

# Data Science in der Instandhaltung

Exemplarisches Vorgehen entlang des Modells CRISP-DM

Harald Piringer & Robin Kühnast-Benedikt & Tobias Gerstmaier & Hubert Biedermann

Geschäftsführer Visplore GmbH & Geschäftsführer ÖVIA & Leader Operational Excellence  
Maintenance INNIO Jenbacher & Präsident ÖVIA

## Management Summary:

Von ABC-Analysen bis hin zu KI-basierten Handlungsempfehlungen stehen Unternehmen vor der Herausforderung datenanalytische Anwendungen unabhängig vom Reifegrad (deskriptiv bis präskriptiv) mit einer hohen Erfolgsquote zu implementieren. Der vorliegende Beitrag zeigt in Form eines Anwendungsfalls, wie das exemplarische Vorgehen entlang eines Datenanalytik-Standards (CRISP-DM Modell<sup>1</sup>) funktionieren kann. Hintergrund ist der Bakterienbefall des Kühlschmierstoffes einer Zentralanlage zur Versorgung von CNC-Bearbeitungszentren, der hohe Wartungs- bzw. drohende Ausfallkosten bedingt. Aufgrund des Risikos wurde zusätzlich Sensorik (Messung pH-Wert und Leitfähigkeit) installiert, um den Zustand besser feststellen zu können. Der Anwendungsfall behandelt ein in der Praxis oft vorherrschendes Problem: es werden die richtigen Daten gemessen – jedoch fehlen Werkzeuge um auch im Fall von einer geringen Zahl an aufgetretenen Störungen sinnvolle Ergebnisse zu erzielen. Der dargestellte Lösungsansatz in diesem Beitrag nutzt die unbefriedigende Ausgangssituation und wandelt das Problem in einen Vorteil um: für die überschaubare Zahl von sechs Störmeldungen lassen sich diese Daten sehr genau untersuchen. Es wurde daher nicht nur nach einfachen Zusammenhängen (direkte Korrelationen) sondern vermehrt nach Mustern (Kausalitäten) gesucht. Es konnte gezeigt werden, dass stets ein sinkender pH-Wert und die steigende Leitfähigkeit die Ausfälle ankündigen. Aufgrund der gewonnenen Erkenntnisse konnten Maßnahmen gesetzt werden, die das Wartungsintervall von 19 auf 49 Tage erweiterten.

---

<sup>1</sup> Siehe Chapman, P. et al. (1999)

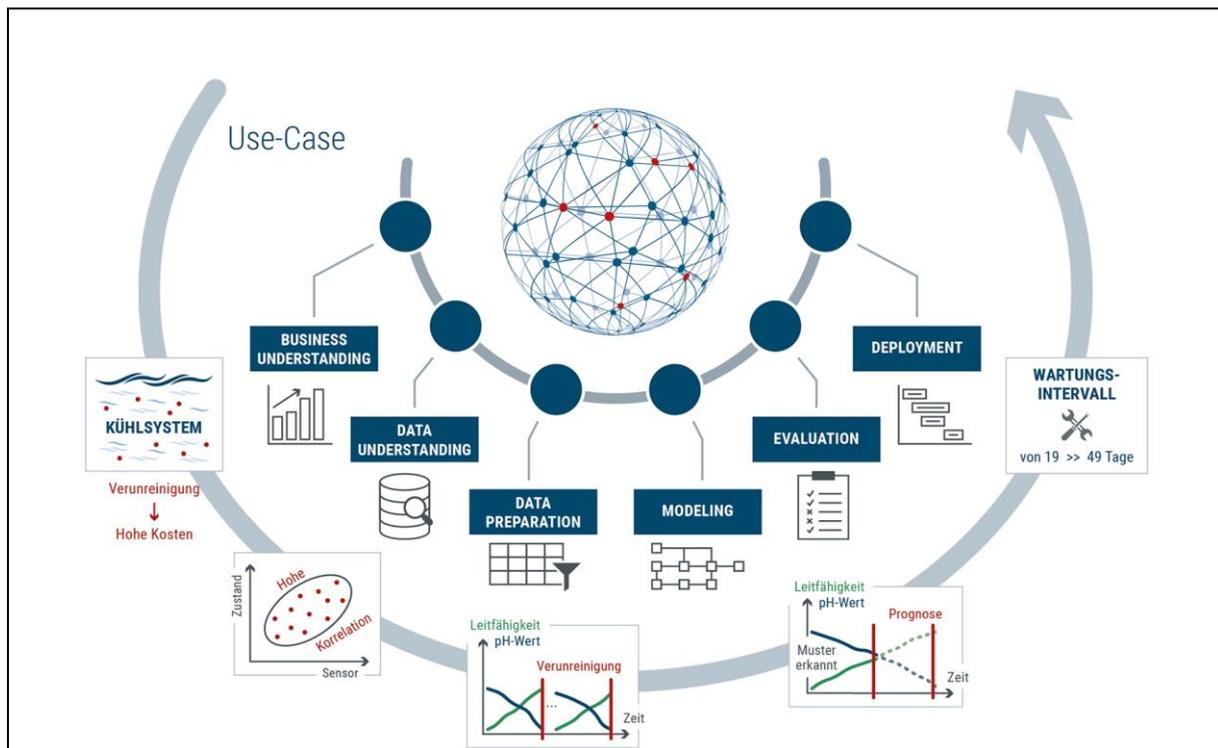


Abbildung 1: Graphical Abstract eines exemplarischen Vorgehens entlang des Datenanalytik-Standards.

## Einleitung

Für produzierende Betriebe äußert sich die Digitalisierung nicht nur als Chance, sondern gleichzeitig auch als Gefahr. Gelingt es nicht, die entstehenden Daten sinnvoll für einen Erkenntnisgewinn zu nutzen, droht aufgrund der hohen Komplexität eine Ohnmacht, die vor allem bei den involvierten Personen Frustration auslöst.

