Konkrete Anwendungsfälle von KI & Big-Data in der Industrie

Leitfaden

www.bitkom.org

bitkom

Herausgeber

Bitkom
Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V.
Albrechtstraße 10 | 10117 Berlin
T 030 27576-0
bitkom@bitkom.org
www.bitkom.org

Projektleitung

Dr. Nabil Alsabah | Bitkom e.V.
T 030 27576-242 | n.alsabah@bitkom.org

Verantwortliche Bitkom-Gremien

AK Artificial Intelligence
AK Big Data & Advanced Analytics

Autoren

Sharif Abdel-Halim | adesso AG
Dr. Hicham Aroudaki | umlaut SE
Paul Balzer | MechLab Engineering UG
Jaroslav Bláha | CellmatiQ GmbH
Friedhelm Bösche | Küttner Automation GmbH
Oliver Brandmüller | Deutsche Bahn AG
Andreas Festl | Virtual Vehicle Research GmbH
Lars Fockele | Postbank AG
Olaf Hein | ORDIX AG
Arian van Hülsen | PTC Inc.
Thorsten Jankowski | Voith Group
Christian Kaiser | Virtual Vehicle Research GmbH
Caroline Kleist | mayato GmbH
Ibrahima Kourouma | Deutsche Bahn AG
Dr. Kim Nilsson | Pivigo
André Rauschert | Fraunhofer IVI
Dr. Matthieu-P. Schapranow | Hasso-Plattner-Institut für Digital Engineering gGmbH
Dr. Achim Steinacker | intelligent views GmbH
Dr. Alexander Stocker | Virtual Vehicle Research GmbH

Satz & Layout

Kea Schwandt | Bitkom e.V.

Titelbild

© OriginalDelivery – stocksy.com

Inhaltsverzeichnis

1 Bei KI und Big-Data mangelt es oft an konkreten Anwendungsfällen

1 Bei KI und Big-Data mangelt es oft an konkreten Anwendungsfällen

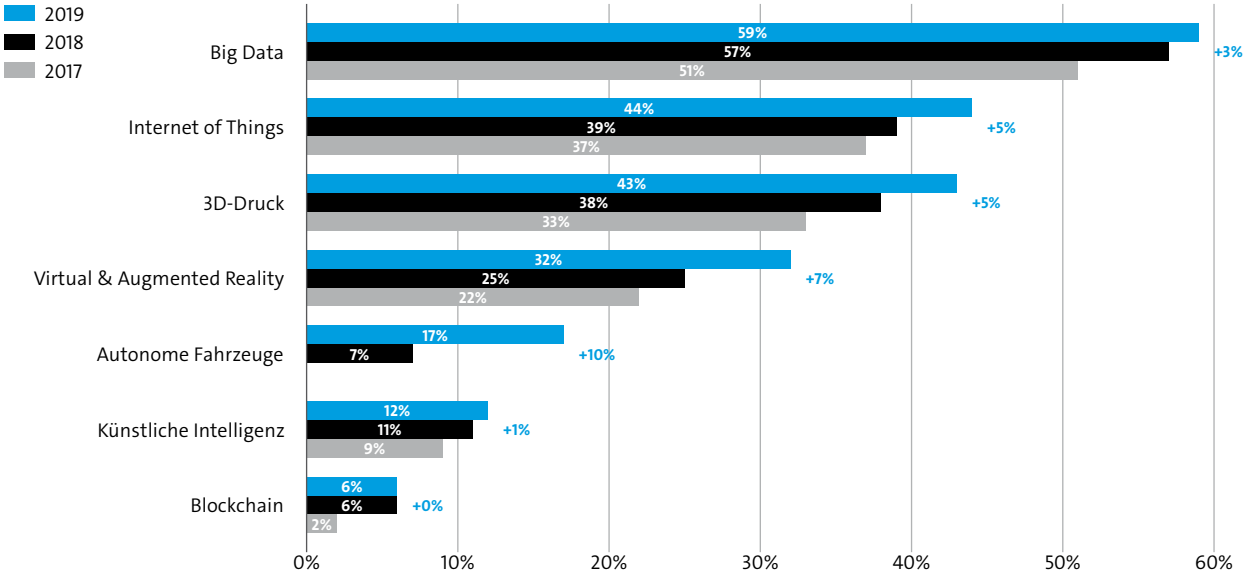
Nabil Alsabah

Im Rahmen der hub.berlin 2019 hat Bitkom Präsident Achim Berg die Ergebnisse einer repräsentativen Bitkom-Studie zur Digitalisierung der Wirtschaft vorgestellt. Berg stellte fest, dass die Digitalisierung den Wettbewerb intensiviert. Der intensivierte Wettbewerb motiviert Unternehmen dazu, ihr Portfolio an Produkten und Dienstleistungen weiterzuentwickeln. Dabei sehen sich Unternehmen mit vielen Herausforderungen konfrontiert. Trotz dieser Herausforderungen stehen Unternehmen der Digitalisierung in ihrer großen Mehrheit aber überaus positiv gegenüber. Der Gestaltungswille ist da.

Auch bei den Schlüsseltechnologien hat die oben zitierte Bitkom-Studie keine Erkenntnisdefizite identifiziert. Manager schreiben Kerntechnologien – wie der Künstlichen Intelligenz – eine bedeutende Rolle für die Wettbewerbsfähigkeit der deutschen Wirtschaft zu. So sagen 8 von 10 Befragten, dass Digitalunternehmen wie Amazon oder Google durch ihre führende Stellung bei KI zu einer ernstzunehmenden Konkurrenz für deutsche Kernindustrien werden. Und fast genauso viele sind der Meinung, dass KI als Technologie entscheidend dafür ist, ob deutsche Unternehmen künftig weiter weltweit erfolgreich sein werden. 6 von 10 Unternehmen sind zudem sicher, dass KI die wichtigste Zukunftstechnologie überhaupt ist.

Top-Technologien kommen nur langsam in der Praxis an

Welche Technologien werden in Ihrem Unternehmen genutzt oder der Einsatz ist geplant / wird diskutiert?



Basis: Alle befragten Unternehmen (2019: n = 606; 2018: n = 604; 2017: n = 505) | Angaben für »aktuell im Einsatz« und »Einsatz geplant oder diskutiert«
Quelle: Bitkom Research

Doch leider lässt die Mehrheit der Befragten den Worten keine Taten folgen. Denn in der Unternehmenspraxis liegt der Anteil der Firmen, die über den Einsatz von KI diskutieren, ihn planen oder umsetzen, bei 12 Prozent. Dies steht im Gegensatz zu Big-Data: Fast 60% der Unternehmen setzen diese Technologie um.

Die Zurückhaltung der Unternehmen bei KI hängt wahrscheinlich mit dem Mangel an konkreten Anwendungsfällen zusammen. Die öffentliche Debatte ist stark von futuristischen Science-Fiction-Szenarien geprägt. Diese haben oft wenig Relevanz für das Alltagsgeschäft der Unternehmen. Auch allgemeine Pauschalaussagen über das Potenzial von KI helfen nicht weiter.

Bitkoms Flaggschiffkonferenz, der Big-Data.AI Summit (#BAS), will diesem Umstand entgegenwirken: An zwei Tagen bringen wir Tausende von Experten zusammen, um sich über konkrete KI- und Big-Data-Lösungen auszutauschen. Die präsentierten Vorträge sind drei Clustern zuzuordnen. (1) Branchenübergreifende Strategien, Technologien und Trends; (2) Branchenspezifische Lösungen; und (3) Gesellschaftliche Herausforderungen.

Am 10. und 11. April 2019 haben über 200 Referenten von Großunternehmen, KMUs und Startups praktische Ansätze für konkrete Probleme vorgestellt. Über 8.000 Besucher haben sich im Rahmen von #BAS19 und hub.berlin mit diesen Lösungsansätzen auseinandergesetzt. Das Feedback war so positiv, dass wir die Referenten aus dem Cluster »Branchenspezifische Anwendungen« um einen schriftlichen Beitrag gebeten haben. Das dabei entstandene Konferenzbuch fasst in zwölf Beiträgen Use-Cases zusammen. Diese decken ein weites Spektrum an Industrien ab: **Autonome Mobilität** (Kapitel 2, 3 und 4), **Zahnmedizin** (Kapitel 5), **Post** (Kapitel 6), **Stahl** (Kapitel 7), **Logistik** (Kapitel 8), **E-Health** (Kapitel 9), **Industrie 4.0** (Kapitel 10 und 12), **Handel** (Kapitel 11) und **Fintech** (Kapitel 13).

Am 1. und 2. April 2020 wird der nächste Big-Data.AI Summit stattfinden. Die Teilnehmer können sich auf weitere inspirierende Anwendungsbeispiele freuen. Zusammen mit unserer Schwesterveranstaltung, hub.berlin, erwarten wir 10.000 Besucher aus Wirtschaft, Politik und Wissenschaft. Der #BAS20 wird die optimale Plattform sein, um die praktische Nutzung von KI und Big-Data in Deutschland und Europa voranzubringen. Seien Sie dabei und lassen Sie sich inspirieren!

In der Zwischenzeit wünschen wir Ihnen viel Spaß bei der vorliegenden Lektüre.

2 Nutzung von Automotive Big Data zur Entwicklung zweier neuer Anwendungen

2 Nutzung von Automotive Big Data zur Entwicklung zweier neuer Anwendungen

Christian Kaiser, Andreas Festl & Alexander Stocker

2.1 Einleitung und Motivation

Ein Fahrzeug ist ein Computer auf vier Rädern. In einem modernen Fahrzeug analysieren zahlreiche Steuergeräte eine enorme Menge an Daten, welche von der verbauten Sensorik während der Fahrzeugnutzung generiert wird, um die Funktion des Fahrzeugs und der darin verbauten Systeme zu gewährleisten. Schon die von einem einzigen Fahrzeug generierten Daten sind für die Entwicklung datengetriebener Anwendungen und Dienste hochgradig spannend. Das wahre Potenzial der Fahrzeugdaten erschließt sich allerdings erst dann, wenn viele Fahrzeuge aus Flotten oder gar alle sich im Straßenverkehr befindlichen Fahrzeuge überhaupt ihren Big-Data-Datenschatz zur Verfügung stellen und eine Kombination von Fahrzeugdaten mit Daten aus anderen Domänen angestrebt wird.

Der vorliegende Beitrag zeigt die Nutzung der von Sensoren generierten Daten in zwei konkreten Anwendungen auf: die Detektion des Fahrverhaltens einzelner Fahrer zum Aufzeigen des Fahrrisikos sowie die Detektion von Schlaglöchern in einem urbanen Straßennetz über mehrere Fahrer und Fahrzeuge hinweg. Der Beitrag beschreibt die dazu erforderlichen Schritte von der Definition der Anwendungen, dem Abgreifen der Daten des Fahrzeugs, der Datenvorverarbeitung, der eigentlichen Datenanalyse und der Generierung von Ergebnissen zur Entscheidungsunterstützung für Fahrer und Infrastrukturverantwortliche.

Zu diesem Zweck wurde eine Big-Data-Analytics-Plattform eingesetzt, welche in dem von der Europäischen Kommission geförderten Projekt EVOLVE entstanden ist und zahlreiche für Data Scientists relevante Open-Source-Tools wie beispielsweise Spark, Kafka, Flink, Elastic, oder Jupyter in einem System vereint. Die beiden im Beitrag beschriebenen Anwendungen wurden als Demonstratoren auf dieser Big-Data-Analytics-Plattform realisiert.

2.2 Datenakquise und Datenstruktur

Für die Datensammlung wird ein durch Virtual Vehicle Research GmbH entwickelter Datenlogger verwendet, welcher auf Basis eines kostengünstigen Einplatinencomputers (BeagleBone Black, vergleichbar mit einem Raspberry Pi) realisiert wurde. Dieser Datenlogger wird mit der On-Board Diagnoseschnittstelle (OBD) eines Fahrzeugs verbunden und startet automatisch mit der Datenspeicherung, wenn das Fahrzeug gestartet und betrieben wird. Er beendet automa-

tisch die Datenspeicherung und fährt herunter, wenn der Motor abgestellt wird. Neben dem Abgreifen klassischer Fahrzeugdaten wie beispielsweise Geschwindigkeit oder Drehzahl enthält der Logger noch zusätzliche Sensorik zur Positionsbestimmung sowie zur Erfassung von Rotation und Beschleunigung.

Die vom Datenlogger produzierten Rohdaten sind tabellarischer Form und enthalten immer genau den Messwert eines Signals zu jedem Zeitpunkt (Tabelle 1). Insbesondere können mehrere Zeilen unterschiedliche Signale zum gleichen Zeitpunkt beschreiben. Zusätzlich variiert die Rate, mit der die Werte erfasst werden, zwischen den Signalen. Innerhalb eines Signals ist diese Abtastrate annähernd konstant, kleinere Abweichungen sind jedoch möglich und üblich. Da der Datenlogger von den Fahrern an unterschiedlichen Positionen im Fahrzeug montiert werden kann, ist das Koordinatensystem der Beschleunigungs- und Rotationssensoren des Datenloggers grundsätzlich nicht automatisch am Fahrzeug ausgerichtet – die x-Achse des Beschleunigungssensors muss daher beispielsweise nicht parallel zur x-Achse des Fahrzeugs verlaufen.

Messzeitpunkt	Signalname	Signalwert
2018-9-13 5:28:36.206089	Motordrehzahl	1500
2018-9-13 5:28:36.226331	Accelerometer-X	0.476
2018-9-13 5:28:36.245312	Geschwindigkeit	39
2018-9-13 5:28:36.268915	Öltemperatur	90
...	...	...

Tabelle 1: Struktur der Rohdaten

Um die durch das Fahrzeug generierten Daten überhaupt für die Entwicklung datengetriebener Anwendungen nutzen zu können, ist ein umfangreicher Datenverarbeitungsprozess nötig, welcher im folgenden Kapiteln beschrieben wird.

2.3 Datenverarbeitungsprozess

Um die für die Services notwendigen Informationen aus den Daten zu extrahieren, müssen drei große Schritte durchlaufen werden: In der Datenvorverarbeitung werden die Daten so vorbereitet, dass eine sinnvolle Weiterverarbeitung erst möglich wird. In den vorbereiteten Daten werden im nächsten Schritt relevante Events wie starkes Beschleunigen, scharfes Abbremsen oder das Überfahren eines Schlaglochs detektiert. Im letzten Schritt, der eigentlichen Analyse, werden die bereits berechneten Bestandteile zusammengeführt und in interpretierbaren Visualisierungen dargestellt.

2.3.1 Datenvorverarbeitung

In diesem Schritt werden zunächst alle Signale nach Ausreißern abgesucht und diese entfernt. Die Signale einiger Sensoren wie etwa des Beschleunigungssensors und des Gyroskops enthalten viel Rauschanteil und müssen geglättet werden. Zusätzlich wird die Struktur der Daten verändert, indem jedes Signal »seine eigene« Spalte bekommt. Damit das möglich wird, müssen alle Signale interpoliert und mit einer regelmäßigen Frequenz abgetastet werden. Das Ergebnis hat wieder tabellarische Form, nun entspricht jedoch jede Zeile genau einem Zeitpunkt und der zeitliche Abstand zwischen den Zeilen ist konstant, beispielsweise 0.1s / 10Hz (Tabelle 2).

Messzeitpunkt	Motordrehzahl	Accelerometer-X	Geschwindigkeit	Öltemperatur	...
2018-9-13 5:28:36.20000	1500	0.477	39	90	...
2018-9-13 5:28:36.30000	1501	0.479	40	90	...
...	...	...	...	...	...

Tabelle 2: Struktur der vorverarbeiteten Daten

Eine besondere Herausforderung stellt die unbekannte Lage des Datenloggers im Fahrzeug dar. Für eine erfolgreiche Datenanalyse ist es unabdingbar, die Richtungen der gemessenen Beschleunigungen und Rotationen zu kennen – diese Richtungen hängen jedoch von der unbekannten Datenlogger-Position ab. Die Messungen müssen also parallel zu den Koordinatenachsen des Fahrzeugs »gedreht« werden (Abbildung 1). Dazu bestimmen wir die Fahrtrichtung, die Straßennormale und die Seitenrichtung aus den Messungen und berechnen daraus eine Rotationsmatrix.

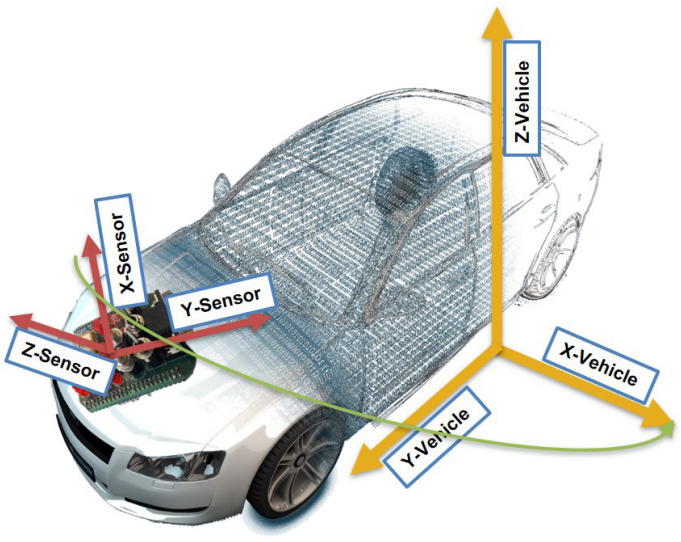


Abbildung 1: Messungen müssen zur Koordinatenachse des Fahrzeugs ausgerichtet werden.

2.3.2 Event-Detektion

In den so vorbereiteten Daten kann nun nach für die Services relevanten Events gesucht werden. Je nach Event-Typ sind dabei andere Signale relevant. Alle Events werden noch nachbearbeitet um beispielsweise zeitlich nahe beieinander liegende Events, die nur durch eine kurze Unterbrechung getrennt werden, zu einem einzelnen Event zusammenzufassen. Die detektierten Events werden mit Wetter und Positionsdaten verknüpft, so dass für jedes Event Zeit und Ort des Auftretens, sowie das zu diesem Zeitpunkt dort herrschende Wetter bekannt ist.

Schlagloch-Events

Dieser Event-Typ zeigt das Überfahren eines Schlagloches (oder ähnlicher Straßenschäden) an. Zur Detektion werden die Beschleunigung normal zur Straße, sowie das »Nicken« des Fahrzeugs (also die Rotation um seine Seitenachse) betrachtet. Wenn sich die Vorderreifen im Schlagloch befinden, liegt die Front des Fahrzeugs tiefer als das Heck, wenn sich die Hinterreifen im Schlagloch befinden, ist es umgekehrt. Dadurch kommt eine typische »Nick-« Bewegung zustande die erkannt werden kann.

Beschleunigungs- und Brems-Events

Zum Erkennen von starken Beschleunigungs- und Bremsvorgängen eignen sich besonders die Signale der Fahrzeuggeschwindigkeit, der Beschleunigung in Fahrtrichtung und der Rotation um die Seitenachse (»Nicken«). Das »Nicken« kommt durch die Änderung der Gewichtsverteilung bei Geschwindigkeitsänderungen zustande: Beim Beschleunigen wandert mehr Gewicht zur Hinterachse – das Heck senkt sich und die Front hebt sich. Beim Bremsen ist es umgekehrt. Diese Bewegungen können erkannt werden. Da jedoch die Erkennung über nur ein einzelnes Signal fehleranfällig sein kann, verwenden wir in unserem Algorithmus immer mehrere Signale, die alle gleichzeitig ausschlagen müssen, um die Detektion auszulösen.

Schnelle Kurvenfahrt-Events

Das schnelle Durchfahren von Kurven ist durch hohe Seitenbeschleunigungen gekennzeichnet. Um diese Information zu nutzen, muss man jedoch zunächst die Kurven selbst detektieren. Dazu wird aus Fahrzeuggeschwindigkeit und Seitenbeschleunigung ein approximativer Kurvenradius an jeder Stelle berechnet. Mithilfe dieser Information können die Kurven selbst gefunden und in weiterer Folger die schnelle Kurvenfahrt detektiert werden.

2.3.3 Implementierung

Der gesamte Datenverarbeitungsprozess wurde auf einer Big-Data-Plattform implementiert, die unter anderem Apache Hadoop, Apache Spark und passende Benutzeroberflächen bereitstellt. Alle Schritte können dort beim Eintreffen neuer Daten vollautomatisch ausgeführt werden.

Zwischenergebnisse werden in eigenen Datasets gespeichert, um die Verwendung in anderen Services oder das Teilen mit anderen Nutzern zu ermöglichen. Abbildung 2 zeigt eine Übersicht der Datenverarbeitungskette.

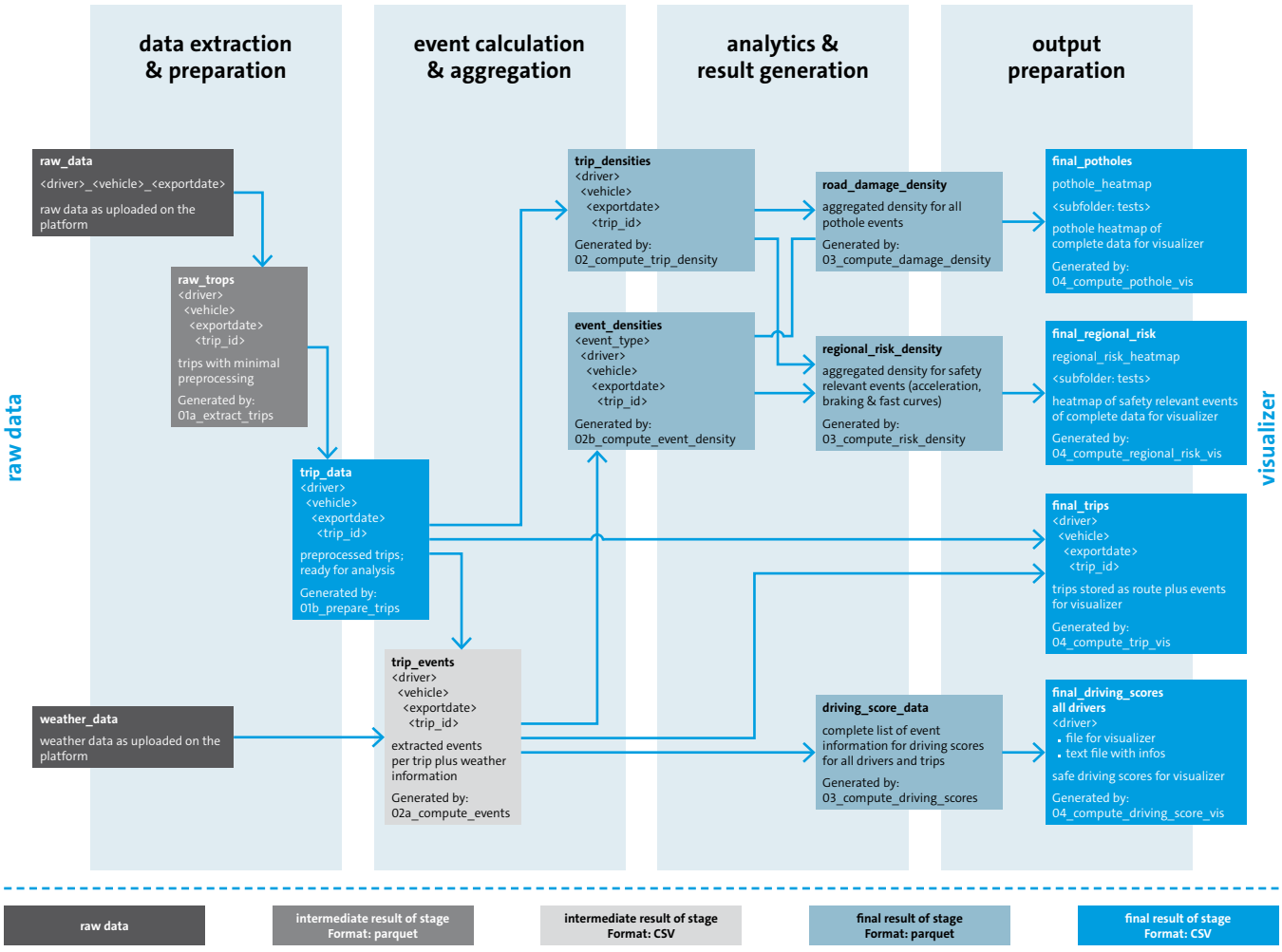


Abbildung 2: Die Datenverarbeitungskette, um aus Rohdaten von Fahrzeugen beispielsweise starke Bremsvorgänge zu identifizieren und in einer Visualisierung anzuzeigen.

2.4 Anwendungen

Durch entsprechende Sammlung, Vorverarbeitung und Analyse der Daten lassen sich Anwendungen erzeugen, die Nutzer aus unterschiedlichen Domänen und Bereichen unterstützen können. Wir möchten zwei Beispiele herausgreifen. Zum einen die Detektion des Fahrverhaltens von einzelnen Fahrern, ein Anwendungsfall der einzelne Fahrer persönlich bei der Analyse der

eigenen Fahrweise unterstützen soll, und zum anderen die Detektion von Schlaglöchern, die insbesondere für Infrastruktur-Betreiber interessant ist.

2.4.1 Detektion von individuellem Fahrverhalten

Die Fahrweise von Autofahrern kann in vielen Facetten sehr unterschiedlich sein (z.B.: gemütlich/vorausschauend/aggressiv, Gang-Wahl, durchschnittliche Streckenlänge und Geschwindigkeit, Aufmerksamkeit, Müdigkeit, etc.), und hat je nach Fahrzeugtyp (Benzin, Diesel, Elektro mit und ohne Rekuperation, Eigengewicht, Leistung) kleinen oder großen Einfluss auf Verbrauch, Verschleiß und die Sicherheit im Straßenverkehr. Um dies abzubilden, benutzen wir alle im Kapitel 3 berechneten Events und die Information, zu welchem Trip und zu welchem Fahrer diese gehören, um einen ›risk score‹ zu errechnen, der angibt, wie sicher ein einzelner Trip war. Je mehr sicherheitsrelevante Events pro Fahrt auftreten (in Relation zur Streckenlänge), desto niedriger der Wert. Einfluss hat jedoch nicht nur die reine Anzahl der sicherheitsrelevanten Events, sondern auch die Umstände, insbesondere das Wetter. So hat eine starke Bremsung bei Regen eine größere Reduktion des Werts zufolge als die gleiche Bremsung auf trockener Straße. Der Wert selbst ist ein statistischer Rang, so bedeutet beispielsweise ein Wert von 56,72%, dass dieser Trip besser als 56,72% aller Trips in der Datenbank ist. In einer Karten-Visualisierung (Abbildung 3) werden dem Fahrer dann zum Beispiel alle Events des Trips mit einer Markierung angezeigt. Blaue Markierungen entsprechen starken Beschleunigungen, während rote Markierungen starken Bremsvorgängen entsprechen.

Driver1
Total riskscore (%): 50.37

	trip ID	risk score (%)	duration (h)	accevents (n)
0	20181119, Trip_069	56.72	0.44	12.0
1	20181119, Trip_066	60.40	0.34	7.0
2	20181119, Trip_065	48.55	0.40	16.0
3	20181119, Trip_064	93.57	0.45	5.0
4	20181119, Trip_063	37.06	0.41	21.0
5	20181119, Trip_062	49.54	0.13	4.0
6	20181119, Trip_061	43.40	0.16	7.0
7	20181119, Trip_059	23.81	0.34	26.0
8	20181119, Trip_058	15.32	0.11	9.0
9	20181119, Trip_056	45.94	0.14	6.0
10	20181119, Trip_055	89.25	0.66	3.0
11	20181119, Trip_054	46.04	0.19	6.0
12	20181119, Trip_053	31.62	0.26	19.0
13	20181119, Trip_052	89.48	0.48	6.0
14	20181119, Trip_051	34.35	0.28	16.0

Abbildung 3: Karten-Visualisierung einer Fahrt (links) und Details der Fahrten von Driver 1, zum Beispiel ist die Fahrt mit trip_ID »Trip_069« weniger Risikoreich als 56.72% aller Fahrten (risk score).

