Roadmap to Implementation of Industry 4.0 in Micro, Small & Medium Enterprises in India

Vaibhav Bisht1,*, V G Sridhar2
1Sibley School of Mechanical and Aerospace Engineering, Cornell University
2School of Mechanical Engineering, Vellore Institute of Technology, Chennai Campus

*vaibhav.bisht.work@gmail.com

Abstract. The paradigm of Industry 4.0 finds major promise for deployment in Indian industries. This paper discusses a qualitative research methodology and strategies for the implementation of Industry 4.0 in MSMEs from various sectors in India. The researchers have collected data from experts from different MSMES through questionnaires and personal interviews. Difficulties and barriers for implementation of industry 4.0 in India have been analysed using the snowball sampling method where the sample units expect different academic and government organisations to develop an apparatus for efficient implementation of the technologies of Industry 4.0. A technical framework for processing real-time big data using the Knowledge Data Discovery (KDD) pipeline for quick decision-making is discussed in this study. The study concludes with the use of the framework for use of process data to improve process efficiency.

1. Introduction

The growing trend of entrepreneurship, supportive government policies, and rising competition for prominence among national as well as international global business entities today have led to significant growth in the share of different Micro, Small and Medium Enterprises (MSMEs) to the economy of the country [1]. These have prompted a greater class of innovation today catering to the diverse goods and services markets of today, striving to make life easier for the common people. With more of such enterprises expected to contribute to the nation’s growth in the coming years, the various elements of the nation’s socio-economic fraternity have an ever-increasing interest in their success. The capable systems of Industry 4.0 today find significant use in empowering the efforts of such undertakings to prosper in the scopeful markets of today [2]. A lack of required workforce constantly affects the operations of these enterprises and the smart systems of Industry 4.0 find suitable use in such instances, performing the desired human role with greater efficiency at reduced costs. Such systems also avail the capability of working in sync with the human worker where the suitable task is assigned to the appropriate entity [3], giving rise to newer concepts like Human-Robot Collaboration (HRC) [4]. The philosophy of Industry 4.0 aims at integrated operations across the value chain, minimising waste and offering provisions for greater innovation by availing more time to the human workforce [5]. Additionally, factors of finances, supply chain management, and project planning are also major considerations for such enterprises, where an integrated structure will promise the multi-fold benefits of increased productivity at minimal costs to the enterprise [6-7].
There exists a lack of use of such systems of Industry 4.0 in MSMEs in India [8] and a framework for their scalable implementation requires detailed narratives concerning the current state of MSMEs in the Indian industrial faction. Surveys are used to collect data relevant to the scope of such a study by collecting information directly from experts operating in the various associated fields using different qualitative and quantitative methodologies. Therefore, the authors adopt a survey-based approach to discuss with various industry experts, the different availabilities and bottlenecks in their operations along with their expectations from the stakeholders for the suitable pathway for the adoption of Industry 4.0 by MSMEs in India.

2. Sampling

The target population identified, as part of the survey shown in Figure 1, for this study comprises all the individuals from the different technical and organisational hierarchies of the Indian industrial faction.

![Diagram](image)

**Figure 1.** Steps in survey on adoption of Industry 4.0 in MSMEs in India

The sampling frame comprises experts from various MSMEs whose scope was relevant to the use of systems of Industry 4.0. It included individuals from the fields of software development, embedded systems, sourcing, supply chain management, manufacturing, product design, chemical engineering, business administration, aerospace engineering, biotechnology, armament technology, wastewater engineering, start-up mentoring, data science, industrial engineering and legal services.

Non-probability sampling has been chosen for the study where the experts from the sampling frame are shortlisted based on their understanding of different concepts of Industry 4.0 like big data analytics, robotics, additive manufacturing and artificial intelligence, making their experience relevant to the scope of the survey. The authors found a limited number of professionals with such exposure...
and therefore, the snowball sampling method was employed. As shown in Figure 2, the initial sample units identified by authors were encouraged to bring other individuals with the required knowledge and experience to create a suitably sized representational sample [9] and obtain an appropriate response rate.

