



Impulsvortrag: Anwendungspotenziale von KI auf Kläranlagen und in Abwassernetzen

Eifeler Abwassertag 2023 der BITControl: Der Pfad zur Energieneutralität in der
Abwasserreinigung

Trier, 14. September 2023
Prof. Dr. Mark Oelmann

MOcons berät und HRW forscht in (ökonomischen) Fragestellungen der Wasser und Abwasserwirtschaft



Kernthemen:

- Preis-/Gebührenmodelle
- Effizienzanalysen
- Prognosen (Kosten, Gebühren, Wasserbedarf ...)
- Studiererstellung
- Digitalisierungsthemen:
 - Makro-Ebene: Digitalisierungsindex für die deutsche Wasserwirtschaft (www.digitalisierungsindex-wasserwirtschaft.de)
 - Meso-Ebene: Digitale Reifegradmodelle für sowohl Wasserver- als auch Abwasserentsorger
 - Mikro-Ebene: Maschinelles Lernen

....

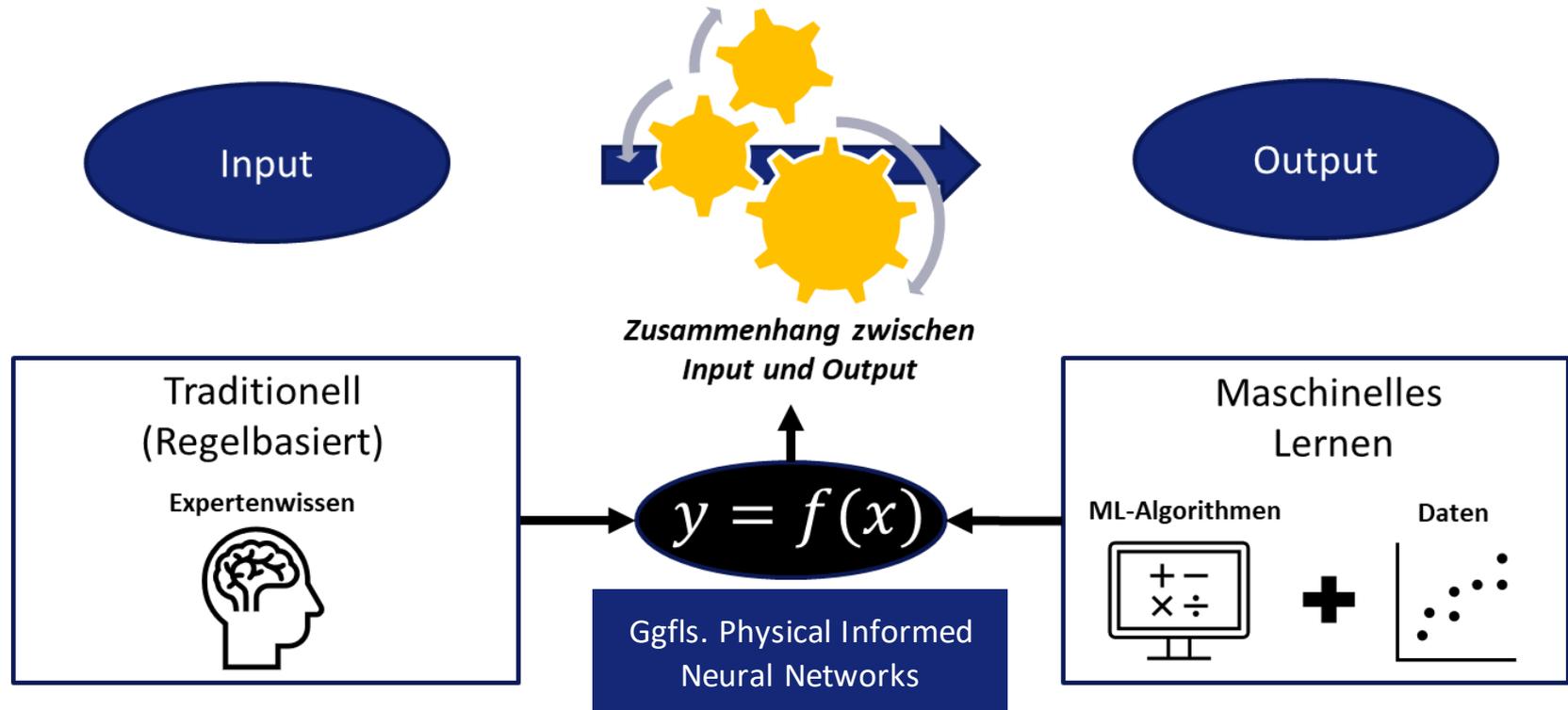
Übersicht

- 1. Was ist Künstliche Intelligenz bzw. Maschinelles Lernen?**
- 2. Schritte in einem Projekt zum „Maschinellen Lernen“**
- 3. Fallstricke: Einblicke in Umsetzungspraxis**
 1. Zur Identifikation eines Anwendungsfalls
 2. Keine Zeit verlieren: Das Thema der Datenverfügbarkeit
 3. Das Zünglein an der Waage: Datenplausibilität
 4. Das Doppel: „Data Scientist“ und „Wasserwirtschaftler“
 5. Nutzung der KI: Von der fertigen Lösung zur Anwendung
- 4. Und jetzt? Womit fange ich an?**

1. Maschinelles Lernen

Was verstehen wir unter Maschinellern Lernen?