Viele Unternehmen begehen dabei den Fehler, die Erwartungshaltung zu hoch anzusetzen und beispielsweise von Beginn an Prognosen oder Simulationen durchführen zu wollen. Sie übersehen dabei, dass die unteren Reifegrade von Analysen (beschreibend oder diagnostizierend) einen ähnlichen Nutzen wie Prognosen oder präskriptive Analysen darstellen und diese Schritte nur in den seltensten Fällen übersprungen werden können. Der vorliegende Anwendungsfall soll exemplarisch das gesamte Vorgehen zeigen, wie Unternehmen bereits mit deskriptiven Analysen einen Mehrwert erzeugen können, der schließlich auch für eine Vorausschau nutzbar wird (s. Abbildung 1).

### Phase 1: Business Understanding

CRISP-DM ist ein Standardvorgehen für Datenanalytik-Projekte. In der ersten Phase gilt es ein Verständnis für den Geschäftsfall und das vorliegende Geschäftsmodell zu erreichen.

Der erste Schritt eines produzierenden Unternehmens in eine digitale Zukunft besteht darin, die wesentlichen Geschäftsprozesse, Objekte und weiteren Aspekte des Betriebs messbar zu machen. Zusätzlich potenziert sich der Nutzen, wenn es gelingt, die digital erfassten Dinge auch noch miteinander zu vernetzen.

### Anwendungsfall: Wiederholte Instandhaltung eines Kühlsystems

Für die Fertigung von Gasmotorkomponenten wie Kurbelgehäusen, Kurbelwellen, Zylinderköpfen und Pleuel betreibt INNIO Jenbacher eine Kühlschmierstoff-Zentralanlage, die rund 70 angeschlossene CNC-Fräs-, Dreh- und Schleifmaschinen mit Emulsionsgemischen zur Kühlung und Schmierung versorgt. Diese Anlage zur Speicherung, Verteilung und Filterung von rund 150 m<sup>3</sup> Kühlschmierstoff („KSS“) besteht aus Tanks, Filtern,

Abscheidekomponenten, Rohrleitungen, Pumpen, Ventilen und der elektrischen und pneumatischen Ansteuerung.

Die eingesetzte Zentralanlage wurde ursprünglich für die Bearbeitung von Stahlkomponenten ausgelegt. Im Zuge von Änderungen an der Produktpalette wurde das Rohmaterial mehr und mehr von Stahl auf Guss umgestellt und der Maschinenpark vergrößert. Bei der Gussbearbeitung entstehen neben groben Spänen sehr feine Graphit- und Eisenpartikel. Diese können sich mit Schmier- und Hydraulikölen, die durch kleine Leckagen in den KSS gelangen, verbinden, zäh klebrige Schichten bilden und das Wachstum von Bakterien und Pilzen fördern. Dieses stellt eine Gesundheits- und Qualitätsbelastung dar und muss folglich kontrolliert und vermieden werden.

Anfang 2021 wird die bestehende KSS-Zentralanlage um eine Aufbereitungs- und Filteranlage („KSS-Pflegeeinheit“) speziell für Gusspartikel erweitert, um durch die Vermeidung von periodisch notwendigen Komplettentsorgungen eine Reduktion von gefährlichen Abfällen gemäß ÖNORM S 2100, eine Verbesserung der KSS-Qualität sowie eine Erhöhung der Betriebssicherheit und Wartungsfreundlichkeit zu erreichen. Seit Juli 2019 misst INNIO Jenbacher minütlich Temperatur, Leitfähigkeit und pH-Wert des umgewälzten Kühl- und Schmierstoffgemischs, um durch datenbasierte Entscheidungen bereits vor der Nachrüstung der KSS-Pflegeeinheit eine Erhöhung der Prozessstabilität zu erreichen.

## Verallgemeinerung des Anwendungsfalls

Dieser konkrete Anwendungsfall spiegelt einige allgemein auftretende Herausforderungen von Prozessdaten wider; insbesondere im Zusammenhang mit Fehleranalyse bei Instandhaltungsthemen. Einige dieser Herausforderungen sind zusammengefasst wie folgt:

1. **Datengröße:** Sensorwerte ändern sich vielfach rasch. Um relevante Muster zu erhalten, müssen die Daten daher meist in recht feinem zeitlichen Raster analysiert werden. Im Anwendungsbeispiel liegen die Daten im Minutenraster über ca. 7 Monate vor, was ca. 300.000 Messwerten pro Sensor entspricht. Für viele Anwendungen sind sogar Sekundenraster üblich, sodass bei Langzeitbetrachtungen rasch Datengrößen von Millionen Werten pro Sensor auftreten. Eine Analyse solcher Datenmengen stellt viele Visualisierungs- aber auch Kalkulationsprogramme - beispielsweise Excel - vor eine große Herausforderung.
2. **Komplexe Struktur:** Sensorzeitreihen folgen vielfach keinen klassischen statistischen Verteilungen, sondern sind geprägt von Sprüngen und Oszillationen sowie Bereichen mit kurz- und langfristigen Trends. Vielfach ändern sich solche Strukturen häufig und plötzlich (z.B. durch Prozessumstellungen), sodass es oft schwer bis unmöglich ist, die Struktur umfassend mathematisch zu beschreiben. Das macht den Einsatz vieler analytischer Methoden zur Herausforderung.
3. **Unsauberkeiten:** Die meisten Sensorzeitreihen haben nicht durchgängig plausible Werte. Auf Grund von Sensorausfällen, Sensordrifts oder auch aus prozesstechnischen Gründen haben viele Sensorzeitreihen phasenweise gar keine oder nur interpolierte Daten, Drifts sowie punktuelle Ausreißer. Für nachgelagerte Analysen ist es vielfach nötig, solche Datenprobleme soweit möglich zu bereinigen, was zeitaufwändig sein kann und typischerweise Domain-Expertise erfordert.
4. **Wenige Störfälle:** Die Signifikanz von Zusammenhängen nimmt im Allgemeinen mit der Anzahl an Beobachtungen zu. Der Datenanalytiker benötigt damit genau das, was die Instandhaltung in ihrer Tätigkeit zu verhindern versucht. In vielen Organisationen sind damit die Störfälle rar - im konkreten Beispiel sind es nur 6. Im Extremfall treten

zu erklärende Fehler ein einziges Mal auf oder sind noch gar nie eingetreten, könnten dies aber in Zukunft tun. Eine Konsequenz ist, dass bei der Beurteilung möglicher Zusammenhänge und Erklärungen mangels statistischer Verallgemeinerbarkeit der Expertise und Einschätzung von Prozessexperten eine noch größere Bedeutung zukommt.

