DICO
Deutsches Institut für Compliance

Datenanalysemethoden für die Compliance

Wesentliche Ziele bei Untersuchungen zur Einhaltung von Compliance sind es herauszufinden, warum Regularien nicht eingehalten wurden, zu überprüfen, ob gerade Auffälligkeiten auftreten, die auf ein potenzielles Fehlverhalten hindeuten, und vorauszusehen, wie hoch die Gefahr einer sich anbahnenden Fehlentwicklung ist. Die Untersuchungen können daher rückwärtsgerichtet auf Geschehnisse schauen, momentane Geschehnisse im Sinne eines Monitorings überwachen oder vorwärtsgerichtet zukünftige Geschehnisse vorhersagen bzw. deren Wahrscheinlichkeiten bestimmen. Diese Betrachtungen können auf der Analyse von gesammelten Daten basieren. Daten können hierbei heterogen sein und sowohl strukturierte als auch unstrukturierte Daten enthalten. Datenanalysemethoden aus dem Bereich des maschinellen Lernens, der Mustererkennung, der Data Science und der visuellen Analytik können zum Einsatz kommen.

Dieser Kurzbeitrag soll die Möglichkeiten der Datenanalyse für die Compliance beleuchten und diskutieren.

Daten

Viele Datenanalysemethoden arbeiten auf strukturierten Daten. Strukturierte Daten sind Daten, die nach einem fest vorgeschriebenen Schema erfasst und gespeichert werden. Ein gängiges Beispiel für strukturierte Daten sind Daten die in Tabellen (oder Spreadsheets) gespeichert werden. Dabei können die Tabelleneinträge numerisch (Zahlenwerte) oder kategorisch (textuelle Beschreibungen) sein. Ein gängiges Beispiel wären Daten zur Ressourcenplanung (Enterprise Resource Planning), wo Informationen über Personal und materielle Ressourcen in tabellarischer Form zusammengefasst sind. Bei Eventdaten spielt zudem die zeitliche Komponente eine wichtige Rolle, gegebenenfalls auch im Zusammenspiel mit einer räumlichen Komponente (Georeferenz). Beispiele hierfür sind Finanztransaktionsdaten oder auch organisatorische Prozesse.

Ein weiterer Aspekt sind Netzwerke, die durch Organisationsstrukturen aber auch durch Beziehungen, Zusammenarbeit und Kommunikation entstehen. So kann beispielsweise die interne Korrespondenz von Mitgliedern einer Organisation oder Unternehmens sowie deren externe Korrespondenz mit Vertretern anderer Organisationen oder Unternehmen in einem Netzwerk zusammengefasst werden, das auch die Stärke der Beziehungen erfassen kann.

Wenn es um die Inhalte von Korrespondenzen wie E-Mails oder Chat-Nachrichten geht, dann liegen hier Textdaten vor. Diese sind der wesentliche Vertreter von unstrukturierten Daten. Die Texte sind meist frei formuliert und unterliegen keiner Struktur (im Gegensatz zu beispielsweise Textdaten in XML-Formaten). Unstrukturierte Textdaten treten nicht nur in Korrespondenzen, sondern insbesondere auch bei Berichten auf.

Betrachtet man komplexe Vorgänge im Rahmen einer Compliance-Untersuchung, dann können Daten unterschiedlicher Quellen zusammengefügt werden, die auch von unterschiedlichem Typ sein können wie beispielsweise Eventdaten und Textdaten. Man spricht dann von heterogenen Daten.

Datenanalyse

Die oben beschriebenen Daten liegen heutzutage meist in digitaler Form vor und können daher unmittelbar als Basis einer Datenanalyse dienen. Da die Datenmengen sehr groß sind und die Daten komplex, können Sie nicht manuell von Menschen durchgeführt werden und es bedarf der Unterstützung durch automatische Analysemethoden, die auf Computern ausgeführt werden.

Ein grundlegendes Problem der Mustererkennung ist es, die vorhandenen Daten in Gruppen hoher Ähnlichkeit (so genannten Clustern) zusammenzufassen. Man spricht hier auch von unüberwachtem Lernen, da dem Erkennen von Clustern keine Vorgaben gemacht werden. Durch das Erstellen von Clustern lassen sich wesentliche Strukturen oder Merkmale in den Daten erkennen. Von Interesse sind dann auch Auffälligkeiten in der Form, dass Daten vorliegen, die nicht einem dieser Cluster zugewiesen werden können. Solche Ausreißer deuten auf Anomalitäten hin, deren Ursache dann untersucht werden kann. Wenn man davon ausgeht, dass Verstöße gegen die Compliance eher eine Ausnahme darstellen, dann können diese als Ausreißer detektiert werden.

