

**ANALISA ALGORITMA KLASIFIKASI *K-NEAREST NEIGHBOR* DALAM
MENENTUKAN NILAI AKURASI TERHADAP KEPUASAN PELANGGAN
(STUDY KASUS PT. TRIGATRA KOMUNIKATAMA)**

Maulana Fansyuri

Universitas Pamulang
dosen02359@unpam.ac.id

Abstrak

Analisa algoritma klasifikasi *K-Nearest Neighbor* dalam menentukan nilai akurasi terhadap kepuasan pelanggan (study kasus PT. Trigatra komunikatama). Perkembangan data tidak terlepas dari kemajuan teknologi informasi yang memungkinkan data dalam jumlah yang besar terakumulasi. Data mining adalah proses pengumpulan informasi penting dari sejumlah data besar yang tersimpan di basis data, gudang data atau tempat penyimpanan lainnya. *Klasifikasi* adalah proses pembelajaran secara terbimbing (*supervised learning*). *Klasifikasi* digunakan untuk memprediksi kelas dari objek yang kelasnya belum diketahui. Metode *klasifikasi* yang umum digunakan antara lain: *Decision tree*, *K-Nearest Neighbor*, *Naïve Bayes*, *Neural Network*, *C4.5* dan *Support Vector Machine*. *K-Nearest Neighbor (KNN)* adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari *query instance* yang baru diklasifikasi berdasarkan mayoritas dari kategori pada *KNN*. Tujuan dari algoritma ini adalah mengklasifikasi object baru berdasarkan atribut dan *training sample*. Data yang digunakan dalam penelitian ini berupa data primer. Dimana data tersebut diperoleh dari hasil kuisioner yang disebarkan kepada pelanggan PT. Trigatra Komunikatama. Obyek dari penelitian ini adalah hal-hal yang berkaitan dengan pelayanan kepada pelanggan. Analisa algoritma klasifikasi *K-Nearest Neighbor (KNN)* ini dilakukan untuk menentukan nilai akurasi dari data kepuasan pelanggan terhadap pelayanan PT. Trigatra Komunikatama. Berdasarkan hasil Analisa, metode *klasifikasi* algoritma *K-Nearest Neighbor* menunjukan bahwa metode ini memiliki nilai keakuratan yang cukup baik. ini dilihat dari nilai *accuracy* dimana algoritma *K-Nearest Neighbor* memiliki nilai *accuracy* sebesar 83,33%.

Kata kunci: Data Mining, Klasifikasi, Akurasi, *K-Nearest Neighbor*

**ANALYSIS OF *K-NEAREST NEIGHBOR* CLASSIFICATION ALGORITHM IN
DETERMINING ACCURACY VALUE ON CUSTOMER PLEASURE
(CASE STUDY ON PT. TRIGATRA KOMUNIKATAMA)**

Abstract

Analysis of the K-Nearest Neighbor classification algorithm in determining the value of accuracy to customer pleasure (case study of PT. Trigatra komunikatama). The development of data is inseparable from the advancement of information technology that allows large amounts of data to accumulate. Data mining is the process of gathering important information from a large amount of data stored in a database, data warehouse or other storage area. Classification is the process of guided learning (supervised learning). Classification is used to predict classes of objects whose classes are unknown. Commonly used classification methods include: Decision tree, K-Nearest Neighbor, Naïve Bayes, Neural Network, C4.5 and Support Vector Machine. K-Nearest Neighbor (KNN) is a method that uses a supervised algorithm in which the results of the new query instance are classified based on the majority of the categories in the KNN. The purpose of this algorithm is to classify new objects based on atribut and training samples. The data used in this study are primary data. Where the data is obtained from the results of the questionnaire distributed to customers of PT. Trigatra Komunikatama. The objects of this study are matters relating to service to customers. The K-Nearest Neighbor (KNN) classification algorithm analysis is performed to determine the accuracy of

customer pleasure data on PT. Trigatra Komunikatama's services. Based on the analysis, the K-Nearest Neighbor algorithm classification method shows that this method has a pretty good accuracy value. This is seen from the accuracy value where the K-Nearest Neighbor algorithm has an accuracy value of 83.33%.

