

KLASIFIKASI OPINI GREEN AND CLEAN KABUPATEN LAMONGAN MENGGUNAKAN ALGORITMA MULTINOMIAL NAIVE BAYES

Agus Setia Budi

*Dosen Program Studi Teknik Informatika, Fakultas Teknik, Universitas Islam Lamongan
 Jl. Veteran No. 53 A Lamongan
 Telp. (0322)324706
 Email: geniusbudi@yahoo.com*

ABSTRACT

This study aims to process and classify an opinion (Opinion mining), opinion is a subjective statement that reflects public sentiment or perception of the entity or activity. Most opinions has not been managed well, if The Opinions properly managed will provide important information can be used to make improvements toward better at an activity or program. This study focuses on the processing of opinions that come from public opinion In Lamongan against LGC program which includes cleanliness, green and financial. The study was divided into two phases, namely the training process to produce data (dataset) to perform the classification process and the subjective (datates). Both processes are aimed to extract attributes and object components that have been commented upon in any document and to determine whether positive or negative comments. The results of the subjective test classification using Multinomial Naive Bayes algorithm has a success rate above 80% classification accuracy when it is matched with the manual classification.

Key Words: *Opinion, Opinion Mining, Multinomial Naive Bayes*

PENDAHULUAN

Pada saat ini pengklasifikasian dilakukan secara manual oleh staf BLH, namun dengan semakin banyaknya opini yang harus diproses maka waktu yang dibutuhkan akan semakin lama dan tingkat akurasi juga semakin berkurang dikarenakan keterbatasan kemampuan manusia dalam memproses data. Ide untuk membuat aplikasi opini secara elektronikpun sudah dilakukan yakni aplikasi opini, namun ide ini masih belum efektif dan belum akurat dikarenakan masyarakat terkadang belum mengarah ke tujuan yang sesuai dengan opini yang disampaikan. Dari permasalahan diatas maka disiplin ilmu teks mining menjadi tolak metode penting yang dimanfaatkan pada penelitian ini. Teks mining memiliki definisi menambang data yang berupa teks dimana sumber data dapat mewakili isi dari dokumen. Metode yang digunakan dalam pengklasifikasian ini adalah Multinomial Naive Bayes.

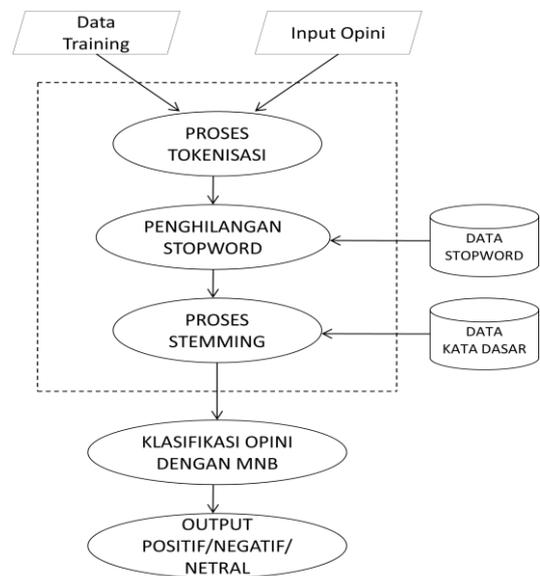
Langkah yang akan di ambil pada penelitian nantinya berawal dari memasukkan informasi berupa data opini dari sistem opini kemudian dilakukan preprocessing terhadap kumpulan opini tersebut, kemudian setelah preprocessing tersebut dilakukan, selanjutnya dilakukan proses pembobotan term dengan algoritma Nazied and andriani, hasilnya akan menjadi bahan training dan pada akhirnya akan dilakukan perhitungan dengan Multinomial Naive Bayes untuk penentuan klasifikasi.

