BIG DATA MINING IN SKILLS 4.0 FOR AMELIORATING I4.0

Dewan Md. Farid[1,2], Aicha Sekhari Seklouli[1], Ouzrout Yacine[1]

[1]*Decision and Information Systems for Production systems (DISP) Laboratory, IUT Lumiere – University Lyon 2, France*
[2]*Department of Computer Science & Engineering, United International University, Bangladesh*

Abstract

An accretion in Skills 4.0 is one of the major challenges for implementing Industry 4.0 in the developing countries. Industry 4.0 is the fourth industrial revolution that's also known as I4.0/I4. Industry 4.0 is an eclectic mix of following technologies: big data analysis, robotics, simulation, system integration, internet of things (IoT), cybersecurity, cloud computing, 3D printing, and augmented reality. Basically, Industry 4.0 transforms industry into intelligent industry (iIndustry) by connecting Industrial Internet of Things (IIoT), which is the network of machines for sharing information and making decisions without any human involvement to produce high-quality products at low costs and changes the traditional production relations among suppliers, producers, and customers. Industry 4.0 uses advance machine learning and data mining techniques to predict future for optimizing the manufacturing process. Skills 4.0 is a skills model to build up Industry 4.0 that what skills we need to know for Industry 4.0. In this paper, we mainly focus on big data analytics in Skills 4.0, e.g. big data integration, processing, and modelling. We have discussed several important challenges and issues for big data mining in Industry 4.0. We lucid the skills that needed to extract knowledge/informative information from big data for decision making to predict the future of any anomalies is supply chain management (SCM) process. The objective of this study is to find the skills for big data analysis in Industry 4.0, and design and develop learning resources and materials to transfer the knowledge from Europe to Asia.

Keywords: Big Data, Industry 4.0, Knowledge Mining, Machine Learning, Skills 4.0.

1 INTRODUCTION

Industry 4.0, simply I4.0 or I4, refers to the *"Fourth Industrial Revolution"* that's the new digital industrial technology for transforming industries into smart/intelligent industries *(iIndustry)* by connecting machines with intelligent robots [1, 2]. It engenders and analysis data across the machines to produce high-quality products at low costs, and amalgamates nine technologies to transform industrial production, which includes: (1) Big Data Analytics, (2) Autonomous Robots/ Robotics, (3) Simulation, (4) Horizontal & Vertical System Integration, (5) Industrial Internet of Things (IIoT), (6) Cybersecurity, (7) Cloud Computing, (8) Additive Manufacturing (such as 3-D printing), and (9) Augmented Reality [3]. Industry 4.0 uses Decision Support Systems (DSS) incorporating with knowledge mining techniques to know what actions need to taken in future that help manufacturers to optimise their operations quickly. The fourth revolution ameliorates the industries with intelligent computing fuelled by data with Machine Learning (ML) and Data Mining (DM). Industry 4.0 connects machines with Internet of Things (IoT) devices to make decisions without any human/experts involvement. Therefore, Industry 4.0 is basically the network of machines that are digitally connected with each others and share information, which makes the industries more productive, efficient, and less wasteful [4].

The Industrial Revolution (Industry 1.0) started in 1784 (the first weaving loom was introduced), which introduced mechanical production of the goods including weapons, tools, food, clothing and housing to the world using water and steam power. In Industry 1.0, hands are replaced by water and steam powered machines, and firms are transformed to factories for increasing production quality, efficiency and scale. In 1870, Industry 2.0 get going by running machines with electrical energy as electricity is considered primary source of power. So, the electrical machines take the place of the water and steam based machines in Second Industrial Revolution for mass production of goods. Also, the concept of assembly line and several production management techniques were introduced in this era. In 1969, Industry 3.0 (Information Technology (IT) & Automation) introduced computers and electronics in industries. The Programmable Logic Controller (PLC) and computers are used to automate an entire production process without any human assistance for mass production of goods with quality; for example using programmable robots in industries. In 2011, Industry 4.0 is introduced from a project titled “High-Tech Strategy 2020” of the German government [5], which basically not looking at a single computer/machine, but the entire network of them; communicating each other and sharing information to make decisions without any human/user involvement. In Industry 4.0, the machines/IoT devices are engendering a huge amount of data and we need advance Data Mining tools & techniques to analysis the data, so that the system can predict the future.

