

Masterarbeit

**Untersuchung von Verifikations- und Validierungstechniken in
Bezug auf ihr Automatisierungspotential im Data Mining**

Christopher Alt

Matrikelnummer: 149937

Studiengang: Wirtschaftsingenieurwesen Master of Science

Ausgegeben am 26.03.2018

Eingereicht am 24.09.2018

Univ.-Prof. Dr.-Ing. Markus Rabe

Dr.-Ing. Dipl.-Inform. Anne Antonia Scheidler

Technische Universität Dortmund

Fakultät Maschinenbau

Inhaltsverzeichnis

Inhaltsverzeichnis.....	I
Abbildungsverzeichnis.....	II
Tabellenverzeichnis.....	III
Abkürzungsverzeichnis.....	IV
1 Einleitung.....	1
2 Grundlagen des Data Mining.....	3
2.1 Einführung in das Data Mining.....	3
2.2 Vorgehensmodelle im Data Mining.....	11
3 V&V-Techniken.....	16
3.1 Verifikation und Validierung.....	16
3.2 V&V-Techniken im Data Mining.....	18
3.3 V&V-Techniken in der Softwareentwicklung.....	21
3.4 V&V-Techniken in der Simulation.....	26
4 Analyse der V&V-Techniken.....	35
4.1 Untersuchung der V&V-Techniken.....	35
4.2 Ableitung von Kriterien zur Bewertung.....	39
4.3 Bewertung der V&V-Techniken.....	45
5 Untersuchung der V&V-Techniken im Data-Mining-Kontext.....	53
5.1 Analyse der V&V-Techniken hinsichtlich der Übertragung ins Data Mining.....	53
5.2 Übertragung ins Data Mining.....	60
6 Praktische Anwendung mit MESC.....	68
6.1 Anwendung der V&V-Techniken im MESC-Vorgangmodell.....	68
6.2 Fazit.....	78
7 Zusammenfassung und Ausblick.....	81
Literaturverzeichnis.....	84
Anhang.....	94
Eidesstattliche Versicherung.....	95

Abbildungsverzeichnis

Abbildung 1: Wissenstreppe (North 2016)	4
Abbildung 2: Data Mining Prozess (Runkler 2015)	6
Abbildung 3: Analyseansätze (Küppers 1999)	7
Abbildung 4: Prozessschritte CRISP-DM.....	11
Abbildung 5: MESc Vorgehensmodell	12
Abbildung 6: Klassifikation von Testverfahren nach Balzert (2009)	21
Abbildung 7: Gruppen der V&V-Techniken in der Simulation (Balci 2007)	27
Abbildung 8: Aufwand.....	39
Abbildung 9: Matrizen-Subtraktion	76

Tabellenverzeichnis

Tabelle 1: Übersicht der Data-Mining-Verfahren.....	10
Tabelle 2: Gegenüberstellung von Alpha-Test und Beta-Test.....	32
Tabelle 3: Gegenüberstellung von automatische und manuelle V&V-Techniken.....	36
Tabelle 4: Beispiel für eine Entscheidungsmatrix	40
Tabelle 5: Übersicht der Kriterien zur Klassifikation der V&V-Techniken hinsichtlich ihres Automatisierungspotentials	44
Tabelle 6: Kriterien-Matrix für das Automatisierungspotential von V&V-Techniken ..	45
Tabelle 7: Bewertungsübersicht der V&V-Techniken im Data Mining hinsichtlich Automatisierung	46
Tabelle 8: Bewertungsübersicht der V&V-Techniken in der Softwareentwicklung hinsichtlich Automatisierung	49
Tabelle 9: Bewertungsübersicht der V&V-Techniken in der Simulation hinsichtlich Automatisierung	51
Tabelle 10: Bewertungsübersicht der V&V-Techniken in der Kriterien-Matrix.....	52
Tabelle 11: Kriterien zur Anwendung von V&V-Techniken im Data Mining.....	56
Tabelle 12: Einsatz von V&V-Techniken im Data Mining	60
Tabelle 13: Anwendung der V&V-Techniken in Data-Mining-Verfahren.....	66
Tabelle 14: Auszug aus dem Datensatz Iris	68
Tabelle 15: Entwickelte Testfälle.....	70
Tabelle 16: Testfälle 1a und 1b.....	75

Abkürzungsverzeichnis

CRISP-DM	Cross-industry standard process for data mining
DM	Data Mining
IT	Informationstechnik
KDD	Knowledge Discovery in Databases
MESC	Musterextraktion in SCs
V&V	Verifikation und Validierung

1 Einleitung

Aktuelle Informationen aus erhobenen Daten stellen eine wichtige Grundlage dar, um flexibel auf neue Trends reagieren zu können. Dabei steht insbesondere die Echtzeitanalyse dieser Daten im Fokus aktueller IT-Forschung. Eine Methode, um Informationen aus Daten zu gewinnen, ist das Data Mining (DM). Im Zuge des Data Minings werden Daten im Rahmen von Vorgehensmodellen verarbeitet und mit geeigneten Algorithmen analysiert. Die Daten müssen möglichst schnell und effizient untersucht werden, um einen Vorteil zu erlangen (Bauernhansl et al. 2014).

Damit ein Wettbewerbsvorteil erzielt werden kann, ist die Qualität der Ergebnisse ein wesentlicher Aspekt. Die Qualitätssicherung mittels Verifikations- und Validierungstechniken (V&V-Techniken) ist deshalb wesentlicher Bestandteil der Data-Mining-Vorgehensmodelle. Allerdings benötigen diese V&V-Techniken oftmals menschliche Akteure. Diese gestalten die Echtzeitanalyse kompliziert. Daher wird auf das Ziel hingearbeitet V&V-Techniken zu automatisieren, um schneller und effizienter die Ergebnisqualität überprüfen zu können.

Zum Nachweis der Richtigkeit von Modellen werden auch in der Softwareentwicklung und der Simulation V&V-Techniken eingesetzt. Durch diese Techniken wird überprüft, ob das untersuchte Modell korrekt transformiert wurde und ob das Modell das Verhalten hinreichend genau wiedergibt (Rabe et al. 2008). Deshalb wird mit einer erfolgreichen Kontrolle durch V&V-Techniken die Richtigkeit und damit die Qualität der Ergebnisse geprüft und gewährleistet.

Ziel dieser Arbeit ist, die Automatisierungspotentiale von geeigneten V&V-Techniken für das Data Mining aufzuzeigen. Die Auswahl der Techniken sowie die Bewertung ihres Automatisierungspotentials werden hierbei mittels eines Kriterienkataloges nachvollziehbar gestaltet. Dazu wird außerdem die Übertragbarkeit von V&V-Techniken aus anderen IT-Bereichen in das Data Mining auf Realisierbarkeit geprüft. Abschließend werden ausgewählte Techniken exemplarisch im Rahmen des Vorgangsmodells MESC (Musterextraktion in SCs) angewandt. MESC beschreibt ein Vorgangsmodell zur Wissensdeckung in Supply Chains (Scheidler 2017).

Da im Data Mining bisher wenige automatisierte V&V-Techniken eingesetzt werden, sondern häufig auf menschliche Akteure zur Kontrolle der Ergebnisse zurückgegriffen wird, werden in dieser Arbeit neben den V&V-Techniken im Data Mining hauptsächlich V&V-Techniken aus der Softwareentwicklung und der Simulation betrachtet und

hinsichtlich ihres Einsatzpotentials im Data Mining untersucht. Auf Grund der schnellen Veränderungen und Neuerungen im Bereich der V&V-Techniken und den Themengebieten der Simulation und der Softwareentwicklung wird im Zuge dieser Arbeit versucht die neuesten Entwicklungen (Stand September 2018) zu berücksichtigen.

Im ersten Schritt der Arbeit wird das aktuelle Vorgehen zur V&V im Data Mining dargestellt und Schwachstellen des aktuellen Vorgehens aufgezeigt. In diesem Zusammenhang werden auch die verschiedenen V&V-Techniken aus Simulation und Softwareentwicklung vorgestellt und analysiert. Aufbauend auf diesen V&V-Techniken werden Kriterien zur Klassifikation hinsichtlich ihres Automatisierungspotentials abgeleitet. Mit Hilfe dieser Kriterien, die in einer Kriterien-Matrix zusammengeführt werden, werden Automatisierungsansätze und -potentiale der ausgewählten V&V-Techniken aufgezeigt.

Im nächsten Schritt werden die betrachteten Techniken hinsichtlich ihres Übertragungspotentials ins Data Mining analysiert. Dabei wird aufgezeigt, ob und wie die Adaption in die Umgebung des Data Minings durchgeführt werden kann. Dazu wird zunächst ein Anforderungsprofil für die Techniken aufgestellt. Die Techniken werden anschließend dahingehend untersucht, in welchem Schritt des Data-Mining-Vorgehensmodells MESC sie zum Einsatz kommen. Des Weiteren wird analysiert im Zuge welcher Data-Mining-Verfahren die Techniken eingesetzt werden können.

Aufbauend auf diesen Erkenntnissen wird im letzten Kapitel eine praktische Demonstration einer spezifischen V&V-Technik in einem Data-Mining-Verfahren innerhalb des MESC Vorgehensmodell exemplarisch angewendet. Abschließend wird aufgezeigt, wie diese exemplarische Anwendung der V&V-Technik automatisiert werden kann.

2 Grundlagen des Data Mining

In diesem Kapitel werden die Grundlagen zum Data Mining vorgestellt. Zentrale Themen sind neben der Erläuterung des Begriffs des Data Minings auch die Abgrenzung zwischen Daten, Informationen und Wissen im Zuge der Wissenstreppe. Des Weiteren werden Data-Mining-Verfahren und Data-Mining-Prozesse erläutert und kurz vorgestellt. Im zweiten Teil dieses Kapitels werden Data-Mining-Vorgehensmodelle dargestellt.

2.1 Einführung in das Data Mining

Seit den 1990er Jahren gewinnt das Data Mining immer mehr an Bedeutung, da die Anzahl und Größe der Datenmengen deutlich wachsen. Diese Datenmengen in den Datenbanken können mit konventioneller Datenabfrage nur schwer analysiert werden. Auch statistische Ansätze zur Analyse sind bei diesen Datensätzen, bestehend aus nominalen und metrischen Attributen, nur bedingt geeignet. Deshalb entsteht eine immer größer werdende Lücke zwischen erhobenen Daten und dessen Nutzen bzw. Verständnis dieser Datenmengen. Um diese Vakanz zwischen Datenmengen und Analysen zu minimieren, ist der Wunsch entstanden, die vorliegenden Daten in möglichst kurzer Zeit in Hinblick auf interessante Muster zu analysieren. Aus diesem Grund ist das Bedürfnis nach Verfahren entstanden, welche diese Muster automatisch erkennen können (Küppers 1999).

Unter anderem mit dieser Problemstellung befasst sich das Forschungsfeld des Data Minings. Im Data Mining wird versucht wertvolle Informationen aus den Datensätzen zu gewinnen. Aus diesen möglichst schnell und effizient analysierten Daten werden nicht selten wichtige Informationen gewonnen, die zu einem Wettbewerbsvorteil führen können (Lassmann et al. 2006; Bauernhansl et al. 2014). Zusätzlich ist die stetige Verbesserung der Leistungsfähigkeit in der Informationstechnik zu beachten. Dadurch ist die gleichzeitige Bearbeitung großer Datenmengen erst möglich. Ende der 1990er Jahren wurden zudem Auswertungsmethoden wie neuronale Netze (weiter)entwickelt. Diese Methoden bilden häufig allein oder in Kombination mit anderen Auswertungsmethoden die Basis für Data Mining Verfahren (Küppers 1999).

Häufig wird als Synonym zu Data Mining ‚Knowledge Discovery‘ verwendet, wobei in der deutschen Sprache hauptsächlich Data Mining verwendet wird (Küppers 1999). In der Literatur wird der Begriff ‚Data Mining‘ mit ‚Datenmustererkennung‘ verstanden

(Bissantz und Hagedorn 2009). Das Data Mining ist ein Teilschritt des Knowledge Discovery in Databases (KDD).

Um den Begriff des Data Mining verstehen zu können, müssen zunächst die Begriffe Daten, Wissen und Informationen voneinander abgegrenzt werden, da diese essentiell für das Data Mining sind. Ein Problem von Unternehmen ist zurzeit nicht, dass zu wenig Daten zur Verfügung stehen, sondern dass diese richtig zu Informationen umgewandelt werden (Malik 2013). Wie in Abbildung 1 deutlich wird, generiert sich Wissen aus Informationen. Informationen wiederum setzen sich aus Daten zusammen. Daraus wird deutlich, dass Wissen weit mehr ist als lediglich eine Ansammlung aus Daten (Hamilton 2008).

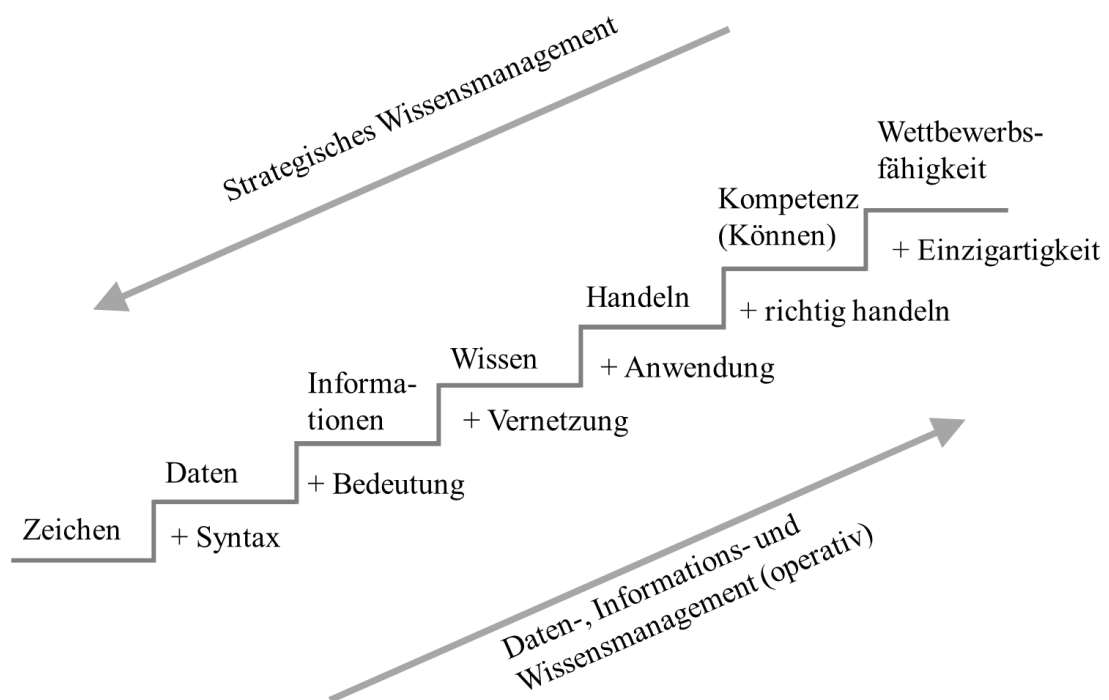


Abbildung 1: Wissenstreppe (North 2016)

Auf der untersten Ebene befinden sich die Zeichen. Diese sind die Basis für alle höheren Ebenen. Von Daten wird gesprochen, wenn Zeichen in einen regelbasierten Zusammenhang (Syntax) gestellt werden. Daten werden in der Informatik als lesbare und bearbeitbare Repräsentation von Informationen verstanden, welche als Zeichenkette dargestellt werden (Krcmar 2015). Werden die Daten in einen Kontext gesetzt, entstehen Informationen. Daten werden kombiniert und interpretiert, sodass diese als Informationen verwendet werden können (Krcmar 2015). Wissen entsteht wiederum aus Informationen, wenn Kenntnisse über deren Zusammenhängen und deren Verwendung vorhanden sind. Dann ergibt sich die Möglichkeit, die vorhandenen Informationen zu verknüpfen

(Steinmüller 1993). Demnach ist Wissen subjektiv. Daher entsteht Wissen erst durch Vernetzung, Kontextbezug von Informationen und Zweckorientierung (Krallmann et al. 2013). Daraus folgt, dass Wissen mehr ist als strukturiert gesammelte Informationen (Hamilton 2008).

Jedoch hat das gewonnene Wissen lediglich dann einen Wert für das Unternehmen, wenn dieses Wissen in Können umgewandelt werden kann. Im Unternehmen spiegelt sich dieses Können in Handlungen und Entscheidungen einzelner Mitarbeiter oder Mitarbeitergruppen der Firma wider. Sofern eine Bereitschaft der Mitarbeiter zur persönlichen Weiterentwicklung vorliegt, besteht die Möglichkeit, durch das Vermitteln von Wissen die individuelle Kompetenz einzelner Mitarbeiter zu steigern. Die Handlungen der Mitarbeiter liefern messbare Ergebnisse, welche das Unternehmen bewerten kann. North bezeichnet den Vorgang, Wissen in konkretes und korrektes Handeln umzuwandeln, als Kompetenz (North 2016). Da Kernkompetenzen von Mitarbeitern kaum zu kopieren sind, entstehen durch diese Kompetenzen neue Möglichkeiten für das Unternehmen. Durch diese neuen Möglichkeiten kann das Unternehmen seine Wettbewerbsfähigkeit erhöhen. Folglich stellt eine Kernkompetenz ein Alleinstellungsmerkmal eines Unternehmens dar.

Somit sind alle Stufen der Wissenstreppe für den Erfolg eines Unternehmens wichtig und das Überspringen einer Stufe sollte vermieden werden. Durch das Auslassen einer Stufe können Probleme bei der Umsetzung von Geschäftsstrategien auftreten (North 2016).

Data Mining kann dabei helfen, aus Daten Wissen zu gewinnen, und kann demnach bei richtiger Anwendung für einen Wettbewerbsvorteil sorgen. Wörtlich übersetzt bedeutet Data Mining Schürfung von Daten. Das Wort ‚Mining‘ stammt ursprünglich aus dem Bergbau, wonach analoger Weise Daten aus Datenmengen ausgearbeitet werden (Petersohn 2009). Demnach werden mit Data Mining Daten geformt und verarbeitet. Das Ziel von Data Mining ist jedoch nicht nur Daten zu formen, sondern auch Informationen und Wissen aus den Daten zu generieren. „Data Mining ist die Extraktion von Wissen aus Daten“ (Cleve und Lämmel 2016). Auch Runkler (2010) definiert Data Mining mit dem Ziel neues Wissen zu extrahieren. Dabei ist hier Wissen als interessante Muster anzusehen. Interessante Muster sind solche, die allgemein gültig, nicht trivial, neu, nützlich und verständlich sind (Runkler 2010). Ein Muster ist als Beschreibung von Daten definiert, die einfacher ist als die Aufzählung der Daten selbst (Fayyad 1996). Das heißt das Ziel von Data Mining ist zunächst das Formen und bearbeiten von Daten um anschließend aus diesen Informationen und Wissen generieren zu können.

Somit ist Data Mining definiert als Prozess zum Entdecken neuer Muster in Daten (Witten et al. 2011). Wie in Abbildung 2 dargestellt, besteht der grundsätzliche Data Mining Prozess aus vier Schritten. Zunächst wird der Prozess geplant und vorbereitet. Darauffolgend werden die Daten vorverarbeitet, um diese in der Datenanalyse verwenden zu können. Anschließend wertet der Data Mining Prozess die Daten aus.

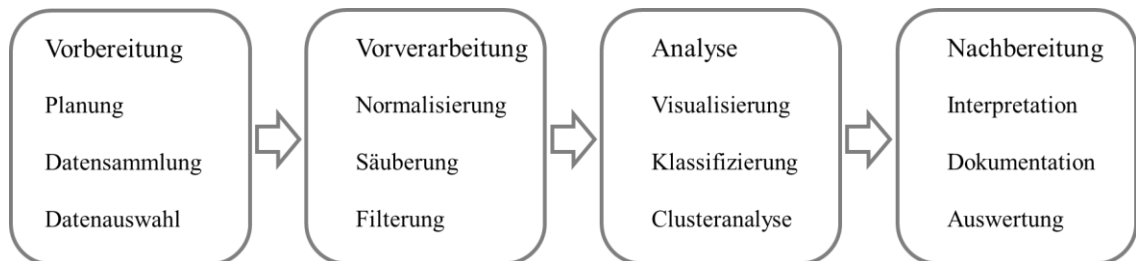


Abbildung 2: Data Mining Prozess (Runkler 2015)

Dabei können die Rohdaten in verschiedenen Strukturstufen vorliegen. Unterschieden wird zwischen unstrukturierten Daten, semistrukturierten Daten und strukturierten Daten (Cleve und Lämmel 2016). Ein Beispiel für die Analyse von unstrukturierten Daten ist das sogenannte ‚Text Mining‘. In Texten werden zunächst scheinbar ohne Struktur Zeichen aneinandergereiht, weshalb von unstrukturierten Daten gesprochen wird. Im Text Mining werden Texte analysiert und um Erkenntnisse aus diesen zu gewinnen (Kwartler 2017). Web Mining befasst sich mit semistrukturierten Daten. Fälschlicherweise wird Data Mining oft nur mit bereits strukturierten Daten in Verbindung gebracht. Allerdings stellt Data Mining lediglich den Oberbegriff dar (Cleve und Lämmel 2016).

Eine notwendige Eigenschaft von Data-Mining-Verfahren ist die Hypothesenfreiheit. Das bedeutet, Muster werden frei von Subjektivität gesucht. Der Anwender des Verfahrens stellt vor dem Verfahren keine zu beweisende Annahme auf. Stattdessen werden die zu überprüfenden Hypothesen aus den Daten selbst generiert. Dadurch wird die Suche nicht im Vorhinein eingeschränkt (Küppers 1999).

Während der traditionelle Ansatz stark von mehreren Benutzern abhängt, wird im Data-Mining-Ansatz vieles von einem Computersystem übernommen (siehe Abbildung 3). Ein großer Unterschied liegt darin, dass im Data Mining die Hypothese nicht von einem Anwender aufgestellt werden muss. Sondern vielmehr wird danach gesucht, was interessant ist (Bigus 1996). Interessant bedeutet, dass eine bestimmte Menge an Objekten diesem Muster folgen. "Das Data-Mining-System generiert aus einem beliebigen Datenbestand die Hypothese selbst" (Küppers 1999).

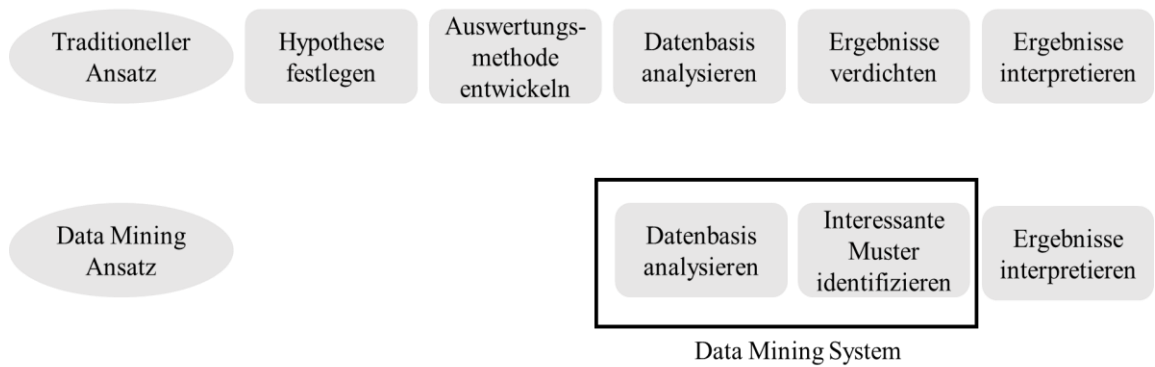


Abbildung 3: Analyseansätze (Küppers 1999)

Der traditionelle Ansatz verfolgt das Ziel zu erforschen ‚Zeige mir was mich interessiert‘, während der Data Mining Ansatz versucht herauszufinden, was allgemein interessant ist.

Als weitere wichtige Eigenschaft in der Definition von Data Mining ist die Forderung nach Automatismus (Perner 2012). Der Benutzereingriff soll durch diese Forderung möglichst gering ausfallen. Das Ziel des Data Mining, große Datenmengen auszuwerten, ist mit möglichst wenigen Benutzereingriffen zu erreichen. Denn jeder Eingriff eines Benutzers verlangsamt unter Umständen diesen Prozess. Allerdings wird ein autonomeres System gleichzeitig auch unflexibler (Küppers 1999). In der Praxis bedeutet dies, dass sehr autonome Systeme nur für bestimmte Anwendungsfälle geeignet sind. In diesem Zusammenhang wird in der Data-Mining-Forschung auch vom Grad der Hypothesenfreiheit gesprochen (Bissantz und Hagedorn 2009).

Demnach sollte ein ideales Data-Mining-System, neben den Kriterien der Ergebnisse, Korrektheit und Verständlichkeit, drei Eigenschaften vereinigen (Küppers 1999):

- Mustergenerierung
- Automatismus
- Interessantheit

Zunächst sollte ein jedes Data-Mining-System den eigentlichen Zweck erfüllen und aus den vorhandenen Daten Muster generieren. Diese Muster müssen jedoch für den Anwender eine Bedeutung haben. Des Weiteren sind automatisch ablaufende Prozessschritte im Data Mining erforderlich. Zusätzlich müssen die Ergebnisse zum einen korrekt sein und zum anderen einen Grad an Interessantheit und Relevanz für die betrachteten Datensätze haben.

Um diese Muster zu identifizieren und aufzuzeigen, werden verschiedene Data-Mining-Verfahren verwendet. Im Folgenden werden mit der Clusteranalyse, der

Assoziationsanalyse und der Klassifizierung wesentliche Data-Mining-Verfahren vorgestellt.

Die Clusteranalyse ist eines der wichtigsten und häufigsten angewendeten Data-Mining-Verfahren. Das Ziel einer Clusteranalyse ist die Identifizierung von Clustern innerhalb der Daten. Bei der Clusteranalyse werden aus einer Gesamtmenge, die aus unterschiedlichen Objekten besteht, ähnliche Objekte zu einer Gruppe (Cluster) zusammengefasst. Die Objekte in den entstehenden Clustern sind möglichst ähnlich, grenzen sich aber von den anderen Clustern deutlich ab (Ester und Sander 2000). Dabei kann das Vorgehen zur Clusterbildung grob in drei Schritte unterteilt werden (Backhaus 2003):

- Auswahl der Variablen: Um die Objekte miteinander vergleichen zu können, müssen geeignete Variablen zur Untersuchung gefunden werden.
- Proximitätsbestimmung: In diesem Schritt werden die Distanzen zwischen zwei Objekten bestimmt. Dazu muss zunächst ein Proximitätsmaß ausgewählt werden. Ein Beispiel dafür stellt die das euklidische Distanzmaß dar. Dabei ist die Distanz zwischen Objekt i und Objekt j wie folgt definiert:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

Auf Grundlage des gewählten Proximitätsmaßes werden die Distanzwerte bzw. Ähnlichkeitswerte bestimmt. Diese Werte beschreiben, wie ähnlich zwei Attribute zueinander sind. Mit Hilfe des Distanzmaßes werden die Objekte einzeln miteinander verglichen und die Distanz zueinander berechnet.

- Clusterbildung: Hier wird ein Clusterverfahren festgelegt und durchgeführt. Um Objekte einem Cluster zuzuordnen, wird zwischen harten und weichen Methoden unterschieden. Bei den harten Methoden (z.B.: k-means) wird jedes Objekt genau einem Cluster zugeordnet, wohingegen eine weiche Methode ein Objekt mehreren Clustern zuordnen kann (Ester und Sander 2000).

Um die Cluster zu finden wurde eine Vielzahl an Verfahren entwickelt. Bacher et. al. (2010) unterscheiden dabei zwischen drei großen Verfahrensgruppen: unvollständige Clusteranalyseverfahren, deterministische Clusteranalyseverfahren und probabilistische Clusteranalyseverfahren (Bacher et al. 2010).

Im hierarchischen Clustering werden Cluster auf der Basis der Distanzen der Objekte zueinander erstellt. Dabei wird zwischen top-down und bottom-up unterschieden. Zur Visualisierung werden Baumstrukturen verwendet (auch Dendrogramm genannt). Der Baum zeigt die Zerlegung der Gesamtmenge kleiner werdender Teilmengen. Die Blätter repräsentieren bei einer top-down Clustering ein Cluster mit nur einem Objekt, während die Wurzel ein Cluster mit allen Objekten darstellt. In einer bottom-up Clustering wird dies umgekehrt dargestellt. Ein Problem des hierarchischen Clusterings stellt der Aufwand zur Analyse der Ergebnisse dar. Denn ein hierarchisches Clusterverfahren präsentiert im Dendrogramm viele Lösungsmöglichkeiten zum Clustern der Objekte. Andere Verfahren wie der k-means-Algorithmus liefern lediglich eine Möglichkeit zur Clusterung.

Das Verfahren der Klassifizierung ordnet die Objekte bereits bestehenden Klassen zu. Darin liegt der Unterschied zwischen der Clusteranalyse und der Klassifizierung (Schlosser und Klages 1976). Vier der wichtigsten Klassifikationsverfahren im Data Mining sind Entscheidungsbäume, Bayes-Klassifikation, nearest-neighbor-Verfahren und neuronale Netze. Die meisten Klassifikationsverfahren im Data Mining bestehen aus zwei Phasen. Zunächst erfolgt die Lernphase anhand von Trainingsdaten und anschließend wird die Klassifikationsphase durchgeführt (Oberhauser und Gödert 2005).

Bei Entscheidungsbäumen durchläuft jedes nicht klassifizierte Objekt den Baum, wobei an jedem Knoten geprüft wird, welchen Pfad dieses Objekt weiterverfolgen soll. Der letzte Blattknoten gibt an welcher Klasse das Objekt zugeordnet wird. Erstellt werden können die Entscheidungsbäume manuell und durch Automatismen des maschinellen Lernens. Der Vorteil der Entscheidungsbäume ist die leichte Verständlichkeit der Klassenzuordnung (Pierson und Porway 2016).

Die Bayes-Klassifikation beruht auf Wahrscheinlichkeiten. Grundlage für die Zuordnung eines Objektes in eine Klasse stellt die Merkmalsausprägung dar. Das betrachtete Objekt wird der Klasse zugeordnet, bei der die Wahrscheinlichkeit der Merkmalsausprägung am größten ist (Walter 2004). Die Merkmalsausprägungen werden mit Hilfe von Trainingsdaten vorher geschätzt. Dabei erfolgt die Einordnung bei einer großen Menge an Daten sehr genau (Pierson und Porway 2016).

