1. Record Nr. UNINA9910878056303321

Titolo	Recent Advances on Soft Computing and Data Mining : Proceedings of the Sixth International Conference on Soft Computing and Data Mining (SCDM 2024), August 21-22, 2024 / / edited by Rozaida Ghazali, Nazri Mohd Nawi, Mustafa Mat Deris, Jemal H. Abawajy, Nureize Arbaiy
Pubbl/distr/stampa	Cham : , : Springer Nature Switzerland : , : Imprint : Springer, , 2024
ISBN	3-031-66965-7
Edizione	[1st ed. 2024.]
Descrizione fisica	1 online resource (451 pages)
Collana	Lecture Notes in Networks and Systems, , 2367-3389 ; ; 1078
Disciplina	006.312
Soggetti	Computational intelligence Engineering - Data processing Artificial intelligence Computational Intelligence Data Engineering Artificial Intelligence
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Intro -- Preface -- Conference Organization -- Contents -- Prediction of OPEC Carbon Dioxide Emissions Using K-Means Clustering and Ensemble Algorithm -- 1 Introduction -- 2 Related Works -- 3 Fuzzy Nearest Neighbor -- 4 Sequential Minimal Optimization -- 5 Logistic Regression -- 6 K-Means Clustering -- 7 Proposed Methodology -- 8 Experimental Setup and Analysis -- 8.1 Assessment Measures -- 8.2 Dataset Description -- 8.3 Simulation Results -- 9 Conclusion -- References -- Detection of Phishing Websites from URLs Using Hybrid Ensemble-Based Machine Learning Technique -- 1 Introduction -- 2 Related Work -- 3 Methodology -- 3.1 Description of the Machine Learning Classifiers -- 3.2 Dataset Description -- 3.3 Model Construction -- 4 Performance Evaluation Metrics -- 5 Result -- 6 Conclusion -- 7 Future Work -- References -- Minimal Data for Maximum Impact: An Indonesian Part-of-Speech Tagging Case Study -- 1 Introduction -- 2 Literature Review -- 3 Methodology -- 3.1 Data Collection -- 3.2 Semi-supervised Learning Preprocessing -- 3.3

