

Application of Kansei Engineering and Data Mining in Developing an Ingenious Product Co-design System

Kittipong Sakornsathien, Sukree Sinthupinyo, and Pongpun Anuntavoranich

Abstract—This research intends to be a part in developing an automatic system that shows the potential of designing a product form by co-designing with the user in order to make it suitable for each consumer by applying Kansei Engineering Technique. However, the operation of KE needs data collection from specific target group in order to devise a formula and interpret the results when applying to study the same target group, or else the accuracy will be reduced. In order to eliminate the drawbacks, Data Mining techniques will be applied with KE system in order to use as an ingenious product co-design system.

The style and preference of each user will be used as a categorizing factor clustering the database into groups with K-means technique, in this research. Each classifying cluster will use its own database in the system processing in order to gain a set of design elements precisely from the system. Decision Tree classification technique is selected for the study. For the model validation, we apply the cross-validation as an unbiased model performance evaluation. In order to build the KE system, the sentiments in this research are indicated by 5 pair-words constituting the system data. The result shows that, this method offers more accurate prediction of design elements comparing to the method which undivided users.

Index Terms—Data mining, machine learning, Kansei engineering, product development.

I. INTRODUCTION

Participatory design and user-centered design admittedly lead to higher efficiency, success and product acceptance [1], [2] but suitable innovative product development process that engenders effective and uncomplicated product co-creation is also an interesting and important issue. In addition, Fuller [3] emphasizes on the importance of consumer diversity that participates in product co-creation. High effective platform and easy to use with low effort constitutes a primary motivation of consumers whether or not they collaborate with the manufacturer in product co-creation. In the same way, the experimentation by Steen [4] found that interactive tools come up with positive results both in communication and collaboration that lead to the design that meets the objective. The application of Data Science tools such as

Recommendation Engines or Smart Algorithms for dynamic pricing facilitates consumers' participation in product development [5].

In this research the authors express the intention to develop product form co-design assistance system which is easy to use. We are convinced that Kansei Engineering (KE) technique that is used to interpret consumer sentiments to as a non-negligible element in design [6], [7] should be applied as a major tool in system development. We are also certain that for diverse usage of form, KE is capable to help designers to plan the communication and to reflect effectively the sentiments in the product to consumers [8]. In the past two decades, KE is recognized widely as a practical tool by several industries. It is also regarded as an effective technique in developing the product to attain the success [9]. However, the operation of KE needs data collection from specific target group in order to devise a formula and interpret the results when applying to study the same target group, or else the accuracy will be reduced. That is the sole restriction of KE technique [10], and the weak point of design assistance system that we aim to apply to different groups. Kuang and Jiang [11] separate database in KE into groups according to target groups that share the same appreciation of the same product prototype. This leads to increasing accuracy of KE. Nevertheless, that way of grouping lifts another restriction when applying to others products. This can be explained by the fact that the target group that appreciates the same product may not feel the same way with others product forms.

With the aim to reduce those Kansei Engineering's system limitations, after reviewing many related studies, the researcher will reinforce the KE system with the competence of machine learning that co-operating with data mining classification technique. Moreover, the style and preference of individual user will be used as a categorized factor clustering the data into groups with K-means algorithm. Each classified cluster will use its own database in the system processing in order to gain set of design elements from the system precisely by classification technique. In this project, the cross-validation method will be used as unbiased model performance evaluation.

II. RELATED WORK

A. Kansei Engineering

The definition of Kansei Engineering as given in Nagamachi [12] stated that Kansei Engineering is seen as "translating technology of a consumer's feeling and image for a product into design elements". It functions as a tool of

Manuscript received August 24, 2018; revised November 7, 2018.

K. Sakornsathien is with the School of Technopreneurship and Innovation Management, Graduate School, Chulalongkorn University, 254 Phayathai Rd., Phatumwan Bangkok, 10330 Thailand (e-mail: kittipong.s@student.chula.ac.th).

S. Sinthupinyo is with the Department of Computer Engineering, Chulalongkorn University, 254 Phayathai Rd., Phatumwan Bangkok, 10330 Thailand (e-mail: sukree.s@chula.ac.th).

P. Anantavoranich is with the Department of Industrial Design, Faculty of Architecture, Chulalongkorn University 254 Phayathai Rd., Phatumwan Bangkok, 10330 Thailand (e-mail: p.idchula@gmail.com).

semantic differential method usually applied to elicit consumers' psychological feelings or preferences about a product in the emotional assessment for KE Technique [13]. Recently, Kansei Engineering approach succeeds in its attempt to get closer to the users through their more grounded association in the early steps of user experience's creation [8]. The system is therefore applied and exploited in various industries [14] including home and interior design by allowing consumers to customize their needs for fashion, Car interior decoration [15], Office Chair [16], Industrial product [17], Material's surface [18], and also many well-known products in the market [19]. Through wide range of advantage and utility, the KE Methods is reinforced in combination with Back Propagation Neural Networks and Genetics Algorithm to develop a Decision Support System (DSS) for industrial design which work autonomously without product designer.

