

KLASIFIKASI DATA PENGADUAN MASYARAKAT PADA LAMAN PESDUK CIMAHU MENGGUNAKAN ROCCHIO

Khusnul Khuluqiyah*, Tacbir Hendro Pudjiantoro, Agung Wahana
Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Jenderal Achman Yani

Jalan Terusan Jenderal Sudirman PO BOX 148 Cimahi - Bandung, Telp./Fax: (022) 6656190

*Email: khuluqiyahkhusnul08@gmail.com

Abstrak

Pelayanan merupakan suatu hal yang utama dalam sebuah organisasi. Pemerintah daerah Kota Cimahi merupakan salah satu instansi pemerintah yang telah memanfaatkan teknologi telekomunikasi yaitu dengan adanya pesan singkat penduduk atau PESDUK. Teknologi ini digunakan untuk mempermudah dan mempercepat penyampaian informasi dari masyarakat kepada Pemerintahan Kota Cimahi baik itu berupa keluhan, pertanyaan, dukungan ataupun laporan. Bertambahnya jumlah data pengaduan masyarakat setiap harinya mengakibatkan banyaknya data yang belum tertanggapi atau belum dibaca. Terbatasnya pengetahuan operator tentang tugas pokok dan fungsi dari setiap dinas juga dapat menyebabkan kesalahan dalam pengklasifikasian pesan. Oleh karena itu kebutuhan pengklasifikasian pesan pengaduan secara otomatis diperlukan demi mengefisiensi waktu dan sumber daya manusia dalam pengelompokan pesan pengaduan. Penelitian dilakukan dengan menggunakan text mining serta menggunakan metode Rocchio dikombinasikan dengan fitur untuk dapat mendeteksi negasi dan pembobotan menggunakan Tf dan Tf-idf. Berdasarkan perhitungan dari data penilaian yang telah diuji, metode Rocchio mampu digunakan untuk melakukan klasifikasi pesan pengaduan masyarakat.

Kata kunci : pelayanan, pengaduan masyarakat, Rocchio, Text Mining

1. PENDAHULUAN

Pelayanan merupakan suatu hal yang utama dalam sebuah organisasi, perusahaan, terutama untuk instansi pemerintah yang harus melayani masyarakat dengan baik. Biasanya suatu lembaga atau instansi pemerintah akan menyediakan sebuah media untuk menyalurkan partisipasi baik itu pengaduan atau pun aspirasi seperti kotak keluhan, SMS, *e-mail*, *web* dan lain-lain.

PESDUK atau Pesan Singkat Penduduk merupakan sebuah wadah yang menampung pengaduan masyarakat Kota Cimahi. PESDUK dibangun sebagai salah satu cara menarik partisipasi masyarakat Kota Cimahi yang menampilkan hasil SMS pengaduan masyarakat yang sudah diolah. SMS yang mengandung unsur politik, rasis dan *spam* tidak akan dimunculkan sedangkan dalam proses pengklasifikasian data SMS membutuhkan waktu untuk memeriksa satu-persatu pesan dan terbatasnya pengetahuan operator tentang tugas pokok dan fungsi dari setiap dinas juga dapat menyebabkan kesalahan dalam mengategorikan pesan. Untuk mengetahui kategori pada data pengaduan di sebuah web secara otomatis tanpa dibaca satu-persatu, perlu dilakukannya pengukuran kemiripan pesan terkait, maka dapat digunakan sebuah proses pengenalan teks dan dokumen. *Text mining* adalah salah satu cara yang dapat mengatasi permasalahan yang telah dijabarkan sebelumnya. Banyak bidang yang bisa dijadikan kajian dalam penerapan teknologi ini salah satunya mengenai klasifikasi berita menggunakan algoritma *Confix-stripping Stemmer* yang berfungsi untuk melakukan proses *stemming* terhadap kata-kata berimbuhan dan merubahnya menjadi bentuk kata dasarnya, dan metode *Naive Bayes Classifier* untuk melakukan proses klasifikasinya (Kurniawan, et al., 2012). Penelitian mengenai klasifikasi teks dengan *naive bayes classifier* untuk pengelompokan teks berita dan abstract akademis mengkaji kinerja NBC untuk kategorisasi teks berita dan teks akademis. Penelitian menggunakan data 1000 dokumen berita dan 450 dokumen abstrak akademik. Seleksi kata dengan minimal muncul pada 4 atau 5 dokumen memberikan akurasi yang paling tinggi (Hamzah, 2012) penelitian lainnya menggunakan metode pohon keputusan. Penelitian ini dibuat sebuah aplikasi tambahan pada laman Kantor Pertanahan Kota Surabaya I yang memanfaatkan teknologi klasifikasi teks untuk mengolah pengaduan masyarakat secara otomatis dengan menggunakan algoritma klasifikasi berbasis pohon keputusan. Pohon keputusan dibangun dengan menggunakan kata kunci dan enam kelas yang