Feature Selection -- 3.4 Classification -- 3.5 Evaluation -- 4 Result and Discussion -- 5 Conclusion -- 5.1 Summary -- 5.2 Future Work -- References -- Alleviating Sparsity to Enhance Group Recommendation with Cross-Linked Domain Model -- 1 Introduction -- 2 Literature Review -- 2.1 Group Recommender System -- 2.2 Cross-Domain Recommender System -- 2.3 Linked Open Data -- 3 Methodology -- 3.1 Experiment Setup -- 4 Result and Discussion -- 5 Conclusion and Recommendation -- References -- Evaluating Deep Transfer Learning Models for Detecting Various Face Mask Wearings -- 1 Introduction -- 2 Literature Review -- 2.1 Deep Learning -- 2.2 Transfer Learning -- 2.3 Existing Works -- 3 Methodology -- 4 Results and Discussion -- 5 Conclusions -- References.
Classification of Stunting Events: Case Study in West Java, Indonesia -- 1 Introduction -- 2 Related Work -- 3 Methods -- 3.1 Dataset Collection -- 3.2 Data Pre-processing -- 3.3 Model Comparison -- 3.4 Model Implementation -- 3.5 Model Evaluation -- 4 Results and Discussion -- 5 Conclusion -- References -- The Effects of Data Reduction Using Rough Set Theory on Logistic Regression Model -- 1 Introduction -- 2 The Basic Theories and Methodology -- 2.1 Rough Set Theory (RST) -- 2.2 Logistic Regression Model -- 3 Implementation Hybrid Classification Approach with LR Analysis and RST -- 3.1 Implementation Hybrid Model on Anemia Data Set -- 3.2 Implementation Hybrid Model on Diabetes Data Set -- 3.3 Discussion -- 4 Conclusion -- References -- Robust Heart Disease Prognosis: Integrating Extended Isolation Forest Outlier Detection with Advanced Prediction Models -- 1 Introduction -- 2 Methodology -- 2.1 Summary of Dataset -- 2.2 Data Preprocessing -- 2.3 Machine Learning Techniques -- 2.4 Deep Learning Algorithm -- 2.5 Evaluation Parameters -- 3 Results Evaluation -- 3.1 Implementing the First Strategy, Which Involves Neither Feature Selection nor Outlier Detection -- 3.2 Implementing the 2nd Strategy: Feature Selection Without Outlier Detection -- 3.3 Employing the 3rd Strategy (Feature Selection and Detection of Outliers) -- 4 Conclusion -- References -- Overlapping Granular Clustering: Application in Fuzzy Rule-Based Classification -- 1 Introduction -- 2 GrC-Fuzzy Logic Models -- 2.1 Granular Clustering -- 2.2 Formation of Fuzzy Logic Rule Base -- 3 Overlapping GrC -- 3.1 R-Value -- 3.2 A New Overlapping Measure During the Iterative Data Granulation -- 4 Case Study and Simulation Results -- 5 Interpretability Index -- 6 Conclusion -- References -- Improved Rough-Multiple Regression for Unemployment Rate Model in Indonesia -- 1 Introduction.
2 Variable Framework and Methods -- 2.1 Multiple Linear Regression -- 2.2 Rough Sets Theory -- 3 Results and Discussion -- 3.1 Descriptive Statistics for Unemployment Rate and Its Variables -- 3.2 Multiple Linear Regression Model for Unemployment Rate -- 3.3 Rough-Multiple Regression Model for Unemployment Rate -- 3.4 Comparison Multiple Regression and Rough-Multiple Regression -- 4 Conclusion -- References -- Utilizing Machine Learning for Gene Expression Data: Incorporating Gene Sequencing, K-Mer Counting and Asymmetric N-Grams Features -- 1 Introduction -- 2 Materials and Methods -- 2.1 Data Pre-processing -- 2.2 Classification Model -- 2.3 Performance Metrics -- 3 Result and Discussion -- 4 Conclusion and Future Work -- References -- Text Sentiment Analysis on VIX's Impact on Market Sentiment Dynamics -- 1 Introduction -- 2 Literature Review -- 3 Methodology -- 3.1 Data Collection -- 3.2 Sentiment Analysis of SnowNLP -- 3.3 Sentiment Index -- 3.4 Pearson's Correlation -- 3.5 Linear Correlation -- 3.6 Granger Causality Test -- 4 Empirical Result and Analysis -- 4.1 Data Description and Cleaning -- 4.2 Text

