P-ISSN : 2715-2448 | E-ISSN : 2715-7199
Vol.1 No.2, 10 July 2020
Buana Information Tchnology and Computer Sciences (BIT and CS)

Implementation of K-Nearest Neighbor Algorithm for Customer Satisfaction

Sutan Faisal [1]
Study Program
Technical Information
Faculty of Engineering Computer
Science, University Buana Perjuangan
Karawang
sutan.faisal@ubpkarawang.ac.id

Nurhayati [2]
Study Program
Technical Information
Faculty of Engineering Computer
Science, University Buana Perjuangan
Karawang
nurhayati@ubpkarawang.ac.id

Abstract*—Customer satisfaction is the company's goal in providing services to its customers. Sewa Camera Cikarang is committed to customer satisfaction. By using the K Nearest Neighbor (KNN) algorithm of this study to analyze customer satisfaction of camera tenants. In this study price, facilities, services and loyalty are input attributes of customer satisfaction. Satisfied and dissatisfied is the result of the output. Increasing customer satisfaction and increasing profits on Cikarang Camera Rentals is the aim of this research. This study using the KNN algorithm obtained accuracy = 98%, recall classification = 86.67%, classification accuracy = 100% and AUC = 0.750. It is expected that the results of this study can be used as a reference for building applications that can facilitate companies in obtaining information about customer satisfaction.
Keywords—*Datamining, classification,* KNN algorithm, customer satisfaction.

Abstrak*—Customer Kepuasan pelanggan merupakan tujuan perusahaan dalam memberikan layanan kepada pelanggannya. Sewa Kamera Cikarang berkomitmen untuk kepuasan pelanggan. Dengan menggunakan algoritma K-Nearest Neighbor (KNN) penelitian ini untuk menganalisa kepuasan pelanggan penyewa kamera. Dalam penelitian ini harga, fasilitas, layanan dan loyalitas merupakan atribut masukan kepuasan pelanngan . Puas dan tidak puas merupakan hasil outputnya. Meningkatkan kepuasan pelanggan dan meningkatkan laba pada Sewa Kamera Cikarang adalah tujuan penelitiian ini. Penelitian ini dengan menggunakan algoritma KNN mendapatkan akurasi = 98%, klasifikasi recall = 86,67%, ketepatan klasifikasi = 100% dan AUC = 0,750. Diharapkan hasil penelitian ini dapat dijadikan acuan untuk membangun aplikasi yang dapat memudahkan perusahaan dalam memperoleh informasi tentang kepuasan pelanggan.

Kata kunci— Pengumpulan data, klasifikasi, algoritma KNN, kepuasan pelanggan.

I. INTRODUCTION

A. Introduction

Along with the high level of human activity to meet the needs and needs of daily life, humans need to release their fatigue with a vacation. Then it needs to be supported with a camera to capture the moment of his vacation. But not everyone has a camera that is good enough to capture the holidays.

Public awareness of the elements of service that can be provided by companies is increasing due to advances in education and a more prosperous economy, as well as the development of science and technology. The importance of service quality provided by service companies and in the form of goods is increasingly being realized by consumers. Each consumer's assessment of the quality of services / services varies depending on how consumers expect the quality of the service / service based on experience [1].

Achieving success in a service business, customer satisfaction must be the basis of management decisions, so management must make increasing customer satisfaction a fundamental goal. In order to provide quality services, the company must continually improve the quality of its human resources and the equipment it leases. This step is important to improve services from time to time.

People who judge whether or not the quality of service is called a consumer. By comparing the services they receive with the services they expect consumers can judge the service. Consumers who are satisfied with the services provided by a company will make these consumers come back again to use the company's services again. Companies that have loyal customers because the company can satisfy their customers. Word of mouth promotion without coercion regarding the services it has received will be carried out by loyal consumers [4].

Tight competition must be faced by companies in the increasingly rapid development of the business world. The customers he has by the company are expected to be maintained forever. To realize this, it is not something that is easily climatic, as business competition is very tight at the moment considering that there are rapid changes that can occur at any time such as changes in customers, competitors or changes in broad conditions that are always dynamic. This requires policy makers to develop a strategy that is able to achieve sales growth targets, increase the company's market share, and achieve capabilities as the basis for sustainable growth. [1].

