

# Application of Machine Learning and Data Mining in Manufacturing Industry

Zhen Song, Shu Luo

School of Mechanical and Electrical Engineering, Southwest Petroleum University, Chengdu 610000, China

## ABSTRACT

With the rise of machine learning in various industries, the traditional manufacturing industry is facing a new disruption, which requires the use of different technologies and tools to achieve its production targets; In this regard, machine learning (ML) and data mining (DM) play a key role. This paper provides a statistical understanding of the main methods and algorithms used to improve manufacturing processes over the past 20 years by dividing them into four main themes: Scheduling, Monitoring, Quality and Failure, presents previous ML research and the latest advances in manufacturing, followed by a comprehensive discussion of existing problem solutions in manufacturing from multiple aspects. It includes tasks (i.e., clustering, classification, regression), algorithms (i.e., support vector machines, neural networks), learning types (i.e., ensemble learning, deep learning), and performance indicators (i.e., accuracy, mean absolute error). In addition, the main steps of database knowledge discovery (KDD) process that should be followed in manufacturing applications are described in detail, and the methods to overcome some problems and the advantages of machine learning applied to manufacturing industry are briefly described. Finally, the paper summarizes and further looks forward to the future development direction.

## KEYWORDS

Machine Learning; Data Mining; Manufacturing Technology; Classification and Clustering

## 1. INTRODUCTION

Machine learning [1] (ML) is an important research field of artificial intelligence, which can help computers to model and accurately predict future events based on experience. The main ML methods fall into two broad categories: supervised learning and unsupervised learning. A typical problem in supervised learning is classification, while unsupervised learning is quite common in clustering. Common classification techniques include neural network, support vector machine and decision tree, and the most widely used clustering technique is k-means. ML technology has been widely and successfully applied in many different fields, such as health, education, wireless sensor networks and finance [2]. This paper Outlines the application of ML technology in manufacturing, where modern manufacturing plants use powerful data acquisition systems to electronically collect and transmit data for almost all processes in an organization. Many manufacturing variables are continuously measured at various stages and their values are stored in an organization's database. This data may be related to the characteristics of the product, the machine, the production line (i.e., which machine uses which setting parameters), the human resources operating the production line (i.e., the level of experience of the workers, the type of shift), the raw materials used in the process, the environment (humidity, temperature, etc.), Sensors connected to the machine (vibration, force, pressure, tension, etc.), machine failure, maintenance, product quality and other important manufacturing factors.

Thanks to technology, the manufacturing industry generates a lot of raw data every day. This large and increasing availability of data has attracted the attention of machine learning concepts. Nearly 20 years ago, machine learning and data mining applications emerged in this field to solve manufacturing problems, intelligent systems that support efficient decision making, planning synchronous production lines, and machine maintenance arrangements can be seen as examples of using ML methods to perform manufacturing tasks. Other examples include failure prediction and its energy consumption estimation, product quality assessment and manufacturing defect detection [3].

The field of machine learning, which includes deep learning, ensemble learning and connected learning, is considered one of the most promising improvements in manufacturing. In addition, its applications in manufacturing include many varieties, from automotive manufacturing to apparel, from semiconductors to many other scientific and engineering fields. Much of the research in machine learning has focused on classification, which is the task of assigning an object to one of the predefined categories. On the other hand, some manufacturing problems belong to the category of clustering. The task of clustering is to divide objects into several groups according to their similarity, which is called clustering [4]. Recently we can see about ML and DM in manufacturing has made new progress in the research, but these new progress on the things is limited, only focus on one aspect, this article the application of machine learning and data mining in manufacturing has carried on the overview of the system, and introduces the research status both at home and abroad, and specify the manufacturing unique advantages and challenges, It opens up a new prospect for future application.

## **2. OVERVIEW**

### **2.1. Overview of ML and DM technologies in manufacturing.**

In manufacturing, machine learning is often more effective than traditional mathematical and statistical models because they are still unable to understand complex relationships between features of data samples or predict unknown eigenvalues of new samples. Because of this situation, ML technology, which is used in a wide range of scientific disciplines, has also been used in manufacturing in recent years. The use of ML and DM technologies is widely recognized in the manufacturing industry because intelligently analyzing data is a valuable resource as it gains new insights that can provide a significant competitive advantage. While it is quite possible to find a master's degree in one discipline (manufacturing or machine learning), few researchers are well integrated in both fields. Therefore, ML in manufacturing is mainly solved by data scientists and manufacturers together. ML technology is applied to the manufacturing data collected by the data technology system. Important features and structures were determined by data analysis; Discover the hidden knowledge, rules and patterns of data through data mining; Machine learning is used to build effective models to train the behavior of manufacturing systems. Data analysis plays an important role in decision making (decision support) in manufacturing. The most common manufacturing tasks used by ML and DM technologies can be listed as scheduling, monitoring, quality assessment, and fault detection. In addition, other manufacturing tasks that benefit from ML capabilities include layout planning, sales forecasting, and process mining [5]. In addition, ML technology is used for many other manufacturing tasks, such as product design, time/cost forecasting, volume forecasting, anomaly detection, and prediction of machine energy consumption.