Als entscheidendes Alleinstellungsmerkmal des Maschinellen Lernens kann die Fähigkeit bezeichnet werden, **selbstständig** aus Beispielen (Daten) zu lernen:



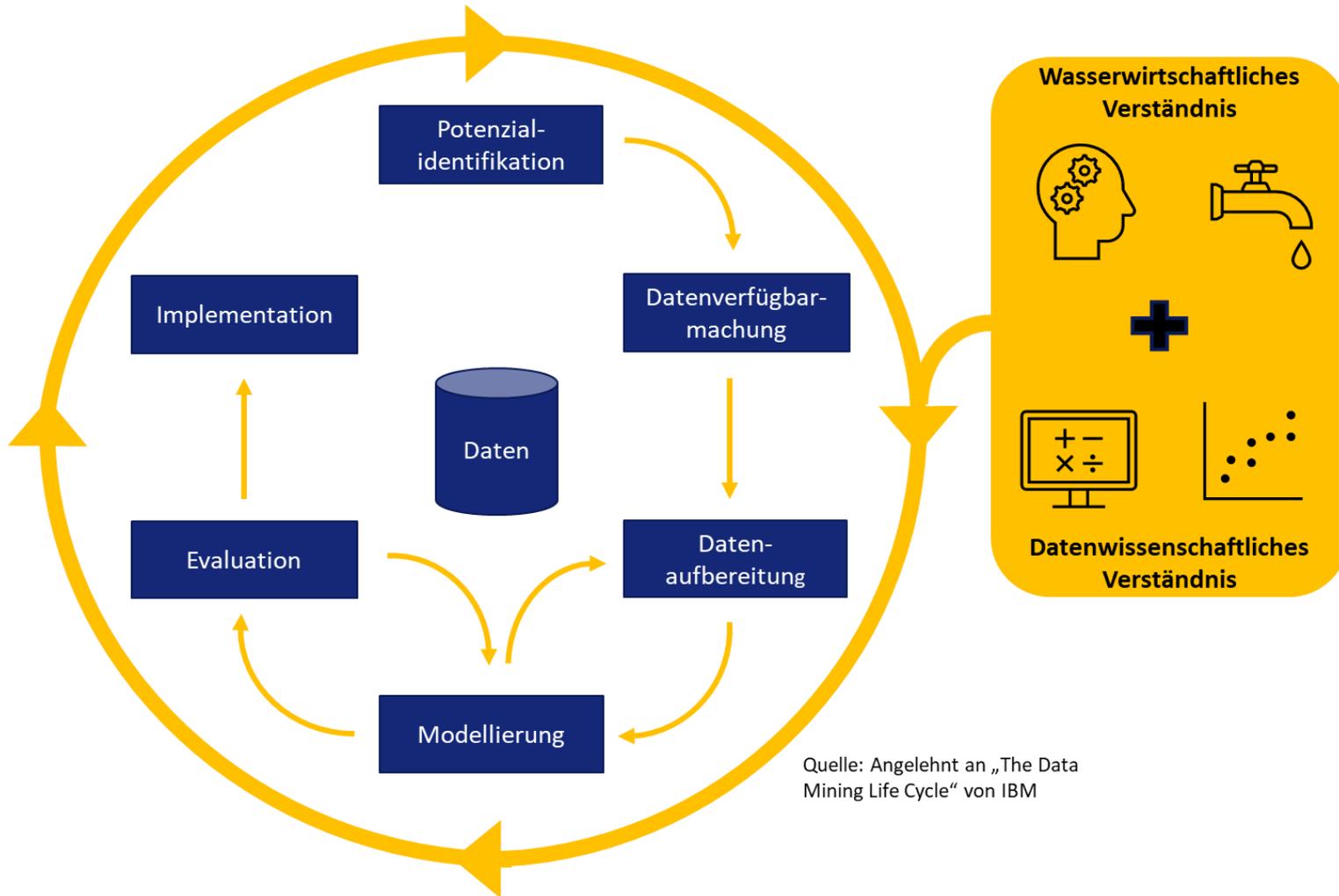
Dabei erlaubt es ML im Allgemeinen, komplexere Probleme deutlich effizienter zu lösen als dies traditionelle Herangehensweisen erlauben würden.

Übersicht

1. Was ist Künstliche Intelligenz bzw. Maschinelles Lernen?
2. Schritte in einem Projekt zum „Maschinellen Lernen“
3. Fallstricke: Einblicke in Umsetzungspraxis
 1. Zur Identifikation eines Anwendungsfalls
 2. Keine Zeit verlieren: Das Thema der Datenverfügbarkeit
 3. Das Zünglein an der Waage: Datenplausibilität
 4. Das Doppel: „Data Scientist“ und „Wasserwirtschaftler“
 5. Nutzung der KI: Von der fertigen Lösung zur Anwendung
4. Und jetzt? Womit fange ich an?

2. Schritte in einem Projekt zum „Maschinellen Lernen“

Schematische Darstellung



Quelle: Angelehnt an „The Data Mining Life Cycle“ von IBM

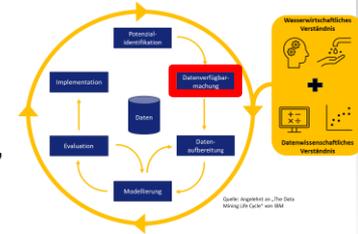
Übersicht

1. Was ist Künstliche Intelligenz bzw. Maschinelles Lernen?
2. Schritte in einem Projekt zum „Maschinellen Lernen“
3. Fallstricke: Einblicke in Umsetzungspraxis
 1. Zur Identifikation eines Anwendungsfalls
 2. Keine Zeit verlieren: Das Thema der Datenverfügbarkeit
 3. Das Zünglein an der Waage: Datenplausibilität
 4. Das Doppel: „Data Scientist“ und „Wasserwirtschaftler“
 5. Nutzung der KI: Von der fertigen Lösung zur Anwendung
4. Und jetzt? Womit fange ich an?

