



**Nonconformity Trend Monitoring -
Feedback-basierter Ansatz zur automatisierten Überwachung von Fehlertrends in
der Produktion**

Bachelorarbeit

Fachbereich I - Wirtschafts- und Gesellschaftswissenschaften
Studiengang Wirtschaftsinformatik (Online)

Erstbetreuer: Prof. Dr. Susanne Glißmann-Hochstein
Zweitbetreuer: Prof Dr. Klaus-Peter Schoeneberg

vorgelegt von: Julia Schöpp
919091
Czarnikauer Str. 19
10439 Berlin
s54931@bht-berlin.de

Abgabetermin: 25. März 2023

Inhaltsverzeichnis

Inhaltsverzeichnis	II
Verzeichnis der Abkürzungen und Akronyme	IV
Symbolverzeichnis.....	V
Abbildungsverzeichnis	VI
Tabellenverzeichnis	VII
1 Einleitung.....	1
2 Theoretische Grundlagen	4
2.1 Nonconformity	4
2.1.1 Der Nonconformance-Report.....	5
2.1.2 Fehlerverläufe als Zeitreihen.....	6
2.2 Trend	9
2.2.1 Trend in der Literatur zu Zeitreihen.....	9
2.2.2 Trend in der Statistischen Prozesskontrolle	12
2.2.3 Zusammenfassung – Relevanter Trend	13
2.3 Visuelles Trend Monitoring	14
2.4 Automatisiertes Trend Monitoring	16
2.4.1 Control Chart Pattern Recognition.....	17
2.4.2 Qualitative Trendanalyse	19
2.4.3 Weitere automatisiertes Monitoring-Methoden für Zeitreihen	20
2.5 Feedback.....	20
2.5.1 Human Machine Teaming.....	21
2.5.2 Methoden des maschinellen Lernens	22
2.6 Zusammenfassung	27
3 Anforderungen des konkreten Anwendungsfalls.....	29
3.1 Methodik	29
3.2 Ziele des Unternehmens	30
3.3 Inputdaten	31
3.4 Wissensextraktion	33
3.5 Implementierung des Ergebnisses	36
4 Umsetzung	38
4.1 Bestimmen und Berechnen der Features	38
4.1.1 Mann-Kendall Test (p-Wert).....	39
4.1.2 Verhältnis zum Mittelwert	41
4.1.3 Exkludierbare Extremwerte	41
4.1.4 Steigung in den letzten n Tagen.....	42
4.1.5 Niveau in den letzten n Tagen	42
4.2 Vornehmen der initialen Klassifizierung.....	43
4.3 Erheben von Feedback	44

4.4	Auswertung des Feedbacks	47
4.5	Optimierung der Klassifizierung	48
4.5.1	Modelle	49
4.5.2	Ergebnisse	51
4.5.3	Bewertung	52
4.6	Implementierung des Ergebnisses	54
5	Fazit	55
A	Interviewleitfaden	58
B	Transkripte	61
C	Ergebnisse der Cross Validation, Tests und Parametersuche	103
	Literaturverzeichnis	106

Verzeichnis der Abkürzungen und Akronyme

ANN	Künstliche Neuronale Netze (engl. <i>artificial neural networks</i>)
ARIMA	Auto-Regressive Integrated Moving Average
BI	Business Intelligence
CAPA	Corrective and Preventive Action
CCPR	Control Chart Pattern Recognition
CSV	Comma Separated Values
DIN	Deutsches Institut für Normung e.V.
DT	Entscheidungsbaum (engl. <i>decision tree</i>)
FN	falsch negativ (engl. <i>false negative</i>)
FP	falsch positiv (engl. <i>false positive</i>)
HMT	Human Machine Teaming
IK	Initiale Klassifizierung
IK-o	Initialer Baum mit optimierten Parametern
ISO	International Organization for Standardization
KNN	<i>k</i> -Nearest Neighbors
KPI	Key Performance Indicator
LOESS	Locally Estimated Scatterplot Smoothing
ML	Maschinelles Lernen (engl. <i>machine learning</i>)
NCR	Nonconformance-Report
o. Jg.	ohne Jahrgang
SPC	Statistische Prozesskontrolle / Statistical Process Control
STL	Seasonal and Trend decomposition using Loess
TN	falsch positiv (engl. <i>true negative</i>)
TP	richtig positiv (engl. <i>true positive</i>)
VB-o	Vereinfachter Baum mit optimierten Parametern
VBDC-o	Vereinfachter Baum optimiert nach <i>demand pattern</i>
XGB	XGBoost

Symbolverzeichnis

β	Parameter der F-Metrik
k	Anzahl der betrachteten Datenpunkte bei KNN Algorithmus, Anzahl der Teilmengen bei der Kreuzvalidierung
n	Anzahl
m	Größe des Fensters bei <i>moving average</i> Verfahren
\bar{x}	Mittelwert von x , Name einer Qualitätsregelkarte
t	Zeitpunkt
y_t	Wert der Zeitreihe zu Zeitpunkt t
S_t	Saisonale Komponente der Zeitreihe zu Zeitpunkt t
T_t	Trendkomponente der Zeitreihe zu Zeitpunkt t
R_t	Residuen der Zeitreihe zu Zeitpunkt t

Abbildungsverzeichnis

Abbildung 2.1: Kategorisierung von Nachfragemustern (Frepple, o. J.)	8
Abbildung 2.2: STL Dekomposition mit Trend Komponente (Perktold et al., 2023).....	11
Abbildung 2.3: Definition eines Trends in der Shewhart \bar{x} Regelkarte (Nelson, 1984, S. 238)	13
Abbildung 2.4: Beispiel eines p Chart (Montgomery, 2009, S. 293).....	16
Abbildung 3.1: Verteilung der Daten nach Kategorie (eigene Darstellung)	32
Abbildung 3.2: Ein in der beispielbasierten Wissensextraktion genutzter Verlauf (eigene Darstellung)	33
Abbildung 3.3: Charakteristika von als interessant bewerteten Verläufen (eigene Darstellung)	34
Abbildung 3.4: Charakteristika von als uninteressant bewerteten Verläufen (eigene Darstellung)	35
Abbildung 4.1: Vorgehensdiagramm zur Umsetzung (eigene Darstellung)	38
Abbildung 4.2: Beispiel eines Verlaufs mit teilweiser Monotonie (eigene Darstellung)	40
Abbildung 4.3: Entscheidungsbaum zur initialen Klassifizierung (eigene Darstellung)	44
Abbildung 4.4: Screenshot des Webtools zur Feedbackerhebung (eigene Darstellung).	46
Abbildung 4.5: Boxplot der Werte für das Feature Mann-Kendall Test (p-Wert) (eigene Darstellung)	48
Abbildung 4.6: Vereinfachter Entscheidungsbaum (eigene Darstellung)	50
Abbildung 4.7: Beispielverlauf mit inkonsistenten Bewertungen (eigene Darstellung)	53
Abbildung 5.1: Beispiel für wenig intuitives Verhalten des Mann Kendall Tests (eigene Darstellung)	57

Tabellenverzeichnis

Tabelle 2.1: Features aus der Literatur zu Control Charts Pattern Recognition	18
Tabelle 2.2: Features aus der Literatur zur Zeitreihenanalyse.....	20
Tabelle 3.1: Anforderungen an die Inputdaten	32
Tabelle 3.2: Interviewergebnisse der vertieften Wissensextraktion	36
Tabelle 4.1: Zuordnung der Features zu Themen der Wissensextraktion	39
Tabelle 4.2: Zuordnung der Features und Grenzwerte des Entscheidungsbaums	44
Tabelle 4.3: Überblick über Labels im Feedback Datensatz	47
Tabelle 4.4: Überblick über Labels im Feedback Datensatz	48
Tabelle 4.5: Vergleich der besten Modelle	52

1 Einleitung

Die American Society for Quality schätzt, dass die Kosten geringer Qualität (engl. *cost of poor quality*) in den meisten Unternehmen 10 - 15 Prozent ihrer gesamten laufenden Kosten ausmachen (Duffy, 2013). Darüber hinaus ist eine hohe Qualität ihrer Produkte für viele Unternehmen ein wichtiges Differenzierungsmerkmal, für andere sogar eine gesetzliche Notwendigkeit, wie z. B. im Sektor Medizin- und Pharmaprodukte (DIN Deutsches Institut für Normung e. V., 2021). Daher ist ein gut funktionierendes Qualitätsmanagement für viele Unternehmen ein wesentlicher Bestandteil ihres Erfolgs.

Fehler möglichst früh zu identifizieren wird als wichtig erachtet, um die Kosten geringer Qualität möglichst weit zu mindern. Man spricht auch von der *Zehnerregel*, die besagt, dass Kosten für jeden Schritt im Fertigungsprozess¹, den sie unentdeckt durchlaufen, zehnmal höher ausfallen (Schmidt-Kretschmer et al., 2006). Um dieses Ziel zu erreichen, hat die Forschung und Praxis der Qualitätssicherung eine ganze Reihe an Methoden entwickelt, um den Status ihrer Prozesse und Produkte stets im Blick zu haben. Darunter fällt unter anderem das Monitoren von Prozess- und Sensordaten, aber natürlich auch die Überwachung von Fehlerdaten selbst – sei es in Form von Key Performance Indicators (KPI), Prozessregelkarten oder ähnlichen Methoden (Dietrich, 2021; Montgomery, 2009). Oft geschieht dies in Form einer Liniengrafik, die den zeitlichen Verlauf der Fehlerzahlen für bestimmte Bereiche, Werke oder Produkte darstellt.

Diese Arbeit beruht auf den Ergebnissen einer Fallstudie, die in einem Unternehmen aus der Medizintechnikbranche durchgeführt wurde. Ausgangspunkt war der Bedarf nach einem Tool, das Fehlerverläufe automatisiert monitort und auf solche Verläufe hinweist, die aktuell einen Trend nach oben zeigen. Aus Gründen der betrieblichen Geheimhaltung wurden Angaben zum Unternehmen und den konkreten Fehlerverläufen weitestgehend anonymisiert.

Um die Bedeutung des in dieser Arbeit behandelten Problems anschaulich darzustellen, bietet es sich an, ein anonymisiertes Beispiel zu betrachten. Angenommen man befindet sich in der Qualitätssicherung eines Unternehmens, das verschiedene hochqualitative Produkte anbietet. Neben der Überwachung der relevanten Prozessparameter durch die Fachbereiche, werden die auftretenden Fehler für die einzelnen Produktgruppen durch verschiedene Gremien gesichtet und diskutiert, um bei Bedarf Maßnahmen einzuleiten. In der letzten Woche war der Fehler *Kratzer auf Display* für das Modelle *A123* auffällig gehäuft. Bei der Untersuchung des entsprechenden Fehlerverlaufs fällt auf, dass der Trend des Fehlers schon seit einigen Wochen leicht steigend ist, bisher jedoch so gering

¹ Grundsätzlich sind die meisten der in dieser Arbeit getroffenen Annahmen auch auf Dienstleistungsunternehmen übertragbar. Da das Unternehmen der Fallstudie ein Produktionsbetrieb ist, steht diese Art des Unternehmens hier im Fokus.

war, dass er nicht weiter aufgefallen ist. In dem Unternehmen werden ca. 40 Produkte gefertigt, darunter selbst produzierte Komponenten anderer Produkte. Im Durchschnitt gibt es pro Produkt ca. 30 verschiedene Arten von Fehlern oder Fehlerbilder, die vorkommen können. Um alle Fehlerverläufe pro Produkt täglich oder wöchentlich zu betrachten und diejenigen zu identifizieren, die weitere Untersuchungen erfordern, müssten in dem Unternehmen also regelmäßig 1200 Verläufe beurteilt werden.

Um diese Anzahl an zu sichtenden Fehlerverläufen sinnvoll bearbeiten zu können, bietet es sich an, den Prozess ganz oder teilweise zu automatisieren. Diese Funktionalität wurde jedoch zur Zeit der Fallstudie von keinem der im Unternehmen verwendeten Softwarepakete im Bereich der Qualitätssicherung angeboten. Ein vergangener Versuch, ein solches Tool zu implementieren, war daran gescheitert, dass das Tool eine zu hohe falsch positiv Rate aufwies und eine schlechte Usability bot. Daher wurde eine Literatursichtung in den Bereichen statistische Prozesskontrolle, Zeitreihenanalyse und qualitative Trendanalyse durchgeführt (vgl. Kapitel 2). Diese Literaturrecherche ergab zwar viele hilfreiche Ansätze, jedoch keine vollständig übertragbare Konzipierung eines solchen Tools.

Das Ziel dieser Arbeit ist somit die Vorstellung der Entwicklung und Implementierung eines automatisierten Fehlertrendmonitorings in einem Produktionsbetrieb der Medizintechnikbranche. Ein automatisiertes Fehlertrendmonitoring wird wie folgt definiert:

Automatisiertes System, das in regelmäßigen Abständen eine definierte Inputmenge an Fehlerverläufen betrachtet und die Nutzenden informiert, wenn ein Muster vorliegt, das mit der kontextspezifischen Definition eines Trends übereinstimmt.

Da für Fehlerrends insbesondere ansteigende Verläufe von Interesse sind, beschränkt sich diese Arbeit auf diese Art von Trend, allerdings lassen sich viele der Überlegungen auch auf abfallende Trends übertragen.

Der Beitrag, der durch diese Arbeit geleistet werden soll, lässt sich in zwei Aspekte untergliedern:

1. Für das Unternehmen der Fallstudie soll ein kontextspezifisches Tool zum automatisierten Monitoring von Fehlerrends entwickelt und implementiert werden.
2. Als Beitrag zur Forschung soll ein auf andere Unternehmen, Kontexte und Bedürfnisse übertragbares Vorgehen zur Einführung eines solchen Fehlertrendmonitorings entwickelt werden.

Das Vorgehen dazu gliedert sich in drei Schritte: In einer Literatursichtung werden zunächst die wichtigsten Begriffe und Konzepte definiert und die theoretischen Grundlagen gelegt. Ein besonderer Fokus liegt dabei auf den Methoden verwandter Literatursparten, wie statistische Prozesskontrolle, Zeitreihenanalyse und maschinelles Lernen, aus denen Ideen und Komponenten für diese Arbeit übernommen werden können. Ergebnis des Kapitels zu den theoretischen Grundlagen soll ein grobes Vorgehensmodell sein, das im zweiten Schritt durch die Ermittlung der unternehmensspezifischen Anforderungen ergänzt wird. Schließlich wird die Umsetzung als Kombination aus theoretischen und praktischen Betrachtungen beschrieben.

2 Theoretische Grundlagen

Um diese Arbeit in ihren wissenschaftlichen Kontext einzubetten, werden in diesem Kapitel die theoretischen Grundlagen zum Thema Nonconformity Trend Monitoring gesichtet. Dazu werden relevante Literatursparten mit dem Ziel gesichtet, Konzepte und Vorgehensweisen zu identifizieren, die bei der Entwicklung einer Lösung von Nutzen sein können. Den Ausgangspunkt bildet der Begriff der *Nonconformity*, der den Rahmen steckt und Bezug zum Qualitätsmanagement herstellt. Anschließend werden die Begriffe *Trend*, *Monitoring* und *Feedback* beleuchtet, immer im Hinblick auf den *Nonconformity*-Kontext. Das Ziel ist das Evaluieren der theoretischen und praktischen Puzzlesteine und Entwickeln eines Möglichkeitsrahmens für die Lösung im konkreten Anwendungsfall.

2.1 Nonconformity

Die DIN/ISO Norm 9000:2015 zu Grundlagen und Begriffen von Qualitätsmanagementsystemen definiert eine *Nichtkonformität* als die “Nichterfüllung einer Anforderung” (DIN Deutsches Institut für Normung e. V., 2015b, S. 40). Eine weitere in der Norm vorgesehene Übersetzung des englischen Begriffs *nonconformity* ist *Fehler*. In der Praxis wird synonym mit *nonconformity* auch oft der Begriff *nonconformance* (oder *nonconformance*) benutzt, insbesondere im Zusammenhang mit dem s. g. *nonconformance report* (vgl. Kapitel 2.1.1).

Prüfung bzw. *Inspektion*, ein weiterer relevanter Begriff der DIN/ISO 9000, beschreibt, wie Nichtkonformitäten festgestellt werden können: “Das Ergebnis einer Prüfung kann Konformität oder Nichtkonformität (3.6.9) oder einen Grad von Konformität aufzeigen.” (DIN Deutsches Institut für Normung e. V., 2015b, S. 56). Daran schließt sich die Frage an, wie mit den festgestellten Nichtkonformitäten umgegangen werden soll, wozu die Norm 9001:2015 zu den Anforderungen an Qualitätsmanagementsysteme dezidierte Vorgaben macht. Dazu gehört insbesondere, auf den Fehler zu reagieren, mögliche Maßnahmen zu beurteilen, geeignete einzuleiten und deren Effekte zu überprüfen (DIN Deutsches Institut für Normung e. V., 2015a). Die Norm teilt Maßnahmen in solche zur Vorbeugung von Nichtkonformität und solche zur Korrektur vor, woraus in der Praxis das Akronym CAPA gebildet wird. Dieses setzt sich zusammen aus den folgenden Begriffen:

1. Preventive Action (Vorbeugemaßnahme): „Maßnahme zur Beseitigung der Ursache einer möglichen Nichtkonformität (3.6.9) oder einer anderen möglichen unerwünschten Situation“ (DIN Deutsches Institut für Normung e. V., 2015b, S. 63)

2. Corrective Action (Korrekturmaßnahme): „Maßnahme zum Beseitigen der Ursache einer Nichtkonformität (3.6.9) und zum Verhindern des erneuten Auftretens“ (DIN Deutsches Institut für Normung e. V., 2015b, S. 63)

Die beiden Definitionen zeigen, dass zur Beseitigung der Ursachen einer Nichtkonformität in den meisten Fällen eine Ursachenanalyse durchgeführt werden muss. Dazu existieren verschiedene in der Praxis verbreitete Methoden, wie z. B. das wiederholte Fragen nach dem Warum (Motschman & Moore, 1999) oder das Ishikawa bzw. Fischgräten-Diagramm (Ishikawa & Loftus, 1990).

Neben der in der Norm skizzierten Vorgehensweise lassen sich weitere Klassifizierungen der zur Behandlung von Nichtkonformitäten notwendigen Maßnahmen vornehmen. Liu & Cheraghi (2006) identifizieren die folgenden Schritte: *monitoring, prediction, prevention, classification, tracking, recovery*. Der erste dieser Schritte, das Monitoring, steht im Fokus dieser Arbeit.

2.1.1 Der Nonconformance-Report

Die DIN/ISO Normen 9000:2015 und 9001:2015 erwähnen nicht explizit den Nonconformance-Report (NCR). Allerdings ergibt sich aus den weiteren Anforderungen der Norm die Notwendigkeit, Nichtkonformitäten systematisch zu dokumentieren. Einerseits sieht die Norm eine Aufbewahrung der Informationen zu Nichtkonformitäten und Korrekturmaßnahmen explizit vor:

„10.2.2 Die Organisation muss dokumentierte Informationen aufbewahren, als Nachweis

- a) der Art der Nichtkonformität sowie jeder daraufhin getroffenen Maßnahme;
- b) der Ergebnisse der Korrekturmaßnahmen“

(DIN Deutsches Institut für Normung e. V., 2015a, S. 48)

Auf der anderen Seite lassen sich die Bewertung und Überprüfung der Maßnahmen zur Ursachenbeseitigung nur vornehmen, wenn Ursachen dokumentiert werden und einer Auswertung zur Verfügung stehen. Insbesondere die Anforderung „Bestimmen, ob vergleichbare Nichtkonformitäten bestehen, oder möglicherweise auftreten könnten“ (DIN Deutsches Institut für Normung e. V., 2015a, S. 48) ist schwer ohne eine systematische Dokumentation der Nichtkonformitäten denkbar. Tatsächlich wird die Funktionalität eines Nonconformance-Reports von vielen IT-Systemen im Bereich Qualitätsmanagement vorgesehen und umgesetzt (vgl. Desmar, 2022; Rahman, 2021;

Rodriguez, 2020). Welche Daten genau in einem NCR erfasst werden ist stark abhängig vom Kontext, in dem er erstellt wird. Denkbar sind Attribute wie Zeit, Ort, Produkt, Prozessschritt, Beschreibung des Fehlers, Ursachen u.v.m. (Mohsen et al., 2019). NCR können dann für Fehlerursachen- oder andere Datenanalyseaufgaben herangezogen werden (z. B. Mohsen et al., 2019). Der Zugang zu digital verfügbaren NCR mit Zeitstempeln kann als Mindestvoraussetzung für ein automatisiertes Nonconformity Trend Monitoring definiert werden.

2.1.2 Fehlerverläufe als Zeitreihen

Aggregiert man Nichtkonformitäten in einer bestimmten Frequenz, z. B. NCR pro Tag oder nichtkonforme Produkte pro Woche, so entstehen Zeitreihen (engl. *time series*). Eine Zeitreihe wird wie folgt definiert:

„Grundsätzlich bezeichnen wir mit einer Zeitreihe eine Folge von oftmals reellen Größen, die in einer diskreten (und eben nicht kontinuierlichen) Zeit anfallen, aufgezeichnet oder verarbeitet werden sollen“ (Kreiß, 2006, S. 1).

Etwas vereinfacht kann eine Zeitreihe als „sequence of observations taken sequentially in time“ (Box et al., 2015, S. 1) beschrieben werden.

Als einer der Vorteile der Betrachtung von Daten als Zeitreihen wird genannt, dass es möglich ist, Veränderungen des Untersuchungsobjektes im Zeitverlauf zu analysieren (Hirschle, 2021, S. 2). Das Haupteinsatzgebiet von Zeitreihen ist dabei die Vorhersage von zukünftigen Daten auf Grundlage der historischen Werte (Box et al., 2015; Hyndman & Athanasopoulos, 2021).

Es gibt keine allgemeingültige Beschreibung einer NCR-Zeitreihe: Ihre Eigenschaften hängen stark vom Kontext, aber auch von den getroffenen Aggregationsentscheidungen ab. Im Folgenden werden einige Annahmen über die in dieser Arbeit betrachteten Zeitreihen transparent gemacht:

1) Die betrachteten Zeitreihen sind Zähldaten-Zeitreihen

Werden NCR pro Zeiteinheit betrachtet, so handelt es sich um s. g. Zähldaten-Zeitreihen, also Zeitreihen, deren Werte nur ganzzahlig und nichtnegativ sein können. Diese Art der Zeitreihen ist verbreitet, aber weniger in der Literatur vertreten als solche, die eine Gaußsche Verteilung annehmen (R. A. Davis et al., 2021). Ihre Eigenschaft besteht unabhängig davon, ob NCR pro Zeiteinheit betrachtet werden, oder ob die im NCR als nichtkonform bezeichnete Summe an Produkten betrachtet wird. Wird jedoch eine

relative Größe betrachtet, wie z. B. die Rate der nichtkonformen Produkte bezogen auf die insgesamt produzierten Produkte in der Zeiteinheit, so trifft diese Eigenschaft nicht mehr zu.

2) Die betrachteten Zeitreihen sind univariate Zeitreihen

Univariate Zeitreihen haben nur eine abhängige Variable, bestehen also aus „[...] fortlaufenden Messungen, die mit einem einzelnen Messinstrument erhoben werden“ (Hirschle, 2021, S. 101). Die abhängige Variable ist dabei die Anzahl der Nichtkonformitäten. Würde man dazu parallel den Temperaturdurchschnitt für die entsprechende Periode betrachten, so handelte es sich um eine multivariate Zeitreihe. Bei univariaten Zeitreihen können Prognosen oder andere Erkenntnisse nicht aus exogenen Variablen gewonnen werden, sondern nur aus den vergangenen Werten der Zeitreihe (Hirschle, 2021).

3) Die Verteilungen der Zeitreihen sind abhängig von den Aggregationsebenen

Die zeitliche Aggregationsebene ergibt sich aus dem Kontext und den konkreten Anforderungen des Anwendungsfalls. Logischerweise nehmen mit zunehmender Granularität bzw. Frequenz der zeitlichen Messungen die Anzahl der Nichtkonformitäten pro Zeiteinheit ab. Ein zweiter wichtiger Faktor ist die fachliche Aggregationsebene: Diese muss so gewählt werden, dass der zeitliche Verlauf eine möglichst große inhaltliche Aussagekraft hat. So ist davon auszugehen, dass es in den meisten Fällen aussagekräftiger ist, die Fehlerverläufe pro Produkt bzw. Produktionsbereich zu betrachten als für alle Produkte gemeinsam, da es hier wahrscheinlicher ist, dass eine gemeinsame Fehlerursache vorliegt. Analog werden sich aus der Betrachtung einzelner Fehlerbilder im Zeitverlauf in den meisten Fällen relevantere Erkenntnisse ableiten lassen als aus der aggregierten Betrachtung aller Fehler gemeinsam.

Je nach Wahl der zeitlichen und inhaltlichen Aggregationsebenen ist zu erwarten, dass

1. ab einer gewissen Granularität eine große Anzahl der Zeitreihen viele Nullwerte aufweisen (so genannte *zero-inflated Models*)
2. die betrachteten Zeitreihen untereinander eine große Variabilität aufweisen.

2.1.2.1 Zero Inflated Models

Die Annahmen einer *zero-inflated* Poisson-Verteilung decken sich mit dem Anspruch einer hoch-qualitativen Produktion oder so genannter „high quality processes“ (Joeke &

Pimentel Barbosa, 2013, S. 407), dass Nichtkonformitäten im Normalfall nicht eintreten sollten:

„under a ZIP [zero inflated Poisson] model one assumes that if no random shocks occur in the process, then the number of non-conformities in a product will be zero; otherwise, the number of non-conformities follows a Poisson distribution with parameter λ “ (He et al., 2011, S. 2).

Diese Art der Verteilung stellt jedoch ein Problem für einige der üblichen Methoden der Qualitätskontrolle dar (vgl. Kapitel 2.3). Daher ist die Frage, ob es sich bei einer Zeitreihe um ein *zero-inflated* Model handelt, relevant zu betrachten.

2.1.2.2 Klassifizierung der Zeitreihen

An die Annahme, dass die aufgrund der getroffenen Aggregationsentscheidungen entstandenen Zeitreihen sich recht stark voneinander unterscheiden können, u. a. hinsichtlich der Anzahl an Nullen oder auch ihres Durchschnitts und der Varianz, schließt sich die Frage nach einer möglichen Kategorisierung der Zeitreihen an. Eine solche Kategorisierung kann genutzt werden, um das Modell zur Behandlung der Zeitreihe möglichst gut an die Charakteristika der konkreten Reihe anzupassen.

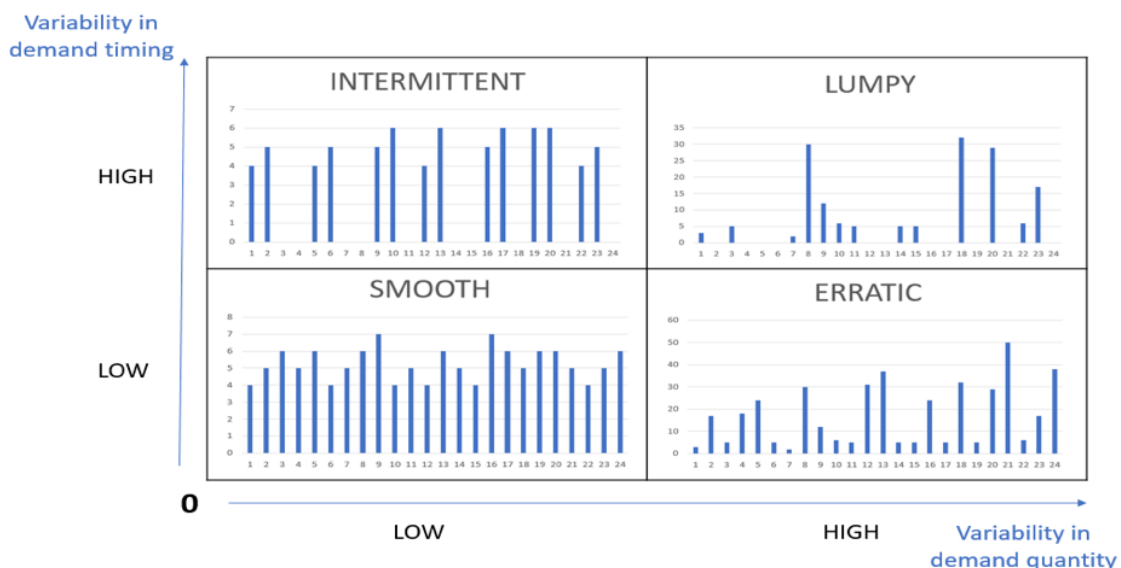


Abbildung 2.1: Kategorisierung von Nachfragemustern (Frepple, o. J.)

Während aus dem Bereich der Analyse von Nichtkonformitätszeitreihen keine solche Kategorisierung bekannt ist, gibt es eine etablierte Kategorisierungsmethode aus dem Bereich der Nachfragevorhersage, s. g. Nachfragemuster oder *demand patterns* (Kostenko & Hyndman, 2006; Syntetos et al., 2005). Die Nachfragemuster werden

anhand der Dimensionen *Nachfrageeintrittsabstände* und *Varianz der Nachfragemenge* in vier grobe Kategorien eingeteilt (Syntetos et al., 2005), wobei die genauen Übergangswerte von einer Kategorie zur anderen umstritten sind (Kostenko & Hyndman, 2006). Die so entstehenden Kategorien werden als *smooth*, *erratic*, *intermittent* und *lumpy* bezeichnet (siehe Abbildung 2.1).

In der Nachfrageforschung werden die Kategorien insbesondere dafür verwendet, geeignete Prognosemodelle für eine Zeitreihe zu identifizieren. Während als *smooth* kategorisierte Reihen mit klassischen Modellen gut prognostiziert werden können, gelten insbesondere *lumpy* Zeitreihen als besonders schwierig vorherzusagen (Kiefer et al., 2021; Mukhopadhyay et al., 2012).

In dieser Arbeit wird argumentiert, dass die beiden Dimensionen *Eintrittsabstand* und *Varianz der Menge* auch für Nichtkonformitätszahlen relevant sind und somit die Kategorisierung auch für die folgende Analyse einen Mehrwert bieten kann. Da es bisher keine gesonderte Forschung zur genauen Eingrenzung der Kategorien im Kontext von Fehlerdaten gibt, werden die Parameter aus der Nachfrageforschung übernommen.

2.2 Trend

Nachdem im vorherigen Kapitel der Kernbegriff der *Nichtkonformität* definiert und Eigenschaften von Nichtkonformitätszeitreihen als Untersuchungsgegenstand dieser Arbeit diskutiert wurden, wird im nächsten Abschnitt der Begriff *Trend* definiert. Dabei wird der Begriff aus zwei Richtungen beleuchtet: Einmal aus Perspektive der in Kapitel 2.1.2 bereits erwähnten Zeitreihenanalyse, und andererseits aus Sicht der in den nachfolgenden Kapiteln stärker vertretenen statistischen Prozesskontrolle (engl. *statistical process control*, SPC).

2.2.1 Trend in der Literatur zu Zeitreihen

Der Begriff *Trend* wird in vielen der einführenden Werke zum Thema Zeitreihenanalyse nicht explizit definiert. Mögliche Gründe dafür sind, dass die Autor*innen ein intuitives bzw. umgangssprachliches Verständnis voraussetzen, oder auch, dass der Begriff je nach Kontext verschiedene Bedeutungen annehmen kann.

Eine mögliche Definition des Trends in einer Zeitreihe ist die der geglätteten Werte, wie in der folgenden Aussage deutlich wird:

„[...] erkennt man, dass ein Trend in gewisser Weise eine Glättung der beobachteten Zeitreihe darstellt. Deshalb kommen auch aus der Statistik bekannte Glättungsmethoden zur Trendschätzung in Frage“ (Kreiß, 2006, S. 10).

Zu den im obigen Zitat genannten Glättungsmethoden gehören insbesondere die des gleitenden Mittels (*moving average*) und das exponentielle Glätten (*exponential smoothing*) (Kreiß, 2006). Beim Glätten mittels *moving average* wird anstelle des Wertes an einem gewissen Punkt der Mittelwert eines Fensters der Größe m , zentriert auf den aktuellen Wert gebildet (Hyndman & Athanasopoulos, 2021). Beim exponentiellen Glätten werden Werte, die sich näher am aktuellen Wert befinden, exponentiell stärker gewichtet als solche, die sich weiter entfernt befinden, da sie als relevanter für die Trendschätzung betrachtet werden (Hyndman & Athanasopoulos, 2021). Ein weiteres Verfahren nennt sich *locally estimated scatterplot smoothing* (LOESS) und beruht darauf, dass für ein lokales Fenster an gewichteten Werten eine lineare oder quadratische Funktion gefunden wird, deren Wert an Stelle des eigentlichen Wertes gesetzt wird (Cleveland, et al., 1990). Die Wahl der Gewichte und der Größe des Fensters können so manipuliert werden, dass der geschätzte Trend mehr oder weniger stark geglättet wird, und damit mehr oder weniger robust gegenüber Ausreißern wird.

LOESS bildet den Hauptbestandteil der s. g. *STL-Dekomposition*, die für *seasonal and trend decomposition using LOESS* steht (Cleveland, et al., 1990). Bei der Dekomposition einer Zeitreihe wird sie in die drei Komponenten Trend, Saisonalität und Residuen zerlegt, wobei der saisonale Anteil z. B. wöchentlich wiederkehrende Schwankungen abbildet und die Residuen der Teil sind, der nicht von Trend und Saisonalität abgedeckt werden (Hyndman & Athanasopoulos, 2021). Nimmt man an, dass diese Komponenten additiv zur finalen Zeitreihe zusammengefügt werden, ergibt sich die folgende Formel einer solchen Dekomposition:

$$y_t = S_t + T_t + R_t$$

Dabei ist y_t der Wert der Zeitreihe bzw. die Daten, S_t die saisonale Komponente, T_t die Trendkomponente und R_t die Residuen-Komponente zu Zeitpunkt t .

Die STL-Dekomposition ist ein populäres Instrument zur optischen Analyse von Zeitreihen, da sie einen guten Überblick über die verschiedenen Faktoren gibt, die den Verlauf einer Zeitreihe beeinflussen (Cleveland, et al., 1990; Hyndman & Athanasopoulos, 2021). Ein Beispiel einer solchen Dekomposition zeigt Abbildung 2.2:

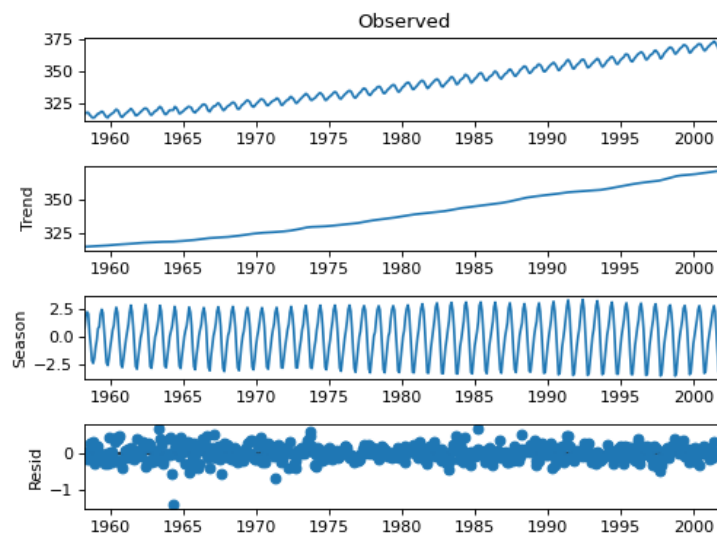


Abbildung 2.2: STL Dekomposition mit Trend Komponente (Perktold et al., 2023)

Während die STL-Dekomposition zur visuellen Analyse einer Zeitreihe sehr nützlich ist, so kann sie bei der Frage, ob ein Trend vorliegt, nicht helfen, da nach der o. g. Definition immer eine Trend-Komponente vorhanden ist. Eine Möglichkeit, die Stärke des Trends in Relation zu den beiden anderen Komponenten zu quantifizieren, wird von Hyndman & Athanasopoulos (2021) vorgeschlagen. Sie beruht auf der Idee, dass bei stark von Trends geprägten Daten die Daten ohne die saisonale Komponente eine stärkere Varianz aufweisen sollten als nur die Residuen. Allerdings gibt diese Metrik keine Auskunft darüber, in welche Richtung der Trend sich bewegt, oder auch ob in dem betrachteten Zeitraum mehrere starke Trends in unterschiedliche Richtungen gab.

Oft wird sich bei der Untersuchung von Zeitreihen der Frage nach einem Trend von der Gegenrichtung aus genähert, nämlich bei der Frage, ob eine Zeitreihe stationär ist. Eine stationäre Zeitreihe wird definiert als eine, deren statistische Eigenschaften – wie Mittelwert oder Standardabweichung – nicht davon abhängen, an welchem Zeitpunkt sie betrachtet wird (Hyndman & Athanasopoulos, 2021). Zeitreihen mit Trend sind demnach nicht stationär. Diese Betrachtung ist wichtig, da einige der klassischen Vorhersagemodelle – insbesondere die ARIMA-Modelle – Stationarität voraussetzen. Neben der visuellen Überprüfung gibt es verschiedene statistische Tests, die die Zeitreihe auf Stationarität überprüfen, dabei aber vor allem auf das Vorhandensein einer bestimmten Art der Nichtstationarität, nämlich der Einheitswurzel (engl. *unit root*) testen (Hyndman & Athanasopoulos, 2021).

Eine andere Blickrichtung auf den Begriff Trend ist, zu betrachten, ob ein monotoner Trend vorliegt, d. h. ein Trend, der sich vorwiegend oder ausschließlich in eine Richtung bewegt. Für diese Frage hat sich vor allem der Mann-Kendall Test durchgesetzt, der auf

der Bewertung von Rangkorrelationen beruht (Kendall, 1948; Mann, 1945). Als Nullhypothese wird dabei angenommen, dass die Werte einer zufälligen Anordnung folgen, während die Gegenhypothese besagt, dass ein monotoner Trend vorliegt (Mann, 1945). Einer der Vorteile des Tests ist, dass er nicht-parametrisch ist, also keine bestimmten Anforderungen an die Verteilung der Daten stellt – im Gegensatz z. B. zu der linearen Regression, die auch eine bewährte Methode zur Abschätzung der Steigung eines Trends ist und die eine Normalverteilung der Daten voraussetzt (Mirabbasi et al., 2020).

Der Mann-Kendall Test ist besonders in der Klimaforschung und Hydrologie sehr populär und wird dort regelmäßig zur Erkennung und Beurteilung von Trends eingesetzt. Mehrere Studien in diesen Bereichen bestätigen dabei seine Effektivität (Alashan, 2020; Loftis et al., 1989; Mirabbasi et al., 2020; F. Wang et al., 2020). Der Test wird jedoch auch in anderen Bereichen eingesetzt, wie z. B. der Überwachung von Gebäudedaten bezüglich des Zustands von Beton (Yang & Han, 2022) oder dem *software aging*, also dem altersbedingten Verschleiß von Software (Machida et al., 2013). Während die genannten Studien dem Test überwiegend eine gute Fähigkeit zur Detektion von Trends attestieren, so wird als einer der Nachteile eine hohe Falschpositivrate genannt (Machida et al., 2013).

Es existieren verschiedene Varianten des Mann-Kendall Tests, die z. B. bei der Bewertung des Trends das Vorhandensein von saisonalen Mustern in der Zeitreihe berücksichtigen können. Entsprechende Softwarepakete bieten diese Varianten standardmäßig mit an (Hussain & Mahmud, 2019).

2.2.2 Trend in der Statistischen Prozesskontrolle

Die statistische Prozesskontrolle oder auch statistische Prozesslenkung beschäftigt sich mit dem Einsatz statistischer Methoden zur Qualitätskontrolle von Produktionsprozessen (Dietrich, 2021; Montgomery, 2009; Western Electric Co, 1958). Einer ihrer Begründer, Walter Shewhart, prägte auch den Namen der als Qualitätsregelkarte bekannten Methode, die im englischen als *Shewhart control chart* bekannt ist (Shewhart, 2015). Die Qualitätsregelkarte wird in Kapitel 2.3 nochmal detaillierter beleuchtet.

Die SPC fragt, ob ein Prozess grundsätzlich unter Kontrolle ist, also ob die Werte innerhalb eines Normbereichs nur zufälligen Schwankungen unterliegen, oder ob ein unnatürliches Muster vorliegt, das auf ein Problem hinweist (Dietrich, 2021). Hier geht es also insbesondere darum, ob eine Ursache vorliegt, die einen Handlungsbedarf begründet. In diesem Kontext kann ein Trend eine Form eines solchen Musters darstellen:

„A trend can be defined as a continuous movement in one direction.“ (Cheng, 1997, S. 669)

Die SPC kennt verschiedene solcher Muster, die oft mit unterschiedlichen Fehlerursachen in Zusammenhang gebracht werden (Nelson, 1985). Ein weiteres Muster, das dem Trend sehr ähnlich ist, sind *shifts*:

“A shift may be defined as a sudden or abrupt change in the average of the process.” (Cheng, 1997, S. 669)

Für die gebräuchlichste Art der Qualitätsregelkarte, die \bar{x} Regelkarte, wurden auch einfache Regeln zur Erkennung einiger Muster definiert, wie z. B. sechs Punkte in einer Reihe stetig ansteigend oder abfallend als Trend (Nelson, 1984, 1985).

Test 3. Six points in a row steadily increasing or decreasing

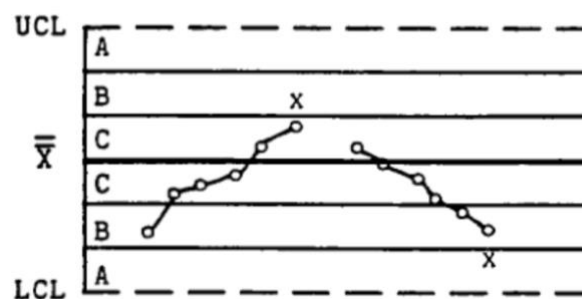


Abbildung 2.3: Definition eines Trends in der Shewhart \bar{x} Regelkarte (Nelson, 1984, S. 238)

Da diese Regel eher eine simple Heuristik als ein statistischer Test ist, überrascht es nicht, dass er in einer späteren Untersuchung als unzureichend für die zuverlässige Erkennung von Trends in Qualitätsregelkarten evaluiert wurde (R. B. Davis & Woodall, 1988).

2.2.3 Zusammenfassung – Relevanter Trend

Die Betrachtungen zum Thema Trend haben gezeigt, dass der Begriff in verschiedenen Kontexten unterschiedliche Bedeutungen erhalten kann. Ausgehend vom Ziel des konkreten Anwendungsfalls, frühzeitig die Fehlerverläufe zu identifizieren, bei denen eine Handlung erforderlich ist, deckt sich das Verständnis mit dem der SPC, nämlich vom Trend als unnatürliches Muster. Welche Verläufe genau unter das Muster *Trend* fallen – und somit als relevanter Trend im Sinne der Aufgabe verstanden werden können – ist dabei jedoch stark vom konkreten Kontext abhängig.

Auch die Interpretation des Trends als eine geglättete und von Saisonalität befreite Darstellung des eigentlichen Verlaufs, und der Test auf Monotonie nach Mann und Kendall können in der Umsetzung des Trend Monitorings von Nutzen sein.

2.3 Visuelles Trend Monitoring

Nachdem die Begriffe Nonconformity und Trend beleuchtet wurden, kann nun der dritte Begriff in die Aufgabenbeschreibung aufgenommen worden. Dieser beschreibt, was mit den Fehlerrends geschehen soll, und das ist – wie bereits erwähnt, das Monitoring.

Zur Definition des Begriffs *Monitoring* kann erneut die DIN/ISO 9000:2015 herangezogen werden, die *Monitoring* (bzw. *Überwachung* auf Deutsch) wie folgt definiert:

„Bestimmung (3.11.1) des Zustands eines *Systems* (3.5.1), eines *Prozesses* (3.4.1), eines *Produkts* (3.7.6), einer *Dienstleistung* (3.7.7) oder einer Tätigkeit

Anmerkung 1 zum Begriff: Bei der Bestimmung des Zustands kann es erforderlich sein, zu prüfen, zu beaufsichtigen oder kritisch zu beobachten.

