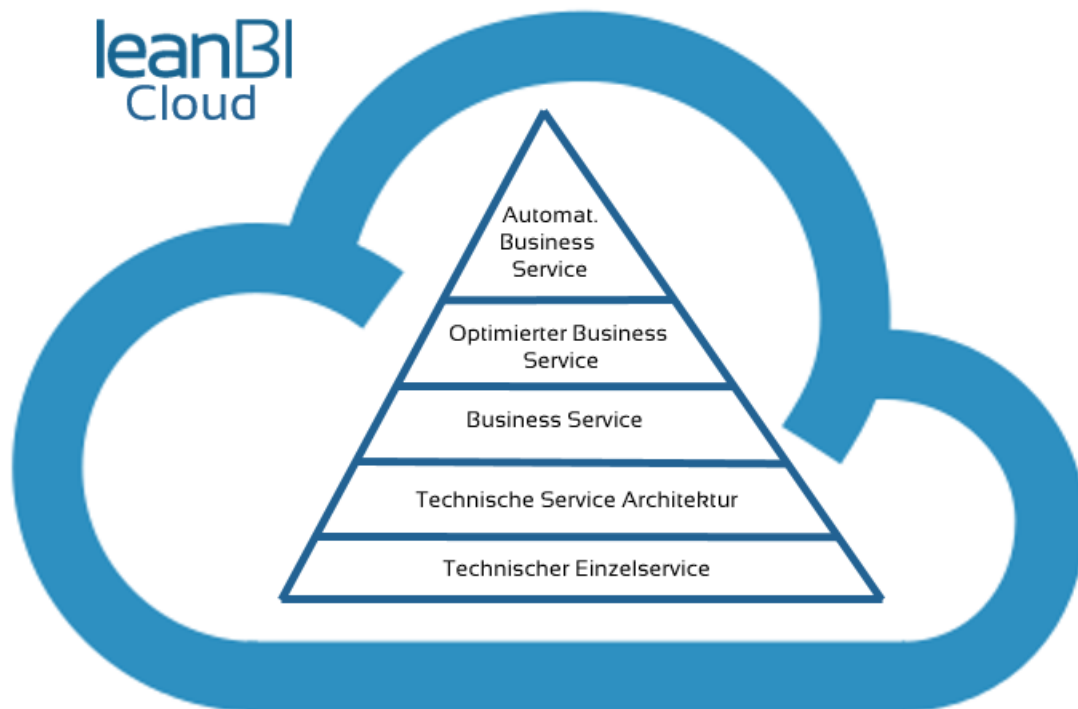


White Paper  
Service Oriented Analytic Cloud in der Industry 4.0



## EINLEITUNG

In diesem White Paper zeigen wir die Zusammenhänge von Analytics 3.0 und Industry 4.0 auf. Der Weg einer Industry 4.0 führt notgedrungen über die Datenanalytik. Dabei spielt „Machine Learning“ eine essentielle Rolle. Um das komplexe Gebiet des Machine Learning in eine Industry 4.0 einzubringen wird ein serviceorientierte Ansatz vorgeschlagen und besprochen.

### 1. Analytics 3.0: Präskriptive Analytik

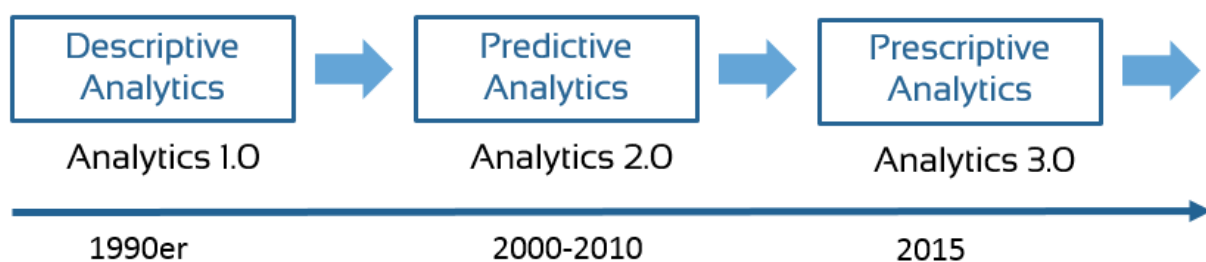


Abbildung 1: Die Entwicklung der Analytik nach Davensport.

**Analytics 3.0** gilt als Folgekonzept nach Big Data. Entwickelt wurde es in den USA von Prof. Thomas Davensport. Man fragt nicht nur „**Was** wird geschehen?“, im Sinne der prädiktiven Analytik, sondern auch „**Warum** wird es geschehen?“, im Sinne der präskriptiven Analytik. Aus „Was“ und „Warum“ ergeben sich neue Wirkungszusammenhänge und Insights in der Ursachen-Wirkungskette. Diese erweiterte Fragestellung ist nicht einfach in der Ausführung. Denn viele Prognoseprobleme basieren auf einer grossen Anzahl von variierenden Inputvariablen. Diverse Algorithmen analysieren Prognosewahrscheinlichkeiten, ohne die eigentliche Wirkungszusammenhänge wirklich erkennbar zu machen. Dafür benötigen wir ganz neue Algorithmen.