B. Ideal Product Design Concept

From the proved model, designing ideal products for positive responses and leading to buying steps depends on two moderate factors which are individual tastes and preferences and situational factors [20]. Bloch [20] presents in his research that moderating factors determining the result of psychological responses to product form consist of situational factors. Another interesting factor of diversity in individual is Individual Tastes and Preferences, including main attributes such as Innate Design Preference, Cultural and Social Context and Consumer Characteristic. Comparative cross-cultural study of craftsmanship assessment for vehicle interior design found no significant differences between two Nationality groups of the subjects that were all males mechanical engineering student [21]. Also, the study by Rajapakse [22] which have found no significant differences in response to the product shape in male and female. This issue is raised in the widely recognized research on product form relation with consumer response written by Bloch [20]. The knowledge is acquired by interdisciplinary approach and used to point out individual basic differentiation factors clearly in response to product form. Previous researches show two types of consumers; those who prefer innovation and those who prefer design. They have positive psychological responses towards modern product forms but different behavioral responses [23]. To conclude, user's tastes and preference are the factor to be concerned in developing the desired product form.

C. Data Mining Techniques

Data mining is the process for sorting the information in large database to predict trend and behavior. This method analyzes existing data to discover patterns of relationship and to gain new information [24]. A wide variation in terms of professional fields, formulations, and data management techniques as well as the possibility to choose important data become necessary to consider for the application of data mining technique [25]. There are many applications using the Data mining techniques, however, in this research, Cluster Analysis and Classification Technique are preferred to conduct the study.

Cluster analysis (CA) or segmentation analysis is one of the inferential methods for data analysis. The analysis is

conducted to illustrate accuracy validation of the model or hypothesis. The objective of cluster analysis consists of the division of objects into homogeneous groups in which two members of the same group are more similar to those of the others groups, according to Data mining process. Moreover, CA is also unsupervised learning technique [25]. As regards of data clustering, two important questions can be posed about (a) the similarity of the objects in each cluster, and (b) the method to identify such similarity. Wierzchoń used cluster analysis of multi-predictors technique to separate and cluster the customers of a bank in credit granting [24]. His study points out that the clustering is accurate and practical. The analysis of Kašćelan [26] illustrates personal preference of investors with data science technique by applying CA in dividing investors who share common pattern of investment into clusters. It is not difficult to see that dividing customers into clusters can double value for the business and enhance potential in management. As for product design, CA is applied to develop a product design platform. The results brought about from the clustering are used to determine the setting of the platform in terms of appropriateness and efficiency [27].

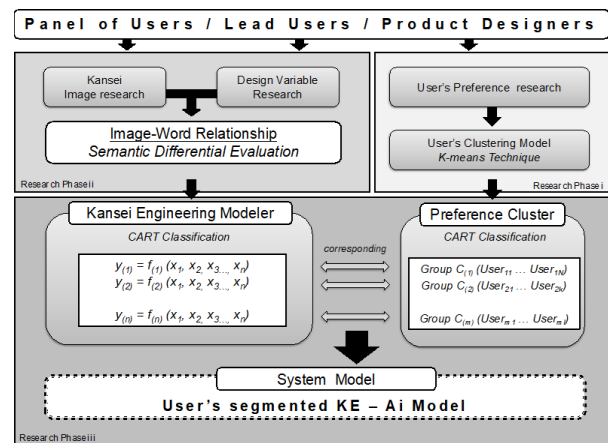


Fig. 1. Detail of proposed research model in this study.

The application of classification and regression tree (CART) or widely known as decision tree helps us indicating the attribute that is used to classify data into clusters [28]. That is because classification and regression tree is the classification methodology that represents modeling process using hierarchical decisions [25]. The structure is in form of tree, as indicated by its name, with a node as attribute together with value-pair as decision point and indication to the next node. Starting with root node, a value-pair is then selected along the branch to end with leaf node.

III. THE PROPOSED MODEL

The researcher followed the idea of Kansei Engineering technique to allow the development of product form co-design assistance system. The reason of this option came from the fact that this technique is engendered during the identification of the relation from Kansei Image Research and Design Variable Research of target customers. Considering the necessity to remove restrictions of Kansei Engineering as mentioned above, the classification of users' data into cluster following their tastes and preferences with

data mining technique was applied to complete Kansei Engineering in such a way that system accuracy will be enhanced. Moreover, the modeling with CART to predict each type of design element from emotional factor constitutes one of the techniques allowing the system to work faster as well as to support AI development and feedback data [29]. Database is therefore continually updated.

The system development can be divided into three main phases as follows:

Phase I. Customer clustering with user's preference research

Phase II. Kansei Image research and Design Variable research

Phase III. KE-Ai model development with Data-mining Classification technique

IV. RESEARCH PHASE I

A. Data and Collection

To be in accordance with the objective of this research, we chose to collect the information with online opinion questionnaires accessible with PC, tablet and mobile phone due to the fact that the questionnaires contain mostly questions with pictures. Attribute groups used in the questionnaires comply with those mentioned in the research on individual tastes and preferences conducted by [20]. The questions are posed about design references, design acumens and social contexts of the respondents. The questions come also in form of pictures. Respondents choose a picture they prefer to test personal preference of picture form. The pictures are also diverse, for example in symmetric or asymmetric forms, golden ratio or non-golden ratio pictures. We also asked the respondents to evaluate the environment factors influence such as career, family members that affect their familiarity with designing. They are asked to provide some personal information that we considered as necessary and corresponds to attributes we need to investigate.

