APPLYING MACHINE LEARNING (ML) IN THE SUPPLY CHAIN: PREDICT THE QUALITY DEFECT PERCENTAGE IN DELIVERY FROM SUPPLIER

APLICAÇÃO DE APRENDIZADO DE MÁQUINA (ML) NA CADEIA DE SUPRIMENTOS: PREVEJA A PORCENTAGEM DE DEFEITOS DE QUALIDADE NA ENTREGA DO FORNECEDOR

APLICACIÓN DE APRENDIZAJE AUTOMÁTICO (ML) EN LA CADENA DE SUMINISTRO: PREDECIR EL PORCENTAJE DE DEFECTOS DE CALIDAD EN LA ENTREGA DEL PROVEDOR

Cited as:
Simão, .L. E. & Santos, A. M. (2024). Applying machine learning (ml) in the supply chain: predict the quality defect percentage in delivery from supplier. Revista Gestão & Tecnologia. Journal of Management and Technology. v 24, nº 1. p .191-214.

Luiz Eduardo Simão
Professor titular do programa de mestrado profissional em administração - gestão, internacionalização e logística (PMPGIL) na Univali
https://orcid.org/0000-0002-6526-9348

Andre Moraes dos Santos
Professor titular da Universidade do Vale do Itajaí do Programa de Mestrado profissional em Administração
https://orcid.org/0000-0002-8605-5234

Scientific Editor: José Edson Lara
Organization Scientific Committee
Double Blind Review by SEER/OJS
Received on 10/04/2023
Approved on 22/02/2024

ABSTRACT

Study Objective: The objective of this article was to describe the process of developing and implementing an ML model in the supply chain to predict the percentage of defective parts prior to supplier delivery.
Methodology/Approach: The methodology used was based on an action research project using a case study approach that described the steps divided into five phases: acquiring data, preparing data, analyzing data, communicating results, and applying results. The machine learning model applied the supplier's performance data related to the supply chain between the years 2021 and 2022.
Originality/Relevance: Despite the growing interest in ML techniques by many companies, challenges remain in known situations and potential applications in building explainable business and decision models. Thus, there is little empirical evidence of the relationship between effective implementation of machine learning (ML) techniques and their real effect on supply chain performance.
Main Results: The results show that, by employing the proposed method, inspection volumes can be reduced by more than 30%, and therefore, economic advantages can be generated by reducing inspections on material receipt.
Theoretical/Methodological Contributions: The main contribution was to demonstrate how the application of ML models can have a positive impact on supplier management process performance. Additionally, the article also describes how to use ML algorithms without the need to write code. Thus, the article can be a possible reference for organizations wishing to use similar ML approaches in their supply chains and improve the quality levels of their suppliers' performance.

Keywords: machine learning; quality; defects; suppliers.

RESUMO

Objetivo do estudo: O objetivo deste artigo foi descrever o processo de desenvolvimento e implementação de um modelo de ML na cadeia de suprimentos para prever a porcentagem de peças defeituosas antes da entrega do fornecedor.
Metodologia/abordagem: A metodologia utilizada foi baseada em um projeto de pesquisa de ação realizado usando uma abordagem de estudo de caso que descreve as etapas divididas em cinco fases: adquirir dados, preparar dados, analisar dados, comunicar resultados e aplicar resultados. O modelo de aprendizado de máquina com a aplicação dos dados de desempenho do fornecedor relacionado à cadeia de suprimentos entre os anos 2021 e 2022.
Originalidade/Relevância: Apesar do crescente interesse nas técnicas de ML por muitas empresas, os desafios permanecem em situações conhecidas e aplicativos em potencial na construção de modelos de negócios e de decisão explicáveis. Assim, há pouca evidência empírica da relação entre a implementação efetiva das técnicas de aprendizado de máquina (ML) e seu efeito real no desempenho da cadeia de suprimentos.

Principais resultados: Os resultados mostram que, ao empregar o método proposto, os volumes de inspeção podem ser reduzidos mais de 30% e, portanto, as vantagens econômicas podem ser geradas pela redução de inspeções no recebimento dos materiais.
Contribuições teóricas/meto dológicas: A principal contribuição foi demonstrar como a aplicação dos modelos de ML pode impactar positivamente no desempenho do processo de gestão dos fornecedores. Além disso, o artigo também descreve como utilizar algoritmos de ML sem a necessidade de escrever códigos. Assim, o artigo pode ser uma possível referência para organizações que desejam usar abordagens de ML semelhantes em suas cadeias de suprimentos e melhorar o desempenho dos níveis de qualidade de seus fornecedores.

Palavras -chave: aprendizado de máquina; qualidade; defeitos; fornecedores

RESUMEN

Objetivo del estudio: El objetivo de este artículo fue describir el proceso de desarrollo e implementación de un modelo de ML en la cadena de suministro para predecir el porcentaje de piezas defectuosas antes de la entrega del proveedor.
Metodología/enfoque: La metodología utilizada se basó en un proyecto de investigación de acción realizado mediante un enfoque de estudio de caso que describió las etapas divididas en cinco fases: adquirir datos, preparar datos, analizar datos, comunicar resultados y aplicar resultados. El modelo de aprendizaje automático aplicó los datos de rendimiento del proveedor relacionados con la cadena de suministro entre los años 2021 y 2022.
Originalidad/relevancia: A pesar del creciente interés en las técnicas de ML por muchas empresas, persisten desafíos en situaciones conocidas y aplicaciones potenciales en la construcción de modelos de negocio y de decisión explicables. Por lo tanto, existe poca evidencia empírica de la relación entre la implementación efectiva de las técnicas de aprendizaje automático (ML) y su efecto real en el rendimiento de la cadena de suministro.
Principales resultados: Los resultados muestran que, al emplear el método propuesto, los volúmenes de inspección pueden reducirse en más del 30%, y por lo tanto, se pueden generar ventajas económicas al reducir las inspecciones en la recepción de materiales.
Contribuciones teóricas/metodológicas: La principal contribución fue demostrar cómo la aplicación de modelos de ML puede impactar positivamente en el rendimiento del proceso de gestión de proveedores. Además, el artículo también describe cómo utilizar algoritmos de ML sin necesidad de escribir código. Por lo tanto, el artículo puede ser una posible referencia para organizaciones que deseen utilizar enfoques de ML similares en sus cadenas de suministro y mejorar los niveles de calidad del rendimiento de sus proveedores.
Palabras clave: aprendizaje automático; calidad; defectos; proveedores.

1. INTRODUCTION

The growing importance of quality for competitive success has been demonstrated over the last 30 years by various companies, which successfully understood and consequently translated customer requirements into final products (Robinson & Malhotra, 2005). Increasing market demand towards higher product and process quality and efficiency forces companies to think of new and innovative ways to optimize their production (Kovacic & Sarler, 2009). Facing these trends, manufacturing companies must deal with rapidly increasing complexity within their manufacturing and business processes to achieve the expected quality of their products. Despite the increasing challenges of rising product variety and complexity and the necessary of economic manufacturing, a comprehensive and reliable quality inspection is often indispensable. In consequence, high inspections volumes turn inspection processes into manufacturing bottleneck (Schmitt & Deuse, 2020).