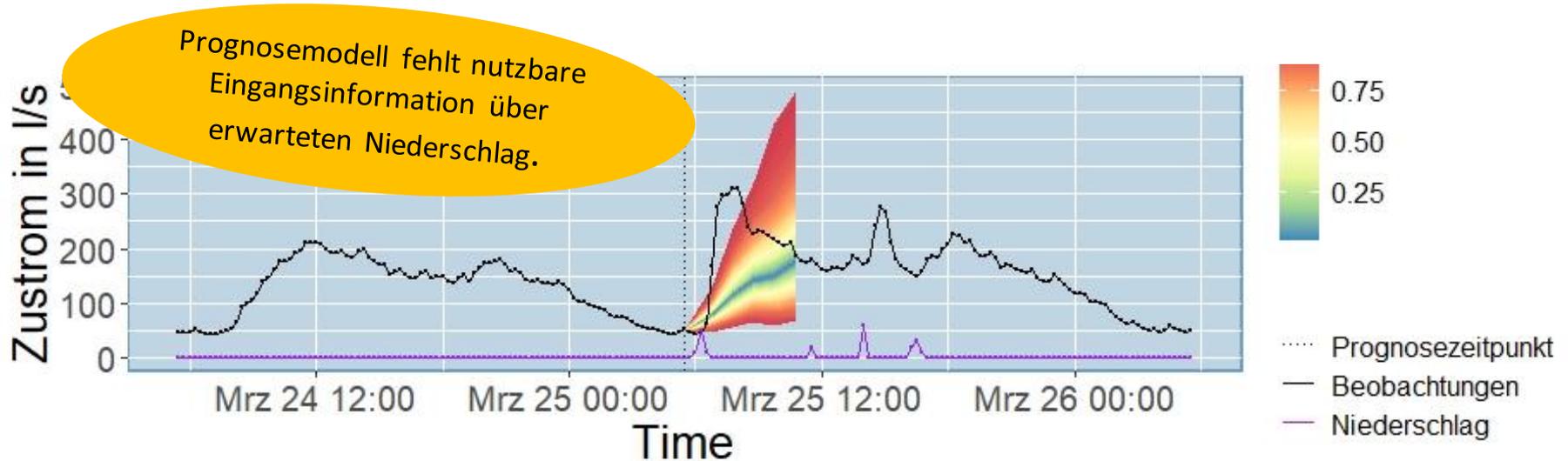
3.2 Keine Zeit verlieren: Das Thema der Datenverfügbarkeit

Sind alle benötigten Daten vorhanden?

Wasserwirtschaftliche Unternehmen verfügen über enorme Datenmengen, die zur Anwendung von ML-Algorithmen verwendet werden können.



↳ Es zeigt sich allerdings, dass oft noch Datenspuren fehlen, die extern beschafft oder durch eine interne Sensornachrüstung erhoben werden müssen.

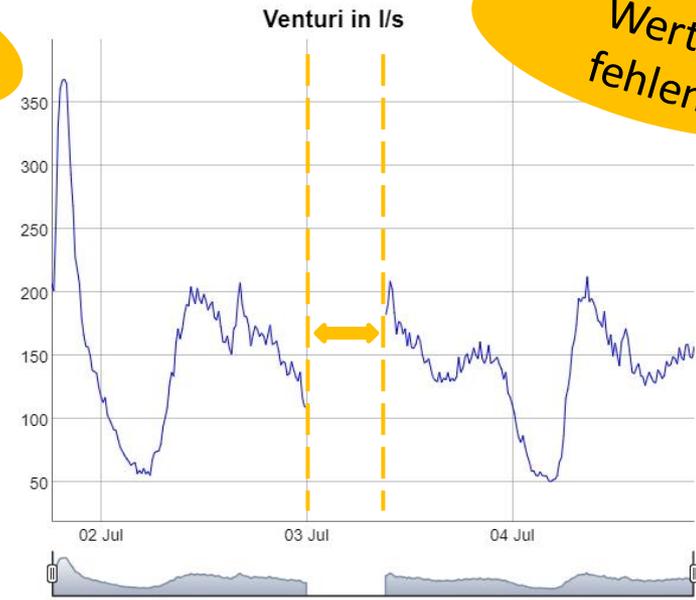
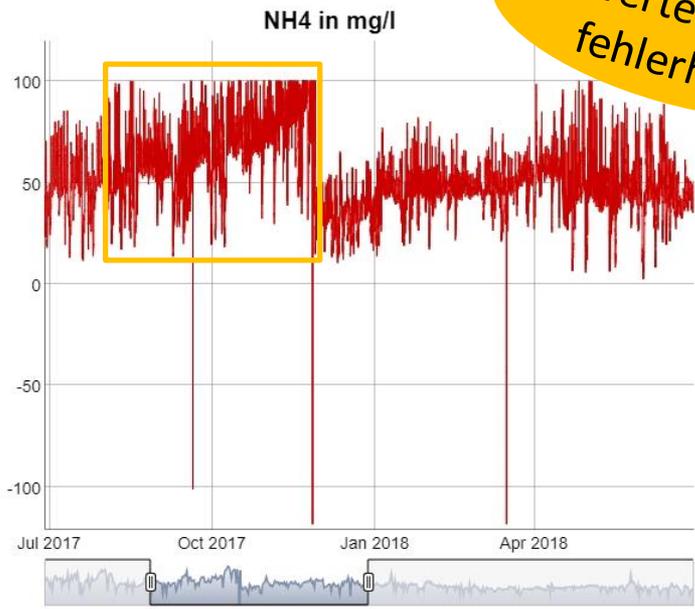
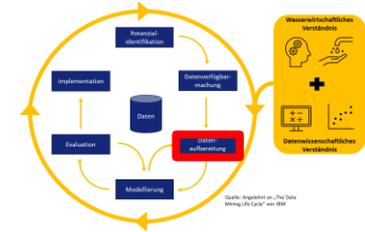


Infolge fehlender Daten kommt es zu nicht unerheblichen Verzögerungen. Und es schafft schnell Frustration...