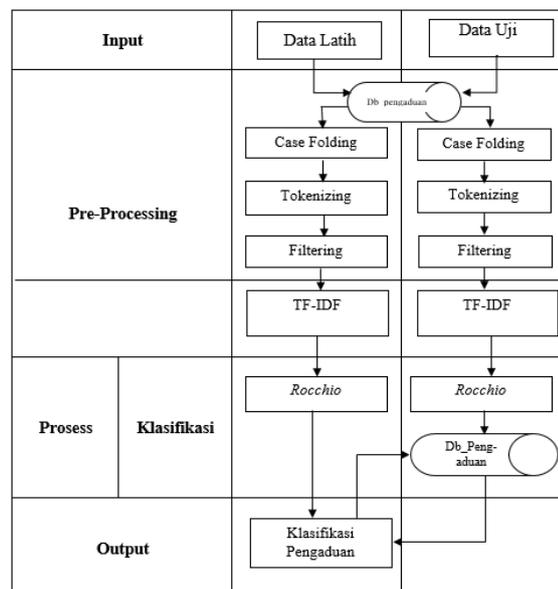
merepresentasikan enam seksi di mana pengaduan harus ditindak lanjuti. (Sulistianingsih, et al., 2015)

Penelitian mengenai penggunaan *Rocchio* untuk mengklasifikasikan dokumen bahasa inggris dengan menggunakan *Chi-Square* untuk memilih fitur. Jumlah seluruh dokumen adalah 21578 dan terbagi kedalam 20 kelas. Pemilihan fitur mampu mengurangi jumlah kata dan kata kunci yang akan digunakan untuk proses klasifikasi. (Ramdani, 2014) dan penelitian lainnya menggunakan teknik *Rocchio*. Teknik ini menggunakan *vector space model* dalam merepresentasikan dokumen, *centroid* dan kueri. Pembobotan dokumen menggunakan nilai *idf-itf* yang telah dilakukan proses normalisasi nilai vektor. Pengklasifikasian data training pada kelas-kelas yang telah ditentukan sangat mempengaruhi hasil rekomendasi dan hasil pencarian. untuk menguji klasifikasi dokumen dilakukan pengujian rekomendasi kelas terhadap data *training*. (Lumbanraja, 2013)

Pada penelitian yang akan dilakukan adalah membuat suatu sistem yang dapat melakukan klasifikasi pesan pengaduan masyarakat menggunakan *text mining* dengan metode *Rocchio*. Hasil dari sistem terdiri klasifikasi berdasarkan dinas yang ada.

2. METODOLOGI

Dalam penelitian ini algoritma atau metode yang digunakan yaitu *Rocchio* yang terdiri dari dua tahap yaitu *Learn* dan *Classify*. *Input* pada sistem ini adalah data dari masalah terkait dengan pengaduan masyarakat berupa data latih dan data uji. Pertama tahap *Learn* (pembelajaran/data latih), data yang menjadi masukan dari tahap ini yaitu data dokumen pengaduan yang sudah diketahui kategorinya. Kemudian *Classify* (pengujian/data uji), data yang menjadi masukan pada tahap ini yaitu data pengaduan yang baru dan belum diketahui kategorinya atau belum di klasifikasi kemudian data masukan ini akan di proses menggunakan metode *Rocchio* sehingga keluaran yang dihasilkan yaitu berupa klasifikasi pengaduan masyarakat berdasarkan kelas yang sudah ditentukan. Berikut ini sistem yang akan dibuat dengan penjelasan dari kedua tahap tersebut dapat dilihat pada Gambar 1. menunjukkan gambaran mengenai sistem yang akan dibangun.



Gambar 1 Gambaran Umum Sistem Klasifikasi Data Pengaduan Masyarakat

2.1. Data Masukan

Dalam metode penelitian ini masukannya adalah data dari masalah terkait dengan pengaduan masyarakat berupa data latih dan data uji. Pertama tahap *Learn* (pembelajaran/data latih), data yang menjadi masukan dari tahap ini yaitu data dokumen pengaduan yang sudah diketahui kategorinya. Kemudian *Classify* (pengujian/data uji), data yang menjadi masukan pada tahap ini yaitu data pengaduan yang baru dan belum diketahui kategorinya atau belum di klasifikasi

2.2. Pra Proses

Tahap ini terdiri dari beberapa tahapan diantaranya adalah

2.2.1. Case Folding

Case folding adalah proses mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter.