Program LGC di Lamongan telah berjalan selama 6 tahun. Selama ini dalam mengevaluasi tujuan program LGC hanya berdasarkan pada sudut pandang pimpinan daerah, kepala BLH dan karyawan tanpa melibatkan komponen yang sebenarnya merupakan objek dari program LGC itu sendiri yaitu warga atau masyarakat.

METODE PENELITIAN

Proses Klasifikasi Data Opini adalah proses klasifikasi data yang diambil dari data opini. Bisa

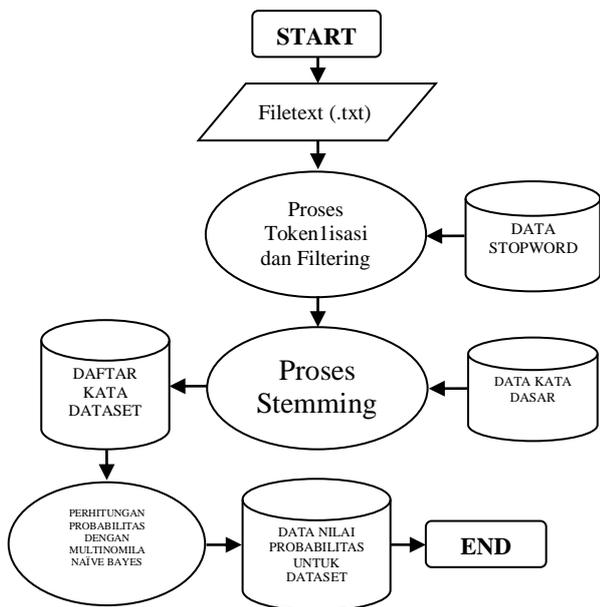
dilihat pada flowchart sistem untuk proses data uji coba opini, perhatikan gambar 1 di bawah ini.



Gambar 1. Flowchat Data Training dan Klasifikasi Opini

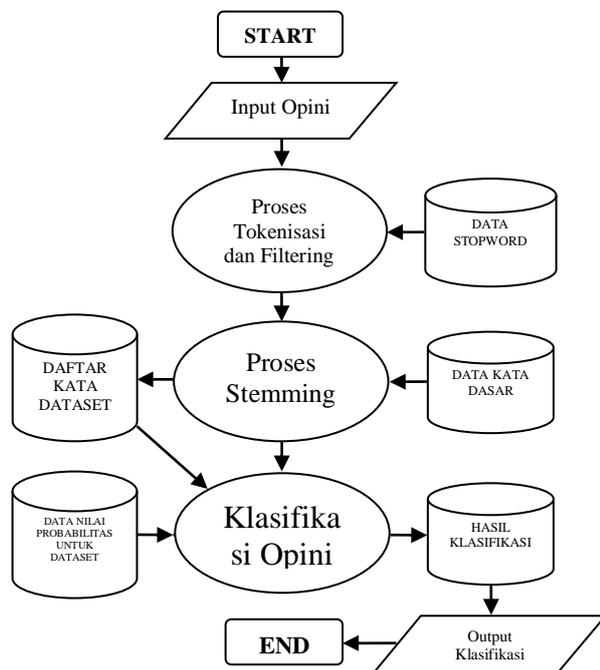
Pada metode penelitian Metode Multinomial Naive Bayes diharapkan dapat membantu penanggung jawab program Badan Lingkungan Hidup (BLH) dan Bupati Lamongan untuk mengetahui opini masyarakat terhadap program LGC, dengan tingkat akurasi ketepatan klasifikasi 80% bila dicocokkan dengan hasil klasifikasi manual.

- Arsitektur sistem dataset atau data training



Gambar 2. Arsitektur Sistem Dataset Untuk Klasifikasi Opini

- Arsitektur sistem datates



Gambar 3. Arsitektur Sistem Datates Untuk Klasifikasi Opini

HASIL DAN PEMBAHASAN

Desain interface merupakan rancangan desain antar muka antara sistem dengan pengguna, desain ini meliputi :

- Desain interface untuk entri data data training secara manual. desain ini digunakan untuk memasukkan data training dengan cara mengetikkan data kuisoner secara langsung tidak melalui file text. Untuk lebih jelasnya bisa lihat pada gambar 4. di bawah ini.



