Machine Learning for Predictive Maintenance of Industrial Equipment

Miguel Clemente Correia
miguel.c.correia@ist.utl.pt

Instituto Superior Técnico, Lisboa, Portugal

November 2018

Abstract

The manufacturing industry is embarking on a progressive level of maturity concerning digitalization. Maintenance is a fundamental segment where competitive advantage can be acquired, through the implementation of quality and data-driven techniques. In manufacturing and especially in the present use case, quality standards are of major importance as they configure a level of difficulty to the global production process. Manufacturers have available a more advanced technology environment, looking at data as an insight generation tool. Currently, the machines are not prepared to offer an integrated perspective of the system, thus being challenging the identification of the root causes for the majority of the failures. The present thesis reports the work developed to address the requirements stated above. Equipment data acquires in this work, a unique interest while it is manipulated to develop an intelligent predictive model adapted to the industry environment. With this aim, it is proposed a novel method for integrating equipment data suppressing the need for a large number of variables and facing uncertainties in the system data flow. In addition, a framework for visualizing patterns and equipment interdependency is developed. The current study provides the manufacturer and the respective maintenance team an effective tool that can be integrated into the existing system providing insight and increasing the conformity with quality standards and production goals.

Keywords: Data-driven analysis, Machine learning, Failure prediction, MTBF, Manufacturing Industry.

1. Introduction

The industrial environment is changing and digital technology is the leading actor in what is considered as the emerging paradigm, Industry 4.0. The idealization of smart manufacturing is supported by the integration of “smart technologies” with standard manufacturing devices as sensors, machines and other equipment. One may affirm that the manufacturing is by itself adjusting to human needs and to the constraints of its supply chain. This fourth revolution can be materialized by attending to cyber-physical systems – a merge between the physical and the digital grades. An example of this systems can be found in the preventive maintenance area, where the condition of a physical equipment and the associated parameters are reflected in a Digital Twin [1]. The combination of artificial intelligence, big data, streaming analytics and machine learning, defines a powerful tool for the manufacturing environment. The maturity level of technology dictates the short number of documents regarding the application of predictive maintenance to the industrial context, the contribution of this research holds the following alignment of achievements:

- A tool for combining asymmetric data sets;
- An apparatus for visualizing patterns and interdependency of machine behaviour;
- An approach to overcome unknown variables;
- A data-driven machine learning classifier adapted to the manufacturing industry.

1.1. Background

The diversification of challenges faced by industries is leading maintenance job to constantly grow to a more mature state, taking advantage of technological evolution to redefine the strategies followed by maintenance teams.

Corrective maintenance is far known as the “only fault repair” approach and it is on the basis of the maintenance ideology. Despite being the earliest maintenance mode, this strategy it's deeply common in industries with reduced complexity and conservative culture. Taking into account the majority of industries, preventive maintenance is the endorsed method to address maintenance needs, performing a regular inspection at well-defined periods of time in order to prevent deterioration of equipment and possible loss of product quality generated by faulty components [2]. Bearing in mind a lack of data regarding the system operation and the absence of intelligence with the capability to process this information, it's acceptable to adopt the previous strategies. However, several companies are realising the true cost of maintenance and the potential they can achieve with a more adapted strategy. For this reason, predictive maintenance is the main hypothesis on this thesis, representing the concept that allows high reliability and enhances economic efficiency.

Predictive maintenance itself is not a replacement for the two methods previous presented, it's essential to have a combination between "run to failure" approach and preventive actions. In this blend of strategies, predictive procedures can complement the results by reducing the number of unexpected failures and identifying periods to apply specific maintenance tasks. This late concept is usually adopted as a tool for maintenance management aiming to prevent unscheduled downtime, as a plant optimization tool helping the definition of production procedures and parameters and also as a reliability improvement tool identifying the deviations in the operative specifications triggering operator actions preventing the potential failure or loss of quality in the production [3]. Usually, both intrusive and non-intrusive methods are used to complement each other, achieving superior results by combining them with maintenance and operational data to build a device capable of estimate equipment lifetime and plan maintenance unexpected activities [4].

Predictive maintenance models are typically developed by considering machine learning techniques as data mining and machine learning classification. In the maintenance environment, the aim of supervised learning is usually to predict a value or a label of one variable such us time to failure or the end of life of a component, and it can be achieved by knowing the values of other variables (production, sensors data and other information). If the variables being predicted are not a value but a category, then the problem is described as classification [5].

