Analysis of factors affecting employee loyalty of PT X in Jakarta region

R P Karimah, S Abdullah and R Setiadi
Department of Mathematics, Faculty of Mathematics and Natural Science (FMIPA), Universitas Indonesia, Depok 16424, Indonesia

Corresponding author’s email: sarini@sci.ui.ac.id

Abstract. Having loyal employees, i.e. those who desire to do their best in their work to achieve the company’s target, is very important for the company. This study is conducted to identify factors that explain the loyalty of employees in PT X, Jakarta. Furthermore, we also aim to generate profiles of the most loyal employees. Questionnaires were used to collect data from the central office and 12 branches of the company, using purposive sampling scheme, giving 467 respondents in total. Data processing was conducted using Partial Least Square (PLS) and Classification and Regression Tree (CART) methods. The results showed that job satisfaction and gender directly influence employees’ loyalty. Moreover, physiological needs, safety needs, love needs, and self-actualization needs indirectly influence loyalty of employees through job satisfaction. Highly loyal employees are those with either high job satisfaction and low physiological needs or low job satisfaction and low level of love needs.

Keywords: CART, employee loyalty, partial least square, purposive sampling, job satisfaction

1. Introduction
Currently, the business world in Indonesia is experiencing rapid growth [1]. This development is marked by the increasing number of companies, resulting in increasingly intense levels of competition between companies. This condition requires the company to have certain advantages in order to compete. These advantages can be obtained by having qualified human resources and have a high loyalty. Human resources are the main driver of the company in achieving company goals. Therefore, it takes human resources that have a high level of loyalty to the company. According to Reichheld in [2] the higher the loyalty of employees in an organization, the easier it will be for the organization to achieve organizational goals set by the organization’s owner. Regarding gender, female employees tend to be more loyal to their company than male employees [3].

One of the factors affecting employee loyalty is job satisfaction [4]. Some research [2, 5, 6] showed a positive relationship between job satisfaction and employee loyalty. Job satisfaction can occur when individual needs are fulfilled [7]. According to [8], humans have five basic needs that form a hierarchy and can only arise when the needs at the previous level have been met. Starting from the lowest level of the physiological needs, the needs of security, love, self-esteem, and self-actualization at the highest level. Maslow stressed that the five levels are not absolute and rigid [8]. There are some exceptions as some consider self-esteem to be far more important than the needs of loving or being loved.

PT X is a banking company that has a desire to be known for its excellent performance, human resources and teamwork. Therefore, PT X requires highly loyal employees for corporate goals to be
This research was conducted to evaluate the loyalty of this company’s employees, in accordance with the company’s goals.

Two main objectives of this research are: to model the simultaneous relationship among factors related to employees’ loyalty in PT X, and further to identify factors that can explain employees’ loyalty and generate the profile of high loyal employees.

2. Data and methods

2.1. Data

Data were collected through questionnaire, distributed to the employees in the head office and 12 branch offices. Purposive sampling scheme was employed, resulting in response from 467 employees.

Data consisted of socio-demographics information, as well as factors such as: employees’ loyalty, job satisfaction, and the fulfilment of need in the aspects of physiological, safety, love, self-esteem, and self-actualization. The fulfilment of each of the needs is by using several items questions, where each item is on a 5-scale Likert measurement. Then, the score on each of the needs factor will be added to produce the score for the corresponding aspect, and each score is then transformed into 2 classes: high and low categories.

Employees’ loyalty is the willingness of employees to work as well as possible and sacrifice to achieve company goals such as willingness to work overtime without overtime payment, committed to advancing the company and so on. Job satisfaction is a positive feeling about a person's work resulting from an evaluation of his or her characteristics including salary, company policy, work situation, and achievement. Physiological needs include the most basic human needs to sustain life physically, including the needs of eating, drinking, water, oxygen, shelter, and rest. This need level is measured by a Likert Scale of 1 to 5 using 4 items. Safety needs include human security needs which include physical security, environment, health, and freedom from worry or fear or anxiety. Love needs include one's need for love, compassion, need to have, the desire to be friends, the need to give and get love and the need to be part of a family, community, and environment. Self-esteem needs include one's need for achievement, acknowledgment, appreciation, and confidence. This need level is measured by a Likert Scale 1 to 5 using 5 items. Self-actualization needs include the need for a person to grow, develop using his or her abilities, and make himself or herself a better person than before.