We obtained answers from 346 respondents, 60 of whom had the degrees related to the design directly, 13 respondents live abroad, 252 respondents in Bangkok and 81 in provinces. According to the results from the questionnaires, there are 97 males, 209 females and 40 others. In terms of mobile phone, 213 respondents use iPhone whereas 123 of them use Android phones and 10 others OS. Concerning age of respondents, we present it by generation; 2 from Gen Z, 91 from Gen Y, 245 from Gen X (the largest) and 9 baby boomers.

B. Data Preparation

Another consequential step in model making is the preparation of the data. Well prepared data produce positive results affecting the model. In general, this step contains sub-steps when it comes to transform and prepare the data according to data type as well as date used to develop the model. The main objective of this step is to prepare and bring the data to completion and to reduce noise causing inaccuracy in model making. RapidMiner Studio Version 8 was used on Macintosh OSX operating system to prepare and conduct the data mining process.

The preparation of the data started with collection and verification of 346 sets of data which was then filtrated and selected. After that, the categorical data was transformed into numeric attributes made ready for CA with binarization methods [30]. At this stage, any missing value was replaced with average value. For more efficiency in model making, we applied "feature selection" to reduced noise attributes, and chose attributes which was the most pertinent to the model [25]. We also opt for wrapper approach with classifier to evaluate the accuracy in predicting attributes via sequential forward selection technique (SFS). The system evaluates the effectiveness with cross-validation method beginning with 0 attribute in the first rotation. Then, the operation is repeated rotation by rotation. Finally, only attribute which illustrated the highest value of efficiency was selected. We have 3 groups of questions with 9 attributes that are regarded as high quality tested by SFS process and will then be used in K-means clustering.

C. Data Clustering

To conduct Cluster Analysis with K-Means method, we are required to determine the amount of K in advance. The number of K has a direct impact on efficiency in clustering. The authors assigned the appropriate amount using internal validation criteria and external validation criteria. For internal validation, the silhouette coefficient was applied in this article to find the average value of the distance from the centroid of the cluster to each unit of the data. Then, the comparison of the average value from clustering test was made from 2 to 20 clusters. After that, the efficiency of clustering was evaluated with "Cluster Distance Performance" operator from RapidMiner program before plotting a graph to find an "inflection point" with highest ratio when comparing two points on the graph [31]. The results from internal validation criteria indicated that the possible position of the inflection point was at $k = 4$.

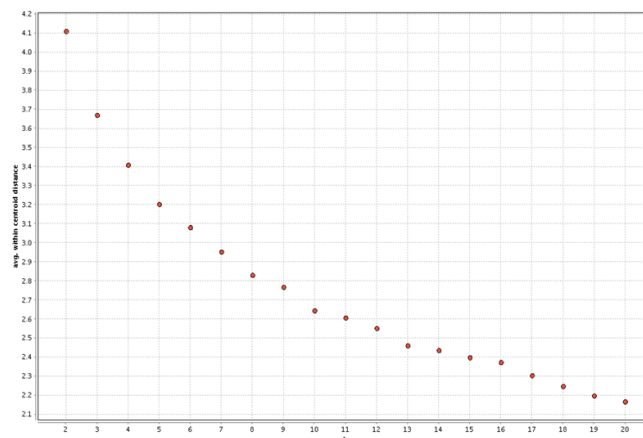


Fig. 2. Average within centroid distance ($k = 2$ to 20).

For external validation, the mixed measure evaluation method with RapidMiner would be used as the tool. The data of attributes from feature selection during data preparation step will also be used. Then, we brought obtained cluster values to label every dataset. After that, the evaluation of efficiency was conducted with 10-folds validation using decision tree technique in classification. Finally, we compared the results of the confusion matrix from $k = 4 / k =$

$5/k = 6$. We could see from the results that 4 clusters ($k = 4$) provide the highest rate of accuracy at 94.22%, followed by $k = 6$ at 89.87% and $k = 5$ at 87.61% respectively. We concluded that 4 clusters provided the lowest rate of classification error.

TABLE I: GROUP OF TASTES AND PREFERENCES ATTRIBUTES AND FORWARD SELECTED ATTRIBUTES

Individual Tastes and Preferences - Attributes			SFS Attributes
Innate Design & Preference	Cultural & Social Context	Consumer Characteristic	
Interior Style	Living Location	Attended Design School	Speaker Style
Watch Style	Gender	Career Influence	Floor Pattern
Speaker Style	Generation	Environment Influence	Unity /Disunity Design
Floor Pattern	Mobile Phone	Family Influence	Organic /Synthetic
Verbal /Nonverbal Style	Gift Style	Design Matching Score	Gender
Asymmetric /Symmetric			Career Influence
Unity /Disunity Design			Environment Influence
Golden /General ratio			Family Influence
Organic/Synthetic			

In the first stage of the research, the final result illustrates that each node constitutes an important attribute in clustering data into labels that we here refer to as the four clusters. The factors used in clustering were the influence of career, environment and family that affect user's design acumen. As for the preference factors including Speaker Style, and Floor Pattern Style, they are applied in decision. The obtained clusters are distinctly different.