Analytik 3.0 beinhaltet neben konventioneller Business Intelligence (BI) Technik, neue Big Data Technologien, die das Streamen grosser Quantitäten von Live-Daten in unterschiedlichsten Datenformate ermöglicht. Die Datenauswertung erfolgt über verteilte Infrastrukturen und In Memory Technologien mit Hilfe von Machine Learning und Data Mining Algorithmen. Im Gegensatz zum konventionellen Data Warehousing bestehen keine Grenzen bezüglich Datenformaten, jegliche lesbare Maschinendaten können verwendet werden, und die Datenmodellierung vereinfacht sich drastisch. Im Zentrum von Analytic 3.0 steht der **Algorithmus**. **Machine Learning** Algorithmen sollen Informationen automatisch aus Daten extrahieren. Dies geschieht ohne Maschine-Mensch Schnittstelle. Um dies zu ermöglichen, werden die Algorithmen mit kleineren Datenmengen (von Menschen) antrainiert und Modelle gebildet. Teilweise sind die Algorithmen auch selbstlernend, d.h. im Laufe der Zeit verbessern sich die Modelle und damit auch die Vorhersagen.

Auch das eigentlich **Data Mining** hat innerhalb der Analytics 3.0 einen wichtigen Stellenwert. Hier steht jedoch immer eine konkrete Person im Entdeckungs- oder Vorhersageprozess. Typischerweise steht die Lösung eines konkreten komplexen Problems im Vordergrund. Man möchte zum Beispiel mit einer Mustererkennung Verständnis über einen komplexen Sachverhalt erhalten, auf den eine grosse Anzahl unbekannter Einflussfaktoren wirken. Data Mining bedient sich vieler Machine Learning Algorithmen und umgekehrt.

Analytics 3.0 ist ein Zusammenspiel von Technologie und Mathematik. Es ist Realität und Zukunft zugleich. Analytics 3.0 wird seit vielen Jahren angewendet, gleichzeitig wird in vielen Universitäten und Unternehmen intensiv geforscht und entwickelt.

Die Anzahl gebrauchsfähiger Algorithmen ist bereits heute schon sehr gross und ist einem starken Änderungsprozess unterlaufen. Viele neue Algorithmen kommen hinzu und bestehende werden laufend verbessert. Ein Grossteil der Algorithmen ist öffentlich und kann über verschiedene Open Source Pakete wie R, Spark (MLlib), Mahout, Weka, usw. bezogen werden. Auch käufliche Produkte mit einer Auswahl von gekapselten Algorithmen sind erhältlich, und reduzieren teilweise die Einstiegshürden zu Analytics 3.0.

Damit die Algorithmen auf den Big Data Technologie optimal funktionieren (massiv parallele Berechnung auf verteilten Rechnern, wird weitere Software benötigt. Auch hier gibt es wiederum Open Source Möglichkeiten oder käufliche Softwareprodukte, die stetig weiterentwickelt werden.

Eins ist klar: Die Möglichkeiten der prädiktiven und präskriptiven Analytik sind noch lange nicht ausgeschöpft.