In 2020, Bu ̈chi et al. [6] investigated and analysed the relationship between the concepts of openness and performance in Industry 4.0. The level of openness to the pillars of Industry 4.0 technologies is evaluated based on the number of technologies used; and the number of value chain stages involved. On the contrary, performance of the Industry 4.0 is estimated by the extent of opportunities business obtain. This study analysed and developed regression models based on the survey data of 231 local manufacturing industries that developed the Industry 4.0 concept in Piedmont (northern Italy); the dataset was collected in year the 2018. The results explicated the opportunities of openness toward Industry 4.0 to achieve higher possibilities in terms of increased production capacity, flexibility, speed, decreased errors and costs, and ameliorated product quality and also ability to encounter customer requirements. Knowledge extraction for decision-making from the previously generated information is one of the major issues to design and develop effective production systems (PSs) in Industry 4.0. In this paper, we mainly focus on big data analytics in Skills 4.0, e.g. big data integration, processing, and modelling. We have discussed several important challenges and issues for big data mining in Industry 4.0. We lucid the skills that needed to extract knowledge/informative information from big data for decision making to predict the future of any anomalies is supply chain management (SCM) process. The objective of this study is to find the skills for big data analysis in Industry 4.0, and design and develop learning resources and materials to transfer the knowledge from Europe to Asia.

The paper is organized as follows. Section 2 introduces Industry 4.0 technologies and architectures. Section 3 discusses teaching and learning strategies in Europe. Section 4 devotes the challenges and issues for Big Data mining in Industry 4.0. Finally, Section 5 collects some conclusions and opens future perspectives.

2 INDUSTRY 4.0

In this digital era of Internet of Everything (IoE), Industry 4.0 amalgamates nine technologies to transform the industrial production for changing traditional production relationships between suppliers, producers and customers [7].

2.1 Industry 4.0 Technologies

In this subsection, we discuss each of the nine technology that form the building blocks of Industry 4.0.

2.1.1 Big Data Analytics

Big Data analytics/mining is the process of extracting knowledge and uncovering hidden patterns from the extremely large-volume of data sets to understand the current and future trends [8]. Big Data mining is one of the major technologies in Industry 4.0 to lucid market trends, customer preferences and other information that’s useful to businesses. Big Data analytics support real-time decision making for building smart factories or *iIndustries* in Industry 4.0 [9]. Big Data refers to 3 V’s: (1) Volume, (2) Variety and (3) Velocity [10]. In Big Data, the volume of data is extremely large that we can’t store all the data into a single computer/ memory; so, mining Big Data is a challenging task. We need parallel and distributed computing for analysing the Big Data. Another issue, the data are multivariate in Big Data; having different types of features in data including numbers, characters, strings, etc. The real-life Big Data is high-dimensional too, and most of the time there is a strong correlations amount the input features [11]. In this digital era of Internet of Things (IoT), we are generating about 2.5 quintillion bytes of data each day, and is expected to reach 44 zettabytes of data by this year of 2020. Therefore, Big Data is streaming data, and the characteristics of data can be changed over the time, for example concept drifting in data streaming environment [12]. We may add few more V’s with Volume, Variety, and Velocity like: Vision - having a purpose/ plan, Verification - ensuring that the data conforms to a set of specifications, and Validation - checking that its purpose is fulfilled [13]. These days, machine learning (ML) is one of the major tools and techniques for mining Big Data. But, the traditional ML classifiers are designed and developed for mining relatively small amount of data [11]. So, we need more scalable and robust ML algorithms for dealing with Big Data like RainForest and BOAT (Bootstrapped Optimistic Algorithm for Tree construction) [30].

2.1.2 Autonomous Robots/ Robotics

Industry 4.0 Robotics plays a principal role in modern manufacturing industry, which refers to industrial manufacturing processes where all equipment and devices are connected with intelligent robots/computers to ameliorate productivity, engender high quality products and also reduce the costs of productions [15]. Programmable Logic Controllers (PLCs) were introduced in Industry 3.0 that's the starting of industrial automation. Intelligent robots basically replacing humans in *iIndustry*, in some cases robots are performing tasks (e.g., packaging & labelling, painting, assembly & disassembly, product inspection & test- ing, mechanical cutting, grinding, and polishing, etc.) without any human assistance; even robots fixing the problems of other robots nowadays. Generally. Industrial robots are classified into six clusters based on mechanical configuration: (1) articulated robots, (2) cartesian robots, (3) selective compliance assembly robot arm (SCARA) robots, (4) delta robots, (5) polar robots, and (6) cylindrical robots. Under other conditions, industrial robots also can be classified based on motion control, power supply control and physical characteristics. In Industry 4.0, Human-Robot Interaction (HRI) is become one of the major research topics to understanding, designing, and evaluating robotic systems that require communication between robots and humans. HRI received considerable attention in the intelligent computational research community and also experts from Industry 4.0 recently. Because of, it enables flexible production in *iIndustry* to rapidly changeover production lines to shorten lead times.

