

# Angepasstes Item Set Mining zur gezielten Steuerung von Bauteilen in der Serienfertigung von Fahrzeugen

Marco Spieß<sup>1</sup> und Peter Reimann<sup>2</sup>

**Abstract:** Qualitätsprobleme im Bereich Fahrzeugbau können nicht nur zum Imageverlust des Unternehmens führen, sondern auch mit entsprechend hohen Kosten einhergehen. Wird ein Bauteil als Verursacher eines Qualitätsproblems identifiziert, muss dessen Verbau gestoppt werden. Mit einer Datenanalyse kann herausgefunden werden, welche Fahrzeugkonfigurationen Probleme mit diesem fehlerverursachenden Bauteil haben. Im Rahmen der domänenspezifischen Problemstellung wird in diesem Beitrag die Anwendbarkeit von Standardalgorithmen aus dem Bereich Data-Mining untersucht. Da die Analyseergebnisse auf Standardausstattungen hinweisen, sind diese nicht zielführend. Für dieses Businessproblem von Fahrzeugherstellern haben wir einen Data-Mining Algorithmus entwickelt, der das Vorgehen des Item Set Mining der Assoziationsanalyse an das domänenspezifische Problem anpasst. Er unterscheidet sich zum klassischen Apriori-Algorithmus in der Beschneidung des Ergebnisraumes sowie in der nachfolgenden Aufbereitung und Verwendungweise der Item Sets. Der Algorithmus ist allgemeingültig für alle Fahrzeughersteller anwendbar. Die Ergebnisse sind anhand eines realen Anwendungsfalls evaluiert worden, bei dem durch die Anwendung unseres Algorithmus 87% der Feldausfälle verhindert werden können.

**Keywords:** Big Data Analytics, Data-Mining, Item Set Mining.

## 1 Einführung

Garantiekosten im Rahmen der Produktgewährleistungspflicht können das Betriebsergebnis eines produzierenden Unternehmens negativ beeinflussen [Ve00]. Im Geschäftsjahr 2017 ist von den Automobilherstellern weltweit eine Summe von 53,37 Mrd. USD an Gewährleistungskosten gezahlt worden [Wo18]. Demnach nehmen das Qualitätsmanagement sowie die Kundenbetreuung in der Produktnutzungsphase einen hohen Stellenwert im Unternehmen ein, um Qualitätsprobleme möglichst frühzeitig zu identifizieren und zu beheben [Ve00, BW16].

Dem in diesem Beitrag behandelten Anwendungsfall liegt ein Qualitätsproblem im Bereich Getriebe von Fahrzeugen zugrunde. Dabei wurde die Kupplung durch verschiedene Messungen als Fehlerursache identifiziert, weshalb diese daraufhin konstruktiv angepasst wurde.

---

<sup>1</sup> Universität Stuttgart, Graduate School of Excellence advanced Manufacturing Engineering (GSaME), Nobelstr. 12, 70569 Stuttgart, Marco.Spiess@gsame.uni-stuttgart.de

<sup>2</sup> Universität Stuttgart, Graduate School of Excellence advanced Manufacturing Engineering (GSaME), Nobelstr. 12, 70569 Stuttgart, Peter.Reimann@gsame.uni-stuttgart.de

Diese angepasste Kupplung soll nun bereits in der Produktion in die Fahrzeuge eingebaut werden, um zukünftige Feldausfälle zu verhindern. Durch vereinbarte Liefermengen kann die neue Kupplungsversion aber nur in jedes zweite Fahrzeug eingebaut werden. Die restlichen Fahrzeuge werden mit einer anderen Kupplung ausgestattet, welche zum Zeitpunkt der Datenanalyse als OK eingestuft wurde. Das Ziel der durchzuführenden Analyse ist die Identifizierung von Fahrzeugen, bei denen die neue angepasste Kupplungsversion eingebaut werden soll, da die andere Kupplung bei diesen Fahrzeugen sonst weitere Probleme verursachen könnte. Demnach soll mittels der Analyseergebnisse eine Steuerung von geeigneten Bauteilen in die individuellen Fahrzeugkonfigurationen erfolgen. Da Fahrzeuge, die zunächst auf der Montagelinie montiert werden, keine Daten über die tatsächliche Nutzung im Kundenumfeld enthalten, reduziert sich die Auswahl der relevanten Datenquellen für die Analyse auf die Konfigurationsdaten von bereits produzierten und sich in Nutzung befindlichen Fahrzeugen. Durch diese Daten wird ein zu produzierendes Fahrzeug in seiner Zusammenstellung beschrieben, z.B. welcher Motor, Getriebe etc. verbaut sind oder ob das Fahrzeug eine Sonderausstattung hat. Die ausfallgefährdeten Fahrzeuge werden durch Kombinationen der ermittelten Konfigurationsmerkmale erkannt und mit der angepassten Kupplung ausgestattet.