2.4.2 Detektion von Schlaglöchern

Auch Straßen altern, bekommen Asphaltrisse, Löcher, verformen sich. Neben der gesetzlichen Pflicht der Straßenerhaltung durch Gemeinden und Städte (Haftung im Schadensfall bei Vorsatz und Grobfahrlässigkeit), möchte man den Bewohnern und Besuchern der Gemeinde/Stadt natürlich auch eine adäquate Infrastruktur bieten. Die Stadt Graz verbraucht ca. 25.000 Tonnen Asphalt und 40.000 Tonnen frostsicheres Material pro Jahr und saniert so 100.000 m² Straßen jährlich [1]. Doch wie findet man heraus, welche Straßen in welchem Grad beschädigt sind und am dringendsten Instandsetzung oder Erneuerung benötigen? Ein Wirtschaftsbetrieb der Stadt Graz schreibt, dass 2008 »90 % des Straßennetzes mit drei eigenen Erfassungsteams visuell aufgenommen« wurden und »nach einem systemisierten Schadenskatalog« nach Schulnotensystem bewertet wurden. Hier zeigt sich auch schon das Problem: So eine Analyse ist kostspielig, zeitaufwändig, nicht vollständig objektiv und erfasst nicht alle Straßen. Hier unterstützt unser zweiter Anwendungsfall, die automatische Detektion und Darstellung von Schlaglöchern/Straßenschäden auf einer Karte. Konkret werden die gefundenen Schäden in einer Heatmap dargestellt – die Farbe korrespondiert zur Sicherheit, mit der ein Schlagloch detektiert wurde (Abbildung 4).

Die Daten stammen von mehreren Fahrzeugen, die privat im Grazer Raum verwendet werden. Diese wurden von uns mit Sensoren ausgestattet: Dadurch haben wir bereits eine hohe Abdeckung der Grazer Innenstadt erreicht. Schlaglöcher werden, wie in Abschnitt 3.2.1 beschrieben, als Event erkannt und zusammen mit GPS-Position, Datum und Uhrzeit gespeichert. Je größer nun der Anteil an Fahrten über den ein und denselben Streckenabschnitt, bei denen ein Schlagloch detektiert wird, ist, desto gravierender ist das Problem. In der Karte zeigt sich das so: Kleinere Probleme sind violett, mittelgroße Probleme grün bis hin zu gelb, und große Probleme rot. Dies lässt sich auch als Tabelle darstellen, um den Straßenzustand anhand von objektiven Zahlen vergleichen zu können, auch wenn die Dämpfung des Fahrgastraumes, wo der Sensor platziert ist, das Ergebnis ein wenig verfälscht. Die einzelnen Problemstellen können punktgenau (soweit es die GPS-Genauigkeit zulässt) kommuniziert und repariert werden.

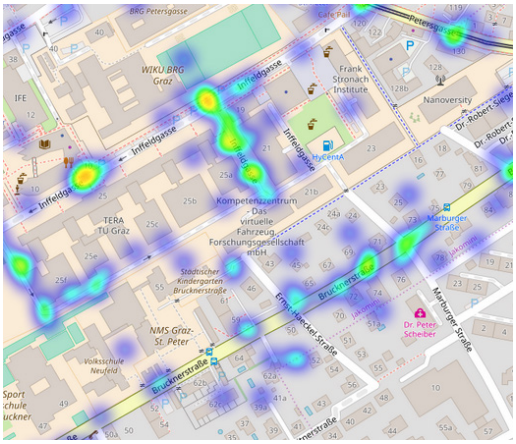


Abbildung 4: Heatmap von erkannten Schlaglöchern/Straßenschäden in einem Teil von Graz, Österreich.

2.5 Zusammenfassung

Der vorliegende Beitrag hat gezeigt, wie von einem Fahrzeug generierte Daten über eine Datenverarbeitungspipeline zwei interessante Anwendungen ermöglichen, die Detektion des individuellen Fahrverhaltens einzelner Fahrer und die Aggregation dieser Daten zu einer Karten-Visualisierung, sowie die Detektion von Schlaglöchern in einem urbanen Straßennetz und die Visualisierung des Ergebnisses in einer Heatmap. Der Beitrag verweist dabei auf die zahlreichen Herausforderungen in der Datenverarbeitung und zeigt dazu Lösungsansätze auf.

2.6 Danksagung

The EVOLVE project (www.evolve-h2020.eu) has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 825061. The document reflects only the author's views and the Commission is not responsible for any use that may be made of information contained therein. The publication was written at VIRTUAL VEHICLE Research Center in Graz and partially funded by the COMET K2 – Competence Centers for Excellent Technologies Programme of the Federal Ministry for Transport, Innovation and Technology (bmvit), the Federal Ministry for Digital, Business and Enterprise (bmdw), the Austrian Research Promotion Agency (FFG), the Province of Styria and the Styrian Business Promotion Agency (SFG).

2.7 Abbildungs-, Tabellen- und Literaturverzeichnis

[1] ↗ **http://www.gestrata.at/publikationen/archiv-journal-beitrage/gestrata-journal-124/gehweg-radweg-und-strasenerhaltung-in-graz**

3 Utilization of Crowd Data to Answer Diverse Questions within the Contexts of Urban Mobility and Smart City

3 Utilization of Crowd Data to Answer Diverse Questions within the Contexts of Urban Mobility and Smart City

Hicham Aroudaki

3.1 Introduction

Since a few years, umlaut SE is using crowdsourcing as a significant addition to classical methods for collecting network performance data and measuring quality of experience from the end user's point of view. Our app-based, on-device data collection engine is working passively in the background of a user's smartphone. With this app-based approach, end-user smartphones become measurement devices delivering invaluable knowledge on how customers use and experience mobile communications. The data collection engine is integrated into more than 800 applications worldwide and generates more than 3,5 billion data samples per day. Basis for integrating the data collection engine is a cooperation between umlaut SE and the application providers. While installing one of their applications the user is normally informed about the additional software by the usage terms and conditions. He has the choice to accept it or not.

For illustration purposes Figure 1 shows the sample distribution within and around the city of Aachen in Germany. The brightness represents the sample density and provides an indication for the frequently driven routes and visited locations.

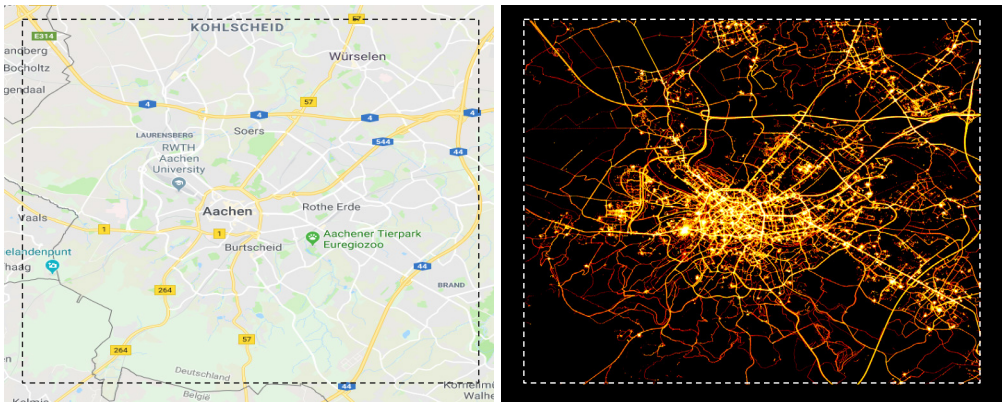


Figure 1: Crowd Data distribution within and around the city of Aachen Germany (width: 12 km, height: 9 km).

The collected data is fully anonymous, no personal identifiers (MSISDN, IMSI, IMEI or IP addresses) are collected. Users are differentiated by a hashed ID derived from the release information of the installed Android operating system.

In addition to the network parameters, Crowd Data provides for each sample valuable additional information on location, speed, and time. This information and its derivatives (acceleration, driven distances, mobility share per km2, etc.) allow to answer a variety of interesting questions related to Urban Mobility and Smart City concepts. In the following, selected use cases are briefly described to illustrate the spectrum of possible applications.

3.2 The analysis approach

Any of the analysis examples described in section 3 starts with the extraction of Crowd Data for a certain period, mostly 6 months. Such periods assure high sample spread as well as indecency from seasonal effects and events. Then, tailored data cleansing steps are applied to correct/remove corrupt or inaccurate data, e.g. very high or low values caused by faulty devices. During an additional step, data is filtered and clustered based on use case specific criteria, which are normally aligned with the customer. Filters can be for example time slots, areas, attributes, or user groups. Users of specific targeted groups can be identified based on different pre-defined features, such as visited Points of Interest (POIs), installed applications, phone call behavior, Bluetooth connections, and much more attributes. User group identification is then followed by the association of the samples to each of the identified groups. Samples are then aggregated following two approaches:

- Aggregation based on a grid of tiles (squares of the same size covering the complete analyzed area), using the sample specific geo-coordinates. The grid size is use-case specific and normally aligned with the client (e.g. 100 m, 200 m).

- Association of each sample to a POI if it was generated within a pre-defined radius around the POI-centre. The radius may vary due to different POI categories.

Once all samples are assigned to tiles or POIs, the corresponding areas (squares, circles) become objects with various attributes, e.g. maximum/average number of users, average duration of stay, maximum/average speed, or frequency of visits per user. Defining these and other attributes helps to rank POIs and tiles to gain indications for further assumptions and decisions. This approach is illustrated in its main steps in Figure 2.

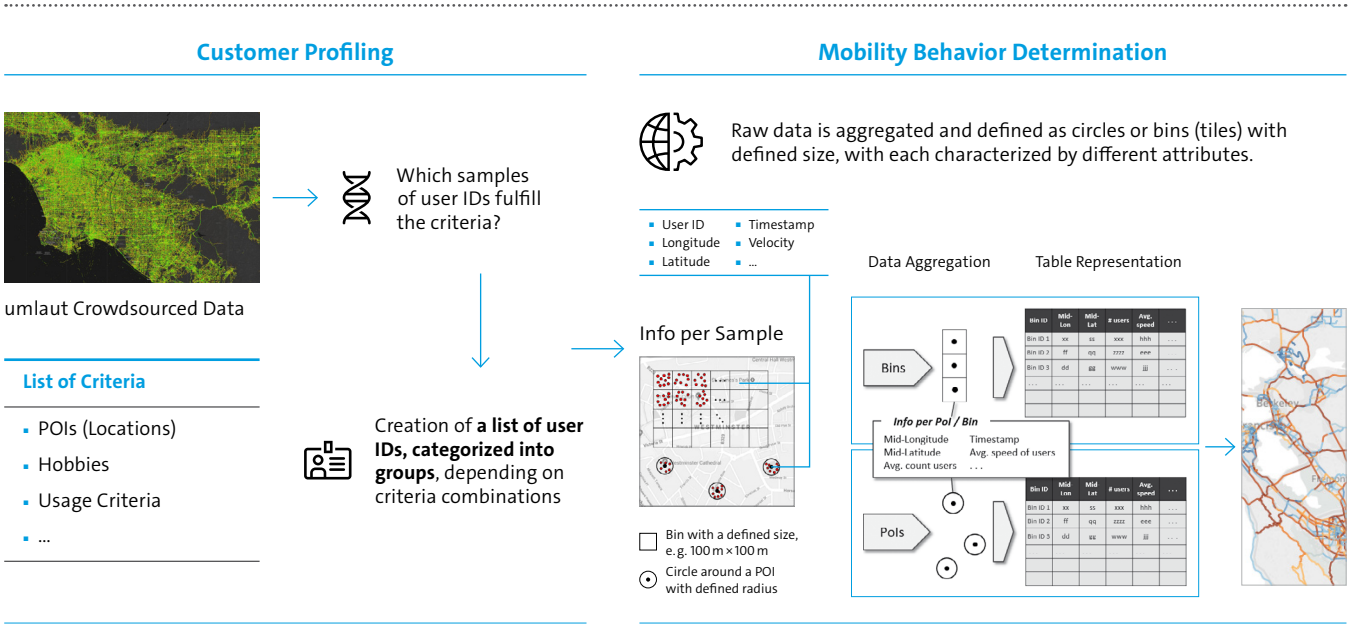


Figure 2: Simplified steps of the analysis approach (customer profiling and aggregation of samples to tiles/POIs).

3.3 Sample Use Cases

umlaut SE used this approach to support specific questions of clients in diverse operational fields, e.g. mobile communications, automotive, infrastructure supplier, and asset and real estate owners. In the following, some selected use cases related to the fields Urban Mobility and Smart City are briefly described. The focus is set on describing the use case and the possible benefits rather than on technical details.

Car Sharing

umlaut SE was approached by a car manufacturer that wants to start a car sharing service in a major city in the USA. For this purpose, umlaut SE Crowd Data offered a convenient way to provide transparency on the mobility behavior of certain target groups, which were specified by the client (travellers, students, freelancers, etc.). Knowing the users' mobility behavior would enable the client to identify perfect locations for car sharing stations that would already consider the potential customers' habits and interests.

Based on pre-defined criteria, such as installed applications or temporal mobility patterns, the collected Crowd Data was further processed to identify the different target groups within the captured users. The aggregation of a group's samples in combination with POIs led to the identification of commonly used user routes (start and end point) and frequently visited locations.

Using this information on each target group, a network of car sharing locations are being planned for set up to attract potential customers by providing a tailor-made solution.

To illustrate some of the results, the left part of Figure 3 shows the count of users (represented by the color) and the average duration of stay (represented by the height) for a gird of tiles covering the complete city. The grid resolution is 200 m x 200 m. The right part of the Figure illustrates the start-end-analysis which was performed for all POIs and tiles in the city, broken-down at user group level.

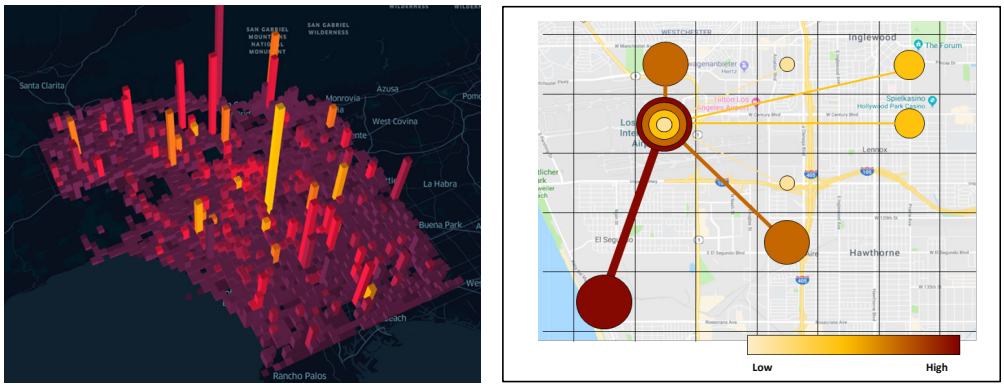


Figure 3: Selected results of the car sharing analysis. Left: user density aggregated in tiles, right: start-end analysis.

Charging Infrastructure for electric vehicles

At the time a premium car manufacturer (OEM: Original Equipment Manufacturer) decided to develop and distribute its own charging infrastructure for electric vehicles within Germany, the umlaut SE Crowd Data has proven to be an appropriate technology to support the decision-making process and the roll-out strategy. The main motivation was to identify suitable locations alongside the daily routes of existing and potential customers of the premium OEM, considering various Points of Interest and time slots. The idea was to make the charging experience as convenient as possible, avoiding customers to interrupt their daily habits or routines.

An extensive research was performed to identify potential customers of the OEM, based on premium hobbies, typical premium locations customers would visit, business or luxury applications premium customers would use, or Bluetooth-identities of selected car brands. In order to consider the most possible scope of features, over 64,000 locations in Germany were defined as relevant POIs and the mobility pattern was analyzed with special focus on these locations. This led to the identification of frequently visited locations, frequently driven routes, both in urban areas and along the highways, including the average duration of stay at the respective locations. This approach supported the customer to narrow down all existing POIs to a reduced and manageable list of promising locations, which became subject to further analysis tasks (e.g. available electricity infrastructure, legal aspects, type and capacity of charging infrastructure, etc.). Thus,

the analysis' results facilitated discussions and negotiations with potential cooperation partners before the roll-out finally starts.

umlaut SE provided the client with a tool, summarizing the analysis results and allowing to dynamically derive strategic results based on set parameters (Figure 4). The tool allows to differentiate between different features (routes, POIs, days, time slots within the day) and attributes (count of users, frequency of visits, average duration of stay, average driven distances, etc.). Furthermore, statistics are showing the fulfillment of each threshold for the visualized parameters in order to estimate the compliance to the thresholds within the considered area.

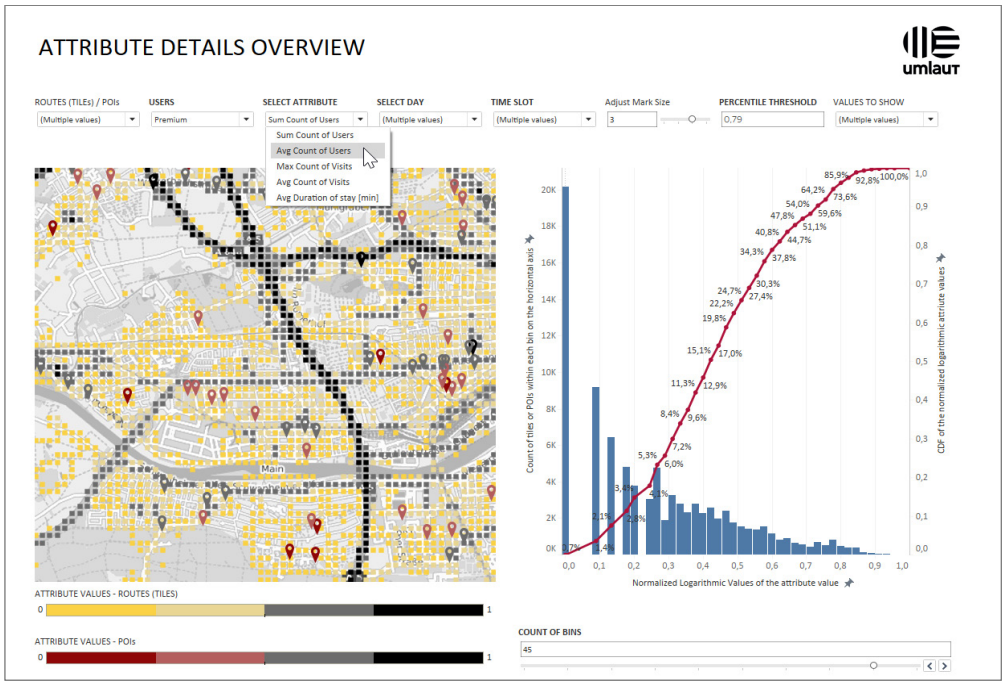


Figure 4: Developed tool for POIs and routes evaluation.

Rescuing 4.0

Efficient emergency care is a key factor in saving lives and can determine the further treatment and scope of a critical disease. In order to respond to extensive demands and the complex problems presented in the daily business of emergency care, umalut SE is working on the development of a new planning tool called »preRESC«. This tool uses available data from the emergency control center (e.g. historical operational information) and integrates new data sources (Crowd Data, geographic and weather information, historical traffic information, street construction points, etc.) to build the base for an in-depth data analysis (see Figure 5). By applying advanced correlation, Clustering and Decision Tree algorithms, a more effective planning of resources is facilitated. The control center dispatcher can gain insights at strategic and operational levels, and therefore better anticipate changes in emergency deployment demands. In addition, the tool

allows better coordination between emergency care teams in different districts. This provides a »digitalized« approach to support the smart emergency protection of the future.

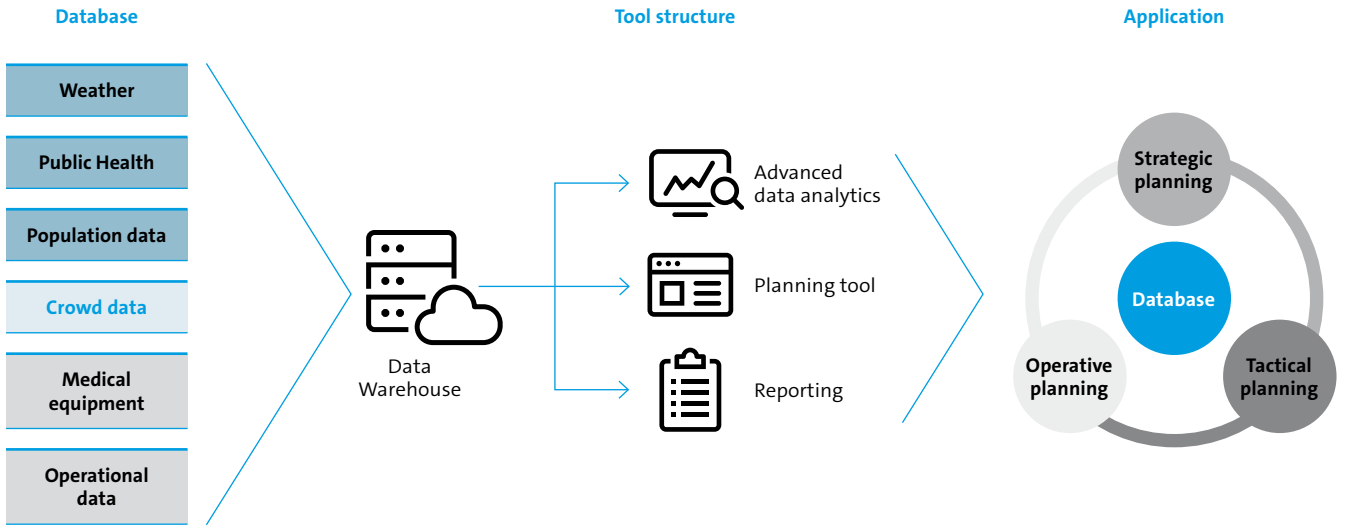


Figure 5: Concept of preRESC.

3.4 Outlook

The experience gained while executing the different use cases helped to identify various aspects for enhancements and optimization to further develop the approach to its full potential. The most relevant aspects are:

- Implement a higher grade of automation in dealing with data management tasks and the adaptation of individual steps to new questions. This would help to execute more complex use cases at a lower level of cost and needed time and efforts.

- Intensify the integration of external data into the overall analysis chain. The consolidation of external data into a common information pool, applicable for different countries and use cases, would facilitate the consideration of additional dimensions of attributes and features.

- Apply machine learning concepts as a second level of analysis. Clustering and pattern recognition analysis after the data aggregation into bins and POIs could help to reveal further geographic and temporal similarities in terms of mobility pattern, thus providing deep dives while answering specific questions.

3.5 List of Figures

4 Cartox² Konnektivitäts-Vorhersage: Ein datengetriebener Deep-Learning-Ansatz

4 Cartox² Konnektivitäts-Vorhersage: Ein datengetriebener Deep-Learning-Ansatz

Paul Balzer & André Rauschert

4.1 Einleitung

Die Automatisierung ist in Fahrzeugen angekommen. Aktuelle Modelle ermöglichen teilautomatisierte Fahrt, einige Hersteller haben vollautomatisierte Fahrzeuge im Zulassungsprozess. Diese nach SAE Level 3 bzw. 4 fahrenden Fahrzeuge meistern komplexe Autobahnsituationen derzeit bis 60 km/h als sogenannte Staupiloten. In den nächsten Jahren ist die Übernahme der Fahraufgabe bis 120 km/h geplant. Hierbei entstehen im Detail herausfordernde Fragestellungen im Spannungsfeld StVO, Mensch-Maschine-Interaktion und Fahrdynamik/-komfort. Es lassen sich beliebig viele Situationen beschreiben und konstruieren, in welchen die Umfeldsensorik der Fahrzeuge nicht ausreichend Informationen für die sichere und gleichzeitig hinreichend zügige Fahrstrategie liefern. Um einiges komplexer sieht die Situation in Innenstädten aus. Hier ist die Sensorik prinzipbedingt nicht in der Lage, verdeckte Hindernisse zu erkennen.

Eine Lösung für intelligente Fahrzeuge der Zukunft ist es, diese mit Funktechnologie (sogenannter Cartox²-Kommunikation) auszustatten und den Verkehrsfluss damit sicherer, schneller und komfortabler zu realisieren.

Die Zukunft des automatisierten und vernetzten Fahrens beginnt mit zuverlässiger Kommunikation zwischen Fahrzeugen, Verkehrsinfrastruktur und anderen Verkehrsteilnehmern. Hierfür ist der WLANp-Standard (nach 802.11p) etabliert, der in Modellprojekten in ganz Europa eingesetzt und getestet wird. Die Fragestellung, die sich bei sicherheitskritischen Anwendungen ergibt: Ist der Empfang oder die Kommunikation möglich, ist die Verbindung und vollständige Datenübertragung schnell und vollständig sichergestellt?

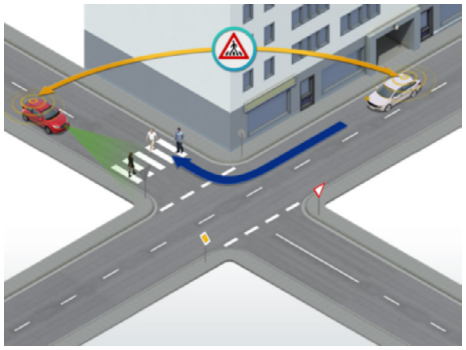


Abbildung 1: Kooperative, vernetzte Fahrzeuge tauschen Informationen über sich und andere Verkehrsteilnehmer aus. Eine sichere Fahrreaktion ist nur möglich, wenn die Datenübertragung fehlerfrei und zuverlässig ist. Woher weiß der Kommunikationspartner, dass die Kommunikation möglich ist?

4.2 Datenfluss von Fahrzeugen über Edge-Clouds ins Cluster

Ein Prädiktionsmodell für Cartox²-Kommunikation kann eine Abschätzung zu den Funkbedingungen zwischen den Kommunikationspartnern liefern. Es ermöglicht das Abschätzen der Übertragungsparameter zwischen Fahrzeugen in Echtzeit. In diesem Kapitel geht es um die Herangehensweise zum Trainieren eines solchen Vorhersagemodells aus real gemessenen Daten.

Messsystem Fahrzeug mit integrierten Cartox² Modulen

Die für die Anwendungsfälle entwickelten Messsysteme erfassen und verarbeiten Informationen zur Cartox²-Konnektivität während des täglichen Fahrbetriebs im Realverkehr. Es werden u.a. folgende Daten von den Fahrzeugen erhoben und gespeichert:

- Fahrzeugdaten (Geschwindigkeit, eigene Position, Fahrtrichtung, ...),
- Kommunikationsdaten (ETSI-ITS Nachrichten, die mittels 802.11p Standard versendet/empfangen werden),
- Positionsdaten sowohl zwischen Fahrzeugen als auch mit Kommunikationsinfrastruktur (Road Side Units [RSU])

Abbildung 2: Beispielhaft von den Fahrzeugen gesammelte Kommunikationsdaten

Datenaufbereitung mittels Cartox²-Big-Data-Infrastruktur

Die Fahrzeuge übertragen die Daten über Edge-Clouds an die Big-Data-Plattform, welche hybride Kommunikations- und Positionierungsdaten zusammenträgt und aufbereitet.

Aus technischer Sicht nutzen die Stream-Processing-Server neueste IoT-Protokolle zur Konsolidierung aller Fahrzeugdaten. Kafka übernimmt das Datenstreaming mittels Konnektoren, um

Daten vom Broker einzulesen und zu Cassandra und HBase zu transferieren. Stream-Apps stehen für die Datenaufbereitung zur Verfügung, wie z.B. Map-Matching, Ride-Detection und Filter-Event-Detection [3]. Cassandra speichert Fahrzeugbewegungen und Filterdaten, um zu erkennen, zu welchem Zeitpunkt Fahrzeuge in der Verkehrsinfrastruktur aktiv waren. Neun Batch-Processing-Server nutzen HBase zum Speichern der Rohdaten, Spark zur Analyse der Rohdaten und Jupyter-Notebook-Systeme zur Datenanalyse.

