

## Perbandingan Metode Naïve Bayes Classifier Dan Holistic Lexicon Based Dalam Analisis Sentimen Angket Mahasiswa

Suprianto<sup>1</sup>, Aida Indriani<sup>2</sup>, Muhammad<sup>3</sup>

<sup>1,2,3</sup> Program Studi Sistem Informasi, STMIK PPKIA Tarakanita Rahmawati

Email: [supri@ppkia.ac.id](mailto:supri@ppkia.ac.id)<sup>1</sup>, [aida@ppkia.ac.id](mailto:aida@ppkia.ac.id)<sup>2</sup>, [muhhammad@ppkia.ac.id](mailto:muhhammad@ppkia.ac.id)<sup>3</sup>

### Abstrak

Analisis sentimen bertujuan untuk menentukan kelas dari data opini mahasiswa yang terdapat dalam form angket mahasiswa berupa kelas positif, negatif, dan netral. Dalam data mining, teknik yang biasa digunakan adalah teknik klasifikasi. Klasifikasi ini digunakan untuk membentuk suatu model yang belum terklasifikasi menjadi terklasifikasi. Metode naïve bayes classifier dan metode holistic lexicon based biasa digunakan untuk melakukan sentimen analisis. Analisis perbandingan suatu metode dilakukan untuk menemukan metode terbaik dalam memecahkan sebuah kasus. Penelitian ini melakukan perbandingan antara metode naïve bayes classifier dan metode holistic lexicon based yang digunakan untuk mengklasifikasikan data komentar angket mahasiswa yang dapat bernilai analisis sentimen berupa kelas positif, kelas negatif, dan kelas netral. Metode naïve bayes classifier menggunakan data latih untuk menentukan kelasnya, sedangkan metode holistic lexicon based menggunakan kamus kata sifat untuk penentuan kelasnya. Dari hasil perbandingan ini dapat disimpulkan bahwa metode naïve bayes classifier memiliki nilai precision dan tingkat accuracy yang lebih baik dibandingkan dengan metode holistic lexicon based.

**Kata kunci:** perbandingan, analisis sentimen, naïve bayes classifier, holistic lexicon based

### Abstract

Sentiment analysis aims to determine the class of student opinion data contained in the student questionnaire form in the form of positive, negative, and neutral classes. In data mining, the technique commonly used is classification techniques. This classification is used to form a model that has not been classified to be classified. The naïve bayes classifier method and the lexicon based holistic method are commonly used to carry out analytical sentiments. Comparative analysis of a method is done to find the best method for solving a case. This study conducted a comparison between the naïve bayes classifier method and the lexicon-based holistic method which was used to classify student questionnaire commentary data that could be valued in sentiment analysis in the form of positive classes, negative classes, and neutral classes. The naïve bayes classifier method uses training data to determine the class, while the lexicon based holistic method uses an adjective dictionary for class determination. From the results of this comparison it can be concluded that the naïve bayes classifier method has a better precision value and accuracy level compared to the lexicon based holistic method.

**Keywords:** comparison, sentiment analysis, naïve bayes classifier, holistic lexicon based

## 1. PENDAHULUAN

Analisis sentimen adalah sebuah proses dalam memahami, mengevaluasi atau menilai pendapat untuk mendapatkan informasi. Analisis sentimen memiliki prinsip dasar yaitu klasifikasi polaritas yang terdapat didalam teks dan dikemukakan menjadi kelas positif, negatif dan netral.

STMIK PPKIA Tarakanita Rahmawati menggunakan angket penilaian mahasiswa dalam rangka melakukan penilaian terhadap dosen. Angket tersebut diisi oleh mahasiswa yang berisi opini positif, negatif, atau netral. Saat ini, perekapan dan klasifikasi hasil angket untuk dosen ini dilakukan secara manual. Pengklasifikasian secara manual memang menghasilkan data yang akurat karena manusia dapat

membedakan dengan tepat apakah kata atau kalimat tersebut bermakna positif, negatif atau netral, namun hal ini tidak efektif dan tidak menutup kemungkinan adanya kalimat ambigu yang sulit diklasifikasikan walaupun oleh manusia sekalipun. Selain itu, pengklasifikasian secara manual membutuhkan banyak waktu dan tenaga.

Berdasarkan penelitian sebelumnya dengan judul “Analisis Sentimen menggunakan Metode Naïve Bayes Classifier pada Angket Mahasiswa”, dari 10 (sepuluh) data uji yang dilakukan diperoleh nilai precision sebesar 75%, recall sebesar 75% dan accuracy sebesar 80% (Sunardi, Abdul, F., & Suprianto, 2018). Pada penelitian ini, penulis menggunakan metode holistic lexicon based sebagai pembanding untuk melakukan analisis sentimen dengan menggunakan data uji yang sama. Tingkat perbandingan dilihat dari nilai precision, recall dan accuracy yang diperoleh dari kedua metode yang digunakan.

