks -- 2.5 Optimization -- 2.6 Metaheuristic Algorithms -- 2.6.1 Genetic Algorithm -- 2.6.2 Particle Swarm Optimization Algorithm -- 2.6.3 Imperialist Competitive Algorithm -- 2.7 Optimization of the Neural Networks Weights Using Optimization Algorithms -- 2.8 Experimental Setup and Method Analysis -- 2.9 Conclusions -- References -- Chapter 3 A Study and Analysis of a Feature Subset Selection Technique Using Penguin Search Optimization Algorithm -- 3.1 Introduction -- 3.2 Literature Review -- 3.3 Proposed Work -- 3.3.1 Pseudocode of the Proposed FS-PeSOA Algorithm --

3.3.2 Discussion -- 3.3.2.1 Hunting Strategy of Penguins -- 3.3.2.2 Fitness Function Evaluation -- 3.3.2.3 Position Update Logic -- 3.3.2.4 Oxygen Update Logic -- 3.4 Result Analysis -- 3.5 Conclusions -- References -- Chapter 4 A Physical Design Strategy on a NoSQL DBMS -- 4.1 Introduction -- 4.2 Motivation Example -- 4.3 Neo4j -- 4.4 Design Guidelines -- 4.5 Physical Design. 4.5.1 Query Rewriting Using Path Redundancy Pattern -- 4.5.2 Query Rewriting Using Minimal Query Pattern -- 4.5.3 Path Materialization -- 4.5.4 Index Creation -- 4.6 Experimental Study -- 4.6.1 Experimental Design -- 4.6.2 Impact of the Proposed Physical Design on Query Performance for a 1 GB Database -- 4.6.3 Impact of the Proposed Physical Design on Query Performance for a 10 GB Database -- 4.6.4 Impact of the Proposed Physical Design on Query Performance for a 100 GB Database -- 4.7 Related Work -- 4.8 Discussion -- 4.9 Future Research Directions -- 4.10 Conclusion -- References -- Chapter 5 Large-Scale Distributed Stream Data Collection Schemes -- 5.1 Introduction -- 5.2 Data Collection Scheme for Distributed TBPS -- 5.2.1 Assumed Environment -- 5.2.1.1 Assumed TBPS Architecture -- 5.2.1.2 Assumed Overlay for Distributed TBPS -- 5.2.2 Proposed Method -- 5.2.2.1 Methodology Principle -- 5.2.2.2 Collective Store and Forwarding -- 5.2.2.3 Adaptive Data Collection Tree -- 5.2.3 Evaluation -- 5.2.3.1 Simulation Parameters -- 5.2.3.2 Simulation Results -- 5.3 Data Collection Scheme Considering Phase Differences -- 5.3.1 Problems Addressed -- 5.3.1.1 Assumed Environment -- 5.3.1.2 Input Setting -- 5.3.1.3 Definition of a Load -- 5.3.2 Proposed Method -- 5.3.2.1 Skip Graphs -- 5.3.2.2 Phase Differences -- 5.3.3 Evaluation -- 5.3.3.1 Collection Target Nodes -- 5.3.3.2 Communication Loads and Hops -- 5.4 Discussion -- 5.5 Related Work -- 5.6 Conclusion -- Acknowledgements -- References -- PART II: Data Design and Analysis -- Chapter 6 Big Data Analysis and Management in Healthcare -- 6.1 Introduction -- 6.2 Preliminary Studies -- 6.3 Healthcare Data -- 6.4 Need of Big Data Analytics in Healthcare -- 6.5 Challenges in Big Data Analysis in Healthcare -- 6.5.1 Capture -- 6.5.2 Cleaning -- 6.5.3 Storage -- 6.5.4 Security -- 6.5.5 Stewardship -- 6.5.6 Querying. 6.5.7 Reporting -- 6.5.8 Visualization -- 6.5.9 Updating -- 6.5.10 Sharing -- 6.6 Collection of Healthcare Data -- 6.6.1 Importance in Healthcare Data Collection -- 6.6.2 Complications and Clarifications of Healthcare Data Collection -- 6.6.3 Current Data Collection Methods -- 6.6.4 Advanced Data Collection Tools -- 6.6.5 Healthcare Data Standards -- 6.6.6 Inferences of Patient Data Collection in Healthcare -- 6.7 Analysis of Healthcare Data -- 6.8 Healthcare Data Management -- 6.8.1 Big Data and Care Management -- 6.8.2 Advantages of Healthcare Data Management -- 6.9 Big Data in Healthcare -- 6.9.1 Big Data and IoT -- 6.9.2 Patient Prophecies for Upgraded Staffing -- 6.9.3 Electronic Health Records -- 6.9.4 Real-Time Warning -- 6.9.5 Augmenting Patient Engagement -- 6.9.6 Using Health Data for Informed Strategic Planning -- 6.9.7 Extrapolative Analytics in Healthcare -- 6.9.8 Diminish Fraud and Enrich Security -- 6.9.9 Telemedicine -- 6.9.10 Assimilating Big Data per Medical Imaging -- 6.9.11 A Method to Avert Pointless ER (Emergency Room) Visits -- 6.10 Future for Big Data in Healthcare -- 6.11 Conclusion -- References -- Chapter 7 Healthcare Analytics: A Case Study Approach Using the Framingham Heart Study -- 7.1 Introduction and Background to the Case Study: Framingham Heart Study -- 7.2 Literature Review -- 7.3 Introduction to the Data Analytics Framework -- 7.3.1 Step 1. Defining the Healthcare Problem -- 7.3.2 Step 2. Explore the Healthcare Data -- 7.3.3 Step 3. Predict What Is Likely to Happen -- or Perform