Quality is defined by the International Organization for Standardization (ISO) as "the degree to which a set of inherent characteristics fulfils requirements" (ISO, 9001:2015). In this context, requirements refer to stated, generally implied, or obligatory needs or expectations. Based on this foundation, product quality in manufacturing can be defined as the extent to which manufacturing supplies are capable of offering products that fulfill customer requirements (Koufteros et al., 2002; Boon-itt, 2010).

The supply of defect-free, high-quality products is an important success factor for the long-term competitiveness of manufacturing companies. In this context, an industrial company has several suppliers that supplies different materials and components with wide quantities, unit prices and total purchase order values. Sometimes a few parts from suppliers fail to pass the delivery quality checks, but the defect percentage doesn't seem to be following a trend. The supply manager would like to predict the defective piece percentage in delivery from the supplier. If the defective piece percentage predicted is below the threshold level, then a quality check will not be performed.

In this study, we investigated whether machine learning has explanatory power for supplier quality prediction problems in the industry. The objective of this article, therefore, is to utilize a predictive model based on supervised machine learning algorithms that allows to predict the supplier quality on delivery on the base recorded process parameters. Using different machine learning algorithms such as Regression Analysis (RA), Support Vector Machine (SVM), AdaBoost (AB) and Artificial Neural Networks (ANN), which allows to interpret the prediction result and enables quality-based process decision support.

The paper is organized as follows. Section 2 introduces the context and the investigated reality. The diagnosis of the problem situation is presented in section 3. Section 4 discuss the problem analysis and the intervention proposal. Still, section 4 present a overview of the defect prediction methodology used and explain the machine learning model. Section 5 show the results and analysis of supplier defect prediction model applied based on a case study. Section 5 concludes and presents the technological contributions,

2. CONTEXT AND THE INVESTIGATED REALITY

The present technological article is based on a case study that was carried out in a medium-sized industrial company in the Medical, Hosptital and Dental Equipment (MHDE) sector. The industry focus of the study is of national origin, was founded in Santa Catarina 40 years ago and is a producer of medical and dental equipment, as well as peripherals and accessories. Due to the confidentiality clause, we will identify it in this article only as an Alfa company.

The company Alfa has a network of resellers spread throughout Brazil in São Paulo, Santa Catarina, Paraná, Rio Grande do Sul, Bahia Minas Gerais and abroad in the United States of America, in addition to a network of technical assistance in all Brazilian states. The company sells to the domestic market, whose main customers are in the Southeast and South, and for export to more than 100 countries, whose main markets are Colombia, USA, Australia and Canada.

In 2021, the company produced 2,325 pieces of equipment and sold 2,224 pieces of equipment at retail, with a market share of 25.7% of the Brazilian market.

In Brazil, the MHDE sector, the production chain of the health sector, represents approximately 2.7% of the industrial GDP, a market of around BRL$ 8.5 billion, with the generation of BLR$ 54.5 million direct and indirect jobs (ABIMO, 2019). The MHDE sector in Brazil had revenues in 2018 of BRL$ 843 million with imports of BRL$ 63 million with exports (ABIMO, 2019).

The Medical, Hospital and Dental Equipment (EMHO) sector is of fundamental importance for the supply of the health products market. The industries that make up the sector have a high degree of innovation in scientific and technological knowledge worldwide, which gives them dynamism in terms of product development and improvement and competitiveness (Morelli, Figlioli, & Oliveira, 2010). Figure 1 shows the distribution of the size of companies in the MHDE sector (ABIMO, 2019).

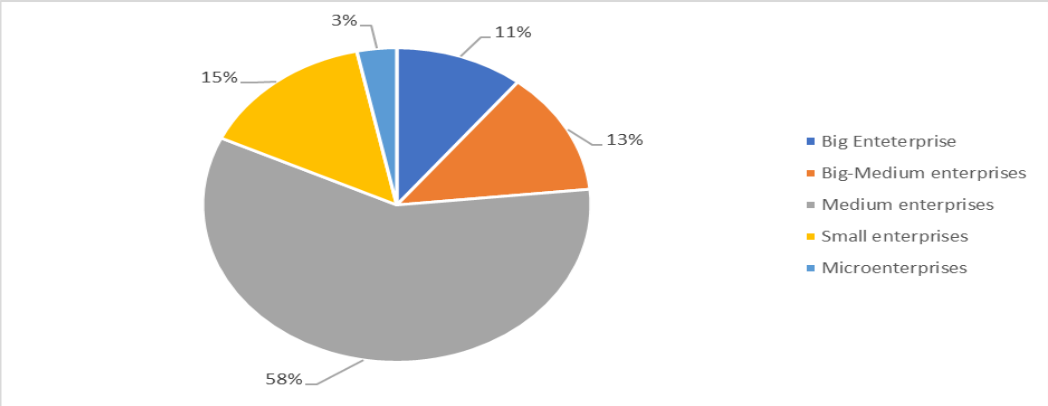

Figure1 - Size of companies in the MHDE sector
Source: ABIMO (2019)

Figure 1 shows a large predominance of medium-sized companies (58.6%), considering annual revenues of BRL$2.4 million to BRL$6 million. Micro and small enterprises (MSEs) together correspond to 18% of companies, which shows their importance for the sector.

It is observed in Figure 2, that the private sector is the largest customer of the sector, which corresponds to 69.62% of the sector's total purchases. The public sector also has good representation with a total of 19.41% of the sector's purchases.

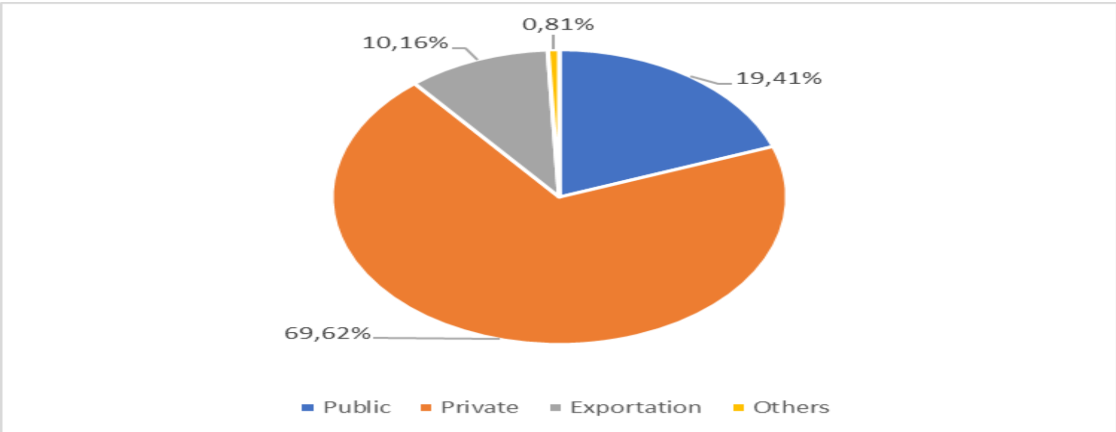

Figure 2 - Buyers in the MHDE sector
Source: ABIMO (2019)

The MHDE sector is inserted in a highly dynamic and competitive environment, which requires companies to invest heavily in the innovation of their products and processes to maintain competitiveness against large foreign companies.

