Journal of Open Innovation: Technology, Market, and Complexity 11 (2025) 100559

Contents lists available at ScienceDirect

Journal of Open Innovation: Technology, Market, and Complexity

journal homepage: www.sciencedirect.com/journal/journal-of-open-innovation-technology-market-and-complexity

Predictive exploratory data analysis of shopfloor CNC machine operation through a machine learning model

Tarique Ameer, Omid Fatahi Valilai [1,*]

School of Business, Social & Decision Sciences, Constructor University Bremen, Campus Ring 1, Bremen 28759, Germany

ARTICLE INFO

Keywords:
Industry 4.0
Predictive model
CNC machine operation
Exploratory Data Analysis

ABSTRACT

Industry 4.0 technologies have provided industries and firms with enormous capacities for data exchange, especially shopfloor sectors. This digitized data exchange process is accumulated exponentially and is necessitating mechanisms for adding value to the business firms. This paper has focused on advanced analytic techniques and machine learning algorithms to predict and classify the problems of the manufacturing environment in connection with the supply chain. The research explores the capabilities of exploratory data analysis in the manufacturing shop floor at the machine level. The findings of this research try to provide:

- Exhibiting the ML algorithms in the manufacturing shop floor for predictive analysis.
- Effective creation of value across the supply chain process by applying predictive analysis.

The paper has conducted a literature review around industry 4.0 trends and the trend for predictive models for shop floor level machine operation with a focus on historical data repositories. A framework has been developed to create value across the supply chain process using data analytics and integrated business decision-making. A shop floor dataset related to the CNC milling machine has been considered for illustrating the framework capabilities which include machining experiments in the CNC milling operation besides the supply chain related data repositories.

1. Introduction

Contemporary Industry 4.0 technologies are directing industries and organizations towards automation and the exchange of data across various sectors, with a primary focus on manufacturing (Alshuaibi et al., 2024; Bussacarini et al., 2023). As a consequence of digitization, there is an exponential accumulation of vast quantities of data, which does not inherently add value to organizations (Miraj et al., 2024). Analytical processes must be applied to these data to generate value by facilitating self-optimization in terms of cost, availability, and resource consumption (Rezapour Niari et al., 2023; Wang, 2018). Based on a survey conducted among 92 manufacturing enterprises aimed to explore the implementation patterns of Industry 4.0 technologies, it has been concluded that a minimal implementation of big data and analytics has been considered by almost all of the studied firms (Frank et al., 2019; Olad and Fatahi Valilai, 2020). Furthermore, the existing analytic networks that gather sensor data from manufacturing shop floor arrangements via IoT lack real-time analytics have been facilitated for interoperability and integrated data analytics (Houshmand and Valilai, 2013; Zaringhalam et al., 2024) and enabling the supply chain level operation planning (Verma et al., 2017; Zhang et al., 2024).

A significant barrier to the adoption of Industry 4.0 technologies and big data analytics in manufacturing is the insufficiency of conventional data analysis methods currently being utilized (Khedr and S, 2024). It is essential to implement advanced analytical techniques and machine learning (ML) algorithms to predict and categorize issues arising within the manufacturing environment for both financial (Chokwitthaya et al., 2020; Molnár-Major and Bóna, 2024) and environmental planning (Fatahi Valilai and Sodachi, 2020). This paper offers a succinct overview of all nine industry pillars, Industry 4.0 Technologies, with a particular

* Corresponding author.
E-mail address: OFatahiValilai@Constructor.University (O. Fatahi Valilai).
[1] 0000-0001-7087-6946

https://doi.org/10.1016/j.joitmc.2025.100559
Received 29 October 2024; Received in revised form 10 May 2025; Accepted 24 May 2025
Available online 6 June 2025

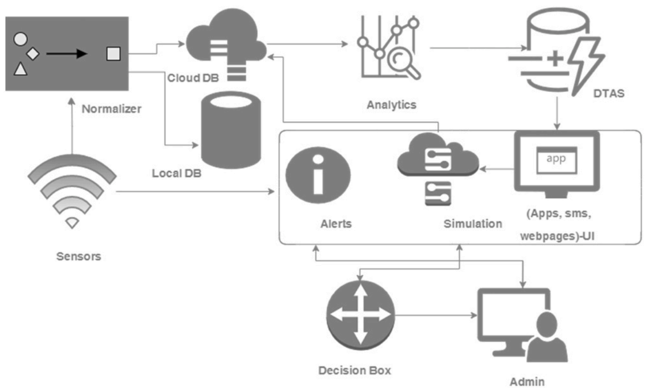

Fig. 1. BDA architecture in industry 4.0.

Fig. 2. ML algorithm for predictive analysis.

focus on big data and analytics (Bussacarini et al., 2023). The principal motivations for conducting this research can be outlined as follows:

- Investigating the potential of data analytics for exploratory data analysis within the manufacturing shop floor at the machine level.
- Demonstrating the application of machine learning algorithms on the manufacturing shop floor for predictive analysis, facilitating the optimization of value creation across the supply chain by systematically analysing data insights and integrating these insights into proactive and comprehensive business decision-making strategies.

2. Literature review

The German government has planned for the platform "*Plattform-i40*" with a 2030 vision for industry 4.0 to bring the digital ecosystem globally capabilities of autonomy, interoperability, and sustainability to illustrate the fundamental process of innovation and transformation in industrial production (Niankara, 2024; Sautter, 2021). This has shaped the building blocks of industry 4.0 encompassing smart sensors, controls, connectivity, and a smart factory (Kumar and Nayyar, 2019). The smart factory requires intelligent manufacturing setup in the context of industry 4.0 consists of sensors, actuators, robots, industrial computers, wireless devices, switches, CPS (Cyber-Physical System), and industrial communication networks that generate big data (Delaram et al., 2021; Erboz, 2017). The proficient exploitation of data derived from the production unit facilitates real-time decision-making. Technological advancements and the abundant availability of data confer a competitive advantage upon manufacturing industries by integrating innovation and enhancing productivity (Khedr and S, 2024). Big data initiatives address the majority of challenges at the organizational level by enabling superior monitoring, measurement, and management. (Vaidya et al., 2018).

BDA (Big Data and Analytics) architecture in industry 4.0 collects the data from the sensors. The sensor's data are normalized in a standard format to remove inconsistencies (Clausen, 2023; Pargmann et al., 2018). The normalized data are stored in the two databases, the local database inside the plant and the master database in the cloud. The data from cloud DB (Database) is passed through analytics, including data preprocessing, trained ML algorithms, or AI (Artificial Intelligence) models. The results derived from the analytics dictate subsequent actions. The Data Transfer and Alert System (DTAS) conveys the outcomes of the analysis to the administrator via website alerts, mobile applications, or SMS. As depicted in Fig. 1, the administrator evaluates the analysis results and implements suitable measures. The administrator's decision is subjected to validation through a simulation model to ensure the desired result. Should the system's condition within the simulation model prove unstable, the administrator must make alternative decisions to achieve stabilization. In instances of decision failure, the data is relayed back to the cloud database, wherein the machine learning algorithm or artificial intelligence model undergoes retraining to prevent recurrence of the same step in the future. (Sharma and Jain, 2019). There are Case oriented discussion of AI/ML methods, data-quality hurdles, and synthetic-data augmentation for PdM (Product data management) that stresses the necessity of rigorous EDA before model building. For example, (Cheng et al., 2022) have synthesised 37 PdM studies, mapping where EDA, anomaly detection, Remaining Useful Life (RUL) estimation and Explainable Artificial Intelligence (XAI) intertwine and highlighted missing frameworks that jointly exploit data driven and knowledge driven approaches. But, there are rare studies like this paper which has tried to integrate EDA and ML frameworks at machine level.

Moreover, EDA (Exploratory Data Analysis) can be assumed to be applied for more detailed and operational prediction (Marodin et al., 2023). EDA helps to maximize the value of data and is necessary to guide and develop research programs (Maass et al., 2018). Univariate EDA includes box plots, dot plots, histograms, and kernel density plots to find the distribution of a single variable. Multivariate EDA helps analyse a large number of variables to detect the trends and clusters. Multivariate EDA incorporates other graphical techniques, for instance, star plots, block plots, parallel coordinate plots, biplots, and spaghetti plots (Sagawa et al., 2024). One of the applications for EDA can be TCM (Tool Condition Monitoring) which is the process of finding tool wear and characterizing it (Elminir et al., 2024; Vijay et al., 2021).

Exploring the EDA initiatives in the literature, (Strazzullo et al., 2024) conducted empirical interviews with 16 practitioners reveal that Industry 4.0 technologies improve data management and time-to-market while also posing capability and security challenges; the authors explicitly referred to EDA concept and highlighted the opportunities for data-driven research. Moreover, (Aspiranti et al., 2023) conducted another Survey of 425 bank executives shows big-data analytics significantly strengthens partner-selection and collaboration-formalisation capabilities within open innovation strategies, offering a transferable model for other sectors exploring data-powered EDA initiatives. In terms of Industry-specific EDA applications, (Rodriguez et al., 2023) used unsupervised EDA including box-plots and K-means tools on turbine SCADA (Supervisory Control and Data Acquisition) streams to surface outliers and hidden fault patterns, enabling data-driven PdM without labelled data which is a direct demonstration of EDA value in discrete-equipment environment. However, the specific application for CNC domain has been an interesting but unapproached domain for this filed. In another related study, (Urbani and Collan, 2020) presented two real-world additive manufacturing cases and a predictive analytics maintenance vision which illustrated how cross-linking EDA with production and supply chain data can unlock new service business models.

