Geochemistry π: Automated Machine Learning Python Framework for Tabular Data

DR. ZHOU ZHANG[1] AND CAN HE[2]

[1]Zhejiang University
[2]National University of Singapor
Presenting Author: zhangzhou333@zju.edu.cn

Although machine learning (ML) has brought new insights into geochemistry research, its implementation is laborious and time-consuming. Here, we announce Geochemistry π, an open-source automated ML Python framework. Geochemists need only provide tabulated data and select the desired options to clean data and run ML algorithms. The process operates in a question-and-answering format, and thus does not require that users have coding experience. After either automatic or manual parameter tuning, the automated Python framework provides users with performance and prediction results for the trained ML model. Based on the scikit-learn library, Geochemistry π has established a customized automated process for implementing classification, regression, dimensionality reduction, and clustering algorithms. 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 ML 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 ML 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.

Geochemistry π
A PYTHON FRAMEWORK FOR GEOCHEMISTRY