FACULTY OF INFORMATICS AND STATISTICS · PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS · DIPLOMA THESIS

Analytical Platform Integrating CleverMiner

Ing. Petr Štípek Supervisor: RNDr. Ing. Petr Máša, Ph.D.

Motivation & Problem

GUHA-based association rule mining with CleverMiner is powerful, but using it requires programming knowledge - putting it out of reach for domain experts without a coding background.

Established applications like LISp-Miner provide outdated experience and do not fully cover the CRISP-DM methodology for data analysis. Other solutions provide different algorithms to association rule mining than GUHA.

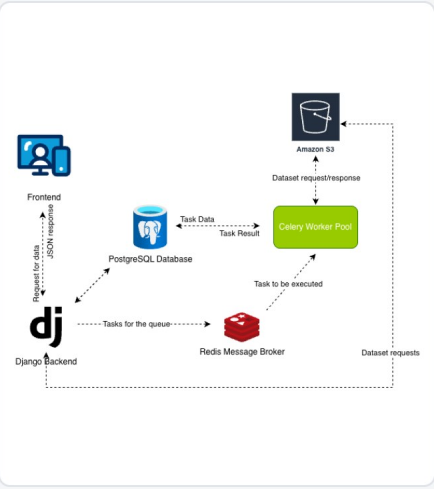
As the industry shows growing demand for low-code and no-code solutions a platform integrating the full CRISP-DM process for the GUHA procedures through CleverMiner library directly fits the possible demand.

The solution needs to be available for students and researchers as well as to the domain experts. Therefore supporting multiple deployment strategies.

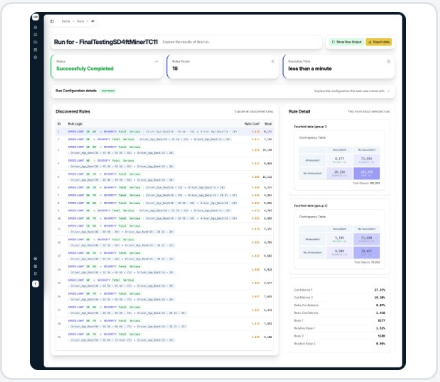
Approach / System Design

A web application with a React frontend, a Django backend, and a Celery-based distributed task execution layer with Redis, supporting all four CleverMiner procedures.

Distributed task execution used to overcome Python's GIL problem, since CleverMiner tasks are CPU heavy and can run for hours.



Delivered Platform



Platform covers the full CRISP-DM workflow: dataset exploration, preprocessing, mining task definition, execution, and result visualisation.

Deployable locally via Docker or on cloud infrastructure.

Supports collaboration between registered users

The attached screenshot shows a result page with all rules found for the selected parameters for the SD4ft-Miner. Each rule is accompanied with statistical data supporting the claim.

Evaluation

User acceptance testing was conducted with students familiar with comparable tools through the MMSP-AV methodology employed throughout development.

The application is set to be used in one introductory subject at the school.

Real-Life Applications

The benefit is democratization of the CleverMiner library and its data analysis capabilities.

The platform consolidates the entire data analysis pipeline into a single application, reducing friction with working in multiple tools in data analysis.

Future Work

Exploring other capabilities in data preprocessing or further improving established strategies.

Developing a cloud-native version without balancing the need for local deployment would further allow for more collaboration, better data handling through sources like S3, and possible introduction of a pay-per-use model for financial stability of the project.

Conclusion

The thesis delivers a full-stack, no-code analytical platform that opens GUHA-based data mining with CleverMiner to domain experts without programming knowledge, integrating the complete CRISP-DM workflow behind an accessible web UI, through multi-deploy strategy, allowing everyone to use the platform.

References:
ŠTÍPEK, Petr. Analytical Platform Integrating CleverMiner. Online. Master's thesis. Praha: University of Economics, Prague. 2026. Available from: https://theses.cz/id/hv6pbb/.
https://github.com/petrstipek/Analytical-Platform-Integrating-CleverMiner.

Ing. Petr Štípek
petr@stip.cz