

Predictive Maintenance und Qualitätsmanagement für chemische Analyselabore (PreMaQLab)

Während prädiktives Wartungsmanagement im Maschinen- -und Anlagenbau inzwischen zum fortgeschrittenen Stand der Technik gehört, gibt es für das Management von Laborgeräten in Analyselaboren keine auch nur ansatzweise vergleichbaren Ansätze und Lösungen. Die Daten aus Laboranalysen und den verwendeten Geräten sind zu heterogen und nicht einheitlich zu interpretieren. Im Rahmen des ZIM-Projekts PreMaQLab wurde vor diesem Hintergrund eine softwaretechnische Lösung entwickelt, um Abweichungen von kritischen Qualitätsparametern frühzeitig erkennen zu können und eine Prognose in Bezug auf die zu erwartende Qualität von Analyseergebnissen zu geben. Dies umfasst sowohl flüssigkeitschromatografische Trennverfahren als auch massenspektrometrische Detektionstechniken.

Für eine breite Anwendung des Verfahrens müssen die in der Regel proprietären Herstellerformate in offene Messdatenformate konvertiert werden. Im Unterschied zu den chromatografischen Daten, die im Wesentlichen zeitaufgelöste Messgrößen wie Druck, Temperatur oder Signalintensität abbilden, weisen massenspektrometrische Daten durch die zusätzliche Dimension des Masse-zu-Ladungs-Verhältnisses (m/z), eine deutlich höhere Komplexität auf.

Zur Optimierung der Anomalieerkennung wurde zunächst ein standardisiertes Vorgehen auf Basis des Isolation-Forest implementiert, bei dem die ersten vier Analyseläufe einer Messreihe pauschal ausgeschlossen wurden. Anschließend erfolgte das Training des Modells mit den verbleibenden Daten bis zu einem definierten Zeitpunkt. Alle nachfolgenden Messungen wurden mit dem trainierten Modell abgeglichen, sodass potenzielle Ausreißer identifiziert werden konnten. Aufgrund dieses Ansatzes war der Trainingsdatensatz jedoch vergleichsweise klein und musste bei Bedarf manuell erweitert werden. Zudem bestand die Gefahr eines Daten-Drifts. Neben dem Isolation-Forest wurden alternative Verfahren wie DBSCAN, One-Class SVM und PCA evaluiert. Diese Ansätze zeigten jedoch ein deutlich geringeres Potenzial für die zuverlässige Anomalieerkennung.

Im weiteren Entwicklungsprozess stellte sich heraus, dass nicht nur die Größe des Trainingsdatensatzes die Ergebnisqualität maßgeblich beeinflusste, sondern auch die parallele Durchführung mehrerer Isolation-Forest-Modelle mit reduzierter Probenzahl. Die Zuverlässigkeit der Anomalieerkennung blieb auch bei kleineren Datensätzen auf konstant hohem Niveau, erforderte jedoch eine höhere Anzahl an Modellen.

Dieser Effekt wurde für die Optimierung des bestehenden Konzepts gezielt genutzt. Wie in Abbildung 1 dargestellt, wird die jeweilige Messserie parallel mehreren Isolation-Forest-Instanzen zugeführt. Obwohl alle Modelle auf demselben Algorithmus basieren, führen die zufällig gewählten Parameterkonstellationen zu unterschiedlichen Klassifikationen: Ein identischer Datensatz kann von einzelnen Instanzen als Ausreißer und von anderen als Normalprobe bewertet werden. Liegt eine

hinreichend hohe Übereinstimmung der Modelle in der Einstufung als Normalprobe vor, wird die zugrunde liegende Kurve als validierter neuer Datensatz in den Isolation-Forest übernommen. Auf diese Weise wird die Vorhersagefähigkeit des Modells iterativ verbessert und kontinuierlich an neue Daten angepasst.

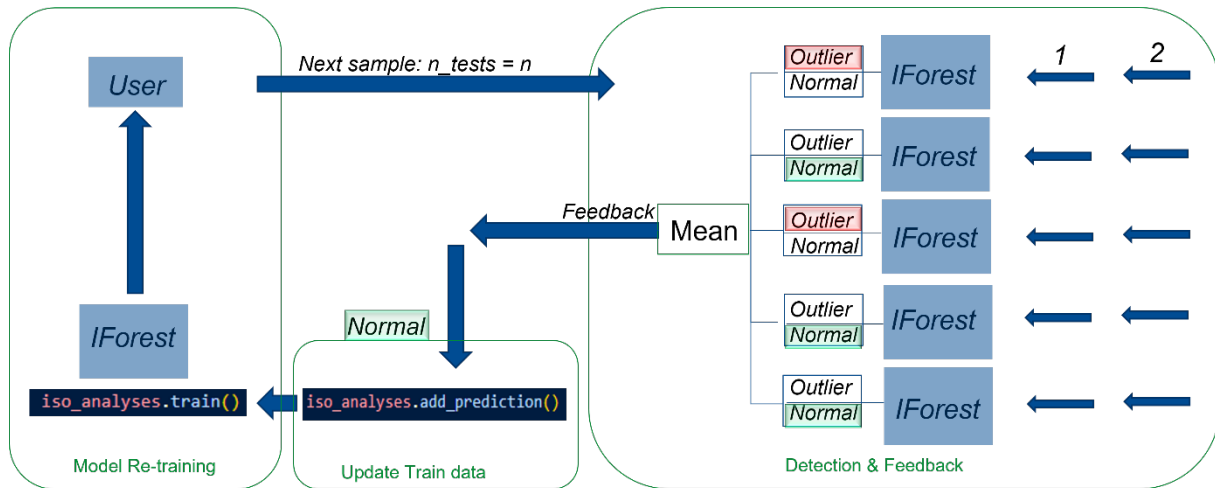


Abbildung 1: Überarbeitetes Konzept zur Anomalieerkennung. Messungen werden fünf parallelen Isolation Forest Modellen zugeführt. Bei einer mehrheitlichen Identifikation als normale Spur wird diese zum weiteren Training der Modelle genutzt.

Zum Ende des Projekts wurde ein browserbasiertes User Interface entwickelt. Die Eingabemaske ermöglicht die Auswahl der zu analysierenden Messdaten sowie der entsprechenden Methode. Darüber hinaus konnten Parameter der Datenaufbereitung und des eingesetzten Machine-Learning-Algorithmus angepasst werden. Die Visualisierung der Ergebnisse erfolgte ebenfalls im Browser, wodurch sowohl die Detektion von Anomalien als auch die Überprüfung des zugrunde liegenden Modells unterstützt wurde. Mit Abschluss des Projekts konnten somit alle vorher definierten Meilensteine eindeutig erreicht werden.