Predicting maintenance work hours in maintenance planning

Khalid, Waqas; Albrechtsen, Simon Holst; Sigsgaard, Kristoffer Vandrup; Mortensen, Niels Henrik; Hansen, Kasper Barslund; Soleymani, Iman

Published in:
Journal of Quality in Maintenance Engineering

Link to article, DOI:
10.1108/JQME-06-2019-0058

Publication date:
2021

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Predicting maintenance work hours in maintenance planning
Waqas Khalid, Simon Holst Albrechtsen, Kristoffer Vandrup Sigsgaard,
Niels Henrik Mortensen, Kasper Barslund Hansen
and Iman Soleymani
Technical University of Denmark, Lyngby, Denmark

Abstract

Purpose
Current industry practices illustrate there is no standard method to estimate the number of hours worked on maintenance activities; instead, industry experts use experience to guess maintenance work hours. There is also a gap in the research literature on maintenance work hour estimation. This paper investigates the use of machine-learning algorithms to predict maintenance work hours and proposes a method that utilizes historical preventive maintenance order data to predict maintenance work hours.

Methodology
The paper uses the design research methodology utilizing a case study to validate the proposed method.

Findings
The case study analysis confirms that the proposed method is applicable and has the potential to significantly improve work hour prediction accuracy, especially for medium- and long-term work orders. Moreover, the study finds that this method is more accurate and more efficient than conducting estimations based on experience.

Practical implications
The study has major implications for industrial applications. Maintenance-intensive industries such as oil and gas and chemical industries spend a huge portion of their operational expenditures (OPEX) on maintenance. This research will enable them to accurately predict work hour requirements that will help them to avoid unwanted downtime and costs and improve production planning and scheduling.

Originality
The proposed method provides new insights into maintenance theory and possesses a huge potential to improve the current maintenance planning practices in the industry.

Keywords: maintenance, machine learning, work hour, planning, estimation, prediction, maintenance management

1. Introduction
Large-scale platforms such as oil and gas have complex maintenance management systems. Moreover, maintenance work is a major contributor to operational expenditure (OPEX)-related costs in large-scale platforms. The downturn in oil prices is pushing operators to reduce OPEX; however, research shows a compound annual OPEX/barrel(bbl) cost escalation rate (CACER) of 2.85% for major oil companies, 5.01% for national oil companies (NOCs) and 5.54% for independent companies from 2007 to 2014 (Ernst & Young, 2015). McKinsey & Company (2014) argue that if the historical trend of increased lifting costs continues, then it will have serious implications for the profitability of the basins because the average cost will exceed USD 100 per barrel within a few years in the United Kingdom Continental Shelf (UKCS). Moreover, Brun et al. (2017) estimate a USD 200 billion performance gap, as most of the offshore platforms operate at approximately 77% of its production potential.
Many operators have realised the need to reduce operational costs. For example, in contrast to the increase of $60 billion in operational costs from 2008 to 2014, the global production cost fell by 29% (USD 40 billion) since 2014 (Cole and Harris-Deans, 2017). From over 200 initiatives, the main reduction (40–50%) in OPEX mainly came from eliminating planned activities in the North Sea (Cole and Harris-Deans, 2017). Here, a question arises: Are these cost reductions sustainable? In a highly competitive and fluctuating market, operators need to find novel ways to reduce OPEX-related costs. According to McKinsey & Company (2014), maintenance- and integrity-related costs went up by 57% between 2009 and 2013 in the North Sea, showing a dire need to reduce non-value-adding activities. In contrast, the effective planning of maintenance work is expected to improve technician wrench time by 14%, thereby reducing both the mean time to repair (MTTR) and the associated costs (McKinsey & Company, 2014).

Accurate work measurement during maintenance planning is crucial. Duffuaa and Ben-Daya (2009, p. 107) define work measurement as ‘a technique to develop time standards for jobs while considering ratings of workers and allowances for personal needs, fatigues and other contingencies’ and lists it as a necessary tool for effective maintenance management. In the job design phase, it is compulsory to include a time estimate (Niebel, 1994), as it is used in forecasting the workload against the available capacity (S. Duffuaa and Raouf, 2015). Erroneous order size estimates may lead to prolonged downtime, increased technician wrench time and increased non-value-adding activities (S.O. Duffuaa and Ben-Daya, 2009; S. Duffuaa and Raouf, 2015). Hence, accurate time estimates are critical for better planning, scheduling and order execution. The methods proposed in the past have not been productive or efficient in complex maintenance environments and can be costly. This includes time study, sampling, predetermined motion time series (PMTS), standard data and estimating (S.O. Duffuaa and Ben-Daya, 2009). Therefore, this paper proposes a new method to estimate work hours in highly complex operational maintenance environments.