Das nearest-neighbor-Verfahren vergleicht die Objekte untereinander und ordnet diese einer Klasse zu. Die Trainingsobjekte bilden dabei die Grundlage für diesen Vergleich. Die Ähnlichkeits- und Distanzmaße geben an, wie weit die beiden Objekte entfernt sind (Lawler 1995).

Neuronale Netze bestehen aus Knoten, die in mehrere Schichten (meist Eingabe-, Zwischen- und Ausgabeschicht) unterteilt sind (Scherer 1997). Jeder Knoten ist mit jedem Knoten aus der anliegenden Schicht verbunden. Die sogenannten Kanten besitzen ein Kantengewicht. Das Kantengewicht gibt an, zu welchem Knoten ein Objekt als nächstes gelangt und somit letztendlich einem Knoten in der Ausgabeschicht zugeordnet werden kann. Ein Knoten in der Ausgabeschicht beschreibt eine Klasse (Rey und Wender 2018). Die Kantengewichte werden mit Hilfe der Trainingsdaten angepasst. Somit sind Ausreißer in den Daten leicht zu entdecken.

Die Assoziationsanalyse sucht in den Datensätzen nach sogenannten Regeln. Diese Regeln beschreiben den Zusammenhang zwischen zwei Merkmalen eines Objektes. Ziel dabei ist die Folgerung von einem Merkmal auf ein anderes Merkmal. Eine typische Anwendung dafür ist die Warenkorbanalyse (Mann 2010). In einem Webshop wird beispielsweise beim Kauf eines Laptops dem Käufer angezeigt, welches Produkt andere Kunden, die denselben Artikel gekauft haben, zusätzlich bestellt haben (z.B. Laptophülle, Maus). Dazu werden der Support und die Konfidenz berechnet. Der Support beschreibt die Anzahl an Beispielen, für die die Regel angewendet werden kann, während die Konfidenz angibt, wie häufig die Regel zutrifft. Abhängig von diesen beiden Werten wird entschieden, ob eine Regel als stark eingestuft wird (Cleve und Lämmel 2016).

In Tabelle 1 gibt eine Übersicht über die drei Data-Mining-Verfahren und fasst diese kurz zusammen. Im folgenden Kapitel werden Vorgehensmodelle vorgestellt, in denen die beschriebenen Data-Mining-Verfahren zum Einsatz kommen.

Tabelle 1: Übersicht der Data-Mining-Verfahren

Data-Mining-Verfahren	Beschreibung
Clusteranalyse	Auf Grundlage der Objekte werden Cluster gebildet, denen die Objekte zugeordnet werden
Klassifikation	Objekte werden in bereits bestehende Klassen klassifiziert
Assoziationsanalyse	Suche nach Regeln, um Zusammenhänge zu beschreiben

2.2 Vorgehensmodelle im Data Mining

Ein Vorgehensmodell gibt einen Überblick über einen Prozess und organisiert diesen. Das erste Vorgehensmodell, welches im Bereich des Data Mining angewendet wurde, ist das ‚From Data Mining to Knowledge Discovery in Databases‘ und wurde von Fayyad et al. (1996) entworfen. Dieses Modell basiert auf den Arbeiten von (Frawley et al. 1992), welche 1992 den Begriff des Data Minings als Erste definierten. Das vierstufige Modell von Berry und Linoff ‚virtuous cycle of data mining‘ konzentriert sich auf die Nutzung des Wissens und die wirtschaftlichen Folgen der Ergebnisse. Des Weiteren existieren Modelle, welche speziell für bestimmte Anwendungen entwickelt worden sind. Jedoch können diese auf andere Anwendungsfälle übertragen werden.

Ein aktuell bekanntes Vorgehensmodell ist das CRISP-DM. Das Data Mining Vorgehensmodell Cross-industry standard process for data mining (CRISP-DM) ist ein branchenüblicher Standardprozess im Data Mining. Dieses Data-Mining-Prozessmodell beschreibt häufig verwendete Ansätze, mit denen Experten Probleme angehen.

Im Vorgehensmodell CRISP-DM ist der Data Mining Prozess in sechs Hauptphasen, gegliedert (Harper und Pickett 2006). Beschrieben ist der Prozessablauf mit den Hauptphasen in Abbildung 4. Dabei ist zu beachten, dass die Reihenfolge der Phasen nicht streng festgelegt ist, sondern entsprechend dem konkreten Fall angepasst werden kann. Die Pfeile in der Abbildung verdeutlichen die häufigsten und wichtigsten Abhängigkeiten zwischen den Hauptphasen. Der Fokus dieses Modells liegt auf dem Projektgeschäft, was durch die erste Phase ‚Geschäftsverständnis‘ deutlich wird (Scheidler 2017).

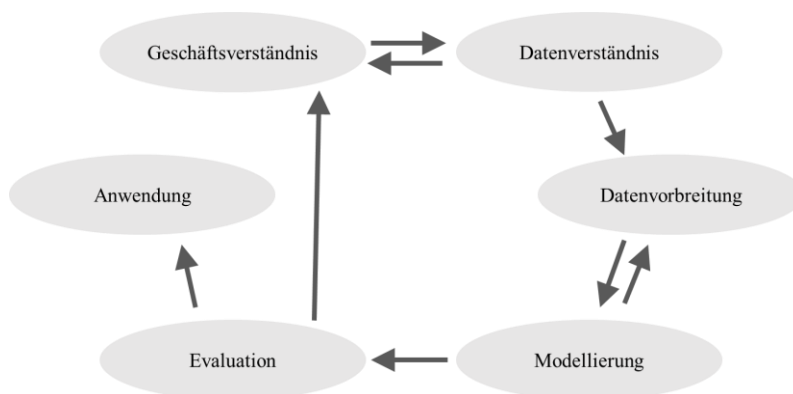


Abbildung 4: Prozessschritte CRISP-DM

Im Jahr 2015 veröffentlichte IBM Corporation eine neue Methodik namens ‚Analytics Solutions Unified Method für Data Mining / Predictive Analytics‘ (auch bekannt als ASUM-DM), die CRISP-DM verfeinert und erweitert (Haffer 2015).

Ein weiteres bekanntes Vorgehensmodell ist das Modell von Hippner und Wilde (Hippner 2001). Dieses ist angelehnt an das an den Marketingumfeld stammende Modell von Fayyad (1996). Das Modell von Hippner und Wilde (2001) besteht aus sieben Phasen, welche hier nicht genauer erläutert werden, da das nachfolgend behandelte Modell MESOC auf diesem Modell beruht. Scheidler (2017) hat dieses Vorgehensmodell auf die Bereiche des Marketings und Supply Chains übertragen. Die Phasen werden anders als im CRISP-DM Vorgehensmodell nacheinander der Reihe nach bearbeitet. Das MESOC Vorgehensmodell ist in sieben Hauptphasen unterteilt, welche im weiteren Verlauf behandelt werden (siehe Abbildung 5).

1. Aufgabendefinition: Zunächst ist eine konkrete Fragestellung zu definieren und festzulegen, um das KDD in der Praxis durchführen zu können. Dafür wird ein Problem unter den gegebenen Bedingungen ausgewählt. Diese Problemstellung wird sowohl unter zeitlichen, technischen als auch fachlichen Randbedingungen formuliert. Typische Randbedingungen in der Produktionslogistik können beispielsweise die räumliche Anordnung der Maschinen oder die verwendeten Transportmittel sein. Mit Hilfe dieser Bedingungen werden Ziele für das Data Mining abgeleitet und anschließend dokumentiert. Jedoch stellt das Hauptziel in jedem Data-Mining-Prozess in der Produktionslogistik die Optimierung der Prozesse dar, wodurch ein Wettbewerbsvorteil entsteht. Die erarbeiteten Ziele und Anforderungen in dieser Phase müssen im Verlauf des KDD berücksichtigt werden und helfen beispielsweise bei der Wahl relevanter Daten und bei der Auswahl des geeigneten Data-Mining-Verfahrens.

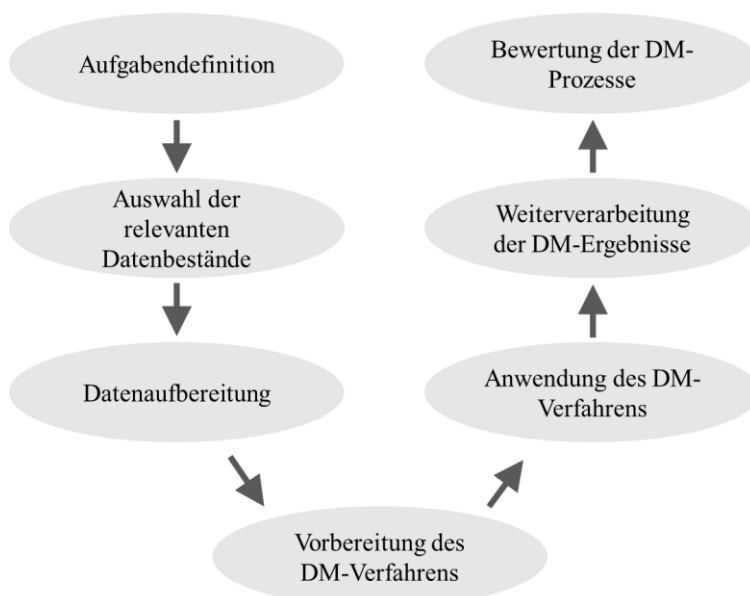


Abbildung 5: MESOC Vorgehensmodell

2. Auswahl der relevanten Datenbestände: Diese Hauptphase des MESC-Vorgehensmodell besteht aus den zwei Schritten der Datenbeschaffung (2.1) und anschließend der Phase der Datenauswahl (2.2). Zunächst werden in der Datenbeschaffung geeignete Datenquellen in Abhängigkeit von der ersten Hauptphase, der Aufgabendefinition, ausgewählt. Die Schwierigkeit besteht darin, die relevanten Daten in den Datenbeständen mit Hilfe von Kontextwissen zu identifizieren, da die Informationen häufig in komplexen Datenkonstrukten und nur selten in einem System vorliegen. Unter Kontextwissen wird Wissen verstanden, welches nicht aus den Daten ableitbar ist. Kontextwissen bezieht sich beispielsweise auf im Unternehmen vorhandene Prozesse und wird in diesem Fall durch Befragung von Experten des Unternehmens gewonnen. Häufig werden mehrere beteiligte Personen hinzugezogen, da das Kontextwissen einer einzelnen Person häufig nicht ausreicht. Grund dafür sind komplexe Konstrukte mit teilweise unabhängigen Systemen, welche von einer Einzelperson schwierig zu überblicken sind. Nach der Erfassung der relevanten Datenbestände erfolgt im nächsten Schritt die Datenauswahl. Auch die Datenauswahl erfolgt mit Hilfe von Kontextwissen mit dem Ziel nur die relevanten Informationen für die Aufgabenstellung aus den Datenbeständen herauszufiltern.

3. Datenaufbereitung: Nach der Beschaffung und Auswahl der Daten folgt die Hauptphase der Datenaufbereitung, welche die zeitintensivste Hauptphase darstellt. Abhängig von der Ausgangslage der Daten sind vier verschiedene Schritte zu durchlaufen. Die ausgewählten Daten stehen häufig in Form einer relationalen Datenbank mit einer Vielzahl an Tabellen zur Verfügung. Um die Daten in einem Data-Mining-Verfahren zu verwenden wird gefordert, diese in exakt einer Tabelle abzubilden. Dazu werden im Schritt der Formatstandardisierung (3.1) die Datenbestände in das vom Data Mining verlangte Format transformiert. Anschließend werden in der Gruppierung (3.2) die Daten unter Berücksichtigung der Fragestellung in fachliche Gruppen gegliedert. Im Schritt der Datenanreicherung (3.3) werden mit Hilfe von Kontextwissen eventuell vorhandene Lücken in Datenbeständen ausgefüllt. Im abschließenden Schritt zur Aufbereitung der Daten der Transformation (3.4) werden fehlerhafte Attribute in den Zeilen berichtigt. Dabei werden unter anderem überflüssige Attribute und Ausreißer in den Datensätzen reduziert.

4. Vorbereitung des Data-Mining-Verfahrens: In der nächsten Hauptphase des MESC Vorgehensmodell wird das Data-Mining-Verfahren vorbereitet. Auch diese Hauptphase ist in vier Schritte unterteilt. Im ersten Schritt wird das zu verwendende Data-Mining-

Verfahren ausgewählt (4.1). Dabei wird aus der Vielzahl an existierenden Verfahren, wie beispielsweise das Clusterverfahren, die Assoziationsanalyse und die Klassifikation, das ideale Verfahren für die Aufgabenstellung unter den Bedingungen ausgesucht. Auch müssen neben dem Problemtyp der zur lösenden Aufgabe unterschiedliche Kriterien wie die Generalisierbarkeit und Interpretierbarkeit der Data-Mining-Werkzeuge berücksichtigt werden. Nach der Wahl des durchzuführenden Verfahrens wird im zweiten Schritt dieser Phase die Werkzeugauswahl (4.2) des Data Minings getroffen. Der Begriff ‚Werkzeug‘ bezeichnet in diesem Zusammenhang beispielsweise die zu nutzende Software (z.B.: SAP BI, RapidMiner), welche unterstützend zum Data Mining eingesetzt wird. RapidMiner ist mit ungefähr 20% die am meisten verwendete Software im Bereich des Data Minings (Weskamp et al. 2014). Weitere Beispiele für eine unterstützende Software im Data Mining sind SPSS und SAP BI. Anschließend folgen fachliche Kodierung (4.3) und technische Kodierung (4.4). Grund dafür ist, dass die anfängliche Kodierung der ausgewählten Daten nicht für das Data-Mining-Verfahren geeignet ist. In der technischen Kodierung müssen die Ausprägungen der Attribute daher in eine geeignete Kodierung für das Verfahren geändert werden.

5. Anwendung des Data-Mining-Verfahrens: Nachdem die Daten erfolgreich für das ausgewählte Data-Mining-Verfahren vorbereitet worden sind, kann in der nächsten Hauptphase die Anwendung der Data-Mining-Verfahren erfolgen. Dabei können auch mehrere Verfahren angewendet werden. Diese können sowohl sequenziell als auch parallel zueinander durchgeführt werden. Unterteilt ist diese Phase in die Schritte der Entwicklung (5.1) sowie des Trainings des Data-Mining-Modells (5.2). In der Entwicklung erfolgt eine Unterteilung der Daten in Trainings-, Validierungs- und Testdaten. Während die Trainingsdaten dabei helfen das Grundmodell zu entwickeln und die Parameter im Modell festzulegen, dienen die Validierungsdaten zur Prüfung der Ergebnisse nach Verwendung der Testdaten. Erfolgt eine negative Validierung, ist die Modellentwicklung zwingend fortzusetzen (Hippner 2001).

6. Weiterverarbeitung der Data-Mining-Ergebnisse: Nach der Anwendung des Verfahrens folgt die Phase der Weiterverarbeitung der Data-Mining-Ergebnisse, welche in zwei Schritte unterteilt ist. Zunächst erfolgt der Schritt der Extraktion handlungsrelevanter Data-Mining-Ergebnisse (6.1). Hierbei werden im Hinblick auf die Fragestellung aussagekräftige Ergebnisse mit Hilfe von technischen Maßzahlen ausgearbeitet. Diese Maßzahlen helfen dabei, den subjektiven Vorgang der Bewertung von Interessantheit der Ergebnisse, zu neutralisieren. Zum Einschätzen der Ergebnisse

bezüglich ihrer Interessantheit helfen Kriterien wie Neuheit, Nützlichkeit, Validität und Verständlichkeit. Als zweiter Schritt in dieser Phase erfolgt der Schritt der Darstellungstransformation der Data-Mining-Ergebnisse (6.2). Hierbei werden die Ergebnisse in eine - sowohl für das eingesetzte Data-Mining-Verfahren als auch für die Aufgabenstellung geeignete - Darstellungsform überführt.

7. Bewertung der Data-Mining-Prozesse: In der letzten Phase des MESC-Vorgehensmodells erfolgt eine Bewertung der Data-Mining-Prozesse. Die Bewertung splittet sich in die Qualitätskontrolle des Data-Mining-Prozesses (7.1) und die Rückführung von Data-Mining-Ergebnissen (7.2) auf. Die Qualitätskontrolle wird mittels (Verifikations- & Validierungs) Maßnahmen durchgeführt. Im letzten Schritt findet eine Rückführung der Ergebnisse statt. Diese können als Grundlage für weitere Data-Mining-Prozesse verwendet werden.

3 V&V-Techniken

In diesem Kapitel werden die Begriffe der Verifikation und Validierung definiert und voneinander abgrenzt. Anschließend wird kurz auf den Begriff der Automatisierung eingegangen und erläutert. Folgend werden zunächst die V&V-Techniken vorgestellt, die bereits im Data Mining eingesetzt werden. Danach werden die V&V-Techniken dargestellt, die nicht im Data Mining zum Einsatz kommen, sondern in der Simulation und der Softwareentwicklung ihre Anwendung finden.

3.1 Verifikation und Validierung

Häufig werden die Begriffe Verifikation und Validierung falsch oder sogar als Synonyme verwendet. Dabei ist die Bedeutung der Begriffe zwar ähnlich, jedoch nicht gleich.

Fragen helfen dabei, die Begriffe zu definieren:

- Verifikation: Fertige ich das Produkt richtig?
- Validierung: Fertige ich das richtige Produkt?

Eine Verifikation ist im Allgemeinen ein Nachweis, dass eine Hypothese wahr ist. Im Zusammenhang mit der vorliegenden Arbeit, der Verifikation von Prozessen, überprüft die Verifikation, ob ein Prozess korrekt von einem Zustand in einen anderen Zustand transformiert wurde. Ähnlich definiert die VDI Richtlinie 3633 für die Simulation Verifikation. Die Verifikation ist ein „formaler Nachweis der Korrektheit des Simulationsmodells“ (VDI-Richtlinie 3633 Blatt 1 2008). Analog dazu untersucht die Verifikation, ob das erstellte Modell das ursprüngliche Modell korrekt wiedergibt. Entsprechend beschreibt Balzert (1999) Verifikation als Beweis von Konsistenz zwischen der Programmimplementierung und seiner Spezifikation. „Verifikation ist die Überprüfung, ob ein Modell von einer Beschreibungsart in eine andere Beschreibungsart korrekt transformiert wurde“ (Rabe et al. 2008).

Jedoch bedeutet eine positive durchgeführte Verifikation nicht, dass das Modell für den einzelnen Anwendungsfall korrekt ist. Eine erfolgreiche Verifikation bedeutet lediglich, dass der Prozess oder das Modell richtig transformiert worden ist. Deshalb ist eine Prüfung über das korrekt angewendete Modell notwendig.

Mit dieser Kontrolle beschäftigt sich die Validierung. Sie überprüft im Gegensatz zur Verifikation, ob das richtige Modell verwendet wird. Dementsprechend wird die Eignung des Modells für den speziellen Anwendungsfall geprüft. „[Die] Validierung ist die

kontinuierliche Überprüfung, ob die Modelle das Verhalten des abgebildeten Systems hinreichend genau wiedergeben“ (Rabe et al. 2008).

Zusammenfassend lässt sich festhalten, dass durch die Verifikation die korrekte Transformation vom ursprünglichen Modell zu einem anderen Modell gewährleistet wird. Durch eine Validierung wird hingegen sichergestellt, dass das verifizierte Modell das richtige Modell ist. Ein richtiges Modell ist in diesem Fall ein Modell, welches die betrachtete Ausgangslage hinreichend genau abbildet.

Ziel dieser Arbeit ist, das Automatisierungspotential von V&V-Techniken zu untersuchen. Deshalb wird im Folgenden der Begriff Automatisierung erläutert. Im Allgemeinen bedeutet ‚automatisch‘, dass etwas ausgeführt wird, ohne dass ein menschliches Eingreifen nötig ist. Führt eine Maschine ein Programm oder eine Funktion automatisch aus, kann der Benutzer davon ausgehen, dass nicht eingegriffen werden muss. In der Regel wird die Maschine den Programmcode ohne Fehler ausführen (Heinrich et al. 2015). Nach DIN „bezeichnet automatisch einen Prozess oder eine Einrichtung, der oder die unter festgelegten Bedingungen ohne menschliches Eingreifen abläuft oder arbeitet“ (DIN 2014).

Grundsätzlich können alle Aufgaben eines Rechenzentrums automatisiert werden. Automatisierung ist ein Kernbestand der IT, die versucht durch Algorithmen Aufgaben zu wiederholen und somit zu automatisieren.(Abolhassan und Kellermann 2016). Daher besteht theoretisch die Möglichkeit, alle Prozesse in einem Prozessor mit einem Prozess zu automatisieren, wenn die benötigten Eingabeinformationen vorhanden sind.

Ein Automatisierungsprozess verarbeitet die Informationen, welche durch Sensoren erfasst werden, und steuert oder regelt den Prozess so, dass die zu erfüllende Aufgabe erledigt wird (Hompel et al. 2008). Ein einfaches Beispiel eines automatisierten Prozesses ist der Tempomat an einem Auto. Erhält der Prozessor die Eingabe den Prozess zu starten, erfasst er die Geschwindigkeit und steuert die Geschwindigkeit des Autos automatisch, also ohne das Eingreifen des Fahrers.

Im Data Mining erscheint die Automatisierung jedoch schwierig. Speziell die Interpretation und Analyse der Ergebnisse scheint nicht automatisierbar zu sein. Nach Bissantz (2009) müssen die gefundenen Regeln und Muster interpretiert werden. „Diese Interpretation erfordert Fachwissen über den Bereich, dem die Daten entstammen.“ (Bissantz und Hagedorn 2009) Petersohn beschreibt Automatismus im Data Mining als

ein Güte Merkmal, stellt jedoch klar, dass völlig autonome Systeme zur selbstständigen Analyse von Datenbeständen zurzeit nicht umsetzbar sind. (Petersohn 2009)

Ziel dieser Arbeit ist eine Potentialanalyse in Bezug des Automatisierungspotentials speziell der V&V-Techniken im Data Mining zu erstellen. Runkler nennt in diesem Zusammenhang die Einbeziehung eines Experten zur Anwendungskontrolle für unabdingbar. Demnach entsteht ein rückgekoppelter Prozess, welcher wiederholt durchlaufen wird, bis zufriedenstellende Ergebnisse erreicht worden sind. (Runkler 2015) Im Folgenden werden verschiedene V&V-Techniken vorgestellt.

3.2 V&V-Techniken im Data Mining

Die Qualitätssicherung ist in Data-Mining-Vorgehensmodellen ein wichtiger Schritt. Dadurch werden die ermittelten Informationen validiert und verifiziert. Mit der Anwendung von V&V-Techniken kann das Wissen, welches aus den Datensätzen generiert wurde, als geprüft angenommen und verwendet werden. Ein Problem bei der Verifikation und Validierung im Data Mining ist zurzeit, dass im zeitintensivsten Schritt im Data-Mining-Prozess, der Datenvorverarbeitung, keine Qualitätssicherung stattfindet (Kurgan und MUSILEK 2006). Im Gegensatz dazu können im Hauptschritt, dem Data Mining, verschiedene Techniken angewendet werden.

Eine Art der V&V sind die zu berechnenden Werte (Konfidenz und Support) in der Assoziationsanalyse, da durch diese Berechnungen nicht relevante und nicht interessante Regeln bzw. Muster nicht weiter beachtet werden. Bei der Assoziationsanalyse sollen die Konfidenz und der Support berechnet werden, um ein gewisses Interessantheitsmaß der Ergebnisse zu gewährleisten. Zur Berechnung der Konfidenz und des Supports werden Algorithmen wie der AIS-Algorithmus, der auf dem Apriori-Algorithmus basiert, und der FP-Growth-Algorithmus angewendet (Han et al. 2004). Aktuell ist der wesentlich effizientere FP-Growth dabei ist, den Apriori-Algorithmus abzulösen (Buneman 1993). Zur Untersuchung eignen sich in erster Linie die statischen Techniken, welche auch in der Simulation (siehe Kapitel 3.4) eingesetzt werden. Dabei sind vor allem die Techniken Korrelationsanalyse, Regressionsanalyse, Chi-Quadrat-Test, Hypothesentests und der Signifikanztest zu nennen.

Bei den Stichprobentests werden vor allem Tests durchgeführt, welche auf der Methode des Resamplings basieren. Resampling ist eine Methode, bei der statistische Eigenschaften von Stichproben bestimmt werden. Charakteristisch für das Resampling

ist, dass die Eigenschaften durch wiederholtes Ziehen von Stichproben aus einer einzigen Ausgangsstichprobe bestimmt werden (Eckstein et al. 1994).

Dadurch wird gleichzeitig auch überprüft, ob Under- bzw. Overfitting vorliegt. Von Underfittung wird gesprochen, wenn beispielsweise bei einer Regressionsanalyse 60% der Trainingsdaten richtig klassifiziert worden sind, in den Testdaten jedoch deutlich mehr. Overfitting beschreibt das Gegenteil. Dementsprechend wird von Overfitting gesprochen, wenn 80% der Trainingsdaten und deutlich weniger in den Testdaten als richtig klassifiziert worden sind (Schendera 2014). Generalisierbarkeit nur anhand der Trainingsdaten ist nicht möglich, da diese bereits zur Erstellung des Modells verwendet wurden. Durch diese Daten kann nur die Korrektheit des Modells für die Trainingsdaten nachgewiesen werden. Für unbekannte Daten kann aber keine Richtigkeit gewährleistet werden (Steiner 2009). Zum Nachweis von Generalisierbarkeit existieren verschiedene Resampling-Methoden. Zu nennen sind hier vor allem die Kreuzvalidierung, die Bootstrapping-Validierung und die Holdoutmethode (Clarke et al. 2009), welche im Folgenden näher beschrieben werden. Neben diesen Techniken finden sich in der Literatur eine Reihe weiterer Techniken, welche im Data Mining zum Einsatz kommen. Da diese jedoch von der Kreuz-Validierung, Split-Validierung oder der Bootstrapping-Validierung abgeleitet sind oder auf diesen basieren, werden diese Techniken hier nicht näher beschrieben. Zu erwähnen sind die Batch-Cross-Validation, Wrapper Split Validation und die Cross-Prediction (Hofman 1990; Kohavi 1995).

Kreuzvalidierung: Das Kreuzvalidierungsverfahren (auch Cross-Validation oder X-Validation genannt) wird meistens bei kleiner Datenmenge mit weniger als 1000 Datensätzen verwendet (Steinlein 2004). Bei dieser Methode werden die Daten in zwei Gruppen eingeteilt. Dabei wird zwischen Test- und Trainingsdaten unterschieden. Während die Trainingsdaten dabei helfen das Modell zu erstellen, dienen die Testdaten dazu, das Modell zu überprüfen. Das Verfahren wird solange mit einer neuen Unterteilung wiederholt, bis jeder Datensatz zumindest einmal in den Testdaten verwendet wurde (Gottermeier 2003; Ross et al. 2009). Durch die Unterteilung der Daten in Test- und Trainingsdaten und das mehrfache Durchlaufen des Prozesses steigt die Durchlaufzeit dieser Validierungsmethode mit Erhöhung der Datensätze deutlich an (Kohavi 1995).

Holdoutmethode: Eine Voraussetzung zur Anwendung der Holdoutmethode (auch Split-Validierung genannt) ist, dass als Grundlage eine große Datenmenge zur Verfügung steht. Für die Einteilung in Trainings-, Validierungs-, und Testdaten wird eine Mindestanzahl

von 1000 Datensätzen verlangt (Hofman 1990). Liegt die Anzahl der Datensätze über dieser Marke, ist die Holdoutmethode den anderen Methoden überlegen. Bei einer geringeren Anzahl an Datensätzen ist zwar auch eine Durchführung dieser Technik möglich, jedoch nur, indem auf Testdaten verzichtet wird. Dadurch kann Overfitting an dieser Stelle entstehen (Bishop 1995). Zur Durchführung der Holdoutmethode werden die gegebenen Daten zunächst in Trainings-, Validierungs- sowie Testdaten aufgeteilt. Mit Hilfe der Trainingsdaten wird das Data-Mining Modell entwickelt. Die Validierungsdaten werden anschließend dazu verwendet, das Modell zu testen und weiterzuentwickeln. Eine Verbesserung kann etwa durch Anpassung der Parameter erfolgen. Schließlich dienen die Trainingsdaten dazu, die Güte des Modells zu überprüfen. Wurde die Güte des Modells nachgewiesen, ist auch die Generalisierbarkeit des Modells bewiesen. Dies liegt daran, dass bei der Holdoutmethode verschiedene unabhängige Daten zum Einsatz kommen (Steiner 2009). Durch die Verwendung der Test- und Trainingsdaten kann bei der Holdoutmethode von Generalisierbarkeit des Modells gesprochen werden, wenn zuvor die Güte des Modells nachgewiesen werden konnte (Steinlein 2004).

Eine Variante für die Verteilung der Daten ist 60% Trainingsdaten, 30% Validierungsdaten und 10% Testdaten (Berry und Linoff 2000). Ähnliche Werte können in der weiteren Literatur gefunden werden. Dort werden Werte von 10% bis 25% an Testdaten und 75% bis 90% sowohl Trainingsdaten als auch Validierungsdaten gesprochen, wobei dort im Verhältnis 4:1 für die Trainingsdaten gewichtet wird (Urban 1998). Ein weiterer Ansatz ist das Zerteilen der Ursprungsdaten in lediglich zwei Gruppen. Während die Trainingsdaten zum Lernen dienen, testen die Testdaten das Modell. Im optimalen Fall werden die Daten in einem Verhältnis von 2:1 aufgeteilt (Cios et al. 2007). Eine ähnliche Einteilung erfolgt im nächsten Ansatz. Dabei werden die Daten auch in zwei Gruppen gegliedert. Jedoch werden diese Validierungsdaten (statt Testdaten) bzw. Trainingsdaten genannt (Breiman 1998). In der Literatur finden sich noch weitere Verteilungsansätze der Daten (Malthouse und Blattberg 2005; Mellish 1995; Zahavi und Levin 1997).

Bootstrapping-Validierung: Die Bootstrapping-Validierung basiert grundsätzlich auf den beiden beschriebenen Techniken. Im Gegensatz zu der Kreuz- und Split-Validierung wird diese Validierung verwendet, wenn die theoretische Verteilung der interessanten Statistik nicht bekannt ist (Efron und Tibshirani 1998). Genauere Erklärung zur Bootstrapping-Validierung werden von Efron (1998) ausgeführt.

3.3 V&V-Techniken in der Softwareentwicklung

Vor einigen Jahrzehnten wurden Teile von Softwarecodes entweder gar nicht getestet oder nur manuell geprüft. Die Integration von Tests in die Entwicklung erfolgte durch die Entwickler, welche die Software an spezielle Tester verteilten. Test-Analysen und A / B – Tests waren Techniken, die meistens nur theoretisch bekannt waren, allerdings in der Praxis nie angewendet wurden (Spinellis 2017).

