International Journal of Engineering & Technology, 7 (2.28) (2017) 312-319

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Using Machine Learning Algorithms on data residing in SAP ERP Application to predict equipment failures

Manu Kohli

School of Informatics and Computing
Indiana University Bloomington,USA
Corresponding author E-mail: kohlim@umail.iu.edu

Abstract

Asset intensive Organizations have searched long for a framework model that would timely predict equipment failure. Timely prediction of equipment failure substantially reduces direct and indirect costs, unexpected equipment shut-downs, accidents, and unwarranted emission risk. In this paper, the author proposes a model that can predict equipment failure by using data from SAP Plant Maintenance module. To achieve that author has applied data extraction algorithm and numerous data manipulations to prepare a classification data model consisting of maintenance records parameters such as spare parts usage, time elapsed since last completed maintenance and the period to the next scheduled maintained and so on. By using unsupervised learning technique of clustering, the author observed a class to cluster evaluation of 80% accuracy. After that classifier model was trained using various machine language (ML) algorithms and subsequently tested on mutually exclusive data sets with an objective to predict equipment breakdown. The classifier model using ML algorithms such as Support Vector Machine (SVM) and Decision Tree (DT) returned an accuracy and true positive rate (TPR) of greater than 95% to predict equipment failure. The proposed model acts as an Advanced Intelligent Control system contributing to the Cyber-Physical Systems for asset intensive organizations.

Keywords: *CBM; Clustering; Condition based monitoring; Corrective Maintenance; Enterprise Resource Planning; Equipment failure; ERP; HANA; Machine Learning; Plant Maintenance; Predictive Maintenance; Reliability Maintenance; SAP*

1. Introduction

Plant maintenance and asset reliability are strategic to organizations operating in manufacturing, healthcare, utilities and other asset-intensive industry sectors [1]. The efficient working of production lines, uninterrupted and distributed power supplies and servicing customers 24X7 is a key to performance and competitiveness of these industries.The importance of maintenance functions can be gauged by the fact that failure of key equipments installed in hospitals and diagnostic industry sectors can severely impact key services if proper upkeep of the maintenance program isn't followed.

Some organizations carry out manufacturing activities directly at the project site because of logistics challenges or contractual requirements. These organizations are responsible for maintaining their own as well as their customer's assets. Faulty maintenance processes in those cases can lead to loss of revenue, reputation and can endanger human life and environment [2].

Unplanned equipment outages can result in increased corrective maintenance and restoration activities leading to unpredictable budget costs, lower equipment life, MTBF (mean time between failures) and frequent breakdowns [3].

The majority of organizations, large or SME (Small and Medium Enterprises), manage their maintenance processes using Information management systems or ERP (Enterprise Resource Planning) applications such as SAP. Deployment of ERP applications such as SAP allows management of business processes, implement controls and segregation of duties across various functions such as procurement, manufacturing, maintenance, invoicing and collection [4] [5].

From the data released by global ERP, in 2013 the global market share of SAP in ERP deployments [6] was at 24 percent. The harmonization of disparate business processes in an organization can be achieved with the usage of SAP application; leading to better inter-functional integration and overall efficiency and productivity improvements by manifold [7].

Corrective, Preventive, predictive and shutdown are classified as the four areas in which Asset maintenance programs operate [8]. A SAP Plant Maintenance application manages all these processes, related data, and resources. The application also provides a provision to integrate cost controls and purchase processes for maintenance programs successfully [9].

However, the SAP application offers negligible functionality to predict the reliability of equipment based on the historical data and condition monitoring parameters that are otherwise present in the application.Although the decision-making and analytic capabilities of the SAP application are limited, the data mining and ML (Machine Learning) techniques can be applied to data residing in the SAP application to enhance the application's decision-making capabilities. ML algorithms have contributed intensively to fields such as medicine [10], operations [11] and finance [12] to resolve business problems. However only limited research studies have been carried out where data related to maintenance processes residing in Information system has been integrated with ML algorithms to improve equipment reliability for asset-intensive industries. The business case of integrating ML algorithms with SAP application is humongus as SAP application is used by 60 percent

Copyright © 2016 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

of the Fortune 2000 companies and about 80% of the Fortune 1000 companies [13].
Therefore author has proposed in this paper an Equipment reliability prototype, a cyber physical system, developed with artificial intelligence (AI) integration that uses historical maintenance data such as spare part usage, preventive, breakdown and other monitoring master and transactional data variables for equipment those are available in SAP application. The recent advances in SAP design and architecture collaborate with our prototype proposal. The SAP application evolution on cloud platform with SAP HANA [14], allows decision making and analysis functions using a huge amount of data present in memory databases at high performance speed. This feature is useful to develop equipment reliability model as large amount of data in the form of historical maintenance records can be accessed at significantly higher speed to perform reliability analysis.
In this paper the author has presented a hypothesis and also provided a data model to test the hypotheses in Research objective and hypothesis formulation section.
The research outcome of the paper proposes an equipment reliability model that would reduce equipment breakdown, increase the workplace safety and cost savings. The proposed model acts as a decision support system that can help maintenance planners to take the necessary corrective actions and pre-emptive measures to prevent equipment breakdowns.

