

IMPLEMENTASI SENTIMEN ANALISIS KOMENTAR CHANNEL VIDEO PELAYANAN PEMERINTAH DI YOUTUBE MENGGUNAKAN ALGORITMA NAÏVE BAYES

Pramana Yoga Saputra¹, Dian Hanifudin Subhi², Fahmi Zain Afif Winatama³

^{1,2,3}Teknik Informatika, Teknologi Informasi, Politeknik Negeri Malang
¹pramanay@gmail.com, ²dhanifudin@polinema.ac.id, ³winatamafza@gmail.com

Abstrak

Pelayanan instansi pemerintah menjadikan tolok ukur dalam menilai tingkat kepuasan masyarakat. Instansi pemerintah di Indonesia perlahan telah memanfaatkan media sosial sebagai sarana komunikasi dengan masyarakat. Karena media sosial mendapatkan tempat yang istimewa pada penggunaan internet di Indonesia. Terutama media sosial, khususnya untuk *Youtube* yang mana sekarang segala sesuatunya dapat diupload setiap saat dan berjuta komentar diunggah pada media berbagi video. Penelitian ini mencoba melakukan Analisa sentimen menggunakan klasifikasi data *Youtube* pada komentar video yang dipublish pemerintah tentang kinerjanya kepada masyarakat. Data tersebut nantinya akan diklasifikasikan menjadi klasifikasi positif, negatif dan netral. Komentar tersebut kemudian diproses menggunakan metode *Naive Bayes Classifier*. Hasil dari pengujian yang dilakukan memperoleh nilai akurasi sebesar untuk KemenPUPR 69.23% dan 64.10% untuk Kemenkeu.

Kata kunci : Pelayanan, Analisis Sentimen, *Naive Bayes Classifier*

1. Pendahuluan

Pelayanan terhadap masyarakat merupakan hal utama yang sangat mendasar bagi suatu instansi pemerintahan ataupun swasta. Sebagai pelayan dari masyarakat pemerintah disini mempunyai peran wajib serta tanggung jawab untuk memberikan pelayanan yang baik dan profesional. Pelayanan tersebut dibutuhkan masyarakat guna memenuhi berbagai macam kebutuhan yang mana hal itu tidak dapat dipisahkan dalam kehidupan manusia.

Pada penelitian terdahulu disebutkan bahwa media sosial digunakan oleh instansi pemerintah untuk menyampaikan dan meningkatkan pelayanan mereka terhadap masyarakat. Di sisi lain, penggunaan media sosial pada pemerintahan adalah untuk membangun sistem pemerintahan yang terbuka dan transparan dengan menampung aspirasi dari masyarakatnya. Di Indonesia instansi pemerintah baik di tingkat pusat maupun tingkat daerah secara perlahan sudah menggunakan media sosial sebagai sarana komunikasi publik, hal ini ditunjukkan dengan disertakannya akun media sosial pada situs web resmi instansi pemerintah. Hal tersebut juga didukung oleh pemerintah pusat di mana pada tahun 2012 telah diterbitkan pedoman pemanfaatan media sosial instansi pemerintah oleh Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Republik Indonesia. Tujuan dari diterbitkannya pedoman pemanfaatan media sosial instansi pemerintah oleh Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Republik Indonesia adalah menciptakan

keterbukaan, komunikasi yang efektif dan interaktif, serta saling menguntungkan antara instansi pemerintah dan pemangku kepentingan dalam penyelenggaraan humas pemerintah.

Dengan adanya tolok ukur tersebut, bisa dimanfaatkan media sosial sebagai obyek untuk dijadikan alat ukur tingkat kepuasan masyarakat tanpa adanya manipulasi data. Di dukung dengan era perkembangan teknologi terutama di Indonesia yang sangat tumbuh sangat pesat saat ini. Terutama media sosial, contohnya *Youtube* yang mana sekarang setiap individu dapat mengekspresikan pendapatnya melalui bentuk video yang dapat sewaktu-waktu dapat diunggah tanpa adanya batasan waktu. Hal ini sangat berguna dan dapat dijadikan studi kelayakan sebagai sebuah nilai tertentu. Terutama berdasarkan kolom Komentar yang disediakan, sehingga dapat dijadikan sebuah informasi Analisis Media Sosial yang diklasifikasikan opini positif maupun negatifnya.

