A Generalized Framework for Adopting Regression-Based Predictive Modeling in Manufacturing Environments

Akinsolu, M. O. and Ziribi, K.

inventions

Article

A Generalized Framework for Adopting Regression-Based Predictive Modeling in Manufacturing Environments

Mobayode O. Akinsolu [1,2,*] **and Khalil Zribi** [1,3]

[1] Faculty of Arts, Science and Technology, Wrexham Glyndŵr University, Wrexham LL11 2AW, UK
[2] Academic Engagement Department, The Alan Turing Institute, British Library, London NW1 2DB, UK
[3] ESEO Angers—Grande Ecole d'Ingénieurs Généralistes, 10 Bd Jean Jeanneteau, 49100 Angers, France
[*] Correspondence: mobayode.akinsolu@glyndwr.ac.uk or m.o.akinsolu@ieee.org

Abstract: In this paper, the growing significance of data analysis in manufacturing environments is exemplified through a review of relevant literature and a generic framework to aid the ease of adoption of regression-based supervised learning in manufacturing environments. To validate the practicality of the framework, several regression learning techniques are applied to an open-source multi-stage continuous-flow manufacturing process data set to typify inference-driven decision-making that informs the selection of regression learning methods for adoption in real-world manufacturing environments. The investigated regression learning techniques are evaluated in terms of their training time, prediction speed, predictive accuracy (R-squared value), and mean squared error. In terms of training time (TT), k-NN20 (k-Nearest Neighbour with 20 neighbors) ranks first with average and median values of 4.8 ms and 4.9 ms, and 4.2 ms and 4.3 ms, respectively, for the first stage and second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, respectively, over 50 independent runs. In terms of prediction speed (PS), DTR (decision tree regressor) ranks first with average and median values of 5.6784×10^6 observations per second (ob/s) and 4.8691×10^6 observations per second (ob/s), and 4.9929×10^6 observations per second (ob/s) and 5.8806×10^6 observations per second (ob/s), respectively, for the first stage and second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, respectively, over 50 independent runs. In terms of R-squared value (R^2), BR (bagging regressor) ranks first with average and median values of 0.728 and 0.728, respectively, over 50 independent runs, for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, and RFR (random forest regressor) ranks first with average and median values of 0.746 and 0.746, respectively, over 50 independent runs, for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process. In terms of mean squared error (MSE), BR (bagging regressor) ranks first with average and median values of 2.7 and 2.7, respectively, over 50 independent runs, for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, and RFR (random forest regressor) ranks first with average and median values of 3.5 and 3.5, respectively, over 50 independent runs, for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process. All methods are further ranked inferentially using the statistics of their performance metrics to identify the best method(s) for the first and second stages of the predictive modeling of the multi-stage continuous-flow manufacturing process. A Wilcoxon rank sum test is then used to statistically verify the inference-based rankings. DTR and k-NN20 have been identified as the most suitable regression learning techniques given the multi-stage continuous-flow manufacturing process data used for experimentation.

Keywords: artificial intelligence; data; data analysis; machine learning; manufacturing; regression; regression learning; supervised learning

Citation: Akinsolu, M.O.; Zribi, K. A Generalized Framework for Adopting Regression-Based Predictive Modeling in Manufacturing Environments. *Inventions* **2023**, *8*, 32. https://doi.org/10.3390/inventions8010032

Academic Editor: Giuseppe Casalino

Received: 14 December 2022
Revised: 7 January 2023
Accepted: 20 January 2023
Published: 29 January 2023

1. Introduction

As the world continues to gravitate to artificial intelligence (AI)-driven societies, where paradigms such as Industry 4.0 [1] and Society 5.0 [2] are gradually and increasingly redefining the manufacturing and production of goods and services, and how we live and interact, respectively, the role of the science of data in the drive for innovation and invention cannot be overemphasized [3,4]. Presently, AI, which in a sense premises on data-driven models and constructs [5], is one of the primary proponents of Industry 4.0 (I4.0) and Industry 5.0 (I5.0) [3]. With a focus on manufacturing, from the first industrial revolution in the 1760s to the present-day I4.0 and I5.0, manufacturing across the globe has always involved workflows and processes that churn out vast amounts of data [3,4]. For example, timestamps and time logs on machine tools and equipment, sensor readings, operational speeds of rotating tools, and measurements of throughput and slags on shop floors have always generated vast amounts of data [6,7]. As a matter of fact, this in a way idealizes big data, where manufacturing big data can be broadly described as data collected at every stage of manufacturing and/or production, including data from operators, equipment, machine tools, process systems, and devices on the shop floor [8,9].

As hinted at above, manufacturing has always been data intensive. However, up until recent times, manufacturing data has not really been used to "automagically" improve or enhance the manufacturing processes that generated them in the first place [9,10]. In other words, other than process monitoring and regulation that are mostly achieved via traditional process visualization and process control using human-machine interfaces (HMIs), manufacturing data is still yet to be used extensively to support data-driven paradigms such as predictive modeling (usually takes the form of predictive maintenance when machine behavior and the life cycle is the context [11,12]) and ubiquitous monitoring and control (usually incorporated in the industrial internet of things (IIoT) where higher levels of connectivity, interoperability and robustness is the goal [13,14]). This apparent gap can be mainly attributed to the purpose-built or fit-for-purpose nature of most of the computational paradigms and analytic frameworks that support manufacturing processes [5].

To lend credence to the above-mentioned, the example of a manufacturing or production process can be used. Take, for instance, a programmable logic controller (PLC) designated for overseeing the operations of a conveyor belt system using a master-slave (M-S) architecture that only needs to power the motor of the conveyor belt system on and/or off based on some measured metrics such as read sensor values when the conveyed item reaches a designated point or landmark along the production line as exemplified in [15]. In a sense, every other data such as the remaining useful life (RUL) of the motor and quality of the product (the item being conveyed in this case such as the bottles and trays in [15]) that are also associated with the process may not really impact on the modi operandi of the PLC-based M-S control architecture. However, if such data are made available and analyzed, inferences drawn from such analysis can go a long way in improving the overall efficiency of the process. Consequently, it is to this end that data science is more relevant than it has ever been before in ushering in a new era of data-driven innovations and inventions in manufacturing environments.

It can be argued that many of the data-driven innovations and inventions expected to be seen in the global manufacturing sector will be mostly based on predictive modeling [3,5,16]. Generally, predictive modeling can be viewed as being synonymous with supervised learning, a form of machine learning or subcategory of AI that can take on the form of classification and/or regression, depending on the nature or type of data that is employed. For most commercial applications, particularly, manufacturing applications as discussed in [17], predictive modeling can also be called predictive analytics. Broadly speaking, predictive analytics is the analysis of available data (e.g., explanatory and response variables in a typical manufacturing process data set as carried out in Section 2.3) and data trends using statistical learning and/or machine learning methods to enable and support the prediction of future trends and/or responses. According to the literature, the global manufacturing industry can be said to be in the early stage of adopting AI in many of their production processes, particularly, machine learning in the form of predictive modeling [5,16]. So, this has necessitated the conceptualization, design, development, and testing of several paradigms by scientists, engineers, and researchers as also carried out in this work.

Presently, contemporary manufacturing technologies now support the embedding of predictive modeling and other AI-based paradigms into many manufacturing and production processes [18–20]. Some of these manufacturing processes include, but are not limited to, maintenance planning, quality control, monitoring and regulation, and robust automation of shop floor operations [18–20]. Hence, there is a growing need to have a generic framework that supports the holistic and granular use of data sets from manufacturing processes to support data-driven innovations and inventions in manufacturing environments. To respond to this need, a generalized framework is presented and investigated in this paper to support predictive modeling in manufacturing environments. More specifically, the following contributions are made:

- Formulation of a generalized framework for predictive modeling in manufacturing environments using data collected from real-time manufacturing processes.
- Validation of the proposed generalized framework by applying several regression learning techniques to undertake predictive modeling using an open-source multi-stage continuous-flow manufacturing process data set.
- Summarized inference-based and statistically verified rankings for the adoption and selection of regression learning techniques to potentially invent predictive modeling-based paradigms for typical manufacturing processes.

The remainder of this paper is organized as follows: Section 2.1 provides a summarized discussion on the relevant related work to further emphasize the usefulness of the work carried out. Section 2.2 details the proposed generalized framework and the main steps required for its implementation. An overview of the case study used to validate the proposed generalized framework is provided in Section 2.3, alongside a description of the associated multi-stage continuous-flow manufacturing process data set. The machine learning problem addressed by the proposed framework in the context of the adopted case study, and the investigated algorithms are detailed in Section 2.4. The experimental setup used for the investigations is presented in Section 3. Results, discussions, and comparisons are detailed in Section 4, and the concluding remarks are provided in Section 5.

2. Materials and Methods

2.1. Relevant Related Work

Machine learning (ML) in the form of data-driven predictive modeling in manufacturing environments is gaining a lot of research interest in recent times [18–20]. Regression or regression analysis or regression learning is one of the widely accepted approaches for predictive modeling in many contexts, including manufacturing [21]. For example, in [7], a generic data analytics system is proposed and demonstrated to be viable for manufacturing production systems. Inventory forecasting, product evaluation, and machine tool condition monitoring are some of the use cases adopted in [7]. For these use cases, manufacturing tasks such as time-series prediction, rules extraction, and general prediction were adequately accomplished. However, in contrast to the work carried out in this paper where several machine learning methods have been investigated, only three machine learning methods were reported on in [7], as part of the experiments and investigations conducted. These methods are random forest, multilayer perceptron, and generalized linear model. Even though these three machine learning methods are arguably among the widely used and popular machine learning methods, there are several other widely-used and popular machine learning methods such as the ones investigated in this paper.

To efficiently evaluate carbon emission from manufacturing processes, particularly machining in manufacturing shop floors and/or workshops, a big data analysis approach involving a data-driven multi-level assessment is proposed in [8] as another instance of adopting ML in manufacturing environments. The work carried out and reported in [8] primarily involves three core stages that provide adequate insights into the pre-processing, correlation analysis, and multi-level data-driven formulation of evaluation indicators or responses of production state data from typical manufacturing shop floors or workshops. Note that the essential steps such as the cleansing of the manufacturing data, partitioning of the manufacturing processes, reduction of the manufacturing data, and extraction of features from the manufacturing data are all adequately captured and detailed in [8], as essential components of the proposed big data analysis approach. However, in contrast to the work presented in this paper, predictive modeling was not explicitly undertaken as part of the big data analysis approach in the manufacturing contexts detailed in [8].

More recently, to have more effective and efficient planning, scheduling, and management with predictive maintenance and product quality in multistage manufacturing or production systems comprising batching machines and finite buffers, a data-driven quantitative method is proposed in [22]. The method proposed in [22] premises on a decision model represented by a Markov decision process whose optimal maintenance policy is desired to satisfy Bellman's optimality equation. To overcome the dimensionality curse and solve the optimality equation, approximate dynamic programming implemented using a deep Q-network algorithm is employed in [22]. Even though the work carried out in [22] adequately and dynamically quantifies the impact of machine breakdowns, maintenance actions, and quality failures on manufacturing or production, only one algorithm was investigated and predictive modeling was not thoroughly explored to make a strong case for harnessing its potential in manufacturing environments in areas such as forecasting and trend analysis of manufacturing or production processes.

Similar work to the endeavor in [22] was reported in [10]. As a supposed improvement on the work carried out in [22], the exploration of optimal maintenance policies also represented by a Markov decision process is further augmented by considering machine stoppage bottlenecks on the shop or plant floors. To do this, two approaches involving the improvement of the worst machine and a random machine, respectively, were considered with a focus on the reduction of the downtime duration. Even though machine stoppage bottlenecks were considered alongside numerical analysis involving comparisons with popular and widely accepted methods by manufacturers in [10] in addition to the work carried out in [22], the drawback of not thoroughly exploring the potential of predictive modeling to make a strong case for its adoption, given the manufacturing and/or production context investigated remains. Hence, this can still be viewed arguably as a relatively grey area of research interest considering the potential impact of the adoption of predictive modeling in manufacturing or production environments.

Therefore, it is observed that currently, the available literature does not offer an in-depth investigative study where several machine learning algorithms or techniques are utilized in the same context to make a strong case for their adoption in predictive modeling in manufacturing environments. For example in [20], only five methods are investigated to make a case for optimal pick-and-place process controls that enhance the efficiency and quality of surface mount technology. The research investigations presented in this paper are devoted to this end, i.e., the investigation of multiple machine learning methods to make a case for their potential adoption in typical manufacturing processes and/or environments.

2.2. The Proposed Framework

The flow diagram of the proposed generalized framework for the adoption of regression learning in manufacturing environments investigated in this work is shown in Figure 1. The framework is both suitable for dynamic and/or static inflows of manufacturing data and some of the steps are iterative. In the context of this work, dynamic data is defined as data collected instantaneously and recursively from manufacturing processes and utilized on the fly for data-driven predictive modeling. In this way, the model is updated continuously and it feeds back, directly, to the manufacturing processes that produced the data used for its construction. It is envisaged that such a model can be employed for trend analysis and predictive maintenance in manufacturing environments. In contrast to dynamic data, static data in the context of this work is also defined as data collected instantaneously from manufacturing processes; however, it is not utilized on the fly for data-driven predictive modeling.

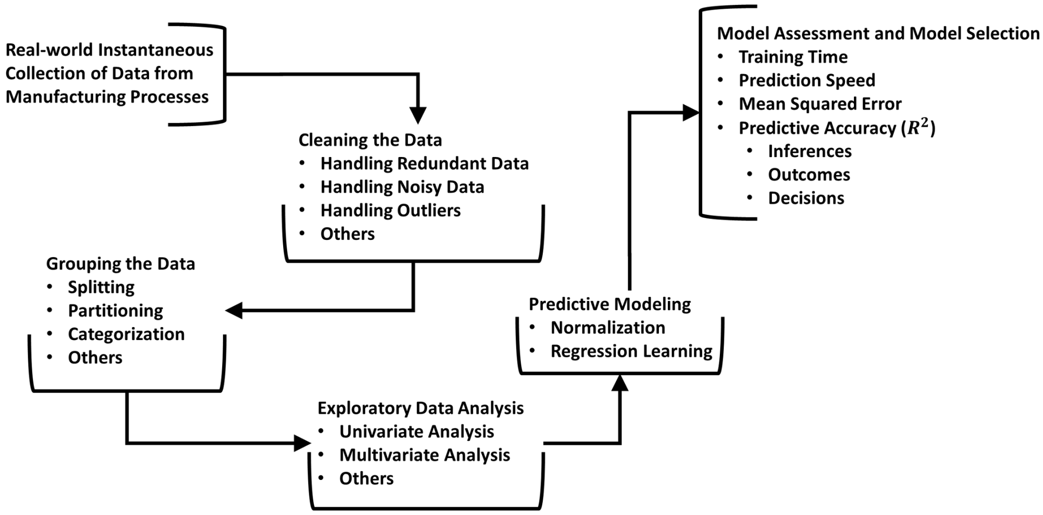

Figure 1. The proposed generalized framework.

