

KLASIFIKASI PENGADUAN MASYARAKAT PADA LAPORGUB MENGUNAKAN ALGORITMA STEMMING PORTER DAN NAIVE BAYES CLASSIFIER

Panggasa¹, Ruri Suko Basuki, M.Kom²

^{1,2}Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang

Jl. Nakula I No. 5-11, Semarang, 50131, 024 3517261

E-mail : 2201104234@mhs.dinus.ac.id¹, rury_sb.at.dosen.dinus.ac.id²

Abstrak

Penerapan teknologi informasi dan komunikasi (TIK) pada Pemerintah Provinsi Jawa Tengah mengalami perkembangan begitu pesat yang ditandai dengan penggunaan website dan sosial media untuk menyebar luaskan informasi mengenai penyelenggaraan pemerintahan dan pelayanan publik kepada masyarakat Jawa Tengah. Agar penyelenggaraan pemerintahan serta pelayanan publik lebih terarah dan tepat sasaran, diperlukan kontrol dari masyarakat terhadap kebijakan, anggaran dan mutu pelayanan Pemerintah Provinsi Jawa Tengah, salah satunya dengan cara memberikan umpan balik berupa pengaduan/aspirasi terkait pelayanan publik kepada Gubernur Jawa Tengah melalui LaporGub. Masyarakat dapat menyampaikan aduan melalui website dengan mengisi biodata dan isi aduan serta melampirkan gambar untuk kategori infrastruktur. Selanjutnya aduan akan di tindak lanjut oleh SKPD yang membidangi maksimal 10 hari kerja dan dapat diperpanjang maksimal 7 hari kerja menurut UU Nomor 14 Tahun 2008 tentang Keterbukaan Informasi Publik. Kecepatan Pemerintah Provinsi Jawa Tengah dalam merespon aduan masyarakat menjadi hal yang sangat penting bagi masyarakat. Sehingga untuk meningkatkan respon terhadap aduan masyarakat pada LaporGub diperlukan pengembangan sistem klasifikasi aduan. Sistem ini dikembangkan dengan metode Prototype menggunakan algoritma Stemming Porter untuk text preprocessing dan Naïve Bayes Classifier. Pada penelitian ini sistem klasifikasi aduan berhasil mengklasifikasikan 500 aduan ke dalam 17 kategori dengan akurasi optimal 84% dengan dokumen uji dan dokumen latih masing-masing 250 dokumen.

Kata Kunci: LaporGub, Klasifikasi, Aduan, Stemming Porter, Naive Bayes Classifier

Abstract

The application of information and communication technology (ICT) in Central Java has developed so rapidly that is characterized by the use of websites and social media to disseminate information on governance and public services to the people of Central Java. In order for governance and public services more focused and targeted, takes control of society on policy, budget and quality of service of the Government of the Province of Central Java, such as giving feedback in the form of complaints / aspiration related public services to the Governor of Central Java through LaporGub. Society can send a complaint through the website by filling the contents of the complaint as well as biographical data and attach the image to the infrastructure category. Furthermore, the complaint will be in the follow-up by SKPD about 10 working days and can be extended up to 7 working days according to Law No. 14 of 2008 on Public Information. Government of the Province of Central Java time consumption in responding to public complaints become very important for the community. So to improve the response to public complaints on LaporGub required the development of a classification system complaints. This system was developed by using a prototyping method and Porter Stemming algorithm for text preprocessing and Naive Bayes classifier. In this study the complaint classification system successfully classifies 500 complaints into 17 categories with optimal accuracy of 84% with test documents and document each train 250 documents.

