CHAPTER 5

STATISTICAL MACHINE LEARNING IN TERMS OF INDUSTRY 4.0 AND INVESTIGATION OF THE IMPACT OF BIG DATA ON THE COMPETITIVENESS OF FIRMS

Kutluk Kağan SÜMER*

*Prof. Dr., İstanbul University, Faculty of Economics, Department of Econometrics, İstanbul, Turkey

E-mail:kutluk@istanbul.edu.tr

DOI: 10.26650/B/SS10.2020.001.05

Abstract

The ultimate goal of Industry 4.0 is to deliver real-time data to network-based information technology systems, which are always connected to machines, components, and ongoing work. They use machine learning and artificial intelligence algorithms to analyze and obtain information from these big data and adjust processes automatically as needed. Statistical machine learning techniques are designed to extract information from existing data. Statistical machine learning is largely based on statistical optimization and forecasting techniques. As a result of the analysis of big data gathered by statistical techniques with statistical machine learning methods, both manufacturers and service sector companies using these new techniques and methods have higher competitive power compared to companies that cannot adapt to these new techniques. In this study, statistical machine learning in terms of Industry 4.0 and the effect of big data on the competitiveness of firms have been investigated.

Keywords: Industry 4.0, Statistical machine learning, Big Data, Competition, Innovation, Innovative solutions.

1. Introduction

The initial Industrial Revolution, which covered the second half of the 18th century and the first half of the 19th century, changed the world into a system where manual methods were replaced by mechanical, motorized production processes and vehicles. After all, the first Industrial Revolution resulted in the mass production and factory system.

After rapidly-passing 200 years, today, the fourth industrial revolution of the industry conquers 21st-century manufacturers - Industry 4.0. The latest digital technologies are used for optimizing and mechanizing the manufacturing, including high-level supply chain processes. The final purpose of Industry 4.0 is to provide real-time data to network-based IT systems by always-connected sensors implanted in machines, parts, and continuous work. They use machine learning and artificial intelligence algorithms to analyze and obtain information from these big data and to automatically adapt processes as required.

The challenges of contemporary production systems are growing complexity, dynamic features, high-dimensionality, and disordered structures. Fast-paced advances in the field of algorithms and increasing the usability of data (for example, due to low-cost sensors and switching to intelligent production) and with the increasing computing power, especially applications for machine learning in manufacturing are increasing rapidly. Today, controlled algorithms provide superiority in most applications in the field of production. However, supervision and learning techniques are rapidly gaining in importance due to the rapid increase in existing data, more and better sensor technologies, and increase in awareness. Even in today's technology, hybrid approaches are actively used. This has made it a necessity to consider Big Data developments in recent years and to use them widely in all areas. Intelligent production systems for many applications in the production area and intelligent machine learning techniques, and big data are powerful tools, and their importance will further increase in the future. With their interdisciplinary nature, machine learning and big data techniques provide great opportunities for new developments in competitive power. However, this interdisciplinary character also maintains its importance as an important risk factor, which is essential for the development of cooperation among different disciplines such as Computer Science, Industrial Engineering, Mathematics, Statistics, and Electrical Engineering simultaneously against today's firms.

Today, the manufacturing industry is experiencing a data increase that has never been seen before. These data are collected raw from semantics, quality, from various forms such as sensor data, environmental data, machine tool parameters from the production line.

Different names can be used for this phenomenon, for example, Industrie 4.0 (Germany), Intelligent Manufacturing (USA), and Intelligent Factory (South Korea). These large amounts of data increase and availability are often referred to as Big Data. In general, the manufacturing industry can be concluded safely, to benefit from extended data accessibility, for example, quality improvement actions, production cost estimation and/or process optimization, a better comprehension of customer needs, etc., for this, it also requires support to address relevant high-dimensionality, complexity, and dynamics.

New developments in some areas, such as mathematics and computer science (e.g. statistical learning) and the existence of easy-to-use, usually freely available (software) tools, offer a great capacity to ensure a sustained grasp of the production area and their growing production data warehouse. One of the most exciting developments is in the field of machine learning (data mining, artificial intelligence, data discovery from databases, etc.). Nonetheless, the area of machine learning is highly diversified and there are a lot of different algorithms, theories, and methods utilizable. For numerous manufacturing practitioners, this constitutes an obstacle to the adoption of these powerful tools and can, therefore, prevent the use of an increasing amount of available data.