To better distinguish between dynamic data and static data in this work, as already explained above, it is good to note that in practice, the first step to undertaking real-world predictive modeling is to utilize static data (as defined above) to build models that can predict responses from manufacturing processes. These models are then fine-tuned and thoroughly evaluated in terms of predictive accuracy, uncertainty, and other performance metrics before they are deployed for on-the-fly data-driven predictive modeling using dynamic data (as defined above). To validate the proposed framework, an open-source multi-stage continuous-flow manufacturing process data set has been used as an initial step to showcase its feasibility. Even though the multi-stage continuous-flow manufacturing process data set has been generated by collecting data and information from dynamic manufacturing processes [23], the models built using the data have not been deployed in real-world manufacturing environments for paradigms such as trend analysis and predictive maintenance. So, the framework proposed and investigated in this work primarily serves as a precursor to real-world physical implementation. The essential steps are described summarily as follows:

- **Step 1:** Data is collected instantaneously and/or recursively from critical manufacturing processes of interest. The collected data is used to create a database, M_{DB}.
- **Step 2:** Metrics for all explanatory and response variables in M_{DB}, E_{DB} and R_{DB}, respectively, are passed through purpose-built functions to generate a clean database, M_{DB}^{C} = [E_{DB}^{C}, R_{DB}^{C}], where unwanted observations or data (e.g., redundant data) are handled correctly and robustly. See Section 2.3.1 for more details on the custom functions adopted in this work.
- **Step 3:** Since the metrics in M_{DB}^{C} are expected to be from the continuous-flow stages of manufacturing processes, M_{DB}^{C} is grouped to ensure metrics collected from the same stage of the manufacturing process are collocated. For example, $M_{DB}^{C_1}$ (first stage) and $M_{DB}^{C_2}$ (second stage) for a two-stage continuous-flow manufacturing process as we have investigated in this work (See Section 2.3).
- **Step 4:** Exploratory data analysis techniques are employed for the univariate analysis and multivariate analysis of the metrics in M_{DB}^{C} to establish their descriptive statistics and relationships. In this work, only descriptive statistics are presented.

- **Step 5:** The explanatory metrics in M^C_{DB} (i.e., E^C_{DB}) are then normalized for regression learning to have a new database (i.e., $E^C_{DB_{norm}}$ as carried out in this work (See Section 2.4)).
- **Step 6:** Predictive modeling is carried out by applying several regression learning techniques to $[E^C_{DB_{norm}}, R^C_{DB}]$ to ascertain the most suitable ones in terms of the mean squared error (MSE), predictive accuracy (R-squared value (R^2)), prediction speed (PS), and training time (TT).
- **Step 7:** To rank all methods, some inferences are drawn a posteriori based on the assessment of the outcomes from Step 6.
- **Step 8:** The inference-based rankings from Step 7 are statistically verified to determine the best regression learning technique(s) to select and adopt for the given manufacturing process data set.

More details about the implementation of the proposed framework are presented in Sections 2.3, 2.4 and 3, where the case study involving an open-source multi-stage continuous-flow manufacturing process data set is further described [23], the procedures in all the steps summarized above are detailed, and the experimental setup is presented, respectively. Note that, for such a data-driven paradigm concocted by the proposed framework, the end product will be a software application that is deployed either as an add-on or a toolbox for existing manufacturing process automation, monitoring, and control software applications. Naively, the software application can be fed data (mostly from sensors interfaced and integrated with the manufacturing process as described in Section 2.3) and using this data, it implements the steps described above for the proposed framework to build and select near-optimal regression learning models for the behavioral analysis and predictive modeling of the manufacturing process that generated its input data. The behavioral analysis and predictive modeling of the manufacturing process can then be adopted for predictive maintenance, trend analysis, and optimal tuning of the manufacturing process.

2.3. Case Study of a Multi-Stage Continuous-Flow Manufacturing Process

A case study involving a multi-stage continuous-flow manufacturing process is used to validate the proposed framework in Section 2.2 and the architecture of the manufacturing process is shown in Figure 2. In contrast to batch production or manufacturing process, the materials being processed in this case study are constantly in motion on the shop floor along the production or manufacturing process line. The main elements of the multi-stage continuous flow manufacturing process are five operational machines ($M1$, $M2$, $M3$, $M4$, and $M5$) and a combiner ($COMB$) that are employed in two stages (see Figure 2). Sensors have been used for data (temperature, pressure, speed, and others) collection [23]. Some of the data collected include (but are not limited to) factory ambient conditions, temperature, pressure, speed, and amperage of the machines. Since the goal of our work is to primarily report on predictive modeling, the manufacturing process and the manufacturing environment used as the case study have not been extensively discussed. However, more details about the manufacturing process and manufacturing environment can be found in [23] or by contacting [24].

Considering the fact that the data collected from the two-stage continuous-flow manufacturing process is over a four-hour operational window, in comparison to the operational hours of real-world manufacturing processes, this may be viewed as being relatively short in some contexts, e.g., continuous manufacturing where the production lines are operational all the time and the available manufacturing process data can be infinitely large. The operating conditions may also be assumed to be relatively stable due to the minimal or no noise observed in the collected data set. In reality, the operational window would be longer and unstable conditions may be observed from time to time, for example, due to mechanical vibrations, and friction noise, after long operational hours of rotating equipment such as motors. Be that as it may, the four-hour operational window can still be argued to emulate the multi-stage continuous-flow manufacturing process reasonably well based on the half-hour to five-hour sampling window recommended in [25] for modeling a typical manufacturing process.

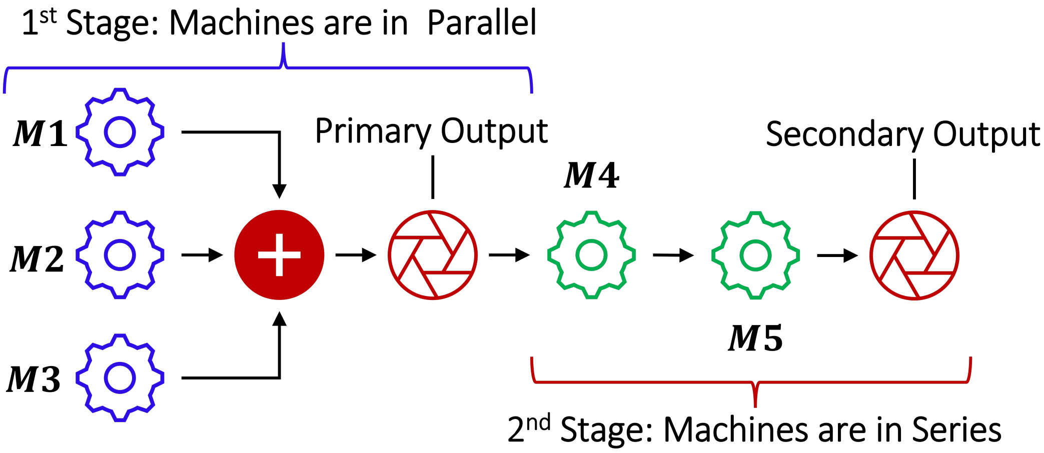
1st Stage: Machines are in Parallel
M1
M2
M3
Primary Output
M4
Secondary Output
M5
2nd Stage: Machines are in Series

Figure 2. Architecture of the manufacturing process used as a case study.

To deploy data-driven predictive modeling systems in real-world manufacturing scenarios, only a subset of the manufacturing data is used per time (and not an infinitely large data set) due to limitations in hardware and computing resources, standard data augmentation techniques can then be employed to extrapolate the data collected over relatively short operational times to account for longer operational hours and better capture the dynamics of the shop floors. Noisy data can be both studied and generated synthetically from the big data to mirror or model current and potential noise patterns in the data set to better account for unstable conditions that may stem from mechanical vibrations and others, after long hours of operation. These approaches require additional studies, investigations, experiments, and possibly a miniaturized manufacturing process model (for example, a model such as the PETRA II [26]) to validate their practicality. Since this is beyond the scope of the work presented in this paper, studies, investigations, and experiments, in this regard, are neither presented nor discussed in this work.

With a focus on the data components of the two-stage continuous flow manufacturing process, the first stage comprises three machines ($M1$, $M2$, and $M3$) that operate in parallel to constitute the parallel stage of the multi-stage continuous flow manufacturing process. $M1$, $M2$, and $M3$ each have eight process variables to account for properties such as temperature, speed, pressure, amperage, and others. The outputs from $M1$, $M2$, and $M3$ are fed into a step that combines the flows. This combiner stage for $M1$, $M2$, and $M3$ has three process variables associated with it) and its output is the primary controlled output. The primary controlled output is measured at 15 unique locations. The combined output from the first stage is then fed to the second stage of the continuous-flow manufacturing process. The second stage comprises two machines ($M4$ and $M5$) that operate in series to constitute the series stage of the multi-stage continuous-flow manufacturing process. $M4$ and $M5$ each have seven process variables to also account for properties such as temperature, speed, pressure, amperage, and others. The output from the second stage is the secondary controlled output and it is also measured at 15 unique locations. The process variables and the controlled outputs used for generating the data set are further discussed in Section 2.3.1.

2.3.1. Data Set Description

The data used for the experimentation and investigations come from a single production or manufacturing process run spanning several hours. Large quantities of this type of data are available from [24], and they are often collected from multiple production lines in various locations to mitigate bias and support the generalization of the data set. In our experiments, a total of over 14,000 observations have been used. If each observation (0_i) is defined as a proper subset of the set, M^C_{DB} (database of the clean data (see Section 2.2)), such that $i \in [1, \ldots, 14{,}088]$, the data set can be described mathematically as follows:

$$0_i \subseteq M^C_{DB} \tag{1}$$

where M^C_{DB} is generated from the data cleaning process detailed as follows:

1 For every observation in M_{DB}, select all the columns for the explanatory variables to have an array or matrix E_{DB} and select all the columns for the response variables to have an array or matrix R_{DB}.

2 For every column vector (E^j_{DB}) in E_{DB}, perform the following operations:

$$E^C_{DB} = \begin{cases} \text{Obtain the mode of } E^j_{DB} & \text{Store as } EMode \\ \text{Determine the frequency of } EMode \text{ in } E^j_{DB} & \text{Store as } EFreq. \\ \text{If } EFreq. \geq THOLD \text{ remove } E^j_{DB} \text{ from } E_{DB} & \text{Otherwise keep } E^j_{DB} \text{ in } E_{DB} \end{cases} \tag{2}$$

3 For every column vector (R^j_{DB}) in R_{DB}, perform the following operations:

$$R^C_{DB} = \begin{cases} \text{Obtain the mode of } R^j_{DB} & \text{Store as } RMode \\ \text{Determine the frequency of } RMode \text{ in } R^j_{DB} & \text{Store as } RFreq. \\ \text{If } RFreq. \geq THOLD \text{ remove } R^j_{DB} \text{ from } R_{DB} & \text{Otherwise keep } R^j_{DB} \text{ in } R_{DB} \end{cases} \tag{3}$$

where $THOLD$ is set to be equal to 6000 and $M^C_{DB} = \{E^C_{DB}, R^C_{DB}\}$.

Note that by setting $THOLD = 6000$ in Equations (2) and (3), if any metric is repeated for 6000 times or more (i.e., representing 40% or more) in any column of the metrics for the explanatory and response variables, such a variable is assumed to be fairly constant and discarded (not included in the data set). After the data cleaning operations, the variables in M^C_{DB} are mathematically described as follows for any ith observation having explanatory variables 0^E_i and response variables 0^R_i in M^C_{DB}:

$$0^E_i = \begin{cases} \{F^1_{AC}, F^2_{AC}\};\ \text{Factory ambient conditions.} \\ \{M1^1_{P_V}, M1^2_{P_V}, M1^3_{P_V}, M1^4_{P_V}, M1^5_{P_V}, M1^6_{P_V}, M1^7_{P_V}, M1^8_{P_V}\};\text{M1 process variables.} \\ \{M2^1_{P_V}, M2^2_{P_V}, M2^3_{P_V}, M2^4_{P_V}, M2^5_{P_V}, M2^6_{P_V}, M2^7_{P_V}, M2^8_{P_V}\};\text{M2 process variables.} \\ \{M3^1_{P_V}, M3^2_{P_V}, M3^3_{P_V}, M3^4_{P_V}, M3^5_{P_V}, M3^6_{P_V}, M3^7_{P_V}, M3^8_{P_V}\};\text{M3 process variables.} \\ \{T^1_{COMB}, T^2_{COMB}, T^3_{COMB}\};\ \text{Combiner stage process variables for M1, M2 and M3.} \\ \{M4^1_{P_V}, M4^2_{P_V}, M4^3_{P_V}, M4^4_{P_V}, M4^5_{P_V}, M4^6_{P_V}, M4^7_{P_V}\};\text{M4 process variables.} \\ \{M5^1_{P_V}, M5^2_{P_V}, M5^3_{P_V}, M5^4_{P_V}, M5^5_{P_V}, M5^6_{P_V}, M5^7_{P_V}\};\text{M5 process variables.} \end{cases} \tag{4}$$

$$0^R_i = \begin{cases} \{S1^1_{C_O}, S1^2_{C_O}, S1^3_{C_O}, S1^4_{C_O}, S1^5_{C_O}, S1^6_{C_O}, S1^7_{C_O}, S1^8_{C_O}, S1^9_{C_O}, S1^{10}_{C_O}, S1^{11}_{C_O}, S1^{12}_{C_O}, \ldots \\ S1^{13}_{C_O}, S1^{14}_{C_O}, S1^{15}_{C_O}\};\text{Primary controlled output.} \\ \{S2^1_{C_O}, S2^2_{C_O}, S2^3_{C_O}, S2^4_{C_O}, S2^5_{C_O}, S2^6_{C_O}, S2^7_{C_O}, S2^8_{C_O}, S2^9_{C_O}, S2^{10}_{C_O}, S2^{11}_{C_O}, S2^{12}_{C_O}, \ldots \\ S2^{13}_{C_O}, S2^{14}_{C_O}, S2^{15}_{C_O}\};\text{Secondary controlled output.} \end{cases} \tag{5}$$

From Equations (4) and (5), it can be seen that there are a total of 43 explanatory variables and 30 response variables. For the predictive modeling of the primary controlled output, the factory ambient conditions (i.e., $[F^1_{AC}, F^2_{AC}]$), the process variables of $M1$, $M2$ and $M3$ (i.e., $[M1^1_{P_V}, \ldots, M1^8_{P_V}]$, $[M2^1_{P_V}, \ldots, M2^8_{P_V}]$, and $[M3^1_{P_V}, \ldots, M3^8_{P_V}]$, respectively) and the combiner stage process variables (i.e., $[T^1_{COMB}, T^2_{COMB}, T^3_{COMB}]$) are the explanatory variables, and the primary controlled output (i.e., $[S1^1_{C_O}, \ldots, S1^{15}_{C_O}]$) is the response to be predicted. The descriptive statistics for these variables are given in Tables 1 and 2. For the predictive modeling of the secondary controlled output, the factory ambient conditions (i.e.,

[F^1_{AC}, F^2_{AC}] and the process variables of $M4$ and $M5$ (i.e., [$M4^1_{P_V}$,…, $M4^7_{P_V}$], and [$M5^1_{P_V}$,…, $M5^7_{P_V}$], respectively) are the explanatory variables, and the secondary controlled output (i.e., [$S2^1_{C_O}$, …, $S2^{15}_{C_O}$]) is the response. The descriptive statistics for these variables are given in Tables 1 and 2.

Table 1. Descriptive statistics of the explanatory variables (non-normalized) from the data set.