Anmerkung 2 zum Begriff: Überwachung ist üblicherweise eine Bestimmung des Zustands eines *Objekts* (3.6.1), die in verschiedenen Stufen oder zu verschiedenen Zeiten durchgeführt wird.“

(DIN Deutsches Institut für Normung e. V., 2015b, S. 55)

Vereinfacht kann Monitoring in diesem Kontext also als regelmäßige Überprüfung der Nichtkonformitätsverläufe hinsichtlich der Frage, ob ein relevanter Trend vorliegt, definiert werden. Da automatisierte Monitoring-Ansätze in Kapitel 2.4 gesondert untersucht werden, liegt der Fokus im Folgenden auf der Frage, welche manuellen bzw. visuellen Monitoring-Methoden es gibt, die für das Monitoring der Fehlerverläufe von Nutzen sein können.

Vorab lohnt es sich jedoch zu betrachten, warum gerade visuelle Methoden zum Trendmonitoring so verbreitet sind. Zurückblickend auf Kapitel 2.2 fällt auf, dass viele der vorgestellten Methoden, insbesondere das Glätten von Trends und die STL-Dekomposition in erster Linie auf die visuelle Interpretation ausgelegt sind. Sein Erfinder begründet z. B. den LOESS-Algorithmus wie folgt:

„The visual information on a scatterplot can be greatly enhanced, with little additional cost, by computing and plotting smoothed points.“ (Cleveland, 1979, S. 829)

Dieses Zitat ist interessant, weil es zweierlei unterstellt:

1. Die visuelle Betrachtung von Daten – hier als Punktwolke – ist ein wichtiger Mechanismus der Datenanalyse.
2. Menschen können bzw. müssen in dieser Aufgabe zusätzlich unterstützt werden, hier durch Glätten der Punktwolke.

Interessant ist die erste der beiden Hypothesen insbesondere, wenn wir die menschliche visuelle Prüfung mit der automatisierten vergleichen, wie das folgende Zitat zeigt:

„Although humans are very good at visually detecting such [trend] patterns, for control system software it is a difficult problem including trend extraction and similarity analysis.“ (Juuso, 2011, S. 162)

Die Wahrnehmungsforschung bestätigt die inhärente Überlegenheit des Liniendiagramms zur menschlichen Wahrnehmung von Trends (Quadri & Rosen, 2022). Kindsmüller & Urbas (2002) untersuchen, von welchen Faktoren die Fähigkeit von Prozessoperateur*innen abhängt, Trendverläufe in Liniendiagrammen korrekt in ihre Einzelkomponenten zu zerlegen, und identifizieren Modelwissen bzw. Training als relevanteste Einflussgrößen. Leider ist wenig darüber bekannt, zu welchen Arten von Fehlschlüssen Menschen bei der Interpretation von Trends in Liniendiagrammen neigen – und wie diese vermieden werden können.

Qualitätsregelkarten

Eine sehr verbreitete Methode des assistierten visuellen Monitorings von Trendverläufen sind die bereits erwähnten Qualitätsregelkarten der statistischen Qualitätskontrolle. Es gibt verschiedene Ausprägungen der Regelkarten, wobei eine grobe Unterscheidung in Regelkarten für Variablen (also z. B. Sensor- oder andere Messdaten eines Prozesses) und Regelkarten für Attribute, wie Rate der nichtkonformen Produkte, getroffen werden kann (Montgomery, 2009). Dabei beschäftigt sich die SPC insbesondere mit Fragen nach der Berechnung von Stichprobengrößen und Kontrollgrenzen. Für die Kontrolle von Nichtkonformitäten sind das so genannte *p Chart* für den Anteil bzw. die Anzahl von nichtkonformen Produkten in einer Stichprobe, bzw. das *c Chart* für die Anzahl an Nichtkonformitäten an einem Produkt (Montgomery, 2009) die wichtigsten Arten von Regelkarten.

Da sich die Standardmethoden für die Berechnung von Regelkarten zum Teil nicht optimal auf Attribut-Regelkarten übertragen lassen (Woodall, 1997) und insbesondere bei der Betrachtung von Nichtkonformitäten zusätzlich das Problem der *zero-inflated* Prozesse besteht (Mahmood & Xie, 2019), wurden verschiedene spezielle Varianten und

Algorithmen zur Modellierung dieser Daten entwickelt (Javadi & Niaki, 2013; Joeques & Pimentel Barbosa, 2013; Khoo, 2004).

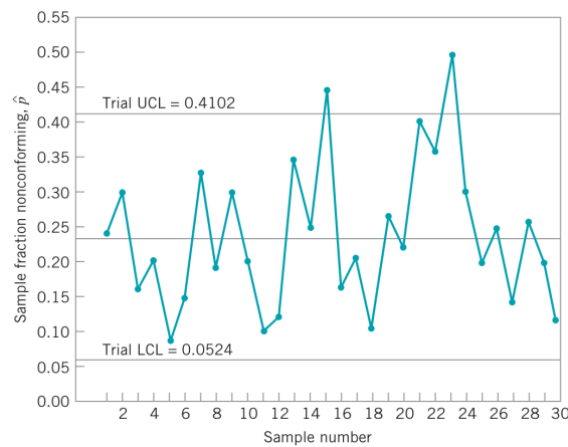


Abbildung 2.4: Beispiel eines p Chart (Montgomery, 2009, S. 293)

Darüber hinaus ist wichtig zu betrachten, inwiefern die in Kapitel 2.1.2 beschriebenen Annahmen zu den hier betrachteten Daten überhaupt mit den Annahmen der Regelkarten übereinstimmen. Eine Einschränkung ist, dass auch bei den so genannten p Charts von einer fixen Stichprobengröße ausgegangen wird, im Gegensatz zu der absoluten Anzahl aller Nichtkonformitäten pro Zeitraum. Außerdem werden die Regelkarten meist nach Stichprobennummer sortiert anstatt nach Datum, was die Effekte von Saisonalität wie z. B. Produktionspausen an Wochenenden umgeht. Eine Möglichkeit wäre also, die Daten so aufzubereiten, dass sie diesem Format entsprechen. Ist dies jedoch nicht erwünscht oder möglich, so sind die Methoden der Regelkarten nicht direkt auf die Fehlerverläufe übertragbar. Dennoch können sie interessante Hinweise liefern.

Auch wenn die Kontrollgrenzen, also das *upper control limit* und *lower control limit* als teilautomatisierte Unterstützung des Monitorings betrachtet werden können, so beruht die Ausgabe von Regelkarten doch immer noch auf der Annahme einer menschlichen visuellen Kontrolle. Daher werden im nächsten Absatz automatisierte Ansätze des Trend Monitorings betrachtet.

2.4 Automatisiertes Trend Monitoring

Automatisierung kann wie folgt definiert werden:

„Übertragung von Funktionen des Produktionsprozesses, insbesondere Prozesssteuerungs- und -regelungsaufgaben vom Menschen auf künstliche Systeme.“ (Voigt, 2018)

Ziele können dabei die Entlastung der Menschen von repetitiven oder starren Tätigkeiten oder die Einsparung von Kosten sein (Voigt, 2018). In diesem Fall soll die Automatisierung dazu dienen, das systematische Monitoring von Trendverläufen zu ermöglichen, das sonst aufgrund seines Umfangs in vielen Unternehmen bzw. Kontexten nicht umzusetzen wäre.

Dazu wird zuerst die Automatisierung der in Kapitel 2.3 betrachteten Regelkarten diskutiert. Anschließend werden mit dem qualitativen Trendmonitoring eine weitere automatisierte Trendanalysemethode, und abschließend andere automatisierte Methoden der Zeitreihenanalyse, die sich nicht in erster Linie mit Trends beschäftigen, beleuchtet.

2.4.1 Control Chart Pattern Recognition

Ziel der Regelkarten der SPC ist, wie in Kapitel 2.3 argumentiert, die visuelle Kontrolle von Prozessdaten im Hinblick auf unnatürlicher Muster, die auf einen Fehler hinweisen können (Hachicha & Ghorbel, 2012). Bereits in den späten 1990er Jahren wurden erste Ansätze entwickeln, die die Sichtung der *control charts* automatisieren sollten. Diese automatisierte Mustererkennung anhand von Qualitätsregelkarten wird in der Literatur unter dem Namen *Control Chart Pattern Recognition* (CCPR) geführt (Hachicha & Ghorbel, 2012). Interessanterweise haben die CCPR-Ansätze oft nicht die komplette Automatisierung des Prozesses zum Ziel, sondern die Unterstützung der zuständigen Prozessingenieur*innen in der Interpretation der Muster, eine Aufgabe, die als anspruchsvoll und trainingsabhängig betrachtet wird (Cheng, 1997; Swift, 1987).

Die CCPR macht sich dabei unterschiedliche Methoden der Mustererkennung zunutze. Auf der einen Seite wird die Codierung von Expert*innenwissen in so genannten wissensbasierten Expertensystemen (engl. *knowledge based expert models*) genutzt, die oft in Form eines Entscheidungsbaums dargestellt werden (Bag et al., 2012; Swift, 1987). Verbreiteter sind allerdings Ansätze, die auf maschinellem Lernen beruhen, hier insbesondere die künstlichen Neuronale Netze (engl. *artificial neural networks*, ANN), die gut für die Aufgabe der Mustererkennung geeignet sind (Hachicha & Ghorbel, 2012). Teilweise werden auch hybride Modelle aus ANN und Entscheidungsbäumen bzw. *expert models* genutzt (Guh, 2005; Ramezani & Jassbi, 2017). In einer vergleichenden Studie ermittelten Avci et al. (2022) jedoch, dass die ANN bei der Identifikation von unnatürlichen Mustern klassischen bzw. einfacheren Klassifikationsmethoden nicht überlegen waren.

Eine weitere Möglichkeit, die Methoden im Bereich CCPR zu unterscheiden ist danach, ob sie als Input für die Mustererkennungsalgorithmen berechnete Kennzahlen zu den

Daten nutzen (s.g. *Features*) oder ob sie auf der Ebene der Rohdaten operieren. Ein Vorteil der Mustererkennung auf der Grundlage von Features ist, dass sie die Größe der Inputvektoren verringert und damit die Trainingszeiten (Pham & Wani, 1997). Es wird jedoch auch argumentiert, dass sie aufgrund der geringeren Dimensionalität besser generalisieren, also Muster erkennen, die auch über die Trainingsdaten hinaus zutreffend sind (Hassan et al., 2003). Dabei müssen die ausgewählten Features in der Lage sein, die relevanten Eigenschaften der Daten ausreichend zu beschreiben. Tabelle 2.1 listet einige Features auf, die in der Literatur zur feature-basierten CCPR genutzt werden.

Relevante Features	Beschreibung	Quelle
Steigung (<i>slope</i>) gemessen durch lineare Regressionslinie, Anzahl der Schnittpunkte des Verlaufs mit der Mittelwert- bzw. Regressionslinie, Steigung der Liniensegmente (Unterteilung der Segmente auf Grundlage v. Überschreitungen des Mittelwerts), Fläche zwischen Verlauf und Mittelwerts- bzw. Regressionslinie und Liniensegmenten	Geometrische Heuristiken zur Beschreibung der sechs häufigsten Muster (Normales Muster, zyklisches Muster, steigender Trend, abnehmender Trend, Shift aufwärts, Shift abwärts)	Pham & Wani, 1997
Statistische Features (Mittelwert, Varianz etc.), insbesondere Gradient des Mittelwerts	Unterscheidung zwischen linearem Trend und plötzlichem Shift auf Grundlage der Größenordnung der Veränderung des Mittelwertes	Hassan et al., 2003
Reduzierte Features von Pham & Wani (1997) plus maximale Variation in kurzem Zeitintervall	Unterscheidung zwischen Shift und linearem Trend durch Codierung der plötzlichen Änderung	Ebrahimzadeh et al., 2013

Tabelle 2.1: Features aus der Literatur zu Control Charts Pattern Recognition

Es stellt sich die Frage, inwieweit die Methoden der CCPR auf den hier vorliegenden Fall des Monitorings von Fehlertrends übertragbar sind. Dabei ergeben sich drei Probleme oder Einschränkungen:

1. Wie in Kapitel 2.3 erwähnt handelt es sich bei den NCR-Zeitreihen nicht um klassische Kontrollchartdaten. Die größte Übereinstimmung besteht mit dem *p chart* zur Darstellung von Fehlerraten pro Stichprobe, allerdings mit der zusätzlichen Schwierigkeit, dass bei einem Teil der Verläufe von einer *zero inflation* auszugehen ist. Die Literatur zu CCPR beschäftigt sich jedoch fast ausschließlich mit den klassischen Shewhart \bar{x} -Charts für Prozessvariablen, die von einer Normalverteilung der Daten ausgehen: In ihrem Review zu CCPR identifizierten Hachicha & Ghorbel (2012) nur zwei aus über einhundert Beiträgen, die sich mit nicht-normalverteilten Daten beschäftigen.

2. Der Fokus der überwiegenden Beiträge im Bereich CCPR liegt in der Unterscheidung der Muster untereinander², während in dieser Arbeit lediglich solche Muster betrachtet werden, die als Trend bezeichnet werden können. Das führt dazu, dass weniger Augenmerk auf die Möglichkeit gelegt wird, innerhalb einer solchen Mustergruppe zwischen signifikanten und nicht-signifikanten Verläufen zu unterscheiden.
3. Viele der Arbeiten zum Thema CCPR basieren auf simulierten, also künstlichen Daten – im Review von Hachicha & Ghorbel (2012) sind es über 90 Prozent. Es ist unklar, inwieweit sich die Methoden, die anhand dieser synthetischen Daten entwickelt wurden, auf Prozessdaten aus der Praxis übertragen lassen.

Eine weitere offene Frage ist die der Sensitivität, wie im folgenden Zitat deutlich wird:

„A natural question that arises at this point is what disturbance levels should be selected to set the recognizers for. This question can be answered based on expertise of the quality control engineer.“ (Al-Ghanim & Jordan, 1996, S. 47)

Wie stark oder relevant die Abweichung des vorliegenden Musters sein muss, um von einem automatisierten Monitoring aufgenommen zu werden, muss demnach auf Grundlage der Fachexpertise im Unternehmen getroffen werden.

2.4.2 Qualitative Trendanalyse

Die qualitative Trendanalyse beschäftigt sich mit Zeitreihen hoher Frequenz, wie sie insbesondere von Sensordaten erzeugt werden, und hat zum Ziel, qualitative Beschreibungen solcher Trends in Echtzeit zur Verfügung zu stellen (Janusz & Venkatasubramanian, 1991). Dazu wird die Zeitreihe in zeitliche Segmente unterteilt:

„In general, qualitative trend analysis is concerned with the segmentation of time series into so called episodes. Such episodes are characterized by means of a start time, an end time, and a unique set of signs for the first and second derivative.“ (Villez & Rengaswamy, 2013, S. 1958)

Basierend auf der qualitativen Trendanalyse wurden Frühwarnsysteme für Prozesse in der Ölproduktion (Zhang et al., 2019) oder der Qualität von Trinkwasser (Tomperi et al., 2016) entwickelt. Für die Überwachung von Fehlerrends ist sie jedoch weniger fruchtbar, da hier nicht die Herausforderungen der Sensordaten – also riesige Datenmengen mit großem Grundrauschen bedingt durch die Messungen – bestehen.

² Eine Ausnahme bildet hier das Paper von Fahmy & Elsayed (2006), das sich explizit mit der Entdeckung linearer Trends auseinander setzt, aber auch von \bar{x} -Charts für Messvariablen ausgeht.

2.4.3 Weitere automatisiertes Monitoring-Methoden für Zeitreihen

Neben der Mustererkennung in Prozessdaten werden Zeitreihen für eine große Zahl weiterer Anwendungsfälle genutzt, für die eigene Methoden entwickelt wurden. Dazu gehören die *outlier-* und *anomaly detection* (Ganapathy et al., 2011; Hyndman et al., 2015; Laptev et al., 2015), die Identifizierung von Wendepunkten (engl. *change point detection*) (Su et al., 2013; Zandi et al., 2011), *clustering* (Fulcher, 2018; X. Wang et al., 2006), Klassifizierung (Fulcher, 2018; Ji et al., 2022) und die Automatisierung von Vorhersagen (Cerqueira et al., 2021; Jombart et al., 2021; Starks & Mercado, 1992).

Auch hier ist ein verbreiteter Ansatz die Betrachtung relevanter Features anstelle der gesamten Rohdaten. Die genannten Vorteile ähneln dabei denen in Kapitel 2.4.1 genannten, darunter die Reduktion der Dimensionalität, größere Robustheit gegenüber fehlenden Werten und bessere Vorhersageergebnisse (X. Wang et al., 2006). Die Auswahl der konkreten Features, die für den einzelnen Anwendungsfall nützlich sind, hängen dabei sowohl von der Aufgabe als auch von der Domäne ab (X. Wang et al., 2006). In Tabelle 2.2 werden einige in der Literatur verwendete Features zusammengefasst.

Feature	Beschreibung	Quelle
Mittelwert	Mittelwert (Mean)	(Hyndman et al., 2015)
Varianz	Varianz	(Hyndman et al., 2015)
Autokorrelation	Wie stark korrelieren Werte einer Zeitreihe mit vergangenen Werten	(Hyndman et al., 2015; X. Wang et al., 2006)
Stärke des Trends	Vgl. Trendstärke nach (Hyndman & Athanasopoulos, 2021) vorgestellt in Kapitel 2.2.1	(Hyndman et al., 2015; X. Wang et al., 2006)
Stärke der Saisonalität	Analog zu Stärke des Trends	(Hyndman et al., 2015; X. Wang et al., 2006)
Lumpiness	Veränderung in der Varianz der Residuen	(Hyndman et al., 2015)
Symmetrie	Schiefe und Wölbung	(X. Wang et al., 2006)
Netzwerkfeatures	Mapping von Zeitreihen in komplexe Netzwerke	(Silva et al., 2022)
Entropie	Quantifizierung der Vorhersagbarkeit einer Zeitreihe	(Fulcher, 2018)

Tabelle 2.2: Features aus der Literatur zur Zeitreihenanalyse

2.5 Feedback

Die bisherigen Kapitel haben sich damit beschäftigt, was unter einem automatisierten Monitoring von Fehlertrends verstanden werden kann. Jetzt soll konkreter auf das *Wie*

eingegangen werden, das bereits im Titel der Arbeit angedeutet wurde, nämlich auf Grundlage von Feedback.

Der Begriff Feedback hat zwei miteinander verwandte Bedeutungen: Im technischen Sinne kann er die Steuerung eines Systems durch Rückmeldung der Ergebnisse bedeuten, und im soziologischen Sinne eine Reaktion auf eine Äußerung oder ein Verhalten oder eine Rückmeldung (Stangl, 2023).

Um die Notwendigkeit dieses Ansatzes zu begründen kann nochmal ein Blick zurück auf die Definitionen von Trend geworfen werden. Hier und an verschiedenen anderen Stellen, z. B. der Festlegung von Sensitivitätsgrenzen (vgl. Kapitel 2.4.1), ergibt sich ein individueller Ermessensspielraum, der sich aus dem fachlichen Kontext begründet. Dabei gibt es zwei Möglichkeiten: Die relevanten Definitionen, Regeln und Parameter können a) vorab bzw. *top down* definiert werden oder b) *bottom up* aus dem Feedback oder den Rückmeldungen der Expert*innen erlernt werden. Nimmt man an, dass das Fachwissen zur Beurteilung von Trends weitestgehend internalisiert ist und auf Erfahrungen beruht (vgl. Hurwitz et al., 2020), so spricht dies für ein Vorgehen nach Ansatz b). Das Erlernen von Mustern aus (Feedback-) Daten entspricht den verbreiteten Definitionen von maschinellem Lernen (ML) (Frochte, 2018; Géron, 2018). Daher werden in den folgenden Abschnitten Methoden aus diesem Bereich beschrieben, die zum automatisierten Monitoring von Fehlertrends genutzt werden können.

2.5.1 Human Machine Teaming

Bevor konkrete Algorithmen des maschinellen Lernens betrachtet werden, lohnt es sich, auf einer Meta-Ebene die Zusammenarbeit von Mensch und Maschine in teilautomatisierten Abläufen zu betrachten, ein Thema, mit dem sich die Literatur zum so genannten *human machine teaming* (HMT) beschäftigt. HMT spielt insbesondere in Kontexten eine Rolle, in denen maschinelle Akteure als autonome Teammitglieder betrachtet werden können, insbesondere im Bereich Militär (Lyons et al., 2019; McDermott et al., 2018), aber auch Medizin (Link et al., 2022). Aus soziologischer Sicht wird dann unter anderem die Frage betrachtet, wie Vertrauen zwischen den menschlichen und maschinellen Akteuren gebildet wird (Lyons et al., 2019).

Versteht man den Begriff etwas weiter können auch verwandte Forschungsgebiete, die sich mit der Inkorporierung von Expert*innenwissen in Modelle des maschinellen Lernens beschäftigen – auch *expert augmented machine learning* genannt – unter das HMT gefasst werden. In diesem Bereich gibt es auch einige Forschungsarbeiten zum Thema Produktion und Qualitätsmanagement (Link et al., 2022; Sylla, 2002). Liegen

ausreichend große Datenmengen vor, kann dem Ansatz von Gennatas et al. (2020) gefolgt werden: Im medizinischen Bereich generierten sie aus Inputdaten mit Hilfe eines *gradient boosting* Algorithmus (vgl. Kapitel 2.5.2) eine große Anzahl an Entscheidungsbäumen, die dann in Regeln konvertiert wurden. Aus den Regeln wurden mit Hilfe statistischer Methoden die relevantesten ausgewählt und damit der Regelsatz auf ein deutlich weniger umfangreiches Maß reduziert. Schließlich wurden die finalen Regeln den Expert*innen zur Kontrolle übergeben. Dieser Ansatz geht jedoch davon aus, dass bereits ein relative großer und gut strukturierter Inputdatensatz vorliegt. Link et al. (2022) beschäftigen sich im Kontext des Qualitätsmanagements mit der Frage, wie Expert*innenwissen genutzt werden kann, um geringe Datenmengen beim Training von ML zu kompensieren. Verallgemeinert betrachtet sieht ihre Vorgehensweise die folgenden Schritte vor:

1. Erstellen eines initialen Modells
2. Inspektion und Bewertung des Modells durch Expert*innen
3. Integration des Expert*innenwissens in das Modell

Schritt zwei und drei können dabei fortlaufend iterativ wiederholt werden. Das Vorgehen hat mehrere Vorteile, unter anderem, dass es bessere Ergebnisse liefert als Modelle, die kein Expert*innenwissen einbeziehen, dass eine höhere Akzeptanz unter den Nutzer*innen erzielt wird und dass der Ansatz bereits im Zuge der Modellerstellung und Optimierung nützliche Ergebnisse liefern kann (Link et al., 2022).

Konkrete Hilfestellung bei der Erstellung von HMT-Systemen gibt der HMT Systems Engineering Guide (McDermott et al., 2018), allerdings mit einem Fokus auf militärische Systeme. Dennoch sind einige methodische Hinweise im Kontext dieser Arbeit hilfreich. So wird beispielsweise erwähnt, dass im Rahmen der Wissenserfassung von Expert*innen (engl. *knowledge audit*) eine bewährte Vorgehensweise ist, kurze, konkrete Beispiele von Entscheidungssituationen anstelle von generischem Wissen abzufragen (McDermott et al., 2018).

2.5.2 Methoden des maschinellen Lernens

Im maschinellen Lernen wird zwischen überwachtem und unüberwachtem Lernen (engl. *supervised / unsupervised learning*) unterschieden. Das Ziel des unüberwachten Lernens kann wie folgt beschrieben werden:

„Wir haben einfach nur eine Menge an Daten und wollen mittels unüberwachtem Lernen versuchen, versteckte Strukturen in unmarkierten Daten zu finden.“
(Frochte, 2018, S. 24)

Der Vorteil dabei ist, dass keine Bewertungen der Daten, so genannte *Label*, vorliegen müssen. Unüberwachtes Lernen wird oft im Bereich des *clusterings* (d. h. Bildung von Gruppen oder Kategorien anhand der Daten) verwendet (Frochte, 2018), auch im Bereich Zeitreihenanalyse (vgl. Kapitel 2.4.3). Für den im Rahmen dieser Arbeit betrachteten Fall böten Methoden des unüberwachten Lernens zwar den Vorteil, dass keine gelabelten Daten generiert werden müssten, andererseits scheinen sie darüber hinaus wenig erfolgsversprechend: Da die Fehlerverläufe auf verschiedenen Ebenen sehr unterschiedlich sein können scheint es schwierig, genau den Cluster von Verläufen zu identifizieren, der aus fachlicher Perspektive als relevant erachtet würden. Darüber hinaus erschwert das unüberwachte Lernen das in Kapitel 2.5.1 geschilderte Vorgehen zur Berücksichtigung von Expert*innenwissen.

Für das überwachte Lernen sind, wie bereits erwähnt, gelabelte Daten notwendig (Frochte, 2018). Das Label, also der korrekte Ausgabewert, dient dem Modell beim Training als Feedback. Es gibt eine Vielzahl von Methoden des überwachten Lernens, die sich u.a. hinsichtlich ihrer Komplexität, Einsatzgebiete und notwendiger Inputdatenmengen unterscheiden. Im Folgenden werden einige Klassifikationsmethoden vorgestellt. Dabei wird insbesondere die Anwendbarkeit auf den vorliegenden Fall und die Interpretierbarkeit als wichtiges Kriterium betrachtet.

2.5.2.1 *k*-Nearest Neighbors

Der *k*-Nearest Neighbors Algorithmus ist eine sehr intuitive Methode: Ein neuer Datenpunkt wird dabei so klassifiziert, wie die Mehrheit der *k* Datenpunkte, die ihm am nächsten liegen (Harrison, 2021). Entscheidend ist dabei neben der Skalierung der Merkmale oder Features die Wahl der passenden Distanzmetrik (Frochte, 2018). Vorteile der *k*-Nearest Neighbors Klassifizierung sind ihre simple und parameterarme Implementierung. Als Nachteil kann in diesem Kontext vor allem betrachtet werden, dass bei vielen Features und relativ wenigen Datenpunkten ein großes Gewicht auf der korrekten Klassifizierung der (wenigen) vorhandenen Datenpunkte liegt, da diese die einzige vorhandene Referenz für die Klassifizierung bilden (Harrison, 2021). Die Interpretierbarkeit ist gegeben, da stets mit der Ähnlichkeit zu existierenden Datenpunkten argumentiert werden kann.

2.5.2.2 Entscheidungsbäume

Auch Entscheidungsbäume (engl. *decision trees*, DT) sind intuitiv gut verständlich. Harrison beschreibt sie wie folgt:

„Ein Entscheidungsbaum funktioniert etwa so wie ein Arzt, der Ihnen eine Reihe von Fragen stellt, um den Grund für Ihre Beschwerden herauszufinden.“
(Harrison, 2021, S. 89)

Für die Erstellung der Bäume gibt es mehrere Algorithmen, die grundsätzlich auf der Idee beruhen, die Teilmengen der Bäume bei jedem Entscheidungsknoten homogener zu machen (Frochte, 2018). Ein signifikanter Vorteil der Entscheidungsbäume ist ihre Interpretierbarkeit: Der Pfad, der für die finale Klassifizierung zuständig ist, kann einfach nachvollzogen werden. Weitere Vorteile sind die Fähigkeit, mit nichtlinearen Zusammenhängen umzugehen sowie die wenig aufwändige Datenaufbereitung (Harrison, 2021). Ein Nachteil ist, dass die Bäume zum *overfitting* neigen, also zum Überanpassen des Modells an den Trainingsdatensatz, anstatt übertragbare Muster zu erlernen (Géron, 2018).

Ein weiterer interessanter Aspekt ist, dass sich die in Kapitel 2.4.1 und 2.5.1 erwähnten Expert*innenmodelle oft als Entscheidungsbaum darstellen lassen. Im Fall, dass die Entscheidungsstruktur grundsätzlich fachlich begründet wird, aber die genauen Entscheidungswerte durch den Lernprozess bestimmt werden sollen, so kann auch eine Gittersuche (engl. *grid search*) eingesetzt werden. Das Vorgehen wird im maschinellen Lernen oft zum Finden der optimalen Hyperparameter eingesetzt. Dabei werden alle Kombinationen an Werten systematisch durchgegangen, die für bestimmte Parameter ausprobiert werden sollen (Géron, 2018). Da das Vorgehen eine Laufzeit von $O(n^d)$ hat, wobei n die Anzahl der möglichen Werte pro Parameter ist und d die Anzahl der Parameter, kann es nur für eine sehr begrenzte Anzahl an Parametern und Werten angewandt werden, garantiert für diese jedoch die optimale Lösung.

2.5.2.3 XGBoost

XGBoost steht für *extreme gradient boosting* und ist ein in der Praxis populärer Algorithmus des maschinellen Lernens (Chen & Guestrin, 2016). Dabei wird ein Ensemble von relativ kleinen Entscheidungsbäumen trainiert, deren Vorhersagen zu einem Gesamtergebnis zusammengefasst werden (Chen & Guestrin, 2016).

Zu den Vorteilen des XGBoost Algorithmus gehören neben seiner guten Performance in der Praxis (Chen & Guestrin, 2016; Géron, 2018), dass der Algorithmus relativ gut mit

kleinen Datenmengen arbeiten kann (Floares et al., 2017). Als Nachteil kann genannt werden, dass das Baumentsemble zwar auf Entscheidungsbäumen beruht, aber deutlich schwieriger zu interpretieren ist als der einfache Entscheidungsbaum.

2.5.2.4 Neuronale Netze

Künstliche Neuronale Netze sind für viele der Durchbrüche im Bereich maschinelles Lernen und künstliche Intelligenz in den letzten Jahren verantwortlich (Géron, 2018), nicht zuletzt auch im Bereich der Zeitreihenanalyse (Hirschle, 2021). Neuronale Netze können komplexe Beziehungen abbilden, benötigen dafür aber eine große Menge an Trainingsdaten (Géron, 2018). Außerdem ist das Ergebnis deutlich schwerer interpretierbar als bei den anderen vorgestellten Methoden (Ebrahimzadeh et al., 2013).

2.5.2.5 Problem der unausbalancierten Klassen

Darüber hinaus gibt es einige Betrachtungen und Methoden, die für alle der diskutierten Methoden von Bedeutung sind, wie das Problem der unausbalancierten Klassen. Bei vielen Klassifizierungsaufgaben ergibt sich das Problem, dass die interessante Klasse deutlich unterrepräsentiert ist. Man spricht dann von unausbalancierten oder unausgeglichenen Klassen (Harrison, 2021). Das kann zu Problemen führen, da die einfachste Möglichkeit für ein Modell eine hohe Vorhersagegenauigkeit zu erreichen ist, immer die Mehrheitsklasse vorherzusagen. Eine Möglichkeit, mit unausgeglichenen Klassen umzugehen, ist die Minderheitsklasse durch *upsampling* zu erweitern, also Datenpunkte der Minderheitsklasse mehrfach zu verwenden, und die Mehrheitsklasse durch *downsampling* zu verkleinern, also Datenpunkte aus dem Datensatz auszuschließen (Harrison, 2021). Ein Nachteil dieses Vorgehens ist jedoch, dass potenziell Informationen verloren gehen oder eine Verzerrung durch die Überbewertung einzelner Datenpunkte entsteht.

2.5.2.6 Preprocessing

Vor der eigentlichen Modellierungsaufgabe müssen die Daten für das maschinelle Lernen meist aufbereitet werden, das sogenannte *preprocessing* oder *cleaning*. Dazu gehört meist die Entscheidung, wie mit fehlenden Daten umgegangen werden soll, sowie die Normalisierung oder Standardisierung von Daten (Frochte, 2018). Außerdem werden kategoriale Daten von vielen Modellen nicht unterstützt, weshalb sie meist mithilfe des so genannten *one hot encodings* auf mehrere binäre Spalten verteilt werden (Frochte, 2018). Eine Spalte, die die Werte grün, gelb und blau enthält, könnte so auf drei Spalten

verteilt werden, die grün, gelb und blau heißen und jeweils als Wert eine eins oder eine null enthalten, je nachdem ob die Kategorie auf den Datenpunkt zutrifft oder nicht.

2.5.2.7 Metriken

Um die Performance eines Modells beurteilen zu können und Modelle untereinander zu vergleichen, braucht es Bewertungswerkzeuge. Ein intuitiver Ansatz ist die Genauigkeit (engl. *accuracy*), die aussagt, welcher Prozentsatz der Datenpunkte korrekt klassifiziert wurde (Harrison, 2021). Allerdings ist diese Metrik, wie bereits erwähnt, bei unausgeglichenen Klassen nicht unbedingt aussagekräftig. Insbesondere bei binären Klassen, also Klassen die nur zwei Ausprägungen haben, lohnt sich eine Betrachtung der Wahrheitsmatrix (engl. *confusion matrix*): Diese unterteilt die Klassifikationsergebnisse in falsch positiv (FP), richtig positiv (TP), falsch negativ (FN) und richtig negativ (TN) (Harrison, 2021). Je nach Anwendungsgebiet kann es besonders relevant sein, eine niedrige falsch negativ oder eine niedrige falsch positiv Rate zu haben – man denke z. B. an Anwendungen im Bereich der Medizin. Um diesen speziellen Anforderungen Genüge zu tun kann das F-Maß genutzt werden, das den Mittelwert aus Präzision (engl. *precision*) und Trefferquote (engl. *recall*) berechnet (Harrison, 2021):

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F_1 = 2 \frac{precision * recall}{precision + recall}$$

Eine Verallgemeinerung des F-Maßes, das auch F_1 -Maß genannt wird, ist das F_β -Maß. Dabei werden Präzision und Trefferquote mithilfe des β -Parameters unterschiedlich gewichtet, wobei ein β größer als 1 die Trefferquote bevorzugt und ein β kleiner als 1 die Präzision (Pedregosa et al., 2011).

2.5.2.8 Validierung

Um zu evaluieren, wie gut ein Modell für die Klassifizierung unbekannter Daten – also Daten, die nicht im Trainingsdatensatz enthalten sind – geeignet ist, ist es üblich, den Datensatz in Trainings- und Testdaten zu unterteilen (Géron, 2018). Die eigentliche Bewertung des Modells erfolgt dann auf Grundlage der Testdaten. Für kleine Datenmengen kann es jedoch passieren, dass einzelne Sonderfälle den Trainings- oder Testdatensatz verzerren und so zu nicht repräsentativen Ergebnissen führen. Daher führt

man in diesen Fällen oft eine so genannte Kreuzvalidierung (engl. *k*-fold cross validation, CV) durch. Bei dieser Methode wird der Datensatz *k* Mal in einen Trainings- und Testdatensatz unterteilt und anhand dieser trainiert und evaluiert (Frochte, 2018). Dann können Mittelwert und Standardabweichung dieser *k* Durchläufe für die verschiedenen Modelle oder Varianten verglichen werden.

2.6 Zusammenfassung

Das Ziel eines Trend Monitorings ist, Probleme systematisch und frühzeitig zu erkennen (vgl. Kapitel 3.2). Damit ist klar, dass die Definition eines Trends nah an der aus dem Bereich SPC sein muss, also Trend als ein unnatürliches Muster. Welche Muster dabei genau gemeint sind – ob monotone Trends, *shifts* oder auch Ausreißer eingeschlossen werden, unterliegt dabei einem gewissen Interpretationsspielraum.

Interessant ist dabei, dass die Frage nach einem relevanten Trend eigentlich eine kausale Frage ist, also ob das beobachtete Muster eine kausale Ursache hat, die es zu beheben gilt, oder ob es sich um zufällige Schwankungen handelt. Diese Frage ist jedoch schwierig zu beantworten, da in einer univariaten Zeitreihe keine weiteren Rückschlüsse auf Ursachen zu finden sind. Eine Möglichkeit wäre, die Zeitreihen mit Daten zu bekannten Ursachen zu korrelieren, z. B. aus CAPA-Prozessen. Das birgt jedoch das Risiko, dass nicht alle Ursachen erkannt, erfasst und dokumentiert wurden, und somit eine verzerrte Datenlage als Grundwahrheit vorliegt.

Stattdessen kann das Problem als Mustererkennungsproblem formuliert werden, wie es z. B. auch in der SPC getan wird. Eine Möglichkeit wäre, die aus der SPC bekannten Muster zu übernehmen und sich auf die zu konzentrieren, die im Rahmen des Fehlertrendmonitorings relevant sind. Dagegen sprechen jedoch die in Kapitel 2.4.1 aufgeführten Gründe. Stattdessen können die aus dem Bereich der SPC und anderen Bereichen der Zeitreihenanalyse bekannten Methoden genutzt werden, um ein für den konkreten Anwendungsfall passgenaues Modell zu entwickeln. Dabei wird als Bewertungsmaßstab die menschliche Fachexpertise herangezogen, was sich vor allem daher anbietet, da im Normalfall eine menschliche Fachperson Nutzer*in des Trendmonitorings sein wird, das Ergebnis sollte also ihren Bedürfnissen angepasst sein. Somit kann das Mustererkennungsproblem in ein Klassifizierungsproblem umgewandelt werden, in dem die Herausforderung ist, die Verläufe so zu klassifizieren wie eine fachliche Expertin sie einordnen würde. Während dies einige Vorteile in der Operationalisierung mit sich bringt, nämlich unter anderem, dass die in Kapitel 2.5.2 vorgestellten Methoden des maschinellen Lernens verwendet werden können, so birgt der Ansatz auch das Risiko der subjektiven Priorisierung und Bewertung.

Als Zusammenfassung der vorhergehenden theoretischen Überlegungen kann also grundlegend das folgende Vorgehen zur Entwicklung eines automatisierten Fehlerrendmonitorings festgehalten werden:

1. Bestimmen der Input-Zeitreihen
2. Bestimmen und Berechnen der relevanten Features
3. Vornehmen einer initialen Klassifizierung
4. Erheben von Feedback (Labeling)
5. Optimieren der Klassifizierung
6. Implementierung des Monitorings
7. Fortlaufende Anpassung

Welche der in diesem Kapitel diskutierten Features und darüber hinaus genutzt werden, und wie genau der initiale Klassifizierungsalgorithmus gestaltet ist, kann dabei auf die genauen Bedürfnisse des Anwendungsfalls angepasst werden. Das Ermitteln dieser Anforderungen ist Inhalt des nächsten Kapitels.

3 Anforderungen des konkreten Anwendungsfalls

Die bisherigen theoretischen Betrachtungen sind so allgemein gehalten, dass sie für die meisten Produktionsunternehmen, die ein Qualitätsmanagement nach ISO 9001 (oder vergleichbar) betreiben, zutreffend sind. Im Folgenden werden die theoretischen Betrachtungen auf den konkreten Kontext des betrachteten Unternehmens übertragen. Es handelt sich dabei um ein Unternehmen, das verschiedene komplexe und qualitativ hochwertige Produkte und Komponenten von Produkten herstellt und dem Qualitätsmanagement einen sehr hohen Stellenwert zumisst. Aufgrund der Anforderungen an die betriebliche Geheimhaltung wurden die konkreten Informationen zum Unternehmen jedoch weitestgehend anonymisiert.

In diesem Kapitel wird die durchgeführte Anforderungserhebung beschrieben. Dadurch werden die Rahmenbedingungen für die in Kapitel 4 folgende Umsetzung geklärt. Insbesondere der erste der in Kapitel 2.6 definierten Vorgehensschritte – die Bestimmung der Inputdaten – wird dabei abschließend geklärt, während für die restlichen Punkte notwendiger Input erfasst wird.

3.1 Methodik

Im Rahmen der Anforderungserhebung wurden verschiedene Methoden eingesetzt, die an die unterschiedlichen Fragestellungen angepasst wurden. Es wurden thematisch vier Bereiche abgefragt:

1. Ziele des Unternehmens und Spezifizierung des konkreten Anwendungsfalls
2. Definition der Inputdaten
3. Wissensextraktion zum Thema Trend
4. Anforderungen an die Implementierung

Bereich 1 dient dabei vor allem dazu, die Motivation, Ziele und Zielgruppen des konkreten Anwendungsfalls zu verstehen. Dazu wurde ein semistrukturiertes Interview (Berg & Lune, 2011) mit einem Vertreter der mittleren Führungsebene im Bereich Qualitätsmanagement geführt, transkribiert, kodiert und ausgewertet³. Das Format der semistrukturierten Interviews wurde gewählt, da sie einerseits zu einem gewissen Grad ein Anpassen an den Verlauf des Gesprächs zulassen und darüber hinaus eine natürlichere Gesprächsatmosphäre bieten, aber andererseits ein methodisches Vorgehen sicherstellen. Dasselbe Vorgehen wurde für den Bereich Inputdaten gewählt. Die so definierten

³ Interviewleitfäden und Transkripte mit Kodierung finden sich in Anhang.

Inputdaten wurden genutzt, um ein Set an Beispielverläufen zu generieren. Mit Hilfe dieser Beispielverläufe wurde dann im Rahmen der Wissensextraktion (vgl. Kapitel 2.5.1) konkrete Entscheidungen qualitativ abgefragt, also Expert*innen gefragt, wie sie einen konkreten Beispielverlauf bewerten würden und wieso. Dazu wurden drei Fachleute befragt, die verschiedene Ebenen und Fachkompetenzen im Bereich Qualitätssicherung des Unternehmens abdecken: Ein Vertreter der mittleren Führungsebene, ein Mitarbeiter der Qualitätssicherung mit eher technischem Hintergrund und ein Mitarbeiter der Qualitätssicherung mit eher fachlichem Hintergrund.

Auch die im Rahmen der beispielbasierten Wissensextraktion gewonnenen Aussagen wurden transkribiert, kodiert und ausgewertet. Unter den so identifizierten Themen fanden sich einige, die einer weiteren Vertiefung bedurften. Also wurden darauf aufbauend noch zwei semistrukturierte Interviews zum Thema *Trend* geführt. Für diese wurden zwei der Interviewpartner*innen befragt, die auch an der beispielbasierten Wissensextraktion teilgenommen hatten.

Die Methodik für die Anforderungserhebung im Bereich Implementierung weicht etwas von den anderen thematischen Blöcken ab. Es wurden, aufbauend auf den Ergebnissen von Kapitel 4, Brainstorming-Methoden angewandt und drei verschiedene Optionen mit ihren Vor- und Nachteilen diskutiert. Hier wurden die Interviewpartner*innen nicht einzeln, sondern in einem Workshopformat interviewt, um im Rahmen der Diskussion zu möglichst konsensualen Ergebnissen zu gelangen. Auch dabei wurde versucht eine Balance aus Teilnehmer*innen der verschiedenen Stakeholdergruppen zu erreichen, inklusive Management, operative Nutzer*innen und technische Verantwortliche. Die Ergebnisse und Leitfäden aller Interviews finden sich im Anhang.

3.2 Ziele des Unternehmens

Die Ziele, die ein Unternehmen mit der Einführung eines Trend Monitorings verfolgt, werden sich in vielen Fällen ähneln, allerdings ist auch hier bereits ein gewisser Spielraum bei den Prioritäten zu erwarten. Vergleichbar zu den in Kapitel 2.4.1 diskutierten Ansätzen der CCPR kann z. B. der Grad, zu dem die Automatisierung zur Unterstützung oder als Ersatz der menschlichen visuellen Kontrolle dienen soll, variieren. Die Ziele zu verstehen ist jedoch Grundvoraussetzung, um sie adäquat erfüllen zu können.

Im Interview zu diesem Thema kristallisierten sich zwei Hauptziele heraus, die unter *verbesserte Abdeckung* und *höhere Geschwindigkeit* zusammengefasst werden können. Der Interviewpartner schilderte die vorhandenen Kontrollmechanismen und -gremien zur Identifizierung von relevanten Fehlerverläufen. Aufgrund der großen Gesamtzahl an

möglichen zu überwachenden Verläufen wird sich dabei auf die Fehler konzentriert, die besonders hohe Kosten verursachen, besonders häufig auftreten oder anderweitig als besonders kritisch beurteilt werden, die so genannten *Top-Fehler*. Das Ziel des automatisierten Monitorings ist daher, auch die Fehler zu identifizieren, die noch nicht zu diesen Top-Fehlern gehören, aber aufgrund ihres Trendverlaufs sich potenziell in diese Richtung bewegen. Der Interviewpartner bezeichnete dies als „ansteigenden Bodensatz“ (Interview „Ziele“ Person 1, vgl. Anhang B).