## 2. Industry 4.0: Die Informatisierung der Industrie

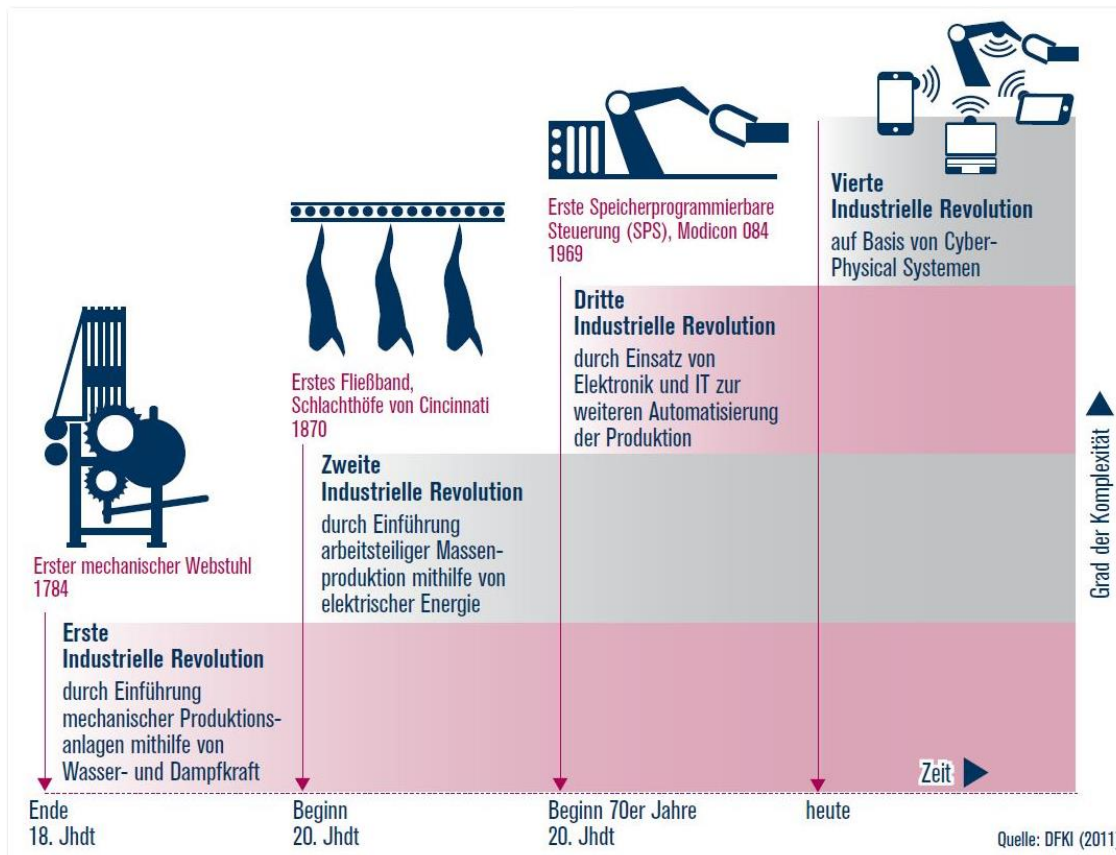


Abbildung 2: Die Entwicklung der Industrie nach „Deutsches Forschungszentrum für Künstliche Intelligenz“ (DFKI)

**Industrie 4.0** als Deutsch/Europäisches Projekt seit 2011 beinhaltet die Informatisierung der Fertigungstechnik. Die für Industrie 4.0 notwendige Automatisierungstechnik soll durch die Einführung von Verfahren der Selbstoptimierung, Selbstkonfiguration, Selbstdiagnose und Kognition intelligenter werden und die Menschen bei ihrer zunehmend komplexen Arbeit besser unterstützen. Damit entsteht eine intelligente Fabrik (**Smart Factory**), die wandlungsfähig und ressourceneffizient ist, und sich optimal in die Geschäftsprozesse einer Unternehmung eingliedert.

Ziel ist Industrieprodukte über Varianten anpassbar zu machen (bis hin zur Losgröße 1), gleichzeitig die Bedingungen einer Grossserienproduktion anzuwenden.

Die Gedanken der Industrie 4.0 sind allgegenwärtig in der Schweizer Industrie, nicht nur in der Fertigungstechnik. Der Reifegrad der Schweizer Industrien bezüglich Industrie 4.0 ist sehr unterschiedlich. Einige Vorreiter führen bereits Fernwartungssysteme, sodass irgendwo auf der Welt installierte Maschinen Daten an die Herstellerfirma liefern und so z.B. Wartungsprozesse rechtzeitig in Gang gesetzt werden können. Andere Betreiber vernetzen ihre über viele Standorte verteilten Anlagen oder ganze Fabriken so zentral, dass ihre Daten gemeinsam auswertbar sind. Aber auch diese Umsetzungen sind nur

erste Schritte in eine Welt der Industry 4.0, denn das „Selbst“, also die logische und physikalische Vernetzung der Maschinen ist erst sehr punktuell am Wachsen. Ein wichtiges Element der Industry 4.0 ist die Entwicklung der Sensortechnik selbst. Beispielsweise liefern die Sensoren im Bereich „Machine Vision“ - also das Gebiet der Bilderfassung, auch im Wellenlängenbereich von IR und Röntgen - vollkommen neue Möglichkeiten der zerstörungsfreien online Qualitätsmessungen, sind aber gleichzeitig sehr datenintensiv. Auch die Spektroskopie wird immer stärker direkt in die Prozesse eingebunden und liefert sehr grosse Datenmengen. Überhaupt zeichnen sich moderne Sensortechniken durch immer grössere Datenströme aus, die zu bewältigen sind. Aus unserer Sicht wird der Analytik innerhalb der Industry 4.0 noch zu wenig Aufmerksamkeit geschenkt. Analytics 3.0 und Industry 4.0 sind zu stark getrennte Welten. Warum? Beide Welten sind komplex und erst teilweise beherrschbar. Die Schnittmenge wird gross sein, nur fehlen heute die Kompetenzen, um beide Welten zu vereinen.

### 3. Analytic Industry 4.0: Zweck

Wenn wir nun die Schnittmenge beider Welten betrachten, dann bezeichnen wir dies als **Analytic Industry 4.0**. Um diese Welten zusammenzubringen wird die **Cloud** von grosser Wichtigkeit werden, denn schliesslich geht es um die Vernetzung von Daten an einem zentralen Ort. Analytic Industry 4.0 ist ein Teilgebiet der Industry 4.0 mit der Konzentration auf den analytischen Teil dieser 4. Industriellen Revolution. Zu welchem Zweck ein solches Teilgebiet? Gehen wir dafür zurück auf die Definition von Industry 4.0 und erkennen die Wichtigkeit der Analytik:

- **Selbstdiagnose / Selbstprognose:** Die Selbstdiagnose hat den Zweck frühzeitig mögliche Maschinenausfälle zu detektieren. Dies geht über die Benachrichtigung von Schwellwertinformationen weit hinaus. Erst mit der Kombination von Messdaten, deren algorithmischer Bearbeitung und Rückführung der daraus abgeleiteten Informationen in den Produktionsprozess zur weiteren physikalischen Bearbeitung, kann eine qualitativ hochwertige Selbstdiagnose erfolgen.
- **Selbstoptimierung:** Die Selbstoptimierung des **Fertigungsprozesses** ist neben der physikalischen Maschinentätigkeit ein mathematischer Optimierungsprozess, der auf Daten basiert. Dahinter stecken nichts anderes als die beschriebenen Algorithmen aus dem Analytics 3.0.  
Die Selbstoptimierung hat zwei Aspekte. Einerseits geht es um die Optimierung des Fertigungsprozesses selber. Daneben steht aber auch das Fertigungsprodukt im Fokus. Eine Selbstoptimierung des Fertigungsproduktes kann man auch als automatisierte Qualitätsoptimierung betrachten. Dafür bedarf es automatisierter Qualitätsmessungen, aus denen grosse Datenmengen entstehen, die zu verarbeiten sind. Damit hier

rechtzeitig Informationen in den Produktionsprozess zurückfließen können, sind grosse und performante analytische Infrastrukturen notwendig.

- **Kognition:** Die Kognition ist die Gesamtheit der geistigen Aktivitäten im Zusammenhang mit Denken, Wissen, Erkennen und Kommunizieren. Mit Analytics 3.0 können Teile der Kognition von Maschinen übernommen werden. Beispielsweise werden Ähnlichkeiten und Strukturen in grosse Datenpools der Industrie mittels Algorithmen übersichtlich dargestellt. Ingenieure gelangen dadurch schneller zu neuen und relevanten Erkenntnissen.

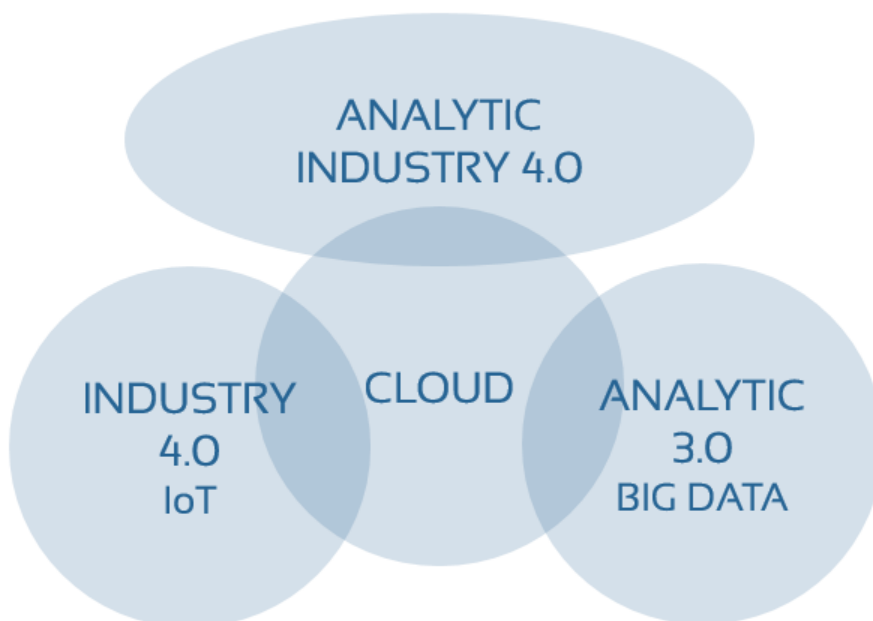


Abbildung 3: Das Zusammenspiel der Kräfte

Es geht also darum, Analytics 3.0 auf die laufende 4. Industrielle Revolution auszurichten. Betroffen ist nicht nur die Fertigungstechnik, sondern auch die Lagertechnik, Verfahrenstechnik, Klimatechnik und Energietechnik. Sowohl die Daten-Infrastrukturen, die Algorithmen, die Tools sind Spezifika dieser Industrien und müssen erarbeitet sein. Eine grosse Rolle spielt aus unserer Sicht die Open Source Bewegung. Wir sind überzeugt, dass über die bestehenden und neuen Open Source Projekte im analytischen und Big Data Bereich das Ziel einer Industry 4.0 nicht nur am schnellsten erreicht wird, sondern auch die beste Kontinuität gewährleistet. Die Open Source Initiativen generieren neue Produkte für die Analytic Industry 4.0.

#### 4. Aktuelle Defizite Analytic Industry

In der Industrie werden seit vielen Jahren Automation und MES (Manufacturing Execution Systeme) eingesetzt. Beide Systeme sammeln Daten von Sensoren, um diese zu verarbeiten.