Gambar 4. Desain Interface Entri Data Training Secara Manual

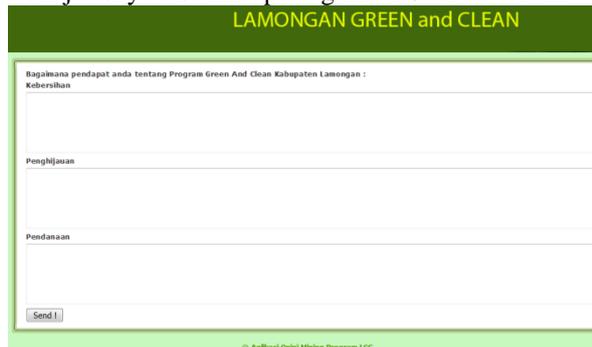
Proses tokenisasi, penghapusan stopwords dan stemming. Data training harus diinputkan kedalam system. Jika file kosong maka ada umpan balik kepada user bahwa data kosong kalau sudah benar memulai teks sama dengan teks dan akan menjadikan huruf kecil dengan karakter sama dengan pemisah antar kata dengan symbol, lalu di proses token, proses penghapusan stopwords dan selanjutnya tampilkan hasil dari proses token dan removal stopwords.

Desain interface untuk entri data data training secara manual. desain ini digunakan untuk memasukkan data training dengan cara mengetikkan data kuisoner secara langsung tidak melalui file text. Untuk lebih jelasnya bisa lihat pada gambar 5. di bawah ini.



Gambar 5. Desain Interface Entri Data Training Secara Manual

Desain interface untuk input opini. Untuk lebih jelasnya bisa lihat pada gambar 6. di bawah ini.



Gambar 6. Desain Untuk Input Opini

Algoritma entri data untuk data training , proses untuk menghasilkan data training file tidak boleh kosong, jika file masih kosong ada kata data masih kosong. Disini kalau sudah benar dan sudah tidak kosong. Mulai memanggil fungsi token untuk memisah kata lalu memanggil fungsi remove stopwords untuk menghapus stopwords atau kata sambung lalu mencari kata dasar sebuah kata proses stemming.

Berikut ini diberikan beberapa sebuah contoh kalimat opini yang sebelumnya tidak terdapat kata hedge, kemudian dibandingkan dengan kata yang didalamnya terdapat kata hedge.

Contoh Inputan Opini yang Bernilai Positif

kebersihan lingkungan warga terjaga dengan baik

penghijauan di lamongan sangat baik
biaya sangat murah sekali

Maka hasil yang diperoleh pada inputan opini yang bernilai positif, bisa dilihat pada gambar 7. di bawah.

kebersihan lingkungan warga terjaga dengan baik	Bersih Positif
penghijauan di lamongan sangat baik	Penghijauan Positif
biaya sangat murah sekali	Pendanaan Positif

Gambar 7. Contoh Hasil Klasifikasi Opini yang Bernilai Positif

Untuk melakukan verifikasi hasil percobaan, dilakukan percobaan sebanyak 10 kali, dataset dari percobaan ini di bagi menjadi dua yakni data latih dan data testing, penentuan data latih dan data testing ini dilakukan secara acak dan merata agar tidak terjadi pengelompokan dokumen-dokumen yang berasal satu kategori tertentu pada sebuah percobaan, data latih di ambil sebanyak 70% dari dataset secara acak dan merata untuk setiap kategori, sedangkan sisanya dipakai sebagai data testing, pembagian data ini akan dilakukan sebanyak 10 kali untuk sepuluh percobaan. Hasil confusion matrik uji coba dari percobaan nantinya di cari nilai akurasi rata-rata. Hasil dari confusion matrix uji coba terbaik ditunjukkan pada Gambar 8.