V&V-Techniken in der Softwareentwicklung können in drei Hauptverfahrensarten eingeteilt werden. Dabei wird zwischen statischen und dynamischen Methoden unterschieden. Des Weiteren kann in diversifizierende Testverfahren klassifiziert werden (Burgdorf 2010). Der Hauptunterschied zwischen den dynamischen und statischen Prüftechniken liegt in der Durchführung der Techniken. Während bei einer statischen Untersuchung lediglich der Quellcode der zu prüfenden Software betrachtet und untersucht wird, führt die dynamische Technik die Software aus und verfolgt den Programmablauf. Dabei ist eine Abweichung von der Soll-Ausprägung als Fehler zu betrachten (Balzert 2009; Vivenzio 2010). Die statischen Testverfahren dienen der Fehlerprävention. Ziel dieser Verfahren ist die möglichst frühe Erkennung und Behebung der Fehler und Abweichungen. Statische Testverfahren werden bereits während der Entwicklungsphase eingesetzt (Spillner und Linz 2012). Bei diversifizierenden Testverfahren werden zwei unterschiedliche Objekte miteinander verglichen. Abbildung 6 zeigt die Einteilung der verschiedenen Testverfahren nach Balzert (2009).



Abbildung 6: Klassifikation von Testverfahren nach Balzert (2009)

In der Literatur finden sich verschiedene dynamische Testverfahren, welche jedoch gemeinsame Merkmale aufweisen. Das Programm wird in der realen Umgebung getestet und dabei mit konkreten Eingabewerten versehen und ausgeführt. Mit Hilfe eines Debuggers ergibt sich die Möglichkeit, in den dynamischen Techniken die Software an verschiedenen Stellen zu pausieren, Variablen auszulesen und je nach Situation zu verändern um das Verhalten des Programms zu testen. Ein weiteres Merkmal der Tests

ist die Eigenschaft, dass alle Tests den Stichprobenverfahren zugeordnet werden können. Damit wird die Korrektheit des getesteten Programmes nicht bewiesen. Kann die Vollständigkeit des Datenflusses des Programms mit den Tests nachgewiesen werden, liegt ein dynamisches Strukturverfahren (auch White-Box-Testverfahren oder Glass-Box-Testverfahren) vor (Balzert 1999).

Kontrollflussorientierte: Das Ziel der Strukturtestverfahren ist die Überprüfung aller Funktionen des Programms. Bei den Strukturtestverfahren ist eine wichtige Gruppe die kontrollflussorientierten Testverfahren. Dabei wird ein Ablaufdiagramm der Software erstellt. Um dieses Testziel zu erreichen, werden Elemente wie Pfade, Anweisungen und Bedingungen verwendet. Weit verbreitete kontrollflussorientierte Testverfahren sind der Anweisungsüberdeckungstest, der Zweigüberdeckungstest und der Pfadüberdeckungstest. Diese drei Tests haben als Ziel, mit einer entsprechenden Anzahl von Testläufen sämtliche Anweisungen und Pfade zu durchlaufen (Balzert 1999). Wie alle strukturorientierten Testverfahren gibt auch das kontrollorientierte Testverfahren keine Regel für die Generierung von Testfällen an. Deshalb muss darauf geachtet werden, dass die Struktur entsprechend abgedeckt wird. Da bei diesem Test der Quellcode betrachtet wird, wird das kontrollorientierte Testverfahren den White-Box-Testverfahren zugeordnet. Im Gegensatz zu den Black-Box-Tests wird bei den White-Box-Testverfahren der Quellcode betrachtet und untersucht. Dazu sind grundsätzlich Kenntnisse über die Funktionsweise des Softwarecodes erforderlich.

Die funktionalen Testverfahren (Black-Box-Testverfahren) sind gekennzeichnet durch die Testfälle, welche durch die Spezifikation des Programms abgeleitet werden. Ein Unterschied zu den White-Box-Verfahren ist, dass keine Kenntnisse über den Softwarecode vorhanden sind, der dementsprechend auch nicht betrachtet wird. (Desharnais et al. 2011). Das zu testende Programm soll für den Tester ein ‚schwarzer Kasten‘ sein. Hieraus leitet sich der Name ‚Black-Box-Testverfahren‘ ab. Ein unabhängiges Team entwickelt dabei Testfälle, ohne den Code des Programms zu kennen (Wichmann und Cox 1992). Damit wird verhindert, dass die Programmierer die Tests um die eigenen Fehler herum entwickeln und zusätzliche Annahmen getroffen werden (APCoQS 2000).

Black-Box-Tests sind wesentlich aufwendiger als White-Box-Tests, bieten aber auch Vorteile gegenüber den White-Box-Tests. Sie verifizieren das Gesamtsystem genauer und testen bei geeigneten Spezifikationen bestimmte Eigenschaften besser (Sneed et al. 2009).

Zur Auswahl der Testfälle bei den Black-Box-Tests stehen dabei verschiedene Verfahren zur Verfügung.

Die Idee des Black-Box-Verfahrens basiert auf dem Grundsatz, dass ein Test des Programmes ausschließlich gegen sich selbst nicht ausreichend ist. Da die Aufgabe eines Programms in der Spezifikation beschrieben wird, sind Tests des Programms gegenüber seinen Spezifikationen dringend erforderlich (Fajardo und Dustin 2007). Um das Ziel der funktionalen Testverfahren zu erreichen, die spezifizierte Funktionalitäten des Programmes zu prüfen, ist eine Untersuchung aller Funktionen des Programms notwendig.

Allerdings ist ein vollständiger Funktionstest normalerweise nicht durchführbar. Die Ableitung geeigneter Testfälle ist die Hauptschwierigkeit beim Black-Box-Testverfahren. Da die Ableitung der Testfälle wegen der hohen Anzahl an Situationen unmöglich ist, müssen stellvertretend eine Anzahl an Testfällen ausgewählt werden. Die Testfälle sind so auszuwählen, dass die Wahrscheinlichkeit, einen Fehler zu finden, hoch ist. Weiterhin soll mit Hilfe der Testfälle alle nicht betrachteten Fälle abgedeckt werden (Liggesmeyer 2009). Zur Bestimmung von geeigneten Testfällen zählt Balzert drei wichtige Verfahren auf: (Balzert 1999)

- Funktionale Äquivalenzklassenbildung
- Grenzwertanalyse
- Test spezieller Werte

Funktionale Äquivalenzklassenbildung: Die funktionale Äquivalenzklassenbildung hat zum Ziel, mit möglichst wenigen Testfällen eine hohe Anzahl an Fehlern zu finden. Mit dem Prinzip ‚Teile und Herrsche‘ wird versucht die Komplexität der Testprobleme durch Unterteilen zu verringern. Am Ende der Zerlegung soll eine einfache Wahl von Testfällen möglich sein. Bei der Zerteilung werden zunächst die Eingabedaten, Ausgabedaten und die Bedingungen der verschiedenen Spezifikationen analysiert. Anschließend werden auf der Grundlage der Ein- und Ausgabedaten Äquivalenzklassen gebildet. Die Testfälle werden schließlich durch Auswahl in jeder Äquivalenzklasse bestimmt (Berghammer 2008). Das zu testende Programm bearbeitet alle Werte einer Äquivalenzklasse. Deshalb sind sämtliche Werte einer Äquivalenzklasse auch repräsentative Stellvertreter für die anderen Werte dieser Äquivalenzklasse. Das Verfahren heißt funktionale Äquivalenzklassenbildung, da die Äquivalenzklassen mit Hilfe der Spezifikationen identifiziert werden (Liggesmeyer 2009).

Grenzwertanalyse: Die Grenzwertanalyse ist eine wichtige und sinnvolle Ergänzung zu den Testfällen, welche durch die Äquivalenzklassen ermittelt worden sind. Grund dafür ist, dass die meisten Fehler in den Grenzbereichen der Klassen auftreten. Deshalb ist eine genaue Prüfung diese fehleranfälligen Bereiche elementar. In der Grenzwertanalyse werden die Grenzen der Äquivalenzklassen untersucht. Dazu werden der genaue Grenzwert und die anliegenden Werte, innerhalb und außerhalb der Klasse, geprüft. Anzuwenden ist die Grenzwertanalyse deshalb nur, wenn geordnete Klassen vorliegen. Folgerichtig entstehen drei Testfälle für jede Grenze, welche jedoch, falls zwei Grenzen identisch sind, zusammengefasst werden können. Da oftmals kein Grenzwert existiert, weil der Wert zur Klasse gehört, reicht eine Betrachtung von zwei Problemfällen in diesen Fällen aus (Spillner und Linz 2012; Ludewig und Lichter 2012).

Test spezieller Werte: Ziel der Tests von speziellen Werten ist fehleranfällige Bereiche zu testen und mit Hilfe von Testfällen zu überprüfen. Zur Auswahl der Testdaten sind die Ansätze ‚zero values-Kriterium‘ und das ‚distinct values-Kriterium‘ am bekanntesten. Beim ‚zero values-Kriterium‘ werden Tests durchgeführt, bei denen Variablen den Wert Null annehmen, denn der Wert Null als Eingabe- oder Ausgabewert ist oft ein Indiz für einen Fehler. Das ‚distinct values-Kriterium‘ fordert eine Zuweisung von verschiedenen Werten an Feldelementen und Eingabedaten, die in engem Zusammenhang stehen. (Balzert 1999)

Zusätzlich ist zu überprüfen, ob ein zustandsbasierter Test durchgeführt werden muss. Dies ist abhängig davon, ob das System zustandsbasiert oder zustandsfrei (auch: gedächtnisbehaftet oder gedächtnislos) ist. Eng gesehen ist jedes Software-System ein gedächtnisbehaftetes System, da ein System über einen Speicher verfügt. Weil jedoch das Gedächtnis nicht immer im Vordergrund der Software steht, wird von einem gedächtnislosen Programm ausgegangen. Da die Speichermöglichkeiten der Software für den Test nicht beachtet werden muss, ist ein zustandsbasierter Test in diesem Fall nicht nötig. Bestimmt die Speichermöglichkeit der Software das Programm maßgeblich, ist ein zusätzlicher Test durchzuführen, welcher den Speicher einbezieht. Eine Möglichkeit, den Speicher mit zu berücksichtigen, ist der zustandsbasierte Test.

Zusätzlich besteht bei dynamischen Verfahren die Möglichkeit die Laufzeiten zu messen und dadurch abschätzen zu können, wie lange der nächste Programmdurchlauf dauert. Dazu misst das Verfahren sowohl die Häufigkeit als auch die Dauer einzelner Code-Teile. Allerdings kann diese Laufzeiteinschätzung erst abgegeben werden, nachdem der Code ausgeführt wurde. Dies kann in manchen Anwendungsfällen, gerade in der

Echtzeitanalyse, zu Problemen führen. Das statische Werkzeug ‚Absint/aiT‘ ist in der Lage, die Dauer eines Programms sehr genau abschätzen zu können (Balzert 2009).

Die statischen Techniken betrachten im Gegensatz zu den dynamischen Testverfahren die Quellcodes der Software und untersuchen diese auf Fehler. Dabei wird das Programm nicht mit Testdaten versehen und nicht ausgeführt. Die Betrachtung des Codes kann und sollte bei allen für die Software relevanten Dokumenten vorgenommen werden. Bei diesen Methoden der Testverfahren ist das Ziel, Fehler und Abweichungen von den vorhandenen Spezifikationen zu ermitteln. Diese möglichst früh identifizierten Abweichungen vom Ideal dienen der Prävention und helfen dabei, aufwendige Verbesserungen am fertigen Code vornehmen zu müssen. Eine Betrachtung kann unter anderem von Personen durchgeführt werden. Als Ergänzung oder Alternative zur Betrachtung durch Personen können unterschiedlich Werkzeuge verwendet werden. Als Werkzeug ist das ‚Program-Slicinig‘ zu nennen. Dabei wird der Code in verschiedene Scheiben unterteilt. Die gängigsten statischen Tests sind Prüfmethode wie Inspektionen, Reviews und Walkthroughs (Balzert 1999; Spillner und Linz 2012).

Bei einem Review wird jeder für den Menschen lesbare Teil der Software in Betracht gezogen. Ein lesbarer Teil kann bei einer Software ein Dokument, ein Modul oder eine Anleitung sein. Um eine Prüfung durchführen zu können, sind Referenzunterlagen notwendig. Diese setzen sich in erster Linie aus Vorgaben und Richtlinien zusammen (Ludewig und Lichter 2012). Auf deren Basis können Listen oder Fragekataloge zur Überprüfung ausgearbeitet werden. Weitere statische Testverfahren werden im nächsten Kapitel im Zuge der Testverfahren in der Simulation ausführlich vorgestellt.

Diversifizierende Tests sind dadurch charakterisiert, dass diese Tests Ergebnisse von verschiedenen Tests miteinander vergleichen. Im Gegensatz zu den bisher vorgestellten Testverfahren, die die Testergebnisse vergleichen, nehmen diese bei diversifizierenden Tests eine geringere Rolle ein. Die am häufigsten eingesetzten diversifizierenden Testverfahren sind der Back-to-Back-Test, der Regressionstest und der Mutationstest. Im Zuge des Back-to-Back-Tests entwerfen unabhängig arbeitende Entwicklerteams ein Objekt mit den gegebenen Randbedingungen. Anschließend werden die Ergebnisse der unabhängig voneinander arbeitenden Teams miteinander verglichen. Zu beachten ist, dass durch die geforderte unabhängige Entwicklung desselben Objektes ohne Kommunikation der Teams hohe Kosten entstehen. Deshalb wird der Back-to-Back-Test in der Praxis in der Regel nicht verwendet, da sich dieser nicht rentiert (Burgdorf 2010).

Mutationstest: Im Gegensatz dazu überprüft der Mutationstest die Tauglichkeit eines bestimmten Testverfahrens für den speziellen Fall. Dazu werden in ein korrektes Objekt absichtlich Fehler eingefügt, die Mutationen genannt werden. Anschließend wird überprüft, ob das zu testende Testverfahren die eingefügten Fehler erkennt (Liggesmeyer 2009). Die Theorie hierzu lautet, dass wenn eine Mutation eingeführt wird, ohne dass sich die Ausgabe des Programms ändert, zeigt, dass der Code, der mutiert worden ist, entweder niemals ausgeführt wurde oder dass der Fehler nicht lokalisiert wurde (Ammann und Offutt 2008). Werden die Mutationen nicht als Fehler erkannt, ist das getestete Verfahren für den vorliegenden Fall ungeeignet.

Durchgeführt wird der Mutationstest, indem eine Reihe von Mutationsoperatoren ausgewählt und nacheinander auf jedes Teilstück des Quellcodes angewendet werden. Mutant wird dabei das Ergebnis der Anwendung eines Mutationsoperators auf das Programm genannt. Ist die Testsuite in der Lage, die Änderungen zu erkennen (ein Test schlägt fehl), ist der Mutant getötet (Cordy 2008). Um einen Mutant zu töten, müssen drei Bedingungen erfüllt sein (Offutt 2011):

- Der Test muss die mutierte Aussage im Quellcode erreichen.
- Die Test-Eingabedaten müssen den Status des Programmes infizieren, indem sie verschiedene für die Mutante und das ursprüngliche Programm verursachen.
- Der inkorrekte Programmzustand muss sich auf die Ausgabe des Programms ausbreiten und vom Test überprüft werden.

Wurden alle drei Bedingungen erfüllt und damit der Mutant eliminiert, wird von einem erfolgreichen Test gesprochen.

Regressionstest: Während des Regressionstestes wird derselbe Testfall mehrfach verwendet (Burgdorf 2010). Dabei wird die Ausgabe eines Testfalles der aktuellen Version des Programms mit der Ausgabe desselben Testfalles im Vorgängermodell des Programms verglichen. Beispielsweise durchläuft der identische Testfall das Programm sowohl in Version 1.045 und 1.046. Sind die beiden Ausgaben der Versionen des Programms identisch, war der Regressionstest erfolgreich (Myers et al. 2004).

3.4 V&V-Techniken in der Simulation

Simulation bedeutet nach der VDI-Richtlinie 3633 das Überführen eines dynamischen Prozesses eines Systems in ein experimentierfähiges Modell (VDI-Richtlinie 3633 Blatt 1 2008). Gründe für die Überführung von Problemen aus der Realität in die Form eines

Modells sind vor allem Kosten- und Sicherheitsgründe. Zunächst einmal können kostengünstigere und sichere Experimente am Modell durchgeführt werden. Die Ergebnisse können anschließend auf die Realität übertragen werden. Jedoch ist das Modell nach der Erstellung mit Hilfe von V&V-Techniken hinsichtlich Vollständigkeit und Korrektheit zu prüfen. Allerdings ist eine vollständige Richtigkeit des Modells in der Simulation mit V&V-Techniken nicht zu gewährleisten. Da die V&V-Techniken als wesentliches Hilfsmittel zur Erkennung von Fehlern beitragen, ist das Ziel der V&V-Techniken in der Simulation das Verhindern falscher Aussagen, da diese zu falschen Folgerungen führen können (Rabe et al. 2008). Durch die Implementierung einer V&V-Technik wird die geplante Simulation zur Überprüfung durchgeführt. Die V&V-Technik führt die Simulation durch, überwacht und analysiert diese hinsichtlich möglicher Fehler. Dies hilft dabei die Simulation zu verbessern und letztendlich Kosten zu sparen. Die V&V-Technik unterstützen dabei zu gewährleisten, dass das Ziel der Simulation erfüllt wird.

Die V&V-Techniken in der Simulation werden in vier Gruppen unterteilt (Balci 2007). Dabei wird, wie in Abbildung 7 aufgezeigt, zwischen den dynamischen, statischen, formalen und informalen Techniken unterschieden. Die vier Gruppen unterscheiden sich in erster Linie in der Komplexität, wobei die formalen Techniken die höchste Komplexität aufweisen. Formale Techniken sind deshalb komplexer als die übrigen Techniken, da diese Techniken aus großen Anteilen an logischen und vor allem mathematischen Formalismen bestehen (Balci 2007). Allerdings wird gleichzeitig mit der gesteigerten Komplexität auch die Objektivität der Techniken erhöht.

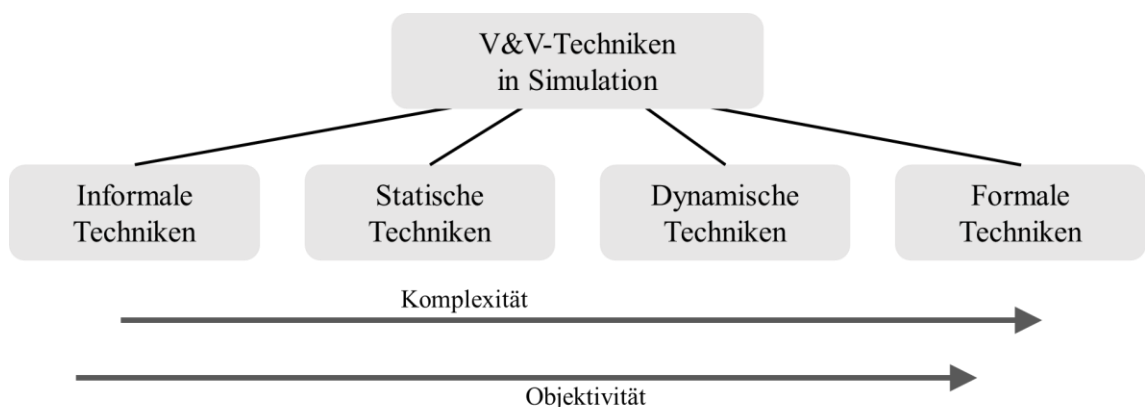


Abbildung 7: Gruppen der V&V-Techniken in der Simulation (Balci 2007)

Die **informalen V&V-Techniken** sind die meistgenutzten Techniken in der Simulation, da aus diesen Techniken formale und strukturierte Abläufe hervorgehen können. Da bei den informalen V&V-Techniken keinerlei formale mathematische Operationen

durchgeführt werden, sind die informalen Techniken durch ihre Subjektivität und das menschliche Denken charakterisiert (Balci 2007). Den informalen Techniken werden

- Audit,
- strukturiertes Durchgehen,
- Review,
- Dokumentenüberprüfung,
- Schreibtischtest,
- Inspektion,
- Turing-Test und
- Validierung im Dialog

zugeordnet. Diese acht Techniken werden im Folgenden erläutert. Die Untersuchungen werden durch Teams von mindestens zwei Personen durchgeführt. Die Hauptunterschiede zwischen diesen Techniken sind die Zusammensetzung der Teams, der Zeitpunkt der Überprüfung und die Verwertung der Ergebnisse.

Bei all diesen V&V-Techniken kommt der Dokumentation eine wichtige Rolle zu. Für eine qualitativ hochwertige Dokumentation sollten neben den Aufzeichnungen über Grundlagen, Randbedingungen und Störfaktoren auch die Qualitätskontrollen an sich dokumentiert werden. Durch eine vollständige Dokumentation des Kontrollablaufes in jeder Phase wird eine Wiederholung der Kontrolle wesentlich vereinfacht.

Audit: Ein Audit kann sowohl in der Softwareentwicklung als auch in der Simulation eingesetzt werden. Mit einem Audit wird geprüft, ob die Prozesse und Anforderungen den Abläufen und Richtlinien entsprechen. Tritt ein Fehler auf, sollte dieser mit Hilfe des Auditpfades bis zum Ursprung des Fehlers zurückverfolgt werden können. Audits helfen dem Management dabei, die Mitarbeiter zu kontrollieren. Die Prüfergruppe besteht aus externen Mitarbeitern, welche nicht im direkten Kontakt mit dem Mitarbeiter und dem Projekt stehen. Sie werden meist in periodischen Abständen durchgeführt oder alternativ bei Erreichung eines Meilensteins des Projekts (Balci 2007).

Strukturiertes Durchgehen: Der Ursprung dieser Technik liegt in der Softwareentwicklung (Balzert 2009). Beim strukturierten Durchgehen treffen sich alle Mitarbeiter des Projektes, gehen jede Programmanweisung durch und ändern diese, bis jeder Mitarbeiter von der Korrektheit überzeugt ist (Rabe et al. 2008). Das strukturierte Durchgehen lässt sich durch Veränderungen auch in der Überprüfung von Phasenergebnissen der Simulation einsetzen (Law 2007). Anders als bei der

Softwareentwicklung überprüfen im strukturierten Durchgehen bei der Simulation Mitarbeiter die Ergebnisse, welche wenig oder sogar gar nicht an dem Projekt mitgearbeitet haben. Damit wird ein neuer Blick auf die Ergebnisse geworfen wird (Law 2007). Das Ziel ist jedoch äquivalent zur Softwareentwicklung. Über jedes zu prüfende Dokument wird Einigung erzielt (Rabe et al. 2008).

Review (Begutachtung): Die Begutachtung unterscheidet sich nur geringfügig von der informalen V&V-Technik strukturiertes Durchgehen. Der Unterschied dieser beiden Techniken liegt bei den Teilnehmern der Techniken. Das Review schließt das Management der Auftraggeber ein und klärt dadurch, ob die Randbedingungen und Ziele korrekt verfolgt werden (Balci 2007). Deshalb liegt der Fokus des Reviews auf der Qualität der Projektdurchführung und der Projektergebnisse (Rabe et al. 2008). Die Technik sieht jedoch vor, dass sich die betroffenen Mitarbeiter einzeln und als Gruppe auf das Review vorbereiten. Der einzelne Mitarbeiter verwendet dazu den Schreibtischtest. Des Weiteren bereitet sich das gesamte Team gemeinsam beispielsweise unter Verwendung des strukturierten Durchgehens vor (Rabe et al. 2008). Eine solche Begutachtung kann zu jedem Zeitpunkt in einem Projekt durchgeführt werden.

Dokumentenüberprüfung: Durch die Überprüfung der verwendeten Dokumente wird Konsistenz, Korrektheit, Eindeutigkeit und Vollständigkeit der Dokumente sichergestellt. Damit besteht keine Möglichkeit, dass durch Veränderungen am Modell das falsche Modell genutzt wird oder andere verwendete Dokumente veraltet sind (Balci 2007).

Schreibtischtest: Beim Schreibtischtest, auch Selbstüberprüfung genannt, liegt der Fokus in der Überprüfung der eigenen Arbeit. Die selbst erbrachte Arbeit soll in Bezug auf Korrektheit, Konsistenz und Eindeutigkeit geprüft werden. Häufig wird der Schreibtischtest in einem frühen Stadium der Entwicklung durchgeführt (Balci 2007). Ein Problem des Schreibtischtests ist, dass eigene Fehler häufig auch beim wiederholten Durchgehen nicht identifiziert werden. Um dem entgegen zu wirken, kann eine weitere Person hinzugezogen werden, um die eigenen Fehler identifizieren zu können, womit diese Technik der Technik Validierung im Dialog gleicht. Deshalb ist zu empfehlen, den Schreibtischtest in Kombination mit einer anderen Technik zu verwenden und einen weiteren Mitarbeiter hinzuzuziehen, welchem die eigene Arbeit vorgestellt und erklärt wird (Rabe et al. 2008).

Inspektion: Die V&V-Technik Inspektion stammt aus der Softwareentwicklung. Jedoch kann diese auch in der Simulation angewendet werden (Balci 2007). Eine Inspektion wird von einem Team von vier bis sechs Mitgliedern für jede Modellentwicklungsphase, wie

etwa der Zieldefinition, durchgeführt und findet nach dem Erreichen eines Meilensteines statt. Eine Inspektion besteht aus fünf Phasen: Überblick, Vorbereitung, Inspektion, Nacharbeit und Nachverfolgung. Im Zuge der Inspektion wird mit Hilfe einer Checkliste nach Fehlern in Teilen der Objekte gesucht. Am Ende einer Inspektion steht ein Protokoll mit aufgedeckten Fehlern und enthält zusätzlich Verbesserungs- und Lösungsvorschlägen. Im Idealfall wird die nächste Entwicklungsphase nach der Inspektion freigegeben.

Diese Technik ist der Technik des strukturierten Durchgehens ähnlich. Eine Inspektion ist ein fünfstufiger Prozess, während das strukturierte Durchgehen nur aus zwei Phasen besteht. Das Inspektionsteam verwendet den Checklistenansatz zum Aufdecken von Fehlern. Dabei ist das in jeder Phase zu verwendende Verfahren formalisiert. Der Prozess der Inspektion dauert, verglichen mit dem strukturierten Durchgehen, länger. Jedoch ist die zusätzliche Zeit gerechtfertigt, da die Analyse mit der Inspektion eine leistungsfähige und kostengünstige Möglichkeit darstellt, Fehler früh in der Modellentwicklung zu erkennen (Balci 2007).

Turing-Test: Beim Turing-Test versuchen Experten das reale System vom Modell zu unterscheiden. Dazu werden den Experten, die das reale System kennen, zwei Ergebnisse vorgelegt. Diese Ergebnisse stammen einerseits vom realen System und die anderen Ergebnisse von einem Modell (Balci 2007). Beide Systeme wurden mit denselben Eingangsvariablen unter denselben Bedingungen durchlaufen. Sind die Experten nicht in der Lage die Ergebnisse dem ursprünglichen System zuzuordnen, wurde das reale System offenkundig ausreichend mit dem Modell abgebildet. Kann der Experte allerdings die meisten Ergebnisse richtig zuzuordnen, muss das Modell mit dem Ziel verbessert werden, dass das reale System besser darstellt wird (Landry und Oral 1993).

Validierung im Dialog: In der Validierung im Dialog werden unter Mithilfe von Experten und späteren Anwendern die Ergebnisse des Modells diskutiert. Dabei können Ergebnisse des ganzen Systems oder Ergebnisse einzelner Phasen Grundlage der Diskussion sein. Bei den Diskussionen steht vor allem die Nachvollziehbarkeit der Ergebnisse im Mittelpunkt (Balci 2007; Sargent 2013). Obwohl durch einen Dialog Fehler schnell entdeckt werden können, sofern der herangezogene Experte einen Fehler im Modell erkennt, besitzt diese Technik einige Schwächen. So könnte der Fachexperte nicht ausreichend Wissen über das reale System besitzen und somit die Ergebnisse des Modells nur unzureichend bewerten. Weiterhin könnte der Experte persönlich nicht

daran interessiert sein, Fehler aufzudecken und das Modell zu verbessern (Hermann 1967).

Statische V&V-Techniken in der Simulation befassen sich mit der Bewertung der Genauigkeit auf der Grundlage von Eigenschaften des Modells und des Quellcodes (MacNair 1989). Zur Durchführung dieser Techniken ist eine maschinelle Ausführung des Modells nicht erforderlich. Statische V&V-Techniken werden häufig mit Hilfe automatisierter Tools angewendet. Durch die Verwendung statischer Techniken besteht die Möglichkeit, viele Informationen über die Struktur des Modells, die Modellierungstechniken, den Kontrollfluss innerhalb des Modells und die syntaktische Genauigkeit zu erhalten (Balci 2007).

Die im Folgenden betrachteten **dynamischen V&V-Techniken** in der Simulation basieren auf entwickelten Modellen, welche bei der Durchführung untersucht werden (Balci 2007). Zu den dynamischen V&V-Techniken zählen

- Akzeptanztest,
- Alpha-Test,
- Beta-Test und
- statistische Techniken

Die Unterschiede bei den Tests liegen in den Zeitpunkten der Durchführung. Der Beta-Test testet die noch nicht fertige und noch in der Entwicklung steckende Beta-Version. Im Gegensatz dazu wird der Alpha-Test an der fertigen Version und mit unabhängigen Personen als Prüfer durchgeführt. Der Akzeptanztest wird letztendlich bei der Übergabe an den Kunden durchgeführt. Hierbei werden vom Kunden die im Vertrag festgelegten Anforderungen und Funktionalitäten getestet.

Akzeptanztest: Beim Akzeptanztest wird das Modell operativ, hinsichtlich der im Vertrag festgelegten Anforderungen, getestet. Dabei werden die tatsächliche Hardware und die tatsächlichen Daten verwendet (Perry 2006). Den Test führt entweder der Kunde selbst, der Entwickler in Anwesenheit des Kunden oder ein unabhängiger Auftragnehmer durch. Der unabhängige Auftragnehmer wird nach der offiziellen Lieferung des Modells und vor der offiziellen Annahme vom Kunden angestellt (Schach 1999).

Alpha-Test: Der Alphatest bezieht sich auf die Alpha-Version des vollständigen Modells an einem Inhouse-Standort, der nicht an der Modellentwicklung beteiligt ist (Beizer 1990).

Beta-Test: Dieser Test testet die Beta-Version des vollständigen Modells. Dabei wird das Modell als "Beta" -Benutzerumgebung unter realistischen Feldbedingungen untersucht (Balci 2007). In Tabelle 2 werden die Unterschiede zwischen Alpha- und Beta-Test abgebildet (Galindo 2018).