Tool Condition Monitoring (TCM) encompasses the handling of process parameters, signal analysis, and decision-making processes. Initially, process parameters are utilized to evaluate data with the aim of extracting tool condition characteristics. Subsequently, these derived features facilitate the assessment of the tool's health status. In TCM, rigorous monitoring is crucial owing to the significance of the signal processing involved, which can be categorized as either direct or indirect methods. In the direct method, tool wear is measured directly; however, the online monitoring of such wear poses significant challenges. (Mathiasen and Clausen, 2024). After the data from sensing of process parameters, preprocessing is applied according to the signal characteristics. Next, the crucial features from these signals are

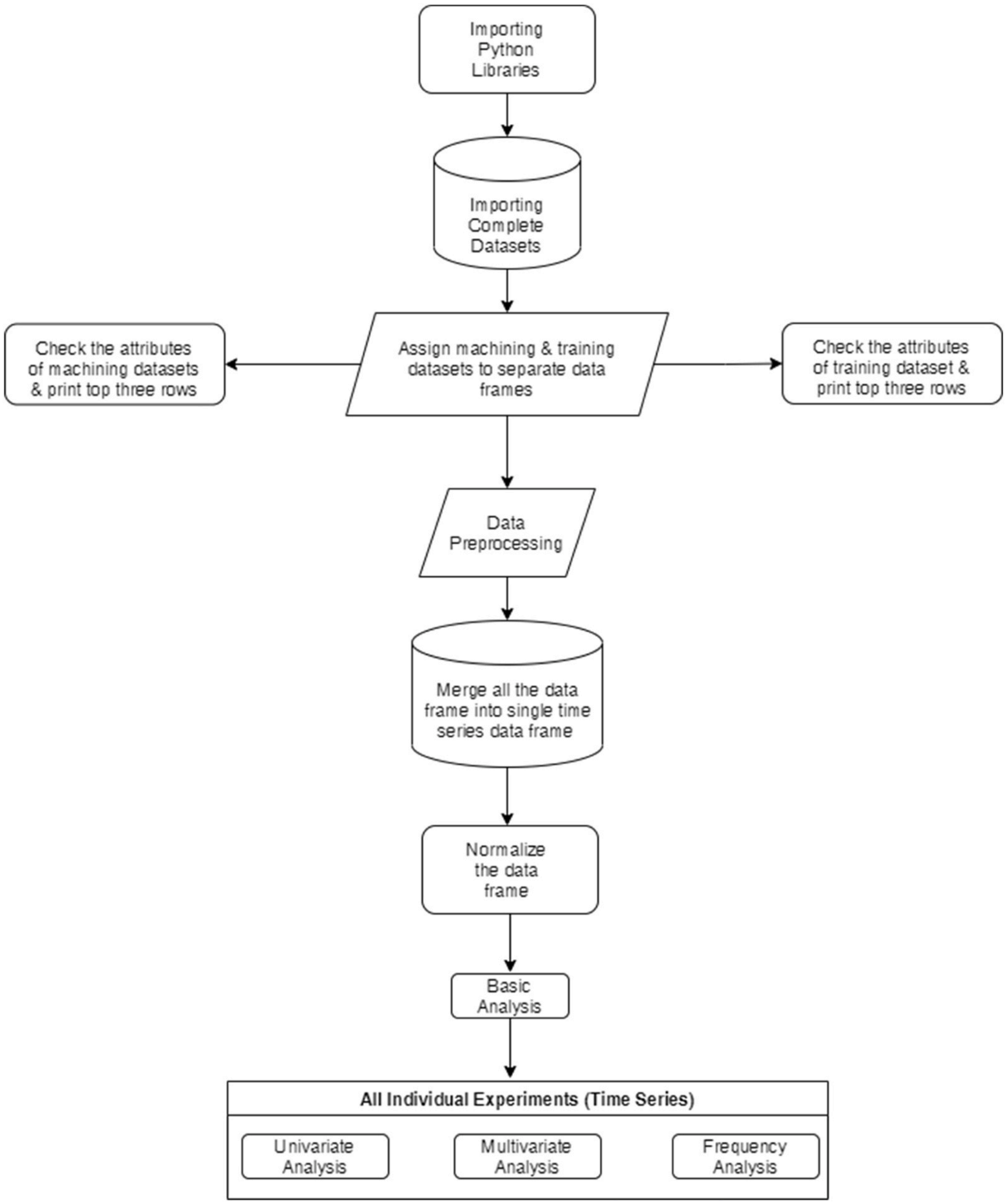

Fig. 3. EDA model designed for exploring shopfloor data.

extracted to access the tool condition. In the end, decision-making supports gaining the proper knowledge about the tool conditions by utilizing fuzzy and neural networks (Pozzi et al., 2024). This can lead to the creation of VVC (Virtual Value Chain) building on Porter's value chain (Mantravadi et al., 2023). The VVC model consists of five steps for the value creation process and enables the value tracing through SC (Supply Chain) (Hasan et al., 2024). While the methodologies for applying Exploratory Data Analysis (EDA) to operational planning have been addressed, the incorporation of these outcomes into higher-level consolidated operations planning, such as supply chain planning, has been neglected, despite recent recommendations for more realistic planning approaches (Sodachi et al., 2024).

3. Proposed solution for predictive exploratory data analysis

This paper introduces a framework development model that elucidates the concept of value creation within the Supply Chain (SC) through the analysis of data insights and the incorporation of these insights into the proactive and integrated business decision-making processes of a manufacturing company. The framework presupposes the existence of Big Data Analytics (BDA) architecture and an IT landscape characterized by horizontal and vertical integration within the organization. To derive insights from data, pertinent data analysis and machine learning (ML) techniques are employed to identify the manufacturing context of the industry, with particular emphasis on sales, production planning and scheduling, inventory planning and warehouse management, and purchasing. In constructing ML algorithms, BDA first undertakes data preprocessing, such as data cleaning or redundancy elimination. The subsequent phase involves feature extraction, which entails deriving meaningful information from a vast array of data. Feature extraction is a critical step, as inappropriate feature selection can lead to the development of a defective model. The next step is to train the model based on extracted features to obtain the desired outcome. Data engineers and data scientists mainly use predictive analytics, prescriptive analytics, and ML algorithms for predictive and pattern discovery. ML algorithm consists of training the model with a predefined dataset to predict the trends in the future as illustrated in Fig. 2.

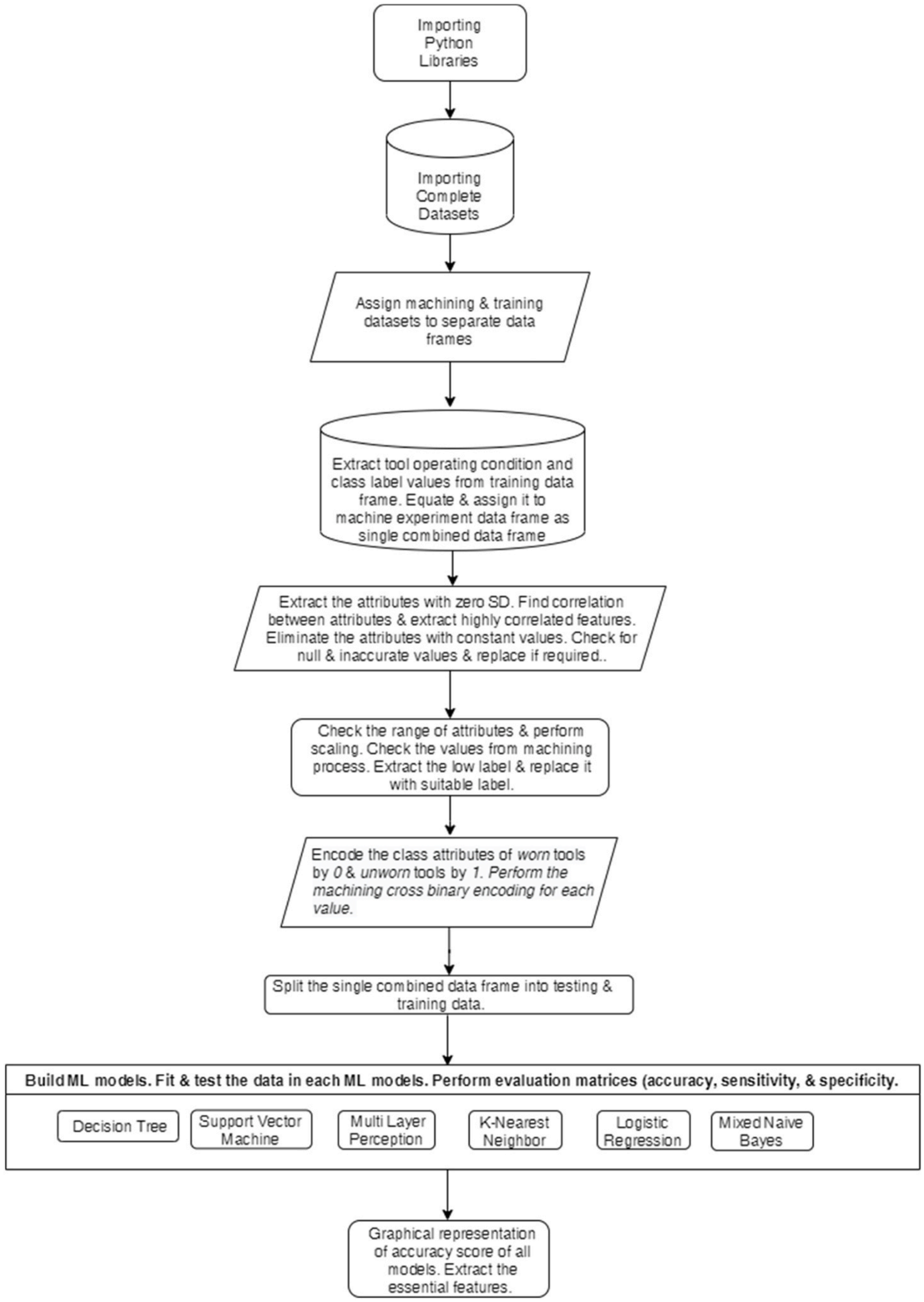

Fig. 4. The steps to develop the ML model and interpretation of accuracy score.

