Jurnal Mantik
Volume 4, Number 3, Nov 2020, pp. 1551-1559 E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik

Credit Card Risk Classification Using K-Nearest Neighbor Weighted Algorithm Based on Forward Selection

[1]Sartika Dewi Purba, [2]Pahala Sirait, [3]Arwin

[1,2,3] Information Technology Department, STMIK Mikroskil
Jl. M.H Thamrin No.140 Kel, Pusat Ps., Kec. Medan Kota, Kota Medan, Sumatera Utara 20212

e-mail: [1]sartikadewipurba85@gmail.com, [2]sirpahala@yahoo.com, [3]arwin@mikroskil.ac.id

ARTICLE INFO

Article history:
Received: 12/07/2020
Revised: 22/08/2020
Accepted: 30/09/2020

Keyword: Credit card risk, W-KNN algorithm, Forward selection

ABSTRACT

One form of credit card risk is non-performing credit cards, which describe a situation where loan repayment approval on credit cards runs the risk of failure. In the classification technique there are several algorithms that can be used, one algorithm that is often used is Weighted k-nearest neighbor (WKNN). This study aims to improve the performance of the Weighted k-nearest neighbor (WKNN) algorithm by applying the forward selection feature that is used to select each unused feature when starting a feature iteration, the results of the study show that by adding forward performance selection of the Weighted k-nearest algorithm neighbor (WKNN) get a better value that is 86.4%, compared to using the Weighted k-nearest neighbor (WKNN) algorithm without a forward selection that is equal to 60.1%.

1. Introduction

Credit is an important catalyst for economic growth and is a core activity of banks around the world [1]. The availability of credit enables households to consume better and enables companies to make investments that cannot be done with their own funds. In allocating or providing loans to potential customers or borrowers, commercial banks and financial institutions must consider the risk of the credit card that will be given, whether the customer will pay the credit card or not (problem credit card) [1].

Non-performing credit cards describe a situation where credit card return approvals run the risk of failure, and indicate that the company will incur potential losses. With the emergence of problem credit cards, the cash turnover rate at the company will be smaller. Even if the credit card problem is very large, then the company's cash circulation can stop and all the positive impacts that can be caused by credit card distribution cannot occur. [2]. For that, companies must plan and analyze credit cards in order to detect possible credit card risks. Credit card risk or often referred to as default risk is a risk due to the failure or inability of the customer to return the loan amount obtained from the company and the interest according to a predetermined or scheduled time period. [3]. Therefore, a reliable evaluation model for credit card risk plays an important role in loss control and revenue maximization [4]. One way of allocating capital to achieve a target is to make predictions about the customer's future ability to pay.

Credit card risk classification methods have played an important role in contemporary banking risk management practices. The classification method contributes to the key to the loan approval process that accurately and efficiently measures the credit card risk level of potential borrowers. This credit card classification method aims to predict future behavior in terms of credit card risk based on past experiences of customers with the same characteristics. The borrower's credit card risk level is associated with the possible risk of default on loans that are approved at a predetermined time. The main task of the credit card classification method is to provide separation between those who fail and those who do not when it comes to credit card payments. Separation ability is the main indicator of the success of a method [5]. To be useful and efficient, credit card scoring models must strike a good balance between classification performance and interpretability [6]. Today, classification performance has become increasingly important for credit card grading, because even a small percentage of the percentage increase means a large amount of profit for financial institutions [7] [8]

One of the most commonly used methods for credit card assessment is the k nearest neighbor (KNN). This method belongs to the category of nonparametric classification methods. It is known that non-parametric classifiers usually suffer from existing outliers, especially in situations of small training sample

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

size [9]. There are many credit card assessment researchers who have used the KNN to assess the risk associated with lending to an organization or individual [10]. Henley and Hand used the KNN classifier to assess consumer credit card ratings by proposing an adapted version of the Euclidean distance metric [11]. KNN is also used as a comparative method to a new method proposed by researchers in credit card scoring [12]. The WKNN method gives different weights to each of the closest neighbors k. Neighbors that are closer together will receive more weight than neighbors that are further away. The weighting process is carried out using the kernel function.

