

Forensic Accounting - Potenziale von Big Data und Predictive Analytics im Rahmen der IT Governance

Rainer Schmidt¹, Michael Möhring², Robert Rieg³, Stefan Maier³, Gianfranco Walsh²,
Ralf-Christian Härting³

¹Fakultät für Informatik und Mathematik
Hochschule München
Lothstraße 64, 80335 München
Rainer.Schmidt@hm.edu

²Wirtschaftswissenschaftliche Fakultät - Lehrstuhl für Allgemeine
Betriebswirtschaftslehre und Marketing
Friedrich-Schiller-Universität Jena, Carl-Zeiß-Straße 3, 07443 Jena,
Michael.Moehring@uni-jena.de, Walsh@uni-jena.de

³Fakultät für Wirtschaftswissenschaften
Hochschule Aalen
Beethovenstraße 1, 73430 Aalen
Robert.Rieg@htw-aalen.de, Stefan.Maier@htw-aalen.de, Ralf.Haerting@htw-aalen.de

Abstract: Die Informationstechnik stellt ein wertvolles Instrumentarium für die Aufdeckung doloser Handlungen und die Prüfung von Regeleinhaltungen im Rechnungswesen bereit. Mit der exponentiellen Zunahme relevanter Daten für das Rechnungswesen, sowohl in strukturierter als auch wenig bis unstrukturierter Form, kommen bisherige Werkzeuge und Verfahren jedoch an ihre Grenzen in Bezug auf die Rechenzeit und Verarbeitungsmöglichkeit. Vor diesem Hintergrund bietet es sich an, zwei bisher nicht verbundene Felder miteinander zu verbinden, die Forensik im Rechnungswesen mit Big Data und Business Analytics. Der vorliegende Beitrag stellt die wesentlichen IT-Entwicklungen vor und diskutiert ihre Potentiale und Einsatzmöglichkeiten für das Rechnungswesen und die IT-Governance. Des Weiteren wird ein Bezugsrahmen für den Einsatz von Big Data und Predictive Analytics vorgestellt, der vor allem großen Unternehmen mit Massendaten, Prüfungsgesellschaften und Prüfungsbehörden Hinweise auf eine prototypische Vorgehensweise der datengetriebenen forensischen Analyse bietet.

1 Ausgangssituation Forensic Accounting

Die Wirtschaftskriminalität in deutschen Unternehmen und der daraus entstehende monetäre Schaden stellen hohe Risiken für Organisationen dar [Pric11]. In Deutschland wurden im Jahr 2012 knapp 82.000 Fälle doloser Handlungen mit einer kumulierten Schadenssumme von ca. 3,7 Mrd. Euro verzeichnet. Dolose Handlungen sind illegale Handlungen, welche durch vorsätzliche Täuschung, Verschleierung oder Vertrauensmissbrauch gekennzeichnet sind [Frei08]. Darunter fallen beispielsweise

Zahlungen, welche Mitarbeiter nicht konform vornehmen und Dritten gegenüber ausbezahlen. Durch sie kann der Unternehmung bzw. Organisation ein erheblicher Schaden entstehen und die Erreichung der Unternehmensziele gefährdet werden [Frei08]. Die Erkennung und das (präventive) Management doloser Handlungen stellt eine wichtige Aufgabe der internen Revision dar [Frei08]. Im Rahmen der IT Governance muss u. a. sichergestellt werden, dass alle IT-gestützten Unternehmensprozesse gesetzlichen und unternehmensspezifischen Regularien (Compliance) entsprechen und Geschäftsziele unterstützen [MeZK03]. Neben innerorganisatorischen Regularien nimmt die Anzahl an gesetzlichen Bestimmungen, wie der Sarbanes-Oxley Act, Basel III, KonTrAG und Solvency II, zu und betrifft damit neben den Geschäftsabläufen auch die IT-Governance [JoGo06]. IT-Forensik stellt damit ein wesentlicher Bestandteil zur Aufdeckung von Compliance-Verstößen im Rahmen der Governance dar [Kirm12]. Unternehmen müssen gemäß ihrer Sorgfaltspflicht [Bund65] und weiteren gesetzlichen Bestimmungen (wie bspw. der Verhinderung von Geldwäsche [Bund08]) betrügerischen Handlungen entgegen wirken. Bei der Abschlussprüfung müssen weiterhin Prüfungshandlungen und Ergebnisse mit Bezug auf betrügerische Handlungen dokumentiert werden [Idw06] [Prüf02].

Dolose Handlungen können auf unterschiedliche Weise erfolgen, wobei oft IT-Systeme genutzt werden [KLSH04] [Herz06] [BeKu12]. So können bspw. Zahlungen über ein Enterprise-Ressource-Planning (ERP) System veranlasst werden oder die Planung und Durchführung von dolosen Handlungen kann in elektronischen Texten wie bspw. E-Mails erfolgen. Da die angebotenen Funktionen und die Menge der verfügbaren Daten in IT-Systemen kontinuierlich steigen [GaRe11], ist auch bei der Suche nach dolosen Handlungen der IT-Einsatz unumgänglich und damit die Durchführung automatisierter IT-gestützter Analyseverfahren unerlässlich.