Zusätzlich werden heterogene Datenströme aus anderen Quellen (bspw. mCloud, OSM, Inspire, Wetter, Behördendaten) in der Cartox² Plattform zusammengetragen.

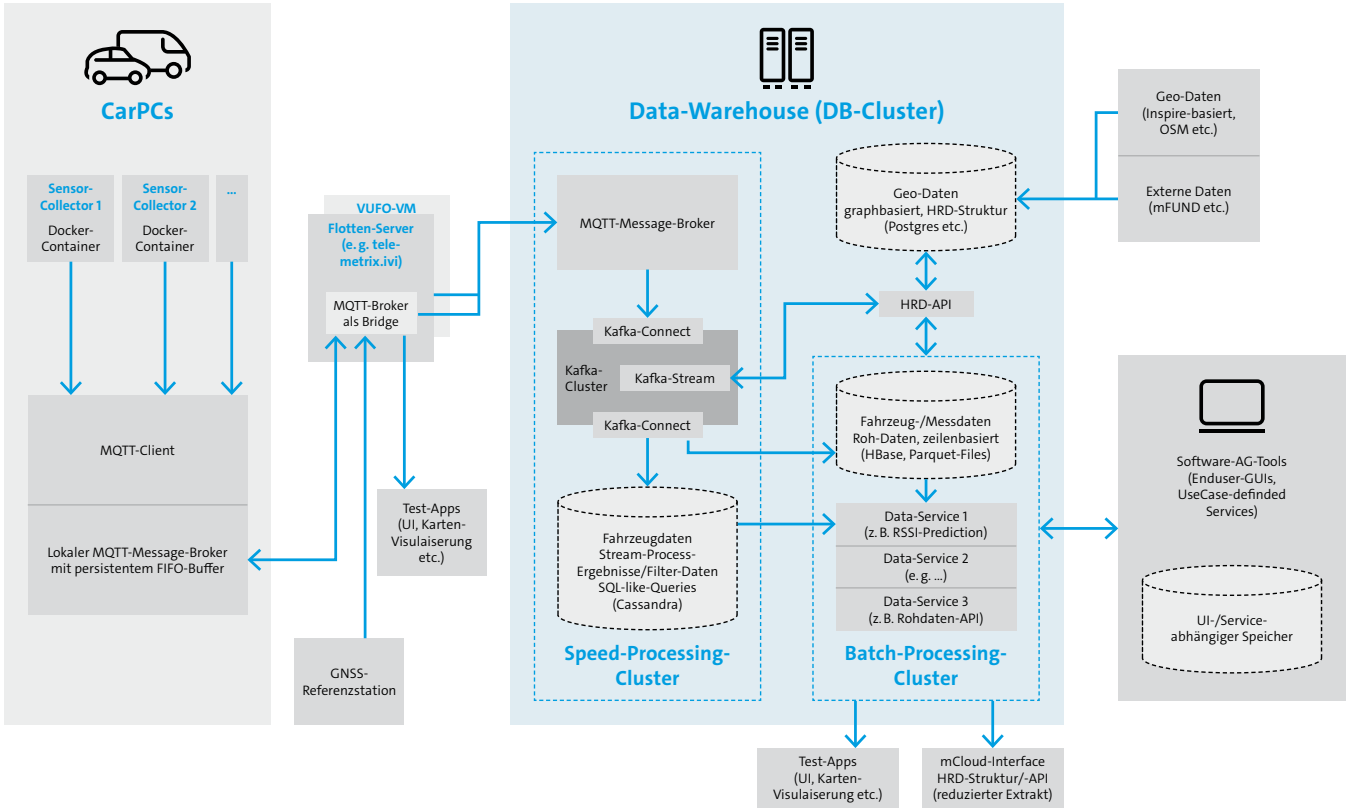


Abbildung 3: Datenfluss von Fahrzeugen (links) in Big-Data-Cluster (Mitte) und Ausgabe der API (rechts)

4.3 Vorhersagemodell für die Fahrzeug-Kommunikation

Cartox² trägt hybride Kommunikations- und Positionierungsdaten mit weiteren Geodaten auf einer Big-Data-Plattform zusammen. Es reichert statische mit Echtzeitinformationen an, um mit Datenanalyse-Algorithmen neuartige Dienste und Geschäftsmodelle bereitzustellen und eine intelligente und sichere Verkehrsinfrastruktur zu gewährleisten.

Abbildung 4: Datenbasis von OpenStreetmap als 3D Rendering, Quelle: Screenshot von demo.f4map.com

Die Ausbreitung von Funkwellen ist stark von der tatsächlichen Umgebung abhängig. Daher muss ein Vorhersagemodell möglichst detaillierte Informationen über die baulichen Gegebenheiten kennen. Die gesammelten Kommunikationsdaten wurden daher mit frei zugänglichen Geodaten angereichert. Beispielhaft seien Gebäude, Straßenarten und Tunnel genannt.

Mittels Deep Learning wurde ein Vorhersagemodell trainiert, welches die zu erwartende Signalstärke (RSSI) in Abhängigkeit der tatsächlichen Gegebenheiten zwischen den Kommunikationspartnern vorhersagt. Anders als gängige Software zur analytischen Berechnung von Funkausbreitungsmodellen (z.B. Raytracern), liefert das neuronale Netz prinzipbedingt eine Vorhersage in Echtzeit.

4.4 Use Cases zur Anwendung des Vorhersagemodells

Use Case	Anwender	User Story
1	Verkehrsunfallforschung	Die Cartox²-Serviceplattform versetzt Akteure aus dem Bereich der Unfallforschung nun in die Lage, bereits bekannte Unfallschwerpunkte nach neuen Kriterien zu bewerten. Wie groß ist das Unfallvermeidungspotenzial von Cartox²-Kommunikation? An welchen Unfallschwerpunkten ist die Installation von Infrastruktur (RSU) sinnvoll?

Use Case	Anwender	User Story
2	Navigationssoftware-Hersteller	Cartox² ermöglicht es Herstellern, neben den klassischen Navigationsoptionen (»schnellste Route« und »kürzeste Route«) zukünftig die Route mit der »besten Konnektivität« als eine weitere Alternative anzubieten. Insbesondere bei hochautomatisierten Fahrzeugen steigen Sicherheit und Effektivität, wenn neben der fahrzeugeigenen Sensorik eine optimale Versorgung mit dynamischen Verkehrs- und Umfelddaten gewährleistet ist.
3	Kommunen und Infrastrukturbetreiber	Die im Cartox²-Projekt entwickelte hybride Funknetzplanung bewertet die Konnektivität und begleitet den Verkehrsinfrastrukturausbau durch eine smarte Fusion aus Mobil- und C2X-Funktechnologien, Positionierungsdiensten und Wetterinformationen – natürlich unter Beachtung der baulichen Randbedingungen.

Am exemplarischen Use Case der Funkplanung wird die Komplexität der Fachexpertise deutlich. Die Funkplanung bildet eine Schlüsselkomponente für die intelligenten Verkehrssysteme der Zukunft. Sie identifiziert vorab Kommunikationsdefizite und prädiziert eine orts- und zeitabhängige Funkabdeckung auf zu erschließenden Gebieten im gesamten Bundesgebiet.

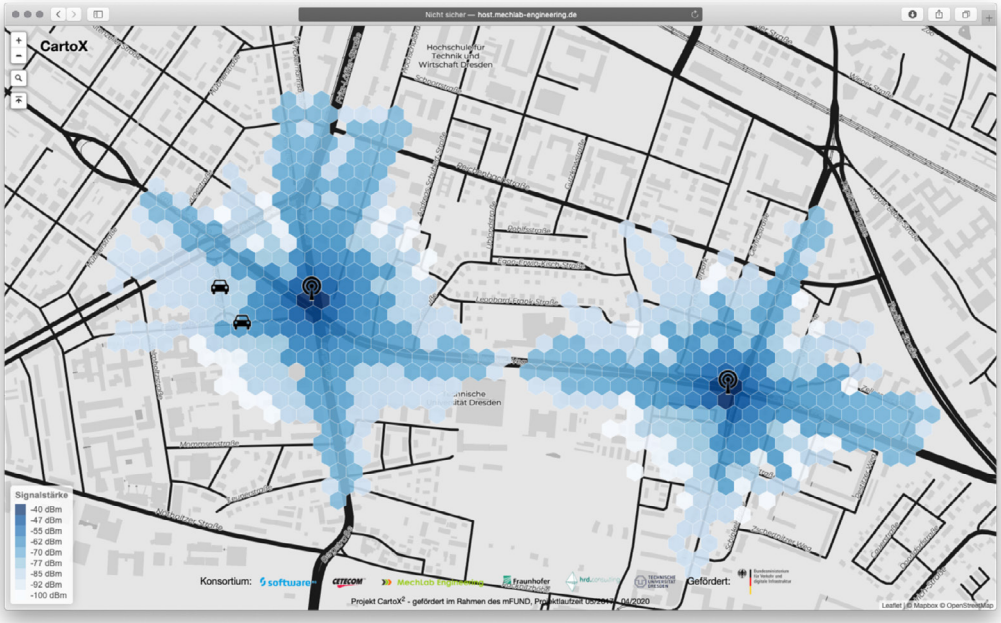


Abbildung 5: Cartox²-Vorhersage der zu erwartenden Signalstärken von WLANp-Road-Side-Units virtuell platziert an zwei Kreuzungen im Stadtgebiet Dresdens.

4.5 Zusammenfassung

Mit der Entwicklung eines Modells zur Vorhersage von Kommunikationsbedingungen ergeben sich erste Use Cases für die Verkehrsunfallforschung, Infrastrukturbetreiber und das konnektivitätsoptimierte Routing. Die Basis dafür wird durch die hochperformante Big-Data-Plattform bereitgestellt, die für Basisdienste des vernetzten und automatisierten Fahrens zur Verfügung steht.

Nutzern ohne Fachexpertise bietet Cartox² eine vereinfachte, anwendungsfreundliche Umgebung, welche den Fokus auf Planung und technischen Betrieb legt. Dies dient nicht nur der Vereinfachung des hochkomplexen Themas, sondern eignet sich ebenfalls hervorragend für webbasiertes verteiltes Arbeiten zwischen Teams an unterschiedlichen Standorten.

Das Projekt Cartox² wurde gefördert durch das Bundesministerium für Verkehr und digitale Infrastruktur.

4.6 Abbildungsverzeichnis

5 Schöne Zähne mit KI

5 Schöne Zähne mit KI

Jaroslav Bláha

5.1 Zahnärzte haben auch Schmerzen

In den vergangenen Dekaden war die Zahnmedizin durch einen hohen individuellen und handwerklichen Aufwand bei der Diagnostik und Zahnreparatur geprägt. Ansätze zur Automatisierung erfolgten nur schrittweise, z.B. in der Produktion von Alignern oder der Einführung von CAD-Software für die weitestgehend manuelle Planung von Zahnersatz oder Implantaten. In allen Teilbereichen der Zahnmedizin erschweren die Vielfalt des menschlichen Körpers, die Menge an möglichen Anomalien und die Komplexität der dreidimensionalen Formen, z. B. für Zahnersatz, die Standardisierung.

Mehrere Faktoren erzeugen aktuell eine kritische Masse, die Zahnmediziner zwingen wird, neue Arbeitsweisen in Betracht zu ziehen. Insbesondere sind dies:

- Ökonomischer Druck durch die allgemeinen Umstände (z.B. Mieten in Großstädten, Vergütung durch Krankenkassen) und die deutliche Zunahme von Zahnarztketten, die ihre Zahnarztpraxen auf Effizienz und gleichmäßiges Qualitätsniveau trimmen müssen.

- Die gravierende Zunahme von bildgebender Diagnostik (mit ca. 10% jährlich) in allen Bereichen der Medizin in Verbindung mit der Novellierung der Strahlenschutzverordnung. Diese erfordert, dass auch Zahnärzte jede produzierte Aufnahme vollständig befunden (d.h. z.B. nicht nur eine Karies identifizieren, sondern auch alle evtl. vorhandenen Zysten und Tumore).

- Die Standesverbände erwarten ein durchgehendes Qualitätsmanagement aller Tätigkeiten, was mit manuellen Methoden einen immensen Aufwand verursacht.

- Nicht zuletzt erwarten auch die Patienten, die aus ihrem täglichen Leben ständig digitalen Prozessen ausgesetzt sind, moderne Arbeitsweisen. So ist eine Wartezeit von mehreren Wochen für einen (mühsam manuell erstellten) Befund einem »Digital Native« kaum mehr zu vermitteln.

Künstliche Intelligenz ermöglicht die Optimierung vieler Arbeitsschritte durch Assistenzsysteme, die dem Zahnarzt die Freiheit geben, sich auf die notwendige handwerkliche Kunst am Patienten zu fokussieren.

5.2 Fallbeispiele

CellmatiQ befasst sich mit der Auswertung von medizinischen Bildern (2/3/4D; alle Modalitäten = Röntgen, DVT, MRT, Foto, ...) mithilfe von Künstlicher Intelligenz. Massive künstliche Neuronale Netze in der Größenordnung von 200 Millionen Neuronen (= numerisch ca. 1/400 des menschli-

chen Gehirns) simulieren das Sehvermögen des visuellen Cortex und können zuverlässig Strukturen und Muster in Bildern identifizieren.

Zusammen mit interessierten Fachärzten hat CellmatiQ initial ca. 20 verschiedene dentale Bereiche bzw. Business Cases identifiziert, in denen KI modernen Zahnärzten Vorteile in einem der o.a. Bereiche liefern wird.

5.2.1 Kephalometrische Analyse

Die kephalometrische Analyse ist das Grundwerkzeug eines Kieferorthopäden oder eines kieferorthopädisch arbeitenden Zahnarztes. In den meisten Ländern ist das Verfahren zu Beginn jeder kieferorthopädischen Behandlung (z.B. für eine Zahnspange) gesetzlich verpflichtend. Dazu müssen in einem speziellen seitlichen Röntgenbild (sog. »FRS«) Landmarken in Knochen und Weichteilen identifiziert werden. Diese werden dann je nach Indikation mit Linien verbunden, aus denen diagnostisch relevante Winkel, Strecken und Relationen berechnet werden. Diese wiederum liefern Aufschluss darüber, inwieweit die Anatomie des Patienten medizinische Normwerte verletzt (z.B. ein Unterbiss), die potenziell eine Therapie erfordern. Die manuelle Auswertung einer Kephalometrie erfordert ca. 10 bis 15 Minuten Aufwand. Das Verfahren ist global weitestgehend standardisiert und wird jährlich weltweit ca. 10 Millionen Mal durchgeführt. Das immense Effizienzpotenzial durch Automatisierung ist offensichtlich.

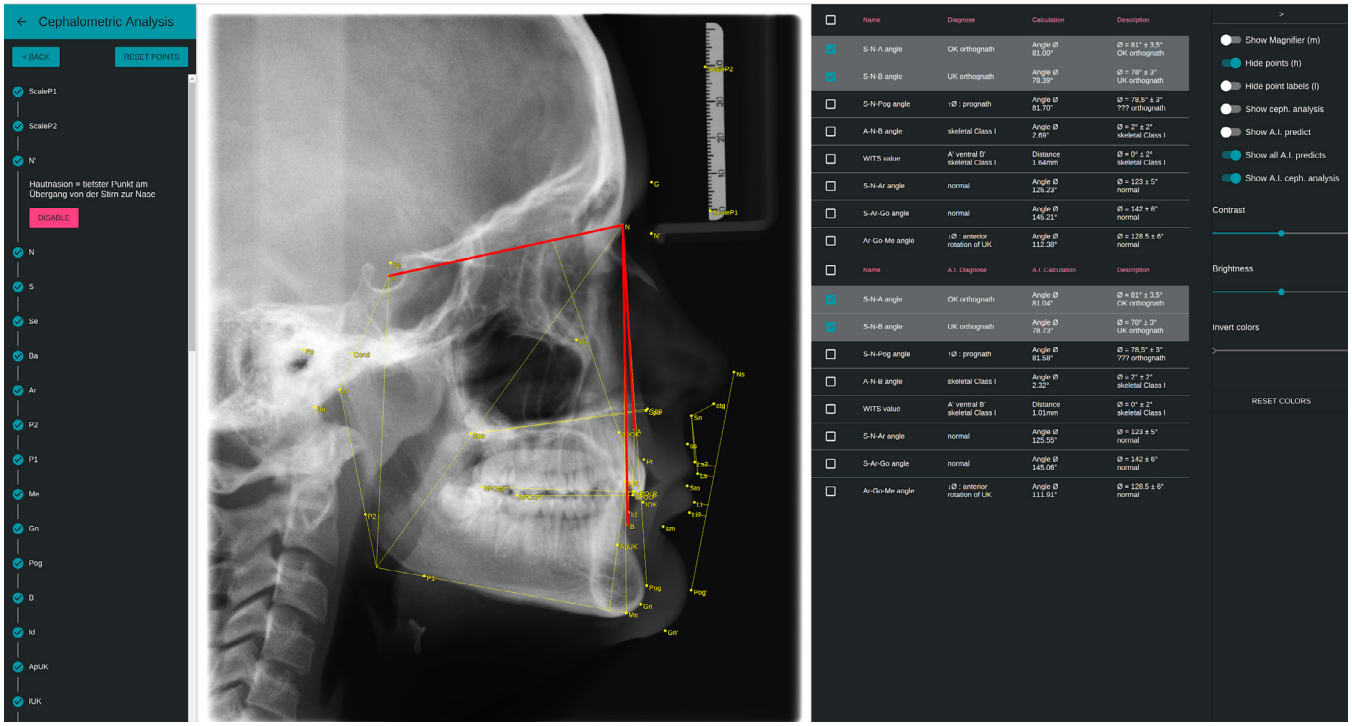


Abbildung 1: DentaliQ ortho: Automatische kephalometrische Analyse mit KI

Die Schwierigkeit der automatisierten Auswertung besteht darin, dass verschiedene Röntgengeräte bei verschiedenen Einstellungen äußerst verschiedene Bilder erzeugen können. Zudem ist die menschliche Anatomie, insbesondere bei größeren Anomalien, sehr heterogen. Die Herausforderung für die KI besteht darin, trotzdem zuverlässig die Landmarken mit einer Präzision deutlich unter 1 mm zu identifizieren.

Abbildung 2: Zuverlässige KI-Analysen auch bei heterogenen Belichtungen und Anatomien

Während ein Kieferorthopäde für eine spezifische Analyse jeweils nur ca. 20 Landmarken verwendet, ist die Obermenge aller Landmarken, die global verwendet werden können, um alle gängigen Verfahren abzudecken, 42.

Die CellmatiQ Software »DentaliQ ortho« identifiziert diese 42 Landmarken aus einem Röntgenbild mit einer Präzision, die vergleichbar ist mit erfahrenen Fachärzten. Damit kann eine vollständige kephalometrische Analyse bei konsistenter Qualität unter einer Sekunde durchgeführt werden – ein Zeitgewinn, der Zahnärzten Freiheiten für z.B. intensiveren Patientendialog und eine sofortige Problemindikation ermöglicht.

5.2.2 Karies-Identifikation und -Klassifizierung

Es muss nicht immer der Bohrer sein! Abhängig von ihrer Eindringtiefe lässt sich eine Kariesläsion auch nicht-invasiv, z.B. durch chemische Versiegelung und Re-Mineralisierung, reparieren. Leider ist das Auffinden von Kariesläsionen und die Bestimmung ihrer Tiefe (nach dem 5-stufigen ICDAS-Schema) extrem schwierig. Gemäß der Literatur erkennen Zahnärzte nur ca. 30 bis 50% der Läsionen korrekt.

»DentaliQ caries« wird Zahnärzte hierbei entlasten und dem Patienten in den sinnvollen Fällen eine Therapie ohne Spritze und Bohrer ermöglichen. Mit diesem Verfahren werden parallel drei Interessengruppen befriedigt: Der Patient erhält eine qualitativ bessere Behandlung bei geringerem Unbehagen, die Zahnärzte erhöhen ihre Kundenbindung und -attraktivität durch moderne Verfahren, und die dentale Chemieindustrie kann ihre Verkäufe erhöhen.

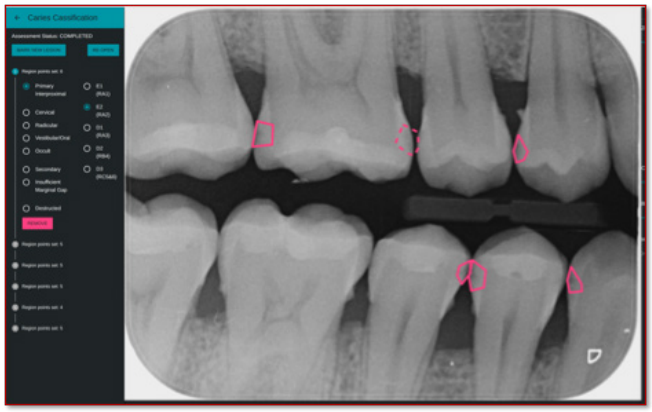


Abbildung 3: KI Karies-Klassifikation: Stufen E1, E2, D1 lassen sich chemisch reparieren

5.3 Ausblick

Nicht nur die o.a. Fallbeispiele demonstrieren, dass KI in der medizinischen Bilddiagnostik einen Paradigmenwechsel hin zur automatisierten Unterstützung des Arztes ermöglicht. Die Philosophie von CellmatiQ ist es explizit nicht, zu versuchen Ärzte zu ersetzen. Dies ist weder aus technologischen Gründen zurzeit machbar, noch ist es aus Sicht der Patienten erstrebenswert. Stattdessen ist es das Ziel von CellmatiQ, bei hochvolumigen Diagnoseprozessen eng begrenzte Arbeitsschritte zu identifizieren, die für den Arzt mühsam, langwierig oder fehlerträchtig sind und in denen die KI-basierte Analyse assistieren kann. Wir betonen das Wort »Analyse« für die KI – eine Diagnose überlassen wir bewusst den Ärzten, denn hierzu ist typischerweise die Auswertung eines Bildes nicht hinreichend, wohl aber wertvoll.

In mehreren medizinischen Bereichen wurden bereits solche Prozesse identifiziert und CellmatiQ arbeitet mit den jeweiligen Experten zusammen, um systematisch Daten zu sammeln sowie KI-Verfahren zu entwickeln und medizinisch zu validieren.

Darüber hinaus sind alle diese Verfahren problemlos auf nicht-medizinische Anwendungsfelder skalierbar. Z.B. kann die Software zur Karies-Identifikation nahezu unverändert eingesetzt werden, um Qualitätsmängel in Produktionsprozessen (z.B. Farbfehler) oder technischen Anlagen (z.B. Rost) zu erkennen. Die Technologie zur Vermessung von Landmarken in Bildern kann gleichermaßen z.B. für röntgenbasierte Qualitätssicherung technischer Anlagen oder die Vermessung von Bauleistungen transferiert werden.

5.4 Abbildungsverzeichnis

6 Postbank Goes Big Data: Wie die Postbank mit Big Data ihre Prozesse optimieren und automatisieren konnte

6 Postbank Goes Big Data: Wie die Postbank mit Big Data ihre Prozesse optimieren und automatisieren konnte

Lars Fockele & Sharif Abdel-Halim

6.1 Einleitung

Vor dem Einsatz von Big-Data-Technologien bestand die Datenlandschaft der Postbank aus vielen verschiedenen »Datentöpfen«, welche nicht in der Lage waren, große, polystrukturierte Daten zu prozessieren und somit Big-Data-Anwendungsfälle abzudecken.

Durch die Einführung eines gemeinsamen Datalakes, in den Daten verschiedener Quellen einfließen, und durch den Einsatz von Big-Data-Technologien konnte die Postbank einige ihrer Kernprozesse optimieren und automatisieren. Eine dieser Optimierungen entstand durch die Einführung einer zentralen Auswertungsplattform im Rahmen der DSGVO, womit es den Datenschutzbeauftragten möglich gemacht wurde, über ein zentrales System alle protokollierten Zugriffe auf Kundendaten durch Postbank-Mitarbeiter auswerten zu können. Da bei der Postbank verschiedene Systeme existieren, auf denen Kundenzugriffe erfolgen, müsste der Datenschutz ohne eine solche zentrale Auswertungsplattform jedes System einzeln einsehen, was die Arbeit deutlich erschwert.

6.2 Postbank-Datenplattform

Die Postbank hat 2015 das erste Hadoop-basierte Big-Data-System produktiv gestellt. Dies geschah in Form eines »Analytics Workplace« (AWP), welcher den Fachseiten zum einen die benötigten Daten, zum anderen die technologischen Möglichkeiten in Form von Big-Data- und Analytics-Tools für Datenanalysen zur Verfügung stellt.

Darauf folgte der Ausbau einer kompletten Datenplattform, die folgende Grundanforderungen erfüllen sollte:

- Verarbeitung von Daten sowohl im Batch- wie auch im Streaming-Modus.
- Bereitstellung der Daten für manuelle Analysen der Data Scientists sowie für fest implementierte, operative Use Cases, auch in Zusammenarbeit mit weiteren Umsystemen außerhalb der Datenplattform.

Funktionale Architektur

Die zugrunde liegende funktionale Architektur gliedert sich in sieben Punkte, wobei die Quellsysteme außerhalb der Verantwortlichkeit der eigentlichen Datenplattform liegen. Den Kern bilden die zentralen Persistenz- und Distributionssysteme Datalake und Data Distribution, welche über die Source Integration mit Quelldaten versorgt werden. Die Entwicklung und Produktivsetzung von Data Driven Products (DDPs), also den Umsetzungen fachlicher Anforderungen, sowie die Analysen auf dem Analytics Workplace können dezentral durch unabhängige Teams erfolgen. Die Data Governance erfolgt übergreifend über sämtliche Teilbereiche der Datenplattform.

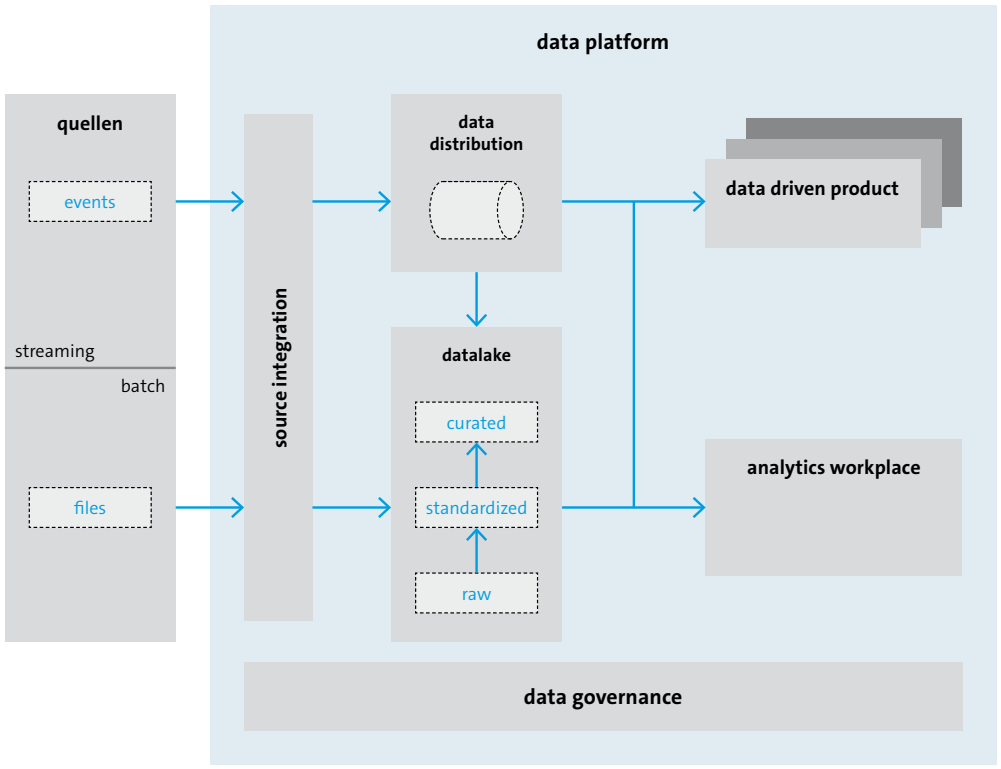


Abbildung 1: Funktionale Architektur Datenplattform

Bereiche der Datenplattform

Quellsysteme: Aktuell beliefern hauptsächlich SAP-Kernbankensysteme die Datenplattform, welche die Fakten dem Source-Integration-Layer in dateiform zur Verfügung stellen. Es werden aber auch Daten über eine Online-Schnittstelle in »Near Realtime« in die Data Distribution übergeben.