The paper is structured as follows: firstly, section 2 presents the most current and state-of-the-art literature; then section 3 outlines paper’s specific contributions. Following this, section 4 gives the research methodology this paper aims to follow. Section 5 details the proposed method. Next, section 6 presents a case study that details the application of the proposed method and conducts a deep analysis of findings. After this, section 7 includes a general discussion based on the findings from the case study and the implications of the study on the theory and industry are discussed in section 8. Finally, section 9 provides a conclusion with limitations and future works.

2. Literature review

This section presents a state-of-the-art literature review on the subject. Due to the non-repetitive nature of corrective maintenance (CM) work, accurate work estimation is not possible. However, Nyman and Levitt (2006) argue that repetitive/preventive maintenance (PM) work can be estimated with a degree of certainty, but with ±15% accuracy, as the nature and scope of the job changes each time it is executed. The literature describes several methods available for work measurement: study, sampling, predetermined motion time system (PMTS), standard data and estimating (S.O. Duffuaa and Ben-Daya, 2009).

Time study involves timing the workers as they perform their jobs and standardising the observed time (S. Duffuaa and Raouf, 2015; Festijo, n.d.). This method is not very useful when maintenance activities are carried out intensively on a daily basis in a complex management system that may require extra, potentially costly, resources.
Work sampling is used to calculate the percentage occurrences of certain activities (S. Duffuaa and Raouf, 2015). These can then be used for non-repetitive work (Buchholz et al., 1996). However, this technique is mainly used for calculating tool time and can be expensive in highly intensive maintenance environments or industries where the work is distributed across wider areas.

PMTS divides a job into already developed standard motions and then adds them together to calculate the total time (Festijo, n.d.). This technique is widely used to estimate manual operations (Alkan et al., 2016; Genaidy et al., 1989). The limitation of this method is its need for highly expert knowledge, which can be very time-consuming in highly intensive and sporadic maintenance environments.

Standard data is another technique in which standards are developed for repetitive work elements in the company and then used to estimate future similar works (S. Duffuaa and Raouf, 2015). However, recording these standards is a cumbersome task.

Estimation is another technique in which experts guess future work based on their past experience (S. Duffuaa and Raouf, 2015). This is an inexpensive technique, but the probability of erroneous estimates is quite high.

These methods require either physical observations of jobs, which is time-consuming and costly, or estimations either based on the experience or standard data of similar jobs that may not be accurate (Baines, 1995; S. Duffuaa and Raouf, 2015). Despite being in use for decades, the efficiency and accuracy of estimations of work order execution time are not optimal in complex systems. Moreover, these methods are decades old, and no research has been done to improve work order time estimation. Hence, there is a need for a better method for work estimation.

Handscomb et al. (2016) argue that technological advances such as advanced analytics and machine learning are disrupting the old ways of working and could serve as a means to increase productivity in the oil and gas sector. A few of these techniques, such as neural network and linear regression, are used in work hours estimation in production (Liu and Jiang, 2005); however, no such techniques exist in maintenance work requirement estimations.