Explanatory Variable	Mean	Median	Minimum	Maximum	S.D.
First Condition: F^1_{AC}	15.3	15.1	13.8	17.2	1.2
Second Condition: F^2_{AC}	23.8	23.9	23.0	24.4	0.4
First Stage: $M1^1_{P_V}$	1242.8	1264.4	231.3	1331.8	95.8
First Stage: $M1^2_{P_V}$	72.0	72.0	71.9	72.5	0.1
First Stage: $M1^3_{P_V}$	72.0	72.0	71.3	72.7	0.4
First Stage: $M1^4_{P_V}$	70.3	72.0	44.4	88.5	5.5
First Stage: $M1^5_{P_V}$	11.1	10.7	10.4	12.2	0.6
First Stage: $M1^6_{P_V}$	409.0	417.0	359.5	487.2	20.5
First Stage: $M1^7_{P_V}$	81.5	81.3	76.3	83.9	0.9
First Stage: $M1^8_{P_V}$	76.0	75.0	69.7	80.0	2.1
First Stage: $M2^1_{P_V}$	202.6	203.3	0.0	266.5	15.1
First Stage: $M2^2_{P_V}$	69.0	69.0	68.7	69.7	0.1
First Stage: $M2^3_{P_V}$	69.1	69.1	67.8	69.9	0.1
First Stage: $M2^4_{P_V}$	73.4	73.4	71.6	74.4	0.4
First Stage: $M2^5_{P_V}$	13.9	13.9	13.8	14.0	0.0
First Stage: $M2^6_{P_V}$	226.1	226.1	218.9	250.6	3.1
First Stage: $M2^7_{P_V}$	76.8	77.0	68.8	77.4	0.9
First Stage: $M2^8_{P_V}$	60.0	60.0	59.6	60.5	0.2
First Stage: $M3^1_{P_V}$	202.4	202.9	0.0	259.1	15.7
First Stage: $M3^2_{P_V}$	78.0	78.0	77.3	78.7	0.1
First Stage: $M3^3_{P_V}$	78.0	78.0	77.7	78.6	0.1
First Stage: $M3^4_{P_V}$	345.1	342.9	321.2	374.3	9.1
First Stage: $M3^5_{P_V}$	13.3	13.4	12.0	14.0	0.4
First Stage: $M3^6_{P_V}$	246.8	247.5	235.1	263.7	6.1
First Stage: $M3^7_{P_V}$	74.1	75.1	65.3	75.5	2.1
First Stage: $M3^8_{P_V}$	65.0	65.0	64.8	65.2	0.1
First Stage: T^1_{COMB}	108.9	105.5	45.3	118.9	5.7
First Stage: T^2_{COMB}	84.9	74.8	53.3	115.2	18.6
First Stage: T^3_{COMB}	80.0	80.0	79.6	80.3	0.1
Second Stage: $M4^1_{P_V}$	360.1	360.0	298.0	393.0	2.3
Second Stage: $M4^2_{P_V}$	360.1	360.0	284.0	396.0	3.0
Second Stage: $M4^3_{P_V}$	17.2	17.0	14.0	25.0	0.9
Second Stage: $M4^4_{P_V}$	322.6	324.0	268.0	327.0	3.7
Second Stage: $M4^5_{P_V}$	17.2	17.0	14.0	25.0	0.9
Second Stage: $M4^6_{P_V}$	309.8	311.0	260.0	326.0	2.9
Second Stage: $M4^7_{P_V}$	187.1	192.0	35.0	216.0	23.7
Second Stage: $M5^1_{P_V}$	310.0	310.0	309.4	310.3	0.0
Second Stage: $M5^2_{P_V}$	290.0	290.0	289.7	290.3	0.1
Second Stage: $M5^3_{P_V}$	269.7	270.0	263.8	270.0	1.0
Second Stage: $M5^4_{P_V}$	242.7	242.7	237.6	245.0	1.6
Second Stage: $M5^5_{P_V}$	245.0	245.0	242.9	245.7	0.1
Second Stage: $M5^6_{P_V}$	63.4	63.4	62.8	66.1	0.4
Second Stage: $M5^7_{P_V}$	154.0	155.6	45.4	159.2	10.3

Table 2. Descriptive statistics of the response variables (non-normalized) from the data set.

Response Variable	Mean	Median	Minimum	Maximum	S.D.
First Stage: $S1^{1}_{CO}$	12.9	13.0	0.0	20.9	0.9
First Stage: $S1^{2}_{CO}$	8.1	13.2	−3.1	19.1	6.9
First Stage: $S1^{3}_{CO}$	11.4	11.3	−4.9	23.5	1.1
First Stage: $S1^{4}_{CO}$	21.3	21.5	0.0	26.2	2.1
First Stage: $S1^{5}_{CO}$	32.9	33.5	−7.7	34.8	3.9
First Stage: $S1^{6}_{CO}$	0.1	0.0	−0.6	5.0	0.6
First Stage: $S1^{7}_{CO}$	1.3	1.6	−0.8	7.0	1.1
First Stage: $S1^{8}_{CO}$	1.1	0.0	−0.8	5.2	1.4
First Stage: $S1^{9}_{CO}$	19.8	20.9	-6.6×10^{-10}	22.5	4.8
First Stage: $S1^{10}_{CO}$	18.0	18.9	-3.6×10^{-3}	20.4	4.2
First Stage: $S1^{11}_{CO}$	7.7	7.8	-1.4×10^{-3}	13.1	1.1
First Stage: $S1^{12}_{CO}$	1.5	0.0	-1.6×10^{-20}	7.5	2.5
First Stage: $S1^{13}_{CO}$	1.2	1.5	-3.5×10^{-4}	4.0	0.7
First Stage: $S1^{14}_{CO}$	2.9	3.2	−1.2	6.9	0.9
First Stage: $S1^{15}_{CO}$	9.9	15.0	−6.6	22.3	7.4
Second Stage: $S2^{1}_{CO}$	11.7	12.8	-9.8×10^{-137}	19.1	3.6
Second Stage: $S2^{2}_{CO}$	6.3	6.5	-3.0×10^{-5}	12.9	1.6
Second Stage: $S2^{3}_{CO}$	10.3	10.9	-3.3×10^{-107}	16.5	2.3
Second Stage: $S2^{4}_{CO}$	19.3	20.6	-6.6×10^{-107}	25.2	4.7
Second Stage: $S2^{5}_{CO}$	2.9	0.0	-1.4×10^{-13}	34.3	9.2
Second Stage: $S2^{6}_{CO}$	2.7	2.7	-3.5×10^{-95}	8.1	0.4
Second Stage: $S2^{7}_{CO}$	0.5	0.6	-3.7×10^{-105}	3.3	0.2
Second Stage: $S2^{8}_{CO}$	2.9	3.0	-1.3×10^{-109}	7.4	0.5
Second Stage: $S2^{9}_{CO}$	18.4	19.7	-3.0×10^{-104}	24.8	5.0
Second Stage: $S2^{10}_{CO}$	11.6	16.6	-3.7×10^{-3}	18.4	7.6
Second Stage: $S2^{11}_{CO}$	7.5	7.9	-5.2×10^{-95}	8.6	1.6
Second Stage: $S2^{12}_{CO}$	5.4	5.6	-2.7×10^{-95}	6.3	1.2
Second Stage: $S2^{13}_{CO}$	2.0	2.1	-6.2×10^{-96}	5.2	0.4
Second Stage: $S2^{14}_{CO}$	3.5	3.5	-1.8×10^{-95}	8.0	0.5
Second Stage: $S2^{15}_{CO}$	7.5	7.9	−3.4	14.3	2.1

From the values reported in Tables 1 and 2, it can be said that the explanatory and response variables all have a significant spread. This is an indication that for every 0_i, the elements all have values that are not necessarily close to their means (i.e., expected values) which are spread over a wider range. To improve the performance and training stability of the regression models to be built using the variables described in Tables 1 and 2, it is essential to transform the features (i.e., specifically, the explanatory variables described in Table 1) to be on a similar scale. In this work, normalization, a popular machine learning approach has been used and it is discussed as follows:

Feature Scaling of the Data Set

Feature scaling is a widely used data pre-processing procedure in machine learning when supervised learning techniques such as regression and classification are employed [27,28]. It is often carried out through the normalization of the range of the explanatory variables or independent variables (also called features) in the data set. Normalization mainly guarantees that the numerical weights contributed approximately and proportionally to the responses or targets by each explanatory variable (feature) are relatively similar [28]. In this way, data redundancy is minimized and the variables have an agreeable metric scale.

In a typical machine learning process, feature scaling can be implemented in a number of ways [27,28]. For example, *z*-scores as carried out in [29] and min-max linear transformation as carried out in [30,31]. In this work, feature scaling has been implemented on the explanatory variables (see Equation (4) and its description) using a min-max linear transformation (normalization) as follows:

$$E_{DB_{norm}}^{C_R^i} = \frac{E_{DB}^{C_R^i} - E_{DB_{min}}^{C_R^i}}{E_{DB_{max}}^{C_R^i} - E_{DB_{min}}^{C_R^i}} \tag{6}$$

where $E_{DB}^{C_R^i}$ is the ith metric in the data column for the Rth variable in E_{DB}^{C} whose minimum and maximum values are $E_{DB_{min}}^{C_R^i}$ and $E_{DB_{max}}^{C_R^i}$, respectively, and $E_{DB_{norm}}^{C_R^i}$ is the normalized value defined as $E_{DB_{norm}}^{C_R^i} \in [0,1]$. After the normalization, the transformed data set ($M_{DB_{norm}}^{C}$) can be described mathematically as follows:

$$M_{DB_{norm}}^{C} = \{E_{DB_{norm}}^{C}, R_{DB}^{C}\} \tag{7}$$

2.4. The Machine Learning Problem and Investigated Methods

In this work, several specialized machine learning methods (including methods having similar model construction methodologies but different hyperparameter settings or control parameter settings) that support outputting multiple response variables for each prediction are investigated. These algorithmic methods are demonstrated to be practical and useful for undertaking predictive modeling in manufacturing environments as exemplified in this work. Using the context of the case study presented in Section 2.3, predictive modeling problem definition, feature scaling of the data set, regression learning, and summarized description of the investigated machine learning methods are detailed as follows:

2.4.1. Predictive Modeling Problem Definition for the Case Study

Briefly, predictive modeling can be best described as a paradigm that adopts both machine learning and data mining techniques to predict or forecast some target data or responses given a set of observations. Considering the case study presented in Section 2.3, to predict the manufacturing process line's outputs (i.e., [$S1_{C_O}^{1}$, ..., $S1_{C_O}^{15}$] for the first stage and [$S2_{C_O}^{1}, \ldots, S2_{C_O}^{15}$] for the second stage), machine learning methods or algorithms that adequately support multi-target data are required. Intuitively, this is because the manufacturing process data set (as described in Section 2.3) contains multi-target data from the two stages (see Figure 2). Note that multi-target learning subsumes many machine learning problems in several disciplines and handles complex decision-making in many real-world applications such as manufacturing processes.

The complexity of multi-target learning mainly stems from its intrinsic multivariate nature that yields complex interactions between its multiple explanatory and response variables. These interactions have not been investigated in this work as exemplified in [30] using typical explanatory variables because it is not within the scope of the work carried out. However, since both the explanatory and response variables in $M_{DB_{norm}}^{C}$ are continuous variables (see Tables 1 and 2, and Section 2.3), regression learning-based predictive modeling must be employed to address the machine learning problem. This is the focus of the work carried out in this paper and regression learning for the case study is discussed in Section 2.4.2 as follows:

2.4.2. Regression Learning for the Case Study

Regression learning is a widely used supervised learning technique for machine learning-based predictive modeling [32]. As a data-driven methodology, it generally takes the form of linear regression analysis for the conventional analysis of data [33]. Primarily, it is used for forecasting outcomes in a methodical and mathematical way that sorts the impact of the independent feature(s) or variable(s) on the associated dependent response(s) or variable(s) in any given data set. The regression learning implementation in this work can be mathematically described as follows for the two stages of the continuous-flow manufacturing process:

$$S1_{C_O}^{m_i} = f(F_{AC}^{norm_i^n}, M1_{P_V}^{norm_i^n}, M2_{P_V}^{norm_i^n}, M3_{P_V}^{norm_i^n}, T_{COMB}^{norm_i^n}, \beta) + e_i \tag{8}$$

$$S2_{C_O}^{m_i} = f(F_{AC}^{norm_i^n}, M4_{P_V}^{norm_i^n}, M5_{P_V}^{norm_i^n}, \beta) + e_i \tag{9}$$

In Equation (8), $S1_{C_O}^{m_i^{m=[1,\ldots,15]}}$, $F_{AC}^{norm_i^{n=[1,2]}}$, $M1_{P_V}^{norm_i^{n=[1,\ldots,8]}}$, $M2_{P_V}^{norm_i^{n=[1,\ldots,8]}}$, $M3_{P_V}^{norm_i^{n=[1,\ldots,8]}}$, $T_{COMB}^{norm_i^{n=[1,2,3]}}$, β and e_i are the primary controlled outputs, normalized factory ambient conditions, normalized process variables for $M1$, normalized process variables for $M2$, normalized process variables for $M3$, normalized process variables for the combiner stage of $M1$, $M2$, and $M3$, unknown parameters (usually, scalar coefficients) and error terms (usually, scalar), respectively, for the ith observation in $M_{DB_{norm}}^{C}$. m is the total number of occurrences of the primary controlled output in $M_{DB_{norm}}^{C}$ and n is the total number of process variables from the factory ambient conditions, each of $M1$, $M2$, and $M3$, and the combiner stage for $M1$, $M2$, and $M3$. Equation (8) is then used to predict $S1_{C_O}$ for new or arbitrary values of F_{AC}^{norm}, $M1_{P_V}^{norm}$, $M2_{P_V}^{norm}$, $M3_{P_V}^{norm}$ and T_{COMB}^{norm}. In Equation (9), $S2_{C_O}^{m_i^{m=[1,\ldots,15]}}$, $F_{AC}^{norm_i^{n=[1,2]}}$, $M4_{P_V}^{norm_i^{n=[1,\ldots,8]}}$, $M5_{P_V}^{norm_i^{n=[1,\ldots,8]}}$, β and e_i are the secondary controlled outputs, normalized factory ambient conditions, normalized process variables for $M4$, normalized process variables for $M5$, respectively, for the ith observation in the $M_{DB_{norm}}^{C}$. m is the total number of occurrences of the secondary controlled output in $M_{DB_{norm}}^{C}$ and n is the total number of process variables from the factory ambient conditions and each of $M4$ and $M5$. Equation (9) is then used to predict $S2_{C_O}$ for new or arbitrary values of F_{AC}^{norm}, $M4_{P_V}^{norm}$, and $M5_{P_V}^{norm}$.

It is good to note that all the machine learning methods investigated in this work have been applied for regression learning, in consonance with the multi-stage continuous-flow manufacturing process data set described in Section 2.3.1. In other words, they have been used for regression analysis. As a result, their respective approximate computational complexities have been estimated using the context of a typical linear regression algorithm. The implementation of linear regression mainly focuses on finding a solution to the matrix or linear algebra problem described as follows:

$$(YY')^{-1} \times (Y'Z) \tag{10}$$

where Y is the matrix holding the predictors or explanatory variables or independent variables and it is a (14,088 × 29) matrix and a (14,088 × 16) matrix for the first stage and second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, respectively, as discussed in Section 2.3.1. Y' is the transpose of Y, and Z is the matrix holding the responses or target variables, or independent variables. Z is a (14,088 × 15) matrix for both the first and second stages of the predictive modeling of the multi-stage continuous-flow manufacturing process.

For the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, Y holds the metrics of the factory ambient conditions (i.e., $[F_{AC}^1,F_{AC}^2]$) and the following process variables (see Table 1): $[M1_{P_V}^1,\ldots, M1_{P_V}^8]$, $[M2_{P_V}^1,\ldots, M2_{P_V}^8]$, $[M3_{P_V}^1,\ldots, M3_{P_V}^8]$ and $[T_{COMB}^1, T_{COMB}^2, T_{COMB}^3]$. For the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, Y holds the metrics of the factory ambient conditions (i.e., $[F_{AC}^1, F_{AC}^2]$) and the following process variables (see Table 1): $[M4_{P_V}^1,\ldots, M4_{P_V}^7]$, and $[M5_{P_V}^1,\ldots, M5_{P_V}^7]$. For the first stage of the predictive modeling, Z holds the metrics of the primary output to control, i.e., $[S1_{C_O}^1,\ldots, S1_{C_O}^{15}]$ (see Table 2). For the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, Z holds the metrics of the secondary output to control, i.e., $[S2^{1}_{C_O}, \ldots, S2^{15}_{C_O}]$ (see Table 2).