Table 1
Weekly product demands.

Products	*Day #1*	*Day #2*	*Day #3*	*Day #4*	*Day #5*	*Day #6*
Current Demand (item)	*107.777*	*10.506*	*33.308*	*108.161*	*57.492*	*12.950*
Forecasted demand (item)	*118.555*	*11.556*	*36.639*	*118.977*	*63.241*	*14.245*

Table 2
Sales decision.

Customer	*Agreements*	*Current Decision*	*Decision after Analysis*
Customer#1	*Service level (%)*	*95 %*	*98 %*
	Order deadline	*20:00 pm*	*14:00 pm*
	Payment Terms	*4 weeks*	*1 week*
Customer#2	*Service level (%)*	*95 %*	*97 %*
	Order deadline	*20:00 pm*	*20:00 pm*
	Payment Terms	*4 weeks*	*1 week*
Customer#3	*Service level (%)*	*95 %*	*97.5 %*
	Order deadline	*20:00 pm*	*17:00 pm*
	Payment Terms	*4 weeks*	*1 week*

Data analytics helps to analyse sales data either by using basic analytics tools or advanced ML algorithms. The historical sales data helps observe trends, customer buying patterns, customers, and product segmentation. After gaining insights and predicting the weekly demand for

Table 3
SCM decisions.

Components	Current Decision		Decision after Analysis	
	Safety Stock (weeks)	Lot size (weeks)	Safety Stock (weeks)	Lot size (weeks)
Pack	2	4	1.8	4.0
Raw Part	2	4	1.6	0.7
Products	Production Interval (days)	Safety Stock (weeks)	Production Interval (days)	Lot size (weeks)
CNC Slot pattern# 1	10	3	3	2
CNC Slot pattern# 2	10	3	4	2
CNC Slot pattern# 3	10	3	5	2
CNC Slot pattern# 4	10	3	3	2
CNC Slot pattern# 5	10	3	4	2
CNC Slot pattern# 6	10	3	5	2

Table 4
Operations decisions.

		Current Decision	Decision after Analysis
Inbound (Raw material warehouse)	Number of Pallet location	900	350
	Number of permanent employees	5	2
	Intake Time (Days)	5	1
Outbound (Finished goods warehouse)	Number of Pallet location	1500	1000
	Number of permanent operators	5	4
Loading operation	Early Inspection (All parts assign to same checking line)	Mixture with a capacity of 300 items in an hour	Mixture with a capacity of 300 items in an hour
Machining operation	Preventive maintenance	No implementation	Implemented
	Training to solve breakdown by employees	No	Yes
Packaging capacity	Conveyer line (Capacity of 300 Packages per hour)	Two shifts (one shift is forty working hours in a week)	Three shifts
	Technology to reduce changeover time	No implementation	Implemented
	Technology to increase the speed of packaging	No implementation	Implemented

all products from sales datasets, the capacity of the current production facility can be accessed. Then, by deciding appropriate production intervals, the lot size, and the safety stock of components and products, production can be optimized remarkably, which eventually helps in adding value for the organization. Analysing historical sales and active inventory datasets using K-mean clustering – an unsupervised machine learning technique helps categorize components and products.

Data insights from sales, production planning and scheduling, inventory planning and warehouse management, and purchasing need to be shared with proper coordination and collaboration for pro-active integrated business decision-making (Delaram et al., 2022). An EDA model is designed to explore the dataset by performing basic, univariate, multivariate, and frequency analysis. The basic analysis as illustrated in Fig. 3, includes machine inputs and outputs; machine input visualizes the distribution of feed rate, clamp pressure, and material count. The machine output looks at the tool wear count (worn or unworn), machine finalized count (yes or no), passed visual inspection (yes or no). The univariate analysis involves machining process counting and observes the actual velocity, feedback current, DC bus voltage of the spindle, X, Y, and Z axes parts. The multivariate analysis reflects the tool condition (worn or unworn) along with the feed rate, clamp pressure, velocity, current, voltage of spindle, X, Y, and Z axes of parts. The multivariate analysis also shows the machine finalized (yes or no) along with the feed rate, clamp pressure, DC bus velocity, feedback current, voltage of spindle, X, Y, and Z axes of parts. The frequency analysis reveals an amplitude variation in the frequency with actual velocity, feedback current, and DC bus voltage of spindle, X, Y, and Z axes of parts.

To construct the machine learning (ML) models depicted in Fig. 4, the combined dataset was initially partitioned into training and testing subsets. A total of 80 % of the dataset was allocated for the training of the ML models. Following the training phase, the model's compatibility with the testing data was assessed. Evaluation metrics for all the models were conducted to ascertain accuracy, sensitivity, and specificity scores. Subsequently, the accuracy of each model was analysed, and the key features were extracted.

4. Experiments and results

To validate the model capabilities, a shop floor dataset related to the CNC milling machine has been used from Kaggle© (Tnani et al., 2022). This dataset includes eighteen different machining experiments with a sampling rate of 100 ms performed on wax blocks in the CNC milling machine. Each experiment produced a finished wax part in S shape. The data was collected using the Rockwell cloud collector agent elastic software from the CNC milling machine in the System-level Manufacturing and Automation Research Testbed (SMART) at the University of Michigan. The general data from all eighteen experiments are given in the training data CSV file that includes experiment numbers, material name, feed rate, and clamp pressure as input features and output predictions are the tool condition, completed machining, and passed visual inspections. The available attributes in the machining experiment from one to eighteen are actual position, actual velocity, actual acceleration, command position, command velocity, command acceleration, current feedback, DC bus voltage, output current, and output voltage of spindle and X, Y, and Z-axis of parts. The command refers to the reference value. The machine experiments consist of attributes measured from four motors present in the CNC machine in a spindle, X, Y, and Z axes.

All experiments draw on the open "CNC _Machining" benchmark published by Bosch Research (Tnani et al., 2022) and mirrored at the University of California Irvine (UCI) Machine-Learning Repository (Dataset ID #752). The archive contains 2 700 labelled files (HDF5 format) that were captured on three brown-field machining centres (M01-M03) across 15 process plans (OP00–OP14) and six semi-annual timeframes. Each file records a tri-axial accelerometer trace sampled at 2 kHz yielding an array of shape (N samples × 3 axes) and is tagged at file level with the binary health label good/bad. The Feature expansion for the present study is conducted by including the controller log that accompanies each cut and extracted process variables ranging from Machine inputs (feed-rate, spindle speed command, clamp pressure, coolant state) and Axis-level traces (actual/command velocity, servo current feedback and DC-bus voltage for X, Y, Z and the spindle) and Discrete quality flags (tool_condition(worn, unworn), finalised (yes, no)).

The applied ML algorithms to analyse ML models have been decision tree, support vector machine, multi-layer perception, k-nearest neighbour, logistics regression, and mixed naive bayes model. Python© 3 notebook supports ML algorithm and builds ML models in which

Table 5
Purchasing decisions.

Procured Raw Materials	*Number of available Suppliers (Worldwide)*	*Supplier Evolution Criteria*	*Current SupplierAgreement*	*Selected SupplierAgreement*
Pack	***Five***	***Quality***	*High*	*High*
		Lead – time	*15 Days*	*5 Days*
		Agreed Delivery – Reliability	*94 %*	*95 %*
		Trade Unit	*Pallet based on a lot size*	*FTL*
		Delivery Window	*4 h*	*1 Day*
		Certification	*No*	*Yes*
		Sustainability Factor	*Low*	*High*
		Associated Risks	*High*	*Low*
		Payment Terms	*4 Weeks*	*8 Weeks*
		Transportation cost	*High*	*Low*
		Supplier Base Price	*Low*	*High*
Raw Part	***Three***	***Quality***	*Poor*	*High*
		Lead – time	*10 Days*	*5 Days*
		Agreed Delivery – Reliability	*94 %*	*95 %*
		Trade Unit	*Pallet based on a lot size*	*Pallet based on a lot size*
		Delivery Window	*4 h*	*1 Day*
		Certification	*No*	*Yes*
		Sustainability Factor	*Low*	*High*
		Associated Risks	*High*	*Low*
		Payment Terms	*4 Weeks*	*8 Weeks*
		Transportation cost	*High*	*Low*
		Supplier Base Price	*Low*	*High*
Sticker	***Seven***	***Quality***	*High*	*High*
		Lead – time	*30 Days*	*10 Days*
		Agreed Delivery – Reliability	*94 %*	*98 %*
		Trade Unit	*Tank (with a capacity of 30000 liters)*	*Tank*
		Delivery Window	*1 Day*	*2 Days*
		Certification	*Yes*	*Yes*
		Sustainability Factor	*Low*	*High*
		Associated Risks	*High*	*Low*
		Payment Terms	*4 Weeks*	*8 Weeks*
		Transportation cost	*High*	*Low*
		Supplier Base Price	*Low*	*High*
Stripe	***Four***	***Quality***	*High*	*High*
		Lead – time	*15 Days*	*10 Days*
		Agreed Delivery – Reliability	*92 %*	*98 %*
		Delivery Window	*1 Day*	*2 Days*
		Certification	*Yes*	*Yes*
		Sustainability Factor	*Low*	*High*
		Associated Risks	*High*	*Low*
		Payment Terms	*6 Weeks*	*8 Weeks*
		Transportation cost	*High*	*Low*
		Supplier Base Price	*Low*	*High*
		Transportation cost	*High*	*Low*
		Supplier Base Price	*Low*	*High*

relevant python libraries can be used as numpy, pandas, matplotlib, seaborn, os, graphviz, sklearn (preprocessing, model_selection, tree, svm, linear_model, neighbors, and matrices), and mixed_naive_bayes. The dataset has been consolidated and all eighteen machining experiments datasets have been assigned to a single data frame and assign training data to a different data frame. The tool operating condition has been extracted from the training data frame and assigned class label values.

