universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Big data analytics in supply chain management"

verfasst von / submitted by

Gabriella Farkas, BA

angestrebter akademischer Grad / in partial fulfilment of the requirement for the degree of

Master of Science (MSc)

Wien, 2019 / Vienna, 2019

Studienkennzahl lt. Studienblatt / Degree program code as it appears on the student record sheet:	A 066 915
Studienrichtung lt. Studienblatt / Degree program as it appears on the student record sheet:	Betriebswirtschaft UG2002
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Karl Franz Dörner, Privatdoz.

Abstract

I have researched a very popular area of nowadays in my thesis, where my goal was to study academic literature and summarize their results in the topic of data analytics and big data within supply chain management.

In the first step, I am trying to clarify the concepts related to the areas of data analytics as well as big data, as well as the relationship between their concepts, sub-parts, methods, and algorithms. This was a particularly challenging task, because there is not existing any generally accepted terminology and taxonomy about these elements of big data analytics yet.

As a next step, I am categorizing the research papers in accordance with the element of a supply chain planning matrix and demonstrating the tools, methods as well as algorithms have been used in order to optimise processes and enhance efficiency at a specific or overall part of the chain.

Finally, I am giving a brief overview in general about the benefits, challenges and hurdles of using these analytics.

Contents

List of Figures

List of Tables

1. Introduction

This master thesis is giving an overview about a highly discussed topic from nowadays life that is data analytics and big data on the field of supply chain management. My aim was to summarise and categorise all of the research papers and books that has written in this topic during the last decade as well as show what kind of benefits, challenges and barriers could bring the implementation of big data analytics with itself.

Due to the fact that it is a relatively young and popular topic several research papers have been issued and the newer ones continuously become public. Therefore, it is impossible to summarise all of the available literature in this topic so I had to select the best works from them. I took such factors into consideration for that like:

- how well know and acknowledged the authors are,
- how many citations the paper have,
- in which journal the paper has appeared,
- how old the paper is or
- whether a specific literature contains a new and value-added information.

In the next lines I would like to give a quick insight into my thesis structure which has three main chapters dealing with the topics of data analytics and big data, supply chain and logistics management and the combination of these two previously mentioned areas. First, I described all the terms that related to data analytics and big data as well as I tried to visualise how they relate to each other. This was a very challenging and consuming job as there is still not existing any generally accepted definitions about them and it has neither clearly defined the structure and hierarchy of these different approaches, methods and algorithms to each other. Because of this issue, I mostly used literatures from recognised scientists and professors who have already put milestones in this field and created a taxonomy by consolidating their works.

I have found important to introduce the term of supply chain management and thoroughly described the elements of the supply chain planning matrix. All these things are in the third chapter of my thesis and they are also completed with operational problems all along the chain that could be efficiently solved by data analytics.

Last, I summarise research papers focusing on data analytics and big data application in supply chain management. After I explained the relevant definitions and characteristics of this type of analytics I started to categorise the papers in accordance with the sub-parts of the supply chain

matrix then I divided the rest academic works into two groups, namely benefits-opportunities versus challenges-barriers of applying big data analytics.

2. Data analytics and big data

We are gathering vast amount of data in our digital world, especially through multimedia and social network, where this information could not be measured in terabytes[1] or petabytes[2] anymore. It is becoming more essential for businesses to understand what are those data hide and step to the path on continuous improvement which will further boost their competitive advantage. As a consequence, companies will be able to forecast customer behaviours more accurately, achieve data-driven customization or drastically enhance business productivity in such sectors as manufacturing, retail, government or healthcare (Min, 2016).

The gained information from these datasets could support decision-making, however it could be incorrect or misleading. This implies to gather data in better quality and quantity with the help of developed techniques. These challenges induce two strategic steps from company perspective: firstly, creating data-reduction mechanism with analytical tools and algorithms, secondly exploiting the integration and manipulation of datasets which contribute to a more effective data processing and storage (Min, 2016).

2.1. Taxonomy of terms

According to all researches up to nowadays we can say that data science is considered to be as the widest group of all the terms in the big data era which includes then all the terms and concepts related to different business intelligence as well as to advanced analytics. So we can differentiate these two greater groups under the umbrella of data science considering such facts like how structured or complex the data need to be processed, from which period data is coming from (e.g. focus on either past or present) or what the goal of data processing is (reporting, visualisation or sophisticated analysis). It is important to mention that business intelligence and analytics have already existed in the 1950s and they have developed continuously together through all these decades. Some researchers divided the evolution of them into three phases. Their techniques, methods and algorithms have become more and more sophisticated in each phase as business intelligence and advanced analytics were applied on even greater and new type of datasets (Chen, et al., 2012).

[1] Terabyte: it is equal ca. one trillion bytes or one thousand gigabytes. To give a good example about it – a one terabyte space could include 300 hours of good quality video or 1000 copies of Encyclopedia Britannica (Quintero, et al., 2015).

[2] Petabyte: it can be same as approximately one thousand terabytes or one million gigabytes. For example, it can hold 500 billion pages of standard printed text (Quintero, et al., 2015).

Business analytics – which is sub part of both business intelligence and advanced analytics - is dealing with such data and problems that usually come up in business life. If it is reporting with the use of historical data and solving less complex problems then we are talking about business analytics within business intelligence, otherwise it belongs to advanced analytics (Evans, 2017). There is the possibility that we divide into more sub-groups the previously mentioned areas like descriptive, predictive and prescriptive analytics. Descriptive analytics is also using data from the past and less complex, sophisticated methods and algorithms to identify some trend in the data set and to predict future "numbers". In this sense, descriptive analytics constitutes a big part of business intelligence and dealing with such popular tasks like SQL analytics, dashboards, metrics reporting or OLAP query (Rozados & Tjahjono, 2014).

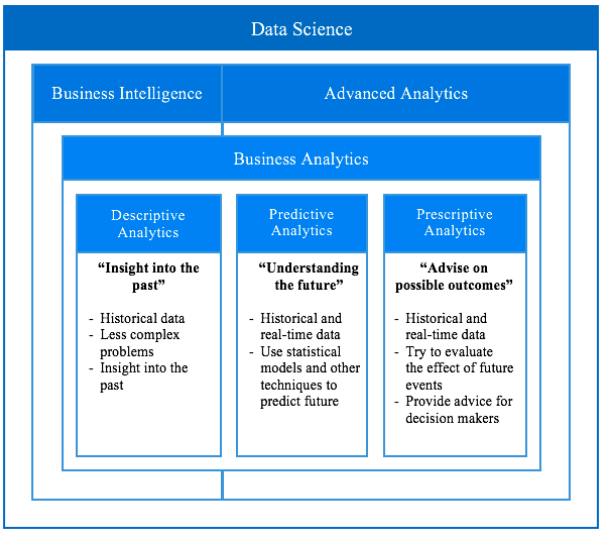

Figure 1: Elements of Data Science

In contrast to descriptive analytics, predictive and prescriptive analytics gather more sophisticated techniques and uses mainly real-time data, which are continuously processed by the algorithms. Shortly, they are using more advanced analytics. Describing predictive analytics, we can say that it applies such algorithms which are able to predict future events and outcomes by connecting statistical analysis with the learning ability of a system. Therefore, machine learning and data mining as two popular phenomena of data science can be mentioned within the framework of predictive analytics. At last, but not least prescriptive analytics constitutes the third sub part of advanced and business analytics. Here, simulation and optimisation techniques are applied on present data in order to predict outcomes as well as effect of future scenarios. The final result of the analysis should provide recommendations and advice for future actions for the decision-maker (Rozados & Tjahjono, 2014).

Finally, I would like to explain in more detail the relation between artificial intelligence (shortly AI), machine learning and data mining as well as locate them in the map of data science. First of all, AI constitutes the greatest group and it doesn't belong fully to data science. It has a common part with that covering predictive and prescriptive analytics, especially methods and algorithms of machine learning and data mining which are element of AI.

When we are talking about AI we need to differentiate between "narrow" artificial intelligence that is applied today by science and deals with carrying out only some specific tasks by machine learning, data mining or natural language processing. The second type is the so called general artificial intelligence, where the aim is to create a machine with a more advanced intelligence able to perform human interactions (Goertzel & Pennachin, 2007). Under machine learning we understand such systems which have the ability to learn from previous patterns discovered in the data set and apply this "knowledge" on similar problems later on. Both AI and Machine Learning are similar, however AI focuses not only learning from examples, definitions or behaviour, but it is also about reasoning and problem solving (Kersting, 2018). Data mining uses different algorithms to produce proper data for later analysis and it gives the basis for machine learning as it is in the need of appropriate information for the learning process. Both data mining and machine learning uses the same algorithms, however machine learning systems can learn on by own, in case of data mining this is led by humans (Brooks & Dahlke, 2017).

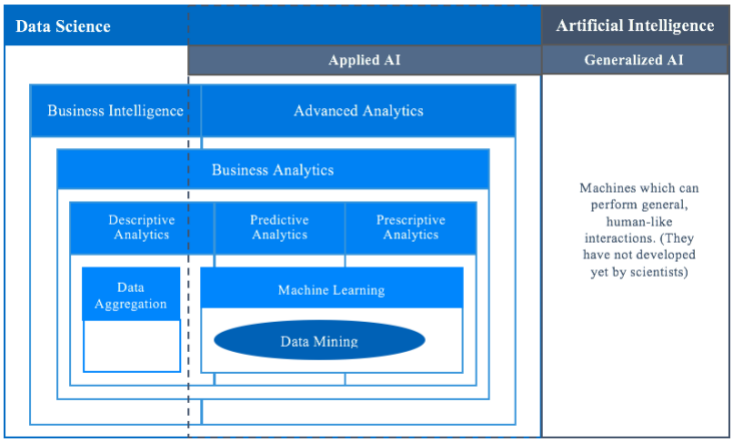

Figure 2: Map of terms relating to Data Science and AI

In the last few lines I would like to explain how big data and big data analytics related to all these terms/concepts I have just written about. It is clear that the techniques, tools and within that the algorithms of data science or AI are working with data. They have the capability to store, process, transform, analyse and gain information from any type of data varying from small to big. If the data considered to be big data according to specific characteristics then all the advanced analytics which work with these datasets named as big data analytics (Bhagat, 2015).

After this brief overview about the relationship among these new phenomena of data science and AI I am going to provide a more detailed summary about all their sub-elements, tasks, techniques and algorithms on the following pages.

2.2. Classification of terms

I dedicate this chapter for the description of all the previously mentioned concepts through the summary of well-known books and research papers published in the past years. The main topics would be data science, AI, business intelligence, advance analytics and business analytics. In

some cases, I am also going into details regarding their tools and techniques when I consider these methods necessary to be classified due to later chapters of my thesis.

2.2.1. Data Science

It applies quantitative and qualitative methods to solve relevant problems and make predictions. It is a group of methods which consist recording, storing, analysing and effectively gaining information from structured and unstructured data (Waller & Fawcett, 2013). It is part of a computers science however instead of including only programing and modelling algorithms it focuses on data analysis that may not only used by computers. It's close to statistics in a sense that its main task incorporating collection, organisation, analysis and presentation of data (Hernán, et al., 2018). However, professionals from diverse field of science considered differently the relation of statistics to data science. Some of them arguing that data science is equal with statistics because it is about analysing data which statisticians originally do for several decades. Others are more convinced about the fact that statistical analysis is dealing with rather small data samples in contrast to data science that aim is using scientific method for creating meaning from large scale raw data (Donoho, 2017). Further critics about why statistics does not relate significantly to data science:

- Statistics has not changed enough to respond quickly and efficiently of computer age problems which data science normally does
- Statistics has rather helped with pre-computer data for the people hypothesis-testing during the centuries, while data science focuses on retrieving, analysing and manipulate information on large data set (Carmichael & Marron, 2018)

Prowost and Facett also emphasise that this science is not just about data mining and if someone wants to be successful in that, one needs to examine the problems from a business point of view too (Provost & Fawcett, 2013).

It includes old and new methods from the field of machine learning, data mining, business intelligence and analytics – see the definition in the following later chapters - and its fundamental goal is to continuously discover new techniques for data analysis. Depending on business goals managers can choose data science either for a more sophisticated analyses or apply business intelligence for a simple reporting or visualisation of the data (EMC Education Services, 2015).

The tasks of data science are building three groups:

- Description: Shortly it is a quantitative summary of certain features of the world where the used techniques ranges from basic calculations (e.g. mean) to more sophisticated methods (e.g. cluster analysis)

- Prediction: This is using data to map inputs to outputs. It starts with easier tasks and simple analysis (e.g. correlation coefficient) and then become more complex (e.g. predict joint distribution of multiple variables).
- Causal inference: This means using data to model certain features of the world if it had been different. Calculations here usually connects to randomized experiments or generalized methods (Hernán, et al., 2018).

2.2.2. Business intelligence & advanced analytics

Researchers started to use the expression of "intelligence" in the 1950s which clearly referred to the artificial intelligence at that time. The term of business intelligence (shortly: BI) became popular from the 1990s in the field of business and IT, later on also the concept of business analytics was created as an essential analytical element of business intelligence. From the beginning of the 3rd millennium big data analytics gained a greater attention. These new phenomena basically mean that experts use such techniques and methods from business intelligence and from analytics that can handle huge and complex datasets. They were and they are not only existing tools developed in the last 50 decades, but more advanced analytical techniques which could enable the storage, management, analysis and visualisation of data. As it can be read in the previous lines existence of business intelligence and analytics (shortly: BI&A) have already covered several decades beginning from the mid 1900s. Some researcher divided into three phases of the evolution of BI&A marked them as BI&A 1.0, BI&A 2.0 and BI&A 3.0:

> The first phase of the BI&A has been developed from data management and warehousing and its methods based on such analytical and statistical techniques from the 1970s and in 1980s. To mention more concrete examples ETL[3], OLAP[4], database query and reporting tools are playing important roles with which companies can extract enterprise specific knowledge from dataset. The statistical analysis and data mining techniques (e.g. association analysis, regression analysis, classification etc.) are focusing on supporting different business fields. Numerous huge IT companies - like SAP, IBM, Microsoft -

[3] Extract, Transform, Load process is the way of get data out of source systems where the company`s operational data has already processed and stored in a normalised database. Then they are placed into business warehouse (Gronwald, 2017).

[4] Abbreviation for Online Analytical Processing that access multidimensional or relational data (example?) from business warehouse for analysis and data mining (Gronwald, 2017).

created BI&A platforms in order to help business in data processing and knowledge extraction (Chen, et al., 2012).

At the beginning of 2000s several new opportunities became available for BI&A due to the emerging trend of internet and web. It provided online platforms for observing costumers' preferences and constantly communicate with the buyers – best examples for that online store of Amazon and eBay. So businesses realised that they could define more accurate the needs and preferences of their clients by analysing their interaction on the web. BI&A 2.0 has been developed that was characterised by web intelligence, web and text mining centred around the analysis of unstructured web content. In contrast BI&A 1.0, techniques haven't been integrated into enterprise IT systems which created a future challenge to invent mature and scalable technologies of text mining, web mining or of social network analysis that could be incorporated accordingly into a business's system (Chen, et al., 2012).

BI&A 3.0. are part of nowadays life where the usage of mobile and other technological devices is rising, and these tools can reach several applications connecting computers and humanities through internet and web. Popular technologies of this new era are the sensor-based internet-enabled devices (like RFID[5]) through which the amount of arrived data are increasing even more and set such challenges for companies like how they could efficiently cope with the processing and exploiting the continuous inflow of sensor information or how they could create a proper integrated commercial system for business intelligence and analytics (Chen, et al., 2012).

Figure 3: Evolution of BI&A (Chen, et al., 2012)

[5] Abbreviation for Radio Frequency Identification and it is also called as a transponder that is attached to objects to count or identify and it has an antenna and a microchip for communication and storing information. Active tags are able to communicate with each other and with RFID reader, which is a transmitter in this system (Jia, et al., 2012).