**Figure 2.** Sample created using snowball sampling

Snowball sampling is suitable for such educational and research studies [10] due to its low costs for adequate representativeness. Considering the broad area of application, bottlenecks and prospective purview of Industry 4.0, a qualitative approach with open-ended questions has been adopted for the authors to dwell deeper into the critical aspects of the study with less computation-heavy statistical methods for identifying the differing discrepancies and correlations. This approach also lets the sample units express themselves without any constraints or bias, as observed in the use of other methods like the Likert scale [11-12]. Multiple individuals are questioned in qualitative research where the salient points (also called themes) from their narratives are collected to conclude their study. An instant comes in such procedures where further interviewing does not add any significant details to the established understanding [13], called thematic saturation [14-15]. A sample size of 50 is taken as [16] have inferred using Monte Carlo simulations that exceeding this limit in qualitative research with open-ended questions adds minimal themes, which are irrelevant to the scope of the study. Participant recruitment strategies for creating the representative sample affect its adequacy and representativeness [17] and the sampling error, where the representative sample does not account for individuals from the target population with roles relevant to the scope of the study seriously affects the result, leading to wastage of time and resources. This was overcome by including individuals at varying levels of the company hierarchies in the sample, from associate engineers to founders and CEOs, holding experience from a few years to over 40 years.

Finalising the sample, different methods of data collection also lead to non-sampling errors like the measurement error, caused due mistake of the responder in understanding the question [18] and nonresponse error, caused because of lack of usable response from responders due to reasons like disinterest and unavailability [19]. Therefore, the data collection methodology was devised to
eliminate such inaccuracies. Formal discussions were conducted with multiple individuals from the representative sample to devise a user-friendly format of the questionnaire and orient them with the structure of the study. A review of literature and narratives of these discussions were used by the authors to devise a mixed-format approach comprising a questionnaire and interviews with all the members of the representative sample. The research comprises open-ended questions and therefore, the interviews enabled the authors to remove the non-sampling errors by clarifying any anomalies that existed in the narrative responses of the individuals.

Table 1. Open-ended questions answered by sample units

| S. No. | Question                                                                 |
|--------|--------------------------------------------------------------------------|
| 1.     | Does your organisation use any of the technologies of Industry 4.0?       |
| 2.     | What are the various challenges to adopting such systems in your view?    |
| 3.     | In which areas do you plan to implement such systems in the coming years?|
| 4.     | What initiatives do you expect for the large-scale adoption of such systems in India? |
| 5.     | Which entities do you expect to act to develop the required measures?     |

The questionnaire shared with the sample units is in Table 1, covering aspects of their experience, problems currently faced in industries and future initiatives.

3. Sample Response

Depending on their organisation’s use of different systems of Industry 4.0, the sample units were divided into four groups. A majority of these fell in Group 1 (“Already using one or more of such technologies”), followed by Group 2 (“Framework under development”) where relevant research & development efforts were underway for subsequent use. Third are institutions part of Group 3 (“May use if required in the future”) where individuals are aware of the availabilities of Industry 4.0 but do not find a pressing need for promoting such a change. Lastly, there were also members from organisations that did not find the need of shifting to the use of such systems, part of Group 4 (“Not going to use”). Since these enterprises had a well-established customer base and significant profit margins, they did need to make any changes in their existing systems. Additionally, they also rely on accustomed supply chains where time spent in reconfiguration could seriously affect the interests of various stakeholders.

The sample units discuss the different bottlenecks found in upgrading their industrial ware, shown in Table 2.

Table 2. Challenges faced by organisations in adoption of Industry 4.0

| S. No. | Description                        |
|--------|------------------------------------|
| 1.     | Lack of human resources            |
| 2.     | High cost of investment            |
| 3.     | Resistance to change               |
| 4.     | Inefficient organisational structure |

A majority believe that a lack of a skilled human workforce is the primary reason for the different issues in the adoption of Industry 4.0 in India. This is broadly attributed to a lack of reference data and
A lack of consulting companies that could help them create their operational models has also been seen to affect employee skill development efforts. Employing new workers with the required skills was also an issue faced by these industries as the many incipient technologies of Industry 4.0 are yet to be explored by the masses.

Catering to the mass-personalisation-based economies of Industry 4.0, industries are required to undergo a major overhaul of their existing resources. Major hardware and software systems are required to be removed or be replaced by other sophisticated apparatus, prompting expenses in logistics, disposal and meeting regulatory standards. Over 10% of those who answered also believe that there exists a certain resistance to change in organisations which is severely affecting their use of such new and efficient systems. Since there is a lack of data showing organisations in India benefitting from these redesigning efforts, companies have a constant fear of major financial losses. Many blue-collar workers are also being seen to lose their jobs in the certain instances of use of these technologies and the resistance to this change would also be coming from workers’ unions and other interest groups. Sample units also discussed how the organisational functioning of their companies is affecting their upgradation efforts and requires an appropriate focus on the aspect of system integration. They have mentioned instances of indecision from the higher levels of their hierarchy and disproportionate distribution of parent-company resources that is severely affecting their undertakings. They say service life is a crucial metric under their consideration and plan to use data to develop a sophisticated predictive maintenance apparatus for the long-term and cost-efficient use of the installed ware. With increased autonomous devices like robots being used on shop floors, they also plan on developing appropriate safety management features that would be important for the safety of the workers as well as installed machinery. Sample units were part of organisations focussing on optimising the Product Service Scheme (PSS) and aim to use these to offer real-time services to the different customers from remote locations.