Figure 1[6] below shows the number of publications using ML and DM technologies between 2000 and 2019. As can be seen from the figure, the number of studies is increasing, indicating the popularity of this topic. The presence of large volumes of manufacturing data is likely to make ML and DM even more important in the coming years. Search for the following keywords in the two sources we looked for (Google Scholar and Web of Science) to get the statistical results given in Figure 1:

(manufacturing) and (machine learning or data mining or supervised learning or unsupervised learning classification or clustering (regression and prediction) or integrated learning or depth or

decision tree or neural network or support vector machine (SVM) or random forests or k - nearest neighbour or naive bayesian or convolution neural network association rule mining sequential patterns mining or text mining and web mining)

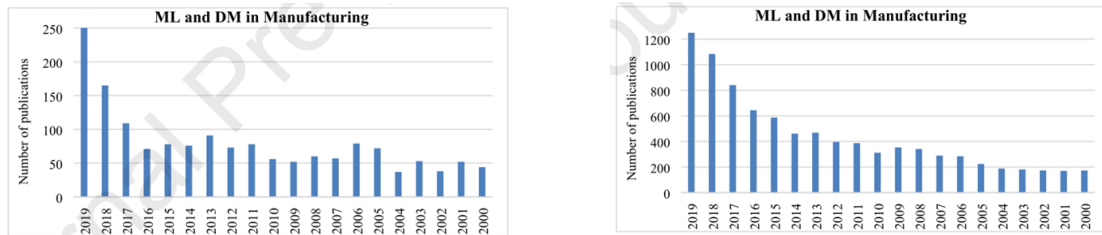


Fig.1 Number of publications per year from (left) Google Scholar and (right) Web of Science related to ML and DM in manufacturing

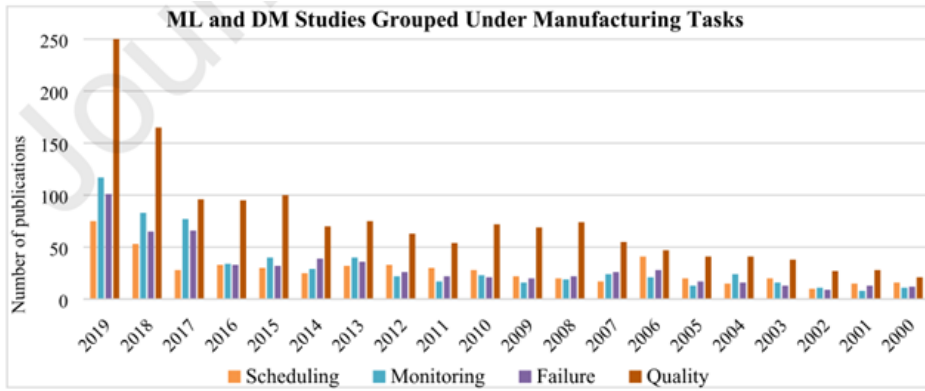
## 2.2. Problems that ML and DM technologies can solve.

Studies on ML and DM can be divided into four main themes [7]:

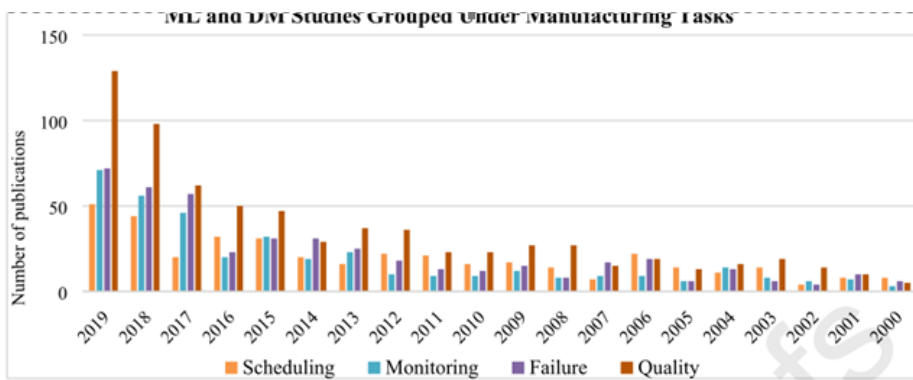
- (1) Scheduling: including order processing, shop scheduling, sequencing, resource allocation, job scheduling and manufacturing planning.
- (2) Monitoring: including decision support system and process monitoring system, to avoid key performance index value deviation, and improve the visibility of the manufacturing system.
- (3) Quality: including product quality prediction, quality improvement in large and complex processes, quality monitoring and control in manufacturing process and diagnostic defect detection.
- (4) Fault: including abnormal situation detection, machine maintenance, fault prediction, equipment monitoring, equipment downtime, and equipment fault information system analysis.

