tdwi VEREIN

TDWI AWARD

Bachelor- und Masterarbeiten im Bereich Analytics und Business Intelligence

2018

Abstracts der eingereichten Arbeiten

INFOMOTION we love data

tdwi.eu

Inhaltsverzeichnis

Vorwort 3

Dritter Platz

Sebastian Dierdorf 4
Fallstudienbasierte Analyse und Bewertung von Methoden des Maschinellen Lernens zur Bestimmung der wirtschaftlich optimalen Losgröße

Simon Frank 6
Cloudbasiertes ETL – Entwicklung & Validierung eines kriterienbasierten Ansatzes zur Bewertung geeigneter Cloud-ETL-Lösungen

Robin Gubela 8
Uplift Modeling – an Introduction of Revenue Modeling to Maximize Campaign Profit

Zweiter Platz

Christian Höcherl 9
Adaption und Evaluation von vortrainierten Konvolutionsnetzen im Anwendungsgebiet der Gewinnerkennung von Rubbellosen auf einer synthetischen Datenbasis

Daniel Jung 12
Leitfaden zur Implementierung einer unternehmensweiten Business-Intelligence

Merlin Knäble 14
Designing Uncertainty-Aware Dashboards for Application Lifecycle Management

Florian Kobylarz 16
Methoden der Datenvisualisierung und ihre Anwendung in der Praxis am Beispiel des Service-Bund

Romana Moser 18
Social Media Analytics in Banking: Measurement of Social Media Activities in Austrian Banks

Erster Platz

Jan Pennekamp 20
Uncovering Doppelgängers in Online Communities

Simon Schmid 22
Next Level User Experience für Self-service BI: sprachgesteuerte Datenanalyse mit Alexa for Business

Daniel Schnappauf 24
Anforderungen und Potenziale von Self-Service-BI im Finanzsektor

Johannes Schwäble 26
Spezifikation endbenutzerfähiger Advanced-Analytics-Lösungen – Entwicklung eines Vorgehenskonzeptes auf der Grundlage einer Einzelfallstudie im Bereich Data & Analytics Mercedes-Benz Cars

Stephan Volkmann 30
Datenmodellierung durch Beispiele unter Verwendung natürlicher Sprache zur Digitalisierung eines Prozesses im Musterbau

Adrian Weber 34
Datenanalyse in nahezu Echtzeit für einen IoT-Showcase

Konstantin Weise 36
Analytics and Big Data in Sales Controlling – Status Quo, Challenges and Opportunities within the DACH Region

Jessica Wewer 41
The Transformational Effects of Predictive Analytics on Planning

Vorwort

Liebe Leser,

zum zwölften Male konnten wir den TDWI Award für hervorragende Abschlussarbeiten auf der TDWI Konferenz in München verleihen. Unser Dank gilt insbesondere der INFOMOTION GmbH, die uns diese Preisverleihung ermöglicht. Wir erhielten 16 hochwertige Einreichungen. Die Themen fokussieren in weiten Teilen Analytics-Ansätze und spiegeln damit den Trend in der unternehmerischen Praxis wider.

Wir gratulieren den folgenden Gewinnern
1. Platz: Jan Pennekamp, Uncovering Doppelgängers in Online Communities
2. Platz: Christian Höcherl, Adaption und Evaluation von vortrainierten Konvolutionsnetzen im Anwendungsgebiet der Gewinnerkennung von Rubbellosen auf einer synthetischen Datenbasis
3. Platz: Sebastian Dierdorf, Fallstudienbasierte Analyse und Bewertung von Methoden des Maschinellen Lernens zur Bestimmung der wirtschaftlich optimalen Losgröße

Jan Pennekamp, RWTH Aachen, trainiert und verwendet ein Modell, um damit nach potenziell relevanten Kunden (Doppelgängers) zu suchen. Dazu identifiziert er die bedeutsamen Datenattribute für alle potenziellen Kunden, die für die Entwicklung des Modells verwendet wurden, und verwendet diese für die Modellausführung. Die Ausgabe ist eine Vorhersage darüber, ob die Personen im Datensatz die gleichen Merkmale wie die definierten Zielpersonen aufweisen. Mit der Vorhersage lassen sich die notwendigen Ressourcen bereitstellen, um die Beziehung zu diesen Kunden, verbunden mit passenden Pflegemaßnahmen, bestmöglich zu fördern. Hier wird Nutzen zur Erreichung der Unternehmensziele durch fortgeschrittene Analytik erzielt, wenn die Kombination von ungewöhnlichen Attributen der wichtigste Faktor bei der Identifizierung eines potenziell hochwertigen Kunden ist.

An der Universität Duisburg-Essen widmete sich Christian Höcherl einer Machbarkeitsstudie der Bildanalyse von Rubbellosen, um diese mit ihren sehr unterschiedlichen Zuständen (von perfekt freigelegt bis nahezu zerstört) korrekt identifizieren zu können, um Gewinne richtig zu interpretieren. Zu diesem Zweck verwendet er Konvolutionsnetze auf einer synthetischen Trainingsdatenmenge und testet die trainierten Netze mit realen Bildern. Dabei zeigte sich, dass die Netze grundsätzlich in der Lage sind, die Aufgabenstellung zu lösen. Die Anwendung auf einem Smartphone mit seinen speziellen Anforderungen stellte jedoch eine besondere Herausforderung dar; bei angemessener Zeitperformance führt dies zu Abstrichen bei der Klassifikationsgüte – ein spannendes und auch für die Praxis interessantes Ergebnis.

In seiner Arbeit untersucht Sebastian Dierdorf, TU Berlin, die Möglichkeit des Einsatzes von Methoden des Maschinellen Lernens zur Bestimmung der wirtschaftlich optimalen Losgröße innerhalb des Produktionsprozesses eines Automobillieferanten. Er fokussiert auf die Nachfragevorhersage, die zur Losgrößenermittlung verwendet wird, um insb. Muster in der kurzfristigen Veränderung der Produktionsaufträge zu identifizieren. Herr Dierdorf leitet Forschungsbedarf aus den Kriterien, Bedeutung der Aufgabe der Losgrößenberechnung, noch vorhandene Mängel in den in der Praxis genutzten Modellen, besondere Eignung von Methoden des Maschinellen Lernens für Prognosen,fehlende Forschung hierzu mit empirischen Datensätzen, ab.

Für die Jury war es fundamental, dass die Problemstellung als Basis für Lösungsansätze sauber herausgearbeitet wurde. Ein weiteres Kriterium des TDWI Awards ist der Bezug der Einreichungen zur Praxis, damit ein entsprechender Impuls zurückgegeben werden kann. In diesem Zusammenhang danke ich meiner Kollegin Prof. Dr. Barbara Dinter, TU Chemnitz, für ihr Engagement, so dass wir uns in der Jury kompetent zu den Arbeiten austauschen konnten.

Machen Sie sich selbst anhand der Abstracts ein Bild über aktuelle Trends und lassen sich durch die erarbeiteten Lösungsansätze inspirieren. Wir hoffen, dass Ihnen das Lesen der Beiträge genau so viel Spaß bereitet wie der Jury und freuen uns schon auf den TDWI Award 2019!

Ihr

Carsten Felden
Vorsitzender des TDWI e.V.

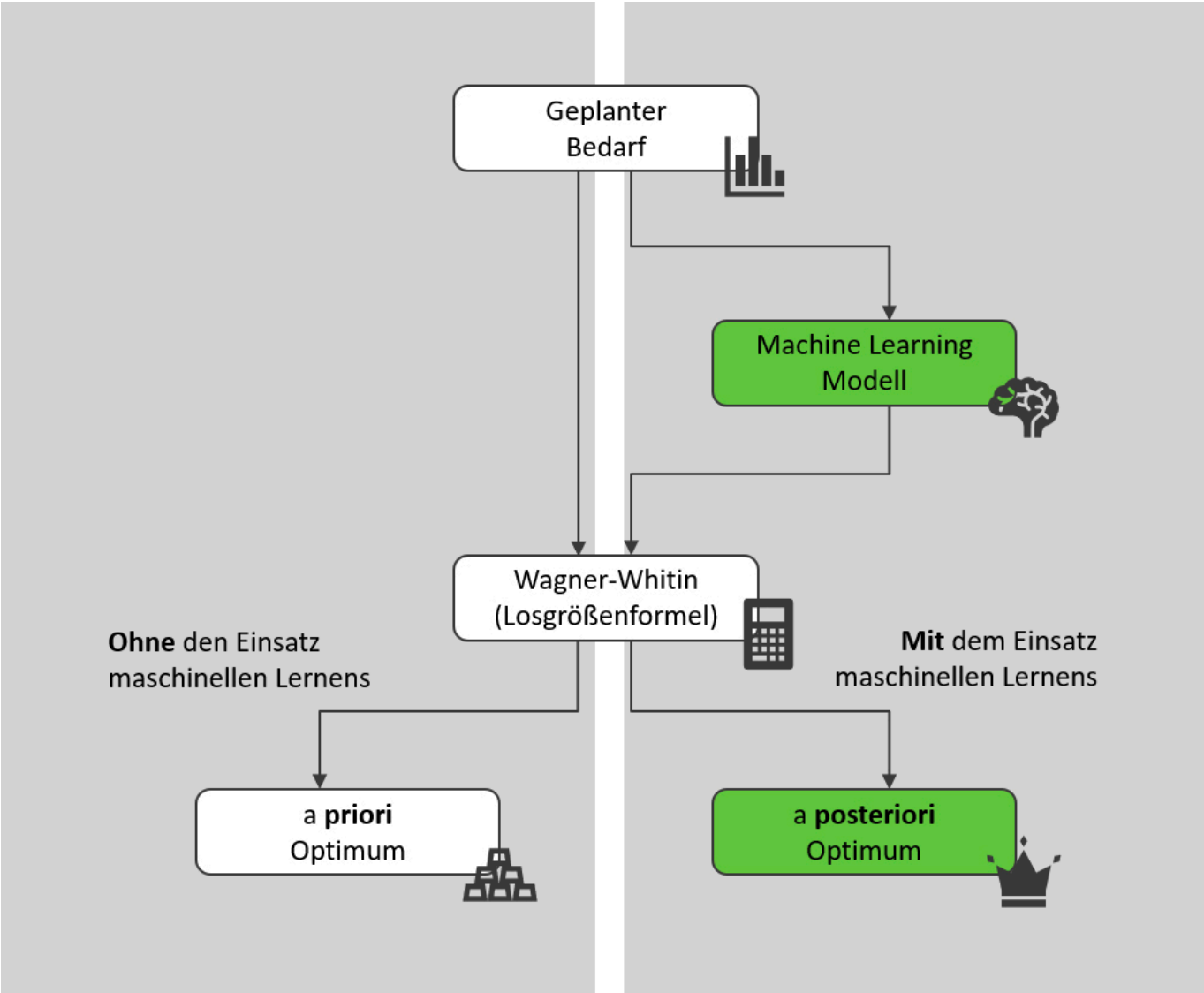
Fallstudienbasierte Analyse und Bewertung von Methoden des Maschinellen Lernens zur Bestimmung der wirtschaftlich optimalen Losgröße

Sebastian Dierdorf

DRITTER PLATZ

Zur Berechnung wirtschaftlich optimaler Bestellmengen und -zeitpunkte existieren verschiedenste Berechnungsverfahren, die basierend auf der Annahme eines deterministischen Bedarfsverlaufs optimale Lösungen versprechen. In der betrieblichen Praxis sorgen jedoch unterschiedliche Gründe dafür, dass kurzfristig von dem ursprünglich geplanten Produktionsprogramm abgewichen werden muss. Folglich handelt es sich bei den im Vorfeld ermittelten Lösungen häufig lediglich um a priori Optima, deren Ungenauigkeiten zu Unter- bzw. Überversorgung von Materialien führen und in der Konsequenz das Nichterfüllen von Kundenwünschen oder eine hohe Kapitalbindung in Form von Beständen zur Folge haben.

Der Untersuchungsgegenstand der Arbeit liegt in der Verifikation der Hypothese, die postuliert, dass der Einsatz eines selbstlernenden Vorhersagemodells Systematiken in den Abweichungen identifiziert und eine verbesserte Vorhersage des tatsächlichen Materialverbrauchs ermöglicht. Im Rahmen einer Fallstudie und unter Verwendung von realen Unternehmensdaten werden Vorhersagemodelle mit Hilfe der Methoden der künstlichen neuronalen Netze, Support Vector Machines und Random Forests generiert. Diese lernen aus den Abweichungsmustern, die über einen Zeitraum von drei Monaten zwischen den geplanten Bedarfsmengen und tatsächlichen Verbrauchsmengen sämtlicher Komponenten des Fallstudienunternehmens entstanden sind. Schließlich werden die Prognosen der Modelle in einer zweistufigen Evaluation miteinander verglichen. Dabei wird zum einen die durchschnittliche Treffgenauigkeit der Vorhersagen bewertet und zum anderen validiert, inwieweit es den klassischen Berechnungsverfahren durch verbesserte Prognosen möglich wird, kostenoptimale Losgrößen zu ermitteln, die nahe dem a posteriori Optimum liegen. Die Ergebnisse zeigen konkrete Steigerungen hinsichtlich der Prognosegüte und deuten in Bezug auf die Losgrößenoptimalität auf weiteres Verbesserungspotential hin.



Die Masterarbeit wurde eingereicht an der Technischen Universität Berlin bei Professor Dr.-Ing. Frank Straube und Tino Herden.

Cloudbasiertes ETL – Entwicklung & Validierung eines kriterienbasierten Ansatzes zur Bewertung geeigneter Cloud-ETL-Lösungen

Simon Frank

Die Masterarbeit wurde in einem Beratungshaus (MT AG) verfasst und setzt sich mit dem Einsatz cloudbasierter Produkte im ELT-Umfeld auseinander.

Der Einsatz von Cloud-Technologien verspricht eine erhöhte Flexibilität und Skalierbarkeit von Ressourcen sowie schnellere Entwicklungszeiten und eine Kostenreduktion der verwendeten Architektur. Diese Vorteile veranlassen sogar immer mehr Unternehmen dazu, eine Cloud-First-Strategie einzuführen. Doch trifft dies auch auf den Bereich der ETL-Verarbeitung zu?

Hauptziel der vorliegenden Arbeit war die Untersuchung der Tauglichkeit von ETL-Prozessen in Verbindung mit der Cloud Technologie. Konkret wurde hierbei überprüft, ob cloudbasiertes ETL mit der Cloud-Technologie in Einklang gebracht und die Vorteile der Cloud auch auf diesem Gebiet angewendet werden können. Zu diesem Zweck wurden diverse Use Cases konzipiert und umgesetzt, anhand derer die verschiedenen Cloud-Produkte analysiert wurden.

Dabei ergab sich, dass ETL in der Cloud es durchaus ermöglicht, die Vorteile der Cloud-Technologie, wie etwa das Kostenmodell, die Skalierbarkeit und die Flexibilität im ETL-Umfeld zu nutzen. Dies konnte durch die Umsetzung der Use Cases und die dadurch aufgezeigten Handlungsmöglichkeiten in diesen drei erwähnten Bereichen validiert werden. Durch den optimalen Einsatz der Cloud-Technologie ist es dem Unternehmen so möglich, zukunftsorientiert auf sich ändernde Gegebenheiten schnell und flexible zu reagieren.

Andererseits musste jedoch festgestellt werden, dass das Cloud-Angebot aktuell sehr unübersichtlich gestaltet ist und die verschiedenen Anbieter proprietäre Dienste ohne einheitliche und wohlgeformte Standards einsetzten. Dies schafft eine gewisse Abhängigkeit von dem Cloud-Anbieter. Zudem können durch das Fehlen einheitlicher Standards die Vorteile auch nicht bei jeder Kombination aus Szenario und Produkt vollkommen ausgeschöpft werden. Hierzu ist es also erforderlich, genau zu analysieren, ob in der aktuellen

Die Masterarbeit wurde eingereicht an der Hochschule Trier
bei Prof. Dr. Wilhelm Steinbuß.

Projektsituation die erwarteten Vorteile mit dem gewählten Produkt erreicht werden können.

