1. Record Nr. UNINA9910877333003321

Autore	Kumar Kukatlapalli Pradeep
Titolo	Data Engineering and Data Science : Concepts and Applications
Pubbl/distr/stampa	Newark : , : John Wiley & Sons, Incorporated, , 2023 ©2023
ISBN	1-119-84199-2 1-119-84198-4
Edizione	[1st ed.]
Descrizione fisica	1 online resource (467 pages)
Altri autori (Persone)	UnalAynur PillaiVinay Jha MurthyHari NiranjanamurthyM
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Cover -- Title Page -- Copyright Page -- Contents -- Preface -- Chapter 1 Quality Assurance in Data Science: Need, Challenges and Focus -- 1.1 Introduction -- 1.1.1 Quality Assurance and Testing -- 1.1.2 Data Science and Quality Assurance -- 1.1.3 Background -- 1.2 Testing and Quality Assurance -- 1.2.1 Key Terminologies Associated With Testing -- 1.3 Product Quality and Test Efforts -- 1.3.1 Testing Metrics -- 1.3.2 How to Improve the Business Value to Products Using Test Automation -- 1.3.3 Data Analysis and Management in Test Automation -- 1.3.4 Data Models in Data Science -- 1.4 Data Masking in Data Model and Associated Risks -- 1.5 Prediction in Data Science -- Case Study -- 1.6 Role of Metrics in Evaluation -- 1.7 Quantity of Data in Quality Assurance -- 1.8 Identifying the Right Data Sources -- 1.8.1 Need to Gather Up-to-Date Data -- 1.8.2 Synthesising Existing Advanced Technologies for Continuous Business Improvements -- 1.9 Conclusion -- References -- Chapter 2 Design and Implementation of Social Media Mining - Knowledge Discovery Methods for Effective Digital Marketing Strategies -- 2.1 Introduction -- 2.1.1 Objectives of the Study -- 2.2 Literature Review -- 2.3 Novel Framework for Social Media Data Mining and Knowledge Discovery -- 2.4 Classification for

Comparison Analysis -- 2.5 Clustering Methodology to Provide Digital Marketing Strategies -- 2.5.1 Status (Text Form) -- 2.5.2 Images (Photos) -- 2.5.3 Video Post -- 2.5.4 Link Post -- 2.6 Experimental Results -- 2.7 Conclusion -- References -- Chapter 3 A Study on Big Data Engineering Using Cloud Data Warehouse -- 3.1 Introduction -- 3.2 Comparison Study of Different Cloud Data Warehouses -- 3.2.1 Amazon Redshift -- 3.2.2 High-Level Architecture of Amazon Redshift -- 3.2.3 Features of Amazon Redshift Cloud Data Warehouse -- 3.2.4 Pricing of Amazon Redshift Cloud Data Warehouse.
3.3 Snowflake Cloud Data Warehouse -- 3.3.1 High-Level Architecture of Snowflake Cloud Data Warehouse -- 3.3.2 Features of Snowflake Cloud Data Warehouse -- 3.3.3 Snowflake Cloud Data Warehouse Pricing -- 3.4 Google BigQuery Cloud Data Warehouse -- 3.4.1 High-Level Architecture of Google BigQuery Cloud Data Warehouse -- 3.4.2 Features of Google BigQuery Cloud Data Warehouse -- 3.4.3 Google BigQuery Cloud Data Warehouse Pricing -- 3.5 Microsoft Azure Synapse Cloud Data Warehouse -- 3.5.1 Microsoft Azure Synapse Cloud Data Warehouse Architecture -- 3.5.2 Features of Microsoft Azure Synapse Cloud Data Warehouse -- 3.5.3 Pricing of Microsoft Azure Synapse Cloud Data Warehouse -- 3.6 Informatica Intelligent Cloud Services (IICS) -- 3.6.1 Informatica Intelligent Cloud Services Architecture -- 3.6.2 Salient Features of Informatica Intelligent Cloud Services -- 3.6.3 Informatica Intelligent Cloud Services Pricing Model -- 3.7 Conclusion -- Acknowledgements -- References -- Chapter 4 Data Mining with Cluster Analysis Through Partitioning Approach of Huge Transaction Data -- 4.1 Introduction -- 4.2 Methodology Used in Proposed Cluster Analysis System -- 4.2.1 Design of Algorithms -- 4.3 Literature Survey on Existing Systems -- 4.3.1 Experimental Results -- 4.4 Conclusion -- References -- Chapter 5 Application of Data Science in Macromodeling of Nonlinear Dynamical Systems -- 5.1 Introduction -- 5.2 Nonlinear Autonomous Dynamical System -- 5.3 Nonlinear System - MOR -- 5.3.1 Proper Orthogonal Decomposition -- 5.4 Data Science Life Cycle -- 5.4.1 Problem Identification -- 5.4.2 Identifying Available Data Sources and Data Collection -- 5.4.3 Data Processing -- 5.4.4 Data Exploration -- 5.4.5 Feature Extraction -- 5.4.6 Modeling -- 5.4.7 Model Performance Evaluation -- 5.5 Artificial Neural Network in Modeling -- 5.5.1 Machine Learning.