Classification Analysis -- 7.3.4 Step 4. Check the Modeling Results -- 7.3.5 Step 5. Optimize (Find the Best Solution) -- 7.3.6 Step 6. Derive a Clinical Strategy for Patient Care and Measure the Outcome -- 7.3.7 Step 7. Update the CDS System -- 7.4 Data Exploration and Understanding of the Healthcare Problem.

7.5 Machine-Learning Model Application -- 7.6 Evaluation of the Machine-Learning Model Results -- 7.7 Conclusion -- 7.8 Future Direction -- Acknowledgements -- References -- Chapter 8

Bioinformatics Analysis of Dysfunctional (Mutated) Proteins of Cardiac Ion Channels Underlying the Brugada Syndrome -- 8.1 Introduction -- 8.2 Results -- 8.2.1 Brief Description of Unique BrS-Related Proteins -- 8.2.2 PIM-Based Analysis of the Unique BrS-Related Proteins -- 8.2.3 Intrinsic Disorder Analysis of the BrS-Related Proteins -- 8.2.4 Kolmogorov-Smirnov Test -- 8.3 Discussion -- 8.4 Materials and Methods -- 8.4.1 Evaluation of Polar Profile -- 8.4.1.1 Weighting of Polar Profiles -- 8.4.1.2 Comparison of Polar Profiles -- 8.4.1.3 Graphics of Polar Profiles -- 8.4.2 Evaluation of Intrinsic Disorder Predisposition -- 8.4.3 Data Files -- 8.4.4 Kolmogorov-Smirnov Test -- 8.4.5 Test Plan -- 8.4.5.1 Polar Profile -- 8.5 Conclusions -- References -- Chapter 9