Keywords: Data Mining, Classification, Accuracy, K-Nearest Neighbor

PENDAHULUAN

Data mining adalah proses pengumpulan informasi penting dari sejumlah data besar yang tersimpan di basis data, gudang data atau tempat penyimpanan (Kusrini, 2010). Perkembangan data tidak terlepas dari kemajuan teknologi informasi yang memungkinkan data dalam jumlah yang besar terakumulasi. Seiring dengan semakin dibutuhkannya data mining, muncul beberapa algoritma klasifikasi untuk memproses data dalam jumlah besar.

Klasifikasi adalah proses pembelajaran secara terbimbing (supervised learning). Klasifikasi digunakan untuk memprediksi kelas dari objek yang kelasnya belum diketahui. Metode klasifikasi yang umum digunakan antara lain: Decision tree, *K-Nearest Neighbor*, Naïve Bayes, Neural Network, C4.5 dan Support Vector Machine (Mawardy, 2017)

Prinsip dasar *K-Nearest Neighbor* (KNN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (*neighbor*) terdekatnya dalam data pelatihan. *K-Nearest Neighbor* bertujuan untuk mengklasifikasikan obyek baru berdasarkan atribut dan *training sample*. Algoritma KNN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari *query instance* yang baru.

Penelitian terdahulu yang terkait dengan topik aplikasi metode klasifikasi *Data mining* dilakukan seperti penelitian yang dilakukan seperti penelitian oleh Edi Permadi Budiyo, Nefita Nikentari, ST., M.Cs dan Sulfikar Sallu, S.Kom., M.kom (2014) dengan judul Analisa Klasifikasi Kadar Karat Emas Menggunakan Metode *K-Nearest Neighbours* (KNN). Pada hasil penelitian tersebut menunjukan akurasi menggunakan algoritma *K-Nearest Neighbours* mencapai 92,93%. Peneliti juga mengambil referensi dari penelitian yang dilakukan oleh Ricky Imanuel Ndaumanu, Kusrini dan M. Rudyanto Arief (2014) dengan judul Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode

K-Nearest Neighbor. Hasil penelitian ini menyatakan sistem yang dibuat dalam penelitian ini cukup layak untuk digunakan dalam prediksi calon mahasiswa meskipun sistem belum menghasilkan tingkat akurasi yang maksimal. Untuk menghasilkan nilai validasi yang maksimal membutuhkan data yang seimbang antara kasus keluar dan aktif.

Oleh karena itu, pada penelitian ini penulis melakukan Analisa algoritma *K-Nearest Neighbor* untuk mengetahui tingkat akurasi kepuasan pelanggan.. Maka penulis mengembangkan sebuah penelitian dengan judul : “Analisa algoritma klasifikasi *K-Nearest Neighbor* dalam menentukan nilai akurasi terhadap kepuasan pelanggan (study kasus PT. Trigatra komunikatama)”.

METODE

Data yang digunakan dalam penelitian ini berupa data primer. Dimana data tersebut diperoleh dari hasil kuisioner yang disebarkan kepada pelanggan PT. Trigatra Komunikatama. Obyek dari penelitian ini adalah hal-hal yang berkaitan dengan pelayanan kepada pelanggan.

Setelah data yang diperlukan diperoleh, kemudian ditentukan variabel-variabel yang akan digunakan dalam penelitian. Variabel respon yang digunakan dalam penelitian ini adalah kepuasan, yaitu pelanggan yang puas dan tidak puas terhadap pelayanan PT. Trigatra Komunikatama. Sedangkan variabel bebas yang digunakan antara lain Realibility, Responsiveness, Assurance, Empathy, Tangibles.