Folgende Einzelbeiträge werden in diesem Beitrag geleistet:

1. Es werden Untersuchungen mit alternativen Data-Mining Techniken auf deren Eignung für die Problemstellung durchgeführt und bewertet. Dabei stellt sich heraus, dass die Anwendung von Klassifikation sowie Clusteranalyse keinen Mehrwert generieren, da sie Standardausstattungen eines Fahrzeugs als Ergebnis erzeugen, die nicht zielführend sind. Die Assoziationsanalyse hingegen ist mit ihrer Eigenschaft, kombinierte Elemente (Item Sets) zu betrachten, am geeignetsten. Dennoch kann mit ihr alleine die Zielstellung nicht erreicht werden, da ausfallgefährdete Fahrzeuge nicht durch einzelne Item Sets, sondern durch die Kombination mehrerer Sets beschrieben werden [HKP12].
2. Im Sinne der Problemstellung schlagen wir einen angepassten Algorithmus vor, um ausfallgefährdete Fahrzeugkonfigurationen aus den Garantiedaten zu identifizieren. Die mit gleichen Konfigurationen zu produzierenden Fahrzeuge werden gezielt mit einem geeigneten Bauteil ausgestattet. Hierfür haben wir das Item Set Mining des Apriori Algorithmus angepasst, damit das Analyseergebnis in der Produktionssteuerung der Serienfertigung genutzt werden kann.
3. Der vorgestellte Algorithmus ist anhand eines realen Anwendungsfalls im Bereich der Produktion von Fahrzeugen evaluiert und als geeignet bewertet worden. Dabei konnten wir Konfigurationsmerkmale identifizieren, mit deren Anwendung 87% der Feldausfälle hätten verhindert werden können.

Dieser Beitrag gliedert sich in fünf Abschnitte. In Abschnitt 2 stellen wir den Anwendungsfall und die relevanten Daten vor. Abschnitt 3 stellt die Untersuchungsergebnisse aus der

Anwendung alternativer Data-Mining Techniken dar. In Abschnitt 4 stellen wir die Grundidee des vorzuschlagenden Algorithmus sowie das damit erzielte Ergebnis vor. Dieser Beitrag endet mit der Zusammenfassung und zukünftiger Arbeiten in Abschnitt 5.

## 2 Anwendungsfall

Der größte Anteil an Garantiekosten ergibt sich durch das vermehrte Auftreten von Qualitätsproblemen beim Kunden im Produktgewährleistungszeitraum. Im Kontext des Qualitätsmanagement werden ausgebaute Schadteile aus dem Feld gezielt in die Befundstellen der Original-Equipment-Manufacturer (OEM) eingesteuert, um sie einer Schadteilanalyse zu überführen [Ve09]. Unter einem Schadteil werden jene Teile verstanden, die im Zuge einer Reparatur ausgebaut und ersetzt werden. Im Kontext der Schadteilanalyse gilt es, durch Methoden aus den Ingenieurwissenschaften wie bspw. Six Sigma oder 8D diejenigen Schadteile zu identifizieren, die eine mögliche Fehlerursache für ein Qualitätsproblem darstellen [JSW17]. Der in diesem Beitrag behandelte Anwendungsfall beschreibt das Problem eines OEM, der die Kupplung als Fehlerursache identifiziert und diese in der Konstruktion angepasst hat. Diese Anpassung muss nun in der Produktionslinie nachvollzogen werden. Die Prämisse hierbei ist die limitierte Teileversorgung durch den Lieferanten, welcher lediglich 50% der Produktion mit angepassten Kupplungen versorgen kann. Durch die Prämisse der Teileversorgung soll die angepasste Kupplung gezielt in ausfallgefährdete Fahrzeuge bereits in der Produktion eingebaut werden. Dabei definieren wir ein Fahrzeug dann als ausfallgefährdet, wenn seine Konfiguration nahezu deckungsgleich ist mit den Konfigurationen bereits ausgefallener Fahrzeuge. Diese Definition ist aus dem behandelten Anwendungsfall abgeleitet, bei dem ein Fahrzeugausfall mitunter auf dessen technische Zusammenstellung zurückzuführen ist. Demnach hat die durchzuführende Datenanalyse das Ziel eine direkte Fehlerabstellung in der Produktion zu ermöglichen.