Keywords: LaporGub , Classification , Complaints , Porter Stemming , Naive Bayes Classifier

1. PENDAHULUAN

1.1 Latar Belakang

Penerapan teknologi informasi dan komunikasi (TIK) pada Pemerintah Provinsi Jawa Tengah mengalami perkembangan begitu pesat yang ditandai dengan penggunaan website dan sosial media untuk menyebar luaskan informasi mengenai penyelenggaraan pemerintahan dan pelayanan publik kepada masyarakat Jawa Tengah. Agar penyelenggaraan pemerintahan serta pelayanan publik lebih terarah dan tepat sasaran, diperlukan kontrol dari masyarakat terhadap kebijakan, anggaran dan mutu pelayanan Pemerintah Provinsi Jawa Tengah. Salah satunya dengan cara memberikan umpan balik (*feed-back*) berupa pengaduan/aspirasi terkait pelayanan publik kepada Gubernur Jawa Tengah melalui LapoGub. Masyarakat yang ingin menyampaikan pengaduan dapat mengisi nama, telepon/email, aduan pada halaman LapoGub. Untuk aduan infrastruktur diwajibkan melampirkan gambar terkini sebagai file pendukung.

Kemudian admin bersama Gubernur akan mengklasifikasi aduan tersebut serta meneruskan kepada SKPD yang membidangi. Selanjutnya aduan yang masuk ke SKPD tersebut akan ditanggapi maksimal 10 hari kerja dan dapat diperpanjang maksimal 7 hari kerja menurut UU Nomor 14 Tahun 2008 tentang Keterbukaan Informasi Publik. Kecepatan dalam mengklasifikasi aduan harus ditingkatkan agar tanggapan oleh SKPD lebih cepat diberikan. Oleh karena itu, perlu adanya sistem yang dapat mengklasifikasi pengaduan secara otomatis dengan teknik *text mining*.

Text mining adalah salah satu teknik yang dapat digunakan untuk melakukan klasifikasi dimana *text mining* merupakan variasi dari *data mining* yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar. Selain klasifikasi, *text mining* juga digunakan untuk menangani masalah *clustering*, *information extraction*, dan *information retrieval*.

Berdasarkan temuan masalah tersebut maka diperlukan sistem yang dapat mengklasifikasi pengaduan secara otomatis sesuai dengan kategori-kategori yang ada. Oleh sebab itu penulis menggunakan Algoritma Stemming Porter dan Naïve Bayes Classifier untuk mengklasifikasi laporan pengaduan masyarakat tersebut.

1.2 Perumusan Masalah

1. Bagaimana mengklasifikasi aduan masyarakat pada LapoGub Pemerintah Provinsi Jawa Tengah secara otomatis?
2. Bagaimana menerapkan algoritma *Stemming Porter* dan *Naive Bayes Classifier* ke dalam sistem klasifikasi aduan?

1.3 Batasan Masalah

1. Aduan yang diklasifikasi berasal dari halaman *website* LapoGub Pemerintah Provinsi Jawa Tengah.
2. Algoritma yang digunakan dalam *text mining* adalah *Algoritma Porter*.
3. Algoritma yang digunakan dalam klasifikasi adalah Naïve Bayes Classifier dan tidak membandingkannya dengan algoritma lain.
4. Data yang digunakan berjumlah 500 aduan masyarakat.
5. Kategori aduan yang ada berjumlah 17 kategorisebagai berikut : infrastruktur, kesehatan, energi, pendidikan, kepegawaian, pertanian, pembangunan daerah, kependudukan, keuangan dan aset, bencana, ekonomi dan industri, sosial masyarakat, lingkungan, pariwisata dan budaya, sektor lain, forkopimda, dan kabupaten/kota.
6. Sistem yang dibangun tidak disatukan dengan halaman LapoGub tetapi dengan membuat *homepage* sendiri.

1.4 Tujuan Penelitian

Tujuan penulisan dalam penelitian adalah

1. Membangun sistem klasifikasi aduan masyarakat secara otomatis.
2. Menerapkan Algoritma *Stemming Porter* dan *Naive Bayes Classifier* untuk mengklasifikasi aduan masyarakat.

1.5 Manfaat Penelitian

Otomatisasi klasifikasi aduan masyarakat pada LapoGub Pemerintah Provinsi Jawa Tengah sehingga aduan masyarakat dapat ditindak lanjuti oleh SKPD dengan lebih cepat.