Automation Lösungen steuern und regeln die Maschinen, Anlagen und Fabriken und überwachen deren Daten. Automation Systeme sind häufig dezentral an den Maschinen und nur selten über ganze Anlagen und Fabriken zentralisiert. Eine Gesamtsicht auf die Anlagen fehlt somit. Bei Anlagen, die jedoch eine solche Zentralisierung der Automation aufweisen, fehlt häufig der prognostische oder präskripte Ansatz. Machine Learning Ansätze in Automation Systemen ist äusserst selten und rudimentär.

Daneben stehen die sogenannten MES (Manufacturing Execution Systeme). Die MES stellen einheitliche Toolsets rund um die Optimierung der Fertigungsprozesse zur Verfügung. Auch hier werden Daten gesammelt, ausgewertet und in den Produktionsprozess zurückgeführt. Die MES beinhalten auch analytische Auswertemöglichkeiten, wobei der Automatisierungscharakter bis heute noch recht eingeschränkt ist. MES hat aber über Industry 4.0 einen neuen Aufwind erhalten. Eine gesamthafte Betrachtung der Daten fehlt auch in diesen Tools, damit auch der Gebrauch von Machine Learning Elementen.

Der Ausbau der Industrien in Richtung Industry 4.0 ist und bleibt eine Kosten/Nutzen-Angelegenheit. Industry 4.0 wird sich erst durchsetzen, wenn die Grundtechnologien in günstiger Form zur Verfügung stehen. Bestehende Anlagen werden dann stetig erweitert oder ausgetauscht. Dieser Prozess wird noch viele Jahre dauern.

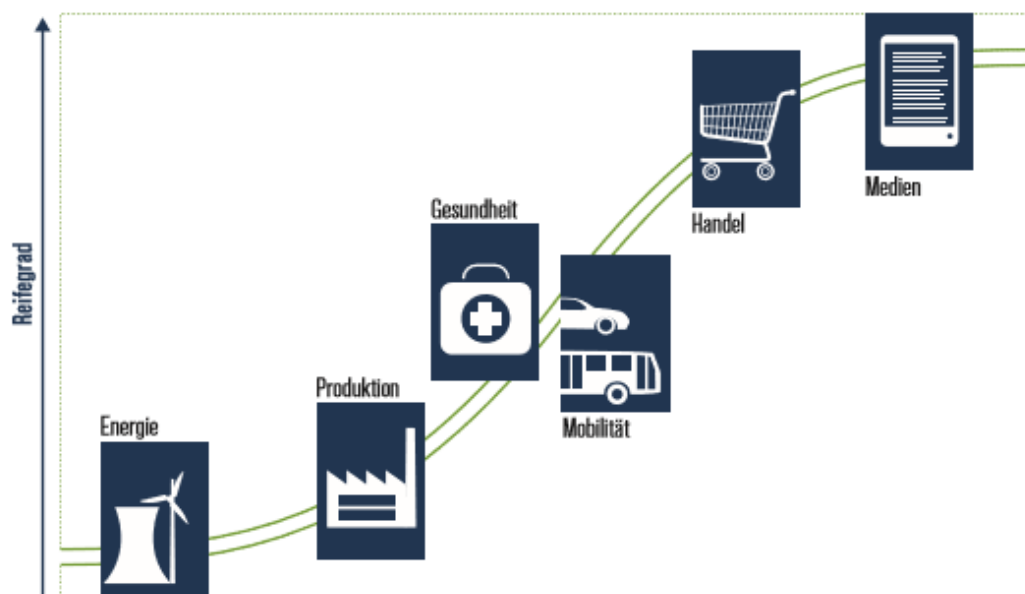


Abbildung 4: Reifegrad digitaler Geschäftsmodelle. Deutsche Akademie der Technikwissenschaften, Smart Service World, März 2014

Betrachtet man Abbildung 4, dann kann man verstehen, warum der Reifegrad in den Branchen Energie und Produktion im Vergleich zu anderen Branchen noch sehr tief ist. Was müssen wir tun, um in diesen Branchen den Ausbau voranzutreiben?



## 5. Service orientierte Architektur für die Industry 4.0

Das IT-Konzept der Service orientierten Architektur (SOA) stammt aus den Anfängen dieses Jahrhunderts und beinhaltet ein Informatik Framework, auf welchem Applikationskomponenten Services über standardisierte Schnittstellen austauschen. Es basiert auf dem Grundprinzip, dass Services wiederverwendbar sind und in verschiedenen Applikationen eingesetzt werden können. Übergeordnet sind die Services entlang eines Geschäftsprozesses (auch Produktionsprozess) angeordnet. SOA ist heute fester Bestandteil der Informatik. SOA ist ein umgesetztes Konzept, wie ein vernetztes System modular aufgebaut wird.

Industry 4.0 verbindet die physische Welt der Produktgenerierung mit dem Netzwerk der Informationswelt. In einer Smart Factory sind die Maschinenelemente miteinander vernetzt und in ständiger Kommunikation. Die Maschinenelemente können sich physisch über verschiedene Standorte verteilen, entsprechend der Wertschöpfungskette eines Produktes. Eine solche Vernetzung zeigt das folgende Bild.



