A data-driven approach to diagnosing throughput bottlenecks from a maintenance perspective

(article starts on next page)

A data-driven approach to diagnosing throughput bottlenecks from a maintenance perspective

Mukund Subramaniyan [a,*], Anders Skoogh [a], Azam Sheikh Muhammad [b], Jon Bokrantz [a], Björn Johansson [a], Christoph Roser [c]

[a] Department of Industrial and Materials Science, Chalmers University of Technology, Gothenburg 41296, Sweden
[b] Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg 41296, Sweden
[c] Department of Management Science and Engineering, Karlsruhe University of Applied Sciences, Karlsruhe 76133, Germany

ABSTRACT

Prioritising maintenance activities in throughput bottlenecks increases the throughput from the production system. To facilitate the planning and execution of maintenance activities, throughput bottlenecks in the production system must be identified and diagnosed. Various research efforts have developed data-driven approaches using real-time machine data to identify throughput bottlenecks in the system. However, these efforts have mainly focused on identifying bottlenecks and only offer limited maintenance-related diagnostics for them. Moreover, these research efforts have been proposed from an academic perspective using rigorous scientific methods. A number of challenges must be addressed, if existing data-driven approaches are to be adapted to real-world practice. These include identifying relevant data types, data pre-processing and data modelling. Such challenges can be better addressed by including maintenance-practitioner input when developing data-driven approaches. The aim of this paper is therefore to demonstrate a data-driven approach to diagnosing throughput bottlenecks, using the combined knowledge of the maintenance and data-science domains. Diagnostic insights into throughput bottlenecks are obtained using unsupervised machine-learning techniques. The demonstration uses real-world machine datasets extracted from the production line. The novelty of the research presented in this paper is that it shows how inputs from maintenance practitioners can be used to develop data-driven approaches for diagnosing throughput bottlenecks having more practical relevance. By gaining these diagnostic insights, maintenance practitioners can better understand shop-floor throughput bottleneck behaviours from a maintenance perspective and thus prioritise various maintenance actions.

1. Introduction

Maintenance is one of the important activities of a production system. The goal of maintenance is to achieve a high degree of machine availability and retain the equipment in proper condition, which jointly helps to meet the production system's target throughput (Swanson, 2001; Li & Ni, 2009). Of the various machines in a production system, availability is mainly constrained in one or two machines affecting the overall system throughput. These machines are called throughput bottlenecks (Goldrat & Cox, 1990). Previous research indicates that prioritising maintenance activities on throughput bottlenecks increases the availability of throughput bottlenecks and hence helps in increasing the throughput from production system (Li, Chang, Ni, & Biller, 2009; Li & Ni, 2009; Ni & Jin, 2012; Gopalakrishnan, Skoogh, & Laroque, 2013).

Two important things in facilitating the planning of maintenance activities in throughput bottlenecks are: 1) throughput bottlenecks need to be identified in a production system, and 2) maintenance-related diagnostic insights need to be obtained on them (Li, Ambani, & Ni, 2009). There has been much previous academic research into developing different methods to identify throughput bottlenecks in production systems (Roser, Nakano, & Tanaka, 2001; Roser, Nakano, & Tanaka, 2002; Sengupta, Das, & VanTil, 2008; Li, Chang, & Ni, 2009; Betterton & Silver, 2012; Yu & Matta, 2016; Li, 2018; Tang, 2019). Identifying throughput bottlenecks leads to the planning of generic maintenance activities on them (such as priortising them for reactive maintenance work orders). However, to facilitate the planning of detailed maintenance activities (e.g., initiating specific maintenance work orders), diagnostic insights are required that explain the possible root causes of

the bottlenecks from a maintenance perspective.

One way to get diagnostic insights and facilitate maintenance decision-making on planning the different maintenance activities into throughput bottlenecks is to analyse unplanned stops. Unplanned stops have been identified as one of the main reasons for the lower availability of bottlenecks (Li et al., 2009; Subramaniyan et al., 2018; Tang, 2019). In industrial practice, unplanned stops are managed by maintenance practitioners (Lee, Lapira, Bagheri, & Kao, 2013) and receive much attention as they contribute directly to lower production line throughput. Current industrial practice is either a) to select unplanned stops based on the experience of maintenance practitioners (Ni & Jin, 2012; Gopalakrishnan, Skoogh, Salonen, & Asp, 2019), or b) conduct a manual Pareto analysis based on the frequency of different unplanned stops (Labib, 2014) stored as event logs in the manufacturing execution system (MES). Given the changing dynamics of production systems, experienced-based decisions may be less accurate. Moreover, the Pareto approach (using frequency as the variable) overlooks infrequent stops of longer duration. Specifically, these stops may cause concern to production teams, and neglecting them can lead to disagreements between production and maintenance teams in a real-world setting (Gu, Jin, & Ni, 2015). It is, therefore, necessary to establish a data-driven approach that can give diagnostic information on bottlenecks, by systematically analysing event logs related to unplanned stop events. The need for such data-driven approaches to facilitating maintenance decision-making was also emphasised by (Holm, 2018; Segura et al., 2018) in a study conducted to identify the demands of future shop-floor teams.

Maintenance decision-making on bottlenecks can be effective if the event-log-based, data-driven approaches can clean and pre-process the raw data, obtain features of interest to provide diagnostic information (such as patterns of different unplanned stops, correlation of stops with product types and stops that are similar in behaviour). Such diagnostic information can be obtained from the event-log data and be visually depicted using a data-driven approach. This can be developed using machine-learning algorithms and visual analytics with a focus to incorporate both expertise from the emerging field of data science and statistics. Developing such approaches necessitate a combination of knowledge from maintenance and data-science practitioners. Maintenance-practitioner knowledge is required in various steps on developing data-driven approaches, such as assessing data quality, being aware of how data is used in machine-learning algorithms and, more importantly, adapting the approaches to suit real-world practice (Harding, Shahbaz, Srinivas, & Kusiak, 2006; Wuest, Weimer, Irgens, & Thoben, 2016; Bokrantz, Skoogh, La, Hanna, & Perera, 2017; Zenisek, Holzinger, & Affenzeller, 2019). Data-science practitioner knowledge will be used in selecting appropriate tools for extracting information from the data (Jordan & Mitchell, 2015; Gil, Honaker, Orazio, Garijo, & Jahanshad, 2019). Such an interdisciplinary approach (with an active feedback loop between maintenance and data-science practitioners) will, therefore, have a significant impact on improving industrial practice because the analysis incorporates the maintenance practitioner's view.

Accordingly, the purpose of this paper is to improve production system throughput by facilitating maintenance decision-making on bottlenecks. The aim is to demonstrate a data-driven approach of working with event-log data to obtain maintenance diagnostic information on throughput bottlenecks by using combined knowledge from the maintenance and data-science domains. The demonstration is based on real-world machine datasets, extracted from the production line. The maintenance diagnostics are obtained by combining the bottleneck identification method based on the active periods as proposed by (Roser, Nakano, & Tanaka, 2001) and unsupervised machine-learning techniques (specifically, k-means clustering technique). Visual analytic tools are then used to depict the clustering results. Such a demonstration will provide useful guidelines for other researchers and practitioners, helping them successfully adapt these data-driven approaches to different settings and production lines. There are three main contributions of the

research presented in this paper: (1) extending the field of data-driven throughput bottlenecks analysis research from detecting to diagnosing the throughput bottlenecks from a maintenance perspective, (2) determining that maintenance practitioners' inputs are necessary to develop a data-driven approach to diagnose the unplanned stops in throughput bottlenecks, and (3) demonstrating how maintenance practitioners' inputs can be used when constructing a data-driven approach for diagnosing throughput bottlenecks using a real-world industrial test study.

2. Theoretical background

Firstly, this section discusses the different bottleneck identification methods, as presented in the literature. Secondly, connected to the desired data-driven approach and its application to obtain maintenance diagnostic insights, the necessary concepts within the unsupervised machine-learning algorithms, especially the clustering techniques, are briefly presented. There is also a brief presentation of common techniques for visualising clustering results. Thirdly, there is a discussion of the need to integrate the maintenance practitioner's perspective in diagnosing throughput bottlenecks.

2.1. Throughput bottleneck identification and diagnosis

Various methods have been developed in the literature to identify throughput bottlenecks in production systems. The underlying logic behind them all is analysing the machines' event-log data. This data provides information on different machine states across a production time (such as producing, downtime, setup time, tool change over time, and so on (Roser et al., 2001)). The event-log data is stored in manufacturing execution systems (MES). Different bottleneck identification methods use data-mining techniques to detect throughput bottlenecks in a production system (the active period method for example (Roser et al., 2001; Roser, Nakano, & Tanaka, 2002), the inactive period method (Sengupta et al., 2008), the turning-point method (Li et al., 2009), the inter-departure time variance method (Betterton & Silver, 2012) and the Overall Equipment Effectiveness method (Tang, 2019)). (Yu & Matta, 2016) proposed a data-driven method to improve bottleneck identification accuracy using statistical methods. Although various methods are proposed in the literature, they focus on identifying throughput bottlenecks in the production system. They provide no diagnostic information explaining why the machines are bottlenecks. Of all the different methods, the active period method has the potential to give diagnostic insights into bottlenecks (Subramaniyan et al., 2018).

2.1.1. Active period method and diagnostic insights into throughput bottlenecks

The active period method was first proposed by (Roser et al., 2001). This method divides events into two categories: inactive and active. "Inactive events" are events when a process waits for another process. This occurs most commonly with the process being starved (waiting for the upstream process) or blocked (waiting for the downstream process). "Active events" are all other events where a process is not waiting for another process, including working, changeover, maintenance, and breakdowns.

Fig. 1 shows the events of the machine across a production time.

When the duration of the active events is computed for all machines in the production system, the one with the highest active duration is the throughput bottleneck machine. The active period method of bottleneck identification has been shown as accurate in identifying throughput bottlenecks (Roser, Nakano, & Tanaka, 2003; Yong-Cai & Qian-Chuan Zhao, 2005). One advantage of the active period method is that it can explain the reasons for throughput bottlenecks. For example, a machine may constitute a bottleneck because it has greater cycle-time variations or greater downtime due to unplanned stops or more changeovers.