3.3 Das Zünglein an der Waage: Das Thema der Datenplausibilität

Datenaufbereitung

Mit Blick in wasserwirtschaftliche Daten zeigt sich, dass Daten durch unterschiedlichste Ursachen verunreinigt werden können:



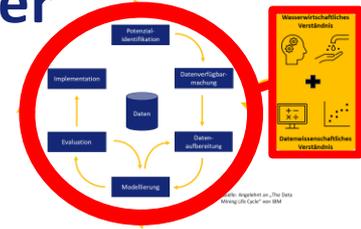
Grund: Verschmutzter Sensor

Grund: Ausfall der Sensorik

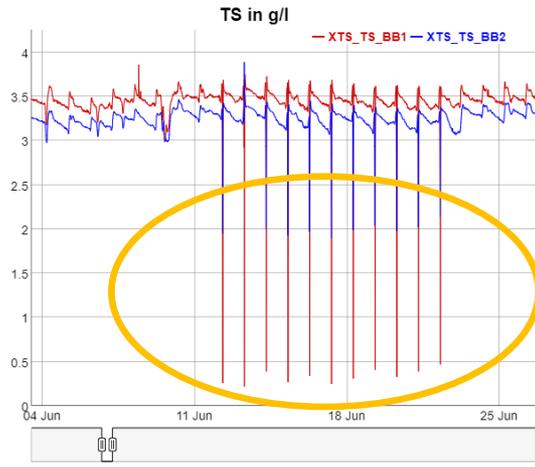
Neben der Identifikation von Anomalien (1. Schritt) ist das (automatisierte) Ersetzen von fehlerhaften und fehlenden Werten (2. Schritt) zentral, um eine brauchbare Datenbasis zur Modellierung zur Verfügung stellen zu können.

3.4 Das Doppel: „Datascientist“ und „Wasserwirtschaftler“

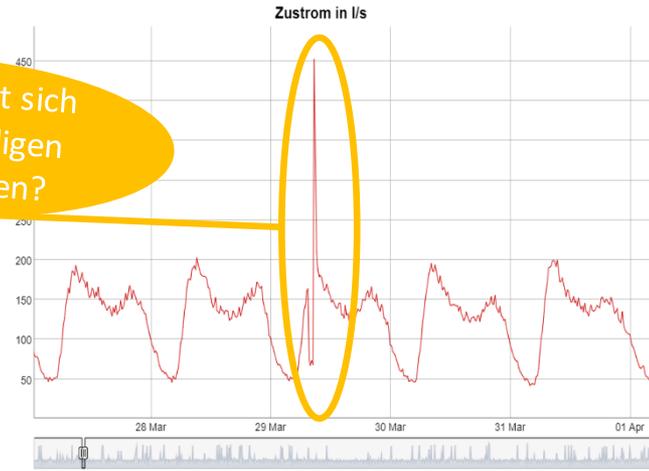
Am Beispiel: Erstellung der Trainingsdaten



Data Scientist kann Auffälligkeiten (Ausreißer) auffindig machen, aber in der Regel nicht alleine vollends erklären!



Was versteckt sich hinter auffälligen Datenpunkten?



- **Grund:** Reinigung der Belüfterelemente
- **Umgang:** Bereinigung der Werte

- **Grund:** Pumpenspiel
- **Umgang:** Bereinigung der Werte

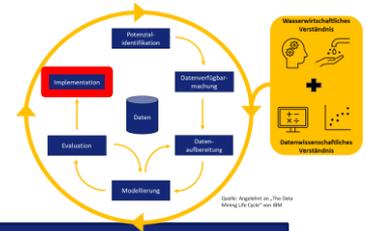
Erklärung der Ausreißer und Einordnung ihrer Auswirkungen in den Gesamtkontext kann nur durch Wasserwirtschaftler erfolgen.

Zur erfolgreichen Umsetzung von ML-Projekten ist zwingend die Einbeziehung der betrieblichen Mitarbeiter nötig! Andernfalls: Garbage in – Garbage out! Aber auch aus einem weiteren Grund...

3.5 Nutzung der KI: Von der fertigen KI-Lösung zur Anwendung

Implementation

Nach erfolgreicher Modellierung gilt es die entwickelte Lösung in den Regelbetrieb einzubinden. **Nicht zu unterschätzen ist:**



Der Faktor Mensch

- Naturegegebene Skepsis gegenüber neuer Technologie schwingt mit und erschwert Akzeptanz.
- Genau deshalb dürfen ML-Ansätze nicht zu viel auf einmal erklären wollen. Sie müssen langsam „wachsen“.
- Output der ML-Lösung muss dem Anwender effizient und effektiv zugänglich gemacht werden.

Anwendung darf keine „Glaskugel“ für den Mitarbeiter sein und sollte nicht zu Informationsflut führen.