3. DIAGNOSIS OF THE PROBLEM SITUATION

It is essential to ensure that only products meeting quality expectations are delivered to customers. However, the complexity of quality control processes increases because of increased product customization and variety (Thalmann et al., 2019; Hirsch et al., 2019).

So, product quality is a fundamental condition for maintaining a business and not just considered a competitive advantage. Organizations' mission is to ensure that the products supplied to customers do not have defects, ensuring total customer satisfaction and thus creating business success. “Good quality reduces rework, scrap and returns costs and, most importantly, good quality generates satisfied customers” (Slack et al., 2002).

However, before and during the production of a particular product, problems may occur, these problems can be generated by several causes like poor quality supplier parts, or causes that are recurrent, and for that, organizations need to be prepared to know how to deal with these problems. According to Aguiar (2014), such problems are generally registered in industrial organizations as non-conformities. Non-conformities are nothing more than non-compliance with the requirements established for the goods and services produced. The consequences of non-conformities or problems that occur are numerous, such as loss of time, material, man-hours, increased number of re-inspections and reworks, waste, reduced productivity, which causes financial losses (Aguiar, 2014).

The company has several suppliers that supplies different materials with wide unit prices and total purchase order values. Sometimes a few parts from suppliers fail to pass the quality checks, but the defect percentage doesn’t seem to be following a trend. The supply manager would like to predict the defective piece percentage in delivery from the supplier. If the defective piece percentage predicted is below the threshold level, then a quality check will not be performed. Let us consider that if the defect piece percentage in the delivery is more than 0.4%, then explicit incoming inspection is needed to ensure the quality of the components of end product. This will help the company to focus on quality checks on only particular purchase orders delivery, to control the end-product quality, and optimize inspection costs. It will also enable us to uncover the variables/parameters which are influencing the defective deliveries of few materials at times and work collaboratively with the supplier to address it.

McGrath (2013) noted that the increasing competition between companies has made it challenging to maintain long-term competitive advantages in several markets. As a result, companies constantly strive to differentiate their products and services from competitors to attract more consumers. To achieve this, significant attention is given to the production area.

4. PROBLEM ANALYSIS AND INTERVENTION PROPOSAL

To understand the supplier quality problem, we first realized a macro analysis, and take out was the total amount of non-conforming parts that are sent by all suppliers, where the incident identification covers all the parts at the edge of the line plus the defective parts found in the stock during an inspection, initiated due to an anomaly detected by the production team.

It is worth mentioning that due to the assured quality policy for the engine components, the company no longer applies receipt audits to verify the quality of the components. In this indicator (see Figure 1) there is a favourable trend towards the reduction of PPM, due to the reduction of incidents and the increase in the volume of company production. With this, there is a 45% and 29% reduction in the PPM index compared to the years 2019 and 2020, as show in Figure 3.

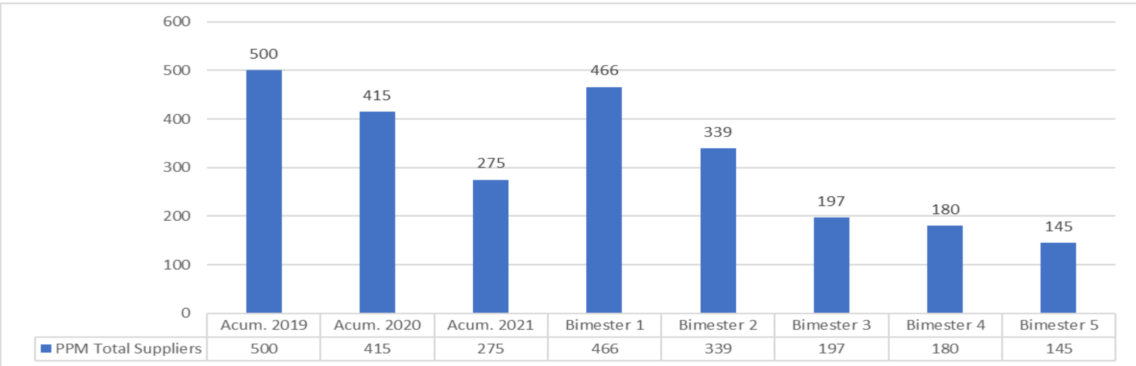

Figure 3 – Total amount of non-conforming parts by suppliers

The problem analysis shows that the main product´s component was a small engine used in all company end products. The analysis of the quality issues is summarized inf Figure 4.

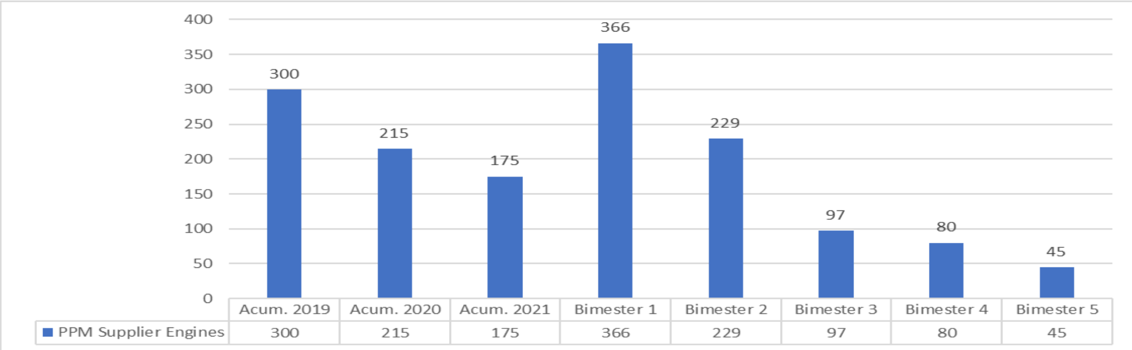

Figure 4 – Total supplier engine quality problems

Most of these quality problems are related to defect supplied direct to the assemble line, as showed in Figure 5.

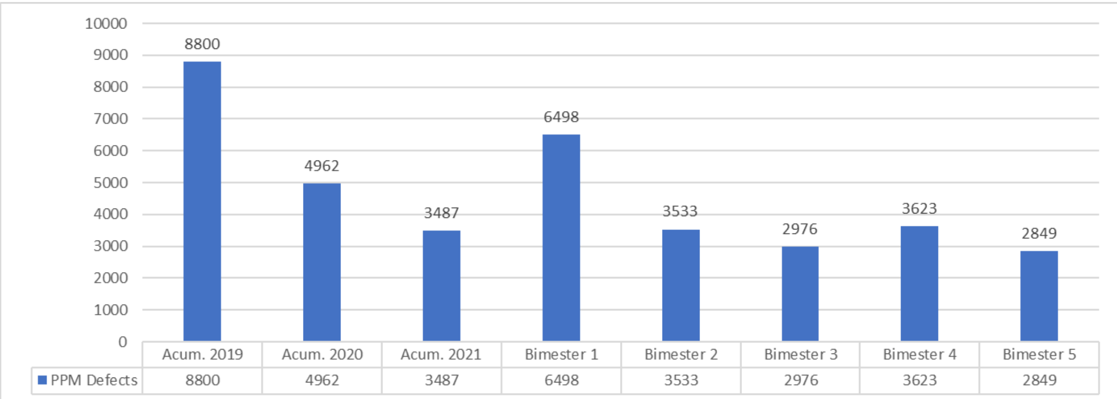

Figure 5 – Total defect detected in assembly line

Despite the improvement in the engine supplier quality results, the supply manager was not satisfied, and he wants to go further. So, to deal with this supplier quality problem, we suggest an approach using four machine learning technics: Linear Regression (LR), Support Vector Machine (SVM), AdaBoost (AB) and Artificial Neural Networks (ANN) to predict the supplier defect by delivery.