2.2.2. Tokenizing

Tahapan *tokenizing/parsing* adalah tahapan pemotongan string *input* berdasarkan tiap kata yang menyusunnya.

2.2.3. Filtering

Filtering adalah tahap mengambil kata-kata penting dari hasil *tokenizing*, dapat menggunakan algoritma *stoplist* dengan membuang kata yang kurang penting atau *wordlist* dengan menyimpan kata penting.

2.2.4. Term Frequency – Inverse Document Frequency

Perhitungan bobot dengan *Term Frequency – Inverse Document Frequency* (TF-IDF) menggunakan kombinasi dari dua nilai, yaitu frekuensi kata dan *inverse* frekuensi dokumen yang didapat dari membagi jumlah dokumen secara keseluruhan dengan jumlah dokumen dimana kata tersebut muncul. Berikut adalah rumus untuk mencari bobot dengan TF-IDF :

$$W_{d,t} = tf_{d,t} \times \log(D/df_t) \quad (1)$$

Dimana :

$tf_{d,t}$ = frekuensi banyak-nya kata ke-t dari kata kunci pada dokumen ke-d

df_t = jumlah dokumen yang mengandung kata ke-t dari kata

$W_{d,t}$ = bobot dokumen ke-d terhadap kata kunci

D = jumlah semua dokumen yang ada di d ke-t

kunci

alam database

Proses
Klasifikasi Rocchio merupakan metode klasifikasi yang merepresentasikan setiap dokumen dalam ruang vektor dan membagi ruang vektor tersebut menjadi beberapa bagian berdasarkan centroid yang ada. Klasifikasi Rocchio mudah dan sederhana. [8] Nilai *centroid* diperoleh dengan menghitung rata-rata vektor pada semua dokumen. *Centroid* kelas c dihitung dengan persamaan:

$$\vec{\mu}(c) = \frac{1}{|D_c|} \sum_{d \in D_c} \vec{v}(d)$$

dengan D_c adalah gugus dokumen di kelas c, $\vec{v}(d)$ adalah vektor kata-kata dalam kelas c, dan $\vec{\mu}(c)$ adalah centroid masing-masing kelas. Salah satu cara untuk menentukan kecocokan dokumen uji terhadap kelas adalah dengan menghitung jarak Euclidean antara kedua titik (x dan y) yang didefinisikan dengan persamaan: dengan rumus berikut:

Dimana,

Similarity (x,y) = nilai *similarity* dokumen uji dan dokumen latih

n = jumlah atribut di dalam dokumen

x_i = bobot dokumen uji pada atribut ke-i

y_i = bobot dokumen latih pada atribut ke-i

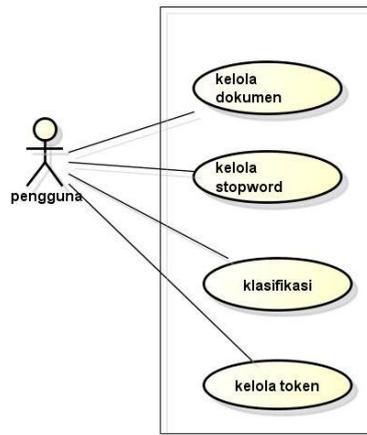
2.3. Keluaran

Keluaran yang dihasilkan yaitu berupa klasifikasi pengaduan masyarakat berdasarkan kelas yang sudah ditentukan

3. HASIL DAN PEMBAHASAN

3.1. Use Case Diagram

Usecase Diagram menggambarkan bagaimana aktor berinteraksi dengan sebuah sistem. *Usecase* dibuat berdasarkan analisa sistem yang sedang berjalan, *business actor* dan *business use case*. use case diagram system klasifikasi dapat dilihat pada gambar.



Gambar 2. Usecase Diagram

3.2. Pembahasan

Pada penelitian ini pembobot dokumen dilakukan dengan menggunakan metode *Term Frequency – Inverse Document Frequency (TF-IDF)*. Sebelum melakukan proses pembobotan setiap dokumen harus melalui tahap text preprocessing yang terdiri dari *case folding, tokenizing, dan filtering*. Adapun proses text preprocessing dapat dilihat pada gambar 2, dan perhitungan TF-IDF dapat dilihat pada tabel 1. pada tabel, D1 menunjukan kategori dinas kependudukan dan pencatatan sipil, D2 dinas kependudukan dan pertamanan, D3dinas perhubungan, D4 dinas pekerjaan umum dan Q merupakan data Uji yang belum memiliki kategori.