2. Literature review

Many researchers have dwelled on the Optimization of various Plant Maintenance processes. Various maintenance processes such as Preventive, breakdown and condition-based maintenance processes are practiced by organizations and each method has both advantages and disadvantages.
When unforeseen equipment breakdown occurs, it creates unplanned maintenance costs, hinders operation and impacts productivity. Both time and money loss are experienced when an unplanned equipment downtime occurs, the example of the adverse effects can be seen in construction industry [15]. Usually, the reactive maintenance costs that are a result of unplanned equipment failure come out to be six to eight times costlier than preventive or predictive maintenance [16].
The failure probability of equipment could be reduced considerably if manufacturer recommended planned maintenance schedule is followed. Planned maintenance approach can be implemented on time-based or performance-based parameters or a combination of both. As a result of planned maintenance process designated set of maintenance tasks and overhaul procedures are performed on the equipment when the stipulated run time or running hours of the equipment have elapsed. More often than not, the scheduled maintenance procedures that are generated as a part of preventive maintenance schedule are not necessary or fail to address future equipment breakdowns [17]. The scheduled maintenance procedures are mainly done based on the maintenance manual that the equipment manufacturer provides. However, the disadvantage of the preventive maintenance method is that the maintenance procedure scheduled specified by manufacturers is fixed based on time and running hours and fails to consider the real-time degradation of the equipment's working condition [18].
The method of predicting asset failure on the basis that the asset will show anomalous readings of essential equipment health parameters such as vibrations, temperature, and pressure and so on before the actual failure is referred as condition based monitoring. The condition based measurements recorded through network of sensors can be utilized for efficient manufacturing and optimized Cyber-Physical Systems where networked machines are able to perform efficiently with higher reliability [19].
Numerous studies have been carried out to develop a predictive maintenance model that can be used to diagnose equipment health in production lines. Young et al. [20] evaluated the use of data mining algorithms on an F-18 Aircraft to improve maintenance procedures. The proposed Data mining model consisted of diagnoses and failures codes along with repair details to design a predictive maintenance model to optimize aircraft run time and reduce the risk of any hazardous technical errors. An innovative prediction and assessment model on the equipment performance was proposed by Liao et al. [21]. Under this model, if the equipment reaches its maintenance threshold it would be routed for maintenance and overhaul to restore the equipment normalcy. Pan et al. [22] recommended that a machine's remaining maintenance life, spare part installation history, and effective age must be considered to predict future degradation rate of the equipment and its parts.
As a part of unsupervised learning, clustering [23] is the process of grouping data sets that are more similar to each other. Risk analysis [24] and Market segmentation [25] processes have been performed by researchers and organizations using the method of clustering. However, not many studies have been directed to examine whether independent clustering algorithms can be applied to predict equipment reliability.
For organization deploying SAP application the plant maintenance function is not only effective in handling multiple maintenance processes such as breakdowns, preventive and condition-based but also stores master and transactional data associated with the equipment [9]. Other SAP applications that integrates with SAP plant maintenance function includes Incident Management that aids in improving workplace safety and reduces the risk for life and property [2].
The use of high-speed computing and integrated memory databases have completely changed the ways of research analytics [26], making it possible to allocate a large amount of data processing on real time basis. With the use of in-memory databases like the SAP HANA, SAP application can perform equipment reliability and monitor equipment health by building a classifier model that includes integration of machine learning algorithms with equipment maintenance data residing in SAP application. With advent of next generation cloud based cyber-physical system dominating in the area of industrial automation, the integration of data residing in SAP application or in SCADA and DCS with Machine learning algorithms with SOA (service oriented architecture) is inevitable [27]. This evolution in technology with highly networked equipment supports the decision based cyber-physical equipment reliability model presented in this paper.

3. Research objective and hypothesis formulation

Reliability maintenance is an effective measure to keep unforeseen equipment breakdowns at bay. Fan et al. [28] carried out a Failure Prediction of equipment that was used in a construction industry and witnessed reasonable success rate by using Time Series Models. Almost all of the equipment will show a specific type of signal or symptom, recorded as condition based measurements through an integrated network of sensors before an oncoming breakdown [29]. It has been found out that reliability models can save more than half of the budget one would spend otherwise on various other maintenance regimes.
Prominent research studies makes a point of the integration of data mining with predictive maintenance systems, however researchers till date have failed to integrate ERP applications with ML algorithms to effectively predict equipment reliability incorporating corrective, preventive and condition-based measurements. Taking this research void as an advantage, the author proposes a Cyber-Physical equipment reliability model that effectively predicts equipment failure, thus providing with a workplace with increased productivity, safety, and reduced equipment downtime.

3.1. Research problem

It has been noted that corrective, preventive, shut-down and breakdown maintenance processes can be easily managed inside

an SAP application. For a preventive maintenance process, the period and maintenance schedules are defined in maintenance plans that trigger the work orders for equipment with stipulated list of maintenance tasks when the predefined run time has elapsed. Based on the equipment running hour data and last maintenance process completed, SAP application can effectively predict the future date of the next planned maintenance using the information defined in a Maintenance plan.

Unplanned maintenance scenarios resulting in corrective maintenance processes can be managed by using SAP application. When a maintenance work order is initiated either because of corrective, preventive or breaks down maintenance processes, the system captures information about maintenance costs, spare parts usage, required workforce, etc.

