Industry 4.0: Data science perspective

Mahyuddin K M Nasution¹,∗
¹Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara, Kampus USU, Padang Bulan 20155, Medan, Indonesia
E-mail: *mahyuddin@usu.ac.id

Abstract. Industry 4.0 is a shift from the previous industry shape. Automation is as a result of technology involving electrical energy. Changes occur and lead to paperless and humanless, but these changes require preparation both in the industrial world and in other supporting worlds. Data is a different side of technology. Data involves many different concepts from the industrial world, even though either data of the industrial world involve the same technology. Specifically, data management is different than industry management. This paper reviews integrated management based on data science, a science that studies the behavior of data. As a result, integration requires data structuring steps from Industry 4.0 and structuring data for Industry 4.0.

1. Introduction

Industry 4.0 is an era of the automation of the production of goods and services, which has led to systemic changes in many sectors and aspects of human life. The presence of various conveniences of information technology has led to industrial automation covering all activities of manufacture. Apart from the physical-cyber system (CPS) [1], the internet of things (IoT) is what makes it possible to create smart factories [2]. The use of sensors, robotics, and computers step by step makes paperless and humanless. On the one hand, the use of human labor is decreasing, but on the other hand, it requires skilled workers with technology to fill the fields of work that suddenly appear. The use of technology to control and monitor manufacturing automatically generates data. Not a few of the tools in question require that data as feedback in machine learning to be able to predict which industrial systems are running. However, data cannot be processed automatically by a computer machine. It needs completion stages, starting from data modeling to data management itself. On the Industry 4.0 side, mastery of information technology is at the core of sustaining activities and increasing performance. On the human side, understanding data is part of mastering information technology [3]. This working paper describes Industry 4.0 from a data science standpoint.

2. Basic Concept

Data, from all sides, has driven the emergence of technology, especially information technology, ranging from simple to complex and sophisticated [4]. The technology includes activities related to input, process, and output, recognizing the interaction between the sender and receiver of information. The technology is at the center...
of the Industrial revolution 4.0, among the technologies are CPS [1], Internet of Things, Augmented reality, Digital Twins, Deep Learning, Edge Computing, Artificial Intelligence, Robotic, Machine to Machine Learning, Big Data Analysis, Cloud Computing, Mobile Internet, etc [5]. This technology aims to produce a smart industrial system. However, the industry has been around for a long time, since the steam engine was invented, and has involved various scientific fields. The field of science concerns natural resources and human resources, namely mathematics, physics, inventory, optimization, electronics, planning, maintenance, business, economics, marketing, and others, both theory, and implementation. Several related scientific fields such as economics, business, planning, maintenance, marketing, and other areas of implementation require change, and some of the basic concepts need to be re-state anew.

According to engineering, industrial systems operate following the underlying technology and techniques. Large machines operate mechanically, although controlled electronically, as in robotics in the automobile assembly industry, where the movements mechanically carried out with high precision. All machines require energy, and it is provided by the energy supply machine, while energy regulation is carried out through control sensors to save the needs of each production machine. Following computer estimates and measurement results by Computer Aided Design (CAD) [6], every robot machine, for example, in a factory, has its movement precision determined. Each robot, however, can make errors when maintenance does not match the age of the machine. Moreover, numerically every robot related to software requires adjustments when the machine is an updated version. Likewise, every machine requires learning, so that production is not only increased but more efficient and optimal. It happens when those machines and technologies are supported by data.

Each machine requires many parameters to operate. Parameters that require appropriate data, usually data requires not only modeling but also standardization. All of this is a data science business.

Data Science is the extraction of knowledge from high-volume data, using skills in computing science, statistics and the specialist domain knowledge of experts [7].

3. A review of the method

Industry 4.0 is a collaboration between science, data, information, and technology with the integration of data science and engineering to produce intelligent industrial systems. Thus, for the industry to become automation, industry 4.0 requires instruments that consider many parameters to accumulate information both inside and outside the system industry [9]. Thus, methods accompany every step of industrial automation. Hence, integration is, therefore, a determinant rationale from a data science perspective.

