26th International Congress of Mechanical Engineering
COBEM 2021
ABCM Associação Brasileira de Engenharia e Ciências Mecânicas
November 22-26, 2021, Virtual Congress, Brazil

COB-2021-0926
BAYESIAN NETWORKS AS PRODUCT CONFIGURATION SYSTEMS (PCSs) TO SELECT ELECTRIC MOTORS COMPONENTS

Bruno Ziegler Haselein
Gabriel da Silva Sezerino
Jonny Carlos da Silva
Federal University of Santa Catarina, Department of Mechanical Engineering, Florianopolis, SC, Brazil
haselein.bruno@gmail.com
gabriel7@weg.net
jonny.silva@ufsc.br

Abstract. *Electric motors are used in virtually any industrial application that requires conversion of electrical energy into mechanical energy due to their high efficiency. However, this product is highly customizable and has to comply with specific configurations according to the customer's needs. Together, these factors result in high demand for industries that design and manufacture the product. The configure-to-order approach allows the user to define the product configuration at the time of the order and the supplier to develop a product that meets the customer's needs. In this context of Design for Mass Customization (DfMC), the use of Product Configuration Systems (PCSs) is a decisive factor in obtaining the benefits of the mass customization approach due to improvements in knowledge management and control of product variants. The objective of this study is to evaluate the use of Bayesian Networks (BNs) as a PCS inference engine to select features of electric motors in a multinational-level company. These networks are considered one of the most effective theoretical models in the fields of knowledge representation and reasoning with uncertainty. The approach combines theory of probabilities and theory of graphs, and has a rigorous mathematical consistency that allows making inferences according to observations and based on the solid rules of probability calculation. Initially, two BNs structures are tested: the naïve Bayes, which uses a fixed structure and therefore dispenses any search engine, and the Tree Augmented Naïve Bayes (TAN), which limits its search space to a set of structures that corresponds to the Bayes naïve plus edges that form a tree whose nodes are the explanatory attributes. Next, the PC algorithm is used to define a general BN structure. This study uses recent data to learn the BNs structures and Conditional Probability Tables (CPTs), and the model evaluation method is cross-validation. The results indicate that it is possible to accurately predict in the proposed network structures the value of a characteristic, e.g., bearing size, given observations about other values. Nevertheless, the accuracy to identify some features, e.g., flange type, is highly sensitive to the network structure. The proposed approach allows the company to identify potential product variants and, consequently, to reduce costs and to better meet customer needs. This work is a preliminary study of a dynamic PCS, in which the characteristic values are automatically suggested to the user based on the probabilities calculated in BNs.*

Keywords: *product configuration systems, Bayesian networks, electric motors, design for mass customization*

1. INTRODUCTION

As a classic estimate of the effects of different phases of the product life cycle, Back *et al.* (2008) suggest that about 70% of the product cost is defined in the early stages of the project. This premise, which emphasizes the importance of assertiveness in decisions taken in the early stages of the product life cycle, implies that companies have to cope with the complexity of developing an organizational system capable of predicting and adapting their product or service to a future market in order to meet each client's specific needs. Accordingly, Design for Mass Customization (DfMC) is a well established way to introduce issues of economy, scope and scale in the early stages of a project (Fettermann and Echeveste, 2011).

DfMC has as principles the use of flexible manufacturing techniques and modularity practices to expand the development of a single product to a family of products, and to efficiently allow the participation of consumers in the development process (Jiao *et al.*, 2003). This modularity can be understood as a special form of design, in which, intentionally, a coupling is created between the product components through the standardization of interfaces (Maribondo 2002 apud Fleig, 2008). A product platform is defined as a set of common components, modules or parts that make up a large number of products (Meyer and Lehnerd, 1997). The possible compositions between the product platform and the modules/components provide a greater amount of product alternatives, and the set of these possible compositions forms the product family (Fettermann and Echeveste, 2011).

The Configure-To-Order (CTO) concept aims to allow the user to define the product configuration at the time of ordering and for the supplier to perform a configuration that meets the needs of the end customer. In this context of DfMC, the use of Product Configuration Systems (PCS) is a decisive factor to obtain the benefits of mass customization due to improvements in knowledge management and control of product variants (Myrodia *et al.*, 2017).

The target company, WEG Electrical Equipment S.A., is a multinational headquartered in Brazil and uses the SAP PCS to offer variants of its products, usually based on the modification of existing products (WEG Electrical Equipment S.A, 2021). In this PCS, 285 characteristics describe the product in a simplified way. After entering some initial values, such as market and product line, the PCS loads the product standard configuration, which is defined through rules established during the product platform development. Later, changes on this standard configuration can be made. Accordingly, the possible values for each feature are classified into standard, optional and special. At the end of the configuration process, special motors are those that have one or more special characteristic values.

Although constraints and dependencies between values must be considered to calculate the maximum number of configurations, it is possible to visualize the problem of combinatorial explosion. For example, at the time of this research, the categorical variable "commercial frame" has 2209 possible values. Furthermore, some variables are continuous, such as output rating. In this case, 5129 distinct values are registered on the database so far. Based on conservative assumptions, the estimated number of possible input combinations is $6{,}22 \times 10^{85}$ (Haselein e Silva, 2019).

In the target company, the Bill Of Material (BOM) is calculated through vertical search in the SAP database and auxiliary functions. Next, the interface between components is verified through an auxiliary expert system called BOM Quality Assistant (Massirer, 2007). Although the combination of all modules, i.e., sets of components and parts, is not feasible, the high degree of customization and the number of possible combinations between components provided by the DfMC approach enables the combinatorial explosion of possible products.

This paper aims to evaluate the applicability of Bayesian Networks (BNs) to identify correlations among characteristic values of the target company's PCS and to suggest configuration restrictions in order to avoid unnecessary material creation, i.e., a set of documents for an item or component. e.g., customer specification, casting mold drawings, CAD and CAM models, product certificates, etc.

