DMAKit: A user-friendly web platform for bringing state-of-the-art data analysis techniques to non-specific users

Medina-Ortiz D.

Contreras S.

Quiroz C.

Asenjo J.A.

Olivera-Nappa Á.

Tremendous advances in different areas of knowledge are producing vast volumes of data, a quantity so large that it has made necessary the development of new computational algorithms. Among the algorithms developed, we find Machine Learning models and specific data mining techniques that might be useful for all areas of knowledge. The use of computational tools for data analysis is increasingly required, given the need to extract meaningful information from such large volumes of data. However, there are no free access libraries, modules, or web services that comprise a vast array of analytical techniques in a user-friendly environment for non-specific users. Those that exist raise high usability barriers for those untrained in the field as they usually have specific installation requirements and require in-depth programming knowledge, or may result expensive. As an alternative, we have developed DMAKit, a user-friendly web platform powered by DMAKit-lib, a new library implemented in Python, which facilitates the analysis of data of different kind and origins. Our tool implements a wide array of state-of-the-art data mining and pattern recognition techniques, allowing the user to quickly implement classification, prediction or clustering models, statistical evaluation, and feature analysis of different attributes in diverse datasets without requiring any specific programming knowledge. DMAKit is especially useful for users who have large volumes of data to be analyzed but do not have the informatics, mathematical, or statistical knowledge to implement models. We expect this platform to provide a way to extract information and analyze patterns through data mining techniques for anyone interested in applying them with no specific knowledge required. Particularly, we present several cases of study in the areas of biology,

biotechnology, and biomedicine, where we highlight the applicability of our tool to ease the labor of non-specialist users to apply data analysis and pattern recognition techniques. DMAKit is available for non-commercial use as an open-access library, licensed under the GNU General Public License, version GPL 3.0. The web platform is publicly available at https://pesb2.cl/dmakitWeb. Demonstrative and tutorial videos for the web platform are available in https://pesb2.cl/dmakittutorials/. Complete urls for relevant content are listed in the Data Availability section. © 2020 Elsevier Ltd

Data mining

Machine learning

Pattern recognition

Statistics

User-friendly web platform

Classification (of information)

Data handling

HTTP

Machine learning

Open source software

Pattern recognition

Web services

Computational algorithm

Data analysis techniques

GNU general public license

Machine learning models

Pattern recognition techniques

Programming knowledge

Statistical evaluation

Statistical knowledge

Data mining