Görz, Schneeberger, and Schmid (2013) define ML as dealing with the computer-aided modeling and realization of learning phenomena. Wenzel, Smit, and Sardesai (2019) state that it is a process that uses experience to improve performance or make concrete predictions. The experience refers to past information, which is provided to the procedure from an electronic data collection. ML involves the design of effective and precise algorithms (Mohri, Rostamizadeh, & Talwalkar, 2012).

The knowledge gained from data can then be generalized and used to solve new problems and analyse previously unknown data. A central role in ML are algorithms, which are responsible for the recognition of patterns and generation of solutions. They can be categorized according to different learning paradigms into (Deng & Li, 2013; Jiang et al., 2017):

•supervised learning,
•unsupervised learning,
•semi-supervised learning,
•reinforcement learning, and
•active learning

Supervised learning refers to training models based on labelled training data. This entails the training of models by taking the expected outcome into account, e.g., the classification group. In unsupervised learning, on the other hand, the model groups are formed automatically based on independently recognized patterns (Mohammed, Bashie, & Khan, 2017). Semi-supervised learning is located between supervised and unsupervised learning. It has gained increasing importance recently, as fully labelled data sets are often not available or can only be generated with high costs. The method of reinforcement learning uses rewards and penalties to improve model performance. Active learning aims at finding useful rather than merely statistical findings. Thereby, instead of using statistical evaluations, the supervising user is asked to provide feedback on a question from which the algorithm should learn in a targeted manner (Berendt et al., 2016). Despite their different approaches, all learning tasks require algorithms to solve the anticipated problem.

Machine Learning methods can be categorized by the task they are solving. Figure 2 show a overview of ML types and methodologies.

As described in Figure 6 the classification and regression are two tasks of supervised learning.

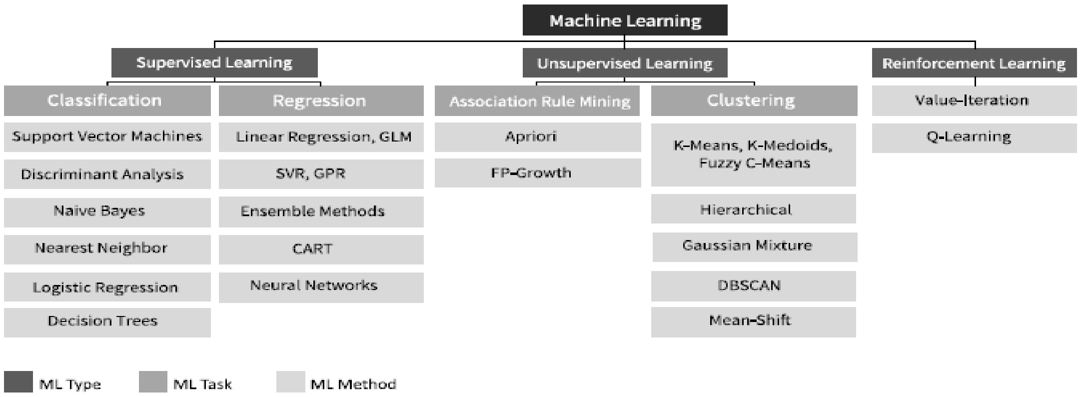

Figure 6 – Overview of ML types and methodologies
Source: Wenzel, Smit & Sardesai (2019)

Examples of tasks in supervised learning are decision trees, neural network and SVR. Examples of tasks in unsupervised learning are clustering and association rule mining. Another common type of learning is reinforcement learning. The focus of this article is on supervised learning.

Figure 2 shows selected methods of ML, grouped by the task they commonly solve. It is important to mention that this selection of methods is far from complete, and some methods can be used to solve more than one task. Also, there exist several variants of algorithms for each method.

ML approaches are widely used in manufacturing and help to achieve a variety of objectives. Some examples that use ML methods to perform manufacturing tasks are data-driven support for decision-making (Cheng et al., 2018; Kujawińska et al., 2018), scheduling (Priore et al., 2018), predictive maintenance (Gashi & Thalmann, 2020; Schmidt & Wang, 2015), fault detection/prediction (Lim et al., 2017; Shao et al., 2017), defect detection/prediction (Zidek et al., 2016; Das et al., 2017; Wang, 2013), quality assessment (Bustillo et al., 2018; Lee et al., 2018; Bai et al., 2018; Schmitt & Deuse, 2020), and condition monitoring (Ren et al., 2018; Zhao et al., 2019; Syafrudin et al., 2018). Some of the most common ML approaches for defect prediction and detection in manufacturing include decision trees (DT), naive Bayes classifiers (NB), support vector machines (SVM), and linear discriminant analysis (LDA) (Dogan & Birant, 2020; Schmitt & Deuse, 2020). For instance, Zhang et al. (2020) employ an SVM in the steel industry to predict defects in the early production stage. Combining multiple ML models can increase predictive accuracy in general, and in the manufacturing industry in particular (Priore et al., 2018; Huang et al., 2018). These combination approaches are known as ensemble learning approaches (Zhou, 2009). Common ensemble learning algorithms include random forest (RF) (Liaw et al., 2002), AdaBoost (Freund & Schapire, 1997), and gradient boosting (Friedman, 2002). For instance, defects on the surface of steel plates were predicted using AdaBoost (Hu et al., 2018). Gandhi et al. (2018) used RF to improve decision support in manufacturing maintenance while aiming for defect prediction. Lingitz et al. (2018) used RF to predict manufacturing lead times. Furthermore, deep learning approaches (e.g., convolutional neural networks, or CNNs) have recently been shown to significantly improve defect prediction/detection in different industrial application settings (Imoto et al., 2018; Lee et al., 2017).

While the recent state of research contains some literature reviews on general applications of ML in manufacturing (Harding et al., 2006; Kusiak, 2006; Wenzel, Smit & Sardesai, 2019), specific reviews with a focus on quality-related applications are rarely found (Köksal, Batmaz & Testik, 2011; Bustillo et al., 2018; Lee et al., 2018; Bai et al., 2018; Schmitt & Deuse, 2020).

The supplier defect prediction methodology used is depicted in Figure 7.

Figure 7 - Overview of the supplier defect prediction methodology

The first step of the methodology was acquired data from ERP. In this step is considered the identification of datasets, retrieve data and query data. The result was data collected of purchase orders information related to purchasing order value (BLR$), purchasing order quantity, po send in advanced of delivery (days) and respective incoming inspection results by defect percent (%) gathered in the ERP. This provided the data points for all the orders over three years period are summarized in Table 1.

The second step was preparing the data. In this step we first explore data to understand the nature of data and preliminary analysis. After that we are pre-processing raw data to clean, integrate to deal with missing data and data standardization.

