Clustering and Association Rule Mining in Commercial Coronary Stent Design: Bare Metal Stent (BMS)

Eriko Rizki Firmansyah[1], Dawi Karomati Baroroh[1,*]
[1]Department of Mechanical and Industrial Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia
*dawi.karomati.b@mail.ugm.ac.id

***Abstract*—Since their first development in the 1980s, more than 100 commercial coronary stents, particularly Bare Metal Stents (BMS), have been produced worldwide. The growing number of commercial brands of coronary stents has led to many variations, making it difficult to distinguish the characteristics between different stents. Additionally, there is limited research examining the relationships between variables that lead to variations in commercial BMS designs. This research aims to use the k-modes algorithm for clustering and the Apriori algorithm for mining association rules, considering five variables: expansion mechanism, material type, form, geometry, and stent thickness. From an analysis of 121 BMS data points, four clusters were formed. The first cluster consists of 51 types of stents, the second cluster comprises 18 types of stents, the third cluster includes 33 types of stents, and the fourth cluster contains 19 types of stents. From the 121 BMS data points, six association rules were identified. Additionally, within the clusters, 15 association rules were formed in the first cluster, 3 in the second cluster, 8 in the third cluster, and none in the fourth cluster. Clustering and association rule mining are expected to facilitate researchers and research teams in developing coronary stent designs.**

***Keywords*—*bare metal stent, clustering, k-modes algorithm, association rule mining, apriori algorithm*.**

I. INTRODUCTION

Heart disease (including coronary heart disease, hypertension, and stroke) is the main cause of death in the United States. Coronary heart disease is one of seven major causes of death in the United States, and kills more than 360,000 people a year [1]. Coronary heart disease is a narrowing of arteries in heart that commonly called restenosis. This is usually caused by formation of fat or calcium deposits called plaque. As time goes by, this plaque is capable of causing total blockage in heart arteries [2]. If the artery is completely blocked, this will cause a heart attack.

Several options for the treatment of coronary heart disease are taking medications, balloon angioplasty, coronary artery stenting, and coronary artery graft surgery [2]. Many patients who undergo balloon angioplasty treatment have restenosis. Restenosis is more common in patients who undergo balloon angioplasty treatment than patients who perform coronary stent. Many patients tend to prefer coronary stent rather than having a graft surgery because the advantage of this method is faster and less surgical treatment to deal with the accumulation of plaque in the coronary arteries. Patients also only take a few days in the hospital and undergo a faster and more comfortable healing period [3]. Therefore, the coronary artery stenting is preferred by the patient. The coronary stents will left in the heart arteries and help to prevent the arteries from restenosis. In its development, there are two types of commercial coronary stents: bare metal stent (BMS) and drug-eluting stents (DES) [4] Since it was first developed in the 1980s, there are more than 100 coronary stents that have been marketed and evaluated worldwide [5]. The growing number of commercial brands of coronary stents, causing many variations that can occur.

According to [5], variations in coronary stent design may be viewed from four major aspects which are type of material, form, fabrication processes, and geometric shapes. These increasingly high coronary stents variations can cause hidden information that can be explored for the improvement process especially in design. In addition, there is very little research examines the relationship between variables that led to variations in commercial BMS designs. To resolve these problems, data mining approach is done. Data mining is able to extract the knowledge or information from the data with large volume so as to produce improvement [6]. Therefore, this research applies data mining approach to extract information from commercial coronary stent products, especially the increasingly varied type of BMS. Data mining is expected to facilitate researchers and research teams in developing the design of coronary stent especially bare metal stent.

II. RESEARCH METHODS

A. *Research Object*

The object of this research is bare metal stent (BMS) produced in the last two decades. The variables that will be considered in this research are expansion mechanism, material type, form, geometry shape, and stent thickness. Variables of expansion mechanism, material type, form, and geometric shape were chosen because according to [5], these four variables are the causes of variations of stents design increase. The stent thickness variable was chosen because at the beginning of the stent development, the main characteristic of coronary stent was the ease of tracking the device down to the target vessel and through the lesion. The advantages of these features are significantly influenced by the thickness of the strut, where the strut with a thinner thickness will become more flexible and reduce cross-sectional profiles [7]. Therefore, the thickness of the stent is also one of the causes of the increasingly varied design of the stents.