Der Faktor Technik

- ML-Lösungen erfordern die Einbettung in eine dafür geeignete IT-Infrastruktur.
- Dabei zeigt die Praxis, dass bestehende IT-Systeme oftmals nicht ohne Weiteres eine Einbettung zulassen.

IT-Infrastruktur muss flexibel, sicher und skalierbar sein.

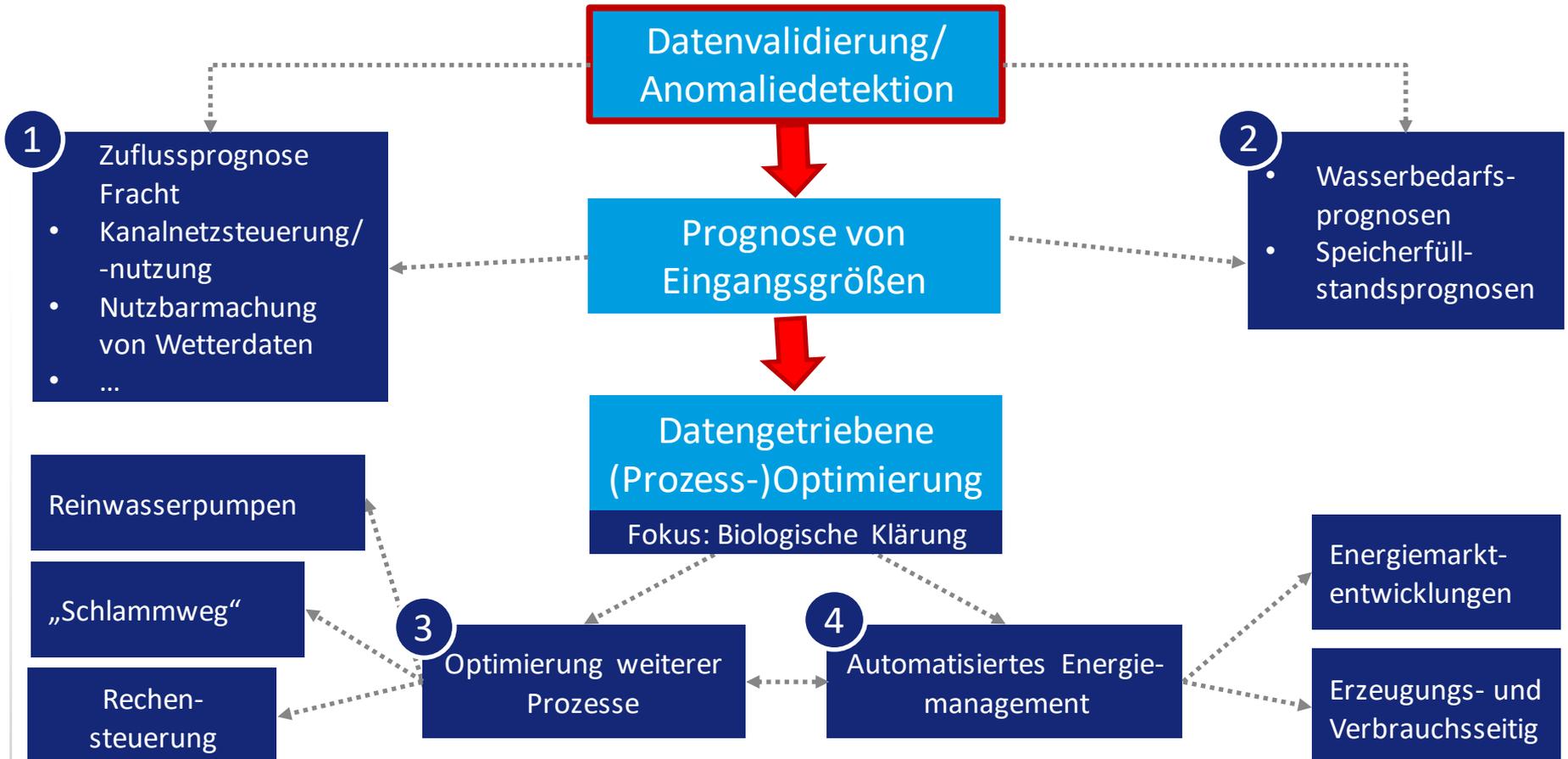
Frühe Einbeziehung der operativ Tätigen ist die Grundvoraussetzung, dass ML-Lösungen überhaupt eine Chance bekommen.

Übersicht

1. Was ist Künstliche Intelligenz bzw. Maschinelles Lernen?
2. Schritte in einem Projekt zum „Maschinellen Lernen“
3. Fallstricke: Einblicke in Umsetzungspraxis
 1. Zur Identifikation eines Anwendungsfalls
 2. Keine Zeit verlieren: Das Thema der Datenverfügbarkeit
 3. Das Zünglein an der Waage: Datenplausibilität
 4. Das Doppel: „Data Scientist“ und „Wasserwirtschaftler“
 5. Nutzung der KI: Von der fertigen Lösung zur Anwendung
4. Und jetzt? Womit fange ich an?

4. Und jetzt? Womit fange ich an?

Mit der Datenvalidierung? Wie langweilig....



4. Und jetzt? Womit fange ich an?

Kategorisierung von Problemen in Daten...

In der Regel werden die Ausgangsdaten den Anforderungen nicht gerecht.

Anomalien sind aber oft zu komplex / subtil für Detektion mit klassischen Verfahren. Bei Arbeit über Schwellwertsetzung bleibt vieles unentdeckt.