(Subramaniyan et al., 2018) adapted the active period method of bottleneck identification and developed an MES-based data-driven
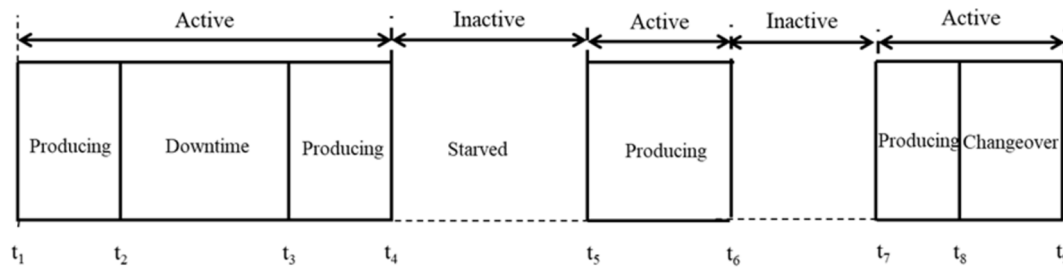
Fig. 1. Active and inactive events across a production time (adapted from (Roser et al., 2001)).

algorithm to identify throughput bottlenecks and give diagnostics insights into them. These are provided as total aggregated durations for the various events which make the machine a bottleneck. Although this is a first step towards understanding the different reasons for bottlenecks, insights into solely the aggregated duration do not aid the decision-making process when planning various maintenance activities. Practitioners still need to use a Pareto analysis (or their implicit knowledge) to manually interpret the different unplanned stops and plan accordingly. However, machine-learning techniques yield more detailed diagnostics; these can assist in planning a range of maintenance activities. One such technique that can be used to obtain diagnostic information on bottlenecks is unsupervised machine learning techniques, specifically, clustering technique.

2.2. Clustering

Clustering is one of the machine-learning algorithms. It belongs to the family of unsupervised machine-learning algorithms that group similar objects into clusters (Sharp, Ak, & Hedberg, 2018). They are also referred to as "segments". Clustering is called "unsupervised" as it groups similar objects not previously classified or labelled. Clustering uses mathematical techniques for multidimensional analysis. Using the variables (or features) of a set of objects, this procedure groups similar objects into clusters. The resulting individual groups contain objects more similar to each other when compared to those outside their group.

Compared to other techniques, such as sorting, Pareto or binning (often used for single variables), clustering can be applied to multiple variables. The best cluster solutions are those with the tightest individual cluster cohesion and the greatest individual cluster separation (Boutsidis, Zouzias, Mahoney, & Drineas, 2015). The three main types of clustering techniques are: 1) hierarchical (finds clusters using previously established clusters), 2) partitional (determines all clusters at once), and 3) Bayesian (generates posterior distribution over the collection of partitional data). In this study, partitional clustering techniques are used, specifically k-means clustering, to demonstrate how MES data can be used to obtain diagnostic information from a maintenance perspective.

2.2.1. K-means clustering technique

K-means is a technique commonly used for clustering purposes, especially for maintenance applications within manufacturing (Carvalho et al., 2019). It works advantageously with large datasets and tends to be more efficient in creating clustering solutions (Dhalmahapatra, Shingade, Mahajan, Verma, & Maiti, 2019). Input to k-means is the set of feature vectors X= {x_1, x_2, $x_{3,...,}x_N$}, the objects and the number of clusters (the "k" in k-means). Feature vectors can be numerical or categorical. The detailed mathematical algorithm of k-means clustering is explained in (Jain, 2010). Typically, k-means uses Euclidean distance (a way of quantifying or measuring similarity) to compute the distance between different points and the cluster centre. The output of k-means clustering is a grouping of the objects into "k" clusters.

Numerous methods are proposed in the literature to find the right number of clusters. The most common methods are: elbow method, silhouette method, X-means clustering, information criterion approach, and practitioner input (Kodinariya & Makwana, 2013). Practitioner input is needed when determining the number of clusters because practitioners can define the number of clusters based on operational constraints and usefulness in operational planning. This study, therefore, uses a combination of elbow method and practitioner input to specify the optimal number of clusters. The elbow method plots the total sum of squared distances within the clusters versus the number of clusters. These plots are also called "scree plots" (Zhu & Ghodsi, 2006). An ideal scree plot arises when, after a drastic decrease in the quantity (total within-clusters sum of the squares), this quantity decreases more slowly as the number of clusters increases. This indicates that the quality of the clustering is not increasing substantially with the increased number of clusters. The optimal number of clusters is when the total within the sum of squares drops radically.

Generally, the clustering will determine the patterns from the datasets and group together data points of similar variance, minimising the cost function (the sum of squares within the cluster). However, the next challenge is to visualise which information from each cluster is useful in better understanding the clusters and facilitating decision-making. This can be done using visual analytics.

2.2.2. Visual analytics of clustering

Visual analytics uses data visualisation techniques to synthesise information and derive insights from datasets (Cheng, 2018). It is used to effectively communicate the results obtained from analysing datasets to practitioners. This allows decision-making to be based on results obtained from practitioners' analyses, thus accelerating the path to insights and decisions (Segura et al., 2018). Visualisation is particularly important in understanding the clustering results and communicating them to practitioners in a way that enables effective action (Cheng, 2018). It is used to demonstrate the existence of the patterns found by clustering techniques. Common techniques used in visualising the information for each cluster based on their features include bar graphs for categorical features (Broeksema, Telea, & Baudel, 2012) and box plots for numerical ones (Blaschko & Lampert, 2008). These techniques allow us to compare different clusters based on different features and allow more detailed decision-making.

2.3. Need to integrate the maintenance practitioner's domain knowledge

The state-of-the-art research efforts in the literature (on developing bottleneck identification methods) are mainly presented from an academic researcher's point of view (Roser et al., 2001; Roser et al., 2002; Sengupta et al., 2008; Li et al., 2009; Betterton & Silver, 2012; Yu & Matta, 2016) and (Tang, 2019). These academic research efforts provide high-level scientific contributions, based on rigorous mathematical analysis and building different bottleneck identification methods. They thus expand the knowledge in the field of throughput bottlenecks and provide an analogical representation of outputs from the application of different bottleneck identification methods. From the maintenance practitioner's perspective, such academic research helps them to learn

new concepts on throughput bottleneck identification. They are then left to adapt these to real-world production systems, but with limited guidance as to the adaptation procedure. This is not an easy task as numerous challenges need to be addressed to successfully adapt the data-driven approaches proposed in the literature.

Applying different throughput bottleneck detection methods developed in the academic literature to the real-world involves multiple stages, such as: extracting suitable data, data cleaning, data processing, data modelling and validating bottleneck results. These stages involve many different choices. In the existing academic research, selections are made from an academic point of view with limited consideration of real-world practice. That said, such practical decisions cannot be made without integrating maintenance-practitioner expertise; using a real-world perspective allows them to better guide their selection. Real-world data is often noisy and needs significant cleaning up. Cleaning can be effectively carried out using input from maintenance practitioner's domain expertise, as they can identify the usable portion of the data and make a quick sanity check on it (Angiulli & Fassetti, 2014). Another example is that, in a production system, many continuous improvement efforts are undertaken by maintenance practitioners (Li, Ambani, et al., 2009). There may also be some structural changes in the production system, which can be difficult to detect using only the machine data. In this situation, a maintenance practitioner's guidance on

how much data to use in identifying bottlenecks can better reflect the real-world system dynamics. Therefore, this study demonstrates how maintenance-practitioner expertise can be used to identify and diagnose throughput bottlenecks from a maintenance perspective.

3. Demonstration of a data-driven approach to maintenance decision-making for throughput bottlenecks

The four-step methodology works with event-log data to obtain maintenance diagnostic information using combined knowledge from the maintenance and data-science domains: 1) data collection and data cleaning, 2) throughput bottleneck identification, 3) diagnostic insights into bottlenecks, and 4) interpretation and decision-making. The details of each step are shown in Fig. 2.