A detailed literature review has been conducted, but work estimation methods than those already explained above do not seem to exist. Other estimation methods comprise cost estimation, such as construction cost, manufacturing cost, life cycle cost, maintenance cost, etc. A few other methods include work estimation for software and web applications and estimations for mineral resource. However, these methods do not incorporate work hour estimation for maintenance jobs. Table 1 contains the summary of the literature review. The search engine DTUfindit was used to conduct the literature review. The terms searched included but were not limited to time estimation, job estimation, work order estimation, resource estimation, estimation and maintenance estimation. This paper proposes a new method that uses machine-learning techniques in complex work management systems (WMS) to accurately estimate the work hour requirements for PM jobs.
<table>
<thead>
<tr>
<th>Source #</th>
<th>Reference</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Staub-French et al., 2003)</td>
<td>A model that supports estimators when doing cost estimations on construction designs.</td>
</tr>
<tr>
<td>2</td>
<td>(Shehab and Abdalla, 2001)</td>
<td>A study of a knowledge-based system that aids inexperienced users to estimate the manufacturing costs of a conceptual product design.</td>
</tr>
<tr>
<td>3</td>
<td>(Heller et al., 2014)</td>
<td>A method for estimating the future life cycle costs of a conceptual design using life cycle assessment data.</td>
</tr>
<tr>
<td>5</td>
<td>(Cao et al., 2012)</td>
<td>Estimating army equipment maintenance hours using bootstrap simulation and an approximated analytical distribution.</td>
</tr>
<tr>
<td>6</td>
<td>(Baines, 1995)</td>
<td>An overall description of different work estimation methods, including time study, sampling technique, estimating technique, synthesis and PMTS.</td>
</tr>
<tr>
<td>7</td>
<td>(Buchholz et al., 1996)</td>
<td>A work sampling technique – posture, activity, tools and handling (PATH) – that is used to estimate the ergonomic hazards of construction work.</td>
</tr>
<tr>
<td>8</td>
<td>(S.O. Duffuaa and Ben-Daya, 2009)</td>
<td>Explanation of several work estimation techniques such as time study, sampling, predetermined motion time series (PMTS), standard data and estimating.</td>
</tr>
<tr>
<td>9</td>
<td>(S. Duffuaa and Raouf, 2015)</td>
<td>Describes work estimation techniques such as time study, sampling, PMTS, standard data and estimating.</td>
</tr>
<tr>
<td>10</td>
<td>(Festijo, n.d.)</td>
<td>Provides a description of and procedure for calculating work measurements such as time study, standard data, work sampling and PMTS; provides the advantages and disadvantages of several work measurement techniques.</td>
</tr>
<tr>
<td>11</td>
<td>(Nyman and Levitt, 2006)</td>
<td>Provides advantages and disadvantages of work measurement techniques such as construction trade estimate, gross estimate, average history, analytical estimates, predetermined motion time, time study and standard data; proposes analytical estimation as the appropriate method for work estimation with sufficient accuracy.</td>
</tr>
<tr>
<td>12</td>
<td>(Liu and Jiang, 2005)</td>
<td>Comparison of three different work hours estimation models: simple linear regression, multiple linear regression and artificial neural network.</td>
</tr>
<tr>
<td>13</td>
<td>(Raadnui, 2000)</td>
<td>Application of PTS – for example, reach, move, grasp, turn, position, etc. in preventive maintenance activities.</td>
</tr>
<tr>
<td>14</td>
<td>(Saito et al., 2018)</td>
<td>A method that uses project-based learning (PBL) based on the cost estimation, benchmarking and risk assessment (CoBra) model to estimate the work hours in software development.</td>
</tr>
<tr>
<td>15</td>
<td>(Ye et al., 2017)</td>
<td>Proposed method for work hour estimation in a flexible manufacturing environment; establishes and uses standard operation unit (SOU).</td>
</tr>
<tr>
<td>16</td>
<td>(Khan et al., 2016)</td>
<td>A locally weighted linear regression (LWLR) technique to estimate the completion time of a job in cloud computing.</td>
</tr>
</tbody>
</table>
2.1 PM order-management system

This section provides a brief description of a PM-based order management system. In complex maintenance management systems, there are hundreds of thousands of pieces of equipment to be maintained. Maintenance plans are made using a computerised maintenance management system (CMMS). Figure 1 represents the process of how these plans are created and monitored in complex maintenance systems.

All pieces of maintainable equipment are assigned a unique identifier, which is usually called a functional location (FL). These FLs are a part of several different systems, for example, processing systems, utility systems, safety systems, etc. The associated system and its criticality are identified based on the business protocols in use. Organizations have already defined standard system maintenance strategies and based on the criticality and function of the system, standard system maintenance strategies are selected for these equipment. Based on the selected system strategies, either a PM strategy is created, or an existing strategy is used if it contains the exact operations, material requirements and maintenance packages (weekly, monthly, yearly, etc.). If there are multiple FLs to be maintained at different dates, then separate maintenance plans are created for each asset. However, if there are several FLs that have the same due dates and the same frequencies for maintenance, then a single maintenance plan is created using a CMMS.

A maintenance plan is pre-planned and contains information such as frequencies, sequence of activities, time, labour requirements, material and tool requirements, etc. Once the maintenance plan is finalised, it is activated in the system. The system monitors the deadline for all FLs, and when the deadline is reached, it generates a PM work order for the corresponding FLs containing the pre-planned maintenance plan. The time estimated to carry out the required work in an order is automatically calculated based on the task list for each of the FLs in the work order. The task list is the list of operations that are pre-defined in the maintenance plan. After execution of the work order, the actual work hours spent on the order are recorded in the system and labelled ‘actual work’.
2.2 PM order-management system

This paper provides a specific contribution to the selection of the attributes to be considered in the methodology, especially in the offshore oil and gas sector. The research proposes that instead of looking at the FLs’ identifiers, the descriptive data of both work orders and the FLs should be considered simultaneously. Moreover, its use of AdaBoost to select an even better sub-set of attributes to improve the performance is of note. AdaBoost (adaptive boosting) is an algorithm that is used in feature selection or attributes selection to improve the performance of machine learning based prediction algorithms. It selects the most accurate and relative attributes to be used in the algorithm when a number of attributes are available and it is difficult to make selection.

3. Methodology

This paper follows the design science research that Hevner et al. (2004) highlight because the research outcome involves information technology (IT) and requires computer algorithms and human/computer interface.

According to Hevner et al. (2004), the scientific research design must produce a viable artefact, the research must address a business problem, the artefact must be evaluated and have a clear contribution, the research must be rigorous and follow a search process and the research must be effectively communicated to business- and technology-oriented audience.