B. Research Stages

Research begins with pre-processing data consisting of data normalization and data discretization. Data normalization is done as input in clustering, while data discretization is done as input in the mining of association rules. Data normalization will be done on variables that have numerical data: stent thickness, while for variables with categorical data that is expansion mechanism, material type, form, and geometric shape assumed to have normal data. Data Normalization is done by use (1).

$$x' = \frac{(x-\mu)}{\sigma} \tag{1}$$

The data discretization is done by the equal width approach that divides the range of variables into intervals of equal width [8]. Data discretization is performed on stent thickness data by dividing stent thickness data into five classes: very thin, thin, normal, thick, and very thick. After data normalization and discretization, the stent will be clustered using k-modes algorithm. All BMS data then will be divided into 85 training sets data for clustering and mining association rules and 36 test set data for validating.

The k-modes algorithm is a data clustering algorithm by simply partitioning data and can be applied to data with numerical or categorical types. The k-modes algorithm has the goal of minimizing the dissimilarity distance or within cluster error expressed as the mode of the whole data, the use of the mode is perform to represent the cluster centre and renew the mode with the most categorical values in each iteration process [9]. The clustering process is done by calculating use (2) and will be done on 1 to n-cluster by replicating 30 times to avoid optimal local solution. Therefore, there will be 2 to n-cluster or stent groups with minimum dissimilarity distance or within cluster error. N-clusters that have been formed will be evaluated to get optimal cluster number for BMS. Cluster evaluation is done by calculating the silhouette index value.

$$J = \sum_{l=1}^{k}\sum_{i=1}^{n}\sum_{j=1}^{r} w_{il} \in \left(x_{ij}, q_{lj}\right) \tag{2}$$

Silhouette index is a clustering evaluation method that is also capable for validating a data, single cluster, or clustered data [10]. The silhouette index values minimize the error values between data in the same cluster and maximize the value of data errors in different clusters. Calculation of silhouette index value and the largest mean value of silhouette index are done by use (3) and (4), respectively.

$$SI_i = \frac{b_i - a_i}{\max\{a_i, b_i\}} \tag{3}$$

$$SI = \frac{1}{k}\sum_{j=1}^{k} SI_j \tag{4}$$

The results of clustering that have been evaluated using the silhouette index will be validated by comparing the error values of the training set data to the centre of each cluster and test set data to the centre of each cluster, so that a valid cluster centre will be obtained. The value of this error will be tested using t-test with hypothesis as follows, where the null hypothesis will be rejected if p-value less than α. The α value used in this statistical test is 0.05.

1. $H_0 : \mu_{training} = \mu_{test}$
2. $H_1 : \mu_{training} \neq \mu_{test}$

After obtaining the optimal number of clusters with each valid cluster centre, the stent will be examined using mining association rules to determine the relationship between variables that cause stent design variation. Association rule mining is done by apriori algorithm. Association rules consist of two items commonly called antecedent and consequent, there are two statistical techniques in describing the affinity or relationship between antecedent and consequent: support and confidence. Support is the proportion or frequency of occurrence of antecedent and consequent relationships in a group of items. Confidence is the proportion of transactions containing antecedent which also contains consequent. Equation of support and confidence can be seen in equation (5) and (6).

$$\textit{Support}\ (A \rightarrow C) = \textit{support}\ (A \cup C) \tag{5}$$

$$\textit{Confidence}\ (A \rightarrow C) = \frac{\textit{support}\ (A \cup C)}{\textit{support}\ (A)} \tag{6}$$

Associated rules that have been established with minimum support and minimum confidence limits will be evaluated to know the most interesting rules by calculating the lift value. A lift value that shows above one indicates that consequent is more common in transactions containing antecedents than transactions that do not contain antecedents [11]. Formulation of lift calculation can be seen in equation (7).

$$Lift(A \rightarrow C) = \frac{confidence(A \rightarrow C)}{support\ C} \tag{7}$$

Association rules that have been evaluated based on lift values will be validated to determine the significance of the relationship between antecedent variables (causal variables) and consequent variables (causal variables). Validation will be done by comparing the value of chi-square observation test with chi-square table at confidence level 0,05. The chi-square value will be calculated by equation (8). The chi-square test will be performed with the following hypothesis, where the null hypothesis will be rejected if the value $\chi^2_{calculate} > \chi^2_{table}$.

$$X^2 = n(lift - 1)^2 \frac{support\ x\ confidence}{(conf - supp)(lift - conf)} \quad (8)$$

1. H_0 : The antecedent and consequent variables have independent relationships
2. H_1 : The antecedent and consequent variables have dependent relationships

III. RESULT AND DISCUSSION

A. Pre-processing Data

Bare metal stents have two categories for expansion mechanism variables, six categories for material type variables, four categories for form variables, and three categories for geometric shape variables. All categories for each variable in BMS can be seen in Table 1. Variables with the numerical type, stent thickness, will be normalized and made discrete, the results of data normalization and discretization are presented in Table 2.