As I have already written in the definition we must differentiate between two major concepts which are business intelligence and analytics where analytics were starting to refer to advanced analytics during the past years. In the following I will shortly compare them:

We find a definition of business intelligence from a managerial view which says that it collects the right information for the right people, in a right time in order to enhance company performance and support decision making. Business intelligence also exits for several years, so during these time companies could collect vast amount of data which exceeded their storage capacities. As a result, they organised data bases into data warehouses which are core element of BI programs nowadays. Originally BI was equal with OLAP and reporting tools but after a while enterprises realised that if they want to use the gathered information effectively it won't be enough to report historical data only, they have to move towards the more adaptive advanced analytics (Bose, 2008). Examining the time horizon, BI could produce reports and dashboards referring to past and present events and most of the time it answers questions which are helpful to examine past cases such as quarterly targets or yearly sales (EMC Education Services, 2015). In contrast to BI advanced analytics uses more sophisticated modelling techniques and focuses rather on predicting future events and explore patterns (Bose, 2008). There is also the possibility to divide BI and analytics into further sub-parts:

- Descriptive analytics try to explore what has happened in the past and make visible trends and expectations (Rozados & Tjahjono, 2014) The main techniques in this group are standard reporting and dashboards, ad hoc reporting, OLAP and visualisation (Zeng, et al., 2011).
- Predictive analytics focuses on the present and analyse real time information from which it tries to predict future events (Rozados & Tjahjono, 2014). Its tools are algorithmic based and some of the technologies used here have the capability to learn from data (Siegel, 2013). The typical examples for that are data mining and machine learning, predictive analytics also includes advanced forecasting and time series methods (very popular with supply chain management and marketing), clustering, plus supervised learning with regression and statistical algorithms (Rozados & Tjahjono, 2014).
- Prescriptive analytics use optimisation and simulation techniques based on data to forecast different outcomes of future scenarios. In contrast to the other two analytics here the future and effect of future decisions in the centre for which prescriptive analytics try to provide advice as well as recommendations. It is embedded uncertainty

and variably as the techniques derives from what/if analysis and game theory (Rozados & Tjahjono, 2014).

2.2.3. Business analytics

It is an application of business intelligence and advanced analytics techniques to solve different business problems. Companies use these organisational procedures and tools to support their data-driven decision-making processes as well as predict outcomes of problem solutions (Trkman, et al., 2010). Business Analytics are considering past and present while using its application and mainly focus on day-to-day operational issues. It has built a large part of data science where the more advanced analytical techniques able to solve complex problems using data of any field of business life. Complexity of analytical techniques has a wide range from the simple analytics dealing with structured, easy-to-handle data till more complicated ones which are used on the area of advanced analytics, like in machine learning (Evans, 2017).

2.2.4. Artificial intelligence

We are talking about artificial intelligence if a machine is able to carry out its tasks independently and on an intelligent way. It doesn't perform task repeatedly which were previously programmed in it, but it can learn from the data which enables it to solve new, different and more complex task as well as to adapt to the environment. AI has been grounded with the invention and usage of first computers, although these machines were not good enough to make decisions only by themselves. However, the scientists were inspired to develop them further and create a machine with logic that can already remember of, learn from and adapt to the information just like a human brain does (Ertel, 2017). By today this phenomenon has developed such a manner that science and practice differentiate between two types of AI: general and "narrow" artificial intelligence. Within the framework of the first one scientists try to create machines with such intelligence that similar or beyond to human's general intelligence. This would mean that an intelligent machine could have understanding, reasoning skills and it could handle different topics, cases and act like a human being. In reality developing such intelligence is a very challenging job so scientists have not yet invented a machine with an intelligence like a young child has (Goertzel & Pennachin, 2007). The second type of artificial intelligence is more common and broadly applied nowadays. Here, the machines have been transformed as a high-functioning systems which could outperform in a specific area for a dedicated purpose. Machine learning and natural language processing are two advanced algorithms of applied intelligence which are not only automating routing task or scanning databases, but these more developed capabilities help machines to come up with recommendations or generate workflows to complete request of users (Ertel, 2017).

2.2.5. Machine learning

Machine learning is a group of algorithms from intelligent software which is able to learn and adapt by itself in contrast to other applications which are explicitly run and developed (by humans). We can identify a connection between data mining and machine learning such as data mining is an information source for the latter one. Both of them use the same algorithms during regression, classification or clustering tasks, but machine learning applications are also capable to learn from the data and then develop themselves further according to the discovered knowledge in the data set. In case of data mining human interaction is necessary to reach a similar result (Brooks & Dahlke, 2017).

Machine learning techniques are divided into two sub-categories: supervised and unsupervised learning. We are talking about supervised learning when there is a training set with a given right answer for every training algorithm (Dangeti, 2017). Here, an important step is the training of the machine when it gets set of training examples as an input in order to learning from them. The supervised learning algorithm study these examples for creating an inferred function in order to map new examples and get the final target value. Regression and classification from data mining techniques stand under the umbrella of supervised machine learning (Harrington, 2012). In case of unsupervised learning the machine works with unlabeled data - without any meaningful and informative tag or class - with a learning algorithm for finding a structure within the data set (Dangeti, 2017). Here, the user is not able to train the machine as the input data do not contain any sample for learning, the training examples are missing. It has to learn from the data set by itself through examining the structure of the data or distribution of elements. Clustering is one best example for it combining data mining and machine learning disciplines (Huang, et al., 2006).

2.2.6. Data mining

It started to develop during the decades after 1990 as a sub-discipline of artificial intelligence. It uses techniques and methods either from area of statistical data analysis and from machine learning. Data mining has not revolutionised the technological field of AI, it has rather created a requirement that machines and applications work on larger data sets and gain more knowledge from them (Ertel, 2017).

Data mining means extracting previously unknown information or knowledge from databases in a more complex, not trivial way (Bose, 2008). Data mining techniques are able to explore such patterns within data, which would never been discovered and known otherwise, as well as they can predict outcomes of future actions (Tan, et al., 2006). It is a key element of a knowledge discovery in databases (shortly: KDD), a sub-part or status of knowledge discovery

where the identification of patterns and hidden relations in a dataset are happening (Bose, 2008). The structure of KDD includes three main steps like data pre-processing, data mining and data post-processing (Tan, et al., 2006).

Figure 4: Main steps of KDD process (Tan, et al. 2006)

The process begins with the collection of input data from different sources, cleaning them and remove noise. Then we come to data mining where the previously mentioned actions take place. Finally, the integration of data mining results into the system closes the second step. Under post-processing we understand the validation of these results for make it sure that only useful and correct results are put into the decision support system (Tan, et al., 2006). The three main step of the iterative and interactive KDD process can also be divided into nine steps starting with defining KDD goals and ending with the integration of the discovered knowledge. At the time when the process closes the effects are measured in a new data mining repositories. As a last step, these results are evaluated and then the KDD process runs again. The nine steps can be described shortly as the followings:

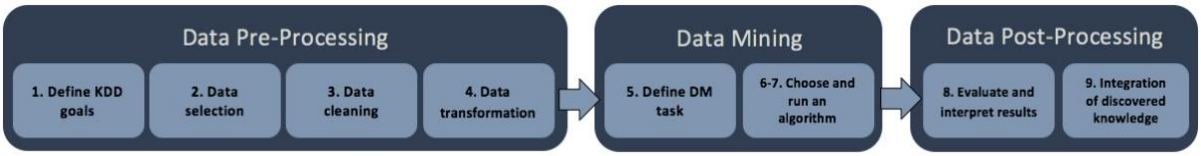

Figure 5: Nine steps of KDD process (Maimon & Rokach, 2005)

First steps are within Data Pre-Processing:

1. The first step includes the preparation of what a user can do with several decisions regarding transformation, algorithms or representation. Here, the users define the environment where the KDD will take place and the goals of the final users.
2. Selecting and creating data set where discovery will run are happening here. More exactly, the users find out which data are available, what kind of data is still necessary to be collected and they also put all the gathered data into one data set which contains important attributes for the processing.
3. As a next step data cleansing takes place in order to enhance the reliability of data. Here the main tasks are handling missing values and cleaning the data out from noise and outliers.

4. Through the transformation, we can get such data that are better used in data mining. The most common applied methods here are dimension reduction (e.g. selecting out features) and attribute transformation (Maimon & Rokach, 2005).

The following steps belongs to Data Mining and the algorithms used during it:

5. The previously defined KDD goals and the actions taken in the previous steps can enable the user to choose a proper data mining task for the knowledge discovery such as clustering, classification, regression etc. Data mining strategy is developed in this phase considering the available data and what extent these data can be used for inductive learning model.
6. After defining the goals and task of data mining the user will chose a concrete algorithm for searching patterns in the data set while considering several indicators (accuracy vs. understandability).
7. As a next step the algorithm is launched and run several times until the user reach a acceptable result (Maimon & Rokach, 2005).

The last two steps fall under the Data Post-Processing phase:

8. The result is going to be evaluated and interpreted in accordance with the goals that were set in the first step. Here, the attention is on the usefulness and comprehensiveness of the model and the documentation of the discovered knowledge for later works.
9. As a final step, the knowledge is started to be implemented and used in another system for further actions. In this last step a user could measure the real efficiency and success of the KDD process. The integration of the discovered knowledge is a big challenge as the previous conditions could quickly change (e.g. data structures, change from static to dynamic data and so on) and the system needs to adapt again to the new environment as rapidly as it can (Maimon & Rokach, 2005).

We can also distinguish among some storage types where the mining usually takes place. The most known are relational database, transactional database, data warehouse, advanced database systems (Han & Kamber, 2000).

- Relational databases or database management systems gathers interrelated data and a set of software programmes through which the information can be reached. The database contains several uniquely named tables with a lot of tuples (like rows or records) and attributes (like columns). Here, we also talk about object for each tuple "identified by a unique key and described by a set of attribute values" in the table. The easiest way to access data and extract information from such repository is the database

query using relational query language (like SQL). It enables the user to carry out relational operations (like join, select) or use aggregate functions (like sum, avarage). Finally, it can be also successfully applied for searching trends and data patterns such as predicting risk in case of different business transactions or detect deviations (Han & Kamber, 2000).

- Transactional database stores such records which have something to do with a transaction. It is common to enlarge the database with additional tables which also connect to the similar transactions (eg. customer ID, sales person ID etc.). It is also neccessary to mention here the market basket data analysis through which it is possible to find out which products are sold well together and increase the sales so (eg. dipers-milk, computer-printer etc.) (Han & Kamber, 2000).
- Data Warehouse is a place where data derived from multiple sources and they are stored by some unique scheme on a single place. Data are organised here according to subject (like customer, supplier, item etc.) and have historical perspective (e.g. give insigt in the last 5 years). Data warehouse is also a place where data cleansing, transformation, integration, loading and periodical data refreshing could take place. The warehouse is usally visualised as a cube with multiple dimensions related to an attribute or set of attributes and the cells in the cube represent a value of some aggregate measure. This physical structure of data warehouse has different level depending on the complexity of stored information. The base cuboid is a cube containing primitiv information while non-base cuboid deals with higher level multidimensional structures. These two cuboids together shape a data cube (Han & Kamber, 2000).
- Advanced database systems were developed during the years to face and solve the new challanges for data mining. These are able to handle spatial data, engineering design data, multimedia data and also time-related data with efficient data structures and scalable methods. Object-oriented and object-relational database systems are usually exploited here to gain useful knowladge from the data set (Pandey, et al., 2015).

Understanding the relation among data mining task and methods we should create a good taxonomy. First of all, we must separate two types of data mining: verification and discovery. Through verification the evaluation of hypothesis takes place and commonly used such methods like traditional statistics, t-test of hypothesis or variance analysis. However, it is weakly connected to data mining as its methods are hardly dealing with discovering new knowladge, they are rather working on testing. In contrast to verification, discovery types of data mining is

dealing with finding new patterns and knowladge from the beginning. Within this we could make a difference between predictive and descriptive task. The techniques of discovery focus on inductive learning in which case the model is able to learn from training examples and used an iductive approach to solve future examples (Maimon & Rokach, 2005).

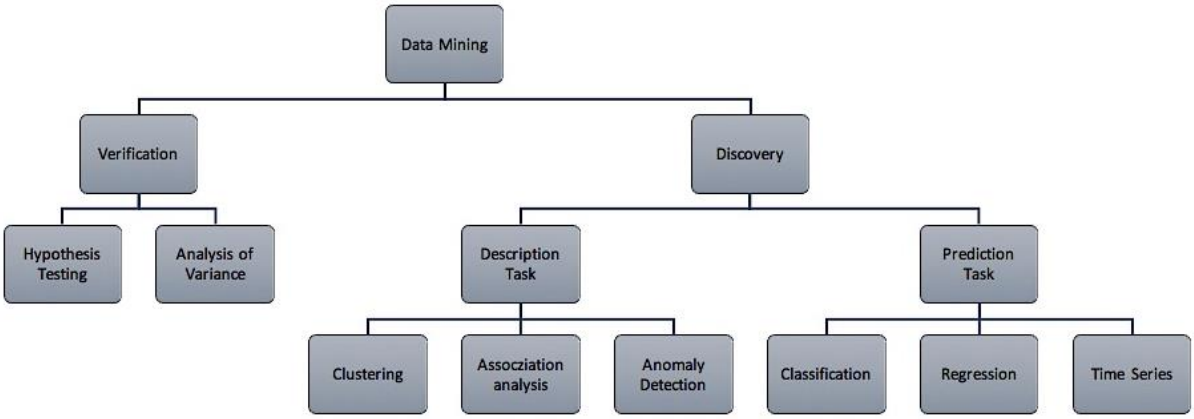

Figure 6: Data Mining with its tasks and methods (Maimon & Rokach, 2005)

Predictive tasks use target/dependent variables and explanatory/independent variables to predict a value of a particular attribute. The following methods are commonly used in business or by academic researchers (Tan, et al., 2006):

- *Regression* is a general analytical method which discovers the relationship among the dependent variable and several other independent variables. With the help of the regression function and the equitation the model attempt to describe how strong the independent variables can explain the dependent one and which of the independent variables have the most intense influence on the outcome. Two commonly used types of regression analysis are linear and logistic regression. The first one usually observes the connection between several input variables and a continuous dependent variable if the model considered to be a linear one. The second one is more able to predict the probability of the outcome with the help of the input variables. Outcome variable has multiple values however in most cases we are speaking about a binary variable (EMC Education Services, 2015).
- *Classification* is a kind of process where a specialist classifies unstructured data into structured datasets. The first phase of the method is the learning process when the analysis on the training data set takes place and then one creates rules and patterns. In the second step there is an evaluation of dataset and a storage of the classification's accuracy (Koturwar, et al., 2015). Classification models are existing in several forms

like decision tree, neural network or Bayesian network. One of the most popular classification model is the decision tree where each node represents a test on an attribute value, then the branches shows the outcome of these tests and the leaves of the tree always indicate a classes or class distributions (Han & Kamber, 2000).

- *Time Series Analysis* is a task from statistics which deals with time series data or trend analysis. Time series data are chronological order of data that were measured through a defined time period in the past and its purpose is to forecast future values. This mostly applies in economics, retail, manufacturing and finance where most specific examples can be listed such as retail sales forecasting, spare parts planning as well as pair trading on the stock exchange (EMC Education Services, 2015).

Another main task can be named as descriptive where the aim is to derive patterns, which explain the relationships in the data. These task are mainly explanatory therefore they require post-processing activity to examine the validity of the results or explain them. We can differentiate three main groups of techniques here (Tan, et al., 2006):

- *Clustering* means grouping of objects according to data that can be found in information. This is an unsupervised process in a sense that an analyst does not set any label beforehand so the grouping happens by any logic that the system finds optimal. The goal of the whole method is to create such groups in which the characteristics of the elements are very similar or there is a strong connection between the group members. Clustering which is a part of explanatory analysis does not use any prediction (EMC Education Services, 2015). I also have to mention shortly the commonly used k-means method within clustering, Here, number of k clusters are chosen and initial values of the centres are assigned. In the next step each element from the data set assigned to a cluster with the closest centre, then there will be a computation of new centres. This will be repeated until the convergence criterion (e.g. squared error) is satisfied (Kurasova, et al., 2014).
- *Association analysis* can be characterised as a data mining task which aim is to explore association rules and representing an attribute-value condition in a data set (Han & Kamber, 2000). It is another unsupervised, descriptive learning method, which finds interesting connections, frequent patterns and causation hidden in the dataset (EMC Education Services, 2015). The typical example of discovering facts often happens together within a certain dataset, just like finding out which products are bought together in a supermarket. Agrawal an his colleague were the first researchers who formulated

association rules mathematically (Agrawal & Srikant, 1994). They defined a set of items and another set of transactions with a subset of items. In the rule they noted the antecedent by X and the consequence by Y. Frequency is a key element of the method which measured in case of appearance of each element in the dataset and comparing it with a minimum support threshold (Fernandez-Basso, et al., 2016). With the help of this rules one can clearly see whether X was observed then Y was also observed (EMC Education Services, 2015).