2.2. Method

2.2.1. Partial least square. Partial Least Square (PLS) is a method used to look at patterns of relationship between latent variables with indicator variables and between latent variables with other latent variables. The relationship between the latent variable and the indicator variable is called the outer model. The relationship between latent variables and other latent variables is called the inner model. PLS does not require any assumption of distribution on the indicator variables. PLS can be used for formative and reflective constructs of latent variables. Reflective construct is the construct of the latent variable measured by the reflection of the indicators. Formative construct is latent variable construct formed by indicators that are the cause of latent variables. The variables involved in the PLS can have nominal, ordinal, interval or ratio measurements [9].

For reflective constructs, the outer model equations can be written as follows [10]:

\[ E(X_{jk} | LV_j) = \lambda_{0jk} + \lambda_{jk} LV_j \]
For formative constructs, the outer model equation between latent variables \( j \) which has \( p \) indicators can be written as follows [10]:

\[
E(LV_j | X_{j1}, X_{j2}, ..., X_{jp}) = v_{0j} + \sum_{k=1}^{p} v_{jk} X_{jk}
\]  

(1)

The inner model equation can be written as follows [10]:

\[
E(LV_j | LV_i) = \beta_{i0} + \sum_{\beta_{ij}} \beta_{ij} LV_i
\]  

(2)

\( X_{jk} \): Value of indicator variable \( k \) of latent variable \( j \)

\( \lambda_{0jk} \): Value of \( X_{jk} \) if value of \( LV_j = 0 \)

\( \lambda_{jk} \): Loading of indicator \( k \) on latent variable \( j \) if the construct of the latent variable is reflective

\( v_{0j} \): Value of \( LV_j \) if value of \( X_{jk} = 0 \) for each \( k \)

\( v_{jk} \): Weight of indicator \( k \) on latent variable \( j \) if the construct of the latent variable is formative

\( LV_j \): Value of latent variable \( j \)

\( \beta_{i0} \): Value \( LV_j \) if value of \( LV_i = 0 \) for each \( i \)

\( \beta_{ij} \): Path coefficient of latent variable \( i \) to latent variable \( j \) if latent variable construct \( j \) influenced latent variable \( i \)

\( LV_i \): Value of latent variable \( i \)

Estimation of parameters in the Partial Least Square (PLS) model is used to estimate path coefficients that relate latent variables to other latent variables, as well as relate latent variables and indicator variables. The parameter estimation method in the PLS model is differentiated between parameter estimation on outer model and inner model. The parameters in the PLS model to be estimated are as follows:

(a) Outer loading is the correlation of the indicator variable and the latent variable of the reflective construct formed on the outer model.

(b) Path coefficient or path coefficient is the amount of influence between latent variables on the inner model.

All parameters in PLS are estimated iteratively; thus, no distribution assumption is needed. After the parameter estimation algorithm in the PLS is executed and model parameters are estimated, model evaluation and significance of the path coefficient can be performed. Model evaluation is conducted for both outer and inner models.

Evaluation of outer models on reflective constructs can be conducted using the convergent validity, discriminant validity, and/or composite reliability. Convergent validity relates to the principle that the measurements of a construct should be highly correlated. Convergent validity test in PLS on reflective indicators is measured by outer loading which is correlation between a latent variable and its indicators. Outer loadings greater than 0.6 are acceptable [11]. Discriminant validity relates to the principle that the indicators of the construct of a different latent variable are not highly correlated. The construct of a latent variable can be said to satisfy discriminant validity if outer loading > cross-loading and cross-loading less than 0.7 [12]. Composite reliability is used to measure the internal consistency of the reflective construct of a latent variable, given by,

\[
\rho_j = \frac{\left(\sum_{k=1}^{p} \lambda_{jk}\right)^2}{\left(\sum_{k=1}^{p} \lambda_{jk}\right)^2 + \sum_{k=1}^{p} (1 - \lambda_{jk}^2)}
\]  

(3)
Composite reliability greater than 0.7 indicates that the indicator variables of the latent variable is reliable [9, 10].