Uraian diatas menunjukkan bahwa perlu dibangun "Analisa Komentar menggunakan algoritma *Naive Bayes Classifier* (NBC) sebagai Tolok Ukur Tingkat Kepuasan Masyarakat Terhadap Pelayanan Pemerintah melalui *Youtube*". Sehingga Pemerintah dalam hal ini sebagai Pelayanan dapat memonitoring penilaian masyarakat terhadap pemerintahan serta untuk sarana perbaikan pelayanan.

2. Tinjauan Pustaka

2.1 Text Mining

Text mining merupakan suatu penambahan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, sesuatu yang tidak diketahui sebelumnya untuk menemukan suatu informasi yang tersirat untuk diekstrak secara otomatis dari data teks yang berbeda-beda. Penerapan teknik *text mining* untuk memecahkan masalah bisnis disebut *text analytics*. *Text mining* hampir sama dengan data mining dalam hal tujuan dan proses, tapi pada *text mining* inputnya adalah file data tidak terstruktur seperti dokumen dalam bentuk word, PDF, text, XML dan sebagainya. *Text mining* dapat digunakan dalam beberapa hal yaitu ekstraksi informasi, *topic tracking*, *summarization*, kategorisasi dan *clustering*.

2.2 Analisis Sentimen

Sentimen analisis merupakan proses mengolah atau mengekstrak data secara tekstual dengan otomatis untuk mendapatkan informasi sentiment yang terkandung dalam suatu kalimat tersebut. Hal ini digunakan untuk mencari pendapat ataupun kecenderungan kalimat yang sifatnya opini terhadap suatu masalah atau obyek oleh seseorang yang sifatnya bisa berisikan opini positif dan negatif.

2.3 Naive Bayes Classifier

Naive Bayes Classifier (NBC) merupakan salah satu algoritma yang digunakan dalam klasifikasi data mining. Klasifikasi sendiri merupakan penentuan sebuah *record* data baru ke salah satu dari beberapa kategori (atau kelas) yang telah didefinisikan. Dan pada *Naive Bayes Classifier* ialah mengadopsi teorema *Bayesian*. Bayes merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan teorema bayes dengan asumsi independensi (ketidaktergantungan yang kuat (naif) [6]. Dengan kata lain, dalam *Naive Bayes* yang digunakan adalah “model fitur independen”.

$$P(C_j|W_i) = \frac{P(W_i|C_j)P(C_j)}{P(W_i)} \quad (1)$$

$P(C_j|W_i)$: Peluang kategori j , ketika terdapat kemunculan kata i

$P(W_i|C_j)$: Peluang kata i masuk ke dalam kategori j

$P(C_j)$: Peluang kemunculan kategori j

$P(W_i)$: Peluang kemunculan kata

Ada banyak cara untuk menghitung $P(W_i|C_j)$, cara paling sederhananya adalah dengan

$$P(W_i|C_j) = \frac{N_{ic}+1}{N_c+V} \quad (2)$$

N_{ic} : Jumlah dokumen latih dengan fitur atribut W_i dengan kategori C_j

N_c : Jumlah dokumen dari kategori C_j

V : Jumlah kategori

2.4 Youtube API

Application Programming Interface (API) sebagai fungsi, kumpulan perintah dan juga protokol yang berguna untuk membantu programmer dalam membangun perangkat lunak dalam sistem operasi tertentu. Khususnya dalam Youtube memungkinkan pengembang mengakses statistik video dan data saluran YouTube melalui dua jenis panggilan, REST, dan XML-RPC. Google sendiri mendeskripsikan Sumber Daya API YouTube sebagai API dan Alat yang memungkinkan Anda membawa pengalaman YouTube ke laman web, aplikasi, atau perangkat Anda. Pada dasarnya fungsi serta perintah pada Youtube API merupakan alat pemanggil *system calls* yang mana berhubungan langsung ke sistem operasinya.

2.5 Crawling

Crawling berasal dari kata “*crawler*” itu sendiri ialah mengambil kumpulan data dimana kemudian diolah sesuai dengan kebutuhan dari *user* tersebut. Fungsi secara detailnya merupakan aplikasi *script* program kemudian melakukan *scanning* terhadap halaman internet yang akan diambil datanya.