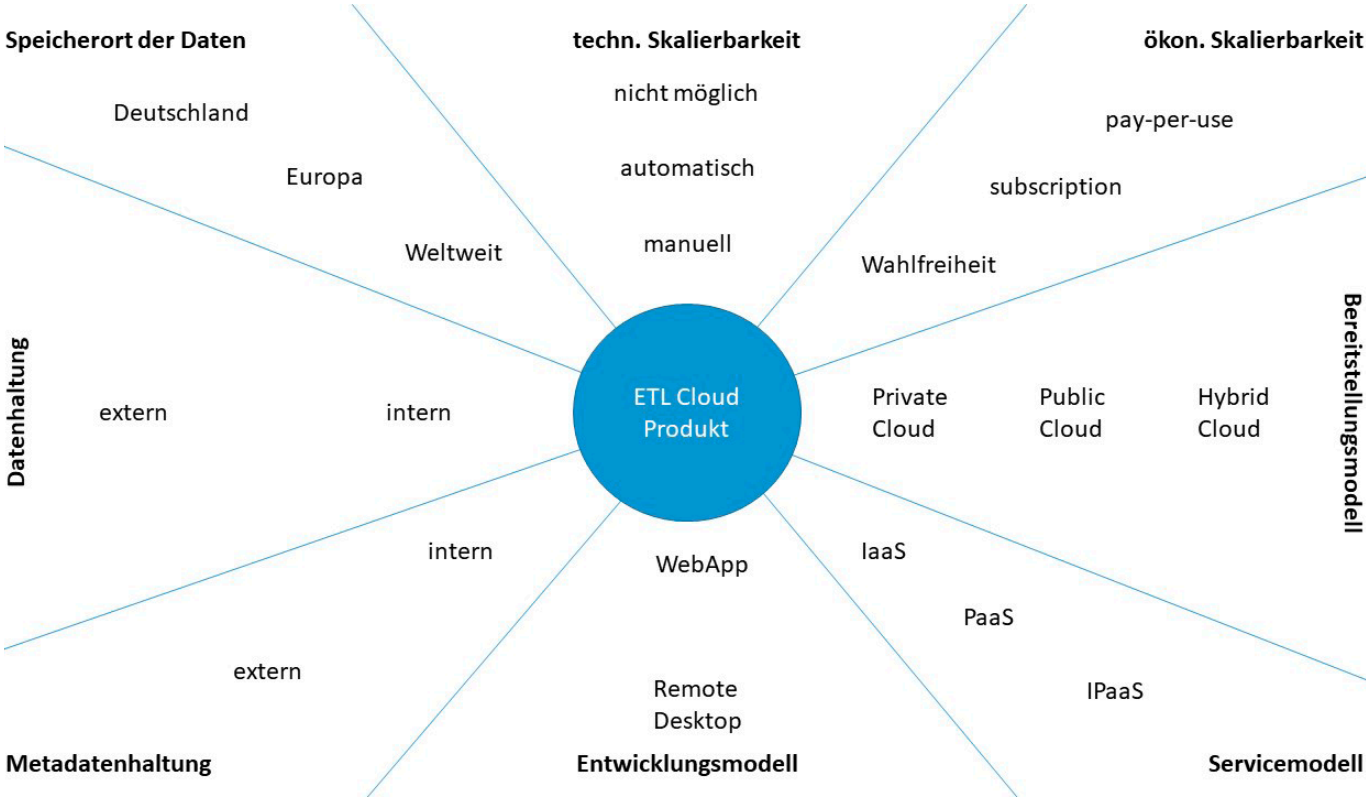
Um möglichst schnell das optimale Produkt zu finden war ein weiteres Ziel die Entwicklung eines kriterienbasierten Ansatzes zur Bewertung von Cloud-ETL-Lösungen. Konkret wurde dabei untersucht, wie die Auswahl eines geeigneten cloudbasierten ETL-Produkts erfolgen kann. Zu diesem Zweck wurde ein Modell zur Produktsegmentierung (kurz PSM) entwickelt, um eine schnelle Entscheidung für Unternehmen zu bieten, welche Produkte für die aktuelle Projektsituation am geeignetsten sind.

In diesem Modell wurden Kriterien ausgewählt, die die verschiedenen Produkte bestmöglich unterscheidbar machen sollen. Dabei wurden zu jedem Kriterium verschiedenen Ausprägungen herausgearbeitet. Je nachdem, welches Projekt durchgeführt werden soll, kann eine bestimmte Kombination der Ausprägungen sich als geeignetste herausstellen. Die nachfolgende Abbildung stellt das Modell mit den verschiedenen Kriterien und Ausprägungen grafisch dar.

Bei der Validierung des Ansatzes wurde festgestellt, dass durch den Einsatz des Modells die Unternehmen in einer ersten Phase effektiv mögliche cloudbasierte ETL-Produkte identifizieren und somit für nachgelagerte Phasen, wie einen möglichen PoC, mehr Zeit und Budget aufbringen können. Außerdem sind Aspekte, wie die Übertragungsdauer über das öffentliche Internet bereits enthalten.

Andererseits musste festgestellt werden, dass für eine genaue Produktwahl alleine das PSM nicht ausreicht und durch die subjektive Komponente der Bewertung schnell falsche Entscheidungen getroffen werden können. Hier gibt es jedoch Abhilfe, indem neben dem PSM in einem Vorgehensmodell auch ein Anforderungsmodell aufgestellt wird, das die Produkte weiter untersucht. Um die Subjektivität zu verringern ist es sinnvoll, eine unabhängige zweite Bewertung durchzuführen und die Ergebnisse gegenüber zu stellen.

Insgesamt wurde in der Arbeit der Schluss gezogen, dass ETL in der Cloud, basierend auf den aktuell zur Verfügung stehenden ETL Produkten, durchaus seine Daseinsberechtigung hat und für bestimmte Projekte eine vielversprechende Alternative zum traditionellen ETL-Ansatz darstellt. Ob cloudbasiertes ETL jedoch tatsächlich für die jeweilige Projektsituation einsetzbar ist, sollte im Einzelfall immer sorgfältig überprüft werden.



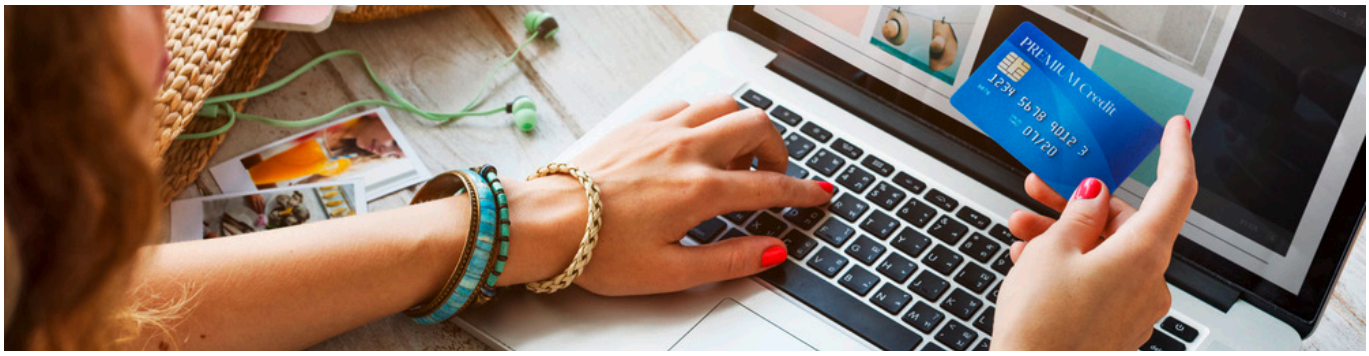
Uplift Modeling – An Introduction of Revenue Modeling to Maximize Campaign Profit

Robin Gubela

In contrast to typical B2C marketing campaigns where response models decide which customers to target, uplift campaigns measure the incremental impact of a treatment based on the prediction of customer-specific behavioral change. Uplift modeling is a novel, yet underestimated discipline that identifies different types of customers and only targets customers that are likely to respond positively (e.g., buy an advertised product) if being incentivized (e.g., receive a sale discount).

Although a growing number of companies and researchers pay attention to this field, the assumption is always maintained that additional profit stems from modeling the number of responders (i.e., conversion). I rebut this assumption by introducing the field of revenue modeling. Revenue uplift models optimize the incremental value of responders and are more effective than conversion models if customers spend heterogenous amounts of money. This is often the case in retail markets with a multitude of diverse products being offered.

I develop a novel approach which increases the effectivity of uplift models by discretizing a differential revenue response. Based on large-scale experiments in e-commerce with thousands of models on millions of observations, I demonstrate that related tree-based revenue uplift models outperform common-practice conversion uplift models and response benchmarks in terms of an expected campaign profit measure.



Die Masterarbeit wurde eingereicht an der Humboldt-Universität Berlin bei Prof. Dr. Stefan Lessmann.

Adaption und Evaluation von vortrainierten Konvolutions-netzen im Anwendungsgebiet der Gewinnerkennung von Rubbellosen auf einer synthetischen Datenbasis

Christian Höcherl

ZWEITER PLATZ

Weltweit stellen Rubbellose ein beliebtes Lotterieprodukt dar. In Deutschland bieten die 16 staatlichen Lotterieanbieter Rubbellose über ihr Verkaufsstellennetz als Teil ihres Lotterieproduktportfolios an. Bei den Rubbellosen handelt es sich um eine Sofortlotterie, d. h. der Kunde kann direkt nach dem Erwerb eines Loses den Gewinn prüfen. Dazu muss eine spezielle Beschichtung auf den Losen durch Rubbeln entfernt und die Zahlen bzw. Symbole nach der Spiellogik interpretiert werden. Diese Logik und die verwendete Symbolik unterscheiden sich dabei von Los zu Los.

In der Regel kann ein Geldgewinn oder ein Freilos gemäß des Gewinnplanes gewonnen werden. Das Los dient als Beleg, durch den sich der Gewinn einlösen lässt. Nieten können entsorgt werden. Es liegt somit in der Verantwortung der Kunden, einen Gewinn bzw. ein Freilos zu erkennen. Ca. 12 Prozent der Gewinnsumme von verkauften Losen werden in den Lotto-Annahmestellen nicht eingelöst. Es wird daher davon ausgegangen, dass eine Teilmenge der nicht abgeholten Gewinne aus einer Fehlinterpretation der Kunden resultiert.

Die Kunden legen die Lose auf unterschiedliche Arten frei. Manche Symbole werden nicht zur Gänze freigerubbelt oder die Oberfläche des Loses ist durch einen scharfen Gegenstand beschädigt worden und hat damit die Gewinnsymbole und –beträge teilweise unkenntlich gemacht. Die Motivation der Masterarbeit stellt die Entwicklung eines gegen diese Herausforderungen robusten Gewinndetektors für Rubbellose dar, der in Form einer mobilen Applikation die Kamera der Kunden-Smartphones nutzt und lokal eine Analyse auf Basis der Bildinformationen durchführt. Das Ergebnis der

Detektion soll dann den Kunden mittels eines Overlays dargestellt werden und ihnen helfen, Gewinne richtig zu interpretieren.

Konvolutionsnetze können auf bestimmte Aufgabenstellungen trainiert werden, sodass u. a. Objekte in zweidimensionalen Bildern detektiert werden können. Sie sind aufgrund ihrer Architektur dazu in der Lage, invariant gegenüber bestimmten Merkmalen zu reagieren und das Wissen auf bisher unbekannte Bilder anzuwenden.

In der Masterarbeit wurde evaluiert, ob sich Objektdetektoren auf Basis von Konvolutionsnetzen für das Anwendungsgebiet der Gewinndetektion auf Rubbellosen eignen, die Detektion in angemessener Zeit lokal auf Smartphones ausgeführt und eine Trainingsdatenmenge synthetisch hergestellt werden können. Schrittweise wurden Hyperparameter verändert und weitere Optimierungen der Trainingsdatenmenge gefunden, die das Detektionsergebnis unter realen Bedingungen verbesserten. Im Ergebnis wurde ein Prototyp entwickelt, der alle Objekte auf den Losen für eine beispielhafte Rubbellosserie lokalisieren und klassifizieren kann. Das Ergebnis der Arbeit ist demnach kein vollständiger Gewinndetektor, sondern stellt eine Vorevaluation in Form einer technischen Machbarkeitsstudie dar.

Unter Anwendung des Prinzips des überwachten übertragenden Lernens wurden vortrainierte Objektdetektoren der Meta-Frameworks ‚Single Shot MultiBox Detector' (SSD), ‚Faster Regions with Convolutional Neural Network Features' (Faster R-CNN) und ‚Region-based Fully Convolutional Network' (R-FCN) auf Basis der beiden Konvolutionsnetze ‚MobileNet' sowie ‚Inception-V2' für die Evaluierung herangezogen und die jeweiligen Klassifizierer auf die eigenen Klassen der Rubbellose angelernt.

Die Trainingsdatenmenge wurde dabei vollständig synthetisch generiert. Hierzu wurden aus Bildern echter Lose einzelne Klasseninstanzen isoliert und neue Bilder mit zufällig gewählten Gewinnsymbolen und -klassen erstellt. Mittels Rotationen, weiterer Verfahren der Datenvermehrungstechniken sowie einer zufälligen Teilwiederherstellung der Rubbelbeschichtung wurde einer Überanpassung entgegengewirkt.

Abbildung 1 zeigt ein für die Trainingsdatenmenge generiertes Beispielbild. Die Trainingsergebnisse wurden anhand einer Testdatenmenge mit realen Bildern evaluiert.

Abbildung 1: Synthetisch generiertes Rubbellos mit drei Gewinnsymbolen und einer Gewinnklasse

Die Aufbereitung der Daten und alle Hilfsskripte wurden mit der Programmiersprache Python, insbesondere mit den Paketen NumPy und den Modulen zur Bildverarbeitung aus der Softwareumgebung SciPy entwickelt. Als Paketmanager für die Erstellung und Trennung virtueller Entwicklungsumgebungen diente die Distribution Anaconda. Das Training der Modelle erfolgte mit dem Framework für maschinelles Lernen Tensorflow respektive Tensorflow Object Detection API auf zwei NVIDIA Geforce GTX 1080 Ti Grafikkarten. Der Android-Prototyp basiert auf der TensorFlow Beispielapplikation ‚TF Detect', die in C++ und Java implementiert ist. Die Ausführungsgeschwindigkeit der Prototypen wurde auf drei Android-Testgeräten unterschiedlicher Generationen gemessen. Es wurden zunächst zwei Referenzmodelle auf Basis von SSD und den Konvolutionsnetzen MobileNet bzw. Inception-V2 übertrainiert, welche dann in einzelnen Schritten optimiert werden konnten.

In den Optimierungsschritten hat sich die Auflösung der Konvolutionsnetze aufgrund der kleinen Symbole auf den Losen als kritischer Hyperparameter herausgestellt, der jedoch wegen des erhöhten Eingabedatenraums einen großen Einfluss auf die Ausführungsgeschwindigkeit der Detektion hat. Die trainierten Modelle ließen sich mit der PASCAL VOC-Metrik und der durchschnittlichen Detektionsgeschwindigkeit auf den Testgeräten mittels Prototypen messen und evaluieren.

Die Masterarbeit wurde eingereicht an der Universität Duisburg-Essen, Wirtschaftsinformatik, insbesondere Business Intelligence, bei Prof. Dr. Peter Chamoni.

Die Genauigkeit des Detektors SSD512-Inception-V2 (Auflösung 512x512 Pixel) in realen Erkennungsszenarien zeigte, dass die Trainingsdatenmenge in der Domäne der Rubbellose vollständig synthetisch hergestellt werden kann. Aus den Versuchen ging hervor, dass die Simulation einer synthetischen Rubbellosschicht die Generalisierbarkeit der Modelle erhöht und auch teilverdeckte Symbole detektiert werden können, solange noch eine hinreichende Anzahl von Merkmalen zur Verfügung steht. In der Abbildung 2 sind beispielhafte Detektionen mittels des Prototyps auf Basis von SSD512-Inception-V2 visualisiert. Im komplett aufgedeckten Zustand werden das Los, alle Gewinnsymbole und die Gewinnklasse korrekt detektiert. Das rechte Bild zeigt ein Detektionsergebnis bei teilverdeckten Gewinnsymbolen.