- *Anomaly Detection* task is a good way to identify such elements in the data set which characteristics are considerably different from other data. These "members" of the set are called anomalies or outliers in practice where the goal is to identify as much them as possible and avoid to label those which are data with normal features. Here, fraud and network intrusion detection or signalling ecosystem disturbances gain an especially important role (Tan, et al., 2006).

2.2.7. Big data

At the end of the 1950s scientists has already used an alternative model for a "synoptic collection of observational data on a global scale" within the framework of an international scientific project. This is considered to be a first big step toward big data anylitics as the project transformed the way data were collected and shared, provided newer analysis technologies, nevertheless put into the public eyes how IT technologies could further be used for data analysis (Ohlhorst, 2012).

The next quick expansion of big data derived from the digitisation of the 1990s, when companies aim was to transform the analogue data to a more readable digital format. That could imply a more structured data sets from which companies could easily identified weaknesses in the processes or more accurately predict future scenarios (Bhagat, 2015). Towards the millennium, Big Business was born and changed the use of big data. It has discovered new opportunities, found relationships and measured efficiencies, on a very similar way as business analysts do nowadays (Ohlhorst, 2012).

Recently big data is referring to an enormous set of unstructured data which requires real-time analysis in contrast to traditional datasets. According to statistics, we are producing data in one million petabytes, which is growing around 40% per year. The most data are produced by mobile devices and social media enterprises such as Google, Facebook, Twitter or other multinationals like Apple (Bhagat, 2015).

As we are speaking about more complex data the methods and the techniques with which we could gain valuable information from them are differing from the general procedures. With the

help of analysing traditional and logical connection among several events derived from information data decision-making has been formed from a static to a dynamic one (Hammer, et al., 2017).

The quick evolution of big data brings several definitions with itself. In the following, I am going to show the most popular and universally accepted terms that were developed by scientific researchers through years.

The term "big data" first appeared in 1998 in a presentation hold by John Mashley with the title of "Big Data and the next wave of InfraStress". The first book, which mentioned this word, originally dealt with data mining and issued in 1998. In case of academic papers, we had to wait two more years when in 2000 Diebold wrote about it in his research (Bhagat, 2015).

Big data is also considered as a situation in which data sets are growing and will continue to grow so rapidly and so large that conventional technologies are not able to handle this size anymore. Acquisition, storage, searching, sharing, analytics and visualisation are especially difficult to manage in case of this vast amount of data. It is also necessary to mention an interesting fact - the term "big data" became as a synonym for business intelligence, business analytics or data mining in many literatures (Ohlhorst, 2012).

The most frequently used term created by Gartner analyst company:

"Big Data is a high –volume, high- velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making." (Gartner, cited in Gandomi & Haider, 2014, pp.138)

This definition is also supported by the Data Protection Party of the European Union which uses big data for exponentially growing amount of data set - any huge data base which used by government, national administrative bodes and companies or analysed extensive by computer algorithms while identifying general trends and connections among them (European Union, 2013).

Andrea DeMauro, Greco and Grimaldi (2016) examined large amount of definitions in their academic papers and arrived to the following conclusions:

> The definition of big data also includes technological and analytical methods, it clearly defines what kind of tools and techniques we need to use to collect data which main characteristics are volume, velocity and variety. They also emphasise that big data has a value, because all the analysis, which could be gained from this data set, produce value-added information.

This definition already reflects the three main characteristics of big data also named as a popular 3 Vs. Later on, other researchers added 2 more Vs to the concept like value and veracity (Demchenko, et al., 2013):

Volume

It simply means the amount of data which is generated on each day by business participants. The amount of data is so large that it's impossible to save and analyse them by using general data processing methods (Demchenko, et al., 2013). There is not an accepted threshold for magnitude but the companies and professionals consider data mainly in volume of terabyte or petabyte as Big Data (McCarthy, et al., 2019).

Velocity

It refers to the speed of data generating, processing and analysing. Big data is also "live" in the sense that data are continuously created and flowing through the system. Parallel to this the speed of processing and interpreting them are also increasing (Demchenko, et al., 2013).

Variety

It refers to the number of types of data. Nowadays more than 50% of the data in the IT systems are unstructured, which means they are not organised into tables they don't have formal structures. As a consequence, it is hard to interpret them or identify any relations at first (Ohlhorst, 2012).

Value

It means the added value to a company as a lot of corporations invest huge amount of money to create their own big data platform in order to generate value for their own businesses (Demchenko, et al., 2013).

Veracity/Validity

This characteristic wants to capture the quality of big data in opposite to Volume features which is considered as a features lack of quality (Demchenko, et al., 2013).

After I have found several definitions for big data as well as the characteristics of big data there is still one question remained open – what makes big data big data? In order to find an answer for that I am summarising the research paper of Kitchin and McArdle (2016) in which the two researchers are dealing with exactly the same problem. The authors are examining the 3Vs characteristics and other attributes of a different type of data sets and trying to define specific characteristics or specific type of data that can clearly fall under the umbrella of big data. In

addition to the *volume*, *velocity* and *variety* they are also considering *exhaustivity*, *extensionality*, and *scalability*. Kitchin has already written about the difference between big data and general, "small" data - he found out that production of small data is happening in a traditional, controlled way using sampling for processing. In contrast to big data, the generation and administration of it considered to be highly inflexible as well as not scalable (Kitchin, cited in Kitchin & McArdle, 2016). In order to find a more exact answer for the question when we are talking about Big Data the researchers examined a data sample included 26 data types, which are considered to belong to the group of big data according to previous literatures and researches. These data types are from seven different domains, such as social media data, website, mobile communication data, data from sensors, camera, transaction process generated data and administrative data. During the research they were examined by each of the previously mentioned characteristics:

- If we are talking about the *volume* of big data it is usually considered to be so huge that the storage capacity of systems must cope with the amount of data in terabytes or petabytes. This could only be realised in clouds covering several systems and locations. However, the research gave the result that in case of some data set it is not necessary to have large storage capacity as the volume of generated data is huge but each of the data are only in bytes/kilobytes. Kitchin and his colleague concluded that in case of this characteristic it is impossible to define a threshold for big data.
- In case of *velocity* big data are normally produced and collected continuously, in real time and not dealing with samples where there is a temporal gap. Here, frequency in generation, recording and handling of data are playing an important role and considered to be as a key characteristic of them.
- The weakest characteristics of all is the *variety* after the examination of the 26 data set. Both small and big data are heterogeneous which means any of them can be structured, semi-structured or unstructured.
- *Exhausivity* could also characterised all of the examined 26 datasets that actually includes such features like fine-grained solution, indexicality and relationality. In every case data go hand in hand with metadata through which it can be identified if data produced by similar tools but in different time or place as well as if there is any relationship among datasets. However, in case of normal data from the 26 datasets sample this relationality is not so obvious in many cases.

- In case of *extensionality*, the researchers observed if the generated big data is variable and required a highly flexible system that is capable to add or remove fields with respect of the collected data. After the research, it became already clear that in some cases set of big data hold the feature of continuality and robustness without the necessity of an adaptable system.
- The last attribute is *scalability* where the main question was if big data requires a system that is able to handle varying amount of data inflow. The research brought a result that most of the cases inflow of data is connected to specific intervals or it is simply continuous, therefore scalability could not unambiguously characterise big data (Kitchin & McArdle, 2016).

Considering all the facts what Kitchin and McArdle (2016) have found during the research we can say that there are several forms of big data and they cannot be characterised by same attributes. However, they are arguing that velocity and exhaustivity are the most important features, which means that if a data holds these two attributes, then we can consider them as big data. Otherwise, it can be misleading an unclear whether we are facing with big data or not.

Transformation of big data to a value-added information was explained thoroughly first in 2016. which was completed at some parts by findings of other researchers, too. The following picture is showing the main steps which will be described more briefly later on (Min, 2016):

Figure 7: Big Data transformation to valuable information (Min, 2016)

Huge number of data do not explicitly imply that the company will *find the right data* for sure. The datasets should be revised and transformed in order to get a valuable information. In order to reveal all of the information within that, one could examine the whole population of the gather data which clearly would be very costly, also time consuming. *Data sampling* is used to bridging this issue which means one can extract a smaller group from the whole amount of data, that considered to be representative, and make conclusions. It is always a current issue how analysts choose the correct sampling method that fits the most to the business entity's problem. When it comes to this step there are two questions which are necessary to be answered: How often and how many data should be collected?

Till now I was written about raw data which has to be transformed to such format that analyst can gain relevant information from them. The authors call this process *data preparation* which includes the following tasks: exploring, cleansing, changing, shaping, screening, profiling, integrating and publishing. After the transformation, experts can carry through efficient analysis, furthermore they will easily recognise patterns within the set.

The fourth step is the *segmentation* which sounds very familiar from marketing, and its important task is to classifies costumers into smaller groups according to some aspects. Here, the enterprises gain insights into customer behaviour, reactions in different situations or buyers' opinion about products. Data segmentation simply means that a dataset is divided into smaller categories by some characteristics (e.g. demographic data), by privacy (e.g. sensitive), by structure (e.g. codified) or by format (e.g. SPSS[6]). The aim is to achieve customisation as well as personalization.

In addition to this, we have *data filtering* for refining Big Data analysis and cutting down the amount of data with a help of specific tools and display only those records in which analysts are interested and meet the criteria previously set up by them.

While a company is continuously collecting this vast amount of information it comes to question how and where it could be stored correctly. It induces the necessity of a new storage and architecture which is more state of the art comparing with a traditional data storage like, data bases, data marts and data warehouses (Bakhsi, 2012).

Data Warehouse is a single space for storing big data. It supports decision makers to find easily the useful information in a such system where the data have already been transformed into an applicable format and documented properly. This relational database already included the subject oriented, time variant, integrated data which is impossible to removed or altered after it is stored in a structured form. The raw data can derive from three sources like external, operational and independent data mart. Their different formats will be integrated into a single, particular one then let them flow into the data warehouse system by the time they are already ready for query, analysis and reporting. Extract-Transform-Load process takes place after the extraction of the data and it prepares the data in a way to avoid inconsistency and shapes to a relational and multidimensional format, and finally loads them into the warehouse.

[6] The acronym SPSS stands for Statistical Package for Social Sciences and broadly used for statistical analysis tool in academic areas (Burns & Burns, 2008).

As a last step business decision makers must *choose the best analytical methods which fits to the collected data* in order to gain better insights into business operations, practices. After the right method was selected the extracted information has to be presented in a more user-friendly format as it still needs to be understandable for those who are not technical experts, but still want to gain knowledge from them (Min, 2016).

2.2.8. Big data analytics

This is the part of advanced analytics, where the experts apply advanced analytic techniques on big data (Bhagat, 2015). Besides just collecting data, it is more important to understand the hidden information of these datasets through analysis. During the process of analytics, there is the possibility to discover unknown, yet useful and valid patterns, relations that have even greater meaning in the business life of decision-makers (Elgendy & Elragal, 2014). One can use such tools, which are able to further reduce cost and save times regarding analytical process, while companies gain valuable information for decisions at the same time. Scientists thrive for such a quick analytical process that we can already name as real-time analysis. The most commonly used techniques, which provide the best insight, are predictive analytics, text mining, data mining, forecasting and optimisation (Russom, 2011).

2.3. Technologies in analytics

As a last part of this chapter I would like to shortly write about the most well know technologies have been used in analytics for decades, plus it is also necessary to mention them for better understanding of later part in my thesis. First, I show what characterises SQL and NoSQL, then I will introduce a Google invention called MapReduce and its open source project, Hadoop.

Standard Query Language (shortly: SQL) and NoSQL are high-level interfaces in order to reach data in database systems. The main differences among them is that SQL uses relational database while not only SQL is a non-relational database management system. In the era of big data, this latest query program draws a bigger attention because of its flexibility, quicker data processing and better performance. Leading internet companies like Amazon, Google or LinkedIn also attempt to exploit these advantages. The following table on the right hand side provides a good summary about the differences between SQL and NoSQL (Shetty & Chidimar, 2016).

	SQL	NoSQL
Data type	Record with same attribute Structured data	Records with different attribute Semi-structured and unstructured data
Structure	Store data in several logical table in order to avoid redundancy and duplication	Store data in a form of flat collections where data is duplicated (key-value pair, JSON data)
Transactions	Working with ACID[7] transactions	Depends on different solutions
Interactions	Declarative query language	Procedural query language

Table 1: Differences between SQL and NoSQL (Shetty & Chidimar, 2016)

It is also necessary to mention one of the most famous innovation connecting to big data phenomena that called MapReduce. It is a programming model for handling large data set with a map function processing key/value pairs. There are more than 10.000 programs, which used MapReduce with different algorithms for text processing, machine learning or graph processing (Dean & Ghemawat, 2010).

Hadoop is an open source project implementation of MapReduce, which help analysts to cope with problems related to this vast amount of data especially when both unstructured and structured data belong to a data set. Furthermore, it provides a support in such situations when analytics are deep and computationally extensive (e.g. clustering) and solve the problems of storing and accessing to the huge amount of data for decision-makers. Further benefits of Hadoop are that it can work on several machines, which do not share any memory, as well as leverage the management from excessive, unclear information by breaking the data into smaller, more manageable pieces (Ohlhorst, 2012).

[7] ACID refers to four characteristics of a data, namely atomicity, consistency, isolation and durability (Yadava, 2007).

3. Supply chain management

In this chapter I am going to describe the concepts of supply chain management as well as summarize the main characteristics, challenges and issues belong to it. In the main part I will write every element of the supply chain and planning matrix thoroughly, structure them to long-, mid – or short-term goals and finally show their importance in the decision-making process.

3.1. Definitions of supply chain management

I would like to start with the definition of supply chain management, which was created by several academic scientist during the years. These terms are different in their extent and detail, however all of them capture the core principles of supply chain management. Although we could find several definitions which created in the last 40 years there is one common thing all of them: they see different elements of production as a part of a larger, complex organisation not as an individual, separate process. Though, supply chain management is considered within an enterprise, its activities are reached far beyond a company's "boarders" as it is based on partnerships from supplier and customer side (Lu, 2011). Without being exhaustive, the most well knows definitions of supply chain management sound as a following:

- One of the most accepted general definition characterises supply chain management as a "set of approaches utilised to effectively integrate suppliers, manufacturers, warehouses, and stores so that merchandise is produced and distributed at the right quantities, to the right locations, and at the right time, in order to minimise system-wide costs while satisfying service level requirements" (Simchi-Levi, et al., 1999, pp.1).
- Supply chain includes all activity which related to the production and delivery of a product, starting from the supplier's supplier to the customers. Lummus and Vokurka (1999) define four main processes - like planning, sourcing, production and transportation – that clearly describes the chain as well as includes the management of supply and demand, purchasing of materials, assembly of products, handling of stocks, processing of purchase orders, distribution and final transport to customers.
- Wisner, Tan and Leon (2015) have written about the general concept of supply chain management at the very beginning of the 21st century. According to him this is a chain of several companies where the goal is to create an acceptable finished product to the client. The functions, processes and activities support the flow of raw materials, products and services along the sourcing-making-delivering chain.

- Chritopher and Towill (2001) define this as a management system and a the network of equipment, tools, and distribution alternatives through which sourcing, shaping of materials to final products and delivery of them to consumers can be realised.

3.2. Structure of supply chain management

The interest toward supply chain management started to increase from 1980s when the companies realised how much benefits are hidden in the cooperation and integration of suppliers. Later on, the growth of this chain enhanced the productivity, profitability and effectiveness of business (Lummus & Vokurka, 1999). During the decades, other external factors – such as globalisation, reduced barriers of international trade or information richness – got a more important role at the creation of supply chains in order to companies could keep up with the increased national and international competition (Gunasekaran, et al., 2004). In the history of supply chain management, the enterprises started to focus on elements within the company and made that efficient and competitive. At the beginning individual businesses competed in the market however this shifted to a competition of interlinked organisations that are involved in a company's supply chain (Toit & Vlok, 2014)

Enterprises realised they can't be successful if they are not working efficiently together with their other business participants along the chain which is a key element to successfully satisfy customer needs. Modern supply chain has good communication and cooperation among its members, but what is more important is flexibility. It is necessary that companies prepare for the quickly-changing individual request of costumers as well as the shortened life-cycle of a product. Therefore, it is recommended to transform the chain to a customer-centric business structure that can enhance responsiveness as well as market performance (Lu, 2011).

Nowadays' rapidly changing word we rather talk about the competition between supply chains than between businesses. This chain consists all of the members who help to fulfil the customers need. The following list includes the participants, such as suppliers, manufacturers, transporters, retailers and customers who close the chain. They are covering different fields of business processes in the system like production, marketing, finance or sales. We also need to mention the decisions and decision making processes which are founded in each part of the chain but different decision horizon regarding time (Chopra & Meindl, 2013).