5.5.2 Biological Neuron Model -- 5.5.3 Artificial Neural Networks -- 5.5.4 Network Topologies -- 5.5.4.1 NARX Neural Network -- 5.5.5 ANN Modeling Using Mathematical Models -- 5.6 Neuron Spiking Model Using FitzHugh-Nagumo (F-N) System -- 5.6.1 Linearization of F-N System -- 5.6.2 Reduced Order Model of Linear System -- 5.6.3 Finite Difference Discretization of F-N System -- 5.6.4 MOR of F-N System Using POD-Galerkin Method -- 5.7 Ring Oscillator Model -- 5.7.1 Model Order Reduction of Ring Oscillator Circuit -- 5.7.2 Ring Oscillator Circuit Approximation Using Linear System MOR -- 5.7.3 POD-ANN Macromodel of Ring Oscillator Circuit -- 5.8 Nonlinear VLSI Interconnect Model Using Telegraph Equation -- 5.8.1 Macromodeling of VLSI Interconnect -- 5.8.2 Discretisation of Interconnect Model -- 5.8.3 Linearization of VLSI Interconnect Model -- 5.8.4 Reduced Order Linear Model of VLSI Interconnect -- 5.9 Macromodel Using Machine Learning -- 5.9.1 Activation Function -- 5.9.2 Bayesian Regularization -- 5.9.3 Optimization -- 5.10 MOR of Dynamical Systems Using POD-ANN -- 5.10.1 Accuracy and Performance Index -- 5.11 Numerical Results -- 5.11.1 F-N System -- 5.11.2 Ring Oscillator Model -- 5.11.3 Reduced Order POD Approximation of Ring Oscillator -- 5.11.3.1 Study of POD-ANN Approximation of Ring Oscillator for Variation in Amplitude of Input Signal and for Different Input Signals -- 5.11.3.2

POD-ANN Approximation of Ring Oscillator for Variation in Frequency -- 5.11.4 POD-ANN Approximation of VLSI Interconnect -- 5.12 Conclusion -- References -- Chapter 6 Comparative Analysis of Various Ensemble Approaches for Web Page Classification -- 6.1 Introduction -- 6.2 Literature Survey -- 6.3 Material and Methods -- 6.4 Ensemble Classifiers -- 6.4.1 Bagging -- 6.4.1.1 Bagging Meta Estimator -- 6.4.1.2 Random Forest -- 6.4.2 Boosting -- 6.4.2.1 AdaBoost.