Ensemble methods were used in the current thesis to develop a machine learning model for classification. These methods are learning algorithms that, from a set of classifiers, perform a weighted vote on their classification and then estimate a class to the new entries. In a simple manner, the ensemble classification combines the prediction of several estimators in a given algorithm, thus improving the robustness of a single classifier [6]. Contained in the category of ensemble methods are the forests of randomized trees. In the case of this methodology, a set of classifiers (decision trees) is created by introducing randomness in the classifier construction as detailed in the following illustration.

$$D = \begin{bmatrix} X_{A1} & X_{B1} & \cdots & y_1 \\ \vdots & \vdots & \vdots & \vdots \\ X_{AN} & X_{BN} & \cdots & y_N \end{bmatrix}$$

$$T_1 = \begin{bmatrix} X_{A12} & X_{B12} & \cdots & y_{12} \\ X_{A34} & X_{B34} & \vdots & y_{34} \\ \vdots & \vdots & \vdots & \vdots \\ X_{A72} & X_{B72} & \cdots & y_{72} \end{bmatrix} T_2 = \begin{bmatrix} X_{A24} & X_{B24} & \cdots & y_{24} \\ X_{A38} & X_{B38} & \vdots & y_{38} \\ \vdots & \vdots & \vdots & \vdots \\ X_{A47} & X_{B47} & \cdots & y_{47} \end{bmatrix} \quad T_3 = \begin{bmatrix} X_{A15} & X_{B15} & \cdots & y_{15} \\ X_{A54} & X_{B54} & \vdots & y_{54} \\ \vdots & \vdots & \vdots & \vdots \\ X_{A62} & X_{B62} & \cdots & y_{62} \end{bmatrix} \quad \cdots$$

Figure 1 | Representation of forests of randomized trees in machine learning classification.

In Figure 1, D is the learning data set fed to the algorithm, while T_1, T_2, T_3 are the randomly generated decision trees. X and y are, respectively, the features and the associated output class. In its learning stage, the algorithm adjusts its parameters to learn from the learning data set to classify the output class y.

On this thesis, predictive maintenance converges with condition base monitoring, applying an intelligent system which can identify patterns and predict potential failures based on progression of the equipment status, operating conditions and maintenance team experience. This approach can be adapted to systems without a substantial volume of information and it's also qualified for the initial stage of equipment's lifetime in which the number of failures is higher, usually characterized by the difficulty in identifying potential problems in the quality control inspections. Therefore, using technics that are non-destructive and adaptable to the installed system, it's possible to take actions to bypass the potential failure of an equipment improving one of the most important performance indicators, the Overall Equipment Efficiency.

2. System Description

The current thesis comprises a study of an industrial process to produce a solid light weighted product subjected to strictly regulated quality standards.

Given the complexity of the global process and all its variables, it was considered data of the two machines with the largest impact on overall downtime. From now on they will be referred as Maker and Packer, respectively the machine that constructs the product by assembling all its components and the one that groups the individual products, also referred as Units, in a pack. As Figure 2 suggests, Maker is fed by product's main component and other secondary constituents. The first element, β, is primarily processed and experiences several treatment steps in the initial stage of the global process, conceiving from the raw material a highly standardized substance with well-defined characteristics. However, particles that compose this component have variable size, weight and properties. Taking as an example, two particles that have experienced equal conditions of processual treatment can vary in ductility, density, and shape. This is mainly due to feedstock traits and confers on the product a blend of properties that gives its unique characteristics. Reasonably, in the system Maker-Packer, the flow is not direct and accounts with a large buffer that can sustain about 15 minutes of continuous packer consumption. Considering that both machines are programmed to produce 14000 units per minute, the Buffer is estimated to hold 200000 units at average capacity. In the Buffer, the individual units have some freedom of movement that could enable the loss of material and moisture, minor deformation and accentuation of previous assembling flaws, hence interfering with the proper functioning of the system.

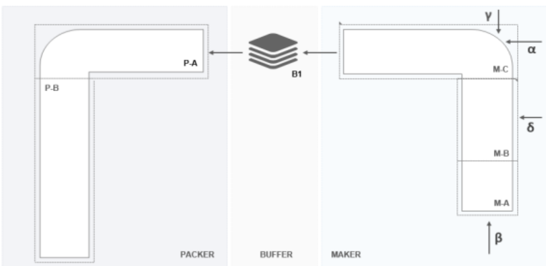

Figure 2 | Process diagram of target section compromising both machines addressed in the study (Maker and Packer) and the intermediate Buffer. Representation of intake areas.

Maker can be subdivided into three stages:

M-A. Admission and distribution of feedstock;
M-B. Construction of preliminary product;
M-C. Formation of the final product (Product A) and the introduction of finishing steps crucial to grant desired properties.