In many mature economies, the contribution of production to GDP has declined over the last decade, becoming a major problem. It is also known that a number of important initiatives have been started to renew the manufacturing sector in recent years. President Obama's announcement of new action plans in 2014 under the title "Enforcement Actions to Strengthen Advanced Manufacturing in America" to further strengthen US manufacturing is an example of these initiatives. Again, the European Union's "Factories of the Future" initiative in 2016 is another example of these initiatives. The challenges facing production now are very different from those experienced before.

There are many pieces of research suggesting the key challenges of production at the global level. It is possible to summarize the key challenges that most researchers address as follows:

- Adoption of sophisticated manufacturing technologies.
- The increasing significance of manufacturing products with high added value.
- Utilizing superior knowledge, knowledge management, and Artificial Intelligence systems.
- Sustainable production (methods) and products.
- Swift and adaptable business capabilities and supply chains.

- Innovation in products, services, and processes.
- Close cooperation between industry and research for adopting new technologies.
- Modern production management standards.

These key challenges emphasize the tendency of the production area to be more complicated and dynamic. This obvious complexity increases not only in the production programs itself but also in the processes of companies and collaborative networks (business) as well as the product to be produced. Adding the difficulty of managing and controlling complexity in manufacturing further increases the uncertainty of the dynamic business environment of existing manufacturing firms.

Artificial Intelligence and machine learning techniques make it compulsory to use techniques to manage complexity, production changes and uncertainties. In particular, in the areas that are most likely to be optimized, for example, monitoring and control, programming and diagnostics, it is seen clearly that increasing the availability of data adds another difficulty: Big amount of available data as well (e.g., sensor data), high dimensionality and diversity (e.g., different sensors due to or related processes) data, besides, production optimization problem of the complementary nature of controls. Determination of product and processing state drive in production systems using machine learning is that the subject new developments are happening. The current candidate methods to overcome some of the challenges of today's complex production systems are machine learning techniques. These data-driven methods can find rather complicated and nonlinear models in data of different types and sources, and then work on raw data by estimation, detection, classification, regression, or estimation.

2. Big Data

Most companies store and use large amounts of information. With the technical advancement that can be seen, particularly in the area of information technology, there has been a significant rise in the demands and emphasis on information storage, analysis, and processing. IBM states that we generate 2.5 million (2.5×10^{18}) bytes of data daily, accordingly over the past two years, 90 percent of the data of the world has been produced. These are in the forms of electronic correspondences, broadcasts on social media, digital images, videos, invoices, sensors, and such, all of which describe the notion of Big Data (IBM, 2013).

It will be best to represent and describe the characteristics of Big Data in this chapter before defining the term of Big Data. It is nearly improbable to attain a precise and uniform definition of Big Data, as this notion is not formally codified or integrated. Here, various acknowledged definitions of Big Data will be presented for a better apprehension of the notion.

A conventional and very common definition was composed by the McKinsey Global Institute: (McKinsey Global Institute, 2011): “Big Data indicates information sets which surpass in size the capacity of standard database program tools to acquire, store, handle, and evaluate.” The above description points out a Big Data-related concern that businesses encounter. The volume and variety of information are so vast and diverse that companies can not operate using conventional systems and instruments that they are used to.

In his article titled “Big Data”, John Gantz explained as follows (Reinsel - Gantz, 2011): “Big Data technologies represent the systems and structures, which are created for efficient data generation from a broad spectrum of information, high-velocity capture, detection, and/or analysis.” The mentioned definition is designated as a worldwide term for immense and complicated technology clusters intending for controlling and analyzing Big Data, which is unorganized information essential to an enterprises' management and growth. One aspect of the whole concept was defined by Douglas Beyer as follows (Laney - Beyer, 2012): “Big data are high-volume, high-speed, and/or high-variety information assets that require new processing forms for advanced decision making, insight exploration, and optimization of the processes”. With this in mind, Big Data signifies sets of data that are not able to be saved or processed employing standard techniques and instruments; this indicates that it is challenging to process them since they cannot be stored in one place, but they must be distributed.