2 TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 LapoGub Provinsi Jawa Tengah

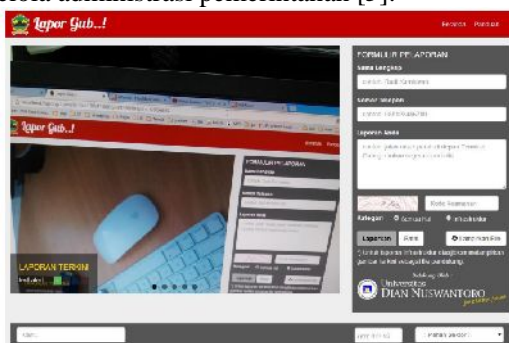
Seiring dengan kemajuan teknologi informasi yang membutuhkan informasi disajikan secara

cepat dan akurat, untuk mencapai hal tersebut tentunya membutuhkan suatu alat, media atau sistem yang dapat mengolah data secara efektif dan efisien yang dapat menghasilkan informasi tersebut. Berdasarkan pemikiran tersebut, Pemerintah Provinsi Jawa Tengah mencoba untuk menerapkan penggunaan kemajuan teknologi guna melayani masyarakat.

Dengan banyaknya keluhan masyarakat yang ingin disampaikan kepada pemerintahan, tentunya akan sulit apabila harus mendatangi kantor pemerintahan dan melalui prosedur yang panjang. Tentunya masyarakat akan merasa keberatan untuk melaporkan kejadian atau masalah yang ada di sekitar mereka. Dan apabila pemerintah tidak segera menindaklanjuti masalah yang ada di masyarakat, dikhawatirkan masyarakat akan menilai buruk kinerja pemerintah [3].

Untuk menjawab permasalahan tersebut Pemerintah Provinsi Jawa Tengah bersama Universitas Dian Nuswantoro melalui PT.Dinustek selaku pengembang meluncurkan LapoGub Pemerintah Provinsi Jawa Tengah pada 7 Mei 2014[2].

Dengan tersedianya aplikasi LapoGub ini diharapkan Pemerintah Provinsi Jawa Tengah menjadi organisasi yang handal dan mendukung pemanfaatan teknologi informasi dalam tata kelola administrasi pemerintahan [3].

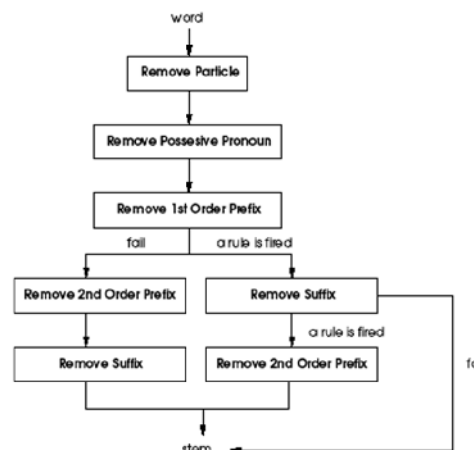


Gambar 1. Tampilan Beranda

2.2 Stemming Porter

Porter Stemmer for Bahasa Indonesia dikembangkan oleh Fadillah Z. Tala pada tahun 2003. Implementasi *Porter Stemmer for Bahasa Indonesia* berdasarkan *English Porter Stemmer* yang dikembangkan oleh W.B. Frakes pada tahun 1992. Karena bahasa Inggris datang dari kelas yang berbeda, beberapa modifikasi telah dilakukan untuk membuat Algoritma *Porter Bahasa Indonesia*[14]. Desain dari *Porter*

Stemmer for Bahasa Indonesia dapat dilihat pada gambar 2.14



Gambar 2. Desain dari *Porter Stemmer for Bahasa Indonesia* [10]

2.3 Naive Bayes Classifier

Algoritma *Naive Bayes Classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat[6]. Dalam penelitian ini yang menjadi data uji adalah dokumen laporan masuk. Ada dua tahap pada klasifikasi dokumen. Tahap pertama adalah pelatihan terhadap dokumen yang sudah diketahui kategorinya. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya.

