



© MicroStockHub / Getty Images / iStock

# Besser entscheiden durch Business Intelligence

Zur Analyse digitalisierter Geschäftsprozesse hat sich das Konzept der Geschäftsanalytik (Business Intelligence) etabliert. Für unstrukturierte Informationen kommen hier Verfahren aus dem Bereich Data-Analytics zur Anwendung. Solche Datenanalysen werden im Studiengang Verwaltungsinformatik der Hochschule des Bundes thematisiert und bereiten Absolventen auf eine Tätigkeit als Data-Scientist vor.

Die Digitalisierung von Verwaltungsprozessen bietet im Rahmen der bundesweiten E-Government-Initiativen, neben dem enormen Beschleunigungs- und Automatisierungspotenzialen digitalisierter Prozesse, ebenfalls Erkenntnisgewinne hinsichtlich zusätzlicher Prozessoptimierungen. Wesentlicher Baustein hierfür bildet die Geschäftsprozessanalytik der Business Intelligence (BI) und der darin enthaltenen Methoden zur Sammlung, Darstellung und Auswertung der zugrunde liegenden Geschäftsdaten sowie

## Kompakt

- Die Digitalisierung der öffentlichen Verwaltung wird auch eine (teil)automatisierte Entscheidungsunterstützung mit sich bringen, die auf Algorithmen der künstlichen Intelligenz basiert.
- Verwaltungsinformatikerinnen und Verwaltungsinformatiker werden als Data-Scientists die notwendige Datenanalyse mit Methoden des maschinellen Lernens sowie der Mustererkennung durchführen und so zur Effizienz- und Laufzeitoptimierung der Geschäftsprozesse beitragen.
- Die automatische Analyse von Texten durch Methoden des Text-Minings hilft bei der raschen Bearbeitung von Vorgängen.

## Autoren

**Prof. Dr.-Ing. Ulrich Lohmann**

ist Leiter des Studienbereichs „Angewandte und Technische Informatik“ im Studiengang Verwaltungsinformatik am Fachbereich Finanzen der Hochschule des Bundes.

ulrich.lohmann@vit-bund.de

**Prof. Dr.-Ing. Jörg Schmittwilken**

ist Leiter des Studienbereichs „Praktische Informatik“ im Studiengang Verwaltungsinformatik am Fachbereich Finanzen der Hochschule des Bundes.  
joeg.schmittwilken@vit-bund.de

**Julian Rommel**

ist Dipl.-Verwaltungswirt (FH) und arbeitet im Informationstechnikzentrum Bund im Arbeitsbereich Unified Communications.

der Extraktion von Information aus diesen Basisdaten. Diese extrahierte Information soll in der Regel eine faktenbasierte Entscheidungsunterstützung zur Optimierung der Prozesse liefern.

Business Intelligence bietet hierzu ein Bündel an Technologien und Methoden zur strukturierten Datenanalyse im Sinne dieser Entscheidungsunterstützung zur Gestaltung effizienz- und laufzeitoptimierter Geschäftsprozesse. Interessante Möglichkeiten ergeben sich zum Beispiel

hinsichtlich teilautomatisierter Vorklassifizierungen von Dokumenten oder Kennzahlen im Rahmen eines effizienten Prozesscontrollings.

Das Anwendungspotenzial der BI ist daher auch in der Verwaltung immens, da die Extraktion von entscheidungsunterstützender Information aus existierenden Rohdaten in vielen Bereichen (und auf den unterschiedlichen Ebenen) der Verwaltung die oben genannten Beschleunigungseffekte erwarten lässt. Die erforderlichen

Analysedaten der unterschiedlichen Verwaltungsprozesse hierfür liegen häufig bereits digitalisiert vor.

Dieses Anwendungspotenzial zeigt sich ebenfalls in den zum Teil hohen Preisgeldern der international ausgetragenen Wettbewerbe im Bereich Data-Analytics und Business Intelligence. Auf den gängigen Data-Analytics-Portalen wie Kaggle & Co. werden die Lösungen zu Fragestellungen der BI und Datenanalyse von renommierten Organisationen und Unternehmen im Rahmen von Data-Analytics-Competitions mit Preisgeldern im hohen sechsstelligen Bereich honoriert. Hier finden unter anderem Wettbewerbe zur prädiktiven Analyse von Kreditwürdigkeiten sowie zur Mustererkennung von Handschriften oder der Gesichtserkennung statt.

