Automatic learning algorithm for troubleshooting in hydraulic machinery

Jose Castro-Puma, Miguel Castro-Puma, Verónica More-Sánchez, Juana Marcos-Romero, Elio Huamán-Flores, Claudia Poma-García, Rufino Alejos-Ipanaque

1Faculty of engineering and management, Universidad Nacional Tecnológica de Lima Sur, Lima, Peru
2School of Economics, Universidad Nacional Federico Villarreal, Lima, Peru
3Faculty of Health Sciences, Universidad Autónoma de Ica, Ica, Peru
4Faculty of Industrial Engineering, Universidad César Vallejo, Lima, Peru
5Faculty of Administrative Sciences, Universidad Nacional del Callao, Lima, Peru

Article Info

Article history:
Received Dec 17, 2021
Revised Sep 6, 2022
Accepted Sep 19, 2022

Keywords:
Algorithm
Ensemble bagged trees
Hydraulic excavators
Machine learning
Maintenance service

ABSTRACT

In Peru, there are many companies linked to the category of heavy machinery maintenance, in which, on the one hand, although it is true they generate a record of events linked to equipment maintenance indicators, on the other hand they do not make efficient use of these data generating operational patterns, through machine learning, that contribute to the improvement of processes linked to the service. In this sense, the objective of this article is to generate a tool based on automatic learning algorithms that allows predicting the location of faults in hydraulic excavators, in order to improve the management of the maintenance service. When developing the research, it was obtained that the algorithm that assembles bagged trees presents an accuracy of 97.15%, showing a level of specificity of 99.04%, an accuracy of 98.56% and a sensitivity of 97.12%. Therefore, the predictive model using the ensemble bagged trees algorithm shows significant performance in locating the system where failures occur in hydraulic excavator fleets. It is concluded then that it was possible to improve aspects associated with the planning and availability of supplies or components of the maintenance service, also optimizing the continuity and response capacity in the maintenance process.

Corresponding Author:
Jose Antonio Castro-Puma
Faculty of engineering and management, Universidad Nacional Tecnológica de Lima Sur
Lima, Peru
Email: jacastrop3@gmail.com

1. INTRODUCTION

The costs related to the maintenance of a machine in general, and the lack of operation of this within the chain of the production process, represent around 15% to 60% of total production costs [1]-[3]. Hydraulic excavators, as well as other machines used in the mineral extraction process, must be permanently monitored so that they do not stop their operation abruptly, since the impact on the production level is significant [4]-[8]. The purpose of the maintenance strategies of a machine is to improve the metrics or indicators of the equipment's operability, such as availability, reliability and maintainability [9]-[11]. An alternative for the early detection of failures in machines that intervene in the mining extraction process is predictive maintenance, which uses historical data of criticality of each component of the machine and that through statistical techniques, it is possible to approximate possible failures [12]-[14]. Today, with the amount of data that can be measured or monitored in machines, it is possible to apply machine learning algorithms in order to generate patterns and trends of operational behavior [15]-[20]. Artificial intelligence, as well as data mining, and the ability to transmit information from

Journal homepage: http://ijeecs.iaescore.com
machines that operate in an industrial environment at this time represent technological tools that contribute to the prediction of failures [21]-[25]. In this fourth industrial revolution or also called industry 4.0, it is relevant that all production processes, as well as the machines and tools involved, are linked to maintenance strategies based on machine learning algorithms, whether classification or regression must guarantee the continuity of the process avoiding failures or unscheduled stops, which generates low productivity and economic losses [26]-[28].

Ensemble bagged trees, is part of the learning of classification, which is a technique of construction of sets, which by means of a data sample several samples are randomly extracted, that is, each variable can be chosen from the original population, so that each variable is equally likely to be selected in each interaction of the process. Once the samples are formed, the models are trained separately, obtaining the final output prediction of all the sub-models. This method is used as a way to reduce the variance of the base estimator (decision tree), by introducing randomization in its construction procedure and then making a set from it [29]-[31]. In the context of the aforementioned, this article aims to determine the classification algorithm and its metrics (sensitivity, specificity, precision and accuracy) for the predictive model of the location of failures in hydraulic excavators, order to improve planning of the maintenance service of these machinery used in mineral extraction production processes in Peru.