The paper is organized as follows. This first section briefly presents the motivations for using BNs as a PCS inference engine. Section 2 summarizes several recent articles on the use of PCSs in industrial applications. Details of the product and mathematical methods are provided in Section 3, followed by tests and results obtained using distinct Bayesian network structures in Section 4. Finally, section 5 summarizes the lessons learned.

2. RELATED WORKS

PCS is a class of Knowledge-Based System (KBS) in which the aim is to convert customer's requirements into a product specification from which the company is able to develop the product (Trentin *et al*., 2013). KBSs developed in the target company are based on forward rules (Massirer, 2007) and data-mining concepts (Haselein and Silva, 2019). Commercial tools that adopt non-sequential programming paradigms are, for example, the Rete algorithm (Forgy, 1982) and the solutions created by the company Configit (Configit, 2021), both highlighting problems of sequential programming, such as duplicate or conflicting information, large number of rules, poor performance and high dependence on the order in which the rules are executed. In addition, Siemens have successfully used Constraint Satisfaction Problems (CSPs) technology for over 25 years (Falkner *et al*., 2016). Also in the work by Hendler *et al.* (2006), the advantages of CSP techniques in the context of product configuration as opposed to rules is evaluated. For example, it is possible to use algorithms to discover interactions not foreseen by experts. However, some issues still need to be addressed, such as how to represent the knowledge of configurations with high complexity, how to deal with user preferences and how to present explanations when unexpected outputs occur.

Belief networks provide several advantages to PCSs. This technique uses probability as a measure of uncertainty, and as information accumulates, knowledge of the true value of unobserved variables usually increases. Furthermore, this AI technique allows us to evaluate how variables from distinct dimensions affect the probability of an outcome (Khan *et al.*, 2018). Moreover, the authors state that data measured on different levels of accuracy can be combined, and these networks can be developed using an incremental and modular approach. Finally, compared to other approaches, their structure is less complex for better communication and explanation. Another useful aspect of BNs lies in the fact that by using Bayes's theorem, one can proceed not only from causes to consequences, but also deduce the probabilities of different causes given the consequences (bidirectional reasoning) (Le *et al.*, 2013). This technique allows combining different sources of knowledge (in our case, experts' knowledge and a database), and the learning mechanism of the network makes the assumptions made by the experts transparent and open to discussion (Uusitalo, 2007). They are an excellent tool for decision analysis because they are able to explain why and how each output was calculated and selected, and because belief networks are solved analytically, they can provide fast responses to queries (Uusitalo, 2007).

There are some challenges on the use of BNs that have to be addressed. The first challenge is the discretization of continuous variables. The usual solution is to discretize the variables and build the model over the discrete domain.

Belief networks can, however, deal with continuous variables using other techniques. While belief models are a useful way of modelling experts' knowledge, it may prove difficult to collect and structure this knowledge. In addition, it may be difficult to get the knowledge out of the experts in a form that can be converted into probability distributions (Uusitalo, 2007). Fortunately, we have a database with a large number of projects, and these distributions will be obtained through mathematical methods

3. MATERIALS AND METHODS

The scope of the study presented in this document is limited to evaluating the applicability of BNs to select one of the modules that compose an electric motor, the drive-end cover (Figure 1), which has enough complexity to serve as a proof of concept of our approach. This module has several parts, e.g., flange, oil seal, bearing, etc. and is described by a large number of features, e.g., flange external diameter, lubricant inlet angle, type of electrical insulation, etc. Moreover, these features are detailed in distinct formats, such as component drawings and datasheets, and are accessed, processed and changed by stakeholders throughout the project using applications such as the BOM Quality Assistant (Massirer, 2007) and SEVC (Haselein e Silva, 2019).

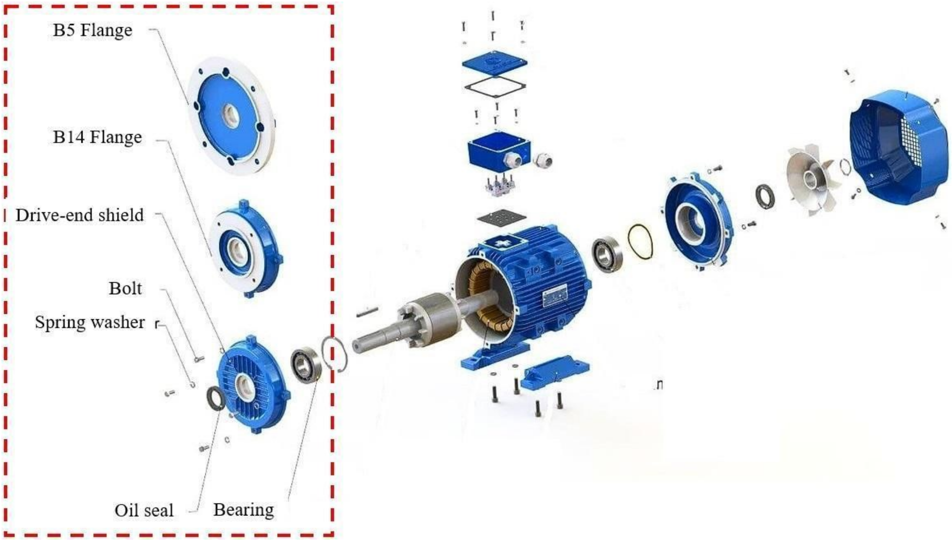

Figure 1. Drive-end shield (Electrical Equipment S.A., 2020).