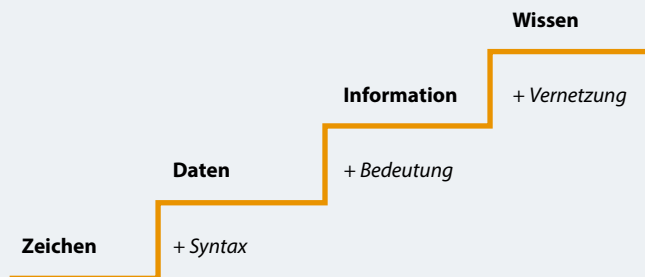
## Der Studiengang Verwaltungsinformatik

Der Studiengang Verwaltungsinformatik der Hochschule des Bundes für öffentliche Verwaltung (HS Bund) vermittelt die erforderlichen Grundlagen der Business Intelligence und der Data-Analytics in entsprechenden Vorlesungsreihen zum Thema (unter anderem in den Studienmodulen „ERP- und BI-Systeme“ sowie „Informations- und Wissensmanagement“). In diesen Vorlesungsreihen wird fundiertes Wissen zu den Methoden, Algorithmen und der Architektur von BI-Systemen als „Rüstzeug des Data-Scientists“ vermittelt. Das erworbene theoretische Wissen über Tools und Algorithmen wird in digitalen Lernumgebungen, den sogenannten „Virtuellen Laboren“ praktisch angewandt. Die Studierenden werden somit befähigt, potenzielle Anwendungsszenarien hinsichtlich der Möglichkeiten von BI praktisch zu evaluieren und die theoretische Methodik praktisch anzuwenden. Virtuelle Labore stehen für eine Vielzahl von Themen zur Verfügung, vom Datenbank-Server über die SAP-Appliance bis hin zum lernenden Neuronetz haben die Studierenden die Möglichkeit, praktische Erfahrungen mit gängiger Software „State of the Art“ zu sammeln.

### Data-Analytics zur Entscheidungsunterstützung

Die Notwendigkeit von Datenanalyse sowie die Relevanz der entsprechenden Ergebnisse lassen sich anhand der Wissensstreppe (North 2016) in Abbildung 1 erläutern: Durch Kenntnis der Syntax von Zeichen entstehen Daten und durch Kenntnis der Bedeutung der Daten entstehen Informationen. Schließlich entsteht Wissen erst durch Vernetzung von Informati-

Abbildung 1: Ausschnitt der Wissenstreppe



Quelle: eigene Darstellung, nach North, 2016

onen. Die begriffliche Trennung und Hierarchie der Konzepte Zeichen, Daten, Information und Wissen zeigen die Notwendigkeit der Datenanalyse zur Wissensgewinnung und somit auch zur Entscheidungsunterstützung. Aktuelle Buzzwords wie künstliche Intelligenz zeigen, dass auch die vierte Stufe der Wissenstreppe, nämlich das Wissen selber, mittlerweile auch von Maschinen repräsentiert werden kann und nicht mehr länger eine rein menschliche Domäne ist. Jedoch werden die Konfiguration und Orchestrierung der im Folgenden vorgestellten Methoden vorerst Aufgabe von Spezialistinnen und Spezialisten sein, die als Data-Scientists bezeichnet werden.

Typischerweise kommen im Bereich Data-Analytics Verfahren der Mustererkennung sowie des maschinellen Lernens zur Anwendung. Beide werden in überwachte und unüberwachte Verfahren differenziert. Algorithmen der Mustererkennung dienen dazu, das Auftreten bekannter oder unbekannter Muster in Daten zu identifizieren. Anwendungen sind beispielsweise die Identifikation von Gesichtern (bekannte Muster) in Bildern oder die Identifikation ähnlicher Krankheitsbilder (unbekannte Muster) anhand der Laborwerte von Blutanalysen. Bekannte Muster müssen den Algorithmen in geeigneter, zumeist mathematischer Form beschrieben werden. Häufig dient Mustererkennung der Klassifikation von Daten, also der Zuordnung der Daten zu bekannten Klassen oder zur Gruppierung in bisher unbekannte Klassen. Die Identifikation von Spam-E-Mails besteht

beispielsweise aus den beiden Schritten der Erkennung typischer Muster wie bestimmter Wortkombination und schließlich der Einstufung (Klassifikation) einer E-Mail als Spam.

Beim maschinellen Lernen (engl. *machine learning*) werden die Muster aus Daten, die für die Mustererkennung notwendig sind, gelernt. Überwachte Lernverfahren benötigen dazu bereits klassifizierte Daten (sogenannter Trainingsdatensatz), anhand derer sie die Muster lernen können. So könnte es beispielsweise von Interesse sein, die Kreditwürdigkeit von Personen basierend auf Kriterien wie Alter, Familienstand, Jahreseinkommen oder der Anzahl und Höhe laufender Darlehen zu lernen. Ein überwacht Lernverfahren benötigt dazu einen Datensatz, der für viele Personen neben diesen Kriterien auch die Zielklassifikation, nämlich deren Kreditwürdigkeit, beschreibt. So ein Trainingsdatensatz könnte aus den Kundendaten eines Kreditinstituts anhand des Tilgungsverlaufs bereits getilgter Darlehen abgeleitet werden. Das aus dem Trainingsdatensatz gelernte Modell kann dann zur Klassifikation unbekannter Daten, also neuer Kunden des Kreditinstituts, genutzt werden.