5. **Erklärbarkeit unsicher:** Ob konkrete Fehlerzustände sich durch Prozessdaten erklären lassen, hängt maßgeblich davon ab, ob die Ursachen durch geeignete Sensoren auch erfasst werden. Ursachen, die außerhalb der vorhandenen Datenbasis liegen, werden durch diese naturgemäß auch nicht erklärbar sein. Ein mögliches Ergebnis einer Analyse kann somit sein, dass relevante Sensoren bislang fehlen und nachgerüstet werden sollten.
6. **Herausfordernde Modellierung:** Obgleich die Prozessdaten eine Erklärung für Fehlerzustände beinhalten, kann es herausfordernd sein, diese geeignet zu modellieren. Viele statistische Methoden zur Klassifikation bzw. Regression sind nicht geeignet, um direkt auf rohe Sensorzeitreihen angewandt zu werden, da sie nur einzelne Zeitpunkte aber keine zeitlichen Muster erfassen. Daher ist es oft nötig, zusätzliche Merkmale zu rechnen, beispielsweise in Form von gleitenden Mittelwerten, Ableitungen, zeitversetzten Datenreihen oder lokalen Frequenzspektren. Der Möglichkeitsraum dafür ist extrem groß, sodass eine qualitative Beurteilung der Daten mit entsprechendem Prozesswissen sehr hilfreich ist, um geeignete Transformationen rasch zu identifizieren.

Eine zusätzliche Herausforderung kann die große Zahl an Sensorzeitreihen sein. In vielen Anwendungsbereichen - etwa dem Facility Management oder bei komplexen Produktionsanlagen - kann es nötig sein, hunderte bis (zig-)tausende Sensorzeitreihen zu berücksichtigen. Das ist im gegenständlichen Beispiel anders, da die Datenbasis hier lediglich aus drei Zeitreihen besteht.

## Zielsetzung

Ziel des Pilotprojekts aus Sicht des Anlagenbetreibers ist eine frühzeitige Erkennung von Bakterien- und Pilzwachstum anhand von Messdaten, noch bevor eine Geruchsbelästigung auftritt. Hierdurch soll ein frühes Gegensteuern, beispielsweise durch eine Konzentrationserhöhung einzelner KSS-Bestandteile, ermöglicht und ein hochdosierter Einsatz von Bioziden seltener notwendig werden.

Ein wesentliches Ziel liegt somit darin, vorausschauende Frühwarnungen zu ermöglichen. Der Weg dorthin führt allerdings über eine hinreichend tiefgehende deskriptive Analyse der historischen Datenbasis. Auf diesem oft vernachlässigten Schritt liegt der Fokus dieses Textes.

## Phase 2: Data Understanding

Der erste Schritt eines Datenanalytik-Projekts besteht in der Sichtung der Daten. Es ist unbedingt erforderlich, durch systematisches Vorgehen ein Datenverständnis zu erreichen, da alle weiteren Schritte auf dieser Basis aufbauen.

Im Folgenden wird beschrieben, wie der konkrete Anwendungsfall von INNIO Jenbacher mit Hilfe der Analysesoftware Visplore<sup>2</sup> gelöst wurde. Zusammengefasst bestand der Ansatz darin, die Zeitbereiche vor den Bakterienbefällen dahingehend zu überprüfen, ob sie Muster aufweisen, die jeweils nur vor den Bakterienbefällen auftreten. Dies erfolgte zunächst qualitativ durch visuellen Vergleich der Prozesszeitreihen und wurde danach quantitativ validiert.

---

<sup>2</sup> Visplore GmbH - [www.visplore.com](http://www.visplore.com) (2020)

## Import und erste Datensichtung

Die Prozessdaten wurden als CSV Datei in Visplore geladen. Daraufhin können die Daten dank vorkonfigurierter Analysewerkzeuge sofort auf vielfältige Art visualisiert und analysiert werden (s. Abbildung 2).