The operating conditions and class label values from the training data frame were aligned and integrated with the machining experiments' data frame to form a unified combined data frame. During the preprocessing stage, attributes exhibiting zero standard deviation were identified and removed. The correlation among attributes was assessed, leading to the extraction of highly correlated features, and attributes with constant values were also eliminated. Subsequently, data were examined for the presence of null and inaccurate values. Following the normalization of the machining experiment data frame, labels characterized by low-value outputs in the machining process were identified and appropriately replaced. The subsequent step involved encoding the class attributes, assigning a value of 0 to worn tools and 1 to unworn tools, and applying binary cross encoding across each value.

To validate the results from ML, the paper has benchmarked every model against two baselines: a majority-class classifier that always predicts "worn" (class prevalence = 52 %), and a stratified random guesser that respects class priors. The majority baseline yields Accuracy = 0.52, Precision = 0.52, Recall = 1.00, Specificity = 0.00, F1 = 0.68 and AUC = 0.50; the random baseline centres on Accuracy ≈ 0.50 and AUC ≈ 0.50. McNemar tests and paired t-tests on Accuracy and F1 across the five folds show that the decision tree's improvements over both baselines are statistically significant at $p < 0.001$, and its margin over the runner-up MLP is significant at $p < 0.05$.

To embed the SC scenario, the business simulation game "Fresh Connection dataset" has been used to create value across the supply chain (MarcusDoubleYou, 2023). The simulation provides an opportunity to manage a virtual enterprise, designated as TFC (The Fresh Connection). The enterprise comprises four primary departments: purchasing, operations, supply chain management (SCM), and sales. An analysis of the current weekly demand of products over a six-month period is conducted, considering promotional offers and horizons. Consequently, a ten percent increase in sales across all products to each customer is projected. The projected demand estimations are communicated to the SCM department to determine feasible lot sizes and to maintain the requisite safety stock of raw materials and finished goods. The operations department examines the capacity of the existing mixing and bottling lines, as well as the inbound and outbound warehouse pallet allocation. The purchasing department identifies potential

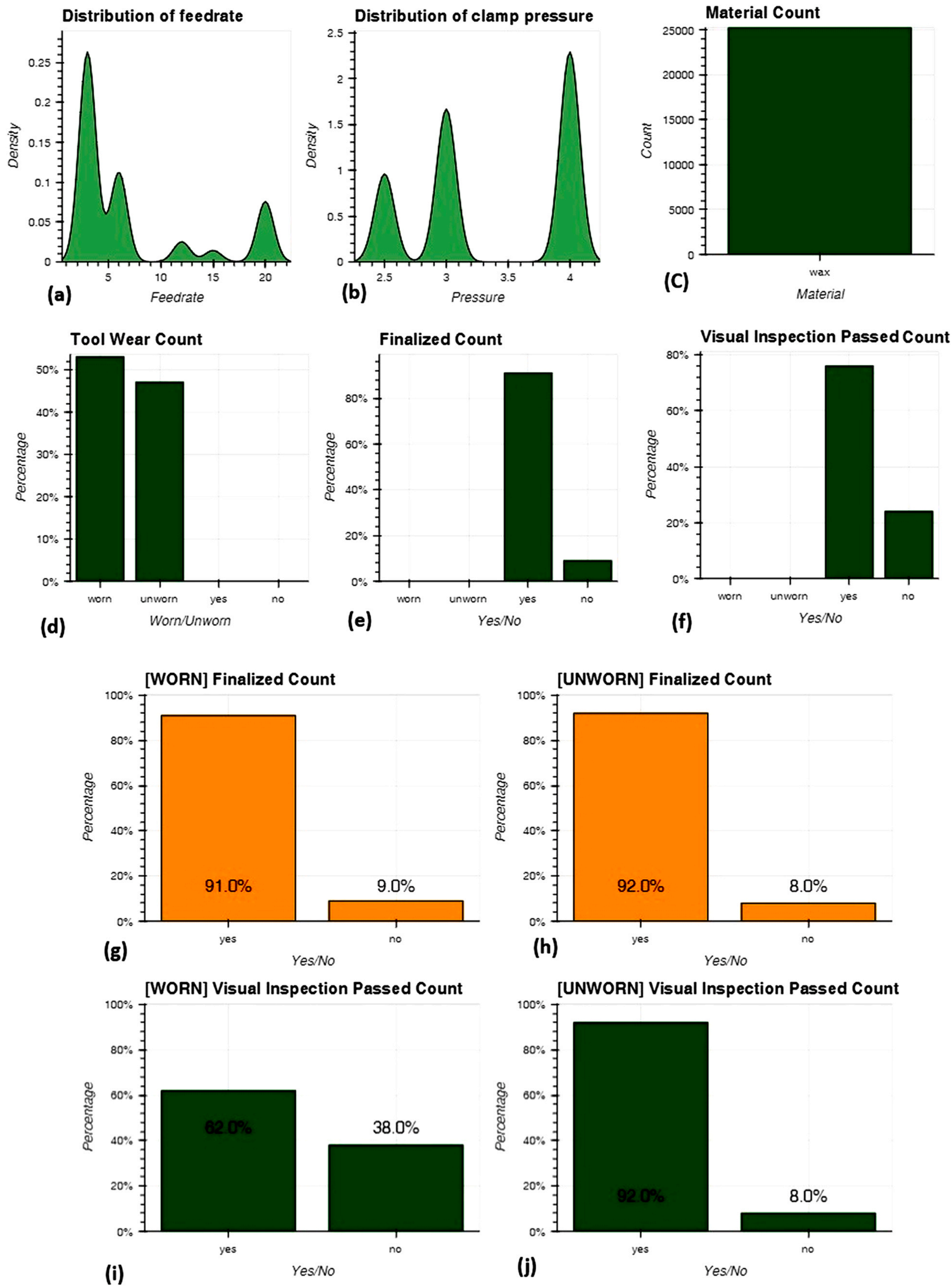

Fig. 5. EDA and the basic analysis for CNC machine tool (a) feedrate effects on density; (b)Clamp pressure effects on density; (c) total machine input; (d) percentage of tool worn scenarios; (e) finalized machining cases; (f) visual inspection quality check; prediction of worn (g) and unworn (h) cases; visual inspection for quality control over (I) worn and (j) unworn samples.

Table 6
Detailed EDA analysis of CNC machine output and tool.

Item	Observation	Technical interpretation	Operational & SCM implications
(a) Feed-rate density	A strongly right-skewed distribution; ~60 % of cuts occur below ≈ 5 mm s^{-1} with diminishing use of mid-range and very sparse high-speed cuts (> 15 mm s^{-1}).	Operators favour conservative settings consistent with fine finishing passes in wax and only sporadically invoke aggressive roughing modes. The minor modes at 8 – 12 mm s^{-1} hint at programme blocks dedicated to semi-finishing.	A narrow effective process window simplifies cycle-time forecasting and inventory buffering. However, the sporadic high-speed bursts align with the current/voltage spikes seen in Fig. 6 and deserve targeted monitoring.
(b) Clamp-pressure density	Three well-separated modes centred ≈ 2.7, 3.3, 4.1 MPa, the latter carrying the highest probability mass.	Distinct fixturing recipes are in use, probably linked to work-piece variants. The dominant 4 MPa peak confirms the reviewers' observation of a "high density of high-pressure clamp."	Because clamping strategy is discretised, it can be encoded as a categorical variable in the ML model, improving interpretability. High-pressure fixtures correlate with improved dimensional accuracy but also higher tool stresses reinforcing the preventive-maintenance focus.
(c) Material count	Entire sample (≈ 25 000 parts) is wax.	Dataset representing prototyping and pattern making.	Stock keeping for raw material is trivial, letting SCM decisions centre on WIP, pallets, and operator allocation (Table 3 & 4).
(d) Tool-wear status	Slight edge for 'worn' (~52 %) over 'unworn' (~48 %).	Data collection deliberately balanced around the wear boundary to avoid class imbalance during model training.	Maintenance policy can be tuned because the historical data span both healthy and degraded regimes almost evenly.
(e) Finalised parts	~90 % of scheduled parts complete machining.	High machine availability and few emergency stops; corroborates the reduced pallet/ operator count in Table 4.	High completion reliability reduces cycle-stock levels and frees pallet slots hence the observed pallet reduction.
(f) Visual-inspection pass-rate	≈ 78 % pass, 22 % fail.	Quality risk still present and inspection is the current gatekeeper.	With a predictive wear model in place, some manual inspections could be replaced or sequenced, further cutting takt time.
(g) & (h) Finalisation split by wear	Pass rates almost identical: 91 % (worn) vs 92 % (unworn).	Tool wear does not raise the probability of an in-process stoppage. Failures during cutting are driven more by fixture or NC errors.	Maintenance scheduling can therefore focus on post-process quality impact rather than machine stop avoidance.
(i) & (j) Visual-inspection pass-rate split by wear	Sharp divergence: only 62 % of 'worn tool' parts pass, versus 92 % for 'unworn'.	Clear causal link that worn tools degrade surface finish/dimensions enough to fail inspection, even though machining completes.	The 30-point drop justifies condition-based tool replacement rather than time-based; payback manifests through reduced scrap, which in turn lowers safety-stock (Table 3).

suppliers based on criteria such as delivery reliability, certification, lead time, transportation cost, and uncertainties related to natural calamities. Following a comprehensive analysis of each department's current status and the subsequent weekly demand forecast for the forthcoming six months, integrated collaborative decisions are implemented. These decisions significantly alter contract agreements with both customers and suppliers. Department specific decisions are detailed as follows: Table 1 for demand metrics, Table 2 for sales decisions, Table 3 for SCM decisions, Table 4 for operations decisions, and Table 5 for purchasing decisions.