In step 1, the event-log data from a real-world production system is collected, cleaned and pre-processed to identify throughput bottlenecks. In step 2, the event-log data is analysed and the throughput bottlenecks are identified using the data-driven algorithm developed by (Subramaniyan et al., 2018). In step 3, maintenance-related diagnostic insights into throughput bottlenecks are obtained using k-means clustering technique. To apply k-means clustering technique, the unplanned stops from the event-log data specific to throughput bottlenecks are extracted and different features are computed. In step 4, the
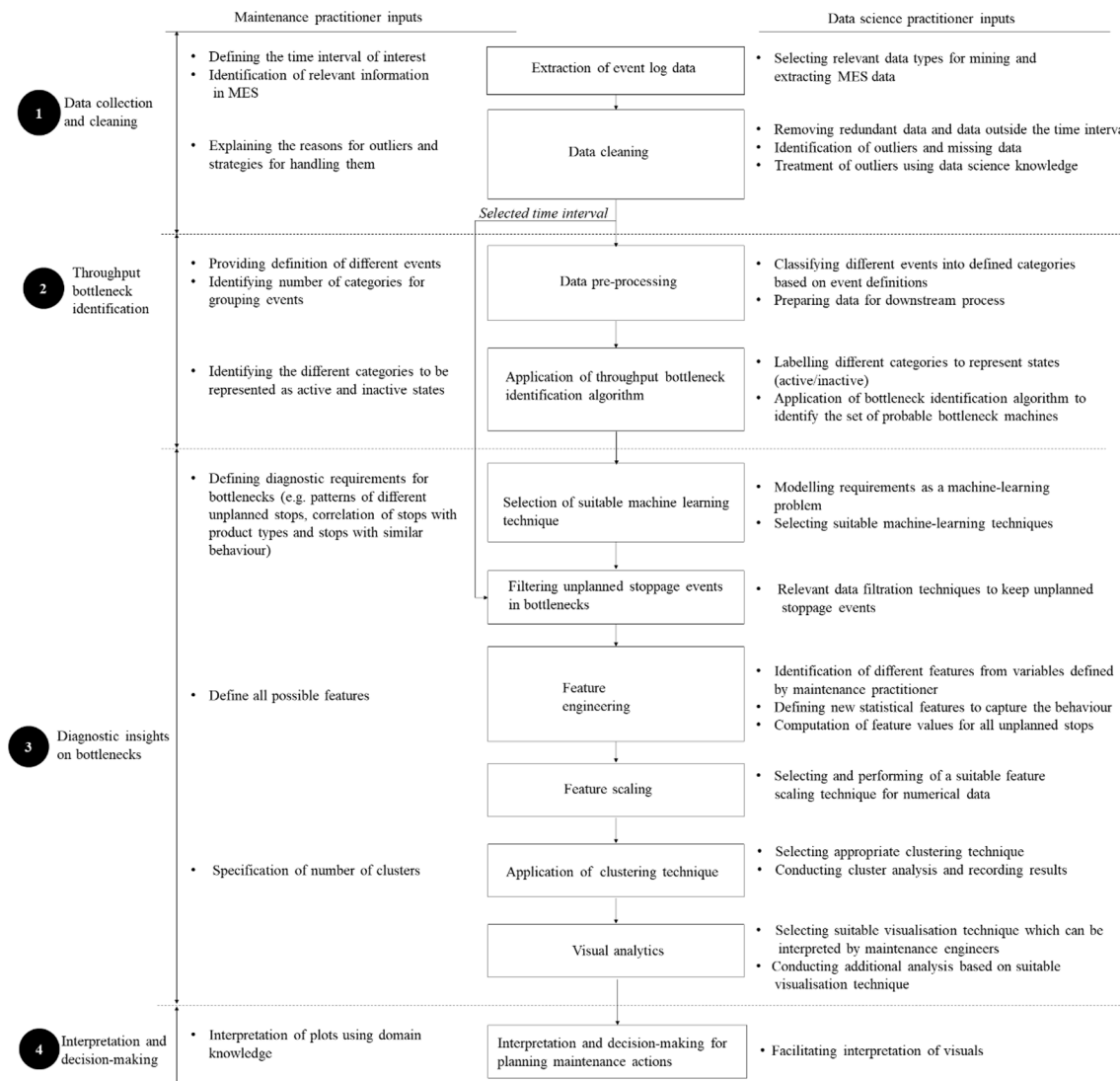

Fig. 2. Data-driven approach.

clustering results are interpreted and the different ways in which the results will facilitate maintenance decision-making are discussed. Steps 1, 2 and 3 are conducted by uploading the event-log dataset into R software (Version 3.4.3) and using libraries such as dplyr, DT, factor-extra, FactoMineR and ggplot2.

Through this demonstration, it is shown that maintenance-practitioner input is necessary for developing data-driven approaches with greater practical relevance. The detailed process and different inputs for each step (using knowledge from maintenance and data science) are summarised in Fig. 2.

3.1. Real-world production system

The demonstration uses event-log datasets extracted from a real-world production line. This section begins with a brief discussion of the real-world production system. It then presents the problem description and explains the need for diagnostic information on bottlenecks from maintenance practitioners in real-world production lines.

3.1.1. Description of the real-world production line

The production line is from an automotive manufacturer in Sweden. In this line, car body parts are welded at five different stations, as shown in Fig. 3. The line starts at Station 10 and ends at Station 80. Each station is connected to an MES which continuously records the different events in the station, plus their timestamps during the production run. The historical event log for each station can be extracted from the MES. A sample event-log record for Station 10 appears in Table 1 and shows the timestamps, event description, duration, and product type.

3.1.2. Problem definition

The maintenance engineers from the production line were looking for event-log-based, data-driven solutions to identify throughput bottlenecks and gain diagnostic insights into them from a maintenance perspective. These engineers had expertise in maintenance practices, but only limited knowledge of using data-science tools to analyse data and facilitate data-driven maintenance decision-making. As academic researchers, the authors had expertise in data-science tools but limited knowledge of the real-world production system presented in this study. The authors and maintenance engineers, therefore, complemented each other within this study, aimed at demonstrating how event-log data can be used to identify and diagnose throughput bottlenecks. In constructing such a data-driven approach, the author's main aim was to identify and understand the various inputs that maintenance engineers must provide.

3.2. Data collection and cleaning

This section begins with a presentation of extracting event-log data from a real-world production system. -

Extraction and collection of event-log data. To collect a production system's event-log data, the time interval within which the event-log data will be used must first be determined. This must be done by maintenance practitioners, who identify throughput bottlenecks for a specific time interval (such as half-year, quarter, month and so on). This definition is based on their own domain knowledge of the production system dynamics. If the interval is not specified by maintenance practitioners, the risk is that too great an interval will be analysed. This may be a poor reflection of the true bottlenecks, due to improvements made in the production system. The type of data which needs extracting from the MES system must then be determined; this cannot be done without

Table 1
Sample event-log data record for Station 10.

Station	Start time	End time	Events	Duration (sec)	Product type
10	03-04-2017 06:42:16	03-04-2017 06:44:24	1754010. AC_Process time	128	A
10	03-04-2017 06:44:25	03-04-2017 06:44:39	1754010. AD_Waiting output	14	A
10	03-04-2017 06:44:40	03-04-2017 06:44:42	1754010. AD_ProcessInput	2	B
10	03-04-2017 06:44:43	03-04-2017 06:45:46	1754010. AC_Process time	63	B
10	03-04-2017 06:45:47	03-04-2017 06:45:54	175010. AD_Waiting output	7	B
10	03-04-2017 06:45:55	03-04-2017 06:47:08	1754010. AC_Roberror	73	B

practitioner input. One advantage of practitioner input in this step is that it aids the selection of relevant data. It means no time is wasted exploring unnecessary data for throughput bottleneck identification and diagnostic information on unplanned stops.

In the production line used in this study, the maintenance engineers wanted to identify the throughput bottlenecks based on five months' worth of data (weeks 14–36 of 2017). Furthermore, the maintenance engineers specified the usable portion of the MES, comprising the event descriptions, time stamps and products produced within the defined time interval for each station. The authors then defined the relevant data types, so as to mine and extract the datasets from the MES. They then uploaded the datasets into the R software for further analysis. The output of this step is the event-log data records for all stations.

Data cleaning: The event-log data cleaning consists of simple, routine tasks such as removing data records made during the unscheduled production time, events recorded multiple times and so on. This can be done without any guidance from practitioners. However, more complex tasks do need maintenance-practitioner input. These include identifying outliers that don't conform to an expected pattern (based on other event-log records in the station) and removing unnecessary data fields and missing data. Handling outliers and missing data is an also an important step in identifying throughput bottlenecks and creating clusters, as outliers can impact the downstream clustering process. Maintenance practitioners can provide the necessary explanations on the outliers and also look into the cause of missing values before designing ways to handle them. For example, outliers which are not representative of the population (such as lack of incoming material, power outages) and those which are insignificant segments within the dataset and of no interest to the throughput bottleneck analysis (such as test runs of new products in the system) can only be identified with maintenance-practitioner input. This type of approach exploits maintenance-domain knowledge to improve the data cleaning process. This, in turn, improves the results when the bottleneck identification algorithm is applied.

In the production line used in this study, the authors conducted routine cleaning activities. They also identified the outliers and missing data, discussed the reasons for them with the maintenance engineers and made joint decisions on handling them by removing them from the event-log dataset. The output of this step is cleaned event-log data for
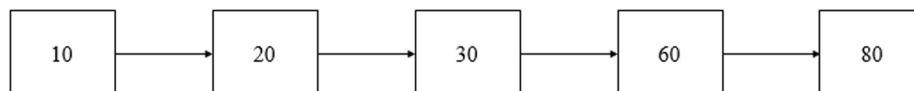

Fig. 3. Layout of the car body welding line.

each station in the production system.

3.3. Throughput bottleneck identification

In this step, the event-log data needs to be pre-processed to make it suitable for the application of the data-driven throughput bottleneck identification algorithm, as proposed by (Subramaniyan et al., 2018). The bottleneck identification algorithm is then applied to find the bottlenecks in the production system.

Data pre-processing. In this step, the cleaned data must be pre-processed to make it suitable to apply the data-driven bottleneck identification algorithm proposed by (Subramaniyan et al., 2018). This step requires maintenance-practitioner input as its tasks involve classifying the various events into different categories based on the definition of the individual event. This includes classification of production-cycle-related events, events representing unplanned stops and so on. The number of categories must be defined by the purpose of bottleneck identification. Usually, bottlenecks are due to greater downtime (of interest to maintenance teams), greater cycle time or greater setup time (of interest to production teams). Classification into these categories helps maintenance practitioners better understand the type of bottlenecks. Maintenance-practitioner input to event classification is important, as decisions about what event needs to be included in which category are best made based on real-world knowledge. For example, in identifying bottlenecks, it is best if practitioners decide whether an event representing a wait for a maintenance technician should be recorded as downtime. This will depend on whether they need the maintenance technician to be deemed part of the system.

In the production line used in this study, maintenance engineers classified all events in the event-log file of all stations into three types: 1) Producing (when the station is engaged in producing a product), 2) Unplanned stops (reflecting the down state of the station), and 3) Others (events representing blockage, starvation, lack of material, waiting). The authors then incorporated the classification into the event-log files of the stations.

Table 2 shows the total number of events for each station and their breakup classified as Producing, Unplanned stops and Others. Table 2 also reveals that, for each station, the number of distinct events representing unplanned stops is high and manual analysis of them is difficult. An example of the events classified into three categories for Station 10 appears in Table 3. The output from this step is the event-log data of all stations in which each event is classified into Producing, Unplanned stops, or Others.

Application of throughput bottleneck identification algorithm. The active and inactive states should be identified before applying the data-driven bottleneck detection algorithm as developed by (Subramaniyan et al., 2018). This step also requires maintenance-practitioner input, in classifying the categories defined in the previous step into active and inactive states.