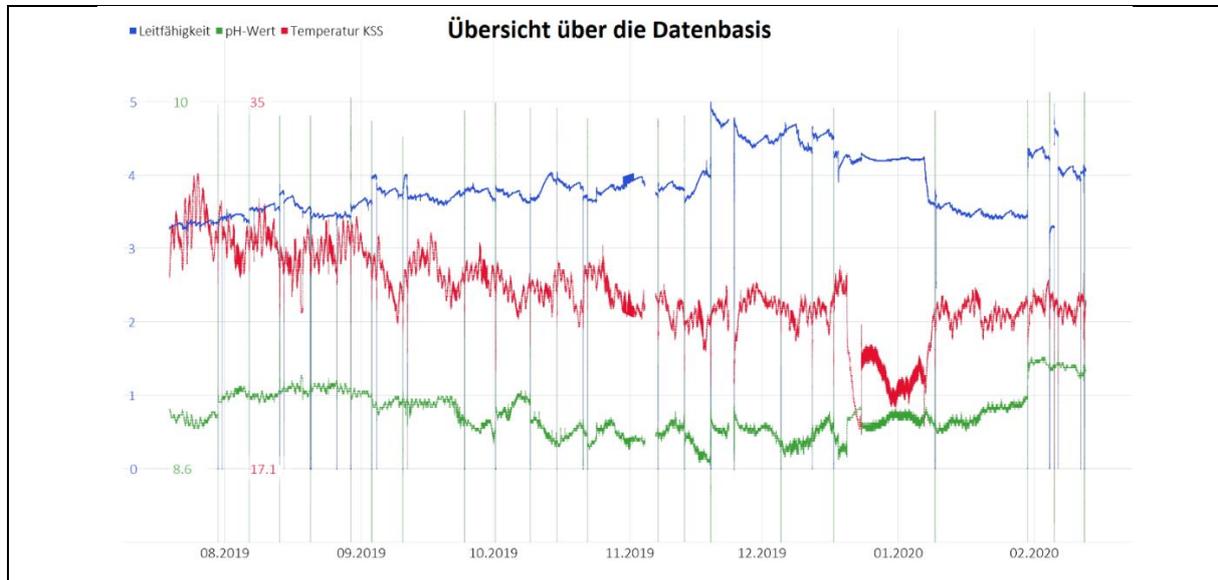


Abbildung 2: Darstellung der drei Messgrößen über den gesamten Beobachtungszeitraum.

Die Temperatur zeigt einen generell abnehmenden Trend über den Beobachtungszeitraum mit einem markanten Absinken rund um den Jahreswechsel und speziell in der Sommerzeit erkennbaren Tagesgängen. Bei näherer Betrachtung kommen wiederholt Zeitbereiche mit raschen Oszillationen von ca. einem Grad Celsius vor, die plötzlich beginnen und enden, gefolgt von deutlich glatteren Verläufen (s. Abbildung 3).

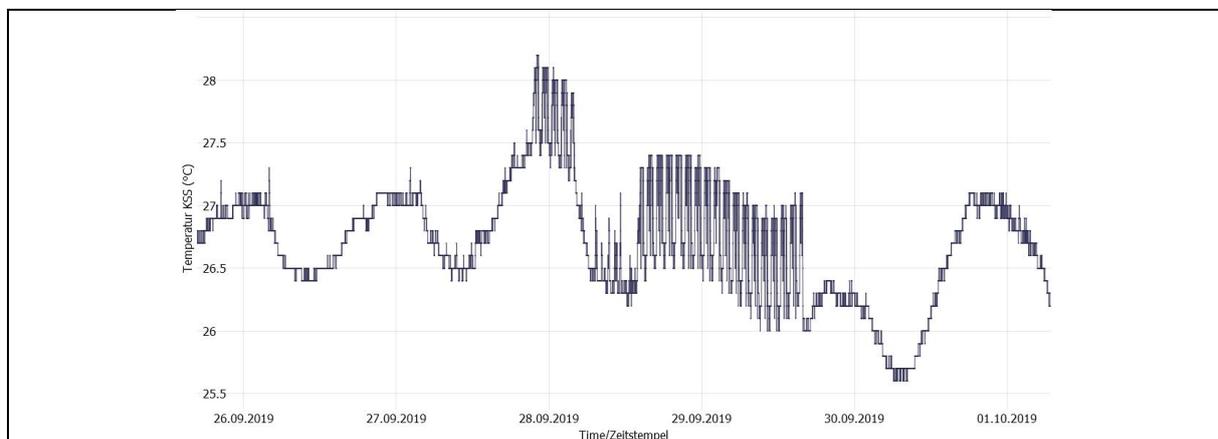


Abbildung 3: Oszillationen der Temperatur.

Bei der Leitfähigkeit ist bis Mitte November ein steigender Trend zu erkennen, gefolgt von einem Sprung und in weiterer Folge absinkenden Trend bis zu einem neuerlichen Sprung Ende Jänner. Der pH-Wert weist dieselben Sprünge mit entgegengesetzten Trends auf. Auch beim pH-Wert kommt es wiederholt zu Zeiten mit raschen Oszillationen.

Die Sichtung der Daten zeigte außerdem zahlreiche ausgeprägte Zacken. Insbesondere sinkt die Leitfähigkeit wiederholt kurzfristig auf Null und der pH-Wert hat wiederholt etwa 15-minütige Ausschläge herab bis auf den Wert 7 gefolgt von Sprüngen auf den Wert 10. Starkes Zooming auf einzelne solcher Ereignisse illustriert gut, dass es sich hierbei in vielen Fällen um

Eingriffe handelt, nach denen sich oftmals die Lage von Prozessgrößen verändert (s. Abbildung 4).

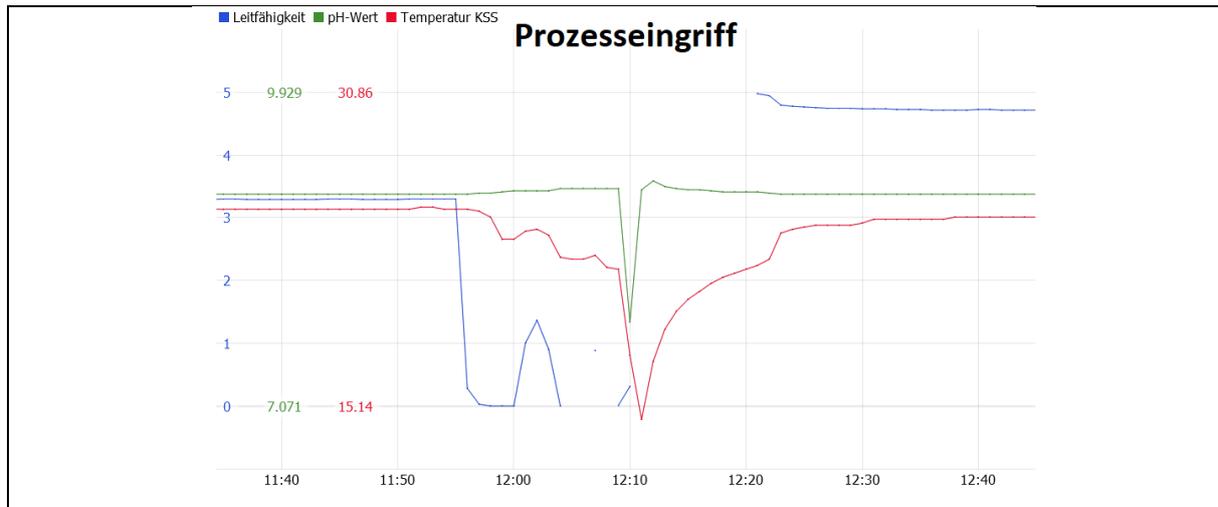


Abbildung 4: Detailansicht eines Eingriffs in den Prozess.

