

## Záverečná karta projektu

Názov projektu Evidenčné číslo projektu **APVV-16-0213****Znalostné prístupy k inteligentnej analýze veľkých dát**Zodpovedný riešiteľ **prof. Ing. Ján Paralič, PhD.**Príjemca **Technická univerzita v Košiciach - Fakulta elektrotechniky a informatiky**

### Názov pracoviska, na ktorom bol projekt riešený

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach  
Fakulta informatiky a informačných technológií, Slovenská technická univerzita v Bratislave

### Názov a štát zahraničného pracoviska, ktoré spolupracovalo pri riešení

Faculty of Medicine in Osijek Chorvátsko, tím vedený doc. Ljiljanou Trtica-Majnaric  
Human-Centered AI Lab, Medical University of Graz, Rakúsko, tím prof. Andreaa Holzinger  
Institute of Atmospheric Physics, AS CR, Česko, tím vedený prof. Ondřejom Santolík  
South African Radio Astronomy Observatory, Južná Afrika, dr. Khadija El Bouchefry  
Department of Earth Sciences, Western University, Kanada, dr. Elizabeth Silber  
Department of Physics, Helsinki University, Fínsko, dr. Maria Gritsevich

### Udelené patenty/podané patentové prihlášky, vynálezy alebo úžitkové vzory, ktoré sú výsledkami projektu

-

### Najvýznamnejšie publikácie (knihy, články, prednášky, správy a pod.) zhrňujúce výsledky projektu – uveďte aj publikácie prijaté do tlače

V kontexte tohto projektu vznikla aj knižná publikácia vydaná v nakladateľstve Springer - J. Paralič, P. Sinčák, P. Hartono, & V. Mařík (Eds.), Towards Digital Intelligence Society: A Knowledge-based Approach. Springer 2021

Najvýznamnejšie články (s ohlasmi):

[1] SARNOVSKÝ, Martin – PARALIČ, Ján: Hierarchical Intrusion Detection Using Machine Learning and Knowledge Model. In: Symmetry : Open Access Journal. Basel (Švajčiarsko): Multidisciplinary Digital Publishing Institute, Roč. 12, č. 2 (2020), s. 1-14 [online]. ISSN 2073-8994 (11 SCI citácií)

[2] VRABLECOVÁ, Petra - BOU EZZEDDINE, Anna - ROZINAJOVÁ, Viera - ŠÁRIK, Slavomír - SANGAIAH, Arun Kumar: Smart grid load forecasting using online support vector regression. In: Computers and electrical engineering. Vol. 65, iss. January (2018), s. 102-117. ISSN 0045-7906 (54 SCI citácií)

[3] LAURINEC, Peter - LUCKÁ, Mária. Interpretable multiple data streams clustering with clipped streams representation for the improvement of electricity consumption forecasting. In: New SZP GEN: Data Mining and Knowledge Discovery, (2018), s. 1-33. ISSN 1384-5810

(10 SCI citácií)

[4] LAURINEC, Peter - LÓDERER, Marek - LUCKÁ, Mária - ROZINAJOVÁ, Viera. Density-based unsupervised ensemble learning methods for time series forecasting of aggregated or clustered electricity consumption. In Journal of Intelligent Information Systems. Vol. 53, iss. 2 (2019), s. 219-239. ISSN 0925-9902 (7 SCI citácií)

[5] MASLEJ KREŠŇÁKOVÁ, Viera - SARNOVSKÝ, Martin – BUTKA, Peter - MACHOVÁ, Kristína: Comparison of Deep Learning Models and Various Text Pre-Processing Techniques for the Toxic Comments Classification. In: Applied Sciences, Vol. 10, Issue 23 (2020). ISSN 2076-3417 (6 SCI citácií)

[6] TRTICA MAJNARIĆ, Ljiljana - BABIČ, František - O' SULLIVAN, Shane - HOLZINGER, Andreas: AI and Big Data in Healthcare: Towards a More Comprehensive Research Framework for Multimorbidity. In: Journal of clinical medicine 2021, 10(4), 1-23. (5 SCI citácií)