Certain condition based measurements for equipment such as temperature, pressure, vibrations and so on can be recorded in SAP application manually or automatically via an interface.

Equipment reliability, however, remains a challenge that any ERP application cannot provide insight on. Therefore the author proposes to bring a model that would integrate historical maintenance data with real-time condition based measurements and a cohesive method using both unsupervised and supervised AI learning techniques in determining equipment reliability with much more accuracy. In theory, this predictive method can predict timely equipment failure and notify the maintenance personnel to do a corrective maintenance procedure to avoid unplanned equipment break down.

3.2. Hypothesis formulation

The author has hypothesized that effective prediction of equipment breakdown process can be carried out by integrating machine learning algorithms and data points compiled from various maintenance processes such as corrective, predictive and condition based measurements.

Based on the data available from previous researchers, the author has drawn some important points to prepare a data model. From the notes, it can be concluded that the data available in the form of preventive maintenance, corrective maintenance, spare parts usage, condition monitoring, and equipment age and past and forthcoming maintenance schedule can significantly contribute to building an equipment reliability model.

The following steps were performed by the author to verify his findings

- With reference of table1, build a data model with the help of extraction algorithms
- According to the instance of origin classify each data sets as corrective, preventive or breakdown
- Group the data sets into clusters mainly three, by their similarity with the help of AI unsupervised clustering techniques and verifying class to cluster evaluation.
- Training, testing and evaluating the outcomes by applying supervised technique of ML algorithms

3.3. Model application and benefits

The various advantages concerning equipment reliability model propositioned in this paper are:

- Iterative nature: The model posed in the paper learn from incremental data sourced from corrective, preventive and conditional measurements residing in SAP application coupled with ML algorithms to propose equipment reliability.
- Flexibility: the reliability model can be enhanced with additional augments from SAP database, third party application or from data historians to design and enhance the model.
- Swift installation: The reliability model can be directly coupled to SAP production database rather than traditional development process that move through development and quality system resulting in reduced downtime and improved test quality.
- Adaptable: Researchers have proposed that any algorithm that can be adhered to SAP application, the same algorithm can be adapted on other IT or ERP application.

The equipment reliability model can be cyber physical system as shown in figure 1 in which data models and features contributing to the reliability model can be identified for the equipment type such as pumps or compressors. The data model can be developed by a team of business programming experts on a cloud platform integrated with functionalities used in SAP application for example to monitor due dates for the next preventive maintenance. The equipment records and data can be integrated daily with the reliability model to predict and notify of an imminent breakdown dynamically. Based on the prediction outcome, a planner can suggest either a Corrective action or initiate an inspection process to monitor equipment's health. Significant advancements have been observed in the field of robotics, specifically with responsive control of robots based on integrated network of cyber-physical systems and sensors [30] The responsiveness of robots has been observed with high success rate in unexplored and uneven terrains, opening the possibility of managing the feedback of equipment reliability model in the form of corrective action or inspection to be performed by the robot [31].

The model proposes a close integration between SAP function, BI (Business Intelligence) and maintenance department in a business organization as shown in Figure 1. Each function has the following responsibilities:

- SAP Application: Use integration of SAP database and standard deadline monitoring functionality with the reliability model. Using classifier model as a decision support system, to predict equipment breakdown so that timely corrective actions can be taken by the maintennce planner.
- Maintenance Department: As a subject matter expert responsible for proposing set of features for an equipment type that may be useful to perform equipment reliability analysis.
- Business Intelligence: to build a data model on cloud platform by implementing ETL (Extracting Transformation and Loading) methodology on SAP application. Performing integration of data from SAP application or 3rd party system using remote function calls (RFC) or web services.

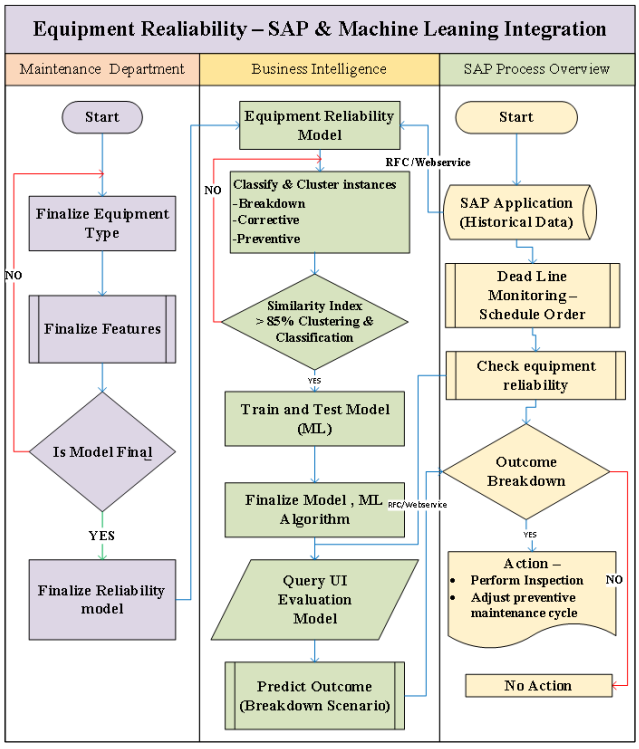


Fig. 1: Equipment reliability model- multiple functions and stakeholders

4. Data model and feature finalization

The data in the reliability model was compiled from multiple tables in SAP application for equipment of type pumps by the author, as shown in Figure 2. The SAP tables in the figure are highlighted in brackets and various conditions initiating the design of the data model are presented in the figure. The objective of the paper was to consider a similar type of equipment; we have opted for pumps in this particular case, to carry out equipment reliability analysis.

