

2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA 2021)

**Porto, Portugal
6 – 9 October 2021**



**IEEE Catalog Number: CFP21DSB-POD
ISBN: 978-1-6654-2100-3**

**Copyright © 2021 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP21DSB-POD
ISBN (Print-On-Demand):	978-1-6654-2100-3
ISBN (Online):	978-1-6654-2099-0

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

TABLE OF CONTENTS

PREDICTIVE MAINTENANCE BASED ON ANOMALY DETECTION USING DEEP LEARNING FOR AIR PRODUCTION UNIT IN THE RAILWAY INDUSTRY	1
<i>Narjes Davari, Bruno Veloso, Rita P. Ribeiro, Pedro Mota Pereira, João Gama</i>	
CLASSIFICATION OF SPACE PARTICLE EVENTS USING SUPERVISED MACHINE LEARNING ALGORITHMS.....	11
<i>Rijad Saric, Junchao Chen, Miloš Krstic, Edhem Custovic, Goran Panic, Jasmin Kevric, Dejan Jokic</i>	
ON THE USE OF CLASS ACTIVATION MAPS IN REMOTE SENSING: THE CASE OF ILLEGAL LANDFILLS	21
<i>Rocio Nahime Torres, Piero Fraternali, Andrea Biscontini</i>	
LEARNING FASHION SIMILARITY BASED ON HIERARCHICAL ATTRIBUTE EMBEDDING.....	31
<i>Cairong Yan, Anan Ding, Yanting Zhang, Zijian Wang</i>	
TWO-PHASE MULTI-ARMED BANDIT FOR ONLINE RECOMMENDATION	39
<i>Cairong Yan, Haixia Han, Zijian Wang, Yanting Zhang</i>	
DISCOVERY OF PATIENT PHENOTYPES THROUGH MULTI-LAYER NETWORK ANALYSIS ON THE EXAMPLE OF TINNITUS.....	47
<i>Clara Puga, Uli Niemann, Vishnu Unnikrishnan, Miro Schleicher, Winfried Schlee, Myra Spiliopoulou</i>	
ASSISTML: A CONCEPT TO RECOMMEND ML SOLUTIONS FOR PREDICTIVE USE CASES	57
<i>Alejandro Gabriel Villanueva Zacarias, Christian Weber, Peter Reimann, Bernhard Mitschang</i>	
SEQUENTIAL DEPENDENCY ENHANCED GRAPH NEURAL NETWORKS FOR SESSION-BASED RECOMMENDATIONS	69
<i>Wei Guo, Shoujin Wang, Wenpeng Lu, Hao Wu, Qian Zhang, Zhufeng Shao</i>	
EFFICIENT MODELING OF DIGITAL SHADOWS FOR PRODUCTION PROCESSES: A CASE STUDY FOR QUALITY PREDICTION IN HIGH PRESSURE DIE CASTING PROCESSES	79
<i>Arnab Chakrabarti, Ravi Prasanna Sukumar, Matthias Jarke, Maximilian Rudack, Paul Buske, Carlo Holly</i>	
OVERFITTING MEASUREMENT OF DEEP NEURAL NETWORKS USING NO DATA	88
<i>Satoru Watanabe, Hayato Yamana</i>	
A PERSONALIZED OVERDRAFT PROTECTION FRAMEWORK	98
<i>Karesia Ramlal, Patrick Hosein</i>	
SOURCE DETECTION ON NETWORKS USING SPATIAL TEMPORAL GRAPH CONVOLUTIONAL NETWORKS.....	106
<i>Hao Sha, Mohammad Al Hasan, George Mohler</i>	
TOWARDS OPTIMIZED ACTIONS IN CRITICAL SITUATIONS OF SOCCER GAMES WITH DEEP REINFORCEMENT LEARNING	117
<i>Pegah Rahimian, Afshin Oroojlooy, Laszlo Toka</i>	