Bisherige Datenanalyseverfahren zur Missbrauchserkennung können die erforderliche Prüfungstiefe immer weniger gewährleisten. So werden isolierte Verfahren der explorativen Statistik sowie zur Jahresabschlussprüfung durch die Menge der zu verarbeitenden Daten überfordert. Sie können zudem die Prüfungen nicht mehr in der verfügbaren Zeit durchführen und haben grundsätzliche Schwierigkeiten bei der Analyse unstrukturierter Daten (wie bspw. E-Mails zwischen Kunden und Lieferanten). Weiterhin ist bei der Analyse unstrukturierter Daten auf die Einhaltung von gesetzlichen Bestimmungen zu achten.

Ziel des vorliegenden Beitrags ist es deshalb, einen neuartigen datengetriebenen Ansatz des forensischen Rechnungswesens im Rahmen des IT-Prüfungswesens vorzustellen. Er stützt sich dabei auf Fortschritte der Datenverarbeitung und -analyse, insbesondere bei den Faktoren Volumen, der Geschwindigkeit und der Varianztoleranz, die unter dem Schlagwort „Big Data“ subsummiert werden. Der Einsatz von neuen technologischen Entwicklungen im Rahmen von Big Data kann die Analyse von großen Datensätzen, unterschiedlichster Struktur in kurzer Zeit ermöglichen. Bisherige Forschung im Kontext des forensischen Rechnungswesens konzentrierte sich auf die Analyse strukturierter Daten ohne Anwendung von Big-Data-Analysen [PLSG10], [Herz06]. Über den Einsatz von Big Data in deutschen Unternehmen liegen nur wenige Forschungsergebnisse und dokumentierte praktische Erfahrungen vor [PLSG10], [Herz06].

Zunächst sollen dazu, Begriffe im Umfeld von Big Data definiert und in ihrem Verhältnis geklärt werden. Danach werden Methoden der datenbasierten Forensik vorgestellt. Es folgt die Einführung eines Bezugsrahmens zur IT basierten Unterstützung forensischer Methoden im Rechnungswesen mit Big Data.

2 Big Data zur betrieblichen Entscheidungsunterstützung

Die verstärkte Nutzung von Daten zur Entscheidungsfindung ist in jüngerer Zeit verstärkt in den Fokus praxeologischer und wissenschaftlicher Diskussionen gerückt [BuCM13]. Auslöser ist die schnelle Verbreitung von Technologien und Konzepten unter dem Schlagwort Big Data [MSMK13]. Häufig werden Beispiele aus dem Marketing herangezogen, um die Vorzüge von Big Data zu belegen [Hays04], [SMMP14]. Insbesondere wird auf die Steigerung der Produktivität, bspw. bei der Neukundenakquisition, durch den Einsatz datengetriebener Entscheidungsverfahren verwiesen [BrHK11].

Im Zusammenhang mit Big Data tauchen eine Reihe weiterer Begriffe auf, die teilweise überdeckend verwendet werden und die es deshalb abzugrenzen gilt. Unter Business Analytics [ChCS12] werden in erster Linie vorhersagende (predictive), aber auch beschreibende (descriptive) statistische Verfahren verstanden, die die Grundlage für Optimierungsmaßnahmen und die Unterstützung für Entscheidungsprozesse liefern. Das Objekt von Business Analytics kann variieren. Wichtige Bereiche sind die Prozess- und die Service-Analytik. Bei der Prozessanalytik werden Größen wie Kosten und Zeiten während der Durchführung von Geschäftsprozessen erfasst. Eine Sonderform stellt das Process Mining [KeMU06] dar, das der Rekonstruktion von Prozessmodellen dient. Die Service-Analytik wird insbesondere im Rahmen der Überwachung von Dienstgütereinbarungen verwendet.

Business Analytics ist häufig eng verbunden mit der Visualisierung von Analyseergebnissen. Business Analytics stützt sich häufig auf eine Kombination von Kompetenzen aus Statistik, Datenbanken und Domänenwissen, die mit dem Begriff Data Science [KeMU06] zusammengefasst wird. Von diesen Begriffen ist außerdem noch Business Intelligence [KeMU06] zu unterscheiden. Hierbei handelt es sich um einen ganzheitlichen Management-Ansatz, der bestrebt ist, alle für die betriebliche Entscheidungsfindung erforderlichen Faktoren zu koordinieren und zu optimieren.

Big Data

Im Allgemeinen wird unter Big Data eine deutliche Steigerung der Verarbeitungsfähigkeit von Daten bezüglich der Dimensionen Volumen (Volume), Geschwindigkeit (Velocity) und Vielfalt (Variety) verstanden [MCBB11]. Big Data löst daher bestehende Konzepte wie Business Intelligence [KeMU06] oder Business Analytics [DaHM10] nicht ab. Big Data ist keine spezifische Technologie und kein Algorithmus, sondern steht vielmehr für das Zusammenlaufen einer Reihe von technologischen Fortschritten. Die dabei auftretenden Steigerungen der Verarbeitungsfähigkeiten sind jedoch so umfangreich, dass

von einer neuen Qualität der Entscheidungsunterstützung gesprochen werden kann [DaBB12]. Typisch für Big Data ist der Verzicht auf die Leistungssteigerung einzelner Rechensysteme (Scale up) und stattdessen die Nutzung von möglichst vielen, parallel arbeitenden, einfachen Systemen (Scale Out) [AgDE11]. Eine häufig angewandte Strategie ist, die Bewegung von Daten zu vermeiden und stattdessen mehrere Berechnungsschritte lokal durchzuführen [AgDE11] .

