Chapter 7

Data Analytics: Industrial Perspective & Solutions for Streaming Data

Mohsin Munir, Sebastian Baumbach, Ying Gu, Andreas Dengel, and Sheraz Ahmed

German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany
firstname.lastname@dfki.de

Over the past few years, a lot of devices and machines around us are becoming 'smart'. Based on the idea of the Internet of Things (IoT), different devices and machines can connect to the internet and communicate with each other. Such internet enabled devices are continuously observing their environment and logging a lot of data in the back-end database. By applying data analytics on the gathered Big Data, smart decisions are made to facilitate the end user according to the current situation. This capability of adaptive decision making actually makes ordinary devices and machines 'smart'. These devices and machines are becoming intelligent by learning about their surroundings from different sources, and develop the ability to avoid unforeseen situations by analyzing that data. In this chapter, we provide a comprehensive overview of how different industrial players are using data analytics to provide better services to their customers and improve their internal processes and workflows. We discuss how different industries use data analytics to gain vital insights for providing better healthcare to public, making homes more secure, increasing crop yield, delivering goods more quickly, reducing the downtime of a machine, avoiding a disease, etc. An overview of different analytics platforms and solutions used in different industries for time series and streaming data are also discussed in this chapter.

1. Introduction

We live in an age where data is becoming a key for success in every field of life. Due to rapid advancements in technology, different players in academics, research, and industry are collecting different types of data, which serve as their driving force. However, the data itself has no meaning until

it is carefully analyzed. Data analytics[a] is a process of analyzing data in order to discover hidden patterns, knowledge, or trends. Many companies are using data analytics widely to add value to their business by analyzing past and current data. The process of data analytics not only includes analysis of the data, but it refers to a complete workflow which starts with data collection and other steps including data cleaning, data preparation, data governance, data analysis/modeling, and finally data visualization. Each of the steps in this workflow is a key for good analytics.

To maximize the output, efficiency, and for added value, there are different data providers, which take care of the initial steps (i.e. data cleaning, preparation, and governance) in the workflow of data analytics by providing data externally. This means that in combination with the in-house data collected by companies, there are many external data sources, which are being utilized to perform data analytics. Furthermore, due to the evolution of technology, the mode of data collection has also evolved. Especially, with the emergence of IoT, which refers to the network of physical devices, vehicles, buildings, and other items embedded with electronics, software, sensors, actuators, and network connectivity that enable these objects to collect and exchange data [1]. Recent studies show that the IoT market is growing and will continue to grow a lot over the next four years. It is expected that a market value of nearly $122 billion will be reached by the year 2022 [2]. Due to these IoT enabled devices, companies are now continuously getting live streams of huge amounts of data. This data is usually collected over a continuous interval of time, which results in time-series data. This IoT based time-series data resulted into the evolution of data analytics where in contrast to traditional data analytics where data cleaning, and analysis were at the core, now it becomes even more difficult to store and maintain the data, which is being collected by these IoT devices.

With the help of data analytics on streaming/time-series data, companies can keep an eye on different aspects, e.g., reducing maintenance costs, avoiding equipment failures, and improving business operations. In addition, retailers, restaurant chains and makers of consumer goods can use the data from smartphones, wearable technologies, and in-home devices to do targeted marketing and promotions – the business side of the IoT's futuristic world of connected consumer gear.

Almost every industrial sector, be it health care, agriculture, manu-

[a]Data analytics and Data mining are used interchangeably nowadays. In this chapter, the term, data analytics is used. In an industrial perspective, data analytics is considered as a complete workflow, which incorporates data mining among other things.

facturing, agriculture, dairy farming, logistics, automotive, etc. is now redefining their products and enabling them to IoT for gaining maximum benefits. In this way, different industrial solutions are also available focusing specifically on data analytics on time-series data. This chapter focuses of providing insight on the different data analytics solutions available in different industrial sectors.

2. Data Analytics in Agriculture

With the increase of the world population and the improvement of living standards, the demand of high quality food is increasing. Agricultural mechanization is playing a vital role to fulfill this need with the help of large-scale production. However, the physical performance of mechanization and mass production is limited with the advancements in the fields of IoT and cloud computing, the devices for data collection and data storage have become affordable and prevalent. In the agriculture sector, data analytics can be applied on the machinery and farming data collected in order to reduce loss, improve efficiency, and lower costs under the condition of unchanged physical properties. This enables a modern farming concept called precision agriculture (PA) or satellite farming. This kind of farming helps the farmer to recognize the variations in the farming land and how to adjust input for different parts of land to optimize the output. A global positioning system (GPS) is the backbone of PA. With the help of GPS, a farmer can identify the exact area where soil conditions vary. In conjunction with the precise location, different measures like air quality, moisture level, field terrain, crop yield, crop maturity, and gas levels are recorded and turned into meaningful information using data analytics.

Data driven decision-making has been extended from the business sector to the agricultural sector. Many large enterprises in the agribusiness are becoming involved in data analytics research and development. They are providing solutions for PA and for a variety of other issues in agriculture. Fierce competition between companies has already begun.

John Deere converted their equipment to the paradigm of IoT to help farmers manage their fleet, reduce down time, and the cost of production. This information is combined with the local weather data, soil data, crop characteristics, and other data sets from different sources. **MyJohnDeere.com** is a platform for data analytics, which provides the possibility to store, analyze, and visualize results on a web-portal (as well as on mobile application called **Mobile Farm Manager**). With the help of such plat-

forms, farmers can figure out when and where to plant which kind of crops, when to plough and when to harvest, and which optimized path should be followed during the work. The right decision can help farmers to improve their efficiency. The data collected during different phases of farming are massive. To take advantage of the collected data, John Deere already steps into big data analysis for the future of farming.

IBM and **SignalDemand** have developed a data analytics system which uses predictive analytics to predict the demand and optimize the margin to meet the needs of different agribusiness companies. While large agricultural enterprises have large datasets, advanced equipment, data scientist, and domain experts at their disposal, the majority of farmers neither have access to such information nor the resources to get benefits from advancements in technology. To help farmers who are working on a small-scale and lack the technology infrastructure, IBM built a back-office network. They supply corn-specific information on a regular basis, along with generalized information on fertilizer and weather conditions to registered farmers via their mobile phones. A farmer can get timely agronomic intelligence simply via automated voice mail or text messages on his mobile phone.

aWhere (an American corporation) collects and analyzes over a billion points of data (which is a pivot element for analysis) from around the globe each day to create unprecedented visibility and insight which is known as Agricultural Intelligence. This intelligence is used for critical decision making from farm level through to national policy [3]. High-quality weather data is combined and analyzed purely for agricultural use. Their major data analytics solutions are **Weather Terrain, Weather Agronomics,** and **Weather Support**.

