

Auto-Anomalieerkennung ML-Neutrainierprozess

Motivation

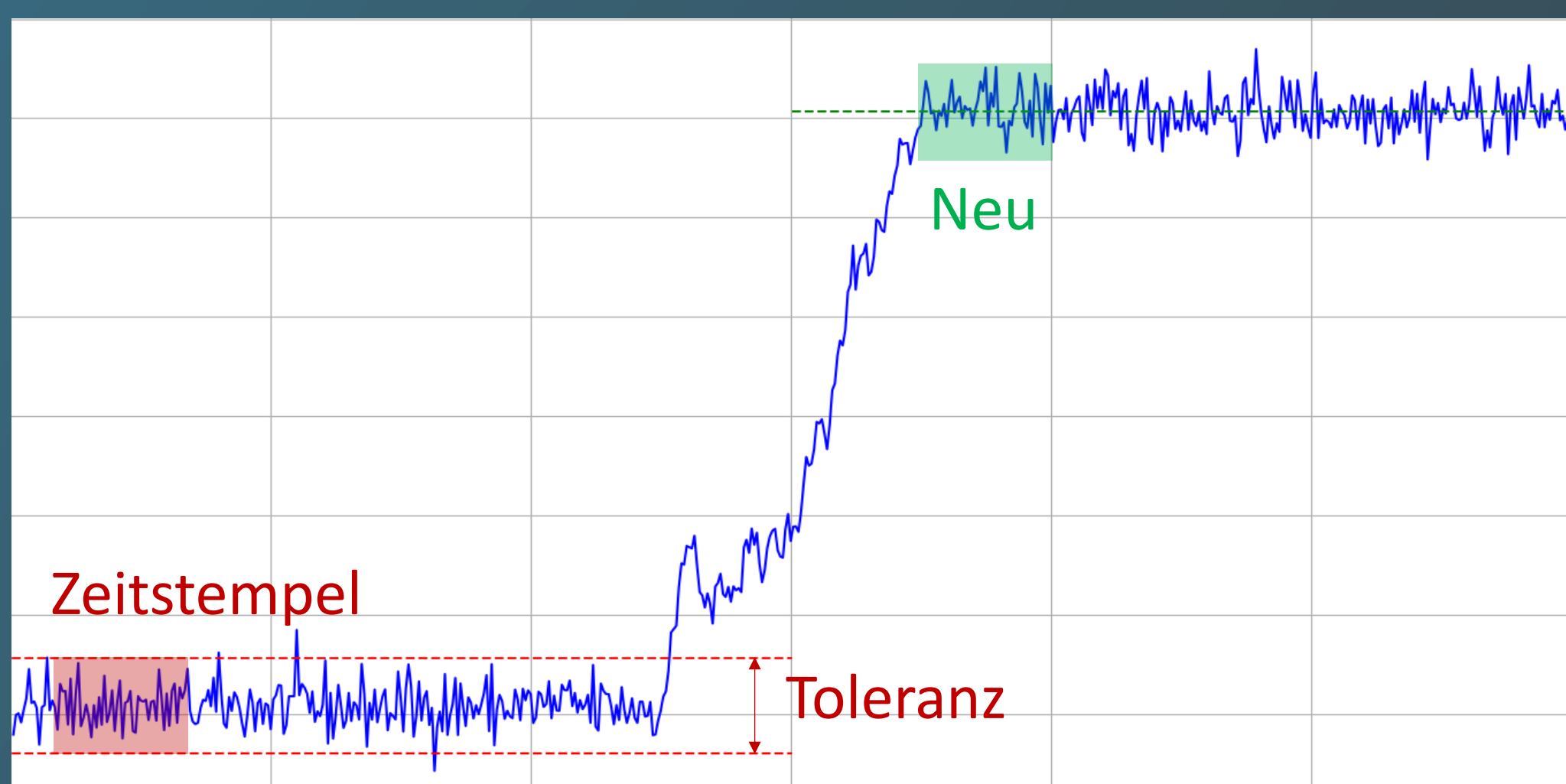
Im **Data Analytics & Observability** Squad der DT-Technik werden Log-Daten mithilfe von drei verschiedenen **Machine-Learning(ML)** Algorithmen auf Anomalien überprüft. Mit der Zeit verlieren die Modelle jedoch an Genauigkeit, was die Qualität der Anomalieerkennung beeinträchtigt.

Ziel: Die Entwicklung einer Code-Erweiterung, die die Genauigkeit der ML-Modelle langfristig erhält und somit eine verlässliche Anomalieerkennung sicherstellt.