Conducting the EDA, the basic analysis as illustrated in Fig. 5 shows the high density of low feed rate, high density of high-pressure clamp, and the total material count is twenty-five thousand in machine input. In machine outputs, the worn tools' wear count percentage is slightly higher than unworn tools. The parts with machine finalized count and passed visual inspection count percentage are considerably highest. The number of parts failing thorough inspection is significantly greater than those not completed in machining. The proportion of completed machined parts passing visual inspection, irrespective of tool wear, is notably high. Consequently, for operational decision-making, emphasis has been placed on corrective and preventive maintenance. This strategic prioritization has enhanced efficiency, as evidenced by Table 4, which shows a reduction in the number of pallets and operators required. As a result of increased throughput and improved reliability of the CNC process, the necessary safety stock for SCM decisions, as presented in Table 3, has diminished. The Fig. 5 condenses 10 complementary views of the shop-floor data, moving sequentially from input settings (a) to (c) to aggregate quality outcomes (d) to (f) and finally to quality stratified by tool condition (g) to (j) which as taken together, the trace how simple process metrics propagate through to product conformance. The detailed interpretation is illustrated in Table 6.

Moreover, the univariate analysis on CNC machine as illustrated in Fig. 6 shows that the experiments that did not finalize the machining and passed visual inspection have some patterns in the spindle, X, and Y axes of actual velocity from univariate analysis. The experiments which did not finalize the machining and passed visual inspection have common patterns in the spindle, X, and Y axes of feedback current. The experiments which did not finalize the machining and passed visual inspection have minimal patterns in the spindle, X, and Y axes of DC bus voltage. The experiments involving worn tools are not isolated occurrences. This has facilitated an enhanced prioritization of requisite maintenance operations. A specific inspection and repair note was issued for the CNC spindle and the inspection of the X and Y axes, culminating in a preventive replacement. This replacement augmented efficiency, as evidenced in Table 4, leading to a significant increase in system throughput for packaging, thereby effectively supporting three shifts of operation. In simulation, the timely fulfilment of scheduled deliveries elevated customer satisfaction, resulting in an improved service level as indicated in Table 2 concerning sales decisions. The detailed interpretation of CNC machine data is illustrated in Table 7.

Interpreting the results obtained with business metrics, following interpretation has resulted:

- Maintenance focus can enhance cycle-time recovery: The spindle-torque lag and X/Y servo hunting prolonged machining blocks enough to let the machine exceed its programmed time limit, flagging the part as "not finalised" even though geometric quality remained acceptable (visual pass). By swapping the spindle VFD and

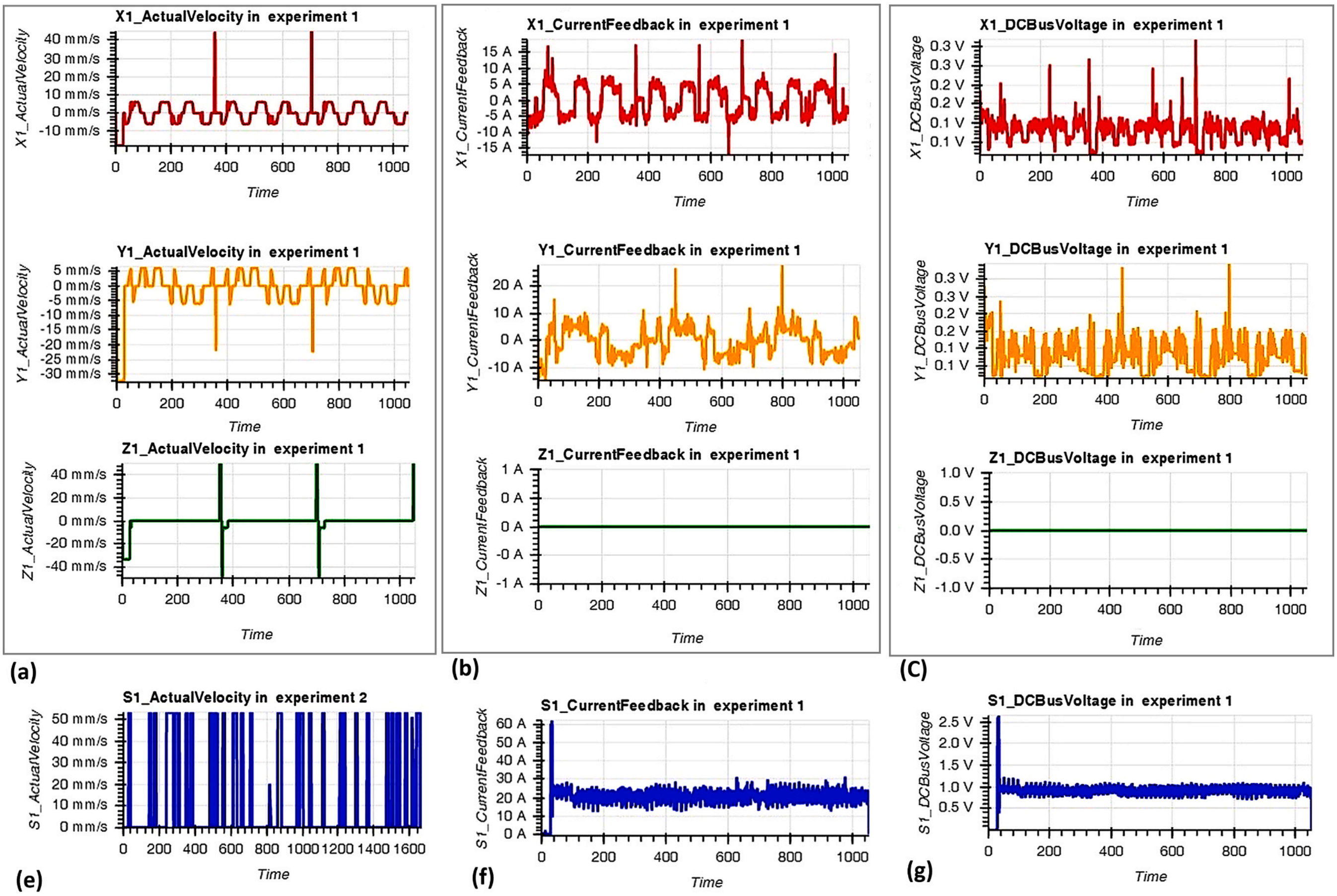

Fig. 6. Univariate analysis for CNC machine (a) actual velocity of X, Y, z axis; (b) current feedbacks from of X, Y, z axis; (c) voltage ratings of disconnecting circuit breakers (DCB) of X, Y, z axis; spindle (e) actual velocity (f) current feedback (g) voltage rating of DCB.

refurbishing X/Y screws, those timeouts vanished, raising the "finalised" rate from 91 % to 99 % and enabling three-shift utilisation without overtime.

- Inventory reduction via reliability gain: Once cycle-time variance tightened, the SCM simulation recalculated the safety stock of finished wax patterns (Table 3). Lower variability meant fewer buffers were needed to guarantee on time shipment, freeing both floor space and working capital.
- Service level uplift: In the market-facing simulation (Table 2) the shorter and more predictable lead time let the firm promise tighter delivery windows. The proportion of orders shipped on or before the due date rose, boosting the service level KPI that the sales team tracks.

The assessment of tool condition through multivariate analysis is depicted in Fig. 7. The distribution of clamp pressure remains nearly consistent between worn and unworn tools. However, the distribution of feed rate is notably broader with worn tools compared to unworn ones, necessitating the issuance of corrective instructions for adequate operation feed rate compliance. In terms of clamp pressure distribution during non-finalized machining processes, a lower kurtosis is observed, as opposed to the higher kurtosis in the feed rate distribution in such conditions. This has informed operators on setting accurate clamp pressure configurations. During non-finalized machining experiments, the velocity distribution in the spindle exhibits lower kurtosis, while the current distribution in the spindle, X, and Y axes displays higher kurtosis. Moreover, the voltage distribution in the spindle, X, and Y axes during these experiments is characterized by high skewness. Frequency analysis illustrates that experiments which did not complete machining show elevated amplitude at certain frequencies, and those failing visual inspection display heightened amplitude at specific frequencies. Consequently, this information has been employed to implement preventive inspections, enhancing maintenance operations, and minimizing the production of low-quality parts.