The Climate Corporation (a San Francisco-based company) examines weather data to provide insurance to farmers who can lock in profits even in the case of drought, heavy rains, or other adverse weather conditions. **FieldView** is their data analytics solution, which combines farmers' field data with real-time and past – soil, crop, and weather data to help them efficiently manage their operations and gain insights into their fields [4]. In addition to the FieldView, they also provide a hardware solution **SeedSense** for Planter Monitoring. Perfect planter performance can be achieved by maximizing planter speed and adjusting vacuum pressure by using SeenSense. It also enables the farmer to sow precisely, maintain depth, avoid compaction, and troubleshoot mechanical problems.

The CropOS is a data analytics platform, which uses machine learning and cloud biology to improve crop performance and help scientists and

breeders with some of the biggest challenges in the agriculture sector. It is developed and maintained by Benson Hill Biosystems, which is an agricultural solutions company. They unlock the global genetic potential of plants to enhance the sustainability of food, feed, fiber, and fuel production [5]. CropOS represents a uniquely powerful platform at the intersection of big data, machine learning, and plant biology. CropOS empowers researchers to significantly increase the yield of major food crops and identify the most promising plant genetics in weeks instead of studying long growing seasons.

CLAAS focuses very much on self-propelled machines developing and producing combine harvesters, self-propelled forage harvesters and tractors [6]. Self-propelled machines are very important especially for crops like wheat, rye, barley, and corn, which have to be harvested at just the right point of maturity. Once this harvest maturity has been reached, the combine harvesters work in the fields day and night. In this process, up to 50 parameters from the reel to the chopper influence the harvest yield. The operator has to continuously monitor and evaluate around a dozen of these parameters. Hardly any operator is capable of keeping an eye on everything and tapping the machine's full potential. To solve this problem, CLAAS also moved toward IoT enabled combine harvesters. In addition to this, an assistance and analytics system is used, which permanently monitors the harvesting process and automatically adjusts the machine setting to the current conditions which is faster and more precise when compared to a human operator. Furthermore, together with the German Research Center for Artificial Intelligence (DFKI), and the Fraunhofer Institute of Optronics, System Technologies, and Image Exploitation (IOSB), CLASS is working on extending the data analytics to improve the performance of mobile work machines with unsupervised anomaly detection algorithms, which can detect unexpected events without any previous domain knowledge.

3. Data Analytics in Healthcare

Similar to agriculture, data analytics is playing a vital role in the advancement of the healthcare sector. With the easy availability of smart devices (including smart watches, smart phones, and smart wristbands), a new dimension of healthcare has emerged – Smart Healthcare. The end-user smart devices are continuously collecting users' data regarding different activities performed over a day, month, or year using different sensors.

Data analytics on smart sensors' data have opened new dimensions of research and applications in Connected or Smart healthcare. Smart health-

care is supporting, and slowly replacing traditional healthcare. By analyzing the streaming data generated by smart wearables, it is possible to see if a user is healthy, or if some preventive measures are required, in order to avoid a potential health problem. Now doctors can remotely examine their patients and suggest treatments on the go. Smart healthcare offers many new possibilities for patients too. Patients can keep updated with their health and fitness data all the time, find other patients suffering with the same disease to discuss various treatments, and easily track the post-surgical needs. The digitization of patient health data encourages the communication and collaboration of all the stakeholders involved in the patient's health. For instance, i) government institutes can use the data to extract different statistics and to make policies as needed, ii) pharmaceutical companies can use the data to track the positive or negative effects of different medicines, iii) doctors can use this data to choose a treatment when a patient has high cardiovascular risk, etc. With smart and connected healthcare, healthcare is shifting from being episodic/reactive to preventive/proactive. Different companies (mentioned below) are providing solutions for connected, smart, or preventive healthcare.

IBM Healthcare is a data analytics solution, which focuses on health monitoring and intervention, analyzing streaming data (such as data generated in ICU), and helping in detecting signs of various changes occurring in a patient's health. The detected early signs are used to generate medical alerts for proactive intervention. It also enables healthcare providers to improve operational performance, reduce cost of care, and counter fraud in healthcare by using integrated data management and analytics. Furthermore, it provides consumer level analytics to understand consumer preferences and behaviors by capturing data from different sources such as claims, clinical history, and social platforms; and then merges all the data into one unified view. It also helps building a predictive model that evaluates the risk of readmission for patients with chronic obstructive pulmonary disease [7]. Researchers at National Institutes of Health (NIH) are using **IBM PureData** System for analytics to unlock new insights from data gathered over decades. With the help of this system, researchers can run analysis on large, complex data sets (both clinical and genomic research data) and generate reports faster than ever before [8].

SAP Real-Time Analytics is a complete solution for patient care, human resources, finance, care collaboration, and healthcare analytics. The big health data collected from electronic health records, research, physician notes, insurance claims, and social media data are used by SAP Real-Time

Analytics to reduce cost and improve quality of care. This solution enables data scientists to separate noise from signals and derive meaningful insights from the data. The unified analytics model transform data from a wide range of sources into actionable information. Seoul National University Bundang Hospital (South Korea) has developed its clinical data warehouse (CDW) using **SAP Data Services** and **SAP HANA**. Their CDW is used to automate the clinical indicators system, gather critical data in real-time, provide instantaneous feedback to clinicians, and provide multidimensional analyses based on patient characteristics, diseases, and location [9].

General Electric (GE) provides many healthcare solutions in general; and some solutions are based on data analytics in the areas of diagnosis, clinical decision-making, and asset monitoring. The GE **Marquette 12SL ECG** analysis program provides diagnostic confidence to care providers by giving fast and reliable cardiac care decisions. In the area of patient monitoring, GE provides **CARESCAPE Central Station** which allows the integration of different medical devices and systems to access patient's historical data. When a patient moves to a care area, this solution enables care providers to perform in-depth analyses and offers clinical decision support. **Centricity Imaging Analytics** is a real-time dashboard which provides visibility into the workflows of the radiology department for increasing department throughput and patient care.

Combined Applications to Reduce Exposure (CARE) (by Siemens) is an analytics solution, which is designed to improve dose monitoring in different interventional radiology systems. The dose of an individual patient is recorded in addition to other data, such as CT-dose index, dose length product, and total recording time. This data is also used to enhance dose reporting and assessment, transparency regarding dose per case, reporting on patient dose history, and cross-institutional reporting.