Dalam algoritma *Naive Bayes Classifier* setiap dokumen direpresentasikan dengan pasangan atribut “x1, x2, x3,...xn” dimana x1 adalah kata pertama, x2 adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori pengaduan. Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (VMAP), dimana persamaannya adalah sebagai berikut :

$$VMAP = \arg \max P(v_j) \prod P(a_i | v_j)$$

Nilai $P(v_j)$ ditentukan pada saat pelatihan, yang nilainya didekati dengan :

$$P(v_j) = \frac{|docj|}{|contoh|}$$

Dimana $|docj|$ adalah banyaknya dokumen yang memiliki kategori j dalam pelatihan, sedangkan $|contoh|$ adalah banyaknya dokumen dalam contoh yang digunakan untuk pelatihan.

Untuk nilai $P(w_k|v_j)$ yaitu probabilitas kata w_k dalam kategori j ditentukan dengan :

$$P(w_k | v_j) = \frac{n_k + 1}{n + |vocabulary|}$$

Dimana n_k adalah frekuensi munculnya kata w_k dalam dokumen yang berkategori v_j , sedangkan nilai n adalah banyaknya seluruh kata dalam dokumen berkategori v_j dan $|vocabulary|$ adalah banyaknya kata dalam contoh pelatihan.

3 METODE PENELITIAN

3.1 Jenis Data

1 Data kualitatif

Jenis data kualitatif yaitu prosedur penelitian yang menghasilkan data tidak dalam bentuk angka, meliputi informasi tentang kata-kata dasar yang terdapat dalam kamus besar bahasa Indonesia, *stoplist* dari penelitian Tala dan dokumen aduan masyarakat laporGub.

2 Data kuantitatif

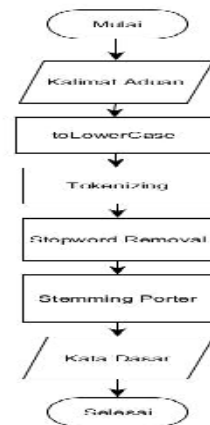
Jenis data kuantitatif yaitu prosedur penelitian yang menghasilkan data dalam bentuk angka seperti angka- angka yang akan dimasukan ke dalam rumus algoritma Naïve Bayes.

3.2 Perancangan Sistem

Pada sistem yang dibangun peneliti terdapat 3 proses utama yang harus dilalui. Pertama *text preprocessing* untuk menghasilkan kata dasar dan frekuensi kata yang digunakan sebagai masukan proses pelatihan dan pengujian. Kedua proses pelatihan untuk menghasilkan probabilitas kata dari setiap kategori yang sudah diketahui. Ketiga proses pengujian untuk menentukan kategori dari dokumen yang diuji. Masing-masing proses akan dirancang menggunakan *flowchart* agar lebih mudah dipahami.

1 Text Preprocessing

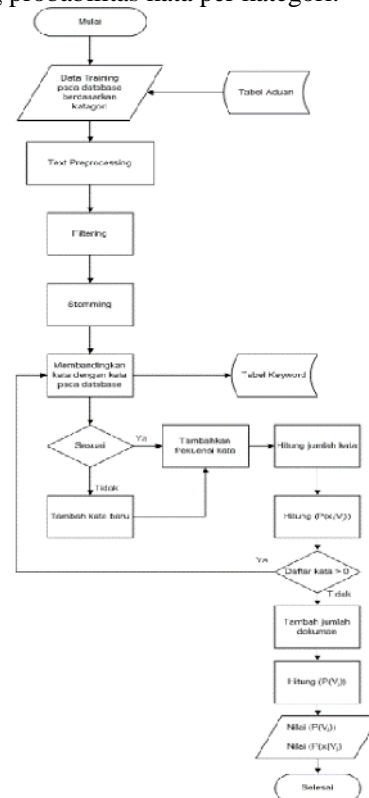
Text preprocessing merupakan tahap awal untuk menguraikan dokumen menjadi kata dasar dan frekuensi yang selanjutnya digunakan sebagai masukan proses pelatihan dan pengujian.