A flanged end-shield is defined in a simplified way by 11 PCS characteristics identified by a domain expert, which are: *drive-end shield type*, *bearing size*, *bearing sealing*, *bearing cap*, *bearing thermal protection*, *greasing system*, *drain*, *flange type*, *flange size*, *frame group*, *vibration sensor*, and *able to grease inlet as-cast part*. However, it is important to highlight that this is a simplified KBS, which will serve as a base model to be incrementally improved. As an example, Table 1 shows the instance of the flange material ID Z13342738.

Table 1. Characteristics and values of a flange cover.

Feature	Value
Material number	Z13342738
End shield type	Drive-end end shield
Frame group	Frame 100
Flange type	Flange C next size up
Flange size	FC-184
Bearing sealing	Lip seal
Vibration sensor	Without sensor
Bearing thermal protection	Without bearing protection
Bearing cap	With bearing cap
Greasing system	With grease fittings
Drain	Threaded drain plug
Bearing size	6206
Able to grease inlet as-cast part	Yes

The database (DB) is composed of more than 2000 recent projects of drive-end cover and, as Table 1 suggests, only categorical data is used. Therefore, we can make the most of the benefits provided by probabilistic KBS, which are discussed in Section 2. It is important to notice, however, that some values are relatively rare. For example, in only 0.83% of the cases the *flange size* is "FC 95" and in 0.14% the *bearing size* is "7309".

One important concept that has to be addressed is the distinction between cast-items and machined-items. In this context, several distinct machined-items can be generated from the same cast-item through machining processes. This distinction will be further explained in Section 4.3. In our project, there are 70 distinct cast-items in the DB.

3.1 Bayesian Networks (BNs)

Bayesian Networks (BNs) are considered one of the most effective theoretical models in the fields of knowledge representation and reasoning with uncertainty. The approach combines theory of probabilities and theory of graphs, and has a rigorous mathematical consistency that allows making inferences according to observations and based on the solid rules of probability calculation (Pearl, 1988).

Figure 2a exemplifies the naive Bayes network structure (Madden, 2009), which uses a fixed structure and therefore does not require any search engine to define its structure, i.e., one node (X_C) is the only parent of all remaining nodes and there are no other connections. Therefore, this structure is a model that assumes that the features X_1, X_2, X_3 and X_4 are conditionally independent, which leads to inaccuracies if this assumption is not true.

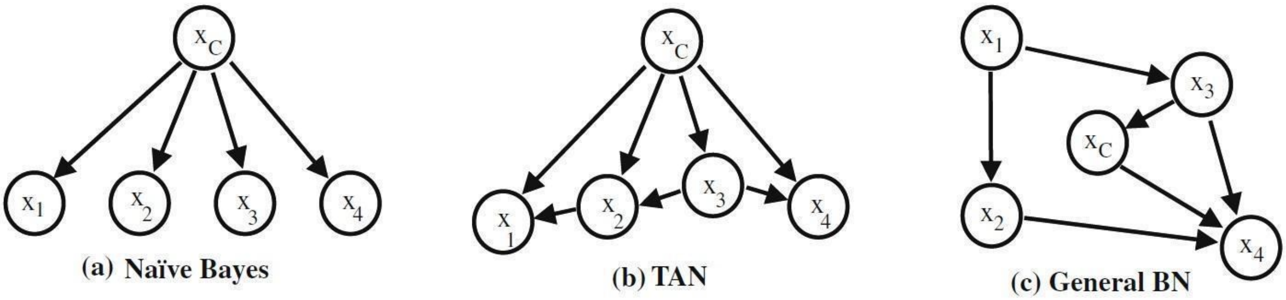

Figure 2. Illustration of naïve Bayes, TAN and general network structures (Madden, 2009).

The Tree Augmented Naïve Bayes (TAN), exemplified in Figure 2b (Madden, 2009) limits the search space to a set of structures that correspond to the naïve Bayes structure plus edges that form a tree whose nodes are the explanatory attributes. Therefore, this structure also defines one node (X_C) as parent of all remaining nodes. However, in contrast to the naïve structure, the TAN does not assume that the variables are independent.

The General BN (GBN) presented in Figure 2c does not defines one node (X_C) as parent of all remaining nodes. Therefore, this structure relaxes even more the assumption made on the naïve and TAN structures learning process. There are several algorithms to learn a GBN based on data, such as K2, Greedy Thick Thinning (GTT) and Bayes search. In our research, a constraint-based algorithm called PC was selected to learn the GBN structure (Tsagris, 2018).

The Conditional Probability Tables (CPTs) are calculated after the structure learning. The Expected-Maximization (EM) algorithm was selected to solve the Maximum-Likelihood Estimation (MLE) problem and to estimate the parameters of the statistical model given data (Mouafo *et al.*, 2016).

In BNs, there are several approaches to define the value of an unobserved feature according to observations on other features. Although some techniques that rely on thresholds and decision nodes may be evaluated in future research, in this preliminary study we chose to use the Most Probable Explanation (MPE) to simply select the value with the highest probability in each node. As discussed in section 4.2., this technique allows the user to visualize how much certainty the network has about each output.

We select well established algorithms in the literature to learn the BNs, and it is beyond the scope of this document to detail the mathematics underlying each approach. For more information about the aforementioned algorithms, we recommend the study published by Madden (2009), Tsagris (2018) and (Mouafo *et al.*, 2016). Finally, it is important to notice that in this proof of concept phase, the network structures and CPTs were calculated using the shell GeNIe Modeler, a Bayesian network development software (BayesFusion, 2020).

In order to evaluate the applicability of BNs to select features of electric motors components, the following steps were followed:

1. The naïve Bayes structure assumes independence between the child nodes. For this reason, this preliminary network structure will be used as a base model, which will be incrementally improved;
2. Elaborate a TAN structure and compare the accuracy and diagnostic value of its nodes with the results obtained with the naïve structure;
3. Create a general BN using the PC and EM algorithm to select features of electric motors drive-end cover;

4. The most appropriate model evaluation method is fold cross-validation (Madden, 2009), which divides the data into two subsets: training and testing. Similarly to the experiments proposed by Madden (2009), in our experiments we chose the k-fold cross validation with fold count = 10;
5. Use a general BN to select cast-moldings items;
6. Critically evaluate the advantages of BNs, e.g., to serve as a tool for decision analysis because they are able to explain why and how each output was calculated and selected, and challenges in our context.