[7] TRTICA MAJNARIĆ, Ljiljana - BEKIĆ, Sanja - BABIČ, František - PUSZTOVÁ, Ľudmila - PARALIČ, Ján: Cluster Analysis of the Associations among Physical Frailty, Cognitive Impairment and Mental Disorders. In: Medical Science Monitor : international medical journal of experimental and clinical research 2020, 26, 1-12. (3 SCI citácie)

[8] BEKIC, Sanja – BABIČ, František – FILIPČIĆ, Igor - MAJNARIC TRTICA, Ljiljana: Clustering of Mental and Physical Comorbidity and the Risk of Frailty in Patients Aged 60 Years or More in Primary Care. Medical Science Monitor 2019, 11(25), 6820-6835 (3 SCI citácie)

[9] MACHOVÁ, Kristína, MIKULA, Martin, GAO, Xiaoying, MACH, Marián: Lexicon-based Sentiment Analysis Using the Particle Swarm Optimization. In: Electronics, Vol. 9, no. 8, 1317, 2020, s.1-22. ISSN 2079-9292 (3 SCI citácie)

[10] BEKIC, Sanja – BABIČ, František – FILIPČIĆ, Igor - MAJNARIC TRTICA, Ljiljana: Clustering of Mental and Physical Comorbidity and the Risk of Frailty in Patients Aged 60 Years or More in Primary Care. Medical Science Monitor 2019, 11(25), 6820-6835 (3 SCI citácie)

[11] SILBER, Elizabeth – NICULESCU, Mihai – BUTKA Peter – SILBER, Raymond: Nitric Oxide Production by Centimeter-Sized Meteoroids and the Role of Linear and Nonlinear Processes in the Shock Bound Flow Fields. In: Atmosphere, Vol. 9, no. 5, 2015, s.1-21, ISSN 2073-4433 (3 SCI citácie)

[12] SARNOVSKÝ, Martin - BEDNÁR, Peter - SMATANA, Miroslav: Big Data Processing and Analytics Platform Architecture for Process Industry Factories. In: Big Data and Cognitive Computing. Vol. 2, no. 1 (2018), p. 3-15, ISSN 2504-2289 (9 SCI citácií)

### **Uplatnenie výsledkov projektu**

V projekte bol vytvorený a validovaný rozšírený znalostný model pre popis tak kontextu cieľovej úlohy (doménové znalosti) ako aj špecifikácie cieľov analýzy dát (popisu samotného procesu analýzy dát (zreťazenie operátorov pri predspracovaní, použitie algoritmov a metód analýzy dát a popis samotných výsledných dátovo-analytických modelov). Tento model umožňuje aj mapovanie konceptov na existujúce technológie pre analýzu dát v prostredí R čo vedie k čiastočnému zautomatizovaniu procesu analýzy dát. Pre účely validácie sme realizovali rozsiahle prípadové štúdie analýzy obchodných a transakčných dát. Boli navrhnuté, implementované a otestované zložené adaptívne predikčné modely, ktoré kombinujú viaceré prediktory a využívajú originálne stratégie na adaptáciu ich váh v čase, aby presnosť predikcie neklesala ani v prípade výskytu konceptového driftu. Ich prínosy boli preukázané pri predikcii časových radov meraní spotreby elektrickej energie z inteligentných meračov.

Ďalšou oblasťou uplatnenia výsledkov projektu je vytvorený softvérový systém na báze modelov dátovej analýzy pre podporu rozhodovania vo viacerých medicínskych oblastiach. Vytvorili sme viaceré popisné aj prediktívne modely s cieľom identifikovať nielen podobné skupiny pacientov, ale aj kľúčové odlišnosti medzi nimi, s cieľom následne aplikovať personalizované liečebné postupy, výsledkom ktorých bude efektívnejšia a kvalitnejšia zdravotná starostlivosť.