In the third step data is analysed using Orange machine learning and data mining software considering five different prediction algorithms (Linear Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), AdaBoost (AB) and Artificial Neural Networks (ANN)) for training (learning with) 80% raw data. With the best algorithm (based on the analyse the performance related to prediction error MSE, RMSE, MAE and R2), we tested the best model with last 20% of raw data to confirm the results. The MSE (mean squared error), MAE (mean absolute error), RMSE (root mean squared error), and R-squared (R2) are mainly used metrics to evaluate the prediction error rates and model performance in regression analysis (Chicco, Warrens & Jurman, 2021):

• MAE (Mean absolute error) represents the difference between the original and predicted values extracted by averaging the absolute difference over the data set (Sammut & Webb, 2010a).

• MSE (Mean Squared Error) represents the difference between the original and predicted values extracted by squaring the average difference over the data set (Sammut & Webb, 2010b).

• RMSE (Root Mean Squared Error) is the square root of MSE and is a measure of the standard deviation of the residuals (Kelley & Lai, 2011).

• R-squared (Coefficient of determination) represents the proportion of variance in the dependent variable that is explained by the independent variable(s). The value ranges from 0 to 1 and is often interpreted as a percentage (Di Bucchianico, 2008).

The fourth step we communicate results by report to supply manager and his team. In this step, we show the results of the model and their error rates of performance in all five different prediction algorithms.

Finally, in the fifth step, based on the results we apply the results to solve the problem stated in the purpose of the study.

In the next section are presented the machine learning model procced using Orange machine learning and data mining software.

4.1 – The Machine Learning Model

Now-a-days, many open-source data mining tools and software are available for use such as the Rapidminer, Waikato Environment for Knowledge Analysis (WEKA), KNIME, R-Programming, Orange, NLTK etc. These tools and software provide a set of methods and algorithms that help in better analysis of data. These tools help in cluster analysis, data visualization, regression analysis, Decision trees, Predictive analytics, Text mining, etc.

In this section we will explain each step of the ML model. and then put the full model at the end of the article for reference.

For deal with the problem, we choose Orange machine learning and data mining software because it is a free open-source machine learning software that people don´t need writing any code. Orange is an open-source machine learning and data mining software written in Python. It has a visual programming front-end for explorative data analysis and visualization and can also be used as a Python library. Orange is a component-based visual programming software for data mining, machine learning and data analysis. Components are called widgets and they range from simple data visualization, subset selection and pre-processing, to empirical evaluation of learning algorithms and predictive modelling (Naik &Saman, 2016).

4.1.1. Dataset

The first step was downloaded the dataset from company ERP and saved it in excel format. The training purchasing dataset sample is summarized in Table 1.

Table 1
Training purchasing order data sample

Purchase Order Value	Purchase Order Quantity	PO in Advance of Delivery	Defect Percent
R$ 820.867,00	536	43	1,43
R$ 156.147,00	674	25	0,34
R$ 328.463,00	846	27	1,57
R$ 650.906,00	736	38	1,06
R$ 661.390,00	648	41	1,37
R$ 67.642,00	741	34	0,11
R$ 49.823,00	517	43	0,09
R$ 413.791,00	802	42	1,54
R$ 237.979,00	615	40	0,26
R$ 71.252,00	867	41	0,37
R$ 710.463,00	295	16	0,37
R$ 710.463,00	295	16	0,03
R$ 13.715,00	577	28	1,26
R$ 939.307,00	820	44	1,03
R$ 856.479,00	333	16	0,91
R$ 354.191,00	817	26	1,43
R$ 221.413,00	606	21	0,34
R$ 319.893,00	794	30	1,57
R$ 155.737,00	352	18	1,06
R$ 895.897,00	473	25	1,37

Source: Company ERP

First, we imported this dataset to orange using the widgets file from widgets from data manipulation in Orange, show in Figure 8.

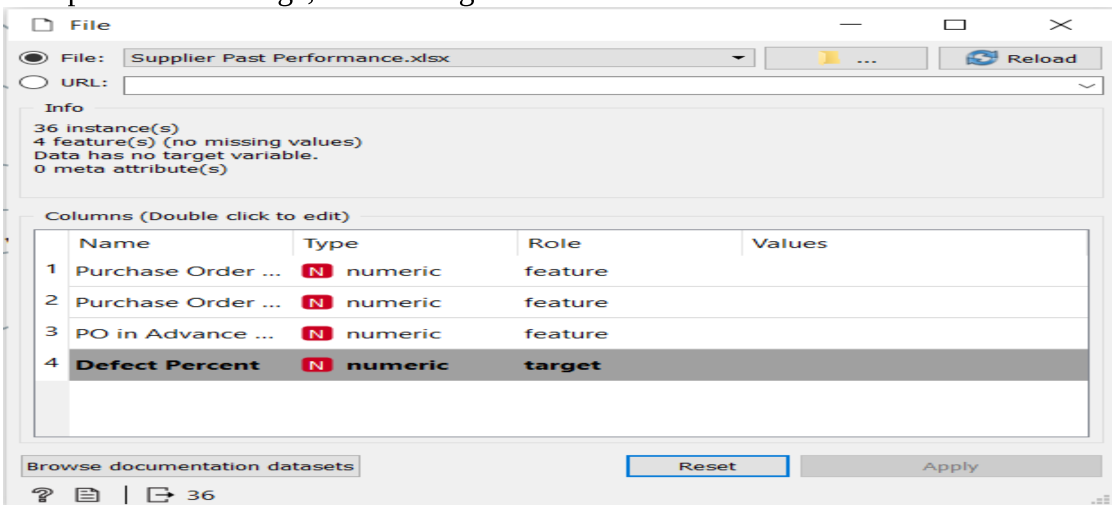

Figure 8 – Data set in Orange data widgets

The dataset has 36 instances with four features: purchase order value, purchase order quantity, po in advance of delivery and defect percent in incoming conference. There is no missing values in dataset. So, all data was using in our machine learning program. As the predict variable is defect percent feature, we need to choose it as target.

Now, we imported this dataset to orange using the widgets data table from the menu data manipulation in Orange, show in Table 2

Table 2
Data table for data visualization

	Defect Percent	Purchase Order Value ($)	Purchase Order QuaNtity	PO in Advance Of Delivery
1	1.43	820867	536	43
2	0.34	156147	674	25
3	1.57	328463	846	27
4	1.06	650906	736	38
5	1.37	661390	648	41
6	0.11	67642	741	34
7	0.09	49823	517	43
8	1.54	413791	802	42
9	0.26	237979	615	40
10	0.37	71252	867	41
11	0.37	710463	295	16
12	0.03	710463	295	16
13	1.26	13715	577	28
14	1.03	939307	820	44
15	0.91	856479	333	16
16	1.43	704104	524	19
17	0.34	76441	271	24
18	1.57	386341	486	16
19	1.06	874261	657	34
20	1.37	672750	807	27
21	0.11	539058	307	23
22	0.09	84740	673	39
23	1.54	759209	843	17
24	0.26	156849	359	17
25	0.37	718231	781	42
26	0.37	544279	292	43
27	0.03	747222	702	23
28	1.26	241531	899	21
29	1.03	173686	870	30
30	0.91	394593	201	20
31	0.10	334482	524	33
32	1.54	181856	300	33
33	0.26	668280	776	36
34	0.37	262389	518	45
35	0.37	103339	893	20
36	0.03	90549	894	32

In the next section we will explain the model algorithms used.