Abbildung 2: Detektionsergebnisse des Prototyps auf Grundlage von SSD512-Inception-V2

Zusammenfassend ließ sich feststellen, dass die in der Arbeit trainierten Objektdetektoren noch nicht für den produktiven Einsatz auf Smartphones im Anwendungsgebiet der Gewinnanalyse geeignet waren und demnach das gesetzte Formalziel von 40 % mean Average Precision (mAP) sowie einer annehmbaren Lauffähigkeit auf einem Smartphone der Mittelklasse nicht erfüllt wurde. Die Inception-V2-Variante bot mit 48,3 % mAP eine ausreichende Genauigkeit für die Gewinnanalyse von Rubbellosen. Jedoch verfügen derzeit nur High-End-Geräte über entsprechende Prozessoren, um eine akzeptable Geschwindigkeit der Überprüfung der Lose zu erreichen. Das schnelle Pendant SSD512-MobileNet war auch auf Smartphones der Mittelklasse ausführbar, jedoch war die Genauigkeit mit 32,5 % mAP für eine Gewinnanalyse von Rubbellosen im Produktiveinsatz noch nicht ausreichend.

Leitfaden zur Implementierung einer unternehmensweiten Business-Intelligence

Daniel Jung

Zunächst werden die beiden Begriffe DWH und BI entsprechend hergeleitet und voneinander abgegrenzt, da sie in der Literatur häufig synonym verwendet werden. Ziel einer BI ist es, aus internen und externen Daten entsprechende Informationen zu ermitteln bzw. Wissen zu generieren, damit Empfänger fundierte Entscheidungen treffen können. Ein vereinfacht dargestellter BI-Ordnungsrahmen dient als Orientierung zur Vorgehensweise beim Aufbau einer BI.

Um den Lesenden in die Lage zu versetzen, selbstständig eine unternehmensweite BI aufzubauen, wird in Kapitel drei eine Vorgehensweise zur Auswahl einer geeigneten BI-Software aufgezeigt. Für einen Einstieg in die vielseitigen Ausgestaltungsmöglichkeiten einer BI ist es notwendig, sich zunächst die aktuellen Einsatzzwecke und Anwendungsbereiche anhand der auf dem Markt befindlichen BI-Software aufzuzeigen. Eine BI-Software lässt sich prinzipiell in die beiden Bereiche Datenmanagement und Anwenderwerkzeuge unterteilen. Jeder Bereich erfüllt dabei unterschiedliche Einsatzzwecke. Im Hinblick auf die Anwendungsbereiche lässt sich festhalten, dass eine BI-Software für den Einsatz im kompletten Unternehmen eine sinnvolle Verwendung finden kann. Den Möglichkeiten von BI sind dabei so gut wie keine Grenzen gesetzt. Die Menge an BI-Software auf dem internationalen Markt ist heutzutage schon beachtlich und wächst stetig weiter. Inwieweit man BI unternehmensweit oder nur in einzelnen Bereichen einsetzen möchte, hängt dabei stark von den individuellen Anforderungen und Bedürfnissen ab. Der Markt für Software ist durch einen schnelllebigen Fortschritt gekennzeichnet. Deshalb sollte man den Anwender-Trends für BI und den allgemeinen Technologie-Trends eine gewisse Beachtung schenken. Denn so erfährt man, welche Themen aktuell breit genutzt werden und welche sich noch etablieren müssen. Der finale Schritt, eine passende Software für ein Unternehmen zu finden, wurde in acht allgemeingültigen Schritten beschrieben. Diese dienen als Orientierung für eine mögliche Vorgehensweise. Festzuhalten ist, dass der Einstufung der eigenen Anforderungen im Hinblick auf die möglichen Anwendungsbereiche eine hohe Beachtung geschenkt werden sollte.

Bevor eine BI implementiert werden kann, sollte man sich zudem mit allgemeinen Themen auseinandersetzen. Darunter ist zunächst das Verständnis der eigenen Datenquellen und der Datenarten zu verstehen. Mögliche Quellen können interner und externer Natur sein. Interne Daten stammen aus intern eingesetzter Business-Software. Externe Quellen sind sehr vielfältig und stellen je nach Anwendungsfall ein enormes Potential dar. Klassische BI befasst sich im Wesentlichen mit strukturierten Daten. Für die Analyse der anderen Datenklassen bedarf es (Big-Data)-Softwarelösungen, mit der sich die Arbeit nicht beschäftigt hat. Des Weiteren sollte man sich mit dem Thema der Historisierung von Stamm- und Bewegungsdaten auseinander setzen. Ebenfalls als wichtig anzusehen sind organisatorische

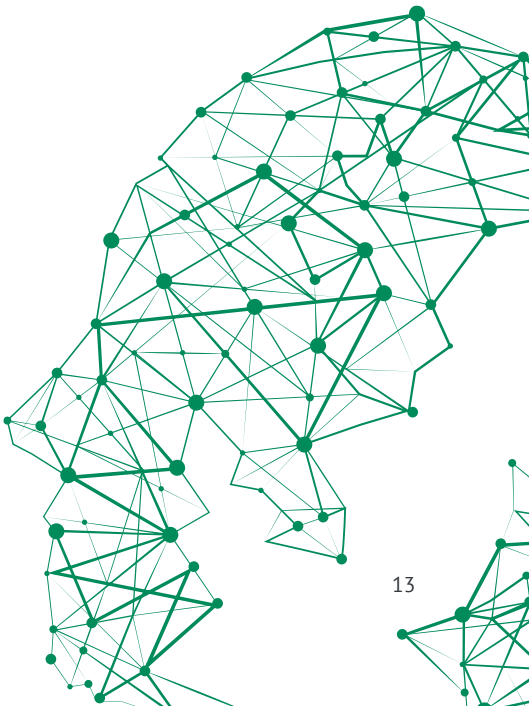
Themen wie die Notwendigkeit von Managementunterstützung und die Auseinandersetzung mit den Auswirkungen auf die Unternehmenskultur. Zudem sollte eine BI-Strategie als Grundlage für zukünftige Schritte als Orientierung dienen. Diese Themen sollten von einer neu zu schaffenden Organisationseinheit, der zentralen BICC, geplant und umgesetzt werden.

Ein weiterer Schritt zur unternehmensweiten BI gemäß dem aufgezeigten Ordnungsrahmen ist das Schaffen einer einheitlichen Datenbasis mit Hilfe der Datenintegration und -bereitstellung. Der zum Befüllen eines DWH notwendige ETL-Prozess dient der Integration der verschiedenen Datenquellen. Für die Speicherung der Daten gibt es mehrere Datenbanksysteme mit unterschiedlichen Funktionsweisen. Man sollte sich der Unterschiede bewusst sein, um sich schließlich für eines der Systeme zu entscheiden. Außerdem muss man sich für eine DWH-Architektur entscheiden, deren Schichten entsprechend der technischen Anforderungen der ausgewählten BI-Software und der eigenen Bedingungen modelliert sein müssen. Dafür sollte man die spezifischen Vor- und Nachteile der Modellierungsmethoden wissen.

Im nächsten Schritt gilt es, die Daten so zu verarbeiten, dass aus diesen sinnvolle Informationen generiert werden, die als Entscheidungsgrundlage Verwendung finden. Neben dem Erstellen von KPIs und Kennzahlen wird für viele BI-Anwendungen die OLAP-Technologie genutzt, um mit Hilfe ihrer Multidimensionalität Analysen zu erzeugen. Sollten Daten der Gegenwart analysiert werden müssen, kann ein Monitoring-System bzw. Realtime-Analytics genutzt werden. Ebenso kann Data-Mining durch bestimmte Verfahren Muster aus Daten aufdecken, die man sonst nicht gesehen hätte.

Der finale Schritt bei der Implementierung einer BI besteht darin, die verarbeiteten Daten und Information für einen Empfänger entsprechend zu visualisieren bzw. diesem einen Informationszugriff zu gewährleisten. Dieser Zugriff kann im Rahmen von SSBI selbstständig durch die Fachanwender erfolgen oder aber durch ein standardisiertes und automatisiertes Reporting geschehen. Der Einsatz von Dashboards ermöglicht einen übersichtlichen Zugriff auf Informationen mit Hilfe von Visualisierungen, welche für den jeweiligen Empfänger zugeschnitten sein sollten. Koppelt man diese Methoden mit Mobile-BI, so können alle Informationen zu jeder Zeit mobil mit sich getragen werden; und zwar dort, wo sie gerade für Entscheidungen benötigt werden.

Abschließend lässt sich sagen, dass das Implementieren einer BI – speziell einer unternehmensweiten BI und somit in allen Anwendungsbereichen – eine sinnvolle Unterstützung für jedes Unternehmen bieten kann. Durch Automatismen und Standardisierungen können erhebliche Entlastungen geschaffen werden, damit sich Unternehmen auf ihr Kerngeschäft konzentrieren können. Dafür bietet BI schon heutzutage eine enorme Vielfalt an sinnvollen Methoden und Techniken, um aus der stetig wachsenden Flut interner und externer Daten die richtigen Informationen gewinnen zu können. Setzt man BI kontrolliert dort ein, wo sie am ehesten benötigt wird und wo sie somit einen schnellen Erfolg erzielen kann, kann ein wesentlicher Vorteil gegenüber den Wettbewerbern generiert werden. Denn Daten sind „the new oil“ der Zukunft.



Die Masterarbeit wurde eingereicht an der Technischen Hochschule Mittelhessen bei Prof. Dr. Harald Ritz.

Designing Uncertainty-Aware Dashboards for Application Lifecycle Management

Merlin Knäble

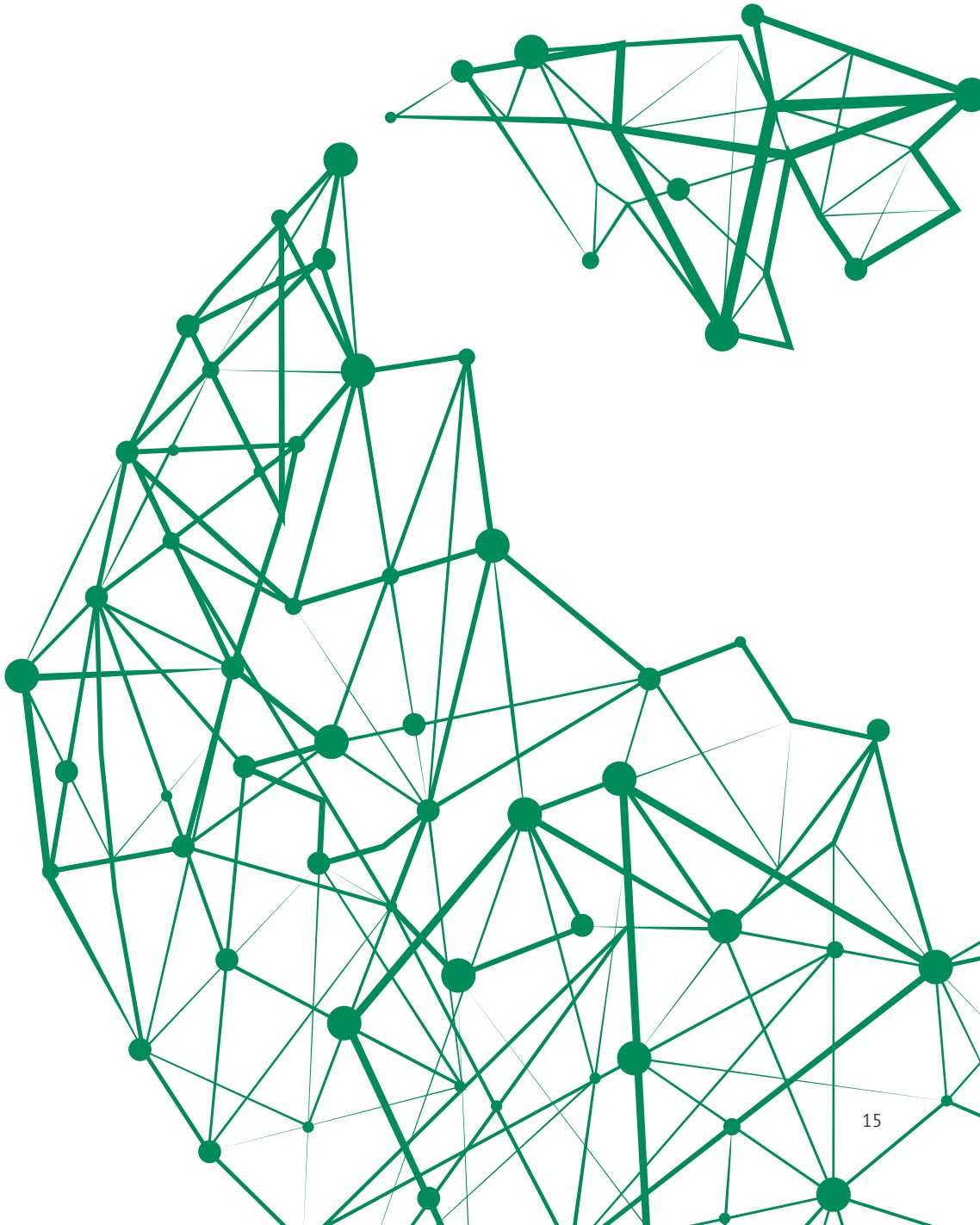
Despite significant advances in risk management methods, the IT industry is still plagued with poorly performing software development projects. Scholars acknowledge the role of data uncertainty in the planning and management efforts of software development projects as a potential area for further improvements. However, current information systems research lacks guidelines on how to measure and visualize data uncertainty within such projects. Thus, the first cycle of a design science project with an industry partner from the automotive sector is conducted, to suggest an uncertainty-aware dashboard based on supervised machine-learning. Dashboards have become a powerful visualization tool to support decision-making, by illustrating relevant information in a space-efficient and ergonomic manner. Centered around the main requirement found in a focus group analysis – uncertainty should not be measured in an abstract KPI but rather in context – we use random forests with classifiers („Are deadlines overrun?") and a regression („How much effort is needed?"). We train two classifiers and one regressor on different items in the data-set and evaluate them with a holdout set. The artifact is evaluated with multiple performance metrics that significantly surpass the estimations from our industry partner.

For the classification of whether a Task (simple implementation effort) or a Build (consisting of many Tasks, customer visible) will be completed before the deadline, we could reach accuracies over 80 %. The respective F1 scores are approximately 83 % and 65 %, the latter being influenced by the highly uneven class distribution of Builds. For the effort regression we could lower the mean squared error by almost 63 %, in comparison to the human estimation (planned effort). The R-squared of 56.02 % of our model may not seem highly sophisticated from an absolute standpoint, but when considering the human estimator its significance becomes clear. The planned efforts baseline R-squared is actually negative, meaning it is worse than using a constant estimator predicting the mean of the dataset. We transfer our predictions to a dashboard created in Microsoft PowerBI and use a calendar in combination with traffic light colors to indicate three levels of uncertainty. Deep drill-down functionality is provided.

From a practical view, the artifact meets the project managers' needs and may be instantiated in the industry partner's software landscape. In particular, the artifact affords managers with features to filter for development projects. To further support this overview, the Builds are also sorted along their planned completion time in an adjacent calendar, emphasized with color to direct an individuals' awareness. As multiple Builds exist for one day, a maximum function was implemented for the assessment. Hence, project managers may less likely overlook risky and uncertain Builds. Furthermore, the artifact's predictive capability offers objective, unbiased insights into data and more accurate forecasts based on the results of past work.

Die Masterarbeit wurde eingereicht am Karlsruhe Institute of Technology (KIT)
bei Professor Dr. Alexander Mädche.

This capability is of specific importance as it would enable the adaption of a more proactive strategy when dealing with projects by preventing potential issues before they manifest. Concluding, the findings may serve as foundation for future efforts addressing data uncertainty and aspects that should be considered when designing related dashboard solutions.



