Journal of Computer Networks, Architecture and High Performance Computing
e-ISSN 2655-9102, Volume 2, No. 1, January 2020, pp 204-210
https://iocscience.org/ejournal/index.php/CNAPC

Data Mining Classification Of Filing Credit Customers Without Collateral With K-Nearest Neighbor Algorithm (Case study: PT. BPR Diori Double)

Jeprianto Sinaga[1], Bosker Sinaga[2]

[1,2]Informatics Engineering Study Program , STMIK Pelita Nusantara, Jl. Iskandar Muda No. 1 Medan, North Sumatra, Indonesia 20154

E-*mail*: *Jeffreysinaga1903@gmail.com, sinagab8@gmail.com*

Abstract- *Unsecured loans are the community's choice for lending to banks that provide Reviews These services. PT. RB Diori Ganda is a regional private banking company that serves savings and loans and loans without collateral for the community. Submission of unsecured loans must go through an assessor team to process the analysis of the attributes that Affect the customer's classification so that credit can be approved, the which is then submitted to the commissioner for credit approval. But what if Reviews those who apply for credit on the same day in large amounts, of course this will the make the process of credit analysis and approval will take a long time. If it is seen from the many needs of the community to apply for loans without collateral, a classification application is needed, in order to Facilitate the work of the assessor team in the process of analyzing the attributes that Affect customer classification. To find out the classification of customers who apply for unsecured loans for using data mining with the K-Nearest Neighbor algorithm. The result of this research is the classification of problematic or non-performing customers for credit applications without collateral.*

Keywords: Data Mining, Customer Classification, K-Nearest Neighbor Algorithm

1. Introduction

Classification is one of the tasks in data mining to predict a future state through state testing in the past with a series of processes or analysis for the calculation of distance closeness that state. Data mining is a process of extracting data or filtering of data by using a data set with a large enough size through a series of processes to obtain valuable information from these data.

Diori PT.BPR Ganda is a private banking company that serves the community savings and loans for medium and supervised by the FSA. Register and problematic customer data already collected by PT. RB Dual Diori stored in the database. Pengelolah companies must be able to predict if any existing customers apply for credit without collateral then must predict whether the customer's classification is problematic or not problematic future in the mortgage payment, the classification prediction aims to avoid the problem of bad loans in the future. The process of data analysis conducted tanpaagunan submitting credit assessor teams are still using manual processes are observed and to sort out the customer's data, so that this process takes a long time or can be called inefficient time.

K-Nearest Neighbor algorithm is a classification method that classifies test data (testing) of the training data (training) based on the test data within the data to several neighbors (neighbor) nearby. K-Nearest Neighbor algorithm often used in the completion of data mining for predictive job classification. This algorithm uses training data (training) who have been there and its output has been known.

2. Theory

2.1 Data Mining

Data mining search containing the desired trends or patterns in large databases to support decision-making at a time when that will come. These patterns are recognized by specific devices that can provide a useful analysis and insightful data that can then be studied more closely, which may use other decision support tools. In the real world, data mining is more than just apply a of the algorithms are simple. Often the data is too crowded and incomplete, if it is justified, slow lorises diiginkan great many patterns will be lost and any pattern detection capabilities will go down. Furthermore, the analysis should decide what kind of mining algorithms are used, apply it into subsets of the data sample, variable, digest the results and the process iterates (Fajar Astuti Hermawati, 2013: 2-3). Employment in data mining can be divided into four groups: the model prediction

Journal of Computer Networks, Architecture and High Performance Computing is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.

(prediction modeling), cluster analysis (cluster analysis), analysis of association (association analysis), and detection of anomalies (anomaly detection).

The prediction model (prediction modeling) This work deals with the creation of a model that can melaukan mapping of any set of variables for each target, and then use that model to give a target on him [newly acquired unan. There are 2 Jenin prediction model that is klasosfokasi and regression. Classification is used for discrete target variable, whereas regression is used for continuous target variable. (Eko Prasetyo, 2014: 5).

Data minng actually part of a series of information on the database search process. Pattern information generated from the data mining process needs to be presented in a form easily understood by the parties concerned. This stage is part of the prose Knowledge Discovery in Databases (KDD) called interpretation. This phase also includes an examination of whether a pattern or information found contradicted by the facts or hypotheses that existed before. (Muchlisin Riadi, 2017). to take more details can be seen in Figure 2.1.