In the production line used in this study, the authors gave the maintenance practitioners a simplified version of the active period method of bottleneck analysis. This allowed the maintenance practitioners to understand the active period method and at the same time helped them to identify the active and inactive states of stations in the production line. The maintenance engineers then classified "Producing" and "Unplanned Stops" as active states of the station, as they cause

blockage and starvation in other stations. "Others" were classified as inactive states, as this reflects the blockage and starvation of the stations. A sample output of this step is shown in Table 4.

The authors then labelled the different categories as active and inactive. They applied the throughput bottleneck identification and diagnostic algorithm as developed by (Subramaniyan et al., 2018), to identify the set of probable throughput bottlenecks in the production system and test its statistical significance. The resulting diagnostic insights into the bottlenecks explore the stations' active states. Fig. 4 summarises the bottleneck identification results from the algorithm, their statistical significance and diagnostic insights into bottlenecks. Fig. 4 shows:

> Set of throughput bottlenecks: Station 20 and Station 60
> Active period percentages: 66.41% (Station 20) and 64.19% (Station 60)
> Diagnostic insights into Station 20: 72.92% (Producing) and 27.08% (Unplanned Stops)
> Diagnostics insights into Station 60: 87.14% (Producing) and 15.86% (Unplanned Stops)

From the outputs of this step, the maintenance engineers inferred that the availability-constrained stations in the production system are 20 and 60. They also understand the contribution of unplanned stops in these stations. Although the engineers can use these types of diagnostic insights to understand the contribution of unplanned stops, further diagnostic information on different stops is required to plan specific maintenance actions.

3.4. Diagnostic insights into bottlenecks

In this step, the maintenance-related diagnostic insights into bottlenecks are obtained from the event-log data using unsupervised machine-learning techniques. The detailed steps for obtaining diagnostic insights are given below:

Selecting suitable machine-learning techniques. Based on the type of diagnostic information on bottlenecks required by the maintenance practitioners, the objective is to model those requirements into a machine-learning problem.

In the production line used in this study, the authors understand the requirements of maintenance engineers on the bottleneck Stations 20 and 60. From the stations' event-log datasets, the maintenance engineers wanted to identify different unplanned stops patterns, the correlation of stops with product types and stops that exhibited similar behaviour. Based on these requirements and the nature of event-log datasets as shown in Table 3, the authors suggested that the requirements could best be modelled as an unsupervised machine-learning problem – specifically, a clustering problem.

Filtering unplanned stop events from the event-log data of bottlenecks. In this step, the unplanned stop events from throughput bottlenecks are filtered to obtain maintenance diagnostic information.

In the production line used in this study, the authors extracted events representing unplanned stops in bottleneck Stations 20 and 60. The events came from the event-log datasets obtained as output from the data-cleaning step.

Feature engineering. The main objective of clustering is to extract patterns to turn the unplanned stops data into knowledge. In this step, unplanned stop features need to be created in order to conduct the clustering process. The most challenging activity in this step is identifying the different possible features; this requires maintenance-practitioner knowledge. Practitioners can better define these features using their domain experience, a useful factor in maintenance decision-making. Thereafter, the set of final features needs to be selected from different possible features, and this can be best done with the data-science knowledge. This allows for determination of whether the features exhibit high correlation, multicollinearity, etc., and also frames

Table 2
Number of events for the stations in the production line.

| Station	10	20	30	50	60	80
Total number of distinct events	44	67	48	18	285	64
Number of events related to Producing	2	4	2	1	1	1
Number of events related to Unplanned Stops	39	52	42	15	274	61
Number of events relating to Others	3	11	4	2	10	2

Table 3

Classification of events of Station 10 into different categories.

Start time	End time	Events	Duration (sec)	Product type	Category
03-04-2017 06:42:16	03-04-2017 06:44:24	1754010.AC_Process time	128	A	Producing
03-04-2017 06:44:25	03-04-2017 06:44:39	1754010.AD_Waiting output	14	A	Others
03-04-2017 06:44:40	03-04-2017 06:44:42	1754010.AD_ProcessInput	2	B	Producing
03-04-2017 06:44:43	03-04-2017 06:45:46	1754010.AC_Process time	63	B	Producing
03-04-2017 06:45:47	03-04-2017 06:45:54	175010.AD_Waiting output	7	B	Others
03-04-2017 06:45:55	03-04-2017 06:47:08	1754010.AC_Roberror	73	B	Unplanned Stop

Table 4

Classification of categories into active and inactive states of Station 10.

Start time	End time	Events	Duration (sec)	Product type	Category	Active/inactive state
03-04-2017 06:42:16	03-04-2017 06:44:24	1754010.AC_Process time	128	A	Producing	Active
03-04-2017 06:44:25	03-04-2017 06:44:39	1754010.AD_Waiting output	14	A	Others	Inactive
03-04-2017 06:44:40	03-04-2017 06:44:42	1754010.AD_ProcessInput	2	B	Producing	Active
03-04-2017 06:44:43	03-04-2017 06:45:46	1754010.AC_Process time	63	B	Producing	Active
03-04-2017 06:45:47	03-04-2017 06:45:54	175010.AD_Waiting output	7	B	Others	Inactive
03-04-2017 06:45:55	03-04-2017 06:47:08	1754010.AC_Roberror	73	B	Unplanned Stops	Active
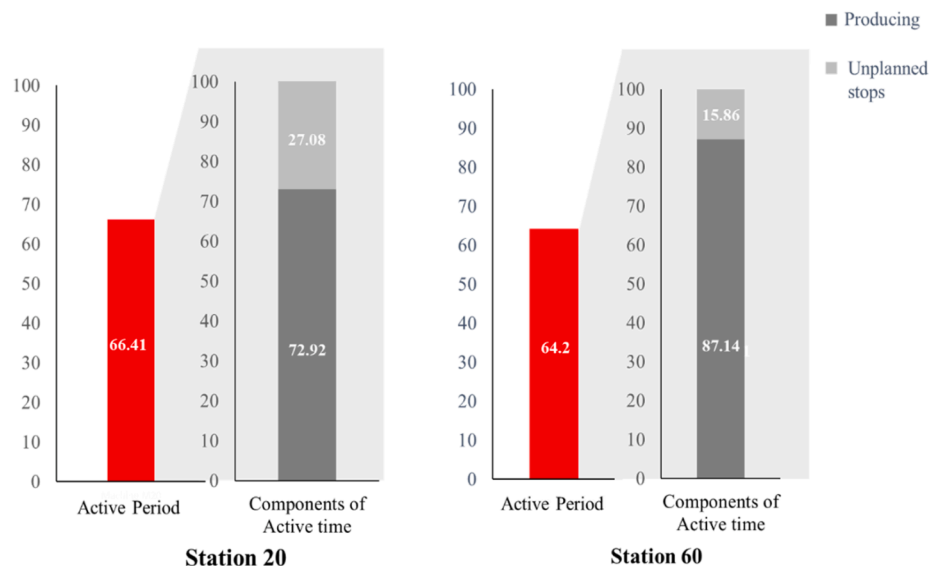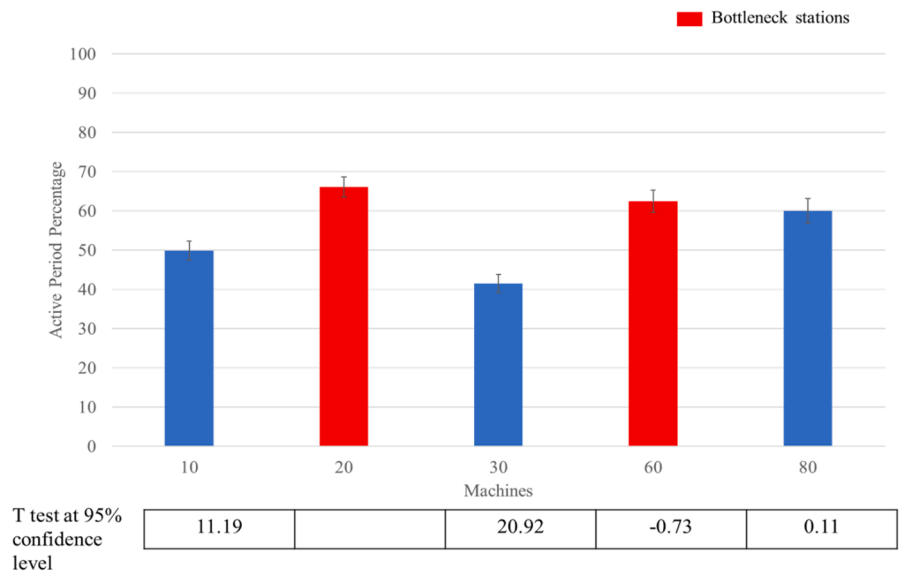
Fig. 4. Throughput bottleneck identification results (figure formats adapted from (Subramaniyan et al., 2018)).

new features that would better capture the variations.

In the production line used in this study, the MES datasets of the stations, unfortunately, did not have many features that can describe unplanned stops. The only direct independent features that can describe unplanned stops available in the MES data were total duration and the product types. Though these two were useful features, they do not explain unplanned stops completely. Therefore, the maintenance engineers and authors worked together to identify other features that could best describe unplanned stop behaviour. Maintenance engineers knowing the total frequency was one independent feature used to describe unplanned stops. Thereafter, new statistical features were created, based on the two independent features, "total duration" and "total frequency". These are, respectively: 1) the standard deviation of the duration, 2) the coefficient of variation of the duration, and 3) mean stop time. Both standard deviation and coefficient of variation of duration measure the variation of duration. However, the coefficient of variation is a relative metric that can be used to compare different unplanned stops, whereas standard deviation is an absolute measurement that cannot be used in this way. The authors, therefore, chose coefficient of variation of duration as one metric, to represent the variation in unplanned stop duration. Overall, the mean stop time and coefficient of duration statistical features will help distinguish the different unplanned stops structurally and will better reveal their behaviour. The definitions of each feature are given below:

Total stop duration: for every product type, total stop duration is the sum of all the time elapses for a particular unplanned stop event type. Stop duration includes the station's waiting time for maintenance personnel to attend the unplanned stop and the time taken to conduct the actual maintenance operations to restore the station.

Total frequency: for every product type, total frequency is the sum of the frequency of a particular unplanned stop event type.