V. RESEARCH PHASE II

The main idea of Kansei Engineering can be explained by the tentative to find any relationship between "sentiment" as known as Kansei image and design elements variable. Thus, the phase 2 of the study consists of the preparation of two types of information describing as "sentiment words" reflecting the closest product form needed in this study. The second information is regarded as "design variable" of "stencil products for interior decoration" which is considered as the case study of this research. In addition, the research process of sub variable was carried out by means of opinion surveyed from experienced designers so as to identify sub variables found in each product type.

A. Kansei Image Research

Sentiment words are seen as one of essential variables when it comes to develop Kansei Engineering system. The process began with the search of words employed in the promotion of stencil products for interior decoration. We obtained about 28 pair-words with nonrecurring meaning. These words was determined by three criteria as shown in the

studied by Jindo [32]; 1) words describing material specifications of the product are cut; 2) jargon words are cut; and 3) words expressing the objective of the product are also cut. Once the criteria were applied to those 28 pairs, 15 pairs were cut and finally we obtained 12 pairs which will be used to evaluate the relation and the appropriateness between words and the product. At the evaluation stage, we asked for a collaboration of one professional designer, one vendor of decoration products and one customer with experience and habit of decoration materials. The score ranged from 1 to 10; where 10 corresponds to the most relevant while 1 corresponds to the least relevant. Finally, only 5 pairs with highest score were used.

The score obtained by the evaluation from the experts demonstrated that these 5 pair-words had got averagely more than 8 points. The pair "normal ----- outstanding" obtained the highest score at 9.67 points followed by "retro ----- modern" and "feminine ----- masculine" at 9.33 points ex-aequo. The next lower points pair words was "Simple ---- Complex". Moreover, "look expensive ----- look cheap" which obtains 8.67 points and ranks fifth together with the others ranking from 1 to 4 was used during opinion survey process of target group towards product form. The information collected from the survey will then be used to find the relationship between variables of each product form sample that finally leads to Kansei Engineering model.

B. Design Variables Research

Regarding the modeling of Kansei Engineering, there are two steps to classify variables including criteria establishment or main variables definition to be used in classification and evaluation of each type of product form using main elements or sub elements with expertise of specialists in related fields [33] from the collection of patterns of stencil product samples in the market. The criterion adopted in the analysis of the elements using in stencil product design is that of 2-dimension design principle including lines, shapes, color weight which are placed in an empty space and others design principles such as placement, harmony, contrast, proportion, balance, focus and unity theory. The researchers divided attributes using to analyze element in product sample design into 11 elements (attributes $X_a - X_k$) with diversity of each element. For example, variable X_{a1} puts focus on "straight line" as an element whereas X_{a2} uses "curve line" as an element in design.

After the selection of stencil image which was regarded as sample product, we obtained 48 images. These images were subsequently submitted to experienced designers to filtrate any redundant pattern in such a way that we arrived at the number of 31 images which would be used in design element analysis afterward. The target group of questionnaire respondents was selected with an objective. The total amount of 33 professional designers was selected by their experience, understanding and knowledge related to design theory so as to evaluate the element in product design and to determine the elements in product form design. The example results surveyed by professional designers indicating design elements have shown in figure 3.

The study in this stage gave us two main parts of information necessary in the development of Kansei


Engineering model which would be used in Classification model afterward.

VI. RESEARCH PHASE III

A. Data and Collection

In this stage of study, the sample is divided into 4 clusters by means of artificial intelligence system model developed in the first research Phase. We collected the results of the evaluation of 5 pair-words analyzed in previous phase with 31 stencil product patterns of each cluster. The results are also used to form data vector including evaluation results of 5 pair-words of each pattern of the product and design elements of such product form in order to develop a classification model using data mining technique through machine learning with RapidMiner program. The process carries out in a prediction model of design elements of stencil product form for interior decoration in each element.

TABLE II: DESIGN ELEMENT ANALYSIS OF PRODUCT PATTERN #1

Product No. 1	Attribute	Design Element	Sub-Attrb	Description
	Xa	Main element	Xa1	Curve line
	Xb	Sub element	Xb3	Curve line
	Xc	Main element size	Xc2	5%-15% of total space
	Xd	Sub element size	Xd2	Smaller
	Xe	Direction and movement	Xe1	Vertical
	Xf	Spacing Technique	Xf2	Negative space
	Xg	Repeating rhythm	Xg1	by form
	Xh	Harmony Theory	Xh1	by form
	Xi	Contrast Theory	Xi3	by direction
	Xj	Balancing	Xj1	Equal
	Xk	Golden proportion	Xk1	Not used in design

The authors collected the data from questionnaires via online survey accessible from PC and mobile devices due to the fact that the questions are in form of pictures mostly. The questionnaire was divided into two parts. The first part consists of questions concerning the influence in users' favorite design and product form. The questions were decisive in such a way that they were used to cluster respondents into 4 groups. The other part of the questionnaire consists of 5 questions about emotion and feeling evaluation of users towards product sample forms. Respondents would see sentiment words in left and right columns with opposite meaning. Respondents can thus choose their desired answer in scale ranging from 1 to 5. The scale is described as follows: 1/ General+, 2/ General, 3/ Moderate, 4/ Outstanding, and 5/ Outstanding+. In a picture, respondents must evaluate five answers. We totalize 285 respondents.