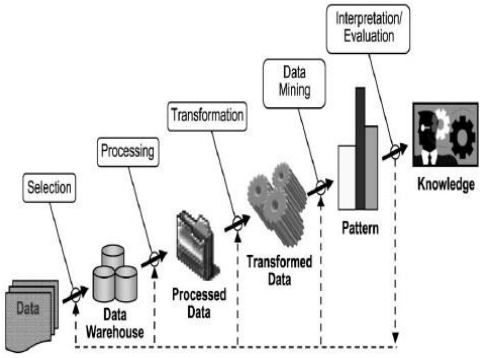


Fig 1. Stages KDD
Source : Muchlisin Riadi, 2017

2.2 Algritma K-Nearest Neighbor

*K-Nearest Neighbor*is one of the Nearest Neighbor method of classification based on the oldest and most popular. Rated K used herein express the number of nearest neighbors were involved in the determination of the class label prediction on the test data. (Eko Prasetyo, 2014: 150).

Jodi Irjaya Kartika, Edy Santoso, and Sutrisno (2017) found a method or algorithm *K-nearst Neighbor*(K-NN) is a method that uses super vised algorithm in which the results of the new queryinstance diklasifikaikan by a majority of the category in the K-nearst Neighbor. The purpose of this algorithm is to classify the new objects based on attributes and sample training. In the algorithm K-Nearest Neighbor (KNN) there are five (5) ways to search for the nearest neighbors are: [1] The Euclidean, [2] Distance Manhattan, [3] The Cosine, [4] The correlation, [5] Hamming distance ,

In this study the means used to search for the nearest neighbors only use the Euclidean distance with the following formula:

$$Similarity\ (p,q) = \frac{\sum_{i=1}^{n} f(pi,qi)qX\ wi}{wi}$$

Keteerangan:
P = The new cases
q = cases exist in storage
n = number of attributes in each case
i = attribute of individuals between 1 and n
f = function i attribute similarity between the case p and q case
w = The weight given to the attributes of the i-th

Characteristics of K-Nearest Neighbor algorithm (KNN) is a supervised learning algorithm that means this algorithm uses existing data and output are known. KNN widely used in applications of data mining, pattern recognition, image processing, etc.The steps of the algorithm calculation K-Nearest Neighbor (K-NN) in resolving a problem is as follows:

a) Menentuan value of k. N Determining the value of k used in klaifikasi not have standard rules, but in this study the value of k used are 3, 5, 7.
b) Calculating the distance between the object / new data on all objects / data yan gtelah in training.
c) Sort the results of these calculations.
d) Determine the nearest neighbors by the minimum distance to K.
e) Use categories majority as object classification / new data,

Journal of Computer Networks, Architecture and High Performance Computing is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.

2.1 Unsecured Loan

The word credit comes from the Latin is credere means trust. The owner of money or goods (creditor) gave keperayaan to the borrower (debtor) for the use of money or goods for a certain time. Such borrowing accompanied by the belief that the debtor can repay the money or goods lent. Thus, the use of a credit is giving money or goods to another person in a specific period of time with collateral or unsecured, with the provision of services of interest or no interest.

According to the Act. No.10 of 1998 on banking, noted that credit is a provision of cash or which can be likened to it, based on the approval or kesepeakatan lending - borrowing between banks and other parties who require the pemijam to repay the debt after a certain period of time with interest. From this it can be concluded uaraian take credit can be given to a person or institution on account of the trust, risk, time, and achievement. Risks include a reference to the fact that the net present value will vary with the value for money when it is returned.

3. Analysis

3.1 Data Analysis Research

The data in this study come from PT. RB Diori Simpang Dual Wire, namely customer data were ever filed and do credit (loan) on the marketing of the region 7. The data obtained and which would be predicted classification has been organizing (performed encoding the data prior to the data the original). Namely the training data and testing data.

1. Training data is customer data that has been applying for credit and classification known. Kolektibilas also decisive klasifikikasinya. If the customer where collectibility between 0 to 2, the customer dklasifikasikan not problematic, and if the customer kolektibilatas between 3 or more, the customer in classifying problematic.

Table 1.
sample Data Training

Marital status	The number of dependents	Last education	Age
Married	There is no	SLTA	21-35 years old
Married	5 people	SD	35-50 years old

Table 2
Sample of training data (continued)

Work	ownership Ruma	monthly income	credit Status
employee	lease	<= 3 jt	No problem
farmer	One's own	> 3 jt <= 6 jt	problematic

2. Data testing is the customer data that would apply for credit without collateral to be classified.

Table 3.
samples Data Testing

Marital status	The number of dependents	Last education	Age
Married	3 people	SLTA	21-35 years old

Table 4
sample Testing Data (continued)

Work	Home ownership	Income Per month	Credit Status
employee	Parents	<= 3 jt	-

The data obtained in this study were 116 credit transaction record whether or not bermaslah bemasalah, obtained from a banking company PT. RB Diori Simpang Dual Wire. All the Attributes in the category valued credit data, as shown in Table 4. Training data consists of eight attributes, in which 7-value attributes (attributes influential), and 1 attribute label. Results obtained data preprocessing is necessary to obtain high quality data,

Journal of Computer Networks, Architecture and High Performance Computing is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.