2.6 Pelayanan Publik

Pada KEPMENPAN Nomor 63 tahun 2004 tersebut, juga telah mengelompokkan tiga jenis pelayanan dari instansi pemerintah serta BUMN/BUMD. Berikut pengelompokan jenis pelayanan yang didasarkan atas ciri-ciri dan sifat kegiatan pelayanan yang dihasilkan :

1. Pelayanan Administratif, merupakan jenis pelayanan berupa pencatatan, penelitian, pengambilan keputusan, dokumentasi, serta kegiatan tata usaha lainnya. Hasil kegiatan yang dilakukan berupa pengurusan sertifikat, ijin-ijin, dan sebagainya
2. Pelayanan Barang, merupakan pelayanan yang berupa penyediaan atau pengolahan barang berbentuk fisik. Hasil kegiatan yang dilakukan merupakan bentuk pelayanan listrik, air bersih, pelayanan infrastruktur dan lain lain.
3. Pelayanan Jasa, jenis pelayanan yang berupa sarana dan prasarana beserta penunjangnya. Bentuk dari hasil kegiatannya ialah pelayanan perbankan, pelayanan pembayaran pajak, pelayanan pengurusan bea cukai, pelayanan pos.

3 Desain dan Metodologi

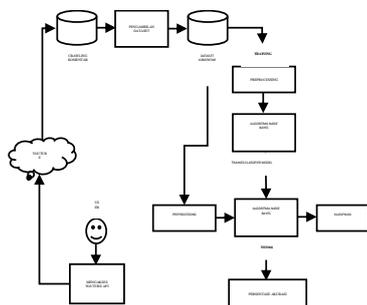
3.1 Analisis

Bagian ini menunjukkan design model beserta struktur implementasi sistem berdasarkan kasus yang dialami.

3.1.1 Deskripsi Umum

Pada penelitian ini secara umum prosesnya meliputi beberapa tahapan yaitu pembuatan proposal, pengumpulan data, pengolahan data, implementasi algoritma *Naive Bayes Classifier* (NBC), pengujian, hingga analisis hasil. Pada tahap pengolahan ada beberapa kegiatan sesuai dengan tahapan yang ada pada data training yang disebut tahap *preprocessing*, yaitu pembersihan data, integrasi data, seleksi data, transformasi data, dan pembentukan *dataset* yang dalam penelitian akan digunakan sebagai data *training* dan data *testing*.

3.1.2 Desain Sistem

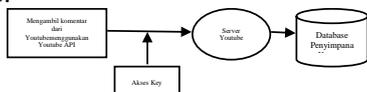


Gambar 1. Alur Desain Sistem

Analisis Sentimen Komentar memiliki beberapa metode yang dapat digunakan. Pada penelitian ini akan menggunakan metode algoritma *Naive Bayes Classifier* dalam melakukan Analisis Sentimen pada Komentar. Diagram *work flow* diatas merupakan alur dari penelitian yang akan dilakukan.

3.2 Pengumpulan Data

Pada pengambilan data penelitian ini diawali dengan penarikan data komentar pada *channel* video pelayanan pemerintah pusat pada Youtube, kemudian disimpan dalam *database*. Proses pengambilan data tersebut dilakukan menggunakan fasilitas *Application Programming Interface* (API). API sendiri merupakan fasilitas yang telah disediakan oleh Youtube. *Application Programming Interface* (API) sendiri mengambil data komentar secara mentah dari *server* Youtube. Untuk dapat mengakses Youtube API tersebut tidak dapat begitu saja menggunakannya, sebagai pengembang diwajibkan untuk registrasi terlebih dahulu agar mendapatkan *Key* Hak Akses, dimana *Key* didapatkan setelah mendaftarkan ke *Developer* resmi Youtube.

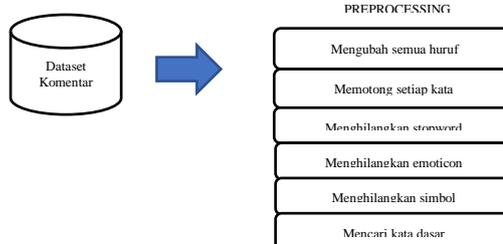


Gambar 2. Proses Pengambilan Data

3.3 Implementasi Sistem

3.3.1 Preprocessing

Merupakan proses dari pengolahan data komentar yang telah diambil menggunakan tahapan *preprocessing*. *Preprocessing* ini dilakukan untuk menghindari data yang kurang sempurna, gangguan pada data, dan data-data yang tidak konsisten.