B. Data Preparation

After the verification of 285 responses by filtrating incomplete questionnaire or outlier, the study finds out that only 191 responses can be really used. The most of respondents are from "generation X" with the number of 135 respondents (70.7%). Whereas 50 respondents (26.2%) describe themselves as "generation Y" and only 2 of them (1%) is "baby boomer". The most of respondents of the questionnaire, 72.3%, live in the capital of Thailand, 24.1%

of the respondents live in provinces and 3.7% (7 respondents) live in a foreign country. When we classified the population into groups with clustering method as described in Phase 1 with RapidMiner in order to divide 4 databases following the objective of model development of users in each cluster, the results stated that the population size of 4 clusters was quite similar. The 2nd cluster and the 4th cluster had the population of 54 and 53 members respectively, followed by the 1st with 48 members and the 3rd cluster with 36 members. This offers real benefit for the preparation of the data and for the analysis of such similar data of each cluster. The number of data rows in each testing cluster varies depending on the number of population in each cluster. In this study, one member of population generated 31 lines resulting from sentiment evaluation of 31 patterns of stencil. Therefore, 48 members in the 1st cluster generate 1,488 (48x31) rows of data for machine learning for example.

C. Data Classification

Classification Technique is one of data mining technique which is widely applied. Data classification is different from clustering analysis by the fact that learning and model development using data classification technique contain variables in form of designated Label in each dataset for machine learning. The machine will learn and determine the process in data classification step by step and the process will be applied to classify other datasets afterward. In the other hand, clustering analysis deals with unlabeled attributes. This research opts for Classification and Regression Tree technique (CART) through "Decision tree" Operator in RapidMiner. There are 3 Steps for model developing in this phase which can be described as follows:

1) Selection of classification technique

It is well known that there are various methods when it comes to classify data with its diversity of advantage and disadvantage. It can also be stated that the accuracy of a model in prediction provides better results in certain type of data than the other type. The authors of this study opt for RapidMiner to test and validate the accuracy of data classification after the comparison of various models in terms of accuracy, strong and weak points considering their application in developing an assistant system of product form selection. The dataset used in this process constitutes the data which has not yet been clustered into 4 groups. The total number of 191 dataset was divided into two datasets; 80% as machine training data and 20% as model testing data. The test was run technique by technique for all 5 selected techniques. After that, the value of accuracy, precision and recall rates was evaluated and compared regarding advantage and disadvantage of each practical technique.

The results of the comparison as illustrated in Table III demonstrate that the first three techniques are Random Forest, Deep Learning, and Decision tree namely. The clusters classified with these three techniques have almost the same rate of accuracy at 83.4%. The highest rate of precision was provided by Decision Tree technique comparing to the others at 86.0%. Finally, Random Forest gave the highest rate of recall comparing to the others at 100%. Nevertheless, when we measured the convenience or difficulty of their function considering the restriction of this research, it is worth stating

that Random Forest and Deep Learning are not appropriate to apply in an easy-to-use and time-saving application prototype developing for testing innovation acceptance.

2) Identify the process of the classification of design element attributes

Due to the fact that it is necessary to classify a great number of variables, the importance of the data classification order is then put into account and has an impact on the accuracy. At this stage, the value of Attribute's Information Gain (IG) was selected to determine the order in attributes classification. The operator called "weight by information gain ratio" in

TABLE III: CLASSIFICATION TECHNIQUE COMPARISON

Technique	Performance (%)				Ease of Application
	Accuracy	Precision	Recall	Average	
Random Forest	83.4	81.1	100.0	88.1	Hard
Deep Learning	83.4	84.7	93.5	87.2	Hard
Decision Tree	83.4	86.0	91.6	87.0	Easy
Naïve Bayes	76.0	80.0	88.3	81.4	Easy
Logistic Regression	74.7	79.6	86.4	80.2	Easy

TABLE IV: CALCULATION OF ACCURACY RATE WITH AVERAGE SPECIFIC WEIGHT METHOD (LOW IG – HIGH IG)

Step	Design Element	Accuracy (Percentage)	Weight	Weight Average
1	Xa	100.00	0.15	15.00%
2	Xb	97.20	0.15	14.58%
3	Xc	96.80	0.15	14.52%
4	Xd	82.90	0.10	8.29%
5	Xe	73.70	0.10	7.37%
6	Xf	68.20	0.05	3.41%
7	Xh	87.00	0.10	8.70%
8	Xi	54.40	0.05	2.72%
9	Xj	87.00	0.10	8.70%
10	Xk	76.50	0.05	3.83%
Total weight average Accuracy				87.12%

RapidMiner program was used to calculate. The results were carried out ranging from the highest to the lowest which was believed to be the appropriate process. However, the test results showed us that the average rate of accuracy arrived at 85.24% with the least accurate rate of prediction of attribute Xa (28%) and Xb (52.53%). Consequently, we had try changing the order of classification, from the lowest value of IG to the highest. The results were unequivocal; the average rate of accuracy in prediction of all variables was at 82.37%, lower than that of the previous process as estimated. Be that as it may, such order has an impact on the accuracy in prediction of variable Xa at a very high rate (100%) and variable Xb at 97.2% which is in accordance to defined objective. The importance is placed on main design elements such as Xa, Xb or Xc with importance weight at 15% each. In the other hand, the others variables are estimated at 5% or 10% of importance weight. After calculating the accuracy rate using average specific weight method of the variable

from the lowest IG value to the highest (shown in table), it can be asserted that the accuracy rate was at 87.12% which was higher than the first method with 78.42% of accuracy rate.