Damit einher geht das zweite Ziel, nämlich ein Zugewinn an Geschwindigkeit. Damit ist insbesondere die Geschwindigkeit gemeint, mit der relevante Fehler identifiziert und Maßnahmen ergriffen werden, bevor sie zu den Top-Fehlern gehören. Der Interviewpartner formuliert dies wie folgt:

„Das heißt, im Endeffekt sagen wir, der große Use Case an dem ganzen Thema ist eben auch, diese [schleichenden] Themen zu identifizieren, die Themen frühzeitig zu identifizieren und damit Qualitätskosten zu sparen oder Kosten von Nicht-Qualität zu sparen, indem man rechtzeitig Maßnahmen einleitet.“ (Interview „Ziele“ Person 1, vgl. Anhang B)

Relevant sind außerdem die Zielgruppen, für die ein solches Tool entwickelt werden soll. In diesem Fall ist dies insbesondere die operative Qualitätssicherung, also die Personen, die auf operativer Ebene für die Qualitätssicherung in bestimmten Produktionsbereichen zuständig sind. Für sie sollen auf täglicher Basis die Ergebnisse des Monitorings so zur Verfügung gestellt werden, dass sie visuell, mit einem möglichst geringen Zeitaufwand und mit einer guten Interaktion mit anderen Tools gesichtet werden können.

3.3 Inputdaten

In Kapitel 2.1.2 wurden einige Annahmen über Fehlerverläufe als Zeitreihen getroffen, die jedoch großen Spielraum für die konkrete Ausgestaltung der Input-Zeitreihen in ein Monitoringsystem lassen. Daher wurden im Interview zum Thema Inputdaten die Erwartungen des Unternehmens in dieser Hinsicht abgefragt. Der Fokus lag dabei insbesondere auf den zu betrachtenden Zeitraum sowie den Aggregationsstufen, die wie in Kapitel 2.1.2 argumentiert eine große Rolle für die Verteilung der Daten spielen.

Das Nonconformity-Management im betrachteten Unternehmen arbeitet mit dem in Kapitel 2.1.1 erwähnten Nonconformance-Report, der daher als Datenquelle für das Tool dient. Tabelle 3.1 fasst die im Interview erhobenen Anforderungen an die Inputdaten kurz zusammen.

Frequenz	Täglich
Zeitraum	Rückschau von ca. 2 Monaten bzw. 60 Tagen
Aggregationsebene	Fehler pro Produkt
Ebenen	Betrachten von zwei getrennten Ebenen des NCR als Fehler
Absolut oder Relativ	Absolut
Summe oder Anzahl	Betrachten der Rückweismenge (Summe)

Tabelle 3.1: Anforderungen an die Inputdaten

Das führt zu einer Inputmenge an Fehlerverläufen im mittleren dreistelligen bis niedrigen vierstelligen Bereich. Da die Produkte sich in der produzierten Menge, und die Fehler sich in ihrer Häufigkeit sehr stark unterscheiden, sind die Verläufe untereinander sehr heterogen, wobei ein großer Anteil der Verläufe als *zero inflated* bezeichnet werden kann. Betrachtet man die Verteilung der Verläufe kategorisiert nach den in Kapitel 2.1.2 vorgestellten *demand patterns*, so ergibt sich das folgende Bild für die Inputzeitreihen eines Tages:

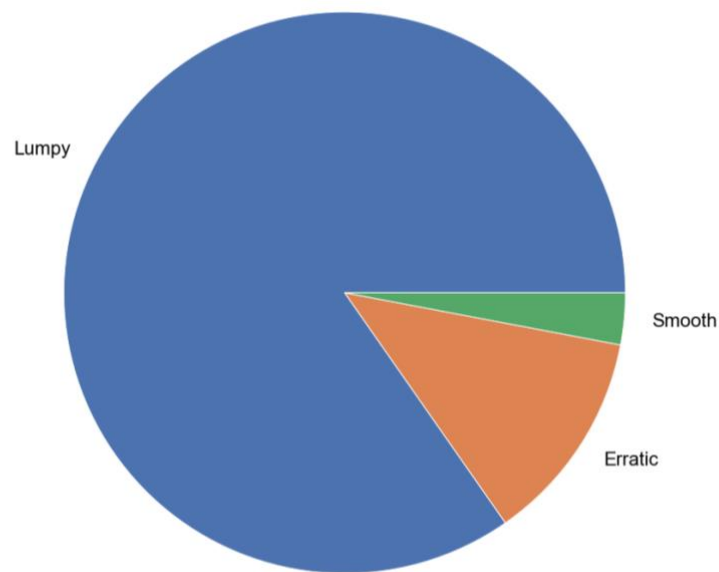


Abbildung 3.1: Verteilung der Daten nach Kategorie (eigene Darstellung)

Die dominierende Kategorie ist *lumpy*, die als schwierig vorherzusagen betrachtet wird. Nur ein kleiner Teil der Fehlerverläufe ist *smooth*, also regelmäßig was sowohl den Abstand zwischen Fehlerereignissen als auch die Varianz angeht. Die Kategorie *intermitted* kommt hingegen überhaupt nicht vor. Zu untersuchen, wie sich das Modell der *demand patterns* anpassen ließe, um besser auf die Fehlerdaten übertragbar zu sein, könnte ein interessanter Ansatz für weiterführende Arbeiten sein.

3.4 Wissensextraktion

Ziel der Wissensextraktion war es, die Intuition und die Fachexpertise der Expert*innen zu extrahieren und somit die Formalisierung in einem regelbasierten System vorzubereiten. Für die beispielbasierte Wissensextraktion wurden Trendverläufe ausgewählt, die eine möglichst große Bandbreite im potenziell relevanten Bereich abdecken. Das bedeutet, dass solche Verläufe, die eine klare Tendenz nach unten aufweisen, nicht ausgewählt wurden, da sie den geringsten Erkenntnisgewinn versprachen. Die Verläufe wurden als Liniendiagramme mit dem Datum auf der x-Achse und der Rückweismenge auf der y-Achse dargestellt. Außerdem wurde die Trendkomponente einer STL-Dekomposition mit einer wöchentlichen Saisonalität (vgl. Kapitel 2.2.1) dargestellt, um die visuellen Informationen zum Trend anzureichern und die Interpretation zu erleichtern (Cleveland, 1979). Die Darstellung wurde analog zu der in der späteren Feedback-Phase gewählt (vgl. Kapitel 4.4), um die Vergleichbarkeit und Einheitlichkeit zu gewährleisten. Ein Beispiel einer solchen Darstellung zeigt Abbildung 3.2⁴. Den Befragten wurden bewusst keine weiteren Informationen, wie Art des Fehlers oder Art des Produkts gegeben, um den Fokus auf die Form bzw. das Muster des Verlaufs zu lenken.

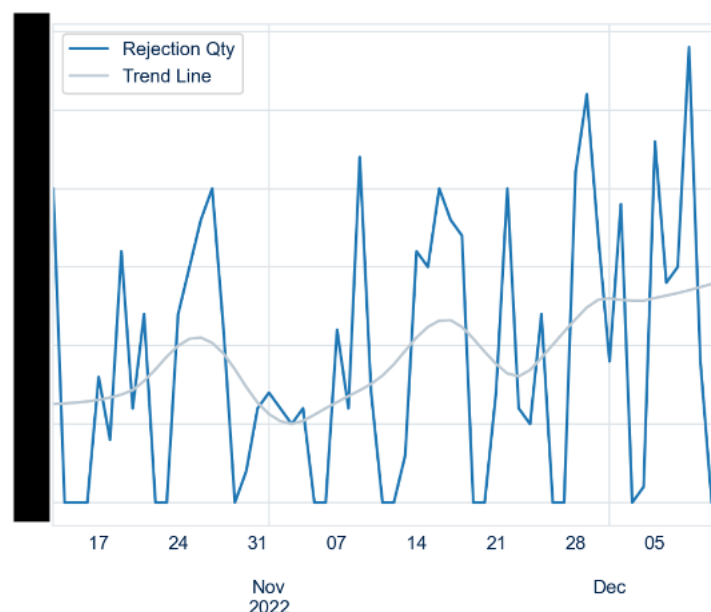


Abbildung 3.2: Ein in der beispielbasierten Wissensextraktion genutzter Verlauf (eigene Darstellung)

Bevor die einzelnen Ergebnisse vorgestellt werden, vorab einige allgemeine Beobachtungen: Innerhalb der einzelnen Interviews konnte beobachtet werden, wie die

⁴ Um den Anforderungen an die Geheimhaltungspflichten genüge zu leisten wurden die absoluten Fehlerzahlen geschwärzt.

Argumentationsmuster stabiler wurden, was sich auch darin äußerte, dass die Personen auf vorherige Beispiele verwiesen. Das zeigt, dass diese Art der Begründung für die Expert*innen eine ungewohnte Aufgabe war, an die sie sich jedoch schnell anpassten. Außerdem wurde bereits bei der kleinen Zahl an Befragten deutlich, dass sich die Bewertungslogiken deutlich unterschieden und die Interviewpartner*innen zum Teil unterschiedliche Muster als relevant beurteilten. Darüber hinaus fiel auf, dass die Expert*innen tatsächlich meist im Einklang mit der kausalen Definition von Trend argumentierten, also ihre Bewertung der Verläufe davon abhängig machten, ob sie von einer Fehlerursache oder einer zufälligen Schwankung ausgingen. Eine weitere interessante Beobachtung ist, dass die Befragten oft prognostisch argumentierten, also ihre Bewertung davon abhängig machten, wie sie den weiteren Verlauf des Fehlers vermuten würden.

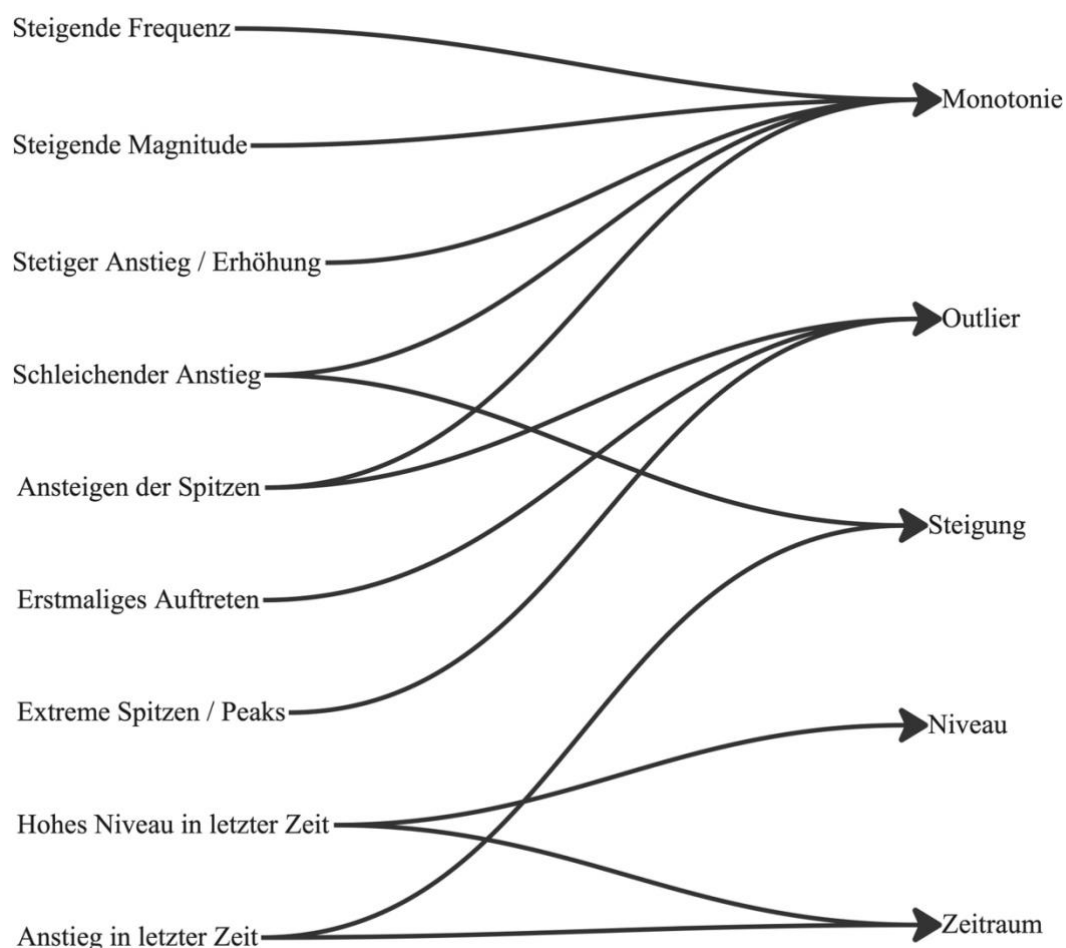


Abbildung 3.3: Charakteristika von als interessant bewerteten Verläufen (eigene Darstellung)

Abbildung 3.3 zeigt Charakteristika oder Themen, die oft im Zusammenhang mit als interessant bewerteten Verläufen erwähnt wurden, gruppiert zu Oberthemen.

Abbildung 3.4 zeigt Charakteristika oder Themen, die oft im Zusammenhang mit als uninteressant bewerteten Verläufen erwähnt wurden, gruppiert zu Oberthemen. Darüber hinaus wurden folgende Charakteristika oder Themen oft im Zusammenhang mit Verläufen erwähnt, bei denen die Expert*innen unentschlossen waren:

- Vermutete Einflüsse der Produktionsmenge
- Wiederholte Spitzen oder extreme Ausschläge

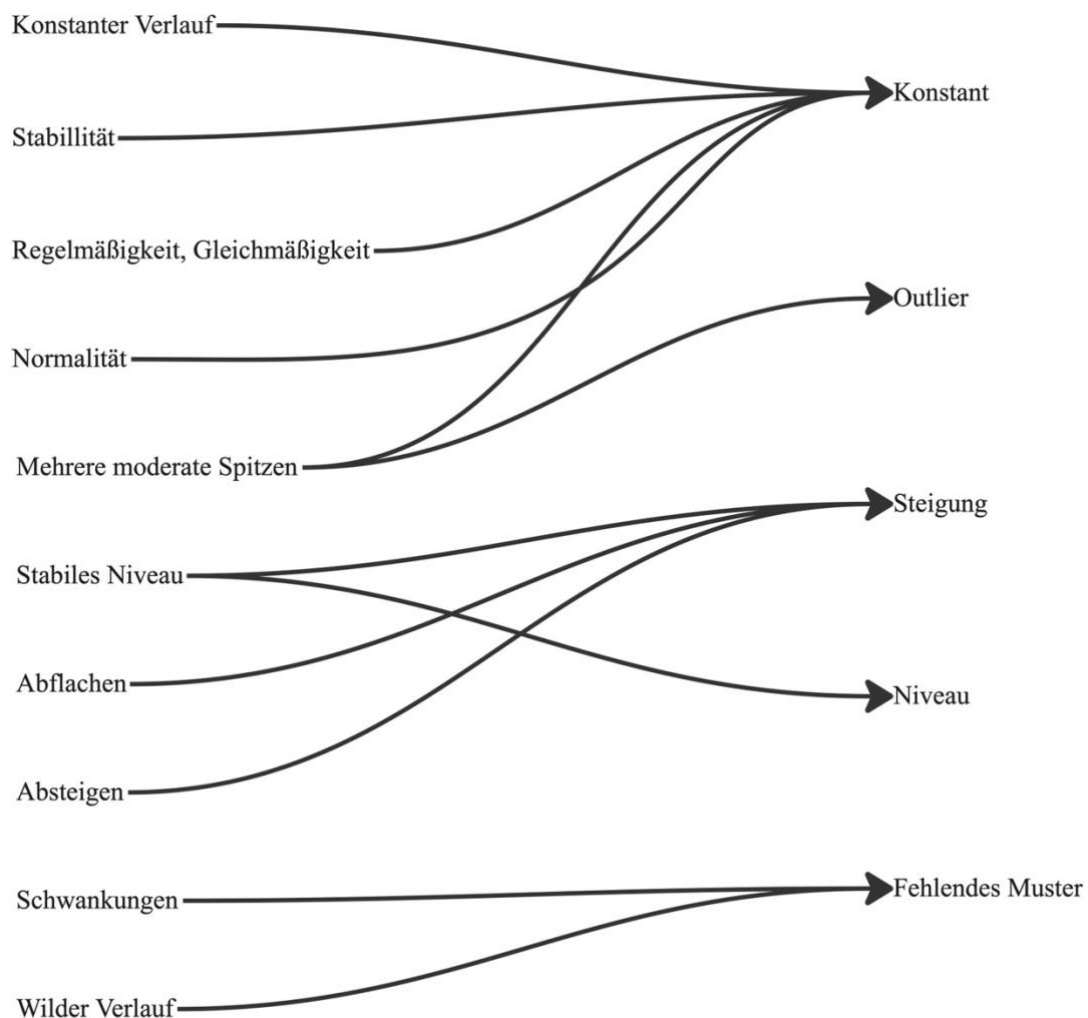


Abbildung 3.4: Charakteristika von als uninteressant bewerteten Verläufen (eigene Darstellung)

Das folgende Zitat zeigt beispielhaft für den in Abbildung 3.2 gezeigten Fehlerverlauf ein Argumentationsmuster:

„Auch den würde ich mir in der Tat anzeigen lassen. Da ich hier vom Trendverlauf her, der hat zwar ein relativ deutliches auf und ab, aber wenn ich mir die Spitzen anschau, steig ich über die letzten 4 Wochen doch konstant immer um ein paar

Stücke an, so dass es durchaus sein könnte, dass wir hier noch weiter nach oben gehen.“

(Interview „Wissensextraktion“ Person 1, Beispiel 22, vgl. Anhang B)

Es wird deutlich, wie zwischen den Schwankungen als ein Indikator für einen eher zufälligen Verlauf, und der monotonen Aufwärtsbewegung als Indikator für einen ursachenbedingten Verlauf abgewogen wird, und die Monotonie überwiegt. Auch die Rolle der Prognose ist hier gut zu erkennen.

Für die folgenden semistrukturierten Interviews im Rahmen der Wissensextraktion wurden insbesondere die Themen Ausreißer und ihre Bedeutung und Identifikation der für die Beurteilung herangezogene Zeitraum, die relevanten Muster und die Anforderungen an die Sensitivität und Spezifität des Tools identifiziert. Die Ergebnisse werden in Tabelle 3.2 kurz zusammengefasst.

Muster	Stetige Anstiege von allen Befragten als relevant betrachtet Uneinigkeit bei Ausreißern oder <i>single events</i> Außerdem identifiziert: U-Muster (relevant), Shift (relevant)
Zeitraum	Flexible Betrachtungsweise Letzte Woche besonders relevant
Ausreißer	Unterschiedlich bewertet in Relevanz In Relation zu Mittelwert fünf- bis zehnfach für Relevanz
Sensitivität und Spezifität	Schwierig zu definieren Verhältnismäßigkeit und akzeptabler Umfang wichtig für Erhalt des Vertrauens Ratio von Richtig Positiv zu Falsch Positiv von 50-70% zu 50-30% akzeptabel

Tabelle 3.2: Interviewergebnisse der vertieften Wissensextraktion

3.5 Implementierung des Ergebnisses

Der Workshop zu den Anforderungen an die Implementierung des finalen Tools wurde nach den in Kapitel 4 beschriebenen Schritte durchgeführt, so dass die Ergebnisse bereits vorlagen. Grundsätzlich wurden jedoch vor allem verschiedene Möglichkeiten für die Gestaltung der Nutzeroberfläche und die damit verbundenen Technologien diskutiert. Als besonders wichtig wurde eine gute Integration in die bestehende Landschaft an Analysetools, und hier insbesondere die existierenden Business Intelligence (BI)-Tools, bewertet. Davon wurde sich sowohl ein geringerer Wartungs- und Schulungsaufwand als auch eine größere Akzeptanz der Nutzer*innen versprochen, die nicht noch ein zusätzliches Tool erlernen und nutzen müssen. Die Vorteile überwogen aus Sicht der Teilnehmenden am Workshop den Nachteil, dass Feedback über die Dashboard-Oberfläche schwieriger zu erheben ist als über ein eigenes Frontend. Als Ziel der Implementierungsphase wurde definiert, die Ergebnisse des Trend Monitorings so im

Data Warehouse des Unternehmens zur Verfügung zu stellen, dass sie für Dashboards im BI-Tool zur Verfügung stehen. Die Integration in diese Dashboards sollte dann von den Verantwortlichen übernommen werden und gehört nicht mehr zum Umfang dieser Arbeit.

4 Umsetzung

Nachdem in Kapitel 3 die Anforderungen im Rahmen des konkreten Fallbeispiels erfasst wurden, wird nun die Umsetzung vorgestellt. Sie leitet sich somit aus den in Kapitel 2 beschriebenen theoretischen Grundlagen angewandt auf den konkreten Anwendungsfall ab. Abbildung 4.1 zeigt nochmal schematisch das in Kapitel 2.6 hergeleitete Vorgehen. Die Aggregation der Inputdaten wurde bereits in Kapitel 3.3 abschließend geklärt, weshalb hier nicht nochmal gesondert auf sie eingegangen wird.

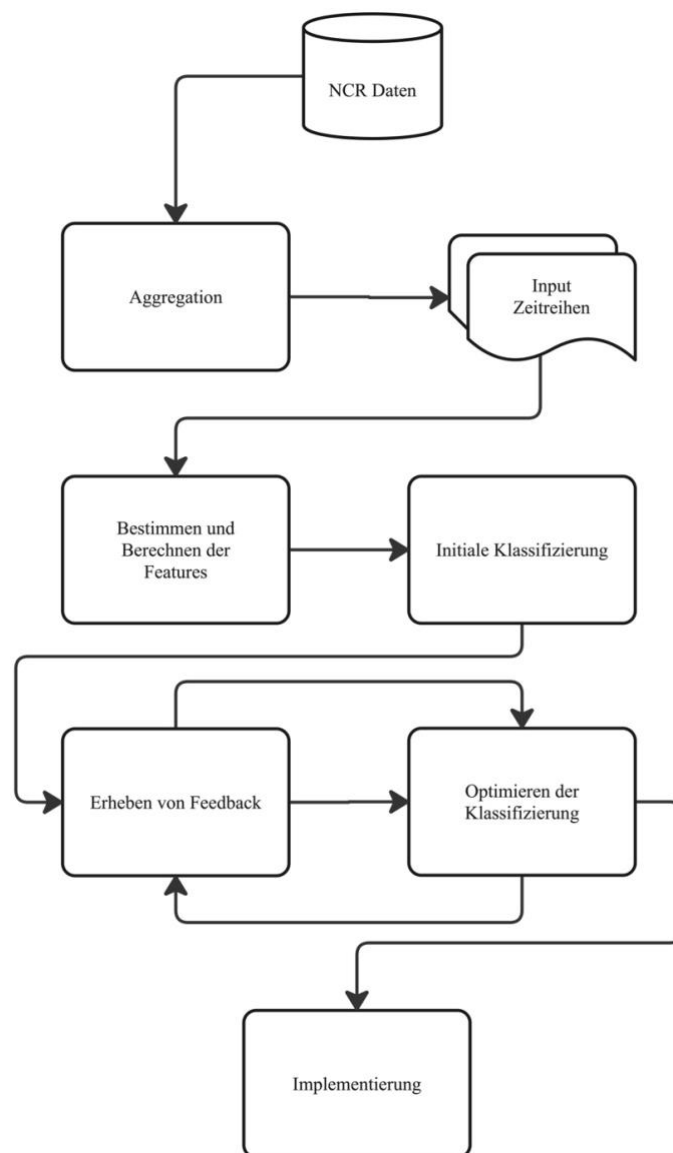


Abbildung 4.1: Vorgehensdiagramm zur Umsetzung (eigene Darstellung)

4.1 Bestimmen und Berechnen der Features

In der Wissensextraktion in Kapitel 3.4 wurden Aussagen und Themen gesammelt, die die Befragten im Zusammenhang mit aus ihrer Sicht interessanten oder uninteressanten

Codes getroffen haben. Das Ziel ist nun einen Satz von Features zu finden, die in der Lage sind, diese Themen möglichst gut zu beschreiben. Dabei ist nicht unbedingt der Anspruch, sofort einen fertigen Regelsatz im Sinne eines klassischen *expert systems* zu definieren, da die finalen Regeln ja auf Grundlage des Feedbacks gefunden bzw. gelernt werden sollen.

Die im Rahmen der Wissensextraktion ermittelten Themen zur Beurteilung der Trends wie dargestellt in Abbildung 3.3 und Abbildung 3.4 sind Monotonie, Outlier, Steigung und Niveau. Das Thema Zeitraum ist ein Querschnittsthema und muss bei allen Aspekten betrachtet werden. Tabelle 4.1 zeigt die Zuordnung der genutzten Features zu den identifizierten Themen, bevor diese im Detail beschrieben werden.

Thema	Features
Monotonie	Mann-Kendall Test (p-Wert)
Outlier	Verhältnis zum Mittelwert Exkludierbare Extremwerte
Steigung	Steigung in den letzten n Tagen
Niveau	Niveau in den letzten n Tagen

Tabelle 4.1: Zuordnung der Features zu Themen der Wissensextraktion

Darüber hinaus wurden einige der klassischen Features (vgl. Tabelle 2.1 und Tabelle 2.2) zur Kontrolle erfasst, wie z. B. der Mittelwert und die Varianz. Ein weiteres wichtiges Feature, das sich eher aus der Literatur als der Anforderungserhebung herleitet, ist das *demand pattern* (vgl. Kapitel 2.1.2) des entsprechenden Verlaufs.

4.1.1 Mann-Kendall Test (p-Wert)

Wie in Kapitel 2.2.1 argumentiert ist der Mann-Kendall Test ein in der Praxis bewährter Test auf Monotonie, der sich in diesem Fall besonders gut eignet, da er nicht-parametrisch ist und damit für die diversen und zum Teil *zero inflated* Inputzeitreihen eignet. Da die Daten aufgrund der Produktionstage eine wöchentliche Saisonalität aufweisen – an den Wochenenden wird im betrachteten Unternehmen deutlich weniger produziert – wurde eine Variante des Mann-Kendall Tests verwendet, der solche saisonale Effekte berücksichtigt.

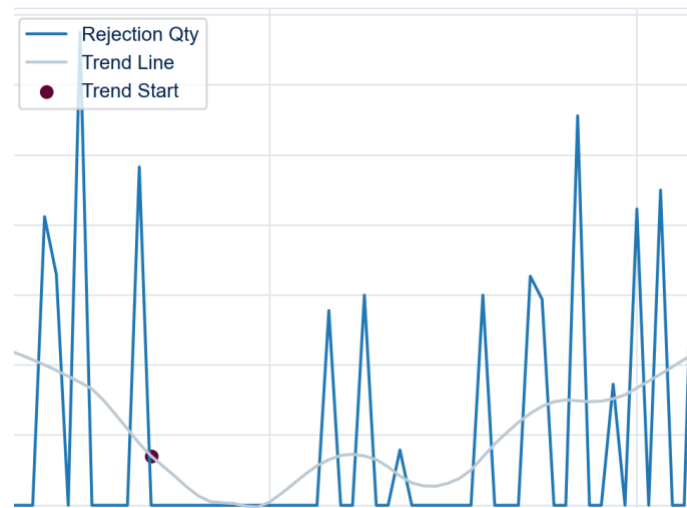


Abbildung 4.2: Beispiel eines Verlaufs mit teilweiser Monotonie (eigene Darstellung)

Es stellte sich dabei jedoch die Frage nach den betrachteten Zeiträumen: Bei einer visuellen Beurteilung eines Trendverlaufs ist man als menschliche*r Betrachter*in flexibel darin, einen gewissen Teil des Verlaufs auszublenden, z. B. einen Zeitraum zu Beginn der Serie, der eine absteigende Tendenz hat. Der statistische Test würde dies hingegen bei der Einschätzung der Monotonie negativ einbeziehen. Abbildung 4.2 zeigt ein solches Beispiel, bei dem bei einer Betrachtung ab einem bestimmten Zeitpunkt (hier markiert mit Trend Start) der Mann-Kendall Test eine Signifikanz bestätigen würde.

Um diese Flexibilität in der Auswahl des betrachteten Zeitraums nachzubilden, wurde der Mann-Kendall Test auf alle möglichen Teilsereien der ursprünglichen Zeitreihe vom aktuellen Zeitpunkt betrachtet angewandt. Die p-Werte aller Tests wurden erfasst und der niedrigste als finaler Wert des Features notiert. Der folgende Codeausschnitt verdeutlicht das Vorgehen.

```
def get_mk_overview(ts, p_thresh = 0.05):
    days_iter = reversed(range(1, ts.index.size+1))
    ps = []
    for ndays in days_iter:
        try:
            trend, p, _ = get_mk(ts, ndays)
            if trend != 'decreasing':
                ps.append(p)
            else:
                ps.append(1)
        except ZeroDivisionError:
            ps.append(1)
    return ps
```

Ein Nachteil dieses Vorgehens ist, dass durch das mehrfache Testen die Aussagekraft des p-Werts sinkt (Farcomeni, 2008). Dem kann dadurch entgegengewirkt werden, dass bei

der späteren Modellbildung der p-Wert, der als Signifikanzgrenze genutzt wird, nach unten korrigiert wird.

4.1.2 Verhältnis zum Mittelwert

Um zu quantifizieren, wie sich der aktuelle Wert der Serie im Vergleich zum Mittelwert verhält, wird der letzte Wert durch den Mittelwert der Zeitreihe ohne diesen letzten Wert geteilt. Ist der Mittelwert der restlichen Serie null, also tritt der Fehler zum ersten Mal im Untersuchungszeitraum auf, wird er stattdessen durch eins geteilt. Das hat den Vorteil, dass der Wert dann leicht interpretierbar ist: Tritt der Fehler zum ersten Mal auf, entspricht der Wert des Features einfach der Rückweismenge. So kann auch gut eine Mindestgrenze definiert werden, die für zum ersten Mal auftretende Fehler gewählt werden muss, um sie als relevant zu kennzeichnen.

Eine Alternative wäre, die Abweichung vom Mittelwert in Standardabweichungen zu betrachten, was für Reihen mit mehreren Spitzen weniger große Werte für das Feature bewirken würde als für solche mit erstmaliger Spitze. Allerdings hätte dies den Nachteil, dass bei erstmalig auftretenden Fehlern der Wert des Features immer gleich wäre, weshalb die andere Variante bevorzugt wurde.

Fachlich betrachtet muss sich ein Outlier oder extremes Event nicht unbedingt auf einen Tag beschränken. So kann die Ursache, die das extreme Event verursacht, wie z. B. eine defekte Maschine, mehrere Tage defekt sein. In den in Kapitel 3.4 erhobenen Definitionen der verschiedenen relevanten Muster wurde die Unterscheidung zwischen *single events* und stetigen Trends eher dadurch begründet, wie leicht die Probleme und ihre Ursachen zu identifizieren sind, als dadurch, dass sie auf einen einzigen Tag beschränkt sind. Daher wurde das Verhältnis zum Mittelwert auf drei Features gestreckt, also der Zeitraum, für den ein Event als Outlier betrachtet wird auf drei Tage erhöht. Die Berechnung erfolgt wie oben geschildert für den Mittelwert der letzten ein bis drei Tage.

4.1.3 Exkludierbare Extremwerte

Das Feature zu den exkludierbaren Extremwerten beruht auf der Überlegung, dass ein sehr hoher Extremwert im hinteren Teil der Serie den Mittelwert in diesem Bereich stark erhöht, obwohl es sich dabei eher um ein extremes Event als einen stetigen Trend handeln würde. Dabei ist wichtig zu bemerken, dass der Mann-Kendall Test sehr robust gegenüber solchen extremen Ausschlägen ist, da er nur auf der Rangfolge der Werte beruht. Aber andere Features wie die Steigung oder das Niveau sind anfällig gegenüber solchen Effekten. Daher wurde für dieses Feature ermittelt, wie oft der Maximalwert durch einen

anderen Wert, wie z. B. den Mittelwert oder den Median der Reihe, ersetzt werden kann, bevor die Zeitreihe über die Signifikanzgrenze des Mann-Kendall Tests fällt. Je höher der Wert, desto robuster ist der Trend. Allerdings ist dieses Feature mit einigen Problemen behaftet, angefangen mit der Frage, welcher Ersatzwert der richtige ist. Außerdem tritt insbesondere für Reihen mit vielen Nullwerten leicht das Problem auf, dass der Wert des Features ins Unendliche geht, da der Ersatzwert immer wieder durch einen anderen Wert größer Null ersetzt wird, für den die Teststatistik jedoch weiterhin eine Relevanz ausgibt.

4.1.4 Steigung in den letzten n Tagen

Das Feature betrachtet einen festgelegten Subzeitraum der gesamten Zeitreihe, dessen Muster bzw. Richtung als besonders relevant betrachtet wird. In der Anforderungserhebung wurde dieser Wert auf ca. eine Woche beziffert. Dies ist sinnvoll, da ein Fehlertrend in einem längeren Zeitraum signifikant am Steigen sein kann, aber in der letzten Woche z. B. stark gefallen sein kann. Dabei läge aus fachlicher Sicht die Vermutung nahe, dass die Fehlerursache behoben oder unter Kontrolle ist und der Verlauf somit nicht mehr relevant.

Um die Steigung der letzten n Tage konkret zu berechnen gibt es verschiedene Optionen, die aber alle darauf beruhen, eine lineare Regressionsgerade nach der Methode der kleinsten Quadrate durch die Datenpunkte zu legen und deren Steigung abzulesen. Die erste Möglichkeit ist, die Gerade durch die Rohdaten zu legen. Diese sind jedoch oft verrauscht und durch saisonale Effekte beeinflusst. Robuster ist daher die Ermittlung der Steigung anhand der Datenpunkte der Trendkomponente einer STL-Dekomposition, die sowohl die Saisonalität berücksichtigt als auch die Daten glättet. Hier gibt es noch die Möglichkeit, eine stark geglättete (robuste) oder weniger stark geglättete Version zu verwenden. Schließlich kann die Robustheit des Werts noch weiter gesteigert werden, indem, ähnlich wie beim Mann-Kendall Feature, die Subserien der letzten Tage betrachtet werden und daraus ein Mittelwert gebildet wird.

Die hier bevorzugte Variante ist die schwach geglättete Trend-Komponente der STL-Dekomposition, da sie einen Kompromiss aus Glättung und zeitnaher Reaktion auf Veränderungen in den Daten bietet.

4.1.5 Niveau in den letzten n Tagen

Für das letzte Feature wurde die Frage, auf welchem Niveau sich der Mittelwert des Fehlerrends im besonders relevanten Zeitraum im Vergleich zur Gesamtserie befindet. Damit werden aus fachlicher Sicht z. B. Situationen beschrieben, bei denen ein leichter

Anstieg zu beobachten ist, dieser sich aber (noch) auf einem so niedrigen Niveau zum als normal betrachteten Niveau des Fehlers befindet, dass er keinen Anlass zur Sorge gibt.

Das Niveau ergibt sich dadurch, dass der Mittelwert der Rückweismengen der letzten n Tage durch den Mittelwert der gesamten Reihe geteilt wird. Zur besseren Vergleichbarkeit werden die Werte zuvor normalisiert. Die Überlegungen dazu, was genau als Inputwerte genutzt wird, wurden hier analog zur Steigung der letzten n Tage geführt. Hier wurde allerdings die stark geglättete bzw. robuste Variante der STL-Dekomposition gewählt, da für das Trendniveau extreme Ausschläge und Schwankungen eine noch kleinere Rolle spielen.

Das Feature ist allerdings kritisch zu betrachten, da einige der als interessant betrachteten Muster gerade darauf beruhen, dass ein Fehler von einem sehr hohen Niveau auf ein niedriges gefallen ist und von dort wieder ansteigt.

4.2 Vornehmen der initialen Klassifizierung

Basierend auf den in Kapitel 4.1 definierten Features konnte ein regelbasiertes System definiert werden, das für die initiale Klassifizierung genutzt wurde. An dieser Stelle soll nochmal betont werden, warum die initiale Klassifizierung vorgenommen wurde anstatt alle Verläufe labeln zu lassen. Dies lässt sich einerseits im Sinne des *human machine teamings* damit begründen, dass bereits in dieser Phase ein Nutzen für die User*innen entstehen soll und damit Vertrauen in das System aufgebaut wird. Andererseits wäre aus ökonomischen- und Ressourcengründen ein Labeln aller Verläufe, auch derer, die klar absteigend verlaufen, nicht sinnvoll.

Der für die initiale Klassifizierung erstellte Entscheidungsbaum beruht auf dem Test auf Monotonie als Kernelement. Davon ausgehend werden die weiteren Features so eingesetzt, dass zwischen den in der Anforderungserhebung als am relevantesten identifizierten Mustern, stetiger Trend und single event, unterschieden wird. Darüber hinaus wurde eine dritte, vorläufige Kategorie eingeführt, nämlich die Kontrollgruppe. Verläufe in dieser Klasse wurden nicht direkt den Expert*innen zum Feedback vorgelegt, da sie grundsätzlich als nicht relevant klassifiziert wurden, aber zum Auffinden von falsch Negativen durch die Autorin stichprobenartig händisch gesichtet und, wo relevant, nochmal von den Expert*innen gelabelt.

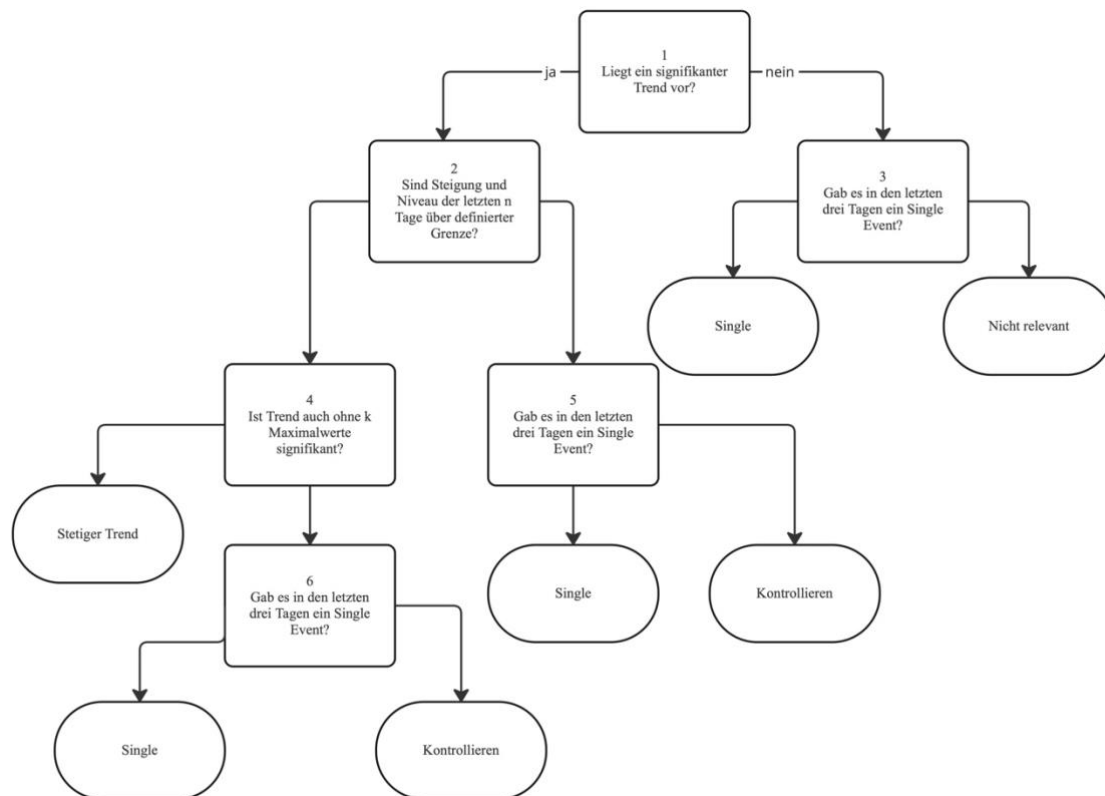


Abbildung 4.3: Entscheidungsbaum zur initialen Klassifizierung (eigene Darstellung)

Neben der Grundstruktur des Entscheidungsbaums mussten die konkreten Grenzwerte für die entsprechenden Features festgelegt werden. Für die initiale Kategorisierung wurden diese so gesetzt, dass möglichst wenig falsch Negative entstehen, also der Fokus auf *recall* (vgl. Kapitel 2.5.2) gelegt. Dadurch ergeben sich im Gegenzug mehr falsch Positive. Dies konnte den Nutzer*innen gegenüber als temporäre Notwendigkeit für die Feedbackphase begründet werden. Tabelle 4.2 zeigt die Zuordnung der Features zu den oben gezeigten Knoten und die gewählten Grenzwerte.

Knoten	Feature	Wert	Begründung
1	Mann-Kendall Test	0.05	Üblicher Grenzwert f. statistische Signifikanz
2	Steigung der letzten n Tage Niveau der letzten n Tage	0.05 0	Mindestens leichte Steigung Sehr weit aus o.g. Gründen
3, 5, 6	Verhältnis zum Mittelwert	6	Vgl. Anforderungserhebung
4	Exkludierbare Extremwerte	2	Bis zu drei als single event

Tabelle 4.2: Zuordnung der Features und Grenzwerte des Entscheidungsbaums

4.3 Erheben von Feedback

Ziel der Feedbackphase war das Erheben eines möglichst aussagekräftigen Datensatzes an gelabelten Verläufen. Dazu wurde jedes Team, das qualitätssicherungsseitig für einen

Bereich bzw. eine Produktlinie der Produktion zuständig ist, gebeten, eine*n oder mehrere Teilnehmer*innen zu bestimmen. So kam eine Gruppe von sechs Personen zusammen, die verschiedenen Hierarchiestufen von operativer Fachkraft bis Teamleiter*in zuzuordnen waren.

In einem Kick-off-Meeting wurde das Projekt vorgestellt, Instruktionen für die Feedbackphase gegeben und Rückfragen beantwortet. Insbesondere wurden die Teilnehmenden auf die folgenden Aspekte hingewiesen:

- Die Feedbackphase wurde für zwei Wochen angesetzt.
- Es wurde keine Anzahl an Bewertungen vorgegeben, sondern die Teilnehmenden wurden gebeten, möglichst jeden Tag alle Verläufe aus ihrem Bereich zu bewerten.
- Es wurde betont, dass die Aufgabe die Bewertung des Verlaufs und nicht des Fehlerbilds sei. Informationen zum Fehlerbild wurden zwar zur Verfügung gestellt (vgl. Abbildung 4.4), um für die Expert*innen einen Mehrwert in ihrer täglichen Arbeit bereitzustellen, sollten jedoch nicht Fokus der Bewertung sein.
- Es wurde betont, dass das Tool in der Feedbackphase voraussichtlich mehr falsch Positive anzeigen würde als später, aber aufbauend auf dem Feedback das Tool an die Bedürfnisse der Expert*innen angepasst werden würde.

Aus technischer Sicht erfolgte die Umsetzung wie folgt: Der in Kapitel 4.2 geschilderte Entscheidungsbaum wurde in Python implementiert. Einmal täglich vor Beginn der regulären Arbeitszeiten wurden durch ein Skript alle NCR der letzten 60 Tage abgefragt, wie beschrieben zu Zeitreihen aggregiert, in Features transformiert und mit Hilfe des Entscheidungsbaums klassifiziert. Die Ergebnisse der Featureberechnung und der Klassifizierung wurden als *comma seperated values* (CSV) Dateien auf dem verwendeten Server abgelegt. Für die Datenerfassung wurde ein Webtool konzipiert, das mit Hilfe des Python Packages *Streamlit* erstellt wurde. Dieses Tool konnte auf die CSV-Dateien zugreifen und die Trends des aktuellen Tages mit direkt aus der Datenbank abgefragten ergänzenden Kontextinformationen visualisieren (vgl. Abbildung 4.4).