Apple CareKit is an open source platform for creating health related apps to regularly track care plans, monitor users' progress, and share their insights. *One Drop* (by Informed Data Systems, Inc.) is an example of such a mobile app created using CareKit. **Apple ResearchKit** enables developers to create apps, which enable researchers and doctors to gather robust and meaningful data for their health related studies, and obtain a complete history of their patients. The real life data collected is used to find physical patterns, correlation between physical history and medication, predict a particular problem, and recommend diet and fitness plans. With the help of ResearchKit and CareKit, researchers use Apple Watch to predict seizures before they actually happen. For instance, *EpiWatch*

(an Apple Watch app by Johns Hopkins University) enables people to accurately track the onset and duration of seizures in real time. A patient sensing an impending seizure launches the app on Apple Watch and an alert is automatically sent to a designated family member or caregiver. Similarly, *Asthma Health* (by Weill Medical College), *Concussion Tracker* (by NYU Langone Medical Center), *GlucoSuccess* (by Massachusetts General Hospital), and *C Tracker* (by Boston Children's Hospital) are examples of such apps which are built on top of CareKit and ResearchKit.

4. Data Analytics in Manufacturing

Data analytics provides a granular approach to diagnose and improve whole manufacturing flaws. It is always in the manufacturers' interest to improve their production processes, product quality, production cycle, and the amount of output per unit of input. Due to the involvement of a number of players and processes in the manufacturing life cycle, it is hard to find the cause of failure or inefficiency exactly. With the growth of Industrial IoT in recent years, everything is going digital and connected. With the help of this digitization and connectivity, a lot of streaming data related to equipment, automation, production lines, systems, and products are generated and stored. Manufacturers can use data analytics to leverage the data collected from on-the-floor factory machinery alongside other traditional (factory logs) and social data. Some of the advantages of using data analytics in manufacturing are to – i) get unexpected insights into different processes, ii) increase accuracy, quality, and yield (amount of output per unit of input), iii) improve the forecast of product supply and demand, iv) enhance the understanding of plant performance across multiple metrics, v) boost the product quality, vi) track all products with defected components, vii) predict machine failure, viii) quantify how daily production impacts financial performance, ix) provide preemptive maintenance and service by continuously monitoring a product instead of fixed term maintenance, and x) identify the root cause of a failure. Figure 1 shows how advance data analytics can help decode and improve complex manufacturing processes.

The main challenges in manufacturing are a lack of collaboration across different departments, disparate systems and data sources, and difficulty in coordinating supply and demand chains. Such challenges, among others are tackled in the solutions provided by different companies using advanced data analytics.

IBM Analytics provides a complete analytics solution to be used in au-

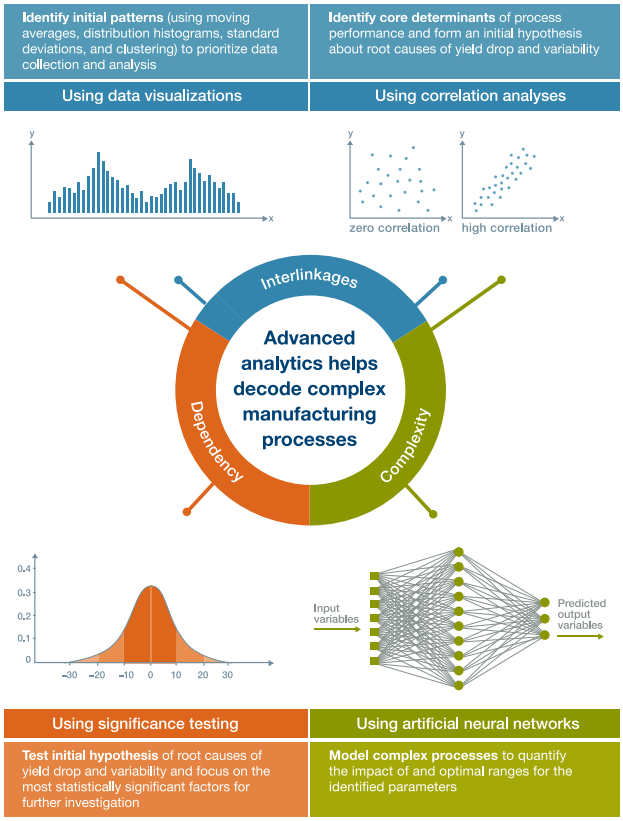

Fig. 1. This figure shows how advanced analytics can be used in streamlining manufacturing value chains by finding the core determinants of process performance [10].

tomotive, defense, chemical, petroleum, energy, aerospace, electronics, and other industries to uncover deeper insights into operations, inventory, market demands, supply chain, and performance [11]. By applying advanced data analytics on aggregated data from different sources (such as different sensors, maintenance logs, and production systems), manufacturers can efficiently achieve their demand, production, and supply requirements; while properly managing all the resources at minimal cost. It can integrate structured as well as unstructured data from different sources. IBM analytics can unveil a number of critical manufacturers questions, such as how operating costs can be reduced while having better project financial performance, how greater visibility into supply chains can be achieved, how the supply

chain's needs can be predicted, and how the maintenance cost can be cut down. It can also uncover insights into customers' behavior, their needs, and market trends to make better business decisions. Nowadays, production assets and consumer products are transmitting vital operational data to backend data warehouses. **IBM Predictive Maintenance and Quality** software solution leverages the data collected from different sources and predicts when a particular asset or machine needs maintenance. In contrast to the traditional scheduled maintenance, predictive maintenance recommends when maintenance is required and when it is not. This type of maintenance helps to keep critical production lines and consumer products running, while saving money and minimizing customer inconvenience. Muller, Inc., USA, is a retailer and manufacturer of metal products. They used *IBM Cognos Business Intelligence, IBM Cognos TM1, IBM SPSS Modeler, and IBM Business Analytics* to pull data from all points of sale, inventory, and ERP systems; so that the employees can view and analyze company data, measure individual performance, and access how their work affects the bottom line [12]. The Vaasan group (a leading bakery operator in Northern Europe) used IBM Analytics to enhance forecasting and inventory management. The solution based on the *IBM Cognos Controller, IBM Cognos Intelligence, and IBM Cognos 8 Planning* enabled the bakery to predict production requirements and helped them prepare for fluctuating orders [13].