As mentioned in the previous section, Decision Tree technique was selected to develop a Classification Model by ordering variables used in clustering beginning with Xk and finishing with Xb, the lowest IG to the highest IG, as described in section 1 and 2. At this stage, the RapidMiner version 8.1 was applied for developing and testing the model with Decision Tree Algorithm following the process shown in Fig. 3

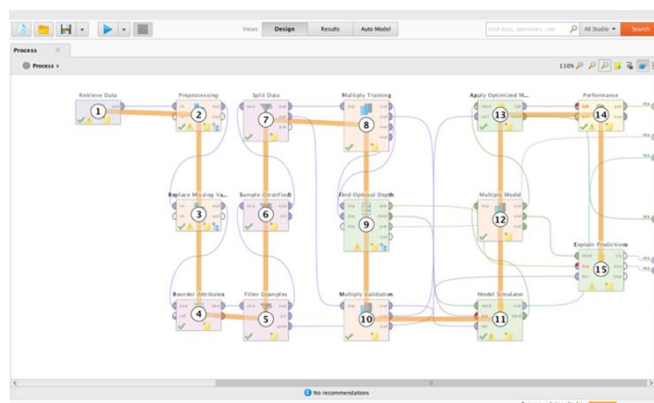


Fig. 3. Steps of a model developing using RapidMiner program.

First of all, the data was entered to the program with a view to develop a tool which supports the verification and the replacement of missing variable value or to detect the anomaly in order to reduce noise affecting the data. After that, the variables were ordered and made sure that the label completed in every data line. In case of missing line, the dataset would be removed. To avoid the problem of overfitting where the model is developed and tested by the same dataset, before the process, the dataset must be divided into two parts. The first part is used for model developing; the other part is kept to effectuate a test of model accuracy (operator #7). The data was divided randomly, and the dataset was copied to operate others tasks simultaneously (operator #8 and #10). The dataset which was randomized was sent to the operator called "Find optimal Depth" (operator #9) using Decision Tree as the main technique of data classification. Within this operator, the test of efficiency was also be effectuated. Consequently, data classification model is made from the previous step. The efficiency enhancement of the model was developed with the operator called "Apply optimized model" (operator #13). Finally, the efficiency of the model was calculated by k-folds validation with the operator called "Performance" (operator #14), the process was repeated 10 times.

Due to the fact that the data used in model developing was divided into 4 groups following the clusters of users, the model developing for design element attributes from Xk to Xb was made in the total number of 4 times according to database.

D. Result

Let us recall the objective of this study, we aim to develop a model capable of predicting different types of design elements which is accurate and meets the customer needs

according to sentiment towards product forms. The result is brought back to create specific design works for users or to select the existing products which are regarded as the most appropriate to customer needs. The developed model consists of 4 sets for 4 groups of customers clustered along customer taste and preference variables. To validate the model, 10-folds cross validation was applied as a tool to test the accuracy of the model for each type of design element variables. The test was conducted separately group by group (according to customer clusters) in the total number of 4 times. The last time was reserved for the model accuracy test as a whole without user clusters. The result as shown in Table V asserts that the accuracy rate of attributes Xa, Xb, Xc which are regarded as important variable group and are weighted more than any others variables at 95% upward. The lowest rate of accuracy of variable can be found in the application of contrast theory (Xi) in designing followed by Positive/Negative Space in Design (Xf).

TABLE IV: CLASSIFICATION ACCURACY OF EACH DESIGN ELEMENTS

	XA	XB	XC	XD	XE	XI	XF	XK	AVERAGE
SEGMENT 1	100.0	97.3	97.9	91.6	74.3	64.2	74.9	74.9	84.4
SEGMENT 2	96.7	97.6	96.0	85.7	71.7	61.3	69.0	74.4	81.6
SEGMENT 3	100.0	97.9	95.5	88.1	73.4	53.7	67.1	77.3	81.6
SEGMENT 4	100.0	97.4	97.0	87.2	78.4	63.0	70.3	77.4	83.8
COMBINED SEGMENT	100.0	97.2	96.8	82.9	73.7	54.4	68.2	76.5	81.2

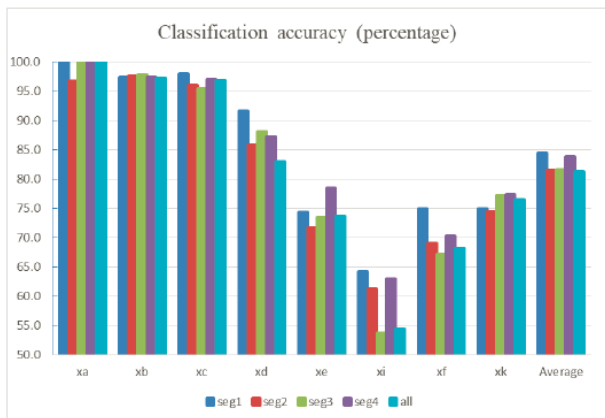


Fig. 4. Classification's accuracy of 4 segment datasets and a mixed dataset.