Figure 2[8] shows the proportional distribution of ML and DM based manufacturing studies by their specific objectives. These categories include planning, monitoring, quality, and fault (failure) detection. Manufacturing project goals may vary from company to company and are not limited to these categories. However, most manufacturing research using machine learning aims to solve these types of problems. As Figure 2 shows, the number of machine learning studies for all manufacturing tasks is increasing. In the past three years in particular, this reinforcement has accelerated. The reasons behind this trend may be publicity and encouragement from the government and multinational production companies, as well as the popularity of Industry 4.0. Figure 2 also reveals the fact that far more studies are expected to improve manufacturing quality than other studies. Add the following precise terms to the above keywords, respectively, and then perform search queries in Scopus and Web of Science to get the results given in Figure 2:

(Quality control or quality prediction or quality assurance or quality management or defect detection or defect prediction) and (fault diagnosis or fault detection or fault prediction or fault classification or (fault analysis) and (process monitoring or condition monitoring or monitoring system) and (scheduling).



(a) Scopus



(b) Web of Science

Figure 2. Number of publications obtained from (a) Scopus and (b) Web of Science, based on targets for the period 2000 to 2019.

Weichert et al.[9] classified the manufacturing data used in ML as follows: qualitative and quantitative data, time series and workpiece related data, controlled and uncontrollable data, current and historical data, measured and simulated data, observable measurements and process state variables. Wang[10] also classified commonly used manufacturing variables in ML, including resource variables, processing variables and working conditions variables.

### 2.3. Multiple machine learning approaches handle manufacturing tasks.

Figure 3[11] shows the general categories of manufacturing responsibilities associated with machine learning approaches. Reference numbers are also attached to publications cited for conducting research to achieve these objectives. Although each manufacturing task is relatively specific to its working conditions, they can be unified under these main groups. Headings for these manufacturing tasks include: Product design, Decision Support, production, process, monitoring, quality, defects, faults/failures, scheduling, layout planning, sales, and energy.

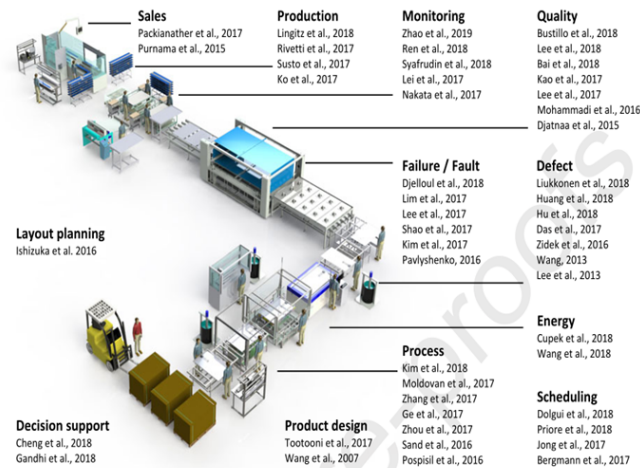


Fig.3 The ML and DM studies were grouped by production task

## 2.4. Data processing technologies (tools, engines, libraries and frameworks) used in the manufacturing industry in recent years

There are a number of data processing tools that provide modeling and prediction capabilities based on ML technology. The literature describes several studies that examine the use of data processing tools in manufacturing. The most widely used free and/or open source data processing tools in manufacturing research (at least for academic purposes) are Weka, R, RapidMiner, KNIME, Orange, Elki, Tanagra, Mallet, and KEEL[12]. These tools allow for easy application in many situations, as well as comfortable adjustment of parameters to improve model accuracy. Python libraries like Keras, Theanos, TensorFlow, Caffe, and Scikit Learn make programming ML relatively easy.

Some of the most common open source data processing engines are Hadoop, Spark, Samza, Flink, and Storm. Machine learning techniques are typically applied to large scale manufacturing data, so they should be able to handle high dimensional data (data sets with more than 20 attributes). Some examples of large distributed machine learning frameworks currently available are MLlib, Mahout, Samoa, H2O, and MLbase[13].

## 2.5. Widely used production data sets

In scientific research, baseline data sets are used to demonstrate the power of proposed methods and to compare the performance of algorithms. One of the most widely used reference data sets related to manufacturing is SECOM[14], which is obtained from semiconductor manufacturing processes. Therefore, the data set consists of manufacturing process variables. Another well-known manufacturing data set, called "Plate Faults," is used in various studies to train machine learning algorithms for automatic pattern recognition. It contains information about steel plate products and therefore mainly includes product variables. Some studies have also processed a very significant benchmark manufacturing data set called "Bosch production Line Performance" to test the classification performance of proposed algorithms[15]. The production line variable is the basic feature of this data set.

## 3. KNOWLEDGE DISCOVERY PROCESS IN MANUFACTURING INDUSTRY

Figure 4[16] shows the overall knowledge discovery process in a database applied to the manufacturing industry. This process usually consists of five main steps: understanding the manufacturing domain, data preparation, machine learning, data mining, evaluation, and presentation.