Source Integration: Sorgt für den Datentransport in den Datalake beziehungsweise in die Streaming-Plattform.

Datalake: Unterteilt in drei Bereiche speichert der Data Lake die Quelldaten:

- absolut unverändert in der RAW-Zone
- normalisiert und schematisiert in der Standardized-Zone
- (optional) vorverarbeitet in der Curated-Zone, falls mehrere DDPs eine gemeinsame, spezielle Sicht auf die Daten benötigen

Es gibt keine direkten Zugriffe auf die Inhalte des Datalakes durch Benutzer. Falls benötigt, werden Datenauszüge nach erfolgreichem Freigabeprozess in die DDPs oder den AWP exportiert.

Data Distribution: Asynchrone Near-Realtime-Versorgung der DDPs oder des AWPs.

Data-Driven-Projects: Fachliche Umsetzungen von Use Cases innerhalb der Datenplattform. Die dadurch gewonnenen Erkenntnisse können per Serviceschnittstellen an Umsysteme (wie zum Beispiel das Onlinebanking) verteilt werden.

Analytics Workplace: Eigenständiges Big-Data-Cluster mit zusätzlichen Analytics- und Machine-Learning-Lösungen für die Data Scientists der Fachseiten. Der AWP wird durch den Datalake oder Umsysteme mit benötigten Daten versorgt.

Data Governance: Ein einheitliches Management der Daten über alle Teilbereiche der Big-Data-Plattform hinweg, zum Erhalt der Datenqualität, der Datensicherheit und des Datenschutzes.

6.3 Anwendungsfall DSGVO – Zentrale Auswertungsplattform

Problemstellung

Ein Bestandteil der im Mai 2018 in Kraft getretenen Datenschutz-Grundverordnung (DSGVO) ist, dass Kunden ein Auskunftsrecht besitzen. Mit diesem Recht können sie die Information einfordern, welche Daten ein Unternehmen über sie speichert, sowie wann und wie auf diese Daten durch Mitarbeiter zugegriffen wurde.

Auf die Postbank bezogen wäre folgendes konkrete Szenario denkbar:

Ein Kunde lässt sich durch einen Kundenberater in einer Filiale beraten. Dabei speichert der Berater personenbezogene Daten des Kunden oder greift auf diese zu. Dieser Zugriff wird in einem Informationssystem protokolliert.

Der Kunde besitzt das Recht, eine Beschwerde einzureichen, beispielsweise wenn er einen unerlaubten Zugriff auf seine Daten vermutet. Er kann einen Anwalt hinzuziehen, der Fall wird anschließend an die Bank gemeldet, wo er durch den Datenschutz geprüft wird.

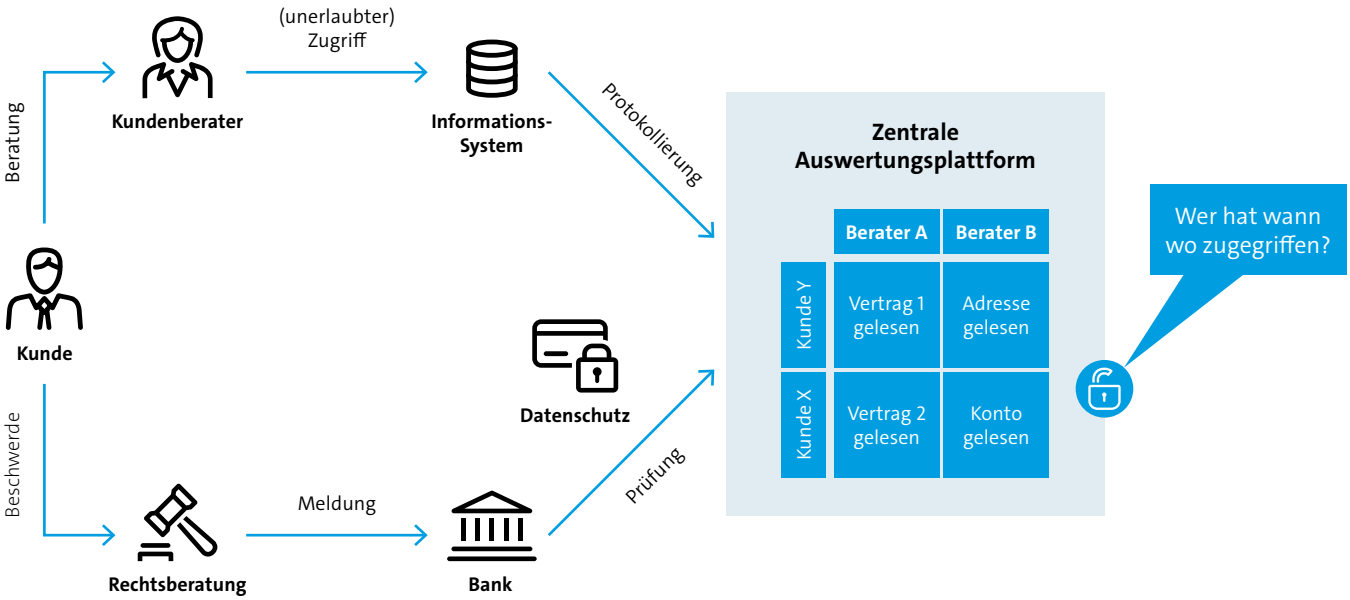


Abbildung 2: Anwendungsfall Zentrale Auswertungsplattform

Das Problem, das hierbei vorliegt, ist, dass bei der Postbank nicht nur ein Informationssystem existiert, in dem Kundenzugriffe erfolgen, sondern viele verschiedene. So müsste der Datenschutzbeauftragte in jedem System einzeln nachsehen, in welcher Form Kundenzugriffe stattgefunden haben. Um ihm die Arbeit zu erleichtern wurde eine zentrale Auswertungsplattform aufgebaut, in der Kundenzugriffe mehrerer Informationssysteme zusammengeführt werden. Auf diese Weise wurde dem Datenschutz eine Plattform zur Verfügung gestellt, auf der er zentral die Zugriffe auf Kundendaten einsehen kann.

Lösung

Es gibt mit der BHW und der Postbank zwei unterschiedliche Mandanten, die jeweils SAP- wie auch Non-SAP-Systeme für die Datenverarbeitung nutzen. Innerhalb dieser sind Zugriffe auf Kundendaten möglich, welche dann durch die einzelnen Systeme automatisch protokolliert werden.

Diese Read-Access-Logs werden in unterschiedlichen Frequenzen und Strukturen in die Landingzone der Datenplattform exportiert. Das Landing-Verzeichnis ist ein NFS-Dateisystem, das als Übergabeverzeichnis dient, wenn Postbank Systeme Daten in den Datalake schreiben wollen. Der Edge Node ist hierbei die Schnittstelle zwischen dem Datalake und den Umsystemen.

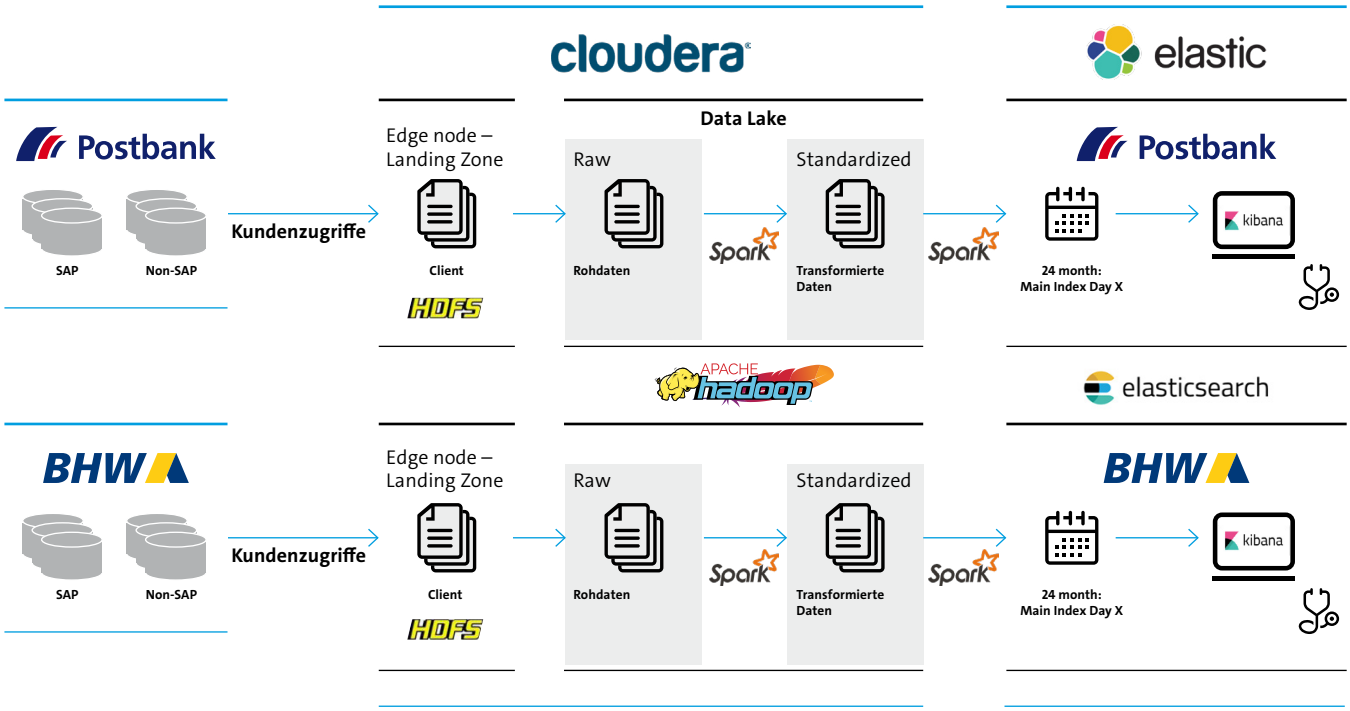


Abbildung 3: Technische Architektur der zentralen Auswertungsplattform

Auf dem Edge Node läuft ein HDFS-Client, der die angelieferten Dateien in den Datalake kopiert, als erstes in die Raw-Zone. Mit Apache Spark finden dann syntaktische Transformationen statt und die Daten werden in die Standardized-Zone geschrieben, in der syntaktisch aufbereitete Daten vorliegen. Syntaktische Transformationen beinhalten beispielsweise das Raustrimmen von Leerzeichen und die Vereinheitlichung von Datumsfeldern.

Von dort aus werden die Daten über Apache Spark in Elasticsearch gespeichert, wo es tagesbasierte Indizes gibt und die Daten 24 Monate vorgehalten werden.

Über Kibana Frontends können die Datenschützer die Datenzugriffe einsehen und auswerten.

6.4 Ausblick

Die Hauptaufgabe der nächsten Zeit wird es sein, die Migration der beiden Big-Data-Plattformen, die in der Deutschen Bank sowie in der Postbank entstanden sind, zu konsolidieren und durchzuführen. Ob die Plattform dann weiterhin On-Premise oder in der Cloud verortet wird, bedarf weiterer Abstimmungen.

Auch in den einzelnen Teilbereichen der Plattform gibt es noch Optimierungsbedarf. So ist ebenso ein Ausbau der Realtime-Fähigkeit der Quellsysteme geplant wie die automatisierte,

toolgestützte Erstellung neuer Schnittstellen für neue Quelldatentypen. Auch die Verarbeitung der Daten innerhalb der einzelnen Datalake-Zonen, wie oben beschrieben, muss noch weiter ausgebaut werden.

Für den Bereich Machine und Deep Learning wird auch die aktuelle Infrastruktur der Plattform überdacht, um durch den Einsatz spezieller GPUs und entsprechendem Hauptspeicher die Performance der ML-Modelle zu steigern.

Ein potenzieller Anwendungsfall für Machine Learning ist im Kontext der zentralen RAL-Auswertungsplattform zu finden: Anomalien in den Kundenzugriffen automatisiert erkennen zu lassen, um so frühzeitig Missbrauch zu registrieren und weiteren Missbrauch präventiv zu verhindern.

6.5 Abbildungsverzeichnis

7 KI-Strategie: Die Entwicklung einer Analytical Roadmap in der Stahlbranche

7 KI-Strategie: Die Entwicklung einer Analytical Roadmap in der Stahlbranche

Caroline Kleist & Friedhelm Bösche

7.1 Einleitung

Artificial Intelligence, Machine Learning, Deep Learning – alles Schlagworte, die aktuell omnipräsent zu sein scheinen. Neben dem akademischen Diskurs rund um die Frage, welcher Begriff nun welche Methoden genau umfasst, geht es in der Praxis aktuell vor allem um die Frage: Wie können wir diese innovativen Methoden für uns als Unternehmen sinnvoll einsetzen? Während sich Unternehmen aus der Banken- und der Versicherungsbranche schon länger der Beantwortung dieser Frage widmen, agieren vergleichsweise konservative Branchen, wie die Stahlindustrie, noch zögerlich. Dadurch, dass in diversen Bereichen und Märkten bereits aktiv Machine Learning Use Cases erfolgsbringend eingesetzt werden, begibt man sich jedoch auch hier allmählich auf die Reise. Doch was sind die ersten Schritte? Das Spektrum an Umsetzungsansätzen scheint unendlich und unüberschaubar: Wie viele Data Scientists braucht man? In welchen Geschäftsfeldern gibt es Anwendungsfälle? Wie definiere ich einen Use Case? Geben unsere Daten solche Analysen überhaupt her? Was für eine Data-Science-Plattform benötigen wir?

Als Data-Science-Beratung begleiten wir viele Unternehmen bei der Beantwortung dieser Fragen. Für die erfahrene Navigation durch die vielzähligen möglichen Antworten hat sich die Entwicklung unserer Analytical Roadmap sehr bewährt. So haben wir im Rahmen unserer Partnerschaft mit der Küttner Automation GmbH, die sich auf die Prozessoptimierung in der Eisen- und Stahlindustrie spezialisiert hat, gemeinsam eine Analytical Roadmap für die Stahlbranche entwickelt, an deren Beispiel im folgenden Beitrag kurz das grundlegende Konzept und die konkrete Herangehensweise bei der Erstellung einer Analytical Roadmap erläutert wird.

7.2 Innovative Methoden in einer traditionsreichen Branche

Die Stahlbranche hat als Basisindustrie eine besonders gewichtige Bedeutung für die (deutsche) Wertschöpfungskette. So ist sie nicht nur wichtiger Zulieferer für den Maschinenbau und die Automobilindustrie, sondern selbst auch Abnehmer für zahlreiche Zulieferbranchen. Vielleicht gerade wegen Ihrer Bedeutung und Erfolgsgeschichte bewegt sie sich in ihrer Gewichtigkeit vergleichsweise jedoch ähnlich schwerfällig wie ihr produziertes Gut, sodass innovative Methoden und Ansätze, wie die Nutzung von Machine Learning, nur sehr langsam Anklang finden. Dieses Phänomen ist nicht ausschließliches Merkmal der Stahlindustrie, sie steht hier jedoch exemplarisch für die produzierenden Mittelständler in Zentraleuropa, die eine eher abwartende und beobachtende Haltung einnehmen, wenn es um die Anwendung von Machine Learning geht. Viele Skeptiker hinterfragen die Sinnhaftigkeit und den Mehrwert von Machine Learning in

ihrem eigenen Umfeld oder empfinden die Thematik als zu abstrakt. Hier sehen wir es als unsere Aufgabe und Mission, Kommunikationsarbeit zu leisten und die Begriffe zu entmystifizieren. Denn, dass die Arbeit mit bzw. die Verwertung von Daten die Zukunft maßgeblich prägen wird, ist unumstritten. Als Unternehmen muss man dieses abstrakte Vorhaben daher zwingend angehen, um den Anschluss hier nicht zu verpassen und das Wertschöpfungspotenzial möglichst früh auszuschöpfen. Unsere Analytical Roadmap dient herbei als Kompass, um auf Basis einer individuellen Ausgangslage den besten Weg zu einer nachhaltigen Data-Science-Infrastruktur zu finden. Neben der Identifikation des günstigsten Startpunkts geht es dabei vor allem darum, individuelle Anforderungen und Potenziale zu identifizieren, um eine unternehmensspezifische Lösung zu finden. Wenngleich es wichtig ist, vorausschauend zu planen, kommt es in einem ersten Schritt darauf an, die Einstiegshürden möglichst niedrig zu halten und durch vergleichsweise einfach implementierbare Use Cases Akzeptanz und Vertrauen im Mitarbeiterkreis zu schaffen. Dies ist insbesondere in Branchen wie der Stahlbranche wichtig, um nicht zu brachiale Veränderungen zu forcieren, denn eben deren skeptische Mitarbeiter sind die zwingend notwendigen Fachexperten und zukünftigen Anwender und Entwickler der Machine-Learning-Lösungen.

Zusammenfassend gilt es daher, dieser Skepsis mit einer klar strukturierten Strategie zu begegnen, die mit schlanken und zügigen Implementierungen aufwartet, zukünftige Skalierbarkeit sichert und die Mitarbeiter und Fachexperten abholt und mitnimmt.

7.3 Herausforderungen bei der (erstmaligen) Implementierung von KI

Die größte Herausforderung ist, wie in dem vorhergehenden Abschnitt beschrieben, sicherlich die Unternehmenskultur und das traditionell wenig agile Umfeld. Hierzu zählen insbesondere auch Skepsis oder übersteigerte Erwartungen an Machine-Learning-Lösungen. Als Treiber in der Nutzung von Machine Learning ist man somit zumeist mit einer unzutreffenden Erwartungshaltung konfrontiert und muss beständiges Erwartungsmanagement betreiben. Hierfür sollte man einen klar formulierten und transparenten Plan verfolgen und nicht in den Verdacht kommen, dass »man halt mal probiert«. Insbesondere erfahrene Metrologen, die sich seit jeher mit den physikalischen Abläufen und Zusammenhängen in Hochöfen beschäftigen, stehen den neuen Ansätzen skeptisch gegenüber und dieser Skepsis begegnet man am effektivsten mit transparenter Kommunikation und Kollaboration.

Analog zur Bedeutung des Benzins für den Verbrennungsmotor ist das Vorhandensein einer soliden Datengrundlage unverzichtbar für die Erarbeitung datengetriebener Machine Learning Use Cases. Die Verfügbarkeit und Eignung der bestehenden Datenbasis stellt sich hierbei als erstes K.O.-Kriterium dar. Oftmals ist es daher zunächst Aufgabe, diese zu überprüfen und vor dem Hintergrund der späteren Anwendung zu bewerten. Hierzu müssen dreierlei fachliche Brücken geschlagen werden: (1) aus den fachlichen Prozessen in deren datenseitige Abbildung, (2) aus den Daten in die mögliche Machine-Learning-Anwendung und (3) vom Machine Learning

wiederum zurück in die fachlichen Prozesse. Die häufigste Herausforderung bildet hierbei die Übersetzungsleistung an den Brückenköpfen und das interne Know-how über die Datengrundlage. Rollen, die das Konzept der Data Ownership manifestieren, sind zumeist noch nicht etabliert. Nicht selten endet eine entsprechende Analyse in der Erkenntnis, dass die Anwendung von Machine Learning auf Basis der vorhandenen Datenbasis aktuell nicht umsetzbar und daher zunächst eine Strategie zur Erfassung der benötigten Daten erforderlich ist. Da der Erfolg von Machine Learning unmittelbar von der Historie bzw. der Größe der vorhandenen Datenbasis abhängt, besteht die Gefahr, den Zug zu verpassen, wenn die benötigten Weichen nicht rechtzeitig gestellt werden.

Die Komplexität zwischen Methodik, Daten und Fachlichkeit und somit die interdisziplinären Anforderungen lassen Machine-Learning-Projekte zu einem sehr vielschichtigen Unterfangen werden, was es wiederum schwer macht, einen sinnvollen Startpunkt zu identifizieren. Beginnt man mit der Plattform, um arbeitsfähig zu sein? Dies ist bei Initiativen, die aus der IT getrieben werden, am häufigsten der erste Gedanke. Doch die Anforderungen an die Plattform sollten unserer Erfahrung nach von den zu bearbeitenden Use Cases abgeleitet werden und die Use Cases wiederum einem zu lösenden fachlichen Problem entspringen. Nur so kann nachhaltig Wert aus den bzw. Vertrauen in die Anwendung von Machine Learning geschaffen werden.

7.4 Use Cases als Dreh- und Angelpunkt der Analytical Roadmap

Um nicht als »nette«, aber erfolglose Innovationsinitiative im Unternehmensarchiv zu enden, muss von Anfang an der Zug zu einem messbaren Ziel vorhanden sein. Dieses Ziel kann beispielsweise in einer Prozessoptimierung und/oder Kostensenkung liegen. Die Machine-Learning-Anwendung muss und kann sich daran messen lassen. Dies beweisen die mannigfaltigen Anwendungen, die schon heute in anderen Branchen erfolgreich produktiv gesetzt sind. Und um messbare Ziele auszumachen, sind klar rechenbare Business Cases das Fundament.

Zur Identifikation von Use Cases führen wir mit allen relevanten Fachabteilungen, wie beispielsweise Prozessingenieuren, Metallogen und der IT Ideation Sessions durch. Hierbei werden zunächst reale Projektbeispiele zur Inspiration erläutert und mannigfaltige Beispiele gezeigt. Im Anschluss erhält jeder Teilnehmer rund 20 Minuten Zeit, um eigene Use-Case-Ideen zu entwickeln. Die Machbarkeit ist dabei zunächst zweitrangig, da es sich hier um eine kreative Phase handelt. Dies ist der erste Kontakt mit Machine Learning, bei welchem es aufs Mitmachen ankommt. Typische Fragestellungen, die die Kreativität anregen sollen, können sein:

- Smart Decisions: Gibt es Aufgaben oder Entscheidungen, bei denen ein Blick in die Zukunft nützlich wäre?

- Robotic Process Automation: Gibt es langatmige, monotone Aufgaben, die automatisiert werden können?

- Leverage Data: Gibt es Daten, die gesammelt werden und aus Ihrer Sicht noch nicht effizient genutzt werden?

Die Ideen werden gesammelt und geclustert und als Use Case Long List zusammengetragen.

Die folgende Abbildung zeigt eine exemplarische Auswahl von Use Cases der gemeinsam mit der Küttner Automation GmbH für die Stahlindustrie erstellten Long List, die allesamt auf den kreativen Ideen der Küttner-Fachexperten beruhen und auf Basis unseres Feedbacks weiterentwickelt und konkretisiert wurden.

Smart Decisions	Robotic Process Automation	Leverage Data
Formstoffverdichtbarkeit	Steuerung und Automatisierung Mischerei	Sales Forecasts
Prognose Produzierte Stahlqualität	Steuerung und Automatisierung Kupolofen	Abgasanalyse
Lastmanagement		Verschleißerkennung
Predictive Maintenance Formanlage		
Computer Vision Formkastenkontrolle		

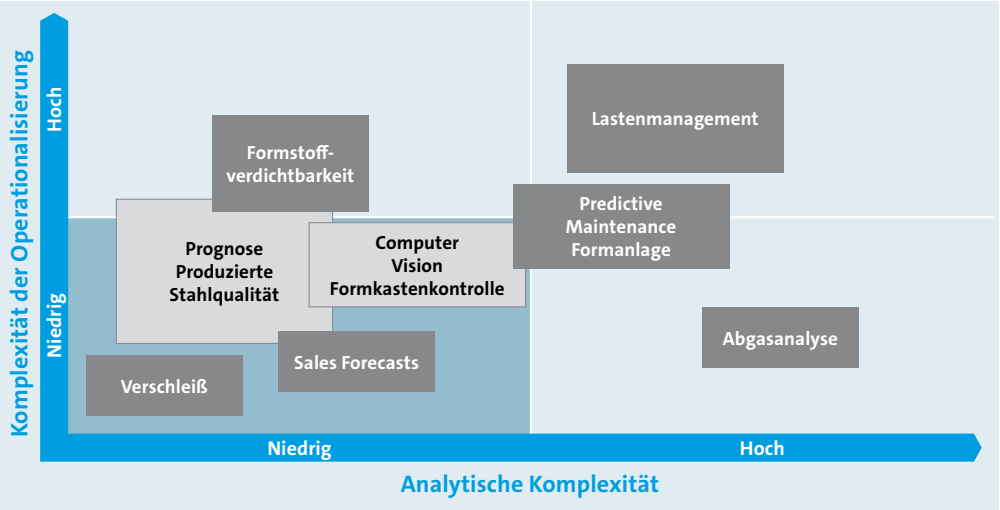
Im Anschluss an die Erstellung der Long List werden die einzelnen Use Cases auf Ihre Machbarkeit überprüft. Dabei werden jegliche benötigten und verfügbaren Datenkategorien zusammengetragen und letztere hinsichtlich ihrer Eignung für Machine-Learning-Anwendungen bewertet. Neben der Nicht-Erfassung elementarer Daten(kategorien) als K.O.-Kriterium werden Aspekte wie die Historisierung, Konsistenz bei der Erfassung und nicht zuletzt der Umfang der Daten berücksichtigt. Am Ende dieses Readiness-Checkups steht eine klare Bewertung, welcher Use Case kurz-, mittel- oder langfristig mit welchen Daten umgesetzt werden kann. Sofern für einen Use Case zwar grundsätzlich Daten verfügbar, diese jedoch lückenhaft sind, kann es vor dem Hintergrund einer Wirtschaftlichkeitsbeurteilung durchaus Sinn machen, die fehlenden Daten nachzuerheben und zu persistieren, um den Use Case mittelfristig umsetzen zu können. Bei Kunden der Küttner Automation GmbH ist dieses Vorgehen unter anderem bei dem Use Case »Steuerung und Automatisierung Kupolofen« zielführend.

Ergebnis dieser Phase ist eine Short List direkt umsetz- und implementierbarer Use Cases.

7.5 KI meistern: Entwicklung der Analytical Roadmap

In einem nächsten Schritt werden sämtliche Use Cases der Short List anhand von Wirtschaftlichkeitsbetrachtungen und des erwarteten Business Values bewertet. Zusätzlich fließen Aspekte der Skalierbarkeit, Übertragbarkeit auf Werke und Aufwand für die Operationalisierung ein. Abschließend werden alle Use Cases transparent anhand ihrer analytischen Komplexität und ihres Operationalisierungsaufwands bewertet, kategorisiert und somit priorisiert.

In der folgenden Abbildung wird der erwartete Business Value über die Größe der Rechtecke abgebildet. Für die gemeinsam mit der Küttner Automation GmbH erstellte Roadmap für die Stahlbranche ergibt sich auf Basis der für den Use Case spezifischen Bewertungen somit folgende Darstellung:



Das Ziel der Analytical Roadmap liegt zunächst in der Identifikation der »am niedrigsten hängenden Früchte« mit dem größten erwarteten Nutzen. Bezogen auf die Analytical Roadmap für die Stahlindustrie (der obigen Abbildung) trifft dies auf die Use Cases »Prognose Produzierte Stahlqualität« und »Computer Vision Formkastenkontrolle« zu, welche entsprechend als erste in die Umsetzung und Implementierung gehen, um als »Leuchtturmprojekte« mit einem hohen ROI das Vertrauen und die Zustimmung der Skeptiker einzuholen.