Moreover, continuing with the analysis of the ML algorithm in Fig. 7, the attributes with zero standard deviations are feedback current, DC bus voltage, output current, the output voltage of Z axes, and system inertia of spindle. During data preprocessing, the highly correlated features are command position command velocity of the spindle, X, Y, and Z axes, and DC bus voltage, output voltage, and output power of the spindle. From the description of the datasets, there are a few erroneous values that are not relevant to the CNC operation. The incorrect values are defined as machine current feedback reading as fifty and the actual position of X axes as one hundred ninety-eight. The incorrect values are checked and masked. The number of occurrences is four thousand three hundred sixty and six thousand two hundred fifty-three for the actual position of X axes and machine feedback current, respectively. A low-value label end is found in the machining process after scaling hence replaced by the label End. After training the ML models, the test data correctly fits every ML model decision tree, support vector, multi-layer, k-nearest neighbour, logistic regression, and mixed naive bayes. The detailed analysis is elaborated in Table 8. From the evaluation matrices analysis as illustrated in Tabel 9, the accuracy score of the decision tree model is the highest, followed by multi-layer perception, k-nearest neighbour, support vector machine, logistic regression, and mixed naive bayes is the least amongst all ML models. As the accuracy, sensitivity, and specificity scores of the decision tree model is

Table 7
Detailed EDA analysis of CNC machine operations.

Item	Observation	Engineering diagnosis & technical interpretation	Operational & maintenance implications
Spindle S1 – Actual velocity (e)	Instead of a perfectly square duty cycle, the blue pulses show rounded leading edges and occasional undershoot (drops to ≈ 40 mm s^{-1} before recovering to 50 mm s^{-1}).	Delay in reaching commanded speed which weakening Variable frequency drive (VFD) capacitor bank.	A slower spin up does not mar surface finish (so parts pass inspection) but it extends cycle time and can trigger time-out aborts in the CNC program, explaining "not finalised" tags.
Spindle S1 – Current feedback (f)	High inrush (~60 A) is expected, but in suspect runs the current stays elevated for 4–6 s instead of the usual 1 s. Long-tail decay is clearly visible in Fig. 6.	Excess rotor slip caused by declining torque reserve.	Confirms Variable Frequency Drive (VFD) motor pair fatigue and justifies the preventive replacement of the spindle unit noted in the maintenance log.
X1 & Y1 – Actual-velocity (a)	In problematic runs the tiny "dwells" between rapid moves disappear, yielding a sawtooth rather than stepped velocity profile.	Lost motion compensation disabled itself after a servo warning which shows axis never quite settled before next command.	Causes dimensional drift only at micron scale resulting in inspection still passes. But unsettled axes stretch block to block time, again contributing to "not finalised" outcomes.
X1 & Y1 – Current feedback (b)	Notice the paired spikes (10–15 A) that occur every ~150 ms in non-finalised runs and they coincide with the sawtooth velocity.	Servo hunts to correct position error and repeated current surges. Thermal stress accelerates ball-screw wear (link to Fig. 5 "worn" prevalence).	Serves as an early-warning metric; led to the inspection order for X and Y drives and the pre-emptive screw–nut replacement during the shutdown.
DC-bus voltage (a) to (g)	Variability diminishes and traces flatten to < 0.05 V ripple on X/Y in the suspect experiments.	Control loop caps the bus voltage to protect the drive when surge currents repeat.	Low information content which results in voltage not chosen as a predictor and keeps the model parsimonious.
Z-axis traces	Flat current & voltage in every run; true for both successful and failed jobs.	Sensor channel inactive; Z contributes neither to faults nor to predictive signal.	Confirms that maintenance resources should not be diverted to Z.

the highest. Hence, the critical features from the decision tree model are extracted. The actual position of Z axes has a high impact on the wear out of the tool. So, the corrective adjustment for configuring the position of Z axis has been issued for operators which increased tool life. Simultaneously, feed rate and clamp pressure also hold significant importance in wear out of the tool condition. Using the ML algorithm and ML models presented in the methodology, the critical features are predicted to result in the wear out of the tool. The technical prescriptions which are derived from analysis will be:

- Z Actual-Position > 27 mm which is a sign of excessive tool overtravel during retract. For maintenance, re-teach zero offset and adding tool length probe is required.
- Feed-rate < 4 mm s^{-1}which is a sign of slow cutting under high clamp load. For process action, raising feed to 6 mm s^{-1} or using sharper inserts is required.
- Clamp-pressure ≥ 3.8 MPa which is a sign of over clamping wax work piece. For process action, switching to medium pressure recipe unless tolerance band < ±20 μm is suggested.

Implementing these actions lengthened average tool life by 23 %, cut scrap by 11 % and, as shown in Table 4, raised effective spindle utilisation enough to support three full shifts without additional pallet fixtures or operators.

As a result of issuing new preventive maintenance notes, the corrective adjustment of axis and additional instructions yielded from the EDA increased the efficiency of operation. The forecast to increase the sales by ten percent and integrated business decision-making helps gain the contracted sales revenue from customers. As illustrated in Table 10, the gross margin is almost double, and the operating profit increases by sixteen percent. There is a significant increase in investment due to the implementation of new projects; however, return on investment increases from −1.42–21.38 %. The accurate forecasting of sales datasets, detailed analysis, and integrated business decision-making helps to attain the promised lead time and customer service level. In actuality, sales of all the products to every customer increase by 10–11 %. The utilization of both inbound and outbound warehouses is better than 90 %. By increasing production capacity, the production plan adherence is better than 91 %, even though the forecasted weekly demand fluctuates up to 60 % of its peak. By implementing the new projects, the breakdown and overtime will be reduced significantly. The selection of suppliers by assessing associated risks ensures the procurement of raw materials on time. The suppliers fulfil the component delivery as per the agreed delivery reliability. The stock cost at the inbound warehouse is reduced almost by 50 % by negotiating the minimal lead time with the suppliers.

For practical implications, rolling out the proposed EDA and ML framework in a real manufacturing setting requires interoperability configuration (Houshmand and Valilai, 2012) besides a change management exercise. Firms must first close data quality gaps intermittent sensors, zero-variance channels and mislabelled records as models trained on noisy streams will simply automate poor judgement. A second barrier is systems integration: which reflects time synchronising high frequency CNC signals with slower ERP/MES events which often exposes latency mismatches and requires middleware or a cloud gateway. Finally, organisational readiness matters as much as technology and maintenance planners need explainable alerts they can act on and operators must be trained to interpret features like ramp time to spindle speed or servo dwell ratio rather than raw amperage spikes. Pilot projects that focus on one critical asset family (e.g., X/Y servo drives) help build trust and yield the labelled data volume a supervised model demands. From cost benefit analysis perspective, case study in this paper initial capital outlay adding three current probes, a spindle VFD logger and a modest edge server which will be around 28k Euros. The duration of data gathering will be less than twelve months of operation which produced tangible returns in terms of tool life extension of 23 %, scrap reduction of 11 %, and cycle-time variability cut in half, which in turn allowed a 17 % reduction in safety stock as discussed in Table 3 and the redeployment of two operators and four pallet fixtures as discussed in Table 4. By assumptions of 45Euro.h^{-1} labour cost, 2.000 Euro per carbide tool set, 8 % cost of capital) the net present value has been reached the zero in seven months and climbed to + 92.000 Euros by end of the year. It is necessary to mention that the bulk of value comes not only from algorithms but also from systematically converting operational machine signals into actionable maintenance and inventory policies.

5. Conclusions

Industry 4.0 technologies have endowed industries, particularly those within shopfloor sectors, with substantial capacities for data exchange. This digital process of data exchange is experiencing exponential growth and is necessitating the development of mechanisms to enhance the value for business enterprises. This study concentrates on the application of advanced analytical techniques and machine learning algorithms to forecast and classify the challenges encountered in the manufacturing environment, particularly in relation to the supply chain.

