

PERANCANGAN SISTEM ANALISIS SENTIMEN KOMENTAR PELANGGAN MENGUNAKAN METODE NAIVE BAYES CLASSIFIER

Evasaria M. Sipayung¹, Herastia Maharani², Ivan Zefanya³

^{1,2,3}Departemen Sistem Informasi, Institut Teknologi Harapan Bangsa, Bandung

¹evasaria@ithb.ac.id, ²herastia@ithb.ac.id, ³ivanzefanya274@gmail.com

Abstract

Grand Royal Panghegar is a company runs in hospitality sector located in Bandung. This hotel faced problem in getting the meaning or conclusion of comments from customers about hotel's products and services, because the amount of the comments reached about 675 comments every year. To overcome the problem, developing a tool named sentiment analysis system. This system supports the hotel to get the meaning from the large comments using Naive Bayes Classifier (NBC) method. This method classified the categories that will reviewed by the hotel and divided by positive and negative sentiment, so hotel can be evaluated by the customer satisfaction to products and services that provided computerized, specific, and systematic. The result from this research is six categories that reviewed with 55 keywords of nouns. From this research got 120 keywords sentiment with 66 of positive sentiments and 54 of negative sentiments. The result of processing from 175 training sets by system can be concluded that the most often sentiment that appear is sentiment positive for 155 comments and 20 comments of negative sentiment. And then for the highest positive sentiment category is hotel room with 73 comments and 17 comments for the highest negative sentiment category. The accuracy of this system to determine the category is 77.14% and the precision to determine the sentiment is 99.12%, recall 72.9%, and accuracy is 75.42%.

Keywords: sentiment analysis, comments, naïve bayes classifier,

Abstrak

Hotel XYZ mengalami kesulitan untuk mendapatkan makna atau kesimpulan dari keseluruhan komentar yang diberikan pelanggan terhadap produk dan layanan hotel dikarenakan banyaknya komentar yang ada, pertahun mencapai 675 komentar. Sistem analisis sentiment analysis system bertujuan untuk membantu pihak hotel dalam mendapatkan makna dari komentar yang banyak dengan menggunakan metode Naive Bayes Classifier (NBC). Metode ini mengelompokkan komentar berdasarkan kategori-kategori yang ditinjau oleh hotel. Komentar dibagi berdasarkan sentimen positif dan negatif, sehingga dapat dievaluasi kepuasan pelanggan terhadap produk dan jasa yang disediakan secara terkomputerisasi dan spesifik. Hasil dari penelitian yang dilakukan mendapatkan enam kategori yang ditinjau dengan 55 keyword kata benda, terdapat 120 keyword sentimen dengan 66 kata sentimen positif dan 54 kata sentimen negatif. Hasil pengolahan terhadap 175 data latih disimpulkan bahwa hasil klasifikasi sentimen yang didapat adalah sentimen positif sebanyak 155 komentar dan sentimen negatif sebanyak 20 komentar. Kategori sentimen positif terbesar adalah kategori kamar sebesar 73 komentar dan kategori dengan sentimen negatif terbesar adalah kategori kamar sebesar 17 komentar. Tingkat akurasi dalam penentuan kategori adalah sebesar 77.14% dan 75.42% dalam penentuan sentimen memiliki tingkat precision 99.12% dan recall 72.9%.

Kata kunci: analisis sentimen, komentar, naive bayes classifier

1. Pendahuluan

Perkembangan bisnis perhotelan kini menjadi lebih mudah dengan adanya pemesanan kamar hotel melalui perusahaan penyedia jasa pemesanan kamar hotel via *online*. Pada saat ini, hotel XYZ sudah bekerja sama dengan perusahaan pemesanan kamar via online, salah satunya adalah agoda.com. Pengguna hotel cenderung memesan melalui agoda.com dikarenakan lebih cepat dan lebih mudah. Pengguna yang memesan melalui agoda.com diundang untuk memberikan *feedback* dan komentar terhadap hotel tersebut. Pelanggan yang memberikan komentar kadang tidak sesuai dengan penilaian yang diberikan. Komentar yang

diberikan oleh salah satu pelanggan ini jika dibaca dan dimengerti makna komentarnya merupakan komentar yang positif, tidak ada satu pun kata yang negatif dalam komentar ini, namun penilaian yang diberikan oleh pelanggan adalah 4,7 dari skala 10,0 yang terbilang kurang baik, sehingga dapat disimpulkan bahwa komentar dengan penilaian yang diberikan tidaklah sesuai atau berlawanan, dari sini dapat dilihat bahwa komentar yang positif tidak mendapat nilai yang baik atau sebaliknya. Namun jika menggunakan bantuan *sentiment analysis* pihak hotel dapat menganalisis dengan mengandalkan komentar yang diberikan dari pelanggan dan kesimpulan berdasarkan komentar yang diberikan berdasarkan makna komentar tidak dari nilai yang diberikan.