Abbildung 5: Die Vernetzung in der Industry 4.0 (Quelle: emedia Juni 2014)

Dieses Bild suggeriert eine wilde Vernetzung der Maschinenteile, Maschinen, Anlagen und Menschen. Genau dies wäre für eine Industry 4.0 sehr kontraproduktiv. Es wird unsere grosse Aufgabe sein, eine serviceorientierte Architektur innerhalb der Industry 4.0 einzuführen. Denn das wird die Umsetzung der Industry 4.0 dramatisch vereinfachen. Der Mehrwert auf dem Information Layer wird über eine Abarbeitung



standardisierter Services generiert. Die Services stehen im Allgemeinen in der Cloud zur Verfügung und werden bei Bedarf von den Maschinenelementen herangezogen. Durch diese Auslagerung der Wertschöpfung auf standardisierte wiederverwendbare Services findet nicht nur eine Ordnung der Prozesse statt, sondern gleichzeitig hilft die Serviceorientierung, die Vernetzung sehr viel einfacher in bestehenden Maschinen und Anlagen- Strukturen einzubauen.

Eine noch sehr grobe Service Orientierte Architektur innerhalb der Analytic Industry zeigt die folgende Abbildung. Der **Machine Learning Service** bietet für verschiedene Einsatzszenarien seine Leistung anderen Services, wie beispielsweise dem **Analytic Service**, zur Verfügung. Ein Analytischer Service gibt eine Ausfallprognose eines Maschinenteils aus vom Machine Learning Service als visuelle Auswertung an eine konkrete Person wie einem Betriebsingenieur weiter. Oder der Machine Learning Service gibt Parameter an den Automation Service weiter, sodass eine Produktionsstrasse besser auf äussere Einflüsse reagieren kann und damit die Qualität des Produktes sich verbessert.

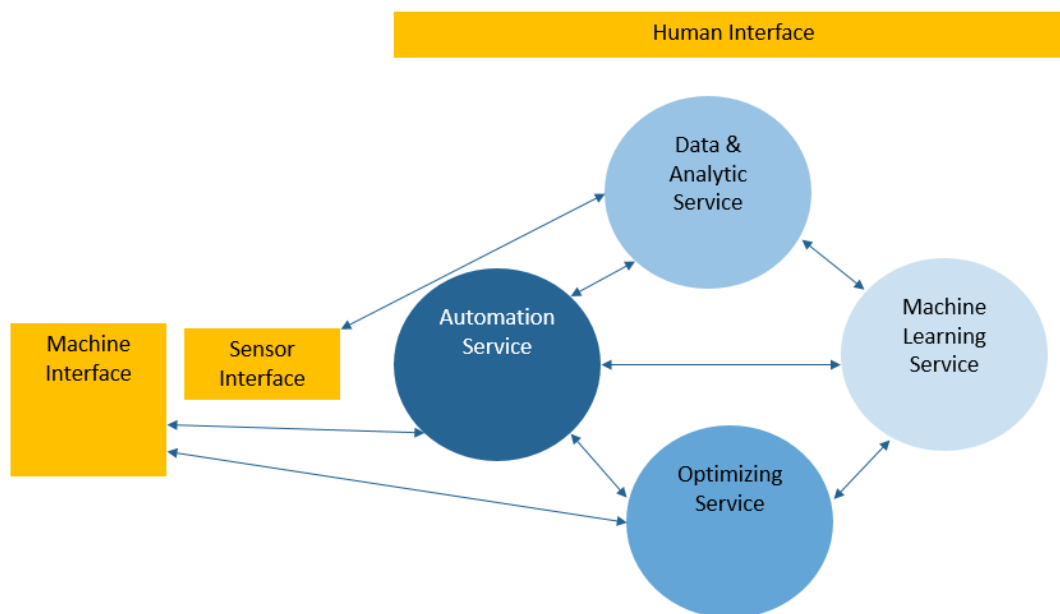


Abbildung 6: Service Orientierte Architektur für Analytic Industry 4.0

Die Erkenntnis ist also, dass die Aufgaben auf der Informationsebene modularisierbar sein sollten, um effizient genutzt zu werden. Die einzelnen Module liegen nicht an einem physischen Ort oder sind an eine physische Einheit gebunden, sondern werden über einzelne Services oder Produkte miteinander vernetzt.