K-Nearest Neighbor

K-Nearest Neighbor (KNN) adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari *query instance* yang baru diklasifikasi berdasarkan mayoritas dari kategori pada KNN. Tujuan dari algoritma ini adalah mengklasifikasi object baru berdasarkan atribut dan training sample. Classifier tidak menggunakan model apapun

untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik *query*, akan ditemukan sejumlah k obyek atau (titik training) yang paling dekat dengan titik *query*. Klasifikasi menggunakan voting terbanyak diantara klasifikasi dari k obyek. algoritma KNN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari *query instance* yang baru. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan *Euclidean Distance* yang di representasikan sebagai berikut :

$$D(A, B) = \sqrt{\sum_{k=1}^d (A_k - B_k)^2} \quad \dots(1)$$

Confusion matrix

Pada penelitian ini, hasil Analisa algoritma diuji menggunakan metode *Confusion matrix*, yaitu suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Rumus ini melakukan perhitungan dengan 4 keluaran, yaitu: *recall*, *precision*, *accuracy* dan *error rate*. *Recall* adalah *proporsi* kasus positif yang diidentifikasi dengan benar. *Precision* adalah *proporsi* kasus dengan hasil positif yang benar. *Accuracy* adalah perbandingan kasus yang diidentifikasi benar dengan jumlah semua kasus. *Error Rate* adalah kasus yang diidentifikasi salah dengan sejumlah semua kasus.

Keakuratan hasil *klasifikasi* dapat diukur dengan menggunakan *confusion matrix*. *Confusion matrix* merupakan media yang berguna untuk menganalisis seberapa baik *Classifier* dapat mengenali tupel dari kelas yang berbeda. Misalkan terdapat dua kelas, maka akan diistilahkan sebagai *tupel positif* dan *tupel negative*. True positif mengacu pada *tupel positif* yang diberi label dengan tepat oleh *Classifier*, sementara true negative adalah *tupel positif* yang diberi label dengan tepat oleh *Classifier*. False positif adalah *tupel negative* yang diberi label tidak tepat, false negatif adalah *tupel positif* yang diberi label dengan tidak tepat. Istilah ini berguna untuk menganalisis kemampuan *Classifier* dan diringkas seperti table berikut.

Tabel 1. Model *Confusion matrix*

	C1	C2
C	<i>true positives</i>	<i>false negatives</i>
C	<i>false positives</i>	<i>true negatives</i>

Misalkan terdapat *Confusion matrix* 2x2 seperti pada table, maka rumus yang akan

digunakan untuk menghitung akurasi adalah sebagai berikut:

$$\text{Akurasi} = \frac{\text{True positif} + \text{true negatif}}{\text{True positif} + \text{true negatif} + \text{false positif} + \text{false negatif}} \times 100\% \quad \dots(2)$$

HASIL DAN PEMBAHASAN

Analisa Penelitian

Data primer pada penelitian ini adalah hasil dari kuisioner yang disebarakan kepada pelanggan PT. Trigatra Komunikatama dan data yang didapat adalah dari 1000 responden. Data ini merupakan data utama dimana data analisisnya ditunjang oleh data sekunder yang didapat dari hasil observasi lapangan dan kajian pustaka untuk memperkuat analisis.