Analisis sentimen adalah sebuah cara yang digunakan untuk mengolah komentar yang diberikan oleh pemesan atau pelanggan melalui berbagai media, mengenai sebuah produk, jasa ataupun sebuah instansi. Permasalahan yang dihadapi Hotel XYZ adalah pengolahan komentar pelanggan terhadap produk dan layanan hotel yang banyak masuk untuk hotel, sehingga pihak hotel mengalami kesulitan dalam menangani komentar tersebut. Oleh karena itu dirancang sebuah sistem yang dapat membantu pihak hotel dalam mengetahui dan mengelompokkan komentar pelanggan, berdasarkan kelompok kategori dan sentimen. Hal tersebut dapat digunakan pihak hotel untuk evaluasi produk dan layanan hotel.

2. Landasan Teori

2.1 Analisis Sentimen

Sentiment Analysis adalah merupakan perpaduan dari *data mining* dan *text mining*, atau sebuah cara yang digunakan untuk mengolah berbagai opini yang diberikan oleh konsumen atau para pakar melalui berbagai media, mengenai sebuah produk, jasa ataupun sebuah instansi. *Sentiment analysis* merupakan sebuah metode yang digunakan untuk memahami, mengekstrak data opini, dan mengolah data tekstual secara otomatis untuk mendapatkan sebuah *sentiment* yang terkandung dalam sebuah opini. Pada *Sentiment analysis* terdiri dari 3 jenis opini, yaitu opini positif, opini negatif dan opini netral, sehingga dengan sentimen analysis perusahaan atau instansi yang terkait dapat mengetahui respon masyarakat terhadap suatu pelayanan atau produk, melalui *feedback* masyarakat atau pun para ahli.

Sentiment mengacu pada fokus topik tertentu, pernyataan pada suatu topik mungkin akan berbeda makna dengan pernyataan yang sama pada subjek yang berbeda, oleh karena itu pada beberapa penelitian, terutama pada review produk, pekerjaan didahului dengan menentukan elemen dari sebuah produk yang sedang dibicarakan sebelum memulai proses *Sentiment Analysis* [5].

2.2 Naive Bayes Classification (NBC)

Naive Bayes dikembangkan oleh Reverend Thomas Bayes pada abad ke 18. Klasifikasi dengan metode *Naive Bayes* secara umum dilakukan dengan pendekatan peluang atau probabilitas. Algoritma ini memprediksi probabilitas di masa depan berdasarkan pengalaman yang sudah ada di masa lalu. Dari sudut pandang peluang, berdasarkan aturan Bayes kedalam kelas c adalah [3]:

$$P(c|d) = \frac{P(d|c)P(c)}{P(d)} \quad \text{persamaan} \quad (1)$$

Pilihan kelas ditentukan dengan menggunakan peluang maksimal dari seluruh c dalam C , dengan fungsi:

$$\operatorname{argmax}_{c \in C} \frac{P(d|c)P(c)}{P(d)} \quad \text{persamaan} \quad (2)$$

Nilai $P(d)$ konstan untuk semua kelas, maka $P(d)$ dapat diabaikan, sehingga menghasilkan fungsi:

$$f_c(d) = \operatorname{argmax}_{c \in C} P(d|c) * P(c) \quad \text{persamaan} \quad (3)$$

3 Analisis dan Perancangan Sistem

3.1 Analisis Sistem

Komentar-komentar yang ditujukan untuk hotel sebagai bentuk *feedback* atau penilaian dari masyarakat mengenai hotel. Komentar-komentar dari pelanggan terhadap produk dan layanan Hotel XYZ secara garis besar dapat dibedakan menjadi 2, yaitu komentar positif dan negatif. Pihak hotel belum melakukan tindakan lebih lanjut mengenai komentar-komentar yang tersebar di situs-situs valid. Tabel 1 yang memperlihatkan banyak komentar pada situs pembookingan yang ada.