Fahrgestellnummer	Feldausfall	Konfiguration
XYZ4711L590	1	C1/C2/C3/C4/C5/C6/C7/C8/C9/C19/.../Cn
XYZ9718L590	0	C3/C6/C7/C8/C14/C34/C45/.../Cn
XYZ4707R790	1	C2/C6/C7/C10//C29/C40/.../Cn
...	...	...

Tab. 1: Exemplarische Datensätze der zu analysierenden Daten

Tab.1 zeigt einen beispielhaften Auszug der zu analysierenden Daten. Darin wird ein Fahrzeug eindeutig über dessen Fahrgestellnummer identifiziert. Ob ein Fahrzeug bereits einen Schaden im Feld bzgl. des betrachteten Qualitätsproblems hatte, wird über das Setzen eines Flags im Datenfeld Feldausfall erkenntlich. Des Weiteren sind im Datenfeld Konfiguration sämtliche Merkmale enthalten, die ein Fahrzeug bzgl. dessen verbauter Ausstattungen näher beschreiben. Dabei werden die Konfigurationsmerkmale als aneinandergereihte Codes der Form C1, C2, C3 usw. gespeichert. Dabei gibt es in unserem Anwendungsfall insgesamt 3.000 disjunkte Code-Elemente. Im Durchschnitt enthält ein einzelnes Fahrzeug in etwa 300

unterschiedliche Codes. Da hierunter auch Merkmale enthalten sind, welche in nahezu jedem Fahrzeug verbaut sind, wie ein Lenkrad, Reifen oder Radio, muss im Ablauf des Algorithmus ein Mechanismus etabliert werden, der diese Standardkonfigurationen erkennt. In unserem Anwendungsfall haben wir in der Analysemenge insgesamt 117.000 Fahrzeuge, wovon 2.000 bereits einen Feldausfall zum behandelten Qualitätsproblem aufweisen. Zusammengefasst ergeben sich aus der vorliegenden Datenstruktur zwei wesentliche datentechnische Herausforderungen für die Datenanalyse:

1. Standardcodes, also allgemeine Ausstattungen wie ein Lenkrad, die sehr viele Fahrzeuge aufweisen und die aber keinerlei Bezug zum Qualitätsproblem haben, können im Rahmen einer Mustererkennung nicht verwendet werden. Werden diese dennoch berücksichtigt, stehen diese in der Ergebnisliste ganz oben. Demnach müssen diejenigen Codes identifiziert werden, deren Kombination zum einen das Schädgeschehen aus den Daten möglichst vollständig abbildet und zum anderen die Menge an noch zu erwartender Feldausfälle möglichst präzise eingrenzt. Gleichzeitig soll die Gesamtmenge an produzierten Fahrzeugen mit diesen Codes maximal 50% betragen.
2. Innerhalb der Fahrzeuge kommt es bzgl. deren Ausstattungen zu einer Schnittmengenproblematik. Beispiel dafür ist Code C1, der in Kombination mit Code C2 in 150 ausgefallenen Fahrzeugen vertreten ist. Die Kombination C1 und C3 tritt hingegen in 100 reparierten Fahrzeugen auf. Da auch alle drei Codes C1, C2 und C3 in einzelnen Fahrzeugen auftreten können, resultiert aus dem Beispiel eine kumulierte Anzahl von 175 statt 250 aufsummierten Reparaturen. Die analytische Herausforderung besteht darin, diejenigen Kombinationen mit dem möglichst größten Anteil am Schädgeschehen zu entdecken, deren Anteile nicht in einer anderen Kombination bereits enthalten sind. Anhand des o.g. Beispiels sollte demnach analysiert werden, ob es weitere Kombinationen gibt, in denen die Codes C1, C2 und C3 enthalten sind und durch Hinzunahme weiterer Codes wie bspw. C4, präzisiert werden können.

### **3 Eignung alternativer Techniken aus dem Stand der Technik**

In diesem Abschnitt werden die Untersuchungsergebnisse hinsichtlich der für die Problemstellung infrage kommenden Techniken aus dem Bereich des Data-Mining vorgestellt [HKP12]. Die Auswahl der aufgeführten Techniken ergibt sich aus der Struktur der zu analysierenden Daten sowie dem Analysezweck aus der Problemstellung. Da die Ausstattungsmerkmale eines Fahrzeugs bereits in einer warenkorbähnlichen Struktur vorliegen, kommt die Assoziationsanalyse als Miningverfahren in Betracht. Ausgehend von der analytischen Zielsetzung, relevante Konfigurationsmerkmale ausfallgefährdeter Fahrzeuge zu identifizieren, sind auch die Klassifikation und Clusteranalyse Gegenstand dieser Untersuchung. Die für die Datenanalysen genutzten Features sind bei allen Techniken jeweils die Konfigurationsdaten. Dies wird damit begründet, da das Analyseergebnis in

