Analisis Sentimen Komentar Masyarakat Terhadap Tempat Digital Printing Menggunakan Metode Naive Bayes

Grafi Maulana[1], Budanis Dwi Meilani[1]
Jurusan Sistem Informasi, Institut Adhi Tama Surabaya[1,2]
e-mail: grafimaulana@gmail.com

ABSTRACT

Sentiment Analysis refers to a classification process of ideas based on certain categories. Basically, it can be employed as a technique for determining something in a certain topic of problem as it serves for investigating the polarity of positive and negative sentiments. The data for sentiment classification can be collected from several comments in the social media and Google review such as from the data of comments in Google Maps. Next, the resulted data were labeled manually and processed by preprocessing method to ease the process of sentiment classification. The data would then have sentiment classification using Naive bayes method to notice the yielded sentiments. After conducting 4 trials within training data 20, 70, 140, 540 and testing data 60, the accuracies gained as follows: the first 65%, second 82%, third 85%, and fourth 88%.

Keywords*: sentiment classification, mining text, Naive Bayes Classifier, Digital Printing*

ABSTRAK

Klasifikasi sentimen merupakan pengelompokan suatu pendapat yang telah ditentukan berdasarkan masing masing kategori. Klasifikasi sentimen bisa dijadikan sebuah cara untuk proses menentukan sesuatu dalam sebuah topik permasalahan. Fungsi dari klasifikasi sentimen adalah untuk mencari tahu polaritas dari sentimen positif, dan negatif. Data klasifikasi sentimen bisa diperoleh melalui beberapa komentar pada media sosial serta pada ulasan Google, salah satunya yaitu data komentar yang ada pada Google Maps. Kemudian data yang sudah diperoleh akan diberi label secara manual dan diolah menggunakan metode preprocessing dengan tujuan untuk mempermudah proses klasifikasi sentimen. Data selanjutnya akan dilakukan klasifikasi sentimen menggunakan metode Naive bayes untuk mengetahui sentimen yang dihasilkan.Uji coba dilakukan sebanyak 4 kali dengan data latih 20, 70, 140, 540 dan data uji 60. Hasil pertama menunjukan akurasi sebanyak 65%, kedua 82%, ketiga 85%, dan keempat 88%

Kata kunci: Klasifikasi Sentimen, Teks Mining, Naive Bayes Classifier, Digital Printing

PENDAHULUAN

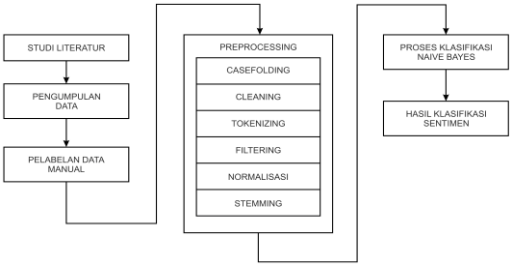
Pada saat ini usaha percetakan atau *digital printing* semakin berkembang pesat dari usaha kecil kecilan sampai toko yang memiliki peralatan lengkap. Percetakan merupakan tempat usaha maupun jasa bergerak dibidang cetak dan desain memiliki peralatan canggih yang mampu menghasilkan barang dengan media bergambar. Banyak ulasan komentar bermunculan pada *Google Maps* mengenai *digital printing* baik itu ulasan baik dan buruk. Komentar yang bermunculan tersebut akan dijadikan bahan evaluasi dan penilaian mengenai *digital printing*. Sentiment *analiysist* atau analisis sentimen dalam bahasa Indonesia adalah sebuah metode cara yang diterapkan untuk mengidentifikasi ekspresi dari sentimen menggunakan teks dan pengkatagorikan sebagai sentimen positif maupun sentimen negatif. Metode yang digunakan dalam penelitian ini yaitu metode *naive bayes.* Metode *naive bayes* bertujuan untuk menghasilkan nilai probabilitas dan peluang kemunculan hasil klasifikasi pada suatu teks komentar. Penelitian analisis sentiment sudah banyak dilakukan diantaranya Classification of community opinion on the use of the Transjakarta bus based on twitter social network using naïve bayes method[1], dan penelitian yang berhubungan dengan metode Naïve Bayes sudah banyak dilakukan diantaranya adalah Aplikasi Data Mining untuk menghasilkan Pola Kelulusan Siswa dengan Metode Naïve Bayes[2], Klasifikasi Identifikasi Faktor Penyebab Ketidaktepatan Masa Lulus Mahasiswa dengan Metode Naïve Bayes Classifier[3].