7.6 Schrittweise Umsetzung und Implementierung der Analytical Roadmap

Ausgehend von der Priorisierung und Planung der Use Cases ergeben sich die Anforderungen an die Architektur, Infrastruktur und an das Team mit den zu besetzenden Rollen und Organisationsstrukturen. Wir empfehlen diese Reihenfolge in der Herangehensweise, obwohl man zumeist eine andere in der Praxis antrifft: zuerst Infrastruktur und dann Use Cases. Doch insbesondere für die spätere Skalierbarkeit über mehrere Märkte hinweg und die Anforderungen an Rechenpower durch Computer-Vision-Projekte gibt es keinen Standard-Toolstack. Daher sollte von Beginn an die spätere Anwendung ebendieser in Form von Use Cases in den Fokus rücken.

Gemeinsam mit der Küttner Automation GmbH werden die Use Cases, »Prognose Produzierte Stahlqualität«, »Formstoffverdichtbarkeit« und »Computer Vision Formkastenkontrolle«, aktuell iterativ und nach klassischen Fail-Fast-Ansätzen umgesetzt. Die Anforderungen für das Team resultieren ebenso aus der Analytical Roadmap. Da die Skepsis aktuell noch überwiegt, sind kommunikationsstarke Mitarbeiter von hoher Bedeutung. Im gemeinsamen Projektteam mit der Küttner Automation GmbH arbeitet aktuell ein vergleichsweise kleines Machine-Learning-Team an den zwei Use Cases, bestehend aus zwei Data Scientists, einem Data Engineer und einem Softwareentwickler für Automatisierung. Das Arbeiten in interdisziplinären Use-Case-Teams schafft zudem Vertrauen und hilft bei der Entmystifizierung von Machine-Learning-Projekten mit handfesten Methoden und Ergebnissen. So wurde der Use Case »Prognose Produzierte Stahlqualität« angesichts der sehr guten Vorhersagegenauigkeit bereits vor einem halben Jahr produktiv gesetzt.

7.7 Fazit

Die zu treffenden Entscheidungen auf dem Weg hin zu datengetriebenen Lösungen durch den Einsatz von Machine Learning scheinen unüberschaubar und unzählbar zu sein. Und zurecht: durch die hohe Interdisziplinarität und Komplexität gibt es verschiedene Möglichkeiten, in die Thematik einzutauchen. Um nicht in den Verdacht zu kommen, »Jugend forscht« zu spielen, wird ein klar strukturierter und strategischer Plan benötigt: eine Analytical Roadmap. Ausgehend von nutzbringenden Use Cases ergeben sich Anforderungen an die Organisation, Ressourcen und die Architektur. Insbesondere in einer konservativen Branche, wie der Stahlbranche, ist stetige Kommunikationsarbeit und die Entmystifizierung von Machine Learning unumgänglich. Zusammenfassend gilt es, der Skepsis mit einem klar strukturierten und strategischen Plan zu begegnen, der mit schlanken und zügigen Implementierungen aufwartet, zukünftige Skalierbarkeit sichert und die Mitarbeiter und Fachexperten abholt und mitnimmt.

8 Personenstromanalyse im Bahnhof

8 Personenstromanalyse im Bahnhof

Ibrahima Kourouma & Oliver Brandmüller

8.1 Einleitung

Mit der Personenstromanalyse erhofft sich die Deutsche Bahn (DB) einen tieferen Einblick in die Vorgänge in einem Bahnhof zu erlangen und Entscheidungsprozesse rund um den Bahnhof zu unterstützen. Im Kontext der Datenanalyse wird vor allem auf die Extrapolation des Personenaufkommens zurückgegriffen. Dabei konnten vorerst zwei Anwendungsfälle identifiziert werden.

Die Ergebnisse der Analyse wollen zum einen genutzt werden, um im Fall eines kleinen Bahnhofs den Personaleinsatz zu optimieren. Da in kleinen Bahnhöfen kein Bahnhofspersonal dauerhaft vor Ort ist, sollen zu den Zeitpunkten, an denen ein ungewöhnlich hohes Personenaufkommen erwartet wird, Einsätze, bei denen beispielsweise Reinigungen oder Mülleimerleerungen erfolgen, stattfinden.

Zum anderen soll In größeren Bahnhöfen versucht werden, Potenziale für die Vermarktung zu erkennen, indem sogenannte Hotspots, sprich Orte im Bahnhof, an den sich viele Personen aufhalten, identifiziert werden.

8.2 Datenquellen: Überblick über die Datenerhebung

Es existiert heute eine Vielzahl an Technologien zur Personenzählung, die auf unterschiedlichster technischer Basis in verschiedener situativer Umgebung Daten in ausreichender Qualität produzieren können. Weiterhin spielen ökonomische Fragen, was Installation und Betrieb angeht, wie auch zunehmend Datenschutzbelange eine Rolle bei der Auswahl einzelner oder auch komplementärer Techniken. Nicht zuletzt hängt die Auswahl auch von der Fragestellung des Datennutznießers ab. Mal reicht es, Trends und Ausnahmen zu erkennen, ein anderes Mal werden möglichst genaue absolute Zahlen benötigt. Ist es wichtig, die Aufenthaltsdauer, womöglich ein Bewegungsprofil zu erfassen oder reicht es aus, für einen Ort die aktuelle Anzahl anwesender Personen abschätzen zu können? Ein weiterer interessanter Aspekt ist die Echtzeitfähigkeit der Erhebung. Für einige Anwendungsfälle werden Daten im Fünf-Sekunden-Raster benötigt, andere wiederum begnügen sich mit einer Auswertung der Historie oder lassen eine Granularität bzw. Verzögerung mehrerer Minuten zu. Es gibt also am Ende nie eine einzig wahre Lösung für die Erhebung der Personenzahlen in öffentlichen oder halböffentlichen Bereichen.

Die gängigen Verfahren teilen sich im Wesentlichen auf nach der Zählung bzw. Abschätzung der Personenzahl im Umfeld des Sensors oder nach der Zählung der Zu- und Abgänge. Eine Umfelderhebung, basierend auf WLAN Daten oder der Auswertung von Videobildern, bietet sich in Bereichen an, wo die aktuelle Personenzahl in einer Menge interessant ist, die aber ggf. nicht scharf abgegrenzt ist. Beispiele sind Wartebereiche vor Informationsschaltern oder auf Bahnsteigen. Die Zugangszählung ist eine gute Methode, wenn es eine begrenzte und klar definierte Anzahl an Zugängen mit den passenden baulichen Strukturen gibt. Exemplarisch sind dies

Tunnelbauwerke oder Räume mit Zugängen durch Türen oder Pforten, aber auch Fahrtreppen oder Treppenzugänge.

Im folgenden Abschnitt erfolgt – ohne Anspruch auf Vollständigkeit – eine Vorstellung der bei der DB erprobten Systeme mit einer kurzen Beschreibung der jeweiligen Einsatzzwecke sowie von Vor- und Nachteilen.

8.3 Umfelderhebung

PaxCounter

Eine Eigenentwicklung der DB auf Basis von WLAN Daten. Gezählt werden Signale unterschiedlicher WLAN-Geräte im Umfeld. Dazu werden Datenpakete, die das Endgerät bei der Suche nach WLAN-Netzwerken sendet, aber auch Pakete bei bestehender Verbindung ausgewertet. Ein Zählzyklus umfasst üblicherweise einen Zeitraum zwischen ein und vier Minuten, dies ist auch die Granularität bzw. der Zeitversatz der Messung. Am Ende des Zählzyklus übermittelt das Gerät über einen schmalbandigen Kanal, z. B. LoRaWAN, eine Zahl der gemessenen Geräte. Die ermittelten MAC-Adressen der Geräte werden auf dem Sensor per Salt- und Hashverfahren anonymisiert, der Salt ändert sich zwischen den Zählvorgängen. Damit ist es weder möglich, die Aufenthaltsdauer einzelner Geräte (und damit Personen) im Messumfeld zu ermitteln, noch können mehrere der Geräte untereinander kommunizieren, um zum Beispiel Personen zu verfolgen. Damit erfüllt die Technik höchste Datenschutzstandards. Das Einsatzgebiet sind eher kleinere Umfelder ohne anderweitige Infrastrukturen, wie Kunden-WLAN oder umfassende Videoanlagen. Der PaxCounter ist aufgrund seiner Bauweise geeignet zum Einbau in vorhandene Geräte, wie zum Beispiel Uhren oder Beleuchtungsanlagen (»Retrofitting«). Die Geräte sind somit äußerst kostengünstig und durch die Verknüpfung mit anderen Anlagen oft auch relativ kostenneutral zu installieren.

WLAN-Infrastruktur

Die Messmethode ist hier ähnlich wie beim zuvor beschriebenen PaxCounter. Allerdings liefern WLAN-Infrastrukturen Daten, die über mehrere Zugangspunkte akkumuliert werden und teils sogar eine ungefähre Ortung erlauben. Das Einsatzgebiet besteht hiermit aus größeren Stationen, an denen dem Kunden eine WLAN-Infrastruktur geboten wird. Da auch Werte wie Aufenthaltsdauer oder teilweise sogar der Weg durch den abgedeckten Bereich ermittelt werden können, ist die Anonymisierung nicht mehr in dem Maße gewährleistet wie beim PaxCounter, der Effekt kann allerdings zum Beispiel durch die Zustimmung der Kunden zur Datenerhebung innerhalb des Anmeldeprozesses an der Infrastruktur mitigiert werden. Damit liefert die WLAN-Infrastruktur zum einen ähnliche Daten wie der PaxCounter, zum anderen, je nach Durchdringung bei der Anmeldung, aber eine Teilmenge der Daten, die um weitere Informationen angereichert sind. Zudem sind die eingesetzten Zugangspunkte mit wesentlich leistungsfähigeren Antennen und Funkeinheiten ausgestattet. Die Datenerhebung ist hier ein Seiteneffekt einer vorhandenen Infrastruktur mit nur geringen Mehrkosten.

Videoanalyse

Die Analyse von Videobildern erlaubt in vielen Fällen recht genaue Daten, setzt aber dazu auf vergleichsweise teure Infrastruktur im Hintergrund. Weiterhin sind sowohl Aspekte hinsichtlich der Datenschutzbestimmungen, selbst wenn die Zählung in den Endgeräten (Datenauswertung in der Kamera) erfolgt, und auf Aspekte der öffentlichen Wahrnehmung beim Einsatz von automatisierter Auswertung von Videobildern in Betracht zu ziehen. Videoanlagen mit geeigneten Blickwinkeln und ausreichender Bildqualität existieren in der Regel nur in größeren Bahnhöfen. Somit ist der Einsatz dieser Variante oft nur in begrenztem Rahmen möglich.

8.4 Zugangszählung

Stereokameras

Stereokameras erfassen, oft von oberhalb eines Türrahmens oder ähnlicher Durchgänge, den engen Erfassungsbereich durchlaufende Personen. Da eine Erkennung der Richtung möglich ist, werden die Durchläufe, auch parallel, nach Zu- und Abgängen getrennt erfasst. Durch die enge Erfassung von oben, oft mit Kameras, die auf den Anwendungsfall optimiert sind, ist der Datenschutz hier gut gewährleistet, die Ergebnisse sind sehr genau. Zum Einsatz kommt die Technik entsprechend dort, wo entweder die Zu- und Abgänge, ggf. nicht aber die Gesamtpersonenmenge relevant sind oder aber ein Raum über sämtliche Zugänge erfasst werden kann. Die Erfassungstechnik kommt daher oft eher im Fahrzeug als im Bahnhof zum Einsatz. Die hohe Genauigkeit macht die Technik trotz des engen Einsatzfeldes interessant. Ökonomisch muss hier der Einbau einer auf den Einsatzzweck spezialisierten Anlage sinnvoll sein.

Zählmatten

Der Einsatz von Zählmatten erfolgt oft temporär, da die Installation relativ simpel ist. Sie können an Orten mit definierten Zugängen eingesetzt werden und liefern hinreichend genaue Ergebnisse für Zu- und Abgänge, allerdings ist der Erfahrungsschatz derzeit noch relativ gering.

8.5 Datenanalyse

Grundvoraussetzungen der Datenanalyse

Bevor überhaupt mit der tatsächlichen Datenanalyse angefangen werden kann, müssen die technischen Rahmenbedingungen hergestellt werden. Denn was bringt eine Datenanalyse, wenn die dafür relevanten Daten fehlen oder von geringer Qualität sind?

Um diese Rahmenbedingungen herzustellen, sind bestimmte Herausforderungen zu bewältigen. Es stellt sich hierbei die Frage, wie die Konsolidierung heterogener Daten aus zahlreichen Quellen erfolgen soll, so dass als Ergebnis ein sogenannter »single point of truth« vorliegt.

Bei näherer Betrachtung der erwähnten Datenquellen ist das Ausmaß dieses Problems deutlich zu erkennen. Alle zurzeit verwendeten PaxCounter nutzen die Netzwerktechnologie LoRa. Für die praktische Nutzung bedarf es einer Art Middleware, die zwischen den LoRa-Gateways und der IoT-Applikation agiert. Die DB greift dabei auf das »The Things Network« (TTN) als offene Plattformlösung zurück, anstatt diese Middleware eigenständig zu entwickeln.

Mittels MQTT kann dann auf die Daten aus dem TTN zugegriffen werden. Jedoch ist MQTT nicht das einzige Kommunikationsprotokoll, welches im IoT-Kontext genutzt werden kann und genutzt wird. Das gleiche gilt für LoRa hinsichtlich der Netzwerktechnologien.

Je nach Einsatzort und Einsatzzweck des Sensors ist das jeweilige Protokoll oder die jeweilige Netzwerktechnologie einer anderen vorzuziehen. Dies bedeutet aber im Umkehrschluss, dass diese Protokolle bei der Datenintegration unterstützt werden müssen bzw. mögliche plattformspezifische Vorgaben berücksichtigt werden müssen, was die Komplexität stark erhöht.

Die zweite Herausforderung stellt die Echtzeitverarbeitung von Datenströmen dar. Die zeitnahe Verfügbarkeit von Daten nimmt eine immer größere Wichtigkeit ein. Applikationen, die darauf basieren, Anomalien zu erkennen und entsprechend Warnnachrichten auszusenden, sind auf eine solche Verfügbarkeit angewiesen. Dauert der Prozess des Verfügbarmachens zu lang, sind Daten beim Ankommen eventuell bereits obsolet.

ETL-Prozess

Das Data Warehouse wird häufig erwähnt, wenn es sich um die Konsolidierung von heterogenen Datenquellen handelt. Jedoch ist in diesem Kontext der ETL-Prozess relevant, bei dem die Daten aus den zahlreichen Quellen extrahiert, diese in ein einheitliches Format transformiert und in das Data Warehouse bzw. in die Datenbank geladen werden.

Es gibt viele Tools, die dabei unterstützen, diesen Prozess umzusetzen. Hierbei hat sich die DB in ihrem Prototypensystem für den »Telegraf« aus dem TICK-Stack entschieden. In der Konfigurationsdatei kann schnell und einfach angegeben werden aus welchen externen Quellen die Daten extrahiert werden sollen.

Standardmäßig werden bereits mehrere Dateiformate, Kommunikationsprotokolle, aber auch Datenbanken unterstützt. Zudem ist es erweiterbar, sodass eigene sogenannte Input Plugins entwickelt werden können, die es erlauben, dass auch nicht standardmäßige unterstützte Datenquellen integriert werden.

Bei der verwendeten Datenbank handelt es sich um die Zeitreihendatenbank InfluxDB, die zurzeit eine hohe Beliebtheit aufweist. Zeitreihendatenbanken sind auf das performante Lesen und Schreiben von hohen Datenmengen ausgelegt, was im IoT-Kontext von besonderer Wichtigkeit ist.

Stream Processing

Hierbei wird der Kapacitor, der ebenfalls Teil des TICK-Stack ist, verwendet. Es liegt eine direkte Anbindung zu der InfluxDB vor, sodass Datenänderungen direkt erkannt werden. Der Kapacitor stellt eine einfach zu verstehende Skriptsprache, die TICKscript genannt wird, zur Verfügung. Diese ermöglicht es, Daten, auch aus verschiedenen Tabellen, miteinander zu kombinieren, zu aggregieren und Berechnungen durchzuführen. Somit ersetzt diese Skriptsprache weitestgehend die Nutzung von komplizierten Datenbankabfragen und erweitert diese um zahlreiche Funktionen.

Automatisierte Datenanalyse

Die lokale Durchführung von Datenanalysen wird als Hindernis angesehen. Für die Unterstützung von Entscheidungsprozessen sind zeitnahe Bereitstellungen der Ergebnisse relevant. Folglich ist die Überlegung aufgekommen, ein System bereitzustellen, welches automatisiert Daten analysiert und die Ergebnisse visualisiert. Auf die Visualisierung kann dann im Anschluss von Mehreren zugegriffen werden. Hierbei wird ebenfalls das automatisierte Trainieren der zugrundeliegenden mathematischen Modelle mit den ankommenden Datenströmen in Betracht gezogen.

Die DB hat sich entschieden, den Kapacitor, der bereits die Möglichkeit anbietet, Datenströme zu verarbeiten, um analytische Modelle zu erweitern. Dies wird mittels sogenannten User Defined Functions (UDF), die im Skript aufgerufen werden können, erreicht. Die UDFs können in verschiedenen Programmiersprachen implementiert werden. Zurzeit werden Golang und Python unterstützt, wobei sich für den Prototypen für Python entschieden wurde, da es reich an Softwarebibliotheken für die Datenanalyse ist. Für die Verwendung weiterer Programmiersprachen müssen Konnektoren implementiert werden.

random_forest

```
dbrp "telegraf"."autogen"

var compute = 8d

var every = 5m

var source = '{"database": "history", "retentionPolicy": "autogen", "measurement": "pax_counter", "fields": ["ble", "wifi"]}'

var mode = '{"calendardays": false, "weekdays": true, "hours": true, "minutes": true, "seconds": false}'

var window = 30d

var frequency = 4m

var fields = '{"fields": ["payload_fields_ble", "payload_fields_wifi"]}'

var data = stream
    |from()
        .measurement('pax_counter_ba')
    |groupBy('dev_id')

data
    @randomforestregression()
        .compute(compute)
        .every(every)
        .fields(fields)
        .frequency(frequency)
        .source(source)
        .timevariables(mode)
        .window(window)

    |InfluxDBOut()
        .create()
        .database('regression')
        .retentionPolicy('autogen')
        .measurement('pax_counter_randomI')
```

Abbildung 1: TICKscript für die Erstellung eines Tasks

Die UDFs stellen eine weitere Abstraktionsebene dar und sind generisch aufgebaut. Sie sind anhand des Präfixes »@« zu erkennen. Ihnen können Parameter übergeben werden, sodass mehrere Varianten der Datenanalyse durchgeführt werden können (siehe Abbildung 1).

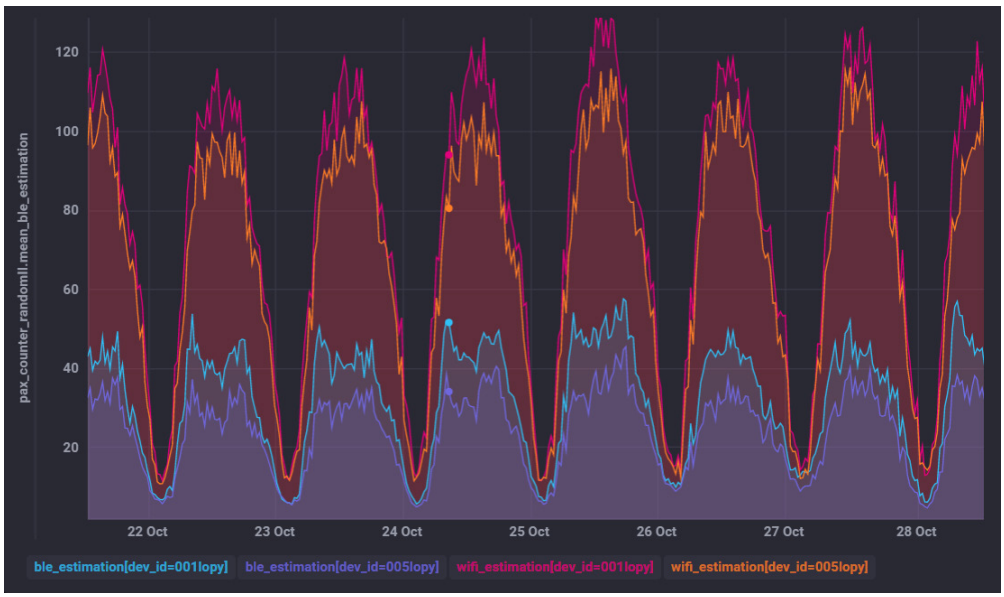


Abbildung 2: Visualisierung der Extrapolation von Zähldaten zweier PaxCounter

Beim Pax-Counter wurde mit der Erstellung einer automatisierten Regressionsanalyse begonnen (siehe Abbildung 2). Neben dem Zeitraum der Trainingsdaten (window) und den Zeitraum der Extrapolation (compute) kann auch definiert werden, welche erklärende Zeitvariablen verwendet werden wollen, sprich es kann eine Auswahl zwischen Sekunden, Minuten, Stunden, Wochentagen und Kalendertagen erfolgen (siehe Abbildung 1). In dem in Abbildung 1 gezeigten Beispiel wird für jedes Gerät für die einzelnen Zähldaten (Bluetooth und WLAN), die mittels der Variable »fields« definiert werden, eine Regressionsanalyse durchgeführt. Die Gruppierung nach einzelnen Geräten wird mittels der Verwendung des Befehls »groupBy« ermöglicht.

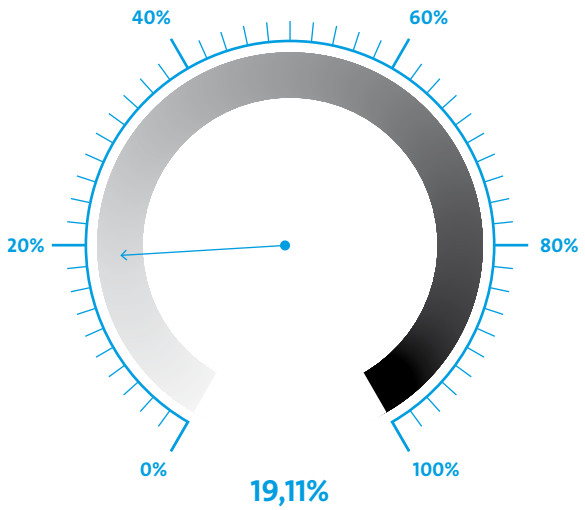


Abbildung 3: Güte als Mittlerer absoluter Fehler (geringer desto besser)

Es können unterschiedliche Kombinationen der Parameter betrachtet werden und es kann getestet werden, welche dieser Kombinationen die besten Ergebnisse liefert, denn gleichzeitig läuft ein weiterer Agent, der kontinuierlich die extrapolierten Daten mit den ankommenden neuen Daten abgleicht und anzeigt, welche Güte die Extrapolation aufweist (siehe Abbildung 3).

Abbildung 4: Bedienung der Oberfläche des Analysesystems

Der vorliegende Prototyp verlangt keine grundlegenden Kenntnisse im Programmieren und spricht somit eine größere Nutzergruppe an. Nachdem ein Task mittels des TICKscripts definiert und aktiviert wurde, muss für die weitere Nutzung nicht viel erfolgen. Je nach Auswahl der Parameter und der Laufzeit des jeweiligen Tasks können bestimmte Zeiträume über eine Datumsauswahl gezielt betrachtet werden (siehe Abbildung 4). Dabei sollte berücksichtigt werden, dass Schätzungen von weit in der Zukunft liegenden Zeiträumen nicht unbedingt aussagekräftig sind.

8.6 Ausblick

Im nächsten Schritt will versucht werden, bestimmte Eventdaten wie beispielsweise Störungen im Bahnbetrieb oder Veranstaltungen in den Analysen zu berücksichtigen, um eventuell bessere Ergebnisse in der Extrapolation zu erzielen.

Zudem wird versucht, den Automatisierungsgrad zu erhöhen. Der jetzige Automatisierungsgrad, der als Automatisierungsgrad 1 bezeichnet werden kann, bietet nur einen Einblick, wie das Personenaufkommen zu einem bestimmten Zeitpunkt aussehen könnte. Im Kontext der Personaldisposition müssen die Entscheidungsträger aus den Ergebnissen der Analysen die Anzahl des zu einzusetzenden Bahnhofspersonals ableiten.

Im Automatisierungsgrad 2 könnten ebenfalls die Daten der Personaldisposition in der Datenanalyse berücksichtigt werden, sodass neben der Extrapolation des Personenaufkommens im Bahnhof auch Vorschläge gemacht werden, wie viel Personal eingesetzt werden soll. Hierfür sind aber historische Daten notwendig, aus denen hervorgeht, wie hoch die Anzahl des Bahnhofspersonals und des Personenaufkommens an jeweiligen Zeitpunkten gewesen ist. Ebenfalls muss dabei berücksichtigt werden, dass an bestimmten Tagen eventuell zu viel oder zu wenig Bahnhofspersonal eingesetzt wurde, was aus den Datensätzen hervorgehen muss.

Im Automatisierungsgrad 3 würde kein manueller Eingriff mehr stattfinden. Das System kommuniziert der Software in der Personaldisposition das extrapolierte Personenaufkommen und führt die Planung der Einsätze automatisiert durch.

8.7 Abbildungsverzeichnis

9 Hand in Hand: Wie KI und Ärzte in der Onkologie zusammenarbeiten

9 Hand in Hand: Wie KI und Ärzte in der Onkologie zusammenarbeiten

Matthieu-P. Schapranow

9.1 Einleitung

In vielen Bereichen unseres täglichen Lebens begegnen uns schon heute Fortschritte durch die Digitalisierung, z.B. beim sekundenschnellen Bestellen von Waren im Internet oder bei der Suche der besten Verbindung, um zum Reiseziel zu gelangen. Aber auch lebenswichtige Disziplinen wie die Medizin profitieren von Datenanalysen in Echtzeit. In der Medizin können zum Beispiel mithilfe neuester diagnostischer Verfahren riesige Datenmengen binnen kürzester Zeit erhoben werden, wie z.B. Laboruntersuchungen oder das persönliche Genom. Doch Mediziner stehen nun vor der Herausforderung, der schier unendlichen Datenmenge Herr zu werden. Hierzu sind neben passenden Methoden auch maßgeschneiderte Analysewerkzeuge erforderlich, die sie bei ihrer täglichen Arbeit unterstützen. [1] Daran arbeiten das Hasso-Plattner-Institut und die Plattform Lernende Systeme gemeinsam. Lassen Sie sich auf eine Reise in eine nahe Zukunft ein und begleiten Sie Herrn Merk ein Stück während seiner Krankheit.

