SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika

https://ejurnal.itats.ac.id/snestik dan https://snestik.itats.ac.id

Informasi Pelaksanaan :
SNESTIK II - Surabaya, 26 Maret 2022
Ruang Seminar Gedung A, Kampus Institut Teknologi Adhi Tama Surabaya

Informasi Artikel:
DOI : 10.31284/p.snestik.2022.2717

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043
Email : snestik@itats.ac.id

Studi Komparasi Metode Klasifikasi K-Nearest Neghbor dan Naïve Bayes dalam Mengidentifikasi Kepuasan Pelanggan Terhadap Produk

Andhika Karulyana Febrian[1], Yulison Herry Chrisnanto[2], dan Pupita Nurul Sabrina[3]

Program Studi Teknik Informatika, Universitas Jenderal Achmad yani[1,2,3]
e-mail: andhikakfebrian@if.unjani.ac.id

ABSTRACT

Someone will feel happy and satisfied when given good service. Satisfaction is a function of the difference between perceived and expected performance. In performing the classification of satisfaction required criteria that support. This study has calculated the classification of customer satisfaction using the KNN and Naïve Bayes methods, these two methods were compared to find out which method is better. Using data from the Kaggle site, 800 data sets were divided into 601 training data sets, and 199 test data sets. With the criteria taken, namely the product name, 1 star, 2 star, 3 star, 4 star, 5 star, and product quality. The testing method in this research is using the Confusion Matrix. This method can measure the performance of a classification problem. The K-Nearest Neighbor method produces an accuracy of 87.84%, this method is supervised which means it requires training data to classify objects that are the closest distance, while Naïve Bayes produces an accuracy of 68.92% smaller than KNN. This shows that the use of the KNN method is better than Naïve Bayes in terms of accuracy.

Keywords: *Customer Satisfaction; Klasifikasi; K-Nearest Neighbor; Naïve Bayes; Product.*

ABSTRAK

Seseorang akan merasa senang dan puas ketika diberikan pelayanan yang baik. Kepuasan merupakan fungsi perbedaan kinerja yang dirasakan dengan yang diharapkan. Dalam melakukan klasifikasi kepuasan diperlukan kriteria yang menunjang. Penelitian ini telah melakukan perhitungan klasifikasi kepuasan pelanggan menggunakan metode *KNN* dan *Naïve Bayes*, kedua metode ini dibandingkan untuk mengetahui metode mana yang lebih baik. Menggunakan data dari situs *Kaggle* sebanyak set 800 data yang dibagi dua menjadi 601 set data pelatihan, dan 199 set data uji. Dengan kriteria yang diambil yaitu nama produk, bintang1, bintang2, bintang3, bintang4,

bintang5, dan kualitas produk. Pengujian metode dalam penelitian ini yaitu menggunakan *Confuision Matrix* metode ini dapat melakukan pengukuran kinerja dari suatu masalah klasifikasi. Metode *KNN* menghasilkan akurasi sebesar 87.84%, metode ini merupakan supervised yang artinya membutuhkan data training untuk mengklasifikasi objek yang jaraknya paling dekat, sedangkan *Naïve Bayes* menghasilkan akurasi 68.92% lebih kecil dibanding dengan *KNN*. Hal ini menunjukan bahwa penggunaan metode *KNN* lebih baik dibandingkan *Naïve Bayes* dari segi akurasi.

Kata kunci: Kepuasan Pelanggan; Klasifikasi; K-Nearest Neighbor; Naïve Bayes; Produk.

PENDAHULUAN

Seseorang akan merasa senang dan puas ketika diberikan pelayanan yang baik. Pada dasarnya, konsep kepuasan merupakan fungsi dari perbedaan antara kinerja yang dirasakan dan kinerja yang diharapkan. Kepuasan pelanggan dapat merangkul perbedaan antara pentingnya tingkat layanan dan hasil yang dirasakan [1]. Persepsi individu tentang hasil produk atau layanan yang terkait dengan harapan pelanggan [2]. Produk adalah apa yang dapat diiklankan dan ditawarkan pasar untuk dibeli dan dijual guna memenuhi kebutuhan dan keinginan pasar. Kualitas memiliki banyak definisi. Karena setiap orang memiliki cara pandang yang berbeda [3].

Istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi atau disebut juga dengan Data Mining, merupakan proses semi otomatis yang menggunakan teknik *statistic*, matematika, *artificial intelegent*, dan *machine learning* yang digunakan untuk mengektraksi dan mengindentifikasi informasi yang bermanfaat, dan juga mencari pola tersembunyi yang sebelumnya tidak diketahui dalam kumpulan data [4]. Teknik pengenalan pola seperti teknik *statistic*, matematika [5], dan juga dapat disebut data yang dialirkan secara dinamis [6], bagian dalam proses [7].

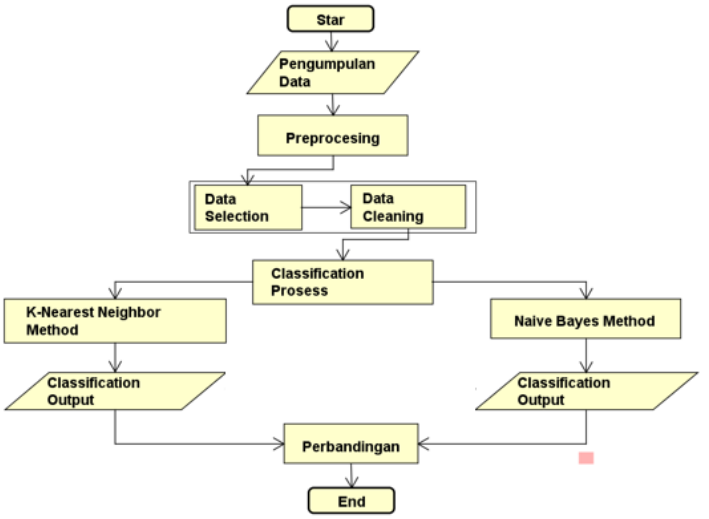
Dalam menentukan kepuasaan pelanggan dibutuhkan kriteria yang dapat menunjang, tidak menggunakan satu kriteria saja. Seperti pada penelitian sebelumnya dalam menentukan kepuasan pelanggan menggunakan kriteria sebanyak empat [8], sistem pendukung keputusan untuk menentukan kepuasan *customer* terdadap pelayanan menggunakan lima kriteria [9]. Untuk melakukan pengklasifikasian atau kepuasaan diperlukan suatu metode yaitu *K-Nearest Neighbor* yang memiliki performa cukup baik. Metode ini merupakan supervised yang artinya membutuhkan data *training* yang digunakan untuk klasifikasi objek dengan jarak terdekat. Penelitian sebelumnya telah melakukan analisis sentiment menggunakan metode SVM dan KNN menghasilkan akurasi sebesar 81.92% [10]. 77.78% untuk model KNN dalam kepuasan pelayanan wali murid sekolah dasar [11], penelitian sebelumnya juga mengatakan penggunaan metode KNN dalam kepuasan pengguna aplikasi *Marketplace* menghasilkan akurasi sebesar 96% [12].

Selain metode KNN, *Naïve Bayes* juga merupakan sebuah metode yang digunakan untuk memprediksi peluang dimasa depan. Keuntungan menggunakan metode ini yaitu menggunakan data latih yang kecil untuk menentukan parameter *mean* dan *varians* dari variabel yang di perlukan [13]. Penelitian sebelumnya telah melakukan klasifikasi kepuasan mahasiswa terhadap pelayanan perguruan tinggi menggunakan metode *Naïve Bayes* menghasilkan akurasi sebesar 96,24% dengan data sebanyak 213 data mahasiswa [14]. Metode *Naïve Bayes* dalam memprediksi kepuasan mahasiswa menggunakan atribut sebanyak empat [15], penggunaan metode *Naïve Bayes* untuk kepuasan pengguna terhadap *Online System* Universitas Advent menghasilkan akurasi sebesar 81.3% [16].

Berdasarkan latar belakang tersebut penelitian ini telah melakukan perbandingan metode *KNN* dan *Naïve Bayes* juga membandingkan keoptimalan dari kedua metode, dengan mengangkat kepuasan pelanggan menggunakan data transaksi untuk menghitung keakuratan data kepuasan pelanggan menggunakan data sebanyak 800 yang mana 601 data ini digunakan menjadi data latih dan 199 menjadi data uji dengan hasil keluran 2 kelas data.