Gambar 3. Flowchart *Text Preprocessing*

2 Pelatihan

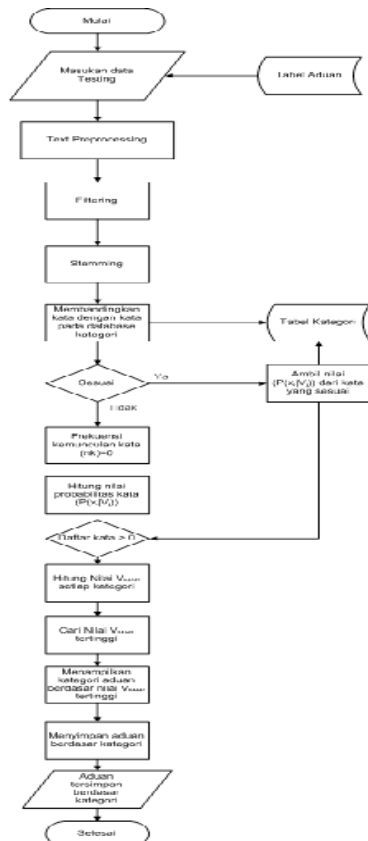
Pelatihan merupakan tahap dimana kata hasil *text preprocessing* dijadikan masukan kemudian dihitung probabilitas kata per kategori.



Gambar 4. Flowchart Pelatihan

3. Pengujian

Pengujian merupakan tahap dimana menguji dokumen yang belum diketahui kategorinya dengan menghitung Vmap kategori dari data pelatihan



Gambar 5. Flowchart Pengujian

4 HASIL DAN PEMBAHASAN

4.1 Analisa Sistem yang Berjalan

Mekanisme pelayanan aduan masyarakat saat ini dapat ditingkatkan kecepatan dan akurasi. Penulis mencoba menerapkan teknik *stemming* dan klasifikasi untuk membangun sistem klasifikasi secara otomatis. Sehingga dapat mempercepat proses tindak lanjut oleh SKPD yang berkenang.

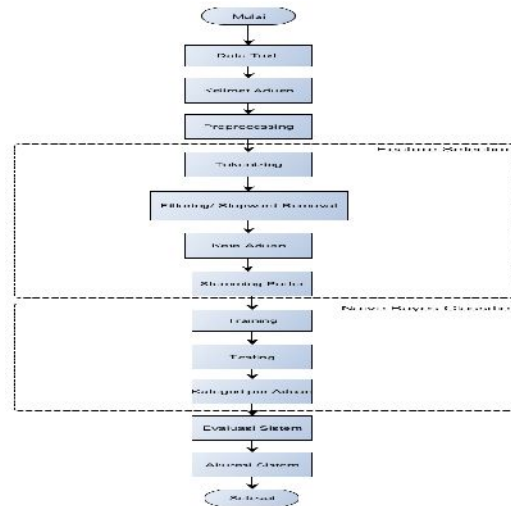
4.2 Alternatif yang Diusulkan

Dengan ditemukan permasalahan yang ada, maka alternatif yang diusulkan adalah membangun sistem klasifikasi otomatis dengan menerapkan *stemming Porter* dengan cara mengubah kalimat aduan menjadi kata dasar sehingga dapat diklasifikasi menggunakan algoritma *Naive Bayes*.

4.3 Framework Penelitian

Penelitian ini menggunakan dua metode yaitu algoritma *Stemming Porter* untuk *text preprocessing* dan *Naïve Bayes Classifier* digunakan untuk mengklasifikasi aduan

masyarakat kedalam 17 kategori yang sudah ditentukan.



Gambar 6. Framework Penelitian

4.4 Penggunaan Stemming Porter Bahasa Indonesia

Berdasarkan algoritma Porter langkah-langkah proses *stemming* adalah sebagai berikut :