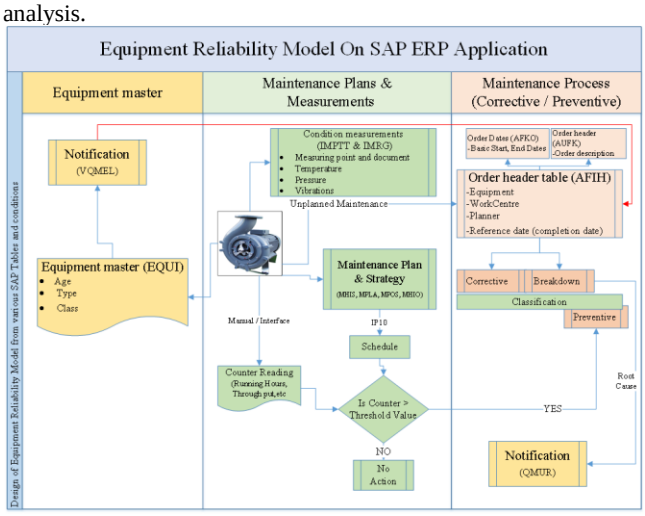


Fig. 2: Equipment reliability model data design

Various master and maintenance transactional datasets were taken from the SAP application for each pump, to design the equipment reliability model. As a step 1, clustering technique was used for these data instances, and the class to cluster evaluation returned with 85% accuracy.

Various types of ML algorithms were employed to predict equipment failure; on the training dataset and then to the test data set to evaluate the robustness of the classifier model. The data model encompasses equipment age, remaining life, condition based measurements, corrective and preventive maintenance schedules and so on. The four steps of data model designing are given below.

• The key features incorporated in the classifier model to predict equipment reliability were compiled by the author taking inputs from a various expert and literature reviews.

• A data model was designed and populated using data extraction algorithms implemented on SAP application.

• The future maintenance date of equipment was taken from SAP preventive maintenance scheduling capability available from IP24 or IP10 transaction in SAP application.

• The integrated model involving various stake holders from SAP, Business Intelligence and Maintenance department was finalized.

4.1. Data extraction and cleaning

The SAP Internet Demonstration and Evaluation System (IDES) application is used by various companies and users mainly for demo and training purpose. The author has extracted data for the required set of features from multiple tables in the SAP IDES application using data extraction algorithms. After that the author has applied data cleaning and transformation to make the data set useful for reliability analysis. Though author has extracted data from various SAP tables, generic or standard SAP extractors can be used to extract the datasets from SAP application as well. The data extraction logic was applied as detailed in Algorithm 1 from the SAP tables as highlighted in brackets in Figure 2.

4.1.1. Data model- equipment reliability design and SAP tables

The author has applied the following data extraction logic for each instance in the data model:

1. Step-1: A plant maintenance work order is sanctioned to carry out multiple maintenance actives such as inspections, corrective maintenance, and preventive maintenance in SAP application. The corrective work order is initiated to restore normalcy in case of breakdowns and the preventive maintenance process is initiated to generate a work order based on elapsed fixed time interval or schedule as recommended by the equipment manufacturer. The work orders were extracted for the equipment type pumps irrespective if they were generated from corrective or maintenance processes. For corrective maintenance order, a further subdivision was performed classifying each order into breakdown and otherwise. Without considering the clustering outcomes, each instance of the data model was assigned as either corrective, preventive or breakdown classification label by the author depending on the type of work order that generated that instance. Only those orders that were technically completed, released and having cost posting were used by the author in the data model. Having a cost posted order ensures that real work was done in that order.

2. Step -2: From the master data of the equipment and associated class, the equipment type (pump) was selected. The lifespan of the pump was assumed to be 15 years [32]. The author has targeted finding the remaining asset life by finding a difference of asset life (15 years) and the date of equipment acquisition and date of the work order.

3. Step -3: Number of days from the work order start date were calculated when the last maintenance job was completed. The calculation was performed by finding the difference of the work order start date and last work order technical completion date for that pump.

4. Step -4: Number of pending days were calculated as a difference of work order start date and the next scheduled preventive maintenance order in a future date. For the scheduled orders that are already generated the actual Start Date of the next release order, or Released Date was taken to perform calculations if the Release date is greater than the Actual Start Date.

5. Step -5 : The condition measurement readings includes the recording of temperature, pressure and so on for the pumps that were used in the analysis. Each condition measurement document has high and low threshold limits defined in SAP application on which measurements are recorded. We observed a high correlation between instances those had higher than the threshold value condition measurements and the work order with breakdown tag. For the feature condition based measurement readings author had found the mean of the upper limit and the lower limit thresholds and recorded percentage deviation of measurement reading from the mean for one week before the order was generated.

The extraction rules as specified in Algorithm 1 were used to prepare the data model from SAP application.

