International Journal of Scientific and Technical Advancements
ISSN: 2454-1532

Improving business productivity through data analytics and machine learning

Ab. Qayoom Sofi [1], Ovass Shafi Zargar [2], Shabir Ahmad Najar [3], Bilal Maqbool Beigh [4], Irshad Ahmad Mir [5], Hashair Nabi [6]

[1] Dept of Computer Applications, Govt. College for Women, M.A. Road,Srinagar, J&K, INDIA-190010
[2] Department of Computer Applications, Government Amar singh College, Srinagar, J&K, INDIA-190010
[3] Dept of Computer Applications, Govt. College for Women, M.A. Road,Srinagar, J&K, INDIA-190010
[4] Dept of Computer Applications, Govt. College for Women, M.A. Road,Srinagar, J&K, INDIA-190010
[5] Dept of Computer Applications, Govt. College for Women, M.A. Road,Srinagar, J&K, INDIA-190010
[6] Dept of Computer Applications, Govt. College for Women, M.A. Road,Srinagar, J&K, INDIA-190010

***Abstract*—**With huge investment in modern Information Technologies, every business organization continuously produces a vast and varied amount of data for use of its different levels of its users/managers, there is an increasing demand/need to manage this big data of a business organization for improving the effectiveness and efficiency of operations of that organization. Hence, there is a need to manage data of business organization to improve its productivity and quality, which can be attained by developing Automatic Systems. The main idea is based on the principle or belief that automatic systems are intelligent systems which learn from the (big) data itself requiring least or no human intervention.

This paper aims to discuss big data analytics and machine learning (Automatic) systems, identify various issues and challenges and the ways in which these can be addressed.

Keywords— Productivity; Big Data; Data Analytics; Data Value Chain; Artificial Intelligence; Machine Learning; Cloud Computing; Automatic Systems.

1. BUSINESS PRODUCITIVITY

According to *Industry Week's* 27th annual survey of CEO's in 1998, productivity is the most important performance indicators for CEO's in their strategic decision-making. Productivity is the quantity of output produced by a system over a period of time to the quantity of resources consumed to produce that output over the same period of time.

For long-term survival of any business, it is essential to manage productivity. Business productivity mainly depends on the ability of the managers to take fast and correct decisions. However, there are several obstacles due to lack of skills or time by the managers to analyze productivity to take necessary corrective actions in time. Such problems can be solved by developing Automatic Systems.

Among the highly complex interacting digital systems of the world, machines and devices must become so intelligent that they must automatically behave in a particular context or react to a certain situation.

In this age of science and information technology, to gain competitive edge and to avoid competitive disadvantage, the businesses must make use of the automatic systems for strategic functioning. Technologies such as automatic systems can help analyze and diagnose the problems as well as recommend solutions. This paper describes how the automatic systems can be successfully used by managers for productivity management/improvement.

2. Research Design:-

To improve the productivity of a business organization by big data analytics employing Automatic systems.

2.1 Big Data and Automatic Systems: An Overview

The huge data that is accumulated continuously from diverse sources and must be analyzed/processed in real time for taking precise (decision) action/response within acceptable time is called **Big Data**.

The big data has following characteristics features. For a problem to be classified as Big data problem, either all or any of the following features need to be satisfied

1. **Volume:** means the size of data which is too big to be handled by the current systems/algorithms.

2. **Velocity:** indicates the rate at which the data is generated/streamed which is faster than the speed at which it can be handled by the traditional algorithms.

3. **Veracity:** implies that the quality of the data is more important than the quantity i.e. Big Data does not necessarily mean high quality data and comes with a lot of overhead data that needs to be filtered to select the quality data required.

Ab. Qayoom Sofi, Ovass Shafi Zargar, Shabir Ahmad Najar, Bilal Maqbool Beigh, Irshad Ahmad Mir, Hashair Nabi, "Improving business productivity through data analytics and machine learning," *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 1, pp. 9-14, 2018.

4. **Variety:** means that data is of different types and may range from structured data that can be organised (relations), semi-structured (XML) to un-structured (video, audio and text etc) that cannot be organised.
5. **Value:** means ability to use the data. The data in its original form may not be usable and needs to be analysed and transformed into a form that is of high value. For example, web site logs cannot be used in its initial form to obtain business value and needs to be analyzed to predict customer behaviour.

