

Penggunaan Metode *NAÏVE BAYES* Dalam Mengukur Tingkat Kepuasan Pengguna Terhadap *Online System* Universitas Advent Indonesia

Yusran Timur Samuel*¹, Kemala Dewi²

^{1,2}Fakultas Teknologi Informasi, Universitas Advent Indonesia
e-mail: *¹ y.tarihoran@unai.edu

Abstrak

Dengan tercapainya kepuasan pengguna dalam menggunakan suatu sistem informasi, merupakan salah satu faktor untuk menentukan sistem tersebut tercapai sesuai harapan. Sistem *online* mulai beroperasi pada tahun 2008 di UNAI guna memudahkan mahasiswa mengakses data akademik. Oleh karena itu penentuan kepuasan pengguna dalam menggunakan sistem *online* di UNAI perlu dibuat sehingga dapat menjadikan suatu ukuran keberhasilan dalam suatu sistem informasi. Metode yang digunakan dalam penelitian ini yaitu menggunakan *data mining* dengan metode klasifikasi *Naïve Bayes* dan menggunakan data yang didapat dari penyebaran kuesioner. Atribut yang digunakan yaitu kualitas konten, relevan, privasi, mudah dioperasikan, kecepatan, daya tarik visual, kelengkapan secara *online*, dan *customer service*. Hasil pengujian menunjukkan bahwa klasifikasi kepuasan pengguna menggunakan metode *naïve bayes* untuk pengujian pertama dengan *full data training* mendapatkan nilai akurasi sebesar 81.3%, pengujian kedua dengan 80% *data training* 20% *data testing* mendapatkan nilai akurasi 80%, pengujian ketiga dengan *cross-validation* mendapatkan nilai akurasi sebesar 78.7 dan pengujian keempat dengan 66% *data training* dan 33% *data testing* mendapatkan nilai akurasi 68.6%.

Kata Kunci: *Data Mining, Naïve Bayes, Klasifikasi, Mengukur Tingkat Kepuasan Pengguna.*

The Use of Naïve Bayes Method in Measuring User's Satisfaction With Adventist University of Indonesia's Online System

Abstract

Achieving user satisfaction in using an information system is one of the factors to determine the system as expected. The UNAI Online System has been operating since 2008 and granting students a convenient access to their academic data. In order to enhance users' satisfaction on using the UNAI Online System, it needs to be done according to the right measurements towards an information system. The methods that were used in this study is data mining along with the Naïve Bayes classifications method and data that were obtained from questionnaires. The attributes used are content quality, relevance, privacy, easily operated, speed, visual appeal, online completeness, and customer services. The test results showed that the first test for users' satisfaction using the Naïve Bayes method scored up to 81.3%. The second test with 80% data training and 20% data testing obtained 80% accuracy value. As for cross-validation test, the score reached 78.7%. Lastly, the test with 66% training data and 33% test data get the accuracy value up to 68.6%.

Keywords: *Data Mining, Naïve Bayes, Classification, Measuring User Satisfaction Levels.*

1. Pendahuluan

Suatu sistem informasi merupakan salah satu perkembangan teknologi masa kini dan juga salah satu cakupan ilmu teknologi informasi yang sangat membantu dalam mempermudah pekerjaan di segala bidang. Kemajuan teknologi informasi dan komunikasi sebagai pendukung proses bisnis juga bisa berdampak pada pengguna dan sistem di UNAI (Universitas Advent Indonesia).

Online System UNAI sudah mulai beroperasi sejak tahun 2008. Sistem ini berguna bagi civitas akademik UNAI untuk mengakses berbagai data akademik hanya dengan mengunjungi *website* <https://www.online.unai.edu>. Dalam pengelolaan *Online System* UNAI, pelayanan yang terbaik selalu diberikan secara berkala, namun sampai saat ini masih sering muncul keluhan terkait penggunaannya. Berdasarkan wawancara dengan pengguna *Online System* UNAI keluhan yang ditemukan berupa informasi yang kurang akurat, terlambat *update* informasi, sering mengalami *error* saat mengakses *website*, terkadang hanya bisa diakses di ruang lingkup UNAI, dan terkadang mahasiswa dapat mengakses NIM dan *password* mahasiswa lain, atau kurang aman dalam data privasi mahasiswa. [1] Adanya pengaruh terhadap pengguna merupakan suatu penentuan dari kepuasan pengguna dan pengaruh terhadap sistem merupakan salah satu pengaruh dalam dampak organisasional dan lembaga pendidikan.