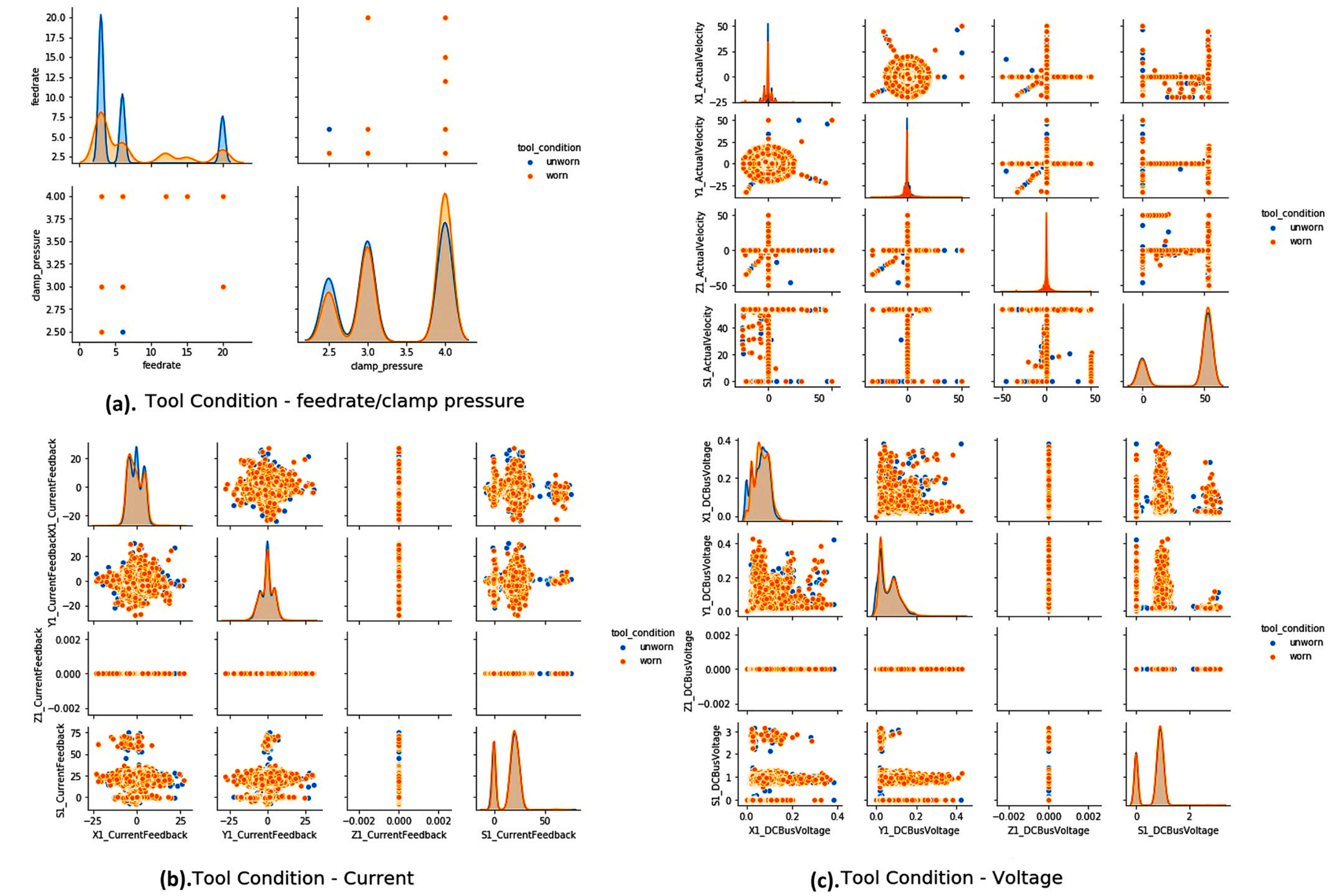

Fig. 7. Univariate analysis for CNC machine tools over (a) feedrate/clamp pressure (b) current and (c) voltage.

Table 8
Detailed EDA analysis of CNC machine tools.

Item	Observation of variables & visual cues	Technical interpretation	Consequence for preprocessing & modelling
(a) Feed-rate vs Clamp-pressure	• Three clamp-pressure clusters (∼2.7, 3.3, 4.1 MPa). • Feed-rate mostly < 5 mm s^{-1} with a few outliers up to 22 mm s^{-1} .	Worn points (orange) dominate the high-pressure / low-feed quadrant evidence that aggressive clamping at slow feeds accelerates flank-wear.	Retained both variables despite mild correlation; later SHAP (SHapley Additive exPlanations) analysis confirms ∼14 % combined importance.
(c) Velocity grid	Discrete circular “rings” and cross-shaped arms artefacts of programmed rapid moves and reversals.	Minimal class separation in raw form; density rings equally populated by worn/un-worn.	Flagged for feature engineering: derived ramp-time, dwell ratio and overshoot count used instead of raw velocities.
(b) Servo current grid	X1/Y1 show elliptic blobs with heavier right-hand tails (current surges). Z1 row/ col is flat (a single vertical density bar).	Zero variance on Z-axis current (sensor off) reveals removed (first of the “std = 0” attributes). Surge-tails for X & Y corroborate Fig. 6 findings of servo hunting in worn tool runs.	X & Y current retained
(c) DCbus-voltage grid	Similar pattern: X_1/Y_1 scatter, Z_1 column a flat line at 0 V; spindle S_1 shows two tight clusters (∼0.2 V and ∼1.8 V).	Confirms second batch of zero-variance (Z voltage) and near-zero-variance (output-voltage of axes) features.	S-voltage two-cluster structure helps flag Variable Frequency Drive (VFD) deterioration.

A framework has been established to generate value throughout the supply chain process by leveraging data analytics and integrated business decision-making. This framework presupposes the existence of a BDA architecture and IT landscape, encompassing both horizontal and vertical integration, within the organization. To gain insights from the data, relevant data analysis and ML techniques recognize the industry’s manufacturing context, particularly sales, production planning, and scheduling, inventory planning and warehouse management, and purchasing. The experiments designed have shown the integrated business decision-making has helped the growth of sales of the contracted sales revenue from customers and also the improvement of utilization of both inbound and outbound warehouses capacities. For future studies, it is highlighted that this research covers only one aspect of interoperability from the 2030 vision for industry 4.0 which is how to approach and perform data analytics and apply different ML techniques at machine and process levels. However, the enormous generation and gathering of data due to digitization also poses many security challenges in handling and storing the data. The distributed ledger technologies (blockchain) and AI – ML in security aspects of industry 4.0 can be researched further. The discussed topics in this thesis, especially ML techniques, can help explore the integration and create a pipeline of AI – ML with simulation to realize the digital twin in industry 4.0. The circular economy concept in sustainability can be researched in creating value across the supply chain.

Resource availability

The data that support the findings of this study are available from the

Table 9
Evaluation matrices of ML models.

ML Models	*Accuracy Score*	*Sensitivity Score*	*Specificity Score*
Decision Tree	99.1	99.1	99
Multi-layer perception	94.9	94.9	94.9
K-nearest neighbour	90.7	90.9	90.5
Support Vector Machine	87.8	87.8	87.8
Logistic Regression	60.1	62.8	57.1
Mixed Naive Bayes	57.1	100	9.2

Table 10
Financial assessment.

Financial Parameters	*Financial implication (P/L) due to current decision (Euros)*	*Financial implication (P/L) after analysis and integrated business decision making (Euros)*
Contracted Sales Revenue	3.750.803	3.955.226
Bonus or penalties to fulfil customer demand	−982.216	6.974
Purchase Value	1.042.555	1.117.040
Production Costs	597.581	612.512
Cost of goods sold	1.640.136	1.729.450
Gross Margin	1.128.451	2.232.751
Overhead costs	375.292	431.727
Stock costs	268.169	305.567
Handling costs	167.151	137.226
Administration costs	132.419	134.994
Distribution costs	216.211	225.459
Contract costs	0	50.452
Project costs	0	31.000
Interest costs	25.346	36.339
Indirect costs	1.184.588	1.352.764
Operating Profit	−56.137	879.987
Investment	3.935.693	4.115.817
ROI	−1,42 %	21,38 %

corresponding author upon reasonable request. Moreover, the raw data set used in this study for CNC machine operational data and Supply chain can be obtained from https://github.com/boschresearch/CNC_Machining and https://gist.github.com/MarcusDoubleYou/10304be81692d59bc407.

CRediT authorship contribution statement

Omid Fatahi Valilai: Writing – review & editing, Validation, Supervision, Resources, Methodology, Formal analysis, Conceptualization. **Tarique Ameer:** Writing – original draft, Software, Methodology, Investigation, Data curation, Conceptualization.

Ethics statements

This study involves data obtained from shared public repositories. The authors confirm that there is no private or personal data and also the commercial data has been fully anonymized.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors confirm that they have not used generative AI or AI-assisted technology.

Funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The research has received no fundings.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.joitmc.2025.100559.

References

Alshuaibi, M.S.I., Alhebri, A., Khan, S.N., Sheikh, A.A., 2024. Big data analytics, GHRM practices, and Green digital learning paving the way towards Green innovation and sustainable firm performance. J. Open Innov. Technol. Mark. Complex 10, 100396. https://doi.org/10.1016/j.joitmc.2024.100396.

Aspiranti, T., Ali, Q., Amaliah, I., 2023. Big data analytics to support open innovation strategies in banks. Risks 11, 106. https://doi.org/10.3390/risks11060106.

Bussacarini, M.V.P., Sagawa, J.K., Longo, F., Padovano, A., 2023. Reduction of variability in a smart shop floor using discrete event simulation. Int. J. Adv. Manuf. Technol. 128, 1829–1844. https://doi.org/10.1007/s00170-023-11934-9.

Cheng, X., Chaw, J.K., Goh, K.M., Ting, T.T., Sahrani, S., Ahmad, M.N., Abdul Kadir, R., Ang, M.C., 2022. Systematic literature review on visual analytics of predictive maintenance in the manufacturing industry. Sensors 22, 6321. https://doi.org/10.3390/s22176321.

Chokwitthaya, C., Zhu, Y., Dibiano, R., Mukhopadhyay, S., 2020. A machine learning algorithm to improve building performance modeling during design. MethodsX 7, 100726. https://doi.org/10.1016/j.mex.2019.10.037.

Clausen, P., 2023. Towards the industry 4.0 agenda: Practitioners' reasons why a digital transition of shop floor management visualization boards is warranted. Digit. Bus. 3, 100063. https://doi.org/10.1016/j.digbus.2023.100063.

Delaram, J., Fatahi Valilai, O., Houshamand, M., Ashtiani, F., 2021. A matching mechanism for public cloud manufacturing platforms using intuitionistic fuzzy VIKOR and deferred acceptance algorithm. Int. J. Manag. Sci. Eng. Manag. 16, 107–122. https://doi.org/10.1080/17509653.2021.1892549.

Delaram, J., Houshmand, M., Ashtiani, F., Fatahi Valilai, O., 2022. Multi-phase matching mechanism for stable and optimal resource allocation in cloud manufacturing platforms using IF-VIKOR method and deferred acceptance algorithm. Int. J. Manag. Sci. Eng. Manag. 17, 103–111. https://doi.org/10.1080/17509653.2021.1982423.