4.1.2. Algorithm-1

Algorithm 1: Data extraction Algorithm

1. Data: table EQUI, EQKT(Equipment Master)
VQMEL; QMUR(Notification)
IMPTT; IMRG (Condition measurements)
AFIH; AFKO; AUFK (Order Master)

Result: Data Model to predict equipment reliability and breakdown

2. Table perform selection initialization and various steps
// Find all pumps from equipment record and their acquisition date

3. While Select All From EQKT where EQKT-EQKTX Contains *PUMP* do
Read EQKT-EQUNR (Equipment Number) and pass EQKT-EQUNR into EQUI-EQUNR and get

EQUI-ANSDT (Equipment Acquisition date)

End

// Find planner group, work center, basic start and end date, Plant, Notification, Maintenance Plan

4. While Select from AFIH passing EQUI-EQUNR into AFIH-EQUNR do

Read AFIH-IWERK (Plant), AFIH-INGRP (Planner group), AFIH-GEWRK (Work Center), AFIH-ADDAT (reference date), AFIH-QMNUM (Notification), AFIH-WARPL (Maintenance Plan)

End

// Fetch Basic start and end date for Maintenance Orders

5. While Select from AFKO passing AFIH-AUFNR do

Read AFKO-GSTRP (Basic Start Date), AFKO-GLTRP (Basic Finish Date)

End

// calculate Condition Monitoring measurements Percentage deviation from mean value

6. While passing EQUI-EQUNR INTO IMPTT-MPOBJ do

Read from IMPTT, IMPTT-MRMAX (Upper Range Limit), IMPTT-MRMIN (Lower Range limit), IMPTT-POINT (measuring point) and pass IMPTT-POINT, into IMRG-POINT and Read IMRG-READG (Measurement Reading)

End

// Conditions for deviations of Condition based Measurements are higher than Normal

7. If IMRG-READG(Measurement Reading) >IMPTT-MRMAX

then High CBM= Mean of 1 week[(IMPTTMRMAX),(IMPTT-MRMIN)]-(IMRG-READG)

// Calculate number of days since last maintenance completed

8. Last Maintenance completed- Perform (AFKO-GSTRP)-(AFIH-ADDAT)

// Remaining Asset Life

9. Asset Life Left= Perform (15X365)-((EQUI-ANSDT)-(AFKO-GSTRP))

// Number of days remaining for next preventive maintenance order

10. Number of days for next Preventive order= Find difference of end date of the order and next Preventive maintenance order (IP10/IP24)

/* iterate over all examples *

11. Exit

Table 1 has the list of features for the final data model. The data model was built by performing from various calculations and measurements from the equipment master and transactional data from SAP tables as mentioned in Algorithm1. The working data was taken from the past records of the equipment such as preventive maintenance, corrective maintenance and breakdown maintenance.

Each instance in the dataset was assigned the classification as either corrective, preventive or break down. The basis of the assignment of that classification is listed by the author in Table 2.

4.2. Data cleaning and finalization

The database made for the model contained 274 instances for 39 distant types of pump, and each instance is linked to 11 features. The classification of the instances came under corrective, preventive and breakdown classification based upon the cause of the maintenance process that started the maintenance order for instance. The data was divided accordingly so that 25% goes into the testing set (52) and 75% goes into the training set (n=205).

Table 1: Data Model: Features used in Equipment Reliability Model

Feature	*Feature description*	*Feature Type*
Plant	The Physical location where equipment is located.	Attribute-Nominal
Equipment	The equipment or the locationtag on which failure is to be predicted.	Attribute-Nominal
Main WorkCtr	Center or a workshop responsible for maintenance of pumps,	Attribute-Nominal
Planner Group	An organization unit or a person who is designated responsible for maintenance of pump.	Attribute-Nominal
Asset Life Left	Calculated based on the asset/equipment life left when the order as generated	Attribute-discrete in 10 bins
Time in days for next Planned Maintenance	Time in days remaining for next closest scheduled preventive maintenance date in future.	Attribute-Discrete in 10 bins
Time in days since last maintenance	Time elapsed since last maintenance was completed. Calculated as a difference of basic start date of the order in the instance and last maintenance order completion date.	Attribute-discrete in 10 bins
Spare part Used	If spare part was used in the maintenance process	Attribute-Boolean(Y/N
Condition based measurements	The percentage deviation of the measure document from the mean value of upper and lower limit.	Attribute-discrete in 10 bins
Root Cause	The cause codes populated in the notification object indicating the reason of equipment failure	Attribute-Nominal
Classification	Either preventive, breakdown or corrective based on work order that generated the instance	

Table 2: Equipment Reliability Analysis: Classification

Classification	*Remarks*
Breakdown	Classify instance as breakdown if breakdown Indicator is allocated to the Notification associated with order VIQMEL-MSAUS=X
Preventive \hline	Classify instance as preventive if the Order is generated Concerning Maintenance plan AFIH-WARPL not equal to Blank
Corrective	Classify instance as corrective if Breakdown Indicator is not allocated, order is not generated from subcontracting process

This filtration is made possible by the use of WEKA resample filter which was set at invert selection and zero replacement filter settings. Each feature in the data model is associated with a feature type that is mentioned in Table 1. The dataset for class distribution is given in Table 3.