By examining Equation (10) and the dimensions of Y and Z, respectively, the matrix product $Y \times Y'$ can be estimated to have a complexity of $\mathcal{O}(29^2 \times 14{,}088)$ and $\mathcal{O}(16^2 \times 14{,}088)$ for the first stage and second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, respectively. The inverse operation of $(Y \times Y')^{-1}$ can be estimated to have a complexity of $\mathcal{O}(29^3)$ and $\mathcal{O}(16^3)$ for the first stage and second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, respectively. The product $Y' \times Z$ can be estimated to have a complexity of $\mathcal{O}(29 \times 14{,}088)$ and $\mathcal{O}(16 \times 14{,}088)$ for the first stage and second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, respectively. Hence, the overall computational complexity can be estimated to be $\mathcal{O}(14, 088 \times 29^2 + 29^3)$ and $\mathcal{O}(14{,}088 \times 16^2 + 16^3)$ for the first stage and second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, respectively. As such for any M_n total number of observations from the manufacturing process having any M_p metrics, the computational complexity may be generalized and estimated as follows:

$$\mathcal{O}(M_n \times M_p^2 + M_p^3) \tag{11}$$

where M_n and M_p will assume their values according to the given linear regression problem.

The regression learning methods or regression techniques investigated in this work are listed in Table 3 and references are also provided for their full descriptions. These methods include variants of multi-output regression (MOR) in which different regression algorithms have been used as estimators and k-nearest neighbors (k-NN) where the implementation is carried out using four different values of k, 1, 5, 10, and 20 to have NN, k-NN5, k-NN10 and k-NN20, respectively.

Table 3. The regression learning methods investigated.

Method	**References**
MOR-based Support Vector Regression (MOR-SVR)	[34–36]
MOR-based Linear Support Vector Regression (MOR-LSVR)	[34–36]
MOR-based *Nu* Support Vector Regression (MOR-NuSVR)	[34,36,37]
MOR-based Gradient Boosting Regression (MOR-GBoostR)	[34,36,38]
MOR-based Extreme Gradient Boosting Regression (MOR-XGBoostR)	[34,36,38,39]
MOR-based Extreme Gradient Boosting Random Forest Regression (MOR-XGBoostRFR)	[34,36,40]
MOR-based Bayesian Ridge Regression (MOR-BRR)	[34,36,41]
k-Nearest Neighbors (k-NN)	[36,42,43]
Random Forest Regressor (RFR)	[36,40,44]
Decision Trees Regressor (DTR)	[36,45]
Extra-Trees Regressor (ETR)	[36,46]
Ridge Regression (RR)	[36,47]
Ridge Regression with Built-In Cross-Validation (RRCV)	[36,48]
Multi-Task Least Absolute Shrinkage and Selection Operator with Cross-Validation (MTLCV)	[36,49]

Table 3. *Cont.*

Method	References
Least Absolute Shrinkage and Selection Operator Model Fit Least Angle Regression (LLARS)	[36,50]
Linear Regression (LR)	[36,51]
Bagging Regressor (BR)	[36,52]
Multilayer Perceptron Regressor (MLPR)	[36,53]

3. Experimental Setup

For the implementation of all the machine learning methods, Scikit-learn library [36,54] and Google Colaboratory development environment [55] have been used. Specifically, the investigated algorithms have been coded using the library functions available from [36,54]. All the settings for the hyperparameters or control parameters for the respective methods are the Scikit-learn library's default settings, except where stated otherwise. The assumption made is that these default settings in the Scikit-learn library are unlikely to be altered by most manufacturing engineers, not experienced in machine learning. Conventionally, most machine learning methods or algorithms show better performance if and/or when their hyperparameters or control parameters are tuned or optimized [56,57]. However, since this is not the aim of this work, this has not been investigated.

In this work, some exceptions regarding hyperparameter tuning investigations can be found in methods such as k-NN (k-nearest neighbor), where different values have been used for the number of neighbors to have different approaches for the same method. Other methods such as RF (random forest) and SVR (support vector regression) that can be argued not to naturally support multi-target regression have been fitted using the multi-output regressor (MOR) in the Scikit-learn library to have different adaptations (or tuning, naively) to allow for multi-target regression using different approaches with similar underlying frameworks. These methods can also be viewed as an exception in terms of hyperparameter tuning. To examine the predictive accuracy of the fitted models from all methods, the validation scheme used is hold-out validation. The choice of this validation scheme is guided by the total number of observations (more than 14,000 (see Section 2.3)). Following some popular approaches for typical hold-out validation schemes [58], 70% of the observations have been used as the training data set and the remaining 30% have been used as the test data set for all experiments.

To compare the regression models built and trained by all methods, the training time (TT), prediction speed (PS), predictive accuracy (R-squared value (R^2)), and mean squared error (MSE) of all methods are used over 50 independent statistical runs. In each independent run, a regression model is built and trained for each of the methods investigated. Note that the 50 independent runs allow for statistical inference and analysis, and comparisons of the efficiencies and robustness of the investigated methods in terms of their performance metrics (TT, PS, R^2, and MSE). Also, by having a sample size of 50, hypothesis tests for probability estimation using a z-statistic can be carried out (See Section 4). All experiments have been carried out on a workstation with Intel 4-core i7-4770K 3.50 GHz CPU and 24.0 GB RAM, except where stated otherwise. Elapsed times reported are elapsed real times from a wall clock.

4. Results and Discussion

In this section, the results of all methods are presented, discussed, and compared in terms of TT, PS, R^2, and MSE. The comparisons are made to establish the most suitable method for each stage of the predictive modeling of the multi-stage continuous-flow manufacturing process. The methods are ranked inferentially using their performance metrics (TT, PS, R^2, and MSE) from their independent runs, and the rankings are statistically verified for validation.

4.1. Training Time

In regression learning, the training time (TT) can be best described as the time taken to build and train a regression model. TT is typically measured in seconds or fractions of seconds (e.g., milliseconds, microseconds, nanoseconds, etc.). From Table 4, the following observations are made for the TT of all methods for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process: (1) The TT of all methods are generally lower than 120 s on average. (2) *k*-NN20 shows the best average and median TT of 4.8 ms and 4.9 ms, respectively, over the 50 independent runs. (3) The average and median TT for NN and *k*-NN10 (4.9 ms and 5.0 ms, and 5.0 ms and 5.1 ms, respectively) over the 50 independent runs are also short and comparable to those of *k*-NN20. (4) MTLCV shows the worst average and median TT of 60,883.6 ms and 60,837.8 ms, respectively, over the 50 independent runs. (5) The average and median TT for BR and MLPR (58,489.6 ms and 58,287.5 ms, and 42,700.9 ms and 42,682.0 ms, respectively) over the 50 independent runs are also long. (6) The robustness of the methods is relative according to their standard deviations over all runs: *k*-NN10 has the lowest standard deviation of 0.3 and MTLCV has the highest standard deviation of 2035.8. So, in terms of TT stability for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, *k*-NN10 can be said to be the most robust among all the methods, while MTLCV is the least robust. The box plots of TT for all the methods are shown in Figure 3, for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process. The findings earlier discussed can be corroborated visually by an examination of the variations in TT for all the methods in Figure 3.

Table 4. Training time (in milliseconds (ms)) for all methods over 50 statistical runs (the first stage).

Method	Best	Worst	Average	Median	S.D.
MOR-SVR	4575.9	6671.0	4827.9	4727.1	400.7
MOR-LSVR	40,305.4	48,648.8	42,700.9	42,682.0	1434.7
MOR-NuSVR	113,936.0	123,388.0	117,689.1	117,561.5	1963.4
MOR-GBoostR	11,168.3	14,176.8	13,158.5	13,318.5	673.0
MOR-XGBoostR	13,315.4	16,089.9	13,712.1	13,588.3	533.6
MOR-XGBoostRFR	14,293.3	15,382.8	14,749.8	14,687.6	239.4
MOR-BRR	319.4	344.6	331.8	331.3	5.5
NN	1.8	6.3	4.9	5.0	0.5
***k*-NN5**	3.5	5.8	5.1	5.2	0.4
***k*-NN10**	3.8	5.4	5.0	5.1	0.3
***k*-NN20**	3.6	5.4	4.8	4.9	0.4
RFR	7006.4	7704.1	7206.6	7170.0	158.5
ETR	1200.7	1245.8	1213.4	1211.5	8.5
DTR	57.4	71.6	60.2	58.6	3.6
RR	5.5	7.7	5.9	5.8	0.4
RRCV	37.2	55.1	44.2	43.7	3.5
MTLCV	56,858.4	65,233.8	60,883.6	60,837.8	2035.8
LLARS	13.3	61.7	15.9	14.3	7.1
LR	22.9	33.3	26.8	26.4	2.4
BR	57,496.6	61,163.2	58,489.6	58,287.5	806.4
MLPR	21,222.5	25,145.9	22,101.6	22,172.1	571.8

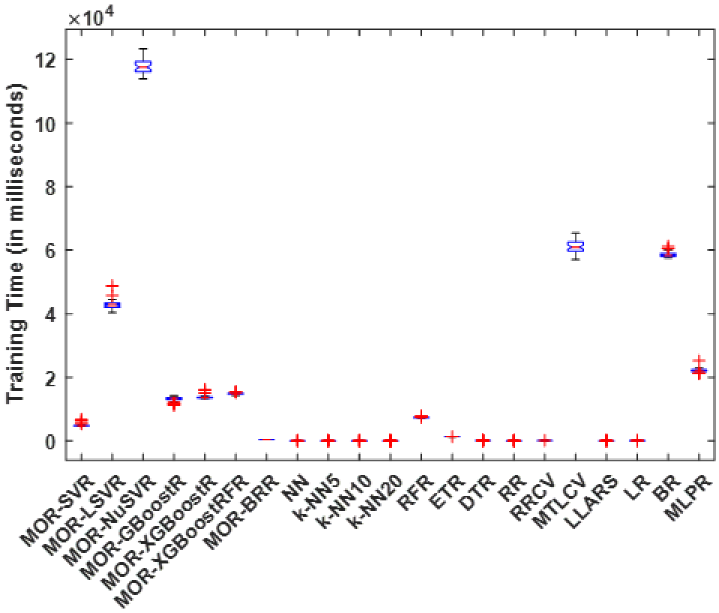

Figure 3. Box plots of TT (over 50 statistical runs) for all methods for the first stage.

In a similar vein to the discussion above for Table 4 and Figure 3, the following observations are made from Table 5 for the TT of all methods for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process: (1) The TT of all methods are generally lower than 91 s on average. (2) k-NN20 shows the best average and median TT of 4.2 ms and 4.3 ms, respectively, over the 50 independent runs. (3) The average and median TT for k-NN10, k-NN5, and NN (4.4 ms and 4.4 ms, 4.5 ms and 4.5 ms, and 4.5 ms and 4.5 ms, respectively) over the 50 independent runs are also short and comparable to those of k-NN20. (4) MOR-NuSVR shows the worst average and median TT of 90,015.0 ms and 89,941.3 ms, respectively, over the 50 independent runs. (5) The average and median TT for MOR-LSVR and BR (39,647.9 ms and 39,768.1 ms, and 34,183.6 ms and 34,043.4 ms, respectively) over the 50 independent runs are also long. (6) The robustness of the methods is relative according to their standard deviations over all runs: k-NN5 and k-NN10 have the lowest standard deviation of 0.2 and MOR-LSVR has the highest standard deviation of 983.8. So, in terms of TT stability for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, k-NN5 and k-NN10 can be said to be the most robust among all the methods, while MOR-LSVR is the least robust. The box plots of TT for all the methods are shown in Figure 4, for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process. The findings earlier discussed can be corroborated visually by an examination of the variations in TT for all the methods in Figure 4.

Table 5. Training time (in milliseconds (ms)) for all methods over 50 statistical runs (the second stage).

Method	Best	Worst	Average	Median	S.D.
MOR-SVR	2255.4	2489.4	2288.4	2281.9	33.8
MOR-LSVR	37,700.3	41,578.9	39,647.9	39,768.1	983.8
MOR-NuSVR	87,112.3	92,142.7	90,015.0	89,941.3	967.9
MOR-GBoostR	6050.4	7306.2	6888.9	6994.2	364.5
MOR-XGBoostR	6132.9	7488.9	6348.6	6313.3	191.8
MOR-XGBoostRFR	5589.7	7371.5	5808.3	5720.9	297.4
MOR-BRR	180.9	201.1	186.4	185.5	4.0
NN	1.6	5.6	4.5	4.5	0.5
***k*-NN5**	3.7	4.8	4.5	4.5	0.2
***k*-NN10**	3.5	4.7	4.4	4.4	0.2
***k*-NN20**	3.1	4.8	4.2	4.3	0.4
RFR	2258.4	2584.2	2308.9	2282.4	72.5
ETR	740.5	845.3	756.7	752.0	16.3
DTR	19.6	22.6	20.2	20.0	0.6
RR	4.5	6.2	4.8	4.7	0.3
RRCV	25.6	32.3	28.0	27.5	1.6
MTLCV	17,352.7	19,314.4	18,337.5	18,385.4	384.9
LLARS	9.9	13.7	11.3	11.1	0.8
LR	13.8	20.9	16.3	15.9	1.7
BR	33,483.8	36,263.5	34,183.6	34,043.4	495.9
MLPR	19,726.4	20,654.0	20,127.9	20,117.5	220.0

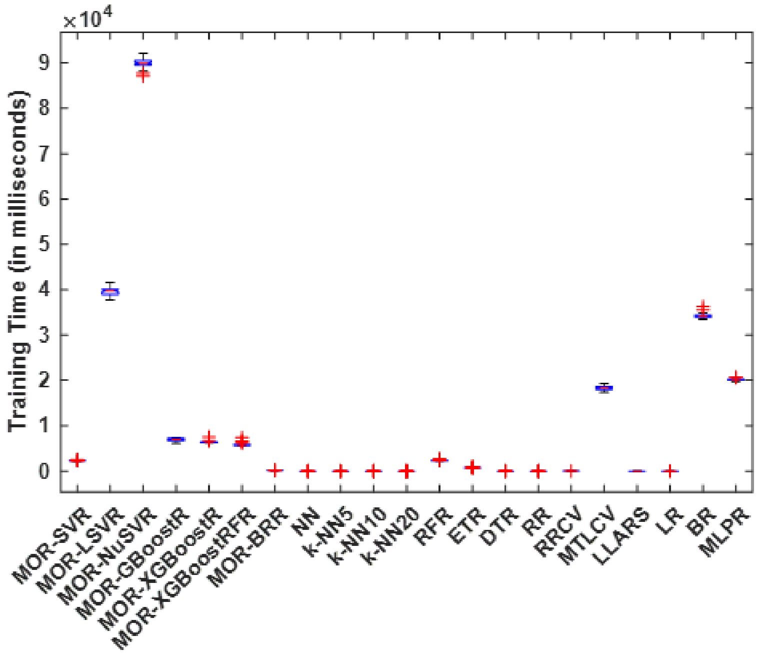

Figure 4. Box plots of TT (over 50 statistical runs) for all methods for the second stage.