TINJAUAN PUSTAKA

Text mining adalah proses dalam mengambil informasi pada teks, informasi yang diperoleh adalah pola serta kecenderungan pembelajaran dalam pola statistik [4]. Menganalisa *text mining* akan dilakukan ketika teks yang akan digunakan dikumpulkan dan disiapkan dulu, selanjutnya diolah pada proses utama. Proses menyiapkan teks disebut juga dengan proses *preprocessing*. Text *preprocessing* berguna untuk mengganti data teks yang tidak tersusun menjadi tersusun rapi[5]. *Preprocessing* diterapkan untuk menghindari data yang kurang sempurna. Tahapan untuk melakukan *preprocessing* yaitu : *casefolding, cleaning, tokenizing, filtering,* normalisasi, *stemming.* Dengan menggunakan *naive bayes classifier* sebagai metode klasifikasi menggunakan teorema *bayes* untuk menghitung probabilitas kategori label berdasarkan dokumen yang telah diketahui [6]. Pengukuran kinerja atau evaluasi dari sistem / metode klasifikasi dilakukan dengan menghitung nilai akurasi, *recall* dan Presisi dengan menggunakan *confusion matrix.* Akurasi merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data. Semakin tinggi nilai akurasi yang diperoleh, hasil klasifikasi semakin baik.[7]

METODE

Berikut alur dan tahapan metodologi penelitian pada gambar 1 dibawah ini :



Gambar 1. Diagram Flowchart

Proses pelabelan yaitu sebuah tahapan untuk memberi label atau kategori pada data komentar yang sudah dikumpulkan.

Tabel 1. Pembagian data uji dan data latih

Kelas atau kategori	Data latih	Data uji	Total
	540	60	600
Positif	270	30	300
Negatif	270	30	300

Preprocessing Teks adalah Proses untuk mengurangi atribut yang kurang relevan pada suatu teks terhadap proses klasifikasi. proses ini bertujuan untuk menata dan mengolah dengan baik sebuah teks yang belum sempurna, sehingga hasil dari proses ini adalah dokumen yang baik. Dan bisa mempermudah proses klasifikasi. Berikut tahapan preprocessing pada tabel 2:

Tabel 2. Contoh proses *Preprocessing*

Kelas	Komentar	Casefolding	Cleaning	Tokenizing	Normalisasi	Filtering	Stemming
Positif	Cepat dan tepat	cepat dan tepat	cepat dan tepat	‘cepat’, ‘dan’, ‘tepat’	‘cepat’, ‘dan’, ‘tepat’	‘cepat’	‘cepat’
Negatif	Percetakan paling lemot sedunia	percetakan paling lemot sedunia	percetakan paling lemot sedunia	‘percetakan’, ’paling’, ‘lemot’, ‘sedunia’	percetakan’, ’paling’, ‘lambat, ‘sedunia’	percetakan’, ‘lambat, ‘sedunia’	‘cetak’, ‘lambat’, ‘sedunia’

Pada proses klasifikasi *naive bayes* dengan menganalisa data latih yang didapatkan dengan dari hasil observasi. Berikut merupakan contoh tahapan perhitungan data training yang terdapat pada tabel 3.

Tabel 3. Contoh hasil probabilitas kata

kosakata	nk		n		jumlah kosakata	nilai probabilitas kata	
	positif	negatif	positif	negatif		positif	negatif
cepat	1	0	6	13	16	0,0909091	0,034482759
kualitas	1	0	6	13	16	0,0909091	0,034482759
oke	1	0	6	13	16	0,0909091	0,034482759
harga	2	0	6	13	16	0,1363636	0,034482759
murah	1	0	6	13	16	0,0909091	0,034482759
cetak	0	1	6	13	16	0,0454545	0,068965517
lambat	0	1	6	13	16	0,0454545	0,068965517
dunia	0	1	6	13	16	0,0454545	0,068965517
murni	0	1	6	13	16	0,0454545	0,068965517
tidak	0	2	6	13	16	0,0454545	0,103448276
order	0	1	6	13	16	0,0454545	0,068965517
pegawai	0	2	6	13	16	0,0454545	0,103448276
lamban	0	1	6	13	16	0,0454545	0,068965517
layan	0	1	6	13	16	0,0454545	0,068965517
buruk	0	1	6	13	16	0,0454545	0,068965517
ramah	0	1	6	13	16	0,0454545	0,068965517

Setelah mengetahui seluruh nilai pada setiap kata, selanjutnya yaitu menghitung probabilitas kata menggunakan rumus

$$P(Ci|Vj) = (nk + 1)/(N + |kosakata|) \quad \text{..... (1)}$$

Contoh perhitungan :
P(cepat | positif) = (nk(cepat | positif) +1)/(N (positif) +|kosakata|)
P(cepat | positif) = (1+1)/(6+16)
P(cepat | positif) = 2/22
P(cepat | positif) = 0,0909091