On the other hand, evaluation of inner models can be conducted through either the path coefficient or the coefficient of determination. Path coefficient $\beta_{jl}$, quantifies the linear relationship between latent variables on the inner model and is obtained using the iteration method. Relationship with significant path coefficients are retained in the model [9]. Another quantity to assess the inner model is the coefficient of determination $R^2$ that measures the variability of the variable, which can be explained by explanatory variables, given by,

$$R_j^2 = 1 - \frac{\sum_{l=1}^{n}(y_{jl} - \hat{y}_{jl})^2}{\sum_{l=1}^{n}(y_{jl} - \bar{y}_j)^2}$$  \hspace{1cm} (4)

where $\hat{y}_{jl}$ is an estimate of the latent dependent variable obtained by the relationship $\hat{y}_{jl} = \sum_{l=1}^{n} \hat{\beta}_{jl}y_{il}$.

2.2.2. Classification and regression tree (CART). Classification and Regression Tree (CART) is a nonparametric statistical method used to perform classification analysis [13]. If the response variable is continuous, CART will produce the regression tree; if the response variable is categorical, CART will generate a classification tree. In this study, the dependent variable and all independent variables are categorical variables with two categories. Therefore, the discussion is devoted to the classification tree.

Classification tree method consists of three main procedure:

1. Splitting - the process of separating the parent nodes into two child nodes through certain splitting rules.
2. Stopping the tree-building process - the process of termination of the creation or formation of tree classification.
3. Class assignment is the process of identifying each node that is formed in a certain class.

The splitting process in the CART method starts from the root node containing all research data. The data in the root node will be split into two child nodes (right node and left node) based on the goodness of split criteria. This criterion will result in grouping on the right node and left node more homogeneous than the root node. In the CART method, the best crusher criteria are measured based on the Gini index, which is calculated as follows,

$$G(t) = \sum_{j=1}^{C} C(ij) p_j p_t$$  \hspace{1cm} (5)

where $p_j = \frac{N_j(t)}{N_t}$ is the proportion of the number of objects entering class $j$ on a node $t$, with $N_j(t)$ being the number of objects in class $j$ on a node $t$, $N_t$ is the sum of all objects in a node $t$, and $C(ij)$ is the weight of misclassification given if a class object $j$ is entered into class $i$ at node $t$. $C(ij) > 0$, $i \neq j$ and $C(i|j) = 0$, $i = j$. The smaller Gini index value on a node indicates that the better the process of separating objects into the class $j$.

Let $X_1, ..., X_n$ be the predictor variable. The parent node is denoted as $t_p$, while $t_R$ and $t_L$ are child nodes (right node and left node). In the CART method, we will look for a predictor variable $X_i$ that maximize the change of Gini index value on the parent node and the child node it creates. In other words, maximize the difference of Gini index value $\Delta G(X_i, t)$ with the following calculation:

$$\Delta G(X_i, t) = G(t_p) - p_L G(t_L) - p_R G(t_R)$$  \hspace{1cm} (6)

where $\Delta G(X_i, t)$ is the difference of Gini index value with $X_i$ the predictor variable acting as a divider, $i = 1, 2, ..., p$, at a node $t$, $G(t_p)$ is the Gini index value in the parent node, $G(t_R)$ is the Gini index value in the child right node, $G(t_L)$ is the value of Gini index in child left node, and $p_R$ and $p_L$ is the proportion...
of the number of objects entering the child right node and child left node. A previously formed child node will be a parent node and will re-splice until no increase in homogeneity in the child node is generated.

Next is the class assignment process. Suppose that the set of nodes produced by a classification tree T is denoted as \( T \), with \( T = t_1, t_2, ..., t_M \), and M is the number of nodes formed on the classification tree including the root node. A \( t_m \) node will be identified as a certain class if the proportion of the number of objects entering the class in the \( t_m \) node is greater than the proportion of the number of objects entering into the other classes on the \( t_m \) node.