4. RESULTS AND DISCUSSION

This chapter summarizes the main results obtained in the naïve, TAN and general network structures to select features and cast-items of electric motors drive-end cover.

4.1 Insights from the naïve Bayes and TAN structures

The elements from the network can be rank-ordered from the most to least informative given the current case (or current query) based on a parameter called Diagnostic Value (DV), also known as entropy reduction. Since the probabilistic reasoning within a belief network is induced by observing evidence, the DV of each node is updated according to new observations (or facts).

The DV is an excellent parameter to identify which are the elements of the PCS that will contribute the most to predict the value of an unknown feature. On the front-end, this information can be used to highlight through, for example, a color code, which is the next feature expected to be filled on the PCS by the user. It is important to mention that it is not mandatory to the user to follow this rank-ordered DV list. Nevertheless, if the user chooses to fill a feature that is not the one with highest DV, the DV values and the order of the features from the most to least informative given the current case will be updated normally.

Figure 3 presents the diagnostic value of the feature *bearing size* to observations in other nodes. Obviously, the DV of *bearing size* to an observation on the node *bearing size* is 100% and, for this reason, this value is not presented below. In this model, which uses a naïve structure, the network is highly responsive to observation on the nodes *frame group* (63, 71, 80, etc.) and *flange size* (C105, FC149, FF115, etc.), both features associated with the motor size, therefore a conclusion factual with design decisions, as the motor size is related to the mechanical loads supported by the bearings.

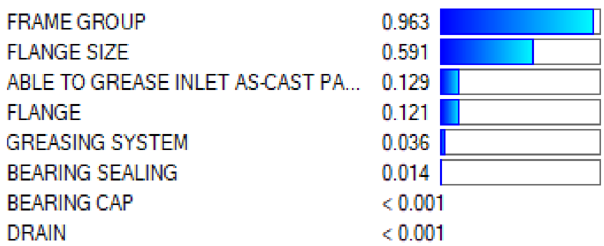

Figure 3. Diagnostic value (DV) of the feature *bearing size* to evidence in other nodes (naïve structure).

In contrast to the nodes *frame group* and *flange size*, *bearing size* (6201, 6202, 2308, etc.) does not depend on whether or not the end-shield has a *drain* or a *bearing cap*. Moreover, *bearing sealing* (v'ring, lip seal, oil seal, etc.) is also a highly independent node, with DV = 1.4%.

Table 2 presents the Accuracy (Acc) in predicting each of the nine features discussed in Section 3 using a naïve Bayes structure. It is noteworthy that two features selected by the domain expert are not evaluated in this experiment, as it was found that these features present only one possible value in the database. The results indicate whether it is possible to predict the value of some features given observations on the other values, i.e., *bearing size*, *able to grease inlet as-cast part* and *frame group* can be predicted with high accuracy. In contrast, *bearing cap*, *bearing sealing* and *drain* have an accuracy of less than 50%.

The DV of the three most relevant features to predict the value of each parent node in the naïve Bayes structure is presented in Table 2. It is observed that *frame group*, *flange size* and *bearing size* are very informative features due to their high DV in several experiments. In contrast, the features with the highest error rate, *bearing cap*, *bearing sealing* and *drain* are little sensitive to facts on other nodes, i.e., when one of these features is the parent node, the DV of the three most informative features is low.

Table 3 presents Acc and DV of the three most informative features in each TAN network elaborated in this experiment. Again, *bearing size*, *able to grease inlet as-cast part* and *frame group* are predicted with low error rate, while *bearing cap*, *bearing sealing* and *drain* have a high error rate. For these features, the accuracy would be better if we just guessed the mode in the database instead of using a naïve and TAN model. Nevertheless, *flange size* and *flange type* showed a significant reduction in the error rate in the TAN structure.

Table 2. Accuracy (Acc) and diagnostic value (DV) in the naïve structure.

Feature of interest	Acc (%)	1st feature	DV (%)	2nd feature	DV (%)	3rd feature	DV (%)
Frame group	100.00	Bearing size	96.5	Flange size	59.1	Able to grease	12.9
Bearing size	99.86	Frame group	96.3	Flange size	59.1	Able to grease	12.9
Able to grease	97.36	Flange size	62.4	Bearing size	56.5	Frame group	56.4
Flange size	83.47	Frame group	40.7	Bearing size	40.6	Flange	20.2
Flange	77.92	Flange size	60.5	Bearing size	3.6	Frame group	3.5
Greasing system	73.89	Able to grease	41.3	Flange size	22.9	Bearing size	19.8
Bearing cap	49.44	Bearing size	2.4	Frame group	2.3	Flange size	1.9
Drain	36.39	Bearing size	0.3	Bearing sealing	0.2	Frame group	0.2
Bearing sealing	13.06	Bearing size	0.7	Flange size	0.7	Frame group	0.6

Table 3. Accuracy (Acc) and diagnostic value (DV) in the TAN structure.