## 6. Cloud als verbindendes Element

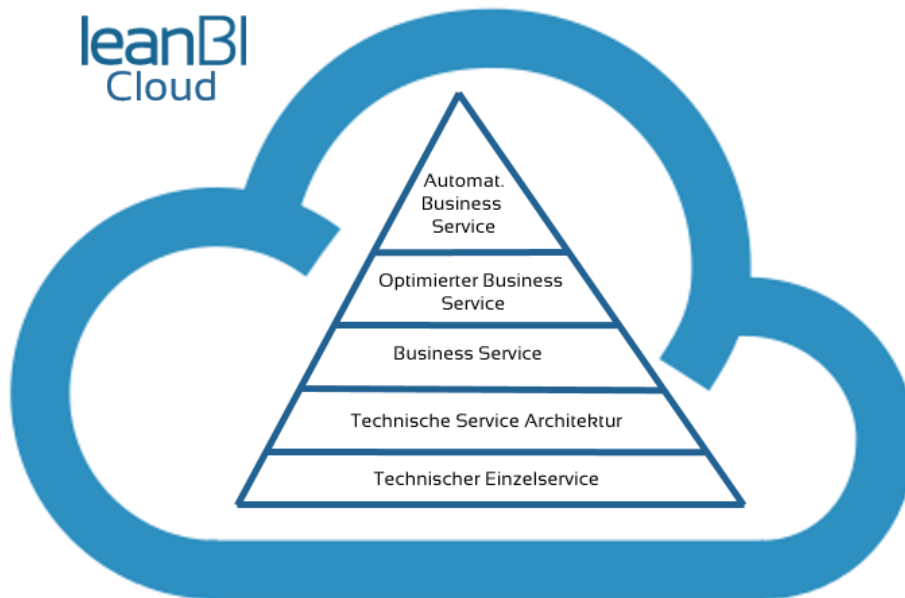


Abbildung 7: Automatisierte Business Services aus der Cloud

Und dabei hilft uns die Cloud. Denn Services können über die Firmengrenzen ausgelagert werden, die nicht als Kernkompetenz einer Unternehmung angesehen werden. Neben den Gründen der Kostensenkung steht die Kompetenz zu einem Service, die sich zentral viel besser erhalten und ausbauen lässt. Der Service läuft innerhalb der Cloud Infrastruktur eines Rechenzentrums, welches sicherheitstechnisch so ausgerüstet ist, wie sich dies nur grosse Unternehmen im Alleingang leisten können. Die Rechenzentren wiederum sind miteinander verbunden, sodass auch einzelne kommunizierende Services sich hochverfügbar und sicher austauschen können. Genauso wie die Wertschöpfung eines Produktes über verschiedene Lokalitäten stattfindet, findet dies auf der logischen Ebene statt.

## 7. Der Machine Learning Service

Zentrales Element des Analytischen Service ist der Machine Learning Algorithmus. Als kompakten Überblick über Machine Learning empfehlen wir die beiden Wikipedia-Seiten:

[http://de.wikipedia.org/wiki/Maschinelles\\_Lernen](http://de.wikipedia.org/wiki/Maschinelles_Lernen)

[http://en.wikipedia.org/wiki/Machine\\_learning](http://en.wikipedia.org/wiki/Machine_learning)

Leider lassen sich Machine Learning Algorithmen selten vollkommen automatisieren. Denn es muss zu jedem Problem ein richtiger Algorithmus angewendet, ein passendes

Modell bestimmt, sowie aus der Datenfülle eine angemessene Teilmenge der Features ausgewählt werden.

In vielen Fällen durchläuft der Algorithmus einen iterativen Prozess, sodass die Modelle mehrmals angepasst werden. So können zwar gewisse Algorithmen sich selber optimieren, diese müssen aber trotzdem zuerst angelernt werden, damit sich die Algorithmen stabil verhalten und gute Resultate liefern.