## 2. METODOLOGI PENELITIAN

Penelitian ini menggunakan langkah-langkah metodologi yang memiliki tahapan pada Figure 1.

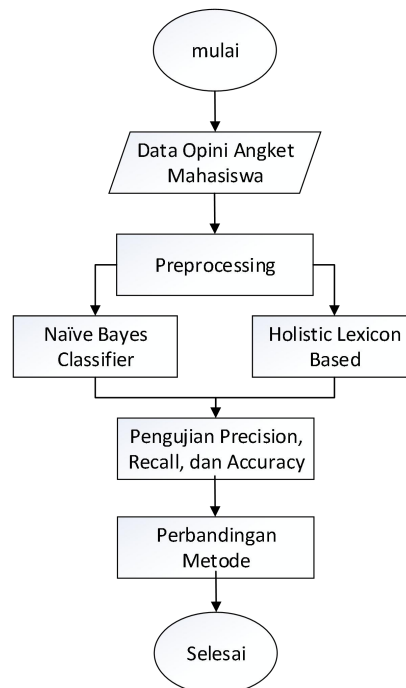


Figure 1. Metodologi Penelitian

Langkah yang dilakukan pada metodologi penelitian dapat dijelaskan sebagai berikut.

### 2.1. Data Opini Angket Mahasiswa

Data opini angket mahasiswa adalah data yang diperoleh dengan cara menyebarkan angket kepada mahasiswa yang dilakukan pada 1 (satu) kali setiap semester. Mahasiswa dapat menuliskan apa saja seputar kampus mulai dari proses belajar mengajar, fasilitas, biaya kuliah dan lain sebagainya. Data angket biasanya dituliskan dalam bentuk esay, dimana mahasiswa bebas menuliskan saran, pernyataan, atau pertanyaan.

## 2.2. Preprocessing

*Preprocessing* adalah proses merubah suatu teks yang tidak terstruktur menjadi lebih terstruktur (Murnawan & Ardiles, S., 2017). Tahapan dalam preprocessing ada 4 (empat) yaitu *case folding*, *tokenizing*, *filtering* dan *stemming*. Proses yang terpenting pada *preprocessing* yaitu *stemming*, karena pada *stemming* dilakukan penghilangan kata imbuhan sehingga menjadi kata dasar (Yoga, D.P., Sigit, S.P., & Nurul, M., 2017).

Pada tahapan *case folding*, terjadi perubahan bentuk kata-kata menjadi *lower case* atau *upper case*. *Tokenizing* yaitu suatu tahapan yang digunakan untuk mengidentifikasi kata pada teks sehingga menjadi potongan-potongan kata. *Filtering* bertujuan untuk membuang kata-kata yang bersifat umum (Wina, W., Herry, R., & Faiza, R., 2016).

## 2.3. Naive Bayes Classifier

Metode *naive bayes classifier* merupakan sebuah metode klasifikasi berdasarkan probabilitas atau kemunculan dengan perkiraan bahwa setiap variabel X memiliki sifat bebas atau berdiri sendiri dan tidak ada kaitannya dengan variabel lainnya. Metode NBC mempunyai dua tahapan dalam proses klasifikasi teks, yaitu tahap pelatihan dan klasifikasi (Amir Hamzah, 2012). *Probabilitas* adalah kemungkinan terjadinya suatu peristiwa antara 0 s/d 1 (wahyudi & fadlil, 2013). Klasifikasi *gaussian naive bayes* dapat digunakan untuk memproses atribut numerik pada layanan jaringan komputer (Fadlil, Riadi, & Aji, 2017). Proses analisis dilakukan pada tahap pelatihan terhadap sampel data yang bisa menjadi representasi dari dokumen. Perhitungan perbandingan antara *term* dengan data testing pada setiap kelas yang ada dapat dilakukan dengan persamaan 1 (Eric Meisner, 2010).

$$P(a_j|v_j) = \frac{n_c + mp}{n + m} \quad (1)$$

keterangan:

n : jumlah *term* pada data latih dimana  $v = v_j$

$n_c$  : jumlah *term* dimana  $v = v_j$  dan  $a = a_j$

p : *probabilitas* setiap kelas dalam data latih

m : jumlah *term* pada data uji

untuk menentukan hasil klasifikasi pada data uji dilakukan perhitungan dengan persamaan 2.