SAP provides multiple solutions in the domain of manufacturing. SAP **Manufacturing Execution System** connects, monitors, and controls different manufacturing operations. With the help of automated data collection, it provides visibility into the manufacturing processes which helps process managers to find and resolve quality issues. Its asset utilization functionality improves overall equipment effectiveness, facilitates predictive maintenance, and minimizes downtime. SAP **Enterprise Resource Planning (ERP)** is an enterprise level system for streamlining the manufacturing, services, sales, finances, and human resource processes. It is composed of different modules, which accelerate the entire manufacturing process, boost sales and customer satisfaction, provide support for administration tasks, streamline and automate financial operations, and provides real-time analytics based on ERP data. SAP **Manufacturing Integration and Intelligence (MII)** is the solution for smart manufacturing which exploit the data collected from Industrial Internet of Things (IIoT). It automates the IIoT and facilitates in manufacturing data transformation and integration. This software is equipped with the **Manufacturing**

Analytics Platform, which provides statistical process control and predictive analytics. It can also identify the root cause of machine downtime and efficiency loss; which makes the maintenance task easy for technicians and helps the operation team to improve efficiency. The SAP **Predictive Maintenance and Service** solution leverages the IoT data to transform reactive maintenance to predictive maintenance. It provides the visibility into manufacturing asset and consumer product health by remotely observing their behavior and patterns. By analyzing the Big Data collectedd, future needs are predicted [14].

Microsoft Azure IoT is a complete suite for connecting IoT devices, collecting IoT data, analyzing the collected data, and mining disparate data [15]. Existing data and systems can also be integrated with new data sources to create new insights and business models. A Predictive Analytics module in Azure provides insight into how a certain product behaves in normal conditions and in other special conditions by finding patterns and correlations in historical and new sensor data. Based on such analytics, this suite is able to provide warning signs, identify where a problem exists, and notify when equipment needs maintenance. With such preemptive warnings, small repairs can be made before big failures occur. It also helps in prioritizing the maintenance task by providing information about which equipment is at high risk. Once an actual root cause of the failure is detected, it can facilitate a technician by recommending the error code (with possible fixes) for that condition. The technician's time of finding the root cause of a failure is saved, now he just has to fix the defective component (with the help of some recommendations about possible fixes). This suite enables manufacturers to remotely monitor their assets, which are deployed outside the factory. Automatic notifications can be triggered on this live data to get real-time asset feedback and maintenance requests.

General Electric (GE) Brilliant Manufacturing is a software suite, which connects people, machines, materials, and processes in IoT. This suite maximizes manufacturing production performance and optimizes operations through advanced real-time analytics. It allows the integration and aggregation of whole manufacturing life cycle data from the beginning till the end. Data driven analytics from disparate manufacturing sources allow manufacturers to take optimal decisions to drive improvements in end-to-end production [16]. This suite includes different products including the following: i) **Efficiency Analyzer** provides an up-to-date view of the entire production process and transforms real-time machine data into action efficiency metrics. Such unified metrics help plant managers to reduce un-

planned downtime, maximize yield, improve production quality, increase flexibility, and maximize team productivity. ii) **Production Quality Analyzer** analyzes data to catch non-conforming events before they occur to help quality engineers to easily identify the problem. iii) **Production Execution Supervisor** digitizes documentation, instructions, orders, and process steps, enabling manufacturers to get the right information at the right time. iv) **Product Genealogy Manager** builds a record of all equipment, raw materials, tools, and personnel which are required to build the finished goods. It helps service personnel to manage services in an efficient way.

Manufacturing Analytics by BOSCH is a solution for analyzing production data. Different types of data such as test, process, and machine data from different sources can be used to improve the production process and product quality while reducing the cost with the help of this suite. This suite can integrate the existing production data with the new data. The predictive models can be applied to real-time data for predictive maintenance and root cause analysis. Data analytics unveils the previously unknown correlations in data and helps manufacturers in gaining new insights. The newly discovered data insights and prediction models can be applied using this suite to automate the analytics process.

SAS provides different solutions to get the best out of the manufacturing life cycle. SAS **Demand-Driven Planning and Optimization** suite improves the supply and demand planning processes. This suite uses analytical insights of demand patterns to help manufacturers in making supply plans, which are aligned with the demand forecast. Production and logistics can also be managed to match the ever-changing customer needs and market dynamics. SAS **Quality Analytics** suite includes data mining and predictive analytic technologies for predictive maintenance and identification of potential problems. It also helps in reducing the total cost of quality by reducing the scrap and rework, and identifying design and production defects. SAS **Field Quality Analytics** helps in making aftermarket service efficient by integrating and analyzing internal and external data sources. It helps in detecting and prioritizing warranty and service issues. SAS **Customer Intelligence 360** collects, analyzes, and reports on customer experiences to improve sales and marketing performance. It provides insight into customer segmentation: which customer groups are more likely to buy which kind of product and why. With the help of such forecasts, advertising and promotion campaigns can be planned and targeted at customer groups [17].

5. Data Analytics in Connected Vehicles

A connected vehicle is a vehicle designed with the capability of connecting to the internet and other connected devices including smart phones, traffic lights, other vehicles on the road, smart home appliances, etc. It is predicted by Gartner, Inc. that by the year 2020, one in five vehicles in the world will have some form of wireless connectivity in them, which adds up to 250 million connected vehicles [18]. The accumulated data based on driver's behavior, car machinery, sensors installed in the car and in the surroundings can leverage data analytics in the following functional areas: autonomous driving, safety, infotainment, well-being of driver's health, vehicle management, mobility management, and smart home integration [19]. Vehicle manufactures like BMW and Volkswagen are making these connected vehicles smart by introducing functionalities like autonomous car parking and emergency assist respectively. Data analytics provides car manufacturers with crucial insights into the vehicle system, behavior of the vehicles in certain conditions, and drivers' patterns. Thousands of components inside the vehicle are continuously logging data. Even if the test driver observes an unexpected shifting characteristic, it is hard for a manufacturer to exactly find the defective component or the contributing components. But, with the help of data analytics, the defective component and the contributing components can be figured out precisely.

Ford and **IBM** are working together to develop a platform which analyzes data collected from a vehicle. Based on the small chunks of vehicular data, this platform can spot patterns, correlations, and trends to help the driver make efficient transportation decisions. Data collected from Ford **Smart Mobility Experimentation Platform** helps their scientists to spot tendencies and behaviors, and their customers to have a better travel experience. They are working on using real-time analytics to learn about a problem on a particular route by taking data feed from different systems [20]. In the domain of predictive maintenance, Ford is working on sending personalized oil change and brake maintenance notifications to drivers. The collected data is statistically analyzed in order to evaluate the maintenance needs for each vehicle separately [21].

Daimler is making their cars and trucks intelligent by enabling them with anticipatory planning. Based on data from different sources, their vehicles are able to operate on an anticipatory basis in which they can foresee different things which the human eye cannot see. Their trucks and buses are equipped with **Predictive Powertrain Control** (PPC), which can

anticipate the terrain and adjust the vehicle accordingly. Based on the 3D map data, PPC adjusts the vehicle speed and gear selection optimally to the topography of the transport route. This control reduces fuel consumption by up to 5% [22].