6.4.2.2 Gradient Tree Boosting -- 6.4.2.3 XGBoost -- 6.4.3 Stacking -- 6.5 Results -- 6.5.1 Bagging Meta Estimator -- 6.5.2 Random Forest -- 6.5.3 AdaBoost -- 6.5.4 Gradient Tree Boosting -- 6.5.5 XGBoost -- 6.5.6 Stacking -- 6.5.7 Comparison with Single Classifiers -- 6.6 Conclusion -- Acknowledgement -- References -- Chapter 7 Feature Engineering and Selection Approach Over Malicious Image -- 7.1 Introduction -- 7.2 Feature Engineering Techniques -- 7.2.1 Methodologies in Feature Engineering -- 7.2.2 Strides in Feature Engineering -- 7.2.3 Feature Extraction -- 7.2.4 Feature Selection -- 7.2.5 Feature Engineering in Image Processing -- 7.2.6 Importance of Feature Engineering in Image Processing -- 7.3 Malicious Feature Engineering -- 7.4 Image Processing Technique -- 7.4.1 Steps Involved in Image Processing Technique -- 7.4.2 Image Processing Task -- 7.4.2.1 Image Enhancement -- 7.4.2.2 Image Restoration -- 7.4.2.3 Coloring Image Processing -- 7.4.2.4 Wavelets Processing and Multiple Solutions -- 7.4.2.5 Image Compression -- 7.4.2.6 Character Recognition -- 7.4.2.7 Characteristics of Image Processing -- 7.5 Image Processing Techniques for Analysis on Malicious Images -- 7.6 Conclusion -- References -- Blog -- Chapter 8 Cubic-Regression and Likelihood Based Boosting GAM to Model Drug Sensitivity for Glioblastoma -- 8.1 Introduction -- 8.1.1 Glioblastoma -- 8.2 Literature Survey -- 8.3 Materials and Methods -- 8.3.1 Methodology -- 8.3.1.1 Generalized Additive Models (GAMs) -- 8.3.1.2 Model-Based Boosting - Boosted GAM -- 8.3.2 Datasets Description -- 8.4 Evaluations, Results and Discussions -- 8.4.1 Akaike Information Criterion (AIC) -- 8.4.2 Adjusted R-Squared -- 8.4.3 Discussion -- Conclusion -- References -- Chapter 9 Unobtrusive Engagement Detection through Semantic Pose Estimation and Lightweight ResNet for an Online Class Environment.
9.1 Introduction -- 9.2 Related Work -- 9.2.1 Analysis for a Classroom Environment -- 9.2.2 Pose Estimation -- 9.2.3 Face Alignment and Landmark Estimation -- 9.2.4 Deep Networks for Emotional Analysis -- 9.3 Proposed Methodology -- 9.3.1 Data Description -- 9.3.2 Facial Detection and Recognition -- 9.3.2.1 Face Detection -- 9.3.2.2 Facial Landmark Detection -- 9.3.3 Emotion Quantification -- 9.3.4 Pose Estimation -- 9.3.4.1 Facial Pose Estimation -- 9.4 Experimentation -- 9.5 Results and Discussions -- Conclusion -- References -- Chapter 10 Building Rule Base for Decision Making - A Fuzzy-Rough Approach -- 10.1 Introduction -- 10.2 Literature Review -- 10.3 Discretization of the Dataset Using Fuzzy Set Theory -- 10.4 Description of the Dataset -- 10.5 Process Involved in Proposed Work -- 10.6 Experiment -- 10.7 Evaluation Result -- 10.8 Discussion -- Conclusion -- References -- Chapter 11 An Effective Machine Learning Approach to Model Healthcare Data -- 11.1 Introduction -- 11.2 Types of Data in Healthcare -- 11.3 Big Data in Healthcare -- 11.4 Different V's of Big Data -- 11.5 About COPD -- 11.6 Methodology Implemented -- Conclusion -- References -- Chapter 12 Recommendation Engine for Retail Domain Using Machine Learning Techniques -- 12.1 Introduction -- 12.2 Proposed System -- 12.2.1 Classification of Suppliers -- 12.2.2 Recommendation for Buyer -- 12.2.3 Forecasting Using ARIMA

Model -- 12.3 Results -- 12.3.1 ARIMA Forecasting -- 12.4 Conclusion -- References -- Chapter 13 Mining Heterogeneous Lung Cancer from Computer Tomography (CT) Scan with the Confusion Matrix -- 13.1 Introduction -- 13.2 Literature Review -- 13.3 Methodology -- 13.3.1 Description of the Data -- 13.3.2 Image Preprocessing -- 13.3.3 Image Segmentation -- 13.3.4 Image Processing -- 13.3.5 Zero Component Analysis (ZCA) Whitening -- 13.3.6 Local Binary Pattern (LBP Feature). 13.3.7 LESH Vector.