A Monitoring Framework for Machine Learning in Healthcare

Adam Ekblom

Developing Machine Learning (ML) models for healthcare is a complicated task. As a result, few ML applications reach the clinic, and fewer show success for their applied purpose. However, novel monitoring frameworks that manage the quirks of ML can apply Model Monitoring, ensuring the effectiveness and safety of healthcare applications. In the future, combinations of novel techniques can further the integration of ML in healthcare.

Clinical solutions of ML and Artificial Intelligence (AI) are tools that would assist clinicians and patients to administer and receive the best possible treatment. However, the road to these solutions is complicated. Up to 85% of all ML and AI projects fail, as most ML models that show promise in the computer lab do not work as expected in real-life clinical settings.
One reason for the large fail rate is the absence, and thereof need, of established development processes and practices that consider the quirks of ML development, deployment, and monitoring for commercial practice, in this case, the real-life clinical setting. With the increasing demand for more effective care, software solutions such as ML applications that increase effectiveness and assist clinicians are needed now more than ever.

With new regulations considering the particularities of ML development, the thesis aims to determine the possibility of implementing model monitoring using an MLOps (Machine Learning Operations) framework for clinical decision support. MLOps is a further development of the established DevOps framework, an agile approach for software development, which is a set of processes for managing models, data, and code for improving performance stability and long-term efficiency in ML systems.
The framework is evaluated by creating a use case with real-life clinical data for training, evaluating, deploying, and monitoring a Deep Learning ML model that predicts the length of stay of a patient admitted to the Intensive Care Unit (ICU). The model's performance is evaluated by altering the clinical data and comparing the performance with statistical metrics.

An end-to-end framework deployed the model and monitored its performance, detecting a significant performance drift where it lost performance. The framework shows promise with clear structure and efficiency. However, there are apparent limitations, such as retraining the model, the disparity of the use case from a real-life clinical setting, and whether the specific framework used is appropriate for the real-life clinical setting. Conclusively, the framework shows promise, and there is much potential for future research utilizing novel ML techniques.