BMW group is also using **IBM Big Data and Analytics** technology to optimize their products, repairs, and maintenance processes. **IBM SPSS** predictive analytics software is used to combine and analyze data from different sources like pre-production sensor data, workshop notes, and numerous test drives of prototypes [23]. In this way, different vulnerabilities can be identified quickly, and eliminated before the model goes into series production. Before this automated process, this evaluation took months to complete. IBM Big Data and Analytics are used to analyze data from all available sources to discover anomalous patterns and predict maintenance needs.

Volkswagen, in collaboration with CSC (a technology solutions and service provider company), use data analytics to support predictive marketing to increase aftermarket service revenues [24]. They combine customer data with vehicle data, and notes written by technicians at the service centers. With the help of that data, they are able to predict upcoming maintenance for specific drivers.

Tesla car manufacturer is collecting data from their connected cars and using telematics to batch stream key data points to backend big data pool. The collected data enable engineers and manufacturing lines to resolve the issues and send back fixes with their over-the-air software updates. They are providing continuously improving customer experience based on the data and analytic views [25].

Audi is also making its vehicles intelligent with a vision to reduce fuel consumption. The predictive efficiency assistant enables the vehicle to slow down or automatically adjust the speed to the conditions in an anticipatory manner. The system analyzes the route topography, speed limits, road users ahead, and navigation data.

Caterpillar, Inc. is the world's leading manufacturer of construction and mining equipment. They have created a new organizational division called Analytics and Innovation (AI) to form a broad and connected analytics ecosystem. The data collected from gigantic machines are used to develop predictive and proscriptive information. This predictive diagnostics is shifting their customers from reactive (repair after failure) to proactive (repair before failure) mode [26]. By using data analytics, they are able to point out inefficiencies in the operation of a particular machine by compar-

ing its operational data with that machine's benchmark data.

6. Data Analytics in Logistics

Logistics service providers move masses of goods from one location to another. A lot of data related to shipments, origin, destination, size, weight, and content are stored per shipment. Some of the advantages of using data analytics in the logistics sector are, i) optimization of delivery time, resource utilization, and geographical coverage, ii) goods storage capacity and required resources forecast, iii) valuable insight into customer sentiment and product quality, and iv) insight into the global flow of goods.

DHL uses big data analytics to make their operations more efficient. Rapid processing of real-time information enables their **SmartTruck** to optimize the delivery route in real-time. Delivery routes are also automatically updated according to traffic conditions. Unsuccessful delivery attempts are avoided in intelligent routing, based on the availability and location information provided by the recipient. SmartTrucks are re-routed on the go, based on the combined analytics of geographical factors, environmental factors, and recipient data [27]. It is important for a logistics company to plan operational capacity in time. The optimal planning cannot be done by neglecting external factors, such as unexpected bankruptcy, a regional outbreak disease, or natural disasters etc. DHL Solutions and Innovation is working on an analytics tool to measure external factors on the expected volume of shipment to make efficient shipment volume prediction. Based on the shipment records, DHL provides an online geo marketing tool **Geovista**, to analyze business potential. This tool provides a sales forecast and local competitor analysis. DHL is also working on a Supply Chain Risk Management Solution which will improve the resilience of logistic providers entire supply chain with the help of predictive analytics on a global scale (by aggregating data from different local sources such as politics, economy, nature, health, etc.).

Amazon was the first company to give recommendations about items in which a user might be interested. Today, it uses different parameters (such as, which items are bought by a particular user before, what he has in his wish list and virtual cart, which items he has rated or viewed, and which items a similar user has bought) to customize the browsing and buying experience. Predictive analytics is used to ensure the right item must be in stock when a customer orders it. Amazon is taking data analytics to a different level with its patent on *Anticipatory Shipping*. The patent is

officially called 'Method and system for anticipatory package shipping'. The idea of anticipatory shipping is to predict who will order what and when, and then ship that item even before it is ordered. Another scenario is also discussed in patent for 'speculative shipping'. In this type of shipping, a package is sent to a geographical area, without completely specifying the delivery address at the time of shipment - the package might remain in near continuous transit on trucks until a customer makes a purchase [28]. In this way, the package is shipped to the customer instantaneously.

7. Data Analytics in Dairy Market

The current trend of automation and data exchange in modern manufacturing is inextricably linked with the production industry as it helps making cars autonomous or factories more productive. Nowadays, not only these industries can benefit from IoT, but one of the oldest sector of mankind, i.e., milk production, is also taking advantage of smart technologies. For a long time, the dairy market has been suffering from low prices, which means that modern technologies and data analytics can neither influence market prices, nor the bargaining power of the dairy, nor the retail industry. However, these new technological trends can help farmers to reduce their production costs and enable them to produce more milk by keeping a keen eye on their cows health.

Effects of the globalized milk market are already noticeable. Farmers are suffering mostly from the extremely sharp fall in prices. The low milk prices make it nearly impossible for farmers to obtain profits, as they are not covering costs. They are forced to optimize their production. Legal requirements and a change in social perception restricted many alternatives, like the prophylactic use of antibiotics in Europe, for optimization [29]. The only chance to raise their economic performance is to reduce costs and increase the efficiency of their production.

The welfare of cows is of enormous importance for farmers because only healthy and happy cows give the maximum amount of milk. The farmers are able to determine the health of their cows themselves, but this is only true for small herd sizes. Farmers lack the time to monitor each cow individually in herds of dozens or hundreds of cows as can be found nowadays [30]. This is why farmers are making more frequent use of tracking systems and data analytics for the automatic health monitoring of their herd.

These tracking systems take advantage of the architecture of modern barns in Central Europe and North America, in which cows can move

around freely. As a result, the everyday movement and activity behavior of cows is an important indicator of their health and whether they are in heat. In general, sick cows move less than cows without any diseases as shown in Figure 2. When cows are in heat, they move much more. The movement behavior is commonly measured with either accelerometers or pedometers embedded into the collar of each cow. These sensors are the central component in these systems as they are measuring the activity and vital parameters of the equipped cows continuously and autonomously.