### Phase 3: Data Preparation

Ausgehend von den Ergebnissen aus der Data-Understanding-Phase gilt es die Daten so zu präparieren, dass die in Frage kommenden, datenanalytischen Methoden einsetzbar sind. Erfahrungsgemäß wird in dieser Phase der meiste Aufwand betrieben, da es sich vielfach um große, heterogene Datenmengen handelt, die noch verdichtet, integriert, verbunden oder anders transformiert werden müssen.

### Definition der Vergleichsperioden

Insgesamt kam es zu sechs Bakterienbefällen, wie aus einer Excel-Datei ersichtlich ist, in der pro Tag einmal erhobene Werte und Maßnahmen festgehalten sind. Der zweite Schritt bestand darin, die Tage mit Bakterienbefällen sowie jeweils zwei Tage davor als Befallsperioden in Visplore zu markieren. Die Wahl von dreitägigen Zeitbereichen war laut Domänenexperte eine plausible Schätzung. Sie stellt einen Kompromiss dar, um einerseits für zukünftige Befälle eine gewisse Vorwarnzeit zu berücksichtigen und gleichzeitig eine Abschwächung möglicher Anzeichen durch eine Wahl zu langer Zeitbereiche zu vermeiden. Die Länge kann jedoch in Visplore auch rasch geändert und somit effizient experimentell ermittelt werden.

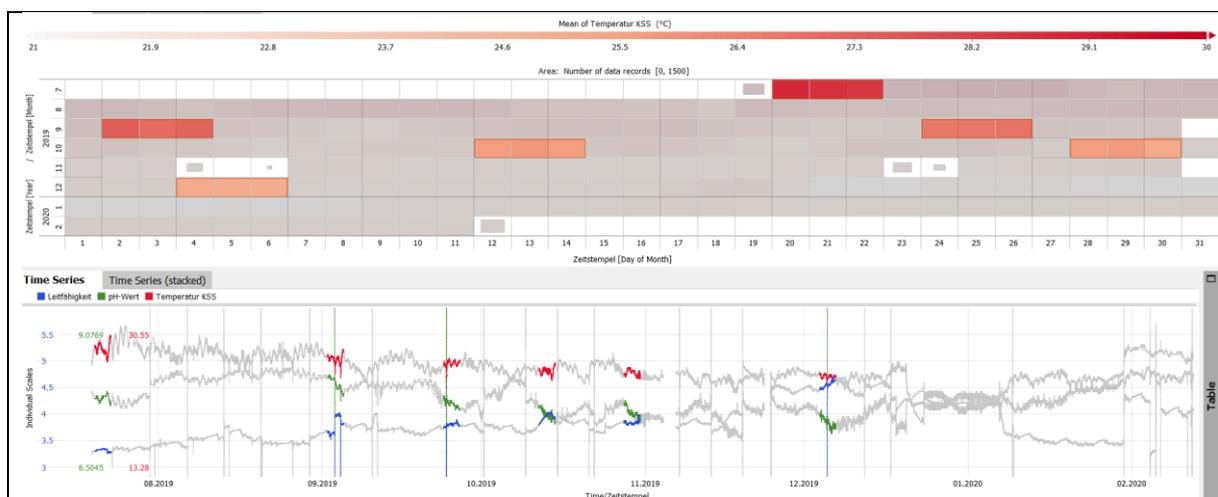


Abbildung 5: Markieren der Befallsperioden in einer Kalenderdarstellung (oben) hebt diese Zeiträume in einer Zeitreihendarstellung hervor (unten).

Zur Definition wurden die betreffenden Tage in einer Kalenderdarstellung interaktiv markiert und diese Auswahl als "Befallsperioden" zur weiteren Verwendung benannt. Dadurch werden diese Zeitbereiche in der Zeitreihendarstellung der drei Messgrößen hervorgehoben und erlauben somit einen ersten qualitativen Vergleich (s. Abbildung 5). Ebenso kann diese Annotation der Daten später für einen quantitativen Vergleich von Befallsperioden herangezogen werden.

## Phase 4: Modelling

In der Modellierungsphase findet die ausgewählte Methode ihren Einsatz. Bei datenanalytischen Anwendungen höherer Reifegrade können komplexe Algorithmen wie Neuronale Netze, Support Vector Maschinen oder Entscheidungsbäume implementiert werden. Die erzielbaren Ergebnisse werden oftmals höher bewertet, als jene von deskriptiven Analysen. In der Praxis ist zumeist das Aufwand-Nutzen-Verhältnis bei beschreibenden Analysen deutlich besser.

## Qualitativer Vergleich

Ziel des qualitativen Vergleichs ist es, durch Betrachten der Muster während der Befallsperioden Hypothesen über mögliche Zusammenhänge mit den Bakterienbefällen zu entwickeln. Ein wesentlicher Vorteil der Visualisierung ist dabei, dass weitgehend vermieden werden kann, vorab Annahmen über die Daten treffen zu müssen.