Tabelle 2: Gegenüberstellung von Alpha-Test und Beta-Test

Alpha-Test	Beta-Test
Stellen die Qualität des Produkts vor dem Übergang zum Beta-Test sicher	Überprüft auch die Qualität des Produktes, stellt aber auch die Benutzbarkeit für die Kunden sicher
Wird von internen Mitarbeitern durchgeführt	Wird von Externen durchgeführt. Oftmals vom Kunden selbst oder von Endnutzern
Wird von den Entwicklern durchgeführt	Wird von und beim Kunden oder Endkunden durchgeführt
Zuverlässigkeits- und Sicherheitstests werden nicht durchgeführt	Zuverlässigkeit und Sicherheit wird getestet
Kritische Probleme werden sofort behoben	Der Großteil der Rückmeldungen und Erkenntnisse werden in zukünftige Versionen des Produkts implementiert
Black-Box-Techniken und White-Box-Techniken werden verwendet	Es werden nur Black-Box-Techniken verwendet

Die **statistischen Techniken** bewerten die Ergebnisse eines Modells. Dies geschieht auf der Grundlage, mit welcher Sicherheit das reale System mit dem Modell abgebildet wird (Rabe et al. 2008). Außerdem können statistische Techniken dazu verwendet werden, die Datengrundlage zu überprüfen. Dazu wird beispielsweise die Korrelation zwischen den Attributen bei einem Datensatz untersucht. Eine weitere Untersuchung, die mit statistischen Techniken durchgeführt wird, ist die Überprüfung, ob Datensätze aus verschiedenen Quellen fehlerfrei zusammengefügt wurden. Auf der Grundlage der Ergebnisse dieser Überprüfung wird entschieden, ob die Daten gemeinsam oder separat voneinander betrachtet und analysiert werden müssen (Law 2007). Statistische Techniken können auch zur Prüfung von Ergebnissen der Modellierung in Verbindung mit Aufzeichnungen des realen Systems als Datengrundlage eingesetzt werden (Rabe et al.

2008). In der Literatur finden sich eine Vielzahl an häufig verwendeten statistischen Techniken. Im Folgenden werden

- die Korrelationsanalyse,
- die Regressionsanalyse
- und der Chi-Quadrat-Test

als statistische Techniken beschrieben.

Korrelationsanalyse: Die Korrelationsanalyse analysiert den linearen Zusammenhang bei zwei Merkmalen. Korrelierend heißt dabei, dass die beiden getesteten Merkmale voneinander abhängig sind. Allerdings ist nur ein linearer Zusammenhang mittels dieser Analyse ermittelbar. Ein weiterer Schwachpunkt dieser Analyse ist, dass die Ursache für die Abhängigkeit der Merkmale durch die Korrelationsanalyse nicht identifiziert werden kann.

Regressionsanalyse: Zur Identifizierung der Ursache kann die Regressionsanalyse angewandt werden, welche Ursachen und Wirkung des festgestellten Zusammenhangs zwischen zwei Objekten ableiten kann (Bosch 2015).

Chi-Quadrat-Test: Um auch nichtlineare Zusammenhänge zwischen zwei Merkmalen feststellen zu können, wird der Chi-Quadrat-Test verwendet. Da nichtlineare Zusammenhänge in der Korrelationsanalyse irrtümlicherweise als schwache Abhängigkeit identifiziert werden können, prüft der Chi-Quadrat-Test die Merkmale bezüglich ihrer stochastischen Unabhängigkeit (Runkler 2015).

Test von Teilmodellen: Um den Test von Teilmodellen durchführen zu können, muss die Möglichkeit bestehen, das zu testende Modell in Teilmodelle in Form einer Top-Down-Teilung zu zerlegen. Dies ist jedoch nur bei hierarchisch aufgebauten Modellen möglich. Während das gesamte Modell ausgeführt wird, werden Daten über alle Eingangs- und Ausgangsvariablen eines jeden Untermodells gesammelt. Diese Daten werden mit den Daten über die Variablen eines jeden Teilmodells verglichen (Balci 2007). Jedoch muss diese Technik für Modelle, welche nicht ausführbar sind, mit anderen Techniken verknüpft werden, weil diese Technik andernfalls keine Ergebnisse liefert (Rabe et al. 2008).

Validierung von Vorhersagen: Grundvoraussetzung zur Verwendung dieses Tests ist die Existenz eines realen Systems zu dem Modell. Das zu testende Modell wird dazu verwendet, eine Vorhersage zu treffen, welche anschließend am realen System überprüft wird. Überprüft wird die Vorhersage durch Messung am realen System oder durch einen

Vergleich mit Daten aus den IT-Systemen des Unternehmens (Sargent 2013). Dadurch dass die Vorhersage immer vor der Beobachtung des realen Systems stattfindet, wird sowohl bewusste als auch unbewusste Manipulation ausgeschlossen (Rabe et al. 2008).

Vergleich mit anderen Modellen: Dieser Vergleichstest kann verwendet werden, wenn mehr als eine Version eines Modells gegeben ist, die dasselbe reale System darstellt (Pressman und Maxim 2015; Sommerville 2016). Dabei werden die Simulationsmodelle mit denselben Eingangsdaten getestet und anschließend die Ausgangsdaten verglichen. Dabei zeigen unterschiedliche Ausgangsdaten Probleme bei der Genauigkeit des Modells an (Balci 2007).

Vergleich mit aufgezeichneten Daten: Für diese Technik müssen genügend früher aufgezeichnete Daten aus dem realen System zur Verfügung stehen. Zunächst werden die vorhandenen Daten in zwei Kategorien unterteilt (Sargent 2013). Die erste Gruppe der Daten dient zur Justierung des Modells, während der zweite Teil der aufgezeichneten Daten zu seiner Validierung dient. Zunächst wird das bestehende Modell mit Hilfe der ersten Gruppe der Daten genauer auf das reale System angepasst. Dazu werden Eingabedaten in das Modell eingegeben und anschließend die Ausgabedaten mit den aufgezeichneten Daten der ersten Gruppe verglichen. Anschließend wird mit der zweiten Gruppe der Daten die Validierung vorgenommen. Je genauer und häufiger die Ausgangsdaten des Modells mit den aufgezeichneten Daten übereinstimmen, umso besser ist das betrachtete Modell (Rabe et al. 2008). Dabei ist zu beachten, dass diese Daten nicht als Validierungsdaten verwendet werden dürfen, wenn Daten Änderungen am Modell zur Folge haben (Sargent 2013). Eine Schwierigkeit bei der Verwendung dieser Technik besteht dabei, dass die verwendeten Daten sowohl zur Justierung als auch zur Validierung aus der gleichen Quelle stammen und somit systematische Fehler nicht ausgeschlossen werden können (Rabe et al. 2008).

Formale V&V-Techniken basieren auf mathematischen Beweisen, die die Korrektheit darlegen. Ist ein solcher mathematischer Korrektheitsbeweis für ein Modell möglich, ist dies das effektivste Mittel zur V&V (Balci 2007). Jedoch besteht in Ausnahmefällen die Möglichkeit lediglich ein Modell mit den formalen V&V-Techniken zu überprüfen (Rabe et al. 2008).

4 Analyse der V&V-Techniken

In diesem Kapitel wird zunächst untersucht, welche Eigenschaften eine V&V-Technik erfüllen muss, damit diese automatisierbar ist. Im Zuge dessen werden auch die Unterschiede einer automatischen und einer manuellen V&V-Technik dargestellt. Auf Grundlage dieser Untersuchungen werden Kriterien abgeleitet, die eine V&V-Technik erfüllen muss, damit diese das Potential für eine Automatisierung aufweist. Anschließend werden diese Kriterien in einer Kriterien-Matrix zusammengeführt. Auf Basis dieser Matrix wird schließlich eine Analyse aller in Kapitel 3 vorgestellten V&V-Techniken hinsichtlich ihres Automatisierungspotentials durchgeführt. Abschließend werden die Ergebnisse der bewerteten Techniken in einer Matrix dargestellt.

4.1 Untersuchung der V&V-Techniken

Diese Arbeit hat zum Ziel, die Automatisierungspotentiale von V&V-Techniken im Data Mining zu untersuchen und aufzuzeigen. Dazu werden zunächst verschiedene Aspekte aufgedeckt, die für eine Automatisierung notwendig sind.

Gründe für die Automatisierung der V&V-Technik sind in erster Linie wirtschaftlicher Natur (Schnieder 1999). Denn kann eine V&V automatisch ablaufen, ist dies besonders für komplexe Modelle lohnend, da Zeit eingespart werden kann. Außerdem kann die Untersuchung durch eine automatisierte V&V an einem komplexen Modell die Personalkosten gesenkt werden. In der Praxis treten trotz der Mithilfe eines Fachexperten Schwierigkeiten bei der manuellen Untersuchung der Modelle auf. Denn häufig ist selbst der Fachexperte nicht in der Lage, die komplizierten Zusammenhänge und logischen Beziehungen in einem komplexen Modell zu verstehen. Des Weiteren ist eine Automatisierung in diesem Fall notwendig, um das gesamte Modell verstehen und anschließend untersuchen zu können. Durch eine mögliche Automatisierung werden die Mitarbeiter nicht nur physisch, sondern auch psychisch entlastet.

Ein weiterer Aspekt ist die steigende Zuverlässigkeit mit einem steigenden Grad an Automatisierung. Durch die Erhöhung der Automatisierung sinkt gleichzeitig die Anzahl auftretender Fehler im Vergleich zu einer nicht automatisierten Durchführung. Dadurch hält die Qualität der V&V im Data Mining konstant ein Niveau (Feja et al. 2013).

Allerdings ist in manchen Praxisfällen die Verwendung einer manuellen V&V alternativlos. Wird von der V&V ein hoher Detaillierungsgrad erwartet, welcher nicht von der automatischen V&V erzielt werden kann, so ist auf diese zu verzichten (Feja et

al. 2013). Ein anderer Grund können neue Entwicklungen und eingearbeitete Ideen in das Modell sein, die dazu führen, dass die manuelle V&V an dieser Stelle besser geeignet ist. Zusammenfassend lässt sich, wie in Tabelle 3 aufgeführt, festhalten, dass automatische V&V durch ihre geringe Fehleranfälligkeit, Zeit- und Kostenersparnis gekennzeichnet sind. Da die automatischen V&V nach einem festgelegten Plan arbeiten, ergibt sich eine geringere Fehleranfälligkeit und gleichzeitig die Möglichkeit der Wiederholung derselben V&V, wodurch verschiedene Modelle auf derselben Grundlage verglichen werden können. Weitere Vorteile der automatischen V&V sind die eindeutige Kommunikationsgrundlage und die bessere Maschinenlesbarkeit der Prozesse (Feja et al. 2013). Zu beachten ist jedoch, dass in bestimmten Fällen die manuelle V&V besser geeignet ist als die automatische.

Tabelle 3: Gegenüberstellung von automatische und manuelle V&V-Techniken

Automatische V&V	Manuelle V&V
Kosten- und Zeitersparnis in der Durchführung	Fehleranfälliger
Wiederholbarkeit und Vergleichbarkeit	Bei großen Modellen nicht praktikabel
Eindeutige Kommunikationsgrundlage	Bei bestimmten Fällen alternativlos
Modelle müssen einen gewissen Grad an Formalisierung aufweisen	Potentiale in der Modellierung bleiben eventuell ungenutzt

Eine bereits automatisch ablaufende V&V-Technik muss nicht grundsätzlich modifiziert werden. Jedoch sollte untersucht werden, in welchen Phasen des Data Mining die Technik zum Einsatz kommen kann. Dabei kann die Möglichkeit bestehen, das Anwendungsgebiet in weitere Phasen auszuweiten. Des Weiteren müssen in den Phasen, in denen diese Technik nicht zum Einsatz kommen kann, andere V&V-Techniken eingesetzt werden, damit eine vollständige automatische V&V gewährleistet werden kann.

Um eine manuelle V&V-Technik zu automatisieren, sind verschiedene Aspekte zu beachten. Grundsätzlich sollte die Technik das Potential besitzen durch Modifikationen automatisierbar zu werden. Grundsätzlich stellen dabei menschliche Akteure zunächst ein

Hindernis dar, weil diese ersetzt werden müssen. Allerdings besteht dazu nicht immer die Möglichkeit. Dadurch können einige Techniken nicht für einen Automatisierungsprozess in Betracht gezogen werden. Des Weiteren geht durch einen eventuellen Ersatz des Mitarbeiters der Aspekt des Fachwissens verloren. Techniken, die durch den Einsatz eines Experten charakterisiert sind, benötigen das Fachwissen dieses Mitarbeiters, um die gewünschte Verifikation und Validierung vorzunehmen. Dies führt unter Umständen dazu, dass die Technik den Sachverhalt nicht hinreichend genau verstehen und analysieren kann.

Um die V&V-Techniken optimal bewerten zu können, müssen auch die Ergebnisse der Techniken bewertet werden. Dabei werden nicht die Ergebnisse von zwei Techniken miteinander verglichen, sondern jede Technik ist getrennt voneinander zu betrachten. Untersucht werden verschiedene Ergebnisse einer Technik sowohl von einem Datensatz als auch von verschiedenen. Gerade bei der Betrachtung von lediglich einem Prozess ist zu prüfen, ob die Technik nach jeder Durchführung zum selben Ergebnis kommt. Wird beispielsweise das Review überprüft, ist das Ziel dieses Kriteriums, dass die Anwender bei einer jeden Prüfung des Prozesses zumindest ein ähnliches Ergebnis erhalten.

Die Objektivität der Technik nimmt unter anderem Einfluss auf das Ergebnis. Objektivität beschreibt die Unabhängigkeit einer Aussage von beeinflussten Bewertungen und Einschätzungen (Becker und Hübner 1976). Denn je größer die Objektivität einer Technik ist, desto sicherer liefert diese bei mehrmaligem Durchlaufen dieselben Ergebnisse. Ein offensichtlicher Schwachpunkt, der zu einer geringen Objektivität führt, ist hierbei das Eingreifen von Mitarbeitern und Experten in einer V&V-Technik. Denn durch das Mitwirken von menschlichen Akteuren werden Meinungen vertreten und fließen in die Bewertung mit ein. Bei der Bewertung mittels Review wird das betrachtete Objekt von Mitarbeitern eingeschätzt. Hierbei liegt zwangsweise eine geringe Objektivität vor. Denn auch wenn sich die Mitarbeiter bemühen, rein objektiv zu bewerten, bleibt stets ein Stück Befangenheit. Allerdings kann durch Vergrößerung des Bewertungsteams die Objektivität in einem gewissen Maß erhöht werden, da durch die gleichzeitig erweiterte Anzahl an Meinungen die verschiedenen Ansichten genauer und kritischer diskutiert werden. Dennoch ergeben sich auch bei Techniken gänzlich ohne menschlichen Akteur Unterschiede. So erklärt Balci (2007) für V&V-Techniken in der Simulation, dass die Objektivität bei formalen Techniken am größten ist. Danach reihen sich absteigend die dynamischen, statischen und informalen Techniken (Balci 2007).

Ein weiterer Aspekt, der die Qualität der Ergebnisse beeinflusst, ist die Zuverlässigkeit. Zuverlässigkeit beschreibt die Eigenschaft einer Technik, wie verlässlich eine zugewiesene Funktion von einem System in einem Zeitraum erfüllt wird. Übertragen auf die V&V-Techniken heißt dies, dass die Techniken eine korrekte Prüfung des Systems ausgeben. In diesem Fall wird die Software-Zuverlässigkeit betrachtet. Diese ist definiert als die Wahrscheinlichkeit, mit der eine Software fehlerfrei in einer spezifischen Zeit und unter individuellen Bedingungen arbeitet (Lyu 1996). Des Weiteren ist Software immateriell. Das heißt, sie ist keinem Verschleiß ausgesetzt wie z. B. Hardware. Demnach ist die Fehlerrate einer Software unabhängig von ihrem Alter und der Anzahl der Anwendungen. In einer Software wird zwischen drei Fehlerarten unterschieden (Dunn 1984):

- Die Anforderungen, die an die Software gestellt worden sind, waren fehlerhaft.
- Die Auslegungen der Anforderungen waren fehlerhaft.
- Die Programmierung war fehlerhaft.

Um eine hohe Qualität der Ergebnisse zu gewährleisten, ist eine hohe Zuverlässigkeit des Systems eine Voraussetzung.

Bei der Untersuchung der Techniken sollte auch die Auswahl der Testfälle genauer betrachtet werden. Dabei ist in erster Linie wichtig, dass die in der V&V-Technik verwendeten Testfälle transparent hergeleitet werden. Denn zu beachten ist, dass möglichst die gesamte V&V automatisiert ablaufen soll. Daher sollte eine Automatisierung der Auswahl der Testfälle zumindest möglich sein. Optimal wäre hierbei, wenn die Testfälle bereits automatisch generiert werden. Bei Techniken, in denen keine automatische Auswahl durchgeführt wird, muss geprüft werden, inwiefern dies automatisiert werden kann. Ist eine Automatisierung nicht möglich, müssen auch bei einer Automatisierung der V&V-Technik menschliche Akteure eingesetzt werden, um Testfälle auszuwählen. Dadurch würde zwar der Hauptteil der V&V-Technik automatisiert ablaufen, die Vorbereitungsphase aber nicht.

Ein weiterer Aspekt, welcher beachtet werden muss, ist der Aufwand, der zu leisten ist. Dieser Aspekt stellt zwar keine notwendige Bedingung für eine Automatisierung dar, allerdings wird diesem Aspekt aus wirtschaftlicher Sicht viel Bedeutung zugestanden. Den Aufwand realistisch einzuschätzen, erscheint schwierig. Deshalb sollte in die Bewertung des Aufwandes das Aufwand-Wirkungs-Modell hinzugezogen werden. Dabei sollte der Aufwand immer im Verhältnis zur Wirkung eingeschätzt und bewertet werden.

Zunächst muss jedoch der Aufwand umrissen werden. Der Aufwand umfasst in dieser Bewertung zwei Aspekte. Zum Aufwand zählen zum einen der Kostenaufwand und zum anderen der Zeitaufwand. Um die Techniken genauer zu bewerten, werden Kosten- und Zeitaufwand wiederum jeweils in Anschaffungsaufwand bzw. Anschaffungskosten sowie den laufenden Aufwand bzw. die laufenden Kosten unterteilt. Abbildung 8 gibt eine Übersicht der Einteilung des Aufwandes.



Abbildung 8: Aufwand

Ein Anteil der laufenden Kosten bei V&V-Techniken sind die Personalkosten. Diese fallen bei jeder Technik zumindest zur Entwicklung und Anpassung an. Hierbei unterscheiden sich die Techniken teilweise erheblich. Allerdings fallen die Personalkosten bei Techniken, welche durch Mitarbeiter und Fachexperten durchgeführt werden, noch deutlicher ins Gewicht, da bei diesen Techniken bei jeder Überprüfung mehrere Mitarbeiter benötigt und bezahlt werden müssen. Ein weiterer Aspekt der laufenden Kosten sind die Materialkosten. Obwohl diese bei den meisten Techniken eine im Vergleich zu den Personalkosten geringe Bedeutung haben, werden diese trotzdem betrachtet und in die Bewertung mit einbezogen.

Der Aspekt des Zeitaufwandes lässt sich grob in zwei Kategorien unterteilen. Dies ist zum einen der Planungsaufwand, welcher vor der Durchführung geleistet werden muss, und zum anderen der Aufwand, welcher für eine Durchführung geleistet wird. Generell benötigen manuelle Techniken mehr Zeit zur Durchführung als automatische Techniken. Jedoch unterscheiden sich auch die automatischen V&V-Techniken in der zeitlichen Dauer deutlich.

4.2 Ableitung von Kriterien zur Bewertung

Im vorherigen Kapitel wurde untersucht, welche Eigenschaften die V&V-Techniken besitzen müssen, damit diese zu automatisieren sind. Dabei wurde untersucht welche Auswirkungen menschliche Akteure, die Objektivität der Technik, die Zuverlässigkeit der Ergebnisse, die Auswahl der Testfälle und der Aufwand auf die Automatisierbarkeit von V&V-Techniken hat. In diesem Kapitel werden, aufbauend auf den Erkenntnissen

aus dem vorherigen Kapitel, Kriterien für die Techniken entwickelt, die diese hinsichtlich ihres Automatisierungspotentials bewerten. Diese Kriterien werden anschließend in einer Kriterien-Matrix zusammengefügt und helfen dabei, die verschiedenen Techniken zu bewerten. Dadurch soll herausgefunden werden, welche der vorgestellten V&V Techniken (siehe Kapitel 3) am besten für weitere Untersuchungen geeignet sind.

Eine Kriterien-Matrix hilft dabei durch mehrere Kriterien, zu einem gemeinsamen Ergebnis zu gelangen. Dazu wird jedes zu bewertende Objekt in eine Zeile und jedes Kriterium in eine Spalte einer Tabelle eingetragen. Zusätzlich füllt das Gesamtergebnis eine Spalte aus. Zur Bewertung werden die betrachteten Objekte nach ausgewählten Kriterien unabhängig bewertet. Das Ergebnis der einzelnen Bewertungen wird anschließend in das zugehörige Feld der Tabelle eingetragen.

Grundsätzlich existieren zwei Arten von Bewertungssystemen. Zum einen wird eine Art Notengebung verwendet. Hierbei wird mit Hilfe von Punktzahlen oder Noten bewertet, wie exakt und gut das Objekt das betrachtete Kriterium erfüllt. Dabei können die unterschiedlichen Kriterien unterschiedlich gewichtet werden. So können Kriterien mehr oder weniger Einfluss auf das Gesamtergebnis haben. Das Gesamtergebnis ergibt sich schließlich aus der Summe der Gewichtungen verrechnet mit den Bewertungen. Die andere Möglichkeit eine Bewertung vorzunehmen, ist das Ampelsystem. Hierbei wird lediglich bewertet, ob das Objekt das betrachtete Kriterium erfüllt. Wird das Kriterium erfüllt, kann beispielsweise mit ‚grün‘ und bei Nichterfüllung mit ‚rot‘ bewertet werden. Auch kann dabei eine Zwischenstufe verwendet werden. Diese kann beispielsweise angewendet werden, wenn das Kriterium ausreichend erfüllt wurde. Je nach Kriterium kann bereits eine einzelne negative Bewertung eine negative Gesamtbewertung zur Folge haben, was als hartes Entscheidungskriterium bezeichnet wird. Tabelle 4 zeigt ein Beispiel für eine Entscheidungsmatrix mit einem Ampelsystem, als Bewertungssystem, bei dem statt grün ‚+‘ und anstatt rot ‚-‘ verwendet wurde.

Tabelle 4: Beispiel für eine Entscheidungsmatrix

	Gesamt-Ergebnis	K1	K2	K3	K4	K5
Technik 1	-	-	-	+	+	-
Technik 2	+	+	+	-	+	+
...						

Die Kriterien in der Matrix, welche für die Bewertung der V&V-Techniken zu Grunde liegen, werden mit einem Ampelsystem bewertet. In wieweit die Bewertung der einzelnen Kriterien in die Gesamtbewertung einfließt, wird im Zuge der Vorstellung der Kriterien erläutert. Der verwendeten Kriterien-Matrix liegen folgende drei Bewertungsstufen zu Grunde:

„+“: Das Kriterium wurde optimal erfüllt

„o“: Das Kriterium wurde ausreichend genug erfüllt

„-“: Das Kriterium wurde nicht erfüllt

Das Ziel der Matrix ist eine aussagekräftige Übersicht mit allen Kriterien und vorgestellten V&V-Techniken zu erstellen, die die Frage beantwortet, welche V&V-Techniken automatisiert werden können. Des Weiteren sollte die Matrix aufzeigen, welche Technik am besten für eine Automatisierung geeignet ist. Durch die Matrix soll die Möglichkeit bestehen, die V&V-Techniken einfach miteinander hinsichtlich ihres Automatisierungspotentials zu vergleichen. Die Fragestellung lautet und worauf die Kriterien abzielen ist, ob die betrachtete V&V-Technik zu automatisieren ist. Demnach stellt die Automatisierbarkeit kein Kriterium dar, sondern ergibt sich aus der Auswertung der Kriterien und stellt das Ergebnis einer jeden Untersuchung dar. Die Kriterien wurden aus Kapitel 4.1 abgeleitet und werden im Folgenden vorgestellt.

Kriterium 1: Menschliche Akteure: Menschliche Akteure stellen für eine Automatisierung eine Schwierigkeit dar. Bei diesem Kriterium wird die Frage beantwortet, ob menschliche Akteure im Zuge der Anwendung der betrachteten Technik beteiligt sind. Mit „+“ wird bewertet, kein menschlicher Akteur in der Technik beteiligt, und „-“, mindestens ein menschlicher Akteur in der Technik beteiligt, bewertet. Die mittlere Stufe „o“ entfällt bei diesem Kriterium. Ein menschlicher Akteur ist für eine manuelle V&V-Technik, die automatisiert werden soll, ein Ausschlusskriterium. Werden in der betrachteten V&V-Technik ein oder mehrere menschliche Akteure eingesetzt, wird dieses Kriterium nicht erfüllt und mit „-“ bewertet. Folglich erhält die Technik in der Gesamtbewertung auch ein „-“, wird als nicht automatisierbar eingestuft und scheidet für weitere Untersuchungen aus. Damit ist dieses Kriterium ein Ausschlusskriterium.

Kriterium 2: Objektivität der Ergebnisse: In vorherigen Kapiteln (siehe Kapitel 4.1) wird beschrieben, welchen Einfluss die Objektivität auf die Qualität der Ergebnisse hat. Die Ergebnisse der Techniken sollten möglichst objektiv sein, damit eine Qualität gewährleistet werden kann. Des Weiteren ergibt sich bei erhöhter Objektivität die

Möglichkeit, Ergebnisse bei Anwendung derselben Technik besser miteinander vergleichen zu können. Durch eine erhöhte Objektivität steigt die Genauigkeit der Technik bei gleichzeitiger geringerer Varianz in den Ergebnissen. Dadurch wird eine hohe Qualität der Ergebnisse einer Technik gewährleistet, wodurch die Technik eine erhöhte Akzeptanz findet. Techniken, welche eine geringere Objektivität aufweisen, können zwar automatisiert werden, allerdings wird durch größere Subjektivität eine differenzierte Bewertung schwierig. Das Kriterium, welches in der Matrix bewertet wird, ist, ob die betrachtete Technik die Ergebnisse mit einer ausreichenden und genauen Objektivität liefert und damit die Qualität der Ergebnisse gewährleistet. Dabei wird die betrachtete Technik mit „+“ beurteilt, wenn die Technik komplett objektiv bewertet und „-“ falls im Zuge der Technik subjektiv bewertet wird. Wird überwiegend objektiv bewertet und nur an wenigen Stellen mit subjektivem Einfluss, wird die Technik in diesem Kriterium mit „o“ bewertet.

Kriterium 3: Zuverlässigkeit der Ergebnisse: Wie in Kapitel 4.1 erläutert, ist eine Zuverlässigkeit der V&V-Technik notwendig, um die Qualität der Ergebnisse zu gewährleisten. Daher muss die Zuverlässigkeit der Technik bei der Bewertung des Automatisierungspotentials berücksichtigt werden. Denn arbeitet eine Technik nicht mit einer Mindestzuverlässigkeit, kann diese zwar automatisiert werden, jedoch fehlt die Grundlage für eine einheitliche Bewertung gänzlich. Dieses Kriterium beantwortet die Frage, wie zuverlässig die V&V-Technik die Ergebnisse liefert. Dabei wird in drei Abstufungen unterteilt. Eine positive Bewertung „+“ heißt, dass die Technik die Ergebnisse sehr zuverlässig liefert. Bei einer negativen Bewertung „-“ werden die Ergebnisse der Technik nicht zuverlässig geliefert und bei einer Bewertung mit „o“ wird das Ergebnis ausreichend zuverlässig geliefert.

Kriterium 4: Testfälle: Auswahl der Testfälle: In Kapitel 4.1 wird der hohe Stellenwert einer automatischen Auswahl der Testfälle beschrieben. Ohne eine automatische Auswahl der Testfälle kann die gesamte V&V-Technik nicht automatisiert werden. Weshalb dieser Aspekt in den Kriterien berücksichtigt wird. Dieses Kriterium prüft, ob eine automatische Auswahl der Testfälle generell möglich ist. Bei der Bewertung dieses Kriteriums kann wie bei Kriterium 1 lediglich mit „+“ oder „-“ bewertet werden. Eine positive Bewertung „+“ bedeutet, dass bei der betrachteten Technik die Auswahl der Testfälle automatisiert ist oder automatisiert werden kann. Dagegen drückt eine negative Bewertung „-“ aus, dass nach den Untersuchungen eine automatisch ablaufende Auswahl

der Testfälle nicht möglich scheint. Eine negative Bewertung dieses Kriteriums hat eine negative Gesamtwertung zur Folge. Daher ist Kriterium 4 ein Ausschlusskriterium.

Kriterium 5: Aufwand: Der im vorherigen Kapitel vorgestellte Aufwand stellt kein Ausschlusskriterium dar. Denn auch wenn eine Technik nur mit sehr großem monetärem oder zeitlichem Einsatz automatisiert werden kann, ist dies theoretisch möglich. Dieses Kriterium soll dabei helfen, die Auswahl für eine Technik zu treffen, die im weiteren Verlauf der Arbeit automatisiert und praktisch angewendet werden soll. Die Fragestellung für dieses Kriterium lautet daher, wie viel Aufwand muss geleistet werden, um die Technik zu automatisieren. Ein Problem stellen die kaum vorhandenen Grundlagen in der Untersuchung dar, weshalb subjektive Einschätzungen in die Bewertung dieses Kriteriums mit einfließen können. Bei geringem Aufwand wird die Technik mit „+“ bewertet, bei akzeptablem Umfang mit „o“ und bei großem Umfang mit „-“.

Tabelle 5 gibt eine Übersicht und eine kurze Beschreibung der vorgestellten Kriterien. Im nächsten Schritt werden diese erarbeiteten Kriterien in einer Matrix zusammengefasst. Dabei ist zu beachten, dass bei einer negativen Bewertung von Kriterium 1 oder Kriterium 4 das Automatisierungspotential dieser Technik nicht gegeben ist. Dabei haben die anderen Kriterien keinen Einfluss, selbst wenn diese alle positiv bewertet wurden. Tabelle 6 zeigt die Kriterien-Matrix, auf deren Grundlage im folgenden Kapitel die verschiedenen V&V-Techniken bewertet werden. K1 steht dabei für Kriterium 1 und das Automatisierungspotential ist das Gesamtergebnis der Kriterien eins bis fünf.