A MULTILAYER NETWORK PERSPECTIVE ON CUSTOMER SEGMENTATION THROUGH CASHLESS PAYMENT DATA	129
<i>Alessia Galdeman, Cheick Ba, Matteo Zignani, Sabrina Gaito</i>	
A NEURAL NETWORK ARCHITECTURE WITH AN ATTENTION-BASED LAYER FOR SPATIAL PREDICTION OF FINE PARTICULATE MATTER.....	139
<i>Luis E. Colchado, Edwin Villanueva, José Ochoa-Luna</i>	
UNSUPERVISED TENSOR BASED FEATURE EXTRACTION AND OUTLIER DETECTION FOR MULTIVARIATE TIME SERIES	149
<i>Kiyotaka Matsue, Mahito Sugiyama</i>	
DECOUPLING AUTOENCODERS FOR ROBUST ONE-VS-REST CLASSIFICATION	161
<i>Max Lübbering, Michael Gebauer, Rajkumar Ramamurthy, Christian Bauckhage, Rafet Sifa</i>	
LEVERAGING LATENT ECONOMIC CONCEPTS AND SENTIMENTS IN THE NEWS FOR MARKET PREDICTION	171
<i>Saeede Anbaee Farimani, Majid Vafaei Jahan, Amin Milani Fard, Gholamreza Haffari</i>	
FEATURE-OPTION-ACTION: A DOMAIN ADAPTION TRANSFER REINFORCEMENT LEARNING FRAMEWORK.....	181
<i>Yunxiao Zhang, Xiaochuan Zhang, Tianlong Shen, Yuan Zhou, Zhiyuan Wang</i>	
MACHINE LEARNING MODEL FOR FRAILTY DETECTION USING ELECTRIC POWER CONSUMPTION DATA FROM SMART METER.....	193
<i>Kijun Kim, Shimpei Ohsugi, Noboru Koshizuka</i>	
NEDRL-CIM:NETWORK EMBEDDING MEETS DEEP REINFORCEMENT LEARNING TO TACKLE COMPETITIVE INFLUENCE MAXIMIZATION ON EVOLVING SOCIAL NETWORKS.....	202
<i>Khurshed Ali, Chih-Yu Wang, Mi-Yen Yeh, Cheng-Te Li, Yi-Shin Chen</i>	
A NEURAL APPROACH FOR DETECTING MORPHOLOGICAL ANALOGIES.....	211
<i>Safa Alsaidi, Amandine Decker, Puthineath Lay, Esteban Marquer, Pierre-Alexandre Murena, Miguel Couceiro</i>	
EXTRACTING CANCER TREATMENTS FROM CLINICAL TEXT WRITTEN IN SPANISH: A DEEP LEARNING APPROACH.....	221
<i>Oswaldo Solarte-Pabón, Alberto Blazquez-Herranz, Maria Torrente, Alejandro Rodríguez-Gonzalez, Mariano Provencio, Ernestina Menasalvas</i>	
TOWARDS TREATMENT PATTERNS VALIDATION IN LUNG CANCER PATIENTS	227
<i>Arturo Redondo, Belén Ríos-Sánchez, Guillermo Viguera, Belén Otero, Roberto Hernández, María Torrente, Ernestina Menasalvas, Mariano Provencio, Alejandro Rodríguez-González</i>	
EXPLAINABLE ARTIFICIAL INTELLIGENCE FOR DATA SCIENCE ON CUSTOMER CHURN.....	235
<i>Carson K. Leung, Adam G. M. Pazdor, Jorglas Souza</i>	
ANYTIME SUBGROUP DISCOVERY IN HIGH DIMENSIONAL NUMERICAL DATA	245
<i>Romain Mathonat, Diana Nurbakova, Jean-François Boulicaut, Mehdi Kaytoue</i>	
A DATA-DRIVEN APPROACH BASED ON TENSOR COMPLETION FOR REPLACING “PHYSICAL SENSORS” WITH “VIRTUAL SENSORS”.....	255
<i>Noorali Raeji Yaneh Sari, Hadi Fanaee-T, Mahmoud Rahat</i>	