2. LITERATURE REVIEW

Industry 4.0, has as one of its fundamental principles to give a relevant value to the data that is produce or generate in a production or service process in order to extract significant information from them they [32], [33]. So also in based on innovation offers great potential in the processes, one of these being the one linked to the bodybuilding sector, which has not ignored this reality so that a large percentage of companies with this line of business are putting into practice to achieve the autonomy of processes in order to reduce time and costs [34]. Artificial intelligence (AI) is part of data science, whose purpose is to condition, process, analyze and reveal through data what lies behind natural, human and social phenomena from a multidimensional, flexible and dynamic perspective [35], [36]. From artificial intelligence it was possible to structure automatic learning algorithms, that is, through automatic learning, computers reach a certain level of autonomy that allows prediction through regression or classification models from the monitoring and acquisition of data from a process or a machine, under any context of work or operation [37].

Liu et al. [38], the author states that, through the historical record of faults in the machinery used in mineral extraction processes, it is possible to predict the operating behavior of the machine, through patterns based on algorithms of machine learning, both supervised and unsupervised. In this regard in [39], the author points out that predictive maintenance strategies in excavation machines traditionally used are based on the collection of historical data manually, however, from the insertion of artificial intelligence to the industrial sector, the techniques and mechanisms aimed at improving the performance of machines are oriented towards the use of neural networks and machine learning. In Alhilali et al. [40], the authors point out that supervised learning consists of an algorithm establishing a behavior pattern from a set of input and output data. In Qarabsh et al. [41], the author points out that in the current context of industry 4.0, industrial maintenance must evolve towards a model that integrates networks with sensors that allow the transfer of electrical signals in real time, supporting the internet of things (IoT) and artificial intelligence. In it is pointed out that ensemble methods try to improve the performance of machine learning models by improving their accuracy in order to solve a particular problem, within this method is the ensemble bagged trees classification algorithm, which is a powerful statistical method for estimating a quantity from a sample of data [42], [43].

3. RESEARCH METHOD

The research design is of a non-experimental type, because initially tests were carried out to search for patterns of the collected data (historical criticality of hydraulic excavator systems) based on various automatic learning algorithms, with the purpose of Identify which of all the analyzed algorithms show better results for sensitivity, specificity, precision and accuracy (algorithm performance metrics). After determining the algorithm with the best performance, the results of the fault classification model were obtained with respect to the system where said fault is located (the systems that make up the hydraulic excavator will be called algorithm classes). Table 1 shows the algorithm classes with their respective coding.

| RN | Code | Class            |
|----|------|-----------------|
| 1  | DS   | Drive system    |
| 2  | RS   | Refrigeration system |
| 3  | IS   | Intake system   |
| 4  | LS   | Lubrication system |

Table 1. Coding of algorithm classes

Indonesian J Elec Eng & Comp Sci, Vol. 29, No. 1, January 2023: 535-544
4. DESCRIPTION AND DEVELOPMENT

4.1. Description

The development of the research was carried out on a study population composed of Caterpillar brand hydraulic excavators with a capacity of 75TN, whose number is 126 excavators. Also, because it was possible to acquire and record data from all the excavators that make up the population, for this investigation it was considered that the sample is equal to the population. Figure 1 shows the architecture of the predictive model determination using the classification algorithm, which aims to locate the system in which faults occur in a hydraulic excavator.

It should be noted that, although the hydraulic excavator is composed of 9 systems, the selection of the four systems called algorithm classes (motor, cooling, admission and lubrication) was made based on the data collected. In which the criticality, frequency of failures and operating conditions of the machinery, that is, resources were selected and directed in the systems where it is most necessary to improve the reliability and availability of the hydraulic excavator. In Figure 2, the traditional testing process for fault detection and identification is shown.