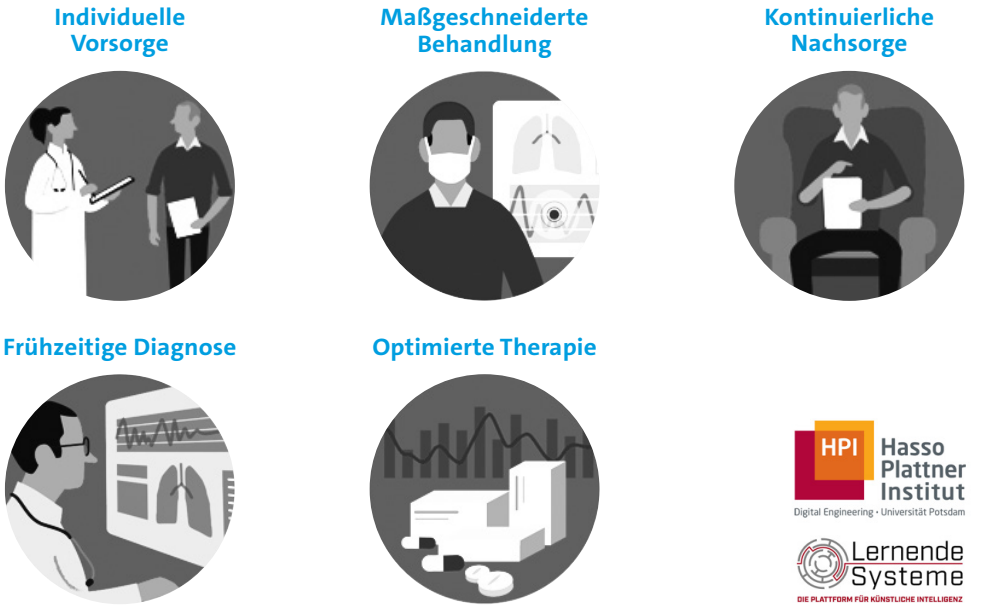


Abbildung 1: Ausgewählte Stationen während der Behandlung von Herrn Merk.

9.2 Das persönliche Gespräch mit einem Arzt: flexibel, wo und wann ich mag!

Jeder kennt ihn: Den »Arzt des Vertrauens«. Der sehr persönliche Austausch mit dem Arzt ist nicht nur ein fachlicher, sondern ebenso ein emotionaler Austausch. Wem sonst vertraut man so persönliche Dinge wie Gefühle an? Bisher stellt das Behandlungszimmer des Arztes den dazu erforderlichen geschützten Raum dar. Hinter verschlossenen Türen können Patienten sich öffnen und ihrem Arzt des Vertrauens sensible Dinge anvertrauen. Doch die Grenzen dieser vertrauten Umgebung wandeln sich gerade in Zeiten ständiger Mobilität. So ebnet die Möglichkeit der Fernbehandlung, z.B. mittels Tele-Konsil, sogar die Schaffung virtueller Behandlungszimmer. Wenn Herr Merk es wünscht, kann er seine Hausärztin von Zuhause aus seinen eigenen vier Wänden kontaktieren oder sie von jedem beliebigen Ort der Welt um Ratschlag bitten. So profitiert er als Bewohner einer ländlichen Region, da bei ihm früher der Weg zum nächstgelegenen Arzt mit signifikanter Reisezeit verbunden war. Aber auch unnötige Wege und lästige Wartezeiten im Wartezimmer können Herr Merk erspart werden und er entlastet gleichzeitig die Umwelt. Dennoch, ernsthafte Erkrankungen erfordern auch künftig den persönlichen Besuch eines Arztes.

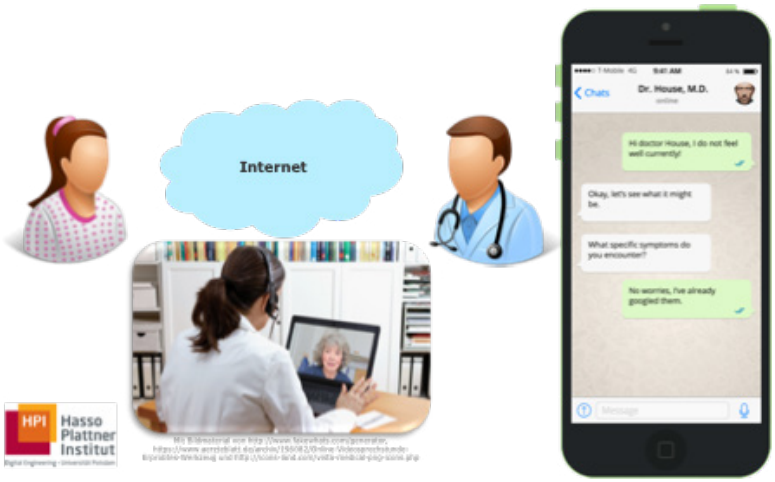


Abbildung 2: Künftig wird das persönliche Gespräch mit dem Arzt an jedem Ort der Welt möglich sein, wenn es der Patient so wünscht.

9.3 Individuelle Vorsorge ermöglicht rechtzeitiges Handeln

Herr Merk leidet seit einiger Zeit an der chronisch-obstruktiven Atemwegserkrankung COPD. Ein KI-Algorithmus nutzt Daten aus seiner digitalen Patientenakte, um den Verlauf und das Risiko für etwaige Folgeerkrankungen zu errechnen. Dazu bildet der KI-Algorithmus eine persönliche Referenzkohorte bestehend aus Daten von ähnlichen Patienten, die z.B. über ein

ähnliches Alter, einen ähnlichen Lebensstil, aber auch über gleiche Vorerkrankungen verfügen. Hierzu greift der KI-Algorithmus auf eine zentrale Forschungsdatenbasis zu, die ganz ohne Personenbezug auskommt. Denn dem Algorithmus geht es nicht um Namen oder Geburtsdaten, sondern um eine möglichst repräsentative Gruppe von Bürgern und deren Krankheitsverlauf, egal ob heute oder vor Monaten. So kann der KI-Algorithmus Herrn Merk nicht nur vergleichend einordnen, sondern auch die Wahrscheinlichkeit für etwaige Folgeerkrankungen und deren Häufigkeit bestimmen. Die errechnete Wahrscheinlichkeit wird in die drei Kategorien »geringes Risiko«, »erhöhtes Risiko« und »sehr hohes Risiko« einsortiert. Dieser Einblick hilft der Hausärztin von Herrn Merk dabei, frühzeitig auf mögliche Risiken aufmerksam zu werden, selbst wenn einer ihrer Patienten an einer seltenen Erkrankung leiden sollte, die sie selbst noch nie zuvor gesehen hat. Sie kann so passende Vorsorgeuntersuchungen initiieren, um den Gesundheitsverlauf von Herrn Merk bestmöglich zu überwachen. Sie rät Herrn Merk zu einem vierteljährlichen Lungenfunktionstest und gibt ihm den Hinweis, dass alle Atemwegsinfekte medizinisch abzuklären sind.

9.4 Frühzeitige Diagnose: KI-gestützte Bildgebung erkennt selbst kleinste Veränderungen

Herr Merk ist seit Wochen durch einen Hustenreiz geplagt, der einfach nicht besser zu werden scheint. Der Lungenfunktionstest zeigt eine reduzierte Lungenfunktion, vermutlich aufgrund des Infekts. Sein Lungenfacharzt erhält zusätzlich durch das KI-System bei der Eingabe der Symptome die Empfehlungen der aktuellen Leitlinie, eine bildgebende Untersuchung der Lunge zum Ausschluss von malignen Veränderungen durchzuführen. Daher wird Herr Merck zum Radiologen überwiesen. Mithilfe eines Minimaldosis-CTs wird die Lunge von Herrn Merk geröntgt. Der Radiologe erhält auf seinem Bildschirm die Bilddaten in digitaler Form und mit Annotationen versehen. Denn die Rohdaten aus dem CT-Gerät wurden bereits durch ein KI-System analysiert, das die Aufnahme von Herrn Merks Lunge mit einer großen Datenbasis früherer Bilddaten abgeglichen hat, z.B. um die Form der Lunge zu identifizieren, das Lungenvolumen zu vermessen und etwaige Änderungen im Gewebe hervorzuheben. So erhält der Radiologe neben dem eigentlichen Bild auch schon zahlreiche Zusatzinformationen, die er ein- und ausblenden kann. So hat der Radiologe bei der Diagnostik erstmals Zugang zu Wissen aus einer Vielzahl historischer Fälle, selbst solcher, die er selbst nicht bearbeitet hat. Es entsteht eine weltweite Sammlung medizinischen Wissens, von der alle profitieren, insbesondere der Patient. Aber auch das KI-System lernt: wenn der Radiologe eine Anpassung an den vorgeschlagenen Ergebnissen vornimmt, fließt diese Information als Feedback zurück in das darunterliegende Modell. So interagieren erstmals Mensch und Maschine und lernen so vom Verhalten des jeweils anderen. Bei Herrn Merk sind sich Mensch und Maschine einig: es gibt einen kleinen Bereich in seiner Lunge, der verdächtig aussieht und besser entfernt werden sollte.

9.5 Maßgeschneiderte Therapie dank aktuellster Leitlinien

Nachdem bei Herrn Merk eine Biopsie des verdächtigen Gewebes durchgeführt wurde, steht die Diagnose fest: es handelt sich um ein sehr frühes Stadium von Lungenkrebs. Zwar ist die Diagnose ein Schock für Herrn Merk und sein familiäres Umfeld, aber der Arzt kann ihn beruhigen. Denn dank der frühen Diagnose sind die Heilungsaussichten sehr gut. Um die passende Therapie für die Behandlung des Bronchialkarzinoms auszuwählen, wird der Arzt durch ein KI-System unterstützt, das Zugang zu den neuesten klinischen Leitlinien und ausgewählten Daten der digitalen Patientenakte von Herrn Merk hat. Dazu extrahiert das KI-System relevante Untersuchungen und Behandlungsdetails im bisherigen Verlauf und gleicht sie mit den aktuellen Leitlinien der Deutschen Krebsgesellschaft ab. Der Arzt spart dabei viel Zeit, denn so können bereits vorliegende Untersuchungsergebnisse automatisch bei der Auswahl möglicher Behandlungsalternativen berücksichtigt werden. Empfohlene Untersuchungen und Behandlungsalternativen zeigt das Unterstützungssystem dem Arzt an, der dann weitere Untersuchungen anfordert, z.B. molekulargenetische Untersuchungen des Tumorgewebes. Die für Herrn Merk passende klinische Leitlinie empfiehlt eine Entfernung des karzinogenen Lungengewebes gefolgt von einer kurzen Chemotherapie. Aufgrund der frühzeitigen Diagnose kann der Chirurg auf eine minimalinvasive Operationsmethode zurückgreifen, die mit zwei kleinen Schnitten entlang der betroffenen Lungenseite auskommt. Während der Operation, bei der das veränderte Lungengewebe vollständig entfernt wird, erhält der Chirurg Hilfe von einem KI-System, das ihm das Tumorgewebe und umgebende Organstrukturen in seiner digitalen OP-Brille markiert, selbst wenn er diese nicht direkt über die Kamera sieht, z.B. weil sie verdeckt liegen. Mithilfe dieser interaktiven Navigationshilfe kann sich der Chirurg bei jedem seiner Schritte sicher fühlen, dass er das Tumorgewebe vollständig entfernt und gesunde Gewebeteile nicht beschädigt werden.

9.6 Therapie nach Maß

Herr Merk hat die Operation sehr gut überstanden und nach einigen Tagen kann er das Krankenhaus wieder verlassen. Der Chirurg hat ihn bereits darauf hingewiesen, dass sich noch eine kurze Chemotherapie anschließt, um auch Tumorzellen zu beseitigen, die sich möglicherweise bereits vom eigentlichen Tumor gelöst haben. Dazu wurde das entfernte Tumorgewebe molekulargenetisch untersucht und es wurden maßgebliche genetische Veränderungen in den Genen NRAS und ERBB2 identifiziert. Die klinische Leitlinie enthält zu dieser Kombination keine speziellen Behandlungsvorschläge und lässt dem Onkologen eine große Wahl an Therapieformen. Der Onkologe nutzt nun das Medical Knowledge Cockpit [3], das Herrn Merk bereits mit ähnlichen Fällen weltweit verglichen hat. Dabei wurden einige Patienten mit genetischen Übereinstimmungen identifiziert und der Onkologe sieht nun deren Behandlungsverlauf. So kann der Onkologe selbst recherchieren und sich einen Überblick über bereits durchgeführte Therapien und deren Erfolg sowie mögliche Nebenwirkungen und Risiken verschaffen. Dieses datengetriebene Vorgehen ermöglicht eine evidenzbasierte Auswahl der Therapieform. Der Onkologe setzt auf eine Kombination mehrere Wirkstoffe, die den Wirkstoff nur an Zellen mit den genetischen Veränderungen des Tumors freigeben. Die zielgerichtete Therapie verschont also alle übrigen

gesunden Körperzellen und ist so deutlich besser verträglich. Erstmals werden die dokumentieren Behandlungsverläufe nicht nur für Abrechnungszwecke, sondern auch für die Behandlung künftiger Patienten eingesetzt und helfen so Menschen weltweit.

9.7 Vorbeugen durch regelmäßige Nachsorge

Herr Merk hat die Operation und die anschließende kurze Chemotherapie gut überstanden und ist seit einigen Monaten frei von Beschwerden. Trotzdem besucht er regelmäßig seine Hausärztin, damit sie ihn im Rahmen der Nachsorge engmaschig überwacht, um etwaige Veränderungen frühzeitig zu identifizieren. Zwar muss Herr Merk für die Nachsorge regelmäßig persönlich bei seiner Hausärztin vorbeischauen, er weiß aber, wie wichtig es war, dass seine Erkrankung in einem sehr frühen Stadium diagnostiziert wurde. Darüber hinaus weiß Herr Merk, dass er auch anderen Menschen mit seinen Behandlungsdaten helfen kann. Dabei hilft ihm die App des »Datenspendeausweises«. In seinem Profil in der App hat Herr Merk hinterlegt, dass er sich für Behandlungen von Lungenkrebs und COPD interessiert. Daraufhin erhielt er eine Nachricht, dass für ein Forscherteam genau diese Kombination besser erforschen möchte, dazu aber Zugang zum Behandlungsverlauf bereits behandelter Patienten benötigt. Nach einem ausführlichen Aufklärungsgespräch durch seine Hausärztin hat sich Herr Merk dazu entschlossen, Daten über seinen Behandlungsverlauf für dieses Forschungsprojekt zur Verfügung zu stellen. Dazu werden die Daten anonymisiert in eine Forschungsdatenbasis transferiert und durch das Forscherteam ausgewertet. Herr Merk bekommt monatlich eine Übersicht, wie oft seine Daten bereits verwendet wurden. Dadurch merkt er, wie wichtig sein Beitrag zur Forschung ist. Sollte es sich Herr Merk künftig doch anders überlegen, kann er seine Einwilligung zur Datennutzung jederzeit selbst über die App zurückziehen. Danach sind seine Gesundheitsdaten für weitere Projekte nicht mehr verwendbar.

9.8 Abbildungs- und Literaturverzeichnis

[1] ↗ **https://hpi.de/**

[2] ↗ **https://www.plattform-lernende-systeme.de/**

[3] ↗ **https://we.analyzegenomes.com/mkc/**

10 Nutzung von KI im Service und bei der Wartung von Großanlagen

10 Nutzung von KI im Service und bei der Wartung von Großanlagen

Thorsten Jankowski & Achim Steinacker

10.1 Motivation für Voith

Voith hat als klassischer Maschinenbauer im April 2016 eine eigene Group Division gegründet mit dem Auftrag, die Digitalisierung im Unternehmen voranzutreiben, damals unter der Digital Solutions, seit Oktober 2018 als Digital Ventures. Damit verbunden waren nicht nur die Bestrebungen, ein neues externes Geschäft zu schaffen, sondern auch interne Mehrwerte durch Digitalisierung voranzutreiben sowie die damit verbundene Digitale Transformation des eigenen Unternehmens.

Als erster Schritt zur Entwicklung eines langfristigen Geschäftsmodells im Bereich Service und Wartung von Großanlagen fand eine Befragung der Mitarbeiter und Kunden von Voith statt. Das Hauptaugenmerk der Befragung lag dabei in der Ermittlung der größten Defizite bzgl. Zeit und Qualität der Produkte und Arbeitsprozesse.

Aufbauend auf den Ergebnissen wurde durch Design-Thinking-Methoden ermittelt, dass die Hauptansätze zur Verbesserung und der Generierung von Mehrwerten in den folgenden Bereichen liegen:

- Bereitstellung von integrierten und einheitlichen Sichten auf alle Daten und Informationen, die bei der Entwicklung, insbesondere aber beim Betrieb und Wartung der Anlagen erforderlich sind, sowohl für die Mitarbeiter von Voith als auch insbesondere für die Kunden.

- Nutzung neuer Medien wie AR und VR für die Wartung und den Betrieb der Anlagen.

Als nächster Schritt wurden externe und interne Pilotprojekte aufgesetzt, um die erarbeiteten Use Cases an lauffähigen Systemen zu testen und detailliertes Feedback der Benutzer zu sammeln. Dabei zeigte sich sehr schnell, dass der Weg von der Idee bis hin zu einem produktiven Produkt oftmals weiter war als erhofft und in den Systemen kein vollständiger »end-to-end«-Prozess abgebildet werden konnte.

Der Hauptgrund dafür lag an der fehlenden Konsistenz des benötigten Datenmodells, die für die Umsetzung eines vollständigen Prozess benötigt wird. Obwohl die erforderlichen Daten in unterschiedlichen Ablagesystemen zwar grundsätzlich vorhanden sind, fehlte es an Verbindungen zwischen den Daten aus den unterschiedlichen Systemen. Sehr schnell zeigte sich zudem, dass der Aufbau eines Data Lakes und die Analyse der Daten mit Verfahren des Maschinellen Lernens nicht ausreichend waren, um die benötigten Strukturen und die Zusammenhänge zu ermitteln. Ein weiteres Problem bestand darin, dass viele Informationen in unstrukturierten Texten und Dokumentationen vorlagen. So ließ sich beispielsweise keine kontextspezifische und

individualisierte Anzeige modularer Inhalte optimiert auf den Benutzer und das verwendete Anzeigegerät umsetzen, da die Informationen nur in PDF-Dateien mit mehreren tausend Seiten Umfang und ohne jegliche Metadaten zur Verfügung standen.

Damit war klar, dass die Umsetzung innovativer Dienste beim Betrieb und der Wartung von Großanlagen nur nach Schaffung der folgenden Voraussetzungen möglich werden würde:

- Aufbau eines zentralen Metadatenmodells zur Harmonisierung aller benötigten Informationen.
- Nutzung intelligenter Analysetechnologien für die Extraktion von Informationen aus unstrukturierten Daten.
- Schaffung einer einheitlichen Kommunikationsinfrastruktur durch Einsatz einer Informations-Middleware.

10.2 Technologische Umsetzung

Wissensnetze als Grundlage »wissensbasierter KI«

Für den Aufbau eines zentralen Metadatenmodells wurde der Ansatz der semantischen Repräsentation des Modells in Form eines Wissensnetzes (KnowledgeGraph) bzw. einer Ontologie gewählt. Obwohl es im aktuellen Hype rund um das Maschinelle Lernen im Moment ein wenig untergeht, decken Ontologien als wissensbasierte KI einen großen Bereich der Künstlichen Intelligenz ab und werden seit vielen Jahren in produktiven Anwendungen eingesetzt. Sie sind in der Lage, Datenmodelle über Systemgrenzen hinweg zu harmonisieren und über Beziehungen die Zusammenhänge zwischen Daten aus verschiedenen Systemen abzubilden. Darüber hinaus können über Regeln die Zusammenhänge und Abbildungen standardisiert beschrieben und damit korrekt umgesetzt werden. Ein weiterer entscheidender Vorteil eines semantischen Modells ist die Verfügbarkeit etablierter Standards für die Repräsentation aller Daten, die bei der Umsetzung der Anwendungsfälle benötigt werden. Steht ein einheitliches Datenmodell zur Verfügung, das die Informationen aus den verschiedenen Silos integriert, kann das Modell sowohl Anfragen von Benutzern beantworten, als auch andere Systeme mit integrierten Informationen versorgen, ohne dass die Systeme oder Benutzer sich darum kümmern müssen, aus welchen Silos die Informationen ursprünglich stammen. Andreas Dengel hat diese Funktion einmal in einer Publikation als »Informationsbutler« beschrieben. [2]

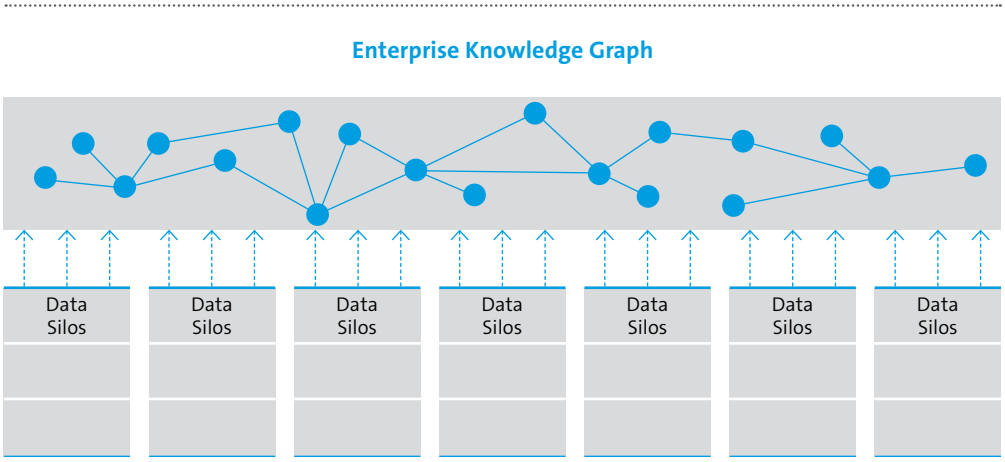


Abbildung 1: Die semantische Repräsentation des Modells in Form eines Wissensnetzes (KnowledgeGraph) bzw. einer Ontologie.

Zum ersten Aufbau eines zentralen Meta Datenmodells für Voith wurden die verschiedenen Systeme analysiert, die produktrelevante Daten in Form von Stücklisten und dazugehörende Dokumentationen enthalten. Grundsätzlich können diese recht einfachen hierarchischen Strukturen problemlos auf Knoten und Kanten in einem Graphen abgebildet werden. Die Regeln für die automatisierte Generierung von Beziehungen zwischen den teilweise redundanten Informationen und die Abbildung von Beziehungen erwiesen sich als recht aufwändig, konnten aber bewältigt werden. Damit steht jetzt ein Modell zur Verfügung, das erstmalig über alle Systemgrenzen hinweg eine vollständige und zuverlässige Sicht auf alle Daten der Voith-Produkte zulässt.

10.3 Strukturierte, halbstrukturierte und unstrukturierte Daten

Die strukturierten Informationen über die Anlagen von Voith, die verbauten Komponenten, Merkmale und Funktionen bilden das Rückgrat des semantischen Modells. Sie liegen in ERP- und PIM-Systemen und können über manuell erstellte Regeln und Transformationen in das Wissensnetz übernommen werden. Für den weitaus größten Teil der benötigten Informationen für die Umsetzung der Anwendungsfälle gilt dies jedoch nicht. Sie sind in Redaktionssystemen, Sharepoints, Laufwerken, Wikis und anderen CMS-Systemen abgelegt, kaum strukturiert und auch nur in Ansätzen mit Metadaten versehen, die eine Einordnung in das semantische Modell ermöglichen würden. Dennoch müssen sie über dieselbe einheitliche Schnittstelle verfügbar sein wie die strukturierten Daten.

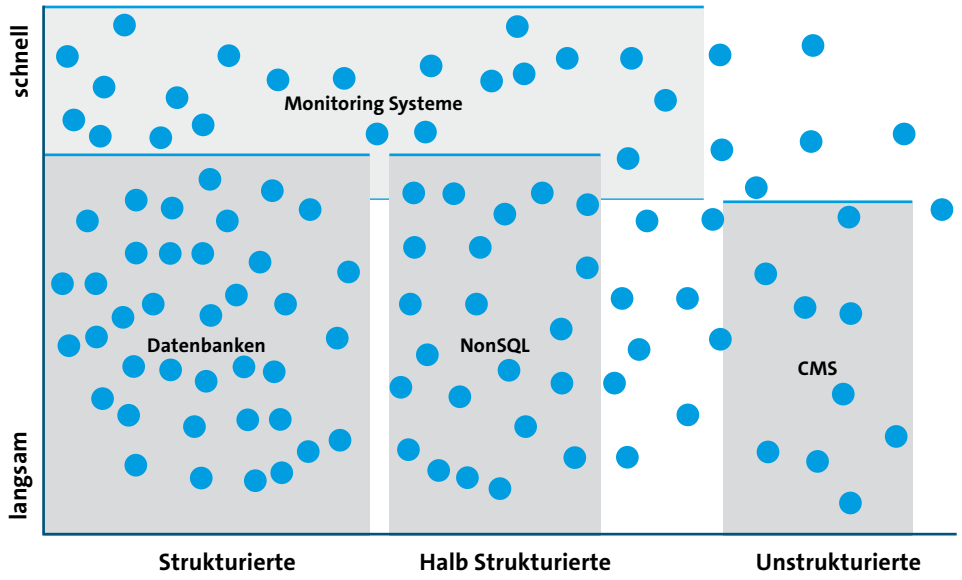


Abbildung 2: Strukturierte, halbstrukturierte und unstrukturierte Daten bilden das Rückgrat des semantischen Modells.

An dieser Stelle werden daher Techniken aus dem Bereich TextMining mit semantischem Textverständnis eingesetzt, um möglichst automatisiert Metadaten aus den unstrukturierten Informationen zu extrahieren. Es gibt zudem erste Überlegungen, mithilfe von Maschinellem Lernen automatisiert Metadaten zu generieren, die zumindest den Typ eines Dokuments (Wartungsanleitung, Benutzerhandbuch, etc. ...) und die Produktlebenszyklusphase, auf die sich das Dokument bezieht, repräsentieren. Dazu ist aber, wie sich gezeigt hat, eine sehr homogene Trainingsmenge erforderlich, sodass aktuell die klassischen Entity-Recognity- und statistischen Klassifikationsverfahren noch bessere Ergebnisse erzielen.

Im nächsten Entwicklungsschritt wurde das Metadaten-Modell erweitert so, dass die Metadaten, die für die Beschreibung der unstrukturierten Dokumentationen verwendet werden, mit aufgenommen werden können. Basis des entwickelten Modells stellt dabei der »Intelligent Information Request And Delivery Standard« (iiRDS) des tekom dar. Dieser Metadatenstandard, der seit Mitte 2019 in der Version 1.0.1 vorliegt, verwendet als Repräsentation ebenfalls ein semantisches Modell und lässt sich damit nahtlos in das eigene Produktmodell integrieren. Damit steht ein System zur Verfügung, mit dem einer der wichtigsten Anwendungsfälle umgesetzt werden konnte, nämlich das »kontextbasierte Suchen & Finden«.

Ein Ingenieur benötigt für die Inbetriebnahme einer Maschine bspw. alle relevanten Zeichnungen für einen definierten Zeitpunkt während der Montage. In der Vergangenheit musste er diese Zeichnungen aus diversen Datensilos manuell zusammentragen und meist eine zusätzliche Liste führen in welcher es galt die Querverbindungen der einzelnen Strukturen abzugleichen. Heute gibt er das Datum ein und bekommt die Information mit zusätzlichen Filterfunktionen, da das Wissensnetz weiß, welche Dokumentenarten oder zusätzliche Informationen zu diesem Datum vorhanden sind.

Als nächster Schritt steht jetzt die Erweiterung des Metadatenmodells für den Umgang und die Verwaltung von Sensordaten an, um einen weiteren Anwendungsfall umzusetzen, nämlich die sensorgestützte Wartung der Voith-Anlagen. Verfahren des Maschinellen Lernens analysieren die von den Sensoren erfassten Rohdaten und liefern im Austausch mit dem semantischen Modell, in dem die Wartungshistorie der Anlage zur Verfügung steht. So kann unter Berücksichtigung vergangener Probleme und mit der Analyse der Sensordaten, die zu den historischen Problemen geführt haben, ein möglicher Fehlerfall frühzeitig erkannt werden. Zudem können noch detaillierte Informationen zur Behebung des Problems oder möglicherweise benötigte Fähigkeiten zur Verfügung gestellt werden. An dieser Stelle wird damit eine Kombination der maschinellen Lernverfahren und der wissensbasierten KI umgesetzt, die eine Erklärbarkeit der Ergebnisse der automatischen Analysen und Schlussfolgerungen liefern kann.