In this case, the research produces an artefact in the form of a methodology that can be used for more accurate estimations of work hour requirements in a complex WMS, where accurate time estimation is a challenge and requires improvement. The proposed methodology is evaluated using a deep case study analysis. A case study is defined as ‘a contemporary phenomenon in its real-life context, especially when the boundaries between phenomenon and context are not clearly evident’ (Dumez, 2015, p. 44; Yin, 1981,
p. 59) and can be used to test a hypothesis (Flyvbjerg, 2006; Levy, 2008). The proposed model provides clear contributions to theory and practice and is described in detail in Section 7. Rigorous methods are used to construct and evaluate the methodology. The study has been designed as a research process, where the methodology aims to solve a specific problem utilising the available means to reach the desired end. Hence, this study follows the guidelines that Hevner et al. (2004) lay out as described in the paragraph above.

4. Proposed method
This section presents a detailed description and step-by-step guide of the proposed method. The framework is developed based on the literature review.

1. Select the equipment type for which the work hours need to be optimised
2. Locate historical PM orders for the same equipment type
3. List all the attributes based on the order and the FLs – the attributes can be of different types, for example, categorical, ordinal, numerical, etc.
4. Convert categorical and ordinal attributes (e.g. strings) using 1-out-of-k encoding into values of 0s and 1s. Suppose there are three different values for a categorical attribute: red, green and blue. To convert this into 1-out-of-k encoding, list a column for each value and assign a data point to one of the columns as per the example shown below. One-out-of-k encoding also known as one-hot encoding and is used in machine learning due to the nature of categorical or ordinal variables (also known as nominal variables). These need to be converted into numerical values in order to be used by machine-learning algorithms (Beck and Woolf, 2000; El Affendi and Al Rajhi, 2018; Feurer et al., 2015; Rodríguez et al., 2018). Encoding the values allows the uniqueness of the individual values to be fed into the regression model.

<table>
<thead>
<tr>
<th>Attribute Value</th>
<th>Red</th>
<th>Green</th>
<th>Blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Red</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Blue</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Green</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

5. Present the data in matrix form along with a target vector representing the actual work used to execute the work order as per the example shown below.

\[ X = \begin{bmatrix} 0 & 2 & \cdots & 1 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 3 & \cdots & 0 \end{bmatrix}, \quad y = \begin{bmatrix} 1.5 \\ 26 \\ \vdots \\ 34.7 \end{bmatrix} \]

The \( X \) matrix represents the orders and their content. The values of the matrix \( X \) represent attribute values. The \( y \)-vector represents the actual hours spent on the orders that are being estimated. Each order may contain multiple functional locations, so the attribute values are summed to represent an order row. An example of constructing an order is given below:

\[
\begin{align*}
\text{Order data} & = 1 \ 0 \ 2 \ - \ - \ - \\
FL_1 \text{data} & = - \ - \ - \ 0 \ 1 \ 1 \ 0 \\
FL_2 \text{data} & = - \ - \ - \ 1 \ 0 \ 1 \ 2 \\
\hline
& = 1 \ 0 \ 2 \ 1 \ 1 \ 2 \ 2
\end{align*}
\]

Thus, a row represents the order attributes and the distribution of the attributes from the set of functional locations of that order. The matrix, or tabular representation, is becoming more widely
used as data manipulation with matrices while using a programming software like Matlab, R or Python becomes easier (McKinney, 2010; Paluszek and Thomas, 2017; Walker et al., 2012).

6. Remove noise and outliers from the data (if required). Outliers are the points in the data that deviate significantly from the normal data points and can drastically affect the results. Many methods and techniques exist to handle the noise and outliers in data sets (Aguinis et al., 2013; Barnett, 1983; Borgelt et al., 2015; Xiaohui Liu et al., 2002).

7. Transform the data so that all attributes have normalised values. The attributes may have different numerical values in the matrix $X$ that may force some algorithms to give more importance to a few of the attributes. This can be done by min–max normalisation, Z-score normalisation or any other suitable method (Himani Goyal, 2015).

8. Use AdaBoost (Freund and Schapire, 1997) for attribute selection to boost the learner’s performance by providing the sub-set of the attributes that best predict the outcome. AdaBoost was introduced by Freund and Schapire in 1995. The algorithm takes in a training set of observations and labels and repeatedly adds given weak and base learning algorithms for a number of iterations. Weights are assigned to each training example; then, with each iteration, the weights of the incorrectly classified examples are increased. This forces the weak learner to focus on the harder examples. The learner then finds a weak hypothesis that is appropriate for the distribution. After receiving the weak hypothesis, the algorithm choses a parameter that measures the importance that is assigned to the weak hypothesis. The final hypothesis is then generated by the weighted majority vote of the weak hypothesis with the final weights (Schapire, 1999).