4.1.2. Model algorithms

The next step was choosing the five algorithms selected (Linear Regression (LR), Support Vector Machine (SVM), AdaBoost (AB) and Artificial Neural Networks (ANN)) to predict the supplier defect by delivery, from the widgets model in Orange, as depicted in Figure 9.

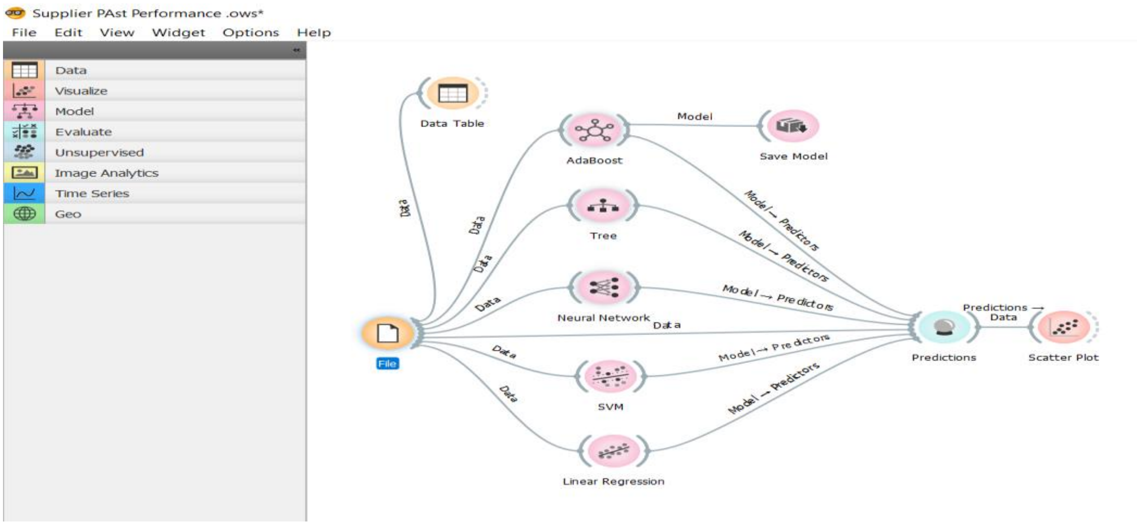

Figure 9 – Predictive algorithms selected

Next, the five algorithms were linked with the data file and with the widget prediction from menu evaluate in Orange. Thus, we can see all five supplier defects predictions five algorithms selected, as show in Table 3.

Table 3
Training defect prediction by algorithms

Predictions

Show probabilities for

	Tree	SVM	Linear Regression	AdaBoost	Neural Network	Defect Percent	:hase Order Value	:hase Order QuaN	ı Advance Of Deli
1	1.13	1.27	0.82	1.43	1.16	1.43	820867	536	43
2	0.80	0.47	0.65	0.34	0.64	0.34	156147	674	25
3	1.56	1.19	0.83	1.57	0.88	1.57	328463	846	27
4	0.90	1.04	0.87	1.06	0.80	1.06	650906	736	38
5	0.90	1.10	0.81	1.37	0.92	1.37	661390	648	41
6	0.08	0.21	0.57	0.11	0.29	0.11	67642	741	34
7	0.23	0.19	0.37	0.11	0.32	0.09	49823	517	43
8	1.56	0.97	0.75	1.54	0.86	1.54	413791	802	42
9	0.21	0.33	0.55	0.26	0.34	0.26	237979	615	40
10	0.23	0.45	0.60	0.37	0.30	0.37	71252	867	41
11	0.20	0.47	0.81	0.20	0.53	0.37	710463	295	16
12	0.20	0.47	0.81	0.20	0.53	0.03	710463	295	16
13	0.80	0.19	0.49	1.26	0.64	1.26	13715	577	28
14	1.31	0.93	1.04	1.03	1.22	1.03	939307	820	44
15	1.13	0.67	0.91	0.91	0.69	0.91	856479	333	16
16	1.50	1.01	0.92	1.43	0.79	1.43	704104	524	19
17	0.32	0.44	0.37	0.34	0.72	0.34	76441	271	24
18	1.50	0.75	0.73	1.57	0.81	1.57	386341	486	16
19	1.13	1.02	0.98	1.06	0.88	1.06	874261	657	34
20	1.31	1.27	1.00	1.37	1.07	1.37	672750	807	27
21	0.21	0.29	0.67	0.11	0.42	0.11	539058	307	23
22	0.08	0.17	0.51	0.26	0.17	0.09	84740	673	39
23	1.31	1.44	1.15	1.54	1.51	1.54	759209	843	17
24	0.32	0.36	0.52	0.26	0.77	0.26	156849	359	17
25	0.20	1.06	0.91	0.37	0.94	0.37	718231	781	42
26	0.94	0.47	0.52	0.37	0.64	0.37	544279	292	43
27	0.20	1.24	1.02	0.03	0.95	0.03	747222	702	23
28	1.15	1.16	0.86	1.26	1.08	1.26	241531	899	21
29	1.15	0.75	0.74	1.03	0.65	1.03	173686	870	30
30	0.94	0.34	0.54	0.91	0.58	0.91	394593	201	20
31	0.21	0.30	0.60	0.10	0.43	0.10	334482	524	33
32	0.94	0.41	0.39	1.54	0.75	1.54	181856	300	33
33	0.90	1.04	0.92	0.26	0.83	0.26	668280	776	36

In the Table 3 the shadow column is the defect percent from dataset and the before column are the prediction for each algorithm.

After applying the supplier defect prediction model on training dataset, we have some results that it will be presented now.

At the beginning of the model building process, the supervised learning algorithms LR, TR, SVM, AB and ANN were trained and parameterized in a coarse parameter optimization on a smaller balanced data sample with 144 data points and validated with a 5-fold cross validation. The achieved results were compared in terms of MSE, RMSE, MAE and R2 (see Table 4).

Table 4
Algorithm's performance

Model	**MSE**	**RMSE**	**MAE**	**R2**
AdaBoost	0.002	0.049	0.015	0.992
Tree	0.064	0.252	0.182	0.799
Neural Network	0.193	0.439	0.376	0.390
SVM	0.213	0.461	0.326	0.329
Linear Regression	0.276	0.525	0.448	0.130

Both RMSE and R2 quantifies how well a regression model fits a dataset. The RMSE tells how well a regression model can predict the value of a response variable in absolute terms while R2 tells how well the predictor variables can explain the variation in the response variable. Even though MAPE, MAE, MSE and RMSE are commonly used in machine learning studies, we showed that it is impossible to detect the quality of the performance of a regression method by just looking at their singular values (Chicco, Warrens & Jurman, 2021). Thus, for comparing the accuracy among different regression models, is RMSE and R2 at the same time. As in table 3, Adaboost has the best performance because the smallest RMSE and the best R2. Finally, we saved the Adaboost model and deployed it in the supplier defect predict. In the next section we will present the results and analysis.