Supply chain strategy and design cover a longer time horizon with several years and build on a long term decision making. It's important to mention that a right decision can only be made if the company considers the uncertainty generated by different unforeseeable market actions. This phase includes questions like – what kind of supply chain configuration a company should

create, how the company could allocate the resources correctly between its sub-parts or whether outsourcing a function or performing it by the company would be the better decision. The middle phase called supply chain planning where the time horizon extends from a quarter to one year. One of the main part of the planning is to make prediction about demand, cost and price for the following quarter (year) as well as decide the inventory policy and delivery schedule – which location serves which customers. This phase also include uncertainty as the business doesn't have a certain information about how the market competition and other extern effects will influence the demand, exchange rate etc. On the other hand, we are talking about a much shorter time horizon here in contrast to the first phase, so the companies are able to react more flexible to unexpected turns in their environment. The literatures refer to the last part as supply chain operations which is already working with weekly or daily time horizon and the focus is on the personal customer requests and the most efficient handling of these. This phase includes the matching of inventory and production parameters to each customer order, also the scheduling of delivery time and managing replenishment orders. Uncertainty occurs in vanishing time thanks to the short time period, which enables the decision makers to react more flexible to unexpected events or customer request as well as to reach better results in performance optimisation (Chopra & Meindl, 2013).

The above mention three phases are also perfectly match to the three levels of a firm's activities: strategic, tactical and operational. Beginning from the first where decisions have long lasting effect through the quarterly/yearly to the day-to-day decision-making scope. Within a supply chain Stadler (2004) distinguishes sub-parts with different issues and these elements can also be connected to one of the three levels:

- Strategic and distribution network configuration means the network of all the production plants, warehouses and customers on a specific geographical region. Decisions made here are part of a strategic level, since the creation, redesign or reorganisation of a distribution network will have an effect for many years ahead. The most important question is how a network can be created which brings reduction in production, inventory and transportation costs.
- Product design and development is also a part of the strategic planning and the decision which are made here can be costly. Creating a product design or redesigning a product involves large expenses - considering especially the manufacturing - and they could increase inventory holding or transportation costs comparing it to other designs, too. The main issues that need to be solved is the timing of redesign in order to reduce

logistics cost or changes on the supply chain system to benefit from the new product design.

- Production planning is considered to be a tactical action in a life of a company and belongs to a mid-term planning and decision making phase usually taking a quarter or a year into account. Here, decision makers need to observe shifts, machine groups, flow lines or any operations in this level in order to identify any bottleneck in advance. At the end, leaders of this department should set up a correct work scheduling, sequence of jobs.
- Distribution planning comes after goods are produced and needed to share them among production and distribution sites. Managers job is to schedule the delivery to customers by considering the route of transport – it can happen through warehouses, cross docking or directly to customers. Another issue is to matching a supply with demand in every period. It means that supply chain expert should be able to coordinate the flow of goods that companies are handicapped with a lesser extent from the shortage of supply.
- Demand planning is part of a mid-term decision making where the job is to forecast demand with the help of famous univariate, multivariate or life cycle methods (e.g. Winter's exponential smoothing model). Here the professionals also add influences into a model which have a high probability to happen in the future (e.g. introduction of a new product) and try to examine its effect on sales.
- Inventory control belongs to the operational level as the retailer has to monitor and predict demand of the customers and the change of the inventory level on a daily or weekly basis. The most important questions of the retailer are: when is the right time to reorder and how much products are necessary to order to reach a low inventory ordering and holding costs? Uncertainty in demand is also need to be consider as well as the impact of forecasting tools.
- Transport planning is usually short-term task as products need to be delivered within days or week in most of the cases. Challenge is here to pay attention specific customer request or labour regulations, for example time windows for delivery and working hours for drivers. Because of these constraints companies facing a vehicle routing problems on a daily basis.

- Demand fulfilment is the last step in the matrix that ended up in an order execution. It is part of a short-term level where due date setting and shortage calculations are weekly or daily jobs of a manager.

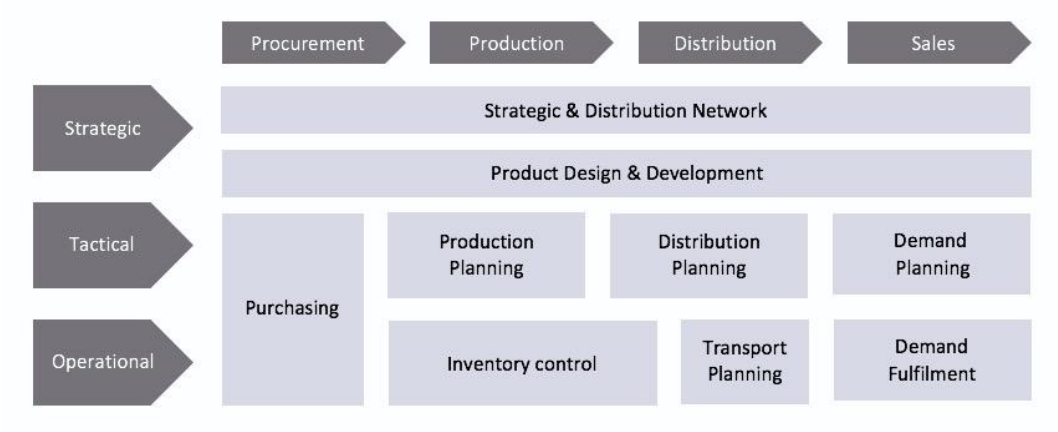

Figure 8: Supply chain planning matrix (Meyr, et al., 2002)

3.3. Challenges in supply chain management

Manufacturing processes is built an essential part of supply chain and provide challenging optimisation problems for businesses. One of the main task here is setting up a proper production planning and scheduling. It means the organisation of production steps and allocation effectively the enterprise's production capacity considering material, timing and place. There are several factors which influencing the processes in manufacturing on production sites so it is hardly possible to completely optimise all of them. In order to cope with them, enterprises create three main categories for manufacturing task ranging from long-term to short-term issues - these are production planning, job assignment and detailed scheduling. While production planning is dealing with to find out how many productions would be optimal to take place within a time period, scheduling is rather focusing on more real-time, daily problems and trying to determine sequencing and timing of machine work or setup duration between different configurations of machines, for example. During the years, several different models are created dealing with single item to multiple products and targeted to minimize the cost of the aggregated production planning and scheduling that includes inventory holding cost, regular production cost backlogging, overtime etc. Furthermore, the job assignment problem as a classic operational research model aims to assign tasks to every worker in a way that total cost can be minimised. The more real-life formulation of the problem called dynamic assignment problem which incorporate time constraint and a possibility for a task to be done by multiple workers (Maimon, et al., 1998).

Inventory management is such area of supply chain where one can face to many optimisation problems. Several methods and algorithms were born during the years in order to handle inventory on a cost effective way and aiming to satisfy costumers needs on a required level. It indicated a technological development where suppliers and retailers are following the inventory related data through a common system which enables a better cooperation within the supply chain (e.g. EDI[8] system, POS[9] data management system).

In case of inventory management, I need to mention the vendor managed inventory concept which is one of the most popular research area in this field. This concept gives a greater power to the supplier as it has a greater control on inventory management in contrast to the retailer or buyer. Supplier mainly pay attention on optimal order quantity or service level requirement where its goal is to create a beneficial transportation and inventory holding cost through which it can optimise its profit. Throughout the years, researcher have invented such models which are closer to real-life and capable to handle real-life instances. This means these algorithms can already work with more suppliers, more retailers in a dynamic model (Bichescu & Fry, 2007).

Transport optimisation constructs such part of supply chain management where the focus is on the proper routing by which the product can arrive in time at the customer. Here, the reserachers mainly deal with vehicle routing problem and its models. They are always containing a fleet of vehicles, which are going from depo to costumers and back. The more sophisticated real-life problems contains intransit points or warehouses where diffrent amount of inventory stocks are available. Furthermore it is typical that these models must handle real life constraints such as customers set a time-window for delivery or drivers are not allowed to work more than 8 hours per day. In this case it is typical that the model must handle the fact that ordered products of some customers must be delivered by more than one vehicle. Besides, real time information about weather, traffic and inventory stock helps companies to easier plan capacity or re-schedule delivery (Cordeau, et al., 2007).

Demand forecast is probably one of the most important part, because it is functioning as a driver for decisions and processes at any stages of the supply chain. Before action is taken in different part of the chain the company has to predict their buyers' need and a future demand as accurately as possible. Usually forecast is considered to be inaccurate but there are several,more

[8] Abbreviation for Electronic Data Interchange which means an exchange of business documents (e.g. purchase invoices, orders, ship notices) among computers mainly of suppliers, retailers, banks and clients (Cohen, 2013).

[9] POS means Point of Sale, which is a technology, carries information about customer payments. Today it is mainly used at check-out counters in retail or grocery stores to observe purchase trend and customer needs (Whitteker, 2014).

advanced method with which this accuracy could be improved and parallel the forecast error could be reduced. Without being exhaustive, the most popular forecasting method used by both researchers and businesses are the following: Static time-series forecasting sticks to historical data and use same historical values for forecasting, so this is a very simple and less accurate method. In contrast to this there are several adaptive forecasting methods where the data about trend, seasonality and level are updated in case of every new observation about demand like in the Moving avarage model, in the Holt's model or in the Winter's model (Chopra & Meindl, 2013).

Here I would like to shortly mention an interesting phenomena in demand forecasting which can highly effect the members of the supply chain on different stages. It is very common in practice that information flow among stages are irregular, incomplete and distorted, there are conflicting objective of some functions as their managers want to reach the highest profit and the most optimised processes only at their own areas. Because of these issues bullwhip effect is formed which means that fluctuation of orders will become greater and greater between the stages. Supply chain participants received disfigured number of demand starting from retailers to wholesalers, then going towards to manufacturers and to suppliers (Chopra & Meindl, 2013). Causes of bullwhip effect is intensively discussed in literatures and the main obstacles can be summarized as the following (Whang & Chuu, cited in Hofmann, 2017):

- Demand signal forecasting is working with aggregated data derived from past. Companies consider them important to predict a future level of inventory as well as achieve a good service level, however historical data is not always reflecting current demand of buyers. As a result, supply chain professionals working with incorrect numbers which will pull through the chain and forecasting result will stay less and less in accordance with real demand numbers (Lee, et al., cited in Hofmann, 2017).
- Another obstacle which could contribute to bullwhip affect - this is such forecasting that based on exaggerated orders. In this case current demand on the market is higher than goods available after production, therefore customers not able to purchase enough to satisfy their needs. If a buyer company is able to predict produced good beforehand then it will order over that amount it needs in reality in the hope of receiving a higher portion from the available products. Again, seller company faces misleading information resulted in poor forecast, inappropriate planning and scheduling (Whang & Chuu, cited in Hofmann, 2017).
- Order batching is a beneficial method for firms to save cost occurring during transaction of transportation or ordering. For this end buyers find financially more

optional to order goods in greater packages and not individually where they can save money with volume/quantity discounts. Unfortunately, this isn't show the correct demand again, since customers only order at specific, unforeseeable time with which inventory planning is hardly possible (Hofmann, 2017).

4. Data analytics in supply chain management

4.1. Definitions

In the first subchapter, I will write about the definitions that related to data analytics and big data application in supply chain management. Waller and Fawcett (2013) has issued one of the most important research paper which deals with data science, big data and predictive analytics in the context of supply chain management. It created a good basis for later research as the authors not only introduce the definitions, but also give examples for the different characteristics of big data by the 3Vs and listed several research questions for future scientists. Waller and Fawcett (2013, pp. 79-80) proposed definitions are listed here:

- *Supply chain data science* is an application of quantitative and qualitative methods from a variety of disciplines combined with supply chain management theory. It considers data quality and availability issues to reach its goals that means solving relevant supply chain management problems and predicting outcomes.
- *Supply chain predictive analytics* is gathered together qualitative and quantitative methods in order to estimate past and future levels of integration of business processes among functions or companies. Furthermore, its aim is to improve supply chain design.
- *Logistics predictive analytics* also uses qualitative and quantitative methods to estimate past and future behaviour of the flow and storage of inventory and predict the associated cost and service levels.

One year another researcher examined several previously issued research paper and summarized the discovered terms of supply chain data analytics. In their search strategy, they looked for publications in EBSCO, Emerald or Scopus. They also completed it with an internet search after keywords like big data analytics, advanced analytics, predictive analytics combined them with a ‘supply chain’ text. They created their own definition for SCM big data analytics based on 87 sources (e.g. journals, books, reviews) after the literature overview and the web search (Rozados & Tjahjono, 2014):

“SCM Big Data Analytics is the process of applying advanced analytics techniques in combination with SCM theory to datasets whose volume, velocity or variety require information technology tools from the Big Data technology stack; leveraging supply chain professionals with the ability to continually sense and respond to SCM relevant problems by providing accurate and timely business insights” (Rozados & Tjahjono, 2014, pp.6)

4.2. 3Vs in supply chain and logistics management

Because of the advantageous utilization of volume, velocity and variety (3Vs) features of big data information and communication technologies are more and more important in supply chain

management. Rapid and real time analysis of large volume of data sets are essential for any company to keep its competitiveness on a dynamic market. Exploiting 3V characteristics of big data supports companies to create an adaptive supply chain that possesses agility in action, responsiveness and resilience. Adaptability feeds from large data sets collected through information and communication devices and find rooms for improvement on areas, like restructuring of business actions, rearrangement of resources or coordination of actions (Rynarzewski & Szymczak, 2016).

I would like to start with the first characteristic, the *volume* which usually set challenges in case of a multi-stage supply chain as the flow of the data and information are happening through several stages which makes problematic to avoid the loss of data.

Velocity get even greater attention among the 3Vs as it can attributable for successfully reducing the bullwhip effect along the chain. Data analytical tools and techniques make it possible to collect, process and analyse datasets with an increased speed in real time. As a consequence, velocity changes the decision making process to be more effective (Hofmann, 2017). Richey and his colleagues conducted a research focusing on big data 3Vs in supply chain and found that increased velocity is considered as an opportunity and an obstacle by supply chain experts at the same time (Richey, et al., 2016).

It is also necessary to mention the last V, *variety* that means data arrives in different forms and structures at the company. RFID readers, sensors and GPS devices gather data in different level of structures, which indicated that the traditional, old way of data storage – which was able to store only one type of data – has been switched to a more up-to-date storage system adapted to big data features.

Other reserachers have also dealt with 3Vs of big data in their research paper and they brought some general examples from supply chain perspective. Volume can be seen as a vast amount of data generated by sensors, bar codes and ERP systems. Collected big data clearly hold variety thanks to diverse sensors at manufacturing sites, retailer shops and facilities. Utilising velocity, as a speed of data collection, can be helpful to faster explore useful knowledge or quickly come to a decision (Benabdellah, et al., 2016).

4.3.Data types used in supply chain and logistics management

A product life is monitored throughout the whole supply chain by sensors and there is a continuous data collection about human resource and processes whom and which related to it. On the different phases of supply chain, the type of gathered data is distinguished, but there is one common thing in them – the tool and sensors can produce up-to-date information and analytical tools are able to carry out real-time analysis. The most information is collected by

Radio Frequency Identification (shortly: RFID), GPS and Point of Sales (shortly: POS) sensors. Beginning with RFID, it can be described as a type of data that plays an important role in inventory management at both manufacturing sites and retailer stores. GPS data support the tracking system of a company and used in transportation control at first. However, there are several studies about the fact that it is good to exploit this kind of data in production, especially in production scheduling too. At finally yet importantly, we arrived to the POS data used at the beginning of the supply chain in order to monitor directly buyers' purchase behaviour as well as improve demand forecast (Myerson, 2007). These three previously mentioned data types are described more thoroughly in the following lines:

- Radio Frequency Identification (RFID) revolutionising how businesses following the life of a product within their supply chain. RFID data are not only used to keep track the product in the system but also identify workers. The tool which enables this for the company called RFID transponder or RFID tag containing a microchip and an antenna and changing information by radiofrequencies within the system.

  One of the biggest advantage of RFID is that it easily identifies location, identity or history of an object as well as analytical tools can receive up-to-date information through the tags. These contribute to the improvement of replenishment processes or inventory management. RFID also brings automatization into the life of a business so the less value-added tasks of employees can be replaced by these sensors through which the system saves time, cost and human resource. Good example for that the RFID replace the task of counting inventory by human workers as well as provide always real-time information about stock levels. Another advantage can be found in retail shops where restocking can be solved way easier as previously. It can be monitors through RFID that how many pieces of a product are left on the shelves after a customer purchases. If the number of available products goes below a threshold the system send automatically a request to inventory management for re-stocking.