Likewise, from these test processes on fault identification data are generated and stored on the location of faults in a historical way, so that a large volume of structured data is generated. And whose utility will be centered on a supervised learning algorithm, to perform a fault classification process with respect to each excavator machine. In Figure 3, the data obtained in the maintenance service process is displayed.

![Figure 1. Architecture of the predictive model determination](image)
Table 2 shows the criticality values obtained in the data collection period of each system. In which it is specified that these indicators refer to failure frequency, mean time between failures (MTBF), medium time to repair (MTTR) and level of criticality. With respect to the frequency of failures, a greater value is presented in the management system, while with respect to the MTBF and MTTR indicators, they are presented in the auxiliary systems and the management system, respectively.

| Systems             | Failure frequency | MTBF   | MTTR | Criticality condition |
|---------------------|------------------|--------|------|-----------------------|
| Drive System        | 832              | 224.81 | 41.23| 3                     |
| Refrigeration system| 572              | 273.43 | 33   | 2                     |
| Intake system       | 348              | 305.76 | 25.3 | 2                     |
| Lubrication system  | 260              | 331.26 | 22.46| 2                     |
| Starting and charging system | 79          | 655.98 | 15.3 | 1                     |
| Fuel system         | 53               | 676.98 | 9.89 | 1                     |
| Exhaust system      | 52               | 678.9  | 9    | 1                     |
| Control system      | 47               | 683.4  | 9.05 | 1                     |
| Auxiliary systems   | 39               | 691.5  | 8.97 | 1                     |

4.2 Development

Through the MATLAB R2021a software and the classification learner and statistics and machine learning toolbox 12.1 tools, the predictive model with the highest accuracy in locating failures in hydraulic excavators is identified. The results generated by the Matlab R2021a software are shown in Table 3. According to Table 3, of all the supervised learning algorithms, the best classification model for the location of failures in hydraulic excavators is given by the ensemble bagged trees algorithm, with an accuracy (validation) of 97.1%. Likewise, the comparative analysis of the classification algorithms is carried out according to their performance metrics (sensitivity, specificity, precision and accuracy). In Figure 4 the ensemble bagged trees algorithm is the one with the best sensitivity value of 0.97, which means that this algorithm is the one that best expresses...
how well the model can detect true positive rates (TPR). This refers to the proportion of positive cases that were correctly identified by the algorithm.

Table 3. Choice of algorithm according to its accuracy

| Class                  | Accuracy  |
|------------------------|-----------|
| Tree: Fine Tree         | 74.8%     |
| SVM: Cubic SVM          | 77.33%    |
| SVM: Fine Gaussian SVM  | 74.66%    |
| Ensemble: Bagged Trees  | 97.1%     |
| Neuronal Network: Medium Neuronal Network | 81.10% |
| Neuronal Network: Wide Neuronal Network | 90.90% |
| Neuronal Network: Trilayered Neuronal Network | 78.00% |

In Figure 4 shows the confusion matrix, in relation to the sensitivity metric, which indicates the rate of true positives (TPR) and the rate of false negatives (FNR) of the predictive model. As shown in Figure 4 in the double support (DS) model, 97.5% of positive samples are correctly classified as positive, while 2.5% of positive samples are erroneously classified as negative in right stance (RS). In the IS model, 95.9% of positive samples are correctly classified as positive, while 4.1% of positive samples are erroneously classified as negative in left stance (LS). In the LS model, 96.7% of positive samples are correctly classified as positive, while 3.3% of positive samples are erroneously classified as negative in IS. And in the RS model, 98.4% of positive samples are correctly classified as positive, while 1.6% of positive samples are erroneously classified as negative in LS.