2. Metode Penelitian

Studi Literatur dan Pemahaman

Peneliti mengumpulkan atribut dari referensi yang berupa memahami dari jurnal, buku atau artikel yang berhubungan dengan penelitian.

Pengumpulan Data

Data yang digunakan untuk kepuasan penggunaan dalam penelitian berupa penyebaran kuesioner. Dengan menyebarkan kuesioner terhadap sampel yang telah ditentukan dan sampel yang digunakan adalah mahasiswa Universitas Advent Indonesia.

Setelah mendapatkan data, penulis kemudian akan mengolah dengan metode *Naïve Bayes* dan mendapatkan hasil yang akurat saat melakukan analisis dalam kepuasan pengguna terhadap *Online System* di Universitas Advent Indonesia.

Tahap Penelitian:

- 1) Pengumpulan atribut untuk digunakan pada saat pengelolaan data.
- 2) Pembuatan kuesioner sesuai dengan atribut yang sudah dikumpulkan.
- 3) Proses data pertama yaitu dengan memisahkan data kuesioner yang sudah terkumpul ke dalam suatu *file*, lalu disimpan di Microsoft Excel 2013.
- 4) Kemudian data yang sudah tersimpan diubah ke dalam bentuk format *.csv*.
- 5) Dengan data *.csv*, kemudian buka *file* menggunakan *software* WEKA, lalu menyimpannya kembali dengan format *.arff*.
- 6) Setelah itu, pilih *tool Classify*, dan pilih penggunaan metode yang akan digunakan. Penulis menggunakan metode *Naïve Bayes* dalam penelitian ini.
- 7) Setelah metode digunakan, selanjutnya langkah yang akan dilakukan yaitu dengan membuka *file* bentuk *.arff* yang sudah disimpan, dan pilih *use training set* untuk mendapatkan hasil dari *data training* dengan penggunaan metode *Naïve Bayes*.
- 8) Hasil akan didapatkan setelah melakukan proses *training* dan proses *testing* dengan menggunakan metode *Naïve Bayes*.

Attribute yang digunakan

Dalam menentukan tingkat kepuasan pengguna *Online System* UNAI, perlunya menggunakan atribut dalam melihat dimensi kualitas suatu *website* [2]

Tabel 1 *Attribute*

INDIKATOR	KETERANGAN
Kualitas Konten	Informasi yang memberikan kegunaan bagi konsumen dan menambah ilmu pengetahuan, lengkap mengenai informasi yang diberikan.
Relevan	Informasi yang disajikan harus tepat pada sasaran atau berguna bagi pemakainya.
Privasi	Adanya Keamanan saat mengunjungi <i>website</i> yang bersifat privasi akan data diri dari pengunjung.
Mudah Dioperasikan	Mahasiswa memahami alur penggunaan <i>Online System</i> .
Kecepatan	<i>Website</i> dapat diakses dengan cepat.
Daya Tarik Visual	Desain yang menarik untuk teks, gambar, suara, dan video terhadap konten di <i>website</i> .
Kelengkapan Secara <i>Online</i>	Layanan dalam kualitas pelayanan terhadap <i>Online System</i> dapat terpenuhi secara <i>online</i> .
<i>Customer Service</i>	Layanan responsif, membantu dan bersedia dengan waktu respons yang cepat terhadap pertanyaan atau keluhan yang dialami mahasiswa.