Once the class assignment is performed, the tree’s performance can be evaluated through its accuracy. Accuracy is defined as the proportion of subjects that are assigned correctly, that is,

\[
\text{Accuracy} = \frac{TP + TN}{P + N}
\]

where TP, TN, P, and N are the number of true positives, true negatives, total positives, and total negatives, respectively, as described in table 1. Positive or negative classification is determined by the researcher. In this study, positive class is the high loyalty group, and negative class is the low loyalty group.

3. Results and discussion

According to figure 1, among the 467 respondents, 54.8% of them were female. Most of the respondents (53.32%) were no more than 30 years old; only 5.6% of them were more than 50 years old. According to their educational level, most of the respondents (72.59%) graduated with a bachelor’s degree, 3.4% graduated with a masters’ degree, and 9.85% of employees had nothing more than a high school diploma.

Table 1. Confusion matrix for classification results.

| Predicted | Observed | Positive | Negative |
|-----------|----------|----------|----------|
|           |          | True Positive (TP) | False Negative (FN) |
| Positive  |          | P         | N        |
| Negative  |          | False Positive (FP) | True Negative (TN) |

(a) (b) (c)

Figure 1. Descriptive of respondents based on (a) gender, (b) age, and (c) educational level.
Characteristics of respondents based on their length of employment, role at work, and monthly salary are represented in figure 2. Most respondents have been working for 5 to 10 years at the company. Based on the role at work, most of them are junior staff. Around 50% of the respondents are those with a salary of up to 5 million; the smallest proportion (around 6%) has a high salary of more than 15 million.

Reliability and validity of items in the questionnaire were also checked, and those that met the assumption were used for further modelling. Model assessment was conducted for inner and outer models, after which model refinement was applied, producing the following results (figure 3).

In figure 3, a rectangle signifies an indicator variable, a circle or ellipse signifies a latent variable and the straight arrows indicate that the variable at the base of the arrow affects the variable at the end of the arrow. It can be inferred from the figure that employees’ loyalty was associated with job satisfaction and gender. Moreover, physiological, security, love and self-actualization needs affect employee loyalty indirectly through job satisfaction.

**Figure 2.** Description of respondents based on (a) employment duration, and (b) role at work.

**Figure 3.** Final model fit obtained after the evaluation of outer and inner models.
Furthermore, it can also be seen that job satisfaction is a factor strongly associated with employees’ loyalty; the more satisfied employees are with their jobs, the more loyal they will be to the company, and females are more loyal than males.

As for job satisfaction, it is strongly associated with the need for self-actualization, in a way that higher needs of self-actualization tend to decrease job satisfaction. Similar results were obtained for physiological needs and security needs, while love needs were positively associated with job satisfaction. This result indicates the need to love and be loved by employees was fulfilled. More insights that can be drawn from the results are:

(a) The most powerful indicator reflects employee loyalty variable is LO3, i.e. the employee is committed to advancing the company.
(b) The most powerful indicator reflects the satisfaction variable is K1, i.e. the employee feels the rules in the workplace are very clear and neat.
(c) The most powerful indicator reflects the self-actualization needs variable is SA2, i.e. the employee can solve the problems that occur in the office.
(d) The most powerful indicator reflects the loving and loving needs variable is LV1, i.e. the employees have a good family.
(e) The most powerful indicator reflects the security requirement variable is S1, i.e. the employee feels secure in their residential / work environment.
(f) The most powerful indicator reflecting the physiological needs variable is F3, i.e. the employee has access to clean drinking water.

**Figure 4.** Profiles for employees based on their loyalty. The numbers underneath the splitting variables (i.e. Job Satisfaction, Love, Physical, Gender, and Safety) represent the improvement on the child nodes due to the corresponding splitting.

