Foreword

Special Issue on Trends in Information Science and Data Analytics

Information Science and Data Analytics gain more and more importance in organizations in today's world. Data is increasingly massive, dynamic and this is a big challenge stimulating research with the goal of turning information overload into a significant opportunity and helping decision making in organizations. Data Analytics assumes a high importance and needs to be supported by highly interactive visual interfaces, tools, and techniques in all aspects of information and computer science, engineering, and business to synthesize information to manage and transform it into knowledge. Data analytics requires interdisciplinary science including statistics, mathematics, knowledge representation, management, and technologies. This special issue is related to development and/or application of data analytics, clearly showing an interdisciplinary approach, and intends to promote the research on the trends and developments in the globally competitive environment as well as to provide indicators for future directions to researchers and practitioners. Eleven papers are accepted in this issue:

The article "Exploiting Machine Learning and Feature Selection Algorithms to Predict Instructor Performance in Higher Education" from Ravinder Ahuja, and Sc Sharma, analyses the performance of the instructor, which is evaluated through feedback collected from students in the questionnaire form. The dataset is taken from the machine learning repository consisting of 5820 records with 33 attributes. Using agglomerative clustering and the k-means algorithms three labels are created and five feature selection techniques are applied to extract essential features, and twelve classification algorithms are applied using Python language. The results showed the highest accuracy value of 99.66%, recall value of 99.66%, precision value of 99.67%, and f-score value of 99.67%.

The article "DDCO - Diversified Data Characteristic-based Oversampling for Imbalance Classification Problems" from Rekha, and Reddy proposes an improved minority samples generation called Diversified Data Characteristic based Oversampling (DDCO) to deal with over-generalization and overlapping problem, an improved minority samples generation is proposed. The performance of the proposed model was evaluated on 14 imbalanced datasets and compared with state-of-the-art methods like SMOTE, Borderline-

SMOTE, ADASYN, MWMOTE using AUC, and F-Measure as the performance measures. The results indicate significant improvement over the state-of-the-art methods.

The article "A comparative study of machine learning models for predicting Length of Stay in hospitals" from Mekhaldi et al. is focused on predicting the Length of Stay (LoS) in a hospital setting, as estimating the LoS on patient' admission helps hospitals in planning, controlling costs and, providing better services. In this paper, the authors consider predicting the LoS as a regression problem for which they have implemented and compared different Machine Learning (ML) algorithms: Multiple Linear Regression (MLR), Support Vector Machines (SVM), Random Forests (RF), and Gradient Boosting model (GBM) using an open-source dataset.

The article "Ensemble Case based Reasoning Imputation in Breast Cancer Classification" from Chlioui, et al. studies the Missing Data (MD) as a drawback that affects breast cancer classification. It presents the impact of using ensemble Case-Based Reasoning (CBR) imputation on breast cancer classification. The authors evaluated the influence of CBR using parameter tuning and ensemble CBR (E-CBR) with three missingness mechanisms (MCAR: missing completely at random, MAR: missing at random and NMAR: not missing at random) and nine percentages (10% to 90%) on the accuracy rates of five classifiers: Decision trees, Random forest, K-nearest neighbor, Support vector machine and Multilayer perceptron over two Wisconsin breast cancer datasets. The findings confirmed that E-CBR yields to better results compared to CBR for the five classifiers.

The article "Big Data Based Knowledge Management vs Traditional Knowledge Management: A People, Process and Technology Perspective" from Sumbal, et al. links the value creation of big data with the knowledge management capability of the organizations in terms of people, processes and technology which play a crucial role in effective knowledge management. This study contributes towards the existing body of knowledge by exploring this linkage of people, process, and technology in relation to big data through the lens of knowledge management, by conducting a qualitative study in the oil and gas industry. The findings reveal that the KM capability of the organizations through big data can be explained through the Complex domain of Cynefin framework which involves probing, sensing and responding in which there are no right answers and instructive patterns (predictive knowledge) emerging from big data could be right or wrong depending upon the complexity of the situation.

The article "Data Science Applied to Marketing: A Literature Review" from Rosário et. al presents a systematic literature review between 2010 and 2020, confirming that the topic of Data Science Applied to Marketing, is a new subject of interest, and a variety of data science methods are applied in the studies, namely: micro-segmentation and real-time application to natural language processing. The findings show that the impact is evident in digital advertising, micro-segmentation and micro-targeting, speed and performance, and real-time experimentation.

The article "Deep Neural Network for Covid-19 Pandemic Recognition Using CT Data" from Slama et al. uses a computerized approach to show that the deep neural network (DNN) is a distinctive method to classify Covid-19 pandemic. Experimental results on various lung CT scan images of different Covid-19 patients, demonstrate the effectiveness of the proposed methodology when compared to the manual scoring of pathological experts. According to the performance evaluation, the authors recorded more than 92% for accuracy of infection detected in ROI scoring over the truths provided by experienced radiologists.

The article "Mgini - Improved Decision Tree using Minority Class Sensitive Splitting Criterion for Imbalanced Data of Covid-19" from Barot and Jethva studies the problem of imbalanced medical data, proposing a modified decision tree algorithm that uses a minority sensitive Gini index called Mgini. According to the authors, in an imbalanced dataset of COVID-19, it is important to focus on the reduction of overall misclassification cost instead of trying improvement in accuracy value. Mgini is useful splitting criteria when the misclassification cost of the minority sample is huge as compared to the majority class. The use of this proposed new Gini index as a splitting criterion in the decision tree reduces the misclassification cost. Mgini based decision tree has higher accuracy and low misclassification cost as compare to the traditional Gini index-based CART algorithm. Their proposed cost-sensitive approach improves imbalanced data classification without the use of data level sampling techniques.