After the individual unit goes through the buffer it's then forwarded to Packer intake zone where it's reorganized along with other units to fit the package geometry. Later, the group of units is directed to the second stage of the machine designed to build the pack. It is possible to define two stages for this operation:

P-A. Admission and reorganization of a set of units;
P-B. Construction of the pack (Product B) and coupling of secondary constituents.

Machines composing the System are current top performers and account with a robust rejecting system based on real-time measurement of well-defined parameters and assessment of product quality. In the case of product's properties doesn't meet quality standards, it will be rejected. It is also essential to bear in mind that machines are programmed to stop if the number of rejected pieces overcome a defined threshold that could vary dependently on the equipment.

2.1. Data characterization

It was considered as sample a historical time span of 4 months that represents 100 days of production. Data were retrieved from a framework that integrates sensors from all connected equipment and provides aggregated operational information. Without treatment, the information used from both machines should be analysed as two independent datasets as a result of an asymmetrical temporal domain. The following conjectures are a consequence of the way data was generated from the system and are valid for both machines:

- Machine events data are a continuous data set grouped in time periods by machine status;
- Each row represents an alteration of machine status from the precedent row, thus a new event;
- Data is arranged in descending order of time;
- No major maintenance operations were performed on the machines;
- No manipulation has altered the data, thus providing a plain exportation of measured variables.

Three status can be identified in the machines as a typical sequence in the data set. *Starting* always follows a stop or the begging of a production, and it is proceeded by *Normal run* after an average time of 30 seconds. Notwithstanding *Ramp-down* status being essential to complete the stop procedure, it isn't necessarily true that it will occur. More severe circumstances usually determine an immediate stop of the machine, independently on the status (*Starting* or *Normal run*). The integration of the two data sets constitutes the first challenge engaged in this thesis as represented in the following illustration:

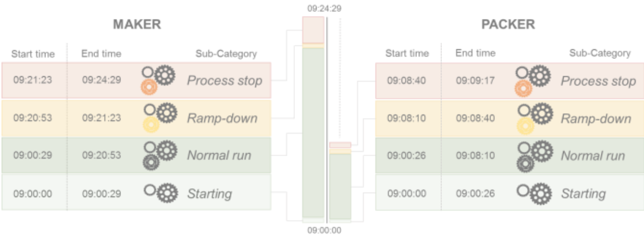

Figure 3 | Transposition of an operational example into the data set structure.

Bearing in mind that each row in the data set characterizes the machine status in each time span, it's correct to assume that each coloured block represents an entry. Therefore, as Figure 3 suggests, even considering that Maker and Packer start their production simultaneously, the periods of one machine does not match necessarily the time windows of the other. From now on, the presented blocks will be referred as Events as they are the reflection on the data of the events occurred during the operational time.

A crucial part of the present thesis concerns the expert knowledge of the people whose daily responsibilities comprise direct contact with the system in study. Operators experience is essential to understand raw data from the machine as they have the expertise to transpose the information recorded into concrete day-to-day situations.

2.2. System stops analysis

The study was centralized in the product that presents higher production, necessarily having more historical data to work on. In addition, its production has the lowest MTBF, a simple indicator that takes into account the total uptime and divides it by the number of stops. For this variety, the value of this indicator referred to the historical period was 19.0 and 6.73 minutes respectively for Maker and Packer. For the same operational time, lower MTBF implies a larger number of stops, thus representing higher variation in machine status that consequently generates more rows.

As previously mentioned, Packer's location in the process is highly sensitive to uniformities in feedstock properties. Given that this feedstock is a result of Maker's *Good Production,* it is fundamental to describe the behaviour of the making machine. With this aim, two dimensions were considered:

- Average rejected units by stop: Number of rejected units that a stop by a given *Stop Reason* rejects in average, providing an estimation on the extent of the damage motivated by the failure.
- Average duration by stop: The amount of time spent by the operators to fix the failure and restore the operation.

The indicators above were determined by equations (1) and (2), respectively.

$$\textit{Average Rejected Units by Stop of Stop Reason X (Units/stop)} = \frac{\sum_{x=0}^{x=n} \textit{Rejected Units}}{n} \quad (1)$$

$$\textit{Average Duration by Stop of Stop Reason X (Units/stop)} = \frac{\sum_{x=0}^{x=n} \textit{Duration}}{n} \quad (2)$$

The following diagram supports the distribution of failures in the two machines granting an intuitive perspective of most distressed areas. Notwithstanding M-B and M-C accounting with a large number of stops, Packer's location P-A1 accounts with more than 50% of all system failures. This section presents a peculiar set of characteristics related to machine stoppages as they have a short average duration of 67 seconds, meaning that an operator is able to quickly identify and mitigate the disturbance.