Table 1. Category of Each Variable

	Variabel			
	Expansion mechanism	Material Type	Form	Geometric Shape
Categories	Balloon Expandable	Stainless Steel	Coil	Open Cell
	Self Expandable	Cobalt-Chromium Alloy	Wire	Close Cell
		Nitinol	Slotted Tube	Hybrid Cell
		Platinum-Iridium Alloy	Modular	
		Platinum-Chromium Alloy		
		Tantalum		

Table 2. Result of Data Normalization and Discretization

Categories	Stent Thickness (millimeter)	
	Original Data	Normalized Data
Very Thin	0.009 until 0.073	-2.214 until -0.403
Thin	0.074 until 0.137	-0.402 until 1.407
Normal	0.138 until 0.202	1.408 until 3.218
Thick	0.203 until 0.266	3.219 until 5.029
Very Thick	0.267 until 0.330	5.030 until 6.840

B. Clustering

The k-modes clustering is done on 85 BMS training set data by calculating within cluster error at 2 < x < n clusters. The visualization of within the cluster error of BMS is presented in Figure 1.

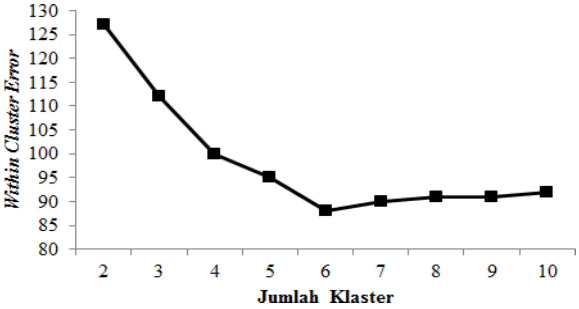


Figure 1. Visualization of Within Cluster Error

Based on Figure 1, within cluster error decreases in the form of two clusters up to six clusters, but within cluster error increases again in the form of seven to ten clusters. The increasing of within cluster error indicates that the total error value between data and each cluster center when seven to ten clusters are formed is higher than when six clusters are formed. Therefore, the BMS clustering process is only done in two to six clusters and will be evaluated to determine the optimal number of clusters. Clustering evaluation is done by calculating silhouette index value. The results of the silhouette index are presented in Table 3 below.

Table 3. Silhouette Index Value

Number of Clusters	**Nth Cluster**						**Average**
	1st Cluster	**2nd Cluster**	**3rd Cluster**	**4th Cluster**	**5th Cluster**	**6th Cluster**	
2 Clusters	0.0955	0.6819					0.3887
3 Clusters	0.1209	0.5276	0.4243				0.3576
4 Clusters	0.3149	0.2979	0.4345	0.5248			**0.3930**
5 Clusters	-0.2718	0.2977	0.5512	0.6748	0.6999		0.3904
6 Clusters	0.3104	-0.1716	0.3975	0.2200	0.7535	0.6963	0.3677

Based on Table 3, the largest average silhouette index value is achieved when forming of four clusters is 0.3930. The silhouette index value for the nth cluster in the formation of the four clusters also shows a positive value, indicating that the data in the ith cluster is solid and the closeness of ith data in the same cluster is closer than in the different cluster. Therefore, based on silhouette index value, the optimal number of clusters generated by training set data of BMS are four clusters. The next stage is to validate the center of each cluster by testing it using t-test with test set data for each cluster. Table 4 shows the validation result of each cluster center.