Alternative: Manuelle Datenvalidierung → diese aber nicht vergnügungssteuerpflichtig...

Arten von Anomalien

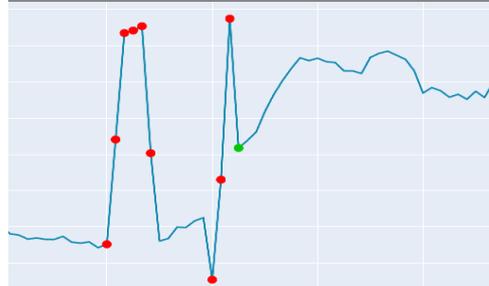
Auffällige Messwerte

Fehlende Werte



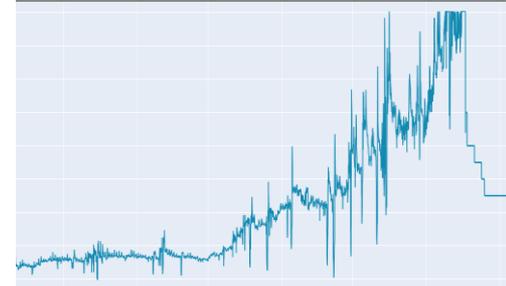
Ersetzung durch plausible Werte

Betriebsanomalien



Beibehaltung je nach Verwendungszweck, aber Identifikation wichtig

Fehlerhafte Messwerte



Sollten in der Regel bereinigt werden + bei Echtzeit Info Betriebspersonal

4. Und jetzt? Womit fange ich an?

Datenvalidierungstool entwickelt – eine Web-basierte Lösung (1)

Aufbereitete
Daten
herunterladen



Übersicht:
Anomalien auf
einen Blick

Beliebiger
Zeitausschnitt
auswählbar

Sensitivität
bestimmbar

Gefundene Anomalien können für jede Zeitreihe eingesehen werden – die Sensitivität der Algorithmen kann bei Bedarf angepasst und die Daten dann aufbereitet werden.

4. Und jetzt? Womit fange ich an?

Datenvalidierungstool entwickelt – zeitnah über Web erreichbar (3)



Daten können auch manuell als Anomalie oder als valide markiert werden - der Benutzer hat das letzte Wort und kann ferner Fachwissen über Anmerkungen verfügbar machen.

4. Und jetzt? Womit fange ich an?

Der unmittelbare Mehrwert: Beispiele aus unseren Erfahrungen.



Beispiel

Unentdeckter Sensordrift führt zu Anstieg des Energieverbrauchs und der –kosten.



Schadensausmaß

1. Direkter Einfluss auf Betrieb.
2. Fällt oftmals erst in 2ter oder 3ter Instanz auf → halbjährliche oder jährliche Energieeffizienzanalyse

Messfehler / fehlende Werte in Eingangsdaten von Prognosemodell führt zu fehlerhafter / keiner Prognose. Gleiches gilt für Optimierung.



1. Prognose ist fehlerhaft bzw. kann nicht erzeugt werden.
2. Optimierung bleibt unter seinen Möglichkeiten.

Wasserver-/Abwasserentsorger müssen immer mehr Berichte schreiben. Neue Mitarbeitende werden u.U. eingestellt und benötigen Daten.



1. Berichterstellung fehleranfällig und extrem ineffizient.
2. Besonders problematisch, wenn Berichte an Behörden gehen.

Anwendung im (Echtzeit-)betrieb erlaubt das schnelle Entdecken betrieblicher Probleme.



1. Untypische Messdaten weisen etwa auf vorgelagerten, kaputten Inputsektor hin.
2. Zu viel Sauerstoff, zu viel Energie...