Table 3: Equipment Reliability Analysis: Classification

Cluster to Class Evaluation	*Preventive*	*Corrective*	*Breakdown*	*Total*
Observed classification	102	51	121	274
Instances correctly clustered (%)	75%	85%	85%	80%
Instances correctly clustered (Number)	76	43	102	221

4.3. Validating results with clustering

The datasets were made to be tested with clustering technique to validate classification outcome with the clustering learning done by unsupervised learning algorithm independently. With clustering technique the robustness of the equipment reliability model is generalized and validated as no classification label is passed to the clustering algorithm. The clustering outcome results can be seen in Table 3 showing an accuracy of 80 %. The higher accuracy of 85% is observed for corrective and breakdown classification which was the primary problem this paper tried to address.

5. Experiment results – implementation of machine learning algorithms

Five ML algorithms namely Logistic Regression, Naive Bayes, Support Vector Machine, Decision Tree and KNN were used by the author as the part of the supervised learning process to build the classifier model. ML algorithms were trained on the training data set, and the tested on the test data set. Ten-fold cross validation was performed to increase accuracy and generalization of the classifier model. Testing and training of the ML algorithms on the data model was performed using the WEKA software. The approach author has taken to compile the data model is shown below in Figure 3. Originating from feature selection, class to cluster evaluation steps, author has built a classifier model integrating various machine learning algorithms with the data model. Suitable steps were taken by the author to negate bias and variance by deploying ensemble methods.

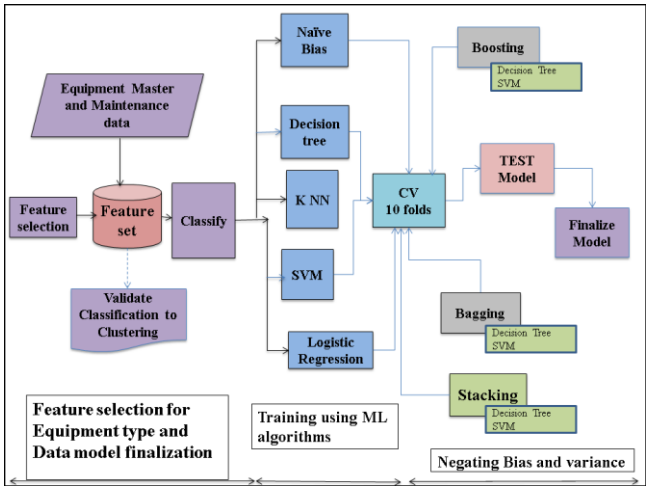


Fig. 3: Implementing ML algorithms to predict equipment failure

6. Experiment results – application of machine learning algorithms on dataset

Salzburg [33] states that cross validation is an efficient method for the reduction of data discrepancies thus improving the reliability and ensuring generalisation of the classifier model. All the ML algorithms such that were evaluated for implementation such as as Logistic Regression (LR), Naive Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT) and KNearest Neighnbour (KNN) were subjected to 10 fold cross validation to improve accuracy and avoid overfitting of the model. Each of the machine learning models was trained on mutually exclusive training dataset (n=205) and test dataset (n=52).
The results recorded are mentioned in Table 4 and 5 for both tests and training sets listing precision, accuracy, recall, F-score of each algorithm and true positive rate (TPR).

Table 4: ML Algorithm Implementation Results - Training Dataset CV 10 Fold

Algorithm	*Accuracy*	*Precision*	*Recall*	*F-Measure*	*TPR*
NB	84.39	.85	.844	.844	.844
DT	80.01	.78	.80	.78	.80
KNN,K=4	84.78	.842	.849	.836	.849
SVM	84.39	.839	.844	.841	.844
SVM Breakdown	91.1	.891	.911	.901	.911
LR	82.44	.812	.824	.814	.824

Table 5: ML Algorithm Implementation Results - Test Dataset

Algorithm	*Accuracy*	*Precision*	*Recall*	*F-Measure*	*TPR*
NB	92.30	.925	.923	.922	.923
DT	94.23	.951	.942	.942	.942
KNN,K=4	78.84	.78	.78	.775	.788
SVM	98.07	.982	.981	.981	.981
SVM Breakdown	99.98	.941	1	.97	1
LR	100	1	1	1	1

The results observed showed that Decision tree, LR, and SVM showed higher accuracy rate of more than 94.5 %. The true positive rate for the model for LR, SVM, and DT was also observed to be more than .95 and .98 for DT. The False positive rate (FPR) was observed to be very low for these algorithms, confirming to the paper objective to correctly and timely predicting equipment failure. Based on the results, it can be assumed that LR, SVM, and Decision tree are suitable to perform equipment reliability and predicting breakdowns for equipment on SAP database. However, to eliminate bias and variance factors, author brought on ensemble methods that are explained in detail in the next section.

7. Ensemble methods: negating bias and variance

The author managed the negation of bias and variance by deploying boosting, bagging and stacking techniques, part of ensemble methods, using the WEKA software tool with tenfold cross-validation. The bagging technique when called builds multiple models independently and draws sample data sets randomly from the data pool with an objective to decrease variance. However, boosting technique adds new model incrementally to the classifier with an objective to manage bias. The stacking technique was also used by the author where SVM and decision tree models were combined to achieve better prediction and classification results.
The analysis of the Ensemble Method results is listed in Table 6. Acquired from the examination above of the classification algorithms revealed that SVM and Decision Tree are appropriate methods to predict equipment failure and contribute significantly to equipment reliability model with high accuracy of close to 100 percent.