### **Súhrn výsledkov riešenia projektu a naplnenia cieľov projektu v slovenskom jazyku (max. 20 riadkov)**

V rámci tvorby znalostných modelov sme sa venovali doménovým ontológiám a ontológiám dátovo-analytických procesov. Vytvorený model bol overovaný pre Priemysel 4.0, medicínu, či elektronické obchodovanie, pričom umožňuje špecifikovať ciele analýzy dát z konceptov, ktoré popisujú samotný proces analýzy dát, vrátane jeho krokov ako predspracovanie dát, použitie algoritmov a vyhodnotenie. V rámci využitia doménových znalostí pre prediktívne modelovanie sme sa zaoberali analýzou dynamických prúdov dát. Navrhnuté metódy boli overené v oblastiach IT bezpečnosti a energetiky. V oblasti analýzy bezpečnosti bol vytvorený znalostný model pre lepšiu klasifikáciu útokov v IT prostredí. V oblasti energetiky sa podrobnou znalostnou analýzou a využitím rôznych stratégií strojového učenia podarilo navrhnúť efektívne modely predikcie spotreby elektrickej energie. V oblasti podpory inteligentnej analýzy medicínskych dát sme sa zamerali na popisné alebo prediktívne dolovanie v dátach s cieľom identifikovať nielen podobné skupiny pacientov, ale aj kľúčové odlišnosti medzi nimi, s cieľom následne aplikovať personalizované liečebné postupy, výsledkom ktorých bude efektívnejšia a kvalitnejšia zdravotná starostlivosť. V prípade doménových/expertných znalostí vstupujúcich do procesu analýzy dát sme sa zamerali na interpretovateľnosť a vysvetliteľnosť rozhodovacích modelov vytvorených pomocou algoritmov a metód strojového učenia alebo umelej inteligencie. Okrem toho sme sa venovali využitiu metód strojového učenia a konceptuálnej analýzy dát v spracovaní textov a vyhľadávaní informácií, s dôrazom na spracovanie prúdov sociálnych médií, detekciu dezinformácií a anti-sociálneho správania používateľov. Takisto sme využili metódy hlbokého učenia a vizualizácie dát pre podporu spracovania astro/geo dát. Vytvorené modely umožňujú detekciu javov v slnečnej koróne a v atmosfére Zeme, či ďalšie klasifikačné úlohy v tejto doménovej oblasti.

#### **Súhrn výsledkov riešenia projektu a naplnenia cieľov projektu v anglickom jazyku (max. 20 riadkov)**

In the area of knowledge models, we focused on domain ontologies and ontologies for data-analytical processes. The designed model, which was tested in domains like Industry 4.0, medicine, or e-commerce, provides the opportunity to specify data analysis goals in terms of concepts describing the data-analytical process and its steps like preprocessing, use of algorithms, and evaluation. According to the usage of domain knowledge for prediction modeling, the main goal was to analyze dynamic data streams. New methods were applied in areas like IT security and energy distribution. The newly designed knowledge model of security threats provides a better classification of security attacks in the IT environment. In the energy industry, the detailed domain analysis and usage of different machine learning strategies lead to new and more effective models for the prediction of electric energy consumption. In support of medical data intelligent analysis, we focused on descriptive or predictive data mining to identify similar groups of patients and key differences between them, with the aim to apply personalized treatment procedures, which will result in more efficient and better healthcare. In the case of domain / expert knowledge entering the data analysis process, we focused on the interpretability and explainability of decision models created using algorithms and methods of machine learning or artificial intelligence. We also focused on applying machine learning and conceptual data analysis methods in text processing and information retrieval, especially for social media processing, detecting disinformation and anti-social behavior of users. We also used deep learning methods and visualization techniques to support astro/geo data processing, where new methods allow physicists to detect events in the Solar corona and Earth's atmosphere, or other classification tasks in this domain area.