INTERNATIONAL

INVITED SESSION SUMMARY

Title of Session:
Data Pre-processing Strategies and Machine Learning

Name, Title and Affiliation of Chair:
Ireneusz Czarnowski, Gdynia Maritime University, Poland
Antonio J. Tallón-Ballesteros, University of Huelva, Spain

Details of Session (including aim and scope):

In current digital world data plays a crucial role and are necessary for working different systems, are important for decision support systems as well as for providing a high effectiveness of different services for society. They are collected in different areas of life and human activity. They come from the world around us, they are generated by various systems, including land and satellite systems. On the other hand it is not important to have the data only, but skillful and effective use of them, for example using machine learning paradigm. Machine learning is an example where data are needed and the machine learning tools are a core of the current intelligent and advanced systems, and they have to deal with this data, process it, so it also means must be able to take them. On the other way, the data must be prepared in proper way. The data are also a part of analytics and reporting level in business, medicine, or in commercial and production space. Thus, they preprocesing is also very important to obtain the desired effect and to be able to analyze these large data at all.

This special session is dedicated on data prep strategies, in context various uses of them, including especially for machine learning. Data prep strategies may be dedicated for a number of activities or processes, from data cleansing, noise detection and elimination, data editing, instance selection, noise reduction, eliminate the outliers and detecting wrong or distorted labels, resampling, feature extraction and selection, to data transformation. Data preparation can be also merged with an elimination of the class imbalance problem, data blending, aggregate data from diverse sources and with data wrangling. Data prep strategies can be merged with processes relatively easy or complex, and can based on simple tools or requiring complex computations. It also is possible, that these strategies require the solution of problems that belong to the group of optimization problems with aim to elimination barriers hidden in the data for the further application of analysis tools. So, it is reason. that the data prep strategies can be also merged with optimization tools.

The topics of interest for this session include, but are not limited to:

- Data science
- Data engineering
- Data selection
- Data editing
- Data cleansing
- Data engineering
- Feature selection and extraction
- Instance selection
- Data normalization
- Data transformation
- Data quality
- Imperfect data
- Data pre-processing
- Stream data preprocessing
- Imbalanced data processing
- Undersampling
- Resampling
- Aggregate data from diverse sources
- Data profiling

- Data wrangling
- Data visualisation
- Data validation
- Data pre-processing technologies
- Optimizing tools for data preprocessing
- Other related topics

Main Contributing Researchers / Research Centres (tentative, if known at this stage):

Website URL of Call for Papers (if any):

https://umg.edu.pl/i.czarnowski/kes_dml2023

Submission of papers: **3 April 2023**
Notification of acceptance: **8 May 2023**
Camera ready papers submission: **29 May 2023**
Authors Registration Deadline: **See on the conference webpage ...**

Conference: **6 - 8 September 2023**

More details is available at the KES 2023 website.

Email & Contact Details:
i.czarnowski@umg.edu.pl
antonio.tallon.diesia@zimbra.uhu.es