Mit Big Data geht eine deutliche Steigerung der Fähigkeit einher, Daten mit großer Vielfalt zu verarbeiten [Bitk12]. So war die betriebliche Datenverarbeitung lange darauf beschränkt, klar strukturierte Informationen, die in Datenbanken abgelegt waren, zu verarbeiten. Durch den Einsatz von Big Data Technologien ist nun auch die Verarbeitung von Daten mit variabler Struktur möglich, d.h. auch Daten zu erfassen, deren Struktur nicht fix ist, sondern immer wieder neu in Erfahrung gebracht werden muss. Typische Einsatzmöglichkeiten von Big Data sind beispielsweise die Erzeugung von Produktempfehlungen die außerhalb der bisherigen Suchvorgänge des Kunden liegen, indem die Ähnlichkeit des Kunden mit anderen Kunden genutzt wird, als auch die Erhöhung der Prognosegenauigkeit [KeMU06]. Auch kann Big Data dazu eingesetzt werden, die Retourenzahl im E-Commerce zu reduzieren [MSKW13]. Inwieweit Big Data Anwendungen für Zwecke des forensischen Rechnungswesens eingesetzt werden können ist in der betriebswirtschaftlichen Literatur bislang wenig erforscht.

3 Methoden der datenbasierten Forensik

Die datenbasierte Forensik [Oliv09] hat sich bisher stark auf Daten aus relativ leicht zu erschließenden Quellen konzentriert. So werden bevorzugt Daten aus betrieblichen Anwendungssystemen wie ERP und CRM herangezogen. Das Schwergewicht bilden dabei transaktionale Daten, die geschäftliche Transaktionen widerspiegeln. Diese sind meist in Datenbanksystemen gespeichert und verfügen über eine klar definierte Struktur und Semantik. Die Einbeziehung weiterer Informationsquellen wird durch zwei Barrieren behindert, eine strukturelle und eine semantische (siehe Abbildung 1). Die strukturelle Barriere bezeichnet den Umstand, dass große Mengen von Daten bisher nicht genutzt wurden, da sie nur in semistrukturierter oder unstrukturierter Form vorlagen, d.h. ihre Verwendung erfordert aufwändige Import- und Aufbereitungsvorgänge [Abit97].

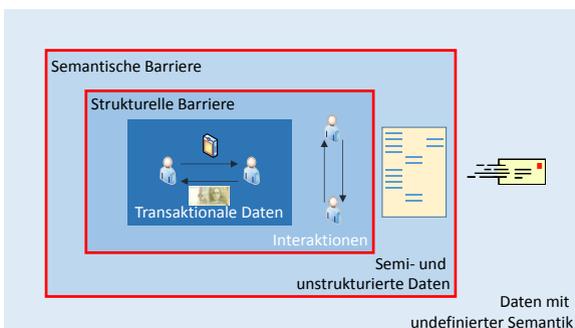


Abbildung 1: Barrieren für die Forensik

Die semantische Barriere verhindert die Nutzung von potentiell nützlichen Daten, da diese keine klar definierte Semantik besitzen. Dieser Umstand wird bspw. bei E-Mails deutlich. Diese sind im Gegensatz zu den durch die Schemadefinition klar beschriebenen Inhalten einer Datenbank kaum automatisch verarbeitbar.

Erweiterung der für die Forensik bereitstehenden Daten durch Big Data

Big Data erweitert deutlich die Menge der für die Forensik bereitstehenden Daten durch Überwindung der strukturellen Barriere und eine deutliche Verschiebung der semantischen Barriere, wie in Abbildung 2 dargestellt. So werden neben den strukturierten Daten beispielsweise solchen aus ERP-Systemen, auch weitere Datenklassen erfasst. Diese umfassen Daten außerhalb von geschäftlichen Transaktionen wie Interaktionen im Vorfeld oder im Nachlauf sowie benutzererzeugte Inhalte. Zu Interaktionen gehören beispielsweise Produkte deren Kauf zwar erwogen, aber dann doch noch verworfen wurde, indem sie aus dem Warenkorb entfernt wurden. Daten die Interaktionen wiedergeben, sind oft nur semi-strukturiert.