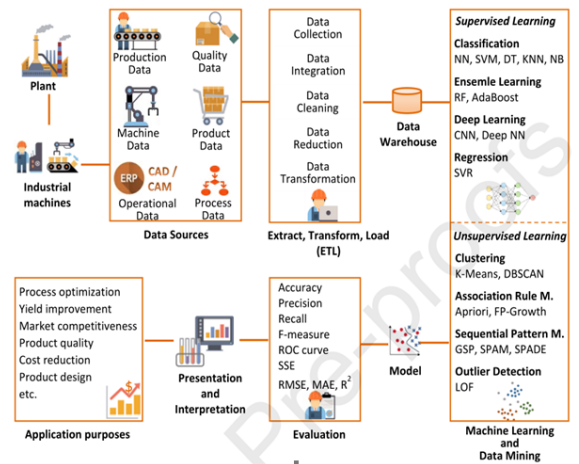


Fig. 4 Monolithic KDD processes are used in manufacturing

As can be analyzed from the figure above, the first stage can be called the design stage. In this phase, the goals of the application, the resources available, the constraints in the mining process, the success criteria for the problem, and the costs and benefits of the application are determined. The second phase is called data preparation and includes data collection, integration, cleansing, reduction, and transformation. Data collection steps include collecting data on many different manufacturing variables such as raw materials, final products or machine adjustments (temperature, pressure, production Settings, time scales, etc.) with the help of sensors or external automatic recorders. Data integration focuses on combining multiple data sources. Data cleaning involves filling in missing values, processing noisy data, resolving inconsistencies in data, considering unbalanced data, and detecting and deleting outliers in data to improve data quality [17]. Data reduction is performed in order to obtain the target data set from the original data without significant loss of information, such as feature selection. The purpose of data transformation processing is to convert data into forms suitable for mining, such as normalization and discretization, when necessary. After the data preparation phase, the data sets are stored in the data warehouse. The third step involves applying appropriate machine learning algorithms to the data in the warehouse to extract patterns/rules or develop models. Supervised learning algorithms can be used for classification and regression problems, while unsupervised learning algorithms can be used for clustering, association rule mining, sequential pattern mining and outlier detection[18]. In the fourth step, the constructed model is evaluated by using appropriate performance indicators. For example, the most commonly used performance indicators for regression problems are root mean square error (RMSE), mean absolute error (MAE) and determination coefficient ( $R^2$ )[19]. The final stage involves interpretation and visualization of patterns, which can be represented as key performance indicators (KPIs) on the dashboard, or alerts in the case of abnormal detection, or predictions obtained through regression models that can be displayed on-screen. Often, some KDD steps require multiple iterations until a satisfactory result is achieved. Finally, the constructed model is incorporated into the manufacturing field. The model should be further modified as new data becomes available. For example, due to the dynamic nature of manufacturing systems, regression models should be updated regularly to maintain their generalization ability. Knowledge discovered during the KDD process can support operator/manager decisions or be directly used to automate improvements in manufacturing systems.

#### 4. APPLICATIONS OF MACHINE LEARNING IN MANUFACTURING

Machine learning tasks can be divided into supervised, unsupervised and reinforcement learning. Although supervised and unsupervised learning techniques are already widely used in manufacturing, accounting for about 90-95% of all applications, reinforcement learning has not been studied as extensively as other techniques. For this reason, this section presents a number of important research roadmaps involving supervised and unsupervised learning in manufacturing.

## 4.1. Supervised learning in manufacturing

Supervised learning aims to learn the mapping between sample input and output pairs. Simply put, a supervised learning algorithm may have many input variables and one output variable. Logically, the number of examples available to learn proportionately affects the predictive ability of supervised learners. Supervised learning is typically used for two separate tasks: categorization and regression. The main difference is that classification is the process of predicting discrete or nominal (classification) values, such as low, medium, and high; Regression is used to predict sequential (numerical quantity) or order values, such as the price of a car. There are many machine learning algorithms that can be used to achieve these goals, each with its own advantages and disadvantages, such as decision trees (DT), neural networks (NN), support vector machines (SVM), K-Nearest Neighbor (KNN), and Naive Bayes (NB).

Because manufacturing data is often characterized by multiple sources (i.e., machine, product, operator), heterogeneity, and noise, some data preprocessing steps such as standardization, attribute construction, feature selection, and elimination of missing values have been considered in many studies. The quality of data mining results depends on the quality of the data. All preprocessing operations need to address various issues related to manufacturing variables such as product, process, and machine variables. At present, because the distribution of manufacturing data is often uneven, the issue of class imbalance has received wide attention in the manufacturing industry. Because of the possible performance loss of ML, the SMOTE technique has been used in a number of studies to overcome this problem [20]. Many manufacturing organizations store processing data in time series and intend to improve quality control by building predictive models on large amounts of time data. Supervised machine learning methods have been widely used to predict and eliminate defects and faults in early production steps in the steel industry [21]. They are also used to create effective models of additive manufacturing. Especially in this area, neural networks are useful for building models with high predictive accuracy. Control procedures in the production process have been improved by supervised learning techniques, such as random forest methods. Because intelligent manufacturing is capable of producing high quality products, many studies have been conducted to create quality prediction models using machine learning methods. Some manufacturing materials or process equipment are prone to deterioration over time. They make the manufacturing process more dangerous. In this regard, supervisory methods have been used to detect internal faults of materials in advance [22].