6. **Variability:** Inconsistency in the big data velocity leads to variation in flow rate of data, which is referred to as variability.

2.2 Sources of Big Data:

Advancement in technologies and digitization of content by the industries has led to the generation of lot of data from different sectors. These sources of data include the Internet, social media etc.

2.3 Big Data Processing:

The process of collection, storage, transportation and exploitation of big data is called Big Data Processing. The main goal of big data processing is to exploit the big data through data analytics for most appropriate and precise action. In other words, it is the process of transforming the raw material (data) available into the valuable actions and is therefore also called **Data Value Chain**, which involves data generation, data collection, data transmission, data pre-processing, data storage, data analysis and decision making.

2.4 Opportunities and Challenges of Big Data:

The main benefit of big data is that with the availability of large number of samples, we can get anything required. However, Big data offers several important challenges/issues, as mentioned below

1. **Volume/Size:** It is impossible to know the exact size of the data as there is continuous increase in the data which means a huge storage capacity is required and it is hard to collect and store.

2. **Velocity:** Since the data is continuously generated at a high speed, a very high computational power for processing the data is necessary.

3. **Veracity:** First, the data source must be trusted. Further, the entire data is not crucial. While only the portion of data that is actually required at a given time is valuable data subset of the original dataset. Further, Data at different places may have different label quality and may have significant level noise- **Inconsistency problem** which can pose management, manipulation, cleansing/classification and transformation challenge/difficulty to the machine learning.

4. **Variety:** The data is generated from multiple sources and is therefore in several forms.

5. **Distributed Nature:** Big data is distributed i.e. the entire data is neither owned by a single owner nor centrally located rather different owners store different parts of the data. Some sources are crucial others are less important. Different owners might warrant different access rights. Further, it is challenging to transport data due to its enormous size. We must aim to leverage data sources without accessing the whole data, and the data can be exploited without need to transport it.

6. **Privacy and Security:** Users now use the cloud computing services offered by the Internet Companies to store their massive personal data and daily activities publicly either due to lack of storage or due to low processing speeds. However, this raises privacy, security and societal issues, which challenges the centralized collection of private data and there is need for **extreme data distribution and data ownership.**

7. It is very difficult to construct one model – "a general model" which can be adapted to other demands given the fact that the computational and storage loads of big data.

Machine learning/Deep learning is the field of artificial intelligence which is most suitable to exploit the opportunities offered by the big data and to overcome challenges posed by it. It is one of the core techniques for data analytics. Machine learning can be used to develop such systems that can process the big data and automatically take an appropriate action in a particular context. An Automatic System may be defined as an application of Artificial Intelligence that can use big data for real time analysis leading to a quick responsive action needed.

The primary goal of automatic systems research is to automatically analyze the big data available in real time to take quick precise action.

2.5 Characteristics of Automatic Systems
An Automatic system has the following characteristics

Ab. Qayoom Sofi, Ovass Shafi Zargar, Shabir Ahmad Najar, Bilal Maqbool Beigh, Irshad Ahmad Mir, Hashair Nabi, "Improving business productivity through data analytics and machine learning," *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 1, pp. 9-14, 2018.

1. It possesses vast data to the minute details from which it quickly learns good feature representations for a given task/object.

2. It optimizes a large array of parameters for its working infeasible and very time consuming for humans to solve difficult problems in a field, as good as if not better than the human experts.

2.6 Benefits of Automatic Systems:

The benefits of the automatic systems are as follows

1. Automatic systems work faster than humans, which means less number of workers are needed. This reduces the cost of the product and increases the output and the productivity.
2. Automatic systems make data available to the people easily, who query the system for some specific advice.
3. Automatic systems are also useful in interdisciplinary domains where multiple automatic systems are required.

2.7 Limitations of Automatic Systems:

The limitations of the Automatic systems are as follows:

1. **Large Data and Storage requirements:** Automatic systems based on machine/deep learning require a lot of training data, which may be cumbersome to collect. The big data also needs big storage.

2. **Large Processing requirements:** Since automatic systems have to analyze and process huge data in real time, a huge processing capability will be required. The major problem lies in data processing and maintenance in addition to the extraction of the data required.