2.1. Key features of Big Data

Even though definitions of Big Data are not coherent, they all refer to three main points (Big Data of 3) that distinguish them from normal data. The first is the volume of data, then the frequency of generating new data, and finally the speed, which is diversity as the form of the data. More features can be added since the Big Data recognition process has just started.

The main characteristic of Big Data is the volume, to put it another way, the data quantity. It is possible to describe it as the physical potential at which it can be utilized (in fundamental classifications, this capacity is gauged between terabytes and petabytes), or measurements can be conducted via the values of the records, transactions, tables or files.

As businesses concentrate on the fact that data is an asset, their reluctance to dispose of data and reduce their size increases. This matter, at the same time, concerns the loss of the relative value of each additional byte. The reverse situation concerns the increase of the relative value of every extra data storage unit needed for data protection.

Velocity can be defined as the production speed of Big Data. At the same time, it can be understood as the time period from transferring data to making the decision by one's own, from the moment the data is received to the moment it is analyzed.

Speed can still be considered a feature that increases its significance, yet it can be presumed that it will have the highest importance compared to other features. This assumption is verified by the reality that many sectors require real-time data processing, and that their judgments almost entirely depend on those data.

Data diversity is the principal constituent of Big Data. Traditional information which is organized is usually called as structured data. A good exemplar of this information is the data stored in the storage structures in databases. In recent years, however, it has been mostly classified as unstructured content and semi-structured content.

There can be numerous types of unstructured content and describing them in depth is not easy. A vast majority of the contemporary formats are included in this group, comprising audio and video data, social network records, blogs, information of geolocation, network click logs, the information accessible on the Internet, and more.

Big Data is aimed at adopting and incorporating all the abovementioned information, integrating those in the form appropriate for subsequent company procedures.

2.2. Using Big Data for the Company's Competitive Edge

Then, why is Big Data or the technology of Big Data so essential to businesses? Above all, they make it easier to comprehend and infer meaning from all information fields in the world. Corporations and organizations have collected and stored data which are parts of each transaction. All that information has been essentially utilized to monitor or predict the future. Nowadays, these data are bursting. It is probable to collect information on every subject.

For instance, customers visiting a company's website, thus, marketing professionals can accumulate information about each consumer who is interested in their brand or product.

All those resources are like a treasure for businesses because they can give the companies a point of view of the consumers' mind. Nonetheless, this necessitates the implementation of new methods, technologies, and mechanisms termed Big Data. The information is here and withholds precious data, you merely need to find a system to investigate how to do it.

It will be appropriate to mention a very real comparison that has been discussed in recent years. Besides, this is where garbage is collected, as well as data, given the issues of polluting

the Earth. It's up to us if we leave the trash on the ground, pick it up and allow it to pollute the planet later, or we take and recycle it. The same problem applies for Big Data, the data in corporations is the same as waste and the same lies, no one uses them, and they just "pollute" the computer. We must use these data for our own benefit. Wherever we look, we are surrounded. If this "garbage" is put into the required form, we can make use of it and raise its value in the future.

Big Data solutions are not only perfect for examining raw structured data but also for examining semi-structured and unstructured data from any sources mentioned in the prior section. Additionally, Big Data solutions are excellent if the entirety or the majority of data requires to be examined, or if exemplification from the data is not as effective as a larger data set.

As reported by McKinsey Global Institute, Big Data is able to create value in five ways (McKinsey Global Institute, 2011):

- It can form a transparent environment by making the new potential more widely available.
- Allows companies to conduct experiments. For example, they can conduct experiments on process changes and examine big quantities of information obtained from these studies in order to recognize the potential efficiency enhancements.
- Big Data is utilizable to produce a more comprehensive client segmentation to individualize the information processing as well as arrange customer-specific services.
- Examination of Big Data might assist humans in making decisions by indicating the latent connections or some concealed perils. For instance, insurance firms might have machines that perform risk or fraud assessments. Decisions of lower complexity might be mechanized utilizing these systems.
- Data may further enable the introduction of innovative business models, goods and services, or upgrade those available. It is possible to employ data on how goods and services are used to originate and refine modern models of products.