The sample units cohesively believe that data to be the most primitive element of Industry 4.0 that widens its scope of use in various unrelated sectors in India and enables them to pursue their discussed undertakings. Since different upcoming MSMEs prioritize operations at minimal costs and workforce, they believe the judicious use of historical and real-time data is a critical element of their success. Considering the scope use of systems of Industry 4.0, current systems in use and the different bottlenecks in their associated operations, the sample units believe various stakeholders should work on creating a sophisticated network of technical consultancies, supervisory committees and training programs to devise a reference framework discussing the procedures and standards for efficient data acquisition and information processing to attain quick decision-making capabilities. The sample units expect the academic fraternity to actively contribute to such research and development efforts due to its availability of different resources that include research laboratories, equipment and staff members with the required expertise. Governments are also expected to play a key role in aiding such efforts due to their authority over varying research grants and regulations.

4. Framework

The authors present a framework consistent with the expectations of the representative sample for use of systems of Industry 4.0 to develop an apparatus for quick decision making where varying technical details required for their functioning and adoption are highlighted. A real-time data flow is a critical element of an interconnected and intelligent work environment [20-21] required for efficient decision-making in the varying aspects of Industry 4.0.
As shown in Figure 3, such a framework synchronically functions on the interaction between the different elements of the value chain that accounts for operations varying from managing logistics in the physical world to processing its associated operational information in the cyber space. The Internet of Things (IoT) finds critical use in the functioning of the scheme of decentralisation in Industry 4.0. It employs a large Wireless Sensor Network (WSN) that collects data from multiple locations in real-time for data acquisition [22]. As business operations expand with greater integration along the value chain, a standardised communication protocol [23] and secure channels for the transfer of operational data [24] are desired. The OPC UA (Open Platform Communications United Architecture) is suitable for such operations due to being an interoperable vendor-neutral protocol [25]. The OPC UA is used with the Time Sensitive Network (TSN) standards for real-time capabilities, called the OPC UA TSN [26]. The TSN promises low latency, desirable transfer rates and deterministic transmission [27-28] critical to use for the industrial environment but lacks usability being an Ethernet-based standard (for wired communication). Therefore, the wireless TSN discussed by [29-31] can be used for devising a wireless real-time OPC UA TSN. It also comes with the Advanced Encryption Standard (AES) and X.509 standard that are robust security apparatuses for the encryption and integrity of data [32].

Big data collected from these autonomous sources is processed through the Knowledge Discovery in Databases (KDD) pipeline shown in Figure 4.

The KDD pipeline primary recognises various patterns in the entered data and creates inferences relevant to the purpose of the operations. Stages in this sequence use different programming languages like python and software like Apache Spark or Apache Hadoop that employ varying pre-defined libraries, functions and logics based on different statistical, probabilistic and machine learning methods.
The selected big data (called features) is a voluminous and heterogeneous dataset that is likely to have varying irregularities due to reasons like error in measurement, noise and other redundancies [33]. This unnecessarily large dataset then requires high computational power and processing time for subsequent mining operations [34]. Since these features are also used for training different machine learning models for data mining, such variations can affect the performance of developed algorithms [35]. Along with correcting the different discrepancies, big data pre-processing is also important to convert the datasets into formats required for processing by the different big data mining algorithms. The features are, therefore, pre-processed to obtain a refined dataset called the feature subset using different pre-processing algorithms for operations like Missing Value Interpretation, Noise Treatment, Discretization & Normalization, Feature Extraction, Feature Selection and Feature Selection & Encoding [36-38].

