



DAMSS

DATA ANALYSIS
METHODS FOR SOFTWARE
SYSTEMS



15th Conference on

DATA ANALYSIS METHODS for Software Systems

November 28–30, 2024

Druskininkai, Lithuania, Hotel “Europa Royale”

<https://www.mii.lt/DAMSS>

Co-Chairmen:

Prof. **Gintautas Dzemyda** (Vilnius University, Lithuanian Academy of Sciences)

Dr. **Saulius Maskeliūnas** (Lithuanian Computer Society)

Programme Committee:

Dr. **Jolita Bernatavičienė** (Lithuania)

Prof. **Juris Borzovs** (Latvia)

Prof. **Janis Grundspenkis** (Latvia)

Prof. **Janusz Kacprzyk** (Poland)

Prof. **Ignacy Kaliszewski** (Poland)

Prof. **Bożena Kostek** (Poland)

Prof. **Tomas Krilavičius** (Lithuania)

Prof. **Olga Kurasova** (Lithuania)

Assoc. Prof. **Tatiana Tchemisova** (Portugal)

Assoc. Prof. **Gintautas Tamulevičius** (Lithuania)

Prof. **Julius Žilinskas** (Lithuania)

Organizing Committee:

Dr. **Jolita Bernatavičienė**

Prof. **Olga Kurasova**

Assoc. Prof. **Viktor Medvedev**

Laima Paliulionienė

Assoc. Prof. **Martynas Sabaliauskas**

Prof. **Povilas Treigys**

Contacts:

Dr. Jolita Bernatavičienė

jolita.bernatavicienne@mif.vu.lt

Tel. (+370 5) 2109 315

Prof. Olga Kurasova

olga.kurasova@mif.vu.lt

Copyright © 2024 Authors. Published by Vilnius University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://doi.org/10.15388/DAMSS.15.2024>

ISBN 978-609-07-1112-5 (digital PDF)

© Vilnius University, 2024

Highly Imbalanced Data Case: Pattern-Guided Feature Selection to Detect Financial Fraud

Dalia Breskuvienė, Gintautas Dzemyda

Institute of Data Science and Digital Technologies
Vilnius University

dalia.breskuviene@mif.vu.lt

Financial fraud represents widespread challenges in the modern financial landscape, necessitating innovative methodologies and technologies to detect and mitigate losses resulting from sophisticated fraudulent tactics.

Fraud detection systems need to respond quickly while maintaining the stability of the financial ecosystem. To address the critical issues identified in fraud detection, we introduce a novel feature selection method, FID-SOM (Feature Selection for Imbalanced Data Using Self-Organizing Maps), specifically designed to overcome the challenges posed by imbalanced data in fraud detection scenarios. Feature selection can significantly enhance classification performance related to accuracy and response time. Given the inherent imbalance in fraud detection data, feature selection must be conducted with high attention. To accomplish this task, we utilize Self-Organizing Maps (SOMs), an artificial neural network designed to organize data into clusters based on similarity, simplifying the complex data landscape typical of financial fraud scenarios. FID-SOM is engineered to address the challenge of high-dimensional data in scenarios characterized by highly imbalanced data. It has been specifically developed to efficiently process and analyze vast, complex datasets commonly encountered in the financial sector. More importantly, it demonstrates adaptability to the dynamic nature of big data environments, thereby making it a practical and reassuring solution for fraud detection.

The proposed method's distinctive aspect lies in forming a new dataset containing the Best-Matching Units of the trained SOM as vectors of attributes corresponding to the initial features. These attributes are

sorted in descending order based on the score calculated using methods like KL-divergence Information Gain and/or Variance. By retaining the required number of attributes, we select features corresponding to those attributes for further analysis. The proposed FID-SOM method has demonstrated its capacity to perform comparably to existing methods and exhibits the potential to surpass them. This potential for superior performance makes FID-SOM a thrilling prospect in fraud detection.