5. RESULTS AND ANALYSIS

Using the best regression model (Adaboost), we tested the data and results are summarized in Table 5.

Table 5

Prediction with data test

Period	AdaBoost	**Purchase Order Value ($)**	**Purchase Order Quantity**	**PO in Advance Of Delivery**
1	0.34	320453.0	223.0	32.0
2	0.34	71252.0	867.0	41.0
3	1.06	710463.0	295.0	16.0
4	1.06	710463.0	295.0	16.0
5	1.26	13715.0	577.0	28.0
6	1.06	939307.0	820.0	44.0
7	1.06	856479.0	333.0	16.0
8	1.06	478402.0	201.0	31.0
9	0.34	163410.0	461.0	15.0
10	0.34	276314.0	273.0	26.0
11	1.06	507904.0	467.0	42.0
12	1.06	484003.0	346.0	26.0

The results of supplier defect prediction test dataset using Adaboost algorithm related to Table 5 is show in Figure 11.

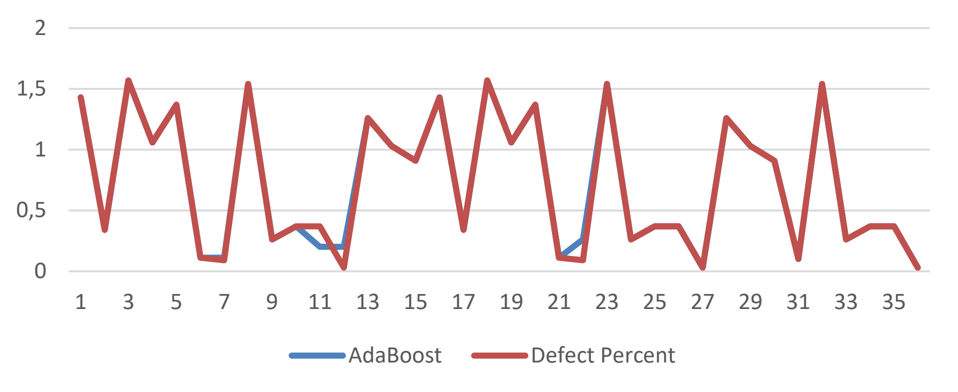

Figure 11 – Supplier Quality Prediction model test results

The graph in Figure 11 show that the model has a high accuracy (95,1%) and because this can be used to predict the supplier defect based on orders. After the validation of the model and approval it, we start to implement in the receiving process. As we already defined that the pilot implementation project will be in purchasing order delivery of engines. So, if defective engines percentage predicted is below the threshold level, then a quality check will not be performed. The initial rule used was consider that if the defect engines percentage in the delivery is more than 0.4%, then explicit incoming inspection is needed to ensure the quality of the engines in the end product. In this step of methodology was the quality defect prediction with new data based in Table 6.

Table 6
Data for predict supplier quality defect

Purchase Order Value	Purchase Order Quantity	PO in Advance of Delivery
BRL$ 320.453,00	2223	32

The application of ML model with Adaboost algorithm using the current supplier purchasing order for engines (see Table 6) resulted in defect prediction of 0.34 %. As this value is bellow of 0.40 % no incoming inspection is needed for this purchasing order.

This ML approach helps company to reducing the incoming inspection about 30% of purchasing orders, and consequently reducing at same level the total cost of supplier products inspection. Thus, this machine learning model allows the company to focus on quality checks on only particular purchase orders delivery (above 0.40 % defect prediction), to control the end-product quality, and optimize inspection costs. This new process also enable company to uncover the variables/parameters which are influencing the defective deliveries of few materials at times and work collaboratively with the supplier to address it.

6. CONCLUSIONS AND TECHNOLOGICAL CONTRIBUTIONS

In this article, we presented how machine learning algorithms can be used to predict the supplier defect percentage quantity in incoming deliveries. To summarize, in this example, we are training the model every time we are running the code to predict the result for new purchasing orders.

With this ML model approach, we helped company to reducing the incoming inspection about 30% of purchasing orders, and consequently reducing at same level the total cost of supplier products inspection.

The main theorical contribution of this technological paper was to present a simple method with five steps to use data analysis and learning method to practical decision related to supplier quality management. Furthermore, the proposed method can be applied in small and medium-sized companies since the investment is low and the model development and implementation needs small code knowledge.

The practical contribution of this work is that the proposed methodology for predicting the quality of incoming components from supplier deliveries problems can serve as a guide for managers to implement or improve different industrial processes in small and medium-sized companies also.

REFERENCES

ABIMO - Associação Brasileira da Indústria de Artigos e Equipamentos Médicos, Odontológicos e Hospitalares e de Laboratórios. (2019). Dados Econômicos 2018. https://abimo.org.br/dados-do-setor/dados-economicos/.

Aguiar, M. C. (2014). Análise de causa raiz: levantamento dos métodos e exemplificação. (Master's thesis). Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial.

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(1), 85.

Berendt, B., Bringmann, B., Fromont, E., Garriga, G., Miettinen, P., Tatti, N., & Tresp, V. (2016). Machine learning and knowledge discovery in databases. In European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases.

Boon-itt, S. (2010). An empirical model of the relationship between manufacturing capabilities: Evidence from the Thai automotive industry. NIDA Development Journal, 59(2), 19–45.
Bustillo, A., Urbikain, G., Perez, J. M., Pereira, O. M., & de Lacalle, L. N. L. (2018). Smart optimization of a friction-drilling process based on boosting ensembles. Journal of Manufacturing Systems, 48, 108-121.
Cheng, Y.-J., Chen, M.-H., Cheng, F.-C., Cheng, Y.-C., Lin, Y.-S., & Yang, C.-J. (2018). Developing a decision support system (DSS) for a dental manufacturing production line based on data mining. Applied System Innovation, 1(2), 17.
Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7, e623.
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, 8-17.
Deng, L., & Li, X. (2013). Machine learning paradigms for speech recognition: An overview. IEEE Transactions on Audio, Speech, and Language Processing, 21, 1060-1089.
Di Bucchianico, A. (2008). Coefficient of determination (R2). In Encyclopedia of Statistics in Quality and Reliability (pp. 291-295). John Wiley & Sons Ltd.
Dogan, A., & Birant, D. (2020). Machine learning and data mining in manufacturing. Expert Systems with Applications, 114060.
Freund, Y., & Schapire, R.E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119-139.
Friedman, J.H. (2002). Stochastic gradient boosting. Computational Statistics & Data Analysis, 38(4), 367-378.
Gandhi, K., Schmidt, B., & Ng, A.H. (2018). Towards data mining based decision support in manufacturing maintenance. Procedia CIRP, 72, 261-265.
Gashi, M., & Thalmann, S. (2020). Taking complexity into account: A structured literature review on multi-component systems in the context of predictive maintenance. Information Systems, Springer International Publishing, 31-44.
Görz, G., Schneeberger, J., & Schmid, U. (2013). Handbuch der Künstlichen Intelligenz. [e-book]. München: Oldenbourg.
Harding, J.A., Shahbaz, M., Srinivas, & A. Kusiak. (2006). Data mining in manufacturing: A review. Journal of Manufacturing Science and Engineering, 128, 969.
Hecker, D., Döbel, I., Petersen, U., Rauschert, A., Schmitz, V., & Voss, A. (2017). Zukunftsmarkt künstliche Intelligenz. Potenziale und Anwendungen. Fraunhofer.
Hirsch, V., Reimann, P., & Mitschang, B. (2019). Data-driven fault diagnosis in end-of-line testing of complex products. 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA), IEEE, 492-503.
Hu, L., Zhou, M., Xiang, F., & Feng, Q. (2018). Modeling and recognition of steel.
Huang, Y.-J., Pan, C.-L., Lin, S.-C., & Guo, M.-H. (2018). Machine-learning approach in detection and classification for defects in TSV-based 3-D IC. IEEE Transactions on Components, Packaging and Manufacturing Technology, 8(4), 699-706.
Imoto, K., Nakai, T., Ike, T., Haruki, K., & Sato, Y. (2018). A CNN-based transfer learning method for defect classification in semiconductor manufacturing. International Symposium on Semiconductor Manufacturing (ISSM), IEEE, 1-3.