P(cepat | negatif) = (nk(cepat | negatif) +1)/(N (negatif) +|kosakata|)
P(cepat | negatif) = (0+1)/(13+16)
P(cepat | negatif) = 1/29
P(cepat | negatif) = 0,034482759

Setelah melakukan proses *training* selanjutnya yaitu melakukan uji data dengan kategori positif 3 dan negatif 1. Untuk mengetahui apakah benar data uji ini termasuk dalam kategori yang diprediksi

Tabel 4. Contoh tabel data uji (label positif)

No	Data komentar	Kategori
1	pegawainya ramah,pelayanan baik. ruangan full ac	positif
2	Mantab,, harga cocok untuk reseller percetakan,, tapi kudu punya kesabaran ekstra ???	positif
3	Pelayanan lambat...	negatif
4	Pelayanan cepat dan professional	positif

Kemudian melakukan perhitungan pada nilai probabilitas berdasarkan nilai setiap kata. Hasil nilai hitung probabilitas akan muncul dengan mengalikan nilai probabolitas kategori dengan probabilitas setiap kata.

Sebelum melakukan proses perhitungan, data uji harus sudah dilakukan proses *preprocessing*. Berikut hasil proses *preprocessing* data uji

Tabel 5. Contoh hasil *preprocessing*

No	Data komentar	Kategori prediksi
1	'pegawai', 'ramah', 'layan', 'baik'	positif
2	'mantap', 'harga', 'cocok', 'reseller', 'cetak', 'ekstra'	positif
3	'layan', 'lambat'	negatif
4	'layan', 'cepat', 'profesional'	positif

Proses selanjutnya mencari nilai probabilitas yang ada pada data uji dengan menggunakan nilai probabilitas kata yang ada pada data latih.

Tabel 6. Contoh hasil probabilitas kata

		p (positif)	p (negatif)
Data uji 1	pegawai	0,0454545	0,103448276
	ramah	0,0454545	0,068965517
	layan	0,0454545	0,068965517
Data uji 2	harga	0,1363636	0,034482759
	cetak	0,0454545	0,068965517
Data uji 3	layan	0,0454545	0,068965517
	lambat	0,0454545	0,068965517
Data uji 4	layan	0,0454545	0,068965517
	cepat	0,0909091	0,034482759

Setelah mencari nilai probabilitas pada data uji kemudian. Hitung probabilitas kategori dengan persamaan :

$$P(Vj) = (N/jumlah) \text{.....(2)}$$

dokumen positif (3) negatif (3) diibagi jumlah seluruh dokumen data latih [6]
Setelah diketahui nilai p(kategori) selanjutnya tahap klasifikasi menggunakan persamaan :

$$P(Vj) \prod_{i=1} P(Ci|Vj) \text{.....(3)}$$

Data uji ke-1 kategori positif

P(uji | positif) = p(positif) * p(pegawai | positif) * p(ramah | positif) * p(layan | positif)
P(uji | positif) = 0,5 * 0,0454545 * 0,0454545 * 0,0454545
P(uji | positif) = 4,69572E-05

P(uji | negatif) = p(negatif) * p(pegawai | negatif) * p(ramah | negatif) * p(layan | negatif)
P(uji | negatif) = 0,5 * 0,103448276 * 0,068965517 * 0,068965517
P(uji | negatif) = 0,000246013

Kesimpulan pada data uji 1 :
Diketahui probabilitas tertinggi terdapat pada kategori negatif dengan memiliki nilai **0,000246013**

Data uji ke-2 kategori positif

P(uji | positif) = p(positif) * p(harga | positif) * p(cetak | positif)
P(uji | positif) = 0,5 * 0,1363636 * 0,0454545
P(uji | positif) = 0,003099174

P(uji | negatif) = p(negatif) * p(harga | negatif) * p(cetak | negatif)
P(uji | negatif) = 0,5 * 0,034482759 * 0,068965517
P(uji | negatif) = 0,001189061

Kesimpulan pada data uji 2 :
Diketahui probabilitas tertinggi terdapat pada kategori positif dengan memiliki nilai **0,003099174**

Data uji ke-3 kategori negatif

P(uji | positif) = p(positif) * p(layan | positif) * p(lambat | positif)
P(uji | positif) = 0,5 * 0,0454545 * 0,0454545
P(uji | positif) = 0,001033058