Companies can achieve a high competitive edge and outperform their competitors by employing Big Data and making use of its advantages. While Big Data is less understood, it offers businesses greater growth potential than traditional technologies. Corporations which are more improved in this sense and appreciate the value of Big Data more quickly, can succeed to get a leadership status in their industries in terms of competition. The significance of this notion should not be overlooked by businesses.

2.3. Industry 4.0 and Big Data

As we said before, manufacturers have been producing plenty of data in terms of real-time production and quality for a while. Nonetheless, there are not sufficient platforms, which can hold various sources of data regarded trash and deduce comprehensive insights to enhance quality, productivity, and such, therefore, it is quite common that these isolated data lakes are “wasted”. To put it another way, the problematic issue is about the capability of efficiently deriving value from data, not about producing and accumulating it.

Figure 1: Sources: The Industrial Internet of Things Volume G1: Reference Architecture, Industrial Internet Consortium

Industry 4.0 big data comes from many and various sources:

- Threshold characteristics such as product and/or machine design data
- Machine processing data obtained from control systems
- Data of product and process quality
- Manual transaction records performed by personnel
- Manufacturing Execution Systems
- Information on production and operating costs
- Fault detection and other installations for system monitoring
- Logistics-related information, including third-party logistics
- Customer data regarding product usage, feedback, and more

Some of those data sources are structured as we have previously explained (such as sensor signals), some are semi-structured (such as manual transaction records), and some are not fully configured (such as image files). However, in most circumstances, most of the data is not used or is used only for very certain tactical objects. A key factor in, generally not strategically, exploiting Industry 4.0 big data is poor interoperability between incompatible technologies, systems, and data types; a second key factor is that traditional IT systems do not store, manipulate and manage those large data volumes at high speeds.

What businesses need, therefore, are state-of-the-art programs which can completely enhance big data generation utilizing machine learning, artificial intelligence, and predictive analytics.

Today, by gathering, evaluating and exchanging information in all main functional areas, production companies are attempting to obtain real business intelligence. Not only are manufacturing technologies more effective in this architecture, but they can also react in time to altering company requirements, including messages from associates and clients.

The model below focuses more on large data and analytical flows at plant and factory levels.

Low (orange) stacks collect, process, and analyze data flow from the production area quickly and scalably. The upper (blue) stacks are likely for large-scale and intense batch analyses applied in cloud-based Big Data frames. It can be seen that the bulk analytical stack, at the same time, receives stored plant/factory big data as an input. In order to optimize production processes and applications, both flow and batch analysis outputs are distributed as information.

Figure 2: Source: Shi- Wah Lin, IIoT (The Industrial Internet of Things) for Smart Manufacturing part 3 - A New Digitalization Architecture, October 16, 2017

2.4. Industry 4.0 Big Data Usage Examples

In 2016, PwC carried out a worldwide survey of the adoption of Industry 4.0 in a wide range of industries, including automotive, electronics, industrial manufacturing, aerospace, and defense and security. Surveyors expected that 2020 Industry 4.0 applications, including big data analytics, would decrease production and operating costs by 3.6%, resulting in cumulative savings of \$ 421 billion.

The following are several chosen real-life instances showing how the Industry 4.0 big data vision can add assessable value to manufacturing businesses:

Combining data of quality and production for improving the quality of production: A semiconductor producer started to associate the single-chip data seized at the end of the production process with the process data previously obtained from the process. The producer can thus detect faulty chips early, and remarkably enhance the quality of the manufacturing process.

Empowered clients: The automotive industry embraces Industry 4.0 with enthusiasm to fulfill customer expectations in a cost-effective way for further economic and digitally connected automobiles. Most use cases of large data to be produced by the connected means

include continuous data exchange with the manufacturing company. In addition to improving the after-sales service to the individual vehicle owner, aggregate data on vehicle performance can be employed to enhance quality processes and future designs.