REDUCING UNINTENDED BIAS OF ML MODELS ON TABULAR AND TEXTUAL DATA.....	261
<i>Guilherme Alves, Maxime Amblard, Fabien Bernier, Miguel Couceiro, Amedeo Napoli</i>	
COMBINING STATIC AND DYNAMIC ANALYSIS TO IMPROVE MACHINE LEARNING- BASED MALWARE CLASSIFICATION	271
<i>Rajchada Chanajitt, Bernhard Pfahringer, Heitor Murilo Gomes</i>	
EXPLAINABLE ANOMALY DETECTION FOR HOT-ROLLING INDUSTRIAL PROCESS	281
<i>Jakub Jakubowski, Przemyslaw Stanis, Szymon Bobek, Grzegorz J. Nalepa</i>	
CENTRALITY-BASED INTERPRETABILITY MEASURES FOR GRAPH EMBEDDINGS	291
<i>Shima Khoshraftar, Sedigheh Mahdavi, Aijun An</i>	
EARLY CLASSIFICATION OF TIME SERIES: COST-BASED MULTICLASS ALGORITHMS	301
<i>Paul-Emile Zafar, Youssef Achenchabe, Alexis Bondu, Antoine Cornuéjols, Vincent Lemaire</i>	
IMPROVING PORTUGUESE SEMANTIC ROLE LABELING WITH TRANSFORMERS AND TRANSFER LEARNING	311
<i>Sofia Oliveira, Daniel Loureiro, Alípio Jorge</i>	
MOTIF2VEC: SEMANTIC-AWARE REPRESENTATION LEARNING FOR WEARABLES' TIME SERIES DATA	320
<i>Suwen Lin, Xian Wu, Nitesh V. Chawla</i>	
DASKDB: SCALABLE DATA SCIENCE WITH UNIFIED DATA ANALYTICS AND IN SITU QUERY PROCESSING	330
<i>Alex Watson, Suvam Kumar Das, Suprio Ray</i>	
ANALYZING AND REPAIRING CONCEPT DRIFT ADAPTATION IN DATA STREAM CLASSIFICATION	340
<i>Ben Halstead, Yun Sing Koh, Patricia Riddle, Russel Pears, Mykola Pechenizkiy, Albert Bifet, Gustavo Olivares, Guy Coulson</i>	
A SEQUENTIAL FRAMEWORK TOWARDS AN EXACT SDP VERIFICATION OF NEURAL NETWORKS	342
<i>Ziye Ma, Somayeh Sojoudi</i>	
DYNAMIC GRAPH CONVOLUTIONAL LSTM APPLICATION FOR TRAFFIC FLOW ESTIMATION FROM ERROR-PRONE MEASUREMENTS: RESULTS AND TRANSFERABILITY ANALYSIS	350
<i>Safa Boudabous, Stephan Cléménçon, Houda Labiod, Julian Garbiso</i>	
BAYESIAN INDEPENDENCE TEST WITH MIXED-TYPE VARIABLES	360
<i>Alessio Benavoli, Cassio De Campos</i>	
ADVERSARIALLY GENERATING RANK-CONSTRAINED GRAPHS	373
<i>William Shiao, Evangelos E. Papalexakis</i>	
HYBRID RECOMMENDER SYSTEM FOR DETECTION OF RARE CASES APPLIED TO PULSAR CANDIDATE SELECTION	381
<i>Di Pang, Katerina Goseva-Popstojanova, Maura McLaughlin</i>	
INTERPRETABLE PREDICTION OF DIABETES FROM TABULAR HEALTH SCREENING RECORDS USING AN ATTENTIONAL NEURAL NETWORK	391
<i>Yuki Oba, Taro Tezuka, Masaru Sanuki, Yukiko Wagatsuma</i>	