ID	RESPONDEN	ALAMAT	UMUR	STATUS PERKAWINAN	RELIABILITY	RESPONSIVENESS	ASSURANCE	EMPATHY	TANGIBLES	VELOCITY
1	PT 1

Gambar 1. Data Hasil Kuisioner

Teknik Pengolahan Data

Setelah data yang diperlukan diperoleh, kemudian ditentukan variable-variabel yang akan digunakan dalam penelitian. Variabel independen yang digunakan antara lain *Realibility*, *Responsiveness*, *Assurance*, *Empathy* dan *Tangibles*, masing-masing variabel memiliki indikato-indikator pendukung dari variabel tersebut. Sedangkan variabel responnya adalah kepuasan, yaitu pelanggan yang puas dan tidak puas terhadap pelayanan. Tahap awal analisa penelitian adalah *preprocessing*. Dari data hasil kuisioner, tiap variabel memiliki beberapa sub variabel, maka sub variabel tersebut dijumlahkan untuk menjadi satu variabel. Setelah dilakukan *preprocessing* data yang didapat sebanyak 1000 *record* kemudian direduksi dengan meminimalkan duplikasi menjadi 150 *record* untuk data *training*. Setelah data di *preprocessing*, kemudian data masing-masing variabel dihitung nilai rata-ratanya untuk menyederhanakan hasil dari masing-masing variable. Kemudian melakukan pengkodegan dari data *training*. *Coding*

adalah proses pemberian kode terhadap data dari kuisioner untuk kelompok ke dalam kategori yang sama dan diberikan symbol (Miftahul, 2013). Setiap variable akan dikategorikan ke dalam 2 kelas yaitu nilai antara 0-3,5 (Rendah), 3,6-5 (Tinggi).

Algoritma KNN

Algoritma K-Nearest Neighbor merupakan bagian dari algoritma *supervised learning*, yakni hasil dari *query instance* yang baru, diklasifikasikan berdasarkan data terbanyak dari kategori yang ada. Kelas yang paling banyak muncul yang akan menjadi kelas hasil klasifikasi. Data yang diambil adalah data yang telah dihitung rata-rata dari data awal. Nilai masing-masing variabel dihitung rata-rata, maka hasilnya akan seperti table berikut:

Table 2 Hasil Perhitungan rata-rata Variabel Data Testing

Reabilit y	Responsivene ss	Assuranc e	Empat y	Tangia b el	Respo n
5,00	4,33	3,67	4,67	4,33	PUAS

Perhitungan dengan melakukan klasifikasi data *testing* apakah hasil perhitungan menggunakan algoritma KNN sesuai dengan data respon dari data testing. Selanjutnya adalah menentukan nilai dari K. Untuk nilai K ditentukan dengan nilai 5.

Selanjutnya adalah menghitung kuadrat jarak *Euclid* (*Euclidean Distance*) dari masing-masing *record* terhadap data *testing* yang diberikan. Untuk menghitung kuadrat jarak *Euclid* menggunakan persamaan:

$$D(X1, X2) = \sqrt{\sum_{i=1}^p (X1i - X2i)^2} \quad \dots(3)$$

Contoh perhitungan untuk data ke 1 terhadap data testing

$$\begin{aligned} D(X1, X2) &= \sqrt{(5 - 4,67)^2 + (4,33 - 5)^2 + (3,67 - 3,67)^2 + (4,67 - 4,67)^2 + (4,33 - 4,33)^2} \\ &= \sqrt{(0,33)^2 + (-0,67)^2 + (0,00)^2 + (0,00)^2 + (0,00)^2} \\ &= \sqrt{0,11 + 0,44 + 0,00 + 0,00 + 0,00} \\ &= \sqrt{0,56} \\ &= 0,75 \end{aligned}$$

Untuk data selanjutnya sampai data ke-150 dihitung dengan cara yang sama seperti perhitungan di atas. Selanjutnya adalah mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak (*Distance*) *Euclid* terkecil. Untuk mengurutkan hanya perlu membuat urutan yang mempunyai data *Euclid* terkecil sampai terbesar. Kemudian mengumpulkan kategori Y (klasifikasi *Nearest neighbor*), dengan cara mengambil data sesuai dengan jumlah K yang telah ditentukan

sebelumnya. Nilai K yang telah ditentukan adalah 5, maka dari data tersebut pilih 5 data yang mempunyai jarak (*Distance*) *Euclid* terkecil.