der Serienfertigung genutzt werden soll, in der zu einem produzierenden Fahrzeug, außer der eindeutigen Fahrgestellnummer und dessen Ausstattung keine weiteren beschreibenden Daten vorhanden sind. Für die Untersuchungen haben wir insgesamt 6.327 Instanzen verwendet. Vor den Analysen haben wir versucht eine manuelle Datenfilterung der Codes durchzuführen indem vermeintliche Standardcodes wie Lenkrad eliminiert werden. Ebenso filterten wir weitere Sachverhalte wie Sitzbezüge und Felgen, bei denen keine technische Relevanz gegeben ist, manuell per Datenbankabfrage aus der Analysemenge heraus. Trotz dieser Datenfilterung können wir dennoch nicht sicherstellen, alle Standardcodes damit ausgeschlossen zu haben. Zumal der Anwender i.d.R. kein Wissen über diese besitzt, muss ein algorithmischer Automatismus entwickelt werden, der die Standardcodes von den relevanten Codes separiert. Unserer Einschätzung nach lässt sich damit die Analysequalität verbessern, denn im Allgemeinen stellt sich heraus, dass Standardcodes nicht eindeutig als solche erkannt werden können. Somit können sie nicht vor einer Analysedurchführung rausgefiltert werden, ohne das Analyseergebnis zu beeinflussen. Aus diesem Grund haben wir die Analysen mit den Codes im Original durchgeführt. Dabei haben wir herausgefunden, dass einzelne Codes in den Daten existieren, die aufgrund ihrer nominalen Bedeutung eine Standardausstattung sind. Dennoch sind sie wichtig, da sie besonders in den bereits ausgefallenen Fahrzeugen vorhanden sind.

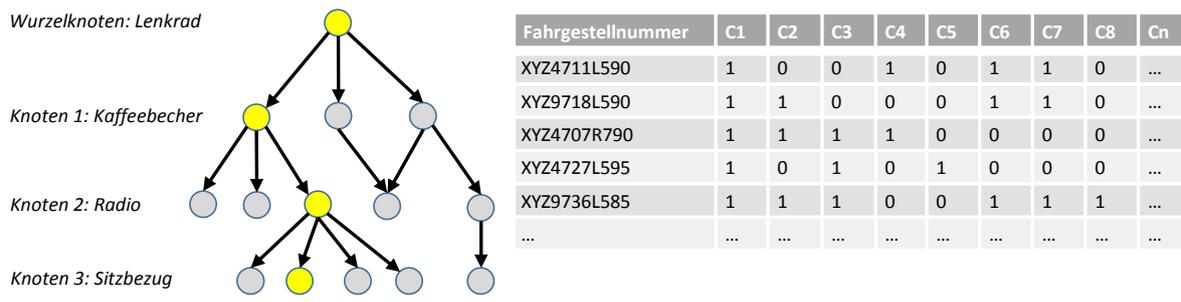


Abb. 1: Klassifikation als Entscheidungsbaum

Im nachfolgenden Teil dieses Abschnitts gehen wir lediglich auf die Assoziationsanalyse detailliert ein. Diese Data-Mining Technik, bzw. der darin vorgelagerte Analyseschritt des Item Set Mining, hat sich im Rahmen unserer Untersuchung als am geeignetsten für die Problemstellung herausgestellt. Für diese Feststellung haben wir die Anwendung einer hierarchischen Clusteranalyse sowie einer Klassifikation als Entscheidungsbaum für die Problemstellung untersucht. Beide Verfahren erzeugen anhand der nominalen Konfigurationsdaten eine Baumstruktur, vereinfacht in Abb. 1 dargestellt, in der jeder Pfad einzeln durchlaufen werden kann. Der stattfindende Test in den Knoten basiert auf der Entscheidung, ob ein Code vorhanden ist oder nicht. Die dafür notwendige Datenstruktur ist in Abb. 1 rechts dargestellt. In diesen Pfaden ist in den oberen Verzweigungen eine signifikant hohe Anzahl von Standardcodes vertreten, die durch die Einbettung in die Baumstruktur, nicht herausgelöst werden können. Durch den o.g. Punkt, dass Standardcodes nicht als solche erkannt werden können, bewerten wir die beiden Data-Mining Techniken bzgl. dieser Problematik als ungeeignet. Diese Bewertung begründet sich darin, dass die Relevanz einzelner Codes nicht ersichtlich ist. Der in Abb. 1 gelb markierte Pfad stellt