Feature of interest	Acc (%)	1st feature	DV (%)	2nd feature	DV (%)	3rd feature	DV (%)
Frame group	100.00	Bearing size	96.5	Flange size	51.9	Able to grease	11.1
Flange	100.00	Flange size	51.8	Bearing size	3.6	Frame group	3.1
Bearing size	99.86	Frame group	85.6	Flange size	57.4	Able to grease	12.9
Flange size	99.72	Frame group	39.3	Bearing size	36.7	Flange	18.2
Able to grease	99.17	Flange size	60.0	Frame group	52.0	Bearing size	51.9
Greasing system	68.61	Able to grease	41.3	Flange size	22.6	Bearing size	17.5
Bearing cap	26.25	Flange size	1.9	Frame group	1.8	Bearing size	1.7
Drain	17.8	Bearing size	0.2	Flange size	0.2	Frame group	0.2
Bearing sealing	4.03	Flange size	0.7	Bearing size	0.5	Frame group	0.4

When we loosen the connection constraints between child nodes imposed by the naïve structure, new links are created and the accuracy to predict the values of *flange size* and *flange type* increase. It is observed that the DV of the first and second most informative features for these parent nodes attenuated in the TAN structure. This phenomenon indicates that the correlation between network variables is of great importance in defining the output node value. Moreover, as shown in Figure 4, the TAN network corroborates with the observed correlation between *frame group*, *bearing size* and *flange size*. Interestingly, the connections between the children nodes are the same in both TAN structures, but the direction of some links have changed.

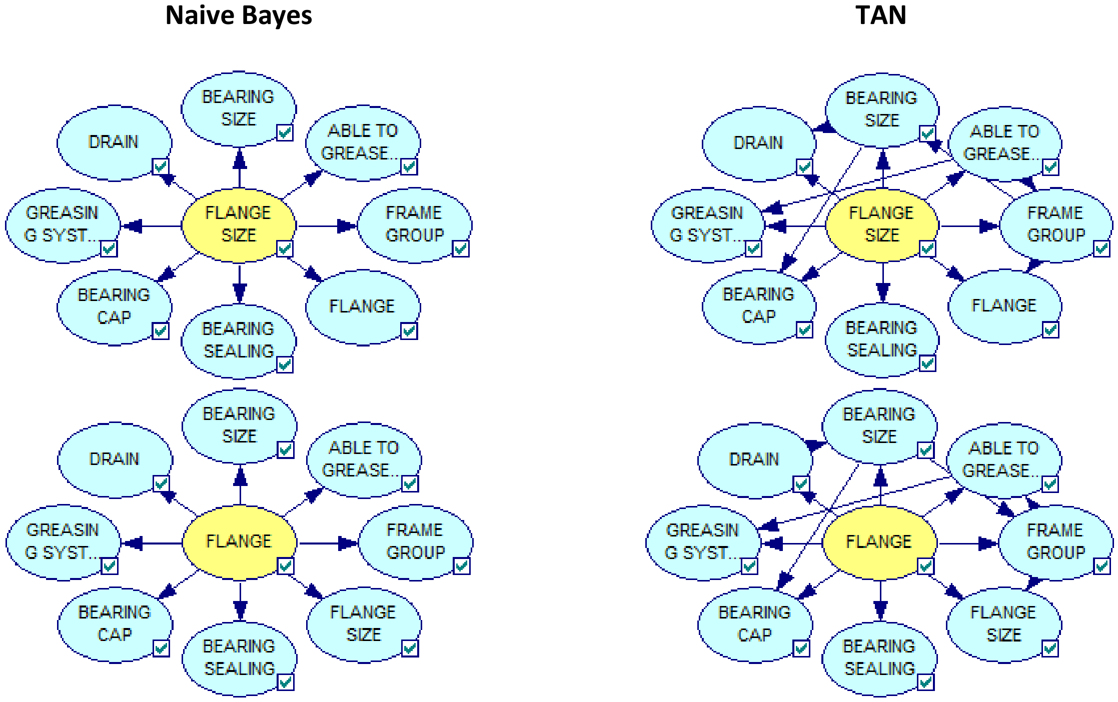

Figure 4. Naïve Bayes and TAN structure for the variables *flange size* and *flange*.

The results shown in the TAN structure makes sense from an engineering perspective. *Frame group*, *flange size* and *bearing size* are features that affect other features, such as *flange type*. In addition, it was expected to see links between these highly informative variables: as aforementioned, *frame group*, *flange size* and *bearing* depends on the mechanical loads. Furthermore, it is known that *bearing cap*, *bearing sealing* and *drain* are independent project variables.

4.2 A general BN for selecting features of electric motors drive-end cover

A general Bayesian network for selecting electric motor features is presented in Figure 5. This network computes the features discussed in section 4.1. However, in order to better illustrate the probabilistic reasoning embedded in BNs, this figure presents the probability (%) of each value in each node. For example, the probability of *bearing size* "6201" is 3%, *bearing size* "6202" is 8%, etc. In this figure, there are no facts introduced by the user, that is, there is no value with probability 100%. Nonetheless, as new facts are introduced, e.g., the user states that *bearing sealing* is "v'ring", the probability distribution in the other nodes will change. In this context, a potential application of BNs is dynamic PCS, in which characteristic values are automatically suggested to the user using probabilities calculated in BNs.