Tabelle 5: Übersicht der Kriterien zur Klassifikation der V&V-Techniken hinsichtlich ihres Automatisierungspotentials

Nr	Kriterium	Beschreibung	Bewertung
K1	Menschlicher Akteur Ausschlusskriterium	Werden bei der Technik menschliche Akteure miteinbezogen?	+ kein menschlicher Akteur – menschlicher Akteur
K2	Objektivität der Ergebnisse	Liefert die Technik eine ausreichende und genaue Qualität der Ergebnisse?	+ objektive Bewertung o größtenteils objektive Bewertung – subjektive Bewertung
K3	Zuverlässigkeit der Ergebnisse	Wie zuverlässig liefert die Technik die Ergebnisse?	+ sehr zuverlässig o ausreichend zuverlässig – nicht zuverlässig
K4	Testfälle Ausschlusskriterium	Ist eine automatische Auswahl der Testfälle möglich?	+ generell möglich – nicht möglich
K5	Aufwand	Wie viel Aufwand muss geleistet werden, um die Technik zu automatisieren?	+ geringer Aufwand o akzeptabler Aufwand – großer Aufwand

Das Automatisierungspotential einer Technik ergibt sich aus den Bewertungen der fünf aufgeführten Techniken. Kriterium 1 und Kriterium 5 stellen dabei harte Kriterien dar. Wurde eines dieser Kriterien negativ bewertet, wird das Automatisierungspotential auch negativ bewertet. Um mit einem positiven Ergebnis bewertet zu werden, benötigt eine Technik mindestens drei positiv bewertete Kriterien. Tabelle 6 gibt das Muster der entwickelten Kriterien-Matrix für die V&V-Techniken wieder. Dabei wird bei Technik 4 deutlich, dass obwohl vier der fünf Kriterien positiv bewertet wurden, das Ausschlusskriterium 1 zur Folge hat, dass die Gesamtbewertung negativ ist. Im nächsten Kapitel werden die V&V-Techniken mit Hilfe dieser Kriterien-Matrix hinsichtlich ihres Automatisierungspotentials analysiert und bewertet.

Tabelle 6: Kriterien-Matrix für das Automatisierungspotential von V&V-Techniken

V&V-Technik	Auto.-potential	K1	K2	K3	K4	K5
Technik 1	+	+	o	o	+	o
Technik 2	+	+	o	+	+	-
Technik 3	-	-	o	-	-	+
Technik 4	-	-	+	+	+	+

4.3 Bewertung der V&V-Techniken

Zunächst werden die V&V-Techniken aus Kapitel 3 hinsichtlich der abgeleiteten Kriterien aus Kapitel 4.2 untersucht, welche im Data Mining angewendet werden. Wie in Kapitel 3.2 beschrieben, findet im Data Mining die Assoziationsanalyse bereits Anwendung. Wie die aufgezeigte Anwendung in einem Webshop zeigt, kann die Berechnung von Konfidenz und Support bereits automatisiert durchgeführt werden. Als Grundlage für die Bewertung der Relevanz der Regel muss lediglich im Vorhinein eine Mindestkonfidenz und ein Mindestsupport festgelegt werden. Beim Übertreffen dieser Werte stuft der Algorithmus die Regeln als interessant ein. Allerdings kann durch die Anwendung dieser Algorithmen nicht die Datenvorbereitung geprüft werden, sondern die Algorithmen setzen eine korrekte Datenvorverarbeitung sogar voraus. Ein Schwachpunkt der Assoziationsanalyse ist, dass eine Mindestkonfidenz angegeben werden muss. Zwar werden im Endeffekt lediglich die Regeln angegeben, welche die höchste Konfidenz vorweisen, jedoch muss die untere Schranke übertroffen werden, damit diese Regel in Betracht gezogen wird. Somit behält die Assoziationsanalyse einen kleinen subjektiven Einfluss.

Wie in Kapitel 3.2 beschrieben, werden die Daten bei den Stichprobentests abhängig vom Verfahren in verschiedene Gruppen unterteilt. Bei allen betrachteten Stichprobentests (Kreuzvalidierung, Holdoutmethode und Bootstrapping-Validierung) werden keine menschlichen Akteure eingesetzt. Auch kann die Auswahl der Testfälle, die durch die Unterteilung der Daten entsteht, automatisiert werden. Da in keiner dieser Stichprobentests subjektive Einflüsse und Voraussetzungen bestehen, sind die Techniken als objektiv zu bewerten. Aufgrund des verschiedenen Einsatzes, bedingt durch die Anzahl an Datensätzen, werden die Techniken in der Zuverlässigkeit der Ergebnisse unterschiedlich bewertet. Grund dafür ist das unterschiedliche Risiko des Overfittings der Modelle in den Techniken. Grundsätzlich besteht die Möglichkeit sowohl die

Kreuzvalidierung als auch die Holdoutmethode zu automatisieren. Auch die Bootstrapping-Validierung, die auf der Kreuzvalidierung und der Holdoutmethode basiert, kann automatisiert werden. Tabelle 7 zeigt eine Übersicht der Bewertungen der V&V-Techniken im Data Mining.

Tabelle 7: Bewertungsübersicht der V&V-Techniken im Data Mining hinsichtlich Automatisierung

V&V-Technik	Auto.-potential	K1	K2	K3	K4	K5
Assoziationsanalyse	+	+	o	+	+	+
Kreuzvalidierung	+	+	+	o	+	+
Holdoutmethode	+	+	+	+	+	+
Bootstrapping-Validierung	+	+	+	o	+	+

Im Folgenden werden die V&V-Techniken der Softwareentwicklung hinsichtlich der Kriterien, die das Automatisierungspotential aufdecken sollen, untersucht. Dabei werden zunächst die dynamischen Testverfahren betrachtet. Die Gemeinsamkeit ist, wie in Kapitel 3.3 herausgestellt wurden, dass das Programm in der realen Umgebung und mit konkreten Werten getestet wird.

Das kontrollflussorientierte Testverfahren (ein White-Box-Verfahren) wird der Gruppe der strukturorientierten Testmethoden zugeordnet und beschreiben ein Ablaufdiagramm eines Computerprogramms. Eine automatische Generierung eines Ablaufdiagramms ist generell möglich. Dabei können Modelling-Programme (z.B. Borland Together Architect, Visual Paradigm) verwendet werden, die aus einem Programmcode ein Klassendiagramm erstellen. Daher können die zu überprüfenden Quellcodes ohne die Mithilfe eines menschlichen Akteurs in ein Ablaufdiagramm abgebildet werden. Auch die Generierung der Testfälle, dass alle Pfade des Diagramms mindestens einmal getestet werden müssen, kann automatisiert werden. Des Weiteren besteht die Möglichkeit, dass kontrollflussorientierte Testverfahren als objektiv und zuverlässig bewertet werden. Somit besteht generell die Möglichkeit, das kontrollflussorientierte Testverfahren und auch generell die White-Box-Verfahren mit angemessenem Aufwand zu automatisieren.

Für Black-Box-Tests besteht generell die Möglichkeit der Automatisierung (Dustin et al. 2001; European Design Automation Association et al. 2013). Somit müssen die verschiedenen Verfahren zur Auswahl von Testfällen analysiert werden. Da in Black-Box-Verfahren das System nicht nur gegen sich selbst getestet wird, sondern gerade durch die Tests der Spezifikationen ausgezeichnet wird, kann von objektiven und zuverlässigen Tests gesprochen werden. Die funktionale Äquivalenzklassenbildung kann

automatisiert werden. Dazu muss der Prozess, der die verschiedenen Daten sinnvoll unterteilt und Klassen zuordnet, automatisiert werden. Wurden die Daten geordnet und unterteilt kann wiederum die Grenzwertanalyse automatisiert werden, da diese die Werte zwischen den verschiedenen Klassen testet. Auch kann das Testen mit speziellen Werten automatisiert werden. Dabei müssen lediglich spezielle Werte wie beispielsweise ‚Null‘ und andere auffällige Werte im Programm getestet werden.

Die statischen Verfahren (siehe Kapitel 3.3) betrachten und überprüfen den Quellcode des Programms auf Fehler, ohne das Programm selbst auszuführen. Dabei werden menschliche Akteure eingesetzt, um die Analyse vorzunehmen. Damit werden die statischen Verfahren bei Kriterium 1 negativ bewertet. Das bedeutet, dass für statische V&V-Techniken kein Automatisierungspotential besteht und diese nicht weiter betrachtet werden.

Bei diversifizierenden Tests werden die Ergebnisse von Tests miteinander verglichen (siehe Kapitel 3.3). Back-to-Back-Tests können in einem Feedback-Loop basierten, selbstorganisierten Mechanismus automatisiert werden (Wotawa et al. 2016). Eine Automatisierung erfordert die Abbildung von Anforderungen auf tatsächliche Designs und die Implementierung von Fachwissen. Daher ist die Gewährung der richtigen Freiheitsgrade im System wichtig, um die Selbstorganisation zu ermöglichen. Back-to-Back-Tests unterstützen diese Anforderung mit einer geeigneten Testmethode, um Fehler aufzudecken. Um Back-to-Back-Tests in selbstorganisierte Mechanismen einbinden zu können, ergeben sich folgende Herausforderungen (Wotawa et al. 2016):

- Bereitstellung von Entscheidern, die in der Lage sind, Entscheidungen im fast unbegrenzten Entscheidungsraum zu treffen.
- Systematische Erzeugung von Testfällen, da wegen des unbegrenzten Zustandsraum keine erschöpfenden Tests möglich sind. Darüber hinaus basieren die meisten selbstorganisierten Algorithmen auf Heuristiken, wodurch ihr Verhalten ziemlich unbestimmt und ihr Zustandsraum ungleichmäßig ist. Daher können allgemeine Testfallauswahlstrategien, die auf strukturiertem Programmverhalten beruhen, nicht verwendet werden.
- Automatisierung der Testdurchführung und -auswertung

Wotawa et. al. (2016) haben einen Ansatz für die automatische Anwendung von Back-to-Back-Tests für selbstorganisierende Mechanismen entwickelt. Dabei unterstützt das Konzept des Back-to-Back-Tests die anspruchsvolle Aufgabe der Entwicklung von

selbstorganisierten Mechanismen in einer gemeinsamen Entwicklung und ist in der Lage, verschiedene Arten von Fehlern in der funktionalen Spezifikation des Systems aufzudecken (Wotawa et al. 2016). Demnach können sowohl Kriterium 1 als auch Kriterium 4 positiv bewertet werden. Daraus lässt sich folgern, dass der Back-to-Back-Test auch für andere Heuristiken und Algorithmen automatisierbar ist.

Um den Mutationstest zu automatisieren, müssen die drei Bedingungen für einen erfolgreichen Test aus Kapitel 3.3 automatisch erfüllt werden. Des Weiteren muss die Modifikation am Quellcode automatisiert werden, damit der Mutationstest in Gänze automatisiert werden kann. Die gezielten Änderungen am Quellcode wie auch die Erstellung von Testfällen scheinen mit einem geeigneten Werkzeug automatisierbar. Auch scheinen die drei Bedingungen erfüllt zu sein. Um dabei sicher zu gehen, müssten je nach System eine Vielzahl an Testfällen durchlaufen werden. Des Weiteren können auch die Generierung und Auswahl der Testfälle automatisiert werden. Dies erscheint jedoch nur mit einer jeweiligen auf das System angepassten Programmänderung möglich. Allgemein besteht aber die Möglichkeit, den Mutationstest zu automatisieren.

Der Regressionstest ist in Echtzeitsystemen nur mit hohem Aufwand anwendbar, da im Grunde bei keinem Test eine Wiederholung gewährleistet werden kann. Kleine Änderungen am System führen bereits zu einem veränderten Verhalten (Sneed et al. 2009). Diese Veränderungen müssen wiederum manuell vorgenommen werden. Durch diesen menschlichen Eingriff kann der Regressionstest nicht in Echtzeit erfolgen. Um dieses Problem zu lösen, kann eine Implementierung von automatischen Teilsystemen erfolgen. Allerdings ist der Aufwand einer solchen Implementierung aus folgenden Gründen hoch und wird selten umgesetzt (Liggesmeyer 2009):

- das Testsystem deckt alle Funktionen des Programms ab
- das Testsystem wird parallel zum Programm entwickelt
- das Testsystem wird zeitgleich mit dem Programm angepasst
- wird eine Änderung an der Hardware vorgenommen, müssen auch die Testergebnisse neu aufgesetzt werden. Gegen die neuen Testergebnisse wird anschließend geprüft

Daher lässt sich der Regressionstest grundsätzlich automatisieren. Jedoch muss diese Automatisierung nach jeder Änderung an der Hardware neu aufgesetzt werden. Deshalb existiert keine Möglichkeit, eine umfassende Automatisierung für einen Regressionstest zu entwickeln, sondern die Automatisierung muss jeweils im Einzelnen lokal vor Ort

entwickelt werden. In Tabelle 8 sind die Bewertungen der V&V-Techniken in der Softwareentwicklung zusammengefasst.

Tabelle 8: Bewertungsübersicht der V&V-Techniken in der Softwareentwicklung hinsichtlich Automatisierung

V&V-Technik	Auto.-potential	K1	K2	K3	K4	K5
Kontrollflussorientierte (White-Box)	+	+	+	o	+	o
Funktionale Äquivalenzklassenbildung (Black-Box)	+	+	+	+	+	o
Grenzwertanalyse (Black-Box)	+	+	+	+	+	o
Test spezieller Werte (Black-Box)	+	+	+	+	+	o
Back-to-Back	+	+	+	o	+	–
Mutationstest	+	+	+	–	+	–
Regressionstest	+	+	o	+	+	–

Die in Kapitel 3.4 vorgestellten V&V-Techniken in der Simulation werden nun untersucht und betrachtet. Informale Techniken folgen formalen und strukturierten Abläufen und werden deshalb häufig in der Simulation eingesetzt. Grundsätzlich sind Vorgänge, die formalen und strukturierten Abläufe folgen, sehr gut für eine Automatisierung geeignet. Bei keiner informalen Technik stellt die Skizzierung der Abläufe ein Problem dar. Das große Hindernis einer Automatisierung einer jeden informalen V&V-Technik stellen allerdings die menschlichen Akteure dar. Informale Techniken sind insbesondere durch ihre Subjektivität und das menschliche Denken charakterisiert und verwenden keine formalen mathematischen Operatoren. Daher findet sich in jeder informalen Technik Menschen wieder, die mit ihrer Erfahrung und ihrem Wissen Bewertungen vornehmen bzw. Meinungen vertreten. Damit wird jede informale Technik bei Kriterium 1 negativ eingestuft und bietet somit keine Möglichkeit zur Automatisierung.

Statistische Techniken prüfen, wie sicher das reale System vom Modell abgebildet wird. Dazu werden verschiedene Berechnungen, wie in Kapitel 3.4 vorgestellt, verwendet. Bei der Korrelationsanalyse werden lineare Abhängigkeiten in Datensätzen überprüft. In dieser Technik werden keine menschlichen Akteure gebraucht und deshalb kann Kriterium 1 positiv bewertet werden. Die Technik arbeitet ausnahmslos objektiv und funktioniert sehr zuverlässig, sodass festgehalten werden kann, dass die Korrelationsanalyse automatisiert werden kann.

Auch die Chi-Quadrat-Test prüft den Zusammenhang in Daten, jedoch nicht linear. Eine Automatisierung dieser Technik erscheint ähnlich realisierbar zu sein wie die Korrelationsanalyse. Der Unterschied liegt einzig in der Abhängigkeitsanalyse. Aber auch diese kann automatisiert werden. Die Regressionsanalyse, die die Ursache der Abhängigkeit ableitet, kann, auf Grundlage der Korrelationsanalyse und des Chi-Quadrat-Tests automatisiert werden. Voraussetzung hierfür ist allerdings, dass der Ursprung der Abhängigkeiten in den Daten dokumentiert ist.

Der in Kapitel 3.4 vorgestellte Test von Teilmodellen fordert eine Unterteilung der Modelle in Teilmodelle in Top-Down-Form. Dies erscheint ohne die Hilfe eines menschlichen Akteurs, der das Modell kennt, nicht realisierbar. In der Literatur findet sich kein passendes Werkzeug, welches den Schritt der Zerlegung eines Modells allgemein automatisiert. Daher ist auf menschliche Akteure zurückzugreifen, womit diese V&V-Technik nicht automatisierbar ist.

Die Validierung von Vorhersagen vergleicht im Wesentlichen die Ergebnisse von Modell und realem System miteinander (siehe Kapitel 3.4). Ein Vergleich kann ohne die Hilfe von einem menschlichen Akteur durchgeführt und automatisiert werden. Auch können die Eingabevariablen in das Modell automatisch erzeugt werden. Gleichzeitig liefert ein Vergleich ein objektives und zuverlässiges Ergebnis. Somit kann die Validierung von Vorhersagen automatisiert werden. Ähnlich erscheint die Bewertung der Kriterien des Vergleichs mit anderen Modellen. Auch dabei steht ein Abgleich der Ergebnisse im Mittelpunkt. Und, genauso wie bei der Validierung von Vorhersagen, müssen Eingangsdaten entwickelt werden.

Auch beim Vergleich mit aufgezeichneten Daten muss ein Vergleich automatisiert werden. Zusätzlich müssen die Ausgangsdaten in Gruppen aufgeteilt werden, wobei ein Teil der Daten zur Justierung des Modells beiträgt. Beide Punkte scheinen ohne Hilfe eines menschlichen Akteurs durchführbar. Durch Zerteilung der Daten und der Justierung des Modells ist die Zuverlässigkeit der Ergebnisse nicht so hoch wie bei den anderen hier betrachteten Vergleichstechniken. Insgesamt kann der Vergleich mit aufgezeichneten Daten automatisiert werden.

Die dynamischen Techniken zeichnen sich in erster Linie durch den Zeitpunkt der Testdurchführung aus (siehe Kapitel 3.4). Im Akzeptanztest wird überprüft, ob der Kunde mit dem Produkt zufrieden ist und die Anforderungen, welche im Vorhinein festgelegt wurden, erfüllt wurden. Diese Kontrolle kann nicht automatisiert werden und erfordert zwingend einen menschlichen Akteur. Ähnlich verfahren der Alpha- und der Betatest, bei

denen gleichermaßen ein menschlicher Akteur gefordert wird. Demnach sind alle dynamischen Techniken nicht automatisierbar (siehe Tabelle 9).

Formale V&V-Techniken sind aufgrund ihrer Basis, welche aus mathematischen Beweisen besteht, nicht nur die effektivste V&V, sondern sind theoretisch auch am geeignetsten für eine Automatisierung. Da allerdings, wie am Ende des Kapitels 3.4 beschrieben, eine Anwendung von formalen V&V-Techniken nur in Ausnahmefällen möglich ist, werden diese aufgrund ihrer geringen Relevanz für die Praxis nicht weiter untersucht und bewertet. Eine mögliche automatische V&V-Technik, welche in der Praxis kaum Anwendung findet, erscheint nicht sinnvoll.

Tabelle 9: Bewertungsübersicht der V&V-Techniken in der Simulation hinsichtlich Automatisierung

V&V-Technik	Auto.- potential	K1	K2	K3	K4	K5
Informale Techniken (Review ect.)	–	–	–	o	–	–
Korrelationsanalyse	+	+	+	o	+	+
Chi-Quadrat-Test	+	+	+	o	+	+
Regressionsanalyse	+	+	+	o	+	o
Test von Teilmodellen	–	–	+	o	–	–
Validierung von Vorhersagen	+	+	+	+	+	+
Vergleich mit anderen Modellen	+	+	+	+	+	+
Vergleich mit aufgezeichneten Daten	+	+	+	o	+	o
Akzeptanztest	–	–	–	o	–	–
Alphatest	–	–	–	o	–	–
Betatest	–	–	–	o	–	–

Tabelle 10 zeigt eine Übersicht aller untersuchten und bewerteten V&V-Techniken. Mit Blick auf das Automatisierungspotential wird deutlich, dass 17 der bewerteten 22 Techniken theoretisch automatisierbar sind. Die Grundlagen der Matrix und die Kriterien wurden in Kapitel 4.2 erläutert und dargestellt.

Tabelle 10: Bewertungsübersicht der V&V-Techniken in der Kriterien-Matrix.

	V&V-Technik	Auto.-potential	K1	K2	K3	K4	K5
Data Mining	Assoziationsanalyse	+	+	o	+	+	+
	Kreuzvalidierung	+	+	+	o	+	+
	Holdoutmethode	+	+	+	+	+	+
	Bootstrapping-Validierung	+	+	+	o	+	+
Softwareentwicklung	Kontrollflussorientierte (White-Box)	+	+	+	o	+	o
	Funktionale Äquivalenzklassenbildung (Black-Box)	+	+	+	+	+	o
	Grenzwertanalyse (Black-Box)	+	+	+	+	+	o
	Test spezieller Werte (Black-Box)	+	+	+	+	+	o
	Back-to-Back	+	+	+	o	+	-
	Mutationstest	+	+	+	-	+	-
	Regressionstest	+	+	o	+	+	-
Simulation	Informale Techniken (Review ect.)	-	-	-	o	-	-
	Korrelationsanalyse	+	+	+	o	+	+
	Chi-Quadrat-Test	+	+	+	o	+	+
	Regressionsanalyse	+	+	+	o	+	o
	Test von Teilmodellen	-	-	+	o	-	-
	Validierung von Vorhersagen	+	+	+	+	+	+
	Vergleich mit anderen Modellen	+	+	+	+	+	+
	Vergleich mit aufgezeichneten Daten	+	+	+	o	+	o
	Akzeptanztest	-	-	-	o	-	-
	Alphatest	-	-	-	o	-	-
	Betatest	-	-	-	o	-	-

Legende:

K1: Menschlicher Akteure

„+“: Das Kriterium wurde optimal erfüllt

K2: Objektivität der Ergebnisse

„o“: Das Kriterium wurde ausreichend genug erfüllt

K3: Zuverlässigkeit der Ergebnisse

„-“: Das Kriterium wurde nicht erfüllt

K4: Testfälle

K5: Aufwand

5 Untersuchung der V&V-Techniken im Data-Mining-Kontext

Die größte Herausforderung in einem Data-Mining-Prozess besteht wie in Kapitel 2 aufgezeigt in der Datenvorverarbeitung, welcher bis zu 80% der Dauer des gesamten Prozesses ausmacht. Daher wird insbesondere in den Prozessen der Datenvorbereitung ein hoher Bedarf an Qualitätssicherung gefordert. Durch die korrekte Anwendung der richtigen V&V-Technik kann die Qualität und Korrektheit in diesen Prozessen sichergestellt werden.

In diesem Kapitel werden zunächst die Techniken, die im vorherigen Kapitel als automatisierbar eingestuft wurden, weiterführend hinsichtlich ihrer Anwendbarkeit im Data Mining untersucht. Da nicht alle Techniken für den Einsatz im Data Mining geeignet sind, werden wiederum Ausschlusskriterien abgeleitet. Die V&V-Techniken, welche in Kapitel 4 als automatisierbar eingestuft wurden, werden daraufhin auf diese Kriterien analysiert. Anschließend wird ein Ansatz aufgezeigt, wie die Transformation in das Data Mining, insbesondere in das Vorgehensmodell MESOC, erfolgen kann.

5.1 Analyse der V&V-Techniken hinsichtlich der Übertragung ins Data Mining

In dieser Arbeit werden nicht nur V&V-Techniken betrachtet und untersucht, welche schon im Data Mining zum Einsatz kommen, sondern auch Techniken, die in der Softwareentwicklung und in der Simulation genutzt werden. Da diese Arbeit zum Ziel hat, die Automatisierbarkeit speziell im Data-Mining-Prozess zu untersuchen, muss die Möglichkeit zum Einsatz der Techniken im Data Mining untersucht werden. Nicht jede V&V-Technik, die im vorherigen Kapitel als automatisierbar eingestuft wurde, ist für einen Einsatz im Data Mining geeignet. Damit diese Techniken im Data Mining angewendet werden können, müssen sie unterschiedliche Voraussetzungen erfüllen. Dabei stoßen die Techniken der Softwareentwicklung und die der Simulation im Data Mining auf unterschiedliche Hindernisse. Grund dafür sind unterschiedliche Ausgangspunkte und Problemzonen, auf denen der Fokus der Untersuchungen liegt. Ein weiterer Aspekt, welcher hier untersucht und diskutiert wird, ist das mögliche Anwendungsgebiet der V&V-Techniken im Data-Mining-Prozess. Damit ist gemeint, in welchem Prozessschritt innerhalb des Data Minings die betreffende Technik eingesetzt werden kann. Dahingehend werden auch die V&V-Techniken untersucht, die bereits im Data-Mining eingesetzt werden. Betrachtet werden im Folgenden nur die Techniken,

welche in Kapitel 4 als automatisierbar eingestuft wurden. Dabei können die betrachteten Techniken grundsätzlich in folgende Kategorien eingeteilt werden.

- Techniken, die bereits im Data Mining eingesetzt und angewendet werden (geeignet für den Einsatz im Data Mining)
- Techniken, die in der Softwareentwicklung oder der Simulation zum Einsatz kommen und ohne Änderungen auch im Data Mining angewendet werden können (geeignet für den Einsatz im Data Mining)
- Techniken, die in der Softwareentwicklung oder der Simulation eingesetzt werden und nicht ohne Modifizierungen im Data Mining eingesetzt werden können (weitere Untersuchungen notwendig)
- Techniken, die nicht im Data Mining eingesetzt werden können (nicht geeignet für den Einsatz im Data Mining)

Von den in Kapitel 3 vorgestellten und in Kapitel 4 untersuchten V&V-Techniken eignen sich folglich nicht alle für den Einsatz im Data Mining. Neben den Techniken, die bereits im Data Mining eingesetzt werden, sind bei den Techniken der Simulation und der Softwareentwicklung verschiedene Aspekte zu nennen, weshalb diese im Data Mining eingesetzt oder nicht eingesetzt werden können. Im Folgenden sind verschiedene Kriterien für den Ausschluss einiger Techniken aufgeführt:

Kriterium 1: Dynamisches Modell: Für die Durchführung einer Simulation sind Modelle, welche zeitliche Zustände in einem dynamischen Modell darstellen, eine gute Option, da diese zeitlichen Zusammenhänge der Realität abbilden können. Jedoch sind diese dynamischen Modelle für Data-Mining-Modelle nicht geeignet. Das liegt daran, dass die Daten zu einem genauen Zeitpunkt erhoben werden müssen. Deshalb sind auch visuelle Techniken (z.B. Monitoring, Animation), die das Modell grafisch abbilden, nicht anwendbar. Auch Ereignisvaliditätstests sind aufgrund der zeitlichen Abfolge der verschiedenen Ereignisse nicht geeignet (siehe Kapitel 3.4).

Kriterium 2: Formale Techniken: Des Weiteren sind auch die formalen V&V-Techniken für die Simulation nur selten in der Praxis verwendbar (siehe Kapitel 3.4). Denn die Korrektheitsprüfung kann nicht einmal auf ein wenig komplexes Simulationsmodell angewendet werden (siehe Kapitel 3.4).

Kriterium 3: Änderung der Eingangswerte: Auch Techniken, welche durch Änderung der Eingangswerte durchgeführt werden, sind in Data-Mining-Modellen nicht geeignet. Darunter zählen unter anderen die Äquivalenzklassenbildung und die Grenzwertanalyse.

Denn die Grundlage für den Einsatz von Data Mining sind Datenbestände, die wegen der Komplexität durch normale Analysen nicht untersucht werden können.

Kriterium 4: Quellcodes: Ein Aspekt, der insbesondere bei den Techniken in der Softwareentwicklung beachtet werden muss, ist die Analyse von Quell- und Softwarecodes innerhalb einer Technik. V&V-Techniken, welche einen Code untersuchen, sind für Data-Mining-Prozesse nicht geeignet. Grund dafür ist, dass diese Techniken keine Allgemeingültigkeit besitzen und damit nicht in jedem Prozess angewendet werden können, da die meisten Data-Mining-Prozesse werden mit einer Standard-Software durchgeführt. Nach einer Studie des Fraunhofer-Instituts greifen lediglich acht Prozent der Prozesse auf eine selbst entwickelte Software zurück (Weskamp et al. 2014). Demnach stehen nur in diesen Fällen Quell- und Softwarecodes zur Verfügung. Codes können dementsprechend nur bei den rund acht Prozent der Prozesse überprüft werden. Zusätzlich muss die selbstentwickelte Software bereits in der Entwicklungsphase geprüft werden. Die erste Prüfung bei der Erstanwendung durchzuführen, reicht nicht aus.

Kriterium 5: Reales System: In der Simulation setzen viele V&V-Techniken die Existenz eines realen Systems neben einem Modell voraus. Einige der Techniken verwenden beide Systeme, um die Ergebnisse miteinander zu vergleichen. Dabei wird geprüft, wie gut das reale System abgebildet wurde. Im Data Mining jedoch existiert kein reales System, welches durch ein Modell abgebildet wird. Daher sind alle V&V-Techniken, die ein reales System voraussetzen und nutzen, für die Anwendung innerhalb eines Data-Mining-Prozesses nicht geeignet. Tabelle 11 gibt einen Überblick der Kriterien, die für einen Einsatz im Data Mining erfüllt sein müssen.

Tabelle 11: Kriterien zur Anwendung von V&V-Techniken im Data Mining

Nr	Kriterium	Beschreibung
K1	Dynamische Modell	Dynamische Modelle sind für den Einsatz im Data Mining nicht geeignet
K2	Formale Techniken	Formale Techniken sind grundsätzlich nicht für den Einsatz im Data Mining geeignet
K3	Änderungen der Eingangswerte	Eingangswerte dürfen innerhalb des Data Minings nicht verändert und modifiziert werden
K4	Quellcodes	Quellcode-Analysen sind für den Einsatz im Data Mining nicht geeignet
K5	Reales System	Reale Systeme existieren im Data Mining nicht und können somit auch nicht untersucht werden

Die in Kapitel 3 vorgestellten V&V-Techniken aus der Softwareentwicklung und der Simulation werden im Folgenden untersucht. Dabei steht die Prüfung im Mittelpunkt, ob diese Technik auch im Data Mining angewendet werden könnte. Des Weiteren werden die verschiedenen betrachteten Techniken miteinander verglichen. In diesem Kapitel werden nur die V&V-Techniken betrachtet und untersucht, die in Kapitel 4 als automatisierbar eingestuft wurden.

In Kapitel 3.3 wurden verschiedene V&V-Techniken vorgestellt, die in der Softwareentwicklung zum Einsatz kommen. Diese wurden in Kapitel 4.3 hinsichtlich ihres Automatisierungspotentials untersucht und bewertet. Die White-Box-Verfahren verlangen einen Einblick in den Quellcode des zu betrachtenden Systems und testen anhand der Erkenntnisse aus dem Code (siehe Kapitel 3.3). Auch wenn eine Automatisierung möglich ist, kann kein White-Box-Verfahren in einem Data-Mining-Prozess angewendet werden, da ein Einblick in den Quellcode des Verfahrens nicht gewährleistet werden kann. Damit scheidet sämtliche White-Box-Verfahren für weitere Untersuchungen aus.