Eine weitere Vorgehensweise wäre die, dass man nach speziellen bekannten Mustern sucht (z.B. Finanztransaktionen großer Beträge in Länder mit gewissen politischen Strukturen). Sind solche Muster bekannt und gab es dafür in der Vergangenheit Beispiele, dann kann man ein überwachtes Lernen durchführen, wo verschiedene Muster in Form von Klassen vorgegeben werden und dann ein System trainiert wird, das es erlaubt, neue Daten zu klassifizieren. So könnte beispielsweise ein Monitoring stattfinden. Man könnte das System trainieren, indem man es mit Daten füttert und dem System mitteilt, ob ein Verstoß gegen die Compliance vorlag oder nicht bzw. auch welches Vergehen vorlag. Es bedarf allerdings einer ausreichend großen Menge an historischen Daten, aus denen man lernen kann.

Bei zeitlichen Prozessen ist die Vorhersage von in Zukunft auftretenden Events ein großes Ziel der Datenanalyse. Hier werden meist mathematische Modelle für die Evolution der Daten verwendet und an historische Daten angepasst, um die momentane Situation einzuschätzen und Vorhersagen für die Zukunft zu tätigen. Bei wiederkehrenden Mustern funktionieren diese Vorhersagen sehr gut, sie können jedoch auch keine unerwartbaren Ereignissen (insbesondere solche, wo es keine Präzedenzfälle gibt) vorhersagen. Gibt es jedoch wiederkehrende Anzeichen für sich anbahnende Compliance-Verstöße, dann könnten diese vorzeitig erkannt werden.

Untersucht man Netzwerke, dann stellen diese mathematisch Graphen dar, wo die Knoten des Graphens beispielsweise einzelne Personen sein könnten und die Kanten zwischen diesen Knoten die Beziehungen (oder Relation) darstellen. Die Kanten können auch Gewichte haben, um die Stärke der Beziehung zu repräsentieren.

Zum Einsatz kommen dann Graphanalysemethoden, die die wesentlichen topologischen Strukturen der Graphen untersucht.

Hier könnte man beispielsweise erkennen, wenn eine Person eines Unternehmens als wesentliche Bezugsperson für außenstehende Unternehmen dient.

Methoden der Datenanalyse widmen sich klassisch meist einem Datentyp, aber die Verarbeitung heterogener Daten gewinnt zunehmend mehr Beachtung.

Visuelle Analytik

Die oben aufgeführten Methoden der Datenanalyse beinhalten automatische Komponenten, die von einem Compliance Officer, der die Analyse durchführt oft als Black Box wahrgenommen werden. Für die Akzeptanz der Methode und das Vertrauen in das Ergebnis, ist es wichtig, den Analyseprozess zu verstehen. Die visuelle Analytik stellt einen nutzerzentrierten Ansatz der Datenanalyse dar. Der Compliance Officer steuert interaktiv den Analyseprozess und bekommt Zwischenresultate zurückgeliefert, mit denen er/sie den Prozess verfeinern kann. Da visuelle Darstellungen für den Menschen intuitiv und effizient zu verarbeiten sind, werden dem Compliance Officer die Informationen in Form von Visualisierungen zur Verfügung gestellt. Die visuellen Darstellungen erlauben dann eine interaktive Exploration und Verfeinerung der Information, wobei man meist mit Überblicksdarstellungen beginnt und während des Analyseprozesses mehr und mehr Details zur Verfügung gestellt werden. Zur Verarbeitung der Daten in den einzelnen Schritten dienen die oben beschriebenen automatischen Komponenten.

Anwendung in der Praxis

Wenn man davon ausgeht, dass die Mitarbeiter eines Unternehmens Fehlverhalten, wie z.B. Bestechungszahlungen oder andere unberechtigte Zahlungen verschleiern wollen, finden sich häufig erste Anzeichen hierfür in den Zahlungsströmen oder auch in der Kreditorenbuchhaltung.