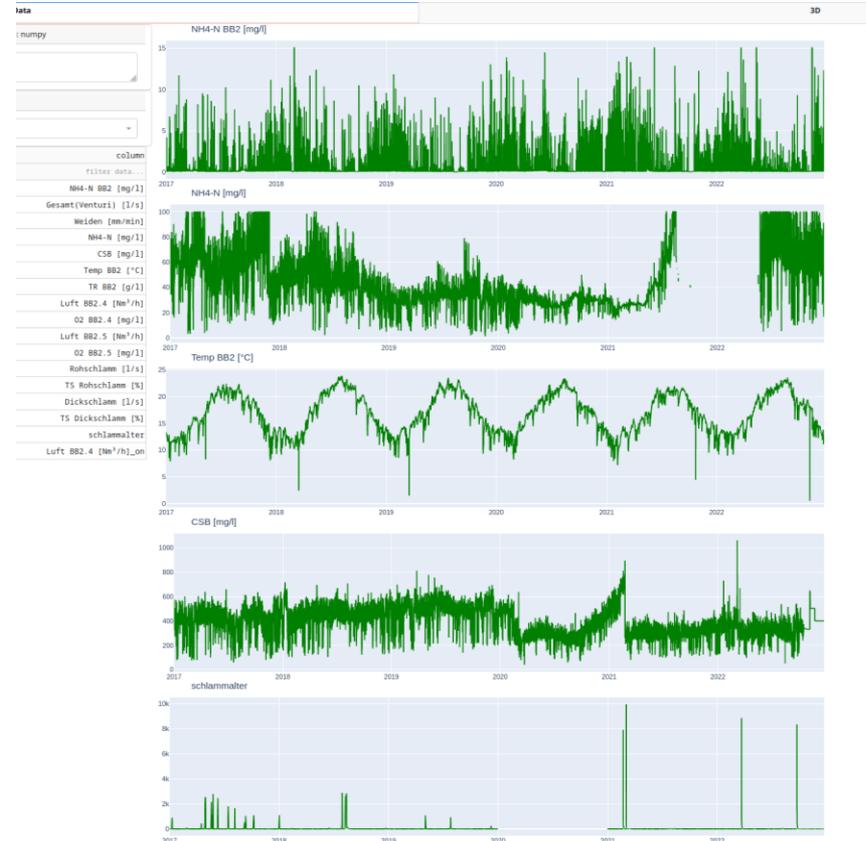
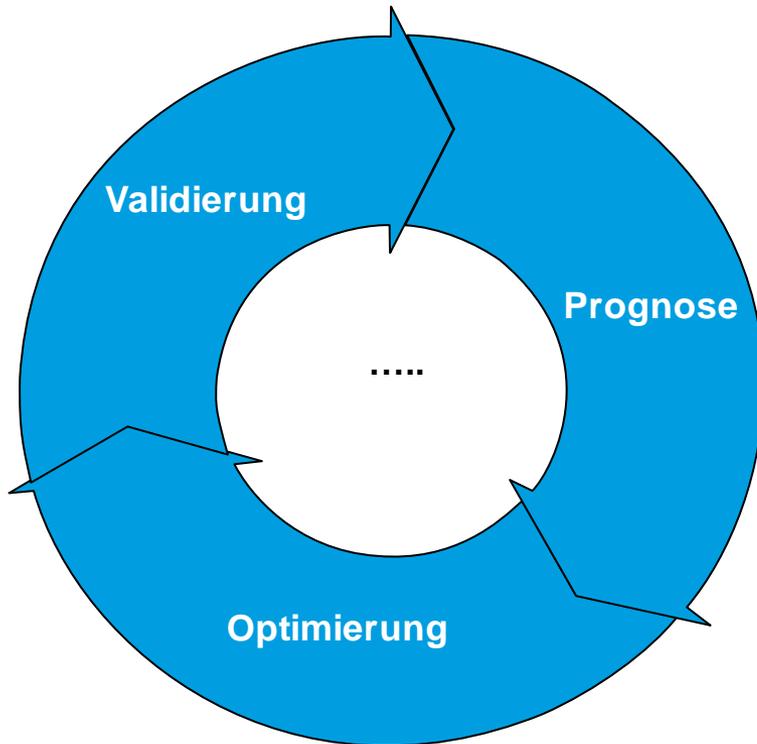
Aus Angst, Fehler nicht rechtzeitig zu entdecken, werden etwa bei Ablaufwerten großzügige Sicherheitsmargen eingerechnet.



1. Führt zu unnötig hohem Energieeinsatz
2. Damit einhergehend hoher CO2-Ausstoß und hohe Kosten.

4. Und jetzt? Womit fange ich an?

Validierung ist nicht immer der erste Schritt...



Die Datenvalidierung ist die – neudeutsch – „No-Regret-Maßnahme“. Im Sinne des Regelkreises bekommt man ein realistisches Gefühl für Digitalisierungsprojekte, vermeidet Frustrationen und Verzögerungen, gewinnt Unterstützung bei und mit den Anwendern und erzielt erste sehr schnelle Erfolge!

Anstatt eines Schlussworts

**Valide Daten sind nicht alles, aber ohne
valide Daten ist Alles Nichts.**



Prof. Dr. Mark Oelmann

- Professur für Wasser- und Energieökonomik (Hochschule Ruhr West (HRW))
- Studiengangsleiter Energie- und Wassermanagement (BWL-Bachelorstudiengang)
- Sprecher Forschungsschwerpunkt „Wasserökonomik und Wasserwirtschaft“
- Geschäftsführender Gesellschafter MOcons GmbH & Co. KG

Tel.: 02 08 / 88 254 - 358

E-Mail: mark.oelmann@hs-ruhrwest.de

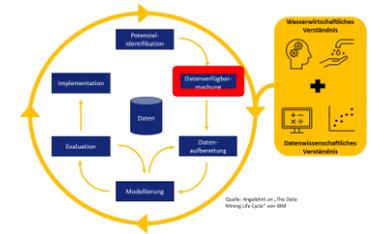
MOcons GmbH & Co. KG · Brandenburg 30 · 45478 Mülheim an der Ruhr
www.mocons.de

3.3 Das Zünglein an der Waage: Das Thema der Datenplausibilität

Datenaufbereitung

Die Nutzung von Daten in Entscheidungsprozessen ist bereits sehr hoch: Z. B. ...

- ... in der Automatisierungstechnik
- ... in datengetriebenen Analyseanwendungen
- ... im strategischen und operativen Monitoring
- ... usw.



Doch sind diese Daten auch ausreichend plausibilisiert und erfüllen die Kriterien der...

1. Nachvollziehbarkeit



2. Richtigkeit



3. Vollständigkeit



Blick in die Praxis zeigt,
Kriterien werden
oftmals nicht erfüllt!