### Anwendung Text-Mining

Eine besondere Art von Data-Analytics ist das Text-Mining, das Algorithmen zur Analyse von un- oder schwachstrukturierten Texten zusammenfasst. Als exemplarische Anwendung von Verfahren des Text-Minings wurde im Rahmen einer Diplomarbeit (Rommel 2018) im Studi-

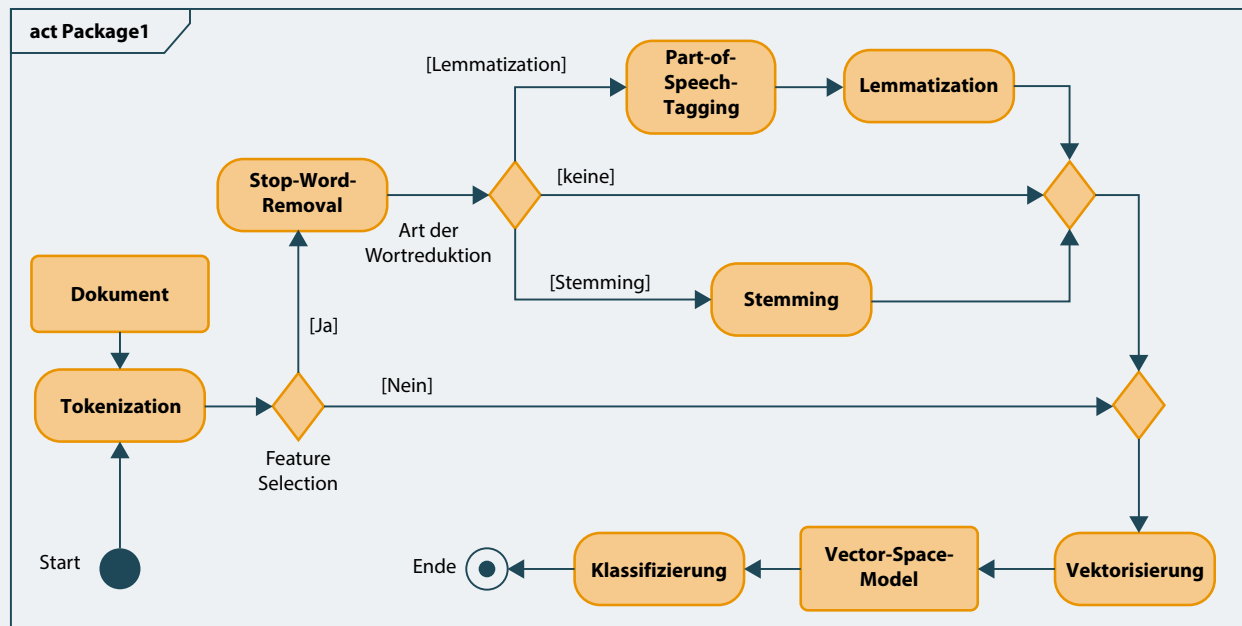
engang Verwaltungsinformatik der HS Bund die Effektivität unterschiedlicher Klassifizierungsverfahren für Texte verglichen. Dabei wurden behördlicher Textdokumente analysiert und unterschiedliche Realisierungen der Klassifikation bewertet. Die Ergebnisse der durchgeführten Untersuchungen werden im Folgenden dargestellt.

Eine Form des Text-Minings ist die Klassifikation von Dokumenten. Das Ziel der Klassifizierung ist es, eine Instanz, also ein Dokument, in eine der vorher definierten Klassen einzuordnen (Feldman/Sanger 2008, S. 64). Dies kann mit Algorithmen aus dem Bereich der überwachten Lernverfahren realisiert werden. Der Algorithmus muss also mit Dokumenten, bei welchen die Klassen bereits bekannt sind, trainiert werden, um anschließend neue und unbekannte Dokumente zu klassifizieren (Marsland 2009, S. 7). Ziel der Diplomarbeit war es, Methoden und Algorithmen zu vergleichen, die geeignet sind, eine solche Klassifizierung von Dokumenten durchzuführen. Der Fokus lag dabei auf der Untersuchung verschiedener überwacht Lernverfahren. Es sollte die Frage beantwortet werden, ob sich die Algorithmen für einen Einsatz in der Praxis der öffentlichen Verwaltung als Klassifizierer von Dokumenten eignen.