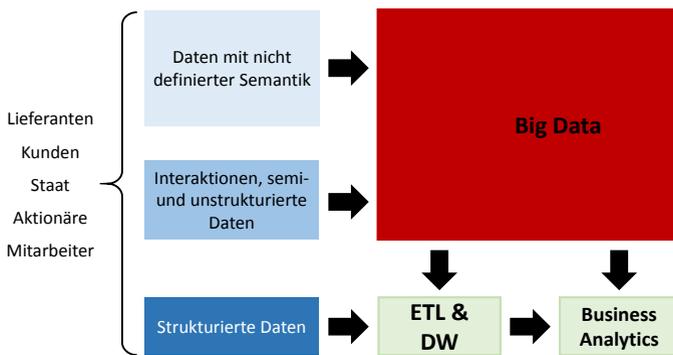


Abbildung 2: Big Data zur Vergrößerung des forensischen Potentials - in Anlehnung an [SMZW14]

Eine deutliche Verschiebung der semantischen Barriere findet bei der Einbeziehung benutzererzeugter Inhalte statt. Sind die bisher diskutierten Daten unter der Kontrolle des Unternehmens, können die Entstehung und Gestaltung benutzergenerierter Inhalte nicht mehr vom Unternehmen gesteuert werden. Erst recht ist es nicht möglich, die Bedeutung der verwendeten Begriffe zu standardisieren und somit einer direkten automatischen Bearbeitung zugänglich zu machen. Es liegt also eine sogenannte semantische Heterogenität [KeMU06] vor. Hierunter ist die unterschiedliche Interpretation von Begriffen in unterschiedlichen Kontexten zu verstehen. Bei den Daten unter der Kontrolle des Unternehmens ist eine eindeutige Interpretation möglich, bei den benutzergenerierten Inhalten ist dies nicht mehr der Fall. So kann es Unterschiede in der unternehmenseitigen Interpretation sowohl zwischen den unternehmenseigenen Daten und den benutzergenerierten Daten, als auch innerhalb der benutzergenerierten Inhalte geben. Letzteres bedeutet, dass der gleiche Begriff innerhalb der benutzergenerierten Inhalte unterschiedlich interpretiert werden muss.

4 Bezugsrahmen zur IT basierten Unterstützung forensischer Methoden im Rechnungswesen mit Big Data

Um dolose Handlungen wie etwa das Veruntreuen von Budgetteilen für organisationsfremde Zwecke zu erkennen, wird nachfolgender Bezugsrahmen zur IT gestützten Forensik im Rechnungswesen mit Big Data konzipiert (siehe Abbildung 3). Dolose Handlungen können im Unternehmen und in Beziehungen mit externen Partnern (wie bspw. Kunden, Lieferanten, Kapitalgebern und Behörden) auftreten [PLSG10] [Herz06]. Im Fokus dieses Bezugsrahmens werden derartige Handlungen in Zahlungsströmen bzw. Buchungen und Äußerungen der Beteiligten in Texten, wie z. B. (elektronischen) Handelsbriefen, untersucht. Auslöser dieser kriminellen Handlungen können externe Geschäftspartner, Fachmitarbeiter oder Manager sein [PLSG10] [Herz06].

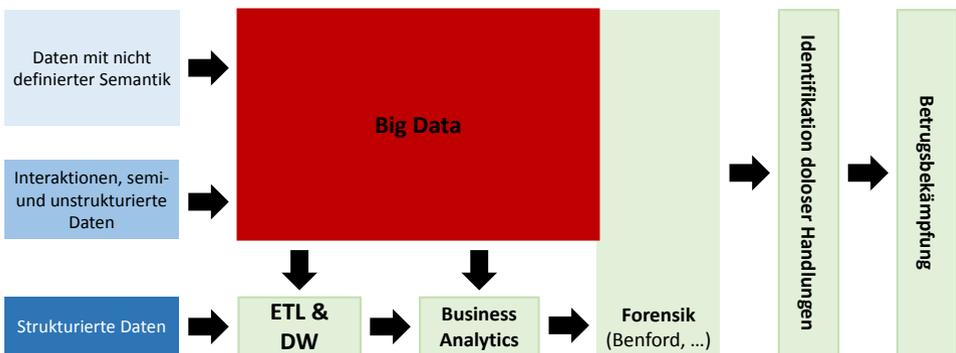


Abbildung 3: Bezugsrahmen zur IT basierten Unterstützung forensischer Methoden in Anlehnung an [SMZW14]

Die Verwendung von Big Data erweitert deutlich die für die Forensik und damit die Identifikation doloser Handlung zur Verfügung stehende Datenmenge wie in Abbildung 3 dargestellt. Wurden früher nur strukturierte Daten verwendet, die über die Zwischenstationen Datawarehouse mit Extract, Transform & Load der Business Analytics zur Verfügung bereitgestellt wurden, kann heute auf ein deutlich größeres Spektrum von Daten zurückgegriffen werden. Auch können Big-Data zugehörige Technologien zur direkten Unterstützung der Forensik verwendet werden. Der Einsatz von Big Data ermöglicht es große Mengen von semi- und unstrukturierten, sowie Daten mit nicht definierter Semantik für die die Forensik zu erschließen. Diese stammen in großen Teilen von externen Partnern, wie Lieferanten, Kunden usw. Somit kann eine durchgängige Erkennung von Betrugsfällen im Rechnungswesen organisationsübergreifend sichergestellt werden. Anschließend werden verschiedene Analysemethoden zur Erkennung doloser Handlungen angewandt. Dies kann für unstrukturierte Daten, wie elektronische Handelsbriefe, mittels Text-Mining bzw. Text-Klassifikationsverfahren erfolgen. Für strukturierte Daten, wie bspw. bei Buchungssätzen, werden Schwellenwert bzw. Musterverfahren (bspw. Benford) angewandt. Zur beschleunigten Verarbeitung dieser großen Datenmengen variabler Struktur, werden Technologien aus dem Big Data

Umfeld (wie bspw. In-Memory-Computing oder Hadoop) angewandt. Diese Verfahren werden zur Steigerung der Robustheit, je nach Anwendungsfall, kombiniert. Auf Basis der Analyseergebnisse können mögliche Betrugsfälle identifiziert und in einem weiteren Schritt diesen entgegengewirkt werden.