Gambar 3 Text Preprocessing kaegori Dinas kependudukan dan pencatatan sipil (D1)

Tabel 1 Term Frequency- Inverse Document Frequency

kata	TF					D	D/df	IDF	W					
	Q	D1	D2	D3	D4				Q	D1	D2	D3	D4	
pembuatan	0	1	0	0	0	1	4	0.6	0	0.602	0	0	0	0
kip	0	1	0	0	0	1	4	0.6	0	0.602	0	0	0	0
daerah	0	1	0	0	0	1	4	0.6	0	0.602	0	0	0	0
cinahi	0	1	0	0	0	1	4	0.6	0	0.602	0	0	0	0
masalah	0	1	0	0	0	1	4	0.6	0	0.602	0	0	0	0
warga	0	0	1	0	3	4	1	0	0	0	0	0	0	0
memberitahukan	0	0	1	0	0	1	4	0.6	0	0	0.602	0	0	0
Lampu	1	0	1	0	0	1	4	0.6	0.602	0	0.602	0	0	0
penerangan	1	0	1	0	0	1	4	0.6	0.602	0	0.602	0	0	0
jalan	1	0	2	2	2	6	0.66667	-0.18	-0.18	0	-0.352	-0.352	-0.352	-0.352
umutn	1	0	1	0	0	1	4	0.6	0.602	0	0.602	0	0	0
Mati	1	0	1	0	0	1	4	0.6	0.602	0	0.602	0	0	0
mohon	1	0	1	0	0	1	4	0.6	0.602	0	0.602	0	0	0
bantuannya	0	0	1	0	0	1	4	0.6	0	0	0.602	0	0	0
diperbaiki	0	0	1	0	0	1	4	0.6	0	0	0.602	0	0	0
jembatan	0	0	0	1	0	1	4	0.6	0	0	0	0.602	0	0
penyebrangan	0	0	0	1	0	1	4	0.6	0	0	0	0.602	0	0
nama	0	0	0	1	0	1	4	0.6	0	0	0	0.602	0	0
tujuannya	0	0	0	1	0	1	4	0.6	0	0	0	0.602	0	0
pencarian	0	0	0	1	0	1	4	0.6	0	0	0	0.602	0	0
alamat	0	0	0	1	0	1	4	0.6	0	0	0	0.602	0	0
kelurahan	0	0	0	0	1	1	4	0.6	0	0	0	0	0	0.602
melong	0	0	0	0	1	1	4	0.6	0	0	0	0	0	0.602
banjir	0	0	0	0	1	1	4	0.6	0	0	0	0	0	0.602
bertambah	0	0	0	0	1	1	4	0.6	0	0	0	0	0	0.602
parah	0	0	0	0	1	1	4	0.6	0	0	0	0	0	0.602
masuk	0	0	0	0	1	1	4	0.6	0	0	0	0	0	0.602
rumah	0	0	0	0	2	2	2	0.3	0	0	0	0	0	0.602
dicor	0	0	0	0	1	1	4	0.6	0	0	0	0	0	0.602
tinggi	0	0	0	0	2	2	2	0.3	0	0	0	0	0	0.602
melebihi	0	0	0	0	1	1	4	0.6	0	0	0	0	0	0.602
Gorong	0	0	0	0	1	1	4	0.6	0	0	0	0	0	0.602
diperbaiki	0	0	0	0	1	1	4	0.6	0	0	0	0	0	0.602
Diperbesar	0	0	0	0	1	1	4	0.6	0	0	0	0	0	0.602
Perdalam	0	0	0	0	1	1	4	0.6	0	0	0	0	0	0.602

3.3. Hasil Perhitungan Klasifikasi

Setelah melalui proses pembobotan, dokumen tersebut akan diklasifikasikan dengan metode *machice learning* yaitu menggunakan metode *Rocchio*. Perhitungan dan Hasil perhitungan klasifikasi dari tabel 2 dan 3.