Naive Bayes

Setelah mendapatkan atribut untuk menilai kepuasan pengguna terhadap suatu website, penulis menggunakan atribut tersebut untuk menilai kepuasan pengguna terhadap *Online System* UNAI dengan penggunaan rumus metode *Naive Bayes* sebagai berikut :

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (1)$$

Keterangan :

X : Data dengan kelas yang belum diketahui

Y : Hipotesis data X adalah suatu kelas spesifik

P(Y|X) : Probabilitas hipotesis Y berdasarkan kondisi X

P(Y) : Probabilitas hipotesis Y

P(X|Y) : Probabilitas X berdasarkan kondisi pada saat hipotesis Y

P(X) : Probabilitas X

Attribute Selected Classifier

Information Gain

Untuk mendapatkan nilai dari *Information Gain* dan *Gain Ratio*, diperlukan untuk melakukan perhitungan nilai *Entropy*. Nilai *entropy* berguna untuk mengukur nilai ketidak murnian dari sekumpulan objek setiap cabang pada suatu atribut.

$$Entropy(S) = \sum_{i=1}^n P_i \log_2 P_i \quad (2)$$

Keterangan :

S = Himpunan Kasus

n = Jumlah Partisi S

P_i = Jumlah sampel kelas i

Setelah mendapatkan nilai *Entropy* langkah selanjutnya adalah perhitungan nilai *Information Gain* dari suatu atribut.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} Entropy(S_i) \quad (3)$$

Keterangan :

S = himpunan kasus

A = atribut

V = menyatakan suatu nilai yang mungkin untuk atribut A

|S_i| = jumlah kasus pada partisi ke-i

|S| = jumlah kasus dalam S *Entropy*

(S_i) = *entropy* untuk sampel-sampel yang memiliki nilai i

Gain Ratio

Untuk mendapatkan hasil dari perhitungan *Gain Ratio*, diperlukan untuk melakukan perhitungan nilai *Split Information*.

$$gain\ ratio = \frac{Gain(S,A)}{split\ information(S,A)} \quad (4)$$

CFS Subset

$$r_{zc} = \frac{kr_{zi}}{\sqrt{k+k(k-1)r_{ii}}} \quad (5)$$

Keterangan :

r_{zc} = Korelasi antara jumlah komponen dan variabel luar (kelas)

k = Jumlah komponen atau atribut

r_{zi} = Rata-rata korelasi antara komponen dan variabel luar (kelas)

r_{ii} = Rata-rata korelasi antar komponen

3. Hasil

Tabel 2 Perbandingan Hasil

Test Option	TP Rate	FP Rate	Precision	Recall	Class	Accuracy
Training Set	0.868	0.361	0.884	0.868	Puas	81.3%
	0.639	0.312	0.605	0.639	Tidak Puas	18.7%
Supplied Test Set	0.864	0.375	0.864	0.864	Puas	80%
	0.625	0.136	0.625	0.625	Tidak Puas	20%
Cross-Validation	0.842	0.389	0.873	0.857	Puas	78.3%
	0.611	0.158	0.55	0.611	Tidak Puas	21.3%
Percentage Split	0.763	0.538	0.806	0.763	Puas	68.6%
	0.462	0.237	0.400	0.462	Tidak Puas	31.4%

Perhitungan menggunakan metode *Naive Bayes* dan melakukan 4 kali pengujian, untuk klasifikasi puas hasil akurasi tertinggi mendapat nilai persentase 81.3% yaitu dengan pengujian *Training Set*. Dengan

penjelasan bahwa perhitungan menggunakan metode *Naïve Bayes* dengan pengujian 150 data sebagai *data training* dan mendapatkan nilai persentase yang sangat baik menunjukkan data dengan kelas puas diklasifikasikan benar dengan data asli.