1. Pertama cari kata yang akan distem dalam kamus kata dasar. Jika ditemukan maka diasumsikan kata tersebut adalah *root word*. Maka algoritma berhenti.
2. Hapus Partikel (“-lah”, “-kah”, “-tah” atau “-pun”). Partikel adalah sejenis kata yang tidak dikelompokkan ke dalam kelas kata *gramatikal infleksi* (*nomina*, *pronomina*, *verba* atau *partikel*).
3. Hapus *Possesive Pronouns* atau kata ganti (“-ku”, “-mu” atau “-nya”). *Possesive Pronouns* adalah kata ganti yang berfungsi untuk menunjukan kepemilikan.
4. Hapus *Derivation Prefix*. Jika tidak ada lanjutkan ke langkah 5.a, jika ada maka lanjutkan ke langkah 5.b.
5. a. Hapus awalan kedua (*Second Order Derivational Prefix*), lanjut ke langkah 6.a
b. Hapus *Inflection Suffixes*. *Inflection Suffixes* adalah huruf atau kelompok huruf yang ditempatkan di akhir kata atau *root* (bentuk paling sederhana (*base form*) dari suatu kata yang tidak dapat diuraikan lagi untuk menghasilkan kata baru. Jika tidak ditemukan maka kata tersebut diasumsikan sebagai *root word*, jika ditemukan maka lanjut ke langkah 6.b
6. a. Hapus *Derivation Suffixes* (“-i”, “-an” atau “-kan”). *Derivation Suffixes* adalah kata kerja

yang terbentuk dari proses pembentukan kata turunan dimana *verb* hasil bentukan tersebut akan memiliki makna yang berbeda dari kata dasarnya. Jika kata ditemukan dalam kamus maka algoritma berhenti dan kata tersebut diasumsikan sebagai *root word*.

b. Hapus awalan kedua (*Second Order Derivational Prefix*). Kemudian kata terakhir diasumsikan sebagai *root word*.

7. Hapus sisipan (*konffix*) untuk kombinasi awalan dan akhiran yang tidak diijinkan. Jika ditemukan maka algoritma berhenti dan kata terakhir diasumsikan sebagai *root word*.
8. Selesai.

Setelah melalui proses *stemming* menggunakan algoritma Porter maka *ouput*-nya dapat dilihat pada Tabel 1 berikut ini :

Id_Kata	Id_Dok	Keyword	Frekuensi
1	1	mohon	1
2	1	tindak	1
3	1	lanjut	1
4	1	dinas	2
5	1	instansi	2
6	1	kait	2
7	1	buat	2
8	1	rambu	2
9	1	lintas	2
10	1	tambah	1
11	1	batas	3
12	1	marka	1
13	1	jalan	4

4.5 Penggunaan Naive Bayes Classifier

Dalam penelitian ini sistem mempunyai 2 tahapan proses yaitu tahapan pertama adalah tahap pembelajaran atau training yaitu tahap pengklasifikasian terhadap aduan yang sudah diketahui kategorinya. Tujuan dari tahap training adalah untuk mencari keyword beserta probabilitasnya yang nantinya akan digunakan pada proses testing.

Sedangkan tahap kedua adalah tahap testing yaitu tahap pengklasifikasian terhadap aduan yang belum diketahui kategorinya.

1. Training

Untuk hasil keseluruhan dapat dilihat pada tabel Tabel 2 berikut ini :

Id_Training	Id_Kata	Probabilitas
1	1	0.00881057
2	2	0.00881057
3	3	0.00881057
4	4	0.0132159
5	5	0.0132159
6	6	0.0132159
7	7	0.0132159
8	8	0.0132159
9	9	0.0132159
10	10	0.00881057
11	11	0.0176211
12	12	0.00881057
13	13	0.0220264

2. Testing

Untuk hasil keseluruhan dapat dilihat pada tabel Tabel 3 berikut ini :

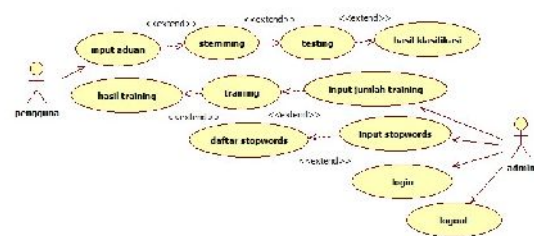
Id_Dok	Kategori	Vmap
3	Kab/Kota	1.5064528375733E-8
3	Infrastruktur	2.3604980424825E-8

Sehingga dapat ditarik kesimpulan bahwa Dokumen 3 memiliki kategori Infrastruktur.