9. Select the machine-learning algorithms for training and testing the data.

10. Divide the data into train and test sets for the algorithms. The train data will be all of the historical work orders, and the test data will be the new work orders upon which the work hours are to be estimated. This paper uses k-fold cross-validation as a proof of concept, as the research is carried out only on the past historical data. K-fold cross-validation shuffles the data to ensure a diverse set of data points to avoid biasing the performance; further, it tests all the data points against the fitted model and then averages the score for all the tests (Polat and Güneş, 2007).

11. Run the model alongside principal component analysis (PCA) (Hotelling, 1933) and measure the performance of each algorithm against the principal components; then, select the best-performing number of principal components $n$.

12. Evaluate the results by using performance indicators such as variance, absolute error, mean square error, etc. and select the best-performing algorithm for work hour prediction.

Table 2 summarises the proposed framework detailed above. As the study produces a viable artefact in the form of a methodology as given above, it follows the first guideline of the design research methodology. Moreover, by proposing this method, the study attempts to solve a business problem with technology and therefore fulfils the second guideline of the design research methodology.

**Table 2. Summary of the proposed framework**

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Select equipment type</td>
<td>(Beck &amp; Woolf, 2000; El Affendi &amp; Al Rajhi, 2018; Feurer et al., 2015; Rodríguez et al., 2018)</td>
</tr>
<tr>
<td>2</td>
<td>Collect historical data</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>List all the attributes</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Convert categorical and ordinal attributes using 1-out-of-k encoding</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Present the data into a matrix form along with the target vector</td>
<td>(McKinney, 2010; Paluszek &amp; Thomas, 2017; Walker et al., 2012).</td>
</tr>
</tbody>
</table>
Remove the noise or outliers from the data (Aguinis et al., 2013; Barnett, 1983; Borgelt et al., 2015; Xiaohui Liu et al., 2002)

Normalize the values between 0 and 1 (Himani Goyal, 2015)

Use ‘AdaBoost’ algorithm for feature selection (Freund & Schapire, 1997; Schapire, 1999)

Select a number of machine learning algorithms for training and testing the data

Divide the data into test and train sub-data (Polat & Güneş, 2007)

Run the model with PCA and select the Principle Components with the highest performance (Hotelling, 1933)

Evaluate the results with different performance indicators and select the best performing algorithm

5. Case study

This section presents a detailed case study. For confidentiality reasons, the name of the case company is not mentioned and is instead termed ‘Oil Company’ throughout this paper. The Oil Company is an oil and gas-producing company that has several offshore oil fields. The study scope has been limited to only PM orders and the equipment types emergency shut-down valve (ESDV) and blow-down valve (BDV). ESDVs and BDVs can have several different functional locations based on their type and physical position across different platforms in the offshore oil fields. The software program Python has been used to handle and process the data in this study. The case study rigorously evaluates the utility, quality and efficacy of the proposed method, also following the third guideline of the design research methodology.

5.1 Data collection and structure

Quantitative data based on work orders and FLs from 2011 to early 2017 are used in this paper. All the data were extracted from the SAP (systems, applications and products) plant maintenance module that is implemented in the Oil Company. The structure of the data is shown in Figure 2. The data are structured into two different data sources: work order data and FL data. Moreover, n different FLs can be linked with only one work order. In total, 804 work orders are used in the analysis. Figure 2 also shows the types of each data set. The description of each attribute is given in Table 3.

Actual work and planned work are not used as attributes in the analysis, but they are used to calculate the accuracy of the predicted work hours estimates and to train and test the algorithms. Moreover, all the ordinal and categorical attributes are converted into numerical ranges of natural number values using 1-out-of-k encoding as described in Step 4, Section 5 to be able to feed into the model. In total, after using 1-out-of-k encoding, 54 attributes are used in the analysis. All the data are presented in a matrix format as mentioned in Step 5, Section 5, where columns represent attributes and rows represent the values of attributes against each work order. Moreover, all the labels are presented in a vector format showing the actual work hours against each work order as per Step 5, Section 5.
Table 3. Description of data variables