The tight competition must be faced by the company in the rapid development of the business world. In general, there are many ways to maintain customers forever, in a very tight

business competition it is very difficult to realize it given the many changes that can occur at any time. Such as changes in customers. Competitors and changes in broad conditions that always change dynamically. This makes policy makers to continue to develop a strategy that can achieve the goals of rental growth, increase market share, and the achievement of capabilities as a basis for sustainable growth [16].

B. Definition of Data Minning

Data mining is data mining that has long been taken from several series of activities when viewed from the point of view, according to [5]. Data mining is an integrated data analysis process that consists of a series of actions based on the definition of the objectives to be analyzed, with data analysis and interpretation of the results.

In recent years data mining has attracted the attention of the public and the world of information systems, because useful information in the form of knowledge generated from large data is needed. Applications ranging from market analysis, fraud detection, and customer retention, to production control and exploration science are generated from information and knowledge. [7].

According to [5], Data mining has the following stages of the process:

1. Defining goals for analysis

   The clearest statement of the problem and the achieved goals are the most important in the correct formulation of the analysis. Determining the method to be used is one of the most difficult parts of the process. There must be no room for doubt or uncertainty and clear goals must.
2. Selection, organization, and preliminary treatment data

   The collection or selection of data needed to be analyzed is done after the objectives are analyzed and identified. The ideal source of data is theata's backup company, a "storage room" of historical data that is no longer used. If there is no data storage, the data market can be created by matching different corporate data sources.
3. Exploration of data analysis and transforming it

   At this stage involves an initial exploration analysis of data, which is very similar to thetechnique *Online Analytical Process* (OLAP). Transformation of the original variables to better understand the phenomena or statistical methods used are carried out at this stage. To highlight anomalous data, different data from other data is used in the analysis of exploration.
4. Specifications of statistical methods

   Statistical methods can be used, as well as many available algorithms, so it is possible to classify already available methods. The choice of method used to prepare the analysis depends on the problem being studied or the type of data available. Different methods are edited into two main classes according to different stages of data analysis, in particular:

   a. Descriptive Method

      To describe groups of data in a concise manner is the main goal of the method. There is no descriptive hypothesis between the available variables. Included in this group are the association method, *log-linear*model, graphical model).

   b. Prediction Method

      The purpose of this class method is to describe one or more variables that are performed by finding classification or prediction rules based on the data. These rules help to predict or classify one or more answers or future variables of the target variables in relation to what is happening with the explanatory or input variable. Included in this method are neural networks, decision trees, and linear and logistic regression models.
5. Data analysis based on the method chosen, which will then be applied to the statistical method to be used then translating into the appropriate algorithm to get the required results based on available data.
6. Evaluation of the methods used and Comparison for the analysis of the final model selection
7. Commentary on the selected model and its use in the *decision-making process.*

C. Clasification and Prediction

Classification and prediction is a method that can make smart decisions. Researchers have now proposed a number of classifications and forecasting methods for machine learning, pattern recognition, statistical research. In this study, we focus on classifying methods in data mining as part of the machine learning process. The form of data analysis that can be used to extract models to predict future trends in data to be predicted is the classification and prediction of data mining. The classification process is divided into two stages, first the learning process in which the classification algorithm is used to analyze training data. is, the results of the presentation of the learning model or classifier in the form of classification rules, the two phases of the classification process, estimating the accuracy of the classification model or classifier from the test data. If the accuracy is accepted, the model is applied to find out the predicted results of new data. Bayesian methods, Bayesian networks, algorithm-based rules, neural networks, vector machine support, mining rules associations, k-nearest neighbors, case-based reasoning, genetic algorithms, rough sets and fuzzy logic are the classification techniques used. Focusing the Nearest Neighbor (KNN) K algorithm in this study.