116 is reduced by eliminating duplicate records to 20 records. Then follows the tranformation to a grouping establishment of training data.

Table 5.
Attributes and Attribute Values

No.	**Attribute**	**Attribute value**
1	Status Perkainan	Single Married Widow widower
2	The number of dependents	There is no 1-2 People > = 3 People
3	Recently Pendidkan	No school SD SMP > = SLTA
4	Age	21-35 Years 35-50 Years 50-60 Years
5	Work	PNS employee entrepreneur farmer
6	Home ownership	One's own Parents Lease / Contract Service
7	monthly income	<= 3 jt > 3 jt <= 6 jt > 6 jt
8	Credit Status	problematic No problem

To measure the distance between the attributes will be given a score between 0 and 1. Where the value of 0 means that if the attribute has no effect, and 1 if the attribute effect. Can be seen in Table 4.4.

Table 6
Attribute weights

No.	**Attribute**	**Weight**
1	Marital status	0.5
2	The number of dependents	1
3	last education	0.8
4	Age	0.5
5	Work	1
6	Home ownership	1
7	monthly income	1

To find a closeness between the attribute value can be seen in Table 3.5.

Table 7.
Distance closeness of attribute values.

No.	**Attribute**	**Attribute Value 1**	**Attribute Value 2**	**Weight**
1	Marital status	Married	Married	0
		Married	Single	1
		Married	Widower	0
		Single	widow	0.5
		Single	Single	1
		Widower widow		0.5
2	The number of dependents	Nothing	There is no	0
		There is no	1-2 people	0.5
		There is no	> = 3	1
		1-2 people	1-2 people	0
		> = 3	> = 3	0
		> = 3	1-2 people	0.5
3	last education	No school	No school	0

Journal of Computer Networks, Architecture and High Performance Computing is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.

		No school	SD	0.5
		No school	SMP	0.5
		No school	> = SLTA	1
		SD	SD	0
		SD	SMP	0.5
		SD	> = SLTA	1
		SMP	SMP	0
		SMP	> = SLTA	0.5
		> = SLTA	> High School	0
4	Age	21-35 years	21-35 years	0
		21-35 years	35-50 years	0.5
		21-35 years	50-60 years	1
		35-50 years	35-50 years	0
		35-50 years	50-60 years	0.5
		50-60 years	50-60 years	0
5	Work	PNS	PNS	0
		PNS	employee	0.5
		PNS	entrepreneur	0.5
		PNS	farmer	0.5
		employee	employee	0
		employees'	entrepreneur	0.5
		employee	farmer	0.5
		entrepreneur	entrepreneur	0
		Wiraswaata	farmer	0.5
		farmer	farmer	0
6	Home	Rental / leasing	Rental / leasing	0
	ownership	Rental / leasing	One's own	1
		Rental / leasing	Parents	1
		Rental / leasing	Service	1
		One's own	Parents	0.5
		One's own	One's own	0
		One's own	Service	0.5
7	monthly	<= 3 jt	<= 3 jt	0
	income	> 3 jt <= 6 jt	> 6 jt	0.5
		> 6 jt	<= 3 jt	1

3.2 Algritma K-Nearest Neighbor

This stage is the manual calculation process to search for a pattern or interesting information in the data selected by using a particular method or algorithm based on the overall KDD process. The algorithm used in this study is the algorithm of K-Nearest Neighbor (KNN) algorithm which has attributes that are initialized as k, the number of neighbors that made reference to the KNN, indigo k is a positive integer, small and odd numbered.

a) Determining the value of k. In this study the k value used is 3.
b) Calculate the distance between the training data and test data (test) using the following formula:

$$Similarity\ (p,q) = \frac{\sum_{i=1}^{n} f(pi,qi)qX\ wi}{wi}$$

1) The proximity of new customers (testing 1) with a record of customers who have already done loans without collateral (the training data 1).

A : Proximity attribute weights marital status (married to get married) = 0
W1 : Weight attribute marital status = 0.5
B : Proximity weighting the number of dependents (none with 3 people) = 1
W2 : The weight attribute is the number of dependents = 1
C : Proximity weights latest education (SLTA with SLTA) = 0
W3 : Weight of the latest educational attributes = 0.5
D : Proximity attribute weights age (21-35 years to 21-35 years) = 0
W4 : Weight attribute age = 0.5
E : Proximity attribute weights work (employees and farmers) = 0.5
W5 : Weight of the job attributes = 0.8
F : Proximity attribute weights home ownership (leasing with one's own) = 1
W6 : Weight attribute homeownership = 0.8

Journal of Computer Networks, Architecture and High Performance Computing is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.