METODE

Metode penelitian perbandingan dua metode yaitu *Naïve Bayes* dan KNN untuk klasifikasi kepuasan pelanggan yang terdiri dari pengumpulan data, pra proses, klasifikasi menggunakan *KNN* dan *Naïve Bayes* ditunjukan pada Gambar 1.



Gambar 1. Metode Penelitian

Pengumpulan Data

Pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan guna mencapai tujuan penelitian. Penelitian ini menggunakan data transaksi produk teh, data diperoleh dari website Kaggle.

A. Pra Proses

Pra proses akan dilakukan sebanyak dua tahapan yaitu *collection* dan *selection* yang digunakan untuk memproses data yang belum siap menjadi data yang siap digunakan pada proses data mining.

1. Collection

Data yang digunakan untuk penelitan ini data transaksi sebanyak 800 data. Data tersebut terdiri dari proses rating produk berupa dokumen excel.

2. Selection

Tahapan ini merupakan tahapan pemilihan data dari banyaknya kumpulan data. Data yang didapatkan dari situs kaggle terdapat 15 parameter seperti *product_id*, *product_name*, *product_company*, *product_flavour*, *category*, *product_size*, *product_price*, *product_rate*, *one_star*, *two_star*, *three_star*, *four_star*, *five_star*, *rivews_number*, dan *product_quality*. Setelah tahapan ini data di selection menjadi enam parameter yaitu nama produk, bintang1, bintang2, bintang3, bintang4, bintang5, dan kualitas_produk.

B. Klasifikasi Menggunakan *K-Nearest Neighbor*

KNN merupakan metode *supervised* dengan kata lain membutuhkan data training untuk mengklasifikasi objek dengan jaraknya yang paling dekat. Proses perhitungan dengan metode KNN data yang telah melalui pra proses akan digunakan menjadi data latih untuk data yang akan diujikan, dengan mengukur kedekatan data yang ada. Perhitungan jarak ini menggunakan rumus *Euclidean Distance*. Ditujnukan pada Persamaan (1).

$$d(x,y) = \sqrt{\sum_{i=1}^{m}(Xi - Yi)^2} \quad (1)$$

Dalam melakukan perhitungan metode KNN terdapat Langkah-langkah yaitu menentukan nilai *K* terdekat, setelah itu menghitung jarak antara data latih dan data uji, kemudian menghitung jarak antara *istance* maka didapatkan hasil perhitungan jarak *Euclidean Distance.*

C. Klasifikasi Menggunakan Naïve Bayes

Pada proses ini perhitungan dengan metode *Naïve Bayes* untuk menghitung probabilitas prior yang digunakan untuk menentukan kelas terhadap kasus yang baru dengan cara menghitung probabilitas posteriornya. Rumus probabilitas prior ditunjukan pada Persamaan (2).

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (2)$$

X	=	Data class yang belum diketahui
H	=	Hipotesis data X merupakan suatu class spesifik
P(H|X)	=	Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)
P(H)	=	Probabilitas hipotesis H (prior prob.)
P(X|H)	=	Probabilitas X berdasarkan kondisi tersebut
P(X)	=	Probabilitas dari X

Langkah pertama menghitung probabilitas prior setelah didapatkan Langkah selanjutnya adalah hipotesis dari class kemudian melakukan perhitungan terhadap probabilitas pada kondisi tertentu dapat disebut juga dengan probabilitas prior. Perhitungan dengan menjumlahkan kasus secara terperinci tiap atribut data, hal ini digunakan untuk mengetahui probabilitas prior.

Langkah kedua yaitu melakukan perhitungan probabilitas posterior digunakan untuk menentukan kelas terhadap kasus baru dalam mengelola data. Setelah mengetahui setiap atribut pada probabilitas yang dirumuskan dengan P(X|H). Langkah selanjutnya yaitu melakukan perhitungan terhadap total keseluruhan probabilitas setiap kelas.