Reduced downtime: Industry 4.0 large data analytics, applicable to many industrial sectors, can reveal guides predicting mechanical or procedural failures before they happen. Machine supervisors can evaluate the processes or machine performances in real-time, and in most circumstances avoid unplanned downtime.

Industry 4.0 recommends predictive production in future industries. The machines are connected as a common community. This evolution requires the use of prediction tools, thus data can be systematically translated into information which is able to clarify uncertainties, and in this way makes more “informed” decisions possible.

This includes industrial big data, that makes manufacturing services and production analytics more important than in previous years, which changed the value proposition of manufacturers. In order to continue with these trends, a systematic framework is proposed for self-recognizing and self-servicing machines. The framework includes the concepts of cyber-physical system and decision support system.

To summarize, the prognostic monitoring system is a trend of intelligent production and industrial big data environment. There are many areas where four key domains are envisaged to be effective in the emergence of the fourth industrial revolution:

- Machine health forecasting will reduce machine downtime and support the ERP system to optimize prognostic information, production management, maintenance timing, and guarantee machine safety.
- The flow of information between production line, enterprise management level and supply chain management makes industry management more transparent and orderly.
- The new industry trend will reduce labor costs and provide a better working environment.
- Finally, energy savings, optimized maintenance schedule, and supply chain management reduce costs.

3. Statistical Machine Learning

In this part of the study, first of all, the primary benefits, difficulties, and requirements of machine learning applications related to production will be explained. Next, the current state of the art in the field of machine learning will be reviewed by focusing again on manufacturing

applications. In this context, different machine learning techniques and algorithms configuration will be developed and presented.

Advantages of the machine learning application over today's production challenges:

Machine learning is often known for its capability to deal with many problems of nature that arise in the field of intelligent production.

It offers soft computing and hybrid artificial intelligence approaches to intelligent production.

The implementation of machine learning techniques has, in the last 20 years, been due to several factors, for example; Increased presence and strength of existing Machine learning tools and the existence of big amounts of complicated data with a small degree of transparency.

The main definition of Machine learning, however, allows computers to solve problems before they are specifically programmed.

At present, machine learning is already in different production areas, for example; optimization, control, and troubleshooting.

Many machine learning techniques (e.g. Support Vector Machine [SVM]) are designed to analyze large amounts of data and handle high dimensionality (> 1000) very well.

However, accompanying considerations, such as possible over-fitting, should be considered. It supports vector machine for machine status monitoring and diagnostics.

If dimensionality arises as a problem despite the possibility of the power of algorithms, there are possible techniques to decrease the dimensions. These decrease the effect of diminishing dimensionality on expected results.

The importance of using machine learning is that in this circumstance, SVM dimensionality is not a practical problem, and thus, the need to reduce dimensionality is decreased. This indicates the possibility of being more flexible in containing apparently unrelated information in manufacturing data, which might be related in particular situations. This might have an immediate effect on the present information gap previously defined.

The application of machine learning in production can lead to the acquisition of patterns from existing datasets, which may form the basis for developing approaches for the future behavior of the system.

This recent information can assist process owners when making decisions, or can be utilized automatically to enhance the system immediately. Finally, the purpose of some machine learning techniques is to identify specific patterns or orders that define relationships.

Considering that machine learning, which is a rapidly changing, dynamic production environment, is a component of artificial intelligence, and the difficulty of inheriting the ability to learn and adapt to changes, the system designer does not need to provide solutions for all possible situations.

Therefore, it provides strong arguments that applying machine learning in production may be beneficial, given the challenge of most first principle models in dealing with adaptability. Learning and adapting automatically from changing environments is a major strength of machine learning.

Machine learning techniques are designed to extract information from existing data.

Alpaydin (2010) emphasizes that stored data will only be useful when analyzed and, for example, converted into information that we can use to make predictions.

This is especially true for production when struggling to obtain real-time data throughout a live production program that has technical, financial, and information-related limitations. This can also be effective in positioning process control points (Wuest, Liu, Lu & Thoben, 2014).