**Table 2.** Percentage of suitability of classification tree.

| Observed | Predicted | Percent correct |
|----------|-----------|-----------------|
| High     | 209       | 133             | 61.10 %         |
| Low      | 30        | 95              | 76 %            |
| Overall Percentage | 51.20 % | 48.80 % | 65.10 % |
Employees’ loyalty and the associated factors that were generated by PLS method will be categorized into two categories: high and low. The cut-off was for categorization was determined based on researchers’ judgement. These categorized variables will further be used as inputs in the classification method based on the CART algorithm. The classification results are shown in figure 4.

According to figure 4, employee’s profiles that have high and low loyalty levels can be described as the following. Highly loyal employees are:

1. 90.7 % of female employees and 82.7 % of male employees have high levels of job satisfaction and have fulfilled their relative physiological and safety needs
2. 63.5 % of employees have low satisfaction levels but have met their love needs.

While for low loyal employees, we obtained the following:

1. 83.3 % of employees with high job satisfaction but have not fulfilled their relative physiological needs.
2. 66.7 % of employees with low job satisfaction and unfulfilled love needs.

Summarizing the classification results in figure 4, we obtain the number of respondents who were highly loyal to the company (based on the results of PLS) and were correctly classified, as well as those who were incorrectly classified. Similar results were also obtained for those with low loyalty. The result is summarized in table 2.

Based on table 2, the established classification tree has a percentage of conformity of 65.1% in describing the employee loyalty profile of PT X in the Jakarta region. Thus, it can be concluded that the model tree classification is good enough.

4. Conclusion
We were able to identify factors that directly affect employee loyalty of PT X in Jakarta region: job satisfaction and gender. Physiological, security, love and self-actualization needs affect employee loyalty indirectly due to their relationship through job satisfaction. Highly loyal employees were those with a high level of job satisfaction and have fulfilled their physiological needs; for employees who had low loyalty, it was the other way around.

Acknowledgments
We would like to acknowledge PT. X and all the staff that participated in this study. We also thank all reviewers for the improvement of this article.

References
[1] Fauzi Y 2016 *Jumlah Wirausahawan RI Bertambah 4 Juta Orang dalam 10 Tahun* available at https://www.cnnindonesia.com/ekonomi/20160819114219-78-152414/jumlah-wirausahawan -rit- bertambah-4-juta-orang-dalam-10-tahun
[2] Mariana R and Irfani H 2017 *Jurnal RAP* 6 193-202
[3] Anna L K 2014 *Mengapa Wanita Lebih Loyal di Suatu Kantor* available at https://lifestyle.kompas.com/read/2014/02/24/1603288/Mengapa.Wanita.Lebih.Loyal.di.Satu .Kantor
[4] Taurisa M C and Intan C R 2012 *Jurnal Bisnis dan Ekonomi* 19 170-187
[5] Basalamah A 2012 *Binus Business Review* 3 493-501
[6] Ningtyas C A and Amir M 2017 *Hubungan antara Kepuasan Kerja dengan Loyalitas Karyawan PT. Pativare Sungai Raya (Kab. Bengkayang) Kalimantan Barat* PhD Dissertation (Surakarta: Universitas Muhammadiyah Surakarta)
[7] Robbins S P dan Judge T A 2005 *Organizational Behavior* 15th edition (New Jersey: Prentice Hall) p. 1393
[8] Maslow A H 1989 A theory of human motivation in *Readings in Managerial Psychology* (Chicago: University of Chicago Press) pp. 20-35
[9] Hartono J 2011 *Konsep dan Aplikasi Structural Equation Modeling Berbasis Varian dalam Penelitian Bisnis* (Yogyakarta: UPP STIM YKPN)
[10] Sanchez G 2013 *PLS Path Modeling with R* (Berkeley: Trowchez Editions) p. 383
[11] Hair Jr J F, Hult G T M, Ringle C and Sarstedt M 2016 *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (California: Sage publications)
[12] Henseler J, Ringle C M and Sarstedt M 2015 *J. Acad. Mark.* 43 115-35
[13] Breiman L, Friedman J, Stone C J and Olshen R A 1984 *Classification and Regression Trees (Wadsworth Statistics/Probability)* (Boca Raton: Chapman and Hall/CRC) p. 368