Tabel 3 *Information Gain*

	RANKED	URUTAN	ATTRIBUTE
Naïve Bayes Info Gain	0.1243	1	Kelengkapan Secara <i>Online</i>
	0.1179	2	Kecepatan
	0.1015	3	Kualitas Konten
	0.0829	4	Daya Tarik Visual
	0.0614	5	<i>Customer Service</i>
	0.0555	6	Relevan
	0.0503	7	Privasi
	0	8	Mudah Dioperasikan

Information Gain dihitung untuk setiap atribut dalam data latih, lalu diurutkan berdasarkan nilai *information gain* dari yang terbesar sampai yang terkecil. Perhitungan yang didapat, bahwa yang terbesar sampai yang terkecil yaitu atribut kelengkapan secara *online*, kecepatan, kualitas konten, daya tarik visual, *customer service*, relevan, privasi dan mudah dioperasikan. Dari nilai *information gain* yang terbesar merupakan atribut yang paling berpengaruh untuk mengukur puas atau tidak puas dalam menggunakan *Online System* UNAI.

Tabel 4 *Gain Ratio*

	RANKED	URUTAN	ATTRIBUTE
Naïve Bayes Gain Ratio	0.1858	1	Kecepatan
	0.1623	2	<i>Customer Service</i>
	0.1566	3	Daya Tarik <i>Visual</i>
	0.1526	4	Kualitas Konten
	0.1459	5	Kelengkapan Secara <i>Online</i>
	0.0688	6	Relevan
	0.0517	7	Privasi
	0	8	Mudah Dioperasikan

Perhitungan yang didapat, bahwa yang terbesar sampai yang terkecil yaitu atribut kecepatan, *customer service*, daya tarik visual, kualitas konten, kelengkapan secara *online*, relevan, privasi dan mudah dioperasikan. Dari nilai *Gain Ratio* yang terbesar merupakan atribut yang paling berpengaruh untuk mengukur puas atau tidak puas dalam menggunakan *Online System* UNAI. *Information Gain* dapat terjadi data bias terhadap atribut yang memiliki banyak nilai unik. Dengan perhitungan *Gain Ratio* dapat mengatasi hal ini. [3]

Tabel 5 *CFS Subset*

	ATTRIBUTE
Naïve Bayes CfsSubset	Kualitas Konten
	Kecepatan
	Daya Tarik Visual
	Kelengkapan Secara <i>Online</i>
	<i>Customer Service</i>

Perhitungan yang didapat, bahwa CFS mengidentifikasi fitur yang relevan terhadap puas atau tidak puas dalam menggunakan *Online System* UNAI yaitu atribut kualitas konten, kecepatan, daya tarik visual, kelengkapan secara *online*, dan *customer service*. Atribut relevan, privasi, dan mudah dioperasikan harus dihapus karena atribut tersebut berkorelasi rendah terhadap kelas puas atau tidak puas.

4. Pembahasan

Pengujian menggunakan *Naïve Bayes* dengan *data training*

(1) *Accuracy*

Correctly Classified Instances	1215	55.965 %
Incorrectly Classified Instances	956	44.035 %
Correctly Classified Instances	122	81.3333 %
Incorrectly Classified Instances	28	18.6667 %

Gambar 1 *Summary Data Training*

Dari gambar di atas mengartikan bahwa data yang diklasifikasikan dengan benar adalah 81.333% sedangkan data yang diklasifikasikan secara tidak benar adalah 18.6667%.

(2) *Confusion Matrix*

```

=== Confusion Matrix ===
  a  b  <-- classified as
99 15 | a = Puas
13 23 | b = Tidak Puas
    
```

Gambar 2 *Confusion Matrix Cross Validation*

Dari gambar 2 *confusion matrix* dapat dilihat bahwa:

((1)) 99 data Puas yang benar diklasifikasikan menjadi data Puas, 15 data Puas yang salah diklasifikasikan menjadi data Tidak Puas.

((2)) 13 data Tidak Puas yang benar diklasifikasikan menjadi data Tidak Puas, 23 data Tidak Puas yang salah diklasifikasikan menjadi data Puas.