Methoden der Datenvisualisierung und ihre Anwendung in der Praxis am Beispiel des Service-Bund

Florian Kobylarz

Verschiedene Studien zeigen, dass die Visualisierung von Managementberichten zunehmend an Bedeutung gewinnt. Allerdings besteht häufig das Problem, dass Berichtsersteller nicht genau wissen, wie man Visualisierungen verwendet, um Informationen effektiv und zielgerichtet präsentieren zu können. Unter dem Begriff des Information-Designs gibt es einige Ansätze zur Nutzung effektiver Visualisierung. Der in Deutschland bekannte Information-Designer Prof. Rolf Hichert hat mit den International Business-Communication-Standards (IBCS), basierend auf den HICHERT®SUCCESS-Rules, Richtlinien zur Visualisierung veröffentlicht, um einen Standard in der Geschäftskommunikation zu etablieren. Daraufhin entstand beim Service-Bund GmbH & Co. KG der Wunsch, ein standardisiertes Visualisierungskonzept aufzubauen, das sich nach diesen Regeln richtet. Daraus wurde die Forschungsfrage der vorliegenden Arbeit abgeleitet: „Wie könnte ein standardisiertes Visualisierungskonzept von Management-Reports beim Service-Bund aussehen und angewendet werden?“ Dabei besteht das Ziel dieser Arbeit darin, ein Visualisierungskonzept nach den Regeln der IBCS aufzubauen und innerhalb des Service-Bund zu überprüfen, ob es angewendet werden könnte.

Zur Beantwortung dieser Frage wurden sowohl qualitative als auch quantitative empirische Daten erhoben. Mithilfe von offenen Expertenbefragungen sollten die unternehmerischen Anforderungen, technischen Limitationen sowie Vor- und Nachteile eines Visualisierungskonzeptes ermittelt werden. Des Weiteren wurde mit einer empirischen Befragung in Form einer Online-Umfrage ermittelt, ob innerhalb des Service-Bunds eine Akzeptanz vorliegt. Dabei werden folgende Punkte nicht in dieser Arbeit berücksichtigt:

- Vergleich verschiedener Visualisierungsansätze
- Vergleich verschiedener Business-Intelligence-Lösungen
- Einführungsstrategien

Nach Einführung in die technischen und betriebswirtschaftlichen Grundlagen wurde im dritten Teil dieser Arbeit versucht, ein Visualisierungskonzept zu erstellen. Auf Basis der IBCS-Regeln und Best-Practice-Ansätzen wurden semantische Notationen und Visualisierungsrichtlinien für den Service-Bund entwickelt und auf Beispielberichte angewendet.

Zur Überprüfung der Anwendbarkeit des erstellten Visualisierungskonzeptes wurde eine Online-Umfrage an die Berichtsersteller und -empfänger des Service-Bund durchgeführt. Von 97 potentiellen Teilnehmern haben sich 45 Personen beteiligt, wobei 32 Personen den Kriterien des Umfragezieles entsprachen (Rücklaufquote von 32,99 %). Den Probanden wurden drei Beispielrichtlinien und -berichte vorgeführt, woraufhin sie Kontrollfragen zu den Berichten beantworten mussten. Am Ende der Umfrage sollten die Teilnehmer einschätzen, wie gut sie mit den Beispielberichten arbeiten konnten und angeben, ob sie sich vorstellen könnten, mit solch einem Reporting-Design arbeiten zu können.

Die Umfrage hat ergeben, dass die Teilnehmer zu 53 % sehr gut bis gut mit den Beispielberichten arbeiten konnten. Zudem können sie sich zu 71,875 % vorstellen, mit dem Reporting-Design zu arbeiten. Bei genauerer Betrachtung der Ergebnisse wurde festgestellt, dass die derzeitigen Berichtsersteller und -empfänger, die das BI-Tool evidanza benutzten, dem Design nicht einhellig zugestimmt haben. Nur 57,17 % der evidanza-Berichtsersteller und 50 % der evidanza-Berichtsempfänger können sich vorstellen, mit solch einem Reporting-Design zu arbeiten.

Auf Grundlage dieser Ergebnisse wird dem Service-Bund empfohlen, nach weiteren Anpassungen und engen Abstimmungen mit den evidanza-Anwendern das Visualisierungskonzept intern zu veröffentlichen und kontinuierlich zu erweitern.

Die Bachelorarbeit wurde eingereicht
an der Hamburg School of Business Administration
bei Gutachterin Dipl.-Math. Kerstin Pahlke.

Social Media Analytics in Banking: Measurement of Social Media Activities in Austrian Banks

Romana Moser

Drawing upon literature, social media involves great potentials for financial services since it bears the possibility to increase the market share in consequence of being more competitively viable and establishing close relations with customers. Besides, financial services can capitalize on social media by improving their customer service and reputation. Due to lacking understanding about effectively measuring the performance of Facebook activities, a comprehensive measurement is required that serves as a basis for drawing meaningful insights from user generated data in order to take appropriate measures in consideration of Facebook objectives. Hence, this master thesis strives for exploring how Austrian banks measure the success of their Facebook activities/pages, based on the increasing importance of measuring social media activities. A qualitative research approach based on expert interviews as research method allowed for exploring how Austrian banks measure the success of their Facebook activities/pages. Qualitative expert interviews with social media experts specialized on Facebook were conducted for identifying relevant issues regarding measuring the performance of Austrian banks' Facebook engagements. For analyzing captured data, a qualitative content analysis approach was applied. The qualitative content analysis as analyzing method, which was basically a content-related structuring of expert answers into categories, allowed for revealing insights relevant for answering the research question.

Besides, this master thesis is based on theoretical concepts highlighting the importance of determining social media objectives when measuring social media efforts followed by a detailed delineation of social media performance measurement and insights indicators from both a financial and non-financial perspective. Besides, distinct data collection methods are outlined serving for gathering user-generated data on social media for subsequent measurement and analysis. Measuring and analyzing social media data, especially with regards to measurement of network structure, interaction, content, and sentiment indicators, is supported by diverse social media mining techniques which aim for analyzing users' social online contributions, determined in detail as a separate part of this thesis. In addition, theoretical concepts are complemented with an explanation how Facebook Insights can be applied for deriving insights about social data in consideration of social media measurement and analytics.

The results of this study allowed for the identification that performance measurement of Facebook activi-

ties/pages differs among Austrian banks. In this connection, frequency of performance measurement and reporting periods constitute main distinctions based on the findings. Whereas performance measurement takes place periodically on a daily, weekly, monthly or event-driven basis, reporting metrics ranges from a weekly basis to merely twice a year. However, when referring to benchmarking metrics provided by Facebook, it seems that Austrian banks are driven by competing companies and make efforts to keep pace with strong competitors based on their metrics. Moreover, the application of follower campaigns & call-to-actions are facilitating to boost performance; however it might be misleading when measuring the performance of their Facebook activities/pages since it could lead to distortions in concerned metrics. In addition, Austrian banks are facing distinct challenges regarding measuring the success of the Facebook activities/pages; lacking time is indicated as their greatest challenge. However, there is still room for improvement concerning current approach of performance measurement of Austrian banks' engagement on Facebook. Besides, Austrian banks are engaging on Facebook for distinct purposes in terms of making use of Facebook as communication and service channel as well as advertising channel, for presence and positioning purposes, and additionally for corporate communications. Moreover, Austrian banks make use of both non-financial and financial metrics for measuring the success of their Facebook activities/pages based on user-generated Facebook data; however, the focus of Austrian banks is primarily on non-financial metrics. Financial metrics in this context merely refer to the costs incurred for promoting activities on Facebook. Regarding analysis of measured metrics, the identification of sentiments plays an important role; likewise, periodical reporting of metrics and drawing conclusions about the metrics are relevant issues when measuring the success of Austrian banks' Facebook activities/pages. Furthermore, experts advise determination of objectives and corresponding key performance indicators as well as the creation of a dashboard for periodic measurement of user-generated Facebook data.

Based on the findings of this research, a performance measurement system was created. It should serve for measuring Facebook objectives and related activities of Austrian banks in order to quantify the success of their Facebook engagements. The performance measurement system constitutes a starting point for measuring the success of Facebook activities/pages at Austrian banks represented by a specific set of metrics which is based on metrics derived from findings complemented with metrics from literature. This ought to allow for a comprehensive measurement of the success of Austrian banks' Facebook activities/pages. Furthermore, this research contributes to knowledge since it both reveals practical insights and relevant issues regarding performance measurement of Facebook activities/pages at Austrian Banks that should be taken into consideration when measuring the success of their Facebook efforts in order to decide up on future actions accordingly. Hence, the performance measurement system represents a foundation for applying social media analytics at Austrian banks and might also be applicable for other financial institutions to measure the effectiveness of their Facebook activities/pages. In this context, investigating relations between Facebook metrics and business metrics which are not based on user-generated Facebook data constitutes a further direction for research.

Die Masterarbeit wurde eingereicht am Management Center Innsbruck
bei Professorin (FH) Dr. Claudia Brauer.

Uncovering Doppelgängers in Online Communities

Jan Pennekamp

ERSTER PLATZ

Auf vielen Diensten und Plattformen im Internet können (anonyme) Nutzer heutzutage ihre Meinung kundtun. Dies umfasst eine Vielzahl von Gemeinschaften, wie z.B. soziale Netzwerke, Blogs und Foren. Aber auch Nachrichtenportale bieten Bereiche in denen Nutzer diskutieren können.

Dieser Meinungsaustausch im Internet beeinflusst auch die öffentliche Meinungsbildung, obwohl sich die teilnehmenden Personen gar nicht kennen und sich der Wirkung auch nicht immer bewusst sind. Jegliche Manipulationen der Meinungsbildung stellen aber eine Gefahr selbst für existierende Demokratien dar. Von Interesse ist unter anderem die Frage, wer dafür Sorge zu tragen hat, dass eine solche Manipulation nicht stattfindet. Mögliche Strafen können für die beteiligten Unternehmen neben Imageproblemen auch wirtschaftliche Schäden hervorrufen. Im September 2017 wurde vom Bundestag ein entsprechendes Gesetz verabschiedet: das Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz – NetzDG). Dieses behandelt jedoch offensichtlichere Manipulationen und findet somit bei unserem untersuchten Ansatz eher weniger Anwendung. Die Meinungsmanipulationen die wir untersuchen sind eher versteckter und für den Laien nicht offensichtlich zu erkennen.

Wie von den Medien in den letzten Jahren immer häufiger berichtet wird, registrieren einige Organisationen eine große Zahl von Nutzern (sogenannte Internet-Trolle) um die öffentliche Meinung gezielt zu beeinflussen. Im Zusammenhang mit vielen politischen Themen (z.B. Ukraine-Krise, Brexit-Abstimmung, oder US-Präsidentenwahl) wurden entsprechende Meldungen veröffentlicht und selbst im Zusammenhang mit der Bundestagswahl gab es Berichte. In der Abschlussarbeit gehen wir gezielt diesem Problem nach. Während klassische Ansätze für die Erkennung von Doppelgängern (verschiedene Benutzeraccounts die von einer Person/Gruppe verwaltet werden) hauptsächlich auf Metadaten (wie z.B. IP-Adressen oder Registrierungsinformationen) zurückgreifen, schaffen wir es durch analytisch extrahierte Merkmale der veröffentlichten Texte zuverlässig alle Pseudonyme eines Autors miteinander zu verknüpfen. Solche Merkmale sind beispielsweise die Frequenz von bestimmten Wörtern oder Zeichen als auch eine bestimmte Art von Schreibfehlern. Gemeinsam betrachtet sind diese Aspekte eindeutig für jeden Autor, allerdings anders als bei z.B. Fingerabdrücken nicht unveränderbar.

Dieser Ansatz kann den Betreibern von Internetplattformen bei Angriffen auf die öffentliche Meinung helfen, auch wenn nutzerspezifische Informationen, wie z.B. IP-Adressen, genutzte Browser und Email-Adressen, bewusst verschleiert wurden.

Im Detail passen wir existierende Ansätze an um öffentliche Kommentare unter Nachrichtenartikeln miteinander zu vergleichen. Bisherige Ansätze wurden meist auf langen Texten angewendet, auf denen es deutlich einfacher ist die benötigten stylometrischen Informationen zu extrahieren, da eine noch viel größere Grundmenge an Text vorliegt.

Zum Anwenden unseres Ansatzes sammelten wir für zwei Monate Nutzerkommentare auf drei großen Nachrichtenplattformen (The Guardian, SPIEGEL ONLINE, ZEIT ONLINE).

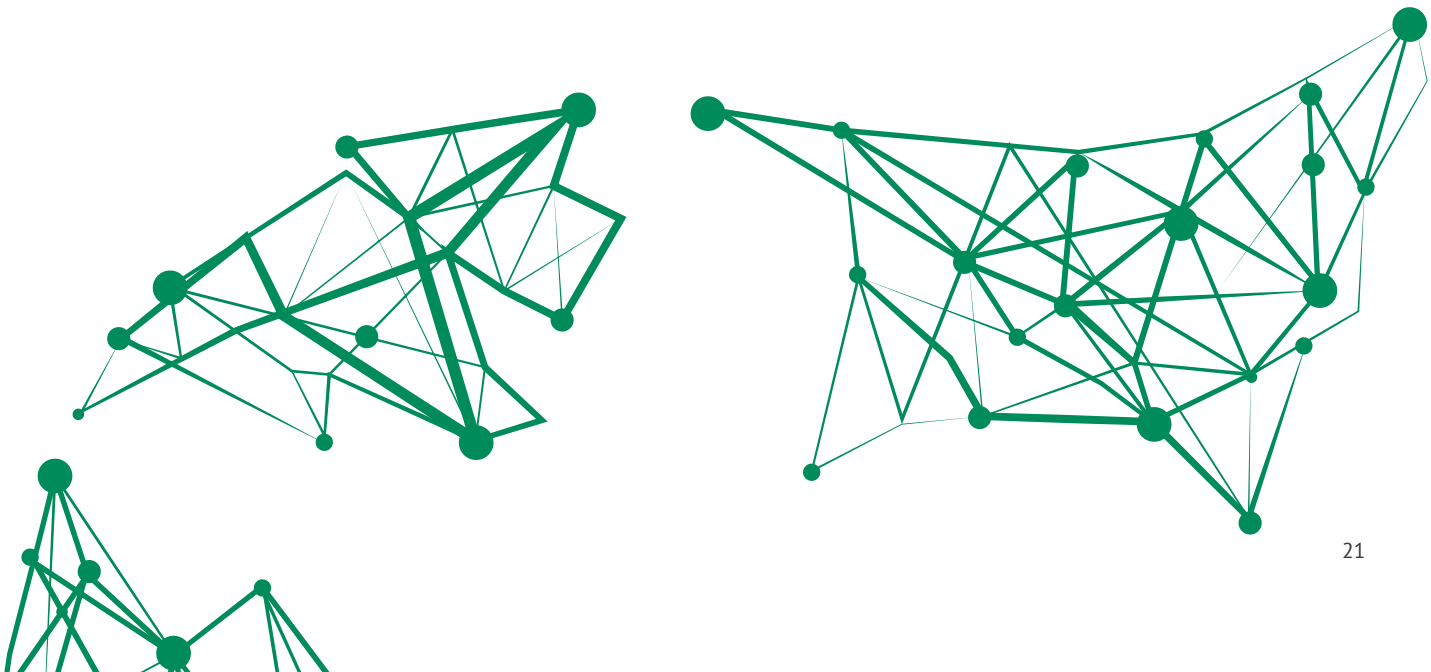
Die Masterarbeit wurde an der Rheinisch Technischen Hochschule bei Professor Dr. rer. Nat. Oliver Hohlfeld eingereicht.

Des Weiteren untersuchen wir, ob dieses Phänomen in bestimmten Bereichen, wie z.B. dem Politikbereich, häufiger auftritt und können schlussfolgern, dass sich Doppelgänger in allen Bereichen auffinden lassen.

Dies zeigt, dass Meinungsmanipulationen im Internet durch Doppelgänger eine reale Gefahr sind, welche nicht ohne weiteres erkannt werden kann.

Außerdem zeigen wir, dass derselbe Ansatz auf Daten (anonymen) sozialer Netzwerke (Twitter und Jodel) nur sehr bedingt anwendbar ist, weil diese Nachrichten normalerweise nur unzureichend stylometrische Details beinhalten. Für eine erfolgreiche Klassifizierung würden wir daher deutlich mehr Nachrichten benötigen um aussagekräftige stylometrische Profile erstellen zu können.