Table 3. Pengurutan 5 data dengan jarak terkecil

No	Reability	Responsiveness	Assurance	Empaty	Tangia b el	Respon	Distance
2	4,67	4,67	3,67	4,67	4,33	PUAS	0,47
46	5,00	4,33	3,67	4,33	4,00	PUAS	0,47
68	4,67	4,33	4,00	4,67	4,00	PUAS	0,58
110	5,00	4,00	3,33	4,33	4,33	TIDAK PUAS	0,58
3	5,00	4,33	3,67	4,00	4,33	PUAS	0,67

Hasil dari table tersebut, dengan menggunakan kategori mayoritas, maka didapat hasil klasifikasi. Dari kelima data tersebut terdapat 4 respon Puas yang berada di baris 2, 46, 68, 3 dan 1 Respon Tidak Puas yang berada di baris ke 110, sehingga mayoritas hasil klasifikasi tersebut adalah dengan respon Puas. Maka hasil klasifikasi sesuai dengan data *testing* dengan nilai respon yaitu Puas.

Confusion matrix

Dengan metode *K-Nearest Neighbor*, berdasarkan data *training*, pada Tabel 4.26 diketahui dari 150 data, 75 diklasifikasikan Puas sesuai prediksi, lalu 21 data diprediksi puas tetapi ternyata tidak puas, 4 data tidak puas sesuai predksi, dan 50 data diprediksi tidak puas ternyata puas.

Table 5. Model *Confusion matrix* untuk *K-Nearest Neighbor*

Accuracy: 83,33 % +/- 9,07 (mikro 83,33%)			
	True PUAS	True TIDAK PUAS	Class precision
Prediksi PUAS	75	21	78,12%
Prediksi TIDAK PUAS	4	50	92,59%
Class Recall	94,94%	70,42%	

KESIMPULAN

Berdasarkan hasil penelitian dari analisa metode klasifikasi *K-Nearest Neighbor* dalam identifikasi kepuasan pelanggan, dapat diambil kesimpulan bahwa analisa metode klasifikasi *K-Nearest Neighbor* memiliki nilai akurasi sebesar 83,33% dan berdasarkan hasil pengolahan algoritma *K-Nearest Neighbor* dari

data pelayanan yang diberikan PT. Trigatra Komunikatama kepada pelanggannya maka hasilnya adalah pelanggannya merasa puas atas pelayanan yang diberikan.

Ruziqa, E. (2014). Analisis Kepuasan Pelanggan Menggunakan Algoritma C4.5.

DAFTAR PUSTAKA

- David Hartanto, S. H. (2014). Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa.
- Dhika, H. F. (2016). Implementasi Algoritma C4.5 Terhadap Kepuasan Pelanggan.
- Edi Permadi, N. N. (2014). Analisa Klasifikasi Kadar Karat Emas Menggunakan Metode K-Nearest Neighbours (KNN).
- Hardiyati. (2010). Analisis Pengaruh Kualitas Pelayanan Terhadap Kepuasan Konsumen Menggunakan Jasa Penginapan (Villa) Agrowisata Kebun di Pagilaran.
- Kristina, Y. (2017, 08 22). Retrieved from [www.linkedin.com: https://id.linkedin.com / pulse / pengaruh – kepuasan – pelanggan – terhadap – profit - consulting-group](https://id.linkedin.com/pulse/pengaruh-kepuasan-pelanggan-terhadap-profit consulting-group)
- Kusrini, L. (2010). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- Mawardy, K. (2017, 03 12). Retrieved from www.i-kuliah.blogspot.co.id.
- Miftahul, N. (2013). Penerapan Data Mining Untuk Menganalisis Kepuasan Konsumen menggunakan Metode C4.5.
- Paulus Wicaksana, D. (2015). Perbandingan Algoritma K-Nearest Neighbors Dan Naive Bayes Untuk Study Data Winconsin Diagnosis Breast Cancer.
- Ricky Imanuel, K. M. (2014). Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode K-Nearest Neighbor.