D. Data Minning Methods

The idea of people already having knowledge in the process of classifying management has already been widely used. But talking about *taxonomy* (Tassein = classify + nomos = science, law) its use as a science of grouping living organisms (*alpha taxonomy*) at first ,has since become a general science group, including the principle of classification (taxonomic schemes). Thus, classification (taxonomy) processes the placement of an object (concept) based on a number of categories, each object (concept) based on ownership. [6].

Four basic components for the classification process:

1. Class: The dependent variable of the model is the categorical variable to represent the 'label' that uses the object after its classification. Examples of lessons are: heart attack, customer loyalty, stellar lesson (galaxy), earthquake lesson (storm), etc.

2. Predictors: Classification of data and based on the classification made from the model represented by the characteristics (attributes) which are independent variables. Examples of such predictors are: smoking, drug consumption, blood sugar, sales frequency, sex status, satellite images, geological record, and wind speed direction, season, etc.
3. Training dataset The data used for the 'training' model to recognize according to class, based on predictions available from the two two component data values before.
4. Testing the dataset: contains new data classifications based on the Model built on, and classifications that are accurate (*model performance*) so they can be evaluated [6].
   a. There are no other attributions in the separate post
   b. There are no *records* inbranch an empty

E, K Neighrest Neighbor

K-Nearest Neighbor (kNN) is included in the *instance-based learning group*. This algorithm is also one of thetechniques *lazy learning*. KNN searches the k group of objects in the training data that is closest (similar) to the object in new data or testing data (similar) to the object in new data ordata *testing* [15]. Case in point, for example it is desirable to find a solution to the problem of a new patient by using a solution from an old patient. To find solutions from new patients, closeness to old patient cases is used, solutions from old cases that have closeness to new cases are used as a solution. There were new patients and 4 old patients, namely P, Q, R, and S (Figure 2). . When there is a new patient, the solution is taken from the case of the elderly patient who has the greatest kinship.

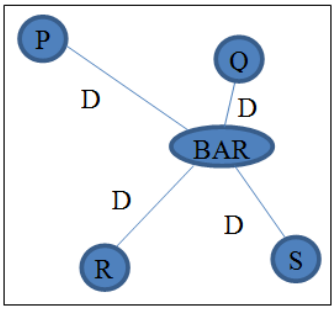


Fig. 1 Ilustrasi KNN

For example, D1 distance between new patients and patient P, D2 distance between new patients and sick Q, D3 distance between new patients and sick R, D4 distance between new patients and patient S. The picture shows that D2 is closest to the new case. Thus, the patient Q solution will be used as a solution for the new patient. (Henny Leidiyana, 2013) *Euclidean* distance and *manhattan distance* (*city block distance*) are ways to measure the proximity between new data and old data (*training data*), the most commonly used is euclidean distance. [2], namely:

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \ldots + (a_n - b_n)^2}$$

Where a = a1, a2, ..., an, and b = b1, b2, ..., bn represents the n attribute values of the two *records*.
For attributes with category values, measurements with *euclidean distance* do not match. Instead, the following functions are used [10]:

Different (a, b) {0 if ai = bi
= 1 besides

where ai and bi are the category values. If the attribute value between the two *records* being compared is the same, the distance value is 0, the meaning is similar, on the contrary, if it is different then the value of proximity is 1, it means it is not similar at all. For example the color attribute with red and red values, the value of proximity is 0, if red and blue then the value of proximity 1. Normalization is done if measuring the distance from attributes that have large values, such as income attributes. Normalization can be done with *min-max normalization* or *Z-score standardization* [10]. If thedata *training* consists of a mixture of numerical and category attributes, the use of min-max normalization is preferred [10]. To calculate the similarity of cases, a formula is used [9]:

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

Note:
P = New cases
q = Cases in storage
n = Number of attributes in each case
i = Individual attributes between 1 to n
f = Function *similarity* attribute i between cases p and case q
w = Weight given to i attribute

E. Evaluation and Validation of Data Mining Prediction Methods

In this study *Cross Validation*, *Confusion Matrix*, and ROC (curves*Receiver Operating Haracteristic*) curve methods are used for evaluation and validation.