Zusammenfassend gibt unsere Arbeit Einblicke in den Kampf gegen Meinungsmanipulation im Internet, der sowohl für die gesellschaftliche Ordnung als auch für in dieser wirkenden Unternehmen von Relevanz ist. Wir liefern dabei einen Baustein um gegen Manipulationen angehen zu können, zumal traditionelle Ansätze zur Erkennung von Doppelaccounts bewusst von Nutzern in die Irre geführt werden können.



Next Level User Experience für Self-service BI: sprachgesteuerte Datenanalyse mit Alexa for Business

Simon Schmid

Die Masterthesis untersucht die Potentiale virtueller Sprachassistenten für den Einsatz im Bereich der Self-service Business Intelligence (BI). Neben der Darstellung konzeptioneller Grundlagen von BI, aktuellen Webtechnologien, sowie virtueller Sprachassistenten, steht die Interoperabilität von Sprachassistenten mit Produkten aus dem Bereich der BI im Zentrum der Untersuchung.

Im konzeptionellen Teil der Arbeit werden zunächst die Entwicklung, Relevanz und die unterschiedlichen Ausprägungen von BI aufgezeigt. Darauf aufbauend beleuchtet der Autor das Thema Self-service BI näher. Hierbei wird das traditionelle BI kritisch reflektiert und gegen die Vorteile von Self-service BI abgewogen. Aus den Eigenschaften von Self-service BI werden abschließend allgemeine Ziele abgeleitet und erläutert.

Die betrachteten Webtechnologien leiten sich aus der Verwertbarkeit für den Prototyp ab. Neben den Cloud Service Modellen werden die Kommunikationsprotokolle http und Websocket zur Kommunikation über das Internet aufgezeigt. Da sich http für den Austausch permanenter Datenflüsse nur bedingt eignet, wird das Websocket Protokoll als Alternative erläutert und genutzt. Es ermöglicht die bidirektionale Kommunikation, indem es eine permanente Verbindung zwischen Client und Server aufbaut und über diese kommuniziert. Der Kommunikationsaufwand ist hierbei geringer als mit http.

Zuletzt erfolgt eine Erläuterung von Sprachassistenten als relevante Webtechnologie. Diese werden beispielhaft anhand des Sprachassistenten Amazon Alexa beschrieben. Zusätzlich zu allgemeinen Informationen zur Funktionsweise des Endgeräts Amazon Echo beschreibt der Verfasser die technische Funktionsweise des Sprachassistenten anhand der konfigurierbaren Software-Komponenten (Interaktionsmodell und Lambda-Funktion).

Im folgenden Kapitel erfolgt eine Betrachtung allgemeiner und technischer Aspekte der Sprachsteuerung. Zunächst werden die unterschiedlichen Systemarten und der Einfluss unterschiedlicher Informationsquellen auf diese erläutert. Aus den Systemarten wurden abschließend allgemeine Komponenten einer Sprachsteuerung abgeleitet.

Kritisch betrachtet wird außerdem die Thematik des Voice User Interface Design. Durch die Besonderheiten der auditiven Schnittstelle sind weitere Anforderungen an das Sprachdialogsystem für eine hohe Akzeptanz der Benutzer zu beachten.

Die Masterarbeit wurde an der Hochschule Pforzheim bei Professor Dr. Frank Morelli eingereicht.

Der Kern der Thesis ist eine Machbarkeitsprüfung (Proof of Concept) zur Realisierung von drei Use Cases. Diese umfassen die Kennzahlenabfrage und Datenexploration, sowie die Interaktion mit Dashboards mittels Sprachsteuerung. Basierend auf den Use Cases werden 7 Anforderungen an den zu entwickelnden Prototyp identifiziert.

Die prototypische Umsetzung wird in der Thesis dokumentiert und mit dem von Amazon veröffentlichten Sprachassistenten Alexa, sowie einem mit SAP BusinessObjects Design Studio erstellten Dashboard implementiert. Die Ausführung der Kennzahlenabfrage erfolgt auf einer SQL-Datenbank. Der entwickelte Prototyp umfasst die Softwarekomponenten des Sprachassistenten, einer Komponente zur Kommunikation mit dem BO-Dashboard über das Internet, sowie einer Middleware zur Durchführung der Kennzahlenabfrage oder der Weiterleitung von Befehlen an das BO-Dashboard.

Die Middleware bildet die zentrale Einheit des Prototyps. Sie empfängt Befehle des Sprachassistenten per http und verarbeitet diese im Fall der Kennzahlenabfrage entweder zu einer entsprechenden SQL-Abfrage oder leitet im Falle der Berichtsteuerung die empfangenen Parameter per Websocket Protokoll an den adressierten Bericht weiter.

Die im Dashboard verbaute Komponente zum Empfang von Steuerungsparametern stellt mittels Websocket Protokoll eine Verbindung zur Middleware her. Aufgrund technischer Gegebenheiten bezüglich der Sicherheit der verwendeten Umgebung von SAP BusinessObjects ist ein initialer Verbindungsaufbau von Seiten der im Bericht verbauten Komponente mit der Middleware notwendig. Über die bestehende Verbindung via Websocket kommunizieren anschließend die Middleware und die verbaute Komponente. Eingehende Steuerungsparameter werden so in der Middleware an die Komponente im BO-Dashboard weitergeleitet. Im Programmcode des Dashboards erfolgt anschließend die Auswertung der Parameter und die Ausführung der entsprechenden Schritte.

Als Ausblick empfiehlt der Verfasser eine Identifizierung weiterer Use Cases im Umfeld der BI und deren Umsetzung mittels Anpassungen und Erweiterungen des bereits entwickelten Prototyps.

Anforderungen und Potenziale von Self-Service-BI im Finanzsektor

Daniel Schnappauf

Die Finanzbranche befindet sich seit der Finanzkrise 2008 im Umbruch. Nachdem die angeschlagene US-amerikanische Bank Lehman Brothers nicht mehr durch andere Bankinstitute gestützt wurde, musste sie Insolvenz anmelden. Dies führte zu großer Unsicherheit auf dem globalen Finanzmarkt und als Folge kamen die Geldströme zwischen den Banken fast vollständig zum Erliegen. Denn es trat schlussendlich eine Art Dominoeffekt auf, da nicht nur Lehman Brothers gewährte Kredite, sondern auch die an Gläubigerbanken von Lehman Brothers vergebene Kredite und schließlich die Kredite von deren Gläubigern abgeschrieben werden mussten. Ein Kollaps konnte nur durch staatliche Bürgschaften für angeschlagene Banken verhindert werden. Infolge der schweren Erschütterungen auf dem Finanzmarkt waren die Notenbanken der USA und der Europäischen Union gezwungen, ihre Leitzinsen auf ein historisch niedriges Niveau abzusenken. Die aus der Bankenkrise resultierende Eurokrise brachte zusätzliche Unsicherheit in die globalen Finanzsysteme. Bereits hochverschuldete Länder wie Griechenland und Portugal erhielten keine Kredite mehr und mussten sich unter den Schutz des Europäischen Rettungsschirms begeben. Diese Maßnahme bewahrte die Länder zwar vor dem Bankrott, verursachte jedoch innerhalb der Europäischen Union einen Einbruch der Wirtschaft. Von diesem Rückgang der Wirtschaftsleistung haben sich die nördlichen Länder Europas zwar weitgehend erholt. In den südlichen Mitgliedsstaaten ist die Lage jedoch noch immer angespannt, weshalb ein Ende der sogenannten Nullzinspolitik der Europäischen Zentralbank in naher Zukunft unwahrscheinlich ist.

Aufgrund dieser Lage wandelt sich das Geschäftsmodell der Banken in gravierender Weise, denn die Niedrigzinsphase lässt die Erträge der Kreditinstitute massiv schwinden. Zwar können Genossenschaftsbanken auf stille Reserven zurückgreifen, allerdings ist auch mittelfristig nicht mit einer Erholung der Zinssituation zu rechnen. Die Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin) empfiehlt daher, neben der Aufdeckung von Kostensenkungspotentialen auch den Ausbau des nicht zinstragenden Geschäfts.

Zum anderen müssen Banken auf die zunehmende Digitalisierung reagieren. Veränderte Kundenbedürfnisse, der Markteintritt flexibler und agil handelnder Unternehmen mit Geschäftsmodellen an der Schnittstelle zwischen Finanzwirtschaft und Technologie – sogenannte FinTechs – und nicht zuletzt gesetzliche Auflagen wie die der Payment Service Directive (PSD2[1]) verursachen tiefgreifende Umwälzungen im Markt und stellen die klassischen Bankinstitute vor immer neue Herausforderungen.

Getrieben durch diese Entwicklungen sind Banken gezwungen, sich neu zu konsolidieren, bisherige Geschäftsmodelle zu überdenken und neue zu erschließen. So entdecken Banken derzeit eine Ressource, über die sie in erheblichem Umfang verfügen – einen

1 PSD2 ist eine gesetzliche Richtlinie, die ab 2018 in Kraft tritt. Finanzinstitute sind dann verpflichtet Drittanbietern den Zugriff auf Daten ihrer Kunden, sofern diese eingewilligt haben, zu ermöglichen.

umfassenden Datenschatz für ihre Kunden. Diese Datenschätze sind historisch gewachsen, da Banken von jeher regulatorisch gezwungen waren umfangreiche Daten – beispielsweise zur Risikobewertung ihrer Kunden – zu erheben. Bisher wurden diese Daten größtenteils nur für diese Zwecke verwendet. Durch gesetzliche Regulatorik ist auch eine angemessene Datenqualität gewährleistet. So werden Banken regelmäßig von der BaFin überprüft und müssen in diesen Audits unter anderem auch eine ausreichend hohe Datenqualität unter Beweis stellen.

Um diese Daten jedoch auch jenseits der Erfüllung regulatorischer Auflagen nutzbar zu machen, müssen organisatorische und technische Kapazitäten aufgebaut werden, die Banken in der Regel bisher nicht hatten. Dabei stellt sich die Entstehung von Schatten-IT in Unternehmen als ein Problem von erheblicher Reichweite dar. Manager der unteren Führungsebenen warten nicht auf benötigte IT-Neuerungen, sondern erstellen eigene Prozesse, Investieren in Tools und stellen sogar Mitarbeiter mit spezifisch passender Qualifikation ein. Die in der Schatten-IT entstandenen Daten bleiben in einem Datensilo und sind so für das Unternehmen nicht verwertbar.

Die Daten aus solchen Silos, aber auch bereits zentralisiert vorgehaltene Daten in der Bank allgemein verfügbar zu machen ist jedoch eine schwierige Aufgabe. Der Bedarf an Daten kann sich aufgrund der Umweltdynamik des heutigen Finanzsektors schnell ändern, sodass eine zentral gesteuerte Informationsverteilung an Grenzen stößt. Vorzuziehen ist daher, den Fachabteilungen selbst die Mittel an die Hand zu geben, Berichte zu erzeugen. Dazu waren bislang spezialisierte Kenntnisse der Structured Query Language (SQL) notwendig, welche nur in begrenzten Umfang vorhanden waren. Self-Service Business Intelligence (SSBI) bietet dagegen die Möglichkeit, dass Fachabteilungen Berichte über eine Desktop oder Webanwendung generieren und abrufen können. Dies würde die Techniker entlasten und so den Weg für tiefergehende Analysen mittels Data Mining frei machen. Jedoch ist die Einführung von SSBI ein komplexes Vorhaben, das insbesondere mittelständische Finanzinstitute vor große Herausforderungen stellt. Die Potenziale und Grenzen einer Einführung von SSBI sollten daher insbesondere von mittelständischen Unternehmen der Finanzbranche sorgfältig abgewogen werden. Zur Untersuchung dieses Sachverhalts wurde ein exemplarisches Einführungskonzept von SSBI in einer Mittelständischen Genossenschaftsbank erstellt. Nach der Analyse der IST-Situation wurden die Schwachstellen identifiziert, die beseitigt werden müssen, um SSBI im Unternehmen einsetzen zu können. Um den IST-Zustand ermitteln zu können wurde auf das Business Intelligence Maturity Modell zurück gegriffen, um anhand der gewonnen Erkenntnisse konkrete Maßnahmen in den Bereichen Organisation, Fachlichkeit und Technik abzuleiten.

Vor dem Hintergrund der in dieser Arbeit gewonnenen Erkenntnisse ist zu schließen, dass SSBI – wenn es isoliert betrachtet wird – keinen signifikanten Vorteil für die Bank erzeugt. Es existiert bereits eine Fülle an Tools, welche zur Automatisierung des Berichtswesens genutzt werden könnten. Auch dürfte der Aufwand, die Power User in den Fachbereichen zu schulen, erhebliche Ressourcen erfordern. Dem Engpass bei der SQL-Erstellung könnte die Bank auch durch personelle Aufstockung von Mitarbeitern mit SQL-Kenntnissen entgegenwirken. Angesichts des zunehmenden Bedürfnisses, Informationen schnell nutzbar zu machen, böte dies jedoch nur ein zeitlicher Aufschub, da die Flut an Daten tendenziell eher zunehmen wird.

Dafür aber, die Daten, wie von Seiten der Bank angestrebt, für tiefergehende Analysen zu nutzen, stellt sich SSBI als wesentliche Voraussetzung dar, insbesondere im Umfeld kleiner und mittelständischer Banken. Die Einführung von SSBI-Tools und -Prozessen schafft den Experten die benötigten Freiräume, um prädiktive Analysen und Modelle erstellen zu können.

Wichtig ist jedoch eine ganzheitliche Betrachtung des Themas. Neben der Einführung benutzerfreundlicher Tools und der Schulung der Mitarbeiter ist es mindestens ebenso wichtig, den Mitarbeitern Plattformen zu bieten, auf denen sie sich zu Problemen und Erfahrungen im SSBI-Umfeld austauschen können. Durch eine agile Vorgehensweise bei der Erstellung der Dashboards kann zudem die Akzeptanz für das neue Dogma der autarken Datenanalyse gefördert werden. So ist gewährleistet, dass dem Anwender in kürzester Zeit neue Daten zur Auswertung zur Verfügung gestellt werden können.

Die Bachelorarbeit wurde eingereicht an der Hochschule für angewandte Wissenschaften Würzburg-Schweinfurt
bei Professor Dr. Frank-Michael Schleif.

Spezifikation endbenutzerfähiger Advanced-Analytics-Lösungen – Entwicklung eines Vorgehenskonzeptes auf der Grundlage einer Einzelfallstudie im Bereich Data & Analytics Mercedes-Benz Cars

Johannes Schwäble

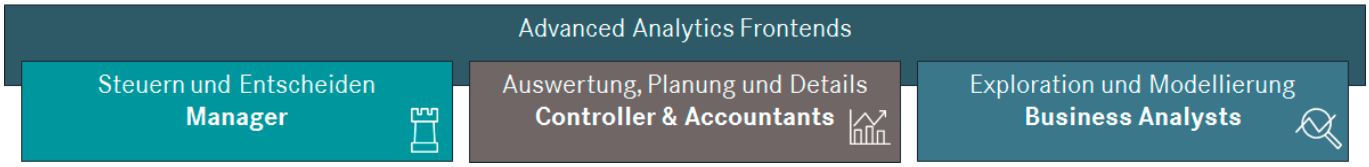
In der Unternehmenspraxis besteht momentan branchenübergreifend in Bezug auf Advanced Analytics-Lösungen Bedarf zur Definition geeigneter Vorgehensweisen, um innovative Anwendungen zu entwickeln und anschließend produktiv einzusetzen. Solche Applikationen ermöglichen datenbasierte Entscheidungsunterstützung und neue Geschäftsmodelle und beinhalten einen Trend von reiner Wertgenerierung zu umfangreichen Applikationen. Die Entwicklung solcher Advanced Analytics-Lösungen unterscheidet sich von BI-Systemen und erfordert die Definition neuer Rollen und Verantwortlichkeiten in einem agilen Projektumfeld. Advanced Analytics umfasst dabei neben Analysefunktionen im Frontend weitergehende Werkzeuge für anspruchsvollere Fragestellungen, z. B. in umfangreicheren, teils unstrukturierten Datenbeständen, und nutzt dazu komplexe statistische Methoden und des maschinellen Lernens und generiert daraus (teil-) automatisiert Erkenntnisse im Bereich von Descriptive, Predictive und Prescriptive Analytics, wobei zunehmend präskriptive Verfahren statt ausschließlicher Vorhersageerstellung zur Anwendung kommen. Dabei besitzen Advanced Analytics-Fragestellungen zunehmend nicht nur dispositiven, sondern auch operativen Charakter,

woraus im Zusammenspiel eine höhere Komplexität resultieren kann.