2.1.3 Simulation

Industry 4.0 Simulation helps smart factories in several ways by designing products, time management, increasing productivity and efficiency in manufacturing, and also in revenue generation in this new industrial age [39, 40]. The role of simulation in Industry 4.0, and the simulation optimisation in the era of Industrial 4.0 open a new lines of research to the practitioners and researchers. Simulation imitates the process of a system to show the eventually real effects of system operation over time. A system should be tested very carefully before the application of a new paradigm. Several types of simulation including discrete event and 3D motion simulation can be performed in various cases to improve the product or process planning. E.g. simulation can be adapted in product development, test & optimisation, production process development and optimisation. In the era of Industry 4.0, simulation can be evaluated as a supportive tool to follow the reflections gathered from various parameter changes and enables the visualisation in decision making. A computer simulation helps to create model a real-life situation on a computer so that it can be studied to see how the system works. By changing parameters we can understand the behaviour of the system. Simulation handles the complexity of industrial systems and provides a strong real-life decision making methodology by incorporating information.

2.1.4 Horizontal and Vertical System Integration

Horizontal and vertical system integrations are blueprints for building *iIndustry* in Industry 4.0 that used by businesses in the same industry or production process. Horizontal integrations help *iIndustry* to expand in size, diversify product offerings, reduce competition, and expand into new markets. It's a business strategy where a company takes over another that operates at the same level of the value chain in an industry. On the other hand, vertical integrations involve the acquisition of business operations within the same production vertical. It can help boosting profit and allow smart factories more immediate access to con- sumers.

2.1.5 Industrial Internet of Things (IIoT)

The Industrial Internet of Things (IIoT) refers to as Internet 4.0 (Industrial Internet) or the Internet of Everything in Industry 4.0, which become one of the most important research area for building *iIndustry* nowadays. IIoT is a system that amalgamate computing devices, sensors, instruments, and other devices in a network to transform business operational processes, including manufacturing and energy management into knowledge for accelerating productivity in industries [18]. IIoT engenders large volume of data, and use cloud computing and Big Data analysis tools and techniques to optimise the process controls and re- duced unplanned downtime. The IIoT applies several technologies such as cybersecurity, cloud computing, edge computing, mobile technologies, machine-to-machine (M2M), 3D printing, advanced robotics, big data, internet of things (IoT), RFID technology, and cognitive computing. For example, with Big Data in IIoT systems, large-volume of streaming data generated by IIoT system can be stored in cloud storage systems and mining the data in cloud to extract knowledge for building decision support system (DSS) in order to boosting up the productivity and efficiency, and reducing the operational costs.

2.1.6 Cybersecurity

Cybersecurity in Industry 4.0 is basically computer/information technology (IT) security, which is a system that protects the computer systems and networks including hardware, software, and data of industries from any kind of inside and outside attacks/intrusions [21, 22, 23, 24]. Industry 4.0 connects Internet of Things (IoT) devices with industrial machines and engenders machine data, so that machines can communicate with each other. In the last decade, the term cybersecurity became very popular issue as information theft is increased a lot, and also it opened a new direction of research due to the growth of Internet of Things (IoT) devices in industries. Cybersecurity can affect business performance with aberration the anomaly of networked manufacturing machines on critical industrial equipments. In 2020, Corallo et al. [25] proposed a classification model of critical assets to be protected against cyber-attacks in the context of Industry 4.0 and potential adverse impacts on business performance due to breaches of cybersecurity. In this study, authors find the relationship between assets and impacts in terms of loss of confidentiality, integrity and availability of data associated with networked manufacturing machines. This work basically support the technical and nontechnical management staff in smart factories for decision making to face the cybersecurity attacks af- fecting the Industry 4.0. Another work also published in 2020 by Lass and Gronau [26] that presented a concept to address safety of manufacturing enterprises as a cyber-physical production systems (CPPS). It's very prime and challenging task to implement cyber-physical systems (CPS) in Industry 4.0. CPS is used to control and monitor both hardware and software components in industrial control system (ICS) in Industry 4.0 to manage industrial process control.

2.1.7 Cloud Computing

Cloud computing is the process of storing and accessing data as well programs online over the Internet (Web); instead of personal computer's (PC's) hard drive (local storage). Cloud is basically metaphor for the Internet. Local storage computing means when we store data on or run programs from the PC hard drive or in a local network, e.g., storing data on a home or of fice network is consider as storing data in local storage device. Any other way, cloud computing is the use of various services over the internet, e.g., Google Drive, Apple iCloud, Amazon Cloud Drive, Dropbox, etc. Cloud computing is generally divided into three services: (1) Infrastructure as a Service (IaaS), (2) Platform as a Service (PaaS) and (3) Software as a Service (SaaS). With cloud computing, we can use hardware and software, access files and applications over the Internet. Generally, cloud computing is bifurcated into public and private clouds. Public cloud is the standard cloud computing framework, which consists of files, applications, storage and services available via Internet, e.g., Gmail. Differently, private cloud is composed of files, applications, storage and services, which are stored and protected within a corporate firewall, e.g., a company using Microsoft Exchange and it can only be accessed by an authorised user through a secure virtual private network (VPN) connection. Cloud computing has infinite storage and computational capabilities, which helps business in all industries adapt to today's rapidly changing technology [27].