10.4 Zusammenfassung

Die Umsetzung der von Voith ausgewählten Anwendungsfälle haben sehr schnell gezeigt, dass sie ohne den Einsatz von Künstlicher Intelligenz nicht realisierbar sind. Der Einsatz eines Data Lakes und die Analyse der Rohdaten mit Verfahren des Maschinellen Lernens stellen dabei einen wichtigen Faktor bei der (Vor-)Verarbeitung der Massendaten dar. Letztendlich muss es aber das Ziel sein, die Daten nicht nur zu analysieren, sondern sie in einem wissensbasierten Kontext abzulegen und damit semantisch auf eine höhere Ebene zu bringen. Damit hat sich gezeigt, dass vor allem der Einsatz der klassischen, wissensbasierten KI erforderlich ist, da nur so die vielbeschworenen »smart services« realisiert werden können, welche die im Umfeld der Industrie 4.0 beschriebenen Anwendungen erfordern.

10.5 Abbildungsverzeichnis

11 Building a Pricing Engine in Five Weeks

11 Building a Pricing Engine in Five Weeks

Kim Nilsson

11.1 Introduction

Not a day goes by without a media story breaking around the opportunities, or dangers, of using data science and AI within society. Some argue that it is all hype, others that we are now beyond hype and into the famous »trough of disillusionment« according to Gartner's »Hype Cycle«. In our experience, there is an element of both within European markets. Most large corporates and multi-nationals today have teams of data scientists mining the company's data for insights, efficiencies, and product innovation, and the Big Data industry is maturing in these segments. However, in smaller businesses, such as SMEs and the »Mittelstand«, data science is yet a novelty and a new challenge. It is also where the greatest opportunities lie.

Many medium to large organisations store large proprietary databases related to sales, operations, finance, and more, and they rarely utilise this data beyond standard reporting and business intelligence. The opportunity to go deeper into the data, combine it in novel ways, and apply sophisticated machine-learning algorithms often leads to immediate gains in revenue or profit. This case study is an example of the powerful potential for data science to impact a medium-sized organisation in a short period of time. The best news of all: it is easily replicable across countless other organisations and industries.

The project described in this case study was a joint partnership between Pivigo and an automotive client in the UK. Pivigo is a data science marketplace and training provider based in London and Berlin. Pivigo supports organisations, large and small, to innovate with data by connecting them with a global pool of data science experts. Pivigo has worked on 200 projects, with 120 organisations, from very large, global companies such as AstraZeneca, Barclays, and KPMG, through SMEs such as BD24 Insurance, to small and nimble startups.

The client is a major force in the distribution of parts in the automotive aftermarket, with £250m turnover and growing. They are the UK and Ireland's largest group of wholly owned businesses and franchises, supplying 30,000 different manufactured car parts to national and local independent garages and workshops.

The particular project described here was completed as part of a training programme called »Science to Data Science« (S2DS), run by Pivigo three times per year. S2DS is a training programme meant to support the transition of academics into new jobs as commercial data scientists. The programme is five weeks full-time, during which the participants (MSc's and PhD's from analytical backgrounds) work on real data science projects in collaboration with partner companies. This gives the participants training »on the job« and commercial experience, and the partners the outcome of the project and a recruitment opportunity.

11.2 The project set-up

The client distributes car parts to local garages of all sizes, and 95% of these sales are done via phone calls from the garage to a sales person within the client. The client offers some 50,000 unique parts from 74 branches, to 18,000 unique customers, and delivery times are normally between half an hour to an hour. The question that the client wanted to solve using their data set was whether their sales were optimal. Hence, the first task was to complete an exploratory analysis of the company's sales history, to generate initial insights and understand buyer behaviour. The client also wanted to explore what variables affected revenue. This was finally followed with a more in-depth attempt at creating a pricing engine, with the ability to make price recommendations in order to maximise revenue and profit. To complete this analysis, the team of four PhD data scientists had a sales history of 20 million transactions at their disposal.

11.3 Results

11.3.1 Understanding buyer behaviour

The first analysis was a simple tree diagram of which customers bought which products. As an initial attempt at customer segmentation, two main categories of buyers were identified: the »Buy at need« customers, and the »Full-stack« customers. In the diagram below, it is clear that some customers purchase fewer brands, and less frequently, whereas others purchase a variety of brands more frequently. As an initial piece of insight, this can allow the client to target different customers with special marketing offers, and also to identify those segments of accounts worth nurturing and spending more time and resources on retaining and growing, and which segments to focus less resources on.

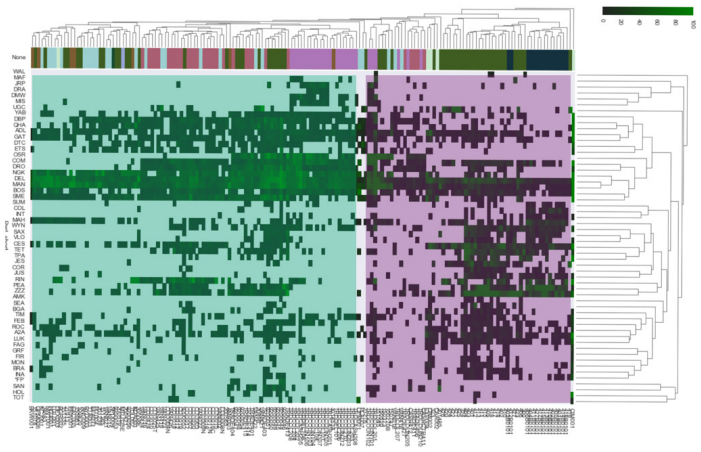


Figure 1: Buyer accounts versus product category. Purple area shows »full-stack« buyers (frequent, multi-brand purchages), green area shows »buy at need« customers (infrequent, single purchases).

11.3.2 Variables affecting revenue

In the next analysis, the team analysed the performance of individual sales branches and created a model to examine the relative performance of the branches. Controlling for factors like line of business, location and branch size, the team ranked the branches with an objective measure of performance. The team found large variations in revenue per sales branch, see Fig. 2.

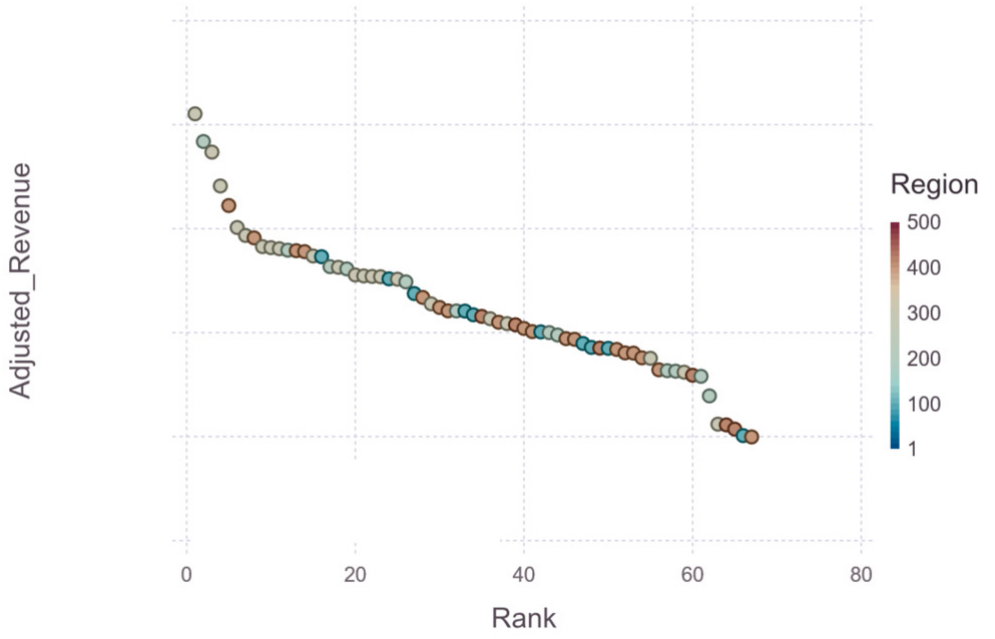


Figure 2: Ranked, adjusted revenue per sales branch. The graph shows clearly a few, extremely successful branches, a long tail of mediumly successful branches, and a few poor performing branches.

Further analysis into what factors are most influential in the success of a branch, and to what extent sales amounts change based on the variation of these factors. They found that factors important to a successful sales team included the number of transactions, the number of sales representatives, and the number of unique parts to sell. Factors not important for revenue generation included the number of unique customers, and the proportion of negotiated discounts.

With these findings the team enabled the client to share best practices across the entire networks of branches to optimise sales. It also highlighted the need for further investigation into what other factors influence the sales of each branch, that may not be evident from the data.

11.3.3 Pricing engine

For the final part of the project, to build a pricing engine, the team targeted the 14% of car parts where price is the main reason a sale takes place. It is important to note that the price of each

part was not fixed, but the sales person always had some room to negotiate the final price with each client. Fig. 3 shows the range of sale price for a selection of parts, and clearly demonstrates the large range of prices that these items had been sold for.

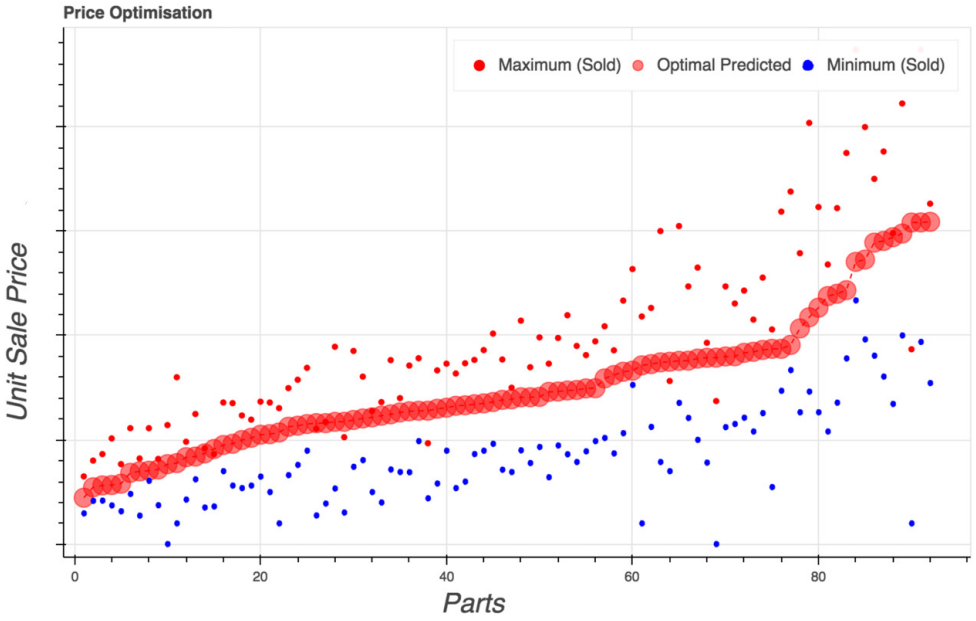


Figure 3: Actual sales prices for a selection of parts. For each part (x-axis), items had been sold for prices ranging from the highest (red dot), to the lowest (blue dot). The larger, red circle shows the optimal price, as calculated by the pricing engine.

It may be tempting to suggest that all parts should be sold at the maximum price, however, this presents a risk that some customers decline to buy. Hence, an optimal price is always somewhere between the maximum and minimum possible sale prices. The team looked at products that had the highest unit price and determined products for which sale probability was less likely correlated to price, and hence where giving customers a discount would not encourage more sales. The ML pricing engine was able to calculate which products should be discounted based on the probability of future demand. The optimal price is shown with large, red circles in Fig. 34.

Based on a test scenario, involving the two major groups of parts responsible for revenue generation, it was shown that implementation of this pricing engine, and a consistent sale at the optimal price, could increase revenues by up to 30%. As this revenue increase would come purely from an uplift in sales, and not from selling more items, this would translate directly to profit for the company, a potential extra profit of £6M over the next five years.

11.4 Conclusion

In just five weeks, a team of four junior data scientists delivered an analysis and pricing engine to the client that could deliver a potential 30% uplift in revenue. For a company of the client's size, this is an extraordinary result, and a true testament to the opportunities that lie within better use of data within mid-to large-size organisations. There are huge benefits to using sophisticated analysis techniques and machine-learning within these traditional businesses, and with the right partner, getting started can be easy. The project also shows the benefits of being agile in innovating with data; breaking a larger programme of change into smaller projects that can deliver immediate value and inform the way forward. This way, it is also possible to start a company's data science efforts with a minimal investment.

It is worth noting that not all projects are incredibly successful. Data science is called »science« for a reason, as it is research and development, and as such they are inherently risky. However, by working through these short, agile data science engagements, it is possible to de-risk the efforts and to course correct if a particular line of inquiry starts to look uncertain to deliver good results. Ultimately, data science can and will touch and transform every business, and those that start early will reap the greatest rewards.

11.5 List of Figures

12 Wie KI Produkte von Industrieunternehmen »smart« machen kann – am Beispiel Bosch Rexroth CytroPac

12 Wie KI Produkte von Industrieunternehmen »smart« machen kann – am Beispiel Bosch Rexroth CytroPac

Arian van Hülsen

12.1 Einleitung

Im Rahmen der digitalen Transformation nähert sich die digitale Welt zunehmend der physischen Welt an. Gerade für Unternehmen in der Fertigungsindustrie ist dies ein wichtiger Schritt in Richtung Produktinnovationen. Ergebnisse dieser Annäherung lassen sich in Konzepten wie z.B. dem digitalen Zwilling ausmachen. Möglich wird dies erst mit der Digitalisierung und neuer Technologie. So werden die Daten aus der physischen Welt durch IoT in die digitale Welt transportiert und im Gegenzug werden digitale Informationen mittels der erweiterten Realität (Augmented Reality, AR) zurück in die physische Welt übertragen.

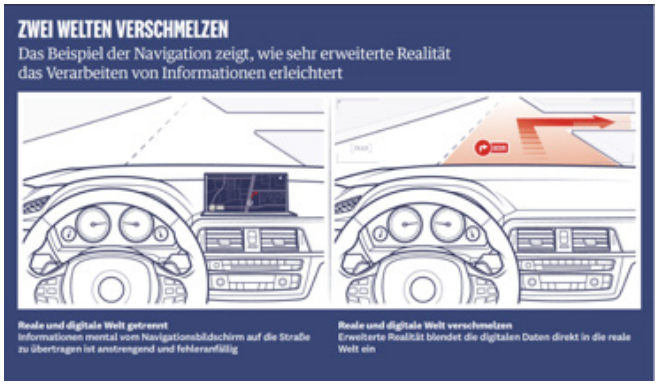


Abbildung 1: Einfaches Beispiel für die Annäherung der physischen und digitalen Welt durch IoT und Augmented Reality.

Das alles macht aber nur dann Sinn, wenn die Realität in der Werkshalle betrachtet wird. Wie passt IoT in den Produktlebenszyklus eines realen Produktes? Anhand von einem Produktbeispiel des PTC-Kunden Bosch Rexroth sehen wir, wie ein vernetztes Produkt die Grundlage für die eigene Überarbeitung bietet und dazu beiträgt, neue Marktsegmente zu definieren und auch dank AR-Technologie eine bessere Verbindung zwischen Hersteller und Kunden zu schaffen. Die AR- und die IoT-Technologie sorgen dafür, dass die Produkte »connected« sind. Durch die Konnektivität lässt sich bereits eine Vielzahl von wertschöpfenden Anwendungsfällen abbilden, wie z.B. das Remote Monitoring. Eine richtige Entfesselung von Wertschöpfung kann aber oftmals erst erfolgen, wenn die gesammelten Daten auch weiterverarbeitet bzw. synthetisiert werden. Man spricht davon, das Produkt »smart« oder intelligent zu machen. Ein sehr wichtiger Bestandteil dieser Intelligenz in den Produkten liegt hierbei in der Anwendung von Künstlicher Intelli-

genz, die dabei hilft, aus den Daten Erkenntnisse zu ziehen. Im Vergleich zu herkömmlichen Produkten ist der Funktionsumfang intelligenter, vernetzter Produkte höher. Mit ihren smarten Eigenschaften sprengen sie die traditionellen Produktgrenzen. Das Beispiel von Bosch Rexroth zeigt sehr eindrucksvoll, wie ein Hydraulikaggregat an verschiedenen Stellen des Produktlebenszyklus mit diesen Technologien Innovation hervorbringen kann.

12.2 Bosch Rexroth und das Hydraulikaggregat CytroPac

Bosch Rexroth setzt auf erweiterte Realität, IoT und KI, um das Design und die Fähigkeiten seines intelligenten, vernetzten Hydraulikaggregats CytroPac zu optimieren.

Hydraulik wird in zahlreichen industriellen Anwendungen eingesetzt und meist überall dort, wo viel und schnell verfügbare Leistung auf beschränktem Raum benötigt wird. Dazu gehören Anwendungen wie Werkzeugmaschinen, Mühlen und Baumaschinen. Hydraulischer Druck wird lokal durch Hydraulikeinheiten (HPU) generiert.

Abbildung 2: Hydraulikaggregat CytroPac

Bosch Rexroth, ein weltweit aktiver Anbieter von Antriebs- und Steuerungstechnik für die Fertigungsindustrie, hat vor kurzem den CytroPac auf den Markt gebracht. Dabei handelt es sich um ein netzwerkgestütztes »Plug & Run«-Hydraulikaggregat, das unter verschiedenen Einsatzbedingungen genutzt werden kann. Mit PTC Creo konnte Bosch Rexroth Motor, Pumpe, Tank, Kühlsystem und Steuerung in einer äußerst kompakten Bauweise verpacken. Des Weiteren wurde der CytroPac von vornherein als intelligentes, vernetztes Produkt entwickelt und mit Sensoren für Druck, Temperatur, Füllstand, Verschmutzungsgrad und Volumenstrom ausgestattet. Die für die Digitalisierung zuständige Abteilung bei Bosch Rexroth hat passend dazu eine App entwickelt, die sich per Plug-in mit der IoT-Plattform ThingWorx verbindet. Damit konnten die Techniker von Bosch Rexroth fortan nicht nur die Vitaldaten von sich im Kundeneinsatz befindlichen Geräten überwachen, sondern auch Nutzungsmuster und Möglichkeiten zur Optimierung erkennen. Daten und Werte, die vor der IoT-Konnektivität nicht vorlagen und höchstens über Simulationen erahnt werden konnten. Für die Korrelationsanalyse wurde hier auf das maschinelle Lernen (Machine Learning), eine Teildisziplin der Künstlichen Intelligenz, gesetzt. Zum Einsatz kam hierbei Thingworx Analytics, eine Lösung zum Automatisieren des maschinellen Lernens. Während in einem typischen KI-Projekt ein Team aus Data Scientists an den Daten-Features, den Algorithmen und der Ergebnisinterpretation arbeitet, hat Thingworx Analytics diesen komplexen Prozess mithilfe des maschinellen Lernens automatisieren können. Es wird also maschinelles Lernen eingesetzt, um maschinelles Lernen auf Daten anzuwenden. Konkret wird hierbei das Supervised Machine Learning Verfahren angewandt. Hierbei entsteht ein enormer Geschwindigkeitsvorteil, und dank der Automatisierung kann später auch die Skalierung von Analysemodellen auf tausende von Produkte im Feld gewährleistet werden.

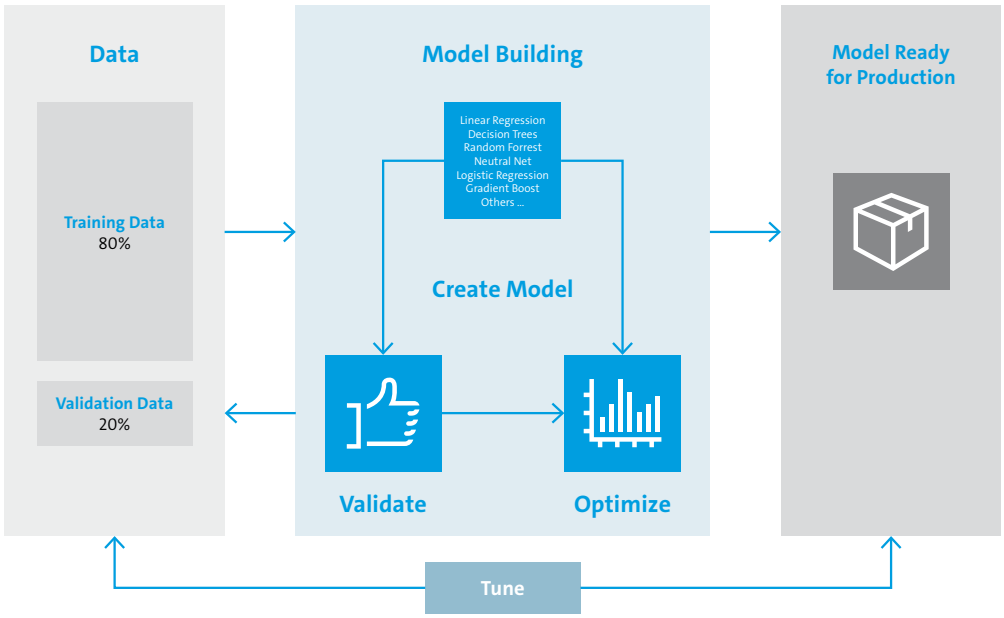


Abbildung 3: Automatisierter Machine-Learning-Prozess mit Thingworx Analytics

Anhand der nun vorliegenden Felddaten konnte eine bestimmte Gruppe an CytroPacs identifiziert werden, die am Rande der zulässigen Temperaturobergrenze arbeiteten. Die Entwickler waren zunächst überrascht, dass solch ein Nutzungsprofil außerhalb des Entwurfsparameterbereichs zu sehen war. Diese Betriebssituation könnte jedoch zu einem Problem für diese besondere Kundengruppe werden. Gleichzeitig bot sie Bosch Rexroth die Chance, den CytroPac zu optimieren, um in diesem potenziell neuen Marktsegment zuverlässig zu funktionieren. Die Frage lautete jedoch, wie die Konstruktion verändert werden muss, um diese neue Anforderung zu erfüllen.

12.3 Neue Erkenntnisse vom Feld erforderten Umdenken im Konstruktionsprozess

Die Ingenieure nutzen die Ergebnisse aus der diagnostischen Analyse der Felddaten und reicherten diese mit physikbasierten Simulationsdaten an. Hierfür kam die Software Simulink von Mathworks zum Einsatz. Damit standen zwei verschiedene Techniken zur Verfügung, um festzustellen, welche Verbesserungen am CytroPac notwendig waren, um dem neuen Nutzungsprofil zu entsprechen. Mithilfe des modellbasierten digitalen Zwillings modellierten sie, wie unterschiedliche Grade an Kühlungseffizienz die Leistung unter den jeweiligen Bedingungen beeinflussen. Beide Analysearten führten zu dem Resultat, dass die Effizienzsteigerung mindestens 30 Prozent betragen müsse, um eine adäquate Leistung zu erbringen.

Diese Anforderung war für die Ingenieure eine echte Herausforderung, schließlich mussten die Kühlungseffizienz derartig gesteigert und trotzdem die Größengrenzen eingehalten werden. Der CytroPac wird mithilfe einer Blechplatte mit drei gebogenen, wassertransportierenden Stahlrohren gekühlt, die eingegossen sind. Der Grad der Kühlung ist durch den Fertigungsprozess begrenzt. Werden die Rohre zu stark gebogen, brechen sie. Die Ingenieure führten eine thermische Analyse der Kühlplatte mithilfe von Creo Simulate durch und stellten fest, dass sie mit der bestehenden Konstruktion nicht in der Lage waren, mehr Hitze abzuführen. War es notwendig, das gesamte Aggregat zu überarbeiten?

Sie kamen zu dem Schluss, dass additiver 3D-Druck einige der Fertigungsprobleme adressieren könne. In der CytroPac-Einheit gab es drei wesentliche Wärmequellen. Die Ingenieure erstellten den Prototyp eines neuen Kühlungsweges, der alle abdeckte. Dies konnte mit keinem der bisherigen Fertigungsprozesse umgesetzt werden, 3D-Druck bot aber die Möglichkeit dazu. Die additive Fertigung ermöglichte darüber hinaus Gitterstrukturen, die man sonst nicht hätte gießen können. Die CAD-Software Creo bietet hier eine spezielle Funktion für die Entwicklung derartiger Gitter an, die die strukturelle Integrität auch bei einem deutlich niedrigeren Gewicht aufrechterhalten. Die Designmöglichkeiten können hierbei äußerst komplex ausfallen. Dieser Bereich im Rahmen der additiven Fertigung ist ein besonders gutes Einsatzgebiet für die Künstliche Intelligenz. Künftig kann mithHilfe von KI autonom ermittelt werden wie die ideale Gitterstruktur aussehen muss, um bestimmte Produkteigenschaften wie Gewicht oder Belastungspunkte zu erfüllen. Man spricht hier von dem Generative Design, welches den klassischen Produktdesignprozess revolutionieren wird.