  All in all, RFID makes the entire supply chain transparent as it helps businesses to follow the life of many products from inventory until selling (Derakhshan, et al., 2007).
- Point of Sale (POS) data are used to measure the consumption of costumers at the beginning of the supply chain, and identify how many and what kind of products are purchased. This helps producers to better understand customers' needs, plus predict future demand. If a supply chain member is placed further from the customer in a

chain, then point of sales data might not as relevant for it as for the seller. However, some researchers proved that if supply chain members contribute in some way to creating or selling the final product then POS data will never be negligible for them (Zhu, 2013). Furthermore, POS data is considered to be more accurate in sensing the actual end-customer activities, so they can reduce the amount of out-of-stock stand as well as better cope with bullwhip effect that usually occurs if a company use rather the purchase order-driven approach (Keifer, 2010).

- Third type of data are coming from GPS devices that mainly used in logistics in their transportation systems to make the flow of information easier and tackle with the challenge of on-time delivery, just in time supply of material or control production processes which can also be independent on transportation. Analytics that based on GPS data can provide real time information for companies about transport interruption, estimated times of arrival or about travelled routes, so production scheduling can be optimised with respect of these information. As it can be seen GPS data are not only useful in transportation, like in cargo tracking, but it also carries valuable information for manufacturing which supports firms to re-schedule more flexible their production processes. Future steps will be probably the integration of a global system of mobile communication, GPS and RFID technologies within one tracking system which could revolutionized logistics and production as this integrated system will contain transport planning, warehouse management and production control based on smart materials at the same time (Klumpp & Kandel, 2011).

4.4. Implementation of data analytics in supply chain management

I find important to briefly describe the implementation steps of data analytics and also presenting some factors that could make this implementation successful in supply chain. Therefore, I dedicate this chapter to give a good overview about this.

4.4.1. Framework for implementing data analytics

Big data analytics has the potential to transform supply chains according to a study of Sanders (2016). He examined several companies with different profile and come to a conclusion that after they went through the transformation caused by application of big data there have been three common features. Firstly, "their efforts are driven to support the companies' strategy", secondly they use "applications along all supply chain functions in a coordinated manner" and the third point is that companies "measure performance through carefully selected metrics" (Sanders, 2016, pp. 37). Framework for the way of implementing big data analytics in supply chain management is also provided with three main steps, namely *segment*, *align* and *measure* embedded in a *continuous improvement cycle*. The picture on the right demonstrates the previously described structure (Sanders, 2016).

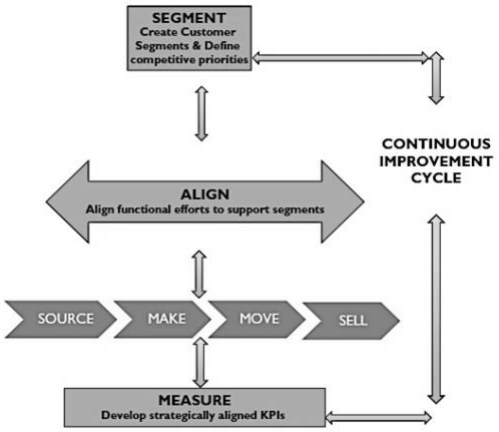

Figure 9: Framework for Big Data Analytics implementation (Sanders, 2016)

In more detailed, company defines segments by analytics and decided a way it wants to compete there at the first stage. When it comes to the competition part segmenting becomes more challenging. First, companies have to define very specific competitive priorities with "target levels for customer service, cost competition, quality, time or responsiveness" by each segment. The next step is to create exact operational requirements from them that leads various supply chain structures, operational strategies or requires different kind of suppliers and transportation. Although every segment has their own objectives, big data analytics can contribute to simultaneously "optimize customer needs and supply chain requirements to serve each segment" (Sanders, 2016, pp. 38).

In the next level, alignment of organisational functions is happening in the way that their efforts should contribute to segment attributes and competitive priorities. It means companies integrates process along the chain where their supply chain functions use their own specific analytics, but they also share intelligence across supply chain members and support joint decision making. Shortly, "alignment integrates the organisation and its supply chain horizontally" and "avoids fragmented efforts" (Sanders, 2016, pp. 39). Last step in the

framework is to measure performance with KPIs designed to segment attributes. Right metrics are key to a success of a company as they correctly reflect in which part is necessary to optimize. In most cases strategically aligned metrics decided by all process members are the best for this purpose as they are not only controlling segment characteristics but also give information about degree of alignment, integration and cross-enterprise cooperation. Framework can be seen as a circle thanks to the continuous improvement cycle as enterprises go repeatedly through the three stages in order to learn and improve their operation progressively. Company metrics and the formed segments with competitive priorities are connected by a feedback loop that helps to shape up segment processes. The author also believes that big data analytics is clearly able to support this step-by-step improvement during the cycle as they make it more efficient to continuously monitor KPIs and detect or mend deviations in real time (Sanders, 2016).

4.5. Application of data analytics along the supply chain planning matrix

There is a possibility to make a classification of the literatures according to the element of the supply chain and operation matrix. I have found and read plenty of research papers in the topic of big data analytics where the researcher focused only one element of the matrix or the model and wrote about big data applications in that specific field.

As we have seen from the supply chain planning matrix in the previous chapter it can be divided into long-, mid-, and short-term decision making and planning phase which can be identified as strategic-, tactical- and operational-level.

4.5.1. Strategic network design

Strategic network design is the part of the long-term planning phase and focusing on the evaluation of different supply chain structures for finding the most profitable as well as effective construction. First research paper that deals with such network configuration was in the topic of humanitarian supply chain management. Most of the cases Non-Governmental Organisations are responsible for designing a proper network for disaster relief, healthcare or education where big data analytics can contribute to process and capacity transformation, facility innovation as well as creating new organisational capabilities in order to easier cope with challenges of the developmental sector in the third world. What the authors could identified that the benefits of the big data characteristics, the 5Vs could be highly exploited in these areas resulting in a very resilient supply chain network which is able to adapt to the rapidly changing world. What is special in this type of supply chain management is the very complex and dynamic environment where the distribution of available equipment and other resources are critical, vital and life-saving. With the help of big data analytics, a NGOs could achieve a

responsive and efficient strategic and distribution network for the long run in the hope of better outcomes on disaster or underdeveloped zones (Prasad, et al., 2016).

Wang, Gunasekaran and Ngai (2016) also dealt with distribution network within supply chain and the application of big data. They identified that traditional network configurations are not able to keep up with a quickly changing supply chain operations and there is a need for redesign of the network structure. Basically distribution operations are working with a vast amount of data which can be exploited to identify the necessary number of distribution centres and the right assignment of clients to them, plus reduce operational cost at the centres. They argued that there is a lack of literature which mention an exact application of analytics in this field, so their goal was to create a powerful model – this was a mixed-integer nonlinear program - which is able to solve supply chain and distribution network problems using big data. The source of data is twofold – first of all, they are deriving from historical data bases, however the author also used behavioural data gathered from social media, web clicks, comments and so on. All of them was essential part for identifying proper customer needs and locations for distribution centres in a supply chain network with more than 2000 stores. The objective of the model is to minimize the total fixed costs, transportation and handling costs of distribution centres as well as decrease the penalty cost which occurs in case of unfulfilled customer orders. With the help of simulation and sensitivity analysis there is a possibility to select better locations for the centres in which case the model uses randomly generated big data sets for customer demand, warehouse operation and transportation. All in all, big data enables firms to find additional information and set up a more complex distribution network.

4.5.2. Product design and development

Another element of a strategic planning the product design and development, which overarch through the whole, supply chain and mainly start at one end with sales and marketing. The product development usually starts with discovering customers' preferences then bring these findings to a manufacturing and design area to create desired products. To accurately understand these needs Bae and Kim (2011) suggested data mining tools, such as association rules and decision tree techniques for exploring big data collected from customers. As it was written in second chapter data mining techniques are able to find previously unknown and useful patterns in big data sets through models that do classification or other algorithms discovering relationship in huge amount of data. The authors test their algorithms to find association rules in large data sets through a case study of purchasing digital cameras. The research grounded on several steps starting from data collection by conducting questionnaire

among customers. Then they tried to gain useful, quality data from the survey and transform them to a format which is applicable in the model. Through the previously mentioned data mining techniques they extracted the necessary information for product development and as a last step they tested the reliability and validity of the integrated rules. With the help of the decision tree the researches came to several results regarding customer opinion of different product types. Furthermore, gaining more sophisticated information by data mining not only provide insight of customers' need but also support manufacturing innovation. Scientists also concluded that results of the research need to be generalized through more experiment with other products as well as exploit greater amount of data and other mining approaches (e.g. genetic algorithms) in order to identify future product development patterns.

There is also another research which aim is to identify an appropriate competence set by big data in order to create a competitive supply chain as well as enhance product development of a company. Tan and his collegues (2015) set up an analytics infrastructure which support managers to generate new product development ideas through identifying necessary competences for the production of new goods. The idea based on a deduction graph model which help firms or departments to combine their competence set with other companies' or production unit's. The whole method can be described as a continuously evolving process where the conjunction of competences happens for the optimisation and it is also eligible to handle more than one decision maker at the same time (Li, 1999). The model can be presented as following:

It has a set of problems which need to be solved, set of needed competence, set of acquired competence and it also contains intermediate skills which job is to connect the needed and acquired competences during the process. Then the model builds up a learning network (graph) by starting from the already acquired set of skills to the needed competences through the intermediate skills (Tan, et al., 2015).

For better understanding the authors provided figures, see below:

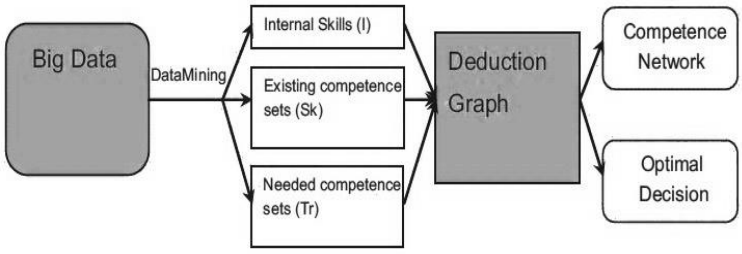

Figure 10: Deduction graph for finding the best competence sets (Tan, et al., 2015)

The idea has been tested through a real case of a glass manufacturing company where the authors tried to identified product development ideas, optimise manufacturing process in a cost effective way and understand how different enterprises can work effectively together. The scientist created up the previously mentioned mathematical model for discover the best learning sequence by generating the greatest profit at the company. After they collected the concrete data for the different sets they were already able to build up the graph which showed which competences can be learned from existing skills and which need to be "purchased". As a final conclusion they argued that using only big data is not enough for proper decision making, they have to relate to some supply chain specific problem for better utilization. Like in this case, these two elements need to be joint for better use (Tan, et al., 2015).

4.5.3. Purchasing

The first element of the mid and short term planning is the purchase and procurement that is highly dependent on the suppliers of a business and has a direct effect on inventory management.

Authors of the scientific paper suggest improvement on the procurement process after they examined the purchasing and sourcing processes by advanced data mining techniques. They believed that decision makers can gain better insight into procurement processes by analytics which help them to come up with a more efficient sourcing strategy.

They used a model with text mining based on clustering for which there was a special program called RapidMiner as a platform. This software is proper for machine learning, data or text mining and other predictive analytics. Database of a real company from the information and communication technology sector was used in the research in order to carry out analysis on the purchase data between the years 2011 and 2014. The company has divided their purchase into three procurement approach types where the related data were stored in different formats and spaces in each of the cases. As each of the transactions contain a lot of unnecessary data the first step was to clear them out and only keep the information about the description of the purchased items, period of transaction, amount spent on the item, supplier and buyer data. Then comes the transformation of the description text into such format that can be handled by the cluster modelling technique. Clustering happens with the K-Means method, which identified the occurrence frequency of words in the records' descriptions. This showed that if a word came up with many times that indicated more transaction, means more purchase of a product or service. However, there are one big disadvantage of text mining as it cannot clearly show how much the cost is that belongs to the different transaction occurrence. It could happen that goods or services with few purchases have high cost while products with high transaction occurrence

have only low values. During the methodology, the authors focused on the supplier side and identified their purchase amount and volume. The results of the two analysis now can complement each other and give a verified and complete information about the procurement process that can be used for further improvement. After the experiment, the authors identified those common goods and services among the purchases that should be aggregate among the purchases and those which should rather purchase by individual project teams instead of the whole company. Besides these, they argued that the conducted analyses could enhance visibility of purchasing goods and services. If the different project teams have access to cluster analysis with their identified group of purchases then they are able to plan more accurately and coordinate with a higher efficiency the long-term procurement plan (Tan & Lee, 2015).

4.5.4. Production planning

As it was discussed earlier production is a part of a mid-term, tactical planning where companies focus on manufacturing activities as well as resource allocation of employees, materials and machine capacities in order to produce the requested demand identified mainly by sales and marketing department.

Lee, Kao and Yang (2014) were among the first scientist who proposed a machine system which could handle industrial big data as well as smart production issues. They argued that companies' current manufacturing system are not able to cope with mass amount of data, so there is a necessity to switch from regular to self-aware, self-learning machines with a capability of self-prediction, self-reconfiguration or self-maintenance. The first issue is that productivity and production quality is heavily dependent on scheduling and task design, but current machines are working passively and reach in any case even when assigned task are not optimal for the machines' condition. Contrarily to this, smart machine system could advise better task arrangement and able to modify operational parameters. Another issue is the lack of adaptive learning and exploitation of big data information. The main cause for that is the absent of prognostics and health management system, therefore health monitoring algorithms are not working perfectly. Furthermore, condition monitoring data are usually processed by algorithms which are incapable for learning or developing themselves along with applying real-time and optimised big data for the analytics. The authors believe that a cyber physical information system[10] would be appropriate to reach a fleet wide information system and set up self-aware

[10] Cyber physical system includes computation with physical processes, and it is an intersection of physical and the cyber. It is characterised by high degree of automation, real –time and securely delivered tasks, it is networked at a multiple scale, and furthermore it has an integration for learning, adaptation and higher performance (Sanislav Teodora & Miclea, 2012).

and self-maintained machines which are able to estimate their health and degradation. They examined a cyber physical system which uses knowledge base and related algorithms instead of the common simulation- or control-oriented ones. Here, knowledge base is built through clustering with the help of adaptive learning and data mining where learning algorithm creates groups according to similarities of the machines (e.g. machine performances or working conditions). While the algorithm is running through the cluster in search of a good health assessment it can end with two different results – it either finds an already existing cluster and updates it or create a new one for the newly found "behaviour". Machines in the same group will have a very same health condition and will require similar interventions. Some of the key benefits of this proposed framework are:

- Coping with unprecedented event becomes more easily for prognostics and health management algorithms
- Companies can optimize production and create energy-saving maintenance scheduling with this new and improved way of health prediction. It may also cut down on machine downtime as well as decrease labour cost.
- Finally, this new smart manufacturing system enables industrial management to become more transparent and achieve a more effective information flow among production lines in supply chain management (Lee, et al., 2014).

There is another interesting topic that discussed by a research paper: RFID technologies used in production and on manufacturing sites. Zhong and his collegues (2015a) propose a method which could support to create a smart, data-driven RFID enabled shop floor manufacturing using a visualisation approach. They named their model RFID-Cuboid which is able to restructure raw data deriving from RFID sensors while considering specific production logic and time series. The invention of the authors belongs to the new Cloud Manufacturing paradigm that enable companies to convert traditional manufacturing resources to smart manufacturing objects. After this transformation the new system contains objects which are already able to sense, react, interact and cooperate applying wireless communication standards. Those data which collected through this way are ordered into a RFID cuboid according to three dimensions: RFID data dimension (x-axis), behaviour dimension (y-axis) and key attributes (z-axis). Each cuboid represents a logistic event and every RFID cuboid are connected together into a chain from which it is clearly see who, what, where and when did an operation. So this chain helps to visualise an entire logistic life cycle. As a next step a logistic trajectory can be pictured then we arrive at a graphical visualisation of logistic operators' performance as well as production

progress. In the last phase, cloud-manufacturing data has already been cleaned from useless information so statistics, machine-learning procedures and prediction models can obtain relevant knowledge for decision-making. More exactly, decision makers are able to successfully exploit this information for creating an intelligent future environment for logistics planning and scheduling, inventory control, logistical task assignment and promotion strategy. At the last step of their research, they examined the feasibility and potential improvement areas of their models on a real-life case from an automotive firm. Concrete findings were that the visualisation approach helped the daily operation of several users. As an example, mined average logistics time is helpful for decision-makers in case of material resource planning and production decisions. Furthermore, the model also showed the performance of different logistics operators and identified that junior employees should rather work on internal logistics operations while senior workers deliver tasks in case of external issues. At last but not least, it also revealed which workers should improve on their efficiency because of high inventory level at specific departments with a noticeably low logistics efficiency. As a future research area the authors suggest to enlarge this model for a global or multi-echelon stage in supply chain and try to cope with an even greater amount of data (Zhong, et al., 2015a).