<table>
<thead>
<tr>
<th>Sr. #</th>
<th>Data Source</th>
<th>Variable Name</th>
<th>Type</th>
<th>Nature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Work Order</td>
<td>Cycle Months</td>
<td>Integer</td>
<td>Categorical</td>
<td>No. of months after which the maintenance is repeated.</td>
</tr>
<tr>
<td>2</td>
<td>Actual Work</td>
<td>Actual Work</td>
<td>Float</td>
<td>Numerical</td>
<td>Actual work in hours used to execute the order.</td>
</tr>
<tr>
<td>3</td>
<td>Planned Work</td>
<td>Planned Work</td>
<td>Float</td>
<td>Numerical</td>
<td>Work in hours that is planned to execute the order.</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Riser ESDV</td>
<td>Boolean</td>
<td>Ordinal</td>
<td>A Boolean parameter, If ESDV has a riser then the value is ‘Yes’ lest it is ‘No’.</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Valve Type</td>
<td>String</td>
<td>Categorical</td>
<td>Type of the valve, for example, ball, globe, butterfly etc.</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Size</td>
<td>Float</td>
<td>Categorical</td>
<td>The size of the valve in inches. For example, 1, 2, 3 inches etc.</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>Asset</td>
<td>String</td>
<td>Categorical</td>
<td>Asset name on which the valve is located</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>System Media</td>
<td>String</td>
<td>Categorical</td>
<td>The media type that is flowing through the valve. For example, gas, oil, water etc.</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>Accessibility</td>
<td>String</td>
<td>Categorical</td>
<td>The type of the offshore oil platform accessibility on which the valve is located. For example, unmanned, accommodation, connected etc.</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>Installation Year</td>
<td>Integer</td>
<td>Ordinal</td>
<td>The year when the valve was installed</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>Platform Type</td>
<td>String</td>
<td>Categorical</td>
<td>The type of the oil platform. For example, wellhead, flare, process, utility etc.</td>
</tr>
</tbody>
</table>

Figure 2. Data structure

5.2 Findings and analysis

Figure 3 shows the actual work and the planned work of all the PM orders executed between 2011 and early 2017 on the ESDVs and BDVs in the Oil Company. Figure 3 represents the output of current estimation practices at the Oil Company. The points on the line are the work orders that have equal planned and actual work hours. The following evaluation metrics were used to calculate the estimation accuracy of the Oil Company.
Figure 3. Current estimates performance of the Oil Company

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>21%</td>
<td>Regression fit, meaning that the regression line has a bad fit and is not a good estimator.</td>
</tr>
<tr>
<td>$</td>
<td>S</td>
<td>$</td>
</tr>
<tr>
<td>$S^2$</td>
<td>147.89</td>
<td>Mean square error, meaning that there are multiple orders with estimates that are far from the planned work.</td>
</tr>
</tbody>
</table>

As per the evaluation metrics values, current estimation practices are not optimal in the Oil Company, and there is a room for improvement.

As per Step 6, the data are cleaned by removing the outliers. In this case, all the data points with invalid information (actual work hours equal to 0 or to a negative value) are considered as the outliers. As per step 7, the data are transformed between the values 0 and 1 to avoid sensitivity of the algorithms. As mentioned in Section 5.1, there were 54 attributes in the data set. Hence, as per Step 8, Section 5, AdaBoost is applied for feature selection to yield a sub-set of attributes that improve the learner’s performance. The result is
shown in Figure 4. The attributes are reduced to only 10 attributes with a high importance value, as listed below:

1. Cycle_months (12): (integer, categorical)
2. Riser_ESDV: (Boolean, ordinal)
3. Type (ball): (string, categorical)
4. Size_in_inches (2): (float, categorical)
5. Size_in_inches (3): (float, categorical)
6. System_media (drain): (string, categorical)
7. System_media (gas): (string, categorical)
8. System_media (multi-phase): (string, categorical)
9. System_media (stabilised oil): (string, categorical)
10. Installation_year: (integer, ordinal)

![Figure 4. Results of the AdaBoost analysis](image)

As per Step 9, the following machine-learning algorithms were selected to train and test the data:

1. AdaBoost (ensemble algorithm)
2. Bagging (ensemble algorithm)
3. BayesianRidge (linear algorithm)
4. ExtraTrees (ensemble algorithm)
5. GradientBoosting (ensemble algorithm)
6. Huber (linear algorithm)
7. Linear (linear algorithm)
8. RandomForest (ensemble algorithm)
9. LinearSVR (SVR (support vector regression) algorithm)
These algorithms are linear, ensemble and support-vector machine (SVM); they provide a good variety of models for fitting the data and thus improve the chances of finding the optimum model.

The data are divided into training and test data, as mentioned in Step 10, Section 5; all the algorithms selected above are run on the data set using k-fold cross-validation. The value of 3 is used for k in this paper. As per Step 11, for \( n = 10 \) (principal components), a score of 76\% is observed for the RandomForest algorithm. The results of all the algorithms are shown in Figure 5.

![Figure 5. Results of all the machine-learning algorithms](image)

As per Step 12, the RandomForest algorithm has the best performance; hence it is selected as the model to be used to predict work hours. Figure 6 represents the actual versus the predicted hours using the RandomForest model; the values of the performance parameters are shown below in Figure 6. It can be seen that the performance of the predicted work hours has significantly improved.
Table 1: Model performance

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$ Regression fit</td>
<td>76.25%</td>
<td>meaning that the regression fit is significantly improved (263% increase)</td>
</tr>
<tr>
<td>$</td>
<td>S</td>
<td>$ Mean absolute error</td>
</tr>
<tr>
<td>$S^2$ Mean square error</td>
<td>44.64</td>
<td>meaning that there are now less orders that are far from the predicted value (70% decrease)</td>
</tr>
</tbody>
</table>