Tabel 2 Perhitungan Rocchio

$D1$	$\sqrt{(0,602 - 0)^2 + (0,602 - 0)^2 + -(0,602 - 0)^2 + (0,602 - 0)^2 + (0,602 - 0)^2 + (0,602 - 0)^2 + (0,602 - 0)^2 + (0 - 0)^2} = 1,394$
$D2$	$\sqrt{(0,602 - 0,602)^2 + (0,602 - 0,602)^2 + -(0,176 - (-0,352))^2 + (0,602 - 0,602)^2 + (0,602 - 0,602)^2 + (0,602 - 0,602)^2 + (0 - 0,602)^2} = 0,627$
$D3$	$\sqrt{(0,602 - 0)^2 + (0,602 - 0)^2 + -(0,176 - (-0,352))^2 + (0,602 - 0)^2 + (0,602 - 0)^2 + (0,602 - 0)^2 + (0 - 0)^2} = 1,358$
$D4$	$\sqrt{(0,602 - 0)^2 + (0,602 - 0)^2 + -(0,176 - (-0,352))^2 + (0,602 - 0)^2 + (0,602 - 0)^2 + (0,602 - 0)^2 + (0 - 0)^2} = 1,475$

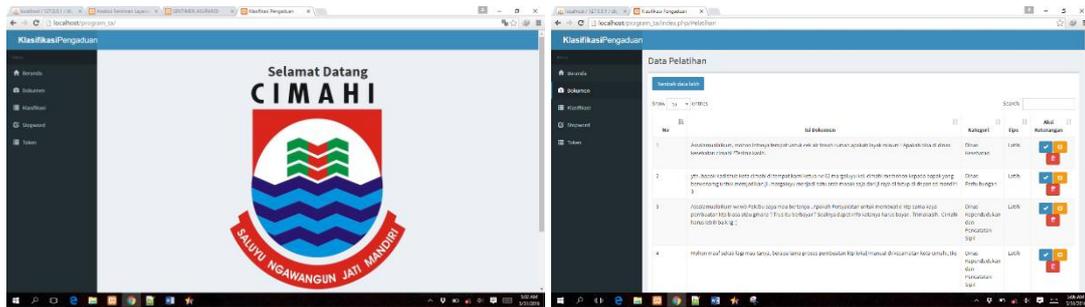
Tabel 3 Hasil Perhitungan Rocchio

D1	1.394
D2	0.627
D3	1.358
D4	1.475

Dari hasil perhitungan di atas didapatkan nilai paling kecil atau rendah yaitu 0.627 yang merupakan pesan jenis D2, maka didapatkan pada dokumen pengujian ini termasuk kedalam jenis pesan berdasarkan kategori Dinas Kebersihan dan Pertamanan.

3.4. Implementasi Antarmuka

Implementasi Antarmuka sistem klasifikasi data pengaduan masyarakat dapat dilihat pada gambar 4.



Gambar 4. Implementasi Antarmuka

4. KESIMPULAN DAN SARAN

4.1. Kesimpulan

Dari hasil kajian ini, maka dapat diambil kesimpulan bahwa kasus penentuan kategori/klasifikasi pesan pengaduan dengan *text mining* menggunakan *Rocchio* dapat dilakukan secara otomatis tanpa harus dibaca satu persatu terlebih dahulu. Sehingga diharapkan dapat membantu pihak operator dalam pengkategorian pesan pengaduan.

4.2. Saran

Saran untuk sistem klasifikasi data pengaduan masyarakat diharapkan dapat dikembangkan kembali dan dapat dijadikan bahan evaluasi untuk penelitian selanjutnya dengan menggunakan metode pembobotan dan klasifikasi yang lain.

DAFTAR PUSTAKA

- Hamzah, A., 2012. Klasifikasi teks dengan Naive Bayes Classifier (NBC) untuk Pengelompokan Teks Berita dan Abstract Akademis. *Seminar Nasional aplikasi Sains & Teknologi Periode III*.
- Kurniawan, B., Effendi, S. & Sitompul, O. S., 2012. Klasifikasi Konten Berita Dengan Metode Text Mining. *Jurnal Dunia Teknologi Informasi*, Volume I.
- Lumbanraja, F. R., 2013. Sistem Pencarian Data Teks Dengan Menggunakan Metode Klasifikasi Rocchio (Studi Kasus:Dokumen Teks Skripsi). *Seminar Semirata*, pp. 219-224.
- Ramdani, A. R., 2014. Klasifikasi Dokumen Dengan Metode Rocchio dengan Pemilihan Fitur Menggunakan Chi-Square. *Jurnal Dunia Teknologi*.
- Sulistianingsih, Y., Kusumawardani, R. P. & Djunaedi, A., 2015. Pengklasifikasian Pengaduan Masyarakat Pada Laman Kantor Pertahanan Kota Surabaya I Dengan Metode Pohon Keputusan. *Seminar Nasional "Informasi Dalam Desain Dan Teknologi*, pp. 93-99.