Data not only trigger technology but also releases methods to analyze data based on hypotheses that dynamically develop according to the demands of society or the interests of researchers. The paradigm for the industrial phenomenon 4.0 is the integration of facilities and infrastructure that are highly dependent on IoT, which makes it possible to connect not only between humans and between machines, but also between humans and machines [10], where the empowerment of theory, science, and data is short into the next section. The considerations draw from four papers
Table 1. Energy demand prediction in Indonesia in 2025 for nine industrial sectors.

| Sector                                                | Energy       |
|-------------------------------------------------------|--------------|
| 1. Food, beverage, and tobacco                        | 72,210 gWh   |
| 2. Fertilizer, chemical, and rubber goods             | 60,232 gWh   |
| 3. Cement industry, and non-metal minerals            | 41,732 gWh   |
| 4. Textile, leather goods and footwear                | 36,050 gWh   |
| 5. Transportation, machinery, and other equipment     | 18,491 gWh   |
| 6. Iron and steel base metal                          | 14,431 gWh   |
| 7. Paper and printed goods                            | 10,212 gWh   |
| 8. The wood and forests products                      | 2,732 gWh    |
| 9. Others                                             | 16,695 gWh   |

at the conference and five articles from journals, where all documents indexed in the Scopus database with titles there are the words both “industry 4.0” and “data science”.

4. The result and discussion

Every industry needs energy [11]. The implementation of Industry 4.0 takes place automatically. Electrical energy is the primary energy in Industry 4.0, where the source may be water, oil, coal, wind, geothermal, nuclear, solar, or others. Generally, the energy source drives an electric generator. For example, oil becomes the energy source for the machines that drive electric generators. The fast flow of water also drives electricity generators, as well as other sources of energy, but the sun can be a direct source of energy stored in storage, such as a battery. For example, the projection of energy required for each industrial sector in Indonesia in 2025 is as shown in Table 1.† Most energy demand comes from the food industry sector, namely 26.6% of the total needs. The second position is the manufacturing sector, which requires 22.0%. The third position relates to the mining industry sector, which requires 15.35%. The clothing industry sector ranks fourth at 13%, in addition to the machinery sector, the metal industry sector, the paper industry sector, and the wood processing industry sector, respectively. Each gWh (gigawatt hours) requires a different energy source and depends on the conversion technology. Therefore, a suitable energy source must be arranged in such a way.

Suppose that the energy for a plant comes from various sources in an integrated manner, namely, fuel to drive industrial machinery, hybrid electricity with a gasoline engine, hybrid electricity with a diesel engine, or electricity with an LPG engine [5]. Electricity comes from the energy available in nature, such as water, wind, or the sun. Thus there are evaluation criteria as follows.

(i) Energy supply [12], \(x_1\), is the annual amount of energy that can supply [13].

The \(a_1\) assignment can be considered based on inventory, for example for total inventory cost (TIC), average inventory \(\frac{Q}{2}\), ordering frequency \(\frac{R}{Q}\), storage cost

† Source: "Buku Perencanaan Kebutuhan Energi Sektor Industri (Industrial Sector Energy Demand Planning Book), Biro Perencanaan Kementerian Perindustrian Republik Indonesia (Planning Bureau of the Ministry of Industry of the Republic of Indonesia), 2012.
per unit time \( C_h = H \), and order fee each time, is obtained

\[ TIC = \frac{1}{2} QC_h + \frac{R}{Q} C_0. \]  

(1)

Eq. (1) also involves the reliability of energy storage and the cost of energy supply.

(ii) Energy efficiency [14], \( x_2 \), i.e. energy efficiency of fuel. The \( a_2 \) setting can consider the following formula:

\[ E_e = \frac{u_{eo}}{e_i} \]  

(2)

where \( E_e \) is efficiency, \( u_{eo} \) is the useful energy output, and \( e_i \) is energy input.

(iii) Air pollution [15], \( x_3 \), the contribution of fuel to air pollution. For example, the determinant, by comparison, is \( a_3 \), where the formula for calculating the concentration of CO in the air is

\[ \mu g/m^3 = (ppm \times BM)/(24.5 \times 10^{-3}) \]  

(3)

where 24.5 is the conversion for 1 mole = 24.5 liters \((25^\circ C, 1 \text{ atm})\), \( BM \) is the molecular weight, for CO, \( BM = 28 \), \( 10^{-3} \) is the conversion from millie liter to liter.

(iv) Noise pollution [16], \( x_4 \), refers to the noise generated during plant operation. For example, the determining factor for \( a_4 \) involves the value of \( leq \) (10 minutes) using the formula \( leq(10 \text{ minutes}) = 10 \log_{10}(120/((\sum_{i=1}^{120} t_i \times 10^{0.1L_i}))) \), where \( leq \) is the equivalent continuous sound pressure level \((\text{dB})\), \( t \) measurement time range, \( L_i \) sound intensity \((\text{dB})\).

(v) Implementation costs [17], \( x_5 \), other alternative production and implementation costs. Determining for \( a_5 \) assesses follows on several considerations.