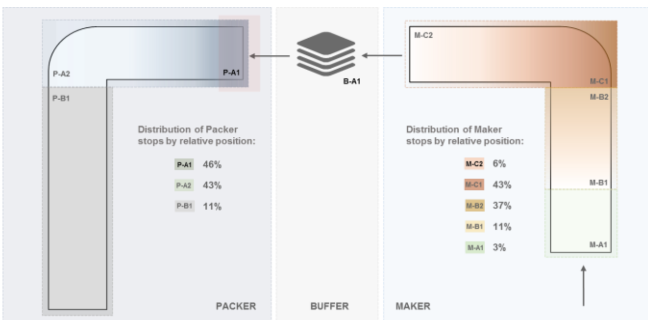

Figure 4 | Process diagram of target section embedded with stop analysis based on relative position. P-A1 is marked in red as the target area.

In the industrial environment, when an obstacle is easily managed, it is typically comprehended as a daily reality, underestimating the impact that can have in the production and in the activity of the operator.
Nonetheless, given its high number of occurrences, these particular failures have an extensive repercussion on the global downtime - close to 2 weeks for each year. In addition, as previously mentioned, one stop can generate a stage of *Ramp-down* and necessarily a *Starting* stage. The two phases of the operation are moments of low production rate and perform a total of 9 days without taking advantage of the capability of the machine in its target speed. P-A1 portrays a bottleneck, where hundreds of Units are forwarded to fit in a space for small dozens in a high-speed performance machine. These circumstances, along with uniformity of feedstock Units, compose the ideal environment for one complication - the blockages. According to the above line of reasoning, P-A1 is an evident choice for the target area giving that:

- Failures in that section are frequent;
- Blockage is mostly motivated by uniformities in feedstock Units;
- Operators have a clear knowledge of the actions they should take to restore the normal operation;
- Total downtime is heavily affected by PA-1 blockages.

The following line-up provides a description of the main challenges addressed considering the focused area:

- Machines' data sets are asymmetric in time;
- Lack of information on sensors data or other variables;
- Unworkability on buffer size estimations;
- High rate of failures near P-A1;
- Defective maker products can also perturbate other sections.

3. Approach

Engaging the consequences of blockages identified as target stops and considering their occurrence is in the intake zone of Packer, it is fundamental to examine preceding variables as potential motivators. The integration of two asymmetric data sets is the primary challenge addressed, settled in consecutive concepts:

1. Sequence Windows;
2. *Apriori* Algorithm;
3. Support Matrix.

The mechanism developed provides an innovative approach to reproduce insights from machine data adapted to overcome conditions of lack of information on independent systems.

3.1. Sequence Windows

Pattern Discovery is embedded in the definition of Data Mining, as it is the detection of similar structures on large data sets. The largest is the data set, more likely it is of having a high content of data distortion and uninteresting patterns, becoming impractical to apply simple statistics on data and consequently paving the way to data science through machine learning and other advanced computer techniques. Notwithstanding, deciding whether the patterns found are or not relevant and pertinent to the circumstances, should take into account the operational context and knowledge from experienced people whose function is directly associated with the subject [7].
As previously mentioned, information of each machine is independent and it is a result of the evolution in the operational status. Accordingly, Maker has events in different time spans than Packer, what makes inconceivable the aggregation of two dimensions: *Start-time* and *End-time.* To conveniently associate this information only the first one was considered. However, this assumption originates a disadvantage: without a timespan, the association of variables is not possible. This limiting opportunity makes attainable to identify patterns in the sequence of Maker events and in the interdependency between equipment that comprise the system. Data from the two data sets was aggregated and organized in descendant order of *Start-time.*
That said, it is appropriate to assume that there is a sequential set of events with different *Stop Reasons* forming a chronological sequence that illustrates the behaviour of both machines regarding their stops.

Figure 5 | Illustration of the application of Sequence Windows technique.

Intending to identify patterns in the sequence of events, the methodology described by *Vilalta et al* [8] references a similar approach as illustrated by Figure 5. The target stops define the beginning of the Lookup Window, a well-defined time span proposed to determine what events occurred from the *Start-time* of the target event until the edge of the window. In the developed concept, the proposed methodology was adapted to engage the thematic of equipment interdependency. As depicted in the representation, the objective was to set a window with a fixed size immediately before a target stop with the ability to identify the preceding Maker events during the given time span. The output contains crucial information to identify patterns in the interdependency of the two machines, as it expresses every event that happened in Maker in a well-defined period preceding Packer target stops, composing the input for the next topic.