Analyse von Zahlungsströmen

Für eine anlassunabhängige Aufdeckung beispielsweise von Bestechungszahlungen aber auch Delikte aus dem Bereich der Geldwäsche lassen sich u.a. folgende Transaktionen für eine Datenanalyse heranziehen:

- Analyse von Zahlungen an Bankverbindungen oder Zahlungsempfänger mit Sitz in Ländern mit hoher bekannter Korruption,
- Analyse von Zahlungen an Bankverbindungen oder Zahlungsempfänger mit Sitz in Ländern mit erhöhtem Geldwäscherisiko,
- Analyse von Zahlungen an Bankverbindungen oder Zahlungsempfänger mit Sitz in Ländern, die als sog. „Steueroasen“ bekannt sind,

- Zahlungen an Geschäftspartner, deren wirtschaftlich Berechtigte nicht bekannt oder verschleiert sind,
- Zahlungen an Empfänger, die nicht in den Stammdaten hinterlegt sind, sondern über CPD-Konten laufen,
- Zahlungen (oder sonstige Zuwendungen) an aus- und inländische Amtsträger oder sonstige sog. PEP (polically exposed persons) und deren näheres Umfeld,
- Abweichungen des Landessitzes der Empfängerbank und des Zahlungsempfängers.

Bei der Analyse von Zahlungsströmen ist es notwendig, über eine klassische Signaturanalyse hinauszugehen, da Auffälligkeiten über mehrere – oft zeitlich nicht direkt aufeinander folgende – Transaktionsströme einzubeziehen sind.

Digitale Analyse der Geschäftsbeziehungen

Unabhängig von der Analyse der Zahlungsströme und der Kreditorenstammdaten, kann auch Augenmerk auf die Analyse der Bewegungsdaten einer Geschäftsbeziehung gelegt werden. Dies passiert i.d.R. bevor die eigentliche Zahlung ausgelöst wird und bietet ebenfalls Potential, um Anzeichen für die Verschleierung frühzeitig zu erkennen.

Beispielhaft sind hier folgende Analysehandlungen zu nennen:

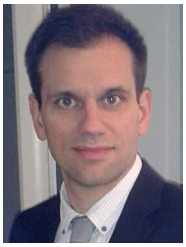
- Analyse der Zahlungen an Geschäftspartner ohne Vertragsgrundlage (auch der Zahlungseingang ohne Geschäftsgrundlage kann auffällig sein),
- Analyse von Zahlungen ohne Rechnung,
- Analyse von Zahlungen ohne Wareneingang oder Leistungsnachweis (z.B. für Beratungsleistungen oder Spesen bzw. sonstige Auslagen),
- mögliche Lieferungen ohne Zahlung,
- mögliche Lieferungen ohne Rechnungen,
- Altersanalyse von Rechnungen mit niedriger Mahnstufe,
- Analyse von Geschäftsbeziehungen mit sehr hohen Rabatten oder ungewöhnlich langem Zahlungsziel,
- Analyse von Kreditlinien auf Auffälligkeiten,
- Betrachtung von außergewöhnlichen Preisunterschieden bei gleichartigen Geschäftsvorfällen,
- Analyse von ungewöhnlichen Gutschriften.

Um die Anzahl der false positive Alarme einzudämmen und die für die Analyse wesentlichen Ereignisse von den uninteressanten Ereignissen zu trennen, kommen hier bereits Algorithmen des überwachten maschinellen Lernens zum Einsatz. Damit kann die Analytics Engine anhand der Entscheidungen des Bearbeiters die zu meldenden Ergebnisse höher priorisieren.

DICO
Deutsches Institut für Compliance

Fazit

Moderne Methoden der Datenanalyse bieten zahlreiche Möglichkeiten, um die Compliance zu unterstützen. Teilweise werden diese auch schon in der Praxis eingesetzt, aber insbesondere im Bereich der Analyse heterogener Daten, wo Daten unterschiedlicher Struktur aus unterschiedlichen Quellen zusammengeführt werden, gibt es noch Potenzial, das nicht erschöpft ist. Solche Analysen sind aber auch durchaus noch Teil der Forschung, so dass sich hier auch für die Forschung interessante Anwendungsgebiete auftun. Zuletzt sei erwähnt, dass dieser Beitrag lediglich die Möglichkeiten diskutiert, die Datenanalysemethoden bieten. Die Gefahren, die damit einhergehen, insbesondere in Bezug auf Datenschutz, stellen einen eigenen Diskussionspunkt dar, der hier nicht eingehender betrachtet wurde.



Prof. Dr.-Ing. Lars Linsen, Westfälische Wilhelms-Universität Münster,
Mitglied des wissenschaftlichen Beirats

Alexander Geschonneck, Partner KPMG Forensic und Co-Leiter des DICO Arbeitskreises "Digitale Transformation"