Forward Selection is a modeling method (linear model development) to find the "best" combination of variables from a group of variables. In the Forward selection procedure, once a variable is entered into the equation it cannot be eliminated. In addition, forward selection can mean entering the independent variable that has the closest correlation with the dependent variable (the variable with the most potential to have a linear relationship with Y). Then gradually enter the next potential independent variable and will stop until there is no more storage and increase algorithm speed, remove irrelevant features, develop and add data quality, speed up learning algorithm running time, develop and add data quality, and improve performance and model accuracy. [13]

In this study, researchers will apply the forward selection-based Weighted k-nearest neighbor (WKNN) weighting method for credit card risk assessment, and the problem for credit card scoring models that need to be considered is the unavailability of real-world credit card data, because credit card data. customers are confidential in most financial institutions and researchers cannot get access to this data. This study will review the performance of Weighted k-nearest neighbor (WKNN) based on forward selection for credit card score analysis compared to Weighted k-nearest neighbor (WKNN) without forward selection using the world credit card dataset published by the University of California Irivine (German dataset, 1994).) https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data) which has indeed been widely used in credit card assessment research so far.

2. Theoretical Basis

2.1 Data Mining

With the increasing number of data in the database, the role of analysts to analyze data manually needs to be replaced with computer-based applications that can analyze data automatically using more complex and sophisticated tools. Basically, data mining is closely related to data analysis and the use of software to look for patterns and similarities in a set of data. Data mining is a basic principle in arranging or sorting large amounts of data and retrieving information related to what is needed, as is usually done by an analyst (Rully, 2009).

Data mining is the extraction of implicit, previously unknown, and potentially useful information from data. The idea is to build a computer program that filters through the database automatically, looking for regularities or patterns. Strong patterns, if found, will likely generalize to make accurate predictions of future data. Anything found to be incorrect: There will be exceptions to every rule and cases that are not covered by any rules. Algorithms must be robust enough to deal with imperfect data and extract imprecise but useful regularities (Ian & Eibe 2005).

2.2 Classification

A very important part of data mining is the classification technique, which is how to study a set of data so that it produces rules that can classify or recognize new data that has never been studied. Classification can be defined as a process for stating a data object as one of the previously defined categories or classes (Zaki etal 2013). Classification is one of the methods that exist in data mining. In the classification, the labels of each class are predetermined. The classification process itself is a process of finding a model or distinguishing classes or data that can be used to predict the class of objects whose class labels are unknown. Classification is a form of data analysis that can describe model extracts from important data (Jiawei, et al., 2012).

Classification is the process of finding a model (or function) that describes and differentiates data classes or concepts that aim to be used to predict the class of objects whose class label is unknown. (Leidiyana Henny, 2013) The classification algorithm that is widely used, namely Decision / classification trees, Bayesian classifiers / Naïve Bayes classifiers, Neural networks, Statistical Analysis, Genetic Algorithms, Rough sets, k-nearest neighbor, Rule Based Method, Memory based reasoning, and Support vector machines (SVM).

Data classification consists of a 2-step process. The first is learning (the training phase), where a classification algorithm is created to analyze training data and then it is represented in the form of a

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

classification rule. The second process is classification, where test data is used to estimate the accuracy of the classification rule. (Leidiyana Henny, 2013)

2.3 Feature Selection

Feature Selection is a process commonly used in Machine Learning where a set of features that are owned by the data are used for algorithm learning. Feature selection has become an active research field in pattern recognition, statistics, and data mining (Oded Maimon, 2010).

Feature selection is one of the most important factors that can affect the level of classification accuracy because if the dataset contains a number of features, the dimensions of the dataset will be large, making the classification accuracy low. The problem in feature selection is dimension reduction, where initially all the attributes are required to obtain maximum accuracy. Four main reasons for reducing dimensions (Oded Maimon, 2010):

a) Decreasing the learning cost or reducing the cost of learning.
b) Increasing the learning performance or increasing learning performance.
c) Reducing irrelevant dimensions or reducing irrelevant dimensions.
d) Reducing redundant dimensions or reducing excessive dimensions.

The main idea of Feature Selection is to select a subset of existing features without transformation because not all features / attributes are relevant to the problem. Even some of these features or attributes are distracting and reduce accuracy. Noisy Features or unused features should be removed to improve accuracy. In addition, many features or attributes will slow down the computation process.