Unternehmen stehen oftmals vor der Frage, wie Advanced Analytics prozessual und organisatorisch umgesetzt werden kann. Zu klären sind die Zielsetzung, Rollen, Verantwortlichkeiten, Prozesse, Plattformen, Methoden und eine generelle Strategie.

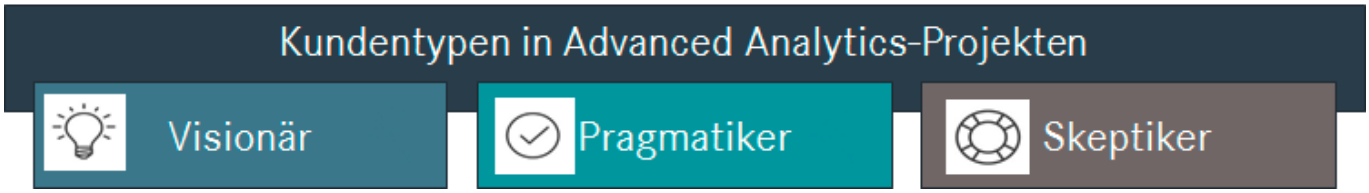
Vor diesem Hintergrund befasst sich die Arbeit mit der Entwicklung endbenutzergeeigneter Advanced-Analytics-Anwendungen von der ersten Idee bis zum Abschluss einer Proof-of-Concept-Phase mit Auslieferung einer ersten funktionsfähigen Lösung als Minimum-viable Product anhand einer Einzelfallstudie im Bereich Data & Analytics Mercedes-Benz Cars der Daimler AG und diskutiert, mit welchen Methoden, Beteiligten, Werkzeugen, Infrastrukturen und Analysen die deren prozessuale Entwicklung erfolgt. Dazu werden unterschiedliche Untersuchungsmethoden verwendet, u.a. Interviews mit Anwendern aus Controlling- und Accountingbereichen, Entwicklern und Projektbeteiligen, Systemanalysen sowie Workshops.

Mögliche Frontends sind in Desktop-Anwendungen, mobile Apps, Dateiexporte und interaktive Teamlösungen unterscheidbar, wobei Funktionalitäten und Freiheitsgrade je nach Zweck und Nutzergruppe variieren. Auf Basis der durchgeführten Interviews die drei Ausprägungen von Advanced Analytics-Frontends für jeweilige Nutzergruppen: Führungskräfte, Controller und Business Analysts, welche im Rahmen einzelner Analyseprojekte für die Kunden bzw. Zielbereiche der Anwendungen erstellt werden. Oftmals kommen die beschriebenen Frontend-Ausprägungen nebeneinander in einem Use Case vor, da sowohl Manager als auch Controller damit arbeiten.



Advanced Analytics Frontends für unterschiedliche Nutzergruppen

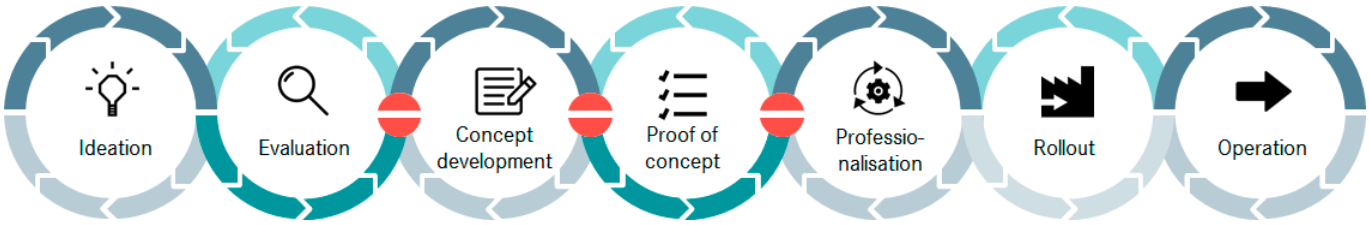
Die Kunden bzw. Projektpartner aus den Fachbereichen sind von den Anwendertypen, d.h. Nutzern, zu unterscheiden, da oftmals Anwendungen für mehrere Nutzergruppen entstehen. Auf Basis der Interviews sind drei Advanced Analytics-Kundentypen in abgrenzbar anhand ihrer Einstellungen und Erwartungen.



Analytics-Kundentypen mit unterschiedlichen Erwartungen und Einstellungen

Diese Unterscheidung nach Anwender- und Kundentypen liefert Mehrwert zum Benutzerverständnis und beeinflusst die Konzeption von Workshops und dienen zur Erstellung von Personas als Entwicklungsgrundlage. Diese Charaktereigenschaften sind während der einzelnen Phasen zu beachten.

Die Nutzung eines agilen, prozessualen Vorgehens gibt im Rahmen eines Projekts eine spezifische Abfolge und Struktur sowie auf Basis von Erfahrung und Wissen geeignete Techniken vor. Dies erhöht die Beherrschbarkeit der Komplexität, führt zu Orientierung und minimiert das Fehlerrisiko. Ein solcher Business Analytics-Prozess eignet sich zur frühen Ermittlung der Kundenanforderungen und deren fortlaufenden Konkretisierung und besteht aus sieben Phasen, beginnend mit der Evaluation, der Evaluation, der Konzeption, der Erstellung eines MVP in der PoC-Phase, der Professionalisierung, dem Rollout und schließlich der Betriebsphase. Diese Arbeit spezifiziert die ersten vier Phasen von der Idee bis zur MVP-Auslieferung und dessen Erprobung. Zu jeder Phase des Prozesses sind einzelne Aufgaben mit verantwortlichen Rollen, erstellte Resultate, geeignete Methoden und mögliche Ergebnisse und Evaluationsmethoden aufgeführt.



generischer Prozessablauf für Advanced Analytics-Projekte

Ein Management-Gremium entscheidet an rot hervorgehobenen Stage Gates über die Fortführung oder den Abbruch von Projekten. Diese Bewertung der Wirtschaftlichkeit ist wichtig, damit verfügbaren Ressourcen für die wichtigeren Use Cases bereitstehen.

Viele Problemstellungen in Analytics-Projekten sind nur in interdisziplinären Teams lösbar, wobei die neue Business Analytics Consultant-Rolle die Projektbearbeitung von der Idee bis zur finalen Lösung erleichtert und die Kunden durch den Prozess begleitet. Die Umsetzung erfolgt iterativ im Rahmen eines Proof of Concepts (PoC). Ziel eines PoC, d.h. eines Pilotprojekts, ist die Evaluation, ob eine Big-Data-Lösung in einem identifizierten Anwendungsfall Mehrwert stiften kann. Schrittweise entsteht dabei aus einem MVP bei ausreichendem Mehrwert und Machbarkeit ein vollumfängliches System. Zur Umsetzung dieser agilen Projekte eignet sich kontinuierliches Anforderungsmanagement mit Hilfe von Scrum und User Stories.

Die Nutzung von Design Thinking und Kreativitätstechniken zur Konkretisierung hat sich dazu bewährt vor allem für neue Themen mit Formalisierungsbedarf, um Probleme frühzeitig zu identifizieren und möglichst viele Ideen zu generieren und evaluieren und eine geeignete Lösung daraus zu entwickeln. Die entstehenden Applikationen sollten benutzergerecht gestaltet sein durch Beachtung von User Experience-Methoden und -Grundsätzen.

Darauf aufbauend zeigen die Ergebnisse, dass Prototyping sich als Vorgehen zur Realisierung von Analyselösungen eignet. Als Hilfsmittel zur Entwicklung endbenutzerorientierter Advanced-Analytics Werkzeuge eignen sich je nach Phase verschiedene Formen von Techniken, u. a. Paper Prototyping, Skizzen, Mock-Ups mit unterschiedlicher Detailierung oder funktionsfähige Teillösungen.

Zur Anforderungserhebung, Schaffung von Verständnis mit Konzeption, Prototyping und Evaluationen dienen je nach Phase verschiedene Techniken. Daraus sind Aktivitäten zu wählen, welche für die genaue Problemstellung Mehrwert liefern.

Dazu sind je nach Entwicklungsstand diverse Prototyping-Methoden in verschiedenen Phasen anwendbar zur benutzerzentrierten, interdisziplinären Ideenfindung sowie Konzepterstellung und verkürzen die Entwicklungszeit.

Die Masterarbeit wurde an der Universität Stuttgart bei Professor Dr. Hans-Georg Kemper eingereicht.

Erst die Erprobung eines MVP im Praxisumfeld macht manche Herausforderungen sichtbar und führt zu realistischeren Einschätzungen zur Umsetzbarkeit.

Ziel der Evaluation am Ende jeder dieser Entwicklungsiterationen ist sowohl die Erprobung einzelner Komponenten als auch des Gesamtsystems, um möglichst früh Rückmeldung zu erhalten aus den Tests kleiner Entwicklungsschritte, die Nutzer besser zu verstehen und eine Bewertung des betriebswirtschaftlichen Nutzens und der technischer Umsetzung anhand definierter Kriterien vorzunehmen.

Wenn die erstellten MVP-Lösungen in den Betrieb überführt werden sollen, müssen dafür Sicherheitsaspekte und rechtliche Fragestellungen zum Betrieb geregelt sein, damit klar ist, welche Daten wofür verwendet und gespeichert werden dürfen.

Durch den Trend zu interaktiven, individuellen Systemen wird neues Expertenwissen wichtig, z. B. zur Spracherkennung oder Frontendprogrammierung. Somit werden sich die notwendigen Entwicklerrollen weiter verändern.

Momentan liefern die Advanced Analytics-Lösungen oft nur Vorschlagswerte, welche die Expertise der Controller ergänzen. Zukünftig sind komplett automatisierte Prozesse technisch denkbar. Die Integration von Artificial Intelligence in Form von Augmented Intelligence in Analytics-Anwendungen, wird die Endanwender zukünftig besser unterstützen, und dabei die Stärken von maschinellem Lernen und menschlicher Erfahrung kombinieren. Dazu ist ein höheres Vertrauen in Daten und Vorhersagen notwendig. Ein Analytics-Prozess kann jedoch nur Grundlagen liefern, um langfristig durch kulturellen Wandel der Organisation neue Arbeitsweisen zu etablieren. Unternehmensweite Analyseplattformen sowohl für Data Scientists als auch für Endanwender können dabei zukünftig das Teilen von Analysemodellen, Daten und generierten Ansichten und Erkenntnissen stark vereinfachen. Als Folge daraus wird die Digitalisierung den kulturellen Wandel im Unternehmen intensivieren.

Information Modeling mit FCO-IM Datenmodellierung unter Verwendung natürlicher Sprache zur Digitalisierung eines Prozesses im Musterbau

Stephan Volkmann

Ziel der Masterarbeit ist der Aufbau eines Datenmodells zur Digitalisierung des Prozesses im beschriebenen Geschäftsbereich. Dazu wird mit der Methode Fully Communication-Oriented Information Modeling (FCO-IM) [ZEH15] ein Informationsmodell erstellt. Das Informationsmodell soll alle im Geschäftsbereich relevanten Fakten und Geschäftsregeln abbilden. Es wird die Grundlage für ein zentrales Informationssystem zur Verwaltung der Auftrags- und Messdaten bilden. Durch Transformation des Informationsmodells in ein physisches Datenmodell kann dieses in einer Datenbank implementiert werden.

Darüber hinaus evaluiert der Autor FCO-IM als Methode zur Erstellung eines Informationsmodells und stellt es dem Entity Relationship Modeling (ERM) [Che76] gegenüber.

Für den Leser der Masterarbeit sollen basierend auf den praktischen Erfahrungen in der Umsetzung mit der Methode FCO-IM mögliche Probleme der Methodik vorgestellt und darauf bezogene Lösungsempfehlungen präsentiert werden.

Informationsmodell

Ein Informationsmodell ist ein konzeptionelles Modell und dient dazu, Verständnis über eine Domäne aufzubauen oder zu vertiefen. Durch Diskussionen über die enthaltenen Informationen können Fehler in Prozessen oder neue Geschäftsmodelle entdeckt werden. Ein physisches Datenmodell hingegen ist in seiner technischen Tabellenform mit Datentypen und Schlüsselbeziehungen unverständlich für Mitarbeiter ohne Fachwissen in diesem Bereich. Fehler in Informationssystemen können häufig auf eine direkte physische Implementierung und unzureichende Untersuchungen der Anforderungen zurückgeführt werden. Daher sollte zunächst ein konzeptionelles Modell angefertigt werden, aus dem anschließend ein physisches Modell entwickelt wird.

FCO-IM gehört zur Familie der Fact-Oriented Modeling (FOM) Methoden. Diese nutzen als Grundlage der Modellierung Verbalisierungen in natürlicher Sprache, geben ein eindeutiges Modellierungsverfahren vor und fokussieren die Modellierung auf elementare Fakten [ZEH15].

Analyse

Um eine Grundlage für die Informationsmodellierung zu schaffen, wird zunächst der Geschäftsbereich des Musterbaus analysiert. Der Autor stellt die Gesamtstruktur des Unternehmens sowie die abzubildenden Geschäftsprozesse vor. Gemeinsam mit dem Fachbereich werden Verbesserungspotentiale erfasst. Weiterhin wird mit Hilfe von Prozessdiagrammen definiert, welche Prozessschritte zukünftig vom neuen Informationssystem unterstützt werden können.

Erstellung

In FCO-IM wird das Verfahren zur Modellierung FCO-IM Procedure [ZEH15] genannt. Die einzelnen Schritte dieses Verfahrens sind in Abbildung 1 dargestellt. In der Masterarbeit beschreibt der Autor ausführlich die schrittweise Erstellung des Informationsmodells für den Musterbau anhand von ausgewählten Teilprozessen. Im Anschluss an die Vorstellung der Ergebnisse werden nach jedem Schritt aufgetretene Besonderheiten der jeweiligen Phase vorgestellt und Lösungen zu Problemen aufgezeigt.

Alle Schritte des FCO-IM Procedures erfordern eine Kommunikation zwischen Modellierer und Domänenexperte. Die Modellierung in FCO-IM erfordert daher eine Projektorganisation, die einen schnellen und mehrmaligen Kontakt zwischen Modellierer und Domänenexperten ermöglicht.

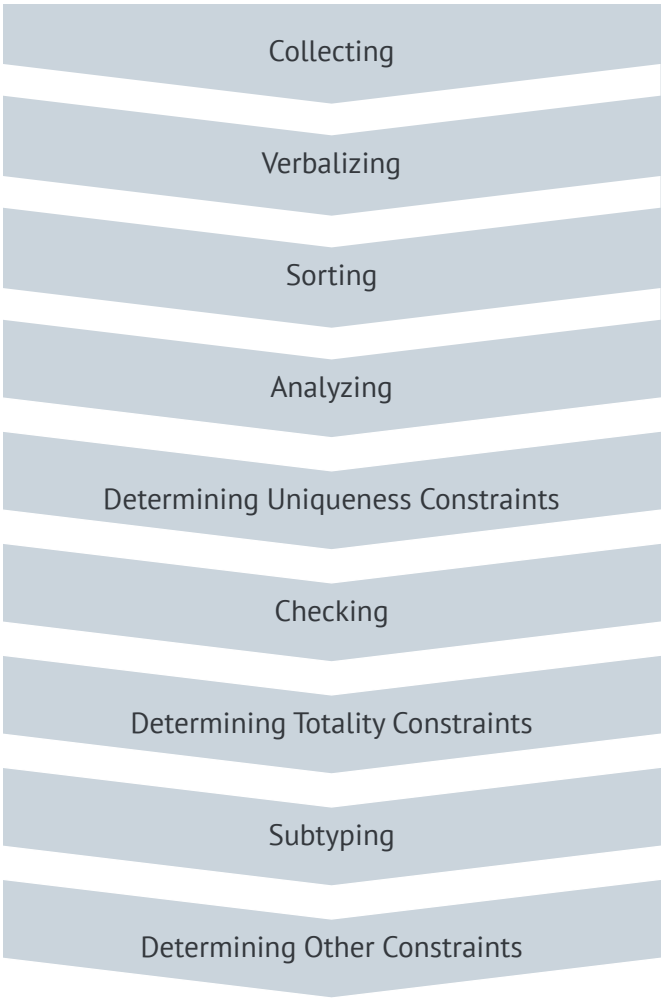


Abbildung 1: FCO-IM Procedure

Die Masterarbeit wurde an der Universität Paderborn
bei Juniorprofessor Dr. Artus Krohn-Grimberghe und Professorin Dr. Leena Suhl eingereicht.