G : Proximity attribute monthly income (<= 3 jt with> 3 jt <= 6 jt) = 0.5
W7 : Weight attribute monthly income = 1
S1 = [(A * W1) + (B * W2) + (C * W3) + (D * W4) + (E * W5) + (F * W6) + (G * W7)] / (W1 + W2 + W3 + W4 + W5 + W6 + W7
= [(0 * 0.5) + (1 * 1) + (0 * 0.5) + (0 * 0.5) + (0.5 * 0.8) + (1 * 0.8) + (05 * 1)] / (0.5 + 1 + 0.5 + 0.5 + 0.8 + 0.8 + 1
= (0 + 1 + 0 + 0 + 0.4 + 0.8 + 0.5) / 5.1
= 2.7 / 5.1
S1 = 0.5294

2) The proximity of new customers (testing 1) with a record of customers who have already done loans without collateral (the training data 2).

A : Proximity attribute weights marital status (married to get married) = 0
W1 : Weight attribute marital status = 0.5
B : Proximity weighting the number of dependents (5 to 3) = 0
W2 : The weight attribute is the number of dependents = 1
C : Proximity weights latest education (elementary to high school) = 0.5
W3 : Weight of the latest educational attributes = 0.5
D : Proximity attribute weights age (35-50 years to 21-35 years) = 0.5
W4 : Weight attribute age = 0.5
E : Proximity job attribute weights (peasant farmers) = 0
W5 : Weight of the job attributes = 0.8
F : Proximity attribute weights home ownership (owned by one's own) = 0
W6 : Weight attribute homeownership = 0.8
G : Proximity attribute monthly income (> 3 jt <= 6 jt with> 3 jt <= 6 jt) = 0
W7 : Weight attribute monthly income = 1

S2 = [(A * W1) + (B * W2) + (C * W3) + (D * W4) + (E * W5) + (F * W6) + (G * W7)] / (W1 + W2 + W3 + W4 + W5 + W6 + W7
= [(0 * 0.5) + (0* 1) + (0.5 * 0.5) + (0.5 * 0.5) + (0 * 0.8) + (0 * 0.8) + (0 * 1)] / (0.5 + 1 + 0.5 + 0.5 + 0.8 + 0.8 + 1
= (0 + 1 + 0 + 0 + 0.4 + 0.8 + 0.5) / 5.1
= 0.5 / 5.1
S2 = 0.0980

c) Sorting distance calculation result data from the closest distance to the most distant (ascending).
Once sorted obtained: S2 =0.0980, S1 = 0. 5294

4. Determining the majority of the results of the test data based on the majority k nearest neighbors. Because nilaik = 3 then taken three nearest distance
5. By using the Nearest Neighbor category most majority then it can be classified customers who apply for credit tapa collateral.

4. Conclusion

Application of the method K-nearset Neighbor ddalam build data mining applications to facilitate team assesr in analyzing the data to predict the classification of problematic or not problematic prospective borrowers / customers who want to apply for payday loans can be known is to partition the old data with new data from the data results test has been obtained. And the results of data mining classification application design to facilitate the work of analysis in analyzing the data in considerable amounts to be analyzed simultaneously with a rapid time to find out the results.

5. Reference

[1] Eko Prasetyo, "Data Mining Processing Data Become Information Using MATLAB", Andi Gresik, Gresik, 2014
[2] Muchlisin Riadi, "Definition, Function, Process and Stages of Data Mining", Available at https://www.kajianpustaka.com/2017/09/data-mining.html
[3] Fajar Astuti Hermawaty, "Data Mining", Andi Yogyakarta, Yogyakarta, 2018.
[4] Norbertus Krisandi, Helmi, Bayu Prihandono, "Determination of Student Achievement using Methods *K-Nearest Neighbor* and Weighted Product (Study Case: SMP Negeri 3 Mejayan "Political Development of Information Technology and Computer Science Vol. 1, No. 5 (May 2017): 352-360.
[5] Smith, Bob. "Application of Data Mining for Business Purposes", Yogyakarta: Graha Science 2007.

Journal of Computer Networks, Architecture and High Performance Computing is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.

[6] Erick Kurniawan "Fast Advanced Visual Basic 2010", Yogyakarta: ANDI OFFSET, Yogyakarta in 2018 ..
[7] Retno Tri Vulandari, "Data Mining Theory and Applications RapidMiner", Gava Media, Yogyakarta, 2017.
[8] US and M. Rosa Shalahudin "Software Engineering" Informatics, Bandung, 2018.

Journal of Computer Networks, Architecture and High Performance Computing is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.