Although all the sensitivity levels shown in Figure 4 are high, it is highlighted that of the 4 classes on which the predictive model acted, the RS class shows the best percentage of sensitivity (98.4%), this means that in this class the predictive model has the greatest ability to discriminate between a true positive rate (TPR) from a false negative rate (FNR). Also, it can be indicated that the percentage of the determined false negative rates are considered as low. In Figure 5 shows the confusion matrix, in relation to the accuracy metric, which indicates the positive predictive value (PPV) and the false detection or false discovery rate (FDR) of the predictive model. As shown in Figure 6 in the DS model, 100% of samples have the probability that a positive and significant finding is true, while 2.4% of the sample has the conditional probability that a false finding reflects a true effect on SR. In the IS model, 96.7% of samples have the probability that a positive and significant finding is true, while 4% of the sample have the conditional probability that a false finding reflects a true effect on SR. In the LS model, 94.4% of samples have the probability that a positive and significant finding is true, while 3.3% of the sample has the conditional probability that a false finding reflects a true effect on IS. And in the RS model, 97.6% of samples have the probability that a positive and significant finding is true, while 1.6% of the sample have the conditional probability that a false finding reflects a true effect on LS. Another aspect to take into account are the levels of accuracy, which turned out to be high, however, of the 4 classes on which the predictive model acted, the DS class shows the best percentage of accuracy (100.0%), this means that in this class the predictive model has the best ability to assess the probability of a significant result reflecting a true difference.
Next, Table 4 shows the results of sensitivity, specificity, precision and accuracy (performance metrics of the algorithm) of the predictive model, for each class. This Table 4 shows that the four metrics show relatively high values in the 4 classes, finding the general average of the specificity with a yield of 99.04%. The precision with a yield of 98.56%, the sensitivity with a yield of 97.12% and the accuracy of the predictive model with 97.15% performance.

Table 4. Ensemble bagged trees classification algorithm metrics

| Class          | Sensitivity | Specificity | Precision | Accuracy |
|----------------|-------------|-------------|-----------|----------|
| Drive system   | 97.54%      | 100.00%     | 99.38%    | 100.00%  |
| Refrigeration system | 95.87%      | 98.90%      | 98.15%    | 96.67%   |
| Intake system  | 96.69%      | 98.08%      | 97.74%    | 94.35%   |
| Lubrication system | 98.36%      | 99.18%      | 98.97%    | 97.56%   |
| Total          | 97.12%      | 99.04%      | 98.56%    | 97.15%   |

Determined the classification algorithm and its metrics (sensitivity, specificity, precision and accuracy). Using the following Figure 6 shows the procedure of the application of the predictive model in the location of failures in hydraulic excavators in a service company. As shown in Figure 6, the application of the predictive model seeks to generate a positive effect in the management of the maintenance service, optimizing its continuity, capacity and availability, in this way a correct and uninterrupted operation will be obtained at a reasonable cost and with correct resources dimensioned.

Figure 6. Procedure for the application of the predictive model in the location of failures

4.3. Discussions

The results are similar to those obtained in [7] where it is observed that the machine learning algorithm correctly classifies the failure data in their respective systems (hydraulic, electrical, motor, mechanical, lubrication and refrigeration). Therefore, when applying this technological tool, it is possible to reduce the time spent in the process of classifying the failure data of the PC4000-6 fleet of machines, making use of a machine learning algorithm with an accuracy of 85%. In this way a support tool can be available to maintenance personnel to enable them to quickly obtain adequate information in order to seek strategies to improve the maintenance management of the PC4000-6 machine fleet.

As indicated in [24], the use of a predictive model using machine learning algorithms is carried out in order to be able to more profitably manage the maintenance of the asset in the operation of the mining trucks, for which a ROC (AUC) of almost 100% (value = 1) allows to visualize the effectiveness of each type of
labeling in the algorithm. In Baptista et al. [12] supervised learning of machine learning was used to predict the state of the induction motor bearings, the model gave a prediction percentage of 80% in serious failures, 97% of minor failures, 82% moderate failures and 100% healthy. In Abdullahi et al. [3] the algorithm programmed in MATLAB allows to detect faults and identify their nature in induction electric motors. In this way, failures are detected in their initial stages, so there is enough time to plan and schedule corrective actions (corrective maintenance), minimizing downtime and the negative effect on production, guaranteeing better quality of repairs. The Hu et al. [6] manages to detect and diagnose the operating modes and faults of a motor by means of machine learning with an accuracy of 98.06%, thereby achieving the correct detection of 13 of the 15 operating modes or failure, providing this This is an advantage to the system, since it would quickly warn a triad of failures with the consequent advance warning and observation status of an evolution towards failure.