In Visplore können Visualisierungen auf vielfältige Weise angepasst werden, was die Aussagekraft stark erhöht. Konkret wurde durch Anpassen der Y-Skalierungen der Fokus der Darstellung auf den Wertebereich abseits der Ausreißer gelegt. Die graphische Hervorhebung der Befallsperioden unterstützt die Wahrnehmung von Gemeinsamkeiten deutlich. Dadurch war auffällig, dass der pH-Wert in allen Befallsperioden einen negativen Trend aufwies; eine erste wesentliche Erkenntnis.

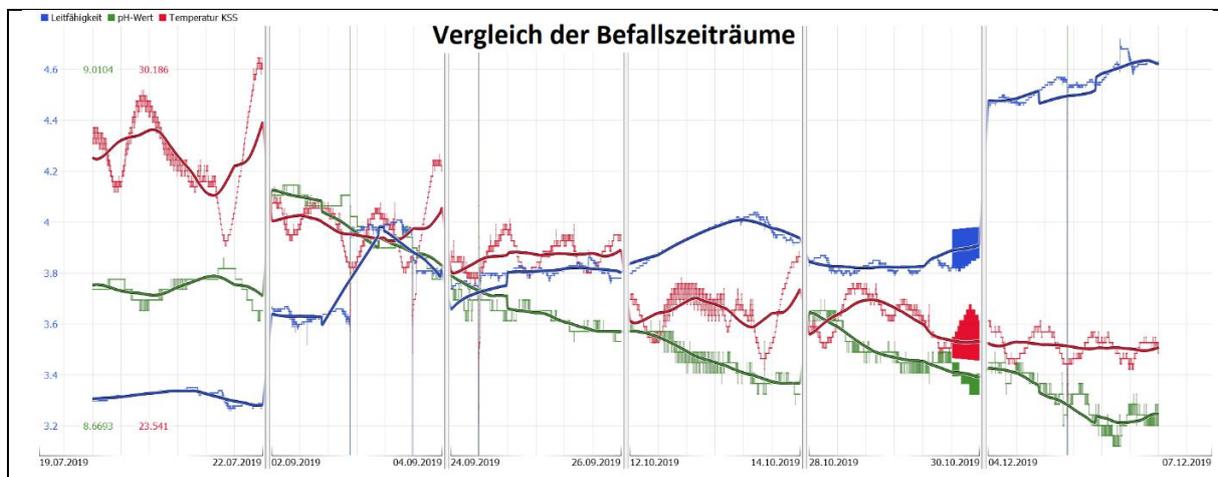


Abbildung 6: Bei einem direkten Vergleich der Befallsperioden zeigen gleitende Mittelwerte des pH-Werts (grün) lokal absinkende Trends und jene der Leitfähigkeit (blau) tendenziell leicht ansteigende.

Um die Befallsperioden noch genauer zu vergleichen, als das im Kontext des gesamten Betrachtungszeitraums möglich ist, wurden die annotierten Zeitbereiche gefiltert und die Zeit dazwischen ausgeblendet (s. Abbildung 6). Zusätzlich unterstützt ein in dieser Grafik eingblendeter gleitender Mittelwert die Wahrnehmung von Trends. Dadurch sind die jeweils sinkenden Trends beim pH-Wert gut zu erkennen. Bei der Leitfähigkeit zeigt der gleitende Mittelwert in einigen Fällen einen schwachen lokalen Aufwärtstrend. Bei der Temperatur ließ sich kein gemeinsames Muster identifizieren. Ergebnis der qualitativen Begutachtung war somit die Hypothese, dass einem Bakterienbefall eine Zeit mit tendenziell sinkenden pH-Wert sowie gleichzeitig leicht ansteigender Leitfähigkeit vorausgeht.

## Phase 5: Evaluation

Die erzielten Ergebnisse müssen evaluiert werden. Für den eingesetzten Algorithmus sind Zielparameter zu definieren, anhand derer eine Bewertung möglich ist. Es werden daher die Ergebnisse von Phase 4 validiert.

Zur Validierung dieser Hypothese wurde zunächst versucht, den Zusammenhang zwischen gleichzeitig auftretendem sinkenden pH-Wert und steigender Leitfähigkeit einerseits und dem Zustand "Befallsperiode" andererseits quantitativ zu erfassen. Zu diesem Zweck wurden direkt in Visplore Zeitreihen berechnet, welche die Stärke des lokalen Trends ausdrücken. Konkret wurden vom pH-Wert und der Leitfähigkeit die geglätteten ersten Ableitungen der gleitenden Mittelwerte mit einer Fenstergröße von 36h errechnet, wozu Visplore hinterlegte Funktionen bereitstellt. Diese Zeitreihen geben unabhängig von der Lage die Tendenz zu einem lokal ansteigenden oder abfallenden Trend an. Abbildung 7 zeigt die beiden errechneten Zeitreihen und hebt die Befallsperioden hervor.

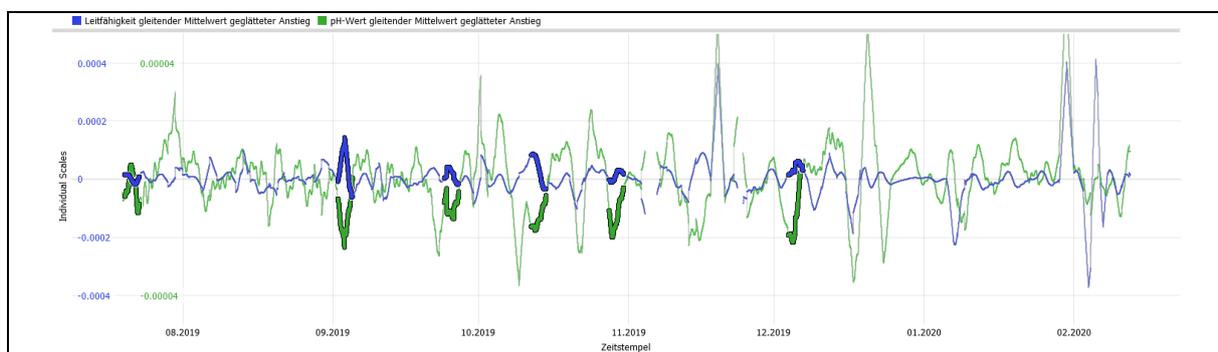


Abbildung 7: Die Befallsperioden in errechneten Zeitreihen, die Trends zu Anstieg bzw. Absinken zeigen.