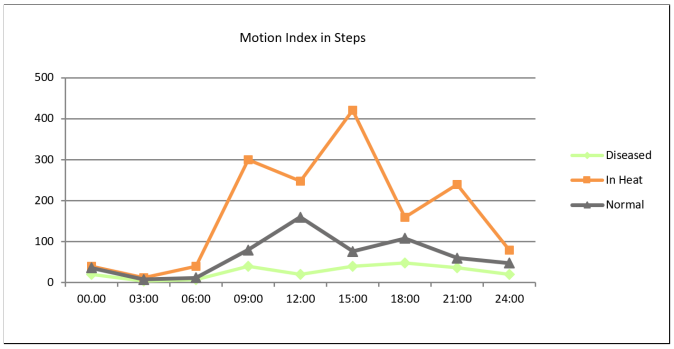

Fig. 2. Simplified movement behavior of cows. Different conditions of cows can be classified based on their movement patterns.

In more recent times, acceleration sensors are used instead of pedometers. They are superior since they cannot only recognize the amount of activity, but also the precise type of movement: walking, running, or lying. From a data perspective, the sensors are just counting steps, which do not tell the farmer anything directly about the health of a cow. However, the number of steps per day is a strong indicator, and it is directly linked to diseases and in heat detection of individual cows in the herd. The smart dairy products are sold by *SCR Europe, Lely, DairyMaster,* and *DeLaval.* They have all placed sensors in the collars of cows and the data is transferred wirelessly to the server station (in most cases, by using proprietary radio standards). By analyzing that data, data analytics provide meaningful information about the cow's health and notify when it is in heat. SCR Europe product named **Heatime** and Lely product named **Qwes-H** also integrate rumination detection. It tells the farmer how much time each cow spends on ruminating which is an essential indicator for their health and whether they are in heat if the average time per day differs significantly [31].

In the year 2014, the first tracking system based on locating cows within

the barns entered the market. **Smartbow** and **CowView** draw the diagnosis from positioning data of cows instead of using pedometers or accelerometers. Both systems utilize an ultra-wide band (UWB) RFID techniques in combination with an approach based on Time Difference of Arrival (TDOA) for locating the cows [32].

Indoor location techniques directly measure the distance traveled by cows instead of indirectly "guessing" them based on step counts or accelerometer values. Data analytics in this case works the same way as for the step count: under certain thresholds, which already had been figured out in studies a priori, cows are marked as in heat while they are classified as diseased above this threshold.

Beyond health monitoring, the determination of being in heat is a very sensitive process as the determination of the correct time is essential for a successful insemination. A failed insemination not only leads to repeated insemination costs, but also results in lower milk production. Nowadays, the insemination of cows in the dairy industry is done synthetically. In contrast to bulls (which can smell the hormones of cows and interpret their behavior), humans can only draw their conclusions based on the interpretation of their behavior. Studies show that the in heat observation plays a time-consuming role – three times a day, 15 minutes of observation are needed for complete heat detection (in addition to the normal working hours in the cowshed) [30]. It is understandable that the farmers need automated heat detection as an alternative to the time-consuming manual observation. The same kind of sensors as used for health monitoring can also be used for in heat detection. Cows in heat, feature a special characteristic in their movement behavior which significantly differs from healthy as well as diseased cows (see Figure 2). This movement behavior can be used to draw conclusions not only about health, but also about being in heat. The tracking systems help farmers to reduce their costs for insemination and again, increase their milk yield.

Now farmers are able to access data about the health and movement behaviors of their herd from their PC, notebook, or smartphone anywhere and at anytime. More importantly, they are notified if a cow shows an abnormal pattern like a reduced feeding behavior. These alarms enable the farmer to look after their cows and call a veterinarian if required before it is too late. Not only the welfare of cows, but also the economic performance of farmers is this improved. Sick cows cause high veterinarian and drug costs for the farmers. Tackling these issues in time also leads to a better yield due to increased milk production. As a result, modern IoT-based products

as well as data analytics improve the quality of dairy products and enable farmers to spend less time in the barn.

8. Data Analytics in Smart Homes

The IT market research company, Gartner predicts that in 2022, there will be more than 500 smart objects in an average family household [33]. The smart home market is now flooded with IoT based devices. Many of the manufacturers are embedding wireless data exchange and interoperability into their devices.

Heating control is one of the areas in smart homes where people can actually save money. Products like **Thermostat+** (by ELV) and **Comet Blue** (by EUROtronic Technology) can easily be installed without even drilling a single hole [34, 35]. Heating control devices are easily plugged onto radiators, and are commonly shipped together with sensor windows (to get the knowledge if the window is close or open) and a gateway. The gateway bridges the heating control devices wirelessly so that a PC or smartphone can control the whole system. These smart devices enable customers to define the rules for temperature by the room, and to control and monitor their heating remotely from anywhere. Customers can specify the required temperature and define different time slots when they are not at home. With the help of data analytics applied on the collected data, people can analyze their habits and behaviors to save energy and more importantly for them – money.

Radio-controlled sockets are cheap and small devices which can be plugged between normal sockets and the device to be powered, such as **Parce One** [36]. They are commonly equipped with Bluetooth 4.0 alias Low Energy and are easily connectible with modern Android or iPhone based smart phones. With the help of these smart sockets, i) customers can (gain the possibility to) monitor the exact power consumption of their electronic devices and, ii) they can define rules when the device gets switched on or off. With these smart sockets, all of the electrical devices can be turned into smart devices by switching them on and off autonomously. The data of the consumed energy can be analyzed per device, which gives customers the possibility to limit the use of a particular device which helps in minimizing the overall energy consumption.

The scope of smart homes is not confined only to the inside area of a home. Gardena is regularly offering new products in order to make gardens and gardening smart [37]. **Gardena's Sensor Control Set** contains a

smart gateway (which has to be installed indoors and connected via Wi-Fi or cable to the network), magnetic valves for taps, and plant sensors. The plant sensor measures temperature, soil humidity, and light intensity. These values can be used to define irrigation profiles. The goal of this application is to automatically identify if the plant needs some water or fertilizer. There is a link between the level of photosynthesis within a plant and its energy supply. Once the soil is dry, the magnetic valve is automatically opened. Customers can fine-tune the irrigation rules, for instance, based on the type of plant. Aquatic plants need more water than a cactus that will survive even if the soil is dry. Environmental factors complicate the data analytics part in this application field. However, a smart irrigation system saves a labor force and more importantly, helps plants to survive even if their owners are not present.

Now, most smart home devices are capable of measuring their surroundings, such as temperature, power consumption, or soil humidity. Additionally, they have the possibility to interact with their environment like switching off devices, activating the heating, or watering plants. What they currently lack is autonomous learning to interact with their environment based on the measured values. Nowadays, the customers still have to manually define some rules for each device. But, there are some systems which are becoming intelligent with the help of analyzing data from different sensors.