Gambar 8. Tahapan Pengolahan Data

3.3.2 Term Frequency

Setelah melakukan *preprocessing*, tahapan selanjutnya yang dilakukan ialah menghitung kata tersebut atau *Term Frequency* yang mana digunakan untuk melanjutkan proses ke dalam perhitungan algoritma *Naive Bayes*. Proses *TF* berdasarkan hasil perhitungan beberapa kata dalam komentar sebagai contoh berikut :

Tabel 1. *Term Frequency*

	D1	D2	D3	D4	D5	D6
kondisi	0	1	0	0	0	0
rakyat	1	2	0	0	0	0
otak	1	0	0	0	0	0
kritis	1	0	0	0	0	0
memikirkan	1	0	0	0	0	0
dampak	1	0	0	0	0	0
manfaatnya	1	0	0	0	0	0
sekarang	0	0	0	1	0	0
ekonomi	0	0	0	2	0	0
pemerintah	0	0	0	0	0	0
Class	P	Ng	Nt	P	Ng	Nt

Keterangan :

- D : Dokumen
- P : Positif
- Ng : Negatif
- Nt : Netral

3.3.3 Algoritma Naive Bayes

Klasifikasi *NBC* sendiri terbagi atas 2 proses yaitu proses *training* dan proses *testing*. Proses *training* digunakan untuk model analisis sentimen yang bertujuan sebagai pedoman klasifikasi dengan data *testing* atau data yang berbeda.

1. Proses Training

$$P(\text{word } i | \text{class}) = \frac{T_{ct} + \lambda}{(N_c + \lambda V)} \quad (3)$$

Keterangan :

- Tct Term Frequency dari kata i dalam kategori C
- Nc Banyak kata keseluruhan dalam kategori C
- V Jumlah kata
- λ Positive constant, biasanya 1. Untuk menghindari nilai 0

2. Proses *Testing*

Pada proses testing dalam algoritma NBC menggunakan persamaan sebagai berikut :

$$Vmap = \frac{argmax}{V_{j \in V}} \prod_i i^n P(x_i, |V_j)P(V_j) \quad (4)$$

Keterangan :

- Vmap Semua kategori yang diujikan V
- j1 Class Positif
- j2 Class Negatif
- j3 Class Netral
- P(Xi|Vj) Probabilitas Xi pada kategori Vj
- P(Vj) Probabilitas dari Vj

3.4 Pengujian

Pengujian yang dilakukan untuk menguji sistem dilakukan dengan 2 langkah pengujian, yaitu pengujian fungsionalitas dan pengujian akurasi sistem. Pengujian bertujuan untuk memastikan bahwa sistem yang dibangun telah berjalan sesuai dengan algoritma *Naive Bayes* yang digunakan :

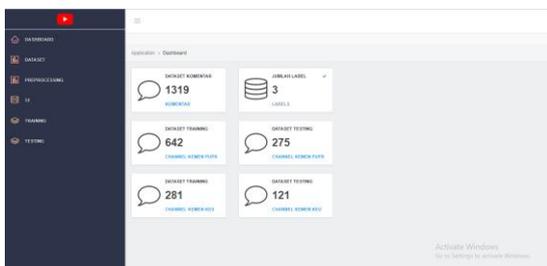
- a. Pengujian Fungsionalitas
- b. Pengujian Akurasi sistem

4 Implementasi dan Pengujian

4.1 Implementasi Database

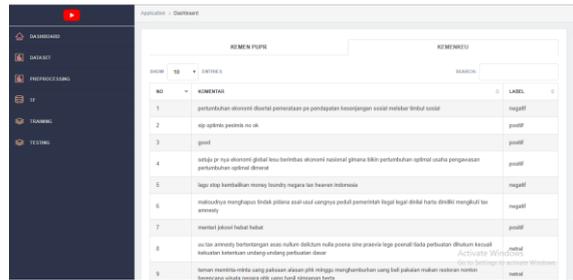
Implementasi *database* merupakan proses dari pengambilan komentar yang kemudian diinputkan ke dalam *database* untuk kemudian diolah. Ada 4 tabel dalam pengimplementasiannya yaitu, tabel *dataset*, tabel *dataset_baku*, tabel *pword* dan tabel *prob*.

4.2 Implementasi Tampilan



Gambar 4. Tampilan Beranda

Berikut ini merupakan bentuk dari tampilan beranda yang dimana terdapat menu *dashboard*, *preprocessing*, *tf*, *training* dan *testing*.



Gambar 5. Tampilan hasil *Preprocessing*

Yang mana merupakan hasil dari *dataset* yang telah di *preprocessing* menghasilkan data yang baku untuk dilakukan proses selanjutnya



Gambar 6. Proses Perhitungan Kata (*TF*)

Proses dimana menghitung kata berdasarkan komentar yang telah dilakukan *preprocessing*



Gambar 7. Proses *Testing*

Tahapan akhir yang dimana hasil dari *dataset* yang telah di *preprocessing* kemudian dilakukan perhitungan *term frequency* yang menghasilkan nilai akurasi dari program tersebut.