$$V_{nb} = \operatorname{argmax}_{v_j \in V} P(v_j) \prod P(a_i|v_j) \quad (2)$$

## 2.4. Holistic Lexicon Based

*Holistic lexicon based* merupakan metode yang dapat mengidentifikasi sentimen dari setiap *opinion words* yang terdapat pada data hasil angket dan dapat menangani permasalahan *multi-opini* di dalam suatu data. Didalam penanganan masalah *multi-opini*, metode ini mengumpulkan seluruh sentimen dari kata opini berdasarkan jarak antara kata *opini* dengan fiturnya. Dengan diterapkannya metode *holistic lexicon based*, sistem ini diharapkan dapat melakukan analisis sentimen dengan cepat, mudah, dan efektif. Proses klasifikasi penentuan data dilakukan pada setiap kata didalam kalimat menggunakan *lexicon based* dengan *corpus*. Kata-kata yang berhasil terklasifikasi sesuai dengan *corpus* akan dilakukan perhitungan nilai sentimen didalam sebuah kalimat pada persamaan 3 dan 4.

$$S_{positif} = \sum_{i \in t}^n \text{Positif score } i \quad (3)$$

$$S_{negatif} = \sum_{i \in t}^n \text{Negatif score } i \quad (4)$$

*Spositif* merupakan kalimat yang memiliki bobot dimana didapatkan melalui penjumlahan dari banyaknya  $n$  skor *polaritas* berupa kata opini berkonotasi positif. *Snegatif* merupakan kalimat yang memiliki bobot dimana didapatkan melalui penjumlahan dari banyaknya  $n$  skor *polaritas* kata opini berkonotasi negatif. Setiap kalimat ini menggunakan bobot sebagai acuan untuk melakukan proses perbandingan, dimana didalam sebuah kalimat jumlah nilai positif (*Spositif*) dan nilai negatif (*Snegatif*) akan diketahui dari tiap- tiap kata penyusunnya. Nilai dari analisis sentimen dapat diperoleh dengan menggunakan persamaan 3 untuk menentukan arah sentimen dengan dibandingkan jumlah nilai negatif, positif, dan netral.

Setelah diketahui semua kata kunci yang mempunyai nilai sentimen, selanjutnya dihitung nilai kemunculan sentimen positif dan negatif mana yang lebih dominan. Jika nilai sentimen positif lebih dominan maka nilai sentimen untuk kalimat tersebut adalah positif, namun jika nilai sentimen negatif lebih dominan maka nilai sentimen untuk kalimat tersebut adalah negatif, namun jika nilainya sama antara sentimen negatif dan sentimen positif maka nilai sentimen untuk kalimat tersebut adalah netral. Penentuan sentimen *value* dapat dilihat pada persamaan 5.

$$\text{Sentimen value} \begin{cases} \text{Positif if } S_{positif} > S_{negatif} \\ \text{Netral if } S_{positif} = S_{negatif} \\ \text{Negatif if } S_{positif} < S_{negatif} \end{cases} \quad (5)$$

## 2.5. Precision, Recall, dan Accuracy

Jawaban dari sebuah informasi yang sesuai dari yang diminta merupakan sebuah tingkat keberhasilan yang disebut *precision*. Tingkat keberhasilan *recall* didapatkan jika dapat menemukan kembali sebuah informasi, sedangkan *accuracy* merupakan tingkat kedekatan dari nilai prediksi dengan nilai yang sebenarnya. Perhitungan akurasi dinyatakan dalam persamaan 6, 7 dan 8 (Puspitasari, A., Santoso, E., & Indriati, 2017).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

## 2.6 Perbandingan Metode

Perbandingan metode adalah melakukan uji coba dengan menggunakan lebih dari satu metode dengan menggunakan data uji yang sama. Pada penelitian ini, penulis membandingkan *metode naïve bayes classifier* dan *holistic lexicon based* dalam melakukan analisa sentimen pada komentar angket mahasiswa. Perbandingan dilihat dari nilai akhir yaitu nilai *precision*, *recall* dan *accuracy*. Data uji yang digunakan berjumlah 10 (sepuluh) data uji yang telah digunakan pada penelitian sebelumnya (Suprianto, Sunardi, A. Fadlil, 2018).

### 3. HASIL DAN PEMBAHASAN

Berikut ini adalah pembahasan dari analisis sentimen menggunakan metode *holistic lexicon based*.

#### 1.1 Preprocessing

Tahapan awal pada metode *holistic lexicon based* adalah melakukan *preprocessing* terhadap kalimat yang terdapat pada angket. Berdasarkan penelitian sebelumnya, beberapa data angket yang digunakan sebagai data uji terdapat pada Tabel 1.