Es wurde die Effektivität der Algorithmen gemessen, sodass auch ein Leistungsvergleich erstellt werden konnte. Die Experimente wurden mit einer Dokumentensammlung durchgeführt, die über 7.500 englischsprachige Manuskripte von Reden enthält, die zwischen 1970 und 2016 bei den UN-Generaldebatten gehalten wurden. Diese stammen aus 199 verschiedenen Herkunftsländern, welche den Klassen entsprechen (Baturu et al. 2017). Bei den getesteten Machine-Learning-Algorithmen handelte es sich um die Support-Vector-Machine, den Random-Forest und das Multi-Layer-Perceptron, eine besondere Art künstlicher neuronaler Netze.

Abbildung 2 zeigt das UML-Aktivitätsdiagramm der einzelnen Schritte der Klassifikation. Bevor die Klassifizierer

Abbildung 2: UML-Aktivitätsdiagramm zum Ablauf der Klassifikation von Textdokumenten



Quelle: eigene Darstellung

die Dokumente verarbeiten können, ist es notwendig, ein Preprocessing durchzuführen. Dabei werden aus den Dokumenten repräsentative Merkmale extrahiert und die Dokumente in Vektoren mit numerischen Werten überführt (Weiss et al. 2010, S. 3 ff.).

Eine Möglichkeit hierfür ist etwa das Bag-of-words-Modell. Dafür werden die Dokumente in Token aufgeteilt, es wird ein Wörterbuch mit allen Token der Dokumentensammlung gebildet und für jedes Dokument gezählt, wie oft der Token darin vorkommt (Feldman/Sanger 2008, S. 68). Um die Aussagekraft dieser Vektoren für die einzelnen Dokumente zu erhöhen, gibt es eine Vielzahl an Möglichkeiten wie beispielsweise das Tf-idf-Maß, Stop-Word-Removal, Stemming, Lem-

matization und die statistische Feature Selection.

In der verwendeten Dokumentensammlung wurden die besten Ergebnisse mit der Darstellung im Tf-idf-Maß und der Nutzung des Stop-Word-Removals erzielt. Die Effektivität der anschließenden Klassifizierung ist in Abbildung 3 dargestellt. Sie wird durch den f1-score (vgl. Sasaki 2007) ausgedrückt. Es zeigt sich, dass es etwa mit dem Multi-Layer-Perceptron möglich ist, voll automatisiert mit einer Genauigkeit von 97,3 Prozent den Dokumenten das richtige Herkunftsland des Redners zuzuweisen.

Da sich die Ergebnisse nicht direkt auf andere Dokumentensammlungen übertragen lassen, sollte für einen Einsatz in der Praxis nochmals eine zusätzliche

Evaluation durchgeführt werden. Dennoch eignen sich anhand der Ergebnisse alle drei getesteten Algorithmen als Klassifizierer für Dokumente. Die Genauigkeit hängt stark vom verwendeten Datensatz ab, ist aber mittels weiterer Preprocessing-Maßnahmen oder anderen Implementierungen von Algorithmen theoretisch noch steigerbar. ■

Literatur

Baturo, A., Niheer D., Slava M. (2017): Understanding "State Preferences With Text As Data", in: Research & Politics 4 (2).  
 Feldman, R., Sanger J. (2008): The text mining handbook, Cambridge.  
 Marsland, S. (2009): Machine learning: An Algorithmic Perspective, Chapman & Hall/CRC.  
 North, K. (2016): Wissensorientierte Unternehmensführung: Wissensmanagement gestalten, Wiesbaden.  
 Rommel, J. (2018): Text Mining in der Bundesverwaltung – Analyse von Algorithmen und Einsatzmöglichkeiten, Diplomarbeit, Hochschule des Bundes für öffentliche Verwaltung.  
 Sasaki, Y. (2007): The truth of the F-measure, School of Computer Science, University of Manchester.  
 Weiss, S. M., Nitin I., Tong Z. (2010): Fundamentals of Predictive Text Mining, London.

Abbildung 3: Effektivität der Klassifizierung

Verfahren	f1-score
Multi-Layer-Perceptron	97,3 %
Support-Vector-Machine	95,7 %
Random-Forest-Algorithmus	94,0 %

Quelle: eigene Darstellung



Business Intelligence



Schneider, A. (2018): Transparenz und Effizienz durch Business Intelligence, in: innovative Verwaltung, 6, S. 14-17, www.springer-professional.de/link/15830280