Tabel 1 Jumlah Komentar Pada Situs Pembookingan Per Tahun 2014

Situs Pembookingan Hotel	Jumlah Komentar
Agoda.com	2700
Booking.com	270
Tripadvisor.com	170

Komentar terbanyak terdapat pada situs agoda.com maka diambil data latih dari situs tersebut dan komentar yang diambil adalah 175 komentar selama tahun 2014.

Pengolahan data latih yang dilakukan terhadap data latih terdiri dari:

1. Klasifikasi Kategori Menggunakan Naive Bayes Classifier
 1. 1. Perhitungan Jumlah Kata Per Komentar
 1. 2. Menentukan Peluang Setiap Kategori pers 1
 1. 3. Menghitung Peluang Bersyarat/likelihood Kategori
 1. 4. Membandingkan Data Berdasarkan Nilai Peluang Tinggi

Terdapat 6 kategori *keyword* unsur produk dan layanan yang didapatkan dari hasil pengolahan data latih yang membentuk 55 kata untuk pengelompokan kategori komentar tertentu, dan juga mendapatkan 120 kata sifat yang dapat menentukan makna dari sebuah komentar yang digunakan dalam mengolah komentar, sehingga dapat diketahui berapa besar peluang setiap peluang dari setiap kategori yang muncul. Hal ini dipengaruhi oleh banyaknya *keyword* yang digunakan. Tabel 2 terdiri dari kategori dan *keyword* yang didapat untuk masing-masing kategori dari data latih.

Tabel 2. Kategori dan Keyword

Kategori	Keyword	Kategori	Keyword
kamar	kamar	fasilitas	fasilitas

	seprai		internet
	tv		parkir
	lantai		lift
	tembok		restaurant
	closet		air
	ranjang		kolam
	sofa		futsal
	meja	pelayanan	pelayanan
	dapur		check
	lemari		staff
	toilet	harga	harga
	air		penawaran
	bathtub		rate
	microwave		promosi
	washtafel	umum	umum
	shower		market
	hairdyer		perabot
	karpel		manajemen
	selimut		interior
	handuk		keamanan
	ac		koran
	kasur		lokasi
	kulkas		suasana
makanan	makanan		lobby
	sarapan		listrik
	rasa		atm
			ruangan
			lingkungan

2. Klasifikasi Sentimen Menggunakan Naive Bayes Classifier

2. 1. Menghitung Peluang Bersyarat/likelihood Sentimen

2. 2. Membandingkan Data Berdasarkan Nilai Peluang Tertinggi

Sentimen dilakukan terhadap peluang positif dan negatif, sentimen positif yang terdapat pada database kata sifat adalah dengan jumlah 66 dan negatif dengan jumlah 54 dengan total 120. Tabel 3 berisi kata sifat positif dan negatif.

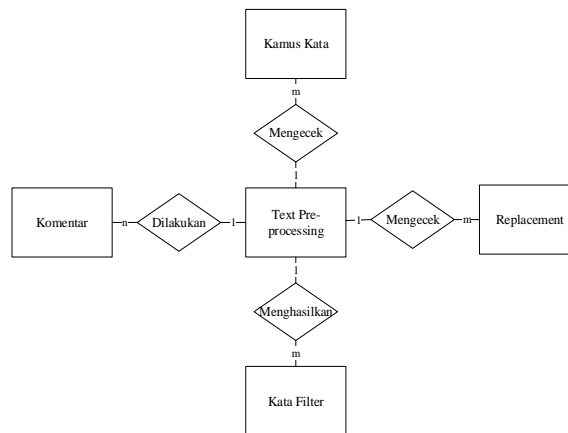
Tabel 3. Kamus Kata Sifat

No	Positif	No	Negatif
1	puas	1	antri
2	dingin	2	bau
3	lengkap	3	berbunyi
4	profesional	4	berdebu
34	romantis	34	kalah
35	terawat	35	kecewa
36	panas	36	kecil
37	ramah	37	kecoa