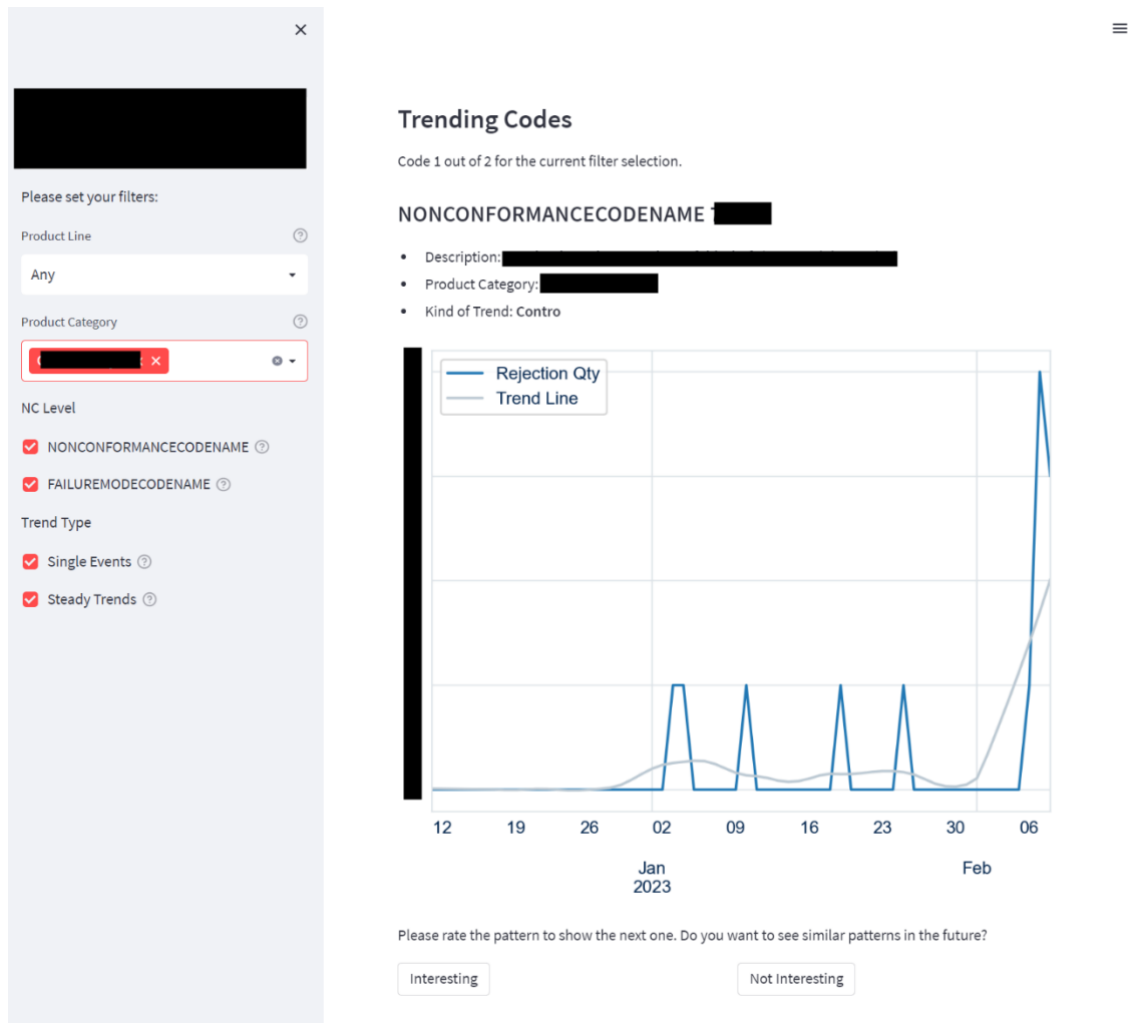


Abbildung 4.4: Screenshot des Webtools zur Feedbackerhebung (eigene Darstellung)

Um die Usability für die Nutzenden zu erhöhen, wurden verschiedene Filtermöglichkeiten angeboten, insbesondere nach Produkt und der übergeordneten Produktlinie, aber auch nach den Fehlerebenen und der Art des Trends, also stetiger Verlauf oder *single event*. Außerdem wurde ein direkter Absprung in das BI-Tool des Unternehmens angeboten, bei dem die Filtereinstellungen übernommen wurden. Das sollte es den Expert*innen vereinfachen, zusätzliche Kontextinformationen wie Produktionsmengen zu ermitteln. Die gefilterten Verläufe wurden einzeln angezeigt, wobei der nächste Code immer dann angezeigt wurde, wenn für den aktuellen eine Bewertung abgegeben wurde. Damit sollte erzielt werden, dass die Verläufe auch wirklich bewertet wurden und nicht nur angesehen. Auch die Bewertungen wurden in eine CSV-Datei geschrieben und für die folgende Analyse (vgl. Kapitel 4.4) genutzt.

4.4 Auswertung des Feedbacks

Vor der Modellbildung wurden die erhobenen Daten explorativ untersucht, um einen ersten Einblick in die Struktur und Besonderheiten der Daten zu erhalten. Davor mussten die verschiedenen Datensätze zur Analyse aufbereitet und zusammengeführt werden. Eine interessante Frage, die sich dabei stellte, war der Umgang mit mehreren und hier insbesondere unterschiedlichen Bewertungen desselben Verlaufs. Es war durch das Design und den Aufbau der Feedbackphase bedingt möglich, dass mehrere Personen denselben Verlauf bewerteten. Da das Feedbacktool als einziges Datum zur bewertenden Person eine Session ID erfasste, war es nur möglich Duplikate aus derselben Session zu entfernen, nicht jedoch zu ermitteln, wie einzelne Personen abgestimmt hatten. Eine Variante wäre, den Mittelwert aller Labels für einen Verlauf zu übernehmen, wobei *interessant* mit 1 und *nicht interessant* mit 0 kodiert wird. Da das Problem dann jedoch nicht mehr als binäre Klassifizierung formuliert werden könnte wurde sich stattdessen dafür entschieden, den Maximalwert als Label zu nutzen. Das bedeutet, wenn eine Person den Verlauf als *interessant* bewertete, wurde er auch so gelabelt, unabhängig davon, wie viele abweichende Stimmen es gab. Eine weitere interessante Frage, die sich an dieser Stelle ergibt, ist zu betrachten, wie enig sich die Befragten in ihrem Abstimmungsverhalten waren. Hier konnte ermittelt werden, dass von 136 Verläufen, die mehr als eine Bewertung erhalten hatten, 66 mindestens zwei unterschiedliche Bewertungen erhalten hatten, was knapp 50 Prozent entspricht. Das kann als Indiz dafür gewertet werden, dass individuelle Einstellungen und Prioritäten – wie schon im qualitativen Teil bemerkt – eine wichtige Rolle bei der Bewertung spielen.

Betrachteter Ausschnitt	Label „Interessant“	Label „Nicht Interessant“
Alle gelabelten Daten	141	803
Von Expert*innen bewertet	131	232
Kontrolldatensatz	10	571
Verläufe des Knotens „nicht relevant“ als Label 0	141	12 173

Tabelle 4.3: Überblick über Labels im Feedback Datensatz

Der gesamte Datensatz umfasst 14 358 Zeilen, von denen jedoch nur 944 eine Bewertung bzw. ein Label erhalten haben, was ca. 6.5 Prozent des Datensatzes entspricht. Ein Großteil der nicht bewerteten Daten entfällt jedoch auf die Verläufe, die als *nicht relevant* klassifiziert wurden und daher nicht betrachtet wurden. Tabelle 4.3 zeigt die Verteilung der Labels auf *interessant* und *nicht interessant* aus verschiedenen Blickwinkeln. Es fällt auf, dass es sich wie erwartet um unausbalancierte Daten handelt, bei denen die positive Klasse unterrepräsentiert ist.

Fehlerklasse	Anzahl
Richtig negativ	571
Falsch positiv	232
Falsch negativ	10
Richtig positiv	131

Tabelle 4.4: Überblick über Labels im Feedback Datensatz

Anschließend wurden die Fehlerraten betrachtet, also die Labels mit den vom Modell getroffenen Vorhersagen verglichen. Für alle gelabelten Daten ergibt sich die in Tabelle 4.4 dargestellte Wahrheitsmatrix, die einem F_1 -Wert von ca. 0.52 entspricht.

Es wurden auch die Werte einzelner Features betrachtet, unter anderem um die Wertespannen für die *grid search* zu ermitteln. Abbildung 4.5 zeigt die Verteilung für eines der Features, den Mann-Kendall p-Wert.

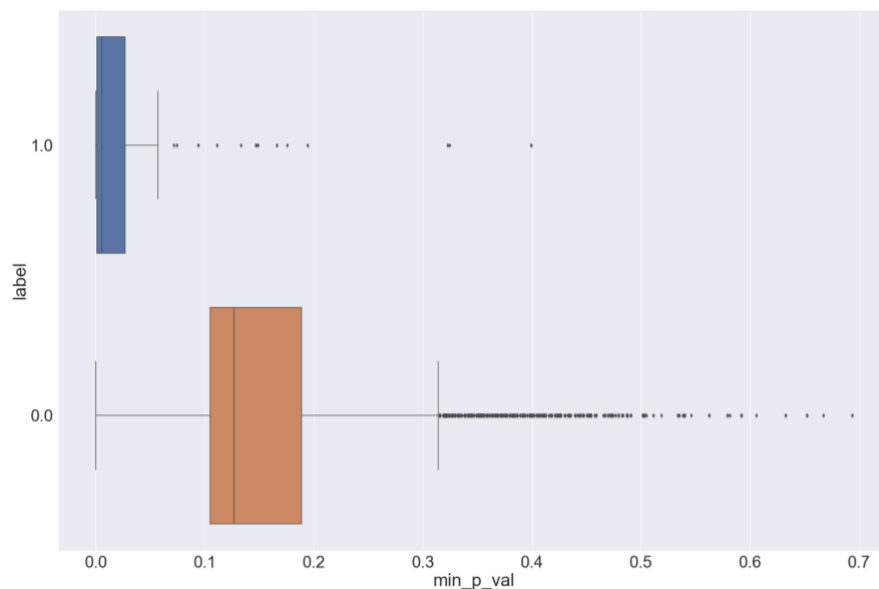


Abbildung 4.5: Boxplot der Werte für das Feature Mann-Kendall Test (p-Wert) (eigene Darstellung)

4.5 Optimierung der Klassifizierung

Das Ziel in diesem Schritt war das Finden des bestmöglichen Klassifizierungsmodells basierend auf Feedbackdaten. Wie in Kapitel 2.5.2 gezeigt gibt es eine Vielzahl an Methoden des überwachten Lernens, die zur Klassifizierung verwendet werden können. Im Rahmen der Optimierung wurden alle der in Kapitel 2.5.2 diskutierten Modelle getestet bis auf die künstlichen Neuronen Netze, da für diese die Größe des Datensatzes zu gering ist.

Um das in Kapitel 4.4 beschriebene Problem der unausbalancierten Daten zu mildern wurde ein *upsampling und downsampling* Ansatz gewählt (vgl. Kapitel 2.5.2.5), das heißt die Stichproben wurden so gewählt, dass der Stichprobendatensatz mit je 200 Datenpunkten der Klasse *interessant* und 200 Datenpunkten der Klasse *nicht interessant* ausbalanciert war. Für alle Modelle außer den regelbasierten Bäumen musste der Datensatz zusätzlich vorbearbeitet werden, wozu eine *preprocessing pipeline* genutzt wurde. Dabei wurden kategoriale Spalten (also die *demand patterns*) mittels *one hot encoding* kodiert und numerische Spalten standardisiert.

Um die Modelle untereinander vergleichen zu können, wurden sie im Rahmen einer fünffachen Kreuzvalidierung trainiert und getestet. Das bedeutet, jedes Modell wurde fünf Mal auf Grundlage unterschiedlicher Subsets des Stichprobendatensatzes trainiert und anschließend bewertet. Die Mittelwerte und Standardabweichungen der fünf Tests wurden sowohl für die Performance auf dem Trainingsanteil als auch dem Testanteil dokumentiert. In Kapitel 3.4 wurde die Sensitivität als ein nicht abschließend geklärtes Thema identifiziert. Daher wurden mehrere Testmetriken angewendet, genauer gesagt neben der Genauigkeit das F_β -Maß mit vier unterschiedlichen Werten für β , nämlich 0.5, 0.7, 1 und 2. Da die regelbasierten Modelle für eine bestimmte Metrik optimiert wurden, wurden sie stets mit dem β -Wert trainiert, mit dem sie auch evaluiert wurden. Für die Genauigkeit wurden die Daten mit einem β -Wert von 1 trainiert.

Abschließend wurden die Modelle mit den ermittelten besten Parametern nochmal auf dem gesamten gelabelten Datensatz anstelle des Stichprobensatzes getestet. Auch wenn dies genau genommen kein Testset ist, da die Datenpunkte der positiven Klasse alle und der negativen Klasse teilweise bekannt waren, so kann dieser zusätzliche Test nochmal Einblicke in die Robustheit der ermittelten Modelle geben.

4.5.1 Modelle

In diesem Kapitel werden die verwendeten Modelle und die gesetzten Parameter kurz beschrieben, die Ergebnisse folgen in Kapitel 4.5.2.

Die **Initiale Klassifizierung (IK)** beschreibt als Baseline das in Kapitel 4.2 beschriebene initiale Modell bzw. die Ergebnisse des in der Feedback-Phase genutzten Modells. Für den **initialen Baum mit optimierten Parametern (IB-o)** wurde für den in Abbildung 4.3 beschriebenen Entscheidungsbaum mittels *grid search* die Grenzwerte für die jeweiligen Entscheidungsknoten gefunden, die die angegebene Bewertungsmetrik maximierten. Dazu wurden für jedes der Features, die von einem der Knoten genutzt wurden, Parameterspannen für die Suche definiert. Diese wurden aus den Verteilungen

der Werte für die einzelnen Features hergeleitet (vgl. Kapitel 4.4). Die Daten wurden dann mit jeder möglichen Wertkombination als Entscheidungswerte klassifiziert und die Bewertungsmetrik berechnet, um die bestmöglichen zu finden.

Es wurde noch ein zweites regelbasiertes Modell getestet, der **vereinfachte Baum mit optimierten Parametern (VB-o)**. Dieser beruht auf der Überlegung, dass man das Entscheidungsmodell aus Abbildung 4.3 deutlich vereinfachen kann, wenn man die Unterscheidung, ob ein Verlauf als *single event* klassifiziert wird oder nicht, an den Anfang stellt, anstatt sie in allen möglichen Zweigen auszuführen. Dann fällt das Feature zu den exkludierbaren Maxima weg. Abbildung 4.6 zeigt schematisch den beschriebenen Entscheidungsbaum. Der Vorteil bei diesem Baum ist neben seiner leichten Nachvollziehbarkeit, dass sich durch den Wegfall eines Features die für die Optimierung benötigte Zeit deutlich verringert. Ein Nachteil ist jedoch, dass dem Grenzwert in Knoten 1 zur Entscheidung, ob ein Verlauf als *single event* klassifiziert wird oder nicht, sehr viel Bedeutung auch für die Klasse *stetiger Trend* zufällt, da ein als Trendverlauf, der über diese Grenze fällt, nicht mehr als solcher klassifiziert werden kann.

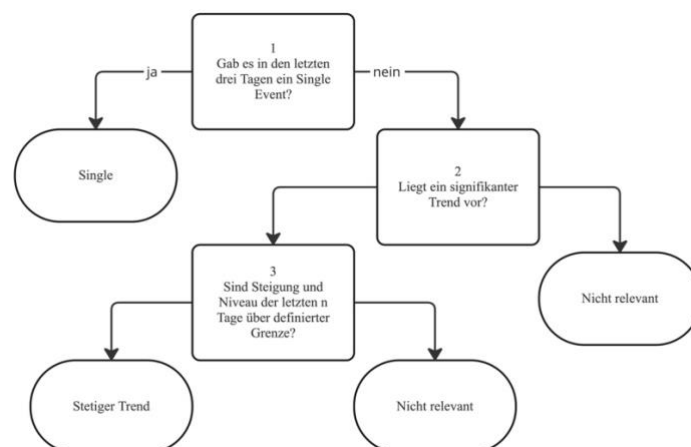


Abbildung 4.6: Vereinfachter Entscheidungsbaum (eigene Darstellung)

Für das dritte regelbasierte System wurde die Erkenntnis genutzt, dass Verläufe der unterschiedlichen *demand patterns* sich stark voneinander unterscheiden, was sich auch auf ihre Bewertung auswirkt. Die Analyse ergab, dass die als *smooth* kategorisierte Verläufe eine deutlich höhere richtig positiv Rate haben als die anderen beiden Kategorien. Darum wurde der **vereinfachte Baum nach demand pattern optimiert (VBDC-o)**, also für die Verläufe jeder Kategorie gesondert die optimalen Parameter ermittelt. Abschließend wurden basierend auf den Ergebnissen und einer tiefergehenden Analyse der Klassifizierungen für das VBDC-Modell die Parameter manuell angepasst (VBDC-m). Dieses Modell wurde nicht in der Kreuzvalidierung, sondern nur auf dem Gesamtdatensatz getestet.

Darüber hinaus wurden die drei regulären Modelle des maschinellen Lernens k Nearest Neighbors (kNN), Entscheidungsbaum (DT) und XGBoost (XGB) getestet. Während für die Modelle kNN und DT die Implementierung des Python Packages *scikit-learn* genutzt wurde (Pedregosa et al., 2011), wurde für den XGB-Algorithmus das Python Package *xgboost* (xgboost developers, 2022) genutzt. Für den **k Nearest Neighbors**-Algorithmus ist die Wahl von k , also die Anzahl der Nachbarn, die zur Berechnung des Labels für den neuen Wert herangezogen werden sollen, relevant. Eine zu hohe Wahl von k führt dazu, dass möglicherweise Datenpunkte berücksichtigt werden, die schon relativ weit von dem neuen Datenpunkt entfernt sind (Harrison, 2021). In diesem Fall wurden für k die Werte 3 und 5 getestet. Für den **Entscheidungsbaum** wurden weitestgehend die Standardeinstellungen für die Hyperparameter übernommen, lediglich der *max_depth* Parameter, der die maximale Tiefe des Baums definiert und das Überanpassen verhindern kann, wurde auf 3 gesetzt. Diese Entscheidung wurde aus fachlichen Überlegungen getroffen. Sie entspricht der Tiefe des initialen Entscheidungsbaums, die als fachlich maximale sinnvolle Tiefe betrachtet wurde. Das **XGBoost-Modell** wurde auch weitestgehend mit den Standardeinstellungen der Hyperparameter implementiert. Neben der Anpassung des *max_depth* Parameters, der wie beim Entscheidungsbaum auf 3 gesetzt wurde, wurde hier noch der Parameter *n_estimators* auf 10 gesetzt. Dieser gibt an, wie viele einzelne Entscheidungsbäume im Ensemble trainiert werden. Auch hier soll der Wert am niedrigen Ende der möglichen Skala das Überanpassen des Modells verhindern.

4.5.2 Ergebnisse

Die gesamte Liste mit den Ergebnissen der Tests kann im Anhang eingesehen werden, an dieser Stelle wird sich auf die wichtigsten Erkenntnisse der Tests beschränkt. Die Genauigkeit aller Modelle, also der Anteil der Datenpunkte, die korrekt klassifiziert wurden, bewegt sich zwischen 68 und 78 Prozent mit Standardabweichungen zwischen 3 und 7 Prozentpunkten. Das Modell mit der schlechtesten Genauigkeit ist das k Nearest Neighbors-Modell mit $k = 5$, die beste Genauigkeit haben das XGBoost-Modell und der pro *demand pattern* optimierte, vereinfachte Entscheidungsbaum, wobei der Baum von den beiden Modellen die geringere Standardabweichung hat. Betrachtet man statt der Genauigkeit die F-Metrik mit einem β -Wert von 1, so liegen die Werte zwischen 0.68 für beide KNN-Modell und 0.79, wieder für die beiden Modelle XGB und VBDC-o. Die Baseline, also die initiale Klassifizierung, erreicht einen Wert von 0.76, etwas besser schneiden der optimierte Baum und der optimierte vereinfachte Baum ab. Der Entscheidungsbaum hat in beiden Metriken sowohl einen Durchschnitt, der schlechter ist als der der besten Modelle, als auch eine hohe Standardabweichung von 7 Prozentpunkten, was dafür spricht, dass das Modell überangepasst und wenig robust ist.

Um die Frage, welches Modell für den Anwendungsfall das Beste ist, zu beantworten, müssen erst die Bewertungskriterien definiert werden. Hier sind neben der Genauigkeit die Robustheit und die Interpretierbarkeit zu berücksichtigen. Unter den regelbasierten Modellen hat das nach *demand pattern* optimierte Modell die besten Werte, weshalb es mit dem XGBoost Modell, das unter den Standardmodellen die besten Werte hat, verglichen wird. Tabelle 4.5 zeigt diesen Vergleich. Dabei wurde auch einbezogen, wie die Modelle sich auf den gesamten gelabelten Datensatz, also nicht die Stichprobe, anwenden lassen. Aus diesem Vergleich geht das regelbasierte System als das zu bevorzugende hervor.

Kriterium	XGBoost	Regelbasiert
Performance	Für alle Metriken die beste oder mit die beste Bewertung	Für alle Metriken unter den Top 3 Bewertungen
Robustheit	In CV weichen Metriken für Trainingsdatensatz stark von Testdatensatz ab Große Standardabweichungen Performance für Gesamtdatensatz deutlich schlechter	In CV weichen Metriken für Trainingsdatensatz mittel bis leicht von Testdatensatz ab Kleinere Standardabweichungen Performance für Gesamtdatensatz unter den besten
Interpretierbarkeit	Mit zusätzlichem Aufwand interpretierbar, Komplexität höher	Sehr intuitiv interpretierbar, leicht in anderen Tools nachstellbar

Tabelle 4.5: Vergleich der besten Modelle

4.5.3 Bewertung

Während in Kapitel 4.5.2 die Modelle insbesondere untereinander verglichen wurden, wird hier die Frage gestellt, wie gut denn das Ziel erreicht wurde, die Klassifizierung an die Bedürfnisse der Nutzenden anzupassen. Betrachtet man dazu zunächst die Metriken, so fällt auf, dass die Verbesserung gegenüber dem initialen Modell sich in einem Bereich von ca. 0.06 Punkten (je nach Metrik) bewegt, was bestenfalls als graduelle Verbesserung bezeichnet werden kann. Um die Ursachen für diese geringe Verbesserung des Modells zu untersuchen, wurden bestimmte Verläufe mit ihren Bewertungen visualisiert. Abbildung 4.7 zeigt eine solche Visualisierung, wobei grüne Punkte Tage mit *interessant* Bewertungen symbolisieren, rote Punkte Tage mit *nicht interessant* Bewertungen und graue Punkte Tage, an denen der Verlauf zwar vom Monitoring als trendend markiert wurde, aber keine Bewertung erhalten hat. Die graue vertikale Linie symbolisiert den Start des Monitorings – davor sind also keine Alarme zu erwarten. Die grüne Linie symbolisiert den Betrachtungsstand, was relevant ist, wenn man sich den Status zu einem bestimmten Tag anschauen möchte.

Was in Abbildung 4.7 deutlich wird ist ein Problem, das schon in Kapitel 3.4 angesprochen wurde, nämlich dass die Befragten zum Teil sehr unterschiedliche

Vorstellungen und Bewertungsmaßstäbe zu haben scheinen. Nur so lassen sich Muster wie das in der Abbildung erklären, bei denen an vier aufeinanderfolgenden Tagen ein Verlauf erst einmal als *interessant*, dann zweimal als *nicht interessant* und abschließend nochmal als *interessant* bewertet wird. Das macht es jedoch sehr schwierig für das Modell, allgemeingültige Muster zu erlernen.

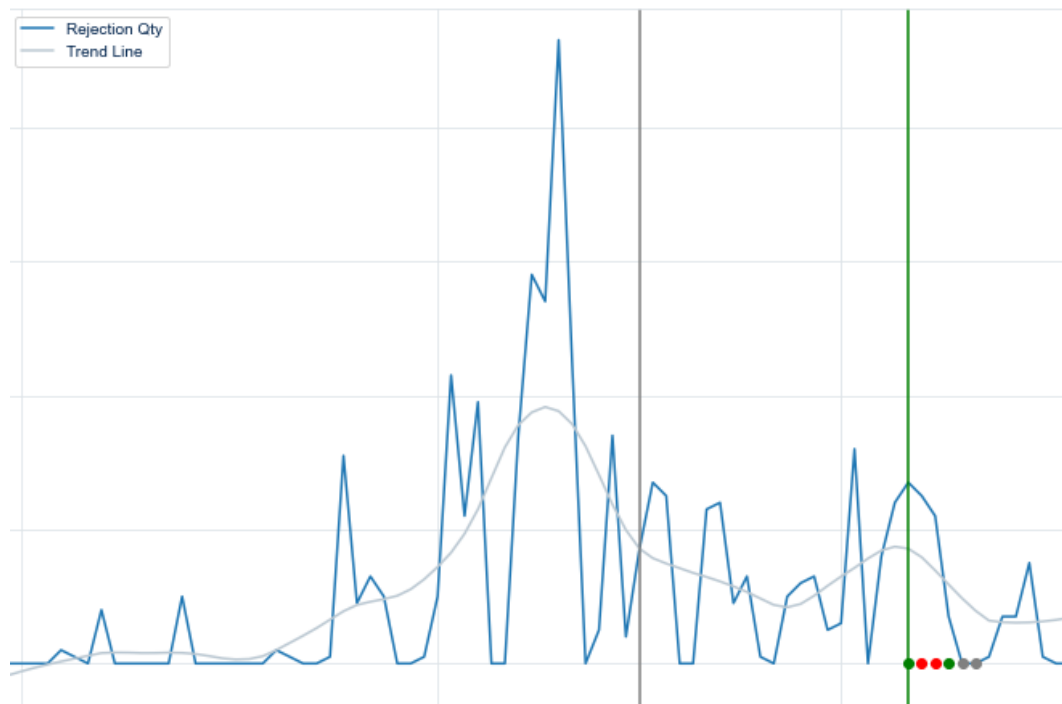


Abbildung 4.7: Beispielverlauf mit inkonsistenten Bewertungen (eigene Darstellung)

Eine weitere Anforderung, die in Kapitel 3.4 erhoben wurde, bezog sich auf die Sensitivität, genauer gesagt auf das als angemessen betrachtete Verhältnis von richtig positiven und falsch positiven Verläufen, die vom Monitoring an einem Tag als relevant klassifiziert werden. Die angestrebte Quote von maximal 50 Prozent falsch Positive lässt sich mit einer Optimierung der regelbasierten Modelle für einen β -Wert von 1 kaum erzielen, weshalb für die Betrachtung des Gesamtdatensatzes der β -Wert von 0.7 für die Klassifizierung genutzt wurde, also mehr Wert auf die Präzision statt auf die Trefferquote gelegt. Für das beste Modell, also das nach VBDC-o Modell optimiert für ein β von 0.7, ergeben sich 115 richtig Positive zu 152 falsch Positiven, was immer noch deutlich über dem gewünschten Verhältnis liegt. Daher wurden stattdessen die Parameter getestet, die sich für ein β von 0.5 ergeben.

Auf der anderen Seite erhöht sich dadurch die falsch Negativanzahl, also die Verläufe, die als *interessant* betrachtet werden könnten, aber von dem Modell nicht als solche klassifiziert werden. Dabei ist wichtig zu bemerken, dass die Label zwar binär vergeben wurden, aber aus fachlicher Sicht Unterschiede bestehen zwischen Verläufen, die auf

jeden Fall relevant sind und solchen, bei denen ein gewisser Interpretationsspielraum besteht. Daher wurden im Rahmen der Fehleranalyse alle durch das präferierte Modell falsch negativen klassifizierten Verläufe wie in Abbildung 4.7 grafisch dargestellt und analysiert. Die Parameter wurden dann im Rahmen dieser Fehleranalyse iterativ leicht angepasst, so dass die Fehler des Modells (also falsch negativ klassifizierte) als möglichst wenig schwerwiegend betrachtet wurden, während die Zahl der falsch positiven möglichst gering sein sollte. Die so gefundenen manuell optimierten Parameter ergaben für den Testdatensatz eine TP-FP Ratio von 97 zu 109, während unter den 40 falsch Negativen keine schwerwiegenden vertreten waren. Das manuelle Modell (VBDC-m) hatte für die β -Werte von 0.5, 0.7 und 1 Werte, die mindestens so gut waren wie die der anderen Modelle (für 0.5 und 0.7 sogar deutlich besser), und erwartungsgemäß für den β -Wert von 2 schlechter war als die anderen Modelle.

4.6 Implementierung des Ergebnisses

Zur Umsetzung der in Kapitel 3.5 beschriebenen Anforderungen wurde im *Snowflake Cloud Data Warehouse* des Unternehmens eine abgespeicherte Prozedur (engl. *stored procedure*) angelegt. Dazu wurde das in Kapitel 4.2 erstellte Skript so angepasst, dass es in dem Cloudsystem lauffähig war, unter anderem musste die Datenabfrage umgestaltet, sowie einige Abhängigkeiten geändert werden und die Ergebnisse wurden statt in eine CSV-Datei direkt in eine neue Tabelle im Data Warehouse geschrieben. Die abgespeicherte Prozedur konnte dann als täglich ablaufender Task angelegt werden. Die Ergebnistabelle steht den Client-Anwendungen im BI-Tool zur Verfügung. Ein Vorteil dieses Designs ist, neben den oben geschilderten Aspekten wie nahtlose Integration in die bestehende Analytics-Infrastruktur, die Historisierung der Daten: Dadurch, dass die Tabelle laufend fortgeschrieben wird, kann z. B. auch untersucht werden, wie oft ein bestimmtes Fehlerbild in dem Monitoring angeschlagen hat.

Statt die in Kapitel 4.5 ermittelten Parameter direkt im Monitoring umzusetzen, wurde ein anderer Weg gewählt: Für das Skript im Data Warehouse wurden die Filter so gesetzt, dass die Sensitivität höher ist, also mehr Ergebnisse und dafür mehr falsch positive Ergebnisse in die Tabelle geschrieben werden. Da jedoch auch alle relevanten Features in der Tabelle verfügbar sind können in der Client-Anwendung abweichende Filter implementiert werden, so kann z. B. auch eine Konfiguration der Sensitivität in mehreren Stufen für die Nutzer*innen angeboten werden.

5 Fazit

In dieser Arbeit wurde ein automatisiertes Fehlertrendmonitoring für ein Produktionsunternehmen mit hohen Qualitätsansprüchen entwickelt und implementiert. Die Entwicklung des Tools barg einiges an Komplexität, was insbesondere durch die folgenden drei Herausforderungen bedingt war:

1. Während die visuelle Beurteilung von Mustern für menschliche Expert*innen eine intuitiv gut zu bewältigende Aufgabe ist, gilt sie für Computer als schwer (Juuso, 2011).
2. Die Definition der relevanten Muster, die als Trends identifiziert werden sollen, hängt stark vom Kontext des Unternehmens ab. Das folgende Zitat aus einem der in Kapitel 3.4 geführten Interviews mit einem Teamleiter in der Abteilung Qualitätssicherung des Fallstudienunternehmens fasst dies gut zusammen:

„Wir sehen es immer wieder und merken es immer wieder, Trend ist nicht eindeutig, Trend ist nichts, was sich hart definieren lässt, das ist sehr viel persönliche Intuition, hat sehr viel mit persönlicher Veranlagung oder Denke zu tun“ (Interview „Trend“ Person 1, vgl. Anhang B).

3. Wie in Kapitel 2.1.2 und Kapitel 3.3 argumentiert wurde, weisen die Inputdaten für das Trend Monitoring, also die Fehlerverläufe, einen hohen Grad an Heterogenität auf, was die Anwendung vieler Methoden erschwert.

Sowohl die Literaturrecherche als auch die Anforderungserhebung im Unternehmen ergaben, dass der Begriff *Trend* Interpretationsspielraum zulässt. Daher wurde ein Vorgehen umgesetzt, das induktiv versucht, das im Unternehmen vorherrschende Verständnis von Trend zu ermitteln, anstatt es vorab festzuschreiben. Dazu wurde das in der Literatur zur Zeitreihenanalyse und Control Chart Pattern Recognition verbreitete Vorgehen verwendet, den Fehlerverlauf in Form von statistischen Kennzahlen oder *Features* zu repräsentieren. Anhand dieser wurde ein initiales Klassifikationsmodell entwickelt, das dann auf Grundlage von Expert*innenfeedback optimiert wurde. Das so gewonnene Modell wurde schließlich im Cloud Data Warehouse des Unternehmens so umgesetzt, dass die Ergebnisse zur weiteren Verwendung in Business Intelligence Anwendungen zur Verfügung steht.

Die beiden Ziele der Arbeit, dem Unternehmen ein Tool zum automatisierten Trendmonitoring zur Verfügung zu stellen und ein Vorgehensmodell zur Entwicklung eines Fehlertrendmonitorings beizutragen, wurden also im Großen und Ganzen erreicht. Dennoch gibt es einige Limitationen und Verbesserungsmöglichkeiten:

- Das Vorgehen basiert auf der Annahme, dass sich durch das Erheben von Feedback zu konkreten Fehlerverläufen eine unternehmensweit gültige Definition eines relevanten Trends herleiten lässt. Die nur marginale Verbesserung der Modellperformance durch die Auswertung des Feedbacks zeigt deutlich, dass diese Annahme nur bedingt zutrifft. Es hat sich gezeigt, dass auch innerhalb des Unternehmens zwischen den Expert*innen sehr unterschiedliche Meinungen darüber herrschen, wie genau ein Trend zu definieren ist. Hier gibt es verschiedene Lösungsansätze. Eine Möglichkeit wäre, das Feedback personalisiert zu erfassen und so personalisierte Modelle zu erstellen. Die Schwierigkeit dabei wäre jedoch, dass pro Person genügend Labels erfasst werden müssten, um das Modell zu trainieren. Das scheint nicht besonders praktikabel, vor allem wenn man Aspekte wie die Personalfuktuation berücksichtigt. Außerdem würde so die gemeinsame Diskussionsgrundlage verloren gehen, da jede Person andere Verläufe angezeigt bekäme. Ein besserer Ansatz könnte sein, das Feedback nicht binär zu erheben, sondern in Form einer Skala, z.B. von eins bis fünf. So könnten Nuancen in der Interpretation besser eingefangen werden und die Klassifikation auf die Verläufe ausgerichtet werden, die als sehr relevant bewertet werden.
- In dem Anwendungsfall wurden unter dem Begriff *Trend* verschiedene Muster zusammengefasst. In diesem Fall wurden die Muster mit den von den Expert*innen verwendeten Begriffen *single event* und *stetiger Trend* übernommen wurde. Für beide Muster mussten somit Features und Relevanzgrenzen bestimmt werden. Es könnten in einem anderen Kontext jedoch auch andere Muster als relevant betrachtet werden, wie der in der Literatur zur statistischen Prozesskontrolle genannte *shift* (vgl. Kapitel 2.3). Eine weitere Schärfung der Unterschiede zwischen den einzelnen Mustern könnte das Monitoring weiter verbessern.
- Auch hinsichtlich der verwendeten Features besteht noch Raum für weitere Forschung. Der Mann-Kendall Test, der eine wichtige Rolle in den erstellten Modellen spielt, ist nicht ohne Schwächen, insbesondere für die hier stark vertretenen unregelmäßigen und null-lastigen Zeitreihen. Abbildung 5.1 zeigt, dass die Ergebnisse des Tests nicht immer mit der menschlichen Intuition übereinstimmen, da die meisten Betrachtenden im rechten Bild durchaus eine Monotonie erkennen würden⁵. Auch einige der in der CCPR vorgeschlagenen Features sind vielversprechend, wurden hier aber nicht getestet, da sie weiter von den in den Expert*innenanforderungen formulierten Kriterien entfernt waren als die verwendeten.

⁵ Die Tests wurden hier mit der Standardvariante des Mann-Kendall Tests durchgeführt. Wählt man die saisonale Variante mit automatisch erkannter Saisonalität, so ergeben sich andere Werte, die den rechten Verlauf als monoton steigend betrachten würden und den linken nicht.

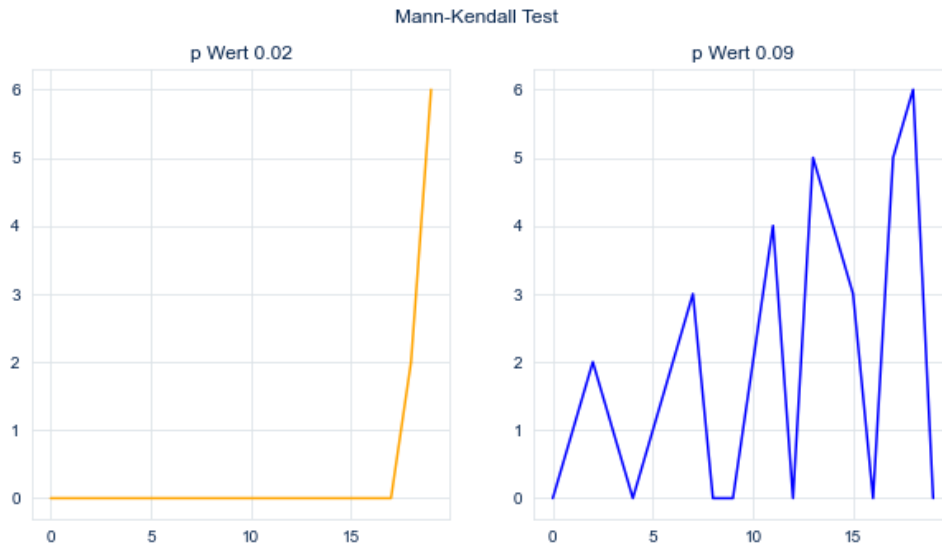


Abbildung 5.1: Beispiel für wenig intuitives Verhalten des Mann Kendall Tests (eigene Darstellung)

- Eine schwierige Fragestellung bei der Entwicklung des Trendmonitors war die Wahl der richtigen Zeiträume. Der wichtigste Indikator ist dabei ohne Frage die fachliche Betrachtung. Allerdings kann, wie für den Mann-Kendall Test auch umgesetzt, ein gewisses Maß an Flexibilität hier einen großen Nutzen bringen. Weitere Überlegungen dazu, wie die Beurteilung der akuten Relevanz – also konkret für diesen Fall die Steigung und das Niveau der letzten n Tage – flexibilisiert werden könnten, würden das Modell möglicherweise verbessern.
- Die *demand patterns* haben sich als nützliches Werkzeug erwiesen, um die sehr heterogenen Verkäufe in grobe Kategorien zu unterteilen. Allerdings waren die Kategorien sehr unausgeglich, eine war sogar überhaupt nicht vertreten. Weitere Forschung dazu, wie sich die Kategorisierung auf den speziellen Fall der Fehlerverläufe übertragen lässt (also *error patterns*) könnte interessante Erkenntnisse liefern.

Ein Vorteil des in dieser Arbeit vorgestellten Ansatzes ist, dass einzelne Aspekte oder Instrumente je nach Kontext und Bedarf des implementierenden Unternehmens flexibel übernommen werden können. Ein Produktionsunternehmen mit ähnlichen Anforderungen und Inputdaten könnte das gesamte Modell samt der gefundenen Parameter übernehmen, während ein anderes nur die grundsätzliche Idee, die Definition aus dem Feedback der Expert*innen herzuleiten, anwenden könnte. Somit ist auch plausibel, dass sich der Ansatz auf Gebiete jenseits der Produktion und Qualitätssicherung übertragen lässt, wie z. B. die oft zitierte Nachfrageforschung.

Anhang

A Interviewleitfaden

Block „Ziele & Use Case“

Trend Monitoring, was genau wird darunter verstanden? Ziele, Zielgruppe, Use Case

- Welche Ziele verfolgt das Unternehmen mit dem Trend Monitoring?
- Welche Zielgruppe oder Zielgruppen sollen das Monitoring nutzen?
- Wie soll das Monitoring genutzt werden?
- Wovon hängt ab, ob die gesteckten Ziele mit dem Trend Monitoring erreicht werden können?
- Wie grenzt sich der Use Case von anderen, bereits implementierten oder noch zu implementierenden Lösungen im Bereich Nonconformance- oder Process- Monitoring ab?
- Warum können existierende Lösungen den Use Case nicht abdecken?

Block „Inputdaten“

- In welcher Frequenz sollen die Daten betrachtet werden (täglich, wöchentlich, stündlich)?
- Wie weit in die Vergangenheit soll bei der Suche nach Trends insgesamt geblickt werden?
- Welche Aggregationsebene soll für das Fehlerbild betrachtet werden?
- Welche Ebenen des NCR sollen betrachtet werden?
- Soll die Anzahl der NCR oder die Rückweismenge betrachtet werden?
- Sollen absolute Zahlen oder Zahlen relativ zur Produktionsmenge betrachtet werden?

Block „Wissensextraktion“

Zu jedem der gezeigten Verläufe lautet die Frage:

- Würdest du diesen oder ähnliche Verläufe im Rahmen eines Trend Monitorings gerne angezeigt bekommen? Findest du den angezeigten Code im Rahmen deiner täglichen Arbeit interessant oder relevant? Wieso oder wieso nicht?

Block „Trend“

- Die definierte Aufgabe des Trend Monitors ist es, Trends automatisiert zu erkennen. Wie wird aus fachlicher Sicht ein Trend definiert?
- Welche Kriterien ziehst du aus fachlicher Sicht heran, um einen Trend zu beurteilen?
- Wie stark gewichstest du bei deiner Betrachtung Ausreißer nach oben oder unten? („mentales Glätten“)
- Wie lange in die Vergangenheit blickst du, um die unmittelbare Relevanz eines Trends für dich zu beurteilen?
- Gibt es aus deiner Erfahrung verschiedene Arten oder Kategorien von Trends, für die du unterschiedliche Bewertungsmaßstäbe heranziehst? Welche Trends sind dir dabei besonders wichtig?
- Gibt es eine Mindestanzahl an Rückweisungen, ab der ein Ausreißer für dich relevant wird?
- Wie wichtig ist es dir, alle möglichen relevanten Trends gemeldet zu bekommen, im Gegensatz dazu, nur die wirklich relevanten Trends gemeldet zu bekommen? Anders formuliert: Wie viele nicht relevante Trends nimmst du in Kauf angezeigt zu bekommen, um einen relevanten Trend angezeigt zu bekommen? Oder: Wie viele relevante Trends bist du bereit zu verpassen, um einen (10) nicht relevante Trends weniger angezeigt zu bekommen?

Block „Implementierung“

Ziele des Workshops: Unter Berücksichtigung der Analyse- und Optimierungsergebnisse,

- Weichen stellen bezüglich weiterer Implementierung eines MVP
- Einigung auf ein grundlegendes Format & Technologie
- Klärung der Anforderungen zur Spezifität und Sensitivität des Systems
- Sammeln von Anforderungen und Ideen zur Etablierung eines kontinuierlichen Feedback- / Verbesserungsprozesses

Teilnehmer*innen:

- Autorin (Moderation)
- Leiter Team Datenanalyse Qualitätssicherung

- MA Team Datenanalyse Qualitätssicherung (zuk. System Owner)
- MA Team Datenanalyse Qualitätssicherung (Key User)

Vorgehen:

- Verwendung eines Whiteboards, Kreativtechniken (v.a. Brainstorming) mit anschließenden Diskussionen
- Präsentieren von 3 Optionen, Brainstorming zu jeder Option (Vorteile, Nachteile, User Stories, Offene Fragen)
- Offenes Brainstorming zu den Fragen Sensitivität/Spezifität, Technologie, Kontinuierliches Feedback
- Zweistufiger Prozess (erst Optionen), da gewählte Option maßgebliche Auswirkungen auf die übrigen Fragen hat

B Transkripte

Interviewte:

- Person 1 (P1): Teamleiter eines Teams in der Abteilung Qualitätssicherung
- Person 2 (P2): Mitarbeiter der Qualitätssicherung mit technischem Hintergrund
- Person 3 (P3): Mitarbeiter der Qualitätssicherung mit fachlichem Hintergrund

Vorbemerkung:

- Antworten wurden zur besseren Lesbarkeit leicht editiert: Dopplungen von Wörtern und „ehm“ wurden gelöscht.
- In eckigen Klammern stehende Worte ersetzen andere, die aus Gründen der Geheimhaltung anonymisiert wurden.
- In geschweiften Klammern wurden sonstige Bemerkungen, wie relevante Gesten etc. vermerkt.