2.1.8 Additive Manufacturing

Additive manufacturing is the process of joining materials to make a three-dimensional object, usually layer upon layer, as opposed to subtractive manufacturing methodologies, e.g., 3D printing. It adds material to create an object using computer-aided-design (CAD) software or 3D object scanners to build up a component in layers by depositing material. Additive manufacturing is using in many industries like: aerospace, medical, transportation, energy, consumer products, etc. Additive manufacturing is really needed in *iIndustry*, because it significantly reduce material waste, reduce the amount of production steps, and reduce the amount of distinct parts needed for an assembly [28]. It basically boosts up innovation in smart factories for creating new kinds of products and prototypes of any complicated parts directly from three-dimensional (3D) computer-aided design. It provides a disruptive transformation in how products are designed and manufactured.

2.1.9 Augmented Reality

Augmented reality (AR) is one of the most promising technologies in the context of Industry 4.0 [29]. It's a technology that superimposes a computer-generated image (CGI) of a real-world environment on a user's view, thus providing computer-generated composite view, sometimes across multiple sensory modalities, including visual, auditory, haptic, somatosensory and olfactory [30]. It adds digital elements to a live view often by using the camera on a smartphone. Augmented reality overlays computer-generated video onto camera-captured video in such a way that the CGI objects appear to have an absolute location in the real world.

2.2 Industry 4.0 Architecture

In this sub-section, we discuss two reference model architectures for Industry 4.0.

2.2.1 *Reference Architectural Model Industry 4.0 (RAMI 4.0)*

The RAMI 4.0 is a three-dimensional reference model that amalgamates all elements with IoT devices in layers and life cycle model [31]. In RAMI 4.0, three axes are describe as: Axis 1 - Hierarchy: The Factory (the hierarchical levels of a manufacturing system networked via cloud), Axis 2 - Product Life Cycle (the life cycle of systems and products), and Axis 3 - Architecture (the IT structure of an Industry 4.0 component) [32]. The hierarchical system in the Industry 4.0 follows the IEC 62264 and IEC 61512 standards, which basically include *Product* and *Internet* with Industry 3.0. The IEC 62264 is an international standard from the International Society of Automation (ANSI/ISA-95) for developing an automated interface between enterprise and control system integration, which consists of the following six parts: (1) object models and attributes of manufacturing operations, (2) object model attributes, (3) activity models of manufacturing operations management, (4) objects models attributes for manufacturing operations management integration, (5) business to manufacturing transactions, and (6) messaging service model. Contrarily, the IEC 61512 standard consists of the following four parts: (1) models and terminology, (2) data structures and guidelines for languages, (3) general and site recipe models and representation and (4) batch production records. The product life cycle in Industry 4.0 is bifurcated into Type and Instance that includes development, production/sales, and service. The architecture in Industry 4.0 has the following six function of layers: (1) Asset Layer (physical things in the real-world), (2) Integration Layer (transition from real to digital world), (3) Communication Layer (access to information), (4) Information Layer (necessary data), (5) Functional Layer (functions of the asset) and (6) Business Layer (organisation and business pro- cesses). RAMI 4.0 maps the key aspects of Industry 4.0. It presents and shows the entire Industry 4.0 solution that integrates IT and automation technology to put them into one single model. RAMI 4.0 model establishes communication between machines and digital word (administration shell). Fig. 1 shows the RAMI 4.0 model.

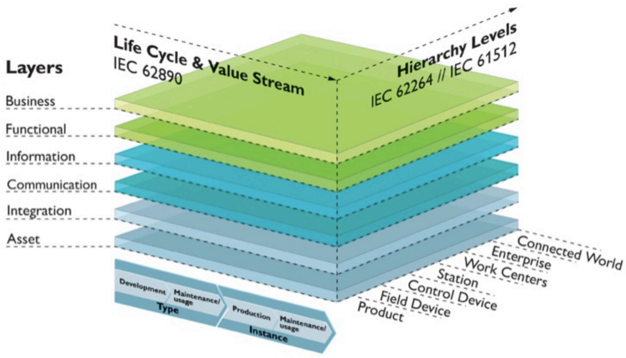

Figure 1. Reference Architectural Model for Industry 4.0 (RAMI 4.0).