Scientists use physical internet[11] for creating an intelligent shop floor manufacturing in another research paper. If we examine more thoroughly this new concept it means an adaptation of networking technology, wireless and cloud-manufacturing in order to create smart manufacturing objects (shortly: SMO) again from production resources, like manufacturers, workers or materials. The authors propose here a physical internet based intelligent manufacturing shop floor and they use RFID technology and wireless communication network to set up their model.

As a first step RFID readers need to be placed areas from row material production areas till finished product receiving sites where each raw material with RFID tags matched to a pallet. Next the materials arrived at a manufacturing shop floors where all of the machines have a stationary RFID reader as well as the internal and external operators carry a mobile reader device. By these processes pallets can be monitored all along the production stages. Another specialty is the transformation of resources to smart manufacturing objects - in a way it was explained in the previous research by Zhong and his colleagues (2015a) – which are able to

[11] Physical Internet is an open, global, interconnected and sustainable logistics system, which is found on physical, digital and operational interconnectivity through interfaces and protocols. In this system modular containers in different size are moved through the multimodal transportation networks and are aggregated at transit sites from different origins to optimise loading on the next segments (Shenle, et al., 2017).

interact and behave with each other in accordance with an earlier defined logistics logic. As a following step, big data generated by SMOs are collected in a data warehouse with the help of a wireless communication network. Then, big data analytics starts to work on the data sets step by step from cleansing through pattern interpretation to knowledge representation. At the very final phase the gained knowledge helps different application in a logistics management, like in real time decision making, knowledge base prediction and logistics knowledge repository. The researcher examined their proposed big data analytics for physical internet based logistics data within the framework of a case study. They successfully identified several behaviours and KPIs which are accurately reflect and evaluate the performance of the operators and the operations. Such indicators are for example the total delivered or shipped number of smart pallets (depends on whether we consider internal or external logistics employee) as well as a total time spent on logistics from raw material to finished product. Considering future research, they suggest to develop a mathematical model for physical internet driven logistics in order to make their proposed model more transparent, accurate and credible (Zhong, et al., 2015b).

4.5.5. Distribution planning

A good delivery scheduling brings higher customer satisfaction, lower delivery attempts plus companies can ensure that delivered products arrive at the time when customers are able to receive them. A distribution concept was built up, which can support the above mentioned advantages of a delivery. This take the sales data of an online stores which is more special case than examining an original store as here the stores are struggling more to reach a good custmer satisfaction in many cases. The authors considered data-intensive tools like global positioning or sensor networks technology. The system contains three elements: the computational center carries out the analytics using the real-time data from the customer and delivery vehicle to optimize the distribution. Costumers are sharing their location data through an application that send them to the center. The delivery vehicles have all the products signed by an RFID code plus other sensors which are sending all the up-dated information about the location. Besides these, the system captures and calculates with other real-time data - such as weather condition, traffic information or road construction – because they could also intensively influence the distribution and rout planning. In more detailed: after the sensors have sent the location information of the vehicles and customers, the computational center provides the three most favourable options for each customers regarding the time and location of the delivery. The clients could choose among the options on the mobile application and then these decisions are automatically sent to the center for recalculating the distribution plan of the vehicles as well as showing the most optimal routes for the drivers. All things considered, big data analytics could

provides an effective delivery scheduling with high customer satisfaction by using data analytics applications (Engel, et al., 2014).

A is writing about advantages of using big data and analytics in supply chain and logistics based on the survey of consulting firm as well as examples based on logistics and commercial companies. One area that benefits from big data and predictive analytics is shipping where companies get the opportunity to utilise better distribution capacity and increase the speed of delivery time after real time information are captured from the dataset. A special use case of the analytics is found in freight transport where the parcel volume analytics can forecast more accurately the expected number of parcels and cargo within the supply chain network. The technology based historical data derived from internal company environment and from Google search, weather forecast or shopping behaviour of online customers, on the other hand. Another potential benefit that was listed by the authors is connected to goods shipment. The main idea is that applying data analytics enable logistics firms to start a delivery to customers prior to their orders. Companies can match group of goods to geographical regions where there is a very high probability of that the customers from different areas will look for products to the assigned group. The aim of the companies is that the customer order arrives when a product is already in transit, so they could significantly decrease delivery time and parallel increase customer satisfaction, plus the number of sales (Leveling, et al., 2014).

Data analytics can also be advantageous is resource planning in a strategic and operational level. It means the configuration of distribution network on a strategic level, while operational level includes capacity planning on a daily or monthly basis. What big data techniques could achieve in both cases is the improvement of the reliability of the planning and optimal matching of available resources and demand by logistics experts. Authors bring advanced regression and scenario modelling techniques as a good example. They can utilize the much higher volume and variety of big data, consequently a business can achieve longer forecast periods or cut down on the risk of long-term infrastructure investment and contracted external capacities. In the past operational tasks were planned accordingly to historical data and personal experience which has changed a lot in the big data era (Mikavica, et al., 2015).

4.5.6. Inventory management

Ittman's research paper (2015) concentrates on the benefits of big data analytics in retailer and vendor collaborations and e-commerce vendors. The more exact benefits are: better visibility for future orders, low in-stock items; reduction the impact of late shipments; predict how profitable special quantity deals are or provide the ability for retailers to see pricing and

allocation strategies where no historical data are available (Ittman, 2015, pp. 6). Most of the companies already have a platform on the internet too, which help them to easily monitor customers shopping behaviour by analysing their clicking and browsing on websites. Furthermore, internet based analytic tools can also generate useful information from web sites' data for managing inventories at different distribution centres and optimise fulfilment strategies. All thing considered, predictive analytics at logistics, e-commerce and predictive-shipping uses recent and historical data at the same time to create "personalised customer engagement" (Ittman, 2015).

The research paper of Zhou with his colleagues (2017) are discussing an issue about the lack of intelligent inventory systems that could properly handle the intense inflow form data as well as effectively do forecasting, carry out anomaly detection and evaluate inventory aging. According to the previous studies, the companies use mainly statistical analysis on historical inventory data that bring a less accurate planning. Considering these problems, the authors came up with a state of the art intelligent inventory system called as iMiner, which is capable to use data mining approaches on enormous data sets and efficiently help out in inventory planning and controlling tasks. Before this program has been created and implemented, the researchers defined all the challenges according to the studies on retail companies that the program must solve. The four key issues were business big data management, inventory forecasting, inventory anomaly detection and inventory aging analysis. Data mining technologies in iMiner programs that were supposed to solve these issues were regression analysis, classification-based learning plus different visualisation tools for easier interpretation of results. The developed models according to the different inventory management areas are the following:

- The goal of a good inventory forecasting is to predict future demand as accurately as possible and reduce inventory loading. Here the authors developed a dynamic prediction model based on machine learning techniques and time series analysis. The algorithm defines the hidden patterns on a stock in/stock out time series by using a regression model based on historical data. In order to make the forecasting more accurate the algorithm also consider other factors such as long-term trend, seasonality or event factors (e.g. sales promotions or holidays). The authors emphasise that the stock in and stock out amounts are highly dependent on each other, so this independency is needed to be handled by the model. (Stock in amounts have to pay attention on stock out in short term in order to avoid excess demand, vice versa to avoid out of stock issue.) From this reason it was completed by a multiple time series prediction which is already able

to capture the dynamic relationship among different time series and make proper forecasting considering the connection between stock in and out.

- Classification-based anomaly detection is applied for better inventory management, which goal was to look for abnormally high fluctuation of stock or sales data. The algorithm was designed to find these anomalous elements in the data set and sign them with labels; these are then converted into a classification problem. The iMiner's job to use this classification model on a training set, memorize the patterns and effectively use what it has learnt on new data arriving on a daily basis (Zhou, et al., 2017).
- Inventory management has also cope with the issue of aging inventory that simply means avoidance of overstocking items as well as cut down on the number of overstocked products. Here, feature selection technique was used, which approach built in a way that it searches for such attributes, which cause overstocking. As a first step a filter model separate the redundant and unnecessary features from the attributes then try to define a candidate set of relevant attributes. After that, the random forest as an ensemble learning algorithm can select the essential attributes, give proper significance to the selected attributes and return the measures of attribute importance This algorithm is basically built up with several decision trees where each tree represents a random subset of features and has the access to a random set of data points (Breiman, 2001). As a result, the program can detect the correlation between the attributes and the overstocking as well as show for the businesses which items should be monitored more intensively in order to successfully tackle with inventory aging (Zhou, et al., 2017).

The authors applied iMiner in a Chinese company getting a good result. The main advantages that this intelligent inventory application has already brought with itself were realisation of large-scale and automatic inventory data analysis, provide an intelligent and efficient decision support system for management and transform the inventory from a demand-driven to data-driven system (Zhou, et al., 2017).

4.5.7. Transport planning

Mikavica and her collegues (2015) summarize the benefits and opportunities of big data in transportation and argue that there are a "great unutilized potential for improving operational efficiency, customer experience and creating new business models". Still, big data analytics are able to provide competitive advantage for any company in the logistics industry, first of all in such fields like last mile optimization. Big data analytics help in the real time optimisation of delivery routes using sensor based detection of transported items. This technique enables

suppliers to automatically change delivery routes according to up-to-date information about traffic conditions. Moreover, the automated control of huge number of randomly moving resources need an "extensive data processing capabilities of big data techniques". As a result, last mile delivery cost could be significantly lower in seldom-populated areas.

A very young paper from the year of 2018 proposed an intelligent transportation system through the vehicle-driving path planning optimization. Considering long term general transport planning model - which already used in the past by scientific researcher for finding optimal routing - are providing good and reasonable results, however if we consider short-term planning the problem becomes more complex with high nonlinearity, time-varying and uncertainty. It creates a special challenge for researcher to design an effective prediction model able to handle the previously mentioned complexity as well as use big data in an optimal manner. One of the experiment's aims was to set up a good traffic network to which the scientist used high-volume GPS data collected by 12.000 taxis in Beijing. There are a few popular methods, which are applied in case of short-term planning and give a good estimation of travel time and vehicle traffic density in case of many scenarios, however still struggling to handle real-time data – these are the historical data-base model, time series model, regression model and machine learning model. The scientists wanted to take a further step and come up with a model which could solve a short-term path planning and traffic flow problem based on the above mentioned GPS data. At first, the authors used a clique-based clustering method to get a more accurate result, so they divided the around 50 GB of GPS data by geographically and timely. As a next step, they predicted traffic flow with the help of an artificial neural network (shortly: ANN), then the improved Dijkstra algorithm has found an optimal path using the traffic speed information derived from ANN based model. An accuracy test was carried out to examine the difference between the predicted results in case of different prediction models and the actual data. The authors concluded that their proposed model give the most accurate traffic situation of all, so it could be used to find an optimal path more effectively (Zhu, et al., 2018).

4.5.8. Demand Planning

Demand planning closes tactical planning phase according to a supply chain planning matrix. A good example for the application of big data analytics comes from the electric vehicle industry where historical traffic and weather data are used to forecast the charging demand of these vehicles. Arias and Bae (2016) issued a paper in which they studied this forecasting problem on a South-Korean market and used clustering and relational analysis, plus a decision tree for creating classification criteria. Scientists used historical clustering technique to find

patterns in traffic data and relational analysis helped to find such factors which affecting these traffic patterns. As a final step, decision tree was ideal to discover the relationship between the created clusters and influencing factors as well as forecasting responses to data. The authors concluded that demand forecasting based on big data could help in future planning of operation profiles in power systems, as analysts will be able to forecast vehicles' charging demand at commercial and residential sites. Furthermore, it can also be seen that the proposed demand-forecasting model contributes for investment and planning decisions of future infrastructure of electronic vehicle charging.

Two years later, Hofmann and Rutschmann (2018) issued a paper in which they have looked for an answer to the question how big data analytics improve demand forecasting. The authors emphasise that finding a good answer for this question is a challenge as other researchers have not found a clear relationship between big data analytic techniques and demand forecasting until nowadays. Moreover, there was a lack in detailed description of methods and applications in this field as well as it was unclear whether big data analytics is suitable for substitution of existing techniques in forecasting. In order to find the opportunities and potential of it the authors have chosen the retail industry and examined the value of big data there. They believe that different types of analytics match better to different types of forecasting (thinking of different time horizons for example), plus there is a necessity of aligning between input, scope or method of certain analytic types for more accurate outcomes. Hoffman and his colleagues (2018) use their own classification for analytic techniques, which resulted in five sub-groups:

- Data Exploration is a self-service analytics technique used directly by business participants for gain insight into business operations. Then, this insight is always discussed with other employees before final validation of results.
- Advanced Analytics is able to give answer in more complex business situations using data mining, statistical methods or machine learning. Here, computer models immediately process unstructured data and combine data sets.
- Interactive Analysis and Planning have a connection to business intelligence that is no more an IT-led consolidation technique but an interactive function available for several users at the same time. Here, we speak mainly about structured data that loaded into an enterprise data warehouse[12] (shortly: EDW). EDW combined with advanced analytic techniques can contribute to the in-load of more extracted data and to better data content.

[12] Enterprise Data Warehouse is a repository that gives analytical information about business processes and core operations. With the help of IT the enterprise-wide business requirements can be easily fulfilled as well as it could also provide a weapon against the competitors (Tupper, 2011).

- Embedded Analytics support databased decision making through automated and analytical processes. This technique is an integration of analytical capabilities, like reporting, predictive analytics or data discovery into business software for CRM[13], ERP or supply chain management applications.
- Stream Analytics' goal is the immediate analysis of incoming data. Here, a specific method is able to analyse data in motion and perform calculations quickly according to previously defined rules.

One of the main findings of the research shows that techniques are used in forecasting with short-, mid- and long-term. Companies can reach an improvement in forecasting results if they choose embedded and stream analytics in case of short-term time horizon, as these methods provide them with data that could influence demand and give information about products in store. Medium-term forecasting should deal with advanced analytics mainly (descriptive and predictive type) and in some cases exploit the benefit of interactive analysis and planning as well as data exploration in order to increase customer insight and trend awareness. Finally, the long-term forecasting is better to apply data exploration and descriptive advanced analytics. Stream analytics with their real-time responses and embedded analytics with their focus on operational decision-making would never be an optimal choice here. When it comes to strategic, long-term forecasting, top management needs to easily reach previously prepared data by advanced analytics techniques (Hofmann & Rutschmann, 2018).

A very young research paper is about a product-in-use big data in demand planning. As the authors focused on an automotive industry this type of collected information also means any type of data related to vehicle from manufacturing till vehicle service and workshop data (e.g. maintenance). They also found a more concrete "vehicle in use data" name for that. After they finished their research, they concluded that product in use data can be exploited in spare parts of demand planning at a greater extent and have good performance effects. Moreover, it has become clear that demand planning in this field is a challenging and complex task due to low frequency, intermittent demand as well as because of large variation in price, criticality and specificity. In addition to this, supply chain structure is also a critical point at the demand variation and can be attributable for the bullwhip effect. The traditional spare part forecast in the automotive industry is working with historical data and time-series forecasting for a non-

[13] CRM is the abbreviation for Customer Relationship Management that is a part of business strategy with customer focus aiming to optimise profitability, revenue and customer satisfaction. It measures costs from marketing, sales and service sides, while applies knowledge about customer needs, behaviour to improve performance (Vogt, 2009).

intermittent demand, but these methods are not appropriate to handle the previously mentioned challenges. The authors argued that causal based methods are more capable in an aftermarket context if there is a good availability of high quality data as well as a good knowledge to identify explanatory variables after data analysis. All in all, the two researchers' aim was to find answers how product-in-use data can be effectively used in an automotive industry and how these could support performance positively. According to their research, they identified three causal-based forecasting methods intensively discussed in other literatures. The first one is the regression-based method, which can easily handle the phase in phase out context, the second one named as reliability based method that matches to the early life cycle phase of the product without any historical data. Finally, the third condition-based techniques can be useful in case of low frequency items. After this categorisation, the researchers divided them into eight sub-groups where they have written about different interventions during the demand planning of the spare parts. These have three main effect on the demand forecasting which are the following:

- Some of the interventions are able to generate item forecast with the usage of demand history through which forecast accuracy will be improved.
- Other interventions have effect on demand planning process as they turn forecast demand to planned distributed demand through which uncertainty could be decreased
- Manual intervention as a sub-group is contribute to alert generation which help supply chain experts to keep attention on abnormal vehicle increase in a specific region or identify outliers by different item categories in demand history (Andersson & Jonsson, 2018).