4.6 Perancangan Sistem

1. Use case

Sistem klasifikasi aduan pada LapoGub ini mempunyai 2 user yaitu Pengguna dan Admin Verifikator. Gambaran diagram use case sistem dapat dilihat pada gambar 7 berikut ini :

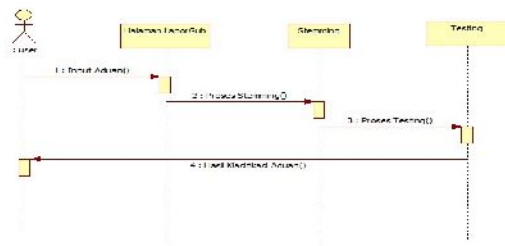


Gambar 7. Use Case Diagram

2. Diagram Sequence

Sequence diagram menggambarkan alur sistem klasifikasi aduan. Proses ini dimulai ketika user memasukkan aduan ke halaman LapoGub kemudian sistem akan menampilkan aduan dan kategori di halaman daftar aduan.

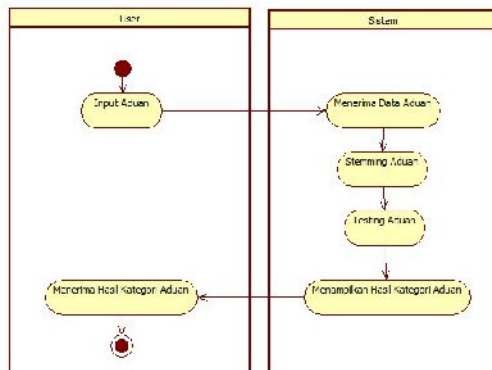
Berikut merupakan sequence diagram dari sistem klasifikasi aduan :



Gambar 8. Sequence Diagram

3. Diagram Aktivitas

Diagram aktifitas ini digunakan untuk menggambarkan berbagai alur aktifitas yang sedang berjalan. Berikut merupakan diagram aktifitas dari sistem klasifikasi aduan :



Gambar 9. Diagram Aktivitas

4.7 Desain Input Output

1 Rancangan Halaman Utama

SISTEM KLASIFIKASI ADUAN PADA LAPORGUB..!

Menu

- Tambah Aduan
- Data Stopword
- Data Kata Dasar
- Data Kamus Kata
- Data Kategori
- Data Training
- Data Testing
- Single Test Aduan
- Logout

Selamat Datang Admin!

Silahkan Gunakan Menu yang Tersedia

2 Rancangan Halaman Training

SISTEM KLASIFIKASI ADUAN PADA LAPORGUB..!

Menu

- Tambah Aduan
- Data Stopword
- Data Kata Dasar
- Data Kamus Kata
- Data Kategori
- Data Training
- Data Testing
- Single Test Aduan
- Logout

Halaman Training

Pilih Jumlah Data Training

Aduan

KD	USER	ADUAN	KAT

3 Rancangan Halaman Testing

SISTEM KLASIFIKASI ADUAN PADA LAPORGUB..!

Menu

- Tambah Aduan
- Data Stopword
- Data Kata Dasar
- Data Kamus Kata
- Data Kategori
- Data Training
- Data Testing
- Single Test Aduan
- Logout

Halaman Testing

Pilih Jumlah Data Testing

Aduan

KD	USER	ADUAN

4.8 Implementasi

Penulis melakukan training dan testing terhadap sistem klasifikasi aduan ini menggunakan data latih sebanyak 100 sampai 400 aduan dan 100 sampai 400 aduan untuk data tes-nya.



Gambar 10. Implementasi Sistem

4.9 Evaluasi Kinerja Algoritma

Tabel 4 Akurasi Uji Dokumen Aduan

Dokumen Latih	Dokumen Uji	Akurasi (%)
400	100	91%
350	150	86%
300	200	85%
250	250	84%
200	300	81%
150	350	74%
100	400	64%

5 KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari pendefinisian masalah, analisa dan penerapan sistem untuk mengklasifikasi aduan pada LaporanGub dengan menggunakan *Naïve Bayes Classifier* dan *Stemming Porter*, maka dapat diambil kesimpulan sebagai berikut :

1. Sistem klasifikasi aduan dapat diterapkan untuk mengklasifikasikan 500 aduan masyarakat ke dalam 17 kategori yang sudah ditentukan sehingga mempercepat dalam proses tindak lanjut oleh SKPD yang membidangi.
2. Sistem klasifikasi aduan dapat diterapkan untuk mengklasifikasikan aduan masyarakat dengan akurasi optimal 84% dengan dokumen uji dan dokumen latih masing-masing 250 dokumen.