2.2.2 *Industrial Internet Reference Architecture (IIRA)*

The Industrial Internet Reference Architecture (IIRA) is a standards-based architectural framework and concepts to design and develop Industrial Internet of Things (IIoT) system in Industry 4.0. IIRA was introduced by Industrial Internet Consortium (IIC) and its version 1.9 published in 2015 [33], which provides information for developing of interoperable IIoT systems, solution and application architectures about wireless communications in *iIndustry* [66, 67]. It identifies and highlights important architectural concerns, concepts and patterns for IIoT systems with four viewpoints: (1) Business, (2) Usage, (3) Functional, and (4) Implementation that adopted from ISO/IEC/IEEE 42010:2011 (http://www.iso- architecture.org/ieee-1471). IIRA provides value for IT managers, and top business management for decision-making. The IIoT is one of the major technologies in Industry 4.0 that is using in all industrial sectors, such as energy, healthcare, manufacturing, public infrastructure, transportation, etc. Implementing IIoT in industries is highly complex and challenging task, because it requires highly diverse system and architecture requirements. IIRA is an open reference architecture that encourages reuse of common and existing system building blocks to reduce project effort, risk, costs and time-to-market. IIRA reduces operational and design cost, and enables plug & play with the IIoT system.

3 TEACHING AND LEARNING

Teaching and learning is the process of gaining new knowledge and skills in a subject or activity that synthesize new understanding from prior learning and new information. It's the process of sharing and transferring knowledge by constructing new ideas and concepts, which helps us to transform knowledge, construct hypothesis and decision making etc. Basically, learning is the process of study and repeat something in order to be able to remember it, e.g. a poem.

3.1 Teaching and Learning Strategies

In this following sub-section, we will discuss the following seven modes of effective teaching and learning strategies that are commonly used in Europe: (1) Lectures, (2) Massive Open Online Courses (MOOCs), (3) Flipped Classroom, (4) Active Learning, (5) Blended Learning, (6) Problem-Based Learning and (7) Work-Based Learning.

3.1.1 Lectures/Courses

The traditional form of teaching in the tertiary-level education often involves lectures being given to a group of students by an academic. A lecture basically an oral presentation to teach students about a particular subject, which is a common teaching method in academia. Lectures are the most impactful, effective and engaging way to promote student learning. Lectures also allow students to ameliorate their interactive and collaborative skills that is difficult to develop in self-study situations. It also helps students to understand the development of the argument/ideas by active participation, and lead to a bigger interest and motivation into the subject/course.

3.1.2 Massive Open Online Courses (MOOCs)

A massive open online course (MOOC) is the process of providing open access (free and open registration) online learning content/course aimed at unlimited interactive participation employing the web, e.g., OpenCourseWare (OCW) by MIT (Massachusetts Institute of Technology). MOOCs provide interactive user forums to build communities for students/learners and academicians using online resources and social networking. Most significantly, MOOCs are very useful for distance education by accessing free educational resources, e.g., lecture notes, videos etc. MOOCs are new pathway in higher education for connecting people to share their learning experiences, because anyone can participate the courses from anywhere in the world, e.g., Coursera (https://www.coursera.org).

3.1.3 Flipped Classroom

The flipped classroom is a pedagogical model, which is type of blended learning where lectures (e.g., video lectures) are introduced/viewed by the students at home before coming to the lecture. The video lectures are one of the important elements in the flipped approach that are developed by the academician and provided to students prior to the lecture. In lecture time, students can discuss and inquire about the lecture content, test their skills by applying knowledge, and do some practical activities. Generally, students watch online lectures, collaborate in online discussions using social media, do some homework and do some activities in the classroom with the guidance of a mentor in a flipped classroom.

3.1.4 Active Learning

Active learning is a learning and teaching method to help students and instructors into effective ways to help everyone engage in meaningful activities based on the concepts about how people learn in a deliberate contrast to passive learning. In active learning, students are actively involved in the learning process and must engage in higher-order thinking, i.e., analysis, synthesis and evaluation. It has two aspects – doing things and thinking about the things students are doing.

3.1.5 Blended Learning

Blended learning also known as hybrid learning, is an approach to education that amalgamates online learning with face-to-face teaching (i.e., traditional classroom teaching), allowing the students to have more control over time, place, path/pace and style of learning. Blended learning amalgamates the strengths of both traditional and online learning methods in order to ameliorate learning experience. Webinars can be a useful tool for blended learning, e.g., Zoom, GoToWebinar, Cisco WebEx, Adobe Connect, Google Hangouts, AnyMeeting etc.

3.1.6 Problem-Based Learning

Problem-based learning (PBL) is a learning by doing teaching method in which complex and authentic real-world problems are solved by the students that helps to develop critical thinking skills, problem-solving abilities, reasoning, communication and self-assessment skills. PBL also provide the opportunities to work in groups (i.e., students work with classmates), finding and evaluating research materials and life-long learning. PBL commonly use in medical education where students learn dealing with real-life medical problems, however this technique can be used in other fields, e.g., education, business, engineering, law, architecture, music and literature etc.

3.1.7 Work-Based Learning

Work-based learning (WBL) is an educational strategy that places students both in the classroom and real-life work experiences (i.e., the workplace) to prepare students for real-world careers. WBL integrates the educational institute curriculum with the workplace to create a different learning paradigm that could be done by doing internships, job shadowing or field trips. WBL helps to students to apply academic and technical skills in terms of professional development.