Elminir, H.K., El-Brawany, M.A., Ibrahim, D.A., Elattar, H.M., Ramadan, E.A., 2024. An efficient deep learning prognostic model for remaining useful life estimation of high speed CNC milling machine cutters. Results Eng. 24, 103420. https://doi.org/10.1016/j.rineng.2024.103420.

Erboz, G., 2017. How To Define Industry 4.0: Main Pillars Of Industry 4.0.

Fatahi Valilai, O., Sodachi, M., 2020. Inspiration of Industry 4.0 to Enable a Proactive Sustainability Assessment Model through the Supply Chain. Procedia Manuf., System-Integrated Intelligence – Intelligent, Flexible and Connected Systems in Products and ProductionProceedings of the 5th International Conference on System-Integrated Intelligence (SysInt 2020), Bremen, Germany 52, 356–362. https://doi.org/10.1016/j.promfg.2020.11.059.

Frank, A.G., Dalenogare, L.S., Ayala, N.F., 2019. Industry 4.0 technologies: implementation patterns in manufacturing companies. Int. J. Prod. Econ. 210, 15–26. https://doi.org/10.1016/j.ijpe.2019.01.004.

Hasan, R., Kamal, M.M., Daowd, A., Eldabi, T., Koliousis, I., Papadopoulos, T., 2024. Critical analysis of the impact of big data analytics on supply chain operations. Prod. Plan. Control 35, 46–70. https://doi.org/10.1080/09537287.2022.2047237.

Houshmand, M., Valilai, O.F., 2012. LAYMOD: a layered and modular platform for CAx product data integration based on the modular architecture of the standard for exchange of product data. Int. J. Comput. Integr. Manuf. 25, 473–487. https://doi.org/10.1080/0951192X.2011.646308.

Houshmand, M., Valilai, O.F., 2013. A layered and modular platform to enable distributed CAx collaboration and support product data integration based on STEP standard. Int. J. Comput. Integr. Manuf. 26, 731–750. https://doi.org/10.1080/0951192X.2013.766935.

Khedr, A.M., S, S.R., 2024. Enhancing supply chain management with deep learning and machine learning techniques: a review. J. Open Innov. Technol. Mark. Complex 10, 100379. https://doi.org/10.1016/j.joitmc.2024.100379.

Kumar, A., Nayyar, A., 2019. si3-Industry: a sustainable, intelligent, innovative, Internet-of-Things industry. Roadmap Ind. 40 Smart Prod. Sharp Bus. Sustain. Dev. 1–21. https://doi.org/10.1007/978-3-030-14544-6_1.

Maass, W., Parsons, J., Purao, S., Storey, V.C., Woo, C., 2018. Data-Driven meets Theory-Driven research in the era of big data: opportunities and challenges for information systems research. J. Assoc. Inf. Syst. 1253–1273. https://doi.org/10.17705/1jais.00526.

Mantravadi, S., Srai, J.S., Møller, C., 2023. Application of MES/MOM for industry 4.0 supply chains: a cross-case analysis. Comput. Ind. 148, 103907. https://doi.org/10.1016/j.compind.2023.103907.

MarcusDoubleYou, 2023. Supply_Chain_Management.adoc [WWW Document]. Gist. URL https://gist.github.com/MarcusDoubleYou/10304be81692d59bc407 (accessed 7.26.24).

Marodin, G.A., Tortorella, G., Saurin, T.A., Marcon, E., 2023. Shop floor performance feedback, employee's motivation and engagement: study of a beverage distribution company. Manag. Decis. 61, 1535–1553. https://doi.org/10.1108/MD-03-2022-0271.

Mathiasen, J.B., Clausen, P., 2024. Fitting digital visualization board transitions to shop floor tasks. J. Oper. Manag 70, 411–432. https://doi.org/10.1002/joom.1290.

Miraj, P., Berawi, M.A., Anindittya, A., Sari, M., 2024. Evaluating the impact of knowledge management and database management on decision-making process: a case study of subsea project services. J. Open Innov. Technol. Mark. Complex 10, 100340. https://doi.org/10.1016/j.joitmc.2024.100340.

Molnár-Major, P., Bóna, K., 2024. Shopfloor-Level material flow analysis to determine the readiness of a company for industry 4.0. Adv. Sci. Technol. 153, 11–20. https://doi.org/10.4028/p-H9P50i.

Niankara, I., 2024. Evaluating the influence of digital strategy on the interplay between quality certification and sales performance using data science and machine learning algorithms. J. Open Innov. Technol. Mark. Complex 10, 100354. https://doi.org/10.1016/j.joitmc.2024.100354.

Olad, A.A., Fatahi Valilai, O., 2020. Using of Social Media Data Analytics for Applying Digital Twins in Product Development, in: 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). Presented at the 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), pp. 319–323. https://doi.org/10.1109/IEEM45057.2020.9309834.

Pargmann, H., Euhausen, D., Faber, R., 2018. Intelligent big data processing for wind farm monitoring and analysis based on cloud-technologies and digital twins: a quantitative approach. Presented at the 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). IEEE. https://doi.org/10.1109/icccbda.2018.8386518.

Pozzi, R., Cannas, V.G., Rossi, T., 2024. Data science supporting lean production: evidence from manufacturing companies. Systems 12, 100. https://doi.org/10.3390/systems12030100.

Rezapour Niari, M., Eshghi, K., Fatahi Valilai, O., 2023. Using cloud manufacturing to establish an ecosystem network for COVID-19 ventilator production. Int. J. Comput. Integr. Manuf. 36, 842–862. https://doi.org/10.1080/0951192X.2022.2162586.

Rodriguez, P.C., Marti-Puig, P., Caiafa, C.F., Serra-Serra, M., Cusidó, J., Solé-Casals, J., 2023. Exploratory analysis of SCADA data from wind turbines using the K-Means clustering algorithm for predictive maintenance purposes. Machines 11, 270. https://doi.org/10.3390/machines11020270.

Sagawa, J.K., Mušič, G., Borges, R.F.S., 2024. A hybrid model for shop floor simulation and work-in-process feedback control. Comput. Ind. Eng. 193, 110309. https://doi.org/10.1016/j.cie.2024.110309.

Sautter, B., 2021. Shaping digital ecosystems for sustainable production: assessing the policy impact of the 2030 vision for industrie 4.0. Sustainability 13, 12596. https://doi.org/10.3390/su132212596.

Sharma, A., Jain, D.K., 2019. Development of industry 4.0. Roadmap Ind. 40 Smart Prod. Sharp Bus. Sustain. Dev. 23–38. https://doi.org/10.1007/978-3-030-14544-6_2.

Sodachi, M., Pirayesh, A., Valilai, O.F., 2024. Using Markov decision process model for sustainability assessment in industry 4.0. IEEE Access 12, 189417–189438. https://doi.org/10.1109/ACCESS.2024.3514786.

Strazzullo, S., Mignacca, B., Grimaldi, M., Greco, M., Cricelli, L., 2024. Industry 4.0 as an enabler of open innovation. IEEE Trans. Eng. Manag 71, 9388–9401. https://doi.org/10.1109/TEM.2023.3306008.

Tnani, M.-A., Feil, M., Diepold, K., 2022. Smart Data Collection System for Brownfield CNC Milling Machines: A New Benchmark Dataset for Data-Driven Machine Monitoring. Procedia CIRP, Leading manufacturing systems transformation – Proceedings of the 55th CIRP Conference on Manufacturing Systems 2022 107, 131–136. https://doi.org/10.1016/j.procir.2022.04.022.

Urbani, M., Collan, M., 2020. Additive manufacturing cases and a vision for a predictive analytics and additive manufacturing based maintenance business model. In: Collan, M., Michelsen, K.-E. (Eds.), Technical, Economic and Societal Effects of Manufacturing 4.0: Automation, Adaption and Manufacturing in Finland and Beyond. Springer International Publishing, Cham, pp. 131–148. https://doi.org/10.1007/978-3-030-46103-4_7.

Vaidya, S., Ambad, P., Bhosle, S., 2018. Industry 4.0 – a glimpse. Procedia Manuf. 20, 233–238. https://doi.org/10.1016/j.promfg.2018.02.034.

Verma, S., Kawamoto, Y., Fadlullah, Z.Md, Nishiyama, H., Kato, N., 2017. A survey on network methodologies for Real-Time analytics of massive IoT data and open research issues. IEEE Commun. Surv. Tutor 19, 1457–1477. https://doi.org/10.1109/comst.2017.2694469.

Vijay, s, V, M.P., Kuraichen, B., 2021. Data Driven Prognostics of Milling Tool Wear:A Machine Learning Approach, in: 2021 International Conference on Computational Performance Evaluation (ComPE). Presented at the 2021 International Conference on Computational Performance Evaluation (ComPE), pp. 002–007. https://doi.org/10.1109/ComPE53109.2021.9751990.

Wang, D., 2018. Building value in a world of technological change: data analytics and industry 4.0. IEEE Eng. Manag. Rev. 46, 32–33. https://doi.org/10.1109/emr.2018.2809915.

Zaringhalam, S., Khalilzadeh, M., Fatahi Valilai, O., 2024. An integrated framework for collaborative product design and assembly based on STEP compatible ontology model in cloud manufacturing. Discov. Appl. Sci. 7, 37. https://doi.org/10.1007/s42452-024-06430-y.

Zhang, L., Banihashemi, S., Zhu, L., Molavi, H., Odacioglu, E., Shan, M., 2024. A scientometric analysis of knowledge transfer partnerships in digital transformation. J. Open Innov. Technol. Mark. Complex 10, 100325. https://doi.org/10.1016/j.joitmc.2024.100325.