5.2 Saran

Berdasarkan hasil penelitian dan pembahasan serta kesimpulan yang telah dikemukakan berikut ini saran yang diharapkan dapat menjadi pertimbangan dan masukan sehingga dapat membantu meningkatkan kinerja sistem klasifikasi. Beberapa saran untuk pengembangan penelitian selanjutnya adalah :

- 1 Penelitian ini menggunakan data pelatihan dengan jumlah dokumen per kategori yang tidak sama, untuk penelitian selanjutnya dapat dikembangkan dengan menggunakan dokumen yang sama dari tiap kategori.
- 2 Penelitian ini hanya menggunakan algoritma *Stemming Porter* untuk *text preprocessing* dan *Naïve Bayes* untuk klasifikasi. Diharapkan penelitian selanjutnya dapat dikembangkan dengan membandingkan akurasi dari *Stemming Porter* dan *Naïve Bayes* dengan algoritma lainnya.

DAFTAR PUSTAKA

- [1] Dodo Priambodo Kresno, "Pembuatan Prototipe Perangkat Lunak Pengklasifikasian Komplain Layananoperasional Pengguna Jasa Kepelabuhananmenggunakan Naïve Bayes Classifier (Studi Kasus Pelabuhan Cabang Tanjung Perak Surabaya)," pp. 19-21, 2014.
- [2] Wisnu Adhi N. (2014, Juni) antaranews. [Online].
<http://www.antaranews.com/berita/440916/gubernur-jateng-buka-situs-pengaduan-masyarakat>
- [3] (2014, Juni) dinustech. [Online].
<http://dinustech.com/laporgub/manual.pdf/>
- [4] M.W. & Kogan, J Berry, *Text Mining Application and theory*. WILEY , United Kingdom, 2010.
- [5] E., Fang, F., Sistla, P., Yu, S. & Meng, W Dragut, *Stop Word and Related Problems in Web Interface Integration.*, 2009.
- [6] Syahril Effendi, Opim Salim Sitompul Bambang Kurniawan, "Klasifikasi Konten Berita Dengan Metode Text Mining, Fakultas Ilmu Komputer dan Teknologi Informasi," *JURNAL DUNIA TEKNOLOGI INFORMASI*, vol. 1, no. 1, pp. 14-19, 2012.
- [7] Boy Utomo Manalu, "Analisis Sentimen Pada Twitter Menggunakan Text Mining".
- [8] Herny Februariyanti, "Perancangan Pengindeks Kata pada Dokumen Teks menggunakan Aplikasi Berbasis Web ," *Jurnal Teknologi Informasi DINAMIK* , vol. 18, no. 10, pp. 161-170, Juli 2013.
- [9] Eko Prasetyo, *Data Mining : Konsep dan Aplikasi menggunakan MATLAB*. Yogyakarta, Indonesia: Penerbit Andi, 2012.
- [10] B. et al Amaliah, "Klasifikasi Voted Perceptron untuk Identifikasi Melanoma," *Seminar Nasional Aplikasi Teknologi Informasi 2011 (SNATI 2011)*, pp. C1-C8, 2011.
- [11] R & Sanger, J Feldman, *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*. New York, USA: Cambridge University Press, 2007.
- [12] J & Kamber, M Han, *Data Mining: Concepts and Techniques*, 2nd ed. San Francisco, USA: Morgan Kaufmann publisher, 2006.
- [13] S.M., Indurkha, N., Zhang, T., Damerau, F.J Weiss, *Text Mining : Predictive Methods fo Analyzing Unstructured Information*. New York, USA: Springer, 2005.
- [14] Fadillah Z Tala, "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia," 2003.
- [15] Abdul Kadir & Terra CH Triwahyuni, *Pengenalan Teknologi Informasi*. Yogyakarta, Indonesia: Penerbit Andi., 2005.