1. Cross Validation
   To predict the error rate standard testing is done. In getting the overall error rate, the training data is randomly divided into several parts with the same comparison then the error rate is calculated section by section, then calculate the average for all error rates

2. Confusion Matrix
   Table 2.1 is the method used, one class is considered positive and the other negative, if the dataset consists of only two classes. Percentage of accuracy of data records that are classified correctly after testing the classification results is the result of evaluation with a confusion matrix that has accuracy, precison, and recall.Accuracy values [7]. The proportion of positive predicted cases that are also true positive on the actual data is called precision or confidence. The proportion of true positive cases that is correctly predicted correctly is called recall or sensitivity. [12].

Table 1 Model Conflusion Matrix

Correct Classification	Classified as	
	+	-
+	True positives	False negatives
-	False positives	True negatives

True Positive is the number of positive records that are classified positively, *false positive* is the number of negative records that are classified positively, *false negative* is the number of positive records classified as negative, *true negative* is the number of negative records classified as negative, then enter the test data. To get the amount of sensitivity (*recall*), *Specifity*, *precision*, and *accuracy* enter the value of the test data into the confusion matrix. *Sensitivity* is used to compare the number of t_pos to the number of positive records, while the comparison of the number of t_neg to the number of negative records is used *precision*. The equation below is used to calculate it 7]:

$$\text{Sensitifity} = \frac{t\_pos}{pos} \quad (3.0)$$

$$\text{Specifity} = \frac{t\_neg}{neg}$$

$$\text{Precision} = \frac{t\_pos}{t\_pos + f\_pos}$$

$$\text{Accuracy} = \text{Sensitivity}\ \frac{pos}{(pos+neg)} + \text{Specifity}\ \frac{neg}{(pos+neg)}$$

Remarks:
t_pos = Number of true positives
t_neg = Number of true negative
p = Number of record positives
n = Number of tuples negatives
f_pos = Number of false positives

3. ROC Curve

Accuracy and visually comparing classifications can be demonstrated by the ROC Curve. Confusion matrix specified by the ROC. Two-dimensional graphics with horizontal lines as false positives and vertical lines as true positive are called ROC (Vercellis, 2009). To measure the difference in performance the method used is generated from the calculation of the area under curve (AUC). The formula used by AUC $\theta^r = \frac{1}{mn}\sum_{j=1}^{n}\sum_{i=1}^{m}\psi\,(xt^r, xj^r)$

Where : $\boldsymbol{\psi}(X,Y) = \begin{cases} 1 & Y < X \\ \frac{1}{2} & Y = X \\ 0 & Y > X \end{cases}$

Description:
X = positive output
Y = negative output

II. METHOD

In this study using rapidminer studio 9.0 testing tools, using the following methodology:

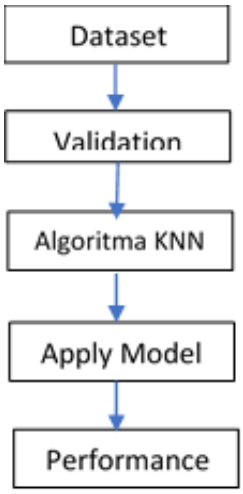


Fig. 2 Methodology Used

A. *Dataset*

Is a collection of data, a database table represented by a dataset, or it could be a data matrix where each particular variable is represented by a column, the amount of data is represented by a row.

The Retrieve operator loads the *RapidMiner* object into the process used in this research. *ExampleSet*, but can also be a Collection or a Model. Data is retrieved this way as well as meta data from the *RapidMiner Object*.

B. *Validation*

The operator used to perform simple validation randomly divides ExampleSet into a training set to set the test and evaluate the model. Split validation to estimate the performance of the learning operator (usually in an invisible data set) is performed by this operator. In practice, it will be shown how accurate a model estimate (learned by certain learning operators).

C. *KNN Algorithm*

In this study an experiment was carried out using the classification method of decamination tree datamining KNN algorithm on customer satisfaction questionnaire data on Cikarang Camera Rental. Data will be processed using the KNN algorithm and produce a model, then the resulting model will be tested Cross Validation which produces accuracy, precision, recall and AUC.