Evaluation

Zur Evaluation von konzeptionellen Modellen wird in dieser Arbeit das Framework von Moody und Shanks [MS94] verwendet und auf die Bedürfnisse des Unternehmens angepasst. In dem Framework werden Qualitätsmetriken dazu genutzt, definierte Qualitätsfaktoren alternativer Modelle in konsistenten und objektiven Verfahren zu vergleichen und somit die Modellqualität zu verbessern.

Mit Hilfe des Frameworks evaluiert der Autor mehrere Informationsmodelle der Modellierungsmethoden FCO-IM und ERM für dieselben Geschäftsprozesse in einer Fallstudie und stellt somit die Qualität der verwendeten Methoden gegenüber.

Die Evaluation zeigt, dass ein Informationsmodell in FCO-IM eine höhere Qualität in den Faktoren Vollständigkeit, Integrität, Verständlichkeit, Korrektheit und Einfachheit besitzt. Im Faktor der Flexibilität zeigen sich noch Schwächen der Methode gegenüber dem ERM. Jedoch wirkt sich die hohe Qualität vom FCO-IM-Modell im Faktor Einfachheit positiv auf die Flexibilität aus, da weniger Elemente bei Änderungen der Anforderungen angepasst werden müssen. Die Modellierungsmethode FCO-IM kann daher als eine bessere Alternative zum ERM empfohlen werden, da durch die Verwendung von Beispielen in natürlicher Sprache und ein vorgegebenes Verfahren zur Modellierung eine höhere Modellqualität erzielt werden kann.

Transformation

Damit das neue Informationssystem das konzeptionelle FCO-IM-Modell als Datengrundlage nutzen kann, muss es in ein physisches Modell transformiert werden. Durch die Transformation ist auch dieses in der Fachsprache der Domänenexperten gehalten. Dadurch wird die Implementierung des neuen Informationssystems unterstützt, da dessen Entwickler die Zusammenhänge der Daten und abzufragende Tabellen und Felder leicht erkennen können und bei Rückfragen direkt mit dem Fachbereich kommunizieren können.

Zur Transformation des FCO-IM-Modells in ein relationales Schema wird der in [BZVdL00] vorgestellte GLR-Algorithmus verwendet. Dieser kann bei einer vollständigen Überprüfung der Modellelemente auf die Einhaltung der FCO-IM-Regeln eine Transformation in ein redundanzfreies relationales Schema garantieren [ZEH15]. Die Erstellung des Informationsmodells, die Überprüfung sowie die Transformation und anschließende Erstellung von SQL-Skripten werden in CaseTalk1 unterstützt. Durch die Ausführung der generierten SQL-Skripte hat der Autor abschließend das Datenmodell auf der Datenbank implementiert.

Durch die Verwendung der Methode FCO-IM hat der Autor für den Geschäftsbereich ein Informationsmodell in hoher Qualität erstellt. Durch anschließende Transformation in das physische Datenmodell und dessen Implementierung ist die Grundlage für das neue Informationssystem gelegt.

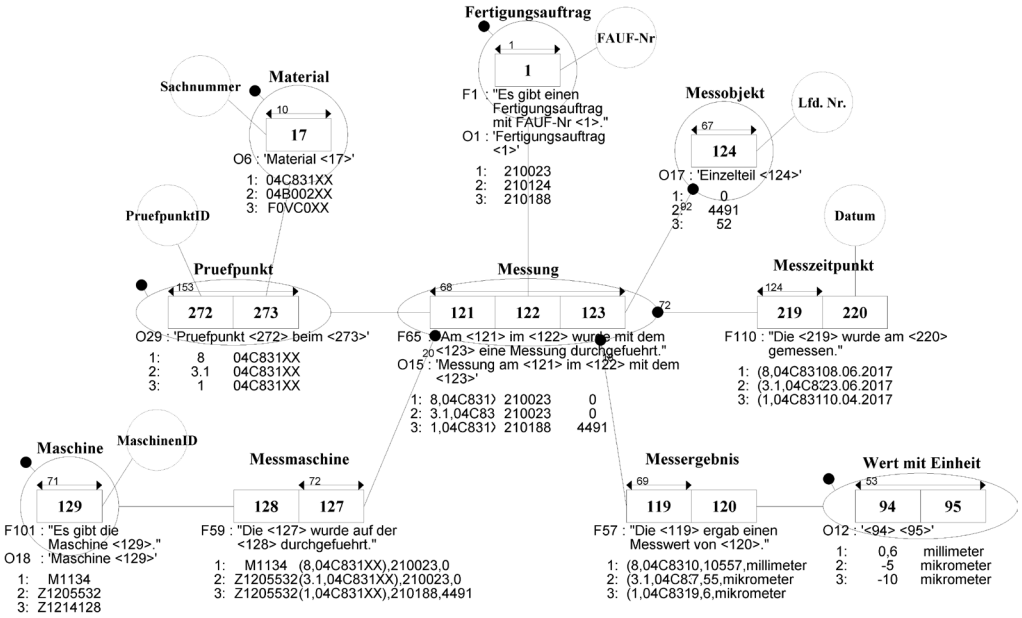


Abbildung 2: Ausschnitt aus dem Informationsmodell

1 CaseTalk Modeling Software https://www.casetalk.com/

Datenanalyse in nahezu Echtzeit für einen IoT-Showcase

Adrian Weber

Die Master Thesis beschäftigt sich mit der Konzeption und Erstellung eines Vorzeigeprojekts aus dem Bereich Internet der Dinge (IoT). Dazu werden Spielzeugautos der Marke Anki OVERDRIVE eingesetzt, welche mit einem Mikroprozessor, Elektromotor, Infrarotsensor und einer Bluetooth-Sendeeinheit ausgestattet sind.

In dem Anwendungsbeispiel sollen nicht nur die Zustände der verschiedenen Fahrzeuge visualisiert werden, sondern auch die Evaluierung einer Steuerung auf Basis von maschinellem Lernen. Im Fokus steht das Aufsetzen einer Lambda-Architektur (LA), mit welcher zum einen die Daten der Fahrzeuge gesammelt beziehungsweise analysiert werden, sowie die Steuerung der Autos ermöglicht wird. Im Bereich Stream-Processing und Monitoring werden einmal kommerzielle und Open-Source Softwareprodukte eingesetzt und anschließend miteinander verglichen. Die Masterarbeit ist in die folgenden Kapitel gegliedert:

- Grundlagen
- Analyse
- Implementierung des Vorzeigeprojekts
- Ergebnisse/Zusammenfassung.

Zuerst werden die wichtigsten Grundlagen für das Verständnis der späteren Lösung kurz vorgestellt. Dazu gehört ein Überblick über die Funktechnik Bluetooth Low Energy (BLE), welche für die Datenübertragung bei den Spielzeugautos eingesetzt wird.

Ebenfalls enthalten in diesem Kapitel ist eine kurze Erläuterung der Lambda-Architektur, bestehend aus den wichtigsten Komponenten und Konzepten dieser Architektur, die besonders gut mit der Datenverarbeitung aus Echtzeit und Langzeitberechnungen umgehen kann.

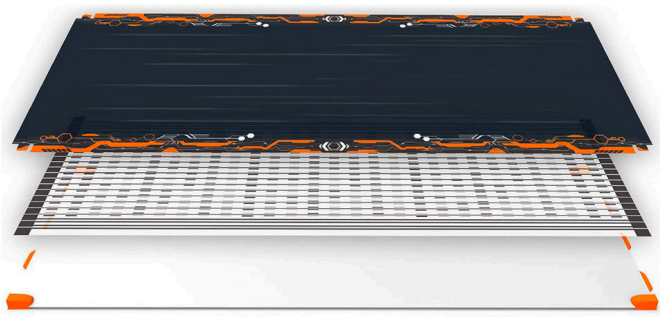
Abschließend ist ein Abschnitt über die unterschiedlichen Kategorien von maschinellem Lernen enthalten. Darin werden die Unterschiede und die bekanntesten Vertreter vorgestellt und mit welchen Kennzahlen die unterschiedlichen Algorithmen miteinander verglichen werden können.

Zu Beginn werden zunächst die Fahrzeuge und die Rennbahn genauer untersucht. Mithilfe einer integrierten Entwicklerschnittstelle können Daten beziehungsweise Nachrichten an die Autos gesendet und empfangen werden. Zunächst wird untersucht, welche Auslöser und Nachrichtentypen existieren und welche Datenfelder in den unterschiedlichen Nachrichten vorhanden sind.

Mit den gewonnenen Daten kann die Topografie der Strecke erstellt werden, mit deren Hilfe sich die Position des Fahrzeugs sowie der Abstand zu anderen Fahrzeugen auf der Strecke ermitteln lassen.

Schließlich wird die Datenqualität der Nachrichten untersucht, dazu gehören beispielsweise Merkmale wie Vollständigkeit, Zeitnähe, Relevanz oder Redundanzfreiheit.

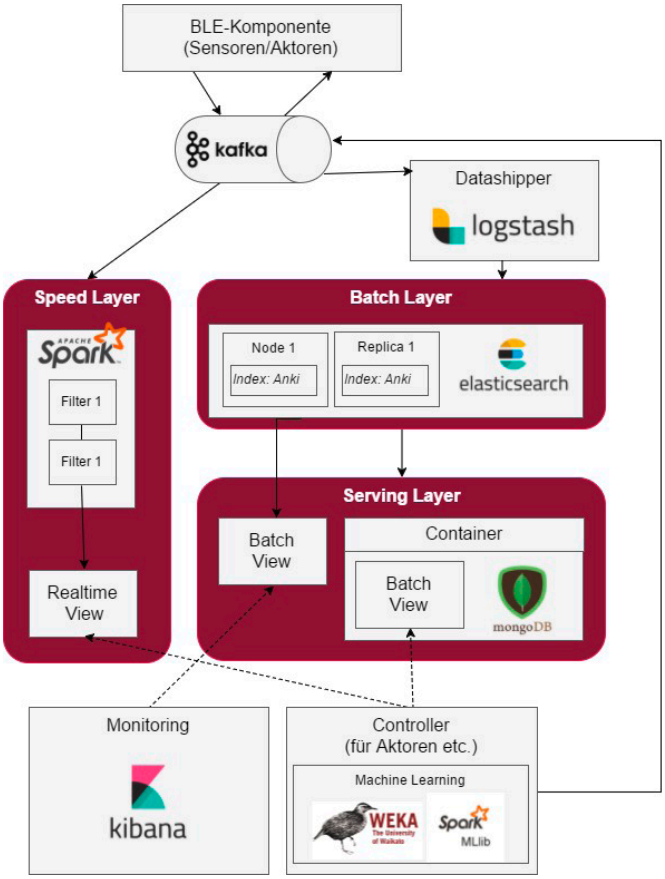
Mit den neu gewonnenen Erkenntnissen werden zwei Szenarien für das Vorzeigeprojekt entworfen: einmal Anti-Kollision und ein Szenario zur Produktionsoptimierung unter Verwendung von maschinellem Lernen (in der Arbeit als Szenario: hohe Übertragungsrate betitelt, da hier die Anzahl übertragener Nachrichten verbessert werden soll).



Streckenelement einer Anki Overdrive Strecke. (Quelle: https://cdn.anki.com/sites/default/_les/car-tech-graphic track v3.jpg)

Im nächsten Kapitel wird dem Leser die Implementierung der Softwarelösung beziehungsweise Lambda-Architektur vorgestellt. Das Kapitel ist gegliedert in vier Abschnitte: Der erste Abschnitt befasst sich mit den allgemeinen Komponenten wie die Fahrzeugsteuerung und die nachrichtenorientierte Middleware (hier Apache Kafka). Darüber hinaus werden die allgemeinen Komponenten der Lambda-Architektur sowie das zugrundeliegende Datenmodel für den Showcase näher beschrieben.

In den zwei darauffolgenden Abschnitten werden einmal die Architektur mit Open Source Lösungen und die Variante mit kommerzieller Software vorgestellt. Dabei wird jeweils auf die verwendeten Softwarelösungen, zum Beispiel ELK-Stack (Elasticsearch, Logstash, Kibana) für Open Source und TIBCO Stream Base für kommerziell, näher eingegangen und gezeigt, für welche Elemente der Lambda-Architektur sie eingesetzt werden können.



Lambda-Architektur mit Open-Source Softwareprodukten.

Im letzten Abschnitt wird konkreter auf die Implementierung der Szenarien mit den zuvor genannten Softwareprodukten eingegangen. Für das Anti-Kollisionsszenario wird gezeigt, wie der Algorithmus zum einen mit Apache Spark (Open Source) und zum anderen mit TIBCO Stream Base (kommerziell) umgesetzt wurde.

Für das Szenario der Produktionsoptimierung wird die Lösung mit maschinellem Lernen genauer vorgestellt. Darüber hinaus wird das Problem mit verschiedenen Algorithmen gelöst und anschließend die bekanntesten Kennzahlen, wie zum Beispiel MAE oder RMSE, der verschiedenen Vertreter präsentiert.

Die Arbeit endet mit einer Zusammenfassung, in welcher nochmals die Open Source mit der kommerziellen Architektur miteinander verglichen und die Anwendung von maschinellem Lernen nochmals kurz zusammengefasst wird.

Die Masterarbeit wurde an der Hochschule Karlsruhe – Technik und Wirtschaft – bei Professor Dr. Uwe Haneke eingereicht.

Analytics and Big Data in Sales Controlling – Status Quo, Challenges, and Opportunities within the DACH Region

Konstantin Weise

Introduction and Research Method

"Every day, we create 2,5 quintillion bytes of data — so much that 90 % of the data in the world today has been created in the last two years alone" (IBM, n.d.-b). From now on, it is expected that these large volumes of data will double every second year (IDC, 2014). The increasing amount of data creates a highly competitive business environment (Baller, Dutta, & Lanvin, 2016, p. 3). Unfortunately, data alone is worthless. Therefore, analytics has to un-leash the potential of big data by transforming data into valuable insights (Gandomi & Haider, 2015, p. 140).

The variety of big data implies a wide range of data formats and sources including structured and unstructured as well as internal and external data (Dash & Subudhi, 2016, p. 4). The large volume and diversity of big data create lots of promising opportunities for organisations (Seufert, 2014, p. 413). Analysing this large amount of structured and espacially unstructured data is a challenging task for organisations (Gandomi & Haider, 2015, p. 137).

However, according to the 2016 "Dreamcar" report published by the International Group of Controlling, the topic "analytics" cannot be disregarded anymore in the field of controlling (Horváth et al., 2016). Within the scope of controlling, it has become crucial to manage and control sales by appropriate analyses because missing sales targets lead to financial deficits and liquidity problems (Havlíček & Břečková, 2013, p. 39).

For that reason, the primary purpose of this master thesis was to identify to what extent analytics and big data have been developed in sales controlling within the DACH region. A detailed literature review aimed to develop a theoretical framework which sets big data and analytics in relation to each other. The empirical part investigated how advanced sales controlling is developed regarding analytics and big data within the DACH region. In this context, the study also intended to identify the main challenges hindering sales controlling to progress analytics and big data further. Based on the problem statement, it was the aim to address the following research questions:

Research Questions:

- What is the current and planned status of analytics and big data in sales controlling within the DACH region?
- Which challenges are hindering the progress of analytics and big data in sales control-ling within the DACH region?