3. **Cost Requirements:** Development of Automatic systems is expensive and time consuming.

4. **Limited Nature:** It may not be possible to automate big data processing using machine learning algorithms in every case.

3. Applications of Automatic Systems:

Automatic systems can be used in a wide variety of areas with initial vast applications mainly in Telecommunication, Retail and Finance. However, owing to its popularity and being successful to facilitate the decision-making process, it is applied in a wide variety of applications as discussed below.

3.1. Healthcare: Aanalysts gather patient information from multiple sources and integrate to develop a reliable treatment system that can reduce the cost and at the same time, improve the quality of treatment.

Obtaining information from external sources such as social media helps in early detection of epidemics and precautionary efforts. After the earthquake in Haiti in January 2010, analysis of tweets helped to track the spread of Cholera in the region (Raghupathi and Raghupathi, 2014).

Artemis, a system developed by Blount et al. (2010), monitors and analyzes the physiological data from sensors in the intensive care units to detect the onset of medical complications, especially, in the case of neo-natal care.

3.2. Telecommunication To improve the customer service, Mobile Service Providers (MSPs) analyze demographic data (gender, age, marital status, and language preferences), customer preferences, household structure and usage details (CDR, internet usage, value-added services (VAS)) to model the customer preferences to offer a relevant personalized service/plan to them by targeted marketing through call or text messages and record, analyze the responses, which improves the services, reduces churn, and increases the revenue of MSPs. Ufone, a Pakistan-based MSP, reduced the churn rate by precisely marketing the customized offers to their customers (Utsler, 2013).

To overcome the limitations of traditional **fraud management system**, MSPs analyze real-time data to minimize the losses due to fraud. Mobileum Inc., one of the leading telecom analytics solution providers, is working towards providing a real-time fraud detection system using predictive analytics and machine learning (Ray, 2015).

3.3. Network Management: MSPs monitor the network speed and manage the entire network to resolve the network problems within few minutes and to improve the customer experience and the quality of service. With wide spread of Smartphones, MSPs analyze real-time location and behavioral data and offer location-based services/context-based services to the customers on demand/request.

3.4. Retail Business firms analyze the data collected from CCTV cameras to observe the age group, gender and purchasing patterns of the customers during weekdays and weekends which helps to decide on the placement of objects and decide on the prices. Accordingly, firms group the items using a well-known data mining technique called Market Basket Analysis (proposed by (Agrawal and Srikant, 1994)), so that a customer buying bread and milk might purchase butter as well. E-commerce firms also use market basket analysis and recommender systems to segment and target the customers by collecting the click stream data, observing behavior and recommending products in the real time.

Ab. Qayoom Sofi, Ovass Shafi Zargar, Shabir Ahmad Najar, Bilal Maqbool Beigh, Irshad Ahmad Mir, Hashair Nabi, "Improving business productivity through data analytics and machine learning," *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 1, pp. 9-14, 2018.

Analytics help the retail companies to manage their inventory. For example, Stage stores, one of the brand names of Stage Stores Inc. which operates in more than 40 American states, used analytics to forecast the order for different sizes of garments for different geographical regions.

3.5. Law Enforcement:
Law enforcement officials analyze past data i.e., type of crime, place and time; social media data; drone and smartphone tracking and try to predict the next crime location. For example, police uses an app called RTM Dx to measure the spatial correlation between the location of crime and features of the environment to prevent crime (Mor, 2014).

Facial analytics is a new technology that examines images of people without violating their privacy, to check child pornography to save the time of manual examination. Various integrated technologies like Artemis and PhotoDNA identify child pornography by comparing files and image hashes with existing files to identify the subject as adult or child.

3.6. Marketing: analytics analyze the consumer behavior, their purchasing patterns and marketing trends, which helps the organizations to analyze their marketing performance to modify/improve the marketing strategies like the positioning of advertisements in a webpage, implementation of dynamic pricing and offering personalized products (Soares, 2012a).

3.7. New Product Development Proper analysis and planning during the development stage minimize the huge risk associated with new product development, increase the customer lifetime value and promote brand engagement if enterprises understand the customer's requirement for a new product, gather ideas for new product and understand the added feature included in a competitor's product by integrating both external sources, i.e., twitter and Face book page and internal data sources, i.e., customer relationship management (CRM) systems. Ribbon UI in Microsoft 2007 was created by analyzing the customer data from previous releases of the product to identify the commonly used features and making intelligent decisions (Fisher et al., 2012).