4 BIG DATA MINING

In this section, we discuss about big data analytics in Skills 4.0 for Industry 4.0.

4.1 Challenges and Issues for Mining Big Data

Machine Learning (ML) is one of the major tools and techniques for mining Big Data. But, the traditional ML classifiers are designed and developed for mining relatively small amount of data. So, we need more scalable and robust ML algorithms for dealing with Big Data like RainForest and BOAT (Bootstrapped Optimistic Algorithm for Tree construction) [14]. In general, real-world Big Data are imbalanced where majority class instances influence over minority class instances, and existing ML algorithms are correctly classify majority class instances and misclassify minority class instances [34]. Also, it’s very hard and time-consuming for labelling unlabelled Big Data in semi-supervised learning, and selecting informative instances (both majority and minority class instances) are really important for boosting up the performance of ML classifiers for mining Big Data [35]. It’s also very necessary to find the similarities amount the instances in Big Data, and clustering Big Data can be very useful for building optimal ML classifiers [36].

4.2 Knowledge Mining from Big Data in Industry 4.0

Big Data with IoT technologies are taking part a vital role in Industry 4.0. In 2020, Sahal et al. [37] explored open-source big data technologies, including big data streaming processing platforms, distributed message queue management systems, big data storage platforms, and streaming SQL engines for Industry 4.0 requirements mapping for predictive maintenance (PM 4.0). Authors found that Big Data analytics reduce operation and maintenance (O&M) costs in PM 4.0. They also identified the functional requirements and operational characteristics for Industry 4.0 as a sample use cases related to the railway transportation and wind turbine energy industries. This study discussed the open-source big data pipeline analysis technologies including data collection, data analysing, data sorting, and data querying. Tiwari et al. [38] reviewed the affect of Big Data analytics research and application in supply chain management between 2010 and 2016. This study surveyed the techniques for data collection, organised and analysed data to present with valuable insights in Industry 4.0 by discussing the role of big data analytics in supply chain management (supply chain analytics). It’s necessary and play an importance role in Industry 4.0 to apply Big Data tools and techniques for dealing with increasing number of data in supply chain management for decision-making. Wang et al. [39] proposed a multi-agent based framework for self-decision making and intelligent negotiation mechanisms to build smart factory by integrating the autonomous agents with Big Data mining based decision support system. This study focuses on two aspects: modelling the smart shop-floor objects such as machines, conveyers, and products as intelligent agents to negotiate mechanism for them to cooperate with each other, and also preventing the deadlocks by improving agents’ decision making and the coordinator’s behaviour. Big data analytics is a principal component in smart manufacturing frameworks. Data analysis and management using Big Data tools and techniques ameliorate manufacturing intelligence that can take timely and accurate decisions.

4.3 Big Data in Skills 4.0

Initially, we need to know about the basic characteristics of Big Data and dimensions of scalability. How can we extract valuable informative information out of Big Data? Need to know about the Big Data systems and programming, e.g., Hadoop, Julia, R programming, Python, Wake, KEEL software etc. Secondly, we need to have the skills to understand the Big Data modelling, Big Data management, designing a Big Data management system. After that we need to focus on Big Data integration and processing, e.g., retrieving Big Data, Big Data integration, processing Big Data, Big Data analytics using Spark etc. Nowadays, dealing Big Data with Machine Learning getting very popular, so we also need to have knowledge about supervised, unsupervised and semi-supervised learning, i.e., classification, regression and clustering algorithms. Finally, knowledge about graph analytics for Big Data, e.g., introduction to graphs, graph Analytics and its techniques, computing platforms for graph analytics etc.

5 CONCLUSIONS

In this paper, we have discussed about Industry 4.0 and its technologies, commonly used teaching and learning strategies applied in tertiary education, and the skills that need for Big Data analytics in Industry 4.0. Big Data mining is one of the major technologies in Industry 4.0, which helps us to extract hidden and informative information for decision making from large volume of business data or industrial data. One of the major contributions of this work is to find the skills, i.e., Big Data in Skills 4.0 that the skills and knowledge we need to have for mining Big Data in Industry 4.0. We have tried to address the challenges for handing Big Data. Implementing and ameliorating Industry 4.0 is a challenging task for developing countries e.g., south Asian countries. Educating graduate students and existing/future industrial works into Industry 4.0 is the real challenge for universities and industries in present time. There should be a strong collaboration between universities and industries to minimize the gap in Skills 4.0.

ACKNOWLEDGEMENTS

We appreciate the support for this research received from the Building Skills 4.0 through University and Enterprise Collaboration (Erasmus+ Shyfte 4.0) project, which aims to develop Skills 4.0 in association with Higher Educational Institutes (HEIs) and industrial Small and Medium-sized Enterprises (SMEs) (http://shyfte.eu).

