Evaluating machine learning models in a SIEM environment

Creating a testbed for systematic evaluation of anomaly detection ML models in the ELK Stack

Students

Matteo Mahler

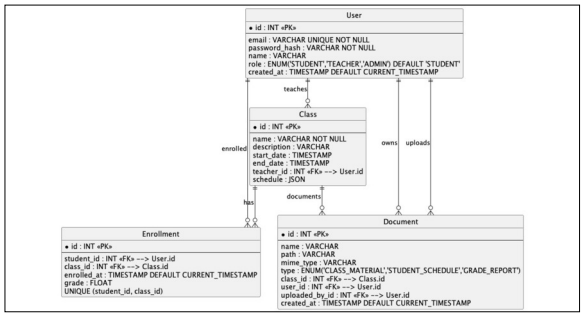
Martin Arendar

Problem: Modern digital services generate large amounts of log data that can be used to identify security relevant anomalies in user behavior. Traditional Security Information and Event Management (SIEM) systems rely mainly on predefined rules to detect suspicious activity. While effective for known attack patterns, these rule based approaches struggle to capture complex or previously unseen behavior.
Machine learning offers a promising alternative, but its practical application is often limited by missing datasets, unclear evaluation procedures, and difficult integration into operational systems.

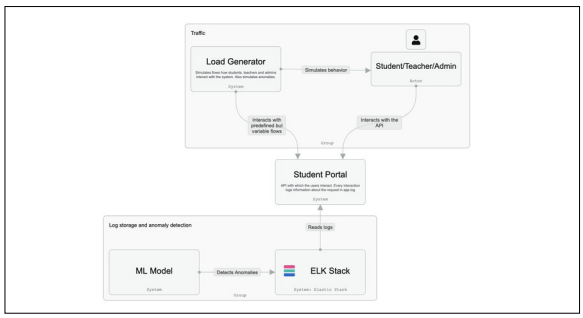
Approach / Technology: This project presents the design and implementation of a testbed for evaluating machine learning based anomaly detection within the Elastic Stack.
Instead of focusing on a single optimized model, the goal is to provide an environment in which different models can be trained, deployed, and compared under identical conditions.
To address the lack of suitable datasets, a synthetic data generation pipeline was developed.
A custom web application simulates realistic user interactions as well as predefined anomalous behavior, producing standardized log data suitable for machine learning.
The generated logs are ingested into Elasticsearch using Elastics data pipeline tools and transformed into session-level features that represent user behavior over time.
Two different machine learning approaches are evaluated on this dataset: Elastic's built-in Data Frame Analytics classifier and a custom Random Forest model trained in Python and integrated into Elastic.
Their performance is compared using standardized evaluation metrics and interactive dashboards.

Result: The results show that the testbed operates reliably and enables consistent model comparison. While Elastics built in solution provides strong baseline results with minimal setup, the custom model highlights the importance of feature engineering and data quality.
Overall, the project demonstrates how machine learning can be systematically evaluated in a SIEM environment and provides a foundation for future work involving improved models, real world data, or real time anomaly detection.

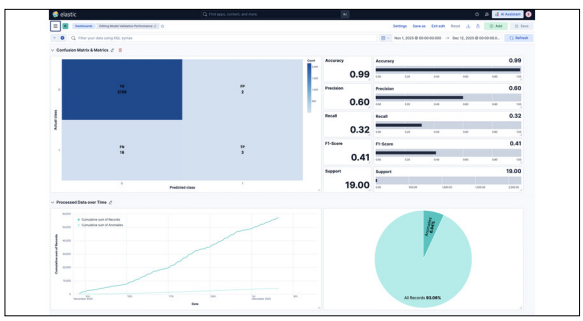
ERD diagram of the backend database
Own presentment



Context diagram of the application landscape
Own presentment



Evaluation dashboard of the custom machine learning model
Own presentment



Advisor
Prof. Dr. Daniel Patrick Politze

Subject Area
Artificial Intelligence, Network and Cloud Infrastructure, Software Engineering