Tabel 1. Data uji angket mahasiswa

No.	Data Angket	Kelas
D1	Dosen sangat sering terlambat mengajar!!!	?
D2	Materi yang diajarkan sudah sangat baik, terimakasih	?

Data pada Tabel 1, dilakukan *preprocessing* sehingga menghasilkan token yang dapat dilihat pada Tabel 2.

Tabel 2. Preprocessing Data Uji

No.	Token
D1	dosen
	lambat
	ajar
	materi
D2	ajar
	baik
	terimakasih

#### 1.2 Kamus Kata Sifat

Sentimen analisis pada metode *holistic lexicon based* dengan melakukan pelabelan terhadap kata yang termasuk kata positif atau kata negatif. Pada penelitian ini, penulis menggunakan 66 (enam puluh enam) kata sifat positif dan 54 (lima puluh empat) kata sifat negatif (Evasaria, M.S., Herastia, M., & Ivan, Z., 2016). Beberapa kata sifat yang terdapat pada kamus kata sifat positif dan negatif ditunjukkan pada Tabel 3.

Tabel 3. Kamus Kata Sifat Positif dan Negatif

No.	Positif	No.	Negatif
1	puas	1	antri
2	dingin	2	bau
3	lengkap	3	berbunyi
4	profesional	4	berdebu
5	terjangkau	5	berisik
6	menarik	6	berkarat
7	mewah	7	berpencar

8	murah	8	bingung
9	pusat	9	bising
10	respon	10	bocor
11	mengesankan	11	buruk
12	membantu	12	debu
13	memadai	13	lambat
14	senyum	14	jorok
15	dekat	15	jutek
..	....	..	....
63	baik	51	merembes
64	memanjakan	52	Noda
65	memuaskan	53	Panjang
66	rendah	54	Usang

### 1.3 Pengelompokkan dan Perhitungan Nilai Sentimen

Pengelompokkan kata pada dokumen kedalam kata positif atau kata negatif dilakukan dengan mencocokkan kata dokumen dengan kamus kata sifat apakah termasuk kata positif atau kata negatif. Pengelompokkan kata berdasarkan kamus kata sifat dapat dilihat pada Tabel 4.

Tabel 4. *Pre processing* Data Uji

No.	Token	Kelompok Kata
	dosen	-
D1	lambat	Negatif
	ajar	-
	materi	-
D2	ajar	-
	baik	Positif
	terimakasih	-

Dari Tabel 4, dilakukan perhitungan score positif menggunakan persamaan 3 dan score negatif menggunakan persamaan 4. Dokumen D1, *score* positif = 0 dan negatif = 1. Dokumen D2, *score* positif = 1 dan negatif = 0.

### 1.4 Penentuan Sentimen Value

Dalam penentuan sentimen *value* untuk setiap dokumen menggunakan persamaan 5. Dokumen D1 memiliki *score* positif < *score* negatif, maka sentimen *valu*nya adalah negatif. Dokumen D2 memiliki *score* positif > *score* negatif, maka sentimen *valu*nya adalah positif. Sentimen *value* untuk 10 (sepuluh) data uji dapat dilihat pada Tabel 5.

Tabel 5. Hasil Sentimen *Value* 10 Data Uji

No.	Sentimen Value
D1	Negatif

D2	Positif
D3	Negatif
D4	Negatif
D5	Positif
D6	Negatif
D7	Positif
D8	Netral
D9	Positif
D10	Negatif

### 1.5 Precision, Recall dan Accuracy

Perbandingan dilakukan untuk mendapatkan nilai *precision*, *recall* dan *accuracy*, maka diperlukan kelas sebenarnya pada 10 (sepuluh) data uji. Perbandingan kelas sebenarnya dengan kelas perhitungan *holistic lexicon based* dapat dilihat pada Tabel 6.

Tabel 6. Kelas Sebenarnya dan Kelas Perhitungan

No.	Kelas Sebenarnya	Kelas Holistic Lexicon Based
D1	Negatif	Negatif
D2	Positif	Positif
D3	Positif	Negatif
D4	Negatif	Negatif
D5	Netral	Positif
D6	Negatif	Negatif
D7	Negatif	Positif
D8	Netral	Netral
D9	Positif	Positif
D10	Negatif	Negatif

Dengan menggunakan persamaan 6, 7 dan 8 diperoleh nilai *precision*, *recall* dan *accuracy* dengan terlebih dahulu menentukan nilai *true positif* (TP), *true negatif* (TN), *false positif* (FP) dan *false negatif* (FN).