6. Discussion

6.1 General discussion

The proposed framework provides the improved work hour predictions for maintenance orders. The machine-learning algorithms try to fit the data points so that the error between the predicted and actual work hours is minimal. The lower error leads to a higher regression fit value ($R^2$). The RandomForest algorithm yielded a 263% increase in the regression fit value, showing a more accurate work hour prediction capability. Moreover, the absolute error $|S|$ and the mean square error $S^2$ also improved. However, the decrease in the mean square error (69%) is greater than the absolute error (16%). This is because the long-term work orders are further away from the predicted values than the short-term work orders. After optimisation, the model forces the regression line to fit the data points such that the long-term work orders are forced to come closer to the regression line as compared to the short-term work orders. This is because in order to maximise the regression fit ($R^2$), the model will try to minimise the numerator ($Actual - Predicted)^2$ in the formula; this has more impact on the long-term orders.

As mentioned in Section 1, the work orders with $\pm 15\%$ tolerance in the predicted and actual work hour differences are considered to be well-estimated work orders. Figure 7 shows this comparison of prediction accuracy with $\pm 15\%$ tolerance between the Oil Company’s current model and the RandomForest model.
Short term is defined as work orders with actual work hours ≤ 20 hours, medium term is defined as the work orders with actual hours between 20–60 and long term is defined as the work orders with ≥ 60 actual work hours.

As can be seen, there is a significant improvement in prediction accuracy, especially for medium-term orders (83% improvement) and long-term orders (600% improvement). However, there is a 60% decrease for short-term work orders; this is due to the previously described effect of the regression fit. Also, the tolerance values for short-term orders are smaller than those for long-term orders; hence, fitting the regression line with the data points (Figure 6) leads to less short-term work orders within the ±15% of the predicted values.

This paper believes that medium- and long-term work orders are most critical from the estimation point of view. They will have a stronger impact on work order planning and the associated resource utilisation, because these work orders tend to use more resources for a longer period of time and by accurate prediction of required work can free up a lot of resources that can be used for other works. With the proposed method, the prediction accuracy of the medium- and long-term orders can be increased significantly. Therefore, the proposed method is a good tool for maintenance work hours estimation.

### 6.2 Attributes, feature selection and PCA

It is worth mentioning that the research was carried out in several iterations. In the first iteration, different attributes were selected than those used in the final iteration. In the first iteration, only the functional location identifiers were considered as attributes, as shown in the example below:

```
  XX . XX . X . 0XX . XXXX . XXX . XXXXX
  Country  Asset  Section  System  SAP Attribute  Valve Type  Tag number
```
A total of 270 attributes were used in the first iteration. This iteration provided the maximum $R^2$-value of 62.1% at $k = 85$ principle component. In the second iteration, all 54 attributes used in the final iteration were considered but without AdaBoost. In this iteration, actual descriptions of the work orders and FLs were considered as attributes. This iteration provided the maximum $R^2$-value of 71.34% at the 46th principle component ($k = 46$). Lastly, in the final iteration, AdaBoost was used to select the sub-set of the features. This increased the performance of the algorithm to 76.25% of the $R^2$-value. Hence, the proposed framework is based on the final improved iteration of the research and provides the best performance.

The comprehensive discussion and case study presented above show research rigour in both the construction and evaluation of the proposed method. Moreover, this method uses existing technologies and satisfies the rules and laws in the problem environment. Lastly, the research has striven to communicate to both technology and management-oriented audiences. Hence, the study also follows the fifth, sixth and seventh guidelines of the design research methodology.

7. Implications for theory and practice

The proposed methodology presented in Section 5 contributes to theory because it proposes a novel approach to estimating work hours. As compared to the already existing methods in the literature, such as time study, sampling, PMTS, standard data and estimating (S. Duffuaa and Raouf, 2015), the proposed methodology removes the limitations of physical observations of jobs that can be costly and the study of the similar jobs to predict work hours. Moreover, the results in Sections 6 and 7 show that the proposed methodology is more accurate and efficient (263% increase in regression fit, 16.2% decrease in mean absolute error, 70% decrease in mean square error and 83–600% increase in prediction accuracy) than the current methodology the case company is using (i.e. estimation based on experience). Moreover, a brief data-enveloped analysis (DEA) has been conducted between two decision-making units (DMUs) – the current estimations and the proposed method – where mean absolute error (MAE) and mean prediction accuracy are used as the two outputs. Figure 8 shows that the DMU that lies below and to the right is more efficient, which in this case, is the proposed method.
Figure 8. Data enveloped analysis (DEA) of the two methods

The implications for practice and the industry are very broad. Maintenance-intensive industries such as oil and gas and chemical industries prioritise correct estimation of the work required to carry out the maintenance jobs, as it is directly linked with OPEX cost and production efficiency. This study argues that the proposed framework will help these industries to make correct work hour estimates. These correct estimations will lead to better planning of regarding orders and resources. Moreover, order planning will require less time and use less resources. Additionally, the method enables correct downtime estimation for production, which will also lead to better production planning. Lastly, due to better estimation, there will be fewer unwanted interruptions in production due to prolonged work.