Abgesehen vom ersten Befall weisen alle Befallsperioden gleichzeitig eine mehr oder weniger stark ausgeprägte lokale positive Spitze bei Leitfähigkeit sowie eine negative Spitze beim pH-Wert auf. Durch Rücksprache wurde später klar, dass zum Zeitraum des ersten Befalls die aufgezeichneten Daten nicht als repräsentativ angesehen werden können, da es danach, jedoch vor dem zweiten Befall, zu einer wesentlichen Umstellung kam.

Zur weiteren Validierung wurden die Verteilungen der errechneten Zeitreihen für die restlichen fünf Befallsperioden mit dem gesamten übrigen Zeitraum verglichen. Hier zeigt sich in Histogramm-Darstellungen für den pH-Wert eine stark nach links verschobene Verteilung, was ebenfalls einem lokalen Absinken entspricht, sowie – etwas weniger stark ausgeprägt – eine nach rechts verschobene Verteilung für die Leitfähigkeit (s. Abbildung 8). Dadurch wird die Signifikanz des Zusammenhangs erhärtet.

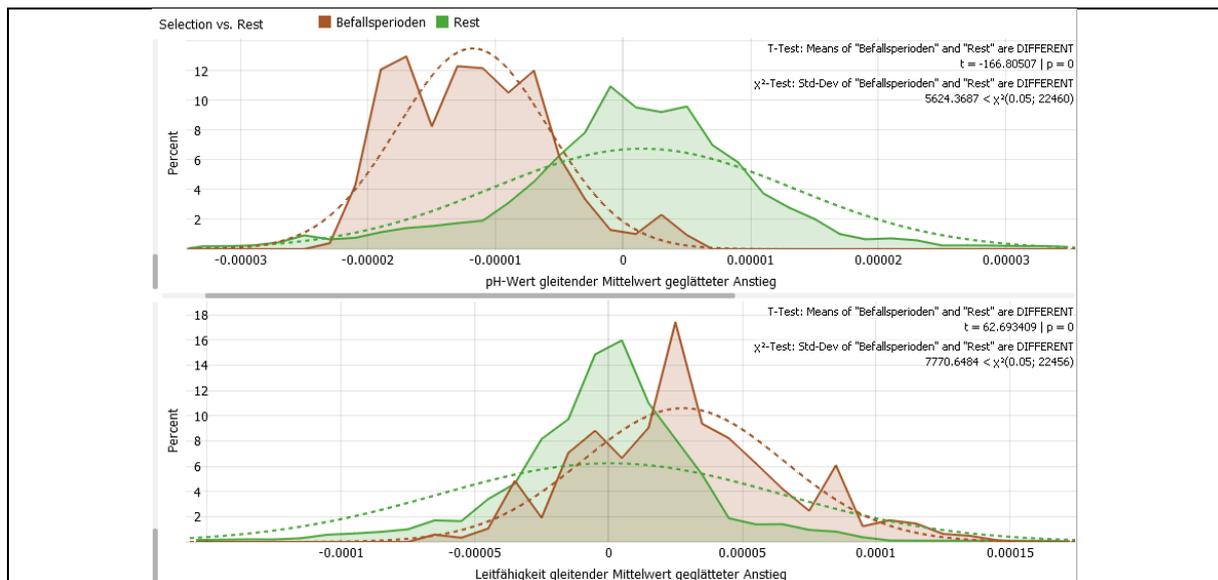


Abbildung 8: Vergleich der Werte-Verteilungen für Zeiten der Befallsperioden (braun) mit den übrigen Zeiten (grün).

Abschließend wurde eine andere Art der Validierung vorgenommen: Die letzte Befallsperiode von 4.12. bis 6.12. wurde als Referenzmuster herangezogen, das über den gesamten Zeitraum durch eine algorithmische Mustersuche gesucht wurde, wie sie mit Visplore ebenfalls möglich ist. Dadurch sollte ermittelt werden, inwiefern ein Auftreten der Muster auch abseits von Bakterienbefällen auftritt. Das Ergebnis illustriert Abbildung 9. Bei allen Befällen bis auf jenem im Juli fallen die durch die Mustersuche gefundenen Zeiträume mit den Befallsperioden entweder zusammen oder gehen diesen wenige Tage voraus. Das legt nahe, dass Befälle gegebenenfalls sogar noch früher erkannt werden können. Umgekehrt wurden das Muster zwei Mal abseits von Befallsperioden erkannt, konkret von 5.8. bis 7.8. sowie von 13.11. bis 16.11. Dieser zweite Zeitraum liegt jedoch knapp vor einem auffälligen Sprung und ist daher gegebenenfalls ebenfalls von Interesse. Nach der letzten Befallsperiode wurde das Muster nicht mehr gefunden.

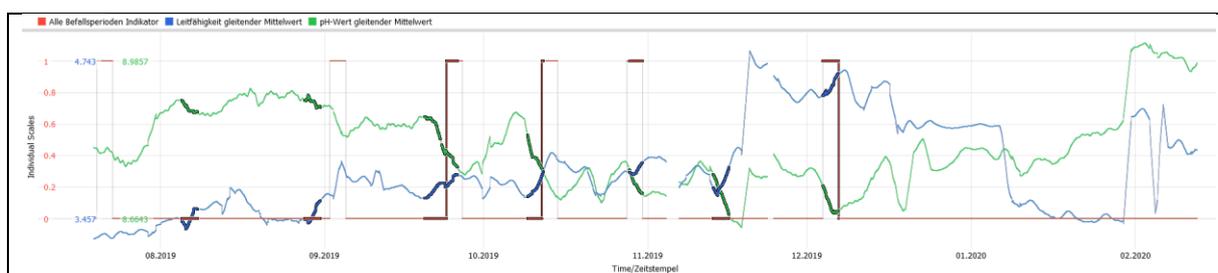


Abbildung 9: Vergleich des Ergebnisses einer automatischen Mustersuche (Hervorhebung) mit den tatsächlichen Befallsperioden (Indikatorzeitreihe).