Apple wanted to change this situation with the development of **HomeKit**: a powerful, interoperable smart home control system which is easy and fast to set up and usable on iOS devices out of the box [38]. Certified vendors and products (which are currently limited in number) can be connected to iOS over Wi-Fi or Bluetooth 4.0. Afterwards, the connected devices can be verbally configured, controlled, and monitored via Apple Siri. Besides the fact that smart home devices have to support Apple's HomeKit and implement its functions into their system, Siri is not yet capable of communicating with people in a way one would expect or wish (to have it). For example, Siri only listens to commands containing the exact name of a device (which has to be defined a priori). General descriptions, which are often used in colloquial and everyday language, are not understood for now.

Vivint is one of the largest home automation companies in North America. Different smart home devices including small appliances, HVAC (Heating, ventilation and air conditioning), security systems, video devices, thermostats, smart doors and locks, smart bulbs, and smoke alarms are con-

nected via Vivint touchscreen panel and make a network of smart devices. That network produces a lot of streaming data, which is stored in Hadoop – an open source framework, for processing and storage of extremely large datasets. They use Datameer (a big data analytics platform) to shorten the time of using raw data for different analytics and actionable intelligence purposes [39]. The collected data is analyzed to better understand the usage patterns of different smart devices, which can be further used to improve the service and reduce energy consumption.

Google Nest offers smart devices including security cameras, thermostats, and smoke detectors. These are devices of daily use which have been in use for ages. But, data analytics and big data have changed the way these devices work. Before becoming 'smart', these devices were used to just record videos, maintain heating to a certain level, and sound the alarm when smoke is detected, respectively. Now, by learning user behavior, Nest's smart thermostat adapts to the user's usage and season changes. It automatically controls the temperature by learning the user schedule. By detecting unwanted events inside and outside a home, and making smart alerts, **Nest Aware** software makes security cameras intelligent. In contrast to the old security cameras which only record the video, Nest's smart security cameras can make custom alerts for the activities a user is interested in. By making the smart notifications, Nest's smoke detectors can tell the user (by speaking or by making mobile notification) in which room there is smoke and gives early warnings to avoid any emergency situation. It can distinguish between steam, food burn, carbon monoxide, and smoke. These smart devices can also be connected to each other to make a home safer and more secure. For example, security cameras, light bulbs, and window shades can work together to give an impression that you are at home when you are away. Or, when a thermostat is set to 'away', it can automatically turn on the security camera. By using data analytics, such smart devices can build up a profile which allows them to intelligently adjust themselves to the environment, minimize human effort, maximize human safety, improve service quality, and save energy [40].

The smart home vision affords many business opportunities, but also faces many challenges. Currently, smart devices are hindered by a lack of interoperability and the communication standard between products designed by different manufacturers. There are different products which are trying to integrate and bridge as many different products, protocols, and wireless standards as possible. **Mediola Gateway V4+** produced by Mediola supports both 433 and 868 MHz [41]. The advantage is that vari-

ous sensors and products of different manufacturers can interoperate which enable customers to mix them in rules and profiles. This works quite well; at least as long as Mediola supports them.

The smart home market is a mix of many different networking technologies and protocols, which are mostly proprietary and not designed for interoperation. All producers in the domain of smart home want a big piece of the cake to consolidate their market position. Thus, they are intending to raise barriers for new producers to enter this market by using proprietary protocols and prevent interoperability between different products. The market will most likely remain technically fragmented through 2020 [33]. From a consumer point of view, their biggest concern is data privacy. There is a need to develop a trust between the service provider and the consumer. It is very important for a consumer that the important information collected about their private life is only used to facilitate them, and not for earning money by selling that information to a third party without the consent of the consumer.

9. Conclusion

This chapter provides an insight into different industrial solutions available for data analytics. In addition to analytics on traditional data, most of these solutions are focusing on the data analytics on streaming/time-series data coming from IoT enabled devices. Almost all fields of life are benefiting from data analytics, including agriculture, healthcare, manufacturing, logistics, crowd analysis, dairy farming, smart homes, etc. This chapter attempts to provide a state-of-the-art in industrial/commercial data analytics solutions available in different fields. A deeper look at these solutions shows that there is still a lot of room and potential for improvement. Especially, most of the existing solutions are based either on traditional statistical-based approaches or to some extent using machine learning. However, almost none of the existing solutions is using the potential of deep learning, which could be very helpful to bring these analytics to the fingertips of data scientists and end users. Big players like Microsoft, Apple, SAP, and IBM have already developed cloud-based solutions, which are very suitable for streaming and time-series data. This already facilitates data scientists and companies a lot in terms of handling and managing the big streaming data from IoT devices. Companies are also making explicit efforts to evolve their analytics methods, which can deal with big data to gain maximum benefit from the collected data.

References

[1] E. Brown. Who Needs the Internet of Things? https://www.linux.com/news/who-needs-internet-things (Sept., 2016). Online; Accessed: 2017-01-01.
[2] General Electric. Smart home market. https://tinyurl.com/gr29h7v (May, 2016). Online; Accessed: 2016-12-14.
[3] J. Corbett. Localized Real-Time Agronomic Weather Data: A Big Data Foundation to Improve and Monitor Agriculture Production. http://www.awhere.com/about/news/localized-real-time-agronomic-weather-data-a%E2%80%9Dbig (June, 2016). Online; Accessed: 2017-02-02.
[4] S. Eathington. Pulling Back the Curtain on Our Innovation Pipeline. http://www.climateinsights.com/2017-innovation-pipeline/ (Jan., 2017). Online; Accessed: 2017-03-02.
[5] Benson Hill Biosystems, Inc. Benson Hill Biosystems Launches CropOS. http://bensonhillbio.com/benson-hill-biosystems-launches-cropostm/ (May, 2016). Online; Accessed: 2017-01-02.
[6] L. Henderson. CLAAS: The Harvesting Specialists. http://www.agrimarketing.com/s/38408 (Jan., 2006). Online; Accessed: 2017-01-02.
[7] IBM Corporation. PinnacleHealth System: Transforming healthcare delivery with insight into the "when" and "why" of readmissions. http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=AB&infotype=PM&appname=SWGE_YT_YT_USEN&htmlfid=YTC03945USEN&attachment=YTC03945USEN.PDF (2015). Online; Accessed: 2016-06-09.
[8] IBM Corporation. National Institutes of Health: Finding cures faster by transforming big data into valuable clinical insight. http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=AB&infotype=PM&htmlfid=IMC14949USEN&attachment=IMC14949USEN.PDF (July, 2015). Online; Accessed: 2016-06-09.
[9] SAP. Seoul National University Bundang Hospital: Transforming Patient Care and Data Access with SAP HANA. http://go.sap.com/documents/2015/09/d688d2a6-417c-0010-82c7-eda71af511fa.html (2015). Online; Accessed: 2016-06-15.
[10] E. Auschitzky, M. Hammer, and A. Rajagopaul. How Big Data can Improve Manufacturing. http://www.mckinsey.com/business-functions/operations/our-insights/how-big-data-can-improve-manufacturing (July, 2014). Online; Accessed: 2016-06-21.
[11] IBM Corporation. IBM Analytics. http://www.ibm.com/analytics/us/en/industry/# (2016). Online; Accessed: 2016-06-22.
[12] IBM Corporation. Using enhanced cognitive analytics to gain a competitive edge by finding valuable answers to questions not yet asked. http://presidionwp.s3-eu-west-1.amazonaws.com/wp-content/uploads/2015/08/Mueller.pdf (July, 2015). Online; Accessed: 2017-04-10.