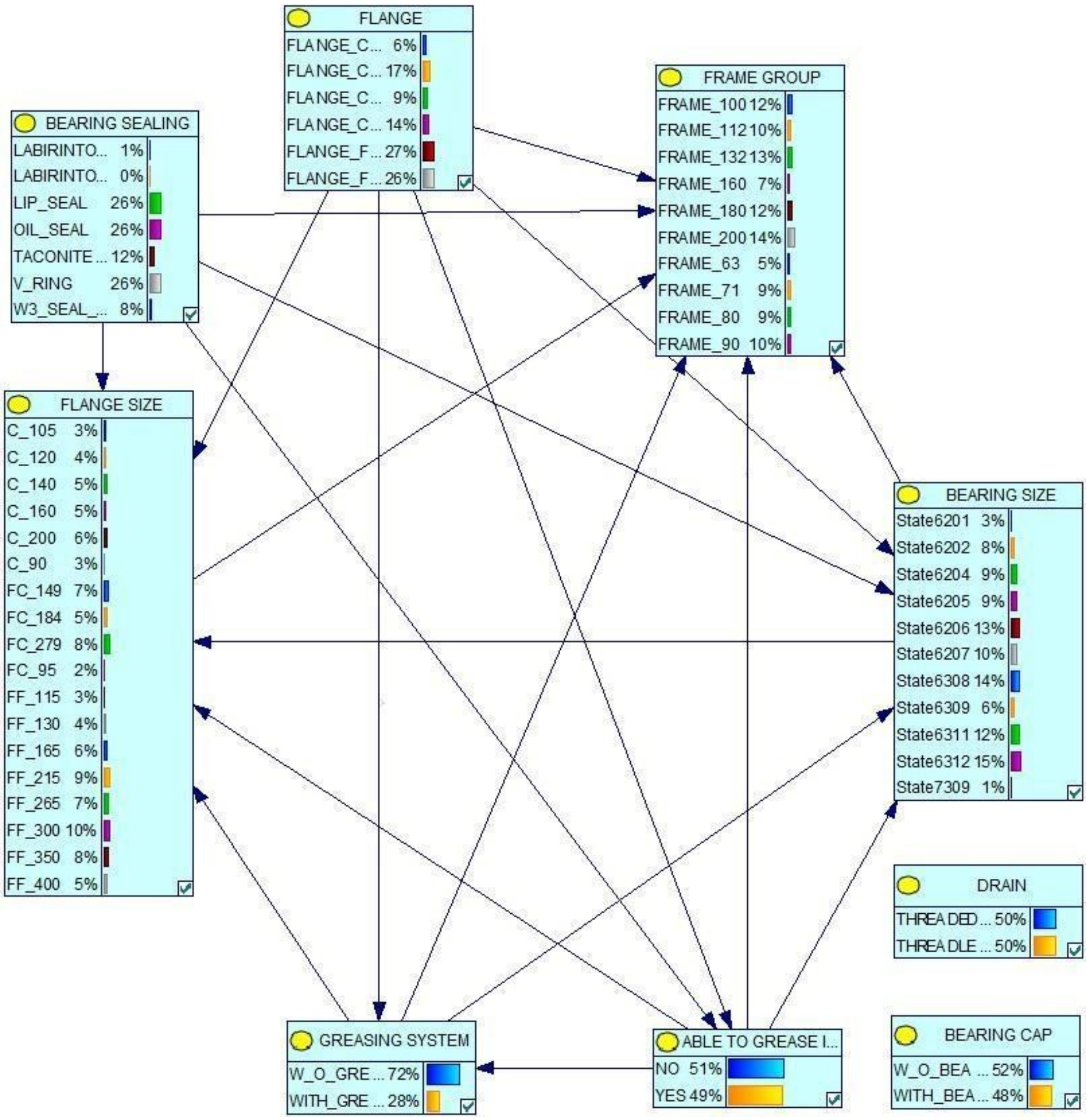

Figure 5. General BN to select features of flanges.

Figure 5 corroborates the insights obtained in section 4.1. The features *frame group*, *flange size* and *bearing size* have links with every node but *drain* and *bearing cap*, which are independent features.

The accuracy obtained in this experiment is similar to the accuracy obtained in the TAN structure, as shown in Figure 6. Additionally, this figure presents the DV of each feature. For example, *frame group* is the most valuable information to predict the *bearing size*, because its DV is 35.5%. It is important to notice that this figure presents only DVs greater than 0.1%. As we can see, although *bearing sealing* has connections with some nodes in the general BN, its DV is very low, which means that information about the *bearing sealing* state does little to define the value of other product features.

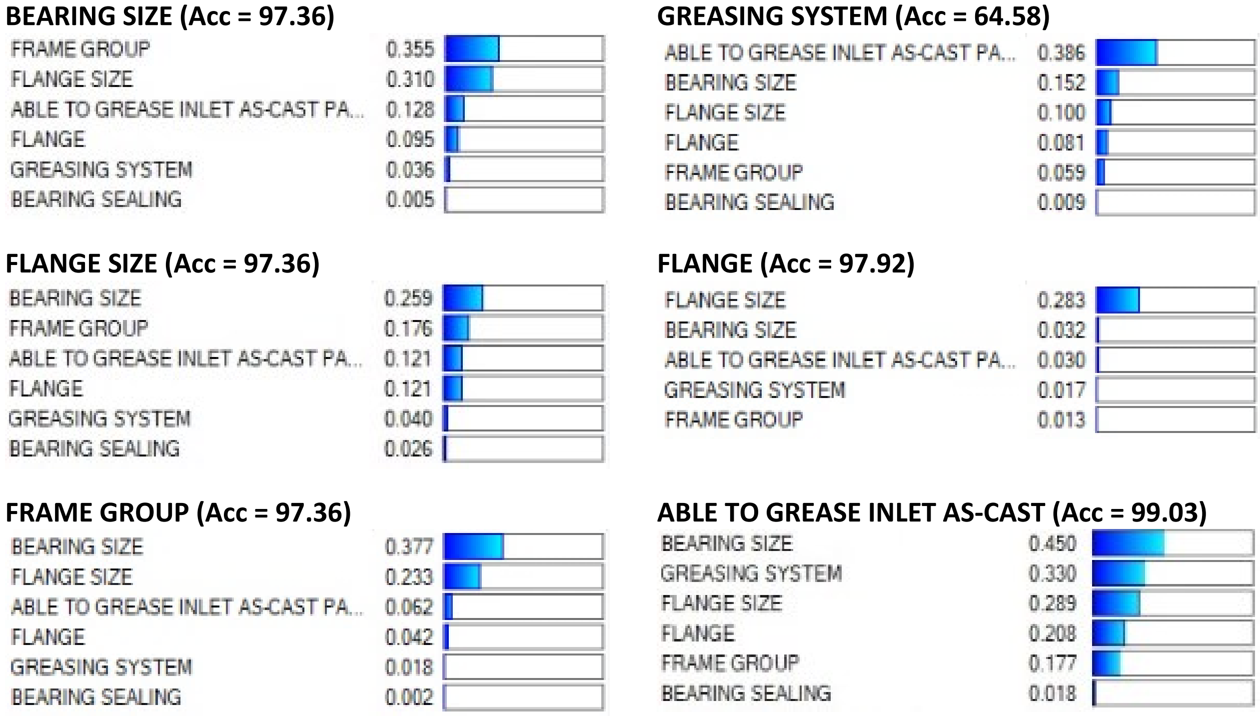

Figure 6. Ranked observations for each feature (GBN to select features of flanges).