This section clearly outlines the contributions of this research and hence confirms that the study fulfils the fourth guideline of the design research methodology.

8. Conclusion, limitations and future work

This paper investigates the research gap in work hour estimation for preventive maintenance work orders, as maintenance jobs contribute a great deal to they OPEX-related costs of any company. The paper proposes a methodology that can be used for better estimations of work hours for preventive maintenance jobs, as these are repetitive in nature. The method takes into account historical work orders and functional locations combined with a number of equipment- and maintenance-related variables; further, it uses machine-learning algorithms to predict work hours. The case study proves that there are significant improvements in the estimations compared to the company’s original estimates. The proposed method contributes to both theory and practice by bringing novel insights to work hour estimation and has huge potential to improve current industry practices. This will lead to better resource utilisation and production planning.
The paper uses real-time data from a company; however, the data set contains some anomalies. For example, some orders had zero or negative work hour estimation. These work orders were removed from the data set because they were considered as noise or outliers. Moreover, more than 60% of the orders had the exact same estimated work hours as the actual work hours, showing bias in the data. This could be due to the maintenance personnel who entered the data into the SAP system. This study proposes that if the bias can be removed from the data, then the improvement in the results will be much higher.

As this research is based only on one case study, it would be useful to apply its method to other case studies. This would enable better verification and evaluation of the proposed methodology’s performance. However, as the case study is representative of very large and complex oil and gas companies, it can be said with confidence that the proposed method is globally applicable. Nonetheless, other researchers are encouraged to test the applicability of the proposed method by conducting more case studies.

9. Acknowledgements
The authors acknowledge the funding received from the Centre for Oil and Gas – DTU/Danish Hydrocarbon Research and Technology Centre (DHRTC). Moreover, the authors acknowledge the efforts of Julia Krogh Agergaard in reviewing and editing the article.

References
Cao, T., Francis, B., and Brealey, N. (2012), "An approach to estimate the probability in meeting army maintenance personnel requirements", Engineering Asset Management and Infrastructure Sustainability, available at: https://doi.org/10.1007/978-0-85729-493-7_7
downturns-upside


Flyvbjerg, B. (2006), "Five misunderstandings about case-study research", *Qualitative Inquiry*, available at: https://doi.org/10.1177/1087900905284363


Himani Goyal, D. (2015), *Normalization of Data in Data Mining*, available at: https://doi.org/10.1.1.698.1746

Hotelling, H. (1933), "Analysis of a complex of statistical variables into principal components", *Journal of Educational Psychology*, available at: https://doi.org/10.1037/h0071325


Raudnui, S. (2000), "Predetermined time system (PTS) as applied for man hour estimation of some basic preventive maintenance activities", *Systems Integrity and Maintenance*.


Schapire, R.E. (1999), "A brief introduction to boosting", in *IJCAI International Joint Conference on Artificial Intelligence*, available at: https://doi.org/citeulike-article-id:765005


Ye, Z., Cui, J., and Zhou, F. (2017), "Man-hour estimation model based on standard operation unit for flexible manufacturing system", in *MATEC Web of Conferences*, available at: https://doi.org/10.1051/matecconf/201710002017

Appendix

Parameters

\( OP_{1a} = \text{Output1(Mean Prediction Accuracy) of current estimate} = 22.62 \)
\( OP_{1b} = \text{Output1(Mean Prediction Accuracy) of proposed method} = 29.04 \)
\( OP_{2a} = \text{Output2(Mean Absolute Error) of current estimate} = 4.57 \)
\( OP_{2b} = \text{Output2(Mean Absolute Error) of proposed method} = 3.83 \)

Variables

\( w_1 = \text{Weight for DMU 1(current estimate)} \)
\( w_2 = \text{Weight for DMU 2(proposed method)} \)
\( eff = \text{Efficiency parameter} \)

Objective

\[ \text{min}(eff) \]

Constraints

For the current estimate

\[ (w_1 \times OP_{1a}) + (w_2 \times OP_{1b}) \geq (eff \times OP_{1a}) \]
\[ (w_1 \times OP_{2a}) + (w_2 \times OP_{2b}) \leq (eff \times OP_{2a}) \]

For the proposed method

\[ (w_1 \times OP_{1a}) + (w_2 \times OP_{1b}) \geq (eff \times OP_{1b}) \]
\[ (w_1 \times OP_{2a}) + (w_2 \times OP_{2b}) \leq (eff \times OP_{2b}) \]
\[ (w_1 + w_2) = 1 \]