3.2. Apriori Algorithm

This method is widely adopted to perform temporal data mining and is known for identifying association rules between frequent events and also providing statistical indicators, as the support value [9].

Support is directly related to the frequency that two events occur in the data set.

$$s\,(A \;=>\; B) = P\,(AB) = \frac{N(AB)}{|D|} \quad (3)$$

Where $N(AB)$ is the number of times that A and B are simultaneously present in a transaction being $|D|$ the universe of all transactions present in the data set [10].

To better understand the fundaments of *Apriori* it is essential to recognize the following concepts:

- Transaction: Set of events associated with a time. Each row in the input data set is a transaction;
- Item: Unit of a transaction. Each event is an item;
- Support (s): Percentage of transactions containing a given item;
- Rule: The rule $-1 => 1000$ with $\boldsymbol{s} = 10\%$ refers that the event with ID -1 occurs before the target event in 10% of the transactions in the data set.

3.3. Support Matrix

Considering the 56 *Stop Reason* that can induce Maker to a stop, there is the same number of rules that can be defined with the employment of the *Apriori* algorithm.

Nonetheless, it is important to be aware that Maker and Packer are separate machines and often present failures that are not related to each other. Therefore, it becomes evident that not all the derived rules are relevant to the analysis and it is essential to consider a threshold of minimum support, s, with the purpose of obtaining only rules that are sufficiently supported by this statistical indicator. Insights from experienced personnel play also an important role deciphering each rule with an effective support. This action provides fundament to the rule, complementing the output with transcription of the operational situation.

The Support Matrix is the combination of the two concepts elucidated in the present chapter (Sequence Windows and *Apriori* Algorithm) and it represents one of the innovative aspects conceived in the formulation of the current work. It is designed to address the adversities on the integration of two asymmetric large data sets and the analysis of the interdependence existing between equipment.

Depicted in Figure 6, the matrix is composed by two axes (size of Lookup Window and ID of Maker Event) and it provides an intuitive perspective of the rules between a pair of items, the target event and the Maker Event. The stronger the correlation, highest is the support and darker the colour set on the respective field. The interpretation is done by considering the concept of Lookup Window detailed earlier and it follows the logic exemplified by the following outcomes:

- The *Stop Reason ID* -2 occurs in 30% of the transactions generated with a Lookup Window of 16 minutes;
- Support associated with *Stop Reason ID* -1 only becomes significant when the Lookup Window is greater than or equal to 17 minutes;
- This set of ten types of Maker Events presents the failures that can occur in Maker and have a significant impact in Packer's functionality given variable time spans from 1 to 20 minutes

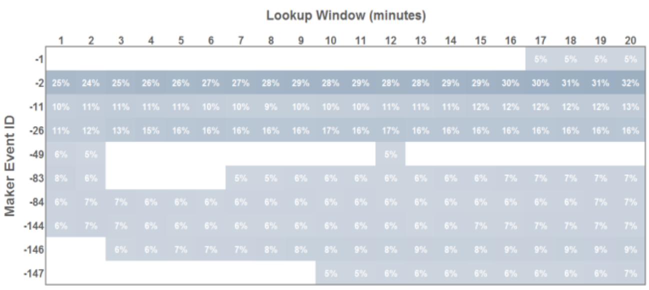

Figure 6 | Support Matrix resultant of the mechanism developed for the integration of both data sets.

Note that minimum threshold was set to a support of 5%, meaning that the set of Maker Events comprises only the ones with support greater than this threshold in at least one of the Lookup Windows.
Support Matrix is, for this reason, a powerful tool in the identification of patterns and correlations between events. One of its main advantages is the versatility shown dealing with incompatible data sets, only requiring one common variable related to time. It provides insightful information on equipment behaviour, creating relations among events, pointing out the statistical strength of the relation.
In the current hypothesis, Support Matrix has another decisive purpose. Its output, the set of Maker Events with a potential association with Packer's target stops, plays an essential role in the following section, which comprises the conception of the intelligent model developed to engage the problem. The performance of the model is highly influenced by the process of variables and features selection, important to build faster and more cost-effective predictors facilitating data understanding and reducing training times.

4. Model development

From operations to process, the application of machine learning systems in the manufacturing industry has been a reality. The integration of this intelligent systems capacitates the manufacturer to understand patterns and generate insights from large data sets. As previously described, supervised learning is related to the prediction of a target function. This form of machine learning is often used in the manufacturing environment as the majority of problems can be divided into two categories: predicting a quantity or predicting a category.