PEAR: A PRODUCT EXPIRY-AWARE AND REVENUE-CONSCIOUS ITEMSET PLACEMENT SCHEME.....	402
<i>Anirban Mondal, Raghav Mittal, Vrinda Khandelwal, Parul Chaudhary, P. Krishna Reddy</i>	
EXPLAINABLE CLUSTERING WITH MULTIDIMENSIONAL BOUNDING BOXES	412
<i>Michal Kuk, Szymon Bobek, Grzegorz J. Nalepa</i>	
CONDITIONAL T-SNE: MORE INFORMATIVE T-SNE EMBEDDINGS	422
<i>Bo Kang, Darío García García, Jefrey Lijffijt, Raúl Santos-Rodríguez, Tijl De Bie</i>	
A GENERIC FRAMEWORK FOR FORECASTING SHORT-TERM TRAFFIC CONDITIONS ON URBAN HIGHWAYS.....	424
<i>Seif-Eddine Attoui, Maroua Meddeb</i>	
PARALLEL MULTI-GRAPH CONVOLUTION NETWORK FOR METRO PASSENGER VOLUME PREDICTION	434
<i>Fuchen Gao, Zhanquan Wang, Zhenguang Liu</i>	
SOFT-CHURN: OPTIMAL SWITCHING BETWEEN PREPAID DATA SUBSCRIPTIONS ON E-SIM SUPPORT SMARTPHONES	444
<i>Patrick Hosein, Gabriela Sewdhan, Aviel Jailal</i>	
EMBED2DETECT: TEMPORALLY CLUSTERED EMBEDDED WORDS FOR EVENT DETECTION IN SOCIAL MEDIA: EXTENDED ABSTRACT	450
<i>Hansi Hettiarachchi, Mariam Adedoyin-Olowe, Jagdev Bhogal, Mohamed Medhat Gaber</i>	
A UNIFIED FRAMEWORK FOR CLOSED-FORM NONPARAMETRIC REGRESSION, CLASSIFICATION, PREFERENCE AND MIXED PROBLEMS WITH SKEW GAUSSIAN PROCESSES	452
<i>Alessio Benavoli, Dario Azzimonti, Dario Piga</i>	
EXPLAINING MULTIMODAL ERRORS IN AUTONOMOUS VEHICLES.....	454
<i>Leilani H. Gilpin, Vishnu Penubarthi, Lalana Kagal</i>	
THE USE OF MACHINE LEARNING TO IDENTIFY THE CORRECTNESS OF HS CODE FOR THE CUSTOMS IMPORT DECLARATIONS	464
<i>Hao Chen, Ben Van Rijnsoever, Marcel Molenhuis, Dennis Van Dijk, Yao-Hua Tan, Boriana Rukanova</i>	
INTERPRETABLE SUMMARIES OF BLACK BOX INCIDENT TRIAGING WITH SUBGROUP DISCOVERY	472
<i>Youcef Remil, Anes Bendimerad, Marc Plantevit, Céline Robardet, Mehdi Kaytoue</i>	
ORDINAL REGRESSION WITH EXPLAINABLE DISTANCE METRIC LEARNING BASED ON ORDERED SEQUENCES: EXTENDED ABSTRACT	482
<i>Juan Luis Suárez, Salvador García, Francisco Herrera</i>	
EXTRACTING INVARIANT FEATURES FOR PREDICTING STATE OF HEALTH OF BATTERIES IN HYBRID ENERGY BUSES.....	484
<i>Mohammed Ghaith Altarabichi, Yuantao Fan, Sepideh Pashami, Peyman Sheikholharam Mashhadi, Slawomir Nowaczyk</i>	
A GENERALIZED OPTIMIZATION EMBEDDED FRAMEWORK OF UNDERSAMPLING ENSEMBLES FOR IMBALANCED CLASSIFICATION	490
<i>Hongjiao Guan, Yingtao Zhang, Bin Ma, Jian Li, Chunpeng Wang</i>	