Table 4. Cluster Center Validation

Nth Cluster	**p-value**	**Interpretation**
1	0.5934	Failed to reject H_0
2	0.3128	Failed to reject H_0
3	0.6322	Failed to reject H_0
4	0.4827	Failed to reject H_0

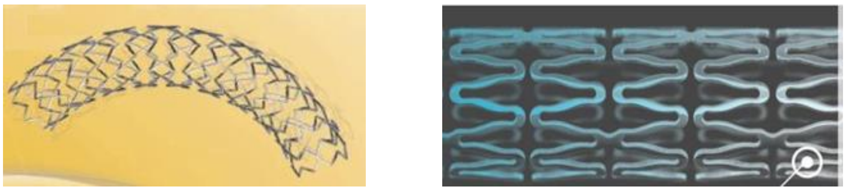
Based on Table 4, all of four clusters show a greater p-value than α, so there is not enough evidence to reject H0. This shows that the average value of data errors of cluster centers for training sets and test sets did not show significant differences. Therefore, the clustering results and all four cluster centers formed are valid. Table 5 shows the cluster center and the number of members on each cluster and Figure 2 shows examples of stent products for each cluster formed.

Table 5. Cluster Center and Number of Member on Each Cluster

Nth Cluster	**Expansion**	**Material**	**Form**	**Geometric**	**Thickness**	**Number of Member**
1	BX	CoCr	Modular	Open	0.065	51
2	BX	CoCr	Slotted	Open	0.081	18
3	BX	SS	Slotted	Open	0.090	33
4	BX	SS	Modular	Open	0.100	19

(a) Flexystar (Sahajanand)

(b) Yukon CC (Translumina)



(c) Gazelle (Biosensor) (d) Cobra PZF (Celonova)

Figure 2. Example of Stent Product: (a) 1st Cluster; (b) 2nd Cluster; (c) 3rd Cluster; and (d) 4th Cluster

C. Association Rule Mining

Association rules mining will be performed on all data of BMS and their clusters using apriori algorithm considering frequent item sets in a single data set. A number of association rules will be formed from a combination of minimum support and minimum confidence that has been determined, and meet the minimum value of lift 1. Table 6 shows an example of the number of rules formed on all data of BMS. The association rules that are formed with the minimum support and minimum confidence values will result in different number of rules as well. Based on Table 6, at a fixed minimum support level, the addition of minimum confidence value has a downward trend in the number of rules and vice versa. The difference in the number of these rules will be evaluated using a comparison of average lift values for each rule to determine the optimal number of association rules, besides evaluation will be performed to determine which association rules have a dependent relationship by performing a chi-square test. The comparison of the average value of lift will be made in graphical form and the rule with the largest average of lift value will be selected.

Table 6. Number of Rules Formed on All Data of BMS

Min. Support	**Min. Confidence**						
	0.30	**0.40**	**0.50**	**0.60**	**0.65**	**0.70**	**0.80**
0.30	**44**	43	38	31	27	23	15
0.35	26	26	24	19	17	16	11
0.40	13	13	11	10	9	8	5
0.45	11	11	9	9	8	7	4
0.50	2	2	2	2	2	2	**1**

The association rules that have been evaluated will be validated with test set data by calculating the support, confidence, and elevator values for each rule. Association rules that have lift values above one show a valid association rule because consequent is more common in transactions containing antecedents than transactions that do not contain antecedents [11]. The number of association rules with combination of minimum support and confidence that have been evaluated and validated is presented in Table 7, while the validated association rules will be presented in Table 8.

Table 7. Number of Association Rules

Description	**Minimum Support**	**Minimum Confidence**	**Number of Association Rules**
BMS	0.30	0.50	6
1st Cluster	0.50	0.75	15
2nd Cluster	0.30	0.90	3
3rd Cluster	0.80	0.30	8
4th Cluster	0.05	0.10	0

Table 8. Association Rules for All Data of BMS

No	**Antecedent**		**Consequent**
1	Expansion : BX	Thickness : Thin	Material : SS
2	Form : Modular		Material : CoCr
3	Material : CoCr		Form : Modular
4	Material : SS		Thickness : Thin

5	Expansion : BX	Material : SS	Thickness : Thin
6	Thickness : Thin		Material : SS

The association rules formed for all data of BMS are 6 rules with min. support 0.3 and min. confidence 0.5, where stainless steel material tend to be affected by stents with BX expansion and thin thickness, whereas cobalt-chromium material is more frequently affected by stents with modular form. The higest number of association rules is formed in the first cluster, indicating that the data in the first cluster are heterogeneous with cobalt- chromium material influenced by stents with BX expansion and very thin thickness. In the second cluster, 3 rules are formed with min. support 0.3 and min. confidence 0.9, where the cobalt-chromium material is affected by the stent with BX expansion and slotted tube form. The third cluster produces 8 rules with min. support 0.8 and min. confidence 0.3, where the stent with slotted tube form tends to be influenced by BX expansion, stainless steel material, and has an open geometry. In the fourth cluster, no rule is established because the data in the fourth cluster is highly heterogeneous and has no dependent relationship.