		HASIL KLASIFIKASI						
		Bersih Positif	Bersih Negatif	Penghijauan Positif	Penghijauan Negatif	Pendanaan Positif	Pendanaan Negatif	
KEMAMPUAN	Bersih Positif	8	2	0	0	0	0	
	Bersih Negatif	3	2	0	0	0	0	
	Penghijauan Positif	0	0	26	1	0	0	
	Penghijauan Negatif	0	0	0	0	0	0	
	Pendanaan Positif	0	0	0	0	14	0	
	Pendanaan Negatif	0	0	0	0	1	1	
	Akurasi Klasifikasi 82%							

		HASIL KLASIFIKASI						
		Bersih Positif	Bersih Negatif	Penghijauan Positif	Penghijauan Negatif	Pendanaan Positif	Pendanaan Negatif	
KEMAMPUAN	Bersih Positif	0,6	0,2	0	0	0	0	
	Bersih Negatif	0,6	0,4	0	0	0	0	
	Penghijauan Positif	0	0	0,94117647	0,05882352	0	0	
	Penghijauan Negatif	0	0	0	0	0	0	
	Pendanaan Positif	0	0	0	0	1	0	
	Pendanaan Negatif	0	0	0	0	0,5	0,5	
	Akurasi Klasifikasi 82%							

Gambar 8. Hasil Uji Coba Matrix

Proses Penentuan jenis data untuk data uji dilakukan dengan memilih jumlah data uji yang akan diklasifikasi, ditunjukkan pada gambar 9.

Gambar 9. Form Untuk Tes Klasifikasi Data Uji

Setelah proses segmentasi ada beberapa proses yang harus dilakukan sebelum melakukan ekstraksi ciri diantaranya proses delasi, erosi, skeletoning dan pelabelan, untuk hasil uji coba ekstraksi ciri menggunakan invarian moment dapat dilihat pada tabel 2. sedang untuk hasil uji coba ekstraksi ciri

menggunakan morfologi dapat dilihat pada tabel 3. berikut ini:

Dari beberapa uji coba terbaik maka bisa di buat tabel data cross validation dan bisa ditunjukkan pada hasil cross validation, pada tabel 2.

Tabel 2. Data Cross Validation Fold=10

Latih Ke	1	2	3	4	5	6	7	8	9	10
Dok Awal	1	51	101	151	201	251	301	351	401	451
Dok Akhir	50	100	150	200	250	300	350	400	450	500
Hasil Prosentase	82%	88%	84%	90%	88%	86%	82%	78%	72%	76%
									Rata2	83%

Pada tabel 2. ditunjukkan nilai akurasi terbesar terjadi pada percobaan ke 4 dengan nilai 90% sedangkan rata-rata akurasi yang terjadi adalah 83%.

Gambar 10. berikut ini merupakan hasil hasil perhitungan confusion matrix terhadap percobaan terbaik.

		HASIL KLASIFIKASI						
		Bersih Positif	Bersih Negatif	Penghijauan Positif	Penghijauan Negatif	Pendanaan Positif	Pendanaan Negatif	
KEMAMPUAN	Bersih Positif	16	0	0	0	0	0	
	Bersih Negatif	1	0	0	0	0	0	
	Penghijauan Positif	0	0	15	1	0	0	
	Penghijauan Negatif	0	0	1	0	0	0	
	Pendanaan Positif	0	0	0	0	13	0	
	Pendanaan Negatif	0	0	0	0	2	1	
	Confusion Matrix							