D. *Apply Model*

Learning algorithm which is the first model trained on *ExampleSet* by other Operators. After that, this model can be applied to another *ExampleSet* called *Apply Model*. To get predictions on data that are not visible or to transform data by applying the preprocessing model is the goal of applying the model.

The model attribute must be compatible with *ExampleSet* where the model is applied. *ExampleSet* Apply The model must have the same number, sequence, type, and role attributes as *ExampleSet* used to generate the model.

E. *Performance The*

The operator is used to evaluate the statistics of a binomial classification task, ie a classification task whose

label attribute has a binomial type. This operator provides a list of performance criterion values from the binomial classification task.
Measure the results of this study using a confusion matrix (accuracy, recall classification, classification accuracy) and the ROC curve.

III. RESULTS AND DISCUSSION

A. *Data Set Analysis Results The data*

Set used is the customer satisfaction questionnaire data set for Cikarang Camera Rental, this data set contains data - information about customer satisfaction questionnaires regarding prices, facilities, services and loyalty. The total data in this data set is 100 records, each of which has 10 attributes including:

1. No (integers, roles: id)
2. Name (polynominals)
3. Price X1 (integers)
4. Facility X2 (integers)
5. Services X3 (integers)
6. Loyalty X4 (integers)
7. Results (binomials: satisfied & dissatisfied)

From the attribute data set above (1 to 7) the training & test process will be carried out using the 10 Fold Cross Validation method, while the 7th attribute will be the target of the results of the classification process. and here we will try to analyze the difference between accuracy and error obtained by comparing the predicted results and results.

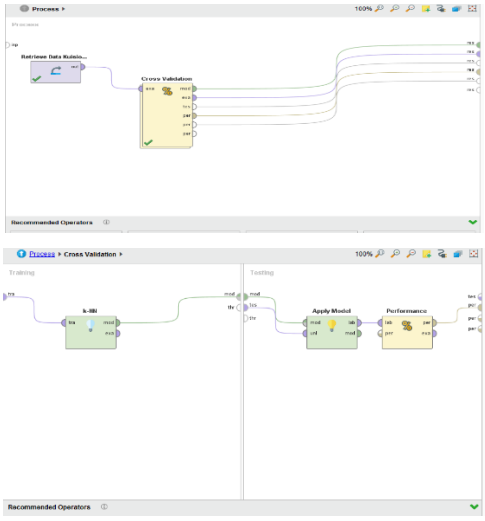


Fig. 3 10 Fold Cross Validation

The questionnaire can be illustrated below:

KUISIONER PENELITIAN
KEPUASAN PELANGGAN TERHADAP SEWA KAMERA CIKARANG

Sehubungan dengan penelitian dosen untuk melakukan tridharma perguruan tinggi dan masukan untuk sewa kamera cikarang, kami mengharapkan bantuan anda untuk memberikan masukan dengan mengisi kuisioner ini dengan keadaan sebenarnya. Bantuan anda sangat berharga untuk penelitian ini. Terima kasih atas waktu dan kebaikan anda untuk menanggapai kuisioner ini.

Nama :
Jenis Kelamin :
Umur :

Isilah kuisioner ini sesuai dengan penilaian anda, dengan tanda (√)

1 : 0 – 40(sangat Kurang 2: 41 – 60 (Kurang) 3 : 61 – 70 (Cukup)
4 : 71 – 80(Baik) 5 : 81 – 90 (Cukup Baik) 6: 91 – 100 (Sangat Baik)

Beri tanda silang (X) pada kotak yang sesuai.