Coefficient of variation of duration: for every product type, coefficient of variation represents a measure of relative variability in the duration of each unplanned stop event type.

Mean stop time: for every product type, mean stop time represents the average stop time of a particular unplanned stop event type.

Product type: represents the product in the station when the unplanned stop event type occurred.

These features will allow maintenance engineers to prioritise the stops for improvement activities. Thereafter, the authors compute the values for different features of unplanned stops. The numerical features are: total stop duration, total frequency, coefficient of variation of duration and mean stop time. In the categorical feature, product type, every product type is turned into a binary variable (also known as one-hot encoding) to allow the application of clustering techniques. An example of features extracted for different stops for station 20 appears in Table 5.

The output of this step is the various unplanned time events with their five features for Stations 20 and 60.

Feature scaling: The next step is to scale the different numerical features. Scaling of the data is an important consideration when preparing it for the clustering process. The idea behind scaling is to put the relative size of the feature values on the same scale. If they are not scaled, the actual data patterns may not be revealed correctly. Because clustering solutions are very sensitive to this kind of difference in scale, the literature shows many techniques for standardising feature values to similar scales. There are various feature-scaling methods such as z-score (also called standard scaling), log transformation and min-max normalisation. When a scaling method is used, the resulting values can then be used without over-weighting the analysis with larger observations about the features. Scaling also eliminates algorithm bias towards these larger observations.

In the production line used in this study, the authors observed that, prior to scaling, the clusters were not well separated. The authors then scaled the numerical features (total stop time, total frequency, coefficient of variation of duration and mean stop time) using three different scaling methods: z-score, min–max, and log transformation. Of all three, the authors observed that the z-score scaling results contained well-separated clusters. Z-score scaling measures the number of standard deviations each value is from the mean. The output from this step is the scaled features of unplanned stop events for Stations 20 and 60.

Application of clustering technique: In this step, natural groups of unplanned stops (based on the different features) must be identified. To do that, the appropriate clustering technique needs to be selected. This is best done by using data-science knowledge.

In the production line used in this study, the authors decided that basic k-means clustering (with Euclidean distance as the default distance measure) would be relatively straightforward to implement. It would allow them to group different unplanned stop events, based on the nature of the problem and feature values.

K-means clustering requires two types of input: 1) collecting features for each type of unplanned stop event, and 2) the number of clusters. The former is obtained from the feature-scaling output. The choice of an optimal number of clusters is heavily dependent on the objective issues of clustering and maintenance management practices. The overall goal is to cluster different unplanned stop events and better understand their nature. The number of clusters should, therefore, be defined based on maintenance practitioners' requirements and expertise. This is done by evaluating the usefulness of producing k clusters to better classify the various unplanned stops. However, maintenance practitioners should also be guided by the fact that choosing too few clusters (which may be interpretable but not fit all the descriptors), or too many (which may be good when characterising features) may not work well operationally. By using maintenance-practitioner input, the process of determining clusters can be balanced, based on the practical value and fit of the data.

In the production line used in this study, the authors used the elbow method to identify the appropriate number of clusters. The results of the elbow method for Stations 20 and 60 are shown in Figs. 5 and 6. From these figures, it can be seen that the distortion diminishes as the number of clusters increases. For Station 20, from Fig. 5, the elbow can be taken between Clusters 1 and 3 as the distortion goes down rapidly between these clusters and after Cluster 3 the distortion goes down slowly. So, the authors recommended that three clusters would be appropriate to group unplanned stops for Station 20. For Station 60, as shown in Fig. 6, finding the elbow point on the curve is a challenge because the curve smoothly decreases from Cluster 1. For this step, the authors, therefore,

Table 5
Features of unplanned stops for Station 20.

| Events	Total stop time (sec)	Total frequency	Coefficient of variation of duration (sec)	Mean stop time (sec)	Product types					
					A	B	C	D	E	F
1754020PCSBodyID.AC_NoDataBodyID	683	9	1.09	75.88	0	1	0	0	0	0
1754022 M3.AC_KeepPosSetAlmCyl2	581	7	0.92	83.00	0	1	0	0	0	0
1754020R04Alarm.AC_ProcEqAirError	627	6	0.25	104.50	0	0	0	0	0	1
1754022 M3.AC_KeepPosSetAlmCyl2	466	4	0.22	116.50	0	0	0	0	0	1
1754020R02Com.AC_R02InterlockReq3_4	302	5	0.26	60.40	0	0	1	0	0	0
1754020R04Alarm. AC_StandCoverStillClosedError	248	2	0.09	124.00	1	0	0	0	0	0
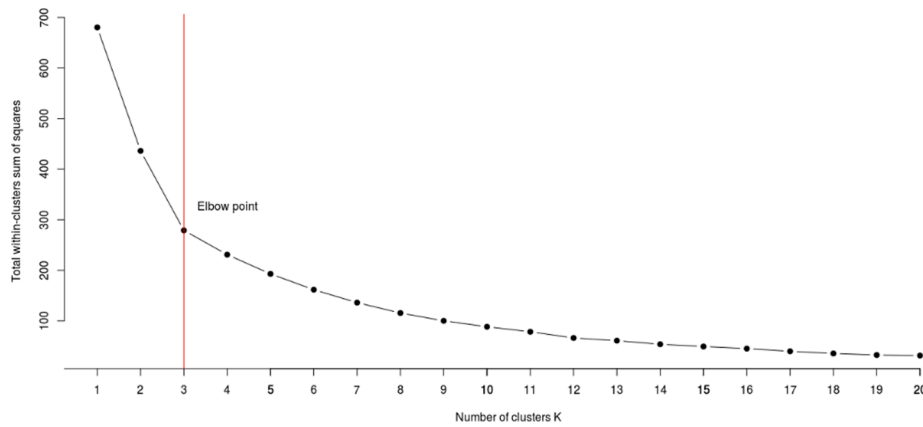
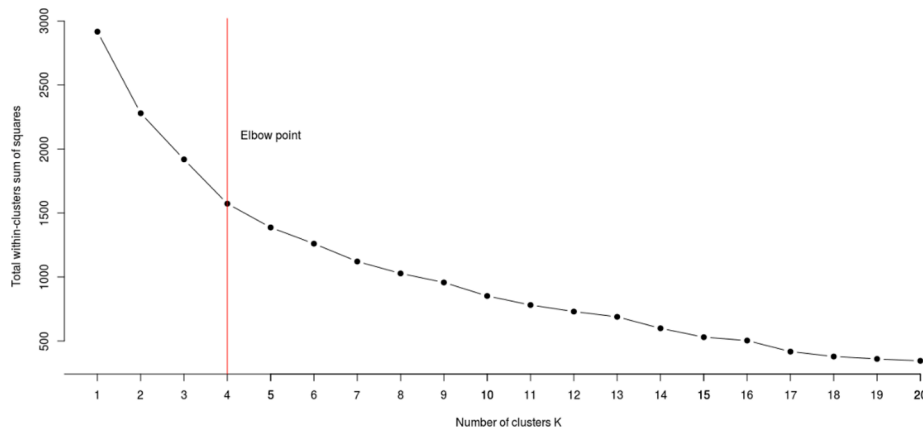**Fig. 5.** Results from elbow method to select number of clusters for Station 20.

Fig. 6. Results of elbow method for selecting number of clusters for Station 60.

used their judgement when interpreting Fig. 6. In doing so, they found that the decrease after Cluster 4 was not as relatively sharp before Cluster 4. The authors, therefore, recommended that four clusters would be appropriate for Station 60, to group unplanned stops. The main goal of clustering unplanned stops in Stations 20 and 60 is to facilitate maintenance decision-making about them. Thus, the authors and maintenance engineers jointly evaluated the number of clusters (discovered using the elbow method) and assessed them from the practical perspective of maintenance operations, based on the activities at each station. From a practical perspective, the number of clusters found from the elbow method for Station 20 was appropriate and reasonable. However, for Station 60, which has more unique unplanned stops (as shown in Table 2), the maintenance engineers and authors jointly decided to select five clusters instead of the four, as previously recommended by authors. This taught the maintenance engineers how cluster analysis could help them think of unplanned stops in terms of clusters.

Once the k has been decided based on the collaborative decision about the outcome of the elbow method with the maintenance engineers, the k and event logs for unplanned stops with scaled features for Stations 20 and 60 are given as an input separate to the k-means clustering. The k-means clustering is then run by the authors. The output of the k-means clustering is the assignment of unplanned stops with a specified cluster. The number of unplanned stops in each cluster is shown in Table 6.

Following this step, the authors extract the cluster numbers for each unplanned stop and add them back to the initial dataset (represented in Table 5), to facilitate visualisation of clustering results. An example of a dataset with added clusters is shown in Table 7.

Table 6
Number of unplanned stops belonging to each cluster for Stations 20 and 60.

Cluster number	Number of unplanned stop events	
	Station 20	Station 60
1	118	1
2	24	196
3	1	35
4	–	375
5	–	8

Visual analytics. The challenge in this step is to gain meaningful insights into the different clusters and how they can facilitate maintenance decision-making. The number of features in the clustering is usually large, so understanding clusters based on these features requires different visualisation techniques. One method is to visualise the clustering results using box and whisker plots for numerical features and bar plots for categorical features. This is a good way to communicate interesting groups of unplanned stops to practitioners. The aim is also to communicate them in a way that is easy for practitioners to interpret and understand. Most practitioners will be interested to know how clusters affect their operational practices. Thus, using box and whisker plots for numerical features and bar graphs for categorical features allows them to easily understand the difference between clusters and gets them there more quickly. It can also help practitioners gain intuition about the data and cluster results.

In the production line used in this study, the post-clustering results must be represented visually to facilitate maintenance decision-making by maintenance engineers. Therefore, the authors created box and

Table 7
Unplanned stops of Station 20 with cluster number added.