Table 6: ML Algorithm Test Results – Ensemble Methods

Algorithm	*Accuracy*	*Precision*	*Recall*	*F-Measure*	*TPR*
SVM Bagging	94.23	.945	.942	.942	.942
DT Bagging	96.15	.966	.962	.961	.962
SVM Boosting	100	1	1	1	1
DT Boosting	100	1	1	1	1
Stack DT +SVM	98.071	.982	.981	.981	.981

The test results of classification models using various ML algorithms are also shown in Figure 4 in a graphical format.

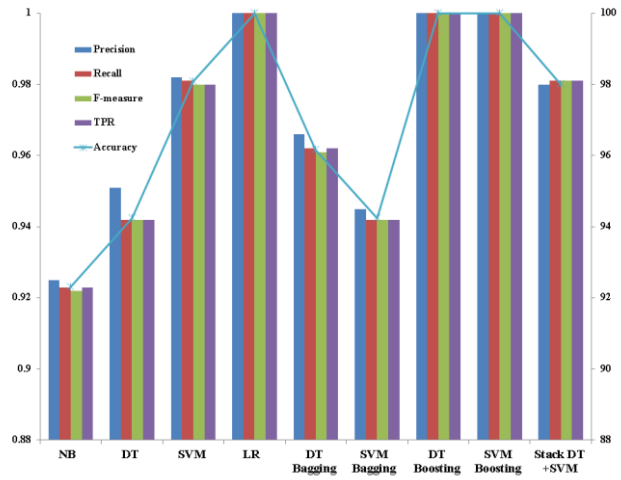


Fig. 4: Test results of various classification models CV 10 fold

8. Conclusion

The present paper proposes a data extraction algorithm and equipment failure prediction model by integrating machine learning algorithms with maintenance datasets existing in SAP application. The author has used data from various tables in SAP application to build a data model that could predict imminent equipment failure with high accuracy. Various transaction and master datasets from the SAP application were used and multiple data manipulations were applied to develop the data model. Various ML models were applied for the prediction of Equipment / Asset failure and tested successfully with an accuracy of more than 95 percent. It can be therefore said that integrating ML algorithms with SAP application datasets can provide a system with high prediction percentage. The proposed equipment reliability model can be generalized to any other IT application not only to SAP as the rules to design data model will remain more or less same. Additional research is recommended that may lead to improved accuracy and generalization of the equipment reliability model by integrating clustering outcomes with supervised ML algorithms to perform classification and predict equipment reliability. A system that can successfully predict machine breakdowns early can ensure improved productivity, lower emission risk and improved safety in workplaces.

Acknowledgement

This is a text of acknowledgements. Do not forget people who have assisted you on your work. Do not exaggerate with thanks. If your work has been paid by a Grant, mention the Grant name and number here.

References

[1] Fraser K, Hvolby HH & Tseng TL, "Maintenance Management Models: A Study of the Published Literature to Identify Empirical Evidence: A Greater Practical Focus is Needed", *International Journal of Quality & Reliability Management*, Vol.32, No.6, (2015), pp. 635-664, available online: https://doi.org/10.1108/IJQRM-11-2013-0185, last visit:26.04.2018

[2] Kohli M, *Incident Management with SAP EHS Management*, Rheinwerk Publishing, (2015), pp:1-132.

[3] Bore CK (2008), Analysis of Management Methods and Application to Maintenance of Geothermal Power Plants. *United Nations University, Geothermal Training Programme*, 1-52.

[4] Hayen R, *SAP R/3 Enterprise Software: An Introduction*, McGraw-Hill/Irwin, (2006), pp:1-192.

[5] Yasar A & Ozer G, "Determination the Factors that Affect the Use of Enterprise Resource Planning Information System through Technology Acceptance Model", *International Journal of Business and Management*, Vol.11, No.10, (2016), pp:91-108.

[6] Columbus L, "Erp market share update: Sap solidifies market leadership", *Forbes*, available online: https://www.forbes.com/sites/louiscolumbus, Retrieved on December, 2013.

[7] Wang Y (2013), Effects of implementation of sap on management accounting: Case: Dongfeng motor corporation, *University of Applied Sciences*, 1-88.

[8] Garrido S, "Types of Maintenance programs", available online: http://www.mantenimientopetroquimica.com/en/typesofmaintenance.html, Retrieved on January, 2017.

[9] Stengl B & Ematinger R, *SAP R/3 Plant Maintenance: Making it Work for your Business*, Pearson Education, (2001), pp:1-368.

[10] Birkner MD, Kalantri S, Solao V, Badam P, Joshi R, Goel A, Pai M & Hubard AE (2007), Creating diagnostic scores using data-adaptive regression: An application to prediction of 30-day mortality among stroke victims in a rural hospital in India. *Therapeutics and clinical risk management* 3, 457-484.

[11] Qiang G, Zhe T, Yan D & Neng Z, "An Improved Office Building Cooling Load Prediction Model Based on Multivariable Linear Regression", *Energy and Buildings*, Vol.107, (2015), pp.445-455.