The conclusion of Andersson and Johnsson (2018) is that product in use data cause positive demand planning performance outcome within a supply chain as there is a clear improvement of forecast accuracy by causal based methods. The authors also provide ideas for future research where scientists could test each interventions through a more exact, single case with quantitative data and examining the implementation of the proposed methods on a real-life example.

SCM Matrix	Analytics, Methods, Techniques	Results	Source
Strategic Network	Big Data Analytics in general	• Responsive network • Better disaster management in humanitarian supply chain	Prasad, et al. (2016); Wang, et al. (2016)
	Mixed Integer Non-Linear Program Sensitivity Analysis	• Minimize total costs including transportation, handling and penalty costs (in case of unfulfilled customer orders) • Find the best location for distribution centre, warehouses in the network	
Product Design & Development	Association Rules and Decision Tree techniques	• Better insight into customer needs • Identify future development patterns	Bae & Kim, (2011) Tan, et al. (2015)
	Deduction Graph Model	• Identify necessary competence sets for product development ideas	
Purchase Planning	Text Mining based on Clustering	• Identify the purchase occurrence if an item • Identify those common goods and services that should be aggregate among the purchases	Tan & Lee (2015)
Production Planning	Clustering Technique Adaptive Learning Algorithm	• Easily cope with unprecedented event • Create energy saving maintenance scheduling • Improved way of health prediction	Leveling, et al. (2014) Zhong, et al. (2015a) Zhong, et al. (2015b)
	Machine Learning from RFID Cuboid data & Cloud Manufacturing	• Support the material resource planning • Better reflect the performance of logistics operators	
Distribution Planning	Different Big Data Analytics in computational center	• Calculating the most optimal delivery routes • Recalculating the distribution plan of the vehicles	Engel, et al. (2014); Leveling, et al. (2014); Mikavica, et al. (2015)
	Parcel Volume Analytics	• Accurately forecast number of cargo and parcels • Match group of goods to geographical regions and start delivery prior to customer orders	
	Advanced Regression Modelling technique	• Achieve more significant predictive value • Optimal matching of available resources and demand	
Inventory Management & Control	Interned based analytical tools	• Low in-stock items, Reduction of late shipments • Optimise fulfilment strategies	Ittman (2015); Zhou, et al. (2017)
	Interned based analytical tools Dynamic Prediction model with Time Series Analysis Classification based Anomaly Detection Feature Selection approach for Aging Inventory	• Capture dynamic relationship between stock in and stock out • Discover abnormally high fluctuation of stock • Detect the correlation between the attributes and the overstocking • Detect the items which should be monitored more intensively	
Transport Planning	Big Data Analytics in Last Mile optimisation	• Real time optimisation of delivery routes using sensor based detection of transported items	Mikavica, et al. (2015) Zhu, et al. (2018)
	Clique-based Clustering Artificial Neural Network Dijsktra Algorithm	• Provide more accurate estimation of traffic flow and travel time • Find an optimal path using the traffic speed information	
Demand Planning	Clustering & Relational analysis with decision tree techniques	• Find factors which affecting traffic patterns • Accurate forecast of vehicles' charging demand at commercial and residential sites	Arias & Bae (2016); Hofmann (2017); Andersson & Jonsson (2018)
	Embedded and Stream Analytics for short-term Advanced Analytics and Data Exploration for mid- and long-term	• Apply higher influence on demand • Get more accurate forecasting results • Increase customer insight and trend awareness	
	Causal-based Forecasting Method (Regression/Reliability/Condition-based)	• Decrease uncertainty in demand of low frequency, intermittent items • Identify abnormality and outliers in specific regions • Increase forecasting accuracy overall	

Table 2: Summary table of data analytics, methods and techniques in supply chain

4.6. Advantages and hurdles of big data and analytics

Here, I set against each other the topics of benefits and challenges of implementing and using big data analytics in general. I found several papers that dealt with exactly this question in general - what could be the advantages, disadvantages as well as challenges of big-data based analytics on the different part of supply chain management?

4.6.1. Benefits and opportunities

A lot of paper deals with the question – What are the benefits of data analytics and working with big data in different field of the supply chain management? Some papers dealt with business cases while in others there was a survey conducted among university students and supply chain professionals in order to find clear answers to this question.

First of all, businesses can achieve improved effectiveness and efficiency and improved product and service quality by successfully applying data analytics. Big data analytics enable firms to gain more accurate information and improve transparency through which they can minimize supply chain and inventory risk, better identify potential problems along the chain and correct them just in time. The analytics can also contribute to reach "higher operational efficiencies" that refer to such benefits like real-time vendor management, on-time service and feedbacks to customer complaints (Wang & Alexander, 2015; Schoenherr & Speier-Pero, 2015). People who use analytics regularly see the greatest benefits in more informed decision making and improvement in the reduction of supply chain costs (Schoenherr & Speier-Pero, 2015).

Furthermore, we can read about how analytics can help in case of personalized service and improve service quality in the third sub-group. It basically enables companies to make analytics from social media, mobile and web data, then they will see what products customers have bought or probably will buy. The authors also added that analysing customer interactions across all (media) channels could also beneficial to create higher customer engagement (Wang & Alexander, 2015).

We can also find benefits of big data analytics in the area of retailer and vendor collaborations and e-commerce vendors. Most of the companies already have a platform on the internet, which help them to easily monitor customers shopping behaviour by analysing their clicking and browsing on websites. Furthermore, internet based analytic tools can also generate useful information from web sites' data for managing inventories at different distribution centres and optimise fulfilment strategies. The more exact benefits are: better visibility for future orders, low in-stock items; reduction the impact of late shipments; predict how profitable special quantity deals are or provide the ability for retailers to see pricing and allocation strategies

where no historical data are available (Ittman, 2015, pp. 6). Data analytics at logistics uses recent and historical data at the same time to create "personalised customer engagement" (Ittman, 2015).

At last but not least, big data analytics enables companies to enhance collaboration with supply chain partners. The data analytical techniques can connect suppliers and customers to a company's big-data-driven system and clearly enhance efficiencies throughout the chain. It is also emphasised that with the help of big data systems they are able to better monitor procurement trends, pull the procurement and then strengthen their bargaining, negotiating power. Just-in-time inventory system can be named as a good example for the benefits of big data. After its appropriate implementation it produces cost savings, reduces stock-outs and associated opportunity costs, plus achieved optimum inventory levels (Richey, et al., 2016).

4.6.2. Challenges and barriers

There were four groups of hurdles created for different companies in one research – namely Needle in a Haystack, Islands of Excellence, Measurement Minutiae and Analysis Paralysis. The aim was to find out what factors are the most challenging companies to take advantage from big data revolution.

The first group is dealing with companies that are not mature enough to fully leverage the advantage of big data analytics but they are running after the latest trends and try to follow the hype. The problem with this approach that employee using analytics before they really understand how they could successfully apply analytics. Although they can find relationship and causation in the dataset they are arriving at false conclusions in most of the cases and just wasting time or money.

The second group, signed as Islands of Excellence, refers to users who are choosing applications to make a specific process optimized. This means that the company faces an issue where employees make only a specific process excellent and it is not connected across the supply chain, therefore it will not have any advantageous effects on other part of the chain.

Measurement Minutiaes are basically such firms which deal with a huge number of metrics and they try to measure everything internal and external. Problem with that they are lost in the metrics and sometimes it is hard for them to choose the correct one to get relevant and informative results. It is a big challenge for these companies to identify which metrics are need to be eliminated and create fewer customized metrics easier to manage.

The last group is the Analysis Paralysis for those ventures, which are complaining about having too much data and not being able to "digesting" them. First of all, it is clear for them that they have to do something with this vast amount of data gathered by Point of Sales terminals,

websites or social media, however they are unable to exploit analytics and technologies which are available for them. Overall, they are in the state of paralysis not even know where to start big data analytics (Sanders, 2016).

Data quality is another issue to that supply chain and logistics managers must pay attention. Here, I would like to shortly summarise the dimensions of data quality, which are accuracy, timeliness, consistency and completeness. The usefulness of data which clearly in connection with its quality could affect decision-making and company cost with a great extent - in case of an organisation poor data quality cost can be 8-12 % of the revenue, while in service industry it is more severe with 40-60 %. The biggest challenge here is that big data are not in common format that makes data transfer among different systems problematic, and after a while, these "complications" slowly make data science, predictive analytics and big data unattractive for firms (Hazen, et al., 2014).

Security obstacles are also coming into light with the inflow of large amount of information including security issues especially in case of data ownership, data storage/accessibility and data privilege (who can access the data) (Richey, et al., 2016).
Regarding data storage, the big question is whether a firm possess a private, internal systems with an appropriate security for all the information. The authors also identified that in emerging and developed countries strict legal regulations have the highest influence on decision making considering supply chain's big data security. On the other hand, countries have a different measure of concern about data sharing and data ownership, which can be a real obstacle for a global, multinational supply chain system. Firms operating in emerging economies are more conscious for protecting customer's privacy and information; what is more, their governments put also a pressure on firms to do so. Even though managers from different nations have different attitude to data security or to data sharing, most of them think data mining with their discovered knowledge could considerably decrease "risk associated, ill-informed decision-making" (Richey, et al., 2016, pp. 726).

It is clear that increased volume of data is advantageous to gather more information about customers and supply chain participants, however firms didn't expect that they have been so rapidly drown by them. First problem is that they do not yet possess an adequate hardware to store all of the inflowing data, plus they are also incapable to select information from them which will be valuable later on, and get a rid of the useless amount. Furthermore, it is quite challenging to correctly discard those data, which will never be useful, and keep the proper ones in data storage, since their "long-term" value is still undiscovered. Nevertheless,

companies must pay attention security and data protection issues too when they decided to throw away data from their systems (Richey, et al., 2016).

Management's mentality is considered as a barrier during the implementation of big data techniques as well. It is still a huge problem that managers bring decision using information based on relationships – like other employees' opinion, which is a less reliable source - rather than harnessing quantitative data results. It is especially a case in firms where the management include people who are working for several decades at the company and not "socialised" in a data-driven business environment, and this means there were no expectation towards them to seek big data out and utilize them. Solution would be to hire personnel who is capable to successfully apply big data both in upper and lower level in the enterprise. The best way is to choose these people from factories, warehouses of the company, since they have more information about the products with that it is easier to reach better quality of big data during the monitoring (Richey, et al., 2016).

5. Conclusions

It takes only few minutes to generate vast amount of data which is continuously happening in today's new era of internet of things. The increase development of technology makes possible that companies process these huge data amount and carry out real time analysis on them. As a consequence, data driven decision-making gain higher popularity and become an essential asset for taking the lead on the market as well as managing businesses successfully. This means that with the usage of advantages of big data and data analytics companies can better reach service level requirements, cut down on financial expenses ad cost of time regarding several business processes or deliver optimisation on such areas, which was unimaginable before.

My aim was to collect and read as many research paper as possible in this topic and then try to find a good categorisation of them, besides that I also show all the challenges, opportunities and drawbacks of this new phenomena.

It was seen that from the three sub categories of business intelligence and advanced analytics – descriptive, predictive and prescriptive analytics - most of the researcher used techniques related to predictive analytics when they set up their own model and algorithms for a specific supply chain problem. Data mining and machine learning algorithms were also in the focus as scientists mainly used cluster, regression and relational analysis in order to make their model functioning on big data set.

Furthermore, smart manufacturing and intelligent shop-floor were two new inventions where the scientists goal was to create machines with artificial intelligence in such way that they could already be able to learn from past actions and become a self-aware, self-maintain tools in production.

There were few cases when the authors came up with less popular methods used in data mining for making predictive analytics work. Good examples were the intelligent inventory system with its classification-based anomaly detection and the dynamic forecasting model with machine learning techniques, which could change a demand-driven inventory system to data-driven ones or artificial neural network model applied to show the most optimal path within transportation planning.

Considering managerial perspective there are also several advantages of using data analytics and exploiting big data along the whole supply chain, beginning from customer demand till purchasing from suppliers. First of all, the decision-makers can gain a deeper insight into customer needs, then analytics can help in better managing stock in and stock out

interdependencies as well as better predicting the health and maintenance of machines, and finally they enhance the visibility of purchasing goods. Taking a look at the two ends of the supply chain it can be said that big data analytics enable businesses to carry out real-time vendor management through which they can strengthen their negotiating power. At the same time, they can also complete on-time service and providing just-in time feedbacks to customer complaints on the other end of the chain.

Nevertheless, some challenges and barriers during the implementation and application of big data analytics were also discovered by the researchers. One of the main hurdles are the lack of common systems and standards at the different supply chain participants and the inappropriate implementation of data protection and security. With the huge inflow of personal private or commercial secret data, the companies must control intensively data sharing and accessibility which would require much more skilled workforce in IT field. Moreover, it is usually a case in immature businesses that they are incapable to choose the proper methods, unable to handle the enormous inflow of data and also struggling with using only quality data for decision-making. Finally, yet importantly, managers' mentality is also a barrier in such cases when the management does not want to accept and see the benefits of the data-driven environment. Consequently, these companies could easily loose market and lag behind their competitors after a while.

Besides all these facts, it has also become clear that there is a lack of literature and good algorithms in some part of supply chain (e.g. strategic and distribution network planning). Therefore, the researchers who has already conducted some experiment in these areas are suggesting some further research questions and motivating their scientific colleagues to try out other data mining approaches on similar optimization problems.

All in all, it is not doubtful that big data analytics is forging ahead with great intensity and a large percentage of companies will allocate their resources to invest in it in the future. Parallel to this, data scientists job will become more popular and other workforce will be substituted by self-working and self-learning algorithms and machines in many business areas, since they will be able to deliver tasks with significantly less mistakes. Today's supply chain leaders need to identify all these opportunities and challenges to know which capabilities of tomorrow they must require from their people in order to make an organisation ready for this new emerging trend.

Bibliography

Agrawal, R. & Srikant, R., 1994. *Fast Algorithms for Mining Association Rules in Large Databases.* San Francisco, Morgan Kaufmann Publishers Inc., pp. 487-499.

Andersson, J. & Jonsson, P., 2018. Big data in spare parts supply chains: The potential of using product-in-use data in aftermarket demand planning. *International Journal of Physical Distribution & Logistics Management*, 16 February , 48(5), pp. 524-544.

Arias, M. B. & Bae, S., 2016. Electric vehicle charging demand forecasting model based on big data technologies. *Applied Energy*, 1 December, Volume 183, pp. 327-339 .

Bae, J. K. & Kim, J., 2011. Product development with data mining techniques: A case on design of digital camera. *Expert Systems with Applications*, 1 August, 38(8), pp. 9274-9280.

Bakhsi, K., 2012. *Considerations for Big Data: Architecture and Approaches,* s.l.: Proceedings of the IEEE Aerospace Conference.

Benabdellah, A. C., Benghabrit, A., Bouhaddou, I. & Zemmouri, E. M., 2016. Big Data for Supply Chain Management: Opportunities and Challenges. *International Journal of Scientific & Engineering Research*, November, 7(11), pp. 20-26.

Bhagat, A., 2015. Understanding Big Data: Framework and Tools for Massive Data Storage and Mining. 3(6), pp. 305-308.

Bichescu, B. C. & Fry, M. J., 2007. Vendor-managed inventory and the effect of channel power. *OR Spectrum*, 29 September, 31(1), pp. 195-228.

Bose, R., 2008. Advanced analytics: opportunities and challenges. 18 September, 109(2), pp. 155-172.

Breiman, L., 2001. Random Forests. *Machine Learning ,* 45(1), pp. 5-32.

Brooks, R. & Dahlke, K., 2017. *Artificial Intelligence vs. Machine Learning vs. Data Mining 101 – What's the Big Difference?.* [Online] Available at: https://guavus.com/artificial-intelligence-vs-machine-learning-vs-data-mining-101-whats-big-difference/
[Accessed 20 November 2018].