		HASIL KLASIFIKASI						
		Bersih Positif	Bersih Negatif	Penghijauan Positif	Penghijauan Negatif	Pendanaan Positif	Pendanaan Negatif	
KEMAMPUAN	Bersih Positif	1	0	0	0	0	0	
	Bersih Negatif	1	0	0	0	0	0	
	Penghijauan Positif	0	0	0,9375	0,0625	0	0	
	Penghijauan Negatif	0	0	1	0	0	0	
	Pendanaan Positif	0	0	0	0	1	0	
	Pendanaan Negatif	0	0	0	0	0,66666666667	0,333333333333	
	Akurasi Klasifikasi							
	Hasil = 90%							

Gambar 10. Hasil Confusion Matrix Uji Coba Terbaik

Dari confusion matrix uji coba terbaik di dapat hasil dan dilakukan perhitungan nilai akurasi sebagai berikut :

Untuk menghitung akurasi data digunakan rumus

$$\text{Akurasi} = \frac{\text{Jumlah klasifikasi secara benar}}{\text{Jumlah semua data klasifikasi}}$$

$$\text{Akurasi} = \frac{172}{200}$$

$$\text{Akurasi} = 0.83 \text{ atau } 83,00\%$$

Dari perhitungan nilai akurasi berdasarkan confusion matrix didapatkan akurasi klasifikasi rata-rata sebesar 83%. Dengan demikian akurasi klasifikasi opini dengan metode multinomial naive bayes memperoleh hasil yang bagus.

KESIMPULAN

Adapun beberapa kesimpulan yang didapatkan antara lain:

1. Data hasil percobaan yang dilakukan untuk mengklasifikasi data opini masyarakat menggunakan multinomial naive bayes yang diusulkan terbukti mampu melakukan klasifikasi dokumen opini masyarakat.

2. Dengan menggunakan metode klasifikasi teks, dari segi kecocokan jenis kelas yang dihasilkan klasifikasi terhadap penentuan jenis kelas yang dilakukan oleh petugas BLH (Badan Lingkungan Hidup) tergolong baik, presentasi kecocokan kelas terhadap 200 data uji rata-rata sebesar 83%.
3. Dengan menggunakan metode klasifikasi teks yaitu metode multinomial naïve bayes, dengan penentuan jenis kelas yang dilakukan oleh petugas BLH tergolong baik, Hasil klasifikasi bila dibanding dengan hipotesa yang diambil, maka sistem ini telah berjalan dengan baik. Terbukti dengan akurasi mencapai 83%
4. Dengan pengklasifikasian data opini masyarakat dengan metode Multinomial Naïve Bayes dapat memudahkan pimpinan dalam melakukan evaluasi program LGC di Lamongan.

REFERENSI

- [1] Alexander Pak, Patrick Paroubek “ Twitter as a Corpus for Sentiment Analysis and Opinion Mining “, Universit´e de Paris-Sud, Laboratoire LIMSI-CNRS, Bˆatiment 508
- [2] Bing Liu, “Opinion Mining”, Department of Computer Science University of Illinois at Chicago.
- [3] Destuardi, Surya Sumpeno (2009), “Klasifikasi Emosi Untuk Teks Bahasa Indonesia Menggunakan Metode Naïve Bayes”, Jurusan Teknik Elektro, Institut Teknologi Sepuluh Nopember Kampus ITS, Sukolilo, Surabaya 6011.
- [4] Erik Boiy; Pieter Hens; Koen Deschacht; Marie-Francine Moens, ”Automatic Sentiment Analysis in On-line Text, Katholieke Universiteit Leuven, Tiensestraat 41 B-3000 Leuven, Belgium
- [5] Jais Hardinal, Bayu Distiawan Trisedya, (2009), “Klasifikasi Dokumen Menggunakan Algoritma Naïve Bayes dengan Penambahan Parameter Probabilitas Parent Category”, Fakultas Ilmu Komputer Universitas Indonesia.
- [6] S.L. Ting; W.H. Ip; Albert H.C. Tsang “ *Is Naïve Bayes a Good Classifier for Document Classification ?* “ International Journal of Software Engineering and Its Application Vol.5, No. 3, University Hung Hum, Kowloon, Hongkong 2011