Bei den dynamischen Verfahren, die sich durch einen Testlauf mit konkreten Eingabewerten auszeichnen, werden weder Quellcodes eines Systems noch ein reales System verlangt. In der funktionalen Äquivalenzklassenbildung muss jedoch die

Möglichkeit bestehen, die Daten in Klassen einzuteilen. Ein Kriterium für den Einsatz im Data Mining ist, dass die Eingangswerte nicht verändert werden dürfen. Allerdings wird sowohl bei der funktionalen Äquivalenzklassenbildung als auch der Grenzwertanalyse genau dies verlangt. Somit scheinen diese beiden Techniken für den Einsatz im Data Mining ungeeignet.

Von den dynamischen Verfahren erscheint das Testen mit speziellen Werten am geeignetsten, um im Data Mining eingesetzt zu werden. Grund dafür ist, dass keine Anforderungen gestellt werden. Im Test werden spezielle Werte, bei denen häufig Probleme und Fehler auftreten, ausgewählt und in Testbereichen getestet. Im Data-Mining-Prozess sind bekannte fehleranfällige Testbereiche ebenfalls interessant zu überprüfen. Eine Möglichkeit zur Anwendung wäre die Verwendung einer selektiven Stichprobe (siehe Kapitel 3). Dabei werden Attribute und Attributsausprägungen geprüft, welche in der Vergangenheit häufig fehleranfällig waren. Um die Technik effektiver zu gestalten, könnte dieses Verfahren mit einem anderen Verfahren verknüpft werden. Das Ziel ist die fehleranfälligen Wertebereiche aufzudecken und gezielt zu testen. Zu beachten ist jedoch, dass die V&V-Techniken, welche auf Expertenwissen zurückgreifen, häufig nicht zu automatisieren sind und somit in dieser Arbeit nicht berücksichtigt werden.

Die statischen Techniken wurden in Kapitel 4 im Unterschied zu den dynamischen Verfahren als nicht automatisierbar eingestuft, weshalb auf diese hier nicht weiter eingegangen wird.

Bei den diversifizierenden Techniken stehen die Modelle mit ihren Ergebnissen im Vordergrund der Untersuchung. Im Zuge des Back-to-Back-Testes werden zwei Modelle getestet. Diese Modelle, welche parallel und unabhängig voneinander entwickelt werden, werden in Bezug auf ihre Ergebnisse verglichen. Eine Durchführung des Back-to-Back-Testes im Data Mining ist daher generell möglich, allerdings in den meisten Fällen wenig sinnvoll. Denn eine parallele und unabhängige Durchführung ohne die gemeinsame Nutzung der Verfahren zur Datenvorverarbeitung ist nicht zielführend. Wird die Datenvorverarbeitung als Basis für verschiedene Verfahren verwendet, kann eine Anwendung in Betracht gezogen werden. Generell ist eine vollständig unabhängige und ohne Austausch des Kontextes und Wissens betriebene Bearbeitung wegen des entstehenden Mehraufwandes für die Fachexperten nicht sinnvoll.

Im Zuge des Mutations-Tests werden Fehler in ein richtiges System eingegeben und dahingehend geprüft, ob diese Fehler gefunden werden (siehe Kapitel 3.3). Daher können mit dem Mutations-Test Testverfahren hinsichtlich ihrer Eignung im konkreten Fall

untersucht werden. Somit kann dieses Verfahren dazu eingesetzt werden, zu überprüfen, ob sich andere Verfahren für den Einsatz im Data Mining eignen. Das Verfahren selbst ist jedoch nicht für den Einsatz im Data Mining geeignet. Mit der Forderung eines realen Systems disqualifiziert sich der Mutations-Test für eine Anwendung.

Der Regressionstest überprüft Teileobjekte des gesamten Objektes. Dazu werden die zu testenden Objekte sowohl vor als auch nach den vorgenommenen Änderungen mit denselben Bedingungen getestet und anschließend verglichen (siehe Kapitel 3.3). Die aufgestellten Kriterien erfüllt der Regressionstest. Daher kann dieser im Data Mining eingesetzt werden. Beispielsweise kann eine Änderung die Veränderung eines Parameters darstellen.

In Kapitel 3 wurden verschiedene V&V-Techniken vorgestellt, die in der Simulation zum Einsatz kommen. Werden die vorgestellten Techniken betrachtet, fällt auf, dass einige dieser Techniken für die Anwendung im Data Mining nicht geeignet sind. Visuelle Techniken sind für den Einsatz in der Simulation nützlich, um Prozesse verständlich darstellen zu können. Im Data Mining ist eine visuelle Darstellung nicht möglich, weil keine dynamische Prozessnachbildung erfolgt. Stattdessen werden statische Modelle verwendet. Die Qualität der Datenvorverarbeitung und -bereitstellung ist in Data-Mining-Vorgehensmodellen ein wichtiges Kriterium. Jedoch erfordert die Durchführung solcher Maßnahmen häufig Kontextwissen. Diese Techniken können häufig nicht vollständig automatisiert werden.

Bei allen informalen Techniken wurde festgestellt, dass diese nicht vollständig automatisierbar sind. Auch wenn diese für einen Einsatz im Data Mining geeignet sind, werden diese hier nicht weiter betrachtet, da eine automatische Durchführung gefordert wird.

Die Gemeinsamkeit der Alpha-, Beta-, und Akzeptanztests ist die Durchführung des Systems. Dadurch sollen mögliche Fehler im Modell aufgedeckt werden (siehe Kapitel 3.3). Der Akzeptanztest erfüllt alle Kriterien und kann im Data Mining eingesetzt werden. Durch den Einsatz eines Akzeptanztestes wird jedoch lediglich überprüft, ob das richtige Data-Mining-Verfahren angewendet wird. Eine Prüfung, ob dieses auch richtig angewendet wird, findet jedoch nicht statt. Da eine automatische Ausführung dieses Tests nicht möglich ist, wird dieser Test nicht weiter vertieft. Tests während der Entwicklungsphase wie Alpha- und Betatest sind auch im Zuge des KDD sinnvoll und notwendig. Allerdings können auch diese nicht automatisiert werden und scheiden daher aus.

Bei den Tests von Teilmodellen wird das Modell in mehrere Teilmodelle unterteilt, welche einzeln getestet werden. Im KDD erscheint dieser Test vor allem in der Datenvorbereitung sinnvoll. Besonders das Prüfen nach Änderungen der Attributsauswahl erscheint beispielsweise sinnvoll. Jedoch kann der Test von Teilmodellen nicht automatisiert werden und findet somit für die weiteren Untersuchungen keine Beachtung.

Wie in der Simulation können auch im Data Mining statistische Techniken zur Validierung des Modells und der Ein- und Ausgangsdaten verwendet werden. Das Modell kann beispielsweise mit der Kreuzvalidierung überprüft werden. Als Alternative bieten sich die vorgestellten statistischen Techniken im Data Mining an. Die in Kapitel 3.4 vorgestellten statistischen V&V-Techniken erfüllen die aufgeführten Kriterien für einen Einsatz im Data Mining. Allerdings können keine Ausgangsdaten überprüft werden, da weder ein Modell noch Ausgangsdaten aus der Vergangenheit zur Verfügung stehen. Trotzdem erscheint ein Einsatz dieser statistischen Verfahren im Data Mining sinnvoll. Mit Hilfe der Korrelationsanalyse können beispielsweise Eingabedaten, welche in einer starken Beziehung zueinander stehen, in der weiteren Analyse berücksichtigt werden.

Die V&V-Technik der Validierung von Vorhersagen prüft das zu prüfende Modell mit Hilfe der im Modell getroffenen Vorhersagen gegenüber dem realen System. Ein ähnliches Vorgehen findet sich im Data Mining, wo Muster aus den Daten gefunden werden und diese Muster eine Regel definieren. Da jedoch Daten aus der Vergangenheit verwendet werden, scheidet die Technik des Vergleichs mit aufgezeichneten Daten für das Data Mining aus.

Die Technik des Vergleichs mit anderen Modellen setzt mehrere Modelle des realen Systems voraus. Diese Technik ähnelt somit dem Back-to-Back-Tests. Da jedoch Modelle eines realen Systems gefordert werden, erscheint diese Technik für den Einsatz im Data Mining ungeeignet. Auch der Vergleich mit aufgezeichneten Daten ist für die Verwendung im Data Mining nicht geeignet, da ebenfalls ein reales System gefordert wird.

Tabelle 12 zeigt eine Übersicht mit allen V&V-Techniken, welche in diesem Kapitel untersucht wurden. Anhand der Tabelle 12 wird deutlich, dass auf Grundlage der durchgeführten Analysen von den untersuchten V&V-Techniken lediglich fünf Techniken sowohl automatisierbar sind als auch im Data Mining eingesetzt werden können.

Tabelle 12: Einsatz von V&V-Techniken im Data Mining

	V&V-Technik	Einsatz im Data Mining
Softwareentwicklung	Kontrollflussorientierte (White-Box)	–
	Funktionale Äquivalenzklassenbildung (Black-Box)	–
	Grenzwertanalyse (Black-Box)	–
	Test spezieller Werte (Black-Box)	+
	Back-to-Back	–
	Mutationstest	–
	Regressionstest	+
Simulation	Korrelationsanalyse	+
	Chi-Quadrat-Test	+
	Regressionsanalyse	+
	Validierung von Vorhersagen	–
	Vergleich mit anderen Modellen	–
	Vergleich mit aufgezeichneten Daten	–

Legende:

„+“: Eignet sich für den Einsatz im Data Mining

„–“: Eignet sich nicht für den Einsatz im Data Mining

5.2 Übertragung ins Data Mining

In diesem Kapitel wird untersucht, wie die V&V-Techniken ins Data Mining übertragen werden können. Dabei werden lediglich diejenigen Techniken aus Kapitel 5.1 betrachtet, die die aufgestellten Kriterien erfüllen und die Techniken, die bereits im Data Mining zum Einsatz kommen und automatisierbar sind. Dies sind zum einen aus der Softwareentwicklung

- der Test spezieller Werte und
- der Regressionstest

und zum anderen aus der Simulation

- die Korrelationsanalyse,
- der Chi-Quadrat-Test und
- die Regressionsanalyse.

Bisher wurde lediglich festgelegt, dass die einzelnen V&V-Techniken im Data Mining eingesetzt werden können. Allerdings besteht ein Data-Mining-Prozess aus vielen Prozessschritten, die sich teilweise erheblich unterscheiden. Dabei kann nicht jede V&V-Technik beliebig in jedem Prozessschritt eingesetzt werden. Auch die V&V-Techniken, welche bereits im Data Mining eingesetzt werden, müssen dahingehend untersucht werden. Im Folgenden werden die V&V-Techniken, die aufgrund der vorherigen Analyse im Data Mining eingesetzt werden, dahingehend untersucht, in welchem Prozessschritt im MESC sie eingesetzt werden können.

Zunächst werden die einzelnen Phasen im MESC-Vorgehensmodell hinsichtlich ihrer Inputs und Outputs analysiert. Denn die V&V-Techniken überprüfen bestimmte Formate. Beispielsweise kann der Test spezieller Werte keine Fragestellungen untersuchen. In der ersten Phase der Aufgabendefinition wird die Aufgabe genau festgelegt (siehe Kapitel 2.2). Eine Überprüfung eines solchen Dokuments leistet keiner der fünf V&V-Techniken. Im Zuge der zweiten Phase werden bei der Datenauswahl aus vielen Daten die relevanten ausgewählt. Diese Daten werden anschließend für das Data-Mining-Verfahren aufbereitet. An dieser Stelle könnten statistische Techniken (Regressionsanalyse, Korrelationsanalyse und Chi-Quadrat-Test) eingesetzt werden. Durch diese Techniken werden die Zusammenhänge in den Datensätzen aufgedeckt, was zur Vorbereitung auf das Data-Mining-Verfahren hilft, indem Fehler vermieden werden. Ein Einsatz der hier betrachteten V&V-Techniken in der dritten Phase, der Datenaufbereitung, ist nicht möglich.

In der Vorbereitung des DM-Verfahrens (Phase 4) wird eine Auswahl für ein DM-Verfahren getroffen und die Daten werden daraufhin vorbereitet. Hier könnte die Kreuzvalidierung eingesetzt werden. Bei einer Anzahl von unter 1000 Datensätzen eignet sich diese Methode hervorragend um das Data-Mining-Verfahren gezielt anzupassen und zu justieren. Die Holdoutmethode kann im Gegensatz zur Kreuzvalidierung bei mehr als 1000 Datensätzen verwendet werden. Sie kann ebenfalls in der Vorbereitung des DM-Verfahrens angewendet werden. Auch die Bootstrapping-Validierung kann in dieser Phase des DM-Prozesses eingesetzt werden.

Im Hauptschritt, der Anwendung des DM-Verfahrens (Phase 5), kann eine grundsätzliche Analyse schwer vorgenommen werden, ob eine V&V-Technik angewendet werden kann, hängt maßgeblich mit dem verwendeten Data-Mining-Verfahren zusammen. Daher wird nachfolgend in diesem Kapitel näher auf die Data-Mining-Verfahren eingegangen. In den

Phasen nach dem Data-Mining-Verfahren (Phase 6 und Phase 7) erscheint keine V&V-Technik einsetzbar zu sein.

Auffallend ist, dass der Großteil der Techniken im MESC-Vorgehensmodell in den Phasen der Vorbereitung auf das Data-Mining-Verfahren angewendet werden kann. Ziel dieser Arbeit ist jedoch eine Untersuchung im Data Mining. Daher wird im Folgenden eine gezielte Untersuchung vorgenommen ob und wo die V&V-Techniken in Data-Mining-Verfahren angewendet werden können. Dabei werden die Data-Mining-Verfahren Clusterbildung, Klassifikation und Assoziationsanalyse untersucht.

Bei der Betrachtung des Data-Minings-Verfahren Clusterbildung, welches in Kapitel 2.1 vorgestellt wurde, fällt auf, dass dort einige Überprüfungen stattfinden können. Wie in Kapitel 2.1 dargestellt, kann das Vorgehen in der Clusterbildung in drei Schritte gegliedert werden. Im ersten Schritt müssen geeignete Variablen zum Vergleich ausgewählt werden. Um eine geeignete Wahl zu gewährleisten, sollte an dieser Stelle eine Untersuchung der Unabhängigkeit der Variablen durchgeführt werden. Auf Grundlage dieser Auswahl wird später das Clustering vorgenommen. Durch eine zu große Abhängigkeit verschiedener Variablen zueinander wird die Clusterbildung verfälscht. Deshalb muss bei der Auswahl der Variablen darauf geachtet werden, dass die Variablen, die bei der Proximitätsbestimmung als Grundlage dienen, keine großen Abhängigkeiten aufweisen. So kann eine ungünstige Auswahl verhindert werden. Würden beispielsweise Variable 2 und Variable 3 eine Abhängigkeit zu Variable 1 aufweisen und Variable 4 und Variable 5 gänzlich unabhängig sein, erscheint eine Auswahl der Variablen 1, 4 und 5 am geeignetsten. Eine Auswahl von Variable 1 und Variable 2 oder 3 könnte zu einer verzerrten und nicht optimalen Clusterbildung führen. Daher ist die Variablenauswahl essenziell für das Verfahren der Clusterbildung. Um in diesem Verfahrensschritt eine Überprüfung der Auswahl vorzunehmen, bieten sich die V&V-Techniken Korrelationsanalyse und Chi-Quadrat-Test an. Beide Techniken untersuchen die Abhängigkeit zweier Variablen zueinander (siehe Kapitel 3.4). Kann keine Abhängigkeit zwischen den betrachteten Variablen nachgewiesen werden, erscheint die Auswahl unter diesem Gesichtspunkt als validiert und verifiziert. Zu beachten ist, dass in diesem Fall die Anwendung der einen Technik die Anwendung der jeweils anderen Technik expliziert nicht ausschließt. Da die Korrelationsanalyse lineare Zusammenhänge und der Chi-Quadrat-Test nicht lineare Abhängigkeiten aufdeckt (siehe Kapitel 3.4), erscheint ein paralleler Einsatz beider Techniken als sinnvoll, und vielleicht sogar als notwendig.

Der Test spezieller Werte (siehe Kapitel 3.3) kann im zweiten Schritt des Clustering der Proximitätsbestimmung und im dritten Schritt der Clusterbildung eingesetzt werden. Da im Schritt der Proximitätsbestimmung Distanzmaß und Ähnlichkeitsmaß berechnet werden, erscheint hier der Test spezieller Werte als sinnvoll. Je nach dem verwendeten Proximitätsmaß (siehe Kapitel 2.1), erfolgt eine unterschiedliche Berechnung der Maße. Dabei sollten insbesondere die schwierig erscheinenden Fälle überprüft werden. Durch die Sicherstellung der fehlerfreien Berechnung der Distanz- und Ähnlichkeitsmaße besonders bei den speziellen Werten kann die Proximitätsbestimmung flächendeckend als fehlerfrei angenommen werden. Auch im letzten Schritt des Clustering erscheint ein Einsatz des Tests spezieller Werte als sinnvoll. In diesem Schritt werden die Cluster gebildet und die Objekte einem Cluster zugeteilt. Ein Anwendungsfall wäre hier beispielsweise in der Zuteilung der Objekte zu einem Cluster. Liegen alle für die Einordnung ausgewählten Variablen exakt zwischen zwei Clustern, erscheint eine Zuordnung nicht eindeutig zu sein. Besonders solche Fälle könnten mit dem Test spezieller Werte überprüft werden. Damit können unsaubere Zuordnungen identifiziert und genauer betrachtet werden.

Der Regressionstest verlangt eine Durchführung mehrerer Programmversionen mit denselben zu Grunde liegenden Daten. In der Clusteranalyse kann diese Technik angewendet werden, indem verschiedene Clusterverfahren (siehe Kapitel 2.1) angewendet werden. Anschließend wird verglichen, worin sich die erstellten Cluster in den verschiedenen Verfahren unterscheiden. Äquivalent können unterschiedliche Proximitätsmaße als Basis für die Berechnungen der Distanz- und Ähnlichkeitswerte miteinander verglichen werden. Mit der Durchführung verschiedener Clusterverfahren bzw. Proximitätsmaße können die Ergebnisse miteinander abgeglichen und Unstimmigkeiten aufgedeckt werden. Allerdings sind dazu mehrere Durchläufe des Data-Mining-Verfahrens nötig, was unter Umständen viel Zeit in Anspruch nimmt.

Die Regressionsanalyse identifiziert in erster Linie die Ursachen von Zusammenhängen zwischen Variablen (siehe Kapitel 3.4). Zwar ist eine Untersuchung von Zusammenhängen zwischen Variablen durch die Korrelationsanalyse und den Chi-Quadrat-Test sinnvoll, aber der Ursprung dieser Zusammenhänge ist in der Clusteranalyse bedeutungslos. Daher findet diese Technik in der Clustering keine Anwendung und ein Einsatz der Regressionsanalyse im Clustering erscheint als nicht sinnvoll.

Bei der Klassifikation werden den Objekten Klassen zugeordnet (siehe Kapitel 2.1). Dabei unterscheiden sich die vier vorgestellten Klassifizierungsverfahren deutlich. Daher muss für jede V&V-Technik jedes Klassifizierungsverfahren genau betrachtet und hinsichtlich eines möglichen Einsatzes analysiert werden. Untersucht werden die vier Verfahren, die in Kapitel 2.1 vorgestellt werden, und die fünf V&V-Techniken, die in Kapitel 5.1 als einsetzbar im Data Mining eingestuft worden sind.

In der Klassifizierung mit Entscheidungsbäumen erscheint die Entscheidungsfindung an jedem Knoten als mögliche Fehlerquelle. Durch die Entscheidung für einen anschließenden Pfad wird die Klassifizierung beeinflusst. Deshalb müssen die Fehler gerade an dieser Stelle minimiert werden. Mit dem Test spezieller Werte können an dieser Stelle Fehler bei problematischen und kritischen Werten identifiziert und unterbunden werden. Der Regressionstest kann angewendet werden, um die Klassifizierung der Objekte zu überprüfen. Dazu sind jedoch mehrere unterschiedliche Entscheidungsbäume nötig, die zu erstellen sind. Grundsätzlich kann der Regressionstest jedoch eingesetzt werden. Im Gegensatz dazu sind Zusammenhänge zwischen Merkmalen eines Objektes uninteressant beim Einsatz von Entscheidungsbäumen, da diese für die Zuteilung eines Objekts in eine Klasse keine Rolle spielen. Daher finden die Korrelationsanalyse, der Chi-Quadrat-Test und die Regressionsanalyse keine Anwendung.

Auch bei der Bayes-Klassifikation spielen Zusammenhänge zwischen Merkmalen bei Objekten keine Rolle. Durch die Zuordnung zu Klassen mittels Merkmalsausprägungen (siehe Kapitel 2.1) finden die Techniken zur Identifizierung von Zusammenhängen (Korrelationsanalyse, Chi-Quadrat-Test und Regressionsanalyse) keine Anwendung. Der Test spezieller Werte kann eingesetzt werden, um die Merkmalsausprägungen einwandfrei zu überprüfen, indem besondere Werte explizit betrachtet werden. Der Regressionstest könnte eingesetzt werden, indem die Trainingsdaten mehrere Male verschieden gewählt werden. So ergibt sich die Möglichkeit (vor allem bei einer kleineren Menge an Daten), dass trotz derselben zu Grunde liegenden Daten unterschiedliche Ergebnisse miteinander verglichen werden können. Dies bietet die Möglichkeit, die verschiedenen Klassifizierungen miteinander zu vergleichen und zu diskutieren.

Die Anwendung der Techniken im nearest-neighbour-Verfahren erscheint schwierig. Der Regressionstest kann wieder mit verschiedenen Wahlen der Trainingsdaten erfolgen. Daraufhin können die Ergebnisse bezüglich der Klassenzuteilung verglichen werden. Die Techniken Korrelationsanalyse, Chi-Quadrat-Test und Regressionsanalyse können nicht angewendet werden. Die Ähnlichkeit von zwei Objekten spielt eine wichtige Rolle in

dieser Technik, aber nicht der Zusammenhang zwischen zwei unterschiedlichen Merkmalen. Mit dem Test spezieller Werte besteht die Möglichkeit, die Ähnlichkeiten der Merkmale zu überprüfen.

In der Klassifikation mit neuronalen Netzen kann aufgrund der Aufteilung unter anderem in Trainingsdaten der Regressionstest angewendet werden. Wie beim nearest-neighbour-Verfahren kann durch die veränderte Auswahl der Testdaten die Richtigkeit und Genauigkeit überprüft werden. Beim nearest-neighbour-Verfahren können so die Kantengewichte verglichen und kontrolliert werden (siehe Kapitel 2.1). Die Techniken, die auf die Zusammenhänge zwischen unterschiedlichen Merkmalen eingehen, finden in diesem Verfahren keine Anwendung. Im Gegensatz dazu kann der Test spezieller Werte angewendet werden, um schwierige Zuordnungen zu kontrollieren.

Zusammenfassend für das Data-Mining-Verfahren Klassifikation ist festzuhalten, dass eine Übereinstimmung beim Einsatz der Techniken vorliegt, obwohl die Klassifikationsverfahren sehr unterschiedlich sind. So kann festgehalten werden, dass der Test spezieller Werte und der Regressionstest in der Klassifikation eingesetzt werden können.

Anders als bei der Clusterbildung und der Klassifikation können in der Assoziationsanalyse die Techniken, die den Zusammenhang von Merkmalen untereinander untersuchen, eingesetzt werden. Die Assoziationsanalyse trifft Vorhersagen auf Basis der vorliegenden Daten (siehe Kapitel 2.1). Diese Vorhersagen können mit Hilfe der Korrelationsanalyse und des Chi-Quadrat-Testes überprüft werden. Gibt die Assoziationsanalyse als Ergebnis einer Warenkorbanalyse aus, wenn Produkt 1 gekauft wird, dass mit 60 prozentiger Wahrscheinlichkeit auch Produkt 2 gekauft wird, sollte dies auf Korrektheit kontrolliert werden. Zum Einsatz kommen in der Assoziationsanalyse der Support und die Konfidenz (siehe Kapitel 2.1). Dazu bietet sich der Einsatz der Korrelationsanalyse und des Chi-Quadrat-Tests als ergänzende Kontrollinstanz an. In den meisten Fällen ist allerdings eine Anwendung beider Techniken notwendig, da die Korrelationsanalyse lineare und der Chi-Quadrat-Test nicht lineare Zusammenhänge identifiziert (siehe Kapitel 3.4). Deshalb ist ein Einsatz beider Techniken zwingend erforderlich, um alle Zusammenhänge identifizieren zu können. Auch die Regressionsanalyse untersucht die Zusammenhänge zwischen verschiedenen Merkmalen. Diese Technik eignet sich dazu, die Ergebnisse der Assoziationsanalyse zu kontrollieren. Insbesondere die Eigenschaft, dass die Regressionsanalyse zukünftige

Werte anhand von anderen Variablen prognostizieren kann, verspricht eine gute Kontrolle.

Im Gegensatz dazu kann der Regressionstest nicht im Zuge der Kontrolle der Assoziationsanalyse eingesetzt werden. Die Assoziationsanalyse folgt einer Vorgehensweise und berechnet den Support und die Konfidenz. Auch erfolgt keine Unterteilung in Trainingsdaten, sodass keine Vergleiche mit Hilfe des Regressionstests vorgenommen werden können. Auch der Test spezieller Werte scheint in der Assoziationsanalyse nicht einsetzbar zu sein. Da lediglich die Zusammenhänge zwischen den Merkmalen identifiziert werden, existieren keine ‚speziellen‘ Werte innerhalb der Analyse. Am Beispiel Warenkorb kann dies erklärt werden. Wenn beispielsweise ein Produkt 1 nicht gekauft und somit mit ‚Null‘ beschrieben wird, hat dies den gleichen Einfluss auf die Analyse, als wenn das Produkt einmal gekauft wird. Genauso können negative und nicht ganze Zahlen berücksichtigt werden. Daher würde der Test spezieller Werte bei der Assoziationsanalyse immer eine positive Rückmeldung geben. Denn ein Fehler wird nicht erwartet und tritt auch nicht ein.

Tabelle 13 zeigt eine Übersicht, welche V&V-Technik in welchem Data-Mining-Verfahren angewendet werden kann. Auffällig ist, dass keine Technik in allen Data-Mining-Verfahren eingesetzt werden kann. Obwohl in Kapitel 5.1 anhand von Kriterien untersucht wurde, welche der Techniken im Data Mining zum Einsatz kommen könnten, unterscheiden diese sich deutlich.

Tabelle 13: Anwendung der V&V-Techniken in Data-Mining-Verfahren

	Clusterbildung	Klassifikation	Assoziationsanalyse
Test spezieller Werte	+	+	–
Regressionstest	+	+	–
Korrelationsanalyse	+	–	+
Chi-Quadrat-Test	+	–	+
Regressionsanalyse	–	–	+

Legende:

„+“: V&V-Technik kann im Data-Mining-Verfahren eingesetzt werden

„–“: V&V-Technik kann nicht im Data-Mining-Verfahren eingesetzt werden

Anhand der Ergebnisse wird deutlich, dass die unterschiedlichen Data-Mining-Verfahren sich bezüglich eines Einsatzes von V&V-Techniken erheblich unterscheiden. Bei der Betrachtung der Verfahren der Klassifikation und der Assoziationsanalyse, wird erkennbar, dass die Techniken, welche in der Klassifikation eingesetzt werden können, in der Assoziationsanalyse nicht eingesetzt werden können. Und die Techniken, die nicht eingesetzt werden können, können in der Assoziationsanalyse eingesetzt werden.

Dies legt nahe, dass eine grundsätzliche Analyse zum Einsatz von V&V-Techniken im Data Mining, wie in Kapitel 5.1 durchgeführt, nicht ausreichend ist. Zwar wurde gezeigt, dass diese Techniken automatisierbar sind und im Data Mining eingesetzt werden können. Jedoch zeigen die Untersuchungen in diesem Kapitel, dass deutliche Unterschiede zwischen den einzelnen Data-Mining-Verfahren existieren. Um diese Unterschiede zwischen den Data-Mining-Verfahren genauer abzugrenzen, ist eine genaue Prüfung der V&V-Techniken in jedem Data-Mining-Verfahren einzeln durchzuführen. Durch die Analyse aller betrachteten V&V-Techniken hinsichtlich ihres Automatisierungspotentials und einer möglichen Einsetzbarkeit in jedem einzelnen Verfahren können außerdem gegebenenfalls auch Techniken eingesetzt werden, welche in den allgemeinen Untersuchungen in dieser Arbeit ausgeschlossen wurden.

6 Praktische Anwendung mit MESC

In diesem Kapitel wird auf Grundlage der Ergebnisse des vorherigen Kapitels 5, in dem untersucht wurde, welche V&V-Techniken in welchem Data-Mining-Verfahren zum Einsatz kommen können, beispielhaft eine V&V-Technik in einem Data-Mining-Verfahren angewendet. Dazu werden Testfälle entwickelt, die an einem Datensatz im Zuge eines Data-Mining-Verfahrens durchgeführt werden. Nach Durchführung der Testfälle wird aufgezeigt und dargestellt, wie diese V&V-Technik automatisiert werden kann. Zum Abschluss des Kapitels wird ein Fazit gezogen, dass die Fragestellungen dieser Arbeit beantworten soll.

6.1 Anwendung der V&V-Techniken im MESC-Vorgangmodell

Im vorherigen Kapitel wurden einige V&V-Techniken eingestuft, um in verschiedenen Data-Mining-Verfahren angewendet zu werden (siehe Kapitel 5.2). Die bisher theoretisch durchgeführten Untersuchungen sollen nun in der Praxis angewendet werden. Die Möglichkeit zum Einsatz und Anwendung von V&V-Techniken besteht im Rahmen des MESC-Vorgehensmodells (siehe Kapitel 2.2). Dazu wurden in Kapitel 5.2 die V&V-Techniken hinsichtlich ihres Einsatzes in den Phasen des MESC-Vorgehensmodells untersucht. Des Weiteren besteht die Möglichkeit, abhängig vom angewendeten Data-Mining-Verfahren, verschiedene V&V-Techniken anzuwenden (siehe Tabelle 13). Darauf aufbauend wird im Folgenden beispielhaft der Test spezieller Werte im Clustering durchgeführt. Als Clustering-Verfahren wurde die hierarchische Clusteranalyse ausgewählt. Zur Hilfe wurde die Open-Source-Cloud basierte Plattform für interaktive Lern- und Data-Mining-Workflows ClowdFlows (Version September 2018) verwendet (Kranjc 2017). Mit Hilfe von ClowdFlows wurde die hierarchische Clusteranalyse (siehe Kapitel 2.1) am Datensatz Iris angewendet. Der Datensatz Iris besteht aus 150 Objekten mit jeweils fünf Merkmalen (FISHER 1936).