(3) *Precision Recall*

	TP Rate	FP Rate	Precision	Recall	Class	Precision	Recall
	0.868	0.361	0.884	0.868	Puas	0.479	0.279
	0.639	0.132	0.605	0.639	Tidak Puas	0.565	0.440
Weighted Avg.	0.813	0.306	0.817	0.813		0.576	0.769
						0.549	0.560

Gambar 3 *Detailed Accuracy*

Dari Gambar 3 dapat dijelaskan bahwa:

((1)) TP Rate mendapatkan nilai 0.868 (Puas) dikalkulasikan dari 99 data Puas dibagi dengan 114 data yang diklasifikasikan menjadi data Puas. TP Rate mendapatkan nilai 0.639 (Tidak Puas) dikalkulasikan dari 23 data Tidak Puas dibagi dengan 36 data yang diklasifikasikan menjadi data Tidak Puas.

((2)) FP Rate mendapatkan nilai 0.361 (Puas) dikalkulasikan dari 13 data Tidak Puas namun diklasifikasikan menjadi data Puas dibagi dengan 36 data yang diklasifikasikan menjadi data Tidak Puas. FP Rate mendapatkan nilai 0.132(Tidak Puas) dikalkulasikan dari 15 data Puas namun diklasifikasikan menjadi data Tidak Puas dibagi dengan 114 data yang diklasifikasikan menjadi data Puas.

((3)) *Precision* dari data Puas adalah 0.884 yang dikalkulasikan dari 99 data Puas dibagi dengan 112 data yang diklasifikasikan menjadi data Puas. Hasil *Recall* adalah 0.868 yang dikalkulasikan dari 23 data Puas dibagi dengan 114 data keseluruhan Puas.

5. Kesimpulan

1. Dari hasil penelitian dan pengujian yang dilakukan penulis, maka metode *Naïve Bayes* dapat digunakan untuk menentukan kepuasan pengguna (Puas dan Tidak Puas).
2. Metode *Naïve Bayes* berhasil mengklasifikasikan 122 data dengan benar dari 150 *data training*, dan menunjukkan bahwa klasifikasi penentuan kepuasan pengguna menggunakan metode *Naïve Bayes* mendapatkan nilai akurasi terbesar dengan pengujian *data training* yaitu dengan persentase 81.3% dibandingkan dengan pengujian yang lain.
3. Dari delapan (8) atribut, ada lima (5) atribut yang mempengaruhi tingkat kepuasan pengguna karena mempunyai tingkat persentase jawaban lebih tinggi yaitu Kualitas Konten, Kecepatan, Daya Tarik *Visual*, Kelengkapan Secara *Online*, dan *Customer Service*.
4. Sedangkan tiga (3) atribut lainnya bukan merupakan faktor yang mempengaruhi tingkat kepuasan pengguna, karena tingkat persentase relatif kecil yaitu atribut Relevan, Privasi, dan Mudah Dioperasikan.

Saran

1. Hal ini diharapkan dapat digunakan untuk meningkatkan akurasi dalam memprediksi tingkat kepuasan pengguna *Online System* UNAI.
2. Dengan didapatnya atribut yang dapat mempengaruhi tingkat kepuasan pengguna, atribut tersebut dapat lebih ditingkatkan lagi.
3. Dilakukan penelitian terkait dengan metode *Webqual* 4.0 untuk mengukur tingkat kepuasan pengguna terhadap kualitas layanan *website*, atau menggabungkan beberapa metode sehingga diharapkan dapat meningkatkan persentase akurasi

6. Daftar Pustaka

- [1] B. Claudya, "Analisis Pengaruh Persepsi Manfaat dan Persepsi Kemudahan Terhadap Sikap Pengguna Sistem *Online* UNAI," Teknik Informatika, Universitas Advent Indonesia, 2018.
- [2] Puspitasari, An Nissaa. 2013. Pengaruh Kualitas *Website* Terhadap Nilai yang dipersepsikan, Kepuasan dan Loyalitas Pelanggan Pada *Online Shop*. Jurnal Ilmu Administrasi.
- [3] A.K.H. Audio, "Klasifikasi Penyakit Hipertensi Menggunakan Algoritma C4.5 Studi Kasus RSU Provinsi NTB," Teknik Informatika, Universitas Sanata Dharma, 2017.