P(uji | negatif) = p(negatif) * p(layan | negatif) * p(lambat | negatif)
P(uji | negatif) = 0,5 * 0,068965517 * 0,068965517
P(uji | negatif) = 0,002378121

Kesimpulan pada data uji 3 :
Diketahui probabilitas tertinggi terdapat pada kategori negatif dengan memiliki nilai **0,002378121**

Data uji ke-4 kategori positif

P(uji | positif) = p(positif) * p(layan | positif) * p(lambat | positif)
P(uji | positif) = 0,5 * 0,0454545 * 0,0909091
P(uji | positif) = 0,045454545

P(uji | negatif) = p(negatif) * p(layan | negatif) * p(lambat | negatif)
P(uji | negatif) = 0,5 * 0,068965517 * 0,034482759
P(uji | negatif) = 0,017241379

HASIL DAN PEMBAHASAN

Pada studi kasus penelitian analisis sentimen komentar masyarakat terhadap tempat digital printing menggunakan metode naive bayes telah dilakukan percobaan menggunakan 4 data latih dan 1 data uji dengan pengujian yang pertama mendapatkan hasil akurasi sebanyak 65%, pengujian yang kedua mendapat hasil 82%, pengujian yang ketiga mendapat hasil 85%, pengujian yang keempat mendapat hasil 88%. Untuk rincian data dapat pada tabel dibawah ini.

Tabel 1. Hasil 4 kali pengujian

Data latih	Data uji	Akurasi		Presisi		Recall	
		Positif	Negatif	Positif	Negatif	Positif	Negatif
20	60	65%	65%	53%	77%	70%	62%
70	60	82%	82%	80%	83%	83%	81%
140	60	85%	85%	83%	87%	86%	84%
540	60	88%	88%	87%	90%	90%	87%

KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan pada Analisis Sentimen Komentar Masyarakat Terhadap Tempat Digital Printing Menggunakan Metode Naive Bayes maka penulis mendapat kesimpulan

yaitu data yang dikumpulkan menggunakan teknik *scraping* lebih cepat dibandingkan dengan cara manual, klasifikasi sentimen komentar opini para pelanggan digital printing menggunakan *text mining* sebagai proses pengolahan data komentar kemudian diklasifikasi menggunakan metode *naive bayes classifier* (NBC) dengan menghasilkan nilai akurasi yang berbeda-beda menggunakan data komentar 600 dan dibagi menjadi 540 data latih serta 60 data uji. Kemudian dilakukan pengujian kepada data latih sebanyak 4 kali pengujian yaitu 20, 70, 140, 540 data latih. Dari ke 4 data latih tersebut nilai akurasi yang paling tinggi diperoleh pada data latih sebanyak 540 dengan nilai akurasi 88%. Dapat disimpulkan bahwa metode *naive bayes classification* ini merupakan metode klasifikasi yang menguji data dengan mengambil nilai porbabilitas dari data uji kemudian dilakukan penjumlahan dan perbandingan sehingga dapat diketahui hasil nilai yang paling tinggi melalui perbandingan nilai. Semakin banyak data latih yang akan di klasifikasi semakin baik hasil prediksi dari data uji.

DAFTAR PUSTAKA

[1]. BD. Meilani, RK Hapsari, IF Novian 2021, Classification of community opinion on the use of the Transjakarta bus based on twitter social network using naïve bayes method, IOP Conference Series: Materials Science and Engineering 1010.

[2]. BD. Meilani, N. Susanti, 2016, Aplikasi Data Mining untuk menghasilkan Pola Kelulusan Siswa dengan Metode Naïve Bayes, Network Engineering Research Operation 1.

[3] BD. Meilani, S. Wahyudiana, AYP Putri, A. Pakarbudi, 2019, Klasifikasi Identifikasi Faktor Penyebab Ketidaktepatan Masa Lulus Mahasiswa dengan Metode Naïve Bayes Classifier, Prosiding Seminar Nasional Sains dan Terapan 1.

[4] Agatha, D., Kusrini, Emha, T.L. (2019). Analisis Pembobotan Kata Pada Klasifikasi Text Mining. Jurnal Teknologi Informasi. (hal 179-184).

[5] C. Sutami, 2015, Perbandingan Metode Klasifikasi Naïve Bayes Classifier dan Lexicon Based, Universitas Widyatama Bandung.

[6] Gilang Jalu Selo, Budi Susanto, 2013, Implementasi Naïve Baysian Classifier untuk kasus Filtering SMS Spam, Jurnal Informatika vol 9.

[7] Arthana, R. (5 April 2019). Mengenal Accuracy Accuracy, Precision, Recall dan Specificity serta yang diprioritaskan dalam Machine Learning. Pada 22 Februari 2021