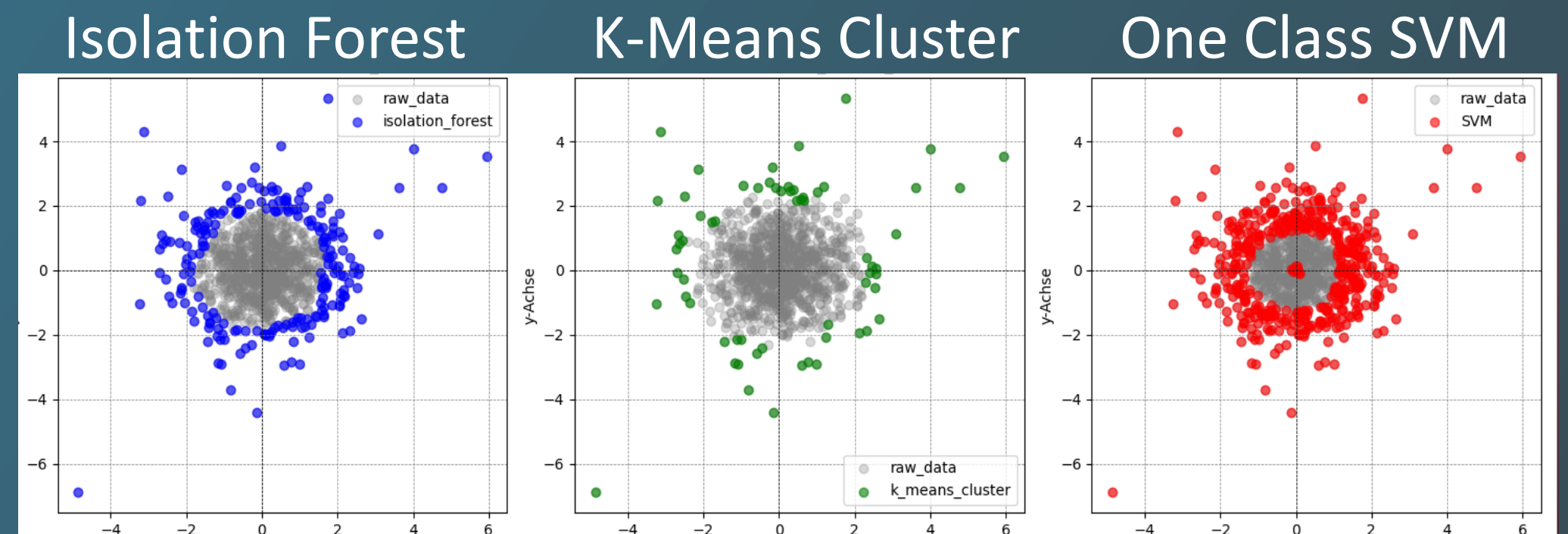
Ansatz

Bei den ersten Designansätzen wurde deutlich, dass die Messung der Performance der ML-Algorithmen eine große Herausforderung darstellt. Der Grund dafür ist, dass unsupervised Learning nicht einfach in „richtig“ oder „falsch“ klassifiziert werden kann. Stattdessen erwies es sich als praktikabler, die Veränderung der Datenbasis zu analysieren, um das Re-training auszulösen – durch die Erkennung von Data Drift.

Datadrift Beispiel



Verwendete ML-Algorithmen



Ergebnisse

Die **Datendrifterkennung** wurde implementiert und sorgt durch die Speicherung und Auslesung von Momentaufnahmen der Daten für die **qualitative Erhaltung der Anomalieerkennung**.

Sie erkennt Veränderungen in den Daten, löst bei Bedarf ein Neutrainig der Modelle aus und ermöglicht den automatischen Austausch des alten Modells durch ein neutrainiertes im Live-Betrieb. Eine vorläufige Bewertungsmethode für die Modelle ist implementiert und wird weiter konzeptioniert, um sie in ihrer Performance vergleichbar zu machen.

Ausblick

Zukünftig soll eine gezielte Bewertung einzelner ML-Algorithmen erfolgen, um ihre Leistungsfähigkeit besser einschätzen zu können. Zudem wird daran gearbeitet, Modelle flexibel an leichte Datendriffs anzupassen, ohne sie vollständig zu verwerfen. Dies soll durch ein spezifisches Nachtrainieren mit einer reduzierten Datenmenge erreicht werden, um die Stabilität und Anpassungsfähigkeit der Modelle zu verbessern.