Im Folgenden werden einige Verfahren beispielhaft kurz erläutert. Die Möglichkeiten der Unterstützung durch Big Data je Verfahren im Rahmen des Bezugsrahmens sind am Ende des Abschnitts beschrieben und in Tabelle 3 dargestellt.

Analyse mittels Text-Mining

Oft werden dolose Handlungen im (elektronischen) Schriftverkehr von den jeweiligen Beteiligten kodifiziert [BBBK11]. Vor allem der elektronische Geschäftsverkehr via E-Mail zwischen Lieferanten und Kunden kann derartige Muster aufweisen. Jedoch stellt die Auswertung von unstrukturierten Daten (wie E-Mails) eine große Herausforderung für Unternehmen und Organisationen dar. Neue technologische Entwicklungen im Big Data Umfeld ermöglichen eine schnellere Verarbeitung von Texten, die semantische Heterogenität aufweisen, bspw. unterschiedliche Begriffsverwendungen. Zur Auswertung und Erkennung von Betrugsfällen kann eine Analyse mittels Text-Mining durchgeführt werden. Als Erweiterung des Data Mining dient Text-Mining der Extraktion von Informationen bzw. Wissen aus unstrukturierten Daten, wie Texten [Tan99]. Um dieses Wissen über mögliche dolose Handlungen aus Texten zu extrahieren, wird nachfolgender Text-Miningprozess nach Hippner und Rentzmann [HiRe06] vorgeschlagen (siehe Tabelle 1).

Prozessschritt	Bezug zur Forensik im Rechnungswesen
1. Aufgabendefinition	Erkennung von dolosen Handlungen in E-Mails
2. Dokumentselektion	E-Mails und Texte von Austauschlaufwerken mit Potenzial für Regelverstöße werden untersucht.
3. Dokumentaufbereitung	Aufbereitung durch Zerlegung der Texte in einzelne Wörter/Wortgruppen unter Berücksichtigung der Stammformreduktion (Stemming) und Löschung von Bindewörtern
4. (Text-) Mining Methoden	Filterung von betrugsrelevanten Wort- bzw. Wortgruppen
5. Interpretation / Evaluation	Interpretation ob ein Betrugsfall vorliegt
6. Anwendung	Ansprache der Beteiligten und Implementation von Gegenmaßnahmen

Tabelle 1: Text-Miningprozess [HiRe06]

Nach Abarbeitung der grundlegenden Prozessschritte kann im Schritt vier des Text-Mining-Prozesses nach Wörtern und Wortgruppen gesucht werden, welche Hinweise auf dolose Handlung darstellen. Dies können bspw. Wörter wie "stehlen", "verschieben", "hinterziehen", "Bestechung" sein. Diese Wörter und Wortgruppen sollten branchenspezifisch durch Analyse erkannter Betrugsfälle extrahiert werden.

Neben elektronischen Handelsbriefen (wie bspw. Lieferantenemails) können auch Lageberichte von Kunden und Lieferanten [HIKH11] anhand von spezifischen Wörtern wie z.B. Personalpronomina wie „wir“, „unserer Einschätzung nach“ etc. ausgewertet werden [HIKH11]. Dadurch können falsche Angaben von Lieferanten und Kunden oder Tochter- sowie verbundenen Unternehmen aufgedeckt werden. Zur Textanalyse werden häufig auch praktikable (Text-)Cluster zusammengestellt und gebildet [BeKu12].

Weiterhin sind bei der Analyse derartiger unstrukturierter Daten (wie E-Mails, Texte auf Firmenlaufwerken) gesetzliche Bestimmungen zu beachten [Schu10]. Je nach Anwendungsfall und vorhergehender Einverständniserklärung der Mitarbeiter ist eine derartige Analyse in der Bundesrepublik Deutschland weitestgehend möglich [Schu10].

Neben der Analyse unstrukturierter Daten können strukturierte Daten über verschiedene Methodiken analysiert werden, welche im nachfolgenden vorgestellt werden.