4.2. Prediction Speed

Generally, in machine learning, the prediction speed (PS) of a method is derived as the total number of predictions made by a method divided by the time taken to make the predictions. It is measured in observations per second (obs/s). The PS of all methods for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process is reported in Table 6. From Table 6, the following observations are made for PS: (1) The PS of all methods are generally higher than 90 obs/s on average. (2) DTR shows the best average and median PS of 4.8691 × 10^6 obs/s and 4.9929 × 10^6 obs/s, respectively, over the 50 independent runs. (3) The average and median PS for RR and RRCV (4.0102 × 10^6 obs/s and 4.1101 × 10^6 obs/s, and 3.8520 × 10^6 obs/s and 3.9050 × 10^6 obs/s, respectively) over the 50 independent runs are also fast. (4) MOR-NuSVR shows the worst average and median *PS* of 9.2834 × 10^1 obs/s and 9.2775 × 10^1 obs/s, respectively, over all 50 independent runs. (5) The average and median *TT* for MOR-LSVR and BR (1.4115 × 10^2 obs/s and 1.4085 × 10^2 obs/s, and 3.4633 × 10^3 obs/s and 3.4946 × 10^3 obs/s, respectively) over the 50 independent runs are also slow. (6) The robustness of all the methods is relative according to their standard deviations over all runs: MOR-NuSVR has the lowest standard deviation of 2.1876 and RR has the highest standard deviation of 3.9870 × 10^5. So, in terms of *PS* stability for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, MOR-NuSVR can be said to be the most robust among all the methods, while RR is the least robust. The box plots of *PS* for all the methods are shown in Figure 5, for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process. The findings earlier discussed can be corroborated visually by an examination of the variations in *PS* for all the methods in Figure 5.

Table 6. Prediction speed (in observations per second (obs/s)) for all methods over 50 statistical runs (first stage).

Method	Best	Worst	Average	Median	S.D.
MOR-SVR	5.6239 × 10^5	4.1334 × 10^5	4.9186 × 10^5	4.8717 × 10^5	2.9146 × 10^4
MOR-LSVR	1.5552 × 10^2	1.2501 × 10^2	1.4115 × 10^2	1.4085 × 10^2	5.9448
MOR-NuSVR	9.6049 × 10^1	8.5853 × 10^1	9.2834 × 10^1	9.2775 × 10^1	2.1876
MOR-GBRoostR	3.4622 × 10^4	2.2161 × 10^4	3.0139 × 10^4	3.0336 × 10^4	2.1214 × 10^3
MOR-XGBoostR	3.5163 × 10^4	2.4821 × 10^4	3.3007 × 10^4	3.3504 × 10^4	1.9354 × 10^3
MOR-XGBoostRFR	1.0518 × 10^5	7.5331 × 10^4	1.0015 × 10^5	1.0069 × 10^5	4.3709 × 10^3
MOR-BRR	5.6375 × 10^5	4.8359 × 10^5	5.1224 × 10^5	5.1097 × 10^5	1.4630 × 10^4
NN	4.2699 × 10^4	2.2612 × 10^4	3.8062 × 10^4	3.8004 × 10^4	3.2043 × 10^3
***k*-NN5**	4.1211 × 10^4	2.6105 × 10^4	3.7383 × 10^4	3.7048 × 10^4	2.8666 × 10^3
***k*-NN10**	3.7018 × 10^4	2.0611 × 10^4	3.2667 × 10^4	3.2526 × 10^4	3.1338 × 10^3
***k*-NN20**	2.9755 × 10^4	2.3212 × 10^4	2.7355 × 10^4	2.7384 × 10^4	1.4434 × 10^3
RFR	8.7096 × 10^4	7.5532 × 10^4	8.4908 × 10^4	8.5385 × 10^4	1.9828 × 10^3
ETR	1.1799 × 10^5	1.0883 × 10^5	1.1583 × 10^5	1.1622 × 10^5	1.5385 × 10^3
DTR	5.2917 × 10^6	3.8139 × 10^6	4.8691 × 10^6	4.9929 × 10^6	3.2706 × 10^5
RR	4.6171 × 10^6	2.5956 × 10^6	4.0102 × 10^6	4.1101 × 10^6	3.9870 × 10^5
RRCV	4.1563 × 10^6	2.3477 × 10^6	3.8520 × 10^6	3.9050 × 10^6	2.9426 × 10^5
MTLCV	3.7941 × 10^6	2.4586 × 10^6	3.4746 × 10^6	3.4643 × 10^6	2.3376 × 10^5
LLARS	4.1514 × 10^6	1.4018 × 10^6	3.6085 × 10^6	3.7444 × 10^6	5.4622 × 10^5
LR	4.1822 × 10^6	2.4105 × 10^6	3.6997 × 10^6	3.7071 × 10^6	3.1664 × 10^5
BR	3.5379 × 10^3	2.9929 × 10^3	3.4633 × 10^3	3.4946 × 10^3	1.0798 × 10^2
MLPR	1.0661 × 10^6	2.2993 × 10^5	8.9940 × 10^5	9.4811 × 10^5	1.6888 × 10^5

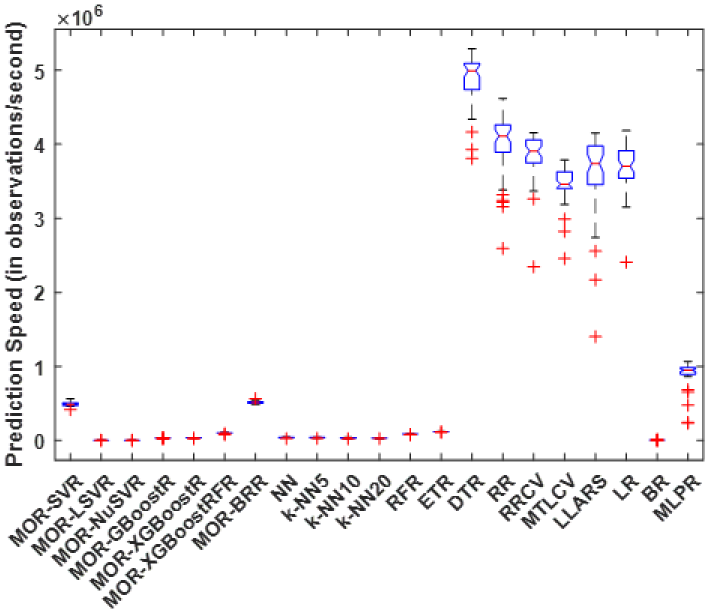

Figure 5. Box plots of *PS* (over 50 statistical runs) for all methods for the first stage.

In a similar vein to the discussion above for Table 6 and Figure 5, the following observations are made from Table 7 for the *PS* of all methods for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process: (1) The *PS* of all methods are generally higher than 130 obs/s on average. (2) DTR shows the best average and median *PS* of 5.6784 × 10^6 obs/s and 5.8007 × 10^6 obs/s, respectively, over the 50 independent runs. (3) The average and median *PS* for RR and LLARS (4.4236 × 10^6 obs/s and 4.1231 × 10^6 obs/s, and 4.4792 × 10^6 obs/s and 4.2340 × 10^6 obs/s, respectively) over the 50 independent runs are also fast and comparable. (4) MOR-NuSVR shows the worst average and median *PS* of 1.3128 × 10^2 obs/s and 1.3213 × 10^2 obs/s, respectively, over the 50 independent runs. (5) The average and median *TT* for MOR-LSVR and BR (1.7654 × 10^2 obs/s and 1.7680 obs/s, and 3.5747 × 10^3 obs/s and 3.5933 × 10^3 obs/s, respectively) over the 50 independent runs are also slow. (6) The robustness of all methods is relative according to their standard deviations over all runs: MOR-NuSVR has the lowest standard deviation of 2.7050 and LLARS has the highest standard deviation of 4.4958 × 10^5. So, in terms of *PS* stability for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, MOR-NuSVR can be said to be the most robust among all the methods, while LLARS is the least robust. The box plots of *PS* for all the methods are shown in Figure 6, for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process. The findings earlier discussed can be corroborated visually by an examination of the variations in *PS* for all the methods in Figure 6.

Table 7. Prediction speed (in observations per second (obs/s)) for all methods over 50 statistical runs (second stage).

Method	Best	Worst	Average	Median	S.D.
MOR-SVR	4.5950 × 10^5	3.5034 × 10^5	4.1594 × 10^5	4.2574 × 10^5	2.2816 × 10^4
MOR-LSVR	1.9021 × 10^2	1.6441 × 10^2	1.7654 × 10^2	1.7680 × 10^2	5.6600
MOR-NuSVR	1.3475 × 10^2	1.2274 × 10^2	1.3128 × 10^2	1.3213 × 10^2	2.7050
MOR-GBoostR	3.9727 × 10^4	2.8796 × 10^4	3.4560 × 10^4	3.4269 × 10^4	2.2557 × 10^3
MOR-XGBoostR	3.8692 × 10^4	2.8567 × 10^4	3.6527 × 10^4	3.6795 × 10^4	1.7434 × 10^3
MOR-XGBoostRFR	1.2194 × 10^5	7.9684 × 10^4	1.1630 × 10^5	1.1776 × 10^5	6.0654 × 10^3
MOR-BRR	6.6072 × 10^5	4.5947 × 10^5	5.7582 × 10^5	5.7859 × 10^5	3.3941 × 10^4
NN	5.1418 × 10^4	4.0254 × 10^4	4.6993 × 10^4	4.7415 × 10^4	2.0383 × 10^3
***k*-NN5**	4.9592 × 10^4	4.0310 × 10^4	4.5902 × 10^4	4.6354 × 10^4	2.0259 × 10^3
***k*-NN10**	4.3716 × 10^4	3.6807 × 10^4	4.0243 × 10^4	3.9825 × 10^4	1.9493 × 10^3
***k*-NN20**	3.4977 × 10^4	2.5228 × 10^4	3.1848 × 10^4	3.1948 × 10^4	1.8785 × 10^3
RFR	8.6406 × 10^4	7.6123 × 10^4	8.4357 × 10^4	8.5108 × 10^4	2.2585 × 10^3
ETR	1.1598 × 10^5	1.0321 × 10^5	1.1374 × 10^5	1.1470 × 10^5	2.7651 × 10^3
DTR	6.0917 × 10^6	4.0165 × 10^6	5.6784 × 10^6	5.8007 × 10^6	4.2400 × 10^5
RR	4.7774 × 10^6	3.2762 × 10^6	4.4236 × 10^6	4.4792 × 10^6	2.2114 × 10^5
RRCV	4.3636 × 10^6	3.3754 × 10^6	4.0458 × 10^6	4.0721 × 10^6	1.9418 × 10^5
MTLCV	3.7229 × 10^6	2.4182 × 10^6	3.3676 × 10^6	3.5491 × 10^6	3.8213 × 10^5
LLARS	4.6237 × 10^6	2.1289 × 10^6	4.1231 × 10^6	4.2340 × 10^6	4.4958 × 10^5
LR	4.3447 × 10^6	2.8044 × 10^6	3.9983 × 10^6	4.0774 × 10^6	3.2755 × 10^5
BR	3.6302 × 10^3	3.3082 × 10^3	3.5747 × 10^3	3.5933 × 10^3	6.4570 × 10^1
MLPR	1.0929 × 10^6	2.4127 × 10^5	1.0069 × 10^6	1.0317 × 10^6	1.2004 × 10^5

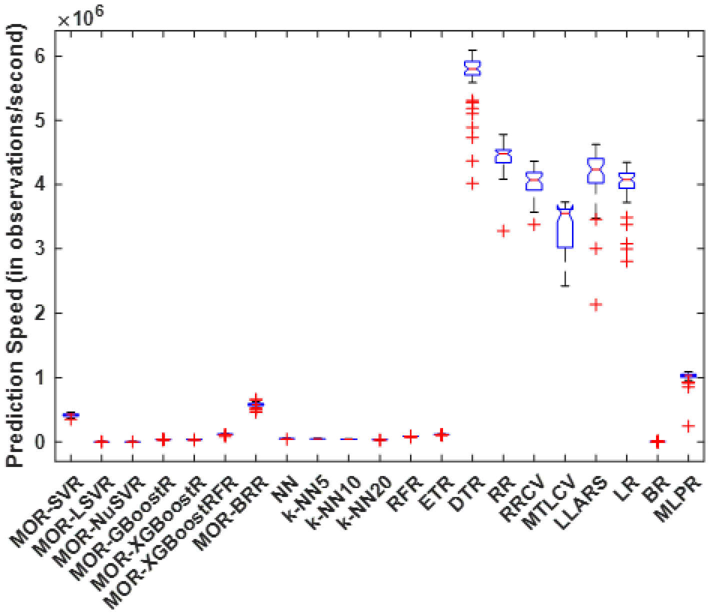

Figure 6. Box plots of *PS* (over 50 statistical runs) for all methods for the second stage.

4.3. R-Squared Statistic

In machine learning, the R-squared value or statistic (R^2) is also called the coefficient of determination and it helps to estimate the goodness of the fit of regression models. In regression learning, it can be derived as follows:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{12}$$

where SSR is the sum of squares regression, SSE is the sum of squares residual error, and SST is the sum of squares total. In Equation (12), $R^2 \in [0, 1]$, indicating that R^2 typically takes on a value between null and unity (i.e., 0% and 100%), where $R^2 = 0$ indicates that the regression model explains none of the variability of the response data around its mean and $R^2 = 1$ indicates that the model explains all the variability of the response data around its mean. In other words, the higher the R^2 value of the regression model, the better the regression model fits the data (in this case the multi-stage continuous-flow manufacturing process data). Note that for some methods, $R < 0$ holds when the regression model is arbitrarily worse. Methods that yielded such models in our experiments have not been presented in this work due to their non-suitability.

The R^2 of all methods for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process is reported in Table 8. From Table 8, the following observations are made for R^2: (1) The R^2 of all methods are generally higher than 0.18 on average. (2) BR shows the best average and median R^2 of 0.728 and 0.728, respectively, over the 50 independent runs. (3) The average and median R^2 for MOR-XGBoostR and MOR-GBoostR (0.673 and 0.676, and 0.663 and 0.664, respectively) over the 50 independent runs are also relatively high and comparable. (4) LLARS shows the worst average and median R^2 of 0.185 and 0.185, respectively, over the 50 independent runs. (5) The average and median R^2 for MOR-SVR and MOR-NuSVR (0.188 and 0.188, and 0.286 and 0.286, respectively) over the 50 independent runs are also small. (6) All methods show relatively good robustness according to their standard deviations over all runs: LLARS has the lowest standard deviation of 0.002 and NN has the highest standard deviation of 0.016. So, in terms of R^2 stability for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, LLARS can be said to be the most robust among all the methods, while NN is the least robust. The box plots of R^2 for all the methods are shown in Figure 7, for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process. The findings earlier discussed can be corroborated visually by an examination of the variations in R^2 for all the methods in Figure 7.

Table 8. R-squared value (R^2) for all methods over 50 statistical runs (first stage).