REFERENCES

[1] E. Oztemel, S. Gursev, “Literature review of industry 4.0 and related technologies,” *Journal of Intelligent Manufacturing*, vol. 31, no. 1, pp. 127–182, 2020.

[2] M. Kerin, D. T. Pham, “A review of emerging industry 4.0 technologies in remanufacturing,” *Journal of Cleaner Production*, vol. 237, pp. 117805, 2019.

[3] B. Chen, J. Wan, L. Shu, P. Li, M. Mukherjee, B. Yin, “Smart factory of industry 4.0: Key technologies, application case, and challenges,” *IEEE Access*, vol. 6, pp. 6505–6519, 2017.

[4] S. Doltsinis, P. Ferreira, M. M. Mabkhot, N. Lohse, “A decision support system for rapid ramp-up of industry 4.0 enabled production systems,” *Computers in Industry*, vol. 116, pp. 103190, 2020.

[5] W. W. H. Kagermann, J. Helbig, W. Wahlster, “High-tech strategy 2020,” Germany (2013).

[6] G. Bu ̈chi, M. Cugno, R. Castagnoli, “Smart factory performance and industry 4.0,” *Technological Forecasting and Social Change*, vol. 150, pp. 119790, 2020.

[7] A. P. T. Pacchini, W. C. Lucato, F. Facchini, G. Mummolo, “The degree of readiness for the implementation of industry 4.0,” *Computers in Industry*, vol. 113, pp. 103125, 2019.

[8] A. L’heureux, K. Grolinger, H. F. Elyamany, M. A. M. Capretz, “Machine learning with big data: Challenges and approaches,” *IEEE Access*, vol. 5, pp. 7776–7797, 2017.

[9] Q. Qi, F. Tao, “Digital twin and big data towards smart manufacturing and industry 4.0: 360-degree comparison,” *IEEE Access*, vol. 6, pp. 3585 – 3593, 2018.

[10] D. M. Farid, M. A. Al-Mamun, B. Manderick, A. Nowe, “An adaptive rule-based classifier for mining big biological data,” *Expert Systems with Applications,* vol. 64, pp. 305–316, 2016.

[11] D. M. Farid, L. Zhang, C. M. Rahman, M. Hossain, R. Strachan, “Hybrid decision tree and naive bayes classifiers for multi-class classification tasks,” *Expert Systems with Applications*, vol. 41, no. 4, pp. 1937–1946, 2014.

[12] D. M. Farid, L. Zhang, A. Hossain, C. M. Rahman, R. Strachan, G. Sexton, K. Dahal, “An adaptive ensemble classifier for mining concept drifting data streams,” *Expert Systems with Applications*, vol. 40, no. 15, pp. 5895–5906, 2013.

[13] T. Siraj, E. Jannat, W. Rasul, M. A. Chowdhury, S. Shatabda, D. M. Farid, “Big data mining in the presence of concept drifting,” *in: international Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, pp. 1–6, 2019.

[14] S. Sabah, S. Z. B. Anwar, S. Afroze, M. A. Azad, S. Shatabda, D. M. Farid, “Big data with decision tree induction,” *in: 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, Island of UKULHAS, Maldives, pp. 1–6, 2019.

[15] R. Goel, P. Gupta, “Robotics and Industry 4.0,” *Springer International Polishing*, Cham, 2020, Ch. A Roadmap to Industry 4.0: Smart Production, Sharp Business and Sustainable Development, pp. 157–169.

[16] M. M. Gunal (Ed.), “Simulation for Industry 4.0 - Past, Present and Future,” *Springer Series in Advanced Manufacturing*, Springer, Cham, 2019.

[17] J. Xu, E. Huang, L. Hsieh, L. H. Lee, Q. S. Jia, C.-H. Chen, “Simulation optimization in the era of industrial 4.0 and the industrial internet,” *Journal of Simulation*, vol. 10, no. 4, pp. 310–320, 2016.

[18] P. Radanliev, D. C. D. Roure, R. Nicolescu, M. Huth, R. M. Montalvo, S. Cannady, P. Burnap, “Future developments in cyber risk assessment for the internet of things,” *Computers in Industry*, vol. 102, pp. 14–22, 2018.

[19] M. Younan, E. H. Houssein, M. Elhoseny, A. A. Ali, “Challenges and recommended technologies for the industrial internet of things: A comprehensive review,” *Measurement*, vol. 151, pp. 107198, 2020.

[20] C. Garrido-Hidalgo, T. Olivares, F. J. Ramirez, L. Roda-Sanchez, “An end-to-end internet of things solution for reverse supply chain management in industry 4.0,” *Computers in Industry*, vol. 112, pp. 103127, 2019.