## Phase 6: Implementierung

In der letzten Phase muss das Ergebnis in den produzierenden Betrieb eingesetzt werden. Dazu muss definiert werden, welche Entscheidung durch das Ergebnis beeinflusst werden kann. Zusätzlich gilt es die damit einhergehenden technischen Herausforderungen zu meistern.

## Ergebnis

Es wurde die Hypothese generiert, dass ein absinkender Trend des pH-Werts bei gleichzeitig ansteigendem Trend der Leitfähigkeit auf einen sich abzeichnenden Bakterienbefall hindeuten. Die Validierung legt nahe, dass es sich dabei um einen signifikanten Zusammenhang handelt. Es wurden nachvollziehbare Algorithmen identifiziert, die eine Basis für die Implementierung einer zukünftig automatischen Überprüfung darstellen. Weiters

können durch die Grafiken diese Ergebnisse und dadurch bedingte Änderungen anschaulich kommuniziert und argumentiert werden.

Abseits dieses Hauptergebnisses wurden durch die interaktive Visualisierung bislang unbekannte Oszillationen in den Messgrößen entdeckt, die ebenfalls einen Ansatzpunkt für eine Prozessverbesserung bieten. Schließlich konnte durch die flexible Darstellung ein guter Überblick über die genauen Abläufe bei einigen Eingriffen in den Prozess erzielt werden. Abschließend ist festzuhalten, dass der Zeitaufwand für die gesamte Analyse bei etwa einer Stunde lag. Speziell bis zur Hypothesengenerierung dauerte es etwa 10 Minuten, da sich alle nötigen Schritte mit wenig Aufwand direkt in Visplore abbilden ließen.

Die Messdaten werden seit März 2020 systematisch auf die Tendenzen „steigende Leitfähigkeit und gleichzeitig sinkender pH-Wert“ hin beobachtet und Gegenmaßnahmen frühzeitig eingeleitet. Im Zeitraum März bis September 2020 lagen durchschnittlich 49 Tage zwischen zwei Zeitpunkten, an denen ein hochdosierter Einsatz von Bioziden notwendig war. Im Vergleichszeitraum März bis September 2019, vor der Einrichtung der Messwertüberwachung, lag dieser Zeitraum bei durchschnittlich 19 Tagen. Es ist zu beachten, dass seit dem Jahreswechsel 2019/2020 auf einen mineralölbasierten Zusatz zum Kühlschmierstoff verzichtet wird und eine automatische Nachdosierung installiert wurde, die ein genaueres Einhalten der Zielkonzentration mit geringeren Abweichungen von den Sollwerten ermöglicht. Diese Maßnahmen dürften neben dem früheren Einleiten von Gegenmaßnahmen ebenfalls zur Erhöhung der Biostabilität beigetragen haben.

## Schlussfolgerung

Dieser Beitrag hat die Anwendung von CRISP-DM an Hand eines konkreten Problems aus der Instandhaltung unter Betonung interaktiver visueller Verfahren illustriert. Über die konkreten Erkenntnisse zum beschriebenen Problemfall lassen sich folgende Aspekte zusammenfassen, die auf andere datenbasierte Problemstellungen übertragbar sind:

- **CRISP-DM** ist ein zielführendes und vielfach bewährtes Vorgehensmodell, um datenbasierte Vorhaben in überschaubare Schritte zu strukturieren. Die konkrete Ausgestaltung einzelner Schritte muss den Daten- und Projektzielen entsprechend angepasst werden. Interaktive Visualisierung kann in vielen Schritten - insbesondere bei den besonders aufwändigen Phasen des Verstehens und Aufbereitens von Daten - einen sehr wichtigen Beitrag leisten.
- **Beschreibende sowie diagnostizierende Analysen** haben allgemein einen oft unterschätzten Nutzen. Sie sind nicht nur wichtige Schritte vor der Erstellung von prädiktiven Modellen, sondern generieren für sich genommen einen wertvollen Beitrag für die Gewinnung von Erkenntnissen über den Prozess. Intuitive grafische Analysen haben dabei den großen Vorteil, dass sie meist einfacher und rascher anzuwenden sind als eine operative Prädiktion und Prozessexperten ohne Data Science Ausbildung wesentlich in die Datennutzung einbinden.
- **Visualisierung** ist eine effektive Antwort auf einige praktische Herausforderungen bei der Nutzung von Prozessdaten. Die bestens ausgeprägte Fähigkeit des Menschen zur Wahrnehmung visueller Muster hilft beim Erkennen von Zusammenhängen trotz vieler überlagerter Störeffekte, Unsauberkeiten und im Allgemeinen auch einen gewissen Zeitversatz. Dadurch können untersuchte Zusammenhänge erkannt und verstanden werden, auch wenn die Struktur des Zusammenhangs komplex ist. Zusätzlich treten durch eine Visualisierung von Rohdaten vielfach unvermutete Aspekte und Anomalien zu Tage, die den Mehrwert einer visuellen Analyse potentiell weiter stark steigern.

## Literaturverzeichnis:

Chapman, P.; Clinton, J.; Kerber, R.; Khabaza, T.; Reinartz, T.; Shearer, C.; Wirth, R. (1999):  
CRISP-DM: Step-by-step data mining guide. S. 76.

Visplore GmbH - [www.visplore.com](http://www.visplore.com) (2021): Visualisierungssoftware: Visplore.