## 4.2. Unsupervised learning in manufacturing

Unsupervised learning is an example of machine learning used to identify regularities and correlations in unlabeled data. Clustering, association rule mining, anomaly (outlier) detection, density estimation and presentation learning can be listed as the most popular unsupervised learning methods [23]. In all of these tasks, the primary goal is to generalize the internal data structure in a way that produces a useful representation without the aid of explicit class tags. Unsupervised learning research is relatively rare, as it is likely to encounter data with category labels in manufacturing.

The cluster divides instances into groups based on their similarity. Clustering is created based on similarity or distance measures that determine how similar or different the data are from each other. The main clustering methods fall into five categories. The partitioning clustering method attempts to decompose the data into  $k$  clusters so that the items in each cluster are closely related to each other. Hierarchical clustering method builds a cluster tree by repeatedly merging smaller clusters into larger ones (clustering) or splitting larger ones into smaller ones (splitting) [24]. Density-based clustering methods attempt to find high-density clusters separated by sparse regions that may differ in size and shape (i.e. not convex, spherical, or elongated). There are also grid-based and model-based approaches.

In the manufacturing industry, cluster analysis is used for pattern recognition, output improvement, quantitative assessment and equipment status diagnosis [25]. In addition, clustering is used to detect product errors, aid the decision-making process, and layout planning problems. In addition, human-computer interaction plays an important role in manufacturing. The necessary operations in this field have been determined by the unsupervised learning model, and the event-driven response of the robot has been formed [26].

## **5. ADVANTAGES AND CHALLENGES OF ML AND DM IN MANUFACTURING**

Data mining has been widely used as a basic tool to discover knowledge from manufacturing databases. The necessary data to be analyzed can be collected throughout the conventional manufacturing process. In manufacturing, data mining provides many competitive advantages, such as higher product quality, lower costs, and improved production processes. It may help automate the knowledge discovery process, a utility that is considered important for the development of knowledge-based systems. There are many areas of manufacturing where machine learning could have a positive impact. First, efficient demand forecasting is highly supported by machine learning. The tendency is to analyze past events to estimate how much product should be produced to meet future demand. Second, the launch of a new product is a process involving machine learning. When launching a new product, use machine learning to track launch success, including sales and customer data. Another effect is price optimization. Manufacturing companies are able to take into account location, seasonality, weather and demand in order to rearrange prices and display products at the best price. It has been observed that ML enables manufacturers to reduce total cycle time and improve resource utilization in some NP-difficult manufacturing problems [27]. In addition, ML provides a powerful approach to continuous quality improvement for complex and large processes. Other advantages of DM and ML in manufacturing are given below; It must be said, however, that the degree of importance of advantage may vary depending on the algorithm chosen. Other advantages are: predictive maintenance; Resource management; Product design; Quality control; Decision support; Optimization; Descriptive analysis; Predictive analysis; Parameter analysis; Market demand analysis; Multi-stage application; Document classification and clustering [28].

Based on the above advantages, machine learning is a useful tool, and when you realize this, some challenges can be overcome. All the tasks that make up the basic steps of the knowledge discovery process may not always be easy to apply. Most researchers agree that the main challenges ML faces in manufacturing are as follows. Learning from and automatically adapting to changing environments is a major advantage of machine learning. Due to the dynamic and rapid changes in the manufacturing environment, ML systems should have the ability to learn and adapt to changes, and system designers need to provide solutions for all possible situations. Another major challenge is getting accurate and relevant manufacturing data, as it has a big impact on the performance of ML algorithms. A common challenge for ML applications in manufacturing is the preprocessing of data, as it has a critical impact on the results. Another key challenge is what ML methods and algorithms to choose. The final major challenge is the interpretation of the results. Information about these and many other challenges can be listed below [29] :

1. Manufacturing data preparation issues: The success of each ML technique depends on the data structure that executes it. However, getting well-organized data in manufacturing is not an easy and quick process. This is because manufacturing data is often characterized by multiple sources (i.e., product, machine, process, operator, raw material, environment, and service data), heterogeneity (i.e., structured or unstructured, syntactic or semantic data), and noise (i.e., incomplete, incorrect, improper, repetitive, and inconsistent data). Prior to the application of ML technology, specific data preprocessing steps (data integration, cleaning, reduction, conversion) are required [30]. Data preprocessing has a crucial impact on the results. However, there are currently no standardized rules

for which data preprocessing techniques should be applied to specific types of manufacturing problems. Identify appropriate techniques through insight and knowledge, or by trying and comparing alternatives.