Likewise, in the research carried out in [10] a precision of 90.3% and 84.5% has been obtained, thus fulfilling the objective of creating an artificial intelligence that self-diagnoses the state of the actuator with a certain precision and potentially more efficient than manual diagnostics could be done by any operator. The aforementioned study demonstrates the great potential of machine learning techniques, and how they can improve the performance of a wide variety of activities. In Jiang et al. [14] it is pointed out that the use of the machine learning technique with MATLAB, manages to improve predictive maintenance, therefore, it also optimizes the availability and service precision of the Komatsu 830E and 930E electric mining trucks. The research determines that the most critical systems in Komatsu trucks are mainly in the electrical propulsion system, specifically in the drive wheels. As Zeng et al. [4], the solution developed using machine learning algorithms allows predicting the appearance of the different failure modes described in the FMECA of a ship's combustion engine. The tasks of prediction and detection of anomalies are totally independent, so the latter can be carried out both for future moments (data from the prediction), present (real time) or past (a posteriori analysis).

4 CONCLUSION

The improvement of continuity management processes, capacity and availability of maintenance services, through technological tools, seek to provide operational support, with an effective cost and with correctly dimensioned resources that achieve the satisfaction of their strategic objectives in the organizations or companies of service, which will be reflected in customer satisfaction. Thus, through the investigation, it was determined that the predictive model with the ensemble bagged trees algorithm grants an accuracy of 97.15% in the location of the system in which the failures in hydraulic excavators occur, thus contributing to planning and availability of resources in the maintenance process, also optimizing the continuity, capacity and availability of the maintenance service. Since, in maintenance services, the demand for components and supplies that could possibly be useful to carry out a change or installation quickly is not anticipated, especially when it comes to a fleet, for this reason the contribution of the ensemble bagged trees classification algorithm since it specifies to 97.15% the location of the system where the fault is found, due to the criticality condition, frequency of failures and operating conditions of the machinery. By performing predictive maintenance, not only can the current status of the machinery be analyzed, but also more precise maintenance can be planned, reducing unplanned production downtime and unwanted costs, due to failures not detected by techniques preventive maintenance, all this generates satisfaction in customer service.

REFERENCES

[1] C. Stenstrom, P. Norrbin, A. Parida, and U. Kumar, “Preventive and corrective maintenance - cost comparison and cost-benefit analysis,” Structure and Infrastructure Engineering, vol. 12, no. 5, pp. 603–617, 2016, doi: 10.1080/15732479.2015.1032983.
[2] P. Soderholm, M. Holmgren, and B. Klefsjo, “A process view of maintenance and its stakeholders,” Journal of Quality in Maintenance Engineering, vol. 13, no. 1, pp. 19–32, 2007, doi: 10.1108/13552510710735096.
[3] A. Abdullahi, N. A. Samsudin, M. R. Ibranhi, M. S. Arpin, S. K. A. Khalid, and Z. A. Othman, “Towards IR4.0 implementation in e-manufacturing: artificial intelligence application in steel plate fault detection,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 20, no. 1, pp. 430-436, Oct. 2020, doi: 10.11591/ijeecs.v20.i1.pp430-436.
[4] Q. Zeng, W. Liu, L. Wan, C. Wang, and K. Gao, “Maintenance strategy based on reliability analysis and FMEA: A case study for hydraulic cylinders of traditional excavators with ERRS,” Mathematical Problems in Engineering, vol. 2020, pp. 1-11, Nov. 2020, doi: 10.1155/2020/2908568.
[5] J. X. Liu, Z. Y. Jiao, F. X. Xian, and W. T. Liu, “Energy recovery and utilization system of excavator boom based on flow re-generation and balance theory,” Journal of the Brazilian Society of Mechanical Sciences and Engineering, vol. 42, no. 1, pp. 35-42, Dec. 2020, doi: 10.1007/s40430-019-2124-x.
[6] P. Hu, J. Zhu, J. Gong, D. Zhang, C. Liu, Y. Zhao, and Y. Guo, “Development of a comprehensive driving cycle for construction machinery used for energy recovery system evaluation: A case study of medium hydraulic excavators,” Mathematical Problems in Engineering, vol. 2021, pp. 1–13, Feb. 2021, doi: 10.1155/2021/8132878.
[7] W. Shen, J. Jiang, X. Su, and H. R. Karimi, “Parameter matching analysis of hydraulic hybrid excavators based on dynamic programming algorithm,” Journal of Applied Mathematics, vol. 2013, pp. 1–10, Nov. 2013, doi: 10.1155/2013/615608.
Automatic learning algorithm for troubleshooting in hydraulic machinery (Jose Castro-Puma)