(vi) Maintenance fee [18], \( x_6 \), cost for alternative machines.

(vii) Machine capabilities [19], \( x_7 \), represent hours of operation.

The multi-criteria analysis process involves many methods from information sources, big data, following data science definition and concept. AHP is used to determine the weight of the evaluation criteria [20], while TOPSIS or VIKOR is used to determine engine ratings with alternative fuels [21]. Therefore, to determine the factor constant \( a_i \), \( i = 1, \ldots, 7 \), each can consider the importance of several comparative information extracted from the information space. However, the importance of integrated measurement of the above criteria involves the objective function by maximizing and minimizing the objective [8], where some variables have an upper limit while some of them have a lower limit, or

\[ \max / \min \sum_{i=1}^{l} x_i \]  

(4)

If the machine has limited resource usage according to the criteria, then there is a constraint function. Composition of constraint in the matrix \( X \) matrix and the matrix \( A \), respectively, namely

\[ AX = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1i} \\ a_{21} & a_{22} & \cdots & a_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ a_{ki} & a_{k2} & \cdots & a_{ki} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \end{pmatrix} \]  

(5)
Figure 1. The rules performance based on number of parameters.

where $x_i \geq, =, \leq 0$.

The formulas are related to the balance between resources and production from a factory. Eqs. (4) and (5), for example, reveal the efficiency of using the resource but holds to static time. However, the dynamics of industrial processes have links to markets and other supporting resources [22]. These changes exist in the information space continuously and require extraction to provide feedback to the integration of the formulation in the program code. Every time a change occurs, it will provide input to decisions from daily to monthly, and this affects the performance of the factory or industry. The information that becomes the input for performance comes from big data, which shows market absorption of industrial production. Suppose $\Omega$ is the information space, there is $\Omega_i \in \Omega, i = 1, \ldots, n$ as parts of the space, where it referred to big data. That is according to their size. Extraction as expressed through data science is a function $\gamma$ that maps $\Omega$ to a composition of information that satisfies formulas such as Eqs. (4) and (5) or $X$, where the mapping is formed through data modeling $\tau$ so that extraction is

$$\gamma : \Omega \rightarrow (X, \tau)$$ (6)

The rules formed in $\tau$ depend on the number of parameters required, in this parameter encoding the processing combination follows the number of parameters with the truth value tested by the program following the complexity stated in the truth graph, see Figure 1.

Also, sensors that control and monitor industrial machines and the environment enter data into the information room via IoT. Data flow becomes big data, which is a source of learning to predict the future of the factory internally, while information from the market determines the future from externally. Each sensor streams data continuously or at least continuously into the information space based on their respective installed functions, namely $\kappa_j, j = 1, \ldots, m$. $m$ is the number of sensors installed in an industrial environment, and through IoT, all the information collected follows their respective patterns flowing into $\Omega$. As a whole it is gathered through the system template filter of big data $\rho$, which places the relationship between industry and information space with

$$\rho : \left\{ \kappa_j | j = 1, \ldots, m \right\} \rightarrow \Omega$$ (7)
Information from $\gamma$ and $\rho$ has a direct impact, and it changes the work system of factories from the use of natural resources, human resources, fuel and energy, production, markets, and others [3]. Through data science, expert systems, and databases integrated into big data, the running of the industry runs efficiently. The role of data science provides feedback to the industry through a variety of features and technologies but also reveals some changes in factory and market performance [23]. Data science, therefore, also provides predictions of targets to be achieved for optimal profit [24].

5. Conclusion

A review of Industry 4.0, according to data science, provides an overview of the integration of technology and data following two distinct paths. The extraction route reveals information resumes from the information room, whereas the reverse is the filtering of data from various tools that operate alongside industrial machines. These two different paths provide support for the implementation of industry 4.0. The two paths are the object of study from data science.

References

[1] D Shetty, S B Henderson 2019 Emerging trends in industry 4.0 with innovative case study of human balance & rehabilitation engineering *Proceedings of the International Conference on Industrial Engineering and Operations Management*. 2019.

[2] M Alabi, A Telukdarie, N J V Rensburg 2019 Industry 4.0: Innovative solutions fot the water industry *2019 International Annual Conference Proceedings of the American Society for Engineering Management and 40th Meeting Celebration: A System Approach to Engineering Management Solution*, ASEM 2019.

[3] C Catal, B Tekinerdogan 2019 Aligning education for the life science domain to support digitalization and industry 4.0 *Procedia Computer Science* 154.