Block „Ziele“ (Person 1)

Person	Inhalt	Kodierung
I	Dann fangen wir mal an mit dem Block Use Cases an und zwar geht es mir darum, wie schon erwähnt, ein bisschen besser die Ziele, Zielgruppen und eben den Use Case zu verstehen. Könntest du erstmal was dazu sagen, welche Ziele durch den Use Case für das Unternehmen abgedeckt werde?	Ziele
P1	Ja sehr gerne. Also vielleicht zum besseren Background oder Verständnis; Grundsätzlich ist es im Unternehmen so, dass sämtliche Nonconformances, also auftretende Fehlerfälle in der Produktion werden bewertet von der Qualitätssicherung aufgrund ihrer Kritikalität und ihre Fehlerursache und potenziell einzuleitenden Maßnahmen, um die Fehler zukünftig vermeiden zu können. Dazu sind im Unternehmen verschiedene Stufen etabliert. Wie diese Fehlerfälle gemonitort werden, das ist zum einen auf einer täglichen Basis durch die sogenannte Fertigungsbegleitung, die diese Nonconformances direkt vor Ort in der Fertigung betrachten, analysieren, abschließen, eskalieren bei größeren Themen und dort auch in täglichen Boardrunden mit der Fertigung besprechen. Dann gibt es darauf aufbauend eine zweite Runde, das sogenannte [wöchentliche Meeting]. Das sieht sich Wochenweise die Top Fehler an, gewichtet nach Fehlerhäufigkeit, nach Kosten und nach Kritikalität; und es gibt als dritte Stufe oben drüber das [monatliche Meeting], das monatlich aggregierte Daten dann auf einem Produkt-Niveau betrachtet und dort nochmal CAPA-Maßnahmen oder ähnliches triggert, die werden allerdings auch schon aus dem [wöchentlichen Meeting] heraus getriggert; aber um eben genau diese Fehlerabstellmaßnahmen zu definieren.	Qualitätssicherung Stufen Gremien Bewertungskriterien: Kosten, Häufigkeit, Kritikalität

	<p>Jetzt ist es grundsätzlich so, dass wir allein, wenn wir die 2 High-Runner Linien, die [Linie A] und die [Linie B] betrachten, bei uns im Unternehmen und ich sehe mir an wie viele Nonconformances gibt es da? Dann hatten wir allein im Jahr 2022 ungefähr [X] Nonconformances, das heißt Records, die erstellt wurden, dass irgendeine Non-Konformität an einem Produkt aufgetreten ist, und das Ganze mit einer Ausprägung von der [n-stelligen] Fehler Ausprägungen also, oder [n-stelligen] Anzahl an verschiedenen Fehlerbildern. Jetzt kann man durch diese Mehrstufigkeit findet man die Top-Themen durchaus raus, das heißt, ein Fertigungsbegleiter sieht durchaus, wenn jetzt irgendein Thema groß aufschlägt. Zum einen oder eben auch wenn's Kosten oder Risiko relevant ist weil dann sind die obergelagerten [...] Gremien auch in der Lage, diese Fehler zu finden. Allerdings, natürlich zum einen, immer mit einem gewissen Verzug. Das heißt, man guckt immer retrospektiv drauf und man agiert natürlich deutlich später, als man vielleicht könnte, das wäre so der erste Use Case zu sagen „Wir wollen früher ins Handeln kommen, um hier um hier früher aktiv zu werden“ und zum anderen sind im Fokus natürlich immer nur die Top Fehler, also die Top Kostenverursacher. Oder die Top Kritikalitätsthemen. Was so ein bisschen unter dem Radar läuft, zwangsläufig, ist, sind gerade so schleichende Trendverläufe, das heißt, ich habe heute vielleicht 3 Non-Konformitäten und morgen 4, übermorgen 5. Das fällt nicht sofort auf, dass hier vielleicht etwas entsteht, was irgendwann einen Fehler wird, ja dann in diese Top Ränge fällt. Mit so einem automatisierten Trend Monitoring würde ich hier rechtzeitig oder einen frühzeitigen Hinweis bekommen „Schau dir das an, schau da schärfer hin, wir scheinen, scheinen hier einen Fehler zu haben der ansteigt und wo man frühzeitiger agieren kann um zu Handeln. Das heißt, im Endeffekt sagen wir der große Use Case an dem ganzen Thema ist eben auch diese Themen zu identifizieren, die Themen frühzeitig zu identifizieren und damit Qualitätskosten zu sparen oder Kosten von Nicht-Qualität zu sparen, indem man rechtzeitig Maßnahmen einleitet. Und gegen das Fehlerbild steuert.</p>	<p>Quantität NCR</p> <p>Top-Themen</p> <p>Geschwindigkeit</p> <p>Proaktives Handeln</p> <p>Schleichende Trendverläufe</p> <p>Geschwindigkeit</p> <p>Kosten Maßnahmen</p>
I	Ok, vielen Dank, du hast ja jetzt auch schon ein paar Zielgruppen erwähnt, aber könntest du nochmal ein paar Worte dazu sagen, welche Zielgruppe oder welche Zielgruppen das Monitoring später nutzen sollen oder werden?	Zielgruppen
P1	Das Monitoring so wie so wie wir's jetzt hier entwickeln, ist in erster Linie wirklich für die für die operative Ebene gedacht. Das heißt für den Fertigungsbegleiter, gegebenenfalls für den [Spezialisten]. Das sind sowas wie die wie die Betreuer der Fertigungsbegleiter, die sind ein bisschen, also ein Fertigungsbegleiter ist meistens für ein oder zwei Fertigungsbereiche zuständig und die [Spezialisten] betreuen mehrere Fertigungsbereiche entweder als Vertretung oder eben auch als Höhe gestellte Stufe. Und es geht genau um diese Stufe, die täglich quasi mit den mit den Rückweisungen arbeiten, hier auf der operativen Ebene wirklich schnell einen Index zu bekommen, dass sie eben nicht nur die [n] Stück sehen, die plötzlich aufgetreten sind, sondern eben genau auch diesen Trend frühzeitig erkennen und nicht erst im Management das Ganze erkannt wird, sondern schon auf der operativen Ebene.	<p>Operativ</p> <p>Täglich</p> <p>Geschwindigkeit</p>
I	Und könntest du nochmal was dazu sagen wie genau sie es nutzen sollen also zum Beispiel, in welcher Frequenz und welche Rolle das so in ihrem gesamten Toolkit vielleicht spielen soll?	<p>Frequenz</p> <p>Rolle</p>

P1	Genau die Idee wär dass wir das Monitoring quasi über Nacht durchlaufen lassen. Immer für den Vortag, so dass morgens, früh morgens wenn der Fertigungsbegleitung oder der [Spezialist] an seinen Tisch kommt, dann sieht er nicht nur seinen Stapel an Rückweisungen, sondern würde quasi die App oder Email oder wie auch immer das dann umsetzen, nutzen, um sich so im Rahmen von 5 - 10 Minuten zu gucken, war irgendwas auffällig am vergangenen Tag, sind irgendwelche Trends aufgeschlagen, wo ich vielleicht mal tiefer reinschauen sollte. Lassen die sich erklären, kenn ich das Thema oder ist hier irgendwas, was mir unbekannt ist, wo ich vielleicht Maßnahmen starten soll?	Nächtlich Interface Zeitaufwand Maßnahmen
I	Ok, vielen Dank, du hast jetzt schon ein bisschen was zu den Zielen gesagt, also insbesondere die, der das Plus an Geschwindigkeit und aber auch bisschen feinmaschiges Monitoring zu haben. Wenn wir jetzt ein bisschen in Richtung Risikoanalyse gehen, wo, wovon hängt in deinen Augen ab, ob diese Ziele erreicht werden können? Also gibt es Faktoren, die da eine besondere Rolle spielen dass diese Ziele dann auch, also sowohl auf diese Usability Seite als auch auf inhaltlicher Seite erreicht werden können?	Ziele Geschwindigkeit Faktoren zur Zielerreichung Usability
P1	Ich denke aus Usability Seite und wahrscheinlich auch aus inhaltlicher Seite ist es wichtig, dass wir hier was schlankes zur Verfügung stellen. Was also, was einfaches und gut verständlich ist. Was, was visuell den Trend auch anzeigt nicht nur Zahlen, sondern eben genau auch ein Trendverlauf in einer Grafik. Ich muss die Möglichkeit haben, relativ schnell in andere Tools abzuspringen, zum Beispiel im [BI-Tool] oder ähnliches, um den Fehler besser zu verstehen. Das ist wichtig, dass wir bei den Trendmonitoring nicht zu viel Inhalt reinpacken, das soll wirklich ein Einstieg sein, ne Einstiegsplattform, wo ich mich schnell durcharbeiten kann und wenn nichts Interessantes ist eben auch in 3 Minuten fertig bin. Wo ich aber genauso die Möglichkeit habe schnell ohne viel Aufwand in andere Tools abzuspringen und den Fehler besser zu verstehen.	Usability Verständlichkeit Visuell Integration mit anderen Tools Zeitaufwand
I	Super, du hattest jetzt ja auch schon erwähnt: Es gibt andere Tools, es gibt andere Use Cases. Wie grenzt sich der Use Case nochmal explizit davon ab oder genauer gefragt warum können die bisherigen Tools eben den Use Case, den wir jetzt versuchen zu bearbeiten, nicht abdecken? Also wenn wir anderen Lösungen im Bereich Nonconformance Monitoring oder auch Process Monitoring im Allgemeinen denken, was ist da deiner Meinung nach der der Added Value den das automatisierte Trend Monitoring da beiträgt.	Abgrenzung andere Tools
P1	Ja. {Verbindungsprobleme} ... Ja, ich glaube der mit Abstand größte Vorteil dem ganzen Thema ist natürlich, aber ich habe es gerade schon gesagt, wir haben wir haben [X] NCR mit, keine Ahnung, ungefähr [Y] verschiedenen Fehlerbildern. Ich kann mir, manuell bin ich nicht in der Lage, mir sämtliche Fehlerbilder anzusehen. Das heißt, ich werde mich immer auf die Top-Themen konzentrieren, manuell, muss die Top-Themen identifizieren und muss dann manuell in Trendverläufe oder ähnliches gehen. Ich werde aber nie diese den Bodensatz oder den ansteigenden Bodensatz betrachten können und die immer erst die Fehler werden bei mir auftauchen, wenn die wirklich kritisch oder teuer werden. Und ich glaub, das ist wirklich der allergrößte Benefit, das was wir aktuell mit den momentanen Tools nicht abbilden können.	Quantität NCR Top-Themen Ansteigender Bodensatz
I	Okay, ähm ja, vielen Dank ich glaube, ich habe jetzt den die Ziele und den Use Case ganz gut verstanden. Danke, oder möchtest du noch was ergänzen zum ersten Block?	

P1	<p>Ja, ich hab' einen Punkt habe ich in der Tat noch vergessen, der ist das ist auch noch sehr interessant. Und zwar gibt es im Unternehmen auch noch die sogenannten [Mechanismus], das sind NCR, oder Nonconformances, die nicht über den Tisch der Qualitätssicherung laufen, weil der Fehler bekannt ist, erkannt ist und man da im System direkt ein automatisches Schließen des NCR quasi implementiert hat. Und ich glaube auch gerade für diese Kategorie ist das Tool sehr, sehr hilfreich, weil wir damit auch ein automatisches Monitoring über diese [Mechanismus] hinbekommen, die sonst mit Sicherheit bei der operativen [Qualitätssicherung] erstmal unten durch rutschen, weil sie diese Teile gar nicht sehen.</p>	<p>Bisher nicht gemonitort</p> <p>durchrutschen</p>
----	--	---

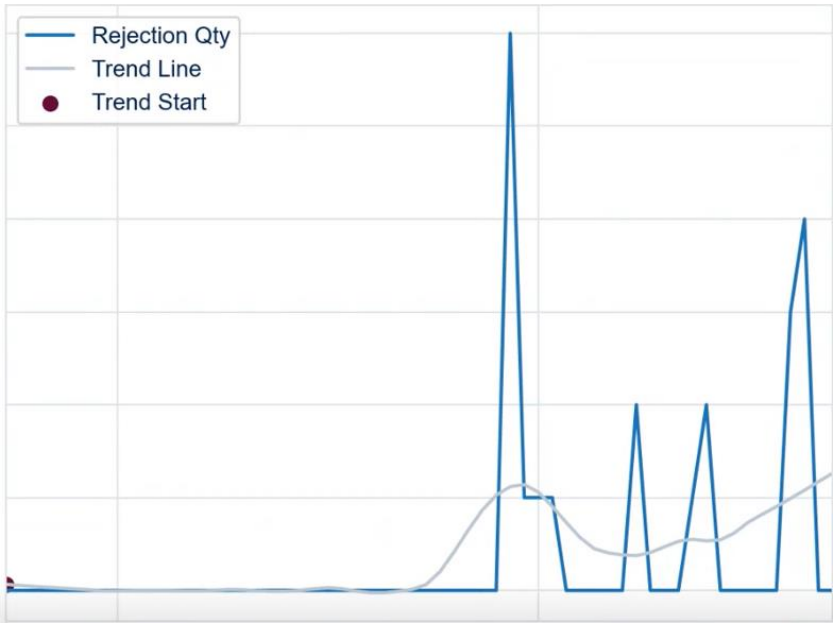
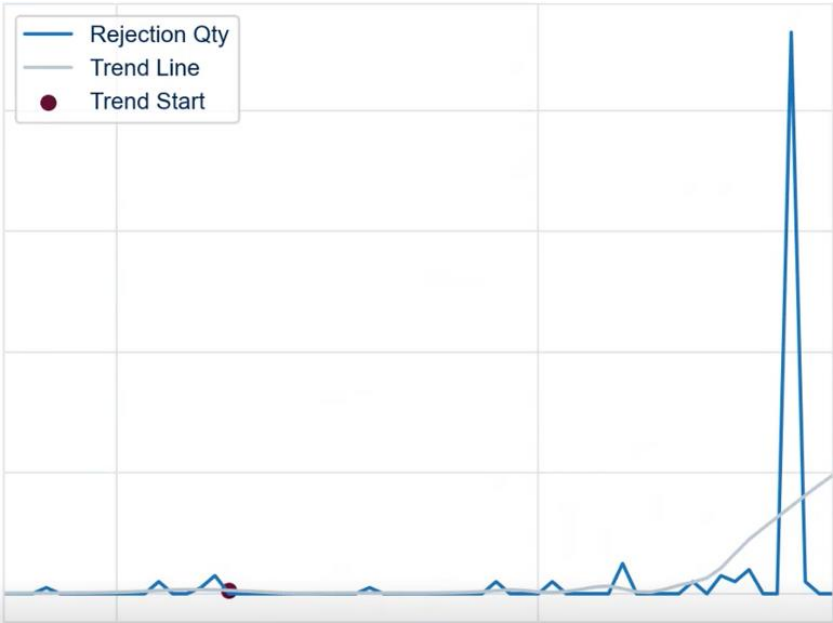
Block „Inputdaten“ (Person 1)

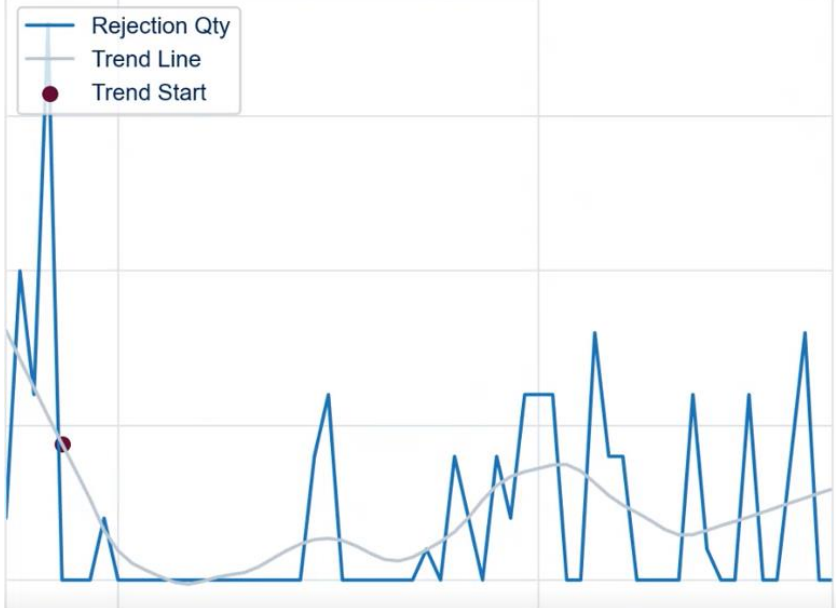
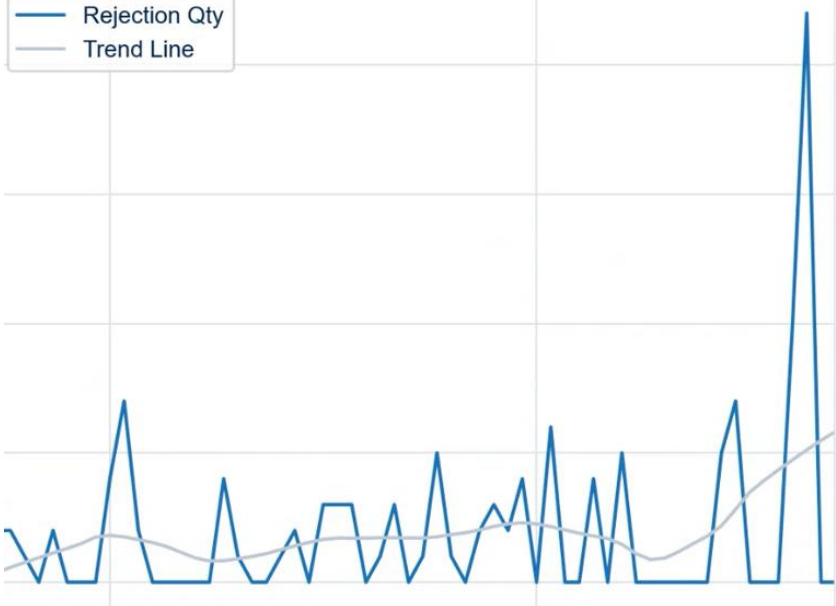
Person	Inhalt	Kodierung
I	Dann würde ich jetzt mal weiter gehen zu den Input Daten weil, wie der Name schon sagt, geht es natürlich um Nonconformance Daten, aber da sind natürlich einige Details zur Implementierung noch offen, angefangen mit der Frage, du hattest bereits erwähnt, das soll einmal nachts durchlaufen, aber die Frequenz, mit der die Daten betrachtet werden sollen, da könnte man sich ja was stündliches, wöchentliches, tägliches, vorstellen also ich glaub kleiner als stündlich zu betrachten, macht kein Sinn. Hab ich das richtig verstanden, dass das Unternehmen hier an ein tägliches Monitoring denkt?	Input Daten Frequenz
P1	Genau, genau das hast du richtig verstanden die Idee wir hier ein tägliches Monitoring zu etablieren was eigentlich so, sag ich mal auf operativer Ebene die, die die aus Erfahrung beste Frequenz ist, um die Themen aus dem Vortag zu Monitoren, ja.	Täglich
I	Und wie weit in die Vergangenheit soll dann die Suche gehen, also wie weit, wie viele Daten für wie weit zurück, also ein Monat, zwei Monate, drei Monate, oder wie weit sollen, sollen sich die Daten insgesamt angeschaut werden?	Zeitraum
P1	Ich denke, für diesen Use Case hier des täglichen Monitorings, ist durchaus ein Zeitraum von 2 Monaten, 60 Tagen sowas ein sehr guter Zeitraum. Da man, sag ich mal so ein bisschen das Bild der Historie bekommt, aber eben zum einen, das nicht mit Daten überfrachtet, in dem man, indem man zu lange nach hinten guckt, auf der anderen Seite aber eben auch nicht nur jetzt die letzten Tage sieht und damit vielleicht wichtige Informationen aus der Vergangenheit verliert.	Zeitraum
I	Ok, dann kommen wir jetzt konkret zu den Daten, also erst einmal, wie weit sollen wir denn runter gehen was die Aggregation angeht, also ich meine, es macht ja nicht so viel Sinn sich alle Fehler zusammen anzuschauen sondern die Frage ist, was hier fachlich Sinn macht, also was ist eine Aggregationsstufe für die ein Verlauf fachlich die meisten Informationen hat?	Aggregation
P1	Also, was wir eigentlich normalerweise betrachten, weil es halt fachlich am sinnvollsten ist, ist die Ebene [Produkt] weil da halt bestimmte Prozesse und Ressourcen dran hängen. Weiter runter gehen macht meines Erachtens keinen Sinn also z. B. nur eine bestimmte Operation anzuschauen, weil ja oft die Fehler nicht da entdeckt werden wo sie entstehen, sondern bei einem Prüfschritt oder bei einer anderen Operation oder so, und weiter hoch zu gehen auch nicht weil dann verschwimmt es, also eine ganze Produktlinie oder so. Und dann natürlich das Fehlerbild, also wären wir dann bei Fehlerbildern pro [Produkt].	Aggregation
I	Ok ähm, und es gibt ja verschiedene Ebenen die in einem Nonconformance Report analysiert oder vermerkt werden. Für die dann auch unterschiedliche Fehler sozusagen hinterlegt werden können. Welche Ebenen sollen denn hier betrachtet werden und wie sollen die betrachtet werden, also getrennt voneinander oder in Kombination?	Ebenen

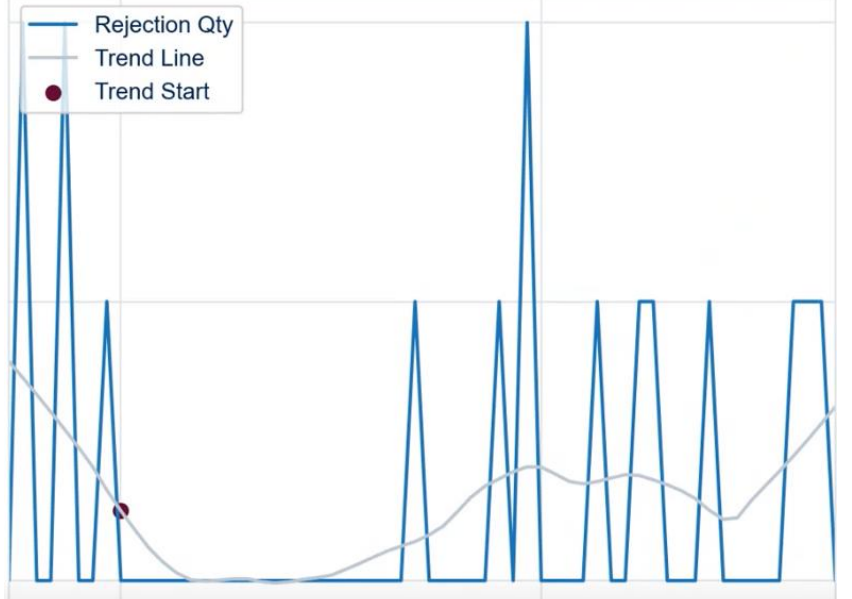
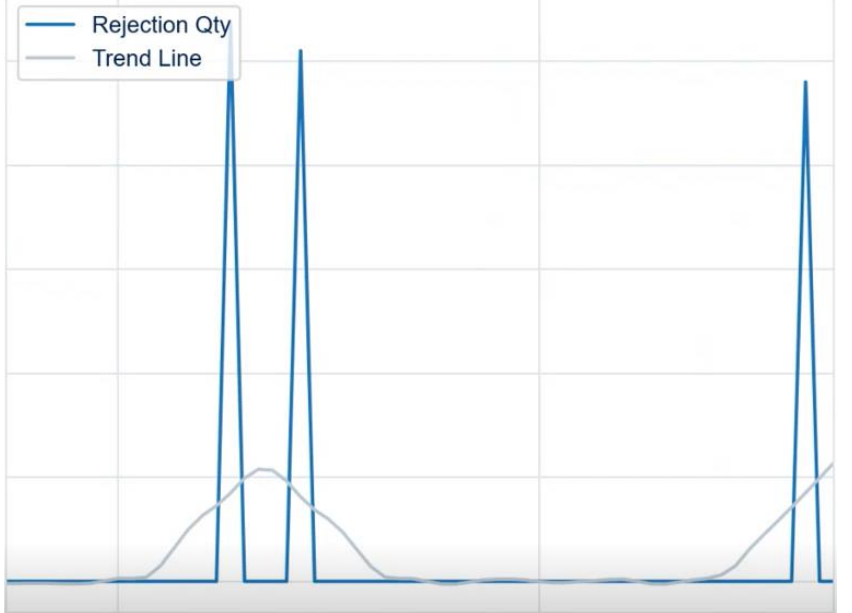
P1	<p>Ja, also grundsätzlich gibt es vier Ebenen bei uns im Nonconformance Management, das heißt die Ebene erste Ebene ist die NC-Ebene also Nonconformance Ebene also wirklich die Ebene der Abweichung, der der Produkt Abweichung vom von von der Spezifikation, die zweite Ebene ist dann die Failure Mode Ebene, also die des der spezifizierte Fehlerfall der von der operativen Qualitätssicherung analysiert wird, also NC Ebene ist das, was die Fertigung entdeckt. Der Prüfer in der Fertigung oder der Fertigungsmitarbeiter und in das System eingibt. Die FM Ebene, die zweite Ebene ist das, was der, was der operative Qualitätssicherer analysiert und dann gibt es noch die dritte und vierte Ebene. Das geht in Richtung Root Cause Analyse, Root Cause Kategorie also sowas wie Prozess, Material, Lieferant wie auch immer ist ein Trend. Ich denke bei uns hier in dem Fall des täglichen, schnellen Monitoren sind die ersten 2 Ebenen interessant. Und es sind beide Ebenen auf unterschiedliche Art und Weise interessant. Die erste Ebene, also die NC-Ebene genau das, was die Fertigung eingibt. Ich möchte natürlich auch das schon sehen, was am Vortag eingegeben ist, was ich vielleicht noch nicht analysiert hab gerade denke stark auch mit dem Fokus auf single events ist das Thema sehr interessant zu sehen ok was ist in der Fertigung aufgetreten, auch wenn ich es vielleicht noch nicht gesehen hab. Und die zweite Ebene, also die FM Ebene ist interessant hinsichtlich dessen was war denn wirklich der Fehlerfall, das passiert durchaus ab und an, dass auch Fehleingaben oder ähnliches stattfinden. Das ist dann korrigiert. Das heißt, ich habe eine verifizierte Ebene, sag ich mal und ich würde die Ebenen auch getrennt voneinander betrachten, um für beides ein Bild zu haben.</p>	<p>Ebenen</p> <p>Geschwindigkeit</p> <p>Analysegrad</p> <p>Trennung Ebenen</p>
I	<p>Okay, super, danke, dann gibt es ja die Möglichkeit, dass wir einfach die Nonconformance-Reports zählen, oder die Rückweismenge also die in den Reports hinterlegte Menge betrachten. Was würdest du hier bevorzugen?</p>	<p>Rückweismenge</p>
P1	<p>Mhm ich würde generell bevorzugen, die, die in den Reports betrachtete Menge heranzuziehen. Und zwar wegen ein paar ja es sind nicht wirklich Sonderfälle, aber es gibt also sagen wir mal für alles, was serialisiert gefertigt wird, spielt es keine Rolle, ob ich da, ob ich da ein Teil, also da ist immer nur ein Teil im NCR. Oder da, wo ich komplette Batches zurückweise, wäre es auch kein, wäre es auch egal. Aber wir haben durchaus Fälle sagen wir mal zum Beispiel im [Bereich A], wo wir [viele] Teile produzieren, danach kommt eine visuelle Kontrolle und wir sehen [einige] davon, haben einen [Fehler A] beim nächsten Tag haben vielleicht nur [wenige] nen [Fehler A] und dann am darauffolgenden Tag haben dann [deutlich mehr] nen [Fehler A]. Wenn ich jetzt immer nur die in den NCR-Anzahl betrachten würde, hätte ich 1-1-1 und würde nichts erkennen. Deswegen sind für solche Fälle durchaus die betrachtete Menge innerhalb des NCR also der schadhaften Teile das interessantere.</p>	<p>Rückweismenge</p>
I	<p>Okay und dann abschließend noch ne Frage die auch da noch relevant ist und zwar ob die Zahlen absolut betrachtet werden oder relativ zum Beispiel zur Produktionsmenge betrachtet werden sollen.</p>	<p>Relative vs. Absolute Zahlen</p>

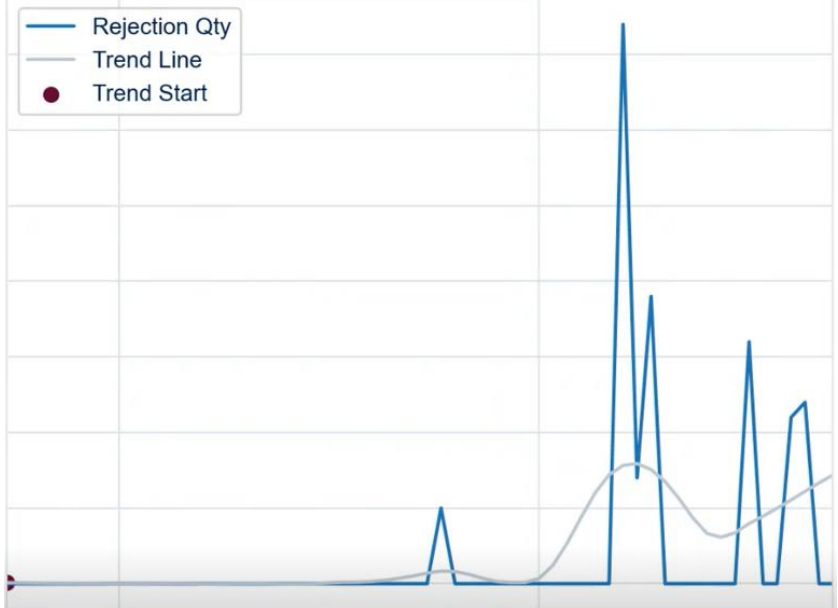
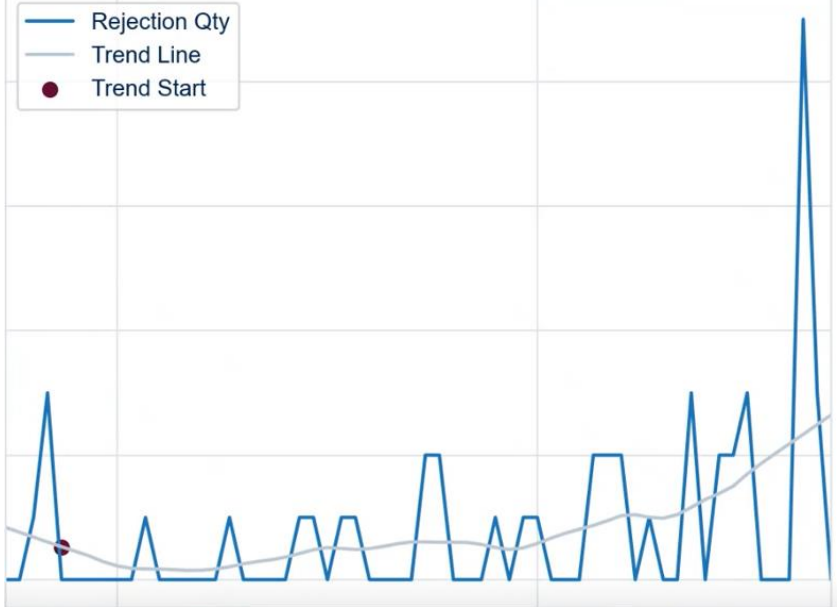
P1	<p>Ja, ja, das ist eine sehr spannende Frage und eine Frage, die auch immer wieder im Unternehmen für Diskussionen sorgt. Aus meiner Sicht ist das gerade hier auf der Basis des, des Daily Monitorings, des täglichen Prüfens eines Trends einzig und allein sinnvoll die absoluten Zahlen zu betrachten. Aus nem relativ einfachen Grund: Die relativen Zahlen brauchen immer einen Bezug zu irgendeiner Produktionsmenge und so ein Fehler, sagen wir mal, eine [Fehler B]. Das kann zum Beispiel an 5 verschiedenen Operationen gefunden werden. Das heißt, ich habe eigentlich 5 Operationen, die ich entweder, wenn ich sehr exakt sein will, wo ich für jede Operation ne Produktionsmenge bräuchte, damit würde ich aber diese 5 wieder aufteilen in 5 verschiedene Fehlerfälle, dann würde mir würde der Trend vielleicht durchflutschen. Oder ich definiere eine Operation, in der ich diese Produktionsmenge ermitteln und sage ok alles was abgeliefert wurde zum Beispiel ist meine Produktionsmenge. Jetzt haben wir aber Durchlaufzeiten die durchaus mal 2, 3 Wochen betragen können. Das heißt wenn da einen Fehler in der [einfließendes Produkt] auftritt und ich nehm' hinten die endproduzierte Menge, die matchen, die haben nichts miteinander zu tun, was an dem Tag abgeliefert wurde und was vorne produziert wird. Das kann im schlimmsten Fall dazu führen, dass ich vielleicht an einem Tag NCRs bearbeite oder erzeugen, wo ich gar keine Produktionsmenge hab, das heißt, meine relative Zeit würde ins Unendliche schießen oder eben auch wenn ich nur 1 produziert hab, dann würde es würde die Zahl sehr hoch werden. Umgekehrt vielleicht kann es natürlich auch genau die umgekehrte Richtung gehen, wenn ich an einem Tag wie an nem [Tag] was produziert habe, da ist gerade aber gerade kein Qualitätssicherer da, der die Rückweisung abschließt. Dann ist da die Quote immer 0 und würde immer den Trend nach unten beeinflussen. Deswegen ist für tägliches Monitoring durch diese Diskrepanz zwischen Auftretenszeit und wo ermittle ich so eine Produktionsmenge immer die absolute Zahl zu bevorzugen.</p>	<p>Absolute Zahlen</p> <p>Produktionsmenge</p> <p>Relative Zahlen</p> <p>Geschwindigkeit</p>
I	Vielen Dank, verstanden. Gibt es zu den Block Inputdaten noch was, was wir noch berücksichtigen sollten oder was dir noch eingefallen ist?	
P1	Zu dem Block Input Daten also ich denke, wir sollten uns komplett auf [System A] und [System A]-gefertigte Produkte konzentrieren, da wir hier ein einheitliches Nonconformance Management haben mit einem einheitlichen System. Wir haben noch ein paar Punkte in Alt-Systemen und das würden wir hier nicht mitberücksichtigen, da da eine ganz andere Datenbasis und Struktur der Daten vorliegen würde.	Datenquelle

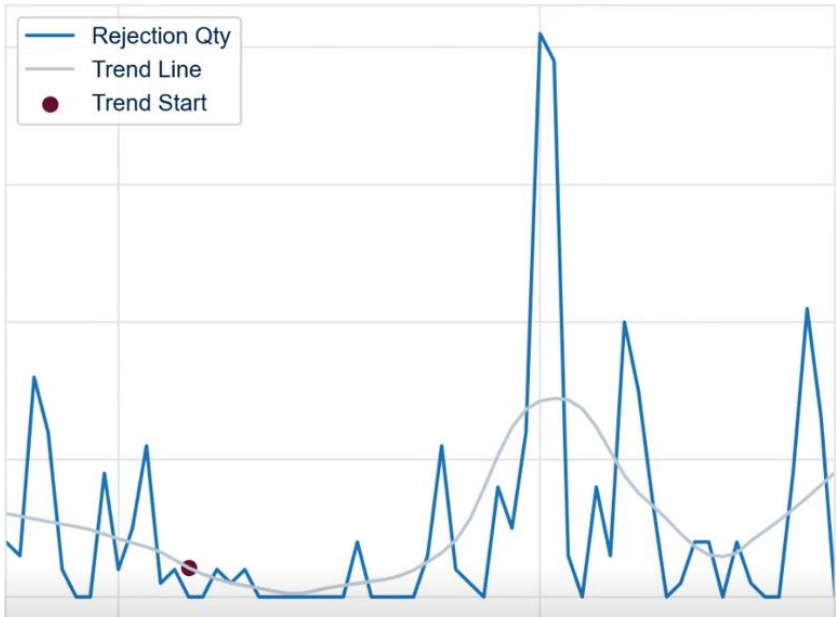
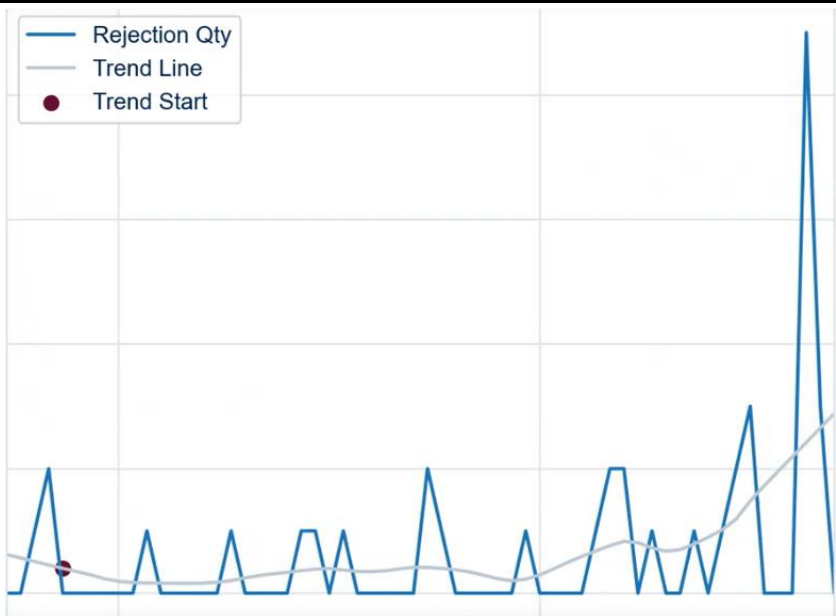
Block „Wissensextraktion“ (Person 1)

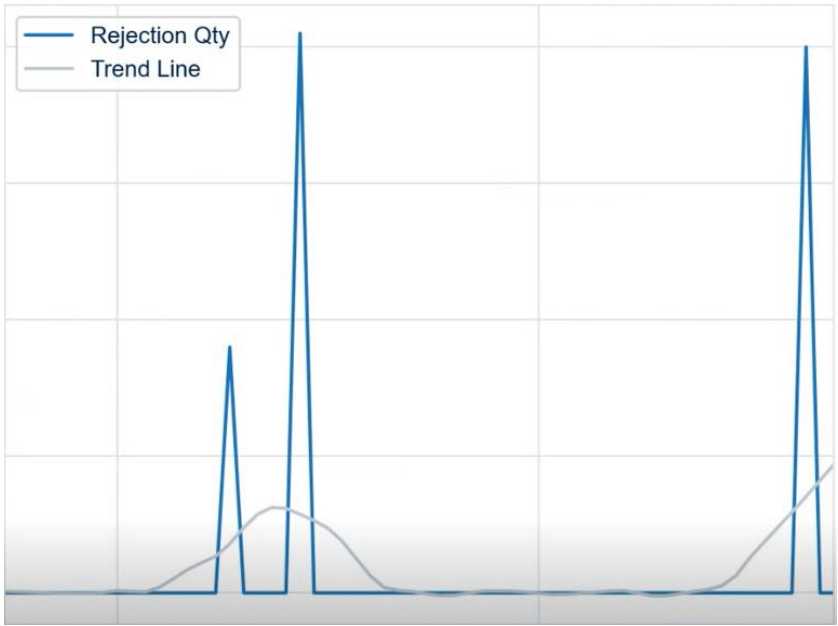
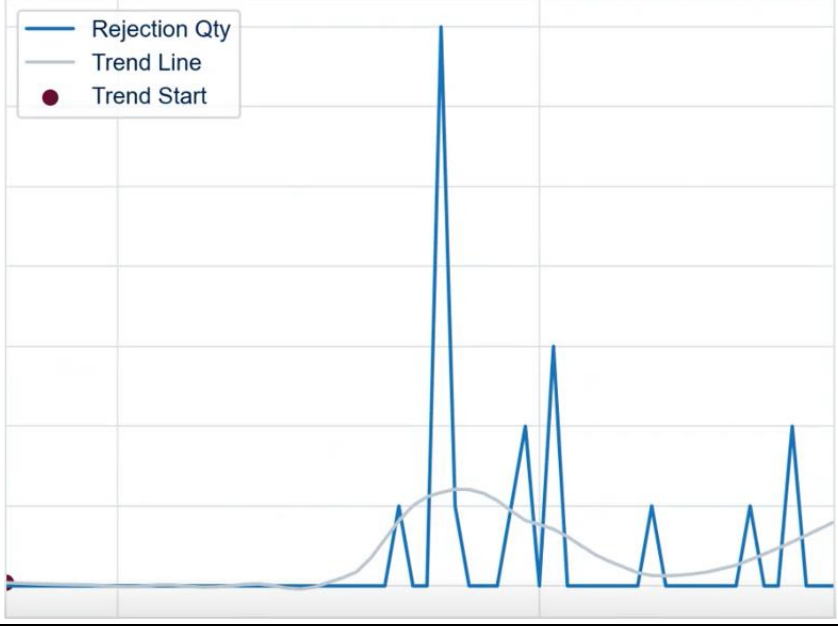
ID	Inhalt / Bild	Kodierung
1		
P1	<p>Ok. Gut ja, ich würde diesen Code als interessant voten. Zum einen scheint das Thema tatsächlich erst Ende November erstmalig oder mehr oder weniger erstmalig aber aufgetreten zu sein, da zwar mit dem höchsten Ausschlag, aber wir sehen uns jetzt doch in einem in einem immer kürzeren Intervall, dass die Fehler auftreten und gegen Ende hin auch wieder deutlich, deutlich ansteigend von der Anzahl. Insofern würde ich den als interessant werten.</p>	<p>+</p> <p>Magnitude Frequenz</p>
2		
P1	<p>Auch diesen Code würde ich als als interessant werten. Und der scheint sonst auf einem sehr niedrigen Niveau zu performen. Und immer mal wieder aufzutauchen. Wir hatten scheinbar vor ein paar Tagen single event. Oder vielleicht singe Event mit mit mit guten 90 in der Ausprägung. Insofern wird wäre das für mich so ein klassischer Fall zu Beobachtung zu gucken war das jetzt wirklich ein single event und einmalig aufgetreten oder haben wir zu erwarten, dass jetzt häufiger mal so ein Peak auftritt.</p>	<p>+</p> <p>single event</p>

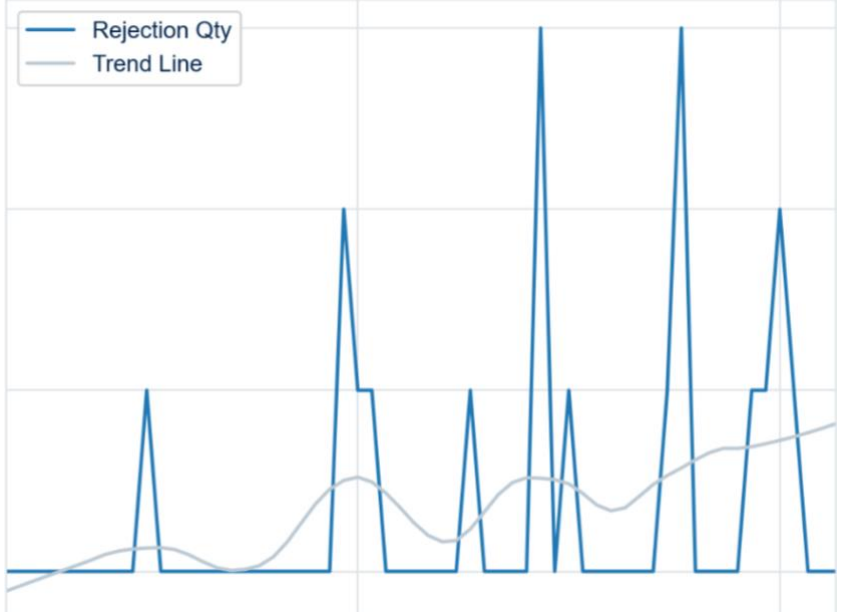
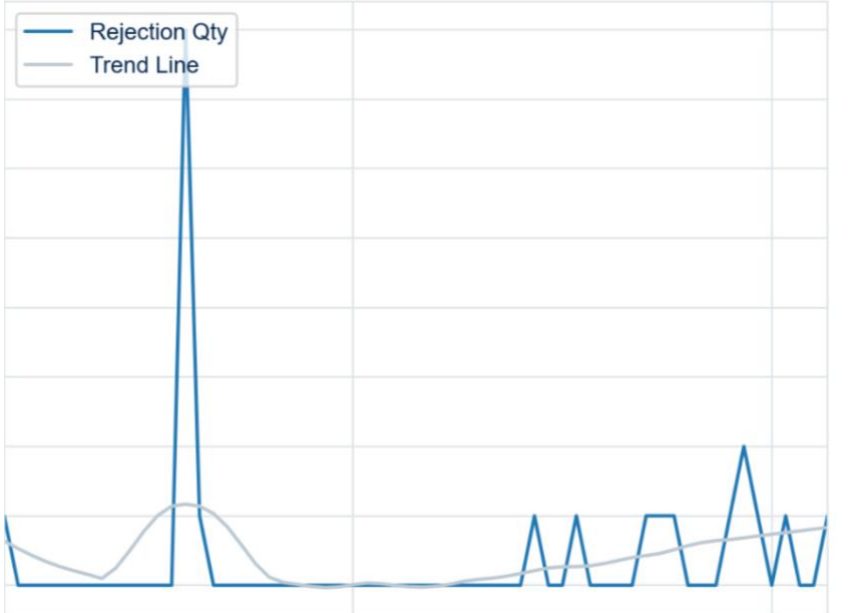
3	 <p>Rejection Qty Trend Line Trend Start</p>	
P1	<p>Der Code ist in der Bewertung etwas schwieriger. Der hatte tatsächlich Anfang November mal einen relativ starken Ausschlag, danach war gefühlt nichts und jetzt ist er seit einem guten Monat eigentlich auf einem sehr konstanten Level. Ähm. Den Code würde ich mir noch mal in der Historie und und in Verbindung mit Produktionszahlen und ähnlichem ansehen, um wirklich bewerten zu können, ob er interessant ist oder nicht. Also wäre für mich aktuell ein Grenzfall.</p>	<p>°</p> <p>Konstant</p> <p>Produktions- menge</p>
4	 <p>Rejection Qty Trend Line</p>	
P1	<p>Den Code würde ich auch erstmal als interessant werten, auch hier haben wir eine relativ konstante Ausprägung innerhalb der letzten Wochen gehabt und ähm dann aber einen deutlichen Peak, um den um 19.12. rum. Auch hier müsste man nochmal betrachten, ob das ob das weiter ansteigt oder nicht, also klassischer Code zu Beobachtung.</p>	<p>+</p> <p>Konstant</p> <p>Peak</p>