[21] D. M. Farid, M. Z. Rahman, C. M. Rahman, “Mining complex network data for adaptive intrusion detection,” in: A. Karahoca (Ed.), *Advances in Data Mining Knowledge Discovery and Applications, InTech*, Croatia, pp. 327–348, 2012.

[22] D. M. Farid, M. Z. Rahman, “Anomaly network intrusion detection based on improved self-adaptive Bayesian algorithm,” *Journal of Computers*, vol. 5, no. 1, pp. 23–31, 2010.

[23] D. M. Farid, J. Darmont, M. Z. Rahman, “Attribute weighting with adaptive nbtree for reducing false positives in intrusion detection,” *International Journal of Computer Science and Information Security*, vol. 8, no. 1, pp. 19– 26, 2010.

[24] D. M. Farid, N. Harbi, M. Z. Rahman, “Combining naïve bayes and decision tree for adaptive intrusion detection,” *International Journal of Network Security & Its Applications*, vol. 2, no. 2, pp. 12–25, 2010.

[25] A. Corallo, M. Lazoi, M. Lezzi, “Cybersecurity in the context of industry 4.0: A structured classification of critical assets and business impacts,” *Computers in Industry*, vol. 114, pp. 103165, 2020.

[26] S. Lass, N. Gronau, “A factory operating system for extending existing factories to industry 4.0,” *Computers in Industry*, vol. 115, pp. 103128, 2020.

[27] G. Terrazas, N. Ferry, S. Ratchev, “A cloud-based framework for shop floor big data management and elastic computing analytics,” *Computers in Industry*, vol. 109, pp. 204–214, 2019.

[28] A. Paolini, S. Kollmannsberger, E. Rank, “Additive manufacturing in construction: A review on processes, applications, and digital planning methods,” *Additive Manufacturing*, vol. 30, pp. 100894, 2019.

[29] K. Lavingia, S. Tanwar, “A Roadmap to Industry 4.0: Smart Production, Sharp Business and Sustainable Development,” *Springer International Publishing*, Cham, Ch. Augmented Reality and Industry 4.0, pp. 143–155, 2020.

[30] M. B. Ibáñez, A. U. Portillo, R. Z. Cabada, M. L. Barrón, “Impact of augmented reality technology on academic achievement and motivation of students from public and private mexican schools. a case study in a middle-school geometry course,” *Computers & Education*, vol. 145, pp. 103734, 2020.

[31] Q. Li, Q. Tang, I. Chan, H. Wei, Y. Pu, H. Jiang, J. Li, J. Zhou, “Smart manufacturing standardization: Architectures, reference models and standards framework,” *Computers in Industry*, vol. 101, pp. 91–106, 2018.

[32] R. Sharpe, K. van Lopik, A. Neal, P. Goodall, P. P. Conway, A. A. West, “An industrial evaluation of an industry 4.0 reference architecture demonstrating the need for the inclusion of security and human components,” *Computers in Industry*, vol. 108, pp. 37–44, 2019.

[33] S.-W. Lin, B. Miller, J. Durand, R. Joshi, P. Didier, A. Chigani, R. Toren- beek, D. Duggal, R. Martin, G. Bleakley, et al., “Industrial internet reference architecture,” Tech. rep., *Industrial Internet Consortium (IIC)*, 2015.

[34] D. M. Farid, S. Shatabda, M. Z. Abedin, M. T. Islam, M. I. Hossain, “Mining imbalanced big data with julia, *in: Julia Conference (JuliaCon)*,” University of Maryland Baltimore, Baltimore, MD, USA, 2019.

[35] S. Jahan, S. Shatabda, D. M. Farid, “Active learning for mining big data,” *in: 21st International Conference on Computer and Information Technology (ICCIT)*, United International University, Dhaka, Bangladesh, pp. 1–6, 2018.

[36] S. S. Khan, S. Ahamed, M. Jannat, S. Shatabda, D. M. Farid, “Classification by clustering (cbc): An approach of classifying big data based on similarities,” *in: Proceedings of International Joint Conference on Computational Intelligence*, Springer Singapore, Singapore, pp. 593–605, 2020.

[37] R. Sahal, J. G. Breslin, M. I. Ali, “Big data and stream processing platforms for industry 4.0 requirements mapping for a predictive maintenance use case,” *Journal of Manufacturing Systems*, vol. 54, pp. 138–151, 2020.

[38] S. Tiwari, H. M. Wee, Y. Daryanto, “Big data analytics in supply chain management between 2010 and 2016: Insights to industries,” *Computers & Industrial Engineering*, vol. 115, pp. 319–330, 2018.

[39] S. Wang, J. Wan, D. Zhang, D. Li, C. Zhang, “Towards smart factory for industry 4.0: a self-organized multi-agent system with big data-based feedback and coordination,” *Computers in Industry*, vol. 101, pp. 158–168, 2016.