III. CONCLUSION

Based on 121 commercial BMS coronary stents, it is known that four clusters are formed and validated from clustering analysis using k-modes algorithm. The first cluster consists of 51 types of stents with a characteristic formed from a type of cobalt-chromium alloy material with a modular form and has a very thin thickness. The second cluster consists of 18 types of stents with a characteristic formed from a type of cobalt-chromium alloy material with a slotted tube form and has a thin thickness. The third cluster consists of 33 types of stents with a characteristic formed from a type of stainless steel material with a slotted tube form and has a thin thickness. The fourth cluster consists of 19 types of stents with a characteristic formed from a stainless steel material type with a modular form and has a thin thickness.

Association rule mining using apriori algorithm is done to determine the relationship between variables that cause variations in BMS stent design. The association rules formed for all data of BMS are 6 rules with min. support 0.3 and min. confidence 0.5, where stainless steel material tend to be affected by stents with BX expansion and thin thickness, whereas cobalt-chromium material is more frequently affected by stents with modular form. Association rule mining is also done on each cluster that has been formed previously, the first cluster has the most association rules- 15 rules with min. support 0.5 and min. confidence 0.75, where cobalt-chromium material influenced by stents with BX expansion and very thin thickness. In the second cluster, 3 rules are formed with min. support 0.3 and min. confidence 0.9, where the cobalt-chromium material is affected by the stent with BX expansion and slotted tube form. The third cluster produces 8 rules with min. support 0.8 and min. confidence 0.3, where the stent with slotted tube form tends to be influenced by BX expansion, stainless steel material, and has an open geometry. In the fourth cluster, no rule is established because the data in the fourth cluster is highly heterogeneous and has no dependent relationship.

This research still needs to be developed by doing clustering with other algorithms to compare whether the number of clusters and cluster centers formed are same or not. Further research is also expected to mine association rules with other algorithms to compare the combination of rules formed and the number of association rules that are formed.

REFERENCES

[1] American Heart Association (AHA), “Heart Disease and Stroke Statistics 2017”, https://www.heart.org (online accessed 5 September 2017).

[2] Boston Scientific, 2017, *Stent: Coronary*, http://www.bostonscientific.com/en-US/products/stents--coronary.html (online accessed pada 18 September 2017).

[3] Auricchio, F. dan Conti, P.M., 2009, *Finite Element Analysis of Coronary Artery Stenting*, Federico Fogorotto, Pavia.

[4] Heart Foundation, 2008, *Coronary Artery Stent*, https://www.heartfoundation.org.au (online accessed 5 September 2017).

[5] Stoeckel, D., Bonsignore, C., dan Duda, S., 2002, A Survey of Stent Design, *Journal of Minimally Invasive Therapy & Allied Technologies*, Vol. 11, No. 4, pp. 137 – 147.

[6] Agard, B. dan Kusiak, A., 2004, Data-Mining-Based Methodology for Design of Product Families, *International Journal of Production Research*, Vol. 2, No. 15, pp. 2955-2969.

[7] O’Brien, B. dan Carroll, W., 2009, The Evolution of Cardiovascular Stent Materials and Surfaces in Response to Clinical Drivers: A Review, *Actabiomatrialia*, Vol. 5, No. 4, pp. 945-958.

[8] Hermawati, F.A., 2013, *Data mining*, Penerbit Andi, Surabaya.

[9] Huang, Z., 1997, *A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data mining*, Cooperative Research Centre for Advanced Computational Systems, Australia.

[10] Prasetyo, E., 2014, *Data mining: Mengolah Data Menjadi Informasi menggunakan Matlab*, Penerbit Andi, Yogyakarta.

[11] Ye, N., 2003, *The Handbook of Data mining*, Lawrence Erlbaum Associates, New Jersey. Zhang, C. dan Zhang, S., 2002, *Association Rule Mining: Models and Algorithms*, Springer, Sydney.