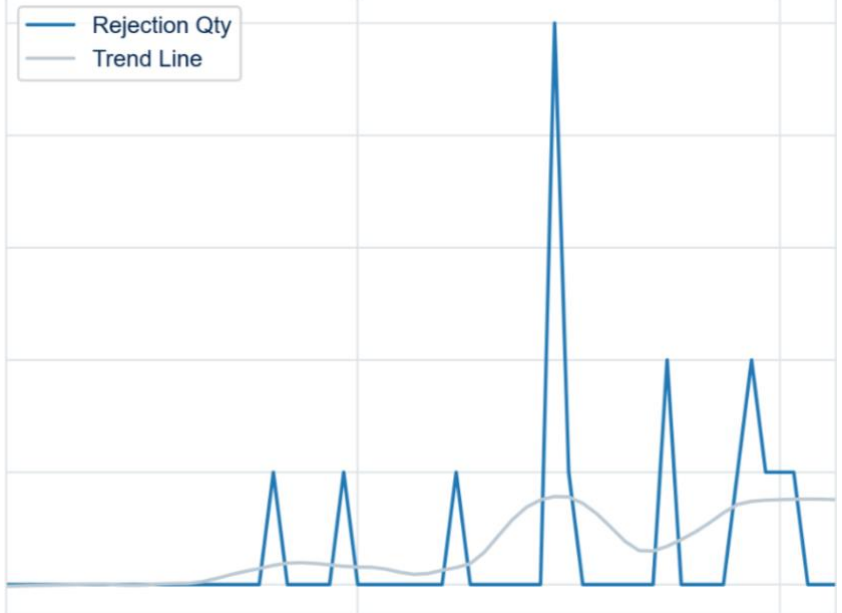
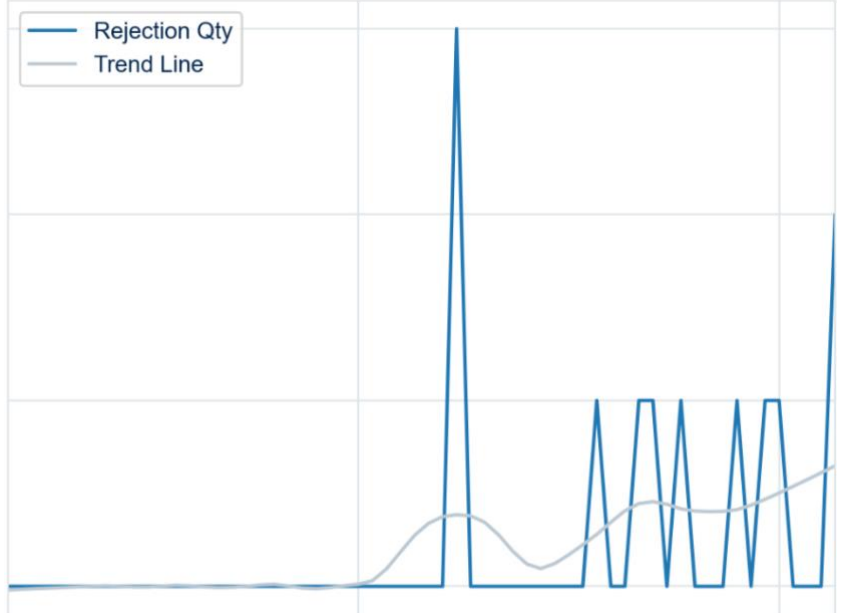
5		
P1	<p>Den Code würde ich aktuell als nicht interessant werten, hier scheinen wir regelmäßig immer mal wieder ein, 2 Teile zu bekommen. Klar gab es hier mal eine Lücke irgendwann Anfang November, die lässt sich vielleicht über die Produktionsmenge erklären oder ähnliches, aber von der Ausprägung und der Anzahl her würde ich den aktuellen nichts sonst werden.</p>	<p>- Regel-mäßig Ursache Aus-prägung Anzahl</p>
6		
P1	<p>Mhm den Code ähm würde ich wahrscheinlich. Auch den ja, den müsste man auch erstmal sich nochmal genauer anschauen, um zu sagen ob das interessant ist. Hier scheinen auch vermehrt immer mal wieder sehr hohe Peaks aufzutreten. Wahrscheinlich handelt sich um Art [Produkt A] oder irgendwas, wo, wo sehr viele Mengen produziert werden, wo es immer mal wieder im Ausschuss gibt. Entweder werden die auch nur sehr unregelmäßig produziert, was die einfachste Erklärung wäre ähm ja. Müsste man sich auf jeden Fall genauer anschauen, um zu beurteilen, aber interessant ist oder nicht.</p>	<p>° Peaks Produktions- menge Ursache</p>

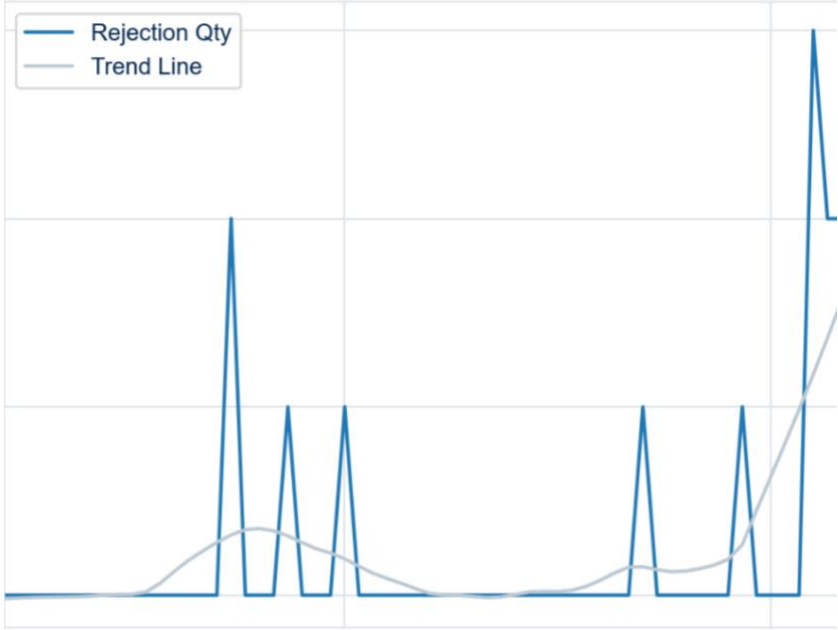
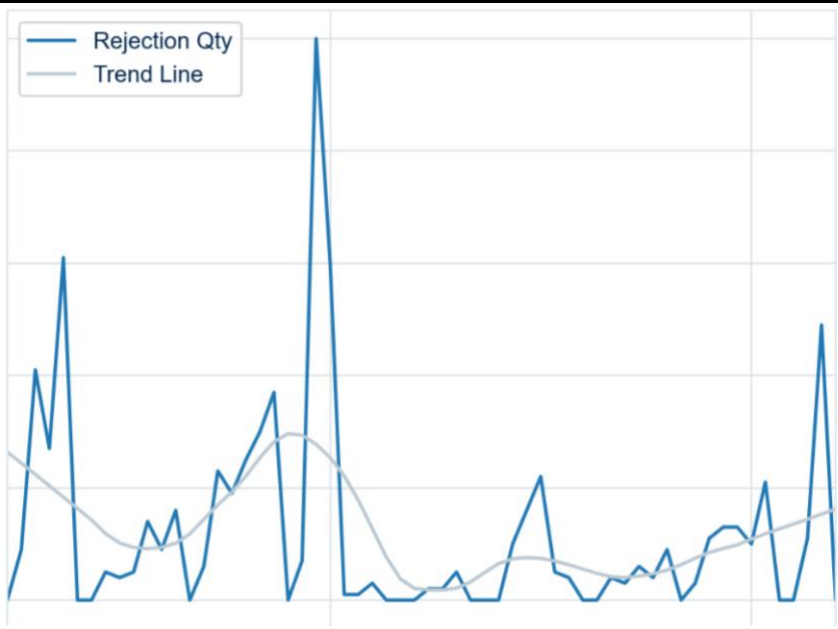
7		
P1	<p>Den Code würde ich auf jeden Fall weiter betrachten, würde ich als interessant werten. Auch hier haben wir einen Fehler, der scheinbar in der Vergangenheit gar nicht aufgetreten ist und jetzt doch vermehrt im Dezember auftritt. Man könnte das Gefühl bekommen, dass es wieder absteigend ist. Nichtsdestotrotz würde ich das erstmal sehen wollen und auch entsprechend bewerten.</p>	<p>+ Erstmalig aufgetreten Frequenz Absteigend</p>
8		
P1	<p>Der Code ist für mich auch relativ deutlich interessant. Wir haben uns über die Wochen immer auf nem sehr niedrigen Niveau und immer mal wieder auftretend bewegt haben jetzt aber doch in den letzten 3 Wochen 2, 3 Wochen hier schon nen deutlichen Anstieg wieder mit dem Peak in 2005 – ähm am 05. Dezember.</p>	<p>+ Anstieg Peak</p>

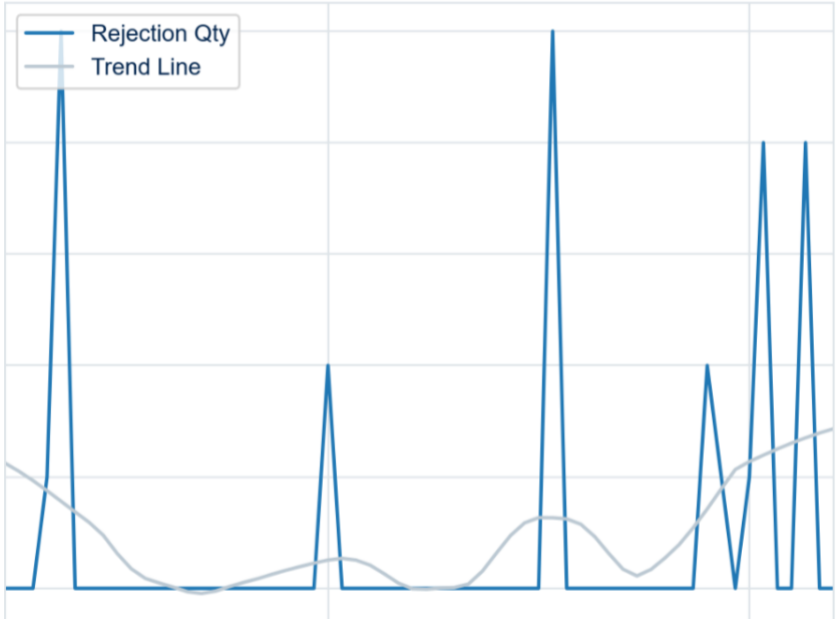
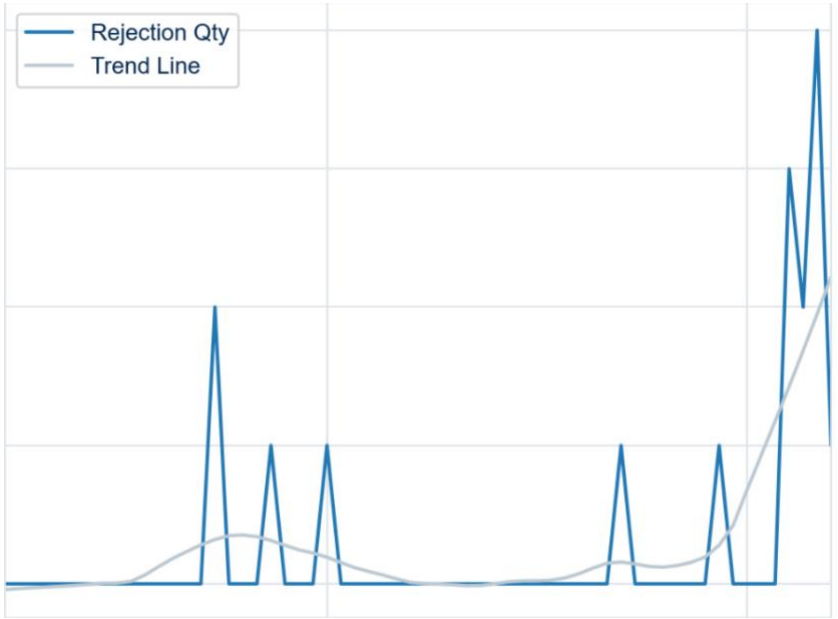
9		
P1	<p>Den Code würde ich auch erstmal als interessant werten. Der hatte zwar gefühlt sein seinen Höhepunkt schon vor vor 3, 4 Wochen und und schien danach wieder abzufallen, allerdings zeigt der der letzte Peak, dass hier durchaus die Möglichkeit besteht, dass wir wieder in nen Anstieg geraten. Deswegen würde ich mir das ansehen wollen, ob man da nen Zusammenhang erkennt.</p>	<p>+ Höhepunkt Abfallend Peak Anstieg Ursache</p>
10		
P1	<p>Der Code ist für mich auch klassisch interessant. Der kommt mir jetzt auch Grad so bekannt vor. Ja, auch ein klarer, klarer Anstieg.</p>	<p>+ Anstieg</p>

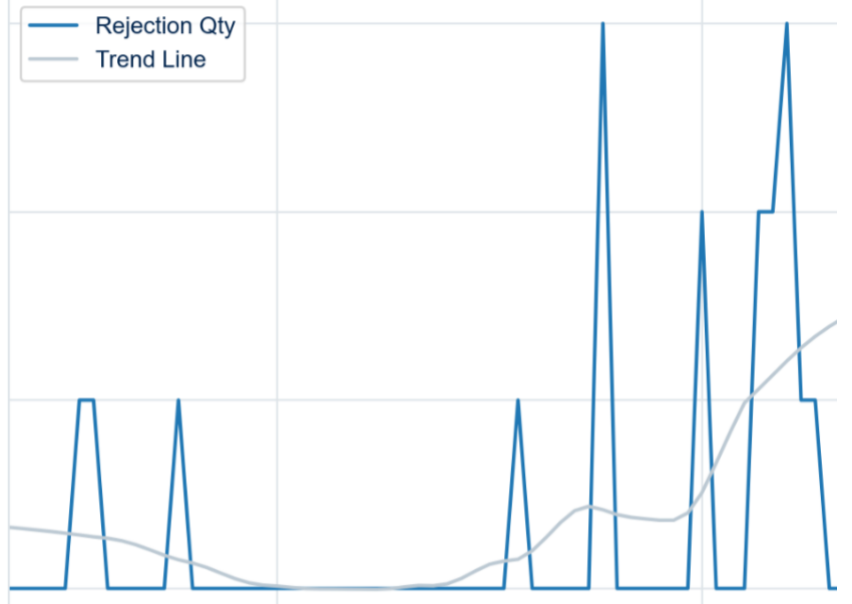
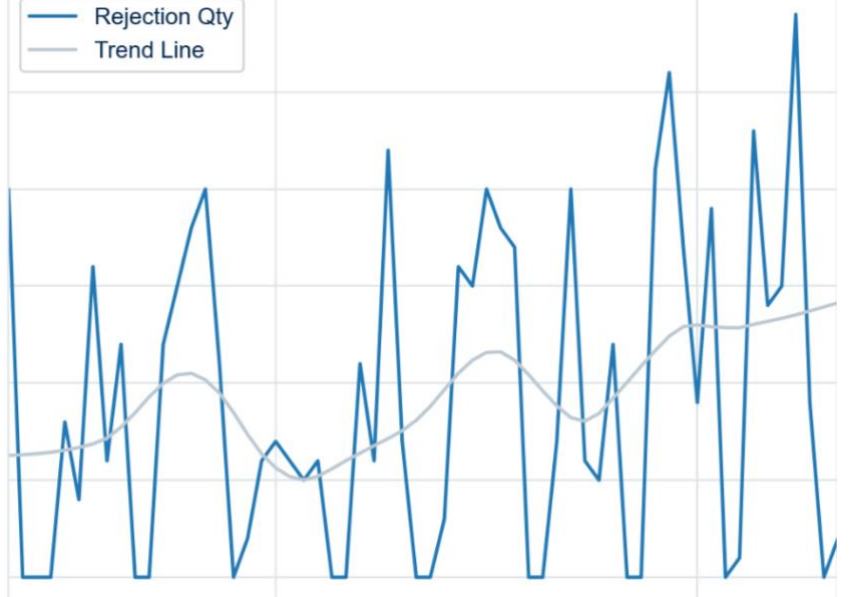
11		
P1	<p>Der ist auch wieder ein bisschen ähnlich wie der, den wir grad mit den Fünfhundertern hatten. Ähnlicher Fall, scheint hier immer mal Peaks zu haben. Müsste man sich ansehen, um was es sich genau handelt, um zu bewerten, ob es interessant ist. Ich würde auf jeden Fall einen Blick drauf werfen.</p>	<p>° Peak Ursache</p>
12		
P1	<p>Ähm. Generell eher interessant, weil wir auch hier wieder in der Vergangenheit scheinbar gar nichts hatten. Das Gefühl wäre ok - Wir haben wir haben hier was, was Abnehmendes. Ähm. Aber um sicher zu gehen, dass es, dass es nicht wieder nach oben wegläuft würde ich einen Blick drauf werfen.</p>	<p>+ Erstmalig aufgetreten Abfallend</p>

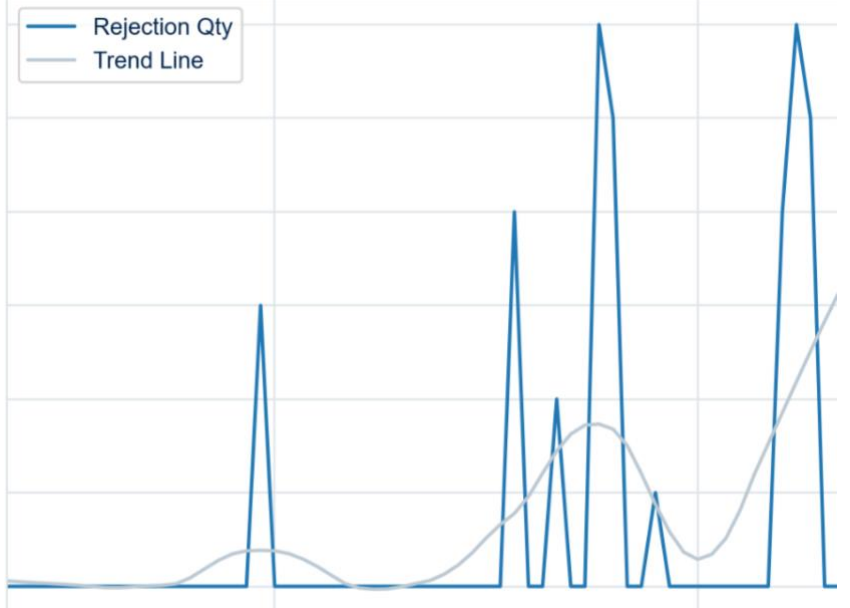
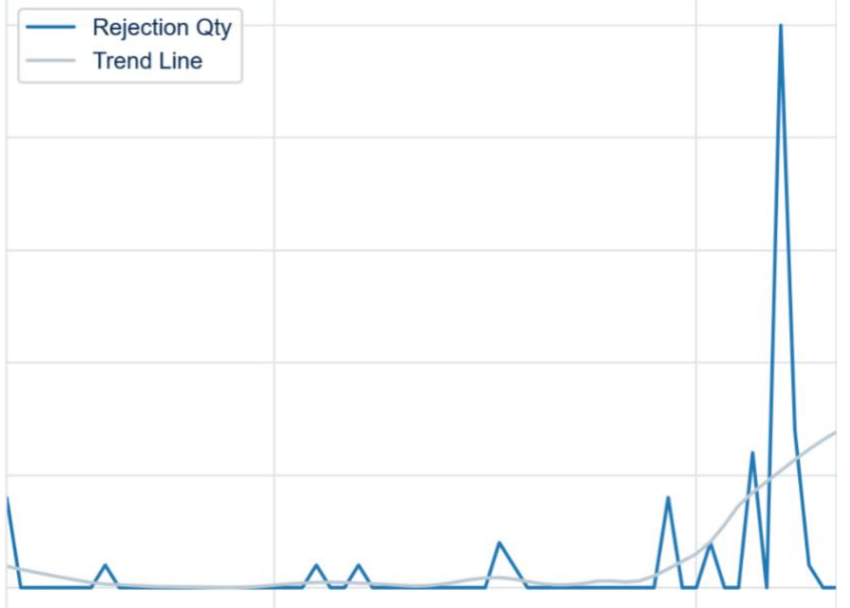
13		
P1	<p>Den kann man durchaus als „nicht interessant“ einstufen, da dass ich die Codes oder die Anzahl immer auf einem sehr stabilen Level zwischen 1 und 3 bewegt, klar, die Statistik zeigt hinten leichten Anstieg, den ich aber so erstmal nicht bewerten würde und weswegen ich ihn auch als nicht interessant ehm einstufen würde.</p>	<p>- Konstant Stabil</p>
14		
P1	<p>Da den Code ähm würde ich mir wahrscheinlich in der Tat anzeigen lassen wollen. Da wir jetzt schon in nen Modus kommen in den letzten Wochen wo er immer wieder auftritt und dann müsste man nochmal, nochmal zumindest verstehen, warum das die Male vorher kaum war, und die letzten Mal oder jetzt jetzt immer mal wieder auftritt, ob sich da irgendwas was einschleicht.</p>	<p>+ Frequenz Ursache Einschleichen</p>

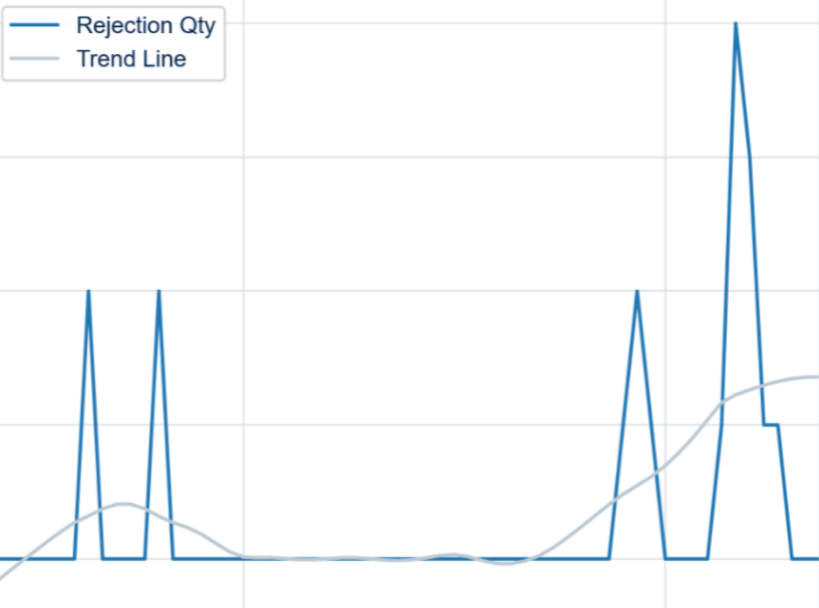
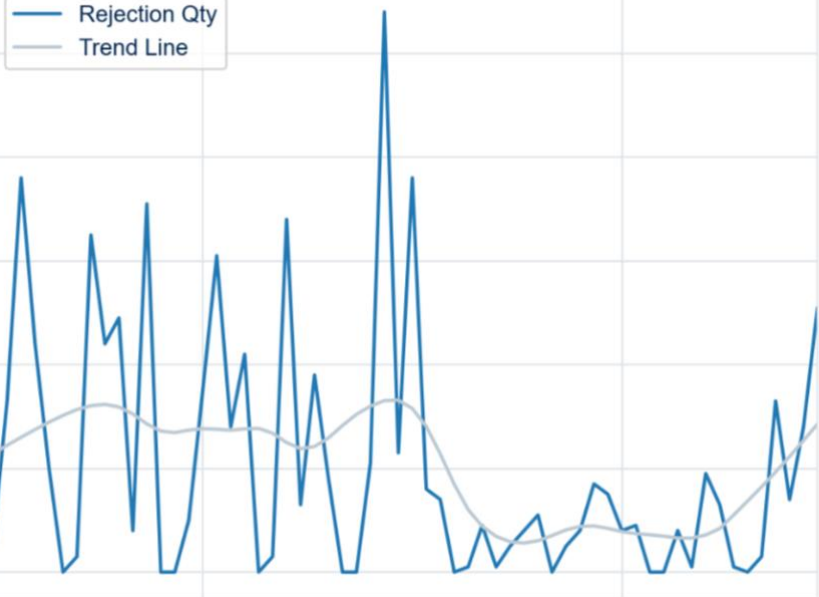
15		
P1	<p>Den würde ich auch als uninteressant klassifizieren. Auch hier haben wir nen Peak der ist ein ganzes Stück her, ansonsten performen wir hier relativ gleichmäßig und in relativ gleichmäßigen Abständen mal mit einem, mal mit 2.</p>	<p>- Peak Konstant Gleichmäßig</p>
16		
P1	<p>Den Code würde ich mir in der Tat wiederum anzeigen lassen, hier haben wir auch in der Vergangenheit gar nichts gehabt und jetzt doch in immer kürzer werdenden Intervallen ein Fehlerbild, auch wieder mit einer leichten Tendenz, auch wenn es hier zwei statt einem Stück sind. Ähm aber den würde ich mir anzeigen lassen.</p>	<p>+ Erstmalig aufgetreten Frequenz</p>

17		
P1	Den Code würde ich in der Tat beobachten. Auch hier haben wir in der jüngsten Vergangenheit gefühlt mehr Teile gehabt, wir hatten dann doch mal einen stärkeren Anstieg, der auf einem höheren Niveau ist, auch wenn Fehler an und für sich relativ gering ist.	+ Jüngst Steigung Niveau
18		
P1	Den würde ich mir aktuell glaub ich nicht anzeigen lassen. Wir sind doch im Mittel sag ich mal auf nem Niveau das das niedriger ist als in der Vergangenheit. Man könnte durchaus sehen, dass hier vielleicht ein leichter Anstieg in den letzten Tagen oder Wochen ist, würde aber bei einem weiteren Anstieg würde der vielleicht auch ohnehin dann nochmal auftreten also momentan würde ich noch nicht berücksichtigen.	- Niveau Anstieg Zeitraum

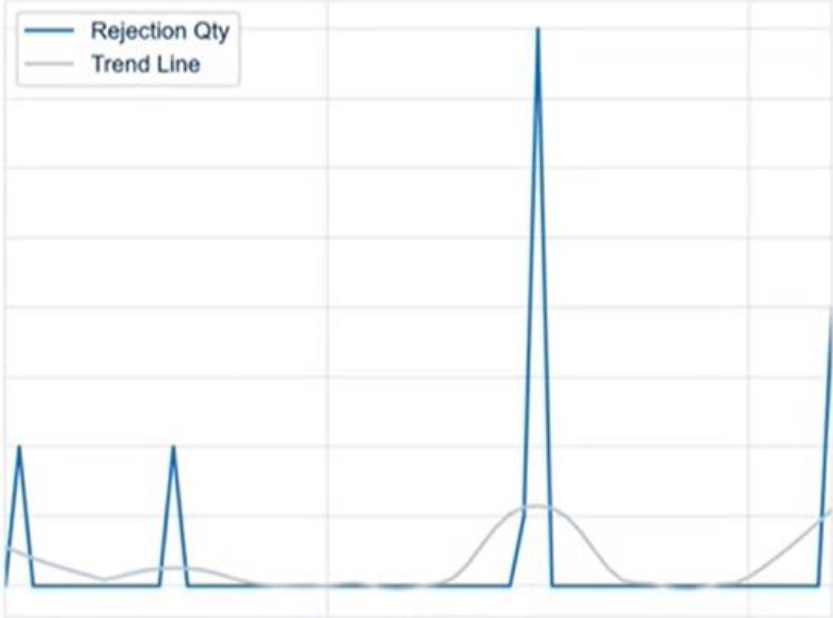
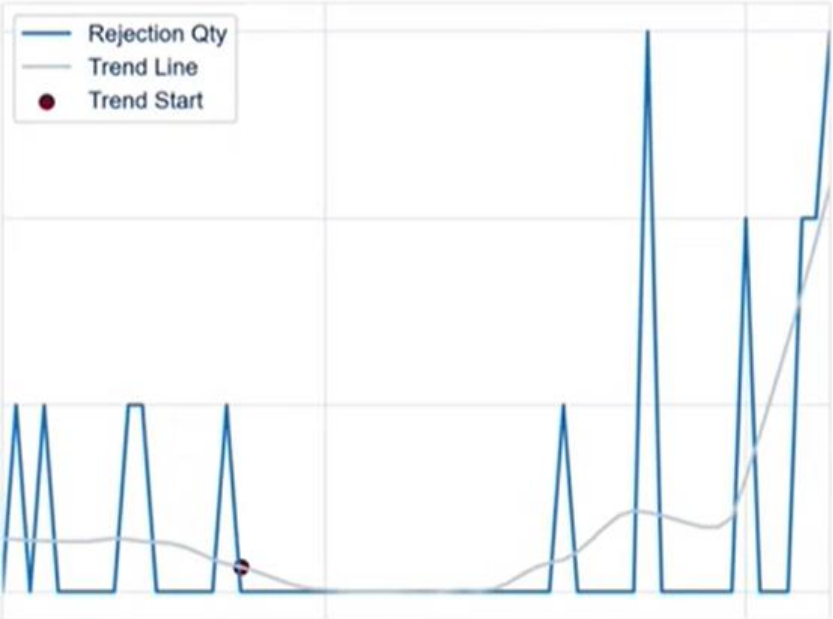
19		
P1	<p>Den würde ich wahrscheinlich momentan auch noch nicht berücksichtigen. Auch hier sehen wir zwar ähnlich wie zu den vorangegangenen, dass man etwas häufigere Frequenz sehen an Bildern in der letzten Zeit, aber es ist trotzdem immer noch auf einem Niveau, wie es vorher auch war. Ich denke bei einem weiteren Anstieg würde er ja aufschlagen, so dass ich momentan nicht beachten würde.</p>	<p>-</p> <p>Frequenz</p> <p>Niveau</p> <p>Anstieg</p> <p>Zeitraum</p>
20		
P1	<p>Den würde ich mir anzeigen lassen. Da sind wir doch die letzten Tage mit einem mit einem klaren Anstieg da und auch mit einer scheinbar täglichen Frequenz wo wir vorher alle paar Tage mal einen hatten.</p>	<p>+</p> <p>Zeitraum</p> <p>Anstieg</p> <p>Frequenz</p>

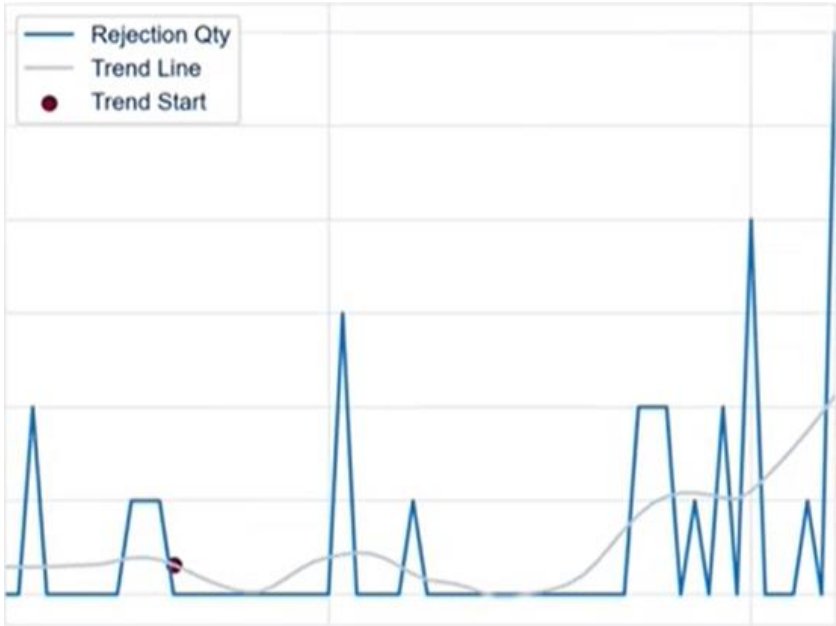
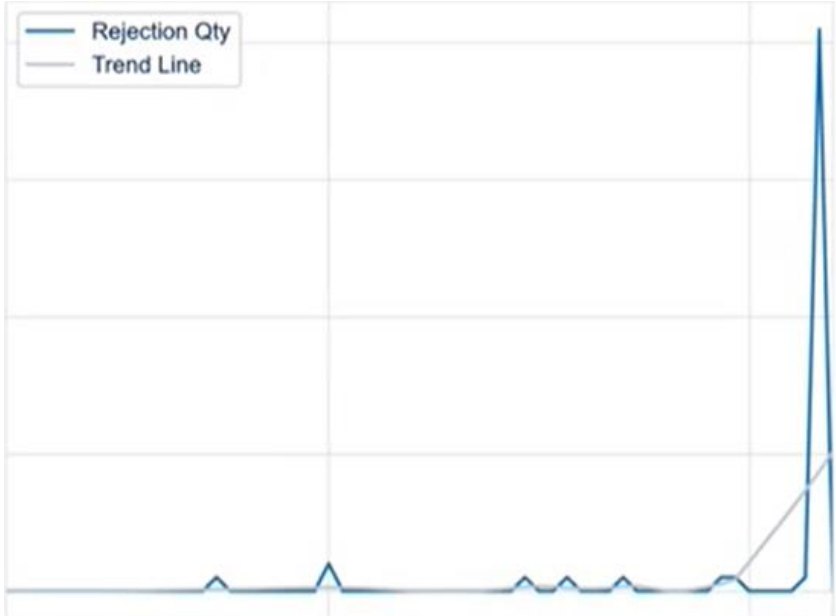
21		
P1	<p>Den würde ich mir auch anzeigen lassen in der Tat ähnliche Argumentation, wir hatten, wir hatten vorher immer mal einen und jetzt haben wir gefühlt über mehrere Tage Fehlerfälle gehabt, und auch mit einer, mit einer teilweise höheren Ausprägung... Könnte sein, dass er, dass es wieder vorbei ist, da die letzten Tage das, das Thema unten ist, aber ich würd ihn mir anzeigen lassen.</p>	<p>+ Frequenz Magnitude Zeitraum</p>
22		
P1	<p>Auch den würde ich mir in der Tat anzeigen lassen. Da ich hier vom Trendverlauf her, der hat zwar ein relativ deutliches auf und ab, aber wenn ich mir die Spitzen anschau, steig ich über die letzten 4 Wochen doch konstant immer um ein paar Stücke an, so dass es durchaus sein könnte, dass wir hier noch weiter nach oben gehen.</p>	<p>+ Varianz Spitzen Zeitraum Monotonie Vorhersage</p>

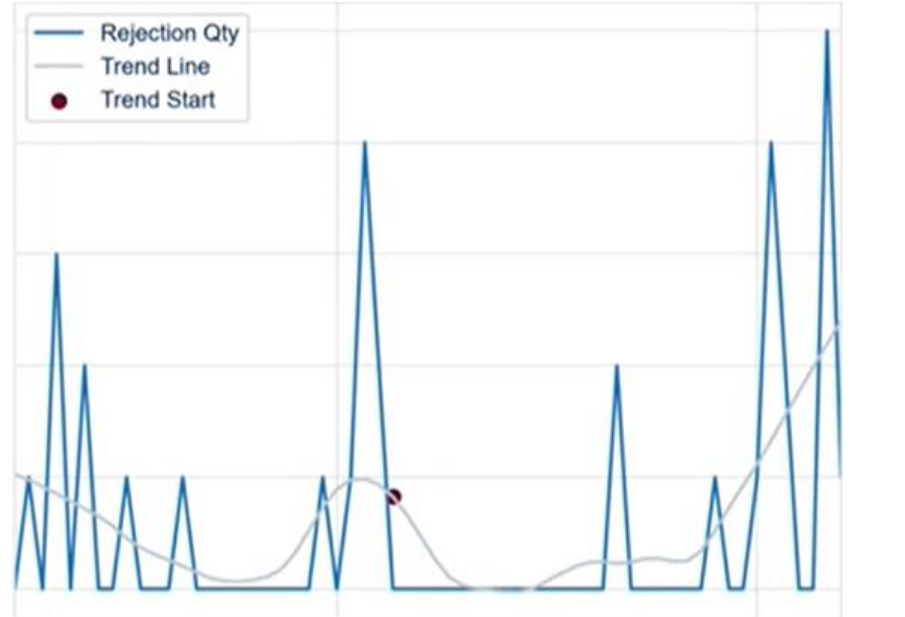
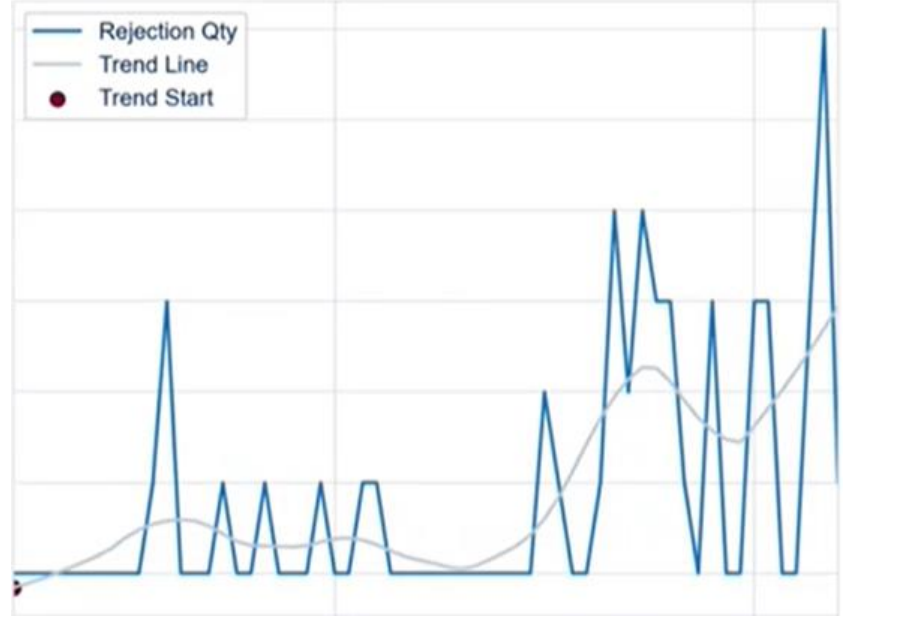
23		
P1	<p>Den würde ich wahrscheinlich in der Tat erstmal rausnehmen. Der hat nen ähnlichen Peak wie der vor ein paar Wochen schon gehabt hat, während er in den letzten Tagen nichts mehr gehabt hat. Ähm wenn jetzt wieder eine häufigere Frequenz auftreten sollte, glaube ich, wird er würde er ohnehin wieder hochgespült werden, sodass ich ihn aktuell ich betrachten würde.</p>	<p>- Wiederholt Zeitraum</p>
24		
P1	<p>Ja, den würde ich mir auch anzeigen lassen. Da waren wir sonst immer auf nem sehr geringen Niveau. Und dann vor ein paar Tagen sind wir hier deutlich nach oben geschossen, auch mit einer mit häufigeren Frequenz.</p>	<p>+ Niveau Magnitude Frequenz</p>

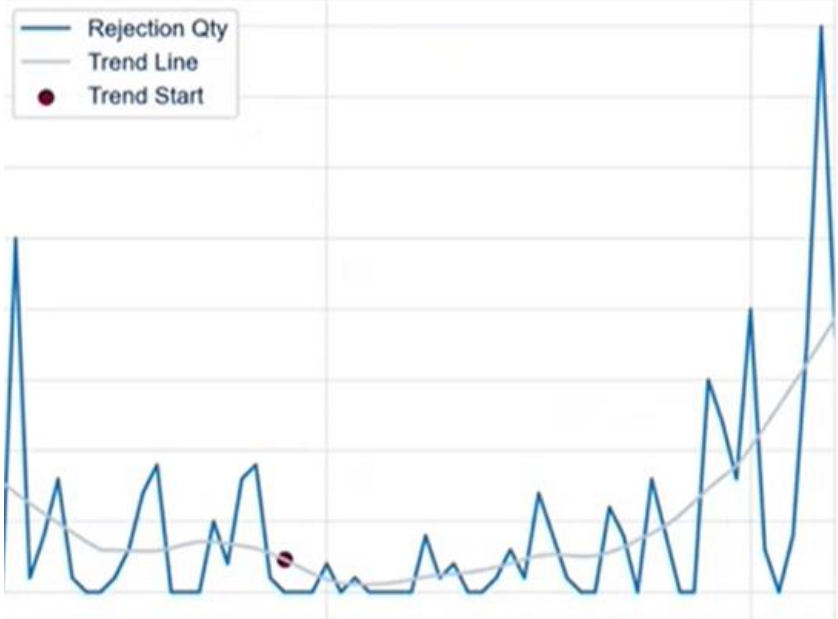
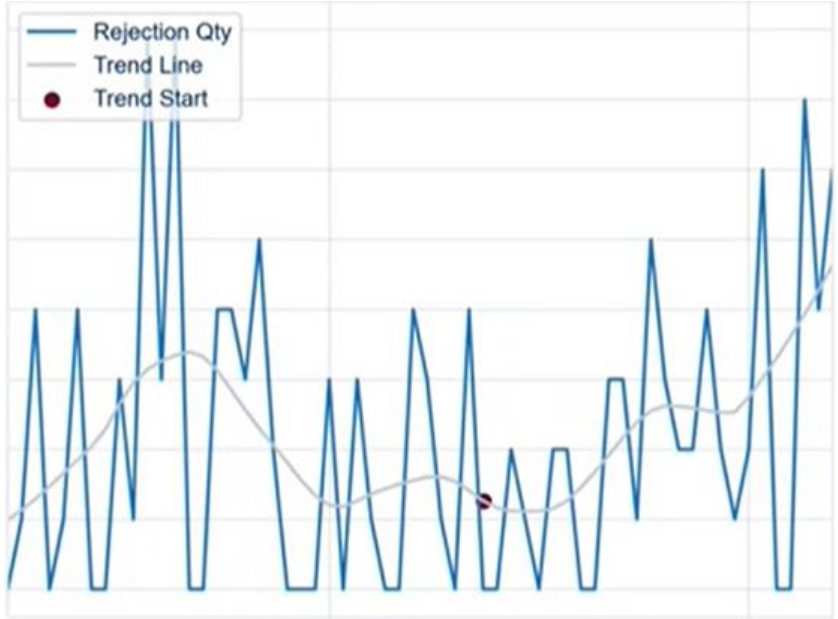
25		
P1	<p>Den würde ich wahrscheinlich erstmal, erstmal raus lassen, ähm der scheint immer mal so ein Peak zu haben. Klar war jetzt der Peak von vor einer Woche oder so etwas höher, aber aktuell würde ich ähm geht es wieder runter, und wenn es wieder wendet, wenn er wieder anschlägt, taucht er wieder auf; aktuell würde ich ihn nicht beachten.</p>	<p>- Wiederholt Peak Zeitraum Abfallend</p>
26		
P1	<p>Das ist ein spannender Code [lacht]. Den ich mir in der Tat anzeigen lassen würde, aus dem Grund ähm, der scheint sehr konstant vor, vor einem Monat immer aufgetreten zu sein, dann war jetzt n Monat fast Ruhe und nichts mehr gewesen, wo es durchaus sein könnte, dass wir irgendwelche Maßnahmen eingeführt haben oder wurden oder ähnliches. Und jetzt sehe ich aber, dass ich wieder einen Anstieg hab vielleicht zurück zu dem zu dem alten Fehlerbild insofern finde ich den interessanteren würde drauf schauen.</p>	<p>+ Konstant Ruhe Zeitraum Maßnahme Anstieg</p>

