

# Monitoring Machine Learning Forecasts for Platform Data Streams

J. Rombouts

Machine learning algorithms hold great promise in many prediction problems. Yet, one of their important drawbacks is that they are computationally too costly to frequently re-train in production environments requiring large-scale forecasting at the same speed as new data batches enter. We propose a data-driven monitoring procedure to answer the important question when the machine learning algorithm should be re-trained. Our decision rule to re-train is based on the streaming forecast loss produced by the machine learning algorithm, where a significant change in forecast loss of the new data batch compared to a well-defined reference batch triggers re-training of the machine learning algorithm to avoid forecast deterioration. As an empirical application, we consider demand forecasting at a last-mile-delivery platform and demonstrate the value of the monitoring procedure for several state-of-the-art machine learning methods including random forest, gradient boosting and lasso.