Warum werden Kriterien so oft nicht erfüllt?

a) Menge und Umfang
der Daten zu groß

b) Echt-Zeit-Anwendbarkeit
der Daten (Zeitdruck)

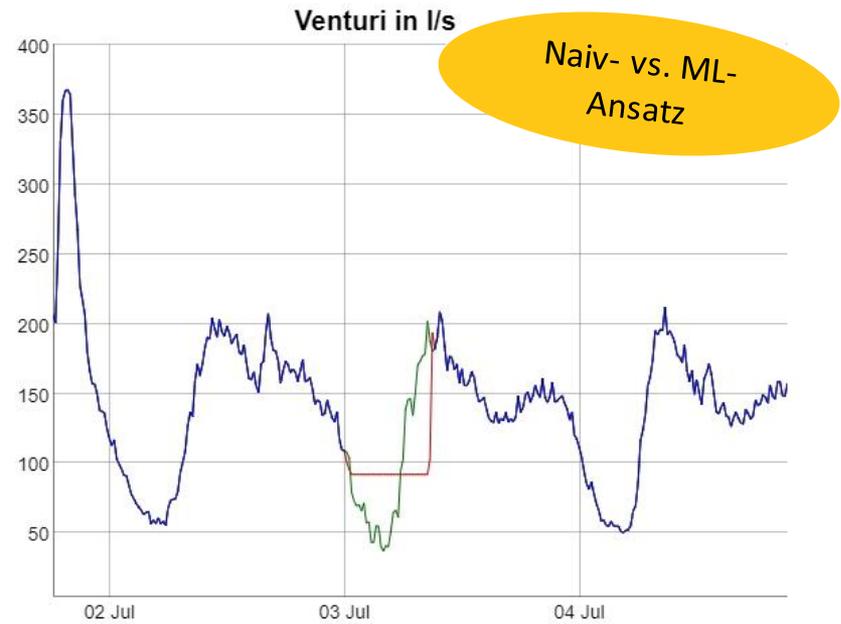
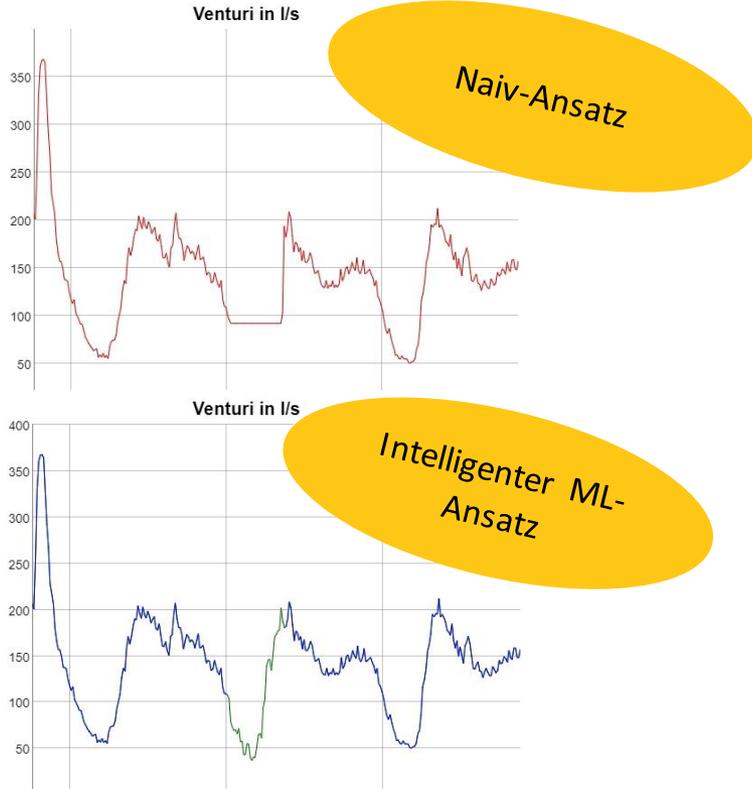
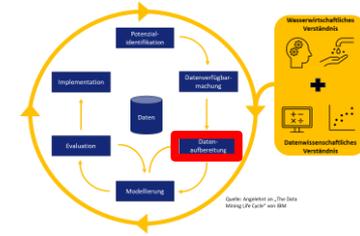
c) Unterschiedlichkeit der
Daten

Doch zunächst: Was ist unter einer nicht verlässlichen Datenbasis zu verstehen?

3.3 Das Zünglein an der Waage: Das Thema der Datenplausibilität

Datenaufbereitung

Zwar existieren in der Automatisierungstechnik Algorithmen, die zur Datenbereinigung eingesetzt werden, diese offenbaren allerdings zum Teil gravierende Mängel...



Bestehende Algorithmen sind oftmals zu naiv (für starke Heterogenität der Daten) und manuelle Plausibilisierung ist zu aufwendig (infolge der Echt-Zeit-Anwendbarkeit und Menge der Daten).

4. Empfehlungen

Was lässt sich für ein wasserwirtschaftliches Unternehmen ableiten?



Daten sind (nahezu) nie vollständig und nie komplett richtig. Entsprechend braucht man immer mehr Zeit, als dies zu Anfang zu erwarten war. Fangen Sie noch früher an als Sie es ohnehin schon vorhatten!



Lernen Sie über die Auseinandersetzung mit Daten, was machbar ist oder auch nicht; was sinnvoll ist oder auch nicht. Entwickeln Sie auf diesen ersten Erkenntnissen eine unternehmensweite Strategie, die mit den Unternehmenszielen im Einklang steht!



Die Wissensverschneidung von „Wasserwirtschaftlern“ und „Data Scientisten“ innerhalb des gesamten Projektzyklus ist essentiell, um Projekte erfolgreich umsetzen zu können. Überlassen Sie ein Projekt nicht nur den „Data Scientisten“. Entwickeln Sie Algorithmen sukzessive und erwarten Sie nicht die „Weltformel“!



ML-Projekte wachsen nicht im luftleeren Raum. Oftmals sind Vorarbeiten mit Blickrichtung auch auf IT-Infrastruktur notwendig, die Sie nicht unterschätzen sollten!