[12] Yu F, "Accounting Transparency and the Term Structure of Credit Spreads", *SSRN Electronic Journal*, Vol.75, No.1, (2005), pp:53-84.

[13] Missourisandt, "Erp share by fortune 2000 companies", available online: http://erp.mst.edu, Retrieved on January, 2017.

[14] Sikka V, Färber F, Lehner W, Cha SK, Peh T & Bornhövd C, "Efficient Transaction Processing in SAP HANA Database", *Proceedings of the 2012 international conference on Management of Data - SIGMOD '12*, ACM, New York, NY, USA, (2012), pp:731-742, https://doi.org/10.1145/2213836.2213946.

[15] Prasad Nepal M & Park M, "Downtime Model Development for Construction Equipment Management", *Engineering, Construction and Architectural Management*, Vol.11, No.3, (2004), pp. 199-210, available online: https://doi.org/10.1108/09699980410535804

[16] Colen IF & Brito J, "Building facades maintenance support system", *Proceedings of The XXX IAHS World Congress on Housing*, Vol.9, No.13, (2002).

[17] Fu C, Ye L, Liu Y, Yu R, Iung B, Cheng Y & Zeng Y, "Predictive Maintenance in Intelligent-Control-Maintenance-Management System for Hydroelectric Generating Unit", *IEEE Transactions on Energy Conversion*, Vol.19, No.1, (2004), pp.179-186.

[18] Tsang AH, "Condition-Based Maintenance: Tools and Decision Making", *Journal of Quality in Maintenance Engineering*, Vol.1, No.3, (1995), pp:3-17, available online: https://doi.org/10.1108/13552519510096350

[19] Lee J, Bagheri B & Kao HA, "A Cyber-Physical Systems Architecture for Industry 4.0-Based Manufacturing Systems", *Manufacturing Letters*, Vol.3, (2015), pp.18-23, available online: https://doi.org/10.1016/j.mfglet.2014.12.001

[20] Young T, Fehskens M, Pujara P, Burger M & Edwards G, "Utilizing Data Mining to Influence Maintenance Actions", *2010 IEEE AUTOTESTCON*, Vol.1, No.5, (2010).

[21] Liao W, Wang Y & Pan E, "Single-Machine-Based Predictive Maintenance Model Considering Intelligent Machinery Prognostics", *The International Journal of Advanced Manufacturing Technology*, Vol.63, No.1, (2012), pp.51-63.

[22] Pan E, Liao W & Xi L, "A Joint Model of Production Scheduling and Predictive Maintenance for Minimizing Job Tardiness", *The International Journal of Advanced Manufacturing Technology*, Vol.60, No.9, (2011), pp.1049-1061.

[23] Wagstaff K, Cardie C, Rogers S & SchrÖdl S, "Constrained K-means Clustering with Background Knowledge", *Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001)*, Williams College, Williamstown, MA, USA, (2001), pp. 577-584.

[24] Murè S, Comberti L & Demichela M, "How Harsh Work Environments Affect the Occupational Accident Phenomenology? Risk Assessment and Decision Making Optimisation", *Safety Science*, Vol. 95, (2017), pp. 159-170, available online: https://doi.org/10.1016/j.ssci.2017.01.004

[25] Kuo R, Ho L & Hu C, "Integration of Self-Organizing Feature Map and K-means Algorithm for Market Segmentation", *Computers & Operations Research*, Vol.29, No.11, (2002), pp. 1475-1493, available online: https://doi.org/10.1016/S0305-0548(01)00043-0

[26] Lake P & Crowther P (2013), In-memory databases. *Concise Guide to Databases, Undergraduate Topics in Computer Science, Springer*, 183-197.

[27] Colombo AW, Karnouskos S & Bangemann T, "Towards the Next Generation of Industrial Cyber-physical Systems", *In Industrial cloud-based cyber-physical systems*, Springer International Publishing, (2014), pp.1-22.
[28] Fan Q & Fan H, "Reliability Analysis and Failure Prediction of Construction Equipment with Time Series Models", *Journal of Advanced Management Science*, Vol.3, No.3, (2015), pp. 203-210, last visit:26.04.2018
[29] Geitner FK & Bloch HP, "Machinery Component Failure Analysis", *Machinery Failure Analysis and Troubleshooting*, Oxford (2012), pp.87-293.
[30] Vladareanu V, Dumitrache I, Vladareanu L, Sacala IS, Tont G & Moisescu MA, "Versatile Intelligent Portable Robot Control Platform Based on Cyber Physical Systems Principles", *Studies in Informatics and Control*, Vol.24, No.4, (2015), pp.409-418, last visit:26.04.2018
[31] Vladareanu L, Tont G, Ion I, Munteanu MS & Mitroi D, "Walking Robots Dynamic Control Systems on an Uneven Terrain", *Advances in Electrical and Computer Engineering*, Vol.10, No.2, (2010), pp:145-152.
[32] Swedepump, Lifespan of Centrifugal Pump, *FLYGT*, (2017), Retrieved from http://www.swedepump.by/files/1587870.pdf
[33] Salzberg SL, "On Comparing Classifiers: Pitfalls to Avoid and a Recommended Approach", *Data mining and knowledge discovery*, Vol.1, No.3, (1997), pp. 317-328, available online: https://doi.org/10.1023/A:1009752403260, last visit:26.04.2018