4.1. Projected Purpose

Developing a machine learning model should invariably begin by the definition of the goals and capabilities that are expected from it. In the current case, the mechanism must address a set of situations widely known in the manufacturing environment:

- Handle noisy data comprising outliers;
- Process large data sets from more than one source;
- Generate perceptible insights;
- Provide a positive impact on production;
- Possibility of scaling the model.

Bearing in mind the analysis taken on the previous chapter and the points stated above, it becomes evident that the problem minded in the current thesis can effectively be handled by applying an estimator for Predicting a Category, also referred as a classifier.
The target stops are a constant reality in the daily operation of the studied line. Nonetheless, in these failures, the job of searching for the defective location and resetting the normal behaviour is done in an effective way. As the operators are familiarized with these blockages, it only takes about one minute to solve them. On the other hand, the impact the target stops have on overall downtime makes inevitable to attend them. It's not only the production that is affected by the number of times these blockages occur, but also the operators' performance should be considered. Looking from the operators' point of view, the blockages occur randomly and with high frequency, interrupting any task that is being performed. Bearing this in mind, it becomes evident that the goal of the machine learning model was to provide an estimation of when a target stop would occur. Transposing from the operation to the model, it means that the conceived classifier should predict an event based on the information from its variables. In the operation, the same trained classifier should be capable of process real-time data and provide a warning when a blockage is predicted.

4.2. Feedstock Data

A typical classification problem usually requires the data to be labelled in a binary form. Considering the present use case, the classifier should correctly identify a Packer blockage in the target area from all set of events, therefore data labelling was performed by attributing 0 or 1 to the set of events (y).
In simple terms, this model is designed for correctly classifying y, given the related attributes X. The events were binarized considering that the target stops are the positive results and the universe of all other stops are the null perspective, as it follows:

Table 1 | Identification of labelled events.

Stop Reason ID	Label	Universe
$y = 1000$	**1**	Target stops
$y > 0 \land y \neq 1000$	**0**	Remaining Packer stops

Therefore, it was considered the following alignment of general variables, given their solid rationalization:

- Packer-Average speed before failure.
- Packer-Total Product Produced before failure.
- Packer-Rejected Production before failure.
- Maker-Total number of rejected Units.

4.3. Time Dimensionality

Without exception, in every engineering problem, time is a reality that cannot be suppressed and often defines the designed approach to engage it. From the beginning to the end, this variable is an absolute entity, present in the Sequence Windows, the construction of the Support Matrix and as further addressed in this chapter, in the model development.
At average capacity, Buffer holds 15 minutes of production, meaning that, when leaving Maker, a Unit takes 15 minutes to get to the target area. Instead of estimating this value based on the difference between the Packer consumption rate and Maker Good Production rate, another approach was developed considering the downtime of the equipment. In the case of Maker presenting a higher downtime than the consecutive machine, the Buffer will drain out faster as Packer is consuming more Units than what are being produced by the subsequent machine. Relying on the earlier declaration, the logic of Lookup Tier was built based on the estimation of the Buffer's size. For each output variable (y), in a time span of 20 minutes preceding that stop, the downtime of both machines was determined and the following relation applied:

$$Downtime\ Packer\ -\ Downtime\ Maker\ (4)$$

As it is evident, the calculated difference, the larger is the Buffer size and the longer it will take for a Unit to course through this equipment. For this purpose, it was created the concept of Lookup Tier – a define time span that starts in $\boldsymbol{t+x_s}$ minutes from the Packer stop $(\boldsymbol{t})$ and ends in $\boldsymbol{t+x_e}$. The conception of tiers of analysis is described by Table 2.

Table 2 | Lookup Tiers for the estimation of Buffer size.

Tier	Downtime Difference (minutes)		Lookup Tier Boundaries (minutes)	
	Minimu	Maximum	Start (x_s)	End (x_e)
1	< -30	- 30	0	5
2	-30	-20	5	10
3	-20	-10	10	15
4	-10	10	15	20
5	10	20	20	30
6	20	30	30	45
7	30	> 30	45	60

4.4. Classifier

As earlier described, *Ensemble Classifiers* holds a group of predictive algorithms developed with the same goal, provide improved generalizability and robustness to the estimator. Present in the context of this group is the *Extremely Randomized Trees* classifier, which is descendant of the widely known category of *Random Forests.* However, in the case of the first classifier (ET), the nodes are split by choosing cut-points fully at random reducing the variance induced in *Random Forests.* In addition, ET uses the all learning sample to grow the trees instead of a bootstrap replica [11].