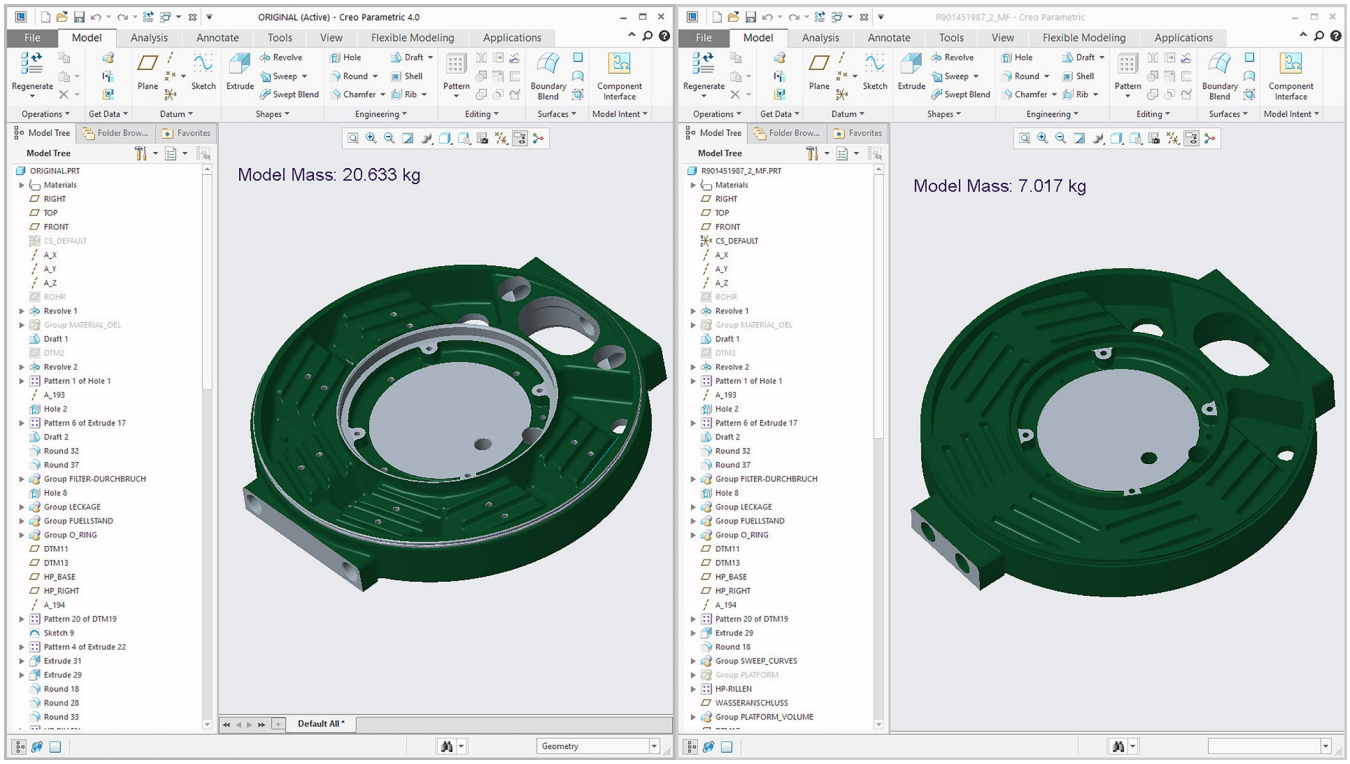


Abbildung 4: Design des Kühlungsweges vor & nach der Optimierung durch additive Fertigung

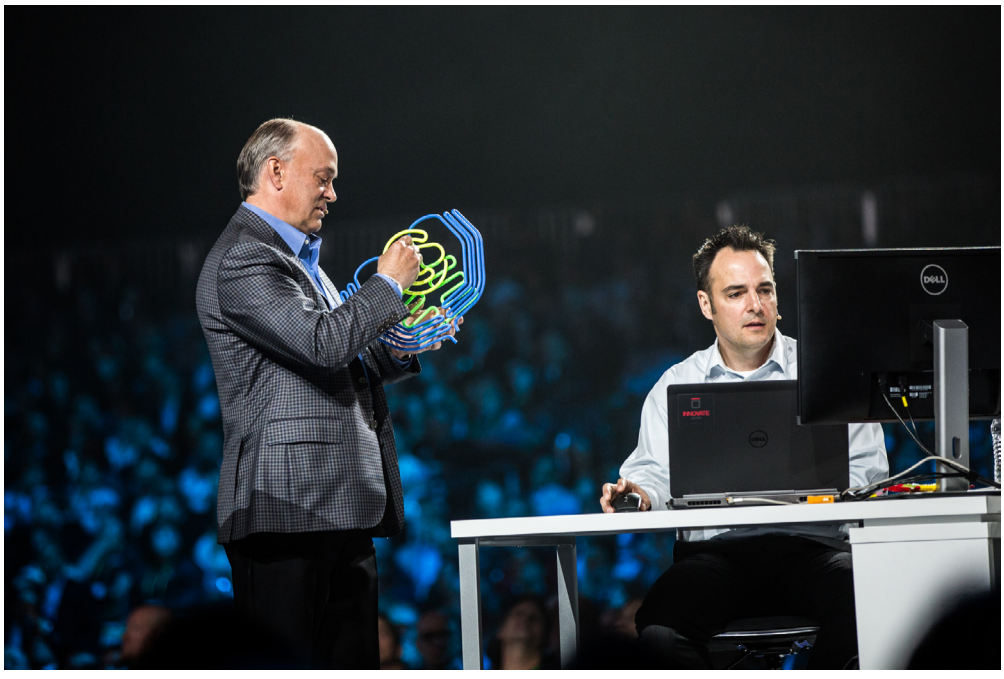


Abbildung 5: 3D-Druck des Kühlungsweges

Im Ergebnis für den Cytropac konnte eine vergleichende thermische Analyse zeigen, dass bei der Kühlungseffizienz eine Steigerung um 43 Prozent erreicht wurde, während bei der Kühleinheit gleichzeitig 66 Prozent an Gewicht eingespart wurden. Damit konnte Bosch Rexroth das neue Marktsegment adressieren.

12.4 Wie auch Vertrieb, Support und Service von IoT und AR profitieren

Die Produktentwicklung ist nicht der einzige Bereich, der von der Verschmelzung von physikalischer und digitaler Welt profitiert. Sie spielt ebenfalls eine Rolle für Vertrieb und Kundensupport. Zudem ist es Bestandteil der IoT-Strategie von Bosch Rexroth, die neuen technischen Möglichkeiten und die daraus gewonnenen Daten unternehmensweit zur Verfügung zu stellen und nicht »nur« die Produktentwicklung zu optimieren.

Ein Beispiel ist die »Sales Recommendation Engine«, eine Lösung zur Generierung von Empfehlungen für Kunden. Sie basiert auf den Daten aus dem Einsatz des CytroPacs sowie den Kunden- und Vertriebsdaten aus Salesforce und nutzt darüber hinaus das automatisierte maschinelle Lernen, um für das Vertriebsteam Up- und Cross-Selling-Optionen bereitzustellen. Wird zum Beispiel festgestellt, dass ein Gerät über mehrere Stunden im Einsatz ist und dabei im kritischen Temperaturbereich arbeitet, könnte ein größeres oder leistungsstärkeres Aggregat möglicherweise besser für die Bedürfnisse des Kunden passen. Werden einzelne Hydraulikaggregate beim Kunden ungewöhnlich lange eingesetzt, macht womöglich ein zusätzliches Gerät mehr Sinn für den Kunden. Mit diesem Wissen und proaktiven Angeboten sorgt das Vertriebsteam für einen besseren Kundenservice und mehr Kundenzufriedenheit. Ein weiteres Beispiel der Analyseergebnisse ist die Entscheidung, ob sich für einen Kunden ein Power-as-a-Service-Angebot eher lohnen könnte als der reine Kauf der Aggregate.

12.5 AR spielt zentrale Rolle beim Kundenkontakt

Die erweiterte Realität wird für die Anwendungsentwicklung bei Bosch Rexroth bereits fest mitbedacht. Aktuelle Beispiele sind eine virtuelle Kundenbroschüre sowie der Ersatz klassischer Bedienungsanleitungen. Creo Illustrate, die technische 3D-Illustrationssoftware von PTC, unterstützt die Erstellung schrittweiser, visueller Prozessbeschreibungen, die beispielsweise erläutern, wie der CytroPac konfiguriert werden kann oder wie eine CytroPac-Pumpe zu reparieren ist. Das Vuforia Studio kann diese Information ohne zusätzlichen Programmierbedarf zum Leben erwecken. So kann der Vertriebsmitarbeiter mittels AR-Brille einen potentiellen Kunden über die verschiedenen Modellvarianten und mögliche Einzelkonfigurationen informieren, die dieser Kunde in dem Moment direkt auf dem Tisch vor sich eingeblendet sieht. Ebenfalls kann sich ein Servicetechniker beim Kunden mit dem jeweiligen Gerät vor Ort verbinden und sich alle notwendigen Leistungsdaten oder Schritt-für-Schritt-Anleitungen für Wartung und Reparatur direkt auf das Gerät einblenden lassen.

Die Notwendigkeit, Anzeigen und Bildschirme im Produkt einzubauen, entfällt in immer mehr Fällen. Ein adaptierbares Dashboard auf dem Telefon oder Tablet oder zunehmend auch AR-Brillen können sämtliche notwendige Statusdaten des Gerätes anzeigen. Die Technologie ist mittlerweile soweit fortgeschritten, dass für diese aktuellen Anwendungen keine Tracker-QR-Codes mehr notwendig sind. Mittels »CAD-Tracking« bekommt der Anwender eine Art Suchanweisung nach einem bestimmten Gerät eingeblendet und sobald er seinen Blick mit der AR-Brille oder Smartphone und Tablet auf das Zielgerät richtet, wird es automatisch erkannt. So ist auch auf den CytroPacs von Bosch Rexroth keine Erkennungsmarke mehr sichtbar. Möglich macht dies erneut die Künstliche Intelligenz. Im Kern benötigt man für die Technologie der erweiterten Realität eine Vielzahl an Bild- und 3D-Objekterkennungsfunktionen. Das Antrainieren dieser Erkennungsfunktionen erfolgt mithilfe der Deep-Learning-Technologie, die wiederum eine Teildisziplin der Künstlichen Intelligenz ist. Hierbei werden tausende von Trainingsbildern in einem komplexen neuronalen Netz trainiert und für die AR-Erkennungsgeräte sowie die Software zur Verfügung gestellt. Dies erlaubt auch das Erkennen von komplexen 3D-Objekten in der physischen Welt, die lediglich über eine digitale 3D-CAD-Information antrainiert wurden.

12.6 Intelligent und vernetzt vom Anfang bis zum Ende

Das Beispiel CytroPac von Bosch Rexroth macht den gesamten Produktlebenszyklus eines intelligenten, vernetzten Produkts vollständig sichtbar, vom ersten Design über Produktion, Auslieferung, Nutzung, Wartung, Änderung und Austausch. Der Einsatz dieser Technologie muss aber nicht erst zwangsläufig in ausschließlich neuen Produkten erfolgen. Bereits im Feld aktive Maschinen und Geräte können ebenfalls mit einem Set an Sensoren, etwa für Temperatur oder Vibrationen, ausgestattet werden. Dabei sind wir auf diesem Feld erst am Anfang der Entwicklung. Es ist davon auszugehen, das vor allem IoT-Technologien zahlreiche weitere Anwendungsfälle zu Tage fördern werden, die dann mit den Möglichkeiten der erweiterten Realität verstärkt Einzug in unseren (visuellen) Alltag halten werden, sowohl für den Geschäfts- als auch den Privatbereich. Bestimmte Nutzungsmuster, wie wir sie heute kennen, wird es in einigen Jahren vielleicht nicht mehr geben, zum Beispiel seitenlange Bedienungsanleitungen, langweilige Produktbroschüren und -präsentationen oder der gewohnte Umgang mit der Haushaltselektronik.

12.7 Abbildungsverzeichnis

13 AI Meets Fintech: Aufbau einer Data Science Pipeline in einer stark regulierten Umgebung

13 AI Meets Fintech: Aufbau einer Data Science Pipeline in einer stark regulierten Umgebung

Olaf Hein

13.1 Motivation

Maschinelles lernen und Künstliche Intelligenz werden für den geschäftlichen Erfolg immer wichtiger. Mit der Hadoop-Plattform, der Anaconda-Distribution und Bibliotheken wie TensorFlow oder scikit-learn kann eine Data-Science-Umgebung sehr leicht aufgebaut werden. In stark regulierten Branchen, wie zum Beispiel der Finanzindustrie, sind vor dem produktiven Einsatz allerdings viele regulatorische und technologische Hürden zu überwinden. Dieser Artikel beschreibt, wie eine Data Science Plattform in einem solchen Umfeld aufgebaut werden kann. Die vorgestellte Lösung basiert auf unterschiedlichen Clustern mit unterschiedlichen Anforderungen an die Schutzziele für die einzelnen Stufen der Data Science Pipeline.

13.2 Flexibilität versus Sicherheit

Für Data Science und AI spielen Flexibilität und Agilität eine entscheidende Rolle. Zuerst einmal werden für Data Science Daten benötigt. Neben möglichst vielen und aktuellen Daten ist es auch wichtig, neue Datenquellen möglichst schnell bereitstellen zu können. Zusätzlich spielt die verwendete Software eine entscheidende Rolle. Es gibt immer wieder neue Bibliotheken und die existierenden entwickeln sich rasant weiter. Um neue Features nutzen zu können, müssen den Analysten die aktuellsten Versionen schnell bereitgestellt werden. Ein weiterer Punkt ist die verwendete Plattform. Dazu gehören zum Beispiel die verwendete Hadoop-Distribution und die Anaconda-Umgebung. Auch bei der Plattform ist es wichtig, dass neue Versionen schnell bereitgestellt werden. Neben Fehlerkorrekturen und neuen Features für die Analysten enthalten neue Versionen auch oft Verbesserungen für den Betrieb und die Sicherheit der Systeme.

Den Anforderungen an die Flexibilität stehen die Anforderungen der Informationssicherheit gegenüber. Für Finanzdienstleister sind unter anderem die Vorgaben der BaFin verpflichtend. Insbesondere die Mindestanforderungen an das Risikomanagement (MaRisk) [1] und die Bankaufsichtlichen Anforderungen an die IT (BAIT) [2] müssen unternehmensweit eingehalten werden. Neben den gesetzlichen Regelungen sind zusätzlich die internen Sicherheitsrichtlinien bindend. Üblicherweise werden die Schutzziele Vertraulichkeit, Integrität und Verfügbarkeit betrachtet. Dabei ist zu beachten, dass diese Ziele nicht nur für die Daten, sondern auch für die verwendeten bzw. selbst erstellten Programme und die genutzte Hardware-Infrastruktur gelten.

Die Data Science Pipeline

Am Anfang der Data Science Pipeline steht die Analyse der Daten und die Entwicklung eines Modells. Dies wird von einem Data-Science-Team durchgeführt (siehe Abbildung 1). Typische Anwendungsfälle aus der Finanzindustrie sind zum Beispiel Betrugsfallerkennung, Umsatzkategorisierung oder Risikobewertung. Am Ende wird das trainierte Modell vom Betriebsteam in der Produktion installiert und betrieben. Die Erkenntnisse aus der Produktion werden in der nächsten Iteration vom Data-Science-Team genutzt, um das Modell zu verbessern. Vor der Nutzung eines Modells in der Produktionsumgebung muss es in existierende Systeme integriert werden. Weiterhin sind Tests der Funktionalität und der Stabilität notwendig. Erst nach erfolgreichen Abnahmetests darf das Modell für die Produktion freigegeben werden. Dieser Schritt wird im Folgenden als Modellhärtung bezeichnet und wird maßgeblich vom Entwicklungsteam durchgeführt.

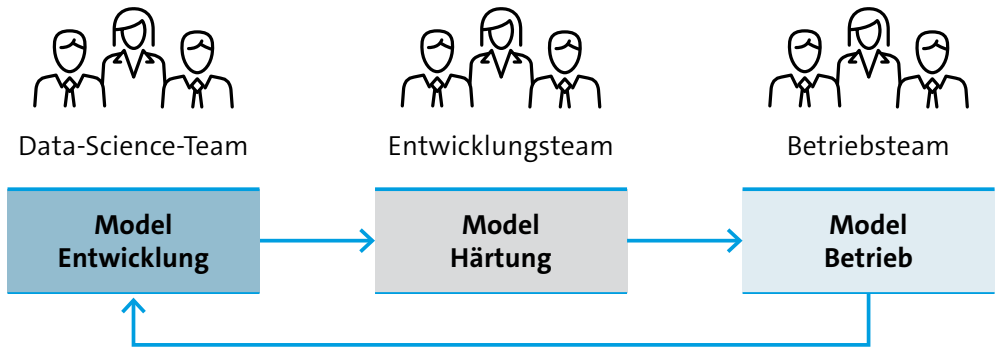


Abbildung 1: Die Data Science Pipeline

Cluster-Trennung

Wie bereits beschrieben, sind unterschiedliche Teams für die unterschiedlichen Stufen der Data Science Pipeline verantwortlich. Die Mitarbeiter in diesen Teams haben unterschiedliche Zugriffsrechte für die Daten und die verwendeten Computersysteme. Weiterhin unterscheiden sich die einzelnen Stufen bezüglich der Anforderungen an die Flexibilität und die Sicherheit. Für den Betrieb des Modells gelten hier sehr hohe Anforderungen an die einzelnen Schutzziele. Um diese zu erreichen, muss auf eine hohe Flexibilität verzichtet werden. Weiterhin ist es notwendig, die Produktivumgebung von anderen Umgebungen zu trennen. Für Finanzdienstleister wird dies in MaRisk AT 7.2.3 explizit gefordert. Um diese Anforderungen zu erfüllen, werden für die einzelnen Stufen der Data Science Pipeline getrennte physikalische Cluster aufgebaut. Diese Cluster unterscheiden sich durch eine unterschiedliche Erfüllung der einzelnen Schutzziele (siehe Abbildung 2) und durch ihre physische Zuordnung zur Test- oder Produktionsumgebung.

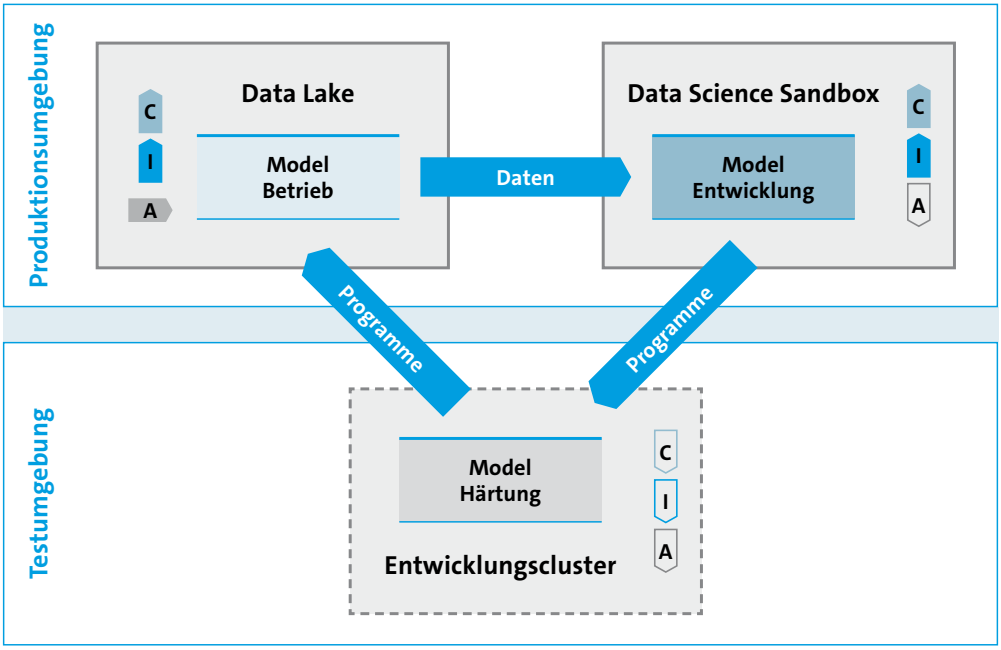


Abbildung 2: Umgebungen & Cluster

Data Science Sandbox

Für die Modellentwicklung wird eine Data Science Sandbox genutzt. In diesem Cluster wird mit schützenswerten Daten gearbeitet. Das können personenbezogene Daten sein. Aber auch für andere Daten, wie zum Beispiel Forschungsergebnisse, kann der Schutzbedarf hoch sein. Durch Anonymisierung oder Pseudonymisierung der Daten kann der Schutzbedarf reduziert werden. Allerdings sind anonymisierte Daten für die Analyse und das Trainieren von Modellen oft nutzlos. Ein Beispiel ist die Kategorisierung von Umsatzdaten, um ein digitales Haushaltsbuch zu erstellen. Bei einer einfachen Anonymisierung könnte die IBAN maskiert werden. Damit ist der direkte Bezug zu einem Konto und somit zu einer Person nicht mehr möglich. Durch die Analyse des Verwendungszweckes ist der Personenbezug in den meisten Fällen aber wieder herstellbar. Wird der Verwendungszweck zusätzlich maskiert, dann sind die Daten für diesen Anwendungsfall unbrauchbar, da der Verwendungszweck für eine korrekte Kategorisierung zwingend notwendig ist.

Neben den Daten müssen auch die entwickelten Algorithmen und Modelle vor unberechtigtem Zugriff geschützt werden. Somit gelten für die Sandbox hohe Anforderungen an die Vertraulichkeit. Ähnlich hohe Anforderungen gelten für die Integrität, denn ein Modell, das mit verfälschten Daten trainiert wurde, kann in der Praxis zu fehlerhaften Ergebnissen führen. Um die Flexibilität zu gewährleisten, wird hier ganz bewusst auf eine hohe Verfügbarkeit verzichtet. Dadurch können zum Beispiel neue Bibliotheken unkompliziert installiert und erprobt werden. Das Risiko, die Stabilität und damit die Verfügbarkeit negativ zu beeinflussen, wird bewusst in Kauf genommen. Die Alternative wären aufwändige und langwierige Tests und Freigabeprozesse für alle Änderungen am Cluster.

13.3 Entwicklungscluster

Der Entwicklungscluster zeichnet sich durch sehr hohe Anforderungen an die Flexibilität aus. Dies hat zur Folge, dass die Schutzziele Vertraulichkeit, Integrität und Verfügbarkeit nur zu einem geringen Teil erfüllt werden können. In der Konsequenz dürfen im Entwicklungscluster keine schützenswerten Daten verarbeitet werden. Das Entwicklungsteam arbeitet mit anonymisierten oder synthetisch generierten Testdaten. In der Praxis haben sich anonymisierte Daten bewährt. Insbesondere bei komplexen Zusammenhängen zwischen einzelnen Datensätzen ist es sehr schwierig, aussagekräftige Testdaten zu erstellen. Synthetisch generierte Daten werden zusätzlich genutzt und sind, zum Beispiel für den Test von Grenzfällen, hilfreich. Die Integrität der entwickelten Software wird über abgesicherte externe Systeme, die Teil der Continuous Integration/Deployment Pipeline sind, sichergestellt.

Data Lake

Der Data Lake ist die Quelle für die Rohdaten des Data-Science-Teams. Weiterhin werden im Data Lake die trainierten Modelle ausgeführt. Hier gelten sehr hohe Anforderungen an die drei Schutzziele. Im Gegensatz zu den anderen Clustern muss hier eine hohe Verfügbarkeit garantiert werden. Das geht nur, wenn ganz bewusst auf Flexibilität verzichtet wird. Alle relevanten Änderungen am Data Lake müssen vorab getestet und abgenommen werden (Vergleiche auch BAIT 6.41).

Umgebungen & Netztrennung

In der Praxis werden IT-Systeme in Netzwerkumgebungen aufgebaut und betrieben. Dabei wird mindestens zwischen einer Produktions- und einer Testumgebung unterschieden. Zwischen diesen Netzen werden Firewalls eingesetzt. Diese verhindern, dass Daten oder Programme unkontrolliert von einem Netz in das andere übertragen werden. Durch die Netztrennung kann das Produktionsnetz als sehr sicher betrachtet werden. Dort sind viele Verfahren für die Einhaltung und Überwachung der Sicherheit etabliert. Ein konkretes Beispiel dafür ist ein Active Directory (AD) für die Authentifizierung und Autorisierung von Benutzern. In einer Testumgebung gibt es auch ein Active Directory. Allerdings wird dieses üblicherweise nicht besonders abgesichert. Somit können dort zwar funktionale Tests durchgeführt werden. Die Vertraulichkeit kann in einer Testumgebung aber nicht gewährleistet werden. Sowohl der Data Lake als auch die Data Science Sandbox haben hohe Anforderungen an die Vertraulichkeit und die Integrität. Daher müssen beide Systeme in der Produktionsumgebung aufgebaut werden. Der Entwicklungscluster wird dagegen in der Testumgebung installiert, um eine maximale Flexibilität zu bieten.

13.4 Weitere Cluster

Mit den drei beschrieben Clustern kann eine Data Science Pipeline auch in einem stark regulierten Umfeld aufgebaut und betrieben werden. Einige wichtige Punkte wurden bisher noch nicht betrachtet. Um die zusätzlichen Anforderungen zu erfüllen, kann es notwendig sein, zusätzliche Cluster aufzubauen.

Backup & Recovery

Ohne ein Backup der Daten kann die Verfügbarkeit eines Systems allenfalls als ›mittel‹ eingestuft werden. Die Hadoop-Mechanismen, wie zum Beispiel die 3-fach Replikation, schützen vor vielen Defekten. Aber insbesondere Fehler, bei denen Daten unbeabsichtigt verändert oder gelöscht werden, können zu einem Datenverlust führen. In diesem Fall hilft ein Backup. Dies kann zum Beispiel durch die Replikation der relevanten Daten in einen weiteren Cluster in der Produktionsumgebung erfolgen.

Plattformentwicklung

Ein weiterer wichtiger Punkt ist die Weiterentwicklung der Plattform selbst. Hierfür ist üblicherweise das Betriebsteam verantwortlich. Bevor eine neue Hadoop-Version produktiv genutzt werden kann, muss diese in jedem Fall getestet werden. Im bisherigen System kann der Entwicklungscluster für diese Tests verwendet werden. Das hat aber zur Konsequenz, dass das Entwicklungsteam nicht mehr ungestört arbeiten kann. Bei kleineren Teams ist das oft noch möglich. In einem großen Umfeld oder aber bei aufwändigen Migrationen wird in der Testumgebung ein separater Cluster für die Plattformentwicklung benötigt.

Wartung

Bei der Weiterentwicklung der Plattform wird üblicherweise so verfahren, dass für die Entwicklung am nächsten Release die Data Science Sandbox und der Entwicklungscluster aktualisiert werden. Sollte jetzt ein schwerwiegendes Problem in Produktion auftreten, dann existiert kein System mehr, das dem Versionsstand des Produktionssystems entspricht. Das kann dazu führen, dass der Fehler nicht reproduzierbar ist. Weiterhin ist ein Test des korrigierten Modells nur eingeschränkt möglich. Zuverlässig lässt sich dieses Problem nur durch einen Wartungscluster mit einem produktionsidentischem Softwarestand lösen. Wenn Tests mit Produktionsdaten durchgeführt werden sollen, dann muss dieser Cluster in der Produktionsumgebung aufgebaut werden, andernfalls kann er auch in der Testumgebung stehen.

13.5 Fazit

Die unterschiedlichen Anforderungen bezüglich Sicherheit und Flexibilität können sehr elegant durch den Einsatz mehrerer unterschiedlicher Cluster umgesetzt werden. Dabei unterscheiden sich die Cluster insbesondere durch die Schutzziele, die erfüllt werden müssen. Für die initiale Planung ist es sehr wichtig, die Netzwerkinfrastruktur zu berücksichtigen. Insbesondere das Schutzziel Vertraulichkeit kann in einer Testumgebung üblicherweise nicht erfüllt werden. Daher gehören die Data Science Sandbox und der Data Lake in jedem Fall in ein abgesichertes Produktionsnetzwerk. Im Gegensatz zum Data Lake wird in der Sandbox aber ganz bewusst auf eine hohe Verfügbarkeit verzichtet. Im Gegenzug erhalten die Analysten eine hohe Flexibilität und viele Freiheiten auf dem System. Weitere Cluster können sehr schnell notwendig werden. Um die Kosten für den Aufbau und Betrieb mehrerer Cluster zu reduzieren, haben sich Virtualisierung der Hardware und Automatisierung der Installation und Konfiguration in der Praxis bewährt.

13.6 Abbildungs- und Literaturverzeichnis

[1] BaFin, Mindestanforderungen an das Risikomanagement (MaRisk)
↗ **https://www.bundesbank.de/de/aufgaben/bankenaufsicht/einzelaspekte/risikomanagement/marisk**

[2] BaFin, Bankaufsichtliche Anforderungen an die IT (BAIT)
↗ **https://www.bundesbank.de/de/aufgaben/bankenaufsicht/einzelaspekte/risikomanagement/bait/bankaufsichtliche-anforderungen-an-die-it-598580**

Bitkom vertritt mehr als 2.700 Unternehmen der digitalen Wirtschaft, davon gut 1.900 Direktmitglieder. Sie erzielen allein mit IT- und Telekommunikationsleistungen jährlich Umsätze von 190 Milliarden Euro, darunter Exporte in Höhe von 50 Milliarden Euro. Die Bitkom-Mitglieder beschäftigen in Deutschland mehr als 2 Millionen Mitarbeiterinnen und Mitarbeiter. Zu den Mitgliedern zählen mehr als 1.000 Mittelständler, über 500 Startups und nahezu alle Global Player. Sie bieten Software, IT-Services, Telekommunikations- oder Internetdienste an, stellen Geräte und Bauteile her, sind im Bereich der digitalen Medien tätig oder in anderer Weise Teil der digitalen Wirtschaft. 80 Prozent der Unternehmen haben ihren Hauptsitz in Deutschland, jeweils 8 Prozent kommen aus Europa und den USA, 4 Prozent aus anderen Regionen. Bitkom fördert und treibt die digitale Transformation der deutschen Wirtschaft und setzt sich für eine breite gesellschaftliche Teilhabe an den digitalen Entwicklungen ein. Ziel ist es, Deutschland zu einem weltweit führenden Digitalstandort zu machen.

Bundesverband Informationswirtschaft,
Telekommunikation und neue Medien e.V.

Albrechtstraße 10
10117 Berlin
T 030 27576-0
F 030 27576-400
bitkom@bitkom.org
www.bitkom.org

bitkom