[13] IBM Corporation. Big Data & Analytics in the Manufacturing Industry: The Vaasan Group. https://www.slideshare.net/IBMBDA/ibm-bda-vassangroupslidesharefinal (2014). Online; Accessed: 2017-04-10.

[14] SAP. Internet of Things. http://go.sap.com/solution/internet-of-things.html Online; Accessed: 2016-06-22.

[15] Microsoft Corporation. Azure IoT Suite. https://www.microsoft.com/en-us/cloud-platform/internet-of-things-azure-iot-suite Online; Accessed: 2016-06-22.

[16] General Electric. GE Launches Brilliant Manufacturing Suite to Help Manufacturers Increase Production Efficiency, Execution and Optimization through Advanced Analytics. https://www.ge.com/digital/press-releases/ge-launches-brilliant-manufacturing-suite (Sept., 2015). Online; Accessed: 2016-06-23.

[17] SAS Institute. Manufacturing. http://www.sas.com/en_us/industry/manufacturing.html Online; Accessed: 2016-06-23.

[18] R. van der Meulen and J. Rivera. A Quarter Billion Connected Vehicles Will Enable New In-Vehicle Services and Automated Driving Capabilities. http://www.gartner.com/newsroom/id/2970017 (Jan., 2015). Online; Accessed: 2016-06-14.

[19] R. Viereckl, D. Ahlemann, A. Koster, and S. Jursch. Connected Car Study 2015: Racing ahead with autonomous cars and digital innovation. http://www.strategyand.pwc.com/reports/connected-car-2015-study (Sept., 2015). Online; Accessed: 2016-06-14.

[20] IBM Corporation. Ford, IBM Use Data Analytics to Find More Efficient Ways to Move. http://www-03.ibm.com/press/us/en/pressrelease/48774.wss (Jan., 2016). Online; Accessed: 2016-06-21.

[21] M. Murphy. Ford Brakes the Mould with Car Maintenance Prediction Algorithm. https://tinyurl.com/m6ky3du (Jan., 2015). Online; Accessed: 2016-06-21.

[22] Daimler AG. Intelligent Driving. https://www.daimler.com/innovation/efficiency/intelligent-driving.html Online; Accessed: 2016-06-18.

[23] IBM Corporation. Leading German car manufacturer boosts customer satisfaction using IBM Big Data & Analytics. https://www-03.ibm.com/press/us/en/pressrelease/43392.wss (Mar., 2014). Online; Accessed: 2016-06-18.

[24] Computer Sciences Corporation. Volkswagen: On the Road to Big Data with Predictive Marketing in Aftermarket. https://snapshot.csc.com/auto/insights/101101-volkswagen_on_the_road_to_big_data_with_predictive_marketing Online; Accessed: 2016-06-16.

[25] P. Gittins. Tesla, big data and industrial disruption. https://www.capgemini.com/blog/insights-data-blog/2015/09/tesla-big-data-and-industrial-disruption (Sept., 2015). Online; Accessed: 2016-06-16.

[26] W. Grayson. Caterpillar's predictive diagnostics will enable more "repairs before failure" in future telematics updates. http://www.equipmentworld.com/caterpillar-predictive-diagnostics-telematics/ (Mar., 2015). Online; Accessed: 2016-06-20.
[27] M. Jeske, M. Grüner, and F. Weiß. BIG DATA IN LOGISTICS: A DHL perspective on how to move beyond the hype. http://www.dhl.com/content/dam/downloads/g0/about_us/innovation/CSI_Studie_BIG_DATA.pdf (Dec., 2013). Online; Accessed: 2016-06-19.
[28] N. Lomas. Amazon Patents "Anticipatory" Shipping To Start Sending Stuff Before You've Bought It. https://techcrunch.com/2014/01/18/amazon-pre-ships/ (Jan., 2014). Online; Accessed: 2016-06-19.
[29] N. Gilbert. Rules tighten on use of antibiotics on farms. http://www.nature.com/news/rules-tighten-on-use-of-antibiotics-on-farms-1.9761 (Jan., 2012).
[30] D. A. Wangler, Aktivitätsmessung zur brunsterkennung – möglichkeiten und nutzen, *Landeskontrollverband für Leistungs- und Qualitütsprüfung Sachsen-Anhalt e.V.* (Nov., 2009).
[31] L. Beringhoff, Kau-sensoren melden die brunst, *top agrar* (2014).
[32] Berkemeier, Ostermann-Palz, and Stöcker, Kuh-navis fr den stall, *Elite Magazin* (2015).
[33] Gartner, Inc. Gartner special report "digital business technologies". http://www.gartner.com/newsroom/id/2839717 (Sept., 2016). Online; Accessed: 2016-06-25.
[34] N. Jurran, Heizkrperthermostat: Eve thermo, *Magazin für Computertechnik c't.* **11**, 48 (May, 2016).
[35] N. Jurran, Bluetooth smart home, *Magazin für Computertechnik c't.* **15**, 59 (June, 2015).
[36] N. Jurran, Homekits anknipser, *Magazin für Computertechnik c't.* **14**, 56 (June, 2016).
[37] S. Hansen, Harke, schlauch, handy, *Magazin für Computertechnik c't.* **13**, 62 (June, 2016).
[38] E. Betters. Apple homekit and home app: What are they and how do they work? https://tinyurl.com/kcx7c84 (June, 2016). Online; Accessed: 2016-06-26.
[39] Datameer, Inc. Vivint Drives Smart Home Automation With Datameer. https://www.datameer.com/wp-content/uploads/2015/09/vivint_testimonial.pdf Online; Accessed: 2017-04-20.
[40] Nest Labs. https://nest.com/ Online; Accessed: 2017-04-24.
[41] S. Hansen, Das gateway – mediolas smart-home-hub, *Magazin fr Computertechnik c't.* **9**, 67 (Apr., 2016).