5. Results

Composed by a set of metrics, the evaluation strategy considers two different perspectives: Data Science and Operational. It is important to bear in mind that the evaluation results are referred to the employment of the validation data set. This subset of the global data was not used in the construction of the model and, for this reason, is the appropriate structure of information to test the performance of the algorithm.

5.1. Data Science Perspective

The concept of probability threshold must be taken into account. In a simple binary classification (1 or 0) the classifier estimates the probability of the output variable being one of the classes considering the respective features. By default, if the probability of attributing class 1 is greater than 50% $(P(\boldsymbol{y}=1)>0.5)$, the algorithm associates the output variable to 1.
However, it is important to contextualize the performance of the algorithm given its application, because in some cases, the machine learning method could be stricter or softer when attributing a class depending on a threshold. Table 3 depicts the results of this application to the validation data set considering a threshold of 0.7.

Table 3 | Overall evaluation metrics.

Metric (Class = 1)	Value (%)
Sensitivity	78.8
Precision	84.3
Accuracy	76.9
F1 Score	81.4
FDR	15.6

5.1.1. ROC Curve

The area under the Receiver Operating Characteristic curve (AUC) depicts the ability of the algorithm to distinguish each class, thus a higher AUC means a better effectiveness when attributing the respective labels, known as separability.

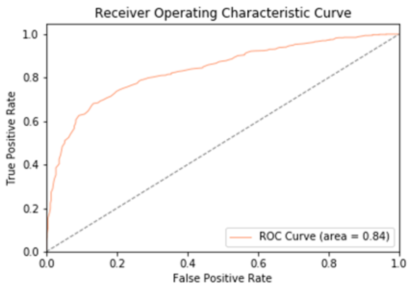

Figure 7 | Representation of ROC Curve and AUC.

5.1.2. Precision - Recall Curve

Precision – Recall Curve is often used to address the influence of imperfections in data such as skew and unbalance. Algorithms that optimize and prosper in the ROC Curve analysis are not guaranteed to perform well under the evaluation of Precision – Recall Curve.

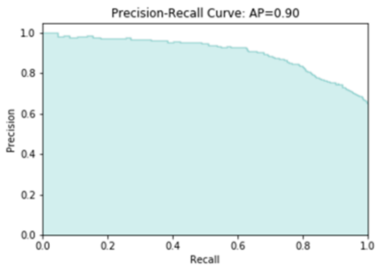

Figure 8 | Representation of Precision – Recall Curve.

5.1.3. TPR - FDR Curve

In manufacturing, it is crucial to be effective and simple when addressing a problem. For this reason, the False Discovery Rate (FDR) is a powerful indicator as it represents the proportion of false positives identified in a real context.

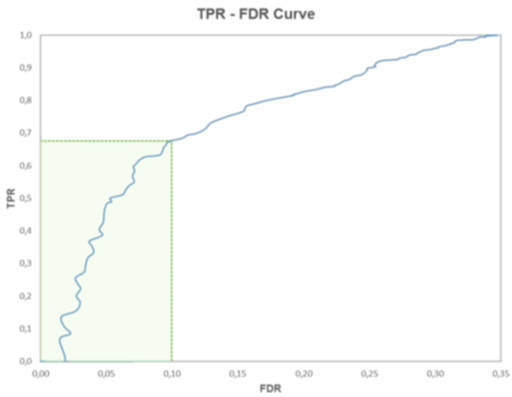

Figure 9 | Representation of TPR – FDR Curve.

A balance between the True Positive Rate and FDR must be taken by identifying the optimal operation point in the curve. In manufacturing, this optimal point is often located in the illustrated green area.

5.2. Operational Perspective

The following alignment presents the effects of the algorithm predictions considering a threshold of 0.78 with a 10% rate of false discoveries. It is essential to consider that the prediction is merely indicative providing the knowledge of when the stop is expected to occur. To reduce the blockages in the target area, is necessary an action from the operators, which can sometimes be not successful. It was considered that in 50% of the cases the experienced personnel can achieve to effectively by-pass the problem.
Therefore, the calculations regarding the achievable MTBF are contemplated in the following alignment and the table summarizing operational data of the packaging machine.

Table 4 | Summary of stats data from Packer in the historical period considered.