RB-CCR: RADIAL-BASED COMBINED CLEANING AND RESAMPLING ALGORITHM FOR IMBALANCED DATA CLASSIFICATION	500
<i>Michal Koziarski, Colin Bellinger, Michal Wozniak</i>	
MISALIGNMENT PROBLEM IN MATRIX DECOMPOSITION WITH MISSING VALUES	502
<i>Sofia Fernandes, Mário Antunes, Diogo Gomes, Rui L. Aguiar</i>	
DEEP LEARNING AND MULTIVARIATE TIME SERIES FOR CHEAT DETECTION IN VIDEO GAMES.....	504
<i>José Pedro Pinto, André Pimenta, Paulo Novais</i>	
TIME-AWARE TENSOR DECOMPOSITION FOR SPARSE TENSORS.....	506
<i>Dawon Ahn, Jun-Gi Jang, U Kang</i>	
INDIVIDUAL SURVIVAL CURVES WITH CONDITIONAL NORMALIZING FLOWS	508
<i>Guillaume Auset, Tom Cifreio, Francois Portier, Stephan Clémenton, Timothée Papin</i>	
ANALYZING CULTURAL ASSIMILATION THROUGH THE LENS OF YELP RESTAURANT REVIEWS	518
<i>Zaiqian Chen, Joonsuk Park</i>	
PREDICTING HYBRID VEHICLES' FUEL AND ELECTRIC CONSUMPTION USING MULTITASK LEARNING.....	528
<i>Venkata Sai Vivek Uddagiri, Shankara Narayanan Bangalore Ramalingam, Mahmoud Rahat, Peyman Sheikholharam Mashhadi</i>	
V2W-BERT: A FRAMEWORK FOR EFFECTIVE HIERARCHICAL MULTICLASS CLASSIFICATION OF SOFTWARE VULNERABILITIES	534
<i>Siddhartha Shankar Das, Edoardo Serra, Mahantesh Halappanavar, Alex Pothen, Ehab Al-Shaer</i>	
3M-TRANSFORMERS FOR EVENT CODING ON ORGANIZED CRIME DOMAIN.....	546
<i>Erick Skorupa Parolin, Latifur Khan, Javier Osorio, Patrick T. Brandt, Vito D'Orazio, Jennifer Holmes</i>	
XPROAX-LOCAL EXPLANATIONS FOR TEXT CLASSIFICATION WITH PROGRESSIVE NEIGHBORHOOD APPROXIMATION	556
<i>Yi Cai, Arthur Zimek, Eirini Ntoutsi</i>	
CONSTRUCTING GLOBAL COHERENCE REPRESENTATIONS: IDENTIFYING INTERPRETABILITY AND COHERENCES OF TRANSFORMER ATTENTION IN TIME SERIES DATA.....	566
<i>Leonid Schwenke, Martin Atzmueller</i>	
ADAPTIVE NOISY DATA AUGMENTATION FOR REGULARIZED CONSTRUCTION OF UNDIRECTED GRAPHICAL MODELS	578
<i>Yinan Li, Fang Liu, Xiao Liu</i>	
ONLINE ENSEMBLE AGGREGATION USING DEEP REINFORCEMENT LEARNING FOR TIME SERIES FORECASTING.....	588
<i>Amal Saadallah, Katharina Morik</i>	
IMPROVED AUTOMATED CASH OPTIMIZATION WITH TREE PARZEN ESTIMATORS FOR CLASS IMBALANCE PROBLEMS	596
<i>Duc Anh Nguyen, Jiawen Kong, Hao Wang, Stefan Menzel, Bernhard Sendhoff, Anna V. Kononova, Thomas Bäck</i>	