2.4 Forward Selection

Forward Selection is a modeling method (linear model development) to find the "best" combination of variables from a group of variables. In the Forward selection procedure, once a variable is entered into the equation it cannot be eliminated. In addition, forward selection can mean entering the independent variable that has the closest correlation with the dependent variable (the variable with the most potential to have a linear relationship with Y). Then gradually enter the next potential independent variable and it will stop until there are no more potential independent variables. (Kusuma Widya Intan, 2011)

The forward selection method is done by entering predictors in stages, this predictor is based on the largest partial correlation. In the forward selection method, the predictor variables that are included in the model can no longer be excluded. The process was terminated when the new predictors could not significantly increase the effect (sig below 0.05) on the response variable. That's why the forward selection procedure is one of the best model selection procedures in regression by eliminating the independent variables that builds the model in stages. There are several ways that can be used in testing with this forward selection method (Khrisna 2017).

In previous research, Forward selection was used to select every unused feature when starting a feature iteration (Zainuddin Sidik, 2019). And in this study also based on the results of classified credit card customers will use the forward selection feature selection to select unused attributes and can be expected to produce a better level of accuracy using the Weighted K-Nearest Neighbor algorithm.

2.5 K-Fold Cross Validation

K-fold cross-validation is widely adopted as the model selection criterion. In K-fold cross validation, folds are used for model construction and hold-out folds are allocated for model validation. This implies more emphasis on model construction than model validation procedures. However, several studies have revealed that more emphasis on validation procedures can result in better model selection. In particular, leave-m-outcross validation with n samples can achieve consistency in variable selection when m / n approaches 1 (Jung, 2017).

K-Fold Cross Validation is a validation technique that divides the data into k parts and then each part will be carried out a classification process. By using K-Fold Cross Validation, k experiments will be carried out. Each experiment will use one testing data and the k-1 part will be the training data, then the testing data will be exchanged with one training data so that for each experiment different testing data will be obtained. (Anto, 2015).

K-Fold Cross Validation is a good alternative to Random Subsampling. In this approach each data is used in the same amount for training and exactly 1 time for testing. The illustration is as follows. Suppose that the dataset is broken down into dbm parts of the same size. First one part of the training is selected while the other is for testing. Next is to swap roles, the part that used to be the training data set is now changed into test data sets, and vice versa. This approach is called two-fold cros-validation. The total error is obtained by adding up the errors obtained from the two processes. Each data has the opportunity to become test data once and once to become training data (Tan etal, 2006).

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

2.6 Weighted K Nearest Neighbor (WKNN)

The WKNN method is a development of the KNN method. This method uses a weighting principle. The weight will be given less to the number of k neighbors who are from the majority class, and vice versa for the minority class. The Weighted k-nearest neighbor (WKNN) method is able to classify well, because this method is suitable for implementation to data that is not evenly distributed (Indriati and Ridok, 2016). The algorithm step in the Weighted k-nearest neighbor (WKNN) method is not much different from the KNN algorithm step, the difference is that there is a weighting for each type / class and the process of calculating the score to determine the classification of the test data (Faldy, 2014).

The steps in the Weighted k-nearest neighbor (WKNN) algorithm according to D.A Adeniyi et al (2016) are as follows:

a) Determine the value of the K variable
b) Calculate the value of the proximity of the test data to the training data using the Euclidean Distanceor Cosine Similiraty Equation (CosSim)
c) Calculating the nearest neighbor, namely by calculating the distance between the training data and the test data using the Euclidean Distance formula, you can use Equation 1:

$$d(x_1, x_2) = \sqrt{\sum_{i=1}^{n} (x_2 - x_1)} \, 2 \qquad (1)$$

Information :
X1 = training data values
X2 = test data value
n = the amount of data
i = data to-i

a) Calculating the nearest neighbor, namely by calculating closeness using the Cosine Simimilary formula, you can use Equation 2:

$$(q, d_j) = \frac{\bar{d}_j \cdot \bar{q}}{|\bar{d}_j| \cdot |\bar{q}|} = \frac{\sum_{i=1}^{m} (w_{ij} \cdot w_{ig})}{\sqrt{\sum_{i=1}^{m} w_{ij}^{2} \cdot \sum_{i=1}^{m} w_{ig}^{2}}} \qquad (2)$$

Information :
Cosine Similarity (q,di) = the similarity value of the q test data with the dj training data
q = test data
dj = training data
$\overrightarrow{d_j \blacksquare \vec{q}}$ = the total result of the vector multiplication between the training data and the test data
$|\vec{d_j}| \blacksquare |\vec{q}|$ = the total result of vector multiplication between normalized training data and test data
wij = weight value i on training data j
wiq = the weight of the value i on the q test data
m = the number of values

b) Sort the results of the distance or proximity calculation into groups that have a similarity or distance.
c) Collect the nearest neighbor classification category
d) The weight calculation used Equation 2.3

$$weight_i = \frac{1}{\left(\frac{Num\left(c_i^d\right)}{Min\left\{Num\left(c_j^d\right) \; j=1,..k*\right\}} \right)^{1/\exp}} \qquad (3)$$