D. Pengujian Metode Menggunakan Confusion Matrix

Metode uji confusion matrix adalah pengukuran kinerja dari suatu masalah klasifikasi yang keluarannya dapat berupa dua kelas atau lebih. Ada empat kombinasi yang berbeda dari nilai prediksi dan nilai aktual. Ada empat istilah yang menggambarkan hasil dari proses klasifikasi confusion matrix: true positives, true negatives, false positives, dan false negatives. Empat metode evaluasi ini digunakan untuk mengukur kinerja.

1. *Accuracy*, persentase jumlah record data yang terklasifikasi dengan benar oleh sistem menggunakan Persamaan (3).
   $$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (3)$$
2. *Precision*, presentase akurasi antara data yang diminta dengan hasil klasifikasi menggunakan Persamaan (4)
   $$Precision = \frac{TP}{TP+FP} \quad (4)$$
3. *Recall*, keberhasilan model klasifikasi *K-Nearest Neighbor* dalam menentukan kembali informasi atau hasil nilai dengan menggunakan Persamaan (5).
   $$Recall = \frac{TP}{TP+FN} \quad (5)$$
4. *F1-Score*, suatu perbandingan nilai rata-rata dari *precision* dan *recall* dengan menggunakan Persamaan (6).
   $$F1 - Score = \frac{2\ x\ Recall\ x\ Precision}{Recall+Precision} \quad (6)$$

HASIL & PEMBAHASAN

Hasil dan pembahasan ini adalah pengujian masing-masing metode menggunakan *Confusion Matrix*. *Confusion Matrix* ini digunakan untuk mengukur performa masalah klasfikasi dimana keluaran penelitian ini dua kelas. Pengujian akurasi ini dilakukan menggunakan 800 data record, 264 merupakan data yang merepresentasikan kelas baik, 536 merupakan data yang merepresentasikan kelas cukup.

Pembahasan I

Pembahasan pertama yaitu menguji performa metode *KNN* dengan akurasi sebesar 87,84%, ditunjukan pada Tabel 1.

Tabel 1. Pengujian Confuision Matrix Metode *KNN*

kelas	*Accuracy*	*Precision*	*Recall*	*F1-Score*
Baik	87,84%	78%	97%	84%
Cukup	87,84%	85%	87%	84%

Pembahasan II

Pembahasan kedua yaitu menguji performa metode Naive Bayes, pengujian ini menghasilkan akurasi sebesar 68,92% metode ini menghasilkan akurasi lebih rendah dibandingkan dengan metode KNN. Ditunjukan pada Tabel 2.

Tabel 2. Pengujian Confuision Matrix Metode *Naïve Bayes*

kelas	*Accuracy*	*Precision*	*Recall*	*F1-Score*
Baik	68,92%	52%	68%	0,58%
Cukup	68,92%	77%	63%	69%

KESIMPULAN

Penelitian ini telah melakukan klasifikasi kepuasaan pelanggan menggunakan dua metode yaitu *KNN* dan Naïve Bayes dan mengukur performa dari masing-masing metode. Data didapatkan dari situs *Kaggle* sebanyak 800 data, dimana data ini dibagi menjadi dua bagian yaitu data latih sebanyak 601 data, dan data uji sebanyak 199. Atribut yang digunakan dalam penelitian ini adalah nama produk, bintang1, bintang2, bintang3, bintang4, bintang5, dan kualitas produk. Pengujian metode menggunakan *Confusion Matrix*. Penggunaan metode *K-Nearest Neighbor* menghasilkan akurasi sebesar 87.84%, selain itu metode *Naïve Bayes* menghasilkan akurasi sebesar 68.92% hal ini menunjukan bahwa penggunaan metode *KNN* lebih baik dibanding dengan *Naïve Bayes*.

DAFTAR PUSTAKA

[1] R. K. Umam, N. P. Hariastuti, I. Teknologi, and A. Tama, "ANALISA KEPUASAN PELANGGAN DENGAN MENGGUNAKAN METODE CUSTOMER SATISFACTION INDEX (CSI) DAN IMPORTANCE PERFORMANCE ANALYSIS (IPA)," pp. 339–344.

[2] R. A. Haryanto, "Strategi Promosi, Kualitas Produk, Kualitas Layanan Terhadap Kepuasan Pelanggan Pada Restoran Mcdonaldâ€™S Manado," *J. Ris. Ekon. Manajemen, Bisnis dan Akunt.*, vol. 1, no. 4, pp. 1465–1473, 2013.