Sentiment Index -- 4.3 Pearson's Correlation -- 4.4 Linear Correlation -- 4.5 Granger Causality Test -- 4.6 Test for Chinese Market -- 5 Conclusion -- References -- Multilevel Monte Carlo Simulation Model for Air Pollution Index Prediction of a Smart Network -- 1 Introduction -- 2 Related Work -- 3 Methods -- 3.1 Monte Carlo Simulation -- 3.2 Multilevel Monte Carlo Simulation -- 3.3 Air Pollution Dataset -- 3.4 Performance Evaluation Metrics -- 4 Results and Discussion -- 4.1 Correlation Analysis -- 5 Conclusion -- References -- An In-Depth Strategy using Deep Generative Adversarial Networks for Addressing the Cold Start in Movie Recommendation Systems -- 1 Introduction -- 2 Related Works -- 3 Research Methodology -- 3.1 Data Preparation. 3.2 Collaborative Filtering with Singular Value Decomposition (CF-SVD) -- 3.3 Generative Adversarial Networks (GANs) -- 3.4 Collaborative Filtering (CF) with SVD and GANs -- 3.5 Content Based Filtering (CB) -- 4 Results -- 5 Conclusion -- References -- Predicting Undergraduate Academic Success with Machine Learning Approaches -- 1 Introduction -- 2 Related Work -- 3 Research Design and Methodology -- 3.1 Dataset Source -- 3.2 Exploratory Data Analysis -- 3.3 Data Preprocessing -- 3.4 Classification Algorithms -- 4 Results and Discussion -- 4.1 Evaluations of Classifiers Using Default Parameters -- 4.2 Model Parameter Optimization by Hyperparameter Tuning -- 5 Conclusion -- References -- Comparative Assessment of Facial Expression Recognition Models for Unraveling Emotional Signals with Convolutional Neural Networks -- 1 Introduction -- 2 Related Work -- 3 Dataset Description -- 4 Methodology -- 4.1 Pre-processing -- 4.2 Feature Extraction -- 4.3 CNN Architecture -- 5 Results -- 6 Discussions and Future Work -- References -- Evaluating Path-Finding Algorithms for Real-Time Route Recommendation System Built using FreeRTOS -- 1 Introduction -- 2 Related Work -- 3 Research Design and Methodology -- 3.1 Adjacency Matrix -- 3.2 Path Finding Algorithms -- 3.3 Real-Time Operating System (RTOS) -- 3.4 Functional Diagram of the Simulated System using FreeRTOS -- 4 Results and Discussion -- 4.1 Validate the accuracy of the recommended route -- 4.2 Performance Evaluation -- 5 Conclusion -- References -- Machine Learning-Based Phishing Website Detection: A Comparative Analysis and Web Application Development -- 1 Introduction -- 2 Literature Review -- 3 Research Design -- 3.1 Dataset Overview -- 3.2 Feature Selection -- 3.3 Detection Techniques Implementation -- 3.4 Performance Evaluation and Comparison -- 3.5 Web Application Development.
4 Results and Discussion -- 5 Conclusion -- References -- Comparative Performance of Multi-level Pre-trained Embeddings on CNN, LSTM and CNN-LSTM for Hate Speech and Offensive Language Detection -- 1 Introduction -- 2 The Architecture of HSOLC Detection Model -- 2.1 Text Embedding Layer -- 2.2 Representation Layer -- 2.3 Output Layer -- 3 Experimental Setup -- 4 Dataset and Results -- 4.1 Results and Discussion -- 5 Conclusion -- References -- Improved Classifier Chain Method Based on Particle Swarm Optimization and Genetic Algorithm for Multilabel Classification Problem -- 1 Introduction -- 1.1 Motivation -- 1.2 Random Label Sequence Ordering (RLSO) -- 2 Related Work -- 3 Method -- 3.1 Dataset -- 3.2 Data Preprocessing -- 3.3 Classification (Proposed Model) -- 3.4 Performance Measures -- 4 Results and Discussion -- 5 Conclusion -- References -- Sentiment Analysis on Umrah Packages Review in Malaysia -- 1 Introduction -- 2 Sentiment Analysis on Social Media -- 2.1 Related Works of Similar -- 2.2 Mobile Phone Reviews from Amazon Using Support Vector Machine -- 2.3 Sentiment Classification of Online Consumer Reviews Using Word Vector

Representation -- 2.4 Online Reviews of Hospitality Services Using Naïve Bayes -- 2.5 Customer Satisfaction Towards Umrah Travel Agencies in Malaysia -- 3 Methodology -- 3.1 Preliminary Study -- 3.2 Data Analysis -- 3.3 Interface and Architecture Design -- 3.4 System Development -- 4 Analysis and Discussions -- 4.1 Naïve Bayes - Gaussian -- 4.2 Naïve Bayes - Multinomial -- 4.3 Support Vector Machine -- 4.4 Random Forest -- 4.5 Analysis -- 5 Conclusion and Recommendations -- References -- Opinion Mining System for Influence Detection Using Machine Learning to Secure Business Reputation -- 1 Introduction -- 2 Related Works -- 2.1 Sentiment Analysis -- 2.2 Supervised Machine Learning Approach -- 3 Methodology.
3.1 Data Preprocessing and Feature Extraction.

Sommario/riassunto

This book explores methods for leveraging data to create innovative solutions that offer significant and meaningful value. It provides practical insights into the concepts and techniques essential for maximizing the outcomes of large-scale research and data mining projects. Readers are guided through analytical thinking processes, addressing challenges in deciphering complex data systems and deriving commercial value from the data. Soft computing and data mining, also known as data-driven science, encompass a diverse range of interdisciplinary scientific methods and processes. The proceedings of "Recent Advances on Soft Computing and Data Mining" provide comprehensive knowledge to address various challenges encountered in complex systems. By integrating practices and applications from both domains, it offers a robust framework for tackling these issues. To excel in data-driven ecosystems, researchers, data analysts, and practitioners must carefully select the most suitable approaches and tools. Understanding the design choices and options available is essential for appreciating the underlying concepts, tools, and techniques utilized in these endeavors.