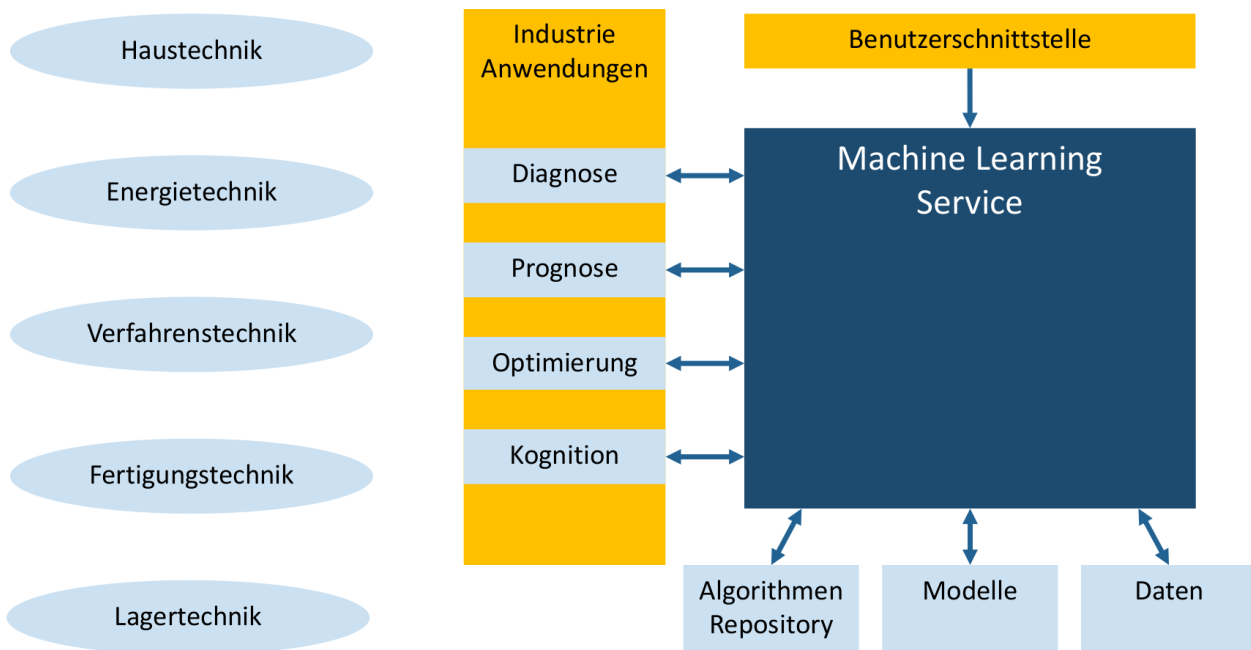


Abbildung 8: Automatisierte Business Services aus der Cloud

Wir verfolgen den Ansatz eines generischen Machine Learning Services. Dieser Service bedient alle analytischen Use Cases der Industrien. Glücklicherweise lassen sich die analytische Use Cases lassen sich auf eine Menge von Basis Use Cases reduzieren.

Dies sind:

### **Diagnose / Prognose:**

Die Auswahl der Algorithmen für Diagnose und Prognose, beispielsweise des Maschinenausfalls oder der Qualität, ist vielfältig; Decision Tree, Neuronale Netze Bayes, Logistic Regression oder Lasso um nur einige zu nennen.

**Optimierung:** Bei einer Kosten/Nutzen Optimierung sollte man auf die Wahrscheinlichkeit beispielsweise eines Maschinenausfalls schliessen können. Einige Algorithmen wie der Bayes Algorithmus können diese Wahrscheinlichkeit abschätzen. Ein Optimierungsalgorithmus liefert dann einen Vorschlag zur Kosten/Nutzen Optimierung. Als einfachere Alternative können die Algorithmen der Diagnose herbeigezogen werden. Auf dieser Grundlage wird erkannt, in welche Richtung die Eingabewerte geändert werden sollten, um das Kosten/Nutzen Verhängnis zu verbessern. Hier können einerseits parlamentarische Modelle helfen. Oder zukünftige

Änderungen der Eingabewerte werden zuerst virtuell (mittels Diagnose-Algorithmen) auf ihre Auswirkungen untersucht. Erst bei positivem Kosten/Nutzen Verhältnis werden die Eingabewerte auch in der realen Welt verändert.

**Kognition:** Im Gegensatz zu Diagnose und Optimierung kann bei Kognition das Ziel zu Beginn unbekannt sein. Sogenanntes unüberwachtes Lernen (engl. unsupervised learning) ermöglicht es Muster in den Daten hervorzuheben und zu erkennen. Dazu werden sogenannte Clustering Algorithmen, wie k-means eingesetzt.

Folgende Inputquellen sind vorgesehen:

**Algorithmen:** Die Algorithmen werden aus einem Repository herausgezogen. Das Repository besteht aus den neuesten Updates, aber geprüften Algorithmen der Open Source Community. So kann für jedes Projekt die Vorhersagekraft verschiedener Algorithmen relativ einfach verglichen und ein passender Algorithmus ausgewählt werden.

**Modelle:** In gewisse Algorithmen, wie Neuronale Netze oder Decision Tree, sind die Modelle implizit enthalten. Sie können mit wenigen Parametern einfach ausgewählt oder sogar automatisch bestimmt werden. Andere Algorithmen (z.B. Bayes oder Logistic Regression) erfordern ein projektspezifisches Modell. Die Bereitstellung eines solchen Modelles ist zeitintensiver und setzt ein hohes Verständnis für die modellierten Vorgänge voraus. Dafür resultiert häufig eine bessere Voraussage-Qualität, insbesondere bei kleineren Datensätzen.

**Datenauswahl** (engl. feature selection): Insbesondere bei kleinen Datensätzen ist es nötig, eine (gut gewählte) Teilmenge durch einen Lernalgorithmus laufen zu lassen. Ansonsten verliert der Algorithmus seine Voraussagekraft wegen Überanpassung (engl. overfitting). Die Gefahr von Überanpassung reduziert sich mit zunehmender Menge an Datensätzen. Damit lassen sich komplexere Modelle rechnen und mehr Features berücksichtigen. Die Qualität der rückgeführten Resultate wird erhöht. Auch kann es hilfreich sein, die Daten vorgängig zu bearbeiten. Beispielsweise sollte bei der Schätzung der kinetischen Energie die Geschwindigkeit vorgängig quadriert werden ( $E_{kin} = 1/2 m v^2$ ).

Der Machine Learning Service hat damit die Aufgabe, für den konkreten Use Case ein passendes Modell zu anzuwenden. Ganz ohne Data Scientist Kompetenzen wird dies nicht geschehen. Aber der Service lässt sich halbautomatisieren und kann den Aufwand eines Data Scientist stark vermindern und die Qualität des Output gleichzeitig erhöhen.

Wie wir auf der Basis von generischen Ansätzen reale Services erarbeiten werden, können sie in weiteren White Papers lesen.