Burns, R. B. & Burns, R. A., 2008. *Business Research Methods and Statistics Using SPSS.* USA: SAGE Publications Ltd.

Carmichael, I. & Marron, J., 2018. Data science vs. statistics: two cultures?. *Japanese Journal of Statistics and Data Science,* 14 May, 1(1), pp. 117-138.

Chen, H., Chiang, R. H. & Storey, V. C., 2012. BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT. *MIS Quarterly*, December, 34(4), pp. 1165-1188.

Chopra, S. & Meindl, P., 2013. *Supply Chain Management: Strategy, Planning and Operation.* 5.ed. s.l.:Peardon Education.

Chritopher, M. & Towill, D., 2001. An Integrated Model for the Design of Agile Supply

Chains. May, 31(4), pp. 235-246.

Cohen, R. P., 2013. *EDI Basics How Successful Businesses Connect, Communicate, and Collaborate Around the World.* Gaithersburg: GXS.

Cordeau, J.-F., Laporte, G., Martin W.P. Savelsbergh, M. W. & Vigo, D., 2007. Vehicle Routing. *Handbooks in Operations Research and Management Science*, Volume 14, pp. 367-428.

Dangeti, P., 2017. *Statistics for Machine Learning.* Birmingham: Packt Publishing Ltd..

De Mauro, A., Greco, M. & Grimaldi, M., 2016. A formal definition of Big Data based on its essential features. *Library Review,* 65(3), pp. 122-135.

Dean, J. & Ghemawat, S., 2010. MapReduce: A Flexible Data Processing Tool. *Communications of the ACM,* January, 53(1), pp. 72-77.

Demchenko, Y., Grosso, P. & de Laat, C., 2013. *Addressing big data issues in Scientific Data Infrastructure,* San Diego : 2013 International Conference on Collaboration Technologies and Systems (CTS).

Derakhshan, R., Orlowska, M. E. & Li, X., 2007. *RFID Data Management: Challenges and Opportunities.* USA, IEEE Xplore.

Donoho, D., 2017. 50 years of Data Science. *Journal of Computational and Graphical Statistics* , 19 December, 26(4), pp. 745-766.

Elgendy, N. & Elragal, A., 2014. *Big Data Analytics: A Literature Review Paper.* St. Petersburg, 14th Industrial Conference, pp. 214-227.

EMC Education Services, 2015. *Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing, and Presenting Data.* 1.ed. Indianapolis: John Wiley & Sons Inc..

Engel, T. et al., 2014. *A Conceptual Approach for Optimizing Distribution Logistics using Big Data.* USA, s.n.

Ertel, W., 2017. *Introduction to Artificial Intelligence.* 2.ed. Switzerland: Springer International Publishing AG.

European Union, 2013. *Opinion 03/2013 on purpose limitation published by Article 29 Data Protection Working Party.* [Online]
Available at: https://www.lexology.com/library/detail.aspx?g=ddf0de93-3ced-4887-bebd-af3ed8f62aa2
[Accessed 22 June 2018].

Evans, J. R., 2017. *Business Analytics.* 2.ed. United Kingdom: Pearson Education Limited.

Fernandez-Basso, C., Ruiz, M. D. & Martin-Bautista, M. J., 2016. Extraction of association rules using big data technologies. July, 11(3), pp. 178-18.

Gandomi, A. & Haider, M., 2014. Beyond the hype: Big Data concepts, methods, and analytics. *International Journal of Information Management,* 35(2015), pp. 137-144.

Goertzel, B. & Pennachin, C., 2007. *Artificial General Intelligence (Cognitive Technologies).* Germany: Springer-Verlag Berlin Heidelberg.

Gronwald, K.-D., 2017. *Integrated Business Information Systems: A Holistic View of the Linked Business Process Chain ERP-SCM-CRM-BI-Big Data.* Berlin: Springer-Verlag GmbH.

Gunasekaran, A., Mcgaughey, R. & Patel, C., 2004. A Framework for Supply Chain Performance Measurement. *International Journal of Production Economics* , Fabruary, 87(3), pp. 333-347.

Hammer, C. L., Kostroch, D. C. & Quirós, G., 2017. Big Data: Potential, Challenges, and Statistical Implications. *Staff Discussion Notes*, September, 2017(6).

Han, J. & Kamber, M., 2000. *Data Mining: Concepts and Techniques.* USA: Morgan Kaufmann Publishers.

Harrington, P., 2012. *Machine Learning in Action.* 1. ed. New York: Manning Publications Co.

Hazen, B. T., Boone, C. A., Ezell, J. D. & Jones-Farmer, L. A., 2014. Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, Volume 154, pp. 72-80.

Hernán, M. A., Hsu, J. & Healy, B., 2018. Data science is science's second chance to get causal inference right. A classification of data science tasks. *CoRR, abs/1804.10846.*

Hofmann, E., 2017. Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect. 55(17), pp. 5108-5126.

Hofmann, E. & Rutschmann, E., 2018. Big data analytics and demand forecasting in supply chains: a conceptual analysis. *The International Journal of Logistics Management,* 29(2), pp. 739-766.

Huang, T.-M., Kecman, V. & Kopriva, I., 2006. *Kernel Based Algorithms for Mining Huge Data Sets Supervised, Semi-supervised, and Unsupervised Learning.* Netherland: Springer-Verlag Berlin Heidelberg .

Hu, H., Wen, Y., Chua, T.-S. & Li, X., 2014. Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. Issue 2, pp. 652-687.

Ittman, H. W., 2015. The impact of big data and business analytics on supply chain management. *Journal of Transport and Supply Chain Management*, 9(1).

Jia, X., Feng , Q., Fan, T. & Lei, Q., 2012. RFID technology and its applications in Internet of Things (IoT). *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, pp. 1282-1285.

Keifer, S., 2010. *Beyond Point of Sale Data - Looking Forward, Not Backwards for Demand Forecasting.* [Online] Available at: http://www.gxs.fr/wp-content/uploads/wp_beyond_point_of_sale_data.pdf [Accessed 11 November 2018].

Kersting, K., 2018. *Machine Learning and Artificial Intelligence: Two Fellow Travelers on the Quest for Intelligent Behavior in Machines,* USA: Front. Big Data.

Kitchin , R. & McArdle, G., 2016. What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 17 February, pp. 1-10.

Klumpp, M. & Kandel, C., 2011. *GPS-BASED REAL-TIME TRANSPORT CONTROL FOR PRODUCTION NETWORK SCHEDULING SIMULATION.* Portugal, The 2011 European Simulation and Modelling Conference.

Koturwar, P., Girase, S. & Mukhopadhyay, D., 2015. *A Survey of Classification Techniques in the Area of Big Data,* India: arXiv.

Kurasova, O. et al., 2014. *Strategies for Big Data Clustering.* Cyprus, 2014 IEEE 26th International Conference on Tools with Artificial Intelligence.

Lee, J., Kao, H.-A. & Yang, S., 2014. *Service innovation and smart analytics for Industry 4.0 and big data environment.* s.l., Elsevier B.V..

Leveling, J., Edelbrock, M. & Otto, B., 2014. *Big Data Analytics for Supply Chain Management.* Malaysia, IEMM.

Li, H.-L., 1999. Incorporating Competence Sets of Decision Makers by Deduction Graphs. *Operations Research*, 1 April, 47(2), pp. 175-344.

Lu, D., 2011. *Fundamentals of Supply Chain Management.* 1. ed. Denmark: Ventus Publishing Aps.

Lummus, R. R. & Vokurka, R. J., 1999. Defining supply chain management: a historical perspective and practical guidelines. *Industrial Management & Data Systems*, 99(1), pp. 11-17.

Maimon, O., Khmelnitsky, E. & Kogan, K., 1998. Optimal Flow Control in Manufacturing Systems - Production Planing and Scheduling. *Applied Optimization*.

Maimon, O. & Rokach, L., 2005. Introduction to Knowledge Discovery in Databases. In: O. Maimon & L. Rokach, 2005.*Data Mining and Knowledge Discovery Handbook.* ed. Boston: Springer, pp. 1-17.

McCarthy, R. V., McCarthy, M. M., Ceccucci, W. & Halawi, L., 2019. *Applying Predictive Analytics: Finidng Value in Data.* 1.ed. Basel: Springer International Publishing.

Meyr, H., Wagner, M. & Rohde, J., 2002. Structure of Advanced Planning Systems. In: *Supply Chain Management and Advanced Planning.* Germany : Springer, Berlin, Heidelberg, pp. 99-104.

Mikavica, B., Kostić-Ljubisavljević, A. & Radonjić Đogatović, V., 2015. *BIG DATA: CHALLENGES AND OPPORTUNITIES IN LOGISTICS SYSTEMS.* Belgrade, s.n.

Min, H., 2016. *Global Business Analytics Models.* New Jersey: Pearson FT Press.

Myerson, J. M., 2007. *RFID in the Supply Chain: A Guide to Selection and Implementation.* 1.ed. New York: Auerbach Publications.

Ohlhorst, F., 2012. *Big Data Analytics: Turning Big Data into Big Money.* USA: John Wiley & Sons Inc..

Pandey, K. K., Yadu, R. K., Dwivedi, A. & Shukla, P. K., 2015. A Analysis of Different Type of Advance database System For Data Mining Based on Basic Factor. *International Journal on Recent and Innovation Trends in Computing and Communication*, 3(2), pp. 456-460.

Prasad, S., Zakaria, R. & Altay, N., 2016. Big data in humanitarian supply chain networks: a resource dependence perspective. *Annuals of Operations Research; S.I.: Big Data Analytics in Operations & Supply Chain Management.*, 4 August.

Provost, F. & Fawcett, T., 2013. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking.* 1.ed. California: O'Reilly Media Inc..

Quintero, D. et al., 2015. *IBM Software Defined Infrastructure for Big Data Analytics Workloads.* 1.ed. USA: IBM Redbooks.

Richey, R. G., Hall, K. L. & Adams, F. G., 2016. A global exploration of Big Data in the supply chain. *International Journal of Physical Distribution & Logistics Management*, September, 46(8), pp. 710-739.

Rozados, I. V. & Tjahjono, B., 2014. *BIG DATAANALYTICS IN SUPPLY CHAIN MANAGEMENT: TRENDS AND RELATED RESEARCH*, Bali: 6th International Conference on Operations and Supply Chain Management.

Russom, P., 2011. BIG DATA ANALYTICS. *TDWI best practices report*.

Rynarzewski, T. & Szymczak, M., 2016. *Changes adn Challenges in the Modern World Economy.* Poznan: Pueb Press.

Sanders, N. R., 2016. How to Use Big Data to Drive Your Supply Chain. *California Management Review*, 58(3), pp. 26-48.

Sanislav Teodora & Miclea, L., 2012. Cyber-physical systems - Concept, challenges and research areas. *Control Engineering and Applied Informatics*, 14(2), pp. 28-33.

Sathi, A. D., 2012. *Big Data Analytics.* Boise: MC Press Online, LLC.

Schoenherr, T. & Speier-Pero, C., 2015. Data Science, Predictive Analytics, and Big Data in Supply Chain Management: Current State and Future Potential. *Journal of Business Logistics*, 36(1), pp. 120-132.

Shenle, P., Ballot, E., Montreuil, B. & Huang, G. Q., 2017. Physical Internet and Interconnected Logistics Services: Research and Applications. *International Journal of Production Research*, 55(9), pp. 2603–2609.

Shetty, D. V. & Chidimar, S. J., 2016. Comparative Study of SQL and NoSQL Databases to evaluate their suitability for Big Data Application. pp. 314-318.

Siegel, E., 2013. *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die.* 1.ed. 2013: Wiley Publishing.

Simchi-Levi, D., Kaminsky, P. & Simchi-Levi, E., 1999. *Designing & Managing the Supply Chain: Concepts, Strategies & Case Studies.* 2.ed. s.l.:McGraw-Hill Higher Education.

Stadtler, H., 2004. Supply chain management and advanced planning—basics, overview and challenges. *European Journal of Operational Research*, 6 May, p. 575–588.

Tan, K. H. et al., 2015. Harvesting big data to enhance supply chain innovation capabilities: Ananalytic infrastructure based on deduction graph. *International Journal of Production Economics*, 5 January, Band 165, pp. 223-233.

Tan, M. H. & Lee, W. L., 2015. Evaluation and Improvement of Procurement Process with Data Analytics. *International Journal of Advanced Computer Science and Applications*, 6(8), pp. 70-80.

Tan, P.-N., Steinbach, M. & Kumar, V., 2006. *Introduction to Data Mining.* 1.ed. Boston: Pearson Education Inc..

Toit, D. D. & Vlok, P.-J., 2014. SUPPLY CHAIN MANAGEMENT: A FRAMEWORK OF UNDERSTANDING. 25(3).

Trkman, P., McCormack, K., Valadares De Oliveira, M. P. & Bronzo, M., 2010. The impact of business analytics on supply chain performance. *Decision Support Systems*, 30 June.

Tupper, C. D., 2011. 20 - The Enterprise Data Warehouse. In: C. D. Tupper, 2011. *Data Architecture: From Zen to Reality.* ed.1. USA: Morgan Kaufmann, pp. 357-368.

Vogt, H., 2009. *Open Source Customer Relationship Management Solutions: Potential for an Impact of Open Source CRM Solutions on Small- and Medium Sized Enterprises.* Hamburg: Diplomica Verlag GmbH.

Waller, M. A. & Fawcett, S. E., 2013. Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, 11 June.

Wang, G., Gunasekaran, A. & Ngai, E. W., 2016. Distribution network design with big data: model and analysis. *Annals of Operations Research; S.I.: Big Data Analytics in Operations & Supply Chain Management.*, 30 June.

Wang, L. & Alexander, C. A., 2015. Big Data Driven Supply Chain Management and Business Administration. *American Journal of Economics and Business Administration*.

Whitteker, W., 2014. *Point of Sale (POS) Systems and Security,* USA: SANS Institute.

Wisner, J. D., Tan, K.-C. & Leon, G. K., 2015. *Principles of Supply Chain Management: A Balanced Approach.* 4. ed. s.l.:Cengage Learning.

Yadava, H., 2007. *The Berkeley DB Book.* New York: Springer Verlag.

Zeng, X., Lin, D. & Xu, Q., 2011. *Query Performance Tuning in Supply Chain Analytics,* China: 4th International Joint Conference on Computational Sciences and Optimization.

Zhong, R. Y. et al., 2015. Visualization of RFID-Enabled Shopfloor Logistics Big Data in Cloud Manufacturing. *International Journal of Advanced Manufacturing Technology,* 84(1-4), pp. 5-16.

Zhong, R. Y., Xu, C., Chen, C. & Hunag, G. Q., 2015. Big Data Analytics for Physical Internet-based intelligent manufacturing shop floors. *International Journal of Production Research,* 55(9), pp. 2610-2621.

Zhou, Q. et al., 2017. *An Advanced Inventory Data Mining System for Business Intelligence.* USA, s.n.

Zhu, D., Du, H., Sun, Y. & Cio, N., 2018. Research on Path Planning Model Based on Short-Term Traffic Flow Prediction in Intelligent Transportation System. *Sensors,* 5 December.18(12).

Zhu, J., 2013. POS Data and Your Demand Forecast. *Procedia Computer Science,* 17(2013), pp. 8-13.

Appendix

Zusammenfassung

Ich erforsche einen sehr populäreren Bereich der heutigen Zeit in meiner Diplomarbeit, wobei mein Ziel ist, die akademische Literatur zu studieren und ihre Ergebnisse zum Thema Datenanalyse und Big Data im Supply Chain Management zusammenzufassen.

Im ersten Schritt versuche ich, die Konzepte zu klären, die sich auf die Bereiche Datenanalyse sowie Big Data beziehen, sowie die Beziehung zwischen ihren Konzepten, Methoden und Algorithmen. Dies war eine besonders anspruchsvolle Aufgabe, da es zu diesen Elementen der Big Data Analytics noch keine allgemein anerkannte Terminologie und Taxonomie existieren.

Im nächsten Schritt kategorisiere ich die Forschungsarbeiten nach dem Element von dem Supply Chain Planning Matrix und demonstriere die Tools, Methoden und Algorithmen, die eingesetzt wurden, um Prozesse zu optimieren und die Effizienz an einem bestimmten oder gesamten Teil der Kette zu steigern.

Abschließend gebe ich einen kurzen allgemeinen Überblick über die Vorteile, Herausforderungen und Schwierigkeiten bei der Verwendung dieser Analysen.