Method	Best	Worst	Average	Median	S.D.
MOR-SVR	0.199	0.176	0.188	0.188	0.005
MOR-LSVR	0.305	0.283	0.295	0.295	0.005
MOR-NuSVR	0.296	0.276	0.286	0.286	0.004
MOR-GBoostR	0.688	0.639	0.663	0.664	0.011
MOR-XGBoostR	0.701	0.633	0.673	0.676	0.016
MOR-XGBoostRFR	0.443	0.414	0.428	0.427	0.006
MOR-BRR	0.327	0.303	0.317	0.317	0.004
NN	0.628	0.564	0.594	0.596	0.016
***k*-NN5**	0.622	0.554	0.592	0.591	0.013
***k*-NN10**	0.594	0.533	0.570	0.572	0.015
***k*-NN20**	0.535	0.505	0.523	0.524	0.007
RFR	0.693	0.620	0.661	0.660	0.015
ETR	0.397	0.366	0.380	0.379	0.006
DTR	0.450	0.338	0.436	0.438	0.015
RR	0.327	0.309	0.317	0.317	0.003
RRCV	0.326	0.310	0.317	0.317	0.004
MTLCV	0.325	0.309	0.317	0.317	0.003
LLARS	0.189	0.180	0.185	0.185	0.002
LR	0.323	0.308	0.317	0.316	0.003
BR	0.747	0.698	0.728	0.728	0.010
MLPR	0.404	0.357	0.376	0.375	0.009

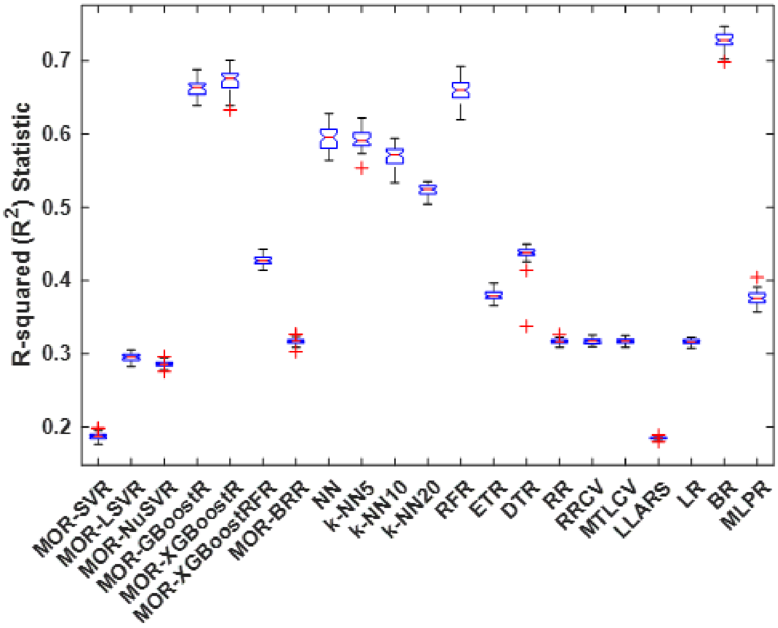

Figure 7. Box plots of R^2 statistic (over 50 statistical runs) for all methods for the first stage.

In a similar vein to the discussion above for Table 8 and Figure 7, the following observations are made from Table 9 for the R^2 of all methods for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process: (1) The R^2 of all methods are generally higher than 0.53 on average. (2) RFR shows the best average and median R^2 of 0.746 and 0.746, respectively, over the 50 independent runs. (3) The average and median R^2 for MOR-GBoostR and MOR-XGBoostR (0.736 and 0.735, and 0.726 and 0.726, respectively) over the 50 independent runs are also high and comparable to those of RFR. (4) LLARS shows the worst average and median R^2 of 0.275 and 0.276, respectively, over the 50 independent runs. (5) The average and median R^2 for MOR-SVR (0.463 and

0.463, respectively) over the 50 independent runs are also small in comparison to other methods. (6) All methods show relatively good robustness according to their standard deviations over all runs: LLARS has the lowest standard deviation of 0.005 and MOR-SVR, MOR-LSVR, and RR have the highest standard deviation of 0.015 each. So, in terms of R^2 stability for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, LLARS can be said to be the most robust among all the methods, while MOR-SVR, MOR-LSVR, and RR are the least robust. The box plots of R^2 for all the methods are shown in Figure 8, for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process. The findings earlier discussed can be corroborated visually by an examination of the variations in R^2 for all the methods in Figure 8.

Table 9. R-squared value (R^2) for all methods over 50 statistical runs (second stage).

Method	Best	Worst	Average	Median	S.D.
MOR-SVR	0.494	0.434	0.463	0.463	0.015
MOR-LSVR	0.574	0.500	0.544	0.544	0.015
MOR-NuSVR	0.580	0.519	0.549	0.55	0.013
MOR-GBoostR	0.765	0.691	0.736	0.735	0.013
MOR-XGBoostR	0.748	0.694	0.726	0.726	0.013
MOR-XGBoostRFR	0.679	0.625	0.657	0.657	0.012
MOR-BRR	0.568	0.517	0.544	0.546	0.012
NN	0.733	0.686	0.706	0.703	0.012
***k*-NN5**	0.729	0.694	0.709	0.710	0.009
***k*-NN10**	0.739	0.686	0.713	0.715	0.013
***k*-NN20**	0.725	0.687	0.708	0.708	0.010
RFR	0.769	0.726	0.746	0.746	0.010
ETR	0.675	0.604	0.648	0.649	0.014
DTR	0.689	0.635	0.662	0.660	0.012
RR	0.568	0.511	0.540	0.541	0.015
RRCV	0.564	0.497	0.538	0.539	0.014
MTLCV	0.573	0.511	0.541	0.543	0.013
LLARS	0.286	0.260	0.275	0.276	0.005
LR	0.569	0.516	0.542	0.543	0.012
BR	0.748	0.709	0.730	0.730	0.010
MLPR	0.621	0.557	0.592	0.592	0.014

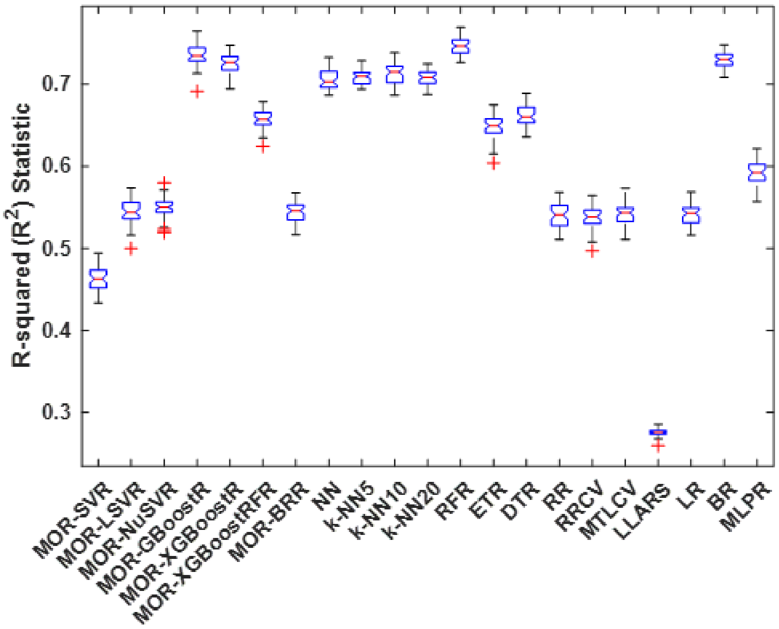

Figure 8. Box plots of R^2 statistic (over 50 statistical runs) for all methods for the second stage.

4.4. Mean Squared Error

The mean squared error (MSE) is another statistical measure used in machine learning to evaluate regression learning models. Mathematically, the MSE of a regression model is the average squared difference between the predicted values by the model and the observed values. In regression learning, it can be derived as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{13}$$

where n is the number of observations, Y_i and $\hat{Y}_i$ are the actual value and estimated or predicted value, respectively, for the ith observation in the data set. Note that in regression learning, MSE is always positive or greater than null such that $MSE \to 0$ depicts the better quality of the regression model and $MSE = 0$ indicates that the regression model is a perfect predictor.

The MSE of all methods for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process is reported in Table 10. From Table 10, the following observations are made for the MSE: (1) The MSE of all methods are generally higher than 2.5 on average. (2) BR shows the best average and median MSE of 2.7 and 2.7, respectively, over the 50 independent runs. (3) The average and median MSE for MOR-XGBoostR and MOR-GBoostR (3.2 and 3.3, and 3.4 and 3.4, respectively) over the 50 independent runs are also low and comparable. (4) MOR-SVR shows the worst average and median MSE of 9.0 and 9.0, respectively, over the 50 independent runs. (5) The average and median MSE for LLARS (8.3 and 8.3, respectively) over the 50 independent runs are also high and comparable to those of MOR-LSVR and MOR-NuSVR (8.1 and 8.1, respectively, for both methods). (6) The robustness of all methods is relative according to their standard deviations over all runs: MOR-GBoostR, MOR-XGBoostR, k-NN5, k-NN10, RFR and BR have the least standard deviation of 0.1, while MOR-SVR and MOR-LSVR have the highest standard deviation of 0.3. So, in terms of MSE stability for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, MOR-GBoostR, MOR-XGBoostR, k-NN5, k-NN10, RFR and BR can be said to be the most robust among all the methods, while MOR-SVR and MOR-LSVR are the least robust. The box plots of MSE for all the methods are shown in Figure 9, for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process. The findings earlier discussed can be corroborated visually by an examination of the variations in MSE for all the methods in Figure 9.

In a similar vein to the discussion above for Table 10 and Figure 9, the following observations are made from Table 11 for the MSE of all methods for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process: (1) The MSE of all methods are generally higher than 3.0 on average. (2) RFR shows the best average and median MSE of 3.5 and 3.5, respectively, over the 50 independent runs. (3) The average and median MSE for MOR-GBoostR and MOR-XGBoostR (3.6 and 3.7, and 3.7 and 3.7, respectively) over the 50 independent runs are also relatively low and comparable. (4) MOR-SVR shows the worst average and median MSE of 10.9 and 10.9, respectively, over the 50 independent runs. (5) The average and median MSE for MOR-LSVR and MOR-NuSVR (10.2 and 10.2, and 10.3 and 10.3, respectively) over the 50 independent runs are also high and comparable. (6) The robustness of all methods is relatively according to their standard deviations over all runs: MOR-GBoostR, MOR-XGBoostR, MOR-XGBoostRFR, NN, k-NN5, k-NN10, k-NN20, RFR and BR have the least standard deviation of 0.1, while MOR-SVR, MOR-LSVR, and MOR-NuSVR have the highest standard deviation of 0.3. So, in terms of MSE stability for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, MOR-GBoostR, MOR-XGBoostR, MOR-XGBoostRFR, NN, k-NN5, k-NN10, k-NN20, RFR and BR can be said to be the most robust among all the methods, while MOR-SVR, MOR-LSVR, and MOR-NuSVR are the least robust. The box plots of MSE for all the methods are shown in Figure 10, for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process. The findings earlier discussed can be corroborated visually by an examination of the variations in *MSE* for all the methods in Figure 10.

Table 10. Mean squared error (MSE) for all methods over 50 statistical runs (the first stage).

Method	Best	Worst	Average	Median	S.D.
MOR-SVR	8.4	9.9	9.0	9.0	0.3
MOR-LSVR	7.4	8.8	8.1	8.1	0.3
MOR-NuSVR	7.4	8.7	8.1	8.1	0.3
MOR-GBoostR	3.1	3.6	3.4	3.4	0.1
MOR-XGBoostR	3.0	3.5	3.2	3.3	0.1
MOR-XGBoostRFR	5.1	6.0	5.6	5.6	0.2
MOR-BRR	6.7	7.4	7.0	7.1	0.2
NN	3.7	4.4	4.1	4.1	0.2
***k*-NN5**	3.8	4.3	4.1	4.0	0.1
***k*-NN10**	3.9	4.6	4.2	4.2	0.1
***k*-NN20**	4.3	5.1	4.7	4.6	0.2
RFR	2.8	3.4	3.1	3.1	0.1
ETR	5.5	6.4	5.9	5.9	0.2
DTR	4.4	6.0	4.9	4.9	0.2
RR	6.5	7.4	7.0	7.0	0.2
RRCV	6.5	7.6	7.0	7.0	0.2
MTLCV	6.6	7.4	7.0	7.0	0.2
LLARS	7.8	8.9	8.3	8.3	0.2
LR	6.4	7.7	7.0	7.0	0.2
BR	2.5	3.1	2.7	2.7	0.1
MLPR	5.5	6.5	5.9	5.9	0.2

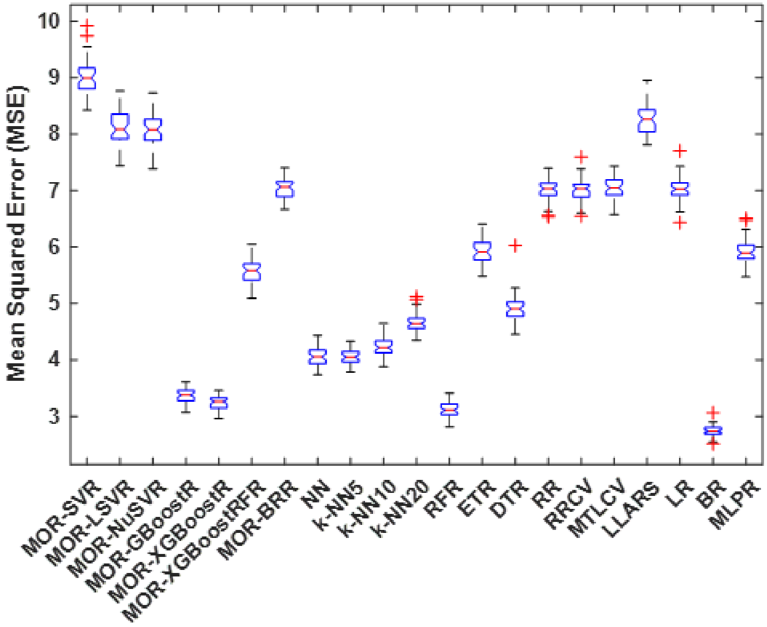

Figure 9. Box plots of *MSE* (over 50 statistical runs) for all methods for the first stage.

Table 11. Mean squared error (MSE) for all methods over 50 statistical runs (the second stage).

Method	Best	Worst	Average	Median	S.D.
MOR-SVR	10.3	11.6	10.9	10.9	0.3
MOR-LSVR	9.6	10.8	10.2	10.2	0.3
MOR-NuSVR	9.5	10.9	10.3	10.3	0.3
MOR-GBoostR	3.3	3.9	3.6	3.7	0.1
MOR-XGBoostR	3.4	4.0	3.7	3.7	0.1
MOR-XGBoostRFR	5.7	6.3	6.0	6.0	0.1
MOR-BR	8.0	9.1	8.5	8.5	0.2
NN	3.7	4.3	4.1	4.1	0.1
***k*-NN5**	3.9	4.4	4.1	4.0	0.1
***k*-NN10**	3.8	4.3	4.1	4.0	0.1
***k*-NN20**	4.0	4.6	4.3	4.3	0.1
RFR	3.2	3.8	3.5	3.5	0.1
ETR	5.6	6.4	6.0	6.0	0.2
DTR	5.0	5.7	5.3	5.3	0.2
RR	8.0	8.9	8.5	8.5	0.2
RRCV	8.0	9.0	8.5	8.5	0.2
MTLCV	8.1	9.0	8.5	8.5	0.2
LLARS	8.6	9.8	9.2	9.2	0.2
LR	8.1	8.9	8.5	8.5	0.2
BR	3.2	3.8	3.5	3.5	0.1
MLPR	6.8	7.9	7.3	7.3	0.2

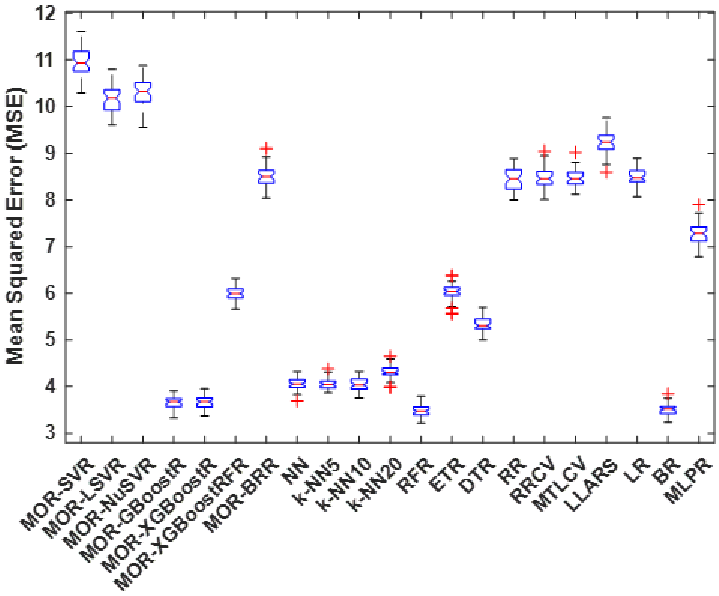

Figure 10. Box plots of MSE (over 50 statistical runs) for all methods for the second stage.