2. Time consuming issue: When the time spent by all machine learning processes is divided into subsegments that include preprocessing, feature extraction, and classification, it is noted that data preparation takes the most time before any data mining algorithm is implemented. This is because the manufacturing data set consists mainly of a large number of events and non-standardized measurements. Sometimes, data preprocessing itself takes a lot of time, accounting for about 50 or 60 percent of the total work on a machine learning project.

In addition to the above, other difficulties are mainly reflected in: data missing problem, data selection problem, manufacturing data imbalance problem, interdisciplinary collaboration challenge, manufacturing data protection and security problem, high dimension problem, etc. [31].

## 6. SUMMARY AND OUTLOOK

Machine learning applications are likely to continue to grow at a higher rate, especially in manufacturing, because computing power is growing day by day and the scale of data available is much larger than it was a few years ago. Big data technology can handle high dimensional data. Particularly given the increasing availability of manufacturing data, they are likely to become more important in the future. The application of association rule data mining in manufacturing industry is more frequent than sequential pattern mining (SPM) [32]. Extended surveys of the SPM, which take into account temporal information about model development, allow manufacturers to respond quickly to time-dependent (temporary) situations. Frequent sequential patterns provide potentially important knowledge for predicting future activity. Text mining has been widely used in many fields in recent years. However, the use of IT in manufacturing is limited. Future work could focus on text mining related to manufacturing. Because the amount of text data from manufacturing suppliers keeps increasing, more than 80% of enterprise information is stored as text [33]. Various text mining methods, such as association rule generation, classification, or clustering, can be implemented to process this data on a large scale and extract valuable manufacturing knowledge. Manufacturing documents can be classified or clustered according to their types, main contents and similarities[34]. Text mining methods can be very helpful in managing these digital document sources. For example, sentiment analysis can be used as a way to survey sentiment in manufacturing-related content or as a tool to help analyze consumer trends. Text mining implementations such as question and answer, expert discovery, sentiment detection, recommendation, part-of-speech tagging, and parsing can be added to improve manufacturing. Manufacturing companies share information about their business (i.e., plans, processes, materials, and technologies) on the network, which helps manufacturers make decisions based on this information. The increasing use of the Internet has had a huge impact on the creation of complex digital manufacturing information[35]. Web mining provides an automatic mechanism to collect data related to manufacturing industry, and extract valuable and understandable business information from these massive Web business data to further provide decision support. However, different terminology used by different manufacturing companies can cause confusion and difficulties in practical and dynamic processes. Therefore, in the future of manufacturing, the theme of web mining will also have more contributions. Recently, some ontology-based information semantic representation methods have been used to organize manufacturing data. However, more contributions are expected in the next few years to develop ontology-based manufacturing systems using ML technology[36]. The ontology-based system provides extraction of semantic relations to improve accuracy and form a better decision support system [37]. New ML algorithms or heuristics can be developed to match manufacturing concepts to ontologies. Several ontologies can be developed for the manufacturing domain. Although supervised and unsupervised learning is widely used in manufacturing, accounting for about 90-95% of all applications, reinforcement learning (RL)

is not as widely studied as in other fields [38]. However, RL provides goal-directed learning without external supervision, ADAPTS to dynamic environments, and provides a framework for understanding and modeling systems in the face of rewards and punishments. RL can help solve complex combinatorial decision-making problems in manufacturing industry, especially various planning and control problems[39]. RL is likely to become more popular in manufacturing in the near future. At present, the clustering research of manufacturing industry usually uses k-means algorithm. However, the k-means++ algorithm can be addressed in future research to improve the speed and accuracy of K-Means. So far, there has been little research (using k-means++ in manufacturing). DBSCAN may soon play a more important role because of its ability to form arbitrarily shaped clusters and handle noise in the data[40]. There are two common methods of data processing: batch processing and stream processing (real-time data processing). Batch data processing is carried out statically. A group of transactions are collected within a period of time and processed to build the model. Meanwhile, the built model is dynamically updated based on the newly accumulated data [41]. Streaming (real-time) data processing provides processing as the data comes in. Most data mining research in the manufacturing field is conducted on batch data. However, recent advances in technology require streaming data processing to gain a competitive advantage in real-time decision making. In real-time or near-real-time data processing, fast response times are critical, so processing times in seconds are acceptable. Real-time data mining and learning in the manufacturing field will be an important and difficult topic for further research [42]. The use of robots is already beginning to find its way into manufacturing. Industrial robots have become a new trend in manufacturing enterprises. They are attracting more and more attention every day. It is estimated that ML technology, the emergence of smart factories, and the use of industrial robots will play a more important role, and applications based on them will increase significantly in the manufacturing industry in the near future [43].