Information :
$Num\,(c_i^d)$ = The amount of training data d in class i
$Num\,(c_j^d)$ = The amount of training data d is in class j, where j is contained in the k set of closest neighbors

Exp = Exponent (exp value is more than 1) Each data that has been calculated the weight value will be used to calculate the score value. Where the result of the weight value will be multiplied by the score result equation.
The score result formula is calculated using Equations 2.4 and 2.5.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

$$Skor(X, C_i) = Weight_i \cdot \left(\sum_{djKNN}((\sqrt{\sum_{i=1}^{n}(x_{zi} - x_{ki})^2}) \cdot \delta(d_j, c_i)\right)$$

or

$$Skor(X, C_i) = Weight_i \cdot \left(\sum_{djKNN}((\sqrt{\sum_{i=1}^{n}(x_{zi} - x_{ki})^2}) \cdot \delta(d_j, c_i)\right)$$

Information :

Weighti = specific gravity / class i

djNKWNN(x) = dj training data on the nearest neighbor set from the x test data

$\sqrt{\sum_{i=1}^{n} (x_{2i} - x_{1i})^2}$ = the distance between test data and training data

$\delta(dj,Ci)$ = will be 1 if the distance value is Ci and 0 if the distance value is Ci.

Sim (q,dj) = Cosine Similiraty value between test data and training data

Ci = type or class i

In previous research, credit card scoring is a quantitative method for evaluating credit card risk from credit card applications. The method used in analyzing credit card risk to help them decide whether or not to apply for a credit card. This method aims to classify future behavior in terms of credit card risk based on past experience submissions. And this study uses the Weighted k-nearest neighbor (WKNN) method for credit card assessment by considering several attributes and available data sets.

3. Tables

In previous research, credit card scoring is a quantitative method for evaluating credit card risk from credit card applications. The method used in analyzing credit card risk to help them decide whether or not to apply for a credit card. This method aims to classify future behavior in terms of credit card risk based on past experience submissions. And this study uses the Weighted k-nearest neighbor (WKNN) method for credit card assessment by considering several attributes and available data sets.

Forward Selection Process:

In this study, the data to be tested has 20 features. The feature selection will be done using the forward selection method.

Step-1: Feature selection

The first feature is a feature that must be inputted (must be selected) in feature selection. Of the 20 features, 1 will be selected as the first feature, and the remaining 19 features will be independent features to be selected. The first selection is done randomly by randomizing the position of the feature. This means that the feature position when inputted, could be in the position of the first feature, but after shuffling is done, other features can become the first feature. From the results of the tests conducted, 16 best features were selected, namely features 1,2,7,8,9,10,11,12,13,14,15,16,17,18,19,20. Then these 16 best features will be used to determine the accuracy value.

In the calculations below, the initial 2 feature tests will be demonstrated. Suppose the 2 features selected are features 2 and 3.

Table 1

Training data with 2 features (feature-2 and feature-5)

DATA to	Feature-2	Feature-5	class
101	0,29	0,07	1
102	0,47	0,11	1
103	0,03	0,04	1
104	0,07	0,09	1
105	0,12	0,12	1
106	0,29	0,64	2
107	0,21	0,34	2
108	0,12	0,32	1
109	0,29	0,41	1
110	0,15	0,06	1
…	...	...	
1000	0,60	0,24	1

Table 2

Testing data (Testing)

Data Training	Features-2	Features-5	Class
1	0 ,03	0,05	1

Step 2: Calculate the classification accuracy with the WK-NN

In the first test, suppose the selected feature is used only feature 2 and feature 5 which consists of 1000 data. Then the data classification accuracy is calculated using the WKNN algorithm.

3.1 K-NN Weighted Process

a) Determine the value of the variable k, say k = 2. Calculating the value of the proximity between the test data and the training data using the Euclidean Distance equation. Calculating the nearest neighbor, namely by calculating the distance between the training data and the test data using the Euclidean Distance formula, you can use Equation 1:

Information :

X1 = value of training data

X2 = test data value

n = amount of data

i = ith data

The following shows the calculation of the distance between Test Data-1 (Data-1) to Training-1 Data (Data-101) for 10-fold:

The distance calculation is done in the same way until the 900th data (1000th data), and it is repeated for all test data (up to the 100th data).