Analyse mittels Verteilungsfunktionen

Das Benford Gesetz [Benf38] [RaTh07] stellt eine Verteilungsfunktion des Auftretens von (führenden) Ziffern in der Natur dar. Danach ist das Vorkommen bestimmter Häufigkeiten von (führenden) Ziffern in der Natur nach folgender logarithmischer Verteilungsfunktion definiert:

$$p(\text{Ziffer}) = \log_{10}\left(1 + \frac{1}{\text{Ziffer}}\right)$$

Formel 1: Benford-Verteilung [Benf38]

Auf Basis der Untersuchung von Buchungstransaktionen bspw. im ERP-System können nun verdächtige Transaktionen in einem definierten Zeitraum ermittelt werden, wenn die führenden Ziffern der jeweiligen Buchungsbeträge weit von der Benford-Verteilung abweichen. Wenn beispielsweise oft Transaktionen unter der zustimmungspflichtigen Grenze von 1000 Euro (bspw. 970 Euro) durchgeführt werden, würde dies aufgrund der Abweichung zum normalen Verteilungsvorkommen (bei Ziffer 9: 4,57%) auffallen und könnte auf eine dolose Handlung hinweisen. Ein derartiges Beispiel ist in nachfolgender Grafik (Abbildung 4) dargestellt. Daten für eine solche Analyse können sowohl direkt aus dem IT-System extrahiert oder über die Verwendung von Datenaustauschformaten wie der XBRL [GeHe04] [DFOP09] etwa mit Daten von Tochterunternehmern gewonnen werden.

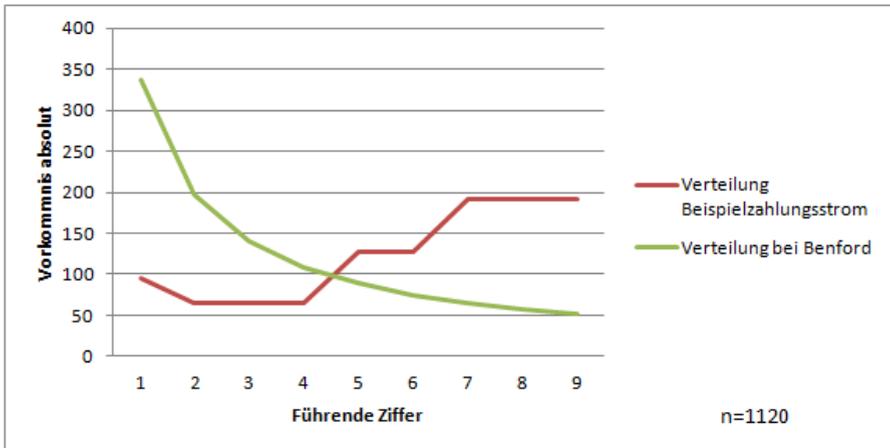


Abbildung 4: Beispielhafte Datenströme in Relation zur Benford-Verteilung

Zur Überprüfung der Passgenauigkeit der realen Zahlungsströme zu der Benford-Verteilung können Anpassungstests implementiert werden [Vogel12]. Die Verarbeitung großer Mengen an Buchungstransaktionen aus verschiedenen IT-System und deren Überprüfung auf die Benford-Verteilung kann durch Big Data Technologien beschleunigt und somit die Betrugsbekämpfung verbessert werden.

Als ein weiteres Testmaß kann zur Überprüfung der Chi-Quadrat-Test für eine Abweichung der Vorkommastelle bei Zahlungen verwendet werden [BHRW07] [BeKu12].

Weitere Analyseverfahren

Weitere daten-getriebene Verfahren zur Mustererkennung bzw. Ermittlung von Schwellenwerten, ab wann eine dolose Handlung vorliegt, sind beispielsweise [KLSH04], [PLSG10]:

- Künstliche Neuronale Netze
- Expertensysteme
- Data-Mining
- Model-based Reasoning

Diese Verfahren können mit bestehenden Methoden kombiniert werden, um die Robustheit zu steigern. Als Ergebnis werden diese in Regelsets definiert um eine einheitliche Vorgehensweise zu ermöglichen [KLSH04], [PLSG10].

Die vorgestellten Verfahren können zum einen durch eine schnelle Verarbeitung größerer Datenvolumina in kürzerer Zeit (wie bspw. bei der Anwendung bei Benford) oder durch die bessere Analyse von unstrukturierten Daten (wie bspw. beim Text-Mining) durch Big

Data Technologien unterstützt werden. Somit können dolose Handlungen besser und schneller als bisher erkannt werden und ihnen präventiv entgegen gewirkt werden. Dieses Nutzenpotenzial (+: hohes Nutzenpotenzial; -: geringes Nutzenpotenzial) ist in nachfolgender Tabelle zusammengefasst.

Analysemethode	Big Data Unterstützung		
	Volumen	Geschwindigkeit	(Hohe) Strukturvarianz
Benford-Analyse	+	+	-
Chi-Quadrat-Test	+	+	-
Text-Mining	+	+	+
Künstliche Neuronale-Netze	+	+	+
Entscheidungsbaum	+	+	+

Tabelle 2: Big Data Unterstützung

Bei der Analyse unstrukturierter Daten wie bspw. E-Mails und Handelsbriefe können Verfahren des Text-Mining unter Berücksichtigung von In-Memory-Computing und neuen Verarbeitungsalgorithmen wie Hadoop zur Aufdeckung doloser Handlungen verwendet werden. Somit kann ein großes Volumina an Daten, unstrukturierter Form in Kürze verarbeitet werden, um betrügerische Handlungen nach gegebenen Muster aufzudecken. Weiterhin können beispielsweise strukturierte Daten (wie etwa Kontoauszahlungen) im ERP-System ebenfalls durch Verwendung von obigen Big Data Technologien auf dolose Handlungen hin schneller als bisher überprüft und präventiv entgegen gewirkt werden. Diese Anwendungsszenarien können gemäß der Kriterien und Methodiken in Tabelle 2 vielfältig ergänzt werden.