The article "Gamification framework for management practice" from Marcão et al. analyses a new business models, to ensure cohesion between teams, studying mechanisms for obtaining high income to support enterprise architecture and the intended business model, which highlights the use of the concept of gamification as one of these mechanisms. In this context, considering project activities, the research identifies how the commitment of stakeholders evolves and if the application of gamification techniques enhances

this variable, taking performance evolution as another variable that could be evaluated.

The article "Data Science Projects in Pharmaceutical Industry" from Pesqueira et al. discuss the relevance of data science in Medical Affairs (MA) functions in the pharmaceutical industry. Tis study analyses pharmaceutical companies who have a data science strategy and the variables that can influence the definition of a data science strategy in pharma companies in opposite to other pharmaceutical companies without a data science strategy. A questionnaire was developed and applied to a sample of 280 pharma companies. The findings indicate that there is good evidence in the empirical relation between Data Science and the strategies of the organization.

The article "Deep Learning based Automated Fruit Nutrients Deficiency Recognition System" from Yogesh studies a deep learning-based method that extracts low level and high-level features such as edges, geometrical, statistical, texture, intensity, etc. The authors studied the validation of the system with test data, to predict the output with the system developed. The processing time is optimized by avoiding fully connected layers which further minimize the requirement of neurons in the network. The convolutional neural network extracts the features of the fruits, Rectified Linear Unit (ReLu) removes the non-fruit pixels. Pooling shrinks, the image by selecting the maximum value of the pixel. The process is repeated until the size of the image is at the desired level. The aim is to identify the objects and recognize them. The foreground region objects are of our interest and being segmented for higher-level image processing. The proposed system attains the accuracy of 99.30 % with the processing time of 3.207 sec.

• Introduction to Guest Editors •

Maria José Sousa (Ph.D. in Industrial Management) is a University Professor at ISCTE, a Research Fellow at Business Research Unit, and a collaborator of IPPS-ISCTE. She is also an expert in digital learning and digital skills, as she has assumed a Post-Doc position from 2016-2018, researching that field, with several publications in journals with high impact factor (Journal of Business Research, Journal of Grid Computing, Future Generation Computer Systems, and others). And is collaborating as an expert in digital

skills, with Delloite (Brussels) for a request of the European Commission in the creation of a new category regarding digital skills to be integrated with the European Innovation Scoreboard (EIS). She was a member of the Coordinator Committee of the Ph.D. in Management at Universidade Europeia. She was also a Senior Researcher at GEE (Research Office) in the Portuguese Ministry of Economy, responsible for Innovation, Research, and Entrepreneurship Policies, and a Knowledge and Competencies Manager at AMA, IP, Public Reform Agency (Ministry of the Presidency and the Ministers Council). She was also a Project Manager at the Ministry of Labor and Employment, responsible for Innovation, and Evaluation and Development of the Qualifications Projects. Her research interests currently are public policies, health policies, innovation, and information science. She has developed major research in the innovation policies with articles published in high-level journals (as the European Planning Studies, Information Systems Frontiers, Systems Research, and Behavioral Science, Computational and Mathematical Organization Theory, Future Generation Computer Systems, and others). She is also the guest-editor of more than 5 Special Issues from Springer and Elsevier. She has participated in European projects of innovation transfer (for example, as Ambassador of EUWIN, and Co-coordinating an Erasmus + project with ATO - Chamber of Commerce of Ankara about entrepreneurship) and is also an External Expert of COST Association - European Cooperation in Science and Technology, and President of the ISO/TC 260 – Human Resources Management, representing Portugal in the International Organization for Standardization.

Alvaro Rocha holds the title of Honorary Professor, and holds a D.Sc. in Information Science, Ph.D. in Information Systems and Technologies, M.Sc. in Information Management, and BCs in Computer Science. He is a Professor of Information Systems at the University of Lisbon - ISEG, researcher at the ADVANCE (the ISEG Centre for Advanced Research in Management), and a collaborator researcher at both LIACC (Laboratory of Artificial Intelligence and Computer Science) and CINTESIS (Center for Research in Health Technologies and Information Systems). He is also President of ITMA (the Information and Technology Management Association), Vice-Chair of the IEEE Portugal Section Systems, Man, and Cybernetics Society Chapter, and Editor-in-Chief of both JISEM (Journal of Information Systems Engineering & Management) and RISTI

(Iberian Journal of Information Systems and Technologies). Moreover, he has served as Vice-Chair of Experts for the European Commission's Horizon 2020 program, and as an Expert at the COST – intergovernmental framework for European Cooperation in Science and Technology, at the Government of Italy's Ministry of Universities and Research, at the Government of Latvia's Ministry of Finance, at the Government of Mexico's National Council of Science and Technology, and at the Government of Polish's National Science Centre.

Ren-Hung Hwang received his B.S. degree in Computer Science and Information Engineering from National Taiwan University, Taipei, Taiwan, in 1985, and the M.S. and Ph.D. degrees in Computer Science from University of Massachusetts, Amherst, Massachusetts, USA, in 1989 and 1993, respectively. He joined the Department of Computer Science and Information Engineering, National Chung Cheng University, Chia-Yi, Taiwan, in 1993, where he is now a Professor and the Chair of the Department of Information Science and Computer Engineering. His research interests include Web 2.0, peer-to-peer applications, ad hoc networks, e-learning, and 3G.