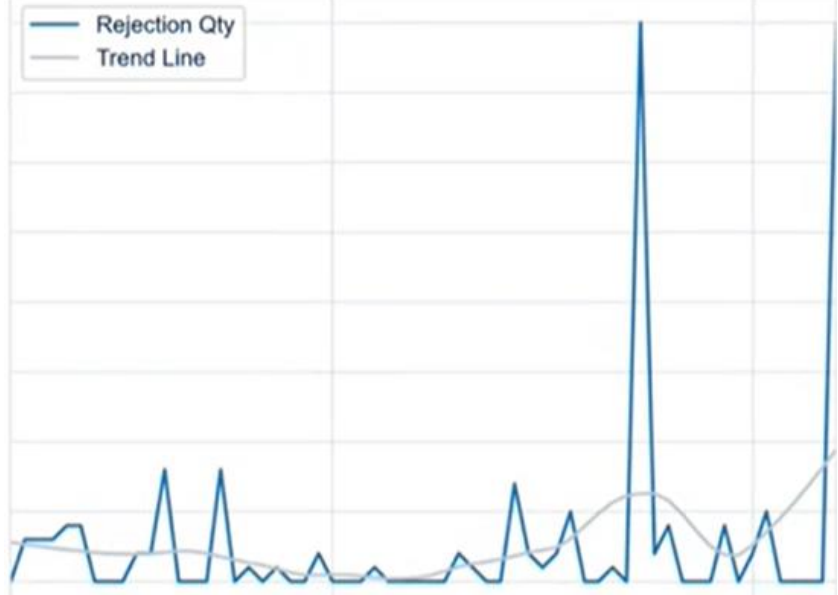
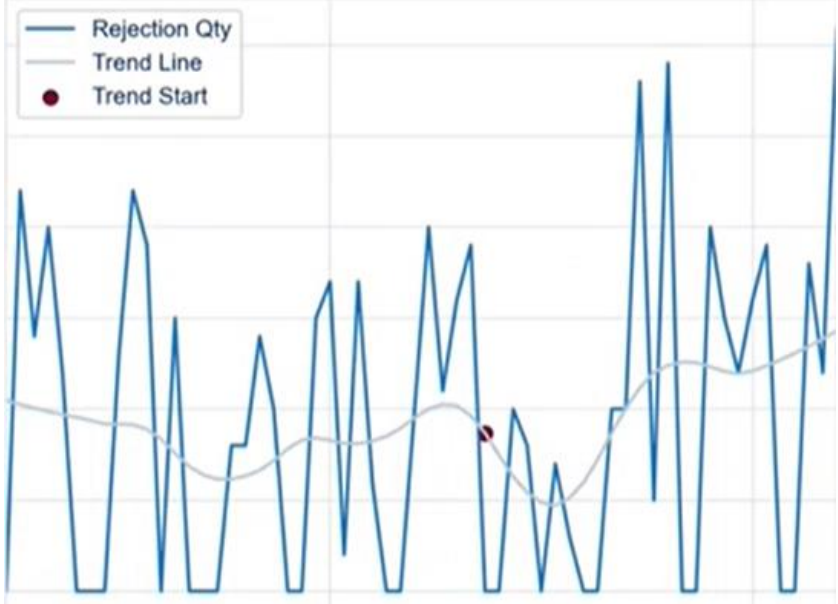
Block „Wissensextraktion“ (Person 2)

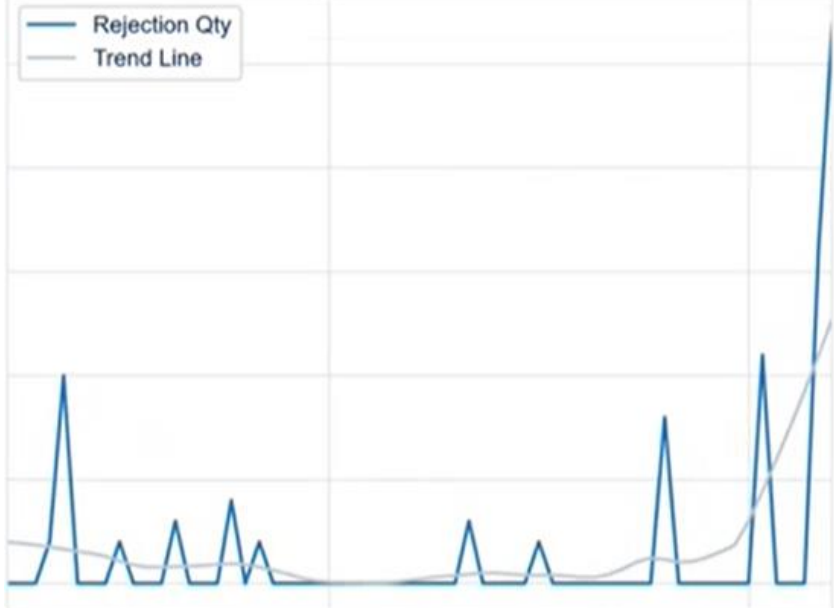
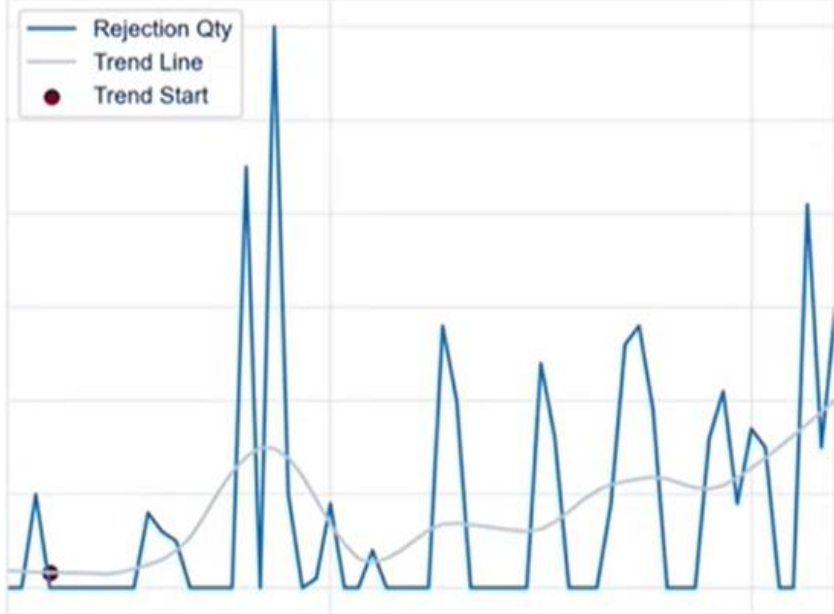
ID	Inhalt / Bild	Kodierung
27		
P2	<p>Nicht interessant, weil das meiner Meinung nach ein ja ein Peak Thema ist also, das heißt, da ist eventuell irgendwie ein gesamter Auftrag oder ne Chargen oder irgendwie sowas ausgefallen und das ist meiner Meinung nach kein Trend wo ich jetzt erwarten würde, dass da jetzt noch mehr ausfällt.</p>	<p>- Peak Vorhersage</p>
28		
P2	<p>Also ne, wenn ich jetzt die Grafik noch mal genau angucke ist das nicht interessant. Man sieht natürlich nen, also ich würde da nen Trend sehen, dass es jetzt auf jeden Fall nach oben geht, aber wir reden hier von 1 auf 3 und insofern ist das jetzt noch keine relevante Anzahl an Rückweisungen, damit das tatsächlich irgendwie interessant wird.</p>	<p>- Anstieg Magnitude Relevanz</p>

29		
P2	<p>So, ja ähnlich wie eben, grundsätzlich würde ich schon einen Trend erkennen, aber wir bewegen uns hier immer noch also dafür, dass jetzt ein Zeitraum von wie was ist das denn Zeitraum? 10 Tage sind das schon da, 2 Monate ah genau. Ja, also, der nimmt, tendenziell würde ich einen Trend sehen, definitiv, aber auch hier ist sind die Zahlen noch zu zu niedrig, als dass man da jetzt wirklich sagen würde das ist jetzt interessant.</p>	<p>- Trend Zeitraum Magnitude</p>
30		
P2	<p>Ist echt schwierig. Nicht interessant, weil es eindeutig nach einem Peak aussieht. Also als Trend nicht interessant, weil es aussieht wie ein, wie ein Peak. Auf der anderen Seite möchte ich wissen, was da los ist. Ich weiß nicht wie ich das jetzt anders ausdrücken soll, also das das ist so, der kommt aus dem Nichts und hat auf einmal 40 Rückweisung insofern, es ist für mich noch kein Trend weil das jetzt irgendwie nicht schleichend kommt, aber durchaus interessant.</p>	<p>- Peak Ursache Erstmalig Nicht schleichend</p>

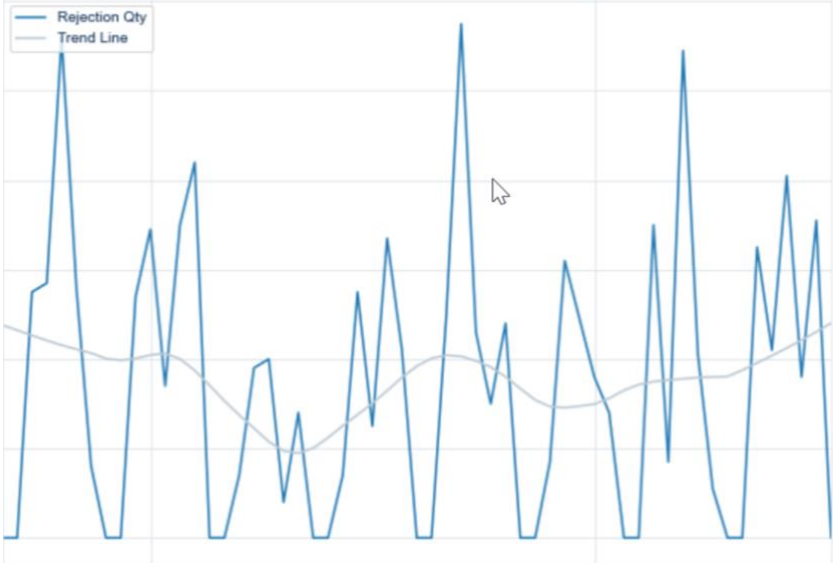
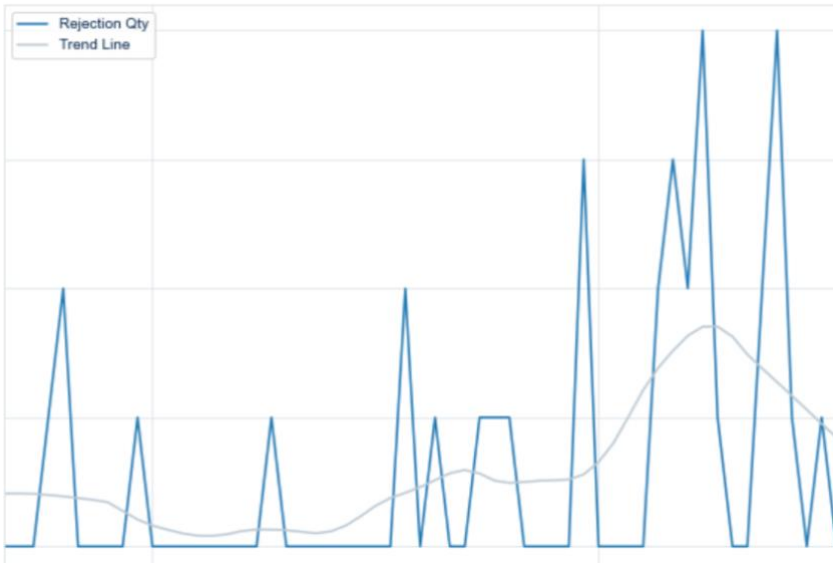
31		
P2	<p>Nicht interessant. Das das Bild ist, ist ein regulärer Trendverlauf von geht ein bisschen hoch, geht ein bisschen runter, geht wieder ein bisschen hoch. Zahlen sind zu niedrig, nicht interessant.</p>	<p>- Konstant Normal Magnitude</p>
32		
P2	<p>Interessant auch wenn, auch wenn es jetzt hier die, die Zahlen immer noch relativ niedrig sind. Da haben wir jetzt schon sozusagen alle 2, 3 Wochen eine Erhöhung. Und das geht jetzt sogar hier in dem aktuellen Zeitraum sogar noch mal nach oben. Das wäre für mich durchaus interessant. Das sieht nach so einem Trend aus, der irgendwie in die falsche Richtung geht, wo man vielleicht mal drauf gucken sollte.</p>	<p>+ Magnitude Monotonie Zeitraum</p>

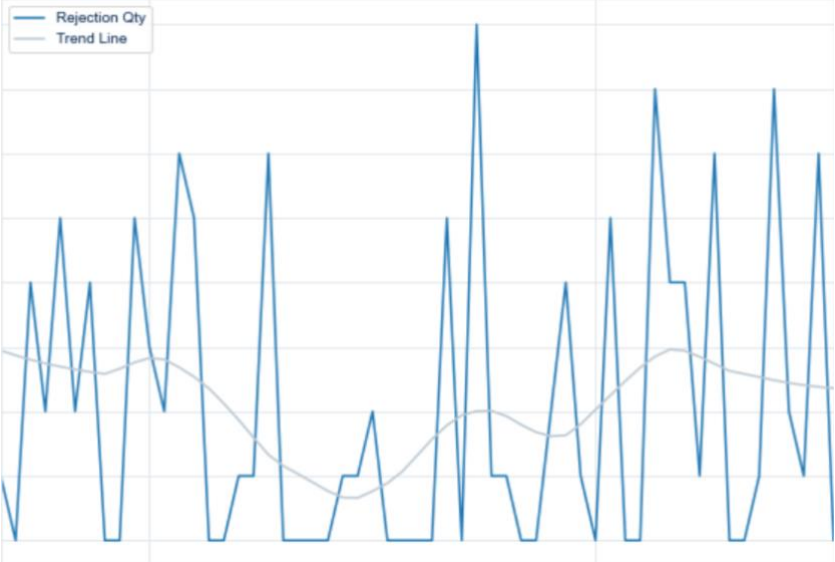
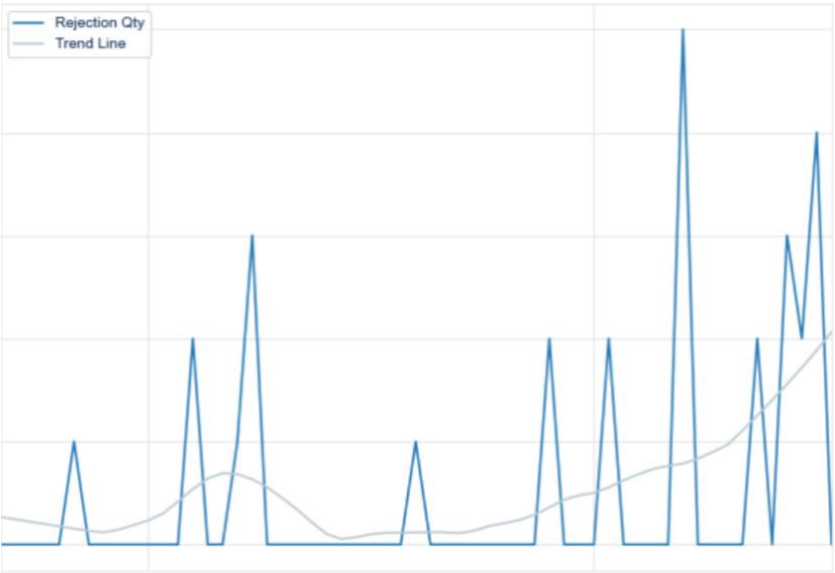
33		
P2	<p>Interessant! Einfach aus dem Grund, dass das zwar auch ein relativ typischer Verlauf ist allerdings diese U-Form irritiert mich ne, du hast irgendwie am Anfang von der Zeitspanne hast du einen Peak, wenn du jetzt 2 Tage später guckst, wirst du den jetzt vielleicht nicht mehr sehen und hättest dann halt aber wieder diesen Peak der nach oben geht, dann muss irgendwie, weiß nicht, was da ist. Auf jeden Fall interessant. Durch diese, die U-Form ist interessant.</p>	<p>+ U-Form Zeitraum Peak Ursache</p>
34		
P2	<p>Nicht interessant, zu wild, nein das ist kein Trend.</p>	<p>- Wild</p>

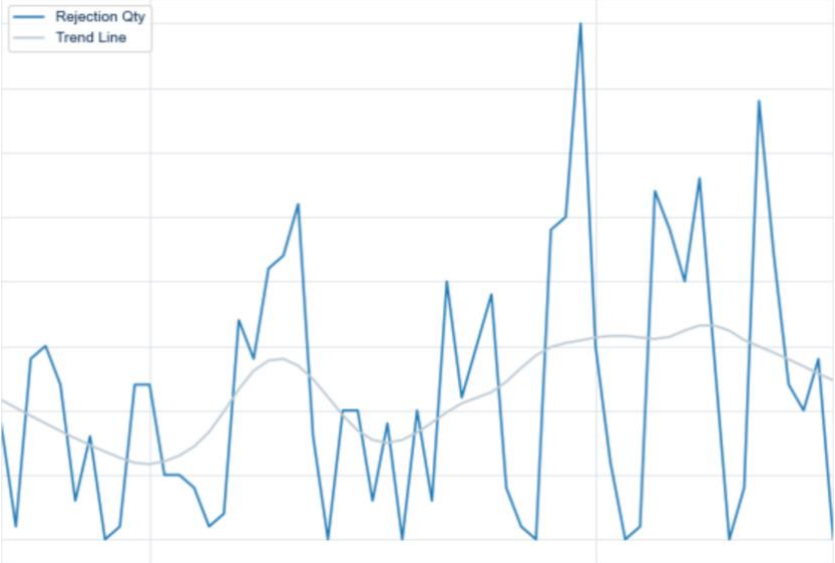
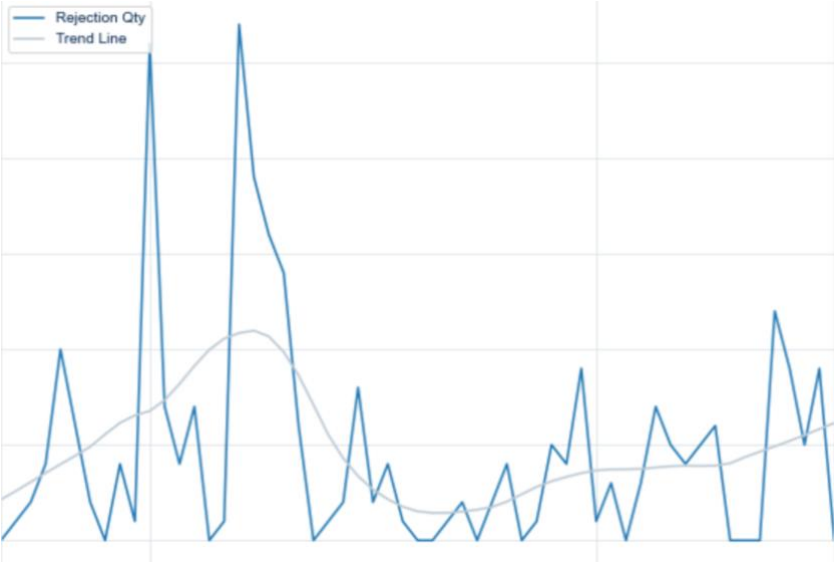
35		
P2	<p>Interessant aber nicht, weil es ein Trend ist, sondern weil es halt, eben weil es 2 extreme Peaks sind die da halt raus Stechen in einem relativ kurzen Zeitraum. Interessant. Aber kein Trend.</p>	<p>+ Peak Extrem Zeitraum</p>
36		
P2	<p>Nicht interessant das ist zu wild. Nein, da ist nix zu erkennen.</p>	<p>- Wild</p>

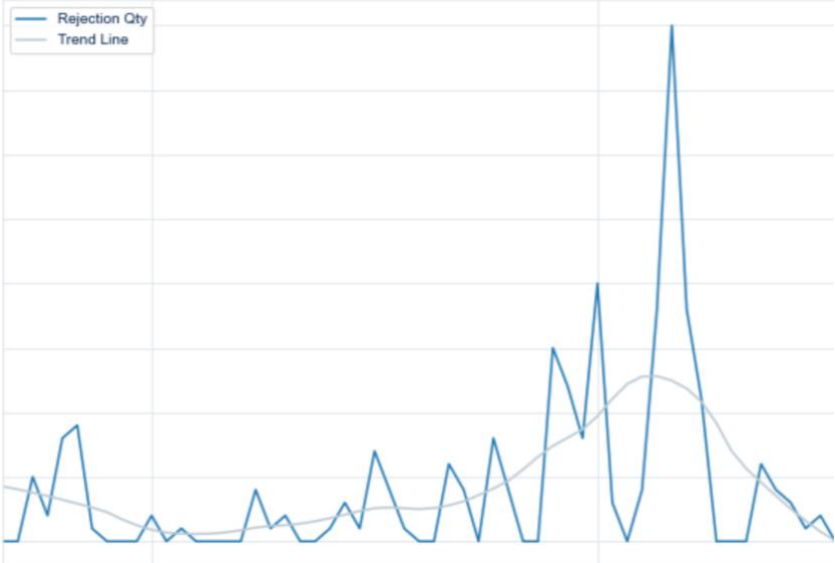
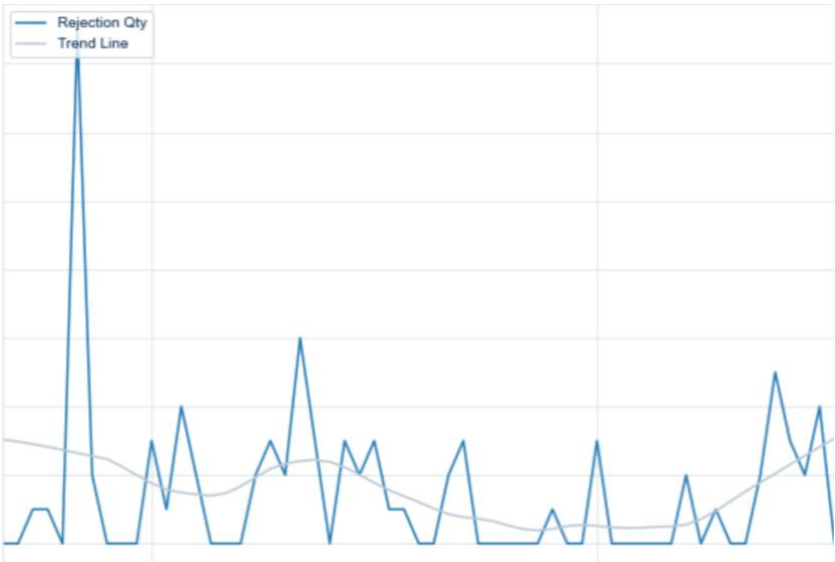
37		
P2	<p>Da haben wir wieder diese U-Form. Die würde ich jetzt durchaus als Trend bezeichnen und auch als interessant. Und es geht hoch, dann geht es irgendwie wieder runter. Jetzt geht es wieder hoch. Das innerhalb von 2 Monaten so. Irgendwas muss doch zwischen den zwischen diesen beiden Daten irgendwas anders sein, als sozusagen da. Das wäre auf jeden Fall interessant.</p>	<p>+ U-Form Abstieg Anstieg Zeitraum Ursache</p>
38		
P2	<p>Ehm. Das würde ich jetzt sagen ist als Fehler natürlich grundsätzlich interessant, aber jetzt nicht als Trend. Du hast da den Peak, dann hast du den zweiten, hoch runter, Verlauf mit ähnlichen Mengen. Würd' ich als Trend würde ich jetzt sagen nicht interessant.</p>	<p>- Konstant Peak Regel- mäßig</p>

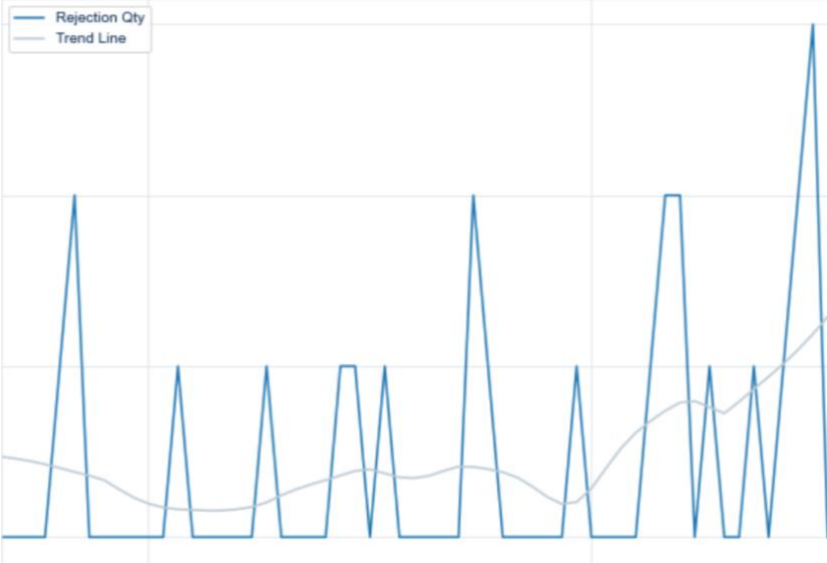
Block „Wissensextraktion“ (Person 3)

ID	Inhalt / Bild	Kodierung
39		
P3	<p>Ja nee würde ich sagen ist nicht interessant weil es geht halt hoch und wieder runter und bewegt sich aber auch halt auf so einem mittleren Niveau. Na also es geht jetzt nicht wirklich, also klar das geht da so ein bisschen hoch vielleicht tendenziell aber so vom Niveau her wo es auch schon mal war, ne, also jetzt nicht höher als sonst würde ich sagen also nein</p>	- Schwankung Niveau
40		
P3	<p>Hm ja, so hätte mich interessiert, wahrscheinlich also so da um den siebten rum wo halt der Peak ist aber jetzt würde ich sagen es ist ja schon wieder am runter gehen, ne, also die geht dann nach unten zuletzt war jetzt nicht so viel los würd ich sagen, also vielleicht ein bisschen früher dann anzeigen, also so wenn es halt wirklich noch am hochgehen ist, ja.</p>	- Peak Abstieg Zeitraum

41		
P3	<p>Nee auch nicht der, also das sieht jetzt so aus als würde der da wieder so ein bisschen hochgehen, aber eigentlich nur weil er da mal kurz unten war, ne, also keine Ahnung was da los war, warum da in letzter Zeit keine Fehler waren aber sonst würde ich sagen ist er jetzt eher wieder auf seinem normalen Niveau so, jetzt nach dem hochgehen ja da hinten schon wieder leichter am runter gehen also nee auch nicht interessant.</p>	<p>-</p> <p>Ursache Niveau Normal Schwankung</p>
42		
P3	<p>Ja den doch den würde ich mir angucken weil also das ist jetzt immer noch auf einem niedrigen Niveau und kann natürlich sein das ist jetzt irgendwie gerade Zufall war hatte jemanden schlechten Tag oder so aber äh sind die Spitzen schon noch ein bisschen höher geworden ne und öfter also sieht dass die Linie schon nach oben geht ne also kann sein dass sie jetzt auch wieder runter gehen würde aber sollte mal vielleicht im Blick behalten</p>	<p>+</p> <p>Niveau Ursache Magnitude Frequenz Steigung Prognose</p>

43		
P3	<p>Also vor einer Woche da wäre, bis dahin wäre er wahrscheinlich interessant gewesen oder hätte man gesagt am hochgehen, oder vielleicht geht es jetzt auch nochmal hoch ne dann ist auch wieder interessant aber jetzt gerade hätte ich gesagt nee da ist er jetzt der abflachen also das scheint irgendwie und wieder unter Kontrolle zu sein aber ja klar zwischenzeitlich ne ist ja ist ja ganz schön hoch gegangen also klar jetzt auch nicht so super ungewöhnlich hoch aber so dass man sich vielleicht angeschaut hätte aber jetzt ist ja sieht so aus als würde er wieder runter gehen</p>	<p>- Zeitraum Anstieg Prognose Abflachen Abstieg Ursache Extrem</p>
44		
P3	<p>Hm schwierig also ich würde natürlich sagen erstmal ist der schon interessant ne weil der da mal so hoch war aber andererseits ist er natürlich jetzt meilenweit davon entfernt also von den Spitzen da früher. Vielleicht wenn man genau hinguckt ist er ja vielleicht wieder so leicht am hochgehen aber würde ich jetzt erstmal nicht als relevant betrachten weil er eben halt noch so weit unten ist ne also wenn der jetzt weiter hochgeht, weiter in Richtung da wo er damals war, dann ja, aber jetzt würde ich sagen erstmal abwarten ob der da stabil unten bleibt</p>	<p>- Zeitraum Wiederholt Maximum Anstieg Niveau Stabil</p>

45		
P3	<p>Oh ja ja klar also voll gut 10 Tagen ne also da hätte man den aber wäre er wahrscheinlich auch angezeigt worden also als er da so klar am hochgehen war, ne also am Anfang hätte man wahrscheinlich noch einmal abgewartet ob das so zufällige Schwankungen sind aber dann ne da geht es dann doch richtig schnell nach oben ist ja fast wie so ne Rampe ähm, aber jetzt ist er ja scheint ja unter Kontrolle zu sein also das ist ja wirklich wieder weit runtergegangen und weiter am runter gehen, ähm ja, das würde ja dann wahrscheinlich wieder auftauchen, ne, dann wenn dann nochmal was passiert, also erstmal nicht interessant.</p>	<p>- Zeitraum Anstieg Schwankung Steigung Abstieg Prognose</p>
46		
P3	<p>Nee würde ich jetzt nicht sagen also ja vielleicht geht er dann leicht hoch, aber das ist glaub ich ne normale Schwankung, also der ist jetzt ja eigentlich nicht höher als sonst, er hatte davon meinen Peak aber jetzt ist er ja eigentlich in einem Bereich wo man sagt das ist Zufall da scheint jetzt nicht groß was im Argen zu sein, also würde ich jetzt sagen würde ich jetzt nicht unbedingt sehen wollen.</p>	<p>- Anstieg Schwankung Niveau Peak Ursache</p>

47		
P3	<p>Uff also bei dem weiß ich nicht, ehrlich gesagt, also klar sieht das jetzt so ein bisschen aus als würde das hochgehen, der letzte Peak da ist 1 höher als die davor immer waren, aber das könnte jetzt halt auch Zufall sein ne, also vielleicht geht er jetzt runter und dann ist eine Woche lang gar nix könnte ich mir jetzt auch vorstellen, weil das jetzt wirklich einfach auf so einem Niveau, da ist immer mal einer oder mal 2, also sieht jetzt für mich nicht so aus als würde das wäre jetzt ein Problem aufgetreten. Könnte ich jetzt eher unter Zufall verbuchen, außerhalb wenn das jetzt wirklich weiter nach oben gehen würde also wenn jetzt dann nochmal nächste Woche jetzt gleich ein höherer käme dann würde ich mir nochmal anschauen aber jetzt erstmal auf dem Niveau glaub ich ist da nicht unbedingt Handlungsbedarf</p>	<p>°</p> <p>Anstieg Peak Ursache Prognose Niveau Konstant Ursache</p> <p>Zeitraum</p> <p>Handlung</p>

Block „Trend“ (Person 1)

Person	Inhalt	Kodierung
I	Okay. Dann könnten wir zum letzten Block kommen, dem Trend. Wir hatten ja uns bereits schon mal ein paar kurz zusammen angeschaut. Die Idee ist jetzt da nochmal ein bisschen in die Tiefe zu gehen. Also Trends sollen ja automatisiert erkannt werden, mein Ziel ist es, in einem ersten Schritt die Intuition oder das fachliche Verständnis der Mitarbeiter in ein regelbasiertes System zu gießen und das dann eben im nächsten Schritt zu nutzen, um Verläufe vorzufiltern. Vielleicht mal ganz allgemein gefragt, aus fachlicher Sicht, welche Kriterien ziehst du denn heran, um dann zu beurteilen, wenn du einen visuellen Trendverlauf von dir hast?	Fachexpertise Regelbasiert Kriterien
P1	Hm, na, die wichtigen Kriterien für mich um einen Trendverlauf zu sehen, ist natürlich, dass ich, sag mal erst mal ganz allgemein gesprochen, dass sich der der Mittelwert über die Dauer, dass der ansteigt, also die Fehlerhäufigkeit, das ist glaub ich so das "1A" Kriterium irgendwo, dass ich, dass ich einen ansteigenden Mittelwert habe. Dann sind durchaus Kriterien interessant wie ich hab einen Fehler gar nicht und habe ihn hab ihn plötzlich hab ich plötzlich in einer starken Ausprägung, das was vielleicht in Richtung "single event" geht. Genauso sag ich mal, wenn die Frequenz der Fehlerhäufigkeit ansteigt, dass kann ein Punkt sein was, was interessant ist, weil ich das natürlich mit dem gemittelten Mittelwert auch wieder rausfinde. Ja, und dann ob ein Trend, sag ich mal oft folgen Trends, vielleicht auch ner aufsteigenden absteigenden Bewegung, die aber trotzdem über Dauer steigend ist. Das heißt da irgendwo so ne, so ne Gerade oder Linie zu haben, die dir das anzeigt eben nicht immer nur den absoluten Trend wäre für mich wichtig.	Kriterien Mittelwert steigt Erstmalig Single Even Frequenz Schwankung Monotonie
I	Vielleicht da gleich mal nachhakend, wie sehr gewichstest du äh, quasi in deinem Kopf, bei der Betrachtung von so einem Trendverlauf, Ausreißer nach unten oder nach oben. Also würdest du da eher auf einer Skala von 1 – 5, wie weit würdest du eher den eigentlichen Verlauf wollen oder wie weit würdest du das in deinem Kopf zu einer zum gerade Trend oder zu nem wirklich glatten Trend glätten?	Glätten
P1	Ich würde die relativ stark glätten, in meinem in meinem Kopf, auf einer Skala von 1 - 10 vielleicht so 7 – 8.	Glätten
I	Okay, ja, okay, gut. Und wir hatten ja damals bei den Input Daten darüber gesprochen, dass wir vielleicht 2 Monate zurückschauen ungefähr. Jetzt kann man aber natürlich in den 2 Monaten ein Trend tendenziell angestiegen sein, aber in den letzten Tagen oder Wochen vielleicht eher wieder zurückgegangen sein. Was ist für dich so ein Zeitraum, den du heranziehen würdest, um wirklich die akute Relevanz von so einem Trend für dich zu beurteilen? Also wie ich die letzten wieviel Tage würdest du nochmal als besonders, ja, wichtig für die Relevanz heranziehen?	Zeitraum
P1	Also wenn ja, wenn wir 60 Tage nach hinten gucken, ich glaube, das sind die letzten 5 Tage ein ganz guter Punkt um dann nochmal ne besondere Relevanz rauszubekommen. So aus dem ersten Bauchgefühl.	Zeitraum
I	Gibt es aus deiner Erfahrung heraus gesprochen verschiedene Arten oder Kategorien von Trends? Die du identifiziert hast und wenn ja, hast du für die unterschiedliche Bewertungsmaßstäbe oder was ist für dich da besonders wichtig?	Muster Maßstäbe

P1	Genau, ich glaub ich hab damals schon mal so ein bisschen angedeutet, also durchaus, wir haben das was vielleicht so single events sind, das ist vielleicht kein klassischer Trend in dem Sinne ist aber einen sprunghafte spontanes Ansteigen eines Fehlerbilds wo ich, sag mal, das ist ein Trend, den ich vielleicht auch ohne ein automatisiertes Monitoring mitbekomme, weil es eben, weil es eben Aufruhr erzeugt, dass da auf einmal Fehler gehäuft oder überhaupt erstmalig auftritt. Ähm da, hab ich damals schon gesagt glaub ich, ist von allem die NC Ebene interessant, um hier ne schnelle Informationen zu bekommen und nicht durch vielleicht Ressourcenengpässe oder weil das erst 2, 3, 4 Tage später bearbeitet wird, dass es auffällt. Und das andere sind natürlich diese, diese stetigen Trends, das heißt das was so was langsam ansteigt, was von Woche zu Woche immer n bisschen mehr wird oder von Tage zu Tag, bis es dann vielleicht irgendwann mal eine Grenze reißen würde, wo jemand alarmiert ist, weil er, weil es eben in eine Top Liste es geschafft hat oder so, wo man aber in dem Fall schon früher eigentlich eine Information haben wollen würde, wo man sich immer wieder sagt, warum haben wir das nicht vorhergesehen? Dass es darüber über Wochen nach oben geht?	single events Sprunghaft Erstmalig Ebene Geschwindigkeit Stetiger Trend Monotonie Top-Fehler Frühwarnung
I	Wenn wir jetzt auf die auf die, was du “single events” genannt hast, schauen. Hast du da eine Mindestanzahl für die du sagst ab der wird es relevant, also lass mich kurz erklären: Sagen wir mal du hast immer so 3 oder 4 Fehler normal im Durchschnitt, wie viel höher muss es sein, damit du es als single event bezeichnen würdest? Oder stellen wir uns vor, wir haben einen Fehler, der in den letzten 3 Monaten nicht aufgetreten ist. Der würde jetzt ab einem Ausreißer von 1 ja ein single event darstellen. Dann würden wir uns allerdings alle Fehler, die nur ganz gelegentlich mal auftreten, sofort als single events angezeigt werden, in einem Trend Monitoring. Kannst du hier so "Pi mal Daumen" eine Grenze sagen, allgemein gesprochen, ich weiß das schwankt wahrscheinlich stark von Produkt zu Produkt, aber wo du sagen würdest, ab da ist für mich eine gewisse Relevanz erreicht?	single events Mindestanzahl
P1	Ja, also wie du schon sagst, das ist glaube ich, in der Tat sehr, sehr schwierig zu definieren, weil es einfach sehr auch von dem Produkt was gefertigt wird abhängt ne, das ist, wenn ich jetzt einen [Produkt A] fertige ist das was anderes als wenn ich [Produkt B] fertige, wo ich eine ganz andere Ausprägung habe, aber ich denke grundsätzlich es so ein so ein so eine Range von 5 - 10 mal ist glaube ich eine ganz gute Rage, um zu sagen da möchte man alarmiert werden.	Mindestanzahl
I	Okay, dann kommen wir jetzt zu einer relativ wichtigen Frage und zwar wie wichtig ist es dir denn, alle möglichen relevanten Trends gemeldet zu bekommen? Im Vergleich dazu nur die wirklich relevanten Trends gemeldet zu bekommen, also ich versuche es noch mal ein bisschen so zu formulieren, dass es vielleicht leichter ist, eine Antwort zu geben und zwar: Wie viele nicht relevante Trends würdest du in Kauf nehmen angezeigt zu bekommen, um einen relevanten Trend angezeigt zu bekommen.	Sensitivität

P1	[Lacht] Hm, wow. Ich glaube das. Ich, ich hole mal ein bisschen aus, OK. Vielleicht auch, um mir selber noch mal zu dem Thema klar zu werden. Grundsätzlich ist es, ist es ja als Arbeitshilfsmittel für die operative Ebene gedacht. Das heißt als zusätzlicher Hinweis "Da könnte was sein". Also zum einen bedeutet das natürlich, wir haben nicht den Anspruch mit dem Tool, dass wir alles finden muss, was potenziell relevant ist, weil wir wollen auch die, den Arbeitsaufwand ja in einem erträglichen Rahmen lassen, auf der anderen Seite tut's in einem gewissen Maße auch nicht weh weil ich relativ schnell sagen kann das interessiert mich nicht so was wegzuklicken oder zu überfliegen, dass ich, dass ich fast eigentlich eher darauf hinaus würde zu sagen dass ich, wenn ich mir, wenn ich mir pro Morgen 10 Trends angucke ist das glaube ich ein durchaus akzeptabler Rahmen. Natürlich muss ich eine gewisse Güte oder Qualität der Trends auch dahinter haben, so dass ich persönlich sagen würde, wenn es ein Verhältnis von 50:50 oder 70:30 ist ja, wäre das für mich gut, also 50 - 70% der Trends werden für mich. Relevant und 50 - 30% wären werden irrelevant.	Operativ Anspruch Arbeitsaufwand Akzeptabler Umfang Verhältnis
I	Du hattest gerade gemeint wenn du morgens eine Anzahl von 10 da hättest, dann wäre das akzeptabel. Würdest du das dann sinnvolles Herangehen finden also eine Maximalgrenze vielleicht festzulegen und daran betrachtet die Auswahl zu treffen?	Akzeptabler Umfang Maximalgrenze
P1	Ja, ich würde glaub ich nicht eine Maximalgrenze machen, das würde ich in der Tat nicht machen, weil das könnte zulasten der Qualität sein, aber ich finde das ist eine ganz gute Ziel-Steuergröße, sag ich mal um, die die Sensitivität der Alarme bisschen zu justieren. Zu sagen ok, wenn ich merke jetzt ich hab so viel ausgeschlossen, dass mir nur noch jeden Tag 3 angezeigt werden, dann kann ich durchaus ein bisschen die Sensitivität hoch drehen. Wenn ich aber merke ich bekomme jeden Tag 50 Alarme, dann wird das keiner da mehr bearbeiten, dann sollte ich die Sensitivität ein bisschen runter drehen, aber ich glaube nicht als alleiniges Merkmal zu sagen, ok auf gar keinen Fall mehr als 10.	Sensitivität Anpassen
I	Ok, das heißt, äh, wir sprechen dann aber von 10 pro Bereich oder pro Produkt?	
P1	Ja, vielleicht, das müsste, das müssen wir uns dann nochmal anschauen. Das könnte durchaus vielleicht sogar fallspezifisch sein. Ich sag mal für den Fertigungsbegleiter, der wirklich der sich wirklich nur 10 anguckt, weil der nur einen Fertigungsbereich oder eine Produktkategorie anguckt oder 3, da könnte es sogar pro Produktkategorie sein. Bei einem, der sich eine komplette Produktlinie anguckt glaub ich wäre es zu viel, wenn 10 pro Produkt wären. Das müssen wir halt nochmal anschauen. Ich denke, ich denke, wenn er in Summe irgendwas so 10 - 20 hat wäre fein.	Umfang
I	Okay, gibt es noch was zum Thema Trend? Was du noch vergessen hast zu sagen oder gerne noch ergänzen würdest?	
P1	Ich glaube was, ist mehr noch so ein generelles Statement noch zum Thema Trend. Wir sehen es immer wieder und merken es immer wieder, Trend ist nicht eindeutig, Trend ist nichts, was sich hart definieren lässt, das ist sehr viel persönliche Intuition, hat sehr viel mit persönlicher Veranlagung oder Denke zu tun. Insofern, wenn wir da was hinbekommen, wo wir was Lernendes haben, wo wir so ein bisschen auf die Intuition der Mitarbeiter mit abbilden, ist uns viel geholfen ja, es ist einfach, es ist nicht schwarz/weiß, Trend.	Individuell Intuition Lernen
I	Sehr gut, dann vielen Dank, das hilft mir sehr weiter.	