dabei eine Konfigurationsausprägung dar. Die in der Abbildung aufgeführten Knoten sind deren jeweiligen Tests. Durch die vorhandene Struktur können Codekombinationen nicht wie beim Item Set Mining unabhängig voneinander untersucht werden, sondern es können nur diejenigen Kombinationen betrachtet werden, die einen Pfad durch den Baum beschreiben. Hingegen werden durch die Baumstruktur jegliche Überschneidungen von Codes mehrerer Konfigurationsausprägungen und daher mehrerer Pfade im Baum vermieden. Aus diesem Grund bewerten wir die Clusteranalyse und Klassifikation in der Schnittmengenproblematik als bedingt geeignet. Nur bedingt deshalb, weil es dafür erforderlich ist, alle Pfade des Baumes von Anfang bis Ende zu durchlaufen. Aufgrund der Vielfalt an möglichen Fahrzeugvarianten beträgt die Anzahl an Ästen nahezu der Anzahl an Instanzen, was nicht zielführend ist, um relevante Codes zu identifizieren.

### 3.1 Assoziationsanalyse

Im Kontext der Assoziationsanalyse haben wir uns für den Apriori-Algorithmus entschieden, da er am bekanntesten ist und wir keine Performancevorgaben haben [AS94].

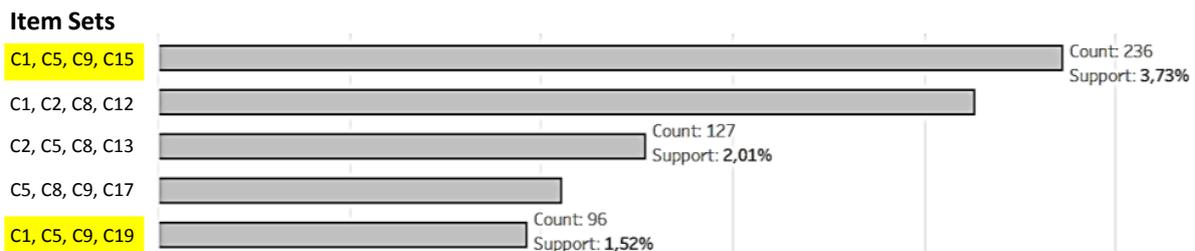


Abb. 2: Ergebnisauszug aus dem Item Set Mining

In Abb. 2 haben wir einen Auszug des visualisierten Ergebnisses aus dem Item Set Mining mit den o.g. 6.327 verwendeten Instanzen aufgeführt. Die erste gelb markierte Kombination ist von ihrer Anzahl dahinterstehender Fahrzeuge am höchsten. Da Abb. 2 das Ergebnis mit allen Instanzen darstellt und nicht zwischen bereits ausgefallenen Fahrzeugen und bisher nicht auffälligen Fahrzeugen unterscheidet, haben wir diese Codekombination gegen unsere Datenbank abgefragt. Dabei stellt sich heraus, dass die Kombination C1, C5, C9, C15 zwar den größten Anteil in den Daten bildet, dennoch sind von diesen 236 Fahrzeugen lediglich 17 bereits im Feld ausgefallen. Bei der untersten Kombination C1, C5, C9, C19 haben wir hingegen ermittelt, dass von diesen 96 Fahrzeugen, 72 davon bereits einen Feldausfall haben. Um die Relevanz aller Codekombinationen bzgl. Feldausfällen zu ermitteln, sind für das Item Set Mining zwei Analysedurchgänge erforderlich und zwar einmal mit den Fahrzeugen mit Feldausfällen und einmal mit den Fahrzeugen ohne Feldausfall. Anhand des Beispiels stellen wir fest, dass wir nur einzelne Codekombinationen als relevant identifizieren können, die lediglich einen Anteil am Schadensgeschehen ausmachen. Um eine vollständige Abbildung der Feldausfälle zu erzielen, ist es erforderlich weitere Kombinationen zu betrachten. Dieser Punkt richtet sich an die Herausforderung der Schnittmengenproblematik, da die Kombination mehrerer Item Sets nicht zwangsläufig zu einem besseren Ergebnis führt.