José Castro-Puma is Bachelor of Mechanical and Electrical Engineering from the National Technological University of Lima Sur (UNTELS). Personnel trained for teamwork and under pressure. Basic English in the language center of the National Technological University of Lima Sur. Training in Gica Ingenieros in the maintenance area, did a maintenance internship as a mechanical assistant in the DS Global company, worked as an Equipment Assistant Engineer in component repair in the San Martin Contratistas Generales company for 2 years in the central workshop of Chilca, currently working as Administrative Assistant in the maintenance area - Trackless in the company Alpayana. He can be contacted at email: jacastrop3@hotmail.com.

Miguel Castro-Puma is Bachelor in Mechanical and Electrical Engineering from the National Technological University of Lima Sur (UNTELS). I did training in the maintenance area at the Gica Ingenieros center, I did my internship at the DS Global company as a Junior mechanic. And I currently work as a Planning Assistant at the San Martin General Contractors company at the Atocongo headquarters. He can be contacted at email: mecastrop@hotmail.com.

Verónica More-Sánchez is a regular professor at the Faculty of Economic Sciences of the Federico Villarreal National University. Professional in Economic Sciences with specialization in Corporate Finance and Administration and Social Management. Her research interests include finance, administration, social management, and education. She can be contacted at email: vmore@unfv.edu.pe.

Juana Marcos-Romero is a Doctor degree in Public Health and doctor in Education. Dean of the Faculty of Health Sciences in The Autónoma de Ica University. professor in research and Head of the Nursing Department of the Hospital Militar Central del Perú. She can be contacted at email: juana.marcos@autonomadeica.edu.pe.
Elio Huamán-Flores is Post-doctor in Didactics of Scientific Research, Doctor in Education, Master in Educational Administration and Bachelor of Education. Coordinator of Intellectual Production and Publications at the Autonomous of Ica university, his research interests include higher education, educational psychology, digital learning, health sciences and social sciences. He can be contacted at email: elio.huaman@autonomadeica.edu.pe.

Claudia Poma-García is a Doctor of Education, Cesar Vallejo University, “UCV” Lima, Peru. Master in Public Management Cesar Vallejo University, “UCV” Lima Peru. Economist Graduated from the Federico Villarreal National University, “UNFV” Lima Peru. College of Economists of Lima: Tuition code: 10049. He has studied a Diploma in Foreign Trade Operations, Safety and Health at Work. She currently works as a University Professor at the Cesar Vallejo University, and at the National University of Callao. She can be contacted at email: cpomagar@ucvvirtual.edu.pe.

Rufino Alcios-Ipanaque is a Doctor of Administration; Master in Education and management; Bachelor of Business Administration; collegiate, graduated with a master’s degree in strategic business administration; diploma in research and University teaching. Administrative audit; municipal regional public management; experience in administrative management in public and private, non-governmental companies. Experience in establishing, interacting and maintaining a good level of management and interpersonal relationships with clients, university teaching experience. He can be contacted at email: ralejosi@unac.edu.pe.