[4] J Para, J Del Ser, A Aguirre, A J Nebro 2018 Decision making in industry 4.0 scenarios supported by imbalanced data classification *Studies in Computation Intelligence* 798.

[5] H Eissa 2020 Unleashing industry 4.0 opportunities: Big data analytics in the midstream oil & gas sector *International Petroleum Technology Conference 2020*, IPTC 2020.

[6] R Perez Fernandez, F J Regueira 2019 What the cad industry can do for the shipyard 4.0 *Sustainable Development and Innovations in Marine Technologies - Proceedings of the 18th International Congress of the International Maritime Association of the Mediterranean*, IMAM 2019.

[7] M K M Nasution, O S Sitompul, E B Nababan 2019 Data Science *Journal of Physics: Conference Series* 1566.

[8] M Sialahi, H Mawengkang, N I Syahputri 2018 Multi-objective model of waste transportation management for crude palm oil industry *IOP Conference Series: Materials Science and Engineering* 300(1).

[9] M A Garcia-Garza, H Ahnnett-Garza, M G Lopez, P Orta-Castañón, T R Kurfess, P D U Coronado, D Gueimes-Castorena, S G Villa, S Salinas 2019 A case about the upgrade of manufacturing equipment for insertion into an industry 4.0 environment *Sensor (Switzerland)* 19(15).

[10] M Compare, P Baraldi, E Zio 2020 Challenges to IoT-enabled predictive maintenance for industry 4.0 *IEEE Internet of Things Journal* 7(5).

[11] F S Tesch da Silva, C A da Costa, C D Paredes Crowato, R Da Rosa Right 2020 Looking at energy through the lens of industry 4.0: A systematic literature review of concerns and challenges *Computer and Industrial Engineering* 143.

[12] M H Sitepu, A R Matondang, M Tryana 2020 Sustainability assessment in construction industry supply network: A review *IOP Conference Series: Materials Science and Engineering* 725(1).

[13] E S Sadirsan, H Siregar, Eriyatno, E H Legowo 2015 Development model of renewable energy
policy for sustainable bio-pellet industry in Indonesia using interpretive structural method

Renewable Energy in the Service of Mankind 1.

[14] M L Singgih, Y Prasetiyawan, Sutikno, D Hartanto, F R Kurniawan, W T Wicaksana 2019
Maintenance management improvement based on reliability centered maintenance II in energy generating industries IOP Conference Series: Materials Science and Engineering 528(1).

[15] A Malloni, D Safu, H Amqam, A Syam, M Hatta, T Sutarti, M Abdul, S Stuyo 2019 Modelling of SO2 and CO pollution due to industry PLTD emission Tello in 2 Makassar Indonesia Journal of Engineering and Applied Sciences 14(2).

[16] I Lubis, M K M Nasution 2017 Probability Model for Designing Environment Condition Journal of Physics: Conference Series 801(1).

[17] A Hidayatno, I Rahman, K R Irminanda 2019 A conceptualization of industry 4.0 adoption on textile and clothing sector in Indonesia ACM International Conference Proceeding Series.

[18] C Carmichael, M Krarti 2010 Greening tenant/landlord processes demonstrating transformation in the industry ASME 2010 4th International Conference on Energy Sustainability, ES 2010.

[19] A D Sari, M I Januar, T Immawan, D S Samsudin 2018 The analysis of kaizen implementation for increasing production capability: A case study of musical instrument industry Journal of Engineering and Applied Sciences 13.

[20] F Fitriaestuti, U Rahmalisa, A S Girsang 2019 Multi-criteria decision making on successfull of online learning using AHP and regression Journal of Physics: Conference Series 1175(1).

[21] M M D Widianta, T Rizaldi, D P S Setyohadi, H Y Riskiawan 2018 Comparison of Multi-Criteria Decision Support Methods (AHP, TOPSIS, SAW & PROMENTHEE) for Employee Placement Journal of Physics: Conference Series 953(1).

[22] B Einabadi, A Baboli, M Ebrahimi 2019 Dynamic predictive maintenance in industry 4.0 based on real time information: Case study in automotive industries IFAC-PapersOnline 52(13).

[23] Y Demchenko, T Wiktorski, J Cuadrado Gallego, S Brewer 2019 EDISON data science framework (edsf) extension to address transversal skills required by emerging industry 4.0 Transformation Proceedings - IEEE 15th International Conference on eScience, eScience 2019.

[24] A Diez-Olivan, J Del Ser, D Galar, B Sierra 2019 Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0 Information Fusion 50.