Block „Trend“ (Person 2)

Per-son	Inhalt	Kodierung
I	Hallo [Person 2]. Also wir möchten ja ein Trend Monitoring implementieren und das Ziel ist es dabei, wie der Name schon sagt, die Trends automatisiert zu erkennen. Jetzt würde ich gerne mit dir nochmal ein bisschen darüber reden, wie du aus fachlicher Sicht oder intuitiv nen Trend definieren würdest. Vielleicht erst mal ganz allgemein gesprochen, wie gehst du denn ran, wenn du einen Trendverlauf optisch siehst? Um zu bewerten, ob es sich dabei um den für dich relevanten Trend handelt oder nicht.	Visuell
P2	<p>Also grundsätzlich ist für mich halt ein Trend, wenn über eine gewisse Zeitspanne tatsächlich, zu erkennen ist, dass sich jetzt in dem Fall die die Fehlermenge tatsächlich auch erhöht, also dass man, dass man sagt ok, man hat halt dieses diese typischen es geht hoch runter und so weiter, aber das es halt insgesamt halt mehr wird ne also dass die, dass die Menge höher wird das wäre für mich, ja, wenn wenn es in die Richtung geht {gestikuliert mit dem Finger eine Linie nach oben} und dann tatsächlich nicht nur irgendwie 1, 2 hoch geht, sondern tatsächlich auch irgendwie ein bisschen, ja die Sprünge ein bisschen höher sind, dann wäre das für mich auf jeden Fall schon mal, ich meine, ein relevanter Trend.</p> <p>Genau so, genauso in die andere Richtung. Wenn Fehler dann wieder verschwinden, wäre das für mich dann halt auch ein Trend, dann halt eben ein absteigender Trend. Also, solche Arten von Trend sind auf jeden Fall immer interessant, weil du immer irgendwie was, anscheinend irgendwie kommt ein Einflussfaktor rein, der dazu führt, dass halt eben die Rückweismengen höher werden.</p> <p>Ansonsten hast du bei Trends oder bei Fehlermengen bezogen auf die Zeit bestimmte Muster, würde ich sagen, die man durchaus betrachtet immer wieder, also das heißt wenn ja, du hast einzelne Ansteigungen oder sowas was, du hast irgendwie ganz wenig Fehler auf einmal hast du ganz viele, so genannte single events, die durchaus auch interessant sind. Meistens werden solche single events aber relativ schnell erkannt und auch recht schnell behoben beziehungsweise die Fehlerursache ist relativ schnell bekannt. Da steckt nicht so viel Analyseaufwand hinter, dann sind natürlich solche Muster, das was wir jetzt eben hatten, die sind so ne “U”s wo du sagst ok, du hast ne längere Zeitspanne und da hast du so eine Art Peak wo es nach oben geht dann geht es wieder runter, also so ein Trend, dann geht es wieder runter und dann geht es irgendwann wieder hoch. Das heißt irgendwie da hast du ja auch wiederkehrende Einflüsse. Das sind durchaus auch interessante Trends, die man sich da angucken kann. Genau und das halt irgendwie immer mit Blick auf einen längeren Zeitraum. Einfach weil du halt durch diese täglichen Schwankungen sowas halt eben schwer beurteilen kannst und deswegen muss man da immer ein bisschen weiter weg gucken, damit man dann halt eben diese Muster im Grunde genommen erkennt. Also für mich sind eigentlich die wesentlichen Trendarten, die für mich interessant sind diese “U”s, und halt längerfristiger Trend, der nach oben geht, oder nach unten beziehungsweise halt Peaks, wo du halt wirklich irgendwie mit einem Mal ganz viele Fehler hast und das dann halt auf einmal nach oben schießt.</p>	<p>Menge</p> <p>Anstieg</p> <p>Steigung</p> <p>Sprünge</p> <p>Richtung</p> <p>Ursache</p> <p>Muster</p> <p>single events</p> <p>Ursache</p> <p>U-Muster</p> <p>Ursachen</p> <p>Zeitraum</p> <p>Schwankung</p> <p>Stetige Trends</p> <p>Peaks</p>

I	Mhm, du hast jetzt du hast das schon so ein bisschen angesprochen, aber ich wollte nochmal explizit nach diesen nach den Schwankungen oder diesen Ausreißern nach oben oder nach unten fragen. Also wenn du jetzt dir vorstellst, dass du das im Kopf, wenn du dir das anschaust, also du sagst: "Das ist jetzt ein Ausreißer, den werde ich jetzt irgendwie weniger berücksichtigen", quasi mental ein bisschen Glätten, wie stark würdest du sagen, vielleicht auf eine Skala von 1 - 10 rechnest du diese Ausreißer, diese Schwankungen im Kopf raus oder sagst du es ist eben auch relevant, wenn wir uns jetzt so einen Verlauf anschauen?	Schwankung Ausreißer Glätten
P2	Ja also, wenn ich, wenn ich mir einen Trendverlauf angucke, also wenn ich wirklich einen längeren Zeitraum angucke und da seh' ich so einen Peak, dann ist es für mich relevanttechnisch, wird der glaube ich mit einer 9 rausgerechnet, weil ich eigentlich davon ausgehe, dass auch die operative diese Peaks halt sehr schnell erkennt und sehr schnell abschließt. Insofern würde ich da jetzt nicht die Notwendigkeit sehen, dass man da jetzt zusätzlich noch irgendwie reagiert. Ja, beantwortet das die Frage?	Peak Glätten Maßnahme
I	Ja, du hattest es ja davor schon recht deutlich gesagt. Wenn du jetzt einen Trend hast, der tendenziell nach oben geht. Wie viele Tage schaust du in die Vergangenheit, um zu sagen, ob der gerade akut für dich relevant ist? Also sagen wir mal, es geht vielleicht seit einem Tag runter oder seit 3, 5, 7 Tagen runter, ab wann würdest du sagen, okay, das ist jetzt ein Trend für mich, schon in die andere Richtung, den muss ich vielleicht akut gar nicht betrachten, also welchem Zeitraum würdest du nochmal so für die akute Relevanz berücksichtigen?	Zeitraum
P2	Kommt immer ein bisschen drauf an. Aber tatsächlich ich glaube, wenn du, wenn du jetzt ein Trend hast, der nach unten geht, also wo du, wo du irgendwie ein Peak hattest, der wo es nach oben gegangen ist und jetzt geht es irgendwie gerade wieder runter ich würd dem glaube ich nicht vor mindestens 2 Wochen trauen, dass der tatsächlich komplett weg ist oder dass er dann komplett runter gegangen ist. Aber ich denke so die diese üblichen Schwankungen kannst du wöchentlich schon ganz gut beurteilen und wenn du jetzt irgendwie einen Trend siehst und du siehst der geht runter, dann hast du ja schon eigentlich irgendwie so ne Woche oder so was wo du siehst, dass der runter geht. Wenn du da nochmal irgendwie ein, 2 Wochen drauf packst, dann kannst du tatsächlich sehen ist er dann weiter runtergegangen oder kommt er eventuell wieder ne, weil das ist ja auch ein typischer Verlauf, dass du, es geht nach oben. Dann geht es kurzzeitig wieder nach unten, weil alles in panisch alle sind super aufmerksam, da muss irgendwas gemacht werden, dann verändert sich irgendwas. Und dann ist einer wieder krank und dann kommt der nächste und der prüft auf einmal wieder schärfer und dann geht es wieder nach oben. Also das sind ganz typische Verläufe eigentlich, das ist ein bisschen runtergeht und dann wieder hoch. Dann vor 2 Wochen würde ich glaub ich nie sagen, der Trend ist jetzt abgeschlossen.	Absteigend Zeitraum Wöchentlich Typisch Schwankung
I	Wenn du jetzt davon ausgehst, dass es ein tägliches Monitoring ist, also du quasi solange der Trend nach oben geht, dann auch jeden Tag den Trend angezeigt bekommen würdest. Würdest du in dem Zeitraum, in dem er nach unten geht, trotzdem sehen wollen?	Zeitraum
P2	Würde ich, also ich würde ihn auf jeden Fall jeden Tag sehen wollen, solange er nach oben geht, definitiv. Und ich würde ihn wahrscheinlich, also jetzt ja, ich persönlich würde ihn wahrscheinlich auch noch gerne sehen wollen, wenn ich schon so ein Tool habe, was mir das automatisch anzeigt, dann würde ich ihn glaub ich auch noch irgendwie ein paar Tage, 3, 4 Tage oder sowas oder dann bis zum Ende der Woche, je nachdem tatsächlich auch beim Runtergehen mir angucken wollen.	Ansteigend Absteigend

I	Du hast auch schon ein bisschen über bestimmte Muster oder Kategorien gesprochen und auch gesagt, dass da verschiedene Bewertungsmaßstäbe ein Stück weit für dich gibt, kannst du noch ein bisschen ins Detail gehen, welche Kategorien oder Muster du da identifiziert hast und wie du die, oder ob du die unterschiedlich bewertet?	Muster
P2	Ja, also für mich sind die Hauptmuster halt tatsächlich, würde ich jetzt sagen, Peaks, so das ist halt wirklich irgendwie der Fehler tritt eigentlich gar nicht auf und auf einmal hast du 80 Rückweisungen oder sowas. Das heißt, die würde ich jetzt nicht als, wie soll ich sagen, konkret gefährlich beurteilen, weil wie gesagt, das ist eine Ausnahmesituation, wo irgendwas schiefgelaufen ist. Vielleicht ist ein [Tablett] runtergefallen oder wie auch immer oder da irgendein Mitarbeiter hatte nen schlechten Tag, und hat den ganzen Tag [Fehler C verursacht]. Das sind Sachen, die lassen sich immer relativ schnell klären. Insofern sind die, wenn ich jetzt sagen würde 10 ist absolut relevant und 0 ist nicht relevant, wären Peaks für mich bei weiß nicht 2 oder 3. Dann hast du Trends, die sehr, sehr plötzlich steigen, also die gehen wirklich ad hoc hoch und bleiben dann auch, also schwanken auch, aber bleiben dann auf einmal relativ weit oben, das sind, das sind auf jeden Fall auf einer Skala, dann die würden schon bei 9 liegen, weil da muss irgendwas Massives passiert sein, dass das auf einmal also irgendeine Veränderung muss es gegeben haben die, die da die dazu geführt hat. Dann, ja dann gibts diese Trends, diese "U"s also wo du wo du irgendwie ein Anstieg hast auf eine gewisse Menge dann geht es wieder runter, oftmals auch nicht lang, sondern es ist irgendwie, ein zweimal geht es hoch, dann geht es wieder runter und dann hast du ne Zeit lang Ruhe, dann geht es wieder hoch. Das sind eigentlich, die gehören mit zu glaube ich zu den ja, schwierigsten Sachen, weil da kommt irgendwie immer wieder was rein man weiß nicht genau, was es eigentlich ist, was verändert sich in der Zeit. Also die sind auf der Skala auf jeden Fall auch mindestens eine 8 oder eine 9 als interessant zu bewerten. Und ja, und dann hast du natürlich die diese schleichenden Geschichten, die irgendwie reinkommen und weiß ich nicht dann vielleicht ja, eine Charge oder so was, die jetzt problematisch ist oder sowas wo du sagst ok die kommt rein und ist vielleicht noch mit einer anderen Charge vermischt also das heißt dann fängt es auf einmal schon an, dass es ein bisschen höher wird und so schleicht sich das dann halt langsam hoch also es gibt ja dann noch so diese schleichenden Trends und die sind tatsächlich eigentlich die, ja das ist fast schon die 10 würde ich sagen, weil das sind die, die dann halt auch ne Zeit lang bleiben, bis sich dann halt dieser, ja, die nicht sofort wieder verschwinden oder sowas, das ist dann meistens keine Mitarbeiter Themen, sondern das sind tatsächlich dann Komponententhemen und die sind dann nicht ganz so einfach, ja, zu finden und vor allen Dingen dann wieder in den Griff zu kriegen. Insofern sind das dann eigentlich fast die, ja, ist das dann die 10 würde ich sagen.	Muster Peak Ursache Maßnahme Shift Ursache Zyklischer Anstieg, U-Muster Ursache Schleichend Maßnahme

I	Okay dann, abschließend möchte ich noch mal so ein bisschen über die Sensitivität reden und ich hab mir überlegt, wie ich das am besten abfragen kann und möchte dich deshalb jetzt fragen, also wie wichtig es dir grundsätzlich alle möglichen relevanten Trends angezeigt zu bekommen im Gegensatz dazu nur die relevanten Trends angezeigt zu bekommen. Also vielleicht ist es einfacher, das zu beantworten, wenn man dich fragst, wie viele nicht relevante Trends du in Kauf nehmen würdest angezeigt zu bekommen, um einen relevanten Trend angezeigt zu bekommen. Also wie viele False Positives würdest du in Kauf nehmen für einen True Positive? Oder wenn dir das anders rum leichter fällt, kannst du auch antworten, wie viele relevante Trends du bereit wärst zu verpassen, um quasi 10 oder einen nicht relevanten Trend weniger angezeigt zu bekommen.	Sensitivität
P2	Gute Frage. Ich muss, ich muss da jetzt leider sagen, es kommt darauf an. Ja, das Ding ist also, tendenziell würde ich durchaus in Kauf nehmen auch mir täglich diverse Trends anzugucken, die für mich nicht relevant ist, wenn ich dann diese die für mich relevant sind, weil die, die für mich relevant sind, oder die, die ich sonst nicht mitbekommen würde, sind die, die teuer werden können oder das sind die, wo man, wo man dann schnell reagieren kann, ne, das sind, das sind die wo ja, wo dann schnell reagiert werden könnte damit nicht irgendeine Kosten entstehen, Verschrottungskosten etc. Also insofern würde ich schon auch einige nicht relevante Trendlinien in Kauf nehmen, damit ich die erwische, die wichtig sind. Wenn ich jetzt allerdings 2 Wochen da jeden Tag reingucke, und ich jeden Tag nur Trends sehe, die mich eigentlich vielleicht gar nicht interessieren oder beziehungsweise wo gar nichts los ist, bin ich mir nicht sicher, inwiefern dann halt sozusagen, ja, man gewillt ist, das zu benutzen, wenn einem dann eh nichts angezeigt wird, ne also dann würde ich lieber, ja, lieber irgendwie ne Info kriegen "Heute gibts nichts". Als mir dann irgendwie, ja, dann muss ich gar nicht erst rein gucken, oder ich mach es auf und sehe ok, heute gibt es keine Codes oder sowas, dann wäre das schon wieder eine andere Geschichte. Weiß jetzt nicht, ob das jetzt so ob du weißt, was ich meine?	Kosten Relevanz Usability
I	Klar, ja.	
P2	Ehm so. Weil dann würde man vielleicht irgendwann ja diesen, wenn man bei dem Tool ist, sag ich jetzt mal, dann würdest du dem vielleicht irgendwann einfach nicht mehr trauen, so das, oder dann würdest du halt sagen ja, was soll ich da rein gucken? Ich krieg sowieso nur das was ich eh schon weiß. Wenn es zu lange dauert. Aber wenn ich, wenn ich reingucke und sage ok, heute gibt es keinen, weil vielleicht die Sensitivität so eingestellt ist, dann weiß ich ok, dann brauche ich gar nicht weiter reingucken, aber wenn ich reingucke, dann klar, dann muss ich mich vielleicht durch 10 Codes durchklicken, hab aber eine gewisse Chance, dass ich da was finde, was ich nicht kenne oder wo ich halt reingucken muss und dann würde ich halt eher benutzen, also insofern kommt halt ein bisschen drauf an. So also ich würde durchaus, wie war das also, ich würde durchaus falsch Positive in Kauf nehmen, um dann halt auch tatsächlich die richtigen Trends zu finden, aber es muss in einem ja weiß ich nicht, angemessenen Verhältnis sein auf Dauer, dass es auch auf Dauer nutzbar ist.	Vertrauen Informationswert Angemessenes Verhältnis
I	Nein, ja, hab ich glaub ich verstanden.	
P2	Ok	
I	Das waren doch schon meine Fragen. Gibt es irgendwas, was ich vergessen hab aus deiner Sicht oder was du noch ergänzen würdest oder was dir noch eingefallen ist?	

P2	Nein. Außer dass ich es total cool finde. Wie gesagt also, lieber ein paar Sachen angucken, die man schon kennt, aber die erwischen, die man halt nicht kennt. Und überhaupt was in der Hand zu haben als dann irgendwann überrascht zu werden, also insofern ja.	Warnung
I	Okay, dann vielen Dank.	

Block „Implementierung“ (Workshopergebnisse)

Optionen



Brainstorming

O1: Weboberfläche mit/ohne Filtermöglichkeiten

+ MVP schnell und einfach umsetzbar	-	! Usability! Wenn Filter, dann mit Cookies / Einstellungen speichern?	?	? Brauchen wir dann eine User Authentifizierung
+ Ohne Filter einheitlich Ergebnisse	- weckt möglicherweise eher Begehrlichkeiten à la "bitte noch xy ergänzen" da prinzipiell fast beliebig erweiterbar	! Keep it simple (Anzahl der Einstellmöglichkeiten)		Führen unterschiedliche Filter zu unterschiedlichen Handlungen? -> Folgen
+ Hintergründe des Alerts durch Anwender dynamisch analysierbar	- Zusätzliche Plattform, die Leute regelmäßig nutzen müssen			? Know-How im Team?
+ Aktiver Einstieg + Individualisierbar (mit)	- kann leicht untergehen im Daily Work (Selbstdisziplin)			

Optionen



Brainstorming

O2: Email-Benachrichtigung (+ Admin Oberfläche zum Einstellen der Benachrichtigungen)

+ Julia Schenke	-	! Ich möchte Emails selbstständig an- und abstellen können und meine Bereiche einstellen können	?
+ kein "Polling" durch den User erforderlich	- Kann schnell zu Missachtung führen (zu viele Mails/Alarmer)	! Ich möchte eine Zusammenfassung mehrerer Alerts in einer Mail bekommen, um die Anzahl der Mails klein zu halten	? E-Mail mit Reportzusammenfassung aller Alerts?
+ Automatisierung ^2	- Spamverdacht - Individualisierung schwierig	! Ich als Anwender möchte je NCFM-Code die Mailhäufigkeit einstellen können (z.B. sofort bei Alert, alle 3h, täglich, ...)	? Know-How im Team?
+ Aktive Information des Users + Vergleichbarkeit ggf besser wegen geringerer Individualisierung	- Leute müssen sich für ihre Alarmer anmelden oder zugewiesen werden -> ggf. höhere Hürde als kurz auf Website	! Inhalt der Mail: a) Alarm b) Alarm code xy c) Alarm code XY und folgender Trend...	

Optionen



Brainstorming

O3: Integration in Qlik



Sensitivität / Spezifität



Brainstorming

Wie bestimmen / festlegen?



Technologie



Brainstorming

Für Webfrontend: Streamlit vs. Flask/Django vs. JS (React?)

Nutzerauthentifizierung und Co. geht mit Streamlit, ist aber hacky

Mini-MVP schnell in Streamlit

Python Dashboard von RecycleIn?

Know How im Team? 🤖

Qlik?

Python, Qlik, SQL
JS???

Kontinuierliches Feedback



Brainstorming

Wie designen?

wenn, dann direkt im Tool mit minimalem Aufwand

Woran kann ich Interesse messen? Weboberfläche vlt. Click auf "Qlik-Absprung"?

? Labeling akut interessant oder hätte interessant sein können?

Korrelation mit tatsächlichen CAPAs

Am ehesten kann man relevante Verläufe erkennen, was ist Vergleichsgruppe?

In Qlik schwierig? Kann ich irgendwelche Reaktionen zurückschreiben?

Feedbackbutton an Sammelemailaddy? Thema für alle Reports? Für Nicht-Qlik-Lösung ggf Labeling einfach dauerhaft weiterlaufen lassen?

C Ergebnisse der Cross Validation, Tests und Parametersuche

Cross Validation

F β mit $\beta = 0.5$

	Mean Training	Std Training	Mean Test	Std Test
IK	0.69	0.01	0.69	0.03
IK-o	0.76	0.01	0.71	0.03
VB-o	0.76	0.01	0.7	0.02
VBDC-o	0.8	0.01	0.75	0.04
DT	0.8	0.01	0.71	0.06
XGB	0.88	0.03	0.76	0.04
KNN k=3	0.86	0.02	0.71	0.05
KNN k=5	0.81	0.01	0.68	0.03

F β mit $\beta = 0.7$

	Mean Training	Std Training	Mean Test	Std Test
IK	0.7	0.01	0.7	0.02
IK-o	0.77	0.01	0.75	0.03
VB-o	0.76	0.01	0.73	0.04
VBDC-o	0.8	0.01	0.75	0.03
DT	0.8	0.01	0.71	0.06
XGB	0.89	0.02	0.77	0.04
KNN k=3	0.87	0.02	0.71	0.05
KNN k=5	0.81	0.01	0.68	0.04

F β mit $\beta = 1$

	Mean Training	Std Training	Mean Test	Std Test
IK	0.77	0.01	0.76	0.03
IK-o	0.79	0.01	0.77	0.04
VB-o	0.79	0.01	0.79	0.04
VBDC-o	0.82	0.01	0.8	0.04
DT	0.8	0.02	0.72	0.07
XGB	0.9	0.02	0.79	0.05
KNN k=3	0.82	0.01	0.68	0.05
KNN k=5	0.82	0.01	0.68	0.05

F β mit $\beta = 2$

	Mean Training	Std Training	Mean Test	Std Test
IK	0.86	0.01	0.85	0.04
IK-o	0.86	0.0	0.86	0.02
VB-o	0.88	0.01	0.87	0.02
VBDC-o	0.89	0.01	0.86	0.03
DT	0.79	0.04	0.73	0.08
XGB	0.91	0.02	0.81	0.07
KNN k=3	0.91	0.01	0.71	0.06
KNN k=5	0.84	0.01	0.67	0.06

Accuracy, Optimierung für $\beta = 1$

	Mean Training	Std Training	Mean Test	Std Test
IK	0.72	0.01	0.72	0.03
IK-o	0.77	0.01	0.75	0.03
VB-o	0.77	0.01	0.76	0.03
VBDC-o	0.8	0.01	0.78	0.04
DT	0.8	0.01	0.71	0.07
XGB	0.9	0.02	0.78	0.05
KNN k=3	0.88	0.02	0.71	0.05
KNN k=5	0.82	0.01	0.68	0.03

Tests an Gesamtdatensatz (Mit Parametern optimiert für $\beta = 0.7$):

	0.5	0.7	1	2	TN	FP	FN	TP
IK	0.41	0.45	0.52	0.71	412	232	6	131
IK-o	0.43	0.48	0.54	0.7	457	187	18	119
VB-o	0.42	0.46	0.53	0.71	437	207	13	124
VBDC-o	0.48	0.51	0.57	0.71	492	152	22	115
DT	0.31	0.34	0.4	0.55	373	271	36	101
XGB	0.36	0.39	0.45	0.6	416	228	31	106
VBDC-m	0.5	0.53	0.57	0.64	535	109	40	97
kNN k=3	0.37	0.4	0.46	0.62	412	232	27	110

Gefundene Parameter

Initiale Klassifizierung (IK-o):

β	MK p-Wert	Steigung	Niveau	Outliers	Single Grenze
0.5	0.03	0.0	0.1	2	10
0.7	0.03	0.0	0.1	1	10
1	0.03	0.0	0.1	1	9
2	0.04	0.0	0.1	1	6

Vereinfachter Baum (VB-o):

β	MK p-Wert	Steigung	Niveau	Single Grenze
0.5	0.02	0.1	0.1	10
0.7	0.03	0.0	0.1	9
1	0.03	0.0	0.1	8
2	0.05	0.0	0.1	6

Vereinfachter Baum nach demand pattern (VBDC-o):

β	demand pattern	MK p-Val	Steigung	Niveau	Single Grenze
0.5	Smooth	0.05	0.0	0.0	5
0.5	Lumpy	0.03	0.0	1.2	9
0.5	Erratic	0.02	0.9	1.0	11
0.7	Smooth	0.05	0.0	0.1	5
0.7	Lumpy	0.03	0.0	1.2	9
0.7	Erratic	0.01	0.5	0.1	5
1	Smooth	0.05	0.0	0.0	5
1	Lumpy	0.03	0.0	0.0	8
1	Erratic	0.01	0.5	0.1	5
2	Smooth	0.05	0.0	0.0	5
2	Lumpy	0.04	0.0	1	6
2	Erratic	0.04	0.1	0.1	5

Vereinfachter Baum nach demand pattern – manuell angepasst (VBDC-m):

β	demand pattern	MK p-Val	Steigung	Niveau	Single Grenze
M	Smooth	0.04	0.1	0.02	5
M	Lumpy	0.03	0.1	1.2	9
M	Erratic	0.02	0.4	1.0	11

Literaturverzeichnis

- Alashan, S. (2020). Combination of modified Mann-Kendall method and Şen innovative trend analysis. *Engineering Reports*, 2(3), e12131. <https://doi.org/10.1002/eng2.12131>
- Al-Ghanim, A., & Jordan, J. (1996). Automated process monitoring using statistical pattern recognition techniques on X-bar control charts. *Journal of Quality in Maintenance Engineering*, 2(1), 25–49. <https://doi.org/10.1108/13552519610113827>
- Avci, U., Bulut, Ö., & Toy, A. Ö. (2022). A Comparative Study of Artificial Intelligence Based Methods for Abnormal Pattern Identification in SPC. In C. Kahraman, A. C. Tolga, S. Cevik Onar, S. Cebi, B. Oztaysi, & I. U. Sari (Hrsg.), *Intelligent and Fuzzy Systems* (S. 417–425). Springer International Publishing. https://doi.org/10.1007/978-3-031-09176-6_48
- Bag, M., Gauri, S. K., & Chakraborty, S. (2012). An expert system for control chart pattern recognition. *The International Journal of Advanced Manufacturing Technology*, 62(1), 291–301. <https://doi.org/10.1007/s00170-011-3799-z>
- Berg, B. L., & Lune, H. (2011). *Qualitative Research Methods for the Social Sciences* (8. Aufl.). Pearson.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.
- Cerqueira, V., Moniz, N., & Soares, C. (2021). VEST: Automatic feature engineering for forecasting. *Machine Learning*. <https://doi.org/10.1007/s10994-021-05959-y>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Cheng, C.-S. (1997). A neural network approach for the analysis of control chart patterns. *International Journal of Production Research*, 35(3), 667–697. <https://doi.org/10.1080/002075497195650>
- Cleveland, R. B., Cleveland, W. S., & McRae, J. E. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1), 3–73.
- Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, 74(368), 829–836. <https://doi.org/10.1080/01621459.1979.10481038>
- Davis, R. A., Fokianos, K., Holan, S. H., Joe, H., Livsey, J., Lund, R., Pipiras, V., & Ravishanker, N. (2021). Count Time Series: A Methodological Review.

- Journal of the American Statistical Association*, 116(535), 1533–1547.
<https://doi.org/10.1080/01621459.2021.1904957>
- Davis, R. B., & Woodall, W. H. (1988). Performance of the Control Chart Trend Rule under Linear Shift. *Journal of Quality Technology*, 20(4), 260–262.
<https://doi.org/10.1080/00224065.1988.11979118>
- Desmar. (2022). *Non-Conformance Report: Everything You Need to Know*. ECLIPSE Suite. <https://www.eclipsesuite.com/non-conformance-report/>
- Dietrich, E. (2021). *Statistische Verfahren zur Maschinen- und Prozessqualifikation* (8., aktualisierte Auflage). Carl Hanser Verlag GmbH & Co. KG.
- DIN Deutsches Institut für Normung e. V. (2015a). *Qualitätsmanagementsysteme – Anforderungen (DIN EN ISO 9001:2015)* [ISO Norm].
<https://dx.doi.org/10.31030/2325651>
- DIN Deutsches Institut für Normung e. V. (2015b). *Qualitätsmanagementsysteme – Grundlagen und Begriffe (DIN EN ISO 9000:2015)* [ISO Norm].
<https://dx.doi.org/10.31030/2325650>
- DIN Deutsches Institut für Normung e. V. (2021). *Medizinprodukte—Qualitätsmanagementsysteme—Anforderungen für regulatorische Zwecke (DIN EN ISO 13485:2021)*. <https://dx.doi.org/10.31030/3223237>
- Duffy, G. (2013). *What is Cost of Quality (COQ)?* / ASQ. <https://asq.org/quality-resources/cost-of-quality>
- Ebrahimzadeh, A., Addeh, J., & Ranaee, V. (2013). Recognition of control chart patterns using an intelligent technique. *Applied Soft Computing*, 13(5), 2970–2980. <https://doi.org/10.1016/j.asoc.2012.02.019>
- Fahmy, H. M., & Elsayed, E. A. (2006). Detection of linear trends in process mean. *International Journal of Production Research*, 44(3), 487–504.
<https://doi.org/10.1080/00207540500310228>
- Farcomeni, A. (2008). A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. *Statistical Methods in Medical Research*, 17(4), 347–388. <https://doi.org/10.1177/0962280206079046>
- Floares, A., Ferisgan, M., & Manolache, F. B. (2017). The Smallest Sample Size for the Desired Diagnosis Accuracy. *International Journal of Oncology and Cancer Therapy*, 2, 13–19.
- Frepple. (o. J.). Demand classification: Why forecastability matters - frepple APS. *Frepple*. Abgerufen 20. Januar 2023, von <https://frepple.com/blog/demand-classification/>
- Frochte, J. (2018). *Maschinelles Lernen*. Hanser.
- Fulcher. (2018). Feature-Based Time-Series Analysis. In *Feature Engineering for Machine Learning and Data Analytics* (S. 87–116). CRC Press.
<https://doi.org/10.1201/9781315181080-4>

- Ganapathy, S., Jaisankar, N., Yogesh, P., & Kannan, A. (2011). An intelligent system for intrusion detection using outlier detection. *2011 International Conference on Recent Trends in Information Technology (ICRTIT)*, 119–123. <https://doi.org/10.1109/ICRTIT.2011.5972372>
- Gennatas, E. D., Friedman, J. H., Ungar, L. H., Pirracchio, R., Eaton, E., Reichmann, L. G., Interian, Y., Luna, J. M., Simone, C. B., Auerbach, A., Delgado, E., van der Laan, M. J., Solberg, T. D., & Valdes, G. (2020). Expert-augmented machine learning. *Proceedings of the National Academy of Sciences*, 117(9), 4571–4577. <https://doi.org/10.1073/pnas.1906831117>
- Géron, A. (2018). *Praxiseinstieg Machine Learning mit Scikit-Learn und TensorFlow* (1. Auflage). O'Reilly.
- Guh, R.-S. (2005). Real-time pattern recognition in statistical process control: A hybrid neural network/decision tree-based approach. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 219(3), 283–298. <https://doi.org/10.1243/095440505X28963>
- Hachicha, W., & Ghorbel, A. (2012). A survey of control-chart pattern-recognition literature (1991–2010) based on a new conceptual classification scheme. *Computers & Industrial Engineering*, 63(1), 204–222. <https://doi.org/10.1016/j.cie.2012.03.002>
- Harrison, M. (2021). *Machine Learning – die Referenz* (1. Auflage, deutsche Ausgabe). O'Reilly.
- Hassan, A., Baksh, M. S. N., Shaharoun, A. M., & Jamaluddin, H. (2003). Improved SPC chart pattern recognition using statistical features. *International Journal of Production Research*, 41(7), 1587–1603. <https://doi.org/10.1080/0020754021000049844>
- He, S., Huang, W., & Woodall, W. H. (2011). CUSUM charts for monitoring a zero-inflated poisson process. *Quality and Reliability Engineering International*, 28(2), 181–192. <https://doi.org/10.1002/qre.1228>
- Hirschle, J. (2021). *Machine Learning für Zeitreihen*. Hanser.
- Hurwitz, J., Morris, H., Sidner, C., & Kirsch, D. (2020). *Augmented Intelligence. The Business Power of Human-Machine Collaboration*. CRC Press.
- Hussain, Md., & Mahmud, I. (2019). pyMannKendall: A python package for non parametric Mann Kendall family of trend tests. *Journal of Open Source Software*, 4(39), 1556. <https://doi.org/10.21105/joss.01556>
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed). OTexts. <https://otexts.com/fpp3/>
- Hyndman, R. J., Wang, E., & Laptev, N. (2015). Large-Scale Unusual Time Series Detection. *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, 1616–1619. <https://doi.org/10.1109/ICDMW.2015.104>

- Ishikawa, K., & Loftus, J. H. (1990). *Introduction to quality control* (3rd ed). 3A Corporation.
- Janusz, M. E., & Venkatasubramanian, V. (1991). Automatic generation of qualitative descriptions of process trends for fault detection and diagnosis. *Engineering Applications of Artificial Intelligence*, 4(5), 329–339.
[https://doi.org/10.1016/0952-1976\(91\)90001-M](https://doi.org/10.1016/0952-1976(91)90001-M)
- Javadi, S., & Niaki, S. (2013). A new uni-attribute control chart to monitor number of nonconformities. *Journal of Optimization in Industrial Engineering*, 6, 79–83.
- Ji, C., Du, M., Hu, Y., Liu, S., Pan, L., & Zheng, X. (2022). Time series classification based on temporal features. *Applied Soft Computing*, 128, 109494.
<https://doi.org/10.1016/j.asoc.2022.109494>
- Joekes, S., & Pimentel Barbosa, E. (2013). An improved attribute control chart for monitoring non-conforming proportion in high quality processes. *Control Engineering Practice*, 21(4), 407–412.
<https://doi.org/10.1016/j.conengprac.2012.12.005>
- Jombart, T., Ghozzi, S., Schumacher, D., Taylor, T. J., Leclerc, Q. J., Jit, M., Flasche, S., Greaves, F., Ward, T., Eggo, R. M., Nightingale, E., Meakin, S., Brady, O. J., null, null, Medley, G. F., Höhle, M., & Edmunds, W. J. (2021). Real-time monitoring of COVID-19 dynamics using automated trend fitting and anomaly detection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1829), 20200266. <https://doi.org/10.1098/rstb.2020.0266>
- Juuso, E. K. (2011). Intelligent Trend Indices in Detecting Changes of Operating Conditions. *2011 UkSim 13th International Conference on Computer Modelling and Simulation*, 162–167. <https://doi.org/10.1109/UKSIM.2011.39>
- Kendall, M. G. (1948). *Rank correlation methods*. Griffin.
- Khoo, M. B. C. (2004). Poisson Moving Average Versus c Chart for Nonconformities. *Quality Engineering*, 16(4), 525–534. <https://doi.org/10.1081/QEN-120038014>
- Kiefer, D., Grimm, F., Bauer, M., & van Dinther, C. (2021, Januar 4). *Demand Forecasting Intermittent and Lumpy Time Series: Comparing Statistical, Machine Learning and Deep Learning Methods*.
<https://doi.org/10.24251/HICSS.2021.172>
- Kindsmüller, M., & Urbas, L. (2002). Der Einfluss von Modellwissen auf die Interpretation von Trenddarstellungen bei der Steuerung prozesstechnischer Anlagen. In *Situation Awareness in der Fahrzeug- und Prozessführung* (S. 152).
- Kostenko, A. V., & Hyndman, R. J. (2006). A note on the categorization of demand patterns. *Journal of the Operational Research Society*, 57(10), 1256–1257.
<https://doi.org/10.1057/palgrave.jors.2602211>
- Kreiß, J.-P. (2006). *Einführung in die Zeitreihenanalyse*. Springer.

- Laptev, N., Amizadeh, S., & Flint, I. (2015). Generic and Scalable Framework for Automated Time-series Anomaly Detection. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1939–1947. <https://doi.org/10.1145/2783258.2788611>
- Link, P., Poursanidis, M., Schmid, J., Zache, R., von Kurnatowski, M., Teicher, U., & Ihlenfeldt, S. (2022). Capturing and incorporating expert knowledge into machine learning models for quality prediction in manufacturing. *Journal of Intelligent Manufacturing*, 33(7), 2129–2142. <https://doi.org/10.1007/s10845-022-01975-4>
- Liu, W., & Cheraghi, S. H. (2006). A Hybrid Approach to Nonconformance Tracking and Recovery. *Journal of Intelligent Manufacturing*, 17(1), 149–162. <https://doi.org/10.1007/s10845-005-5518-9>
- Loftis, J. C., Ward, R. C., Phillips, R. D., & Taylor, C. H. (1989). *Evaluation of trend-detection techniques for use in water-quality monitoring programs* (PB-90-100058/XAB). Colorado State Univ., Fort Collins, CO (USA).
- Lyons, J. B., Wynne, K. T., Mahoney, S., & Roebke, M. A. (2019). Chapter 6 - Trust and Human-Machine Teaming: A Qualitative Study. In W. Lawless, R. Mittu, D. Sofge, I. S. Moskowitz, & S. Russell (Hrsg.), *Artificial Intelligence for the Internet of Everything* (S. 101–116). Academic Press. <https://doi.org/10.1016/B978-0-12-817636-8.00006-5>
- Machida, F., Andrzejak, A., Matias, R., & Vicente, E. (2013). On the effectiveness of Mann-Kendall test for detection of software aging. *2013 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*, 269–274. <https://doi.org/10.1109/ISSREW.2013.6688905>
- Mahmood, T., & Xie, M. (2019). Models and monitoring of zero-inflated processes: The past and current trends. *Quality and Reliability Engineering International*, 35(8), 2540–2557. <https://doi.org/10.1002/qre.2547>
- Mann, H. B. (1945). Nonparametric Tests Against Trend. *Econometrica*, 13(3), 245–259. <https://doi.org/10.2307/1907187>
- McDermott, P., Dominguez, C., Kasdaglis, N., Ryan, M., Trahan, I., & Nelson, A. (2018). *Human-Machine Teaming Systems Engineering Guide* (Nr. MP180941). The MITRE Corporation. <https://apps.dtic.mil/sti/citations/AD1108020>
- Mirabbasi, R., Ahmadi, F., & Jhajharia, D. (2020). Comparison of parametric and non-parametric methods for trend identification in groundwater levels in Sirjan plain aquifer, Iran. *Hydrology Research*, 51(6), 1455–1477. <https://doi.org/10.2166/nh.2020.041>
- Mohsen, O., Mohamed, Y., & Al-Hussein, M. (2019). Data Analytics Application for Non-Conformance Reports in a Cabinet Manufacturing Facility. *Modular and Offsite Construction (MOC) Summit Proceedings*, 341–348. <https://doi.org/10.29173/mocs112>

- Montgomery, D. C. (2009). *Introduction to Statistical Quality Control* (6. Aufl.). John Wiley & Sons, Inc.
- Motschman, T. L., & Moore, S. B. (1999). Corrective and preventive action. *Transfusion Science*, 21(2), 163–178. [https://doi.org/10.1016/S0955-3886\(99\)00088-0](https://doi.org/10.1016/S0955-3886(99)00088-0)
- Mukhopadhyay, S., Solis, A. O., & Gutierrez, R. S. (2012). The Accuracy of Non-traditional versus Traditional Methods of Forecasting Lumpy Demand. *Journal of Forecasting*, 31(8), 721–735. <https://doi.org/10.1002/for.1242>
- Nelson, L. S. (1984). The Shewhart Control Chart—Tests for Special Causes. *Journal of Quality Technology*, 16(4), 237–239. <https://doi.org/10.1080/00224065.1984.11978921>
- Nelson, L. S. (1985). Interpreting Shewhart \bar{X} Control Charts. *Journal of Quality Technology*, 17(2), 114–116. <https://doi.org/10.1080/00224065.1985.11978945>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Perktold, J., Seabold, S., & Taylor, J. (2023). *Statsmodels.tsa.seasonal.STL* — statsmodels [Software Documentation]. statsmodels.tsa.seasonal.STL. <https://www.statsmodels.org/dev/generated/statsmodels.tsa.seasonal.STL.html#statsmodels.tsa.seasonal.STL>
- Pham, D. T., & Wani, M. A. (1997). Feature-based control chart pattern recognition. *International Journal of Production Research*, 35(7), 1875–1890. <https://doi.org/10.1080/002075497194967>
- Quadri, G. J., & Rosen, P. (2022). A Survey of Perception-Based Visualization Studies by Task. *IEEE Transactions on Visualization and Computer Graphics*, 28(12), 5026–5048. <https://doi.org/10.1109/TVCG.2021.3098240>
- Rahman, R. (2021, März 31). *A Comprehensive Guide to Non-Conformance Reporting*. QISS Quality Management Software | QIA. <https://www.qi-a.com/a-comprehensive-guide-to-non-conformance-reporting-ncr>
- Ramezani, J., & Jassbi, J. (2017). A Hybrid Expert Decision Support System Based on Artificial Neural Networks in Process Control of Plaster Production – An Industry 4.0 Perspective. In L. M. Camarinha-Matos, M. Parreira-Rocha, & J. Ramezani (Hrsg.), *Technological Innovation for Smart Systems* (S. 55–71). Springer International Publishing. https://doi.org/10.1007/978-3-319-56077-9_5
- Rodriguez, J. (2020, Dezember 30). *Non-Conformance Report: How to Report a Quality Issue*. LiveAbout. <https://www.liveabout.com/non-conformance-report-how-to-report-a-quality-issue-844987>

- Schmidt-Kretschmer, M., Gries, B., & Blessing, L. (2006). *Bug or Feature? Möglichkeiten und Grenzen des Fehlermanagements in der Produktentwicklung*. 17. Symposium „Design for X“, Neunkirchen.
- Shewhart, W. A. (2015). *Economic Control of Quality Of Manufactured Product*. Martino Fine Books.
- Silva, V. F., Silva, M. E., Ribeiro, P., & Silva, F. (2022). Novel features for time series analysis: A complex networks approach. *Data Mining and Knowledge Discovery*, 36(3), 1062–1101. <https://doi.org/10.1007/s10618-022-00826-3>
- Stangl, W. (2023). *Feedback – Online Lexikon für Psychologie & Pädagogik*. <https://lexikon.stangl.eu/4603/feedback>
- Starks, S. A., & Mercado, J. L. (1992). *Automated system trend monitoring based on time series analysis*. 1825, 351–355. <https://doi.org/10.1117/12.131543>
- Su, W., Zhu, Y., Liu, F., & Hu, K. (2013). On-line outlier and change point detection for time series. *Journal of Central South University*, 20(1), 114–122. <https://doi.org/10.1007/s11771-013-1466-2>
- Swift, J. A. (1987). *Development of a knowledge based expert system for control chart pattern recognition and analysis* [The Oklahoma State University]. <https://shareok.org/handle/11244/14411>
- Sylla, C. (2002). Experimental investigation of human and machine-vision arrangements in inspection tasks. *Control Engineering Practice*, 10(3), 347–361. [https://doi.org/10.1016/S0967-0661\(01\)00151-4](https://doi.org/10.1016/S0967-0661(01)00151-4)
- Syntetos, A. A., Boylan, J. E., & Croston, J. D. (2005). On the categorization of demand patterns. *Journal of the Operational Research Society*, 56(5), 495–503. <https://doi.org/10.1057/palgrave.jors.2601841>
- Tomperi, J., Juuso, E., & Leiviskä, K. (2016). Early warning of changing drinking water quality by trend analysis. *Journal of Water and Health*, 14(3), 433–442. <https://doi.org/10.2166/wh.2016.330>
- Villez, K., & Rengaswamy, R. (2013). A generative approach to qualitative trend analysis for batch process fault diagnosis. *2013 European Control Conference (ECC)*, 1958–1963. <https://doi.org/10.23919/ECC.2013.6669494>
- Voigt, P. D. K.-I. (2018, Februar 19). *Definition: Automatisierung* [Text]. <https://wirtschaftslexikon.gabler.de/definition/automatisierung-27138>; Springer Fachmedien Wiesbaden GmbH. <https://wirtschaftslexikon.gabler.de/definition/automatisierung-27138/version-250801>
- Wang, F., Shao, W., Yu, H., Kan, G., He, X., Zhang, D., Ren, M., & Wang, G. (2020). Re-evaluation of the Power of the Mann-Kendall Test for Detecting Monotonic Trends in Hydrometeorological Time Series. *Frontiers in Earth Science*, 8. <https://www.frontiersin.org/articles/10.3389/feart.2020.00014>

- Wang, X., Smith, K., & Hyndman, R. (2006). Characteristic-Based Clustering for Time Series Data. *Data Mining and Knowledge Discovery*, 13(3), 335–364. <https://doi.org/10.1007/s10618-005-0039-x>
- Western Electric Co. (1958). *Statistical Quality Control Handbook by Western Electric* (2nd Edition). Mack Printing Company.
- Woodall, W. H. (1997). Control Charts Based on Attribute Data: Bibliography and Review. *Journal of Quality Technology*, 29(2), 172–183. <https://doi.org/10.1080/00224065.1997.11979748>
- xgboost developers. (2022). *Python Package Introduction—Xgboost 1.7.3 documentation*. https://xgboost.readthedocs.io/en/stable/python/python_intro.html
- Yang, X., & Han, X. (2022). Mann–Kendall-Based Concrete Failure Trend Analysis and Its Implementation for Dynamic Building Monitoring. *Buildings*, 12(8), Art. 8. <https://doi.org/10.3390/buildings12081165>
- Zandi, F., Niaki, S. T. A., Nayeri, M. A., & Fathi, M. (2011). Change-point estimation of the process fraction non-conforming with a linear trend in statistical process control. *International Journal of Computer Integrated Manufacturing*, 24(10), 939–947. <https://doi.org/10.1080/0951192X.2011.608720>
- Zhang, L., Cai, S., & Hu, J. (2019). An adaptive pre-warning method based on trend monitoring: Application to an oil refining process. *Measurement*, 139, 163–176. <https://doi.org/10.1016/j.measurement.2019.03.013>

Abschließende Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit selbstständig, ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

A handwritten signature in blue ink, reading 'J. Schiöpp'. The signature is written in a cursive style with a large initial 'J'.

Berlin, den 23. März 2023