Although it is logical to carefully select control points regardless of which data are useful, given the analytical power of machine learning techniques, it may not be possible to obtain information from previously futile data. This may end up with the capacity to collect further data in the whole production schedule. It is a clear question of whether this is useful or not. Considering the ability of machine learning to manage high-dimensional data, the technical side of examining extra data is not difficult. Nevertheless, with regard to data capture, especially the ability to collect data can still be a problem. When data are obtained, the identification of status drives is not considered problematic and is not repeated frequently in very high dimensional situations.

The table below provides a review of the theoretical capabilities of machine learning methods to address the principal challenges in manufacturing applications (requirements) (Table 1).

Table 1: Source: Thorsten Wuest, Daniel Weimar, Christopher Irgens & Klaus- Dieter Thoben
Machine learning in manufacturing: Advantages, challenges, and applications, Journal of Production
& Manufacturing Research, Volume 4, 2016 - Issue 1

Table 1. Summary of suitability of ML techniques in manufacturing application.

Manufacturing requirement	Theoretical ability of ML to meet requirements
Ability to handle high-dimensional problems and data-sets with reasonable effort	Certain ML techniques (e.g. SVM) are capable of handling high dimensionality (>1000) very well. However, accompanying issues like possible over-fitting has to be considered (Widodo & Yang, 2007; Yang & Trewn, 2004)
Ability to reduce possibly complex nature of results and present transparent and concrete advice for practitioners (e.g. monitor XX and parameter YY at checkpoint ZZ)	ML may be able to derive pattern from existing data and derive approximations about future behavior (Alpaydin, 2010). This new information (knowledge) may support process owners in their decision-making or used to automatically improve a system
Ability to adapt to changing environment with reasonable effort and cost. Ideally a degree auf 'automated' adaptation to changing condition	As ML is part of AI, and thus be able to learn and adapt to changes, 'the system designer need not foresee and provide solutions for all possible situations' (Alpaydin, 2010). Learning from and adapting to changing environments automatically is a major strength of ML (Lu, 1990; Simon, 1983)
Ability to further the existing knowledge by learning from results	ML can contribute to create new information and possibly knowledge by, e.g. identifying patters in existing data (Alpaydin, 2010; Pham & Afify, 2005)
Ability to work with the available manufacturing data without special requirements toward capturing of very specific information at the start	ML techniques are designed to derive knowledge out of existing data (Alpaydin, 2010; Kwak & Kim, 2012). 'The stored data becomes useful only when it is analyzed and turned into information that we can make use of, for example, to make predictions' (Alpaydin, 2010)
Ability to identify relevant process intra- and inter-relations & ideally correlation and/or causality	The goal of certain ML techniques is to detect certain patterns or regularities that describe relations (Alpaydin, 2010)

3.1. Structuring machine learning techniques and algorithms

In recent years, machine learning has been used within a wide range of research and applications. This has led to various subfields, algorithms, theories, application fields and such. Different researchers choose different approaches to construct the area. The following figures illustrate these algorithms (Figure 1, 2, 3).

Figure 3: An overview of tasks and main algorithms in DM (Corne et al., 2012)

Figure 4: Classifications of main ML techniques according to Pham and Afify (2005)

Figure 5: Structuring of ML techniques and algorithms

As a result, machine learning is largely based on statistical optimization and forecasting techniques.

Basically, these techniques can be grouped under three headings:

1. Uncontrolled machine learning: The definition is that there is not any feedback given by an external teacher/knowledgeable specialist within uncontrolled learning. It has introduced the rule that the algorithm itself, eg. if examples of conceptual integrity of the attributes are not tagged (no known labels and no correct corresponding outputs), is probably uncontrolled learning. The aim is to explore categories of objects which are unknown by clustering. Particularly in the context of Big Data, uncontrolled techniques are becoming more and more important. Nevertheless, as with the production practice, the basic hypothesis is that knowledgeable specialists can give feedback on the sorting of states to define the learning set to train the algorithm.

2. Reinforcement learning: Reinforcement learning is described by the supply of educational information by the environment. A digital boost signal provides data about the performance of the system in a particular sequence. Another descriptive characteristic is that the student should try to find out the actions giving the optimal outputs (digital amplification signal) by testing instead of being told. This distinguishes reinforcement learning from most other methods of machine learning. However, reinforcement learning is regarded by some researchers as “a special form of supervised learning”.

