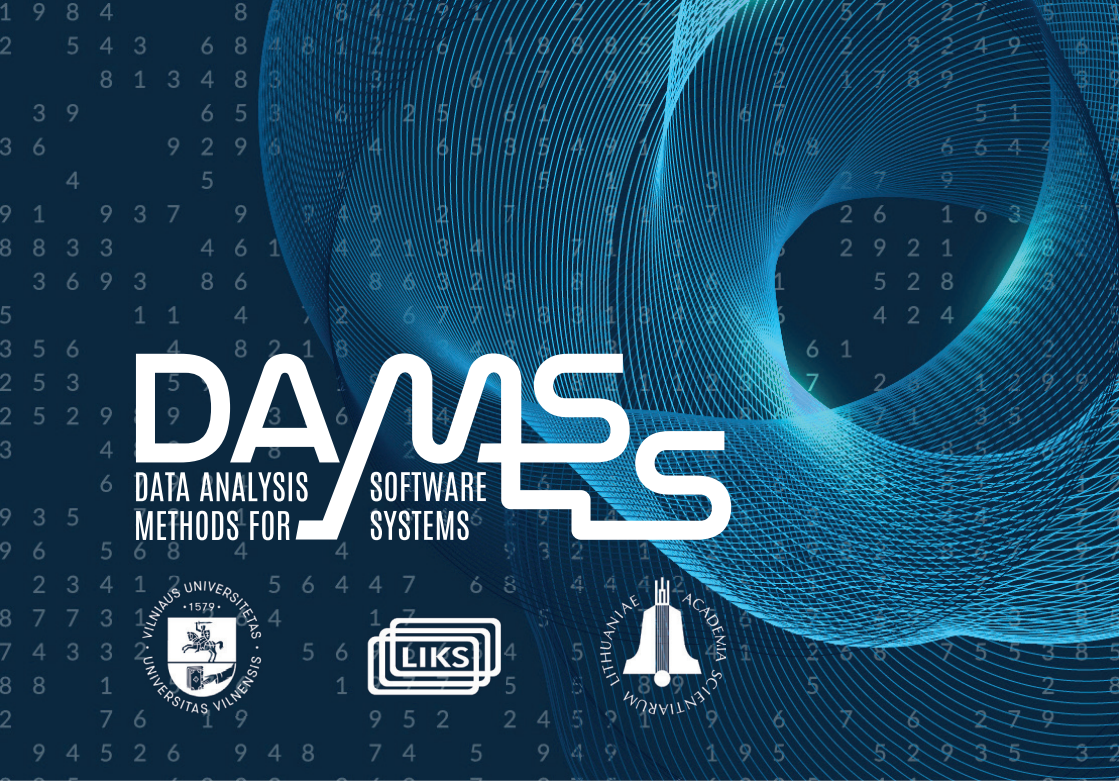

15th Conference on

DATA ANALYSIS METHODS
for Software Systems

November 28–30, 2024

Druskininkai, Lithuania, Hotel "Europa Royale"
https://www.mii.lt/DAMSS

Contacts:

Dr. Jolita Bernatavičienė
jolita.bernataviciene@mif.vu.lt
Tel. (+370 5) 2109 315
Prof. Olga Kurasova
olga.kurasova@mif.vu.lt

https://doi.org/10.15388/DAMSS.15.2024
ISBN 978-609-07-1112-5 (digital PDF)

Fault Detection in Solar Power Plants Using Energy Production Data

Dominykas Vilčinskas, Lukas Voveris, Jolita Bernatavičienė

Institute of Data Science and Digital Technologies
Vilnius University

dominykas.vilcinskas@mif.stud.vu.lt

Addressing the critical need for timely identification of faults in solar power plants is essential, as even minor malfunctions can lead to substantial electricity losses over time, reducing overall efficiency and profitability. This study focuses on the analysis of energy production data from a solar power plant in Lithuania, consisting of 143 strings distributed across 12 inverters, collected over a 19-month period. To facilitate the analysis, 16 key features were extracted from each string's time series data during the preprocessing phase. These features capture the global structure of the data, transforming each time series into an object with 16 representative attributes. This transformation made the data more suitable for further analysis, allowing for more effective identification of anomalies. Various statistical and machine-learning techniques were applied to detect systems exhibiting abnormal behaviour. The combination of Principal Component Analysis (PCA) and Alpha-Hull methods was used to reduce dimensionality and identify anomalous systems by isolating points outliers. Alongside this, other machine learning algorithms, such as Isolation Forest (iForest) and Local Outlier Factor (LOF), were employed to detect anomalous systems. The results suggest that these methods can potentially identify solar energy generation systems exhibiting abnormal behaviour, with a combined anomaly score offering a comprehensive assessment of string performance. This approach provided valuable insights into which systems were underperforming or exhibiting abnormal behaviour. In addition to the aforementioned techniques, the study also utilised Random Sample Consensus (RANSAC) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) methods to construct detailed fault profiles, which enabled a more in-depth analysis of each

system's performance, offering further confirmation of previously identified outlying systems. This combination of approaches demonstrated its potential for fault detection in solar power plants.

Acknowledgment: The conference participation is funded by EPAM.