International Organization for Standardization. (2015). Quality management systems - Requirements (ISO standard no. 9001:2015). https://www.iso.org/obp/ui/#iso:std:iso:9001:ed-5:v1:en

Jiang, C., Zhang, H., Ren, Y., Han, Z., Chen, K.-C., & Hanzo, L. (2017). Machine learning paradigms for next-generation wireless networks. IEEE Wireless Communications, 24, 98-105.

Kelley, K., & Lai, K. (2011). Accuracy in parameter estimation for the root mean square error of approximation: Sample size planning for narrow confidence intervals. Multivariate Behavioral Research, 46(1), 1-32.

Koufteros, X. A., Vonderembse, M. A., & Doll, W. J. (2002). Examine the competitive capabilities of manufacturing firms. Structural Equation Modelling, 9(2), 256-282.

Kovacic, M., & Sarler, B. (2009). Application of the genetic programming for increasing the soft annealing productivity in steel industry. Materials and Manufacturing Processes, 24, 369-374.

Kujawińska, A., Rogalewicz, M., Muchowski, M., & Stańkowska, M. (2018). Application of cluster analysis in making decision about purchase of additional materials for welding process. Smart Technology, Springer, 10-20.

Kusiak, A. (2006). Data mining: manufacturing and service applications. International Journal of Production Research, 44(19), 4175-4191.

Lee, J., Noh, S.D., Kim, H.-J., & Kan, Y.-S. (2018). Implementation of cyber-physical production systems for quality prediction and operation control in metal casting. Sensors, 18(5), 1428.

Lee, K. B., Cheon, S., & Kim, C. O. (2017). A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes. IEEE Transactions on Semiconductor Manufacturing, 30(2), 135-142.

Liaw, A., Wiener, M., et al. (2002). Classification and regression by randomforest. R News, 2(3), 18-22.

Lim, H.K., Kim, Y., & Kim, M.-K. (2017). Failure prediction using sequential pattern mining in the wire bonding process. IEEE Transactions on Semiconductor Manufacturing, 30(3), 285-292.

Lingitz, L., Gallina, V., Ansari, F., Gyulai, D., Pfeiffer, A., Sihn, W., & Monostori, L. (2018). Lead time prediction using machine learning algorithms: A case study by a semiconductor manufacturer. Procedia CIRP, 72, 1051-1056.

McGrath, R. G. (2013). Transient advantage. Harvard Business Review, 91(6), 62-70.

Mohammed, M., Bashier, E. B. M., & Khan, M. B. (2017). Machine learning: Algorithms and applications. CRC Press, Boca Raton.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). Foundations of machine learning. Cambridge, MA: The MIT Press. Retrieved from https://pdfs.semanticscholar.org/e923/9469aba4bccf3e36d1c27894721e8dbefc44.pdf.

Naik, A., & Samant, L. (2016). Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime. Procedia Computer Science, 85, 662-668.

Priore, P., Ponte, B., Puente, J., & Gómez, A. (2018). Learning-based scheduling of flexible manufacturing systems using ensemble methods. Computers & Industrial Engineering, 126, 282-291.

Ren, L., Sun, Y., Cui, J., & Zhang, L. (2018). Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. Journal of Manufacturing Systems, 48, 71-77.
Ren, L., Sun, Y., Cui, J., & Zhang, L. (2018). Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. Journal of Manufacturing Systems, 48, 71-77.
Robinson, C. J., & Malhotra, M. K. (2005). Defining the concept of supply chain quality management and its relevance to academic and industrial practice. International Journal of Production Economics, 96(3), 315-337.

Robinson, C. J., & Malhotra, M. K. (2005). Defining the concept of supply chain quality management and its relevance to academic and industrial practice. International Journal of Production Economics, 96(3), 315–337.

Sammut, C., & Webb, G. I. (2010). Mean absolute error. In Encyclopedia of Machine Learning (pp. 652).

Sammut, C., & Webb, G. I. (2010). Mean squared error. In Encyclopedia of Machine Learning (p. 653).

Schmidt, B., & Wang, L. (2015). Predictive maintenance: literature review and future trends. In The International Conference on Flexible Automation and Intelligent Manufacturing (FAIM), 23–26 June 2015, vol. 1, University of Wolverhampton, UK, The Choir Press, pp. 232-239.

Shao, S.-Y., Sun, W.-J., Yan, R.-Q., Wang, P., & Gao, R. X. (2017). A deep learning approach for fault diagnosis of induction motors in manufacturing. Chinese Journal of Mechanical Engineering, 30(6), 1347-1356.

Slack, N., Chambers, S., & Johnston, R. (2002). Administração da Produção (2nd ed.). São Paulo: Atlas.

Syafrudin, M., Alfian, G., Fitriyani, N. L., & Rhee, J. (2018). Performance analysis of iot-based sensor, big data processing, and machine learning model for real-time monitoring system in automotive manufacturing. Sensors, 18(9), 2946.

Syed-Yusof, S. M., Sun, W.-J., Yan, R.-Q., Wang, P., & Gao, R. X. (2017). A deep learning approach for fault diagnosis of induction motors in manufacturing. Chinese Journal of Mechanical Engineering, 30(6), 1347-1356.

Thalmann, S., Gursch, H., Suschnigg, J., Gashi, M., Ennsbrunner, H., Fuchs, A. K., ... & Kappl, G. (2019). Cognitive decision support for industrial product life cycles: A position paper. In COGNITIVE 2019: The Eleventh International Conference on Advanced Cognitive Technologies and Applications (pp. 3-9).

Thalmann, S., Gursch, H., Suschnigg, J., Gashi, M., Ennsbrunner, H., Fuchs, A. K., Schreck, T., Mutlu, B., Mangler, J., Kappl, G., et al. (2019). Cognitive decision support for industrial product life cycles: a position paper. COGNITIVE 2019: The Eleventh International Conference on Advanced Cognitive Technologies and Applications, 3-9.

Wang, K.-S. (2013). Towards zero-defect manufacturing (zdm)-a data mining approach. Advances in Manufacturing, 1(1), 62-74.

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, 104258.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115, 213-237.

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).