No	Aspek yang dinilai	Skor					
		1	2	3	4	5	6
Harga							
1	Sewa Kamera Cikarang harganya relative terjangkau	☐	☐	☐	☐	☐	☐
Fasilitas							
2	Fasilitas penjaian pada sewa kamera cikarang lengkap dan berfungsi dengan baik (kamera, tripod, lensa dll)	☐	☐	☐	☐	☐	☐
3	Tempat duduk bersih dan rapih	☐	☐	☐	☐	☐	☐
4	Metode pembayaran yang diterapkan pada sewa kamera cikarang	☐	☐	☐	☐	☐	☐
Pelayanan							
5	Karyawan memberikan penjelasan yang baik mengenai produk sewa kamera cikarang	☐	☐	☐	☐	☐	☐
6	Persyaratan peminjaman pada sewa kamera cikarang	☐	☐	☐	☐	☐	☐
7	Jam buka operasional sewa kamera cikarang	☐	☐	☐	☐	☐	☐
8	Promo yang sering dilakukan pada sewa kamera cikarang	☐	☐	☐	☐	☐	☐
Loyalitas							
9	Secara keseluruhan anda lebih sering memilih sewa kamera cikarang dari pada tempat yang lain	☐	☐	☐	☐	☐	☐
10	Jika ada seseorang yang meminta saran ke anda, anda akan merekomendasikan sewa kamera cikarang	☐	☐	☐	☐	☐	☐
11	Anda akan mengatakan hal hal yang positif mengenai sewa kamera cikarang kepada orang lain	☐	☐	☐	☐	☐	☐
12	Jika anda sewa kamera anda akan menempatkan sewa kamera cikarang sebagai pilihan anda	☐	☐	☐	☐	☐	☐

Bagaimana kepuasan anda tentang harga, fasilitas dan , pelayanan sewa kamera cikarang ?

PUAS
TIDAK PUAS

Fig. 4. Questionnaire Form

Data that has been processed using MS Excel

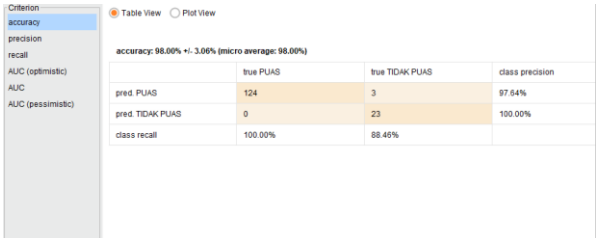
No	Nama	Harga X1	Fasilitas X2	Pelayanan X3	Loyalitas X4	Hasil
1	Anwar	6	4.33	4.00	3.50	PUAS
2	Maulana	6	4.33	4.00	3.50	PUAS
3	Budiman	5	4.33	3.75	3.50	PUAS
4	Geofany	5	4.33	3.75	3.00	PUAS
5	Fiki Ananda	5	4.33	3.75	3.00	PUAS
6	Haryanto	2	3.00	3.00	2.50	TIDAK PUAS
7	Rizky Narezka	5	4.33	4.25	3.75	PUAS
8	Aditia	5	4.00	4.50	3.75	PUAS
9	Agil	5	4.33	3.75	4.00	PUAS
10	Fadilah	3	3.00	3.75	2.75	TIDAK PUAS
11	Purwati	5	4.33	3.50	3.75	PUAS
12	Nurhajjah	5	3.67	3.50	4.25	PUAS
13	Umay	5	4.67	4.00	3.50	PUAS
14	Jesica	3	3.67	3.25	2.75	TIDAK PUAS
15	Krismonga	5	4.00	3.50	3.50	PUAS
16	Marzuki	3	3.00	2.75	2.75	TIDAK PUAS
17	Akbar	3	3.67	2.50	2.50	TIDAK PUAS

Fig. 5. Data Questionnaire that has been processed

B. *Experiment and Evaluation Results*

In this experiment there are 7 attributes which will be trained and 2 values that indicate the target (classification) on the 7th attribute, which means the KNN algorithm is initialized, 6 input attributes and 1 output attribute.
The results of this study are:

1. Confusion Matrix Table



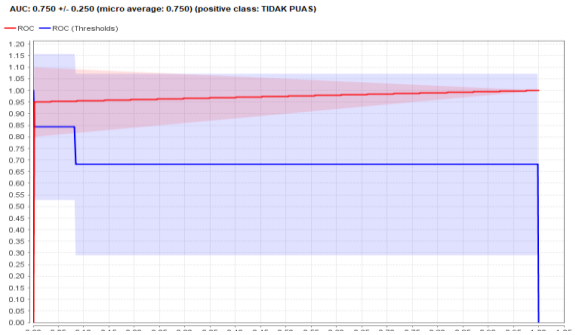
number of True PUAS (TP) is 124 records classified as True Positive 124 records and False Negative (FN) of 0 records . Next 26 records for True Dissatisfaction (TTP) are classified as True Positive 23 records and False Negative as many as 3 records.