## REFERENCES

- [1] Yang Xiaojing, Zhang Fudong, Hu Changbin. Review of Machine Learning Research [J]. Market of Science and Technology Economy, P. 40-42.
- [2] Bai Yuying, Li Chunchun, Sun Zhiqiang, Chen Hao. Deep neural networks for Manufacturing quality prediction. Proceedings of the 8th IEEE Health Management Conference on Predictive Systems, 2017,07, 09/p. 115-186.
- [3] Han Jie, Chi Shuai. Applying machine learning methods to predictive manufacturing. Proceedings of the 8th International Conference on Future Networks (ICUFM), 2016,07,05. p. 176-189.
- [4] Huang Yong, Pan Can, Lin Nan. Machine Learning Method for 3-DIC Defect Detection and Classification based on TSV [J]. 2018,10. p. 125-136.
- [5] Ahmadinia,M.,Alinejad-Rokny,H.,&Ahangarikiasari,H.(2014). Data aggregation in wireless sensor networks based on environmental similarity: A learning automata approach. JournalofNetworks,9(10), p. 2567-2573.
- [6] Ahmad, A., & Dey, L. (2007). A k-mean clustering algorithm for mixed numeric and categorical data. Data & KnowledgeEngineering,63(2), p. 503-527.
- [7] Zheng Yongshi, Zhang Tianwei. Online Analysis of Intelligent Factory Based on Text Mining [J] ICIC Express, p. 71-89.
- [8] Arellano-Espitia, F., Delgado-Prieto, M., Martinez-Viol, V., Saucedo-Dorantes,J., &Osornio-Rios, R.(2020). Deep-learning-based methodology for fault diagnosis in electromechanical systems. Sensors, 20(14), p.3929.
- [9] Bergmann, S., Feldkamp, N., & Strassburger, S. (2017). Emulation of control strategies through machine learning in manufacturing simulations. Journal of Simulation, 11(1), p.38-50.
- [10] Cheng, Y., Chen, M., Cheng, F., Cheng, Y., Lin, Y., & Yang, C. (2018). Developing a decision support system (DSS) for a dental manufacturing production line based on data mining. In Proceedings of 4th IEEE International Conference on Applied System Innovation 2018 (ICASI 2018), Chiba, Japan, 13- 17 April 2018, p. 638-641.
- [11] Alfaro-Cortes, E., Alfaro-Navarro, J., Gamez, M., & Garcia, N. (2020). Using random fores to interpret out-of-control signals. Acta Polytechnica Hungarica, 17(6), p.115-130.
- [12] Das, B., Pal, S., & Bag, S. (2017). Torque based defect detection and weld quality modelling in friction stir welding process. Journal of Manufacturing Processes, 27, p. 8-17.