DEEPHATEEXPLAINER: EXPLAINABLE HATE SPEECH DETECTION IN UNDER-RESOURCED BENGALI LANGUAGE.....	605
<i>Md. Rezaul Karim, Sumon Kanti Dey, Tanhim Islam, Sagor Sarker, Mehadi Hasan Menon, Kabir Hossain, Md. Azam Hossain, Stefan Decker</i>	
A FRAMEWORK FOR STATISTICALLY-SOUND CUSTOMER SEGMENT SEARCH	615
<i>Sihem Amer-Yahia, Laure Berti-Equille, Abdelouahab Chibah</i>	
AN OVERVIEW ON SUPERVISED SEMI-STRUCTURED DATA CLASSIFICATION	625
<i>Lijun Zhang, Ning Li, Zhanhuai Li</i>	
ACTIVE LEARNING FOR IMBALANCED DOMAINS: THE ALOD AND ALOD-RE ALGORITHMS.....	635
<i>Mayukh Bhattacharjee, Hema Sri Kambhampati, Paula Branco, Luis Torgo</i>	
RESGCN: ATTENTION-BASED DEEP RESIDUAL MODELING FOR ANOMALY DETECTION ON ATTRIBUTED NETWORKS.....	645
<i>Yulong Pei, Tianjin Huang, Werner Van Ipenburg, Mykola Pechenizkiy</i>	
MLIFE: A LITE FRAMEWORK FOR MACHINE LEARNING LIFECYCLE INITIALIZATION	647
<i>Cong Yang, Wenfeng Wang, Yunhui Zhang, Zhikai Zhang, Lina Shen, Yipeng Li, John See</i>	
LEARNING USER PREFERENCES WITHOUT FEEDBACKS.....	649
<i>Wei Zhang, Chris Challis</i>	
MIGRATE ON-PREMISES REAL-TIME DATA ANALYTICS JOBS INTO THE CLOUD.....	651
<i>Huijun Wu, Xiaoyao Qian, Hulya Pamukcu Crowell, Tushar Singh, Aleks Shulman, Prashil Bhimani, Abhishek Maloo, Chunxu Tang, Yao Li, Lu Zhang, Chris Ulherr</i>	
NON-TECHNICAL LOSSES DETECTION IN ENERGY CONSUMPTION FOCUSING ON ENERGY RECOVERY AND EXPLAINABILITY: EXTENDED ABSTRACT	653
<i>Bernat Coma-Puig, Josep Carmona</i>	
ZIPLINE: AN OPTIMIZED ALGORITHM FOR THE ELASTIC BULK SYNCHRONOUS PARALLEL MODEL	655
<i>Xing Zhao, Manos Papagelis, Aijun An, Bao Xin Chen, Junfeng Liu, Yonggang Hu</i>	
ATLANTIC REPUTATION AI: FOUR YEARS OF ADVANCING AND APPLYING A SOTA NLP CLASSIFIER.....	657
<i>Viet Yen Nguyen, Andréa B. Duque, Jules Belveze, Lukas D. Baker, Astrid Harsaae, Pedro Pessoa, Inês Oliveira</i>	
A FULL-STACK MACHINE LEARNING ENVIRONMENT FOR RAPIDLY EVOLVING INDUSTRY APPLICATIONS.....	659
<i>Kayla Robinson, Chris Billman, Muktesh Masih, Kevin Rose, Xi Wang, Kyle Hundman</i>	
CAUSALYSIS: CAUSAL MACHINE LEARNING FOR REAL-ESTATE INVESTMENT DECISIONS	662
<i>Rodrigo Rivera-Castro, Evgeny Burnaev</i>	
TENSOR DECISION TREES FOR CONTINUAL LEARNING FROM DRIFTING DATA STREAMS	665
<i>Bartosz Krawczyk</i>	
OPPORTUNISTIC RESOURCE ALLOCATION DRIVEN BY PATTERN RECOGNITION TO ATTAIN HIGH AVAILABILITY IN VISION BASED EDGE COMPUTE SYSTEMS	667
<i>Muhammad Sibtain Hamayun, Olga Ivina</i>	

MINING MAVERICKS - A DATA-DRIVEN APPROACH TO DETECT SPEND LEAKAGE	670
<i>Priyadarshi, Anish Chaugule, Maitreya Natu</i>	
PREDICTING BATCH PROCESS TO PREVENT BUSINESS OUTAGES	672
<i>Satya Samudrala, Neha Behl, Vikrant Shimpi, Deepa Vaidyanathan, Maitreya Natu</i>	
RESOLVING THE MESSAGE RIDDLE: A MULTI-PRONGED APPROACH TO INFER TROUBLE TICKETS	674
<i>Vikrant Shimpi, Aakash Patel, Siddartha Kshirsagar, Maitreya Natu, Monika Bhave</i>	
SURVEILLANCE OF AIRBORNE PLANT DISEASE DISSEMINATION AT CONTINENTAL SCALE USING AIR MASS TRAJECTORY ANALYSIS AND NETWORK THEORY	676
<i>Andrea Radici, Davide Martinetti, Daniele Bevacqua</i>	

Author Index