Tabelle 14: Auszug aus dem Datensatz Iris

sepal length	sepal width	petal length	petal width	iris
4.300	3.000	1.100	0.100	Iris-setosa
4.400	2.900	1.400	0.200	Iris-setosa
4.400	3.000	1.300	0.200	Iris-setosa
...

In Tabelle 14 sind drei Objekte des Datensatzes Iris beispielhaft aufgeführt. Der Datensatz Iris dokumentiert Beobachtungen und Messungen an Schwertlilien. Dabei geben die ersten vier Merkmale die Höhen und Breiten des Kelchblattes sowie des Kronblattes wieder. Das fünfte Attribut zeigt die Klasseneinteilung der Objekte in die drei Klassen. Für die Clusteranalyse gilt der Datensatz Iris als Paradebeispiel und wird häufig für Untersuchungen verwendet.

Auf Grundlage des Datensatzes und der erfolgten Clustering wurde anschließend der Test spezieller Werte durchgeführt. Dazu wurden nacheinander verschiedene spezielle Werte in Form von Testfällen aufgezeichnet. Dabei wird nicht nur von speziellen Werten gesprochen, wenn Zahlen und Buchstaben gemeint sind, sondern auch Sonderzeichen und leere Zellen. Denn auch nach der Datenvorbereitung inklusive einer angewendeten V&V-Technik können solche Fälle auftreten. An die Testfälle wurden Erwartungen und Ergebnisse geknüpft, welche der Test ergeben sollte. Beispielsweise wurde bei einem Testfall erwartet, dass die zu berechnende Distanz zu ‚0‘ korrekt angegeben wird. Mit Hilfe dieser Testfälle wurde der Ursprungsdatensatz Iris für jeden Testfall verändert. Anschließend wurde für jeden Testfall ein erneutes Clusterverfahren durchgeführt und ermittelt, inwiefern die Manipulation der Daten die Distanzen der Objekte und damit die Clusterbildung beeinflusst. Dabei wurde dokumentiert, inwiefern die erwarteten Ergebnisse eingetroffen sind oder welche Abweichungen der Testfall lieferte.

Tabelle 15 zeigt die zwölf durchgeführten Testfälle und erläutert diese kurz. Zusätzlich gibt die Tabelle einen Überblick, ob die erwarteten Ergebnisse eingetroffen sind und ob der Test erfolgreich war. Zum einen wurden spezielle Zahlen wie null und negative Werte getestet, um sicher zu stellen, dass auch hierbei korrekte Distanzen zwischen den Attributen berechnet wurden. Des Weiteren wurden Sonderzeichen daraufhin überprüft, ob diese korrekt eingeordnet werden.

Im Folgenden werden die einzelnen Testfälle dargestellt und erläutert. Die Testfälle 1a und 1b testen das Verhalten des Data-Mining-Verfahrens bei ‚0‘. Während Test 1a Null in einem Merkmal einsetzt, in dem ausschließlich weitere Zahlen vorkommen, wird bei 1b eine Null in ein Merkmal getestet, wo ausschließlich Buchstaben und Zeichenketten vorkommen. Beide Tests liefern das erwartete Ergebnis. Im Testfall 1a wird jeweils die Distanz zwischen ‚0‘ und dem Wert des betrachteten Attributs berechnet. Bei Testfall 1b wird ‚0‘ als Zeichen erkannt und liefert beispielsweise die Distanz von ‚a‘ zu ‚0‘ von eins.

Tabelle 15: Entwickelte Testfälle

Testfall	Testwert	Erklärung	Erwartung an Testergebnis	IST-Test-Ergebnis	Test erfolgreich?
1 a	0 1 2 3	Prüfen auf Besonderheit bei „0“ bei Zahlen	Berechnung der Distanz zu „0“	Die Distanz zu „0“ werden berechnet	Ja
1 b	A B 0 C	Prüfen auf Besonderheit bei „0“ bei Zeichenketten	„0“ wird als Zeichenkette erkannt	„0“ wird als Zeichenkette erkannt	Ja
2	+1 2 3 4 5	Prüfen auf Besonderheit bei „+“	Berechnung der Distanz zu „1“	Die Distanz zu „1“ werden berechnet	Ja
3	-1 1 2 3 4	Prüfen auf Besonderheit bei „-“	Berechnung der Distanz zu „-1“	Die Distanz zu „-1“ werden berechnet	Ja
4	9999999 99999999	Prüfen auf Besonderheit bei großen Zahlen	Abscheiden der überschüssigen Ziffern	Ab acht Stellen wird die Zahl gerundet; Rundungsfehler treten auf	Nein
5	0.1 0.2 0.3 0.4	Prüfen auf Besonderheit bei „.“	Transformation von „Komma“ zu „Punkt“	Komma wird zu Punkt transformiert	Ja
6	2.900555 2.900111	Prüfen auf Besonderheit bei mehr als drei Nachkommastellen	Runden auf drei Nachkommastellen	Die Zahlen werden auf drei Nachkommastellen gerundet	Ja
7	Null 0 1 2 3 4	Prüfen ob „Null“ als „0“ erkannt wird	Keine Transformation zu „0“	Keine Transformation von „Null“ zu „0“	Ja

Testfall	Testwert	Erklärung	Erwartung an Testergebnis	IST-Test-Ergebnis	Test erfolgreich?
8 a	,leer' 1 2 3 4	Prüfen von Besonderheiten bei leeren Feldern bei Zahlen	Für ein leeres Feld wird „0“ eingesetzt	Bei Merkmal 1: Zeile wird gelöscht Bei weiteren Merkmalen „?“ wird eingesetzt	Nein
8 b	,leer' A B C D	Prüfen von Besonderheiten bei leeren Feldern bei Zeichenketten	Ein leeres Feld wird als eigene Zeichenkette wahrgenommen	Bei Merkmal 1: Zeile wird gelöscht Bei weiteren Merkmalen „?“ wird eingesetzt	Nein
9	0 ? 2 4 A ? B C	Testen ob „?“ als normales Zeichen erkannt wird	„?“ wird als normales Zeichen betrachtet	Wird nicht als normales Zeichen betrachtet Distanz zwischen Attribut 1 und Attribut 1 ungleich 0	Nein
10	0 ! 2 4 A ! B C	Prüfen ob sich „!“ verhält wie „?“	„!“ wird als normales Zeichen betrachtet	„!“ wird als normales Zeichen betrachtet	Ja
11	ä a b c d ß a b c € a b c	Prüfen ob „ä“, „ß“ und „€“ als normale Zeichen erkannt werden	„ä“, „ß“ und „€“ werden als normale Zeichen betrachtet	„ä“, „ß“ und „€“ werden als normale Zeichen betrachtet	Ja
12	A a b c	Groß- und Kleinschreibung prüfen	Unterscheidung zwischen „A“ und „a“	Unterscheidung zwischen „A“ und „a“	Ja

In den Testfällen 2 und 3 wird auf Besonderheiten bei ‚+‘ und ‚-‘ geprüft. Dazu werden in ein Merkmal Zahlen mit dem jeweiligen Vorzeichen eingesetzt. Dabei liefern beide Testfälle die erwarteten Ergebnisse. Während ‚+‘ korrekterweise weggelassen wird und die Distanzen korrekt berechnet werden, werden bei Testfall 3 die Distanzen richtig zum negativen Wert berechnet.

Testfall 4 untersucht das Verhalten bei großen Zahlen. Hierbei wird überprüft, ob die überschüssigen Ziffern abgeschnitten werden oder ob keine obere Grenze besteht. In diesem Fall werden die Zahlen nicht abgeschnitten, sondern gerundet, sobald diese aus acht oder mehr Ziffern besteht. Dabei fällt insbesondere auf, dass desto ungenauer gerundet wird, je größer die Zahl wird. Beispielsweise wird bei 9.999.999.999 auf 10.000.000.000 aufgerundet, während bei einer Stelle mehr auf 99999997952 abgerundet wird. Dies führt zu Ungenauigkeiten in der Berechnung der Distanz zwischen den Objekten.

Dezimalzahlen werden in diesem Datensatz mit einem Punkt geschrieben. Bei Testfall 5 wird getestet, ob ein Unterschied auftritt, wenn statt eines Punktes ein Komma gesetzt wird. Der Test zeigt, dass hierbei kein Unterschied auftritt. Das gesetzte Komma wird automatisch zu einem Punkt transformiert. Daraufhin werden anschließend die Distanzen richtig berechnet und angegeben.

Testfall 6 überprüft das Verhalten bei Dezimalzahlen, die mehr als drei Nachkommastellen aufweisen. Wie erwartet werden die Zahlen auf drei Nachkommastellen gerundet und nicht die überschüssigen Stellen abgeschnitten. Auch die Berechnung der Distanzen erfolgt ausnahmslos korrekt.

Im nächsten Testfall wird getestet, ob das Wort ‚Null‘ als ‚0‘ erkannt und transformiert wird. Ziel des Tests ist zu prüfen, ob mehrere Zeichen als Wort erkannt werden oder lediglich als Zeichenkette ohne Bedeutung. Wie erwartet findet keine Transformation statt, sodass hier lediglich eine Zeichenkette erkannt wird.

In den nächsten beiden Testfällen wird die Ausführung des Verfahrens bei leeren Zellen geprüft. Dazu wurden sowohl in Testfall 8a als auch in 8b Merkmalsausprägungen eines Attributs nicht ausgefüllt, sodass die Zelle in der Tabelle leer ist. Testfall 8a betrachtet dabei leere Zellen bei einem Merkmal, in dem ausschließlich Zahlen vorkommen, während Testfall 8b das Verhalten innerhalb von Merkmalen mit Zeichenketten betrachtet. Bei Testfall 8a wird erwartet, dass automatisch ‚0‘ eingesetzt wird und die Distanz zu ‚0‘ berechnet wird. Innerhalb von Testfall 8b wird die Berechnung der Distanz

zu einer bisher nicht verwendeten Zeichenkette erwartet. Dies hat zur Folge, dass die anderen Attribute, eine Distanz von Eins zu diesem Attribut aufweisen. Allerdings liefern beide Testfälle dieselben Ergebnisse. Befindet sich die leere Zelle in der ersten Spalte, so wird das gesamte Attribut gelöscht. Des Weiteren werden keine Distanzen dieses Attributs und zu diesem Attribut berechnet. Findet sich die leere Zelle in einer anderen Zelle als in der ersten Spalte, so wird ein ‚?‘ eingesetzt. Um das Fragezeichen näher zu untersuchen, wurde in Testfall 9 geprüft, ob ein ‚?‘ als normales Zeichen erkannt und verwendet wird. Jedoch ergibt die Untersuchung, dass das Fragezeichen nicht als normales Zeichen erkannt wird. Vielmehr stellt das ‚?‘ das Programm vor Probleme. Die Distanzen zwischen zwei Attributen sind nicht nachzuvollziehen. Während die Distanz in verschiedenen Tests zwischen zwei Attributen bei Eins liegen müsste, liegt diese wesentlich höher. Aus weiterführenden Tests wird nicht ersichtlich, wie die Distanz von und zu Fragezeichen berechnet wird. Ein Fehlverhalten des Verfahrens wird bei der Betrachtung der Distanz zwischen Attribut 1 (Attribut bei dem ein Merkmal mit ‚?‘ ausgewiesen ist) zu sich selbst deutlich. Ein Attribut sollte zu sich selbst, da alle Merkmale gleich sind, zu jeder Zeit die Distanz von Null aufweisen. Jedoch weist die Distanz von Attribut 1 zu sich selbst den Wert 0,791 auf. Dies zeigt, dass ClowdFlows mit der Verwendung von ‚?‘ Schwierigkeiten hat und keine korrekten Distanzberechnungen durchführen kann. Dieses Ergebnis verdeutlicht auch, dass der Test spezieller Werte im Data Mining eingesetzt werden kann. Durch die Anwendung des Tests wurde ein Fehler im Verfahren aufgedeckt, was auf ein Fehlverhalten innerhalb des Verfahrens schließen lässt.

Analog zu Testfall 9 wird nun in Testfall 10 geprüft, ob ein Ausrufezeichen ähnliche Auswirkungen auf die Berechnung der Distanzmaße hat wie ein Fragezeichen. Jedoch zeigt das Ergebnis wie erwartet, dass das ‚!‘ als normales Zeichen erkannt und die Distanzen korrekt berechnet werden. Ein ‚!‘ hat genauso wie ein ‚b‘ die Distanz von 1 zu ‚a‘. Dies ist korrekt, da die Zeichenketten als unterschiedlich erkannt werden und dabei nicht differenziert wird, ob Buchstaben oder Zeichen vorliegen.

In Testfall 11 werden besondere Zeichen getestet. Hierbei wird überprüft, ob die getesteten Zeichen als normale Zeichen verwendet werden und ob keine Schwierigkeiten bei der Anwendung auftreten. Der Test liefert erfolgreiche Ergebnisse. Die Zeichen werden ohne Besonderheit in die Distanzberechnung einbezogen. Lediglich einzelne Zeichen können nicht angezeigt werden.

Bei Testfall 12 wird überprüft, ob innerhalb des Verfahrens zwischen Groß- und Kleinschreibung unterschieden wird. Das Ergebnis des Verfahrens liefert die Erkenntnis, dass zwischen Groß- und Kleinschreibung korrekten Weise unterschieden wird. Dabei zeigt sich, dass ‚a‘ und ‚A‘ dieselbe Distanz (1.000) zueinander aufweisen wie ‚a‘ zu ‚b‘. Deshalb kann davon ausgegangen werden, dass ‚A‘ nicht als großes ‚a‘ interpretiert wird, sondern dass keine Gemeinsamkeiten bestehen.

Die erfolgreiche Durchführung des Tests spezieller Werte zeigt, dass eine Durchführung im Data Mining in der Praxis möglich ist. In diesem Fall wird durch die Ergebnisse und Erkenntnisse deutlich, dass das Programm ClowdFlows nicht gänzlich fehlerfrei ist. Wie in den Testfällen 8 und 9 gezeigt wird, ergeben sich bei bestimmter Werteeingabe Fehler im getesteten Verfahren. Dies zeigt, dass das hier ausgeführte Clusterverfahren technisch nicht einwandfrei funktioniert. Der durchgeführte Test kann auf jedes beliebig andere Clusterverfahren im Data Mining angewendet werden. Dazu müssen lediglich die zwölf Testfälle geprüft werden. Da die Daten zu einer Clusteranalyse in Tabellenform vorliegen und weiterverarbeitet werden, kann bei der Anwendung dieses Tests wie gezeigt vorgegangen werden. Somit kann der Test spezieller Werte generell im Clusterverfahren im Data Mining eingesetzt werden.

Die Durchführung des Tests wurde manuell durchgeführt. Allerdings soll diese Arbeit das Automatisierungspotential der Techniken aufzeigen. In Kapitel 4 wurde der Test spezieller Werte als automatisierbar eingestuft. Demnach ist die Möglichkeit gegeben, den hier entwickelten Test mit den zwölf Testfällen für das Clusterverfahren zu automatisieren.

Da die gesamte Entwicklung und Programmierung der Automatisierung den Rahmen dieser Arbeit übersteigen würde, wird im Folgenden eine Möglichkeit zur Automatisierung skizziert und aufgezeigt. Die Basis der Überprüfung der unterschiedlichen Clusterverfahren stellt ein SOLL – IST – Vergleich in Form einer Matrizen-Subtraktion dar. Dazu wird in das zu testende Verfahren eine Tabelle in einem Dokument eingelesen und überprüft, ob das Verfahren die Distanzen zwischen den Attributen korrekt berechnet und ausgibt. Die SOLL – Werte werden dazu mit den IST – Werten verglichen. Dazu wird in einer Matrix-Subtraktion die Matrix der IST - Werte von der Matrix der SOLL - Werte abgezogen. Unterscheidet sich ein Wert in dem Ergebnis – Matrix von Null, gilt der Test als nicht bestanden und das Clusterverfahren weist Fehler auf.

Die Testfälle sollen, wie beschrieben, in einem Dokument bereitgestellt werden. Darin sollen die zwölf Testfälle in Form von einer Tabelle abgebildet werden. Das Ziel sollte sein, alle zwölf Testfälle in möglichst einer Tabelle zu vereinen, damit lediglich eine Datei und somit nur ein Import nötig ist. In Tabelle 16 sind beispielhaft die Testfälle 1a und 1b für das Dokument dargestellt.

Tabelle 16: Testfälle 1a und 1b

ID	Merkmal 1	Merkmal 2
1	2	A
2	0	A
3	1	A
4	1	0
5	1	B
...

Das Dokument besteht aus zwei Spalten. Während die Attribute bei Merkmal 1 ausschließlich mit Zahlen beschrieben werden, sind die Merkmalausprägungen in Merkmal 2 Zeichen und Zeichenketten. Die erste Spalte von Tabelle 16 dient lediglich als Hilfe zur Beschreibung in dieser Arbeit und wird nicht im Dokument benötigt. Die Attribute 1 bis 3 repräsentieren Testfall 1a. Hierbei wird geprüft, ob eine ‚0‘ in Zusammenhang mit Zahlen korrekt interpretiert und die Distanzen zwischen den Attributen korrekt ausgerechnet wird. Äquivalent dazu findet sich in den Attributen 3 bis 5 Testfall 1b wieder. Attribut 3 wird dabei für die beiden ersten Testfälle gleichermaßen verwendet. Darin wird die richtige Anwendung von ‚0‘ in Verbindung mit Zeichen überprüft. In den weiteren Zeilen werden die restlichen Testfälle betrachtet. Beispielsweise wird für die Testfälle 2 und 3 jeweils nur eine Zeile benötigt, die das jeweilig betrachtete Vorzeichen mit einbeziehen. Ergänzende Zeilen mit ‚+1‘ bzw. ‚-1‘ und ‚A‘ reichen aus, um mit Hilfe der Distanzen zwischen den Attributen den Erfolg des Testfalls feststellen zu können. Testfall 2 gilt demnach als erfolgreich, wenn die Distanz zwischen der Zeile mit ‚+1‘; ‚A‘ und Zeile 3 genau 0 ist. Damit wird gezeigt, dass, wie in Tabelle 15 im Testfall 2 beschrieben, das ‚+‘ korrekterweise weggelassen wird. Als Dateityp für das zu importierende Dokument erscheint zum einen eine Textdatei (.txt) und zum anderen eine CSV-Datei (.csv) (Comma-separated values) geeignet, da die Möglichkeit eines Imports solcher Dateitypen wahrscheinlich ist. Deshalb ist eine Bereitstellung der Testfälle in beiden Dateiformaten erstrebenswert.

Wurde das Dokument mit den Testfällen importiert und das Clusterverfahren auf die Objekte angewendet, werden anschließend die Distanzen betrachtet. Dazu werden die einzelnen Distanzen zwischen den Attributen mit den SOLL – Werten verglichen. Da die Distanzen zwischen den Attributen in Form einer Matrix angegeben werden, sollten auch die SOLL – Werte in Form einer Matrix angegeben werden. Die Matrizen beinhalten alle Distanzen, die zwischen den Objekten liegen. Beide betrachteten Matrizen weisen im Wert x_{ij} die Distanz zwischen Objekt i zu Objekt j auf und sind symmetrisch.

In Ergänzung zu Tabelle 16 müssen die Distanzen von Objekt 1 zu den anderen Objekten in der Matrix wie folgt übereinstimmen:

- Zu Objekt 1: 0.000
- Zu Objekt 2: 1.414
- Zu Objekt 3: 1.414
- Zu Objekt 4: 1.000
- Zu Objekt 5: 1.000

Die hier angegebenen Distanzen entsprechen den Werten, welche bei korrekter Berechnung und technischer Umsetzung bei euklidischer Distanz gegeben sein müssten. Dabei sollte beachtet werden, dass verschiedene Distanzmaße als Grundlage für die Distanzen möglich sind. Die beschriebene Matrix bezieht sich auf das nicht normierte euklidische Distanzmaß. Stimmen alle Distanzen mit den oben angegebenen Werten überein, arbeitet das Data-Mining-Verfahren technisch korrekt.

Zum Vergleich der Matrizen wird eine Matrizen-Subtraktion durchgeführt (siehe Abbildung 9). Dazu wird die IST – Matrix von der SOLL – Matrix abgezogen. Die Ergebnis - Matrix gibt anschließend die Differenzen in den jeweiligen Distanzen im Wert y_{ij} an. Der Wert y_{ij} repräsentiert dabei die Differenz des Wertes aus der Soll – Matrix und dem Wert x_{ij} aus der IST – Matrix. Entspricht ein Wert in der Ergebnis - Matrix nicht null, liegt eine Abweichung zwischen den beiden Matrizen vor und der Test verlief nicht erfolgreich.

$$\begin{array}{ccc}
 \text{SOLL – Matrix} & & \text{IST – Matrix} & & \text{Ergebnis – Matrix} \\
 \left(\begin{array}{cccc}
 0.000; 1.414; 1.414; 1.000; 1.000 & \dots & x_{1m} \\
 \dots & & \dots \\
 x_{n1} & \dots & x_{nm}
 \end{array} \right) & - & \left(\begin{array}{ccc}
 x_{11} & \dots & x_{1m} \\
 \dots & & \dots \\
 x_{n1} & \dots & x_{nm}
 \end{array} \right) & = & \left(\begin{array}{ccc}
 y_{11} & \dots & y_{1m} \\
 \dots & & \dots \\
 y_{n1} & \dots & y_{nm}
 \end{array} \right)
 \end{array}$$

Abbildung 9: Matrizen-Subtraktion

Zur Automatisierung des Vergleiches wird die Summe der Beträge aller Werte der Ergebnis-Matrix berechnet und im Wert S angegeben. In einem erfolgreich verlaufenden Test sind alle Werte in der Ergebnis-Matrix 0 und somit auch die Summe aller Werte 0. Bei einem nicht erfolgreich verlaufenen Test weist mindestens ein Wert in der Matrix eine Abweichung gegenüber 0 auf. Die Summe der Beträge dieser Werte ist somit ungleich null. Der Betrag der einzelnen Werte ist in diesem Fall notwendig. Der Grund liegt daran, dass sowohl der Wert in der IST – Matrix als auch der Wert in der SOLL – Matrix größer sein kann.

$$S = \left| \sum_{i=1}^n \sum_{j=1}^m y_{ij} \right|$$

Bei der Subtraktion können somit als Ergebnis positive als auch negative Werte herauskommen. Um dem Fall entgegenzuwirken, dass zwei Fehler sich innerhalb der Addition aufheben, werden an dieser Stelle Beträge verwendet.

Somit reicht dieser eine Wert aus, um zu überprüfen, ob der Test erfolgreich verlaufen ist. Wird als Ergebnis 0 angegeben, verlief der Test erfolgreich und das Data-Mining-Verfahren weist keinen Fehler auf. Ist jedoch der Wert ungleich 0, wurden in dem untersuchten Data-Mining-Verfahren Fehler entdeckt und das Verfahren arbeitet nicht einwandfrei.

Jedoch muss neben der Prüfung der korrekten technischen Durchführung auch getestet werden, ob das Data-Mining-Verfahren korrekt arbeitet. Im zu testenden Clusterverfahren ist deshalb zu prüfen, ob der vorliegende Datensatz richtig geclustert wird. Dazu müssen sowohl die richtigen Cluster gefunden werden als auch die Objekte korrekt den richtigen Clustern zugeordnet werden. In dem betrachteten Datensatz Iris zeichnet das fünfte Merkmal eines jeden Objektes das richtige Cluster eines jeden Objektes aus. Dadurch kann durch einen Vergleich der entwickelten Cluster des Verfahrens mit der jeweiligen Zuordnung der Objekte und den Merkmalen der Objekte die Richtigkeit im Verfahren nachgewiesen werden. Um eine fehlerfreie Anwendung des Verfahrens zu gewährleisten, wurde ein Vergleich der erstellten Cluster mit den ordnungsgemäßen Clustern unter speziellen Bedingungen vorgenommen. Mit dem Ziel zu zeigen, dass das Data-Mining-Verfahren richtig clustert, wurden zufällig fünf Datensätze zunächst herausgenommen. Objekt 1, Objekt 13, Objekt 74, Objekt 102 und Objekt 147 wurden dazu dem Ursprungsdatensatz entwendet und für den späteren Test gespeichert. Bei der Durchführung des hierarchischen Clusterverfahren zeigte sich, dass

die richtigen Cluster gefunden und die Objekte korrekt zugeordnet wurden. Anschließend wurden die fünf herausgenommenen Objekte zur Clustering hinzugefügt. Auch diese wurden korrekt den jeweiligen Clustern zugeordnet. Dadurch wurde nachgewiesen, dass das Data-Mining-Verfahren richtig arbeitet. Aufgrund der entdeckten Fehler mit Hilfe des Tests spezieller Werte liegen jedoch Fehler in der technischen Durchführung vor, die aufgrund des nachgewiesenen richtig arbeitenden Verfahrens nicht in der Anwendung liegen können.

6.2 Fazit

Dieses Kapitel dient dazu, die Ergebnisse der Arbeit kurz wiederzugeben. Das erste Ziel war die Aufdeckung von Automatisierungspotential von V&V-Techniken. Dieses Ziel wurde erreicht, indem die vorgestellten V&V-Techniken mit einer Kriterien-Matrix hinsichtlich ihres Automatisierungspotentials untersucht wurden. Die hergeleiteten Kriterien der Matrix prüfen, ob menschliche Akteure im Zuge der Technik eingesetzt werden, die Objektivität und die Zuverlässigkeit der Ergebnisse, die Entwicklung der Testfälle im Zuge der V&V-Technik und den Aufwand, der geleistet werden muss, um die betrachtete Technik zu automatisieren. Dabei stellen die Kriterien ‚menschlicher Akteur‘ und ‚Entwicklung der Testfälle‘ Ausschlusskriterien dar. Gründe dafür sind zum einen, dass Techniken, welche einen menschlichen Akteur beinhalten nicht automatisiert werden können und zum anderen, dass die Auswahl der Testfälle automatisch ablaufen muss, da diese nicht manuell ausgewählt werden dürfen. Auf Grundlage dieser Matrix wurde jede einzelne V&V-Technik hinsichtlich ihres Automatisierungspotentials bewertet. Festzuhalten ist, dass alle informalen Techniken aufgrund des Einsatzes von menschlichen Akteuren nicht automatisiert werden können. Außerdem weisen alle vorgestellten Techniken aus dem Data Mining als auch aus der Softwareentwicklung ein Automatisierungspotential auf. Als Ergebnis dieser Untersuchung kann festgestellt werden, dass viele der untersuchten V&V-Techniken theoretisch das Potential besitzen, automatisiert zu werden. Lediglich fünf der bewerteten 22 Techniken wurden als nicht automatisierbar eingestuft.

Das Ziel der Prüfung des Automatisierungspotentials im Data Mining wurde erreicht und im Zuge von Kapitel 5 validiert. Die V&V-Techniken, welche als automatisierbar eingestuft wurden, wurden hinsichtlich ihrer Möglichkeit zum Einsatz im Data Mining geprüft. Grund für die Prüfung ist, dass nur wenige der bewerteten Techniken bereits im Data Mining zum Einsatz kommen. Auf Grundlage der Analyse erscheinen die Kriterien,

ob in den Techniken ein dynamisches Modell benötigt wird, ob formale Techniken vorliegen, ob Änderungen an den Eingangswerten notwendig sind, ob Quellcodes und reale Systeme benötigt werden, am wichtigsten. In der Softwareentwicklung erfordern viele V&V-Techniken einen Quellcode, welcher untersucht wird. Da im Data Mining kein Quellcode untersucht werden soll, können viele der Techniken aus der Softwareentwicklung im Data Mining nicht eingesetzt werden. Viele der V&V-Techniken aus der Simulation bilden zeitliche Zustände in einem dynamischen Modell ab. Diese sind für einen Einsatz im Data Mining nicht geeignet. Daher wurden lediglich fünf Techniken (Test spezieller Werte, Regressionstest, Korrelationsanalyse, Chi-Quadrat-Test und Regressionsanalyse) sowohl als automatisierbar als auch als für einen Einsatz im Data Mining geeignet eingestuft.

Die Techniken können jedoch nur in den drei Phasen des MESC-Vorgehensmodell eingesetzt werden. Die statistischen Techniken (Regressionsanalyse, Korrelationsanalyse und Chi-Quadrat-Test) können innerhalb der Datenauswahl (Phase 2) zum Einsatz kommen. In Phase 4 können die Bootstrapping-Validierung, Kreuzvalidierung und die Holdoutmethode (V&V-Techniken, die bereits im Data Mining eingesetzt werden) angewendet werden. Alle diese Techniken können im Zuge der Anwendung des Data-Mining-Verfahrens eingesetzt werden. Für die Analyse hinsichtlich eines Einsatzes in einem Data-Mining-Verfahren muss jede V&V-Technik bezüglich jedes Data-Minings-Verfahrens getestet werden, denn die Data-Mining-Verfahren Clusterbildung, Klassifikation und Assoziationsanalyse unterscheiden sich in ihrem Vorgehen deutlich. Die Ergebnisse der Untersuchung (siehe Tabelle 13) zeigen, dass sich die verschiedenen Data-Mining Verfahren bezüglich eines Einsatzes von V&V-Techniken erheblich unterscheiden. Daraus wurde gefolgert, dass eine generelle Untersuchung wie in Kapitel 5.1 nicht ausreichend ist. Obwohl diese Techniken automatisierbar sind und im Data Mining eingesetzt werden können, existieren deutliche Unterschiede zwischen den einzelnen Data-Mining-Verfahren, sodass eine detaillierte Untersuchung hinsichtlich jedes einzelnen Verfahrens sinnvoll ist.

Abschließend wurde durch die praktische Anwendung des Tests spezieller Werte in der hierarchischen Clusterbildung bewiesen, dass ein Einsatz dieses Tests im Data Mining möglich ist und damit die in Kapitel 4 und 5 theoretischen Überlegungen bei dieser Technik stimmen. Die im Zuge der Anwendung des Tests spezieller Werte entwickelten zwölf Testfälle können mit Hilfe der aufgezeigten Struktur automatisiert und auf andere Verfahren angewendet werden. Dazu müssen die Testfälle in Form einer Tabelle

eingelassen und durchgeführt werden. Anschließend werden mittels einer Matrix-Subtraktion die Distanzen zwischen den Objekten überprüft.