Vielmehr ist es notwendig die daraus resultierende Abdeckung von Felddausfällen sowie die Anzahl aller mit diesen Kombinationen produzierten Fahrzeuge zu ermitteln. In Bezug auf die datentechnische Herausforderung der Standardcodes bewerten wir das Item Set Mining als bedingt geeignet, da die erzeugten Item Sets unabhängig voneinander betrachtet werden können. Im Gegensatz zu anderen Data-Mining Verfahren wie der Klassifikation entsteht hierbei keine Strukturierung, weshalb einzelne Items gezielt ignoriert werden können, ohne das Analyseergebnis weiterer Item Sets zu beeinflussen. Bei der Klassifikation hingegen wird damit die gebildete Struktur des Entscheidungsbaumes zerstört, wodurch die Ergebnisse obsolet werden. Dennoch ist für das Item Set Mining ein vorgelagerter Schritt zu entwickeln um die minimale Menge an einzelnen Codes zu bestimmen, mit denen das Schadgeschehen abgebildet werden kann. Um auf die Herausforderung der Schnittmengenproblematik einzugehen, betrachten wir dazu das erste und letzte Item Set in Abb. 2. Hierbei wollten wir wissen, wie viele Fahrzeuge es mit beiden Sets zusammen gibt. Das Resultat daraus sind 251 Fahrzeuge. Somit stellen wir fest, dass sich 81 Fahrzeuge durch die Vereinigungsmenge der beiden Vierer-Item Sets charakterisieren lassen. Daraus lässt sich schließen, dass die Codes, welche die beiden Item Sets unterscheiden, anteilig jeweils in beiden Item Sets enthalten sind. In diesem Beispiel sollte demnach ein Fünfer-Item Set gebildet werden, um die Schnittmengenproblematik der beiden Vierer-Item Sets aufzulösen. Daraus ergibt sich die Folgeproblematik, dass eine Entscheidung darüber getroffen werden muss, an welchem Punkt die Erzeugung von weiteren Item Sets gestoppt werden soll. Diese Entscheidung sollte durch testweises Kombinieren erzeugter Item Sets erfolgen. Dabei wird die eingangs erwähnte Zielgröße der Teileversorgung von maximal 50% als zu erreichende Abbruchbedingung herangezogen. Der Standard Apriori-Algorithmus kann diese Anforderung nicht erfüllen, weswegen wir das Item Set Mining bzgl. der Schnittmengenproblematik als ungeeignet bewerten. Ergänzend zum Item Set Mining haben wir die Ergebnisse aus der Erzeugung von Assoziationsregeln auf Basis der bereits berechneten Item Sets untersucht. Das Fazit dazu ist, dass die Regeln nicht den Analysezweck erfüllen, da mit diesen lediglich eine Wenn-Dann-Beziehung innerhalb der Codes aufgezeigt wird. Die Erzeugung der Item Sets als vorgelagerter Analyseschritt brachte hingegen mehr Erkenntnisse über die vorliegenden Daten. Aus den Sets konnten wir per Ranking einzelne Merkmale finden, welche nicht in Kombination mit Standardcodes ausgewiesen werden und gleichzeitig in einer signifikant hohen Menge an ausgefallenen Fahrzeugen enthalten sind.

### **3.2 Zusammenfassung der Untersuchungsergebnisse**

Nachstehende Tab. 2 führt die Ergebnisse der untersuchten Standard Data-Mining Techniken auf deren Eignung bzgl. datentechnischer Herausforderungen auf. Dabei haben wir festgestellt, dass sich lediglich Teilaspekte der Vorgehensweisen bedingt für jeweils eine Herausforderung eignen. Die Conclusio der Untersuchung ist, dass sich das Item Set Mining für die Problemstellung grundsätzlich eignet, da es die erzeugten Item Sets unabhängig von einer notwendigen Struktur betrachtet. Somit können die von den Item Sets repräsentierten Teilmengen einzeln analysiert und bewertet werden. Aus den durchgeführten Untersuchun-

Datentechnische Herausforderung	Assoziationsanalyse (Item Set Mining)	Clusteranalyse (Hierarchisch)	Klassifikation (Entscheidungsbaum)
Standardcode-problematik	Bedingte Eignung	Ungeeignet	Ungeeignet
Schnittmengen-problematik	Ungeeignet	Bedingte Eignung	Bedingte Eignung

Tab. 2: Gegenüberstellung von Data-Mining-Techniken bzgl. analytischer Herausforderungen

gen ergeben sich Anforderungen, die im Kontext einer domänenspezifischen Anpassung des Standard Algorithmus erforderlich sind. (1) Vor der ersten Item Set Erzeugung soll algorithmisch geprüft werden, welche Codes überhaupt notwendig sind. Damit beschneiden wir den Ergebnisraum der Item Set Erzeugung, um somit minimale Item Sets zu identifizieren, mit denen wir trotzdem eine nahezu 100% Abdeckung ausgefallener Fahrzeuge erhalten. (2) Nach jeder Erzeugung von Item Sets, soll getestet werden, ob die erzeugten Item Sets bereits ausreichen, um die Prämisse von maximal 50% produzierten Fahrzeugen mit diesen Codes zu erfüllen. Hierfür wird, ausgehend von einem Startelement geprüft, welche nachfolgenden Item Sets in das Ergebnisset geschrieben werden sollen. Zielsetzung dabei ist, dass mit dem Hinzufügen eines Sets in die Ergebnisliste, die Abdeckung der Feldausfälle stetig steigt. Das Unterschreiten der maximalen Anzahl produzierter Fahrzeuge bildet dabei die Abbruchbedingung des Algorithmus.