Discrimination of Healthy Skin, Superficial Epidermal Burns, and Full-Thickness Burns from 2D-Colored Images Using Machine Learning -- 9.1 Introduction -- 9.2 Literature Review -- 9.2.1 Skin Burns -- 9.2.2 Causes of Burn Injuries -- 9.2.3 Burns Category -- 9.2.4 Burn Assessment Techniques -- 9.2.4.1 Clinical Assessment -- 9.2.4.2 Blood Perfusion Measurement -- 9.3 Machine Learning -- 9.3.1 Convolutional Neural Networks -- 9.3.1.1 Convolution Layer -- 9.3.1.2 Pooling Layer -- 9.3.1.3 Output/Classification Layer -- 9.3.2 Training a ConvNet -- 9.3.3 Common ConvNet Models -- 9.3.3.1 AlexNet -- 9.3.3.2 GoogleNet -- 9.3.3.3 VGGNet -- 9.3.3.4 Residual Network -- 9.4 Goals and Methodology -- 9.4.1 Image Acquisition and Preprocessing -- 9.4.2 Feature Extraction and Classification -- 9.5 Results and Discussion -- 9.5.1 Terms Related to Contingency Table -- 9.5.2 Classifier Performance -- 9.6 Conclusions -- References.

Chapter 10 A Study and Analysis of an Emotion Classification and State-Transition System in Brain Computer Interfacing -- 10.1 Introduction -- 10.2 Literature Review -- 10.3 Proposed Work -- 10.3.1 Classification Processes -- 10.3.1.1 SVM Classifier -- 10.3.1.2 KNN Classifier -- 10.3.1.3 Random Forest Classifier -- 10.3.2 State-Transition Machine -- 10.3.2.1 Proposed Algorithm of Emotional State Transition Based on Channel Value for a Fixed Time Interval -- 10.4 Result Analysis -- 10.4.1 Requirement -- 10.4.2 Result Comparisons of SVM, KNN, and Random Forest Classifiers -- 10.4.3 SVM Polynomial Kernel Performance Analysis -- 10.4.4 Analysis of the State-Transition Machine -- 10.4.5 Comparison with Previous Works -- 10.4.6 Computational Complexity -- 10.5 Conclusion -- Acknowledgment -- References -- PART III: Applications and New Trends in Data Science -- Chapter 11

Comparison of Gradient and Textural Features for Writer Retrieval in Handwritten Documents -- 11.1 Introduction -- 11.2 Literature Review -- 11.3 Adopted Features -- 11.3.1 Local Binary Pattern -- 11.3.2 Histogram of Oriented Gradients -- 11.3.3 Gradient Local Binary Pattern -- 11.3.4 Pixel Density -- 11.3.5 Run Length Feature -- 11.4 Matching Step -- 11.5 Experimental Evaluation -- 11.5.1 Evaluation Criteria -- 11.5.2 Experimental Setup -- 11.5.3 Retrieval Results -- 11.6 Discussion and Comparison -- 11.7 Conclusion -- References -- Chapter 12

A Supervised Guest Satisfaction Classification with Review Text and Ratings -- 12.1 Introduction -- 12.2 Related Literature -- 12.2.1 Guest Satisfaction and

Online Reviews -- 12.3 Methodology -- 12.3.1 Data Description and Analysis -- 12.3.2 Data Cleaning -- 12.3.3 Latent Semantic Analysis -- 12.3.4 Classifiers and Performance Measures -- 12.4 Experimental Results -- 12.4.1 Features Related to Guest Satisfaction. 12.4.2 Hotel Guest Satisfaction Prediction.

Sommario/riassunto

"The aim of this book is to provide an internationally respected collection of scientific research methods, technologies and applications in the area of data science. This book can prove useful to the researchers, professors, research students and practitioners as it reports novel research work on challenging topics in the area surrounding data science. In this book, some of the chapters are written in tutorial style concerning machine learning algorithms, data analysis, information design, infographics, relevant applications, etc. The book is structured as follows: Part 1: Data Science: Theory, Concepts, and Algorithms This part comprises five chapters on data Science theory, concepts, techniques and algorithms. Part II: Data Design and Analysis This part comprises five chapters on data design and analysis. Part III: Applications and New Trends in Data Science This part comprises four chapters on applications and new trends in data science"--