4.3 A general BN to select materials (cast items) of electric motors drive-end cover

One of the objectives of this research is to evaluate the use of BNs to select materials of cast items of electric motors drive-end cover. As we can observe in the general BN presented in Figure 7, the node *as-cast item*, which contains the ID of 70 cast-items, connects with 8 of the 9 network nodes. Naturally, it is possible to present the probability (%) of each value in each node similarly to figure 5. However, in Figure 7 we present only the nodes due to the large number of values in the node *as-cast item*.

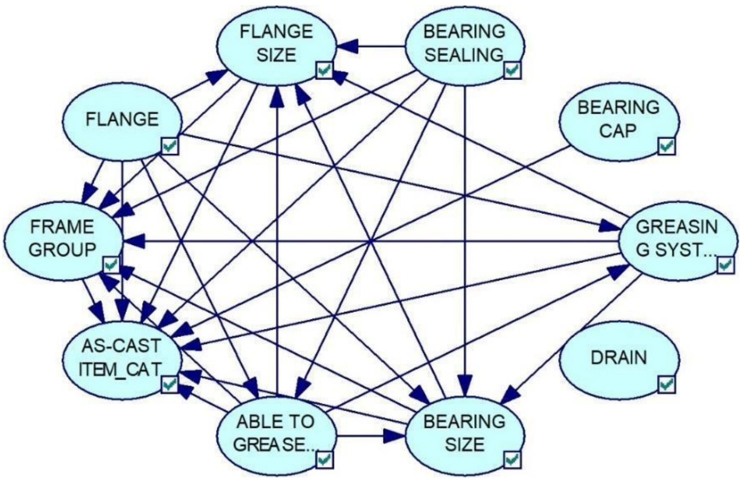

Figure 7. Resulting GBN to select materials of flanges.

This network achieved an accuracy of 83.61% in identifying cast items in the DB, an important output from an engineering perspective because it reduces the number of documents generated throughout the project and decreases lead-time. Additionally, another useful result of this BN is that now we are able to rank the features from the most to least informative to identify cast items. This information can be used to define the order in which the features are displayed in the PCS. As Figure 8 indicates, the most important feature is *frame group*, followed by *flange size*.

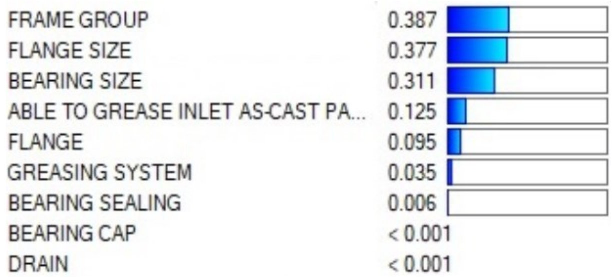

Figure 8. Ranked observations to select materials of flanges (Acc = 83.61%).

BNs are an excellent tool for decision analysis because they are able to explain why and how each output was selected, and the approach uses probability as a measure of uncertainty, as discussed in Section 4.2. Moreover, this white-box technique allows the user to visualize the confusion matrix, as shown in Table 4, and therefore to get insights about product variants. In this example, Table 4 presents the confusion matrix to predict the *bearing size*, which allows us to determine the Acc for each value of each feature. As we can see, errors occurred, for example, to identify if the bearing size is “6202” or “6204”.

Table 4. Bearing size: confusion matrix.

Predicted						
6201	6202	6204	6205	6206	**…**	**Actual**
2	0	0	0	0	…	6201
0	6	4	0	0	…	6202
0	0	22	0	0	…	6204
0	0	0	24	0	…	6205
0	0	0	0	32	…	6206
…	…	…	…	…	…	…

5. CONCLUSIONS

In the DfMC context, the use of PCSs is a decisive factor in obtaining the benefits of mass customization due to improvements in knowledge management and control of product variants. Accordingly, this paper presented a preliminary research on the use of BNs as PCS inference engine to select features of electric motors in a multinational-level company. The naïve Bayes and TAN structures made it possible to identify the main dependencies between features that may be selected by the client. Next, a general BN to select features of flanges was developed using the PC algorithm and tested through cross-validation. Finally, a general BN to select cast items was implemented.

The results indicate that the features *frame group*, *bearing size* and *able to grease* can be accurately predicted in the naïve and TAN BNs. However, the TAN structure increases the accuracy to predict the values of the features *flange size* and *flange type*, which indicates that the correlation between network variables is of great importance in defining the output node value. The accuracy to select features using the general BN is similar to the obtained in the TAN structure, and the GBN achieved an accuracy of 83.61% to select cast items. Moreover, it was possible to identify features that have high diagnostic value to define the product configuration. The knowledge about this DV can be used to suggest which are the next features expected to be filled on the PCS by the user and to order the features in the PCS screen. For example, the most important feature to select cast items is *frame group* (DV = 38.7%). Another potential application of BNs is a dynamic PCS, in which characteristic values are automatically suggested to the user using probabilities calculated in BNs. The features *bearing cap*, *bearing sealing* and *drain* are highly independent from other nodes. Thus, a complementary approach should be evaluated, e.g., the use of rules to load the standard value for these features according to the product line. Moreover, it may be interesting for the company to offer several options to the client because these features have little impact on the product BOM.