## 4 Lösungsansatz und Ergebnisdiskussion

In diesem Abschnitt stellen wir den Grundgedanken des Algorithmus vor, den wir für die Problemstellung, unter Berücksichtigung der datentechnischen Herausforderungen, entwickelt haben. Um die relevanten Codes für eine nahezu 100% Abdeckung bereits ausgefallener Fahrzeuge zu identifizieren und dabei die Prämisse von 50% Teileversorgung einzuhalten, sind im Ablauf des Apriori-Algorithmus, die in Abschnitt 3.2 aufgeführten Anforderungen, notwendig. In Abb. 3 haben wir die Anpassungen am Apriori-Algorithmus exemplarisch dargestellt. Im Anwendungsbeispiel haben wir einen Input von 117.000 Instanzen, wovon 2.000 einen Feldausfall haben. Zu jeder Instanz sind die Konfigurationsdaten enthalten. Der zu unterschreitende Parameter beträgt 50%.

Der erste Schritt ist die Bestimmung der kleinstmöglichen Codemenge durch den Algorithmus A1, mit der 100% der Feldausfälle abgedeckt werden können, die gleichzeitig in einer minimalen Anzahl von produzierten Fahrzeugen enthalten sind. Hierfür testet A1 iterativ, beginnend bei den Codes mit den minimalsten Stückzahlen, ob sich bereits mit diesen das Schädgeschehen abbilden lässt. Mit jeder Iteration werden weitere Codes in die Betrachtung miteinbezogen, bis alle Fahrzeuge mit Feldausfällen abgebildet werden. Damit lassen wir algorithmisch bestimmen, welche einzelnen Codes tatsächlich benötigt werden, um aus

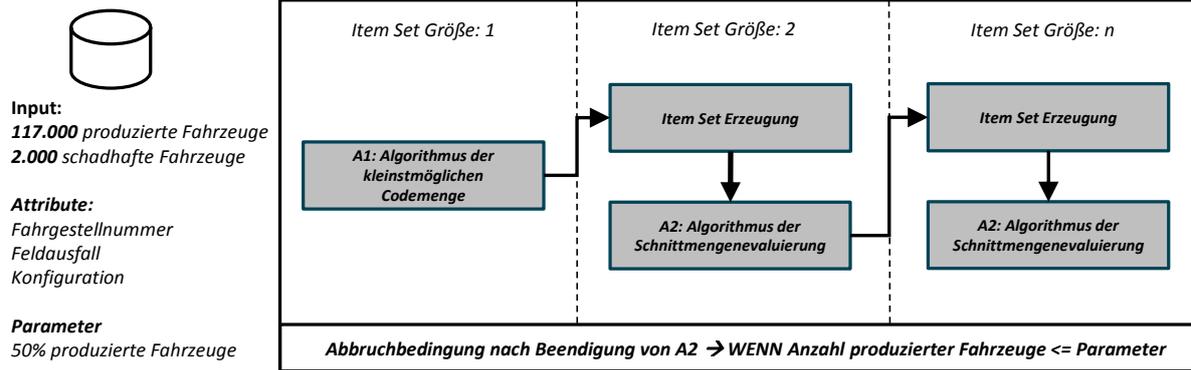


Abb. 3: Ablauf des vorgeschlagenen Algorithmus für das angepasste Item Set Mining

diesen anschließend Item Sets zu bilden. Somit löst A1 die Standardcodeproblematik und lässt sich zu Anforderung (1) einordnen.

In unserem Anwendungsbeispiel haben wir als Ergebnis von A1 25 disjunkte Codes über alle Instanzen hinweg erhalten. Die anschließende Item Set Erzeugung erfolgt nach dem Standardverfahren, jedoch werden nur Item Sets erzeugt, zu denen mindestens ein Fahrzeug mit Feldausfall existiert. Der darauffolgende Algorithmus A2 evaluiert die Item Sets bzgl. deren Schnittmengen und kumuliert in einem iterativen Prüfverfahren, welche Item Sets miteinander kombiniert werden müssen. Ziel dabei ist es, eine Abfolge von Item Set Kombinationen zu entdecken, die zusammen eine möglichst vollständige Abdeckung der Feldausfälle erreichen. Ebenso wird in A2 die von dieser Abfolge betroffene Fahrzeuganzahl berechnet. Ist die Anzahl größer als der gewählte Parameter, erfolgt erneut die Erzeugung von Item Sets der nächsten Größe. Folglich lässt sich durch A2 die Schnittmengenproblematik lösen und zur Anforderung (2) einordnen, indem die Fahrzeuganzahlen aus den Vereinigungsmengen der erzeugten Sets berechnet werden.