Events	Total stop time (sec)	Total frequency	Coefficient of variation of duration (sec)	Mean stop time (sec)	Product types						Cluster number
					A	B	C	D	E	F	
1754020PCSBodyID.AC_NoDataBodyID	683	9	1.09	75.88	0	1	0	0	0	0	1
1754022 M3.AC_KeepPosSetAlmCyl2	581	7	0.92	83.00	0	1	0	0	0	0	1
1754020R04Alarm.AC_ProcEqAirError	627	6	0.25	104.50	0	0	0	0	0	1	1
1754022 M3.AC_KeepPosSetAlmCyl2	466	4	0.22	116.50	0	0	0	0	0	1	1
1754020R02Com.AC_R02InterlockReq3_4	302	5	0.26	60.40	0	0	1	0	0	0	1
1754020R04Alarm. AC_StandCoverStillClosedError	248	2	0.09	124.00	1	0	0	0	0	0	1

whisker plots for each of the numerical features at Stations 20 and 60 and a bar plot for the product-type categorical feature. The box plots for numerical features (total frequency, total stop time, coefficient of variation in the duration and mean duration) and bar plots for categorical features for Stations 20 and 60 are shown in Figs. 7 and 8. These figures are not meant to be interpreted directly. The authors used this to demonstrate how cluster results are visualised. The interpretation and decision-making are carried out in conjunction with the maintenance practitioners, as shown in the next section (Section 3.5).

3.5. Interpretation and decision-making

The generated box plots for numerical features and bar plots for categorical features can be used to examine the distinguishing characteristics of each cluster and identify substantial differences between them. The comparison of different clusters provides decision support for maintenance decision-making on throughput bottlenecks. There are numerous ways to interpret the plots and interpretation is highly dependent on the overall goal, nature of the production system and machines, and the maintenance practitioners' expertise. Generally, the
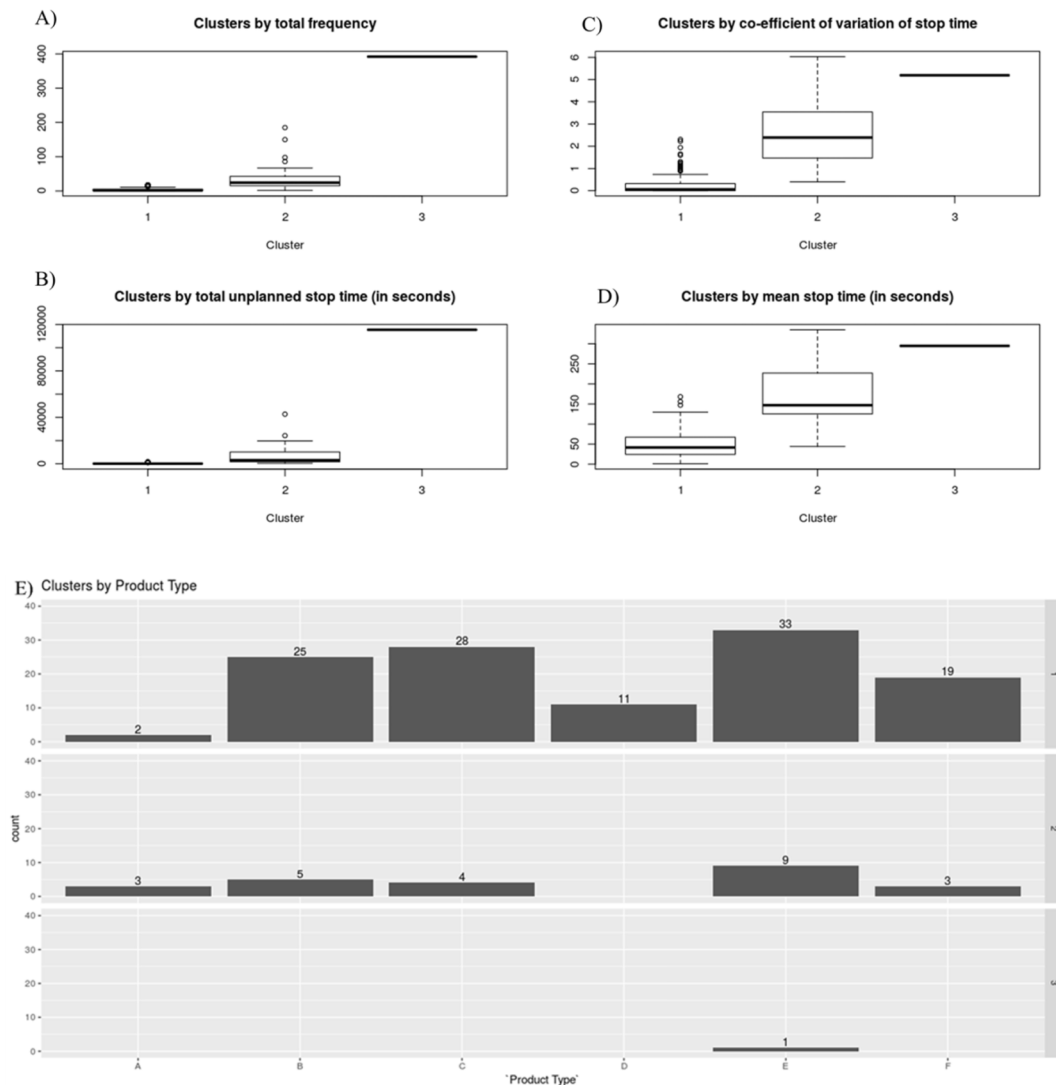

Fig. 7. Feature-based box plots for Station 20. A) Represents the spread of the total frequency for unplanned stops within each cluster. B) Represents the spread of the total stop time for unplanned stops within each cluster. C) Represents the spread of the coefficient of variation for unplanned stops within each cluster. D) Represents the spread of the mean stop time for unplanned stops within each cluster. E) Represents how many times a specific product-related stop is captured in each cluster.
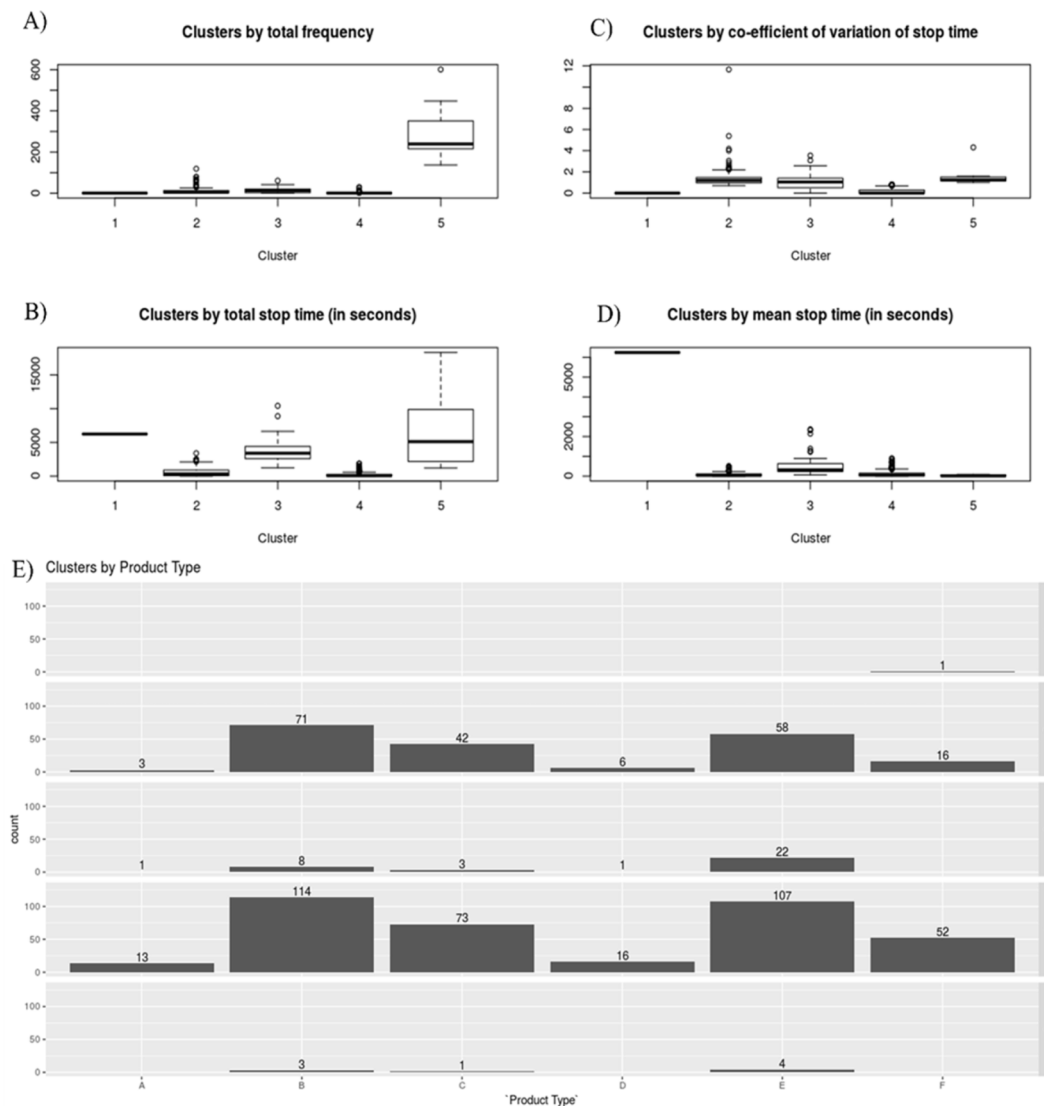
Fig. 8. Feature-based box plots for Station 60. A) Represents the spread of the total frequency for unplanned stops within each cluster. B) Represents the spread of the total stop time for unplanned stops within each cluster. C) Represents the spread of the coefficient of variation for unplanned stops within each cluster. D) Represents the spread of the mean stop time for unplanned stops within each cluster. E) Represents how many times a specific product-related stop is captured in each cluster.

different plots reveal the possible relationships in the different clusters. With the maintenance practitioners' domain knowledge, these clusters can be analysed in different ways and be prioritised for maintenance activities.