Uptime (hours)	1257
Total number of stops	11211
Historical MTBF (minutes)	6.73
Historical target stops	9486
Target stops predicted (threshold = 0.78)	6450
Achievable Uptime	1317
Achievable target stops	3225
Achievable MTBF	9.9
Achievable Increase in MTBF	47%

6. Conclusion

To address the machine learning problem engaged in the current work, two goals were defined. Primarily it should provide insights on machine behaviour that are understandable and can be transposed to operational environment and then the prediction target stops based on the production data of the making machine.

6.1. Discussion

These goals were attained in the stage of data processing and feature selection by developing algorithms capable of interpret the data. Based on statistical theories, it was conceived the Support Matrix tool. Promoted by *Apriori* algorithm and integrated with the logic of the Lookup Windows, this framework provided a set of outputs fundamental to pattern recognition.

In this case, the Support Matrix provided the set of Maker *Stop Reasons* with solid relation to Packer's target stops. The main challenges accomplished on the study are listed below:

1. Combine asymmetric data sets;
2. Retrieve Maker attributes, relevant to the model;
3. Produce a solid range of features from the few available variables;
4. Properly estimate the influence of Buffer by considering its size;
5. Determine the right classifier for an effective prediction of the target stops.

Expertise of operators and maintenance personnel was essential to define the features that were more representative of the operational reality. As it is unattainable the calculation of Buffer size, it is not possible to determine the amount of time spent by a Unit during this course. To overcome the stated obstacle, the concept of Lookup Tiers was introduced in the data preparation stage. Time is a fundamental dimension to be considered in problems of this nature and, in the current use case, represents a distinctive factor in feature generation to blockage prediction. The classifier that best fitted this purpose was the *Extremely Randomized Trees*, proposed by *Geurts et al* [11]. Testing data was applied to tune the model's hyper-parameters, while the validation data set was used to evaluate its performance. Justified by the context of the problem and the requirements for a low rate of false discoveries, the threshold of 0.78 allows a FDR lower than 10%, maintaining the TPR at 68%. From an Operational perspective, the reduction of target stops is the central objective. However, it is important to note that the algorithm only indicates the probability of the occurrence of these blockages and could not have a direct impact on machine performance. The aim is to alarm the operator so he can prevent the failure. Therefore, the analysis of the effectiveness of the algorithm in the industrial context could only be attained by assuming that only 50% of the identified stops could be avoided by the operator. Considering only the achievable reduction of target stops an increase in 47% was determined for the historical MTBF of Packer. The importance of adapting the data to the problem context and the insights from the experienced personnel is a considerable output regarding the application of machine learning in an operational environment.

6.2. Future work

Further analysis should be carried to identify possible changes that could benefit the work environment. In the author's perspective, the step further should be taken by integrating online data with the concept of Internet of Things applied to machine learning. The adoption of the suggested approach is vital to prove the integrity of this work, providing a wider knowledge concerning the predictive maintenance groundwork.

References

[1] H. Lasi, P. Fettke, H. G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," *Bus. Inf. Syst. Eng.*, vol. 6, no. 4, pp. 239–242, 2014.

[2] H. Wang, X. Ye, and M. Yin, "Study on Predictive Maintenance Strategy," *Int. J. u- e-Serv. Sci. Technol.*, vol. 9, no. 4, pp. 295–300, 2016.

[3] R. K. Mobley, "Impact of Maintenance," in *An Introduction to Predictive Maintenance (Second Edition)*, 2002, pp. 1–22.

[4] M. W. Gary R. Garrow, Charles P. Newton, III, Patrick E. Weir, P. West II David, "Performing predictive maintenance on equipment - Patent No. US 6,738,748 B2," 2009.

[5] T. A. Marques, S. T. Buckland, D. L. Borchers, E. Rexstad, and L. Thomas, "International Encyclopedia of Statistical Science- Distance Sampling," pp. 398–400, 2011.

[6] T. G. Dietterich, "Ensemble Methods in Machine Learning," in *Proceedings of the First International Workshop on Multiple Classifier Systems*, 2000, pp. 1–15.

[7] D. M. Hand and H. Smyth, "Principles of Data Mining," in *Principles of Data Mining*, vol. 30, no. 7, 2001, pp. 621–622.

[8] R. Vilalta and Sheng Ma, "Predicting rare events in temporal domains," *2002 IEEE Int. Conf. Data Mining, 2002. Proceedings.*, pp. 474–481, 2002.

[9] J. Nakamura, "Predicting Time to Failure of Industrial Machines with Temporal Data Mining," *Masters Sci. - Univ. Washingt.*, 2007.

[10] H. Road and S. Jose, "Fast Algorithms for Mining Association Rules," in *Proceedings of the 20th International Conference on Very Large Data Bases*, 1994, pp. 487–499.

[11] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, 2006.