The average accuracy rate of classification of all variables of the 1st cluster was at 84.4% (the highest) followed by 83.8% in the 2nd cluster. The data of all respondents without clustering provided the result of the lowest accuracy rate at 81.2%. However, the average accuracy rate of all models was more than 80% and considered apt for applying to develop product form co-design assistance prototype system as defined by the objective.

In order to evaluate the application of the model from this study, the researchers developed a system prototype of product form co-design assistance which had been used the algorithm from the developed KE-Ai model. The test was randomized in different target groups by questionnaire. The respondents gave their answers ranging from 1 to 5 after trial on the prototype of the proposed system so as to evaluate by 51 prototype testers. The selected results of the questionnaire were shown in Table V. It represents that the model is accuracy enough to classify the respondents into one of the 4 clusters such a way to evaluate the performance of the

clustering model. The result shows that the average score is at 4.00. For the performance of Design Elements classification, the result from the question number 2 which is related to product recommendation efficiency from users' sentiments is scored at 3.98. Therefore, we can interpret that the developed KE-Ai model works with satisfaction and can be extended to further development

TABLE V: EVALUATION OF THE SYSTEM PROTOTYPE

Question	Frequency (N=51)					Mean	S.D.
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
Prototype system can sort you into user clusters considering your tastes and preferences	1	0	13	21	16	4	0.872
(Percentage)	1.90%	0.00%	25.00%	40.40%	30.80%		
Prototype system can recommend you the samples of desired stencil products	0	4	6	28	13	3.98	0.836
(Percentage)	0.00%	7.70%	11.50%	53.80%	25.00%		

VII. CONCLUSION

This study explains the application of data mining technique combining with Kansei Engineering technique to predict possible design elements of a product which is appropriate for users. The sentiments are indicated by 5 pair-words constituting the system data. In addition, the users are divided into 4 clusters with different tastes and preferences such a way to increase the accuracy of the prediction. The model developing uses 4 datasets from users in 4 clusters. The developed model leads to the creation of a prototype of product form co-design assistance system. The result from the evaluation of the prototype system points out that the prediction, users and product recommendation using sentiments expression are satisfactory. The conclusion can be drawn from this research as follows:

- 1) Classification with style & Design preference criteria by k-means technique engenders 4 different clusters of users. The factors impacting on classification could be enumerated as the familiarity of design by career, family members, close friends including environment. There are also factors like personal preference regarding design elements. It all boils down to stating that the 4 clusters possess different design acumen.
- 2) Through the comparison of various classification techniques to predict design elements using collected dataset, it is worth saying that Random forest and Deep Learning techniques obtain the higher score in terms of precision and accuracy. However, Decision Tree technique offers an advantage of being fast and easy to use. Thus the researcher selected Decision Tree technique for this study. This technique ranked third in terms of accuracy and precision, the score was lower than the others technique by 1.1% and 0.2% respectively.
- 3) For the test, the model is divided into 4 sets and was used to predict 4 clusters of users. This method offers more accurate prediction of design elements comparing to the

method which undivided users.

The validation result of the model of prediction of design elements is at a very high rate, the score is more than 80%. The verification of the operation of system prototype offers the correspondent result: 4 out of 5. However, a closer look into the accuracy in predicting design element variables reveals that they obtain the score less than the others variables. Our feeling is that design theory which is classified to as element is hard to interpret. Even though it is one of the most used standard theories, the application or the decision to use such theory for a product form still presents variation comparing to others design elements including main form, secondary form used in design which is clearer. As a result, an element which can be clearly classified and indifferent to ambiguous element exerts an impact on model accuracy. In addition, it is said to be more advantageous for the operation of the system as long as such design element does not have highest importance or direct influence on product form studied in this research.

REFERENCES

[1] M. J. Muller, "Participatory design: The third space in HCI," *Human-Computer Interaction: Development Process*, 2003, vol. 4235, pp. 1051-1068.

[2] D. A. Norman and S. W. Draper, "User centered system design, new perspectives on human-computer interaction," *Annals of Physics*, 1986, p. 544.

[3] J. Füller, "Virtual co-creation of new products and its impact on consumers' product and brand relationships," *Academy of Management Annual Meeting Proceedings*, 2010, vol. 8, pp. 1-6.

[4] M. Steen *et al.*, "Using interactive model simulations in co-design: An experiment in urban design," *CoDesign*, 2013, vol. 9, pp. 2-16.

[5] A. Gandhi, C. Magar, and R. Roberts, "How technology can drive the next wave of mass customization," *MC Kinsey & Company's Business Technology Journal*, 2014, p. 1-8.

[6] M. Nagamachi, "Kansei Engineering: An ergonomic technology for product development," *International Journal of Industrial Ergonomics*, 1995, vol. 15, p. 1.

[7] S. Ishihara, *et al.*, "An analysis of Kansei structure on shoes using self-organizing neural networks," *International Journal of Industrial Ergonomics*, 1997, vol. 19, pp. 93-104.

[8] A. Gentner *et al.*, *Kansei Design Approaches for the New Concept Development Process*, 2012.

[9] S. Schütte, "20 affective meaning: The Kansei engineering approach," *Engineering*, 2008, pp. 477-496.

[10] S. H. Hsu, M. C. Chuang, and C. C. Chang, "A semantic differential study of designers' and users' product form perception," *International Journal of Industrial Ergonomics*, 2000, vol. 25, pp. 375-391.