The simplified model presented in this research will serve as a base model to be incrementally improved. Only few variables were used in the Bayesian networks. As discussed in Section 1, the target company PCS has 285 variables, which will be used in future versions. For example, this simplified network does not include *poles* and *output rating*. Furthermore, in the next versions of the model, algorithms such as GTT and Bayes search will aid in identifying relationships among variables and, therefore, to elaborate a structure that better emulates the product structure. There are several structure learning and CPTs learning parameters, e.g., significance threshold and max adjacency size, which may be tuned to improve the results. Furthermore, a greater number of iterations during the learning process will refine the model. Finally, there are several rare events in the database, e.g., the value of *bearing size* “7309” occurs in only

0.14% of cases, which represents noise during the learning processes. Expanding the database, i.e., having more cases, will help to mitigate this phenomenon.

6. ACKNOWLEDGEMENTS

We would like to show our gratitude to Federal University of Santa Catarina (UFSC), CAPES and WEG Electrical Equipment S.A.

7. REFERENCES

BayesFusion, 2020. GeNIe 3.0.5905.0 (32-bit). Built on 2020-11-05. Available at <https://www.bayesfusion.com>.
CONFIGIT, 2020. Integration of Configit ACE® and Aras Innovator.
Falkner, A., Friedrich, G., Haselböck, A., Schenner, G., Schreiner, H., 2016. "Twenty-five years of successful application of constraint Technologies at Siemens". AI Magazine, pp. 67-80.
Fettermann, D. C., Echeveste, M. E. S., 2011. "Desenvolvimento de produto para customização em massa: alternativas para o setor de móveis modulados". Espacios. 32v. pp. 1-12.
Forgy, C. L., 1982. "Rete: A fast algorithm for the many pattern/many object pattern match problem". Artificial Intelligence, p. 17-37.
Haselein, B. Z., Silva, J. C., 2019. "Knowledge-Based System for Electric Motors Compliance Verification: A Design for Mass Customization Scenario," 25th ABCM International Congress of Mechanical Engineering. Uberlandia, MG, October 20-25, pp. 1-10.
Hendler, J., Kitano, H., Nebel, B., 2006. "Handbook of Constraint Programming". Elsevier: 1ª Ed.
Fleig, A. M. 2008. "Sistematização da concepção de produtos modulares: um estudo de caso na indústria de refrigeração". Dissertação de mestrado em engenharia mecânica - UFSC. Florianópolis, SC.
Jiao, J., MA, Q., Tseng, M. M., 2003. "Toward High-Added Products and Services: Mass Customization and Beyond. Technovation". Ed. 23. pp. 809-821.
Khan, B., Khan, F., Veitch, B., Yang, M., 2018. "An operational risk analysis tool to analyze marine transportation in Arctic waters". Reliability Engineering and System Safety 169, pp. 485-502. https://doi.org/10.1016/j.ress.2017.09.014.
Le, T., Liu, L. Liu, B., Tsykin, A., Goodall, G., Satou, K., Li, J., 2013. "Inferring microRNA and transcription factor regulatory networks in heterogeneous data". BMC bioinformatics, 14. 92, pp 1-14. 10.1186/1471-2105-14-92.
Meyer, M. H., Lehnerd, A. P., 1997. "The Power of Product Platforms: Building Value and Cost Leadership". The Free Press, New York, NY.
Madden, M. G., 2009. "On the classification performance of TAN and general Bayesian networks". Knowledge-Based Systems 22, pp. 489-495. doi:10.1016/j.knosys.2008.10.006.
Massirer, D. A., 2007. "Sistema Especialista Protótipo para Auxílio à Verificação da Estrutura de Produto de Motores Elétricos". Dissertação de Mestrado, Programa de Pós-Graduação em Engenharia Mecânica. UFSC, Florianópolis.
Myrodia, A., Kristjansdottir, K., Hvam L., 2017. "Impact of Product Configuration Systems on Product Profitability and Costing Accuracy". Computers in Industry. Ed 88. pp. 12-18. DOI http://dx.doi.org/10.1016/j.compind.2017.03.001.
Mouafo, S., Vaton, S., Courant, J-L., Gosselin, S., 2016. "A tutorial on the EM algorithm for Bayesian networks: application to self-diagnosis of GPON-FTTH networks". IWCMC 2016: 12th International Wireless Communications & Mobile Computing Conference, Sep 2016, Paphos, Cyprus. pp. 369-376, 10.1109/IWCMC.2016.7577086. hal-01394337.
Pearl, P., 1988. "Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference". San Mateo, CA: Morgan Kaufmann.
Trentin, A., Perin, E., Forza, C., 2013. "Sales configurator capabilities to avoid the product variety paradox: Construct development and validation". Computers in Industry, Ed. 64, pp. 436-447.
Tsagris, M., 2018. "Bayesian network learning with the PC algorithm: an improved and correct variation". Applied Artificial Intelligence. Vol. 33, pp. 1-20. DOI: 10.1080/08839514.2018.1526760.
Uusitalo, L., 2007. "Advantages and challenges of Bayesian networks in environmental modeling". Ecological modelling, 203(3-4): 312-318. https://doi.org/10.1016/j.ecolmodel.2006.11.033.
WEG Electrical Equipment S.A., 2021. "WEG em números". Available at <https://www.weg.net/institutional/BR/pt/weg-in-numbers>. Last access: June 11, 2021.
WEG Electrical Equipment S.A., 2020. "W22 Three-Phase Electric Motor Technical Catalog NEMA Market. Catalog". Jaragua do Sul, pp. 1-44.

8. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.