The transformed data is the structured integrated dataset that is mined in the subsequent stage. Mining primarily mines the pre-processed data to recognise the salient patterns and is largely performed in two kinds of analytics:

4.1 Descriptive analytics
This method comprises the use of data of the past to determine what has happened in the current scenario. The various methods and appropriate tools used are presented in Table 3.

| Table 3. Methods adopted in descriptive analytics |
|-----------------------------------------------|
| S. No. | Method               | Description                                           | Tools                                                      |
|--------|----------------------|-------------------------------------------------------|------------------------------------------------------------|
| 1.     | Clustering           | Grouping of data on the basis of common traits.       | 1. k-means Clustering                                       |
|        |                      |                                                       | 2. Expectation–Maximization Clustering                     |
|        |                      |                                                       | 3. Agglomerative Hierarchical Clustering                   |
| 2.     | Association Rule     | Identification of relations between different data    | 1. Association Rules                                        |
|        | Discovery            | using the if-then patterns.                           | 1. Generalized Sequential Pattern (GSP) algorithm           |
|        |                      |                                                       | 2. SPADE (Sequential PAttern Discovery using Equivalence    |
|        |                      |                                                       | classes) algorithm                                         |
| 3.     | Sequence Discovery   | Discovery of statistically pertinent patterns from    | 1. Generalized Sequential Pattern (GSP) algorithm           |
|        |                      | sequential data.                                      | 2. SPADE (Sequential PAttern Discovery using Equivalence    |
|        |                      |                                                       | classes) algorithm                                         |

4.2 Predictive analytics
This system uses current and historical details to predict what will happen in the future. The various methods and appropriate tools used are presented in Table 4.
Table 4. Methods adopted in predictive analytics

| S. No. | Method                  | Description                                                                 | Tools                                      |
|--------|-------------------------|-----------------------------------------------------------------------------|--------------------------------------------|
| 1.     | Classification          | Grouping of data by comparing traits with information in created available database. | 1. Naïve Bayes Classifier                 |
|        |                         |                                                                             | 2. Bayesian Networks                      |
|        |                         |                                                                             | 3. J48 Algorithm                          |
| 2.     | Regression              | Use of relations between variables to predict the future states.             | 1. Lasso Regression                       |
|        |                         |                                                                             | 2. Multivariate Regression                |
|        |                         |                                                                             | 3. Multiple Regression                    |
| 3.     | Time Series Forecasting | Prediction of future states by identification of patterns in data ordered by time. | 1. Autoregressive Model                   |
|        |                         |                                                                             | 2. Moving Average Model                   |
|        |                         |                                                                             | 3. Simple Exponential Smoothing           |

Several unrelated patterns identified in these data mining operations are filtered out using pattern evaluation. Techniques like calculation of classification error and evaluation of association rules via numerical indicators are performed to assess these patterns and extract ones relevant to the user’s requirement. These results are subsequently used to integrate relevant nuggets of information for better understanding. The extracted patterns relevant to the scope of this study are converted to different forms comprehensible by machines, or humans using different techniques of visualisation for subsequent use. This information is then visualised in the form of tables, narrative reports and graphs using different summarization methods like sentence scoring and for the various technical and organisational requirements of planning, maintenance and restructuring as part of the final stage of knowledge exploitation.

5. Results and Discussion

Academia has a major role in enhancing the capabilities of various MSMEs concerning the discussed framework. Since, organisations also require greater collaboration among individuals and departments, elements of system integration in Industry 4.0 find major use in such upcoming enterprises. The qualified manpower can be used for training members of MSMEs with the discussed technical skills by organising different seminars, workshops and conferences. Considering the lack of knowledge of the mentioned technical details, stakeholders in academia can also work on developing alternative easier to use data processing algorithms and logics with similar capabilities. The governments can also promote research on such technologies by proposing greater grants to the members of the academia. The taxes and duties on their operations can also be eased to ensure greater capital for their efforts. The academia and governments can also collaborate to modify curriculums in academic institutions to equip the workforce of the future with these required skills. Additionally, relevant teachings on entrepreneurship should also be included for individuals to gain greater insights into integrating their resources as stakeholders in MSMEs. Innovation studios can also be developed where many of such stakeholders can work together in intellectual environments, promoting ingenuity by the use of various sophisticated resources.
6. Conclusion

Individuals with varying experience from diverse fields of interest are identified and the critical aspects of the adoption of Industry 4.0 by MSMEs in India using a survey have been discussed in this study. Various stakeholders primarily require an apparatus for technical consultation concerning the use of such technology. Data is inferred to be the most critical element of a working framework with significant use and therefore, a technical framework for the acquisition and processing of big data is presented. The Knowledge Data Discovery (KDD) pipeline finds suitable to use in this regard, considering the broad areas of use in the expanding outlook of MSMEs in India. Academia and governments play a crucial role in the varying upgradation efforts and their prospective offerings have also been discussed in detail. Subsequent research is now required on the use of this framework in various sectors for further analysis of the various pros and cons of their implementation.