In the production line used in this study, Figs. 7 and 8 indicate that different clusters behave in different ways. Moreover, the clusters show some variability with respect to each feature; the k-means clustering can pick that up. The authors and maintenance engineers jointly interpreted different plots and explored which clusters could be prioritised for maintenance actions at Stations 20 and 60. Although there are numerous ways to interpret the plots, the summary below shows one way of interpreting the different clusters of Stations 20 and 60 for maintenance actions.

Station 20: As shown in Fig. 7A, the more frequent unplanned stops tend to fall in Cluster 3. Moreover, Cluster 3 also has the stops with a greater total unplanned stop time (as shown in Fig. 7B) and a higher coefficient of variation in duration (as shown in Fig. 7C). There again, it can be seen from Table 6 that Cluster 3 has only one type of unplanned stop. This stop only happens when the station produced product type E (as seen in Fig. 7E). In summary, it can be concluded that the unplanned stop in Cluster 3 happens very frequently only for product type E and

that, whenever it happens, it has a high variability of duration. Therefore, Cluster 3 has the most influence in reducing the availability of the station compared to the stops in other clusters. To improve availability, maintenance actions may be initiated. These include: a) training maintenance technicians to prioritise this stop and handle it efficiently, b) standardising the tasks towards restoring the station and reducing the variability in stop time, c) initiating technical diagnosis to identify why this stop happens in product E, and d) exploring solutions to phase out this stop.

It is interesting to note that, after Cluster 3, Cluster 2 has unplanned stops with a high coefficient of variation in stop time, as seen in Fig. 7D. Cluster 2 has 24 unplanned stops (Table 6). The variability in Cluster 2 comes mainly from product types E and B, as shown in Fig. 7E. This is an interesting cluster to examine after Cluster 3. Maintenance actions may be directed at unplanned stops mainly connected with product type E in this station. Cluster 1, which has 118 unplanned stops (Table 6) may be of least importance when it comes to initiating maintenance actions, as the stops are relatively infrequent (Fig. 7A). The total stop time is low (Fig. 7B), the mean stop time is lower (Fig. 7D) and the stop times are not widespread compared to other clusters (Fig. 7C). This indicates that Cluster 1 has mainly short unplanned stops and thus the least reduction

of availability for the station.

Station 60: as can be seen from Fig. 8A, Cluster 5 has the highest frequency of unplanned stops compared to other clusters. Cluster 5 also has a significant spread of total stop time (Fig. 8B) but, relatively speaking, a lower mean stop time compared to other clusters (Fig. 8D). There again, Cluster 5′s variation of duration is not very high, as can be inferred from Fig. 8C. Also, Cluster 5 has only eight stops as reported in Table 6; these are mainly related to product types E and B. In summary, prioritising this cluster for maintenance actions may have a major influence in improving the availability of Station 60. Next to Cluster 5, Cluster 3 (which has 35 stops) may be interesting to look at, to improve the availability of the station as its relative spread of total stop time, mean stop time and variation of stop time are greater. The majority of stops in Cluster 5 occur when the station is producing product type E. However, Cluster 1 had only one stop. This occurred just once when running product type F. Detailed analysis may be conducted on this stop to check whether the actions taken were appropriate and ensure this stop does not occur again.

In summary, the collaborative effort between the maintenance engineers and authors on visualising the clustering results may serve as decision support for maintenance decision-making. The maintenance engineers may also hold a brainstorming session with maintenance technicians/operators, to identify the root causes of different stops in the prioritised clusters and create an action register for corrective action. Moreover, by analysing the stops in different clusters, broader issues can be revealed, such as those related to process standardisation and automation design (if any). An action plan can then be initiated to design out certain specific stops. Moreover, maintenance engineers may also name the clusters which summarise the findings. For example, Cluster 3 at Station 20 may be named as a high-frequency, high-duration, high-variation of duration cluster (as shown in Fig. 7). Such methods can help engineers to quickly understand the basic makeup of the cluster. The unplanned stop events in each cluster can then be studied in detail and specific maintenance activities initiated. Overall, the different plots guide the maintenance engineers to look at unplanned stops in different ways, ask appropriate questions to the maintenance technicians and initiate maintenance activities.

4. Discussion

This study demonstrates a data-driven approach to working with event-log data and gaining maintenance diagnostic information on bottlenecks, using the combined knowledge of the maintenance and data-science domains. This demonstration is effected by using event-log datasets extracted from a real-world production system. The academic and practical contributions of such a demonstration are explained in this section.

4.1. Academic contributions

Compared to previous studies on throughput bottleneck analysis (Roser et al., 2001; Sengupta et al., 2008; Li et al., 2009; Betterton & Silver, 2012) and (Subramaniyan et al., 2018), the proposed study advances on previous ones by providing diagnostics on bottlenecks from a maintenance perspective. Maintenance-related diagnostic information is obtained by extracting the different features of unplanned stops and applying unsupervised machine-learning based k-means clustering technique. Feature-wise plots are then constructed to aid understanding of different clusters and maintenance decision-making on handling unplanned stops. In the main, such an approach reduces the ambiguity between production and maintenance practitioners, when it concerns reducing unplanned stops in bottlenecks and improving throughput. This type of approach will enable joint production and maintenance planning. The solution presented in this paper is aligned with the industry's need to develop data-driven approaches for maintenance decision-making, as presented by Holm (2018) and Segura et al. (2018).

While the existing literature acknowledges the advantages of integrating domain knowledge when developing data-driven approaches (Harding et al., 2006; Wuest et al., 2016), no concrete examples are demonstrated on how this process could be realised in the context of throughput bottleneck analysis. The demonstration given in this paper contributes to the recent, ongoing discussions among academic researchers as to how the practitioners' expertise can be integrated when developing data-driven machine-learning approaches to decision-making (Jordan & Mitchell, 2015; Gil et al., 2019). Such approaches will be more relevant to developing solutions to real-world problems and can thus increase the use of scientific outcomes in the manufacturing industries. Moreover, this paper highlights how academic researchers working on developing data-driven approaches should give specific consideration to those steps requiring practitioner input. Such considerations will be useful to maintenance practitioners as they adapt the approaches to real-world settings.

4.2. Practical Contributions

The current practice by maintenance practitioners in the industry of prioritising maintenance activities in bottlenecks is based on the Pareto analysis of unplanned stop frequency. Using a Pareto-based approach on a single feature (Li & Ni, 2009; Labib, 2014) omits the effects of other features. Compared to existing approaches, this paper proposes that complex MES event-log datasets can be systematically explored to summarise the behaviour of unplanned stops and uses machine-learning-based clustering techniques to identify which ones show similar behaviour. This will lead to better diagnostic insights into throughput bottlenecks. Clustering enables the larger group of unplanned stops to be broken into a set of smaller clusters, based on a set of features. Feature-based visualisation of clustering results helps visualise those stops showing similar behaviour, but which cannot be recognised by Pareto analysis. It may be anticipated that the stops in each cluster will behave in similar fashion. This makes the operations involved in managing unplanned stops much easier. The data-driven approach presented in this paper can readily be converted into an algorithm and integrated with MES data. Using that algorithm, maintenance practitioners can periodically analyse throughput bottlenecks and plan maintenance actions to improve availability and, hence, system throughput.

This paper's demonstration of a step-by-step data-driven approach to diagnosing throughput bottlenecks gives industrial maintenance managers a higher level of insight when planning projects for developing data-driven approaches, especially in the era of big data, to facilitate maintenance decision-making. In the demonstration given in this paper, the authors shared data-science expertise in developing data-driven approaches. However, in industrial practice, the development of data-driven approaches is usually carried out by data-science practitioners with expertise in data engineering relating to statistics and machine-learning. In such cases, data-science and maintenance practitioners need to work as a team, complementing each other with their expertise and developing data-driven approaches like those emphasised in (Jordan & Mitchell, 2015). Working together, practitioners from data science and maintenance can use the right technologies and techniques to solve the right problems. Also, instead of data-science practitioners attempting to build highly sophisticated data-driven approaches, the experience of maintenance practitioners can be consolidated within the data-driven process. Thus, a balance is struck between sophisticated data-driven approaches and a data-driven approach that adds value. Such an approach will also make the practitioner aware of the limits of data-driven methods and foster judicious use of data-driven decision-making. This type of interdisciplinary approach will facilitate the institutionalisation of data-driven approaches in companies. It will also enhance the acceptance of insights obtained from practitioners' approaches, thus enabling data-driven maintenance decisions.

4.3. Limitations and future work

Some working limitations must be considered when implementing a data-driven approach to obtain maintenance diagnostics on bottlenecks, as demonstrated in this study. All unplanned stops on machines must be monitored, with their time stamps. The features of each unplanned stop event used in this study are total frequency, total stop time, coefficient of variation of duration, and mean duration. No other features were included. This is because no other features were extractable from the MES. However, other features such as safety, maintenance action logs, necessary skill level for the maintenance team to address each stop event, spare parts and criticality may be added to gain more enriched diagnostic insights; this will be factored into future work. To do this, manufacturing companies are encouraged to store more detailed features of their unplanned stop events. Moreover, the approach proposed in this paper can be further enhanced by the addition of sensor-based information from machine components. The fusion of such information from the sensor with machine-level event-log data may enable further use of machine-learning approaches to gain deeper diagnostic insights and lead to more accurate planning of maintenance activities.

5. Conclusions

In a production system, maintenance decision-making to improve the availability of throughput bottlenecks is a complex process. To facilitate it, maintenance practitioners need to know two things: 1) the throughput bottlenecks in the production system, and 2) maintenance-related diagnostic insights into bottlenecks. The existing research efforts from the literature focus on developing methods to detect throughput bottlenecks. This paper extends the research on throughput bottlenecks from detecting the throughput bottlenecks to diagnosing them from a maintenance perspective. This has been achieved by constructing a step-by-step, data-driven approach using the event-log data of the underlying production system. Maintenance related activities are mainly focused on unplanned stops of the bottleneck machines. The proposed approach provides a basis for studying the behaviour of the unplanned stops in bottlenecks using unsupervised machine-learning based k-means clustering technique. The usability and effectiveness of the constructed data-driven approach is demonstrated on a real-world production system. Also, within the domain of data-driven approaches for maintenance decision-making, this study has highlighted the necessity of maintenance practitioners' inputs, especially in steps such as data cleaning, data pre-processing and feature engineering. Whenever the practitioners wanted to diagnose the throughput bottlenecks, each of the proposed data-driven steps needs to be executed and when executing each of these steps, practitioners' inputs are necessary. The constructed data-driven approach helps practitioners to plan specific maintenance actions to improve the availability of the bottlenecks and hence throughput from the system.