4.5. Ranking of All Methods

To straightforwardly rank all methods based on inference, positions or points are assigned to each method using their average performances (over the 50 independent statistical runs) for training time (TT), prediction speed (PS), R-squared values (R^2) and mean squared error (MSE). The points are assigned in descending order of performance to have: 21, 20, 19, 18, . . . , 1. In this way, the best method(s) will have 21 points and the worst method(s) will have one point. Tables 12 and 13 show how the overall points (sum of all points or the positions) obtained by each method can be used to deduce the overall ranking for the first and second stages of the predictive modeling of the multi-stage continuous-flow manufacturing process, respectively. These results are further corroborated in Figures 11–13 where the ranking (in points) of all methods using the average values of their performance metrics (i.e., TT, PS, R^2, and MSE) over 50 statistical runs are shown for the first and second stages of the predictive modeling of the multi-stage continuous-flow manufacturing process.

Table 12. Ranking (in points) of all methods (the first stage).

Method	*TT*	*PS*	R^2	*MSE*	Overall
MOR-SVR	1	13	2	10	26
MOR-LSVR	3	2	4	4	13
MOR-NuSVR	4	1	3	1	9
MOR-GBoostR	18	5	19	8	50
MOR-XGBoostR	19	7	20	7	53
MOR-GBoostRFR	12	11	12	6	41
MOR-BR	8	14	8	12	42
NN	16	9	17	20	59
***k*-NN5**	17	8	16	18	59
***k*-NN10**	15	6	15	19	55
***k*-NN20**	14	4	14	21	53
RFR	20	10	18	9	57
ETR	11	12	11	11	45
DTR	13	21	13	13	60
RR	6	20	6	17	49
RRCV	9	19	9	14	51
MTLCV	5	16	7	2	30
LLARS	2	17	1	16	36
LR	7	18	5	15	45
BR	21	3	21	3	47
MLPR	10	15	10	5	40

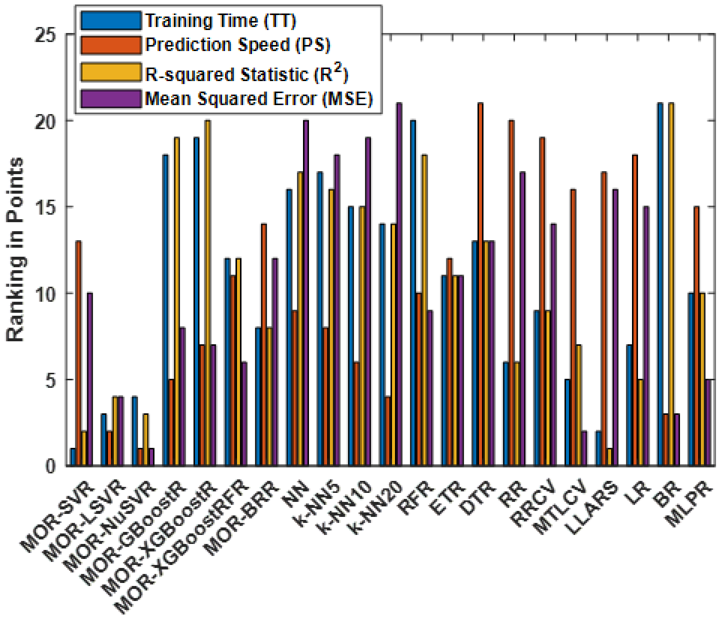

Figure 11. Ranking (in points) of all methods using the average values of their performance metrics (i.e., *TT*, *PS*, R^2, and *MSE*) over 50 statistical runs for the first stage.

For the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, the overall rankings in Table 12 indicate that DTR ranks first with a total of 60 points, NN and *k*-NN5 both rank second with 59 points in total each. MOR-NuSVR ranks last with 9 points in total. This suggests that the overall performances of DTR, NN, and *k*-NN5 are better for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, in comparison to the other methods investigated. In a similar vein, for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, the overall rankings in Table 13 indicate that *k*-NN20 ranks first with a total of 64 points, DTR and RFR rank second and third, with 61 points in total, and 60 points in total, respectively. MOR-LSVR and MOR-NuSVR both rank last with 15 points in total each. This suggests that the overall performances of *k*-NN20, DTR, and RFR are better for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process, in comparison to the other methods investigated. If the average number of points obtained by each method for the two stages of the predictive modeling of the multi-stage continuous-flow manufacturing process is used to select the overall best method, then DTR having an average of 60.5 points will be the overall best method for the predictive modeling of both stages of the multi-stage continuous-flow manufacturing process. To statistically verify that the overall performances of DTR and *k*-NN20 rank better than the other methods for the first and second stages of the predictive modeling of the multi-stage continuous-flow manufacturing process, respectively, hypothesis tests are carried out in the next subsection using the Wilcoxon test [59].

Table 13. Ranking (in points) of all methods (the second stage).

Method	TT	PS	R^2	MSE	Overall
MOR-SVR	1	13	2	10	26
MOR-LSVR	3	2	8	2	15
MOR-NuSVR	4	1	9	1	15
MOR-GBoostR	18	5	20	6	49
MOR-XGBoostR	19	6	18	7	50
MOR-XGBoostRFR	12	12	12	8	44
MOR-BRR	8	14	7	12	41
NN	16	9	14	19	58
***k*-NN5**	17	8	16	18	59
***k*-NN10**	15	7	17	20	59
***k*-NN20**	14	4	15	21	64
RFR	20	10	21	9	60
ETR	11	11	11	11	44
DTR	13	21	13	14	61
RR	6	20	4	17	47
RRCV	9	18	3	13	43
MTLCV	5	16	5	5	31
LLARS	2	19	1	16	38
LR	7	17	6	15	45
BR	21	3	19	3	46
MLPR	10	15	10	4	39

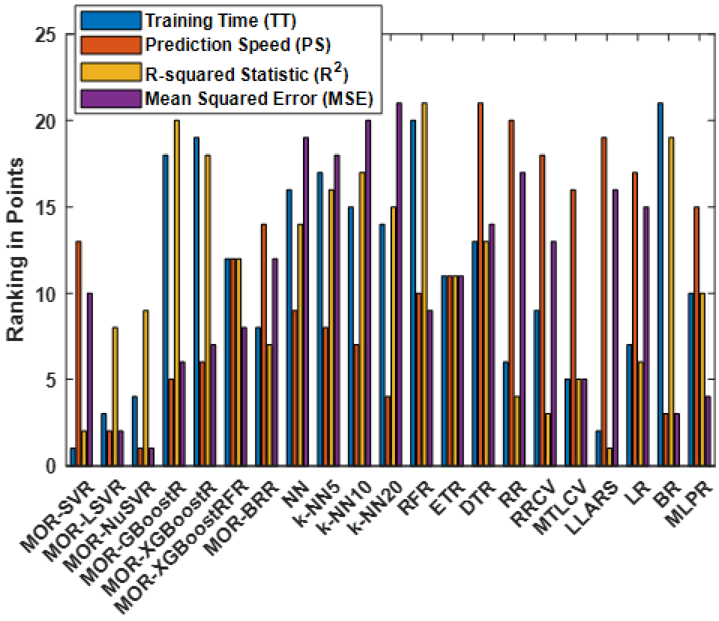

Figure 12. Ranking (in points) of all methods using the average values of their performance metrics (i.e., TT, PS, R^2, and MSE) over 50 statistical runs for the second stage.

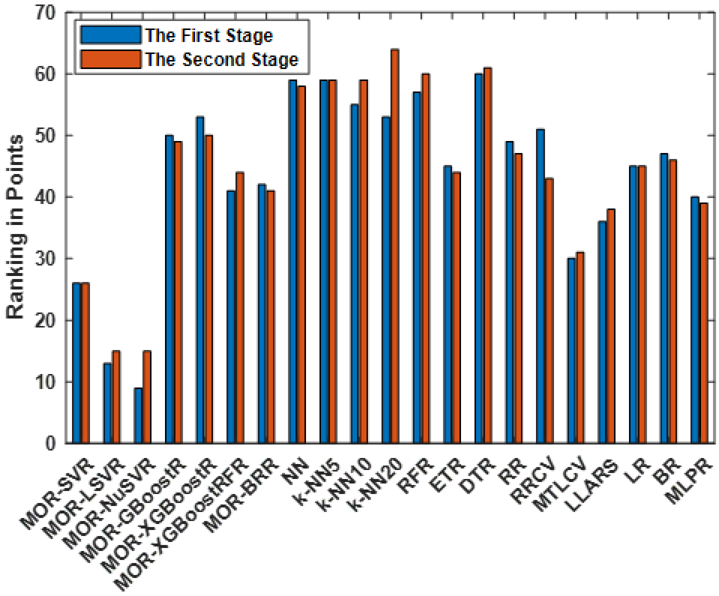

Figure 13. Overall ranking (in points) of all methods using the average values of their performance metrics (i.e., TT, PS, R^2, and MSE) over 50 statistical runs.

4.6. Hypothesis Test

A Wilcoxon test [59] is the hypothesis test used in this work to statistically verify the inference-based rankings detailed in Section 4.5. To carry out the test, the results obtained by all methods for training time (TT), predictive accuracy (R^2), prediction speed (PS), and mean squared error (MSE) over the 50 independent statistical runs for each stage of the predictive modeling of the multi-stage continuous-flow manufacturing process are used as data samples. In this instance, the null hypothesis for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process is that the data samples of DTR (the best method according to the inference-based ranking in Table 12) and the other methods have equal medians at 5% significance level (i.e., 95% confidence level). In a similar vein, the null hypothesis for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process is that the data samples of k-NN20 (the best method according to the inference-based ranking in Table 13) and the other methods have equal medians at 5% significance level (i.e., 95% confidence level). For the two stages of the predictive modeling of the multi-stage continuous-flow manufacturing process, if the resultant probability value (p-value) of the hypothesis test is less than or equal to 0.05, then it can be said that a strong evidence exists against the null hypothesis and it is therefore rejected.

The results for the hypothesis tests are shown in Tables 14 and 15, it can be said that TT, PS, R^2, and MSE rejected the null hypothesis in all the cases according to the p-values reported in Tables 14 and 15, except for NN and k-NN5 where the p-values for R^2 are greater than 0.05 for the second stage of the predictive modeling of the multi-stage continuous-flow manufacturing process. From the p-values reported in Table 14, it can be said that the TT, PS, R^2, and MSE of DTR over the 50 independent runs are significantly better than the TT, PS, R^2, and MSE of other methods for the first stage of the predictive modeling of the multi-stage continuous-flow manufacturing process. In a similar vein, the p-values reported in Table 15 also indicate that the TT, PS, and MSE of k-NN20 over the 50 independent runs are significantly better than the TT, PS, and MSE of other methods for the second stage of the predictive modeling. Also, from Table 15, it can be deduced that the R^2 of k-NN20 over the 50 independent runs are significantly better than the R^2 of all the other methods except for NN and k-NN5. These exceptions can be easily attributed to the stochastic nature and similarity of these methods and the number of statistical runs used for the experimentation. Therefore, the rankings earlier established in Tables 12 and 13 can be said to be statistically valid.

Table 14. Hypothesis test: DTR v other methods (the first stage).

Method	TT **(p-Values)**	PS **(p-Values)**	R^2 **(p-Values)**	MSE **(p-Values)**
MOR-SVR	7.0661×10^{-18}	7.0645×10^{-18}	7.0454×10^{-18}	7.0629×10^{-18}
MOR-LSVR	7.0661×10^{-18}	7.0645×10^{-18}	7.0486×10^{-18}	7.0629×10^{-18}
MOR-NuSVR	7.0661×10^{-18}	7.0645×10^{-18}	7.0533×10^{-18}	7.0629×10^{-18}
MOR-GBoostR	7.0661×10^{-18}	7.0645×10^{-18}	7.0486×10^{-18}	7.0629×10^{-18}
MOR-XGBoostR	7.0661×10^{-18}	7.0645×10^{-18}	7.0549×10^{-18}	7.0629×10^{-18}
MOR-XGBoostRFR	7.0661×10^{-18}	7.0645×10^{-18}	2.4714×10^{-9}	2.2652×10^{-16}
MOR-BRR	7.0661×10^{-18}	7.0645×10^{-18}	7.0517×10^{-18}	7.0613×10^{-18}
NN	7.0661×10^{-18}	7.0645×10^{-18}	7.0613×10^{-18}	7.0629×10^{-18}
k-**NN5**	7.0661×10^{-18}	7.0645×10^{-18}	7.0581×10^{-18}	7.0613×10^{-18}
k-**NN10**	7.0661×10^{-18}	7.0645×10^{-18}	7.0613×10^{-18}	1.1412×10^{-17}
k-**NN20**	7.0645×10^{-18}	7.0645×10^{-18}	7.0517×10^{-18}	1.9649×10^{-8}
RFR	7.0661×10^{-18}	7.0645×10^{-18}	7.0533×10^{-18}	7.0629×10^{-18}
ETR	7.0661×10^{-18}	7.0645×10^{-18}	1.3462×10^{-16}	5.0133×10^{-17}
RR	7.0661×10^{-18}	7.2697×10^{-15}	7.0374×10^{-18}	7.0597×10^{-18}
RRCV	7.0661×10^{-18}	1.8002×10^{-16}	7.0422×10^{-18}	7.0629×10^{-18}
MTLCV	7.0661×10^{-18}	7.0629×10^{-18}	7.0311×10^{-18}	7.0629×10^{-18}
LLARS	9.5300×10^{-17}	8.0023×10^{-17}	7.0311×10^{-18}	7.0613×10^{-18}
LR	7.0661×10^{-18}	5.0144×10^{-17}	7.0502×10^{-18}	7.0629×10^{-18}
BR	7.0661×10^{-18}	7.0645×10^{-18}	7.0597×10^{-18}	7.0629×10^{-18}
MLPR	7.0661×10^{-18}	7.0645×10^{-18}	1.3474×10^{-16}	5.9759×10^{-17}

Table 15. Hypothesis test: *k*-NN20 v other methods (the second stage).