Dari tabel 5, diperoleh nilai TP = 7, TN = 0, FP = 3, dan FN = 0. Dari nilai TP, TN, FP dan FN diperoleh nilai *precision* sebesar 70%, nilai *recall* 100%, dan nilai *accuracy* 70%. Perbandingan nilai *precision*, *recall* dan *accuracy* untuk metode *naïve bayes classifier* dan *holistic lexicon based* ditunjukkan pada Tabel 7.

Tabel 7. Nilai Perbandingan Metode

Jenis	Naïve Bayes Classifier	Holistic Lexicon Based
Precision	75%	70%
Recall	75%	100%
Accuracy	80%	70%

Pada Tabel 7, dapat diketahui bahwa metode *naïve bayes classifier* yang digunakan



pada penelitian sebelumnya memiliki nilai *precision* dan *accuracy* lebih baik dari pada metode *holistic lexicon based*.

#### 4. PENUTUP

Dari hasil pembahasan dan pengujian yang dilakukan terhadap 2 metode maka dapat disimpulkan bahwa metode *naïve bayes classifier* menggunakan data latih untuk melakukan analisis sentimen, sedangkan metode *holistic lexicon based* menggunakan kamus kata sifat untuk menentukan analisis sentimen terhadap komentar data angket mahasiswa. Dari masing-masing perhitungan yang telah dilakukan, metode *naive bayes classifier* memiliki tingkat nilai *precision* sebesar 75% dan nilai *accuracy* sebesar 80% lebih tinggi dibandingkan metode *holistic lexicon based* dengan tingkat nilai *precision* sebesar 70% dan tingkat nilai *accuracy* sebesar 70%. Dengan demikian dapat disimpulkan bahwa metode *naive bayes classifier* masih lebih baik daripada metode *holistic lexicon based* pada tingkat *precision* (kelas sebenarnya) dan *accuracy* (akurasi).

#### 5. DAFTAR PUSTAKA

- [1] A Fadlil, I Riadi, S Aji, 2017. DDoS Attacks Classification using Numeric Attribute-based Gaussian Naive Bayes. Japan International journal of advanced computer science and applications, Vol. 8, No. 8, page 42-50.
- [2] Amir Hamzah. 2012. Klasifikasi Teks dengan Naïve Bayes Classifier (NBC) untuk Pengelompokan Teks Berita dan Abstract Akademis. Prosiding Seminar Nasional Apikasi Sains & Teknologi (SNAST) Periode III, p.B269-B277. Yogyakarta.
- [3] DT Yuwono, A Fadlil, S Sunardi, 2017, penerapan metode forward chaining dan certainty factor pada sistem pakar diagnosa hama anggrek *coelogyne pandurata*, Banjarmasin, klik-kumpulan jurnal ilmu komputer 4 (2), 136-145.
- [4] E M Sipayung, H Maharani, I Zevanya, 2016. Perancangan sistem analisis sentimen komentar pelanggan menggunakan metode *naive bayes classifier*. Jurnal sistem informasi, Vol. 8, No. 1.
- [5] Eric Meisner. 2010. Naïve Bayes Classifier Example.
- [6] Murnawan, A Sinaga, 2017. Pemanfaatan Analisis Sentimen Untuk Peningkatan Popularitas Tujuan Wisata. Jurnal Penelitian Pos dan Informatika, Vol. 7 No. 2, page 109-120.
- [7] Puspitasari, A., Santoso, E., & Indriati, I. Klasifikasi Dokumen Tumbuhan Obat Menggunakan Metode Improved K-Nearest Neighbor. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 2, no. 2, p. 486-492, agu. 2017. ISSN 2548-964X.
- [8] Sunardi, A Fadlil, Suprianto, 2018. Analisis sentimen menggunakan metode *naïve bayes* pada angket mahasiswa. Jurnal sains dan teknologi, vol. 10, No. 2, page 1-9.
- [9] W Witanti, H Rahmanto, F Renaldi, 2016. Pembangunan sistem temu balik informasi (Information Retrieval) dalam pemilihan pemain sepak bola berkualitas



di Indonesia berbasis analisis sentimen. Seminar Nasional Teknologi Informasi dan Komunikasi 2016, No. 34.

- [10] Y D Pramudita, S S Putro, N Makhmud, 2018. Klasifikasi Berita Olahraga Menggunakan Metode Naïve Bayes dengan Enhanced Confix Stripping Stemmer. Jurnal Teknologi dan Ilmu Komputer, Vol. 5 No. 3, page 269-276.