[11] H. Jiang *et al.*, "A methodology of integrating affective design with defining engineering specifications for product design," *International Journal of Production Research*, 2014, pp. 1-17.

[12] M. Nagamachi, "Kansei engineering — a new ergonomic consumer-oriented technology for product development," *International Journal of Industrial Ergonomics*, 1995, vol. 15, no. 1, pp. 3-11.

[13] Y. C. Lin, C. C. Chen, and C. H. Yeh, "Intelligent decision support for new product development: A consumer-oriented approach," *Applied Mathematics & Information Sciences*, 2014, vol. 8, no. 6, pp. 2761-2768.

[14] O. Adelabu and T. Yamanaka, "Kansei as a function of aesthetic experience in product design," *Industrial Applications of Affective Engineering*, 2014, Springer International Publishing, pp. 83-95.

[15] M. Nagamachi and A. M. Lokman, "Kansei innovation: Practical design applications for product and service development, in Industrial innovation series," 2015.

[16] T. Jindo, K. Hirasago, and M. Nagamachi, "Development of a design support system for office chairs using 3-D graphics," *International Journal of Industrial Ergonomics*, 1995, vol. 15, no. 1, pp. 49-62.

[17] S. Boqiang and S. Changzheng, "Emotional optimized design of electro-hydraulic actuators," *Sensors & Transducers*, 2014, vol. 177, no. 8, pp. 191-195.

[18] M. Bergman *et al.*, "Surface design methodology – challenge the steel," presented at International Conference on Metrology and Properties of Engineering Surfaces, 2014.

[19] S. Schutte *et al.*, *Affective Meaning: The Kansei Engineering Approach*.

[20] P. H. Bloch, "Seeking the ideal form: Product design and consumer response," *Journal of Marketing*, 1995, vol. 59, pp. 16-29.

[21] J. F. Petiot and B. Yannou, "Measuring consumer perceptions for a better comprehension, specification and assessment of product semantics," *International Journal of Industrial Ergonomics*, 2004, vol. 33, pp. 507-525.

[22] R. P. C. J. Rajapakse *et al.*, "A study on gender-kansei of three-dimensional geometric shapes," *International Journal of Biometrics*, 2012, vol. 4, p. 388.

[23] Y. Truong *et al.*, "Consumer response to product form in technology-based industries," *Journal of Product Innovation Management*, 2014, vol. 31, no. 4, pp. 867-876.

[24] S. T. Wierchoń and M. A. Kłopotek, *Modern Algorithms of Cluster Analysis*, 2018, p. 421.

[25] C. C. Aggarwal, *Data Mining: The Textbook*, Springer International Publishing, 2015, p. 746.

[26] L. Kaščelan, V. Kaščelan, and M. Jovanović, "Analysis of investors' preferences in the Montenegro stock market using data mining techniques," *Economic Research-Ekonomska Istraživanja*, 2014, vol. 27, pp. 1-20.

[27] Z. Dai and M. J. Scott, "Product platform design through sensitivity analysis and cluster analysis," *Journal of Intelligent Manufacturing*, 2007, vol. 18, pp. 97-113.

[28] J. J. Armstrong *et al.*, "K-means cluster analysis of rehabilitation service users in the home health care system of Ontario: Examining the heterogeneity of a complex geriatric population," *Archives of Physical Medicine and Rehabilitation*, 2012, vol. 93, pp. 2198-2205.

[29] J. Jiao, Y. Zhang, and M. Helander, "A Kansei mining system for affective design," *Expert Systems with Applications*, 2006, vol. 30, pp. 658-673.

[30] N. Alldrin, A. Smith, and D. Turnbull, *Clustering with EM and K-means*. University of San Diego, California, 2003, pp. 261-95.

[31] S. Salvador and P. Chan, "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," in *Proc. International Conference on Tools with Artificial Intelligence*, 2004, pp. 576-584.

[32] T. Jindo, K. Hirasago, and M. Nagamachi, "Developing a qualitative sense," *International Journal of Industrial Ergonomics*, pp. 49-62.

[33] C. C. Wei, M. Y. Ma, and Y. C. Lin, "Applying Kansei Engineering to decision making in fragrance form design," *Smart Innovation, Systems and Technologies*, 2011, pp. 85-94.



development process.

Kittipong Sakornsathien was born in Bangkok, Thailand, 1979. He holds a bachelor degree in economics from Chulalongkorn University, Thailand, a master degree in business administration from Assumption University, Thailand. He is a Ph.D. candidate in School of Technopreneurship and Innovation Management at Chulalongkorn University, Thailand. His research interest is in data mining and co-design in product



Sukree Sinthupinyo holds a bachelor degree, a master degree and a Ph.D. in computer engineering from Chulalongkorn University, Thailand. He is an associate professor at Chulalongkorn University. His main areas of research are artificial intelligence, machine learning & pattern recognition, and engineering.



innovation, and product development.

Pongpun Anantavoranich holds a bachelor degree in industrial design of Chulalongkorn University, Thailand, a master degree (fine arts) of Pratt Institute, USA, and a Ph.D. (ceramic design education) of the Ohio State University, USA. Currently he is working as a lecturer and he has been an assistant professor doctor in the Department of Industrial Design, Chulalongkorn University. His research areas are product design,