5 Fazit und Diskussion

Durch Anwendung von IT-basierten Analyseverfahren können dolose Handlungen in Unternehmungen besser aufgedeckt werden als bisher. Damit kann die IT-Governance besser Geschäftsziele und Compliance-Bestimmungen einhalten und (IT-)Risiken minimieren. Weiterhin ermöglicht die Anwendung von Big Data Technologien große Volumen an Daten in kürzerer Zeit kosteneffektiv zu analysieren und auch unstrukturierte Daten (wie bspw. E-Mails zu Kunden und Lieferanten) hinsichtlich Compliance-Verletzungen zu analysieren. Bei der Anwendung dieser Verfahren ist jedoch zu beachten, dass einerseits nicht alle Betrugsfälle fehlerfrei erkannt werden und andererseits alle erkannten Fälle direkt einer dolosen Handlung entsprechen. Zum einen können neuartige Muster auftauchen, die die Algorithmen bisher nicht erkennen. Zum anderen können diese Verfahren nie alle Transaktionen fehlerfrei klassifizieren. Daher kann es vorkommen, dass normale Transaktionen als Betrug gekennzeichnet werden. Entsprechend sollte je nach Gewichtung des potenziellen Betrugsfalls eine manuelle Überprüfung stattfinden. Weiterhin sollte jede Unternehmung vorab prüfen, ob eine automatisierte Überprüfung und die damit verbundene IT-Integration wirtschaftlich sind. So können Transaktionen

aus Produktionssystemen (wie bspw. Betriebsdatenerfassungssysteme) ggf. keinen hohen Beitrag zu Betrugserkennung von Transaktionen liefern, wie dies etwa Accounting-Systeme können. Weiterhin sind bei der Analyse von unstrukturierten Daten wie Texte aus E-Mails oder Firmenlaufwerken gesetzliche Bestimmungen zu beachten und gegebenenfalls vorab Einverständniserklärungen einzuholen. Zukünftig sollten die Verfahren in praktischen Szenarien implementiert und auf branchenabhängige Determinanten untersucht und angepasst werden. So haben beispielsweise Transaktionen in Finanzunternehmen andere Schwellenwerte zur Betrugserkennung als dies bspw. in mittelständischen Maschinenbauunternehmen der Anwendungsfall ist.

Literaturverzeichnis

- [Abit97] Abiteboul, S.: Querying semi-structured data. Springer (1997).
- [AgDE11] Agrawal, D. et al.: Big data and cloud computing: current state and future opportunities. Proceedings of the 14th International Conference on Extending Database Technology. pp. 530–533 ACM (2011).
- [BBBK11] Bitzer, B. et al.: Tax Fraud & Forensic Accounting: Umgang mit Wirtschaftskriminalität. Springer DE (2011).
- [BeKu12] Berwanger, J., Kullmann, S.: Interne Revision. Springer DE (2012).
- [Benf38] Benford, F.: The law of anomalous numbers. Proc. Am. Philos. Soc. 551–572 (1938).
- [BHRW07] Boenner, D.-K.A. et al.: „STAAN: Standard Audit Analysis“–Teil II. ZIR Z. Interne Revis. Erich Schmidt Verl. Berl. Ausgaben. 6, 2006, 03 (2007).
- [Bitk12] BITKOM: Big Data im Praxiseinsatz – Szenarien, Beispiele, Effekte, http://www.bitkom.org/files/documents/Big_Data_BITKOM-Leitfaden_Sept.2012.pdf, (2012).
- [BrHK11] Brynjolfsson, E. et al.: Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? (2011).
- [BuCM13] Bughin, J. et al.: Ten IT-enabled business trends for the decade ahead | McKinsey & Company, http://www.mckinsey.com/insights/high_tech_telecoms_internet/ten_it-enabled_business_trends_for_the_decade_ahead.
- [Bund08] Bundesrepublik Deutschland: Gesetz über das Aufspüren von Gewinnen aus schweren Straftaten. (2008).
- [Bund65] Bundesrepublik Deutschland: Aktiengesetz. (1965).
- [ChCS12] Chen, H. et al.: Business intelligence and analytics: From big data to big impact. MIS Q. 36, 4, 1165–1188 (2012).
- [DaBB12] Davenport, T.H. et al.: How “Big Data” is Different. MIT Sloan Manag. Rev. 54, 1, 21–24 (2012).
- [DaHM10] Davenport, T.H. et al.: Analytics at Work: Smarter Decisions. Better Results Harv. Bus. Sch. Press Boston MA. (2010).
- [DFOP09] Debreceny, R. et al.: XBRL for interactive data. Springer (2009).
- [Frei08] Freidank, C.-C.: Corporate Governance und interne Revision: Handbuch für die Neuausrichtung des Internal Auditings. Erich Schmidt, Berlin (2008).
- [GaRe11] Gantz, J., Reinsel, D.: Extracting value from chaos. IDC IView. 1–12 (2011).
- [GeHe04] Gehra, B., Hess, T.: XBRL—Ein Kommunikationsstandard für das Financial Reporting. Control. Manag. 48, 6, 403–405 (2004).
- [Hays04] Hays, C.L.: What Wal-Mart Knows About Customers’ Habits, <http://www.nytimes.com/2004/11/14/business/yourmoney/14wal.html>, (2004).
- [Herz06] Herzog, H.: Risk & Fraud Management als erweiterte Managementaufgabe. ZRFG–Risk Fraud Gov. 1, 2006, 11–15 (2006).
- [HiRe06] Hippner, H., Rentzmann, R.: Text mining. Inform.-Spektrum. 29, 4, 287–290 (2006).