- [13] Choudhary, A., Harding, J., & Tiwari, M. (2009). Data mining in manufacturing: a review based on the kind of knowledge. *Journal of Intelligent Manufacturing*, 20, p.501–521.
- [14] Deng, H., Diao, Y., Wu, W., Zhang, J., Ma, M., & Zhong, X. (2019). A high speed D-CART online fault diagnosis algorithm for rotor systems. *Applied Intelligence*, 50, p.29-41.
- [15] Ferreira, R., Sabbaghi, A., & Huang, Q. (2020). Automatic geometric shape deviation modelling for additive manufacturing systems via Bayesian neural networks. *IEEE Transactions on Automation Science and Engineering*, 72, p.584-598.
- [16] Arif, F., Suryana, N., & Hussin, B. (2013). A data mining approach for developing quality prediction model in multi-stage manufacturing. *International Journal of Computer Applications*, 69(22), p.5-40.
- [17] Bustillo, A., Urbikain, G., Perez, J., Pereira, O., & de Lacalle, L. (2018). Smart optimization of a frictiondrilling process based on boosting ensembles. *Journal of Manufacturing Systems*, 48(PartC), p.108-121.
- [18] Forero-Ramirez, J., Restrepo-Giron, A., & Nope-Rodriguez, S. (2019). Detection of internal defects in carbon fiber reinforced plastic slabs using background thermal compensation by filtering and support vector machines. *Journal of Nondestructive Evaluation*, 38(1), Article Number 33.
- [19] Zidek, K., Maxim, V., Pitel, J., & Hosovsky, A. (2016). Embedded vision equipment of industrial robot for inline detection of product errors by clustering–classification algorithms. *International Journal of Advanced Robotic Systems*, 13(5), p.1-10.
- [20] Amruthnath, N., & Gupta, T. (2018). A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. In *Proceedings of the 5th International Conference on Industrial Engineering and Applications (ICIEA)*, Singapore, 26-28 April 2018, (pp. 355-361).
- [21] Buddhakulsomsiri, J., & Zakarian, A. (2009). Sequential pattern mining algorithm for automotive warranty data. *Computers & Industrial Engineering*, 57(1), p.137-147.
- [22] Wang, J., Wang, K., Wang, Y., Huang, Z., & Xue, R. (2018). Deep Boltzmann machine based condition prediction for smart manufacturing. *Journal of Ambient Intelligence and Humanized Computing*, 21 April 2018, p.1-11.
- [23] Traini, E., Bruno, G., D'antonio, G., & Lombardi, F. (2019). Machine learning framework for predictive maintenance in milling. *IFAC-PapersOnline*, 52(13), p.177-182.
- [24] Bai, Y., Sun, Z., Deng, J., Li, L., Long, J., & Li, C. (2018). Manufacturing quality prediction using intelligent learning approaches: A comparative study. *Sustainability*, 10, article no 85, p. 1-15.
- [25] Cupek, R., Ziebinski, A., Zonenberg, D., & Drewniak, M. (2018). Determination of the machine energy consumption profiles in the mass-customised manufacturing. *International Journal of Computer Integrated Manufacturing*, 31(6), p.537-561.
- [26] Han, J., & Chi, S. (2016). Consideration of manufacturing data to apply machine learning methods for predictive manufacturing. In *Proceedings of the Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*, July 5-8, 2016, Vienna, Austria, (pp. 109-113).
- [27] Herrera, A., Stoyanov, S., Bailey, C., Walshaw, C., & Yin, C. (2019). Data analytics to reduce stop-onfail test in electronic manufacturing. *Open Computer Science*, 9, p.200-211.
- [28] Huang, J., Segura, L., Wang, T., Zhao, G., Sun, H., & Zhou, C. (2020). Unsupervised learning for droplet evolution prediction and process dynamics understanding in inkjet printing. *Additive Manufacturing*, 35, 101197.
- [29] Jung, Y-S., & Chang, T-W. (2018). Text mining based online news analysis about smart factory. *ICIC Express Letters, Part B: Applications*, 9(6), p.559-565.
- [30] Moldovan, D., Cioara, T., Anghel, I., & Salomie, I. (2017). Machine learning for sensor-based manufacturing processes. In *Proceedings of the 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, Cluj-Napoca, Romania, 7-9 Sept. 2017, (pp. 147-154).
- [31] Jong de, W. A., Rubrico, J., Adachi, M., Nakamura, T., & Ota, J. (2017). Big data in automation: Towards generalized makespan estimation in shop scheduling problems. In *Proceedings of the 13th IEEE Conference on Automation Science and Engineering (CASE)*, Xi'an, China, August 20-23, 2017, (pp. 1516-1521).
- [32] Kao, H., Hsieh, Y., Chen, C., & Lee, J. (2017). Quality prediction modeling for multistage manufacturing based on classification and association rule mining. In *Proceedings of the 2nd International Conference on Precision Machinery and Manufacturing Technology, ICPMMT 2017*, Pingtung, Taiwan, 19-21 May 2017, MATEC Web of Conferences, 123, 21 September 2017.
- [33] Kim, A., Oh, K., Jung, J., & Kim, B. (2018). Imbalanced classification of manufacturing quality conditions using cost-sensitive decision tree ensembles. *International Journal of Computer Integrated Manufacturing*. 31(8), p.701-171.
- [34] Lee, H., Kim, Y., & Kim, C. (2017). A deep learning model for robust wafer fault monitoring with sensor measurement noise. *IEEE Transactions on Semiconductor Manufacturing*, 30(1), February 2017, Article number 7744687, p.23-31.

- [35] Lei, Q., Shao bo, L., & Jing kun, C. (2017). Online monitoring of manufacturing process based on autoCEP. *International Journal of Online Engineering*, 13(6), p.22-34.
- [36] Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, 115, p.213-237.
- [37] Zhang, X., Kano, M., Tani, M., Mori, J., Ise, J., & Harada, K. (2020). Prediction and causal analysis of defects in steel products: handling nonnegative and highly overdispersed count data. *Control Engineering Practice*, 95.
- [38] Wen, L., Gao, L., & Li, X. (2019). A new deep transfer learning based on sparse auto-encoder for fault diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(1), p.136-144.
- [39] Wuest, T., Irgens, C., & Thoben, K. (2014). An approach to monitoring quality in manufacturing using supervised machine learning on product state data. *Journal of Intelligent Manufacturing*, 25(5), p.1167- 1180.
- [40] Wang, Q., Jiao, W., Yu, R., Johnson, M., & Zhang, Y. (2019). Modeling of human Welders' operations in virtual reality human-robot interaction. *IEEE Robotics and automation letters*, 4(3).
- [41] Tootooni, M.S., Dsouza, A., Donovan, R., Rao, P.K., Kong, Z., & Borgesen, P. (2017). Classifying the dimensional variation in additive manufactured parts from laser-scanned three-dimensional point cloud data using machine learning approaches. *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, 139(9), Article number 091005.
- [42] Tian, Y., Fu, M., & Wu, F. (2015). Steel plates fault diagnosis on the basis of support vector machines. *Neurocomputing*, 151(1), p.296-303.
- [43] Rostami, H., Dantan, J., & Homri, L. (2015). Review of data mining applications for quality assessment in manufacturing industry: Support Vector Machines. *International Journal of Metrology and Quality Engineering*, 6, p.1-18.