Item Sets	Kumulierte Feldausfälle	Kumuliertes Produktionsvolumen	Kum. Anteil an Feldausfällen	Kum. Anteil am Produktionsvolumen
C1, C6, C7	1750	57.500	87,50%	49,15%
C1, C8, C12	1400	42.000	70,00%	35,90%
C4, C9, C13	850	25.000	42,50%	21,37%

Tab. 3: Ergebnisdarstellung aus dem vorgeschlagenen Algorithmus

Der Algorithmus endet mit dem Erreichen des initial gewählten Parameters. Das Ergebnis des Algorithmus ist somit eine Liste von Item Sets, die als Abfolge kombiniert werden, um einen möglichst hohen Anteil von Feldausfällen abzudecken. In Tab. 3 ist solch ein Ergebnis dargestellt. In diesem Beispiel sind drei Item Sets identifiziert und deren kumulierten Anteile zu Feldausfällen und Stückzahlen absteigend sortiert worden. Diese Kombinationen können mittels OR-Operator kombiniert werden, um den Verbau von Bauteilen, in den damit zu identifizierenden Fahrzeugen, zu steuern. Für unseren Anwendungsfall konnten

wir durch die Anwendung des Algorithmus, der prototypisch in T-SQL implementiert ist, eine für die Produktionssteuerung geeignete Abfolge von Codekombinationen generieren [MNK15]. Diese haben wir Anfang Juli 2018 mit der o.g. Anzahl Instanzen erzeugt und mit den Felddausfällen, die bis zum Datenstand 30.11.2018 aufgetreten sind, abgeglichen. Somit haben wir das Analyseergebnis mit den bis dahin neuen Daten als Validierungsdaten getestet. Dabei kam heraus, dass wir 87% der bisher nicht betrachteten Felddausfälle mit unserer Abfolge abdecken. Diese Abfolge wird ab dem 01.12.2018 zur gezielten Bauteilsteuerung in der Serienfertigung genutzt.

## 5 Zusammenfassung und Ausblick

Für die Problemstellung, ausfallgefährdete Fahrzeugkonfigurationen für den Verbau einer angepassten Kupplung zu identifizieren, konnten wir eine domänenspezifische Lösung erarbeiten. Diese Lösung kann für alle OEMs genutzt werden, da die Datenbasis allgemein gehalten ist. Diese wird aus den Garantie- sowie den Produktionsstammdaten bereits produzierter Fahrzeuge gebildet. Mit der Identifizierung ausfallgefährdeter Fahrzeugkonfigurationen können wir die Erfahrungen aus den Gewährleistungsdaten zur Rückkopplung in die Serienfertigung nutzen. Somit können im Feld als Qualitätsproblem erkannte Fehler gezielt in der Serienfertigung vermieden werden. In zukünftigen Arbeiten wird die vorgestellte Problemstellung formalisiert sowie die Anpassungen des Algorithmus in Form eines Pseudocodes detailliert. [BW16]

## Literaturverzeichnis

- [AS94] Agrawal, Rakesh; Srikant, Ramakrishnan: Fast Algorithms for Mining Association Rules in Large Databases. In: VLDB'94, Proceedings of 20th International Conference on Very Large Data Bases, September 12-15, 1994, Santiago de Chile, Chile. S. 487–499, 1994.
- [BW16] Brunner, Franz J.; Wagner, Karl W.: Qualitätsmanagement. Leitfaden für Studium und Praxis. Hanser, München, 2016.
- [HKP12] Han, J.; Kamber, M.; Pei, J.: Data mining. Concepts and techniques. Elsevier/Morgan Kaufmann, Amsterdam, 2012.
- [JSW17] Jung, B.; Schweißer, S.; Wappis, J.: 8D - Systematisch Probleme lösen. Hanser, München, 2017.
- [MNK15] Mertins, D.; Neumann, J.; Kühnel, A.: SQL Server 2014. Galileo Press, Bonn, 2015.
- [Ve00] Verband der Automobilindustrie e.V. (VDA): , Jahresbericht, 2000.
- [Ve09] Verband der Automobilindustrie e.V. (VDA): , Schadteilanalyse Feld, 2009.
- [Wo18] Worldwide Automobile Warranties, <https://www.warrantyweek.com/archive/ww20180816.html>, Stand: 11.11.2018.