Table 3

Distance of Euclidean Data-1 to Training data (data 101-1000)

Data Training	Feature-2	Feature-5	Distance with D-1	Class
101	0,29	0,07	0,265	1
102	0,47	0,11	0,446	1
103	0,03	0,04	0,013	1
104	0,07	0,09	0,060	1
105	0,12	0,12	0,113	1
106	0,29	0,64	0,649	2
107	0,21	0,34	0,340	2
108	0,12	0,32	0,284	1
109	0,29	0,41	0,447	1
110	0,15	0,06	0,118	1
…	…	…	...	
1000	0,60	0,24	0,600	1

b) Sort the results of the calculation of distance or closeness to groups that have similarity or distance.

Table 4

Ranking of Train data with the smallest distance

Data Training	Feature-2	Feature-5	Distance with D-1	Rank	Class
712	0,03	0,05	0,002	1	2
519	0,03	0,05	0,002	2	1
248	0,03	0,05	0,002	3	1
486	0,03	0,05	0,002	4	2
793	0,03	0,05	0,003	5	1
696	0,03	0,05	0,004	6	1
493	0,03	0,05	0,004	7	1
474	0,03	0,05	0,004	8	1
330	0,03	0,05	0,006	9	1
664	0,03	0,04	0,007	10	1
916	0,65	1,00	1,133		

For k = 2, it means that the neighbor data from the 1-test data are 712 data which are in class 2 and data 519 which are in class 1.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

c) Collect the nearest neighbor classification category
In the training data which consists of 900 data, namely the 101st data to the 900th data, there are 625 data that are in class 1, and 275 data that are in class 2.

d) The weight calculation used Equation 2.3, which is for k = 2 neighbors.
Weight class 1:
weight Class 2:

e) Each data that has been calculated the weight value will be used to calculate the score value, where the result of the weighted value will be multiplied by the score result equation.
The score result formula is calculated using Equations 2.4 and 2.5.

Information :
Weight = specific gravity / class i
djNKWNN (x) = dj training data on the nearest neighbor set from the x test data
= distance between test data and training data
$\delta\ (dj, Ci)$ = will be worth 1 if the value of distance is ∈ Ci and is worth 0 if the value of the distance is ∉Ci.
Sim (q, dj) = Cosine Similiraty value between test data and training data
Ci = type or class i
Then:
For the 1st neighbor (k = 2) from table 4.5, namely 712 data who are in class 2, the score is obtained:
For the 2nd neighbor (from 2 neighbors, k = 2), namely the data 519 who are in class 1, the score is obtained:

Then perform class predictions for the test-1 data, namely by comparing the weighted scores that have been calculated. It can be seen that the class 1 weight score of 0.00008 is smaller than the class 2 weight score of 0.00012, so it is concluded, the 1-test data is predicted to enter class 2.

The calculation is continued for the next test data until the 100th test data. The test is repeated against the second 100 data (2nd fold), and so on until the 10th fold.Then the accuracy of each fold is calculated as validation, namely by calculating the predicted amount of test data. correctly divided by a lot of test data. The average accuracy of the 10 folds becomes the final accuracy.

Step 3:

Then the next 1 feature is selected, so that the current data input consists of 3 data features. Then the calculation of the accuracy of the data classification is done again using the WKNN algorithm. Then the accuracy obtained in step 3 is compared with the accuracy obtained from the previous step. If the accuracy after the second feature is added is better than the accuracy before it was added, then the newly added feature will be preserved (feature selected). If the accuracy after the next feature is added is not better than the accuracy before it was added, then the newly added feature will be reduced (the feature is not selected).

Step 4:

Repeat step 3 until the last feature is entered. The features that are retained will be the selected features and will be used as prediction models. The accuracy of credit card data classification in the test scenario is presented in table 6.