4.5 Pengujian

Hasil pengujian yang dilakukan terhadap penilaian Implementasi Sentimen Analisis pada Komentar *Channel Video Pelayanan Pemerintah di Youtube* Menggunakan *Naive Bayes*.

- a. Pengujian fungsional

Pengujian fungsional dilakukan dengan cara menjalankan setiap fitur aplikasi dan melihat kesesuaian hasil yang terjadi dengan hasil yang dicapai. Tabel 2. Menunjukkan bagaimana pengujian saat dilakukan apakah berjalan dengan sesuai atau tidak.

Tabel 2. Pengujian Fungsional

No.	Pola Pengujian	Hasil Pengujian	Status
1.	Load <i>Dataset</i>	<i>Dataset</i> dapat ditampilkan sesuai dengan yang terdapat dalam <i>database</i>	Sesuai
2.	<i>Preprocessing</i>	<i>Preprocessing</i> menampilkan hasil dari <i>preprocessing dataset</i>	Sesuai
3.	Perhitungan <i>TF</i>	<i>TF</i> menampilkan hasil dari perhitungan <i>dataset</i> yang telah melalui <i>preprocessing</i>	Sesuai
4.	<i>Training</i>	<i>Training</i> menampilkan data untuk proses <i>training</i>	Sesuai
5.	<i>Testing</i>	<i>Testing</i> menampilkan data untuk proses <i>testing</i>	Sesuai
6.	Pengujian	Pengujian menampilkan tingkat akurasi dari klasifikasi	Sesuai

b. Pengujian akurasi sistem

Pengujian akurasi sistem dilakukan dengan cara menghitung nilai dari *accuracy* dan *precision*. Rumus untuk menghitung nilai *accuracy* sebagai berikut.

$$\frac{\sum v}{n} * 100 \tag{5}$$

Keterangan

- v : Jumlah data benar
- n : Jumlah dokumen

Sedangkan untuk menghitung nilai *precision* menggunakan rumus sebagai berikut.

$$\frac{\sum vP}{nP} * 100 \tag{6}$$

Keterangan

- vP : Jumlah data positif, negatif atau netral benar
- nP : Jumlah dokumen positif, negatif atau netral

Pengujian dilakukan dengan jumlah data *training* yang berbeda. Tabel 3. akan menyajikan hasil pengujian akurasi sistem.

Tabel 3. Pengujian Akurasi Sistem

P	Acc	Prec 1	Prec -1	Prec 0
1	69.23	73.73	54.55	77.27
2	64.10	72.09	53.49	70,00

Keterangan :

- P1 : Pengujian KemenPUPR
- P2 : Pengujian KemenKeu
- Acc : Accuracy
- Prec 1 : Precision Positif
- Prec -1 : Precision Negatif
- Prec 0 : Precision Netral

5 Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan proses pengujian dan analisis yang telah dilakukan, dapat diambil kesimpulan sebagai berikut :

1. Hasil akurasi klasifikasi algoritma *Naive Bayes Classifier* pada *channel* KemenPUPR menghasilkan nilai 69.23% persen dan pada *channel* Kemenkeu menghasilkan nilai 64.10%.

5.2 Saran

Berdasarkan kesimpulan penelitian di atas. Saran yang dapat diberikan dari hasil penelitian untuk pengembangan sistem ini ke depannya bisa dilakukan dengan metode algoritma lain dengan kasus yang sama.

Daftar Pustaka:

Amonik, A. Jivane, N, Bhandari, M. Venkatesan, Dr. M. 2015. "Twitter Sentiment Analysis of Movie Reviews using Machine Learning Techniques". *Internasional Journal of Engineering and Technology (IJET)* Vol 7 No 6 Dec 2015 - Jan 2016.

Feldman, R & Sanger, J. 2007. *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press : New York.

Sipayung Evasaria, M. Maharani, Herastia Zefanya Ivan, "Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode *Naïve Bayes Classifier*," Departemen Sistem Informasi Institut Teknologi Harapan Bangsa Bandung, 2016.

Suryadharma Bertung, Susanto Dwi Tony, "Faktor Penerimaan Media Sosial Instansi Pemerintah di Indonesia," Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh November, 2017

Suyanto. (2017, Mei). *Data Mining untuk Klasifikasi Dan Klasterisasi Data*. Bandung: Penerbit Informatika.