3. Supervised machine learning: In manufacturing practice, usually, supervised machine learning methods are implemented because the problems are intense in data, yet less in knowledge. In addition, supervised machine learning can benefit from data collection in manufacturing for statistical process control purposes, and based on that this data is mostly tagged, supervised machine learning is in the manner of learning from examples provided by a knowledgeable external auditor. Once an algorithm is selected, it is trained to utilize the training data set. To assess the capability to fulfill the intended task, the trained algorithm is assessed via a set of evaluation data. Dependent on the achievement of the algorithm trained by the evaluation algorithm, parameters can be set to optimize achievement if the achievement is already good. If the performance is unsatisfactory, the procedure should be restarted in a prior stage, based on the actual performance.

In principle, 70% of the data set is employed as a training data set, 20% as an assessment data set (for setting parameters – e.g. bias) and 10% as test data.

4. Conclusion

The ultimate goal of Industry 4.0 is to deliver real-time data to network-based information technology systems, which are always connected to machines, components, and ongoing work. They use machine learning and artificial intelligence algorithms to analyze and obtain information from these Big Data, and adjust processes automatically as needed. Statistical machine learning techniques are designed to extract information from existing data. Statistical machine learning is largely based on statistical optimization and forecasting techniques. As a result of the analysis of Big Data gathered by statistical techniques with statistical machine learning methods, both manufacturers and service sector companies using these new techniques and methods have higher competitive power compared to companies that cannot adapt to these new techniques.

Big Data brings several advantages to companies. It makes the information inside more transparent. It provides a wider, deeper and more accurate insight. Therefore, it also improves decision-making. It allows companies to create a more complex and complete image of their customers and therefore, offers more accurately designed products and services.

References

- Akay, D. (2011). Grey relational analysis based on instance based learning approach for classification of risks of occupational low back disorders. *Safety Science*, 49(1), 1277–1282.
- Alpaydin, E. (2010). *Introduction to machine learning* (2nd ed.). Cambridge, MA: MIT Press
- Corne, D., Dhaenens, C., & Jourdan, L. (2012). Synergies between operations research and data mining: The emerging use of multi-objective approaches. *European Journal of Operational Research*, 221(1), 469–479.
- E. Y. Andieva, A. A. Kapelyuhovskaya (2017) New approaches to digital transformation of petrochemical production; *AIP Conference Proceedings*, 20–76.
- Laney, D. Beyer, M. (2012). The Importance of Big Data: A Definition. Retrieved from <https://www.gartner.com/doc/2057415>
- Lee, J. & Kao, H. A. (2014). Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment. *McKinsey Global Institute*. (2011). Big Data: The next frontier for innovation, competition, and productivity. Retrieved from http://bigdatawg.nist.gov/MGI_big_data_full_report.pdf
- Michael Schuldenfrei (2019) Big Data Challenges of Industry 4.0, Retrieved from <https://www.datanami.com/2019/04/25/big-data-challenges-of-industry-4-0/>
- Milan Kubinaa, Michal Varmusa, Irena Kubinova; (2015) Use of big data for competitive advantage of company; 4th World Conference on Business, Economics and Management, WCBEM; *Procedia Economics and Finance*, 26(1), 561–565.
- Pham, D. T. & Afify, A. A. (2005). Machine-learning techniques and their applications in manufacturing. Proceedings of the Institution of Mechanical Engineers. Part B: *Journal of Engineering Manufacture*, 219(1), 395–412.

- Reinsel, D. & Gantz, J. (2011). Extracting value from chaos. Retrieved from <http://www.emc.com/collateral/analyst-reports/idcextracting-value-from-chaos-ar.pdf>
- Shi-Wah Lin, (2017) IIoT for Smart Manufacturing part 3 – A New Digitalization Architecture, October 2017, 16,
- Thorsten W., Weimer, D., Irgens, C., & Thoben, K. D. (2016). Machine learning in manufacturing: advantages, challenges, and applications. *Production & Manufacturing Research*, 4(1), 23–45.