Table 5
Accuracy of WKNN classification results by Forward Selection For k = 2 to 30

Value k	Accuracy 10-fold
2	**86,4**
3	66,7
4	84,9
5	70,0
6	73.1
7	70,0
8	70,0
9	70,0
10	70,0
11	70,0
12	70,0

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Value k	Accuracy 10-fold
13	70,0
14	70,0
15	70,0
16	70,0
17	70,0
18	70,0
19	70,0
20	70,0
21	70,0
22	70,0
23	70,0
24	70,0
25	70,0
26	70,0
27	70,0
28	70,0
29	70,0
30	70,0

4. Graphics Content

For testing with validation using a variation of the value k with validation 10-? Fold. The test results show that the accuracy results for the greater k value are not increasing and always around 70% accuracy. For all tests, the best accuracy is obtained always at the neighbor value k = 2. From the variation of testing, the best overall accuracy is obtained at 86.4%.

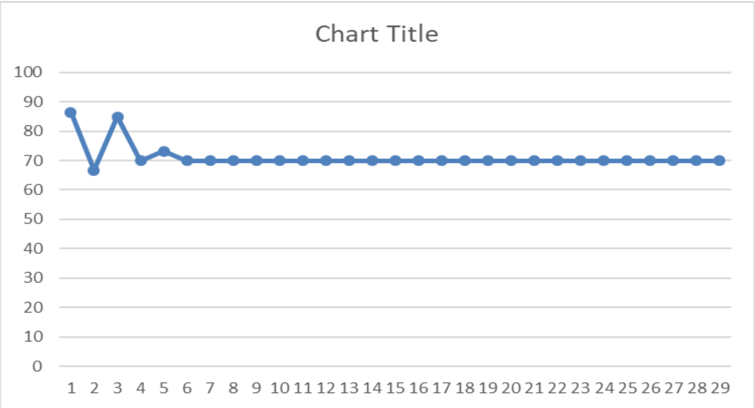

Fig 1 Graph of Classification Accuracy for the variation in the value of k = 2 up to 30 with 10-fold validation

The accuracy obtained from the WKNN classification with forward selection got a better value, namely 86.4%, compared to using the Weighted KNN without forward selection which was 60.1%. From testing, it can be obtained an increase in accuracy of 26.3%. From these results it can be concluded that the accuracy of WKNN with forward selection is higher than using WKNN alone. Graph of accuracy results for testing with variations in k values with 10-fold validation can be seen in Figure 2.

5. References

[1] Z. Zhang, F. Wu, H.J.W. Zandvliet, B. Poelsema, H. Metiu,& M.G. Lagally, "Paper Title," *Physical Review Letter*,vol. 74,pp. 3644-3677. 1995.

[2] G. Smith, "Paper Title" (to be published).

[3] L. Weiss, Instruction to Authors, Elsevier Publishing, http://www.elsevier.com/authors.html, 1999, retrieved May 13, 2010.

[4] C.H. Perry, F. Lu, F. Namavar, N.M. Kalkhoran, &R.A. Soref, “Paper Title” *In Proceeding of ICACSIS 2011*, pp. 456-541, 2010.
[5] M.J. Adams, B.J. Briscoe, &S.K. Sinha, In: D. Dowson, C.M. Taylor, T.H.C. Childs, M. Godet, G. Dalmas (Eds.), *Dissipative Processes in Tribology*, Tribology Series, vol. 27, Elsevier, Amsterdam, p.223, 1994.
[6] D. Palik (Ed.), *Handbook of Optical Constants of Solids II*, 3rd ed., Academic Press, New York,p.151, 1991.
[7] R. Ramos, “Thesis Title,” Ph.D Thesis, College van Dekanen, University of Twente, The Netherland, 1992.
[8] S. Badu, “Thesis Title,” B.S Thesis, Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Indonesia, 1990.
[9] H. Yamagishi, A. Hiroe, H. Nishio, K. Miki, K. Tsuge, & Y. Tawada, U.S. Patent No. 5264710, 23 Nov. 1993.
[10] J. Cleveland, Spring Constant Update, Digital Instruments, Santa Barbara, 1996 [if a website address available, it could be included in italic].
[11] R.D. Nicholson, International Structures In Nickelbased Transition Joints After Long Term Service, Report RD/M/N1131, Central Electricity Generating Board, Marchwood, 1980.
[12] Joint Committee on Powder Diffraction Standards, Powder Diffraction File, ASTM, Philadelphia, Card 4301027, 1967.
[13] Anon, 19-th Annual Book of ASTM Standards Part 17, ASTM, Philadelphia, p.636, 1969.
[14] R. Stumpf, X. Gonze, &M. Scheffler, Fritz-Haber Institute Research Report, 1990, unpublished.
[15] D.H. Smith, Physics Department, Chicago University, Chicago, U.S.A., private communication, 1986.