[3] E. F. S. Simanjorang, "Analisa Faktor-Faktor yang Mempengaruhi Keputusan Pembelian Kopi pada Warkop On Mada Rantauprapat," *Ekon. Bisnis Manaj. dan Akunt.*, vol. 1, no. 1, pp. 91–101, 2020.

[4] V. Novalia and R. Goejantoro, "Perbandingan Metode Klasifikasi Naive Bayes dan K-Nearest Neighbor The Comparison Method Of Classification Naive Bayes and K-Nearest Neighbor (Case Study : Employment Status Of Citizen In Kutai Kartanegara Regency 2018

),” vol. 11, pp. 159–166, 2020.

[5] M. Reza Noviansyah, T. Rismawan, and D. Marisa Midyanti, “Penerapan Data Mining Menggunakan Metode K-Nearest Neighbor Untuk Klasifikasi Indeks Cuaca Kebakaran Berdasarkan Data Aws (Automatic Weather Station) (Studi Kasus: Kabupaten Kubu Raya),” *J. Coding, Sist. Komput. Untan*, vol. 06, no. 2, pp. 48–56, 2018

[6] N. A. Rahmawati and A. C. Bachtiar, “Analisis dan perancangan sistem informasi perpustakaan sekolah berdasarkan kebutuhan sistem,” *Berk. Ilmu Perpust. dan Inf.*, vol. 14, no. 1, p. 76, 2018.

[7] A. Data *et al.*, “oleh 72150001,” 2019.

[8] G. Gusrianty, D. Oktarina, and W. J. Kurniawan, “Sistem Pendukung Keputusan Dengan Metode Promethee Untuk Menentukan Kepuasan Pelanggan Penjualan Sepeda Motor Bekas,” *Sistemasi*, vol. 8, no. 1, p. 62, 2019.

[9] M. Dahria, S. N. Arief, I. Santoso, and R. Kustini, “Sistem Pendukung Keputusan Menentukan Tingkat Kepuasan Customer Terhadap Pelayanan Jasa Kebersihan Di Pt. SASMenggunakanMetode Fuzzy Asosiative Memory,” *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 3, no. 1, p. 1, 2020.

[10] B. S. Amalia, Y. Umaidah, and R. Mayasari, “Analisis Sentimen Review Pelanggan Restoran Menggunakan Algoritma Support Vector Machine Dan K-Nearest Neighbor,” *SITEKIN J. Sains, Teknol. dan Ind.*, vol. 19, no. 1, pp. 28–34, 2021.

[11] P. Astuti, “Penerapan Aplikasi Sistem Komparasi Metode K-Nearest,” vol. 10, no. 4, pp. 333–345, 2017.

[12] S. Hanifah, I. Indriati, and M. Marji, “Analisis Sentimen Kepuasan Pengguna Pada Ulasan Aplikasi Marketplace Menggunakan Metode BM25F dan Neighbor-Weighted K-Nearest Neighbor,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 10, pp. 10309–10315, 2019

[13] I. dan A. Mutiara, “Penerapan K-Optimal Pada Algoritma Knn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan Ip Sampai Dengan Semester 4,” *Klik - Kumpul. J. Ilmu Komput.*, vol. 2, no. 2, pp. 159–173, 2015.

[14] M. Siddik, Y. Desnelita, and Gustientiedina, “Penerapan Naïve Bayes untuk Memprediksi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademis,” *J. Infomedia*, vol. 2, no. 4, pp. 89–93, 2019.

[15] D. R. Sari, D. Hartama, I. S. Damanik, and A. Wanto, “Penerapan Metode Naive Bayes dalam Memprediksi Kepuasan Mahasiswa Terhadap Cara Pengajaran Dosen,” *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 287, 2019

[16] Y. T. Samuel and K. Dewi, “Penggunaan Metode NAÏVE BAYES Dalam Mengukur Tingkat Kepuasan Pengguna Terhadap Online System Universitas Advent Indonesia The Use of Naïve Bayes Method in Measuring User’s Satisfaction With Adventist University of Indonesia’s Online System,” pp. 147–153, 2008

[17] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, “Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa indonesia,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 427, 2018.

[18] P. Pranowo, “C7_13_semnasteknomedia _2015,” 2018.