3.8. Banking Analyze demographic details, behavioral data and financial employment of the customers to predict their investment worthiness to target specific customer segments.

3.9. Energy and Utilities Smart meters measure the consumption of water, gas and electricity at regular intervals and generate a huge amount of data which is analyzed in real time to reveal energy consumption pattern, instances of electricity thefts and price fluctuations.

3.10. Insurance companies analyze updated profiles of changes in wealth, customer risk, home asset value and other data inputs to tailor personalized insurance plan for each customer.

Insurance companies now analyze driver risk and policy. They collect driving data of customers such as miles driven, routes driven, time of day, and braking abruptness using sensors in their cars and compare individual driving pattern and driver risk with the statistical information available such as peak hours of drivers on the road to develop a personalized insurance plan.

3.11. Education: It is used to assess the academic performance and learning patterns of the students in real time for immediate feedback and can therefore help teachers to assess and improve/modify their teaching pedagogy according to the need. Further, patterns about dropouts, average, below average and above average students can be predicted.

3.12. Other sectors: Realizing the significance of the competitive advantage, every organization is now enhancing the analytics skills and therefore it is now applied in almost all other sectors like construction, material and other sciences.

4. Conclusion

With huge IT infrastructure already in place, a huge amount of data is generated continuously at an ultra high speed from variety of sources that must be processed at the right time/real time for taking right timely action. If this big data of a business organization is automatically analyzed and processed, to form an automatic system for that organization that may be used by the different level managers of the organization for improving their decision making process about their core activities and hence lead to the overall development in productivity of the organization, by enhancing the performance and quality of that system. With massive personal data stored using the services of cloud computing, the machine learning algorithms can be used to target users to offer/advertise goods and services as per their likings and disliking based on the analytics of the data provided, which can be automated for faster and better results.

This paper is an endeavor to simplify and describe big data management for effective business and discuss how automatic system applications can help to improve business productivity. Examples of existing applications/prototypes illustrate the possibilities and the strengths of the automatic systems to improve productivity in the business world.

5. References

1. Zhi-Hua Zhou, Nitesh V. Chawla, Yaochu Jin, and Graham J. Williams, Big Data Opportunities and Challenges: Discussions from Data Analytics Perspectives.
2. Abhay Kumar Bhadani, Dhanya Jothimani, Indian Institute of Technology Delhi, India, Big Data: Challenges, Opportunities, and Realities.
3. Mrs.J.Sukanya, S. Vijaya Kumar, Applications of Big Data Analytics and Machine Learning Techniques in

Ab. Qayoom Sofi, Ovass Shafi Zargar, Shabir Ahmad Najar, Bilal Maqbool Beigh, Irshad Ahmad Mir, Hashair Nabi, "Improving business productivity through data analytics and machine learning," *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 1, pp. 9-14, 2018.

ISSN: 2454-1532

Health Care Sectors, International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 6 Issue 7 July 2017, Page No. 21963-21967

4. Hirak Kashyap, Hasin Afzal Ahmed, Nazrul Hoque, Swarup Roy, and Dhruba Kumar Bhattacharyya, Big Data Analytics in Bioinformatics: A Machine Learning Perspective JOURNAL OF LATEX CLASS FILES, VOL. 13, NO. 9, SEPTEMBER 2014.
5. Bill Haffey, Predictive Analytics and Machine Learning: An Overview, April 27, 2012,
6. GOOGLE ANALYTICS: How Analytics and Machine Learning Help Organizations Reap Competitive Advantage.

Ab. Qayoom Sofi, Ovass Shafi Zargar, Shabir Ahmad Najar, Bilal Maqbool Beigh, Irshad Ahmad Mir, Hashair Nabi, "Improving business productivity through data analytics and machine learning," *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 1, pp. 9-14, 2018.

ISSN: 2454-1532

Ab. Qayoom Sofi, Ovass Shafi Zargar, Shabir Ahmad Najar, Bilal Maqbool Beigh, Irshad Ahmad Mir, Hashair Nabi, "Improving business productivity through data analytics and machine learning," *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 1, pp. 9-14, 2018.