2. Pervormance Vector

PerformanceVector

```
PerformanceVector:
accuracy: 98.00% +/- 3.06% (micro average: 98.00%)
ConfusionMatrix:
True:   PUAS    TIDAK PUAS
PUAS:   124     3
TIDAK PUAS:     0       23
precision: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: TIDAK PUAS)
ConfusionMatrix:
True:   PUAS    TIDAK PUAS
PUAS:   124     3
TIDAK PUAS:     0       23
recall: 86.67% +/- 20.82% (micro average: 88.46%) (positive class: TIDAK PUAS)
ConfusionMatrix:
True:   PUAS    TIDAK PUAS
PUAS:   124     3
TIDAK PUAS:     0       23
AUC (optimistic): 1.000 +/- 0.000 (micro average: 1.000) (positive class: TIDAK PUAS)
AUC: 0.750 +/- 0.250 (micro average: 0.750) (positive class: TIDAK PUAS)
AUC (pessimistic): 0.950 +/- 0.150 (micro average: 0.950) (positive class: TIDAK PUAS)
```

3. ROC Curve



IV. CONCLUSION

The classification method using the KNN algorithm is very good for determining the correctness of classification in data mining. Evidenced by the results of accuracy = 98%, classification recall = 86.67%, Classification precision = 100% and AUC = 0.750.

REFERENCES

[1] Abdul Rohman, *Model Algoritma K Nearest Neighbour (KNN) Untuk Prediksi kelulusan Mahasiswa*, Universitas Pandanaran Semarang, 2015.

[2] Bramer, Max. *Principles of data mining*. Vol. 180. London: Springer, 2007.

[3] Basuki, Achmad dan Syarif, Iwan. 2003. *Modul Ajar Decision Tree.* Surabaya : PENS-ITS.

[4] Deddy Setyawan, "Analisis Kepuasan Pengguna Jasa Transportasi Taksi Untuk Meningkatkan Loyalitas," Universitas Diponegoro, 2010.

[5] Giudici, Paolo, and Silvia Figini. *Front Matter*. John Wiley & Sons, Ltd, 2009.Applied Data Mining for Business and Industry.

[6] Gorunescu, Florin. *Data Mining: Concepts, models and techniques*. Vol. 12. Springer Science & Business Media, 2011.

[7] Han J, Kamber M. 2001. *Data Mining : Concepts and Techniques.* Simon Fraser University, Morgan Kaufmann Publishers.

[8] Henny Leidiyana, 2013. *Penerapan Algoritma K Nearest Neighbor untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor*. Jurnal Penelitian Ilmu Komputer, *System Embedded & Logic.*

[9] Kusrini&Luthfi,E.T. 2009. Algoritma Data Mining. Yogyakarta : Andi Publishing.

[10] Larose, D.T, 2006. *Discovering Knowledge in Data: An Introduction to Data mining*. John Willey &Sons, Inc.

[11] M Rizki Ilham, Purwanto. 2016. *Implementasi Datamining Menggunakan Algoritma C 4.5 Untuk Prediksi Kepuasan Pelanggan*. UDINUS Semarang.

[12] Powers, David Martin. *"Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation*." (2011).

[13] Rapid-I GmbH. (2008).*Rapidminer-4.2-tutorial*. Germany: Rapid-I.

[14] Resty Mardiana, *"Faktor – Faktor Yang* Memperngaruhi *Kepuasan Pengguna Jasa Taksi Blue Bird," Jakarta*, Universitas Gunadarma, 2010.

[15] Sachdeva, M., Zhu, S., Wu, F., Wu, H., Walia, V., Kumar, S., ... & Mo, Y. Y. (2009). p53 represses c-Myc through induction of the tumor suppressor miR-145. *Proceedings of the National Academy of Sciences*, *106*(9), 3207-3212.

[16] Tan S, Kumar P, Steinbach M. 2005. *Introduction To* Data *Mining*. Addison Wesley.