References

[1] Srinivasan S and Kunjangad K 2019 Proc PICMET ’19: Technol. Manage. World Intell. Syst. 1-11.
[2] Seving A, Gür S and Eren T 2018 Processes 6(12) 264.
[3] Galin R and Meshcheryakov R 2019 IOP Conf. Ser.: Mater. Sci. Eng. 537 032073.
[4] Matheson E, Minto R, Zampieri EGG, Faccio M and Rosati G 2019 Robotics 8(4) 100.
[5] Sony M 2018 Prod. Manuf. Res. 6(1) 416-432.
[6] Mukerjee S 2018 J. Global Entrepreneurship Res. 8 28.
[7] Nardo M., Forino D and Murino T 2020 Prod. Manuf. Res. 8(1) 20-34.
[8] Kumar R, Singh R K and Dwivedi YK 2020 J. Cleaner Prod. 275 124063.
[9] Naderifar M, Goli H and Ghalaie F 2017 Strides Dev. Med. Educ. 14(3) 67670.
[10] Elfil M and Negida A 2017 Emergency 5 (1) e52.
[11] Bishop PA and Herron RL 2015 Int. J. Exercise Sci. 8(3) 297-302.
[12] Kreitchmann RS, Abad FJ, Ponsoda V, Nieto MD and Morillo D 2019 Front. Psychol. 10.
[13] Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, Burroughs H and Jinks C 2018 Qual. Quantity 52 1893–1907.
[14] Lowe A, Norris AC, Farris AJ and Babbage DR 2018 Field Methods 30(8) 1-17.
[15] Guest G, Namey E and Chen M PLoS ONE 15(5) e0232076.
[16] Tran VT, Forcher R, Tran VC and Ravaud P. J. Clin. Epidemiol. 82 71–78.e2.
[17] Raza S, Faheem M and Guenes M 2019 Int. J. Commun. Syst. 32(15) e3758.
[18] Ponto J 2015 J. Adv. Pract. Oncol. 6(2) 168-171.
[19] Badvaz M J. Off. Stat. 26(1) 25–42.
[20] Lindner JR, Murphy TH and Briers GE J. Agric. Educ. 42(4) 43-53.
[21] Kabugoa JC, Jounelaa SLJ, Schiemannn R and Binder C 2020 Int. J. Electr. Power Energy Syst. 115 105508.
[22] Zhong RY, Xu X, Klotz E and Newman ST 2017 Eng. 3 616-630.
[23] González I, Calderón AJ, Figueiredo J and Sousa JMC 2019 Electron. 8(5) 510.
[24] Tawalbeh L, Muheidat F, Tawalbeh M and Quwaider M 2020 Appl. Sci. 10 4102.
[25] Panda SK, Majumder M, Wisniewski L and Jasperneite J 2020 Proc 2020 25th IEEE Int. Conf. Emerging Technolo. Factory Autom. (ETFA) 1143-1146.
[26] Li Y, Jiang J, Lee C and Hong SH 2020 IEEE Access 8 200100-200111.
[27] Li Q, Li D and Sun X 2020 IOP Conf. Ser. Mater. Sci. Eng. 782 042015.
[28] Nasrallah PDFA, Thygaturu AS, Alharbi Z, Wang C, Shao X, Reisslein M and ElBakouy H 2019 IEEE Commun Surv. Tutorials 21(1).
[29] Fischer C, Krummacker D, Karrenbauer M and Schotten HD 2021 Inf. 12(1) 12.
[30] Khorov E, Levitsky I and Akyildiz IF 2020 IEEE Access 8 88664-88688.
[31] Kang Y, Lee S, Gwak S, Kim T and An D 2021 Energies 14 4497.
[32] Mourtzis D, Angelopoulos K and Zogopoulos V 2019 *Procedia CIRP* **84** 265-270
[33] Mhamdi F and Elloumi M 2008 *Proc. 2008 Second Int. Conf. Res. Challenges Inf. Sci.* 427-432
[34] Tsai CW, Lai CF, Chao HC and Vasilakos AV 2015 *J. Big Data* **2** 21
[35] Gupta S and Gupta A 2019 *Procedia Comput. Sci.* **161** 466-474
[36] García S, Gallego SR, Luengo J, Benítez JM and Herrera F 2016 *Big Data Anal.* **1** 9.
[37] Meena MJ, Chandran KR, Karthik A and Samuel AV 2012 *Expert Syst. Appl.* **39**(5) 5861–871
[38] Kang H 2013 *Korean J. Anesthesiol* **64**(5) 402–406