5	terjangkau	5	berisik	38	available	38	Kelupas
6	menarik	6	berkarat	39	Baik	39	kotor
7	mewah	7	berpencar	40	menyenangkan	40	kropos
8	murah	8	bingung	41	nyaman	41	kuning
9	pusat	9	bising	42	responsif	42	kusam
10	respon	10	bocor	43	modern	43	laba
11	mengesankan	11	buruk	44	pengertian	44	lama
12	membantu	12	debu	45	strategis	45	lambat
13	memadai	13	dibenahi	46	suka	46	lembab
14	senyum	14	ditingkatkan	47	sesuai	47	macet
15	dekat	15	gangguan	48	kedap	48	masalah
16	bersahabat	16	gelap	49	rapih	49	mati
17	cukup	17	goyang	50	sebanding	50	menunggu
18	tanggap	18	jelek	51	hangat	51	merembes
19	teliti	19	jorok	52	fungsi	52	noda
20	putih	20	jutek	53	besar	53	panjang
21	bersih	21	pengap	54	berdebu	54	usang
22	dapat	22	penuh	55	aman		
23	variasi	23	putus	56	asik		
24	mendukung	24	rendah	57	baru		
25	enak	25	rusak	58	cekatan		
26	cepat	26	sempit	59	ekonamis		
27	komplrit	27	susah	60	elegan		
28	luas	28	terbatas	61	lancar		
29	didukung	29	tersendat	62	leluasa		
30	representatif	30	tersumbat	63	lezat		
31	bagus	31	tinggi	64	memanjakan		
32	tengah	32	tua	65	memuaskan		
33	siap	33	usam	66	rendah		

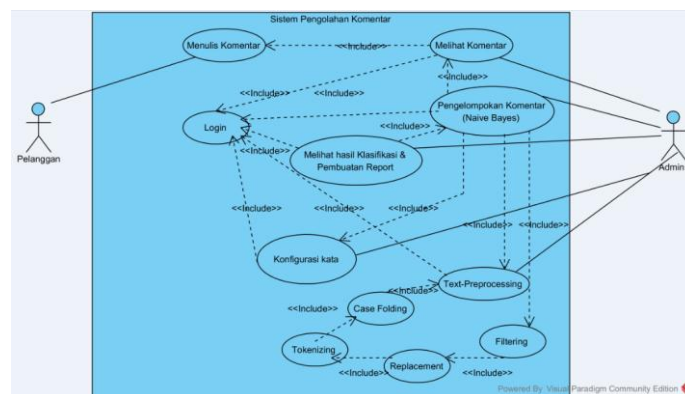
3.2 Perancangan Sistem

Gambar 1 ERD (*Entity Relationship Diagram*) sistem analisis sentimen yang dikembangkan terdiri dari 5 entitas yaitu: Komentar (atribut terdiri dari: id_komentar, nama_pengirim, tanggal_waktu, isi_komentar), Text Pre-processing (atribut terdiri dari: id_filter, case_folding, tokenize, replacement, filtering, kata_filter), Replacement (atribut terdiri dari: id_replacement, kata_tidak_baku, kata_baku), Kamus Kata (atribut terdiri dari: id_kamus, keyword, kata_sifat, kata_negasi), Kata Filter (atribut terdiri dari: id_filter, kata_filter)



Gambar 1 ERD Sistem

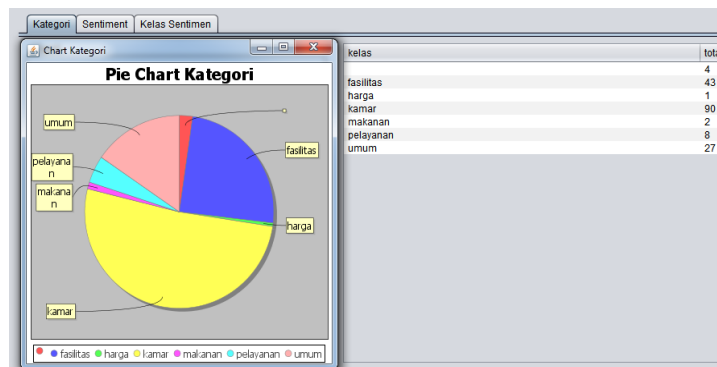
Use Case Diagram dari sistem yang dikembangkan pada gambar 2 terdiri dari melihat komentar, konfigurasi kata, pengelompokan komentar yang termasuk kedalam *text pre-processing*, melihat hasil klasifikasi, dan pembuatan *report*.