Method	TT **(p-Values)**	PS **(p-Values)**	R^2 **(p-Values)**	MSE **(p-Values)**
MOR-SVR	7.0661×10^{-18}	7.0661×10^{-18}	7.0629×10^{-18}	7.0629×10^{-18}
MOR-LSVR	7.0661×10^{-18}	7.0661×10^{-18}	7.0549×10^{-18}	7.0629×10^{-18}
MOR-NuSVR	7.0661×10^{-18}	7.0661×10^{-18}	7.0533×10^{-18}	7.0629×10^{-18}
MOR-GBoostR	7.0661×10^{-18}	1.9602×10^{-9}	1.8852×10^{-15}	7.0629×10^{-18}
MOR-XGBoostR	7.0661×10^{-18}	1.8349×10^{-18}	2.5288×10^{-10}	7.0629×10^{-18}
MOR-XGBoostRFR	7.0661×10^{-18}	7.0661×10^{-18}	7.0613×10^{-18}	7.0629×10^{-18}
MOR-BRR	7.0661×10^{-18}	7.0661×10^{-18}	7.0613×10^{-18}	7.0629×10^{-18}
NN	1.3613×10^{-5}	7.0661×10^{-18}	0.2385	1.8024×10^{-13}
k-**NN5**	9.6788×10^{-7}	7.0661×10^{-18}	0.7512	1.3559×10^{-18}
k-**NN10**	0.0013	7.0661×10^{-18}	0.0223	1.5532×10^{-12}
RFR	7.0661×10^{-18}	7.0661×10^{-18}	7.0549×10^{-18}	7.0629×10^{-18}
ETR	7.0661×10^{-18}	7.0645×10^{-18}	7.0645×10^{-18}	7.0629×10^{-18}
DTR	7.0661×10^{-18}	7.0661×10^{-18}	7.9652×10^{-18}	7.0629×10^{-18}
RR	3.9816×10^{-15}	7.0645×10^{-18}	7.0613×10^{-18}	7.0629×10^{-18}
RRCV	7.0661×10^{-18}	7.0645×10^{-18}	7.0597×10^{-18}	7.0613×10^{-18}
MTLCV	7.0661×10^{-18}	7.0629×10^{-18}	7.0597×10^{-18}	7.0613×10^{-18}
LLARS	7.0661×10^{-18}	7.0661×10^{-18}	7.0502×10^{-18}	7.0613×10^{-18}
LR	7.0661×10^{-18}	7.0645×10^{-18}	7.0581×10^{-18}	7.0613×10^{-18}
BR	7.0661×10^{-18}	7.0661×10^{-18}	4.9533×10^{-15}	7.0629×10^{-18}
MLPR	7.0661×10^{-18}	7.0661×10^{-18}	7.0645×10^{-18}	7.0629×10^{-18}

5. Conclusions

A generalized framework for the predictive modeling of manufacturing processes via regression learning is presented and validated using an open-source data set for a multi-stage continuous-flow manufacturing process in this work. Specifically, the practicality of the framework is demonstrated by investigating several regression-based supervised learning techniques using the given data set. The investigated methods are appraised using the training time (TT), prediction speed (PS), predictive accuracy (R-squared statistic (R^2)), and mean squared error (MSE) obtained by each method over 50 independent statistical runs for the predictive modeling of the two stages of the multi-stage continuous-flow manufacturing process. The methods are then ranked inferentially according to the statistics of their metrics in terms of TT, PS, R^2, and mean squared error (MSE)

obtained by each method over the 50 independent statistical runs. Based on the inference-based rankings, DTR (decision tree regressor) and *k*-NN20 (*k*-nearest neighbour with 20 neighbors) have been identified as the most suitable regression learning techniques for the predictive modeling of the first stage and second stage of the multi-stage continuous-flow manufacturing process, respectively. These rankings are also verified statistically using a Wilcoxon rank sum test to validate the inferences drawn. Even though the experiments and outcomes in this work are not exhaustive due to the broad and stochastic nature of regression learning techniques, they can adequately serve as a guide and benchmark for the adoption and/or invention of regression learning-based paradigms in present-day and future manufacturing environments. In the future, a real-world paradigm where regression-based predictive models are employed on the fly in the manufacturing process will be investigated.

Author Contributions: Conceptualization, M.O.A.; methodology, M.O.A. and K.Z.; software, M.O.A. and K.Z.; validation M.O.A. and K.Z.; formal analysis, M.O.A.; investigation, M.O.A. and K.Z.; resources, M.O.A. and K.Z.; data curation, M.O.A. and K.Z.; writing—original draft preparation, M.O.A. and K.Z.; writing—review and editing, M.O.A.; visualization, M.O.A. and K.Z.; supervision, M.O.A.; project administration, M.O.A.; funding acquisition, M.O.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The APC for the publication of this work was fully supported by the editorial office.

Data Availability Statement: The open-source multi-flow continuous stage manufacturing data set used for the experiments in this paper is available at: https://www.kaggle.com/majidtabkhpaz, accessed on 18 October 2022.

Acknowledgments: The authors would like to especially thank the editorial office for the invitation to submit our manuscript and the APC waiver provided. The authors would also like to thank Liveline Technologies, U.S.A., for the availability of the open-source multi-stage continuous-flow manufacturing process data set.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bajic, B.; Rikalovic, A.; Suzic, N.; Piuri, V. Industry 4.0 Implementation Challenges and Opportunities: A Managerial Perspective. *IEEE Syst. J.* **2021**, *15*, 546–559. [CrossRef]
2. Shiroishi, Y.; Uchiyama, K.; Suzuki, N. Society 5.0: For Human Security and Well-Being. *Computer* **2018**, *51*, 91–95. [CrossRef]
3. Javaid, M.; Haleem, A. Critical components of Industry 5.0 towards a successful adoption in the field of manufacturing. *J. Ind. Integr. Manag.* **2020**, *5*, 327–348. [CrossRef]
4. Malige, A.; Korcyl, G.; Firlej, M.; Fiutowski, T.; Idzik, M.; Korzeniak, B.; Lalik, R.; Misiak, A.; Molenda, A.; Moroń, J.; et al. Real-Time Data Processing Pipeline for Trigger Readout Board-Based Data Acquisition Systems. *IEEE Trans. Nucl. Sci.* **2022**, *69*, 1765–1772. [CrossRef]
5. Akinsolu, M.O. Applied Artificial Intelligence in Manufacturing and Industrial Production Systems: PEST Considerations for Engineering Managers. *IEEE Eng. Manag. Rev.* **2022**, *51*, 1–8. [CrossRef]
6. Miller, W.J.; Poole, M. Advanced CIM environment for manufacturing data analysis. *IEEE Trans. Semicond. Manuf.* **1993**, *6*, 128–133. [CrossRef]
7. Zhang, H.; Wang, H.; Li, J.; Gao, H. A generic data analytics system for manufacturing production. *Big Data Min. Anal.* **2018**, *1*, 160–171. [CrossRef]
8. Zhang, C.; Ji, W. Big Data Analysis Approach for Real-Time Carbon Efficiency Evaluation of Discrete Manufacturing Workshops. *IEEE Access* **2019**, *7*, 107730–107743. [CrossRef]
9. Cui, Y.; Kara, S.; Chan, K.C. Manufacturing big data ecosystem: A systematic literature review. *Robot.-Comput.-Integr. Manuf.* **2020**, *62*, 101861. [CrossRef]
10. Feng, M.; Li, Y. Predictive Maintenance Decision Making Based on Reinforcement Learning in Multistage Production Systems. *IEEE Access* **2022**, *10*, 18910–18921. [CrossRef]
11. Costello, J.J.A.; West, G.M.; McArthur, S.D.J. Machine Learning Model for Event-Based Prognostics in Gas Circulator Condition Monitoring. *IEEE Trans. Reliab.* **2017**, *66*, 1048–1057. [CrossRef]
12. Ayodeji, A.; Wang, Z.; Wang, W.; Qin, W.; Yang, C.; Xu, S.; Liu, X. Causal augmented ConvNet: A temporal memory dilated convolution model for long-sequence time series prediction. *ISA Trans.* **2022**, *123*, 200–217. [CrossRef] [PubMed]

13. Boyes, H.; Hallaq, B.; Cunningham, J.; Watson, T. The industrial internet of things (IIoT): An analysis framework. *Comput. Ind.* **2018**, *101*, 1–12. [CrossRef]
14. Aebersold, S.A.; Akinsolu, M.O.; Monir, S.; Jones, M.L. Ubiquitous Control of a CNC Machine: Proof of Concept for Industrial IoT Applications. *Information* **2021**, *12*, 529. [CrossRef]
15. Siva Vardhan, D.S.V.; Narayan, Y.S. Development of an automatic monitoring and control system for the objects on the conveyor belt. In Proceedings of the 2015 International Conference on Man and Machine Interfacing (MAMI), Bhubaneswar, India, 17–19 December 2015; pp. 1–6. [CrossRef]
16. Çınar, Z.M.; Abdussalam Nuhu, A.; Zeeshan, Q.; Korhan, O.; Asmael, M.; Safaei, B. Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. *Sustainability* **2020**, *12*, 8211. [CrossRef]
17. Djurdjanovic, D.; Lee, J.; Ni, J. Watchdog Agent—An infotronics-based prognostics approach for product performance degradation assessment and prediction. *Adv. Eng. Inform.* **2003**, *17*, 109–125. [CrossRef]
18. Liu, C.; Tang, D.; Zhu, H.; Nie, Q. A Novel Predictive Maintenance Method Based on Deep Adversarial Learning in the Intelligent Manufacturing System. *IEEE Access* **2021**, *9*, 49557–49575. [CrossRef]
19. Liu, H.; Yuan, R.; Lv, Y.; Li, H.; Gedikli, E.D.; Song, G. Remaining Useful Life Prediction of Rolling Bearings Based on Segmented Relative Phase Space Warping and Particle Filter. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–15. [CrossRef]
20. He, J.; Cen, Y.; Alelaumi, S.; Won, D. An Artificial Intelligence-Based Pick-and-Place Process Control for Quality Enhancement in Surface Mount Technology. *IEEE Trans. Components Packag. Manuf. Technol.* **2022**, *12*, 1702–1711. [CrossRef]
21. Dogan, A.; Birant, D. Machine learning and data mining in manufacturing. *Expert Syst. Appl.* **2021**, *166*, 114060. [CrossRef]
22. Cui, P.H.; Wang, J.Q.; Li, Y. Data-driven modelling, analysis and improvement of multistage production systems with predictive maintenance and product quality. *Int. J. Prod. Res.* **2021**, *60*, 1–18. [CrossRef]
23. Liveline Technologies. Multi-Stage Continuous-Flow Manufacturing Process. 2019. Available online: https://www.kaggle.com/datasets/supergus/multistage-continuousflow-manufacturing-process (accessed on 15 October 2022).
24. Liveline Technologies. Convert Your Manufacturing Assets into an Intelligent, Autonomous System Using AI-Based Process Controls. 2022. Available online: https://www.liveline.tech/ (accessed on 5 December 2022).
25. Pach, F.; Feil, B.; Nemeth, S.; Arva, P.; Abonyi, J. Process-data-warehousing-based operator support system for complex production technologies. *IEEE Trans. Syst. Man Cybern.- Part A Syst. Hum.* **2006**, *36*, 136–153. [CrossRef]
26. LJ Create. PETRA II Advanced Industrial Control Trainer. 2023. Available online: https://ljcreate.com/uk/engineering/petra-ii-advanced-industrial-control-trainer/ (accessed on 7 January 2023).
27. Nargesian, F.; Samulowitz, H.; Khurana, U.; Khalil, E.B.; Turaga, D.S. Learning Feature Engineering for Classification. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17), Melbourne, Australia, 19–25 August 2017; pp. 2529–2535.
28. Wan, X. Influence of feature scaling on convergence of gradient iterative algorithm. *J. Phys. Conf. Ser.* **2019**, *1213*, 032021. [CrossRef]
29. Sangodoyin, A.O.; Akinsolu, M.O.; Pillai, P.; Grout, V. Detection and Classification of DDoS Flooding Attacks on Software-Defined Networks: A Case Study for the Application of Machine Learning. *IEEE Access* **2021**, *9*, 122495–122508. [CrossRef]
30. Akinsolu, M.O.; Sangodoyin, A.O.; Uyoata, U.E. Behavioral Study of Software-Defined Network Parameters Using Exploratory Data Analysis and Regression-Based Sensitivity Analysis. *Mathematics* **2022**, *10*, 2536. [CrossRef]
31. Sangodoyin, A.O.; Akinsolu, M.O.; Awan, I. A deductive approach for the sensitivity analysis of software defined network parameters. *Simul. Model. Pract. Theory* **2020**, *103*, 102099. [CrossRef]
32. Liu, P.; Lv, N.; Chen, K.; Tang, L.; Zhou, J. Regression Based Dynamic Elephant Flow Detection in Airborne Network. *IEEE Access* **2020**, *8*, 217123–217133. [CrossRef]
33. Shohani, R.B.; Mostafavi, S.A. Introducing a New Linear Regression Based Method for Early DDoS Attack Detection in SDN. In Proceedings of the 2020 6th International Conference on Web Research (ICWR), Tehran, Iran, 22–23 April 2020; pp. 126–132. [CrossRef]
34. Borchani, H.; Varando, G.; Bielza, C.; Larranaga, P. A survey on multi-output regression. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2015**, *5*, 216–233. [CrossRef]
35. Montesinos López, O.A.; Montesinos López, A.; Crossa, J. Support Vector Machines and Support Vector Regression. In *Multivariate Statistical Machine Learning Methods for Genomic Prediction*; Springer: Cham, Switzerland, 2022; pp. 337–378.
36. scikit-learn. scikit-learn Machine Learning in Python. 2022. Available online: https://scikit-learn.org/stable/ (accessed on 15 October 2022).
37. Platt, J. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Adv. Large Margin Classif.* **1999**, *10*, 61–74.
38. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [CrossRef]
39. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
40. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
41. Tipping, M.E. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **2001**, *1*, 211–244.
42. Dudani, S.A. The Distance-Weighted k-Nearest-Neighbor Rule. *IEEE Trans. Syst. Man Cybern.* **1976**, *SMC-6*, 325–327. [CrossRef]
43. An efficient instance selection algorithm for k nearest neighbor regression. *Neurocomputing* **2017**, *251*, 26–34. [CrossRef]

44. Ho, T.K. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; Volume 1, pp. 278–282.
45. Xu, M.; Watanachaturaporn, P.; Varshney, P.K.; Arora, M.K. Decision tree regression for soft classification of remote sensing data. *Remote Sens. Environ.* **2005**, *97*, 322–336. [CrossRef]
46. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
47. McDonald, G.C. Ridge regression. *Wiley Interdiscip. Rev. Comput. Stat.* **2009**, *1*, 93–100. [CrossRef]
48. Nokeri, T.C. Advanced Parametric Methods. In *Data Science Revealed*; Apress: Berkley, CA, USA, 2021; pp. 45–53.
49. Liu, K.; Xu, B.; Kim, C.; Fu, J. Well performance from numerical methods to machine learning approach: Applications in multiple fractured shale reservoirs. *Geofluids* **2021**, *2021*, 3169456. [CrossRef]
50. Hesterberg, T.; Choi, N.H.; Meier, L.; Fraley, C. Least angle and $\ell 1$ penalized regression: A review. *Stat. Surv.* **2008**, *2*, 61–93. [CrossRef]
51. Cohen, P.; West, S.G.; Aiken, L.S. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*; Lawrence erlbaum associates, Inc.: Mahwah, NJ, USA, 2003.
52. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
53. Negnevitsky, M. *Artificial Intelligence: A Guide to Intelligent Systems*; Pearson Education: Harlow, UK, 2005.
54. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
55. Google Research. Colaboratory. 2022. Available online: https://colab.research.google.com/ (accessed on 15 October 2022).
56. Isabona, J.; Imoize, A.L.; Kim, Y. Machine Learning-Based Boosted Regression Ensemble Combined with Hyperparameter Tuning for Optimal Adaptive Learning. *Sensors* **2022**, *22*, 3776. [CrossRef] [PubMed]
57. Passos, D.; Mishra, P. A tutorial on automatic hyperparameter tuning of deep spectral modelling for regression and classification tasks. *Chemom. Intell. Lab. Syst.* **2022**, *223*, 104520. [CrossRef]
58. Chicco, D. Ten quick tips for machine learning in computational biology. *BioData Min.* **2017**, *10*, 1–17. [CrossRef]
59. Wilcoxon, F. Individual comparisons by ranking methods. In *Breakthroughs in Statistics*; Springer: New York, NY, USA, 1992; pp. 196–202.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.