- [HIKH11] Hlavica, C. et al.: Fraud: Aufdeckung und Prävention durch Forensic Accounting. Tax Fraud & Forensic Accounting. pp. 176–246 Springer (2011).
- [Idw06] IDW, P.: 261: Feststellung und Beurteilung von Fehlerrisiken und Reaktionen des Abschlussprüfers auf die beurteilten Fehlerrisiken. Prüfungsstandard Inst. Wirtsch. Stand. 6, (2006).
- [JoGo06] Johannsen, W., Goeken, M.: IT-Governance–neue Aufgaben des IT-Managements. HMD-Prax. Wirtsch. 43, 250 (2006).
- [KeMU06] Kemper, H.-G. et al.: Business Intelligence: Grundlagen Und Praktische Anwendungen: Eine Einführung in Die IT-Basierte Managementunterstützung. Springer DE (2006).
- [Kirm12] Kirmes, R.: Private IT-Forensik und private Ermittlungen, zwei Seiten einer Medaille? eine Analyse der Begriffe, Rollen und legalen Betätigungsfelder für private IT-Forensik, zugleich Grundlegung für ein Berufsrecht der privaten IT-Forensik. Eul (2012).
- [KLSH04] Kou, Y. et al.: Survey of fraud detection techniques. Networking, sensing and control, 2004 IEEE international conference on. pp. 749–754 (2004).
- [MCBB11] Manyika, J. et al.: Big data: The next frontier for innovation, competition, and productivity. McKinsey Glob. Inst. 1–137 (2011).
- [MeZK03] Meyer, M. et al.: IT-Governance. Wirtschaftsinformatik. 45, 4, 445–448 (2003).
- [MSKW13] Möhring, M. et al.: Big Data – neue Möglichkeiten im E-Commerce. Wirtsch. Manag. 2013, 2, 48–56 (2013).
- [MSMK13] Möhring, M. et al.: Big Data – Implikationen für die Betriebswirtschaft. Wist - Wirtsch. Stud. 2013, 8, 456–458 (2013).
- [Oliv09] Olivier, M.S.: On metadata context in Database Forensics. Digit. Investig. 5, 3–4, 115–123 (2009).
- [PLSG10] Phua, C. et al.: A comprehensive survey of data mining-based fraud detection research. ArXiv Prepr. ArXiv10096119. (2010).
- [Pric11] Price Waterhouse Coopers: Studie zur Wirtschaftskriminalität 2011: Kommissar Zufall deckt am meisten auf, <http://www.pwc.de/de/risiko-management/studie-zur-wirtschaftskriminalitaet-2011-kommissar-zufall-deckt-am-meisten-auf.jhtml>.
- [Prüf02] Prüfungsstandard, I.D.W.: Abschlußprüfung bei Einsatz der Informationstechnologie (IDW PS 330). Wirtsch. 55, 21 (2002).
- [RaTh07] Rafeld, H., Then Bergh, F.R.: Digitale Ziffernanalyse in deutschen Rechnungslegungsdaten. ZIR. 2, 26–33 (2007).
- [Schu10] Schuster, F.P.: IT-gestützte interne Ermittlungen in Unternehmen – Strafbarkeitsrisiken nach den §§ 202a, 206 StGB. Z. Für Int. Strafdogmatik. (2010).
- [SMMP14] Schmidt, R. et al.: Big Data as Strategic Enabler - Insights from Central European Enterprises. In: Abramowicz, W., Kokkinaki, A. (Hrsg.). 17th International Conference on Business Information Systems (BIS) (in press). Springer LNBIP. pp. 50–60 Springer Berlin Heidelberg, Larnaca, Cyprus (2014).
- [SMZW14] Schmidt, R. et al.: Towards a Framework for Enterprise Architecture Analytics. EDOC Conference Workshop, 2014. EDOC'14, , Ulm (2014).
- [Tan99] Tan, A.-H.: Text mining: The state of the art and the challenges. Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases. pp. 65–70 (1999).
- [Voge12] Vogel, J.: Die Güte von Anpassungstests auf Benfordverteilung Arbeitsbericht Nr. 2012-01, Juli 2012. Technische Universität Ilmenau Fakultät für Wirtschaftswissenschaften Institut für Wirtschaftsinformatik, Ilmenau (2012).