Damit wurden die in Kapitel 1 beschriebenen Ziele dieser Arbeit erreicht. Im Zuge dieser Arbeit wurde aufgezeigt, welche V&V-Techniken automatisiert und gleichzeitig im Data Mining angewendet werden können. Die beispielhafte Ausführung einer der Techniken belegt die vorherigen theoretischen Untersuchungen.

7 Zusammenfassung und Ausblick

In dieser Arbeit wurde das Automatisierungspotential von Verifikations- und Validierungstechniken im Data Mining untersucht. Dazu wurden in Kapitel 2 zunächst die begrifflichen Grundlagen im Data Mining definiert und Data-Mining-Vorgehensmodelle vorgestellt. Anschließend wurden in Kapitel 3 bekannte V&V-Techniken beschrieben, die im Data Mining bereits zum Einsatz kommen. Darüber hinaus wurden auch V&V-Techniken aus der Softwareentwicklung und aus der Simulation dargestellt.

Diese vorgestellten V&V-Techniken wurden in Kapitel 4 hinsichtlich ihres Automatisierungspotentials analysiert. Im Zuge dessen wurde zunächst untersucht, welche Bedingungen eine V&V-Technik erfüllen muss, um automatisiert werden zu können. Darauf aufbauend konnten fünf Kriterien hergeleitet werden, die für eine Automatisierung erfüllt sein müssen. Diese fünf Kriterien wurden in einer Matrix zusammengefasst, mit deren Hilfe die in Kapitel 4 vorgestellten V&V-Techniken hinsichtlich ihres Automatisierungspotentials bewertet wurden. Tabelle 10 zeigt die Ergebnisse der Untersuchungen und Bewertungen. Kapitel 5 analysiert die als automatisiert eingestuften V&V-Techniken hinsichtlich eines Einsatzes im Data Mining. Im Zuge der Untersuchung wurde speziell auf das Vorgehensmodell MESC eingegangen. Dabei wurde deutlich, dass ein Einsatz von V&V-Techniken in den Phasen nach dem Data-Mining-Verfahren nicht darstellbar ist. Allerdings kann ein Großteil der in Kapitel 5.1 überprüften Techniken im Zuge der Vorbereitung des Data-Minings-Verfahrens eingesetzt werden. Die Analyse hinsichtlich des Einsatzes in den Data-Mining-Verfahren ergab, dass eine Anwendung der V&V-Technik maßgeblich von dem verwendeten Data-Mining-Verfahren abhängt. Tabelle 12 in Kapitel 5.2 gibt einen Überblick über einen möglichen Einsatz der V&V-Techniken in den verschiedenen Data-Mining-Verfahren. Daher kann keine allgemeine Aussage darüber getroffen werden, welche der V&V-Techniken in Data-Mining-Verfahren eingesetzt werden können. Dies ist aufgrund der sehr unterschiedlichen Vorgehensweisen der Verfahren jeweils abhängig vom jeweils betrachteten Verfahren. Im Zuge eines Data-Mining-Vorgehensmodells kann im Gegensatz dazu eine Aussage getroffen werden. Tabelle 12 in Kapitel 5.1 zeigt, welche im Zuge dieser Arbeit betrachteten und analysierten V&V-Techniken sowohl ein Automatisierungspotential aufweisen als auch in einem Data-Mining-Vorgehensmodell eingesetzt werden können. Zu sehen ist, dass lediglich fünf der untersuchten Techniken automatisiert und im Data Mining eingesetzt werden können.

In Kapitel 5.2 wurde analysiert, welche dieser Techniken in welcher Phase des MESC-Vorgehensmodell angewendet werden können. Aufgrund der großen Unterschiede der Data-Mining-Verfahren wurde jede Technik hinsichtlich ihres Einsatzes in jedem einzelnen Data-Mining-Verfahren untersucht (siehe Tabelle 13).

Der Test spezieller Werte wurde in Kapitel 6.1 auf das Data-Mining-Verfahren Clustering übertragen. Die durchgeführten Testfälle mit ihren Ergebnissen (Tabelle 15) zeigen, dass dieser Test im Clustering angewendet werden kann. Die Testfälle 8 und 9 weisen Fehler im Zuge der Anwendung des hierarchischen Clusterings in ClowdFlows nach. Diese Fehler belegen eindeutig eine fehlerhafte Durchführung der technischen Anwendung des Clusterverfahrens. Damit wurde gezeigt, dass die Überlegungen in Kapitel 5.2 bezüglich einer Anwendung im Clusterverfahren korrekt sind.

Auch konnte in Kapitel 6.1 dargestellt werden, wie eine Automatisierung des Tests spezieller Werte im Rahmen des Data-Mining-Verfahrens Clustering vorgenommen werden kann. Dazu wird zunächst eine Datei importiert, die einen Datensatz mit den vorgestellten Testfällen enthält. Auf diese Objekte in dieser Tabelle wird das Clusterverfahren angewendet. Anschließend werden die Distanzen zwischen den Objekten mit den SOLL – Distanzen mittels einer Matrizen-Subtraktion verglichen. Ist dabei die Summe der Beträge aller Werte nicht 0, weist das betrachtete Data-Mining-Verfahren einen Fehler vor und arbeitet nicht einwandfrei.

Im Zuge der Untersuchungen in Kapitel 6.1 konnte zudem nachgewiesen werden, dass das Clusterverfahren richtig durchgeführt wurde. Neben der Prüfung der technischen Durchführung des Clusterverfahrens musste parallel dazu die Überprüfung des Clusterings an sich vorgenommen werden. Dazu wurde geprüft, ob das Clusterverfahren die Objekte dem korrekten Cluster zuordnet, wenn fünf Objekte des Datensatzes fehlen. Da das Clustering die Objekte richtig zuordnete und anschließend die herausgenommenen fünf Objekte auch korrekt zuordnete, wurde nachgewiesen, dass das Clusterverfahren richtig durchgeführt wird. Damit wurde zusätzlich gezeigt, dass die aufgedeckten Fehler in der technischen Umsetzung des Clusterverfahrens liegen müssen.

Durch diese Untersuchungen wurde gezeigt, dass der Test spezieller Werte im Data-Mining-Verfahren Clustering angewendet werden kann. Damit wurden die theoretischen Analysen und Bewertungen in den vorherigen Kapiteln in einem Fall praktisch durchgeführt und bewiesen.

Weitere Untersuchungen, die an diese Arbeit anknüpfen, könnten sich mit der praktischen Umsetzung der Anwendung von anderen Techniken ins Data Mining beschäftigen. In dieser Arbeit wurde lediglich der Test spezieller Werte im Data-Mining-Verfahren Clustering angewendet. Einerseits könnte der Test spezieller Werte auch in der Klassifikation praktisch angewendet werden und im Zuge dessen aufgezeigt werden, wie diese Ausführung automatisiert werden kann. Andererseits zeigt Tabelle 13 weitere Techniken, die in den verschiedenen Data-Mining-Verfahren angewendet werden könnten. Eine praktische Ausführung dieser Anwendungen erscheint denkbar. Des Weiteren könnte eine detaillierte Untersuchung der verschiedenen V&V-Techniken hinsichtlich ihres Einsatzes in der Datenvorbereitung und Vorbereitung des Data-Mining-Verfahrens im Zuge des MES-C-Vorgehensmodells vorgenommen werden. Hier wäre sowohl eine genaue Untersuchung hinsichtlich der Einsatzmöglichkeiten als auch der Automatisierbarkeit im Rahmen einer weiteren Arbeit denkbar. Auch wäre eine praktische Umsetzung der aufgezeigten Automatisierung des Tests spezieller Werte mit den vorgestellten Testfällen im Clusterverfahren möglich.

Aktuelle Untersuchungen und Paper zeigen die Aktualität dieses Themas. Beispielsweise sieht Chan (2018) den traditionellen Ansatz der Prüfung durch die Echtzeitwirtschaft überholt. Eine Innovation des traditionellen Prüfprozesses erscheint für Chan notwendig. Durch den Einsatz von Technologie und Automatisierung wird eine kontinuierliche Prüfung als Nachfolge gesehen, die die Effektivität und Effizienz erhöhen (Chan et al. 2018).

Aus meiner Sicht wird die Automatisierung von V&V-Techniken zunehmend an Bedeutung gewinnen. Daher werden weitere Untersuchungen, wie oben aufgezeigt, zu diesem Thema folgen.

Literaturverzeichnis

Abolhassan, Ferri; Kellermann, Jörn (Hg.) (2016): Effizienz durch Automatisierung. Das „Zero Touch“-Prinzip im IT-Betrieb. Wiesbaden: Springer Gabler. Online verfügbar unter <http://gbv.ebib.com/patron/FullRecord.aspx?p=4441827>.

Ammann, Paul; Offutt, Jeff (2008): Introduction to software testing. 1. publ. Cambridge: Cambridge Univ. Press.

APCoQS (Hg.) (2000): Quality software, 2000. First Asia-Pacific Conference on Quality Software : proceedings : October 30-31, 2000, Hong Kong, China. Asia-Pacific Conference on Quality Software; Institute of Electrical and Electronics Engineers. Los Alamitos, CA, Piscataway, N.J.: IEEE Computer Society. Online verfügbar unter <http://ieeexplore.ieee.org/servlet/opac?punumber=7095>.

Bacher, Johann; Pöge, Andreas; Wenzig, Knut (2010): Clusteranalyse. Anwendungsorientierte Einführung in Klassifikationsverfahren. 3., erg., vollst. überarb. und neu gestaltete Aufl. München: Oldenbourg.

Backhaus, Klaus (2003): Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. 10., neu bearb. und erw. Aufl. Berlin: Springer (Springer-Lehrbuch).

Balci, Osman (2007): Verification, Validation, and Testing. In: Jerry Banks (Hg.): Handbook of simulation. Principles, methodology, advances, applications, and practice. New York, Norcross, Ga.: Wiley, S. 335–393.

Balzert, Helmut (1999): Lehrbuch Grundlagen der Informatik. Konzepte und Notationen in UML, Java und C++, Algorithmik und Software-Technik, Anwendungen ; mit 2 CD-ROMs. Heidelberg: Spektrum Akad. Verl. (Lehrbücher der Informatik).

Balzert, Helmut (2009): Lehrbuch der Softwaretechnik. Basiskonzepte und Requirements-Engineering. 3. Aufl. Heidelberg: Spektrum Akad. Verl. (Lehrbücher der Informatik).

Bauernhansl, Thomas; Hompel, Michael ten; Vogel-Heuser, Birgit (Hg.) (2014): Industrie 4.0 in Produktion, Automatisierung und Logistik. Anwendung, Technologien, Migration. Wiesbaden: Springer Vieweg (SpringerLink). Online verfügbar unter <http://dx.doi.org/10.1007/978-3-658-04682-8>.

- Becker, Werner; Hübner, Kurt (Hg.) (1976): Objektivität in den Natur- und Geisteswissenschaften. 1. Aufl. Hamburg: Hoffmann und Campe (Kritische Wissenschaft).
- Beizer, Boris (1990): Software testing techniques. 2. ed. New York: Van Nostrand Reinhold.
- Berghammer, Rudolf (2008): Ordnungen, Verbände und Relationen mit Anwendungen. Studium. 1. Aufl. Wiesbaden: Vieweg+Teubner Verlag / GWV Fachverlage GmbH Wiesbaden (Leitfäden der Informatik).
- Berry, Michael J. A.; Linoff, Gordon (2000): Mastering data mining. The art and science of customer relationship management. New York, NY: Wiley Computer Publ.
- Bigus, Joseph P. (1996): Data mining with neural networks. Solving business problems from application development to decision support. New York: McGraw-Hill (Computing McGraw-Hill).
- Bishop, Christopher M. (1995): Neural networks for pattern recognition. Oxford: Oxford University Press; Clarendon Press.
- Bissantz, Nicolas; Hagedorn, Jürgen (2009): Data Mining (Datenmustererkennung). In: *Wirtsch. Inform.* 51 (1), S. 139–144. DOI: 10.1007/s11576-008-0108-z.
- Bosch, Karl (2015): Großes Lehrbuch der Statistik. Reprint.
- Breiman, Leo (1998): Classification and regression trees. Repr. Boca Raton: Chapman & Hall.
- Buneman, Peter (Hg.) (1993): Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data ; Washington, DC, May 26 - 28, 1993. Association for Computing Machinery; ACM SIGMOD International Conference on Management of Data; Annual Conference. Association for Computing Machinery Special Interest Group on Management of Data. New York, NY: ACM Press (SIGMOD record, 22.2).
- Burgdorf, Florian (2010): Eine kunden- und lebenszyklusorientierte Produktfamilienabsicherung für die Automobilindustrie. Zugl.: Karlsruhe, Karlsruher Inst. für Technologie, Diss., 2010. Karlsruhe: KIT Scientific Publ.
- Chan, David Y.; Chiu, Victoria; Vasarhelyi, Miklos A. (2018): Continuous Auditing. A Book of Theory and Application. Bingley: Emerald Publishing Limited.

- Cios, Krzysztof J.; Kurgan, Lukasz A.; Pedrycz, Witold; Swiniarski, Roman W. (2007): Data Mining. A Knowledge Discovery Approach. Boston, MA: Springer Science+Business Media LLC.
- Clarke, Bertrand; Fokoue, Ernest; Zhang, Hao Helen (2009): Principles and theory for data mining and machine learning. New York, NY: Springer (Springer Series in Statistics).
- Cleve, Jürgen; Lämmel, Uwe (2016): Data mining. Berlin, Boston: De Gruyter Studium.
- Cordy, James R. (Hg.) (2008): Eighth IEEE International Working Conference on Source Code Analysis and Manipulation, 2008. SCAM 2008 ; 28 - 29 Sept. 2008, Beijing, China ; proceedings. IEEE Computer Society; IEEE International Working Conference on Source Code Analysis and Manipulation; SCAM. Piscataway, NJ: IEEE. Online verfügbar unter <http://ieeexplore.ieee.org/servlet/opac?punumber=4637522>.
- Desharnais, Jean-Marc; Abran, Alain; Suryn, Witold (2011): Identification and analysis of attributes and base measures within ISO 9126. In: *Software Qual J* 19 (2), S. 447–460. DOI: 10.1007/s11219-010-9124-5.
- DIN (Hg.) (2014): DIN IEC 60050-351 "Internationales Elektronisches Wörterbuch: Teil 351: Leichttechnik. Berlin: Beuth.
- Dunn, Robert H. (1984): Software defect removal. New York: McGraw-Hill.
- Dustin, Elfriede; Rashka, Jeff; Paul, John (2001): Software automatisch testen. Verfahren, Handhabung und Leistung. Berlin, Heidelberg: Springer (Xpert.press).
- Eckstein, Peter; Götze, Wolfgang; Hartl, Friedrich; Rönz, Bernd; Strohe, Hans Gerhard (1994): Lexikon Statistik. Hg. v. Bernd Rönz und Hans Gerhard Strohe. Wiesbaden: Gabler Verlag. Online verfügbar unter <http://dx.doi.org/10.1007/978-3-322-91144-5>.
- Efron, Bradley; Tibshirani, Robert (1998): An introduction to the bootstrap. [Nachdr.]. Boca Raton, Fla.: Chapman & Hall (Monographs on statistics and applied probability, 57).
- Ester, Martin; Sander, Jörg (2000): Knowledge Discovery in Databases. Techniken und Anwendungen. Berlin, Heidelberg, s.l.: Springer Berlin Heidelberg.
- European Design Automation Association; Design, Automation & Test in Europe Conference & Exhibition; DATE (2013): Design, Automation & Test in Europe

Conference & Exhibition (DATE), 2013. 18 - 22 March 2013, Grenoble, France ; proceedings. Piscataway, NJ: IEEE.

Fajardo, Jose; Dustin, Elfriede (2007): Testing SAP R/3. A manager's step-by-step guide. Hoboken, N.J.: J. Wiley & Sons (Safari Books Online).

Fayyad, Usama M. (Hg.) (1996): Advances in knowledge discovery and data mining. Menlo Park, Calif.: AAAI Press.

Feja, S.; Witt, S.; Speck, A. (2013): Automatische Validierung von Prozessmodellen. Kiel, 20.12.2013. Online verfügbar unter http://www.fh-kiel.de/fileadmin/data/wirtschaft/ProzessManagementForum2012/PMF_2013/Feja_Witt_automatische_Validierung.pdf, zuletzt geprüft am 28.06.2018.

FISHER, R. A. (1936): THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS. In: *Annals of Eugenics* 7 (2), S. 179–188. DOI: 10.1111/j.1469-1809.1936.tb02137.x.

Frawley, W. J.; Piatetsky-Shapiro, G.; Matheus, C. J. (1992): Knowledge discovery in databases: an overview. In: *AI Magazine* (3).

Galindo, Daniel dos Santos (2018): Software Quality. Concepts and Practice. Hoboken, USA: IEEE Computer Society Inc.

Haffer, Jason (2015): Have you seen ASUM-DM? Hg. v. SPSS Predictive Analytics. IBM. Online verfügbar unter <https://developer.ibm.com/predictiveanalytics/2015/10/16/have-you-seen-asum-dm/>.

Hamilton, Patrick (2008): Wege aus der Softwarekrise. Verbesserungen bei der Softwareentwicklung. Berlin, Heidelberg: Springer-Verlag.

Han, Jiawei; Pei, Jian; Yin, Yiwen; Mao, Runying (2004): Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach. In: *Data Mining and Knowledge Discovery* 8 (1), S. 53–87. DOI: 10.1023/B:DAMI.0000005258.31418.83.

Harper, Gavin; Pickett, Stephen D. (2006): Methods for mining HTS data. In: *Drug discovery today* 11 (15-16), S. 694–699. DOI: 10.1016/j.drudis.2006.06.006.

Heinrich, Berthold; Linke, Petra; Glöckler, Michael (2015): Grundlagen Automatisierung. Sensorik, Regelung, Steuerung. Wiesbaden: Springer Vieweg (Lehrbuch).

- Hermann, Charles F. (1967): Validation problems in games and simulations with special reference to models of international politics. In: *Syst. Res.* 12 (3), S. 216–231. DOI: 10.1002/bs.3830120306.
- Hippner, Hajo (Hg.) (2001): Handbuch Data Mining im Marketing. Knowledge Discovery in Marketing Databases. 1. Aufl. Braunschweig: Vieweg (Business Computing).
- Hofman, H. J. (1990): Die Anwendung des CART-Verfahrens zur statistischen Bonitätsanalyse von Konsumentenkrediten. In: *Zeitschrift für Betriebswirtschaft* (60), S. 941–962.
- Hompel, Michael ten; Büchter, Hubert; Franzke, Ulrich (2008): Identifikationssysteme und Automatisierung. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg (VDI-Buch).
- Kohavi, R. (1995): A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, Montreal (Kanada)*, S. 1137–1143.
- Krallmann, Hermann; Bobrik, Annette; Levina, Olga (2013): Systemanalyse im Unternehmen. Prozessorientierte Methoden der Wirtschaftsinformatik. 6., überarb. und erw. Aufl. München: Oldenbourg-Verl.
- Kranjc, Janez (2017): clowdflows. Hg. v. Janez Kranjc. Online verfügbar unter <http://clowdflows.org/>, zuletzt geprüft am 21.09.2018.
- Krcmar, Helmut (2015): Informationsmanagement. 6., überarb. Aufl. Wiesbaden: Springer Gabler.
- Küppers, Bertram (1999): Data Mining in der Praxis. Ein Ansatz zur Nutzung der Potentiale von Data Mining im betrieblichen Umfeld. Zugl.: Graz, Techn. Univ., Diss., 1998. Frankfurt am Main: Lang (Europäische Hochschulschriften Reihe 5, Volks- und Betriebswirtschaft, 2373).
- Kurgan, Lukasz A.; MUSILEK, PETR (2006): A survey of Knowledge Discovery and Data Mining process models. In: *Knowl. Eng. Rev.* 21 (01), S. 1. DOI: 10.1017/S0269888906000737.
- Kwartler, Ted (2017): Text mining in practice with R. Hoboken, NJ: John Wiley & Sons Ltd.

- Landry, Maurice; Oral, Muhittin (1993): In search of a valid view of model validation for operations research. In: *European Journal of Operational Research* 66 (2), S. 161–167. DOI: 10.1016/0377-2217(93)90310-J.
- Lassmann, Wolfgang; Schwarzer, Jens; Rogge, Rolf (Hg.) (2006): *Wirtschaftsinformatik. Nachschlagewerk für Studium und Praxis*. Wiesbaden: Gabler. Online verfügbar unter <http://dx.doi.org/10.1007/978-3-8349-9152-2>.
- Law, Averill M. (2007): *Simulation modeling and analysis*. 4. ed., internat. ed. Boston, Mass.: McGraw-Hill (McGraw-Hill series in industrial engineering and management science).
- Lawler, Eugene L. (Hg.) (1995): *The traveling salesman problem. A guided tour of combinatorial optimization*. Reprint. Chichester: Wiley (Wiley-interscience series in discrete mathematics and optimization).
- Liggesmeyer, Peter (2009): *Software-Qualität. Testen, Analysieren und Verifizieren von Software*. 2. Aufl. s.l.: Spektrum Akademischer Verlag.
- Ludewig, Jochen; Lichter, Horst (2012): *Software Engineering. Grundlagen, Menschen, Prozesse, Techniken*. s.l.: dpunkt.verlag.
- Lyu, Michael R. (1996): *Handbook of software reliability engineering*. New York: McGraw Hill (Computing McGraw Hill).
- MacNair, Edward A. (Hg.) (1989): *1989 Winter Simulation Conference proceedings. December 4 - 6, 1989, The Capital Hilton Hotel Washington, DC. Winter Simulation Conference*. New York, NY: Assoc. for Computing Machinery.
- Malik, Fredmund (2013): *Führen, leisten, leben. Wirksames Management für eine neue Zeit. Limitierte Sonderausg.* Frankfurt/Main: Campus-Verl.
- Malthouse, Edward C.; Blattberg, Robert C. (2005): Can we predict customer lifetime value? In: *Journal of Interactive Marketing* 19 (1), S. 2–16. DOI: 10.1002/dir.20027.
- Mann, Manfred (2010): *Assoziationsanalyse*. München: GRIN Verlag GmbH.
- Mellish, Chris S. (Hg.) (1995): *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence. International Joint Conference on Artificial Intelligence (IJCAI)*. San Mateo, Calif.: Morgan Kaufmann.

- Myers, Glenford J.; Badgett, Tom; Thomas, Todd M.; Sandler, Corey (2004): The art of software testing. 2nd ed. Hoboken, N.J: John Wiley & Sons (Business Data Processing, v.28).
- North, Klaus (2016): Wissensorientierte Unternehmensführung. Wissensmanagement gestalten. 6., aktualisierte und erweiterte Auflage. Wiesbaden: Springer Gabler.
- Oberhauser, Otto; Gödert, Winfried (2005): Automatisches Klassifizieren. Entwicklungsstand - Methodik - Anwendungsbereiche. Frankfurt am Main: Lang (Europäische Hochschulschriften Reihe XLI, Informatik, 43).
- Offutt, Jeff (2011): A mutation carol: Past, present and future. In: *Information and Software Technology* 53 (10), S. 1098–1107. DOI: 10.1016/j.infsof.2011.03.007.
- Perner, Petra (Hg.) (2012): Machine learning and data mining in pattern recognition. 8th international conference, MLDM 2012, Berlin, Germany, July 13 - 20, 2012 ; proceedings. MLDM; International Conference on Machine Learning and Data Mining. Berlin: Springer (Lecture notes in computer science Lecture notes in artificial intelligence, 7376). Online verfügbar unter <http://dx.doi.org/10.1007/978-3-642-31537-4>.
- Perry, William E. (2006): Effective methods for software testing. 3rd ed. Indianapolis, IN: Wiley (Safari Books Online).
- Petersohn, Helge (2009): Data Mining. Verfahren, Prozesse, Anwendungsarchitektur. Zugl.: Leipzig, Univ., Habil, 2004. München, Wien: Oldenbourg.
- Pierson, Lillian; Porway, Jake (2016): Data Science für Dummies. Unter Mitarbeit von Tobias Häberlein. 1. Auflage. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA (... für Dummies).
- Pressman, Roger S.; Maxim, Bruce R. (2015): Software engineering. A practitioner's approach. 8. ed. New York, NY: McGraw-Hill Education.
- Rabe, Markus; Spiekermann, Sven; Wenzel, Sigrid (2008): Verifikation und Validierung für die Simulation in Produktion und Logistik. Vorgehensmodelle und Techniken. Berlin, Heidelberg: Springer (VDI-Buch).
- Rey, Günter Daniel; Wender, Karl F. (2018): Neuronale Netze. Eine Einführung in die Grundlagen, Anwendungen und Datenauswertung. 3., überarbeitete Auflage. Bern: Hogrefe.

- Runkler, Thomas A. (2010): Data Mining. Methoden und Algorithmen intelligenter Datenanalyse ; mit 7 Tabellen. 1. Aufl. Wiesbaden: Vieweg + Teubner (Studium).
- Runkler, Thomas A. (2015): Data mining. Modelle und Algorithmen intelligenter Datenanalyse. 2., aktualisierte Auflage. Wiesbaden: Springer Vieweg (Lehrbuch).
- Sargent, R. G. (2013): Verification and validation of simulation models. In: *Journal of Simulation* 7 (1), S. 12–24. DOI: 10.1057/jos.2012.20.
- Schach, Stephen R. (1999): Classical and object-oriented software engineering with UML and C++. 4. ed., international ed. Boston: WCB McGraw-Hill (McGraw-Hill international editions).
- Scheidler, Anne Antonia (2017): Methode zur Erschließung von Wissen aus Datenmustern in Supply-Chain-Datenbanken. 1st ed. Göttingen: Cuvillier Verlag (Schriftenreihe Fortschritte in der IT in Produktion und Logistik, v.1).
- Schendera, Christian F. G. (2014): Regressionsanalyse mit SPSS. 2. korr. und aktualisierte Aufl. München: De Gruyter (De Gruyter Studium).
- Scherer, Andreas (1997): Neuronale Netze. Grundlagen und Anwendungen. Wiesbaden: Vieweg+Teubner Verlag (Computational Intelligence).
- Schlosser, Otto; Klages, Helmut (1976): Einführung in die sozialwissenschaftliche Zusammenhangsanalyse. Reinbek bei Hamburg: Rowohlt (Rororo-Studium Sozialwissenschaft, 89).
- Schnieder, Eckehard (1999): Methoden der Automatisierung. Beschreibungsmittel, Modellkonzepte und Werkzeuge für Automatisierungssysteme. Wiesbaden: Vieweg+Teubner Verlag (Studium Technik).
- Sneed, Harry M.; Baumgartner, Manfred; Seidl, Richard (2009): Der Systemtest. Von den Anforderungen zum Qualitätsnachweis. 2., aktualisierte und erw. Aufl. München: Hanser.
- Sommerville, Ian (2016): Software engineering. Tenth edition, global edition. Boston, Columbus, Indianapolis, New York, San Francisco, Hoboken, Amsterdam, Cape Town, Dubai, London, Madrid, Milan, Munich, Paris, Montreal, Toronto, Delhi, Mexico City, São Paulo, Sydney, Hong Kong, Seoul, Singapore, Taipei, Tokyo: Pearson (Always learning).

- Spillner, Andreas; Linz, Tilo (2012): Basiswissen Softwaretest. Aus- und Weiterbildung zum Certified Tester ; Foundation Level nach ISTQB-Standard. 5. überarb. und aktualisierte Aufl. Heidelberg: dpunkt Verl. (Safari Tech Books Online).
- Spinellis, Diomidis (2017): State-of-the-Art Software Testing. In: *IEEE Softw.* 34 (5), S. 4–6. DOI: 10.1109/MS.2017.3571564.
- Steiner, Viviana (2009): Modellierung des Kundenwertes. Ein branchenübergreifender Ansatz. Mit einem Geleitwort von Christian Homburg. Zugl.: Mannheim, Univ., Diss., 2008. 1. Aufl. Wiesbaden: Gabler (Gabler Edition Wissenschaft Schriftenreihe des Instituts für Marktorientierte Unternehmensführung (IMU), Universität Mannheim).
- Steinlein, Uwe (2004): Data Mining als Instrument der Responseoptimierung im Direktmarketing. 1st ed. Göttingen: Cuvillier Verlag.
- Steinmüller, Wilhelm (1993): Informationstechnologie und Gesellschaft. Einführung in die Angewandte Informatik. Darmstadt: Wiss. Buchges.
- VDI-Richtlinie 3633 Blatt 1 (2008): VDI-Richtlinie 3633 Blatt 1. Simulation von Logistik-, Materialfluss und Produktionssystemen. Berlin: Beuth.
- Vivencio, Alberto (2010): Testautomation mit SAP®. SAP Banking erfolgreich einführen. 1. Aufl. s.l.: Vieweg+Teubner (GWV).
- Walter, Jörg Andreas (2004): Datamining. 1st ed. Göttingen: Cuvillier Verlag.
- Weskamp, M.; Tamas, A.; Wochinger, T.; Schatz, A. (2014): Einsatz und Nutzenpotenziale von Data Mining in Produktionsunternehmen. Stuttgart: Fraunhofer IPA.
- Wichmann, B. A.; Cox, M. G. (1992): Problems and strategies for software component testing standards. In: *Softw. Test. Verif. Reliab.* 2 (4), S. 167–185. DOI: 10.1002/stvr.4370020403.
- Witten, Ian H.; Frank, Eibe; Hall, Mark A. (Hg.) (2011): Data mining. Practical machine learning tools and techniques. 3. ed. Amsterdam: Elsevier/Morgan Kaufmann (/Morgan Kaufmann series in data management systems]).
- Wotawa, Franz; Nica, Mihai; Kushik, Natalia (2016): Testing Software and Systems. Cham: Springer International Publishing (9976).

Zahavi, Jacob; Levin, Nissan (1997): Applying neural computing to target marketing.
In: *J. Direct Mark.* 11 (4), S. 76–93. DOI: 10.1002/(SICI)1522-
7138(199723)11:4<76::AID-DIR10>3.0.CO;2-D.

Anhang

Elektronischer Anhang A: Datensatz IRIS

Elektronischer Anhang B: Clustering in ClowdFlows

Eidesstattliche Versicherung

Name, Vorname

Matrikel-Nr.

Ich versichere hiermit an Eides statt, dass ich die vorliegende Masterarbeit mit dem Titel „*Untersuchung von Verifikations- und Validierungstechniken in Bezug auf ihr Automatisierungspotential im Data Mining*“ selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Ort, Datum

Unterschrift

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden. (§ 63 Abs. 5 Hochschulgesetz - HG -)

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft. Die Technische Universität Dortmund wird ggfs. elektronische Vergleichswerkzeuge (wie z.B. die Software „turnitin“) zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Ort, Datum

Unterschrift