

Geochemistry π Open Source Software: Machine Learning for Geochemists Who Don't Want to Code

ZHOU ZHANG

Zhejiang University

Geochemistry π is an open-source automated machine learning Python framework. Geochemists need only provide tabulated data (e.g. excel spreadsheet) and select the desired options to clean data and run machine learning algorithms. The process operates in a question-and-answering format, and thus does not require that users have coding experience. Version 0.7.0 includes machine learning algorithms for regression, classification, clustering, dimension reduction and anomaly detection. After either automatic or manual parameter tuning, the automated Python framework provides users with performance and prediction results for the trained machine learning model. Based on the scikit-learn library, Geochemistry π has established a customized automated process for implementing machine learning. The Python framework enables extensibility and portability by constructing a hierarchical pipeline architecture that separates data transmission from algorithm application. The AutoML module is constructed using the Cost-Frugal Optimization and Blended Search Strategy hyperparameter search methods from the A Fast and Lightweight AutoML Library, and the model parameter optimization process is accelerated by the Ray distributed computing framework. The MLflow library is integrated into machine learning lifecycle management, which allows users to compare multiple trained models at different scales and manage the data and diagrams generated. In addition, the front-end and back-end frameworks are separated to build the web portal, which demonstrates the machine learning model and data science workflow through a user-friendly web interface. In summary, Geochemistry π provides a Python framework for users and developers to accelerate their data mining efficiency with both online and offline operation options. All source code is available on GitHub (<https://github.com/ZJUEarthData/geochemistrypi>), with a detailed operational manual catering to both users and developers (<https://geochemistrypi.readthedocs.io/en/latest/>).