CRediT authorship contribution statement

Mukund Subramaniyan: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Data curation. **Anders Skoogh:** Conceptualization, Methodology, Writing - review & editing, Supervision, Project administration, Resources, Funding acquisition. **Azam Sheikh Muhammad:** Methodology, Validation, Writing - review & editing. **Jon Bokrantz:** Conceptualization, Methodology, Writing - review & editing, Visualization. **Björn Johansson:** Methodology, Writing - review & editing, Resources. **Christoph Roser:** Conceptualization, Methodology, Writing - review & editing.

Acknowledgments

References

Angiulli, F., & Fassetti, F. (2014). Exploiting domain knowledge to detect outliers. *Data Mining Knowledge Discovery, 28*, 519–568. https://doi.org/10.1007/s10618-013-0310-5

Betterton, C. E., & Silver, S. J. (2012). Detecting bottlenecks in serial production lines – a focus on interdeparture time variance. *International Journal of Production Research, 50*(15), 4158–4174. https://doi.org/10.1080/00207543.2011.596847

Blaschko, M. B., & Lampert, C. H. (2008). Correlational Spectral Clustering. In IEEE Conference on Computer Vision and Pattern Recognition (pp. 1–8).

Bokrantz, J., Skoogh, A., La, D., Hanna, A., & Perera, T. (2017). Data quality problems in discrete event simulation of manufacturing operations. *Simulation: Transactions of the Society for Modeling and Simulation International, 94*(11), 1009–1025. https://doi.org/10.1177/0037549717742954

Boutsidis, C., Zouzias, A., Mahoney, M. W., & Drineas, P. (2015). Randomized dimensionality reduction for κ-means clustering. *IEEE Transactions on Information Theory, 61*(2), 1045–1062. https://doi.org/10.1109/TIT.2014.2375327

Broeksema, B., Telea, A. C., & Baudel, T. (2012). Visual Analysis of Multidimensional Categorical Datasets. In S. Bruckner, S. Miksch, & H. Pfister (Eds.), Eurographic Conference on Visualization 2012 (Vol. 31).

Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers and Industrial Engineering.* https://doi.org/10.1016/j.cie.2019.106024

Cheng, C. Y. (2018). A novel approach of information visualization for machine operation states in industrial 4.0. *Computers and Industrial Engineering, 125*(May), 563–573. https://doi.org/10.1016/j.cie.2018.05.024

Dhalmahapatra, K., Shingade, R., Mahajan, H., Verma, A., & Maiti, J. (2019). Decision support system for safety improvement: An approach using multiple correspondence analysis, t-SNE algorithm and K-means clustering. *Computers and Industrial Engineering, 128*(June 2018), 277–289. https://doi.org/10.1016/j.cie.2018.12.044

Gil, Y., Honaker, J., Orazio, V. D., Garijo, D., & Jahanshad, N. (2019). Towards Human-Guided Machine Learning. In Proceedings of the 24th International Conference on Intelligent User Interfaces (pp. 614–624). ACM.

Goldrat, E., & Cox, J. (1990). *The Goal: A Process of Ongoing Improvement (Third Revi).* Great Barrington, MA: North River Press.

Gopalakrishnan, M., Skoogh, A., & Laroque, C. (2013). Simulation based planning of maintenance activities in the automotive industry. In R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, & M. Kuhl (Eds.), Proceedings of the 2013 Winter Simulation Conference (pp. 342–353).

Gopalakrishnan, M., Skoogh, A., Salonen, A., & Asp, M. (2019). Machine criticality assessment for productivity improvement Smart maintenance decision support. *International Journal of Productivity and Performance Management Machine.* https://doi.org/10.1108/IJPPM-03-2018-0091

Gu, X., Jin, X., & Ni, J. (2015). Prediction of Passive Maintenance Opportunity Windows on Bottleneck Machines in Complex Manufacturing Systems. *Journal of Manufacturing Science and Enginnering, 137*(June 2015), 1–9. https://doi.org/10.1115/1.4029906

Harding, J. A., Shahbaz, M., Srinivas, & Kusiak, A. (2006). Data mining in manufacturing: A review. *Journal of Manufacturing Science and Engineering, 128*(4), 969. https://doi.org/10.1115/1.2194554

Holm, M. (2018). The future shop-floor operators, demands, requirements and interpretations. *Journal of Manufacturing Systems, 47*, 35–42. https://doi.org/10.1016/j.jmsy.2018.03.004

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245).

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies, 1*(6), 90–95.

Labib, A. (2014). A decision analysis model for maintenance policy selection using a CMMS. *Journal of Quality in Maintenance Engineering, 10*(3), 191–202. https://doi.org/10.1108/13552510410553244

Lee, J., Lapira, E., Bagheri, B., & Kao, H. An. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing Letters, 1*(1), 38–41. https://doi.org/10.1016/j.mfglet.2013.09.005

Li, L. (2018). A systematic-theoretic analysis of data-driven throughput bottleneck detection of production systems. *Journal of Manufacturing Systems, 47*, 43–52. https://doi.org/10.1016/j.jmsy.2018.03.001

Li, L., Ambani, S., & Ni, J. (2009). Plant-level maintenance decision support system for throughput improvement. *International Journal of Production Research, 47*(24), 7047–7061. https://doi.org/10.1080/00207540802375705

Li, L., Chang, Q., & Ni, J. (2009). Data driven bottleneck detection of manufacturing systems. *International Journal of Production Research, 47*(18), 5019–5036. https://doi.org/10.1080/00207540701881860

Li, L., Chang, Q., Ni, J., & Biller, S. (2009). Real time production improvement through bottleneck control. *International Journal of Production Research, 47*(21), 6145–6158. https://doi.org/10.1080/00207540802244240

Li, L., & Ni, J. (2009). Short-term decision support system for maintenance task prioritization. *International Journal of Production Economics, 121*(1), 195–202. https://doi.org/10.1016/j.ijpe.2009.05.006

Ni, J., & Jin, X. (2012). Decision support systems for effective maintenance operations. *CIRP Annals - Manufacturing Technology, 61*(1), 411–414. https://doi.org/10.1016/j.cirp.2012.03.065

Roser, C., Nakano, M., & Tanaka, M. (2001). A Practical Bottleneck Detection Method. In B. Peters, J. Smith, D. Medeiros, & M. Rohrer (Eds.), *Proceedings of the 2001 Winter Simulation Conference* (pp. 949–953). Arlington, VA, USA: IEEE. https://doi.org/10.1109/WSC.2001.977398.

Roser, C., Nakano, M., & Tanaka, M. (2002). Shifting bottleneck detection. In E. Yucesan, C.-H. Chen, J. Snowdon, & J. Charnes (Eds.), Proceedings of the 2002 Winter Simulation Conference (Vol. 2). Doi: 10.1109/WSC.2002.1166360.

Roser, C., Nakano, M., & Tanaka, M. (2003). Comparison of bottleneck detection methods for AGV systems. In S. Chick, S.P.J.D. Ferrin, & M.D.J. (Eds.), Proceedings of the 2003 Winter Simulation Conference (pp. 556–564).

Segura, Á., Diez, H. V., Barandiaran, I., Arbelaiz, A., Álvarez, H., Simões, B., … Ugarte, R. (2018). Visual computing technologies to support the Operator 4. 0. *Computers & Industrial Engineering*, 1–8. https://doi.org/10.1016/j.cie.2018.11.060

Sengupta, S., Das, K., & VanTil, R. P. (2008). A new method for bottleneck detection. In S. Mason, R. Hill, L. Monch, O. Rose, T. Jefferson, & J. Fowler (Eds.), *Proceedings of the 2008 Winter Simulation Conference* (pp. 1259–1267).

Sharp, M., Ak, R., & Hedberg, T. (2018). A survey of the advancing use and development of machine learning in smart manufacturing. *Journal of Manufacturing Systems, 48* (Part C), 170–179. https://doi.org/10.1016/j.jmsy.2018.02.004

Subramaniyan, M., Skoogh, A., Salomonsson, H., Bangalore, P., Gopalakrishnan, M., & Sheikh Muhammad, A. (2018). Data-driven algorithm for throughput bottleneck analysis of production systems. *Production & Manufacturing Research, 6*(1), 225–246. https://doi.org/10.1080/21693277.2018.1496491

Swanson, L. (2001). Linking maintenance strategies to performance. *International Journal of Production Economics, 70*, 237–244. https://doi.org/10.1016/S0925-5273(00)00067-0

Tang, H. (2019). A new method of bottleneck analysis for manufacturing systems. *Manufacturing Letters*. https://doi.org/10.1016/j.mfglet.2019.01.003

Wuest, T., Weimer, D., Irgens, C., & Thoben, K. (2016). Machine learning in manufacturing : Advantages, challenges, and applications. *Production & Manufacturing Research, 4*(1), 1–23. https://doi.org/10.1080/21693277.2016.1192517

Yong-Cai, Wang, & Qian-Chuan Zhao, D.-Z. Z. (2005). Bottlenecks in production networks: An overview. *Journal of Systems Science and Systems Engineering, 14*(3), 347–363.

Yu, C., & Matta, A. (2016). Data-driven bottleneck detection in manufacturing systems: A statistical approach. *International Journal of Production Research, 54*(21), 6317–6322. https://doi.org/10.1080/00207543.2015.1126681

Zenisek, J., Holzinger, F., & Affenzeller, M. (2019). Machine learning based concept drift detection for predictive maintenance. *Computers and Industrial Engineering*. https://doi.org/10.1016/j.cie.2019.106031

Zhu, M., & Ghodsi, A. (2006). Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Computational Statistics & Data Analysis, 51*, 918–930. https://doi.org/10.1016/j.csda.2005.09.010