Gambar 2 Use Diagram Sistem

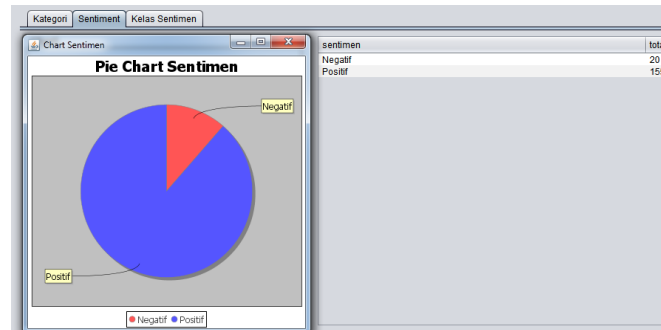
4. Implementasi

Implementasi dari perancangan yang sudah dibuat dengan bahasa pemrograman Java dan database MySQL. Gambar 3 merupakan implementasi antar muka hasil klasifikasi kategori.



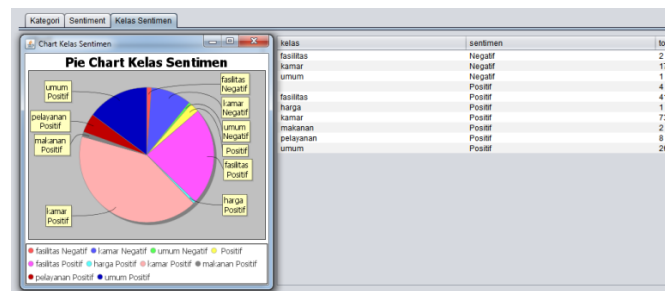
Gambar 3 Hasil Klasifikasi Kategori

Gambar 4 merupakan implementasi antar muka hasil klasifikasi sentimen.



Gambar 4 Hasil Klasifikasi Sentimen

Gambar 5 merupakan implementasi antar muka hasil klasifikasi kategori sentimen.



Gambar 5 Hasil Klasifikasi Kategori Sentimen

Seluruh fungsi-fungsi yang ada pada sistem analisis sentimen ini telah melalui proses pengujian. Berdasarkan pengujian tersebut, diketahui bahwa setiap fungsi pada sistem dapat memproses data serta memberikan respon yang sesuai dengan kasus yang ada pada pengujian seperti data input yang benar, salah atau tidak lengkap. Kemampuan sistem dalam melakukan analisis dengan mengelola komentar dengan menghasilkan *accuracy* kategori, dan untuk sentimen *precision*, *recall*, dan *accuracy*.

5 Kesimpulan

Berdasarkan penelitian tugas akhir yang telah dilakukan, maka didapat hasil penelitian sebagai berikut:

1. Dari hasil analisis yang dilakukan didapat 6 kategori hasil tinjauan dari data latih, yaitu kamar, makanan, pelayanan, fasilitas, harga, dan umum. Dari kategori tersebut digolongkan kembali menjadi 55 keyword.
2. Hasil dari perhitungan sistem analisis sentimen dengan metode NBC ini dari 175 data latih maka didapat 3 hasil klasifikasi sebagai berikut:
 - a. Klasifikasi kategori yang tertinggi adalah kamar sebesar 90 komentar dan yang terendah adalah untuk kategori harga sebesar 1 komentar

- b. Klasifikasi sentimen positif sebesar 155 komentar dan sentimen negatif sebesar 20 komentar
 - c. Klasifikasi kategori dengan sentimen positif tertinggi adalah kategori kamar sebesar 73 komentar dan untuk kategori dengan sentimen negatif terendah adalah kategori kamar sebesar 17 komentar.
3. Perhitungan kehandalan sistem analisis sentimen dengan metode NBC ini dari 175 data latih dapat dibagi menjadi dua bagian, perhitungan klasifikasi kategori dan perhitungan klasifikasi sentimen, untuk hasil dari akurasi sistem memiliki nilai *accuration* kategori 77.14% dan untuk *precision* sentimen 99.12%, *recall* sentimen 72.9%, dan *accuration* sentimen 75.42%, sehingga sistem mampu mengembalikan dokumen dan kecocokan data yang tinggi.

DAFTAR PUSTAKA

- [1] Bing, Liu. 2012. *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers.
- [2] Jurafsky, Dan. 2012. *Natural Language Processing*. Stanford University.
- [3] Lamurias, Andre Francisco.M. 2014. *Identifying Interactions Between Chemical Entities in Text*. Desertation Universidade De Lisboa.
- [4] Ting, S.L, W.H. Ip, Albert H.C. Tsang. 2011. *Is Naive Bayes a Good Classifier for Document Classification?*. International Journal of Software Engineering and Its Applications.
- [5] Barber, I. 2010. Bayesian Opinion Mining. [Online]. Tersedia di: <http://phpir.com/bayesian-opinion-mining> [diunduh: 15 Juni 2014].