To answer the research questions a deductive approach and a quantitative research method by collecting data through a questionnaire in the DACH region were applied (Saunders, Lewis, & Thornhill, 2016, p. 164). Furthermore, persons who fulfil responsibilities in sales controlling and who are located in Germany, Austria or Switzerland constituted the target group for this research. Data were gathered through an online questionnaire as well as a "delivery and collection questionnaire", consisting of six parts and 19 questions, which was also pre-tested. The sample includes 84 completed questionnaires, which is a reasonable amount for the capacities of one researcher and the specific field of sales controlling in the DACH region. To finally answer the present research questions, the analysis of data and presentation of results were conducted with descriptive statistics using the software SPSS and the spreadsheet software Excel.

Theoretical considerations: Big Data and Analytics

IBM (n.d.-a) states that "data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few". Furthermore, all these different and new emerging data sources do not only increase the volume of data. Moreover, these data also increase the variety of data formats. Beside structured data, big data refers to all variances of semi-structured and unstructured data. A common concept for describing big data is called "Three Vs", volume, velocity, and variety (Chen, Chiang, & Storey, 2012, p. 1182). Furthermore, Nair & Narayan (2011) established the "Booz Big Data Framework" (p. 4), which categorises different types of big data by two dimensions: data format and data source. Additionally, the framework displays an evolutionary process for utilizing big data.

In the next step, analytics has to turn all different types of data into valuable information for organisations and users. Analytics emerged in the late 2000s as a key component of BI (Davenport, 2006, p. 102). The most cited definition of analytics is probably the one from Davenport (Loughborough University, 2013). The author states that analytics is the "extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management" (Davenport & Harris, 2007, p. 7). Furthermore, various authors state that analytics should support decision making by using data and its insights (Davenport & Harris, 2007, p. 7; Herschel & Davis, 2016, p. 15; Horváth et al., 2016, p. 1). Following this approach, analytics can be distinguished into four stages: descriptive, diagnostic, predictive, and prescriptive (Bitkom, 2014, p. 61; Horváth et al., 2016, p. 1). These four stages were combined into the "Analytics Maturity Model", which was introduced by Gartner (Puget, 2015).

Matrix Model for Analytics and Big Data

For the reason that analytics is complementary to data (Varian, 2008), both dimensions, big data and analytics, are interrelated. Unfortunately, previous frameworks, models, and studies focus solely on one dimension.

To discover the status quo of sales controlling within analytics and big data, the comprehensive framework "Matrix Model for Analytics and Big Data" was developed within this study (Figure 1).

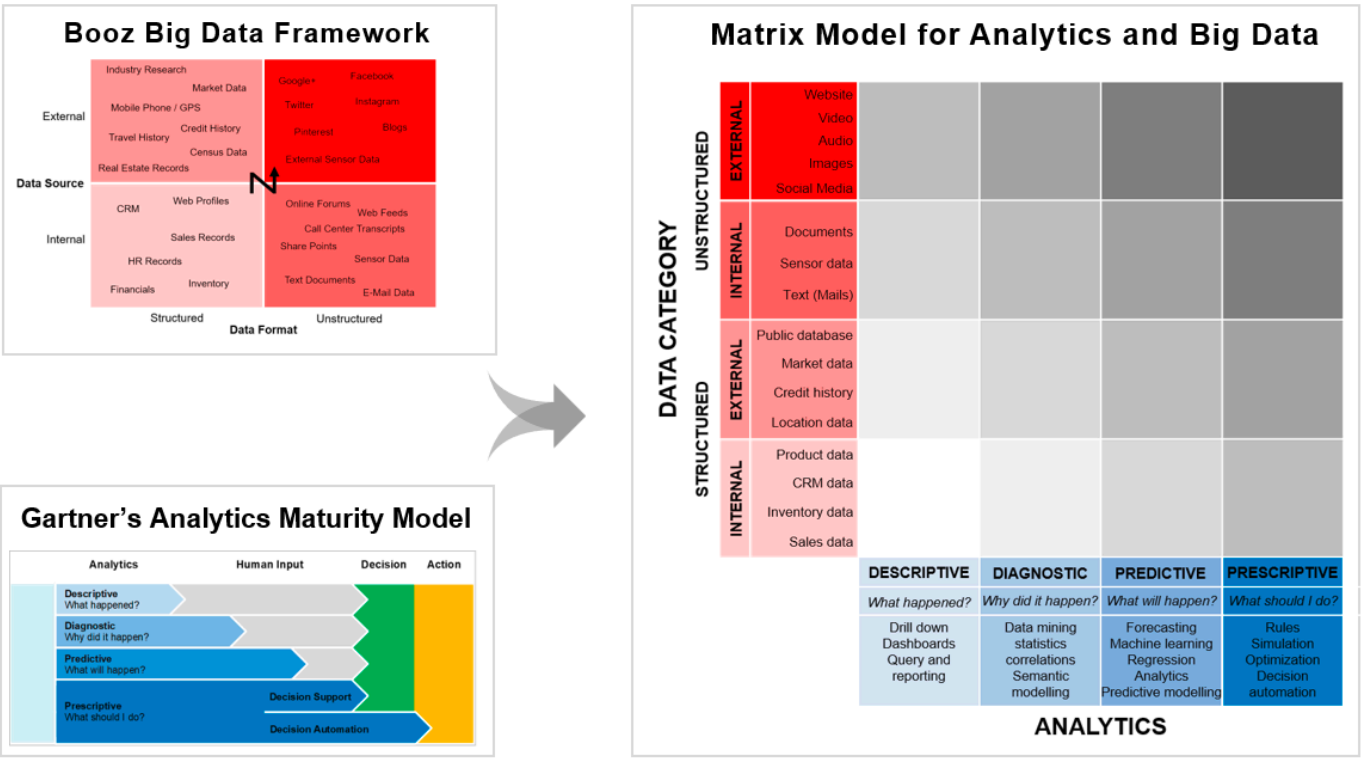


Figure 1: Development of "Matrix Model for Analytics and Big Data"

Overall, the "Matrix Model for Analytics and Big Data" results in 16 different analytics and big data sections. The more sections are in use, the more advanced the progress of big data and analytics is. Moreover, the most developed status is shown in the top right corner, when external-unstructured data is processed for prescriptive analytics.

Emperical Findings

The quantitative research shows that sales controlling is not yet utilising the full potential of analytics and big data. The reason is that sales controlling primary analyses structured data, whereas external or unstructured data types are analysed by 20 or less percent of the participants. As displayed in Figure 2, the only exception within external data is market data, which is analysed by 43 percent. Additionally, the newly introduced "Matrix Model for Analytics and Big Data" reveals that the practical application of analytics decreases from descriptive towards prescriptive in each of the four data categories.

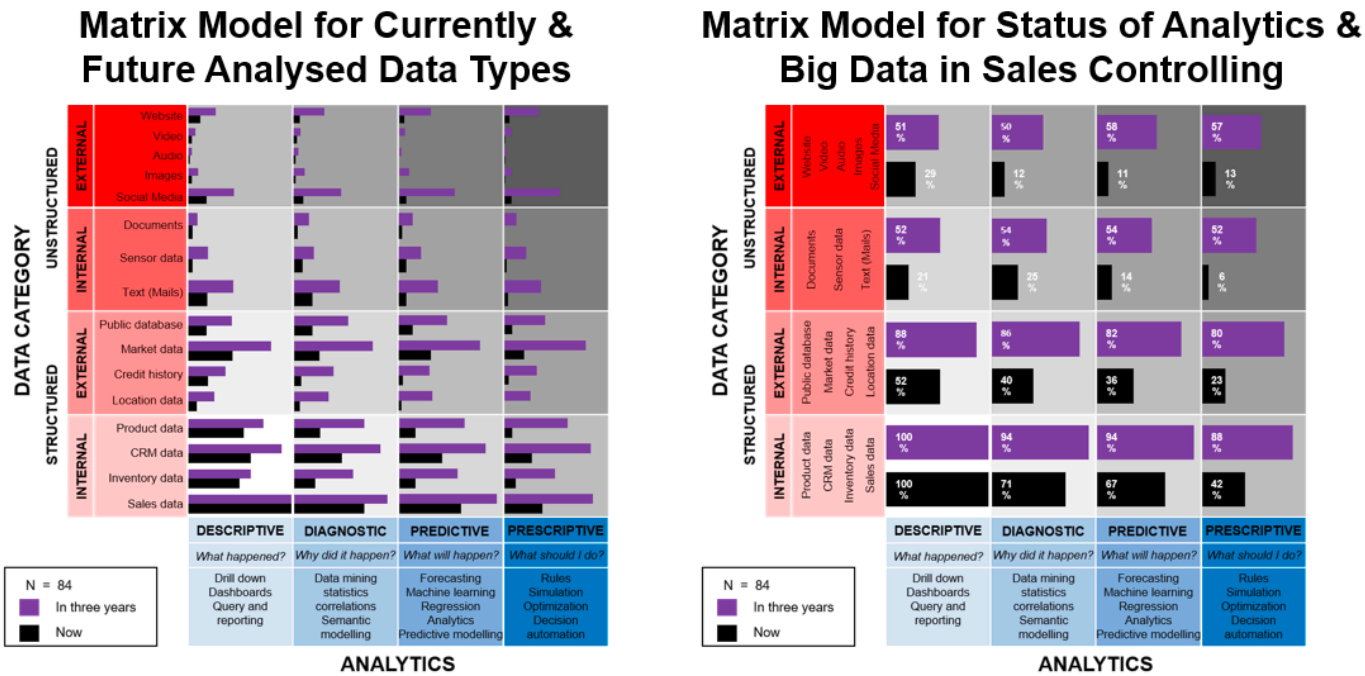


Figure 2: Matrix Model – Results

Furthermore, Figure 2 illustrates that 100 percent of participants descriptively analyse internalstructured data. However, only five of 16 sections of the Matrix Model are currently exploited by at least 50 percent of the participants. These sections are in the lower left corner of the Matrix Model and do not exceed structured data and predictive analytics. Conversely, nine of 16 sections are rarely utilised, which confirms the finding that the majority within sales controlling is currently not ready to make use of unstructured data and prescriptive analytics. Additionally, it can be stated that sales controlling is more focused on developing the four analytical levels for structured data, rather than developing capabilities for utilizing "big" data opportunities derived from external and unstructured data types further. For these reasons, sales controlling is currently taking advantage of improving reporting and forecasting, but it does not completely leverage the opportunities to enhance customer analysis, to improve pricing, to extend market and competitive analysis.

Another finding is that sales controlling might be more capable of taking advantage of analytics and big data opportunities in the future. The reason is that more participants are striving to use different sections and data types in three years. Nearly all representatives of sales controlling want to use all four analytical levels for structured data. In contrast, only half of the participants plan to analyse unstructured data in the future, which means that 50 percent of participants limit themselves by not planning to fully exploit the opportunities of big data and analytics in sales controlling.

Additionally, this study shows that the lack of knowledge and insufficient capabilities to handle big data are the main reasons hindering sales controlling to progress analytics and big data. To overcome this challenge, the Pearson Chi-Square Test illustrates that smaller companies are more capable to analyse external as well as unstructured data. Consequently, bigger organisations can learn from smaller companies how to handle big data. In addition, it becomes clear that participants in a sales controlling department more often use predictive analytics compared to their colleagues in a central controlling department. This master thesis also revealed the implication for researchers and managers that the distinctive variables "current level of importance" and "familiarity with best practice examples" are of major importance for the exploitation and progress of analytics and big data in sales controlling within the DACH region. In addition, the Pearson Chi-Square Test shows that participants are further developed within the Matrix Model when they evaluate analytics as important and are familiar with best practice cases. Finally, the Spearman's rank-order correlation outlines that the importance of analytics and the familiarity with best practice cases also leads to less strongly perceived challenges. Concluding these findings, it can be suggested to clearly illustrate the importance of analytics within the company and to stay informed about best practice examples to fully exploit the opportunities of analytics and big data within sales controlling in the DACH region.

Die Masterarbeit wurde am Management Center Innsbruck im Master Programm International Business & Management bei Professorin (FH) Dr. Claudia Brauer eingereicht.

The Transformational Effects of Predictive Analytics on Planning

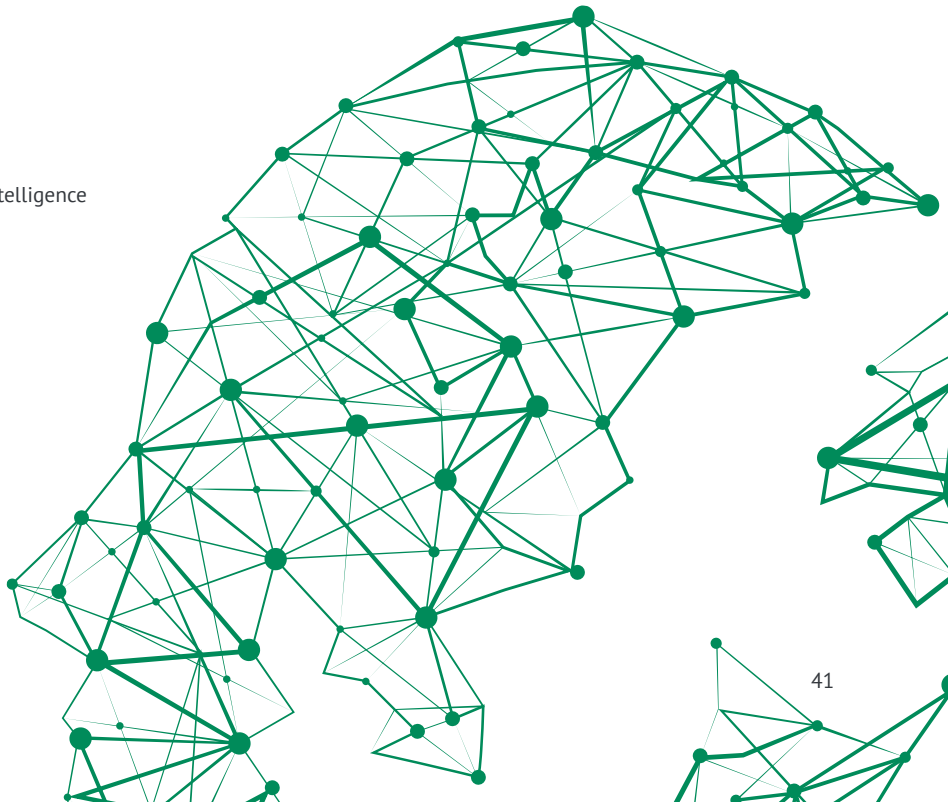
Jessica Wewer

As organizations face an ever-increasing complexity in its business activities, planning has developed to one of the most important functions within a company. In order to improve the business planning an increasing number of companies relies on the implementation of analytics. Thereby, there are not only internal but also several external factors, which may affect financial measures of a company but which cannot be controlled for directly.

Most of the forecasting research available is based on qualitative analyses of the current status of business analytics applications in organizations only. The thesis, therefore, presents a case study developing a predictive analytics model by identifying the most appropriate statistical methods in order to analyze the effect of several external factors on the KPIs net sales, cost of sales, research and development expenses and working capital of an automotive supplier. Thereby, various combinations of countries and macro and microeconomic factors are tested for using the statistics program "R".

With an average deviation of 6.19% and 7.68% respectively, the statistical methods Autoregressive-Moving Average with Exogenous Terms (ARMAX) and Extreme Gradient Boosting provide the best forecasting results for the KPIs. It becomes obvious that out of more than 200 external indicators, the BCI of Estonia seems to be best suited to forecast the test period.

Die Masterarbeit wurde an der Universität Duisburg-Essen am Lehrstuhl für Wirtschaftsinformatik insbes. Business Intelligence bei Professor Dr. Peter Chamoni eingereicht.



TDWI e.V.
Lindlaustr. 2c
53842 Troisdorf
Tel. +49 (0) 2241 2341 212
Fax +49 (0) 2241 2341 199
E-Mail info@tdwi.eu

tdwi.eu