

PEMBUATAN PROTOTIPE PERANGKAT LUNAK PENGKLASIFIKASIAN KOMPLAIN LAYANAN OPERASIONAL PENGGUNA JASA KEPELABUHANAN MENGUNAKAN *NAÏVE BAYES CLASSIFIER* (STUDI KASUS PELABUHAN CABANG TANJUNG PERAK SURABAYA)

Dodo Priambodo Kresno

Sholiq

^{1) 2)} S1/Jurusan Sistem Informasi, STIKOMP Surabaya,

¹⁾ Email: dd_graff@yahoo.com;

²⁾ Email: sholiq@stikom.edu

Abstract: Quality of service activities port have to ever imported to reach good government mention, one of them is by accepting customer feedback in the form of complain operational service. By exploiting technology internet and mobile phone, complain operational service will be processed and classified automatically based on lable categorize and delivered to relevant divison in Tanjung Perak port branch office by using one of algorithm *Machine Learning* that is *Naïve Bayes Classifier*. Result of research test-drive by using complain data, indicate that the prototype of classification system complain operational service of port services can classified complain quickly and precisely and serve the purpose of reference in system development classification complain manifestly.

Keywords: complaint classification, operational service of port service, naïve bayes classifier

Komplain layanan operasional jasa kepelabuhanan dapat didefinisikan sebagai salah satu cara dimana pelanggan (pengguna jasa kepelabuhanan) dapat berkomunikasi secara langsung dengan pihak perusahaan. Agar layanan operasional jasa kepelabuhanan dapat bekerja dengan baik dan benar, maka kontrol terhadap mutu pelayanan mutlak harus dilakukan. Salah satunya dengan cara menerima umpan balik (*feed-back*) dari pelanggan pengguna jasa kepelabuhanan berupa komplain layanan operasional jasa kepelabuhanan.

Saat ini, pengguna jasa dapat menyampaikan komplain secara lisan ataupun tertulis dengan cara datang langsung pada bagian Pusat Pelayanan Satu Atap (PPSA) kantor Pelabuhan Cabang Tanjung Perak. Komplain yang masuk oleh petugas akan diklasifikasikan berdasarkan kategori dan divisinya, dimana komplain diklasifikasikan secara manual berdasarkan kategori dan divisi terkait. sehingga proses pengklasifikasian komplain cenderung lambat dan komplain terabaikan.

Dengan kondisi ini, kesulitan yang dihadapi adalah bagaimana mengklasifikasikan komplain secara cepat dan tepat. Dari permasalahan tersebut, maka selanjutnya akan dibuat suatu perangkat lunak sistem pengklasifikasian komplain layanan

operasional menggunakan *naïve bayes classifier* berbasis web dan didukung fasilitas SMS.

Tujuan penelitian ini adalah

1. Membangun suatu sistem pengklasifikasian komplain layanan operasional pelanggan (jasa kepelabuhanan) berbasis web dan didukung fasilitas *Short Message Services* (SMS).
2. Menerapkan algoritma *Naïve Bayes Classifier* untuk mengklasifikasikan komplain layanan operasional jasa kepelabuhanan berdasarkan kategori.
3. Menerapkan algoritma *Naïve Bayes Classifier* untuk mengklasifikasikan komplain layanan operasional jasa kepelabuhanan berdasarkan divisi terkait.
4. Menerapkan algoritma TF-IDF untuk mendukung penerapan algoritma *Naïve Bayes Classifier* dalam mengklasifikasikan komplain operasional dengan data teks yang tidak terstruktur.

Algoritma Naïve Bayes Classifier untuk Klasifikasi Teks.

Pada *Naïve Bayes Classifier* (NBC) setiap dokumen (komplain) direpresentasikan dalam pasangan atribut (a_1, a_2, \dots, a_n) di mana a_1 adalah kata pertama, a_2 kata kedua dan seterusnya (Wibisono,

2005). Sedangkan V adalah himpunan kategori komplain (keluhan, saran dan keberatan). Pada saat klasifikasi, pendekatan Bayes akan menghasilkan label kategori yang paling tinggi probabilitasnya (V_{MAP}) dengan memasukkan atribut ($a_1, a_2 \dots a_n$)

$$v_{MAP} = \arg \max_{v_j \in V} P(v_j | a_1, a_2 \dots a_n) \dots \dots \dots (1)$$

(Wibisono, 2005)

Teorema Bayes menyatakan :

$$P(B | A) = \frac{P(A | B)P(B)}{P(A)} \dots \dots \dots (2)$$

(Wibisono, 2005)

Menggunakan teorema Bayes ini, rumus 3 dapat ditulis :

$$v_{MAP} = \arg \max_{v_j \in V} \frac{P(a_1, a_2 \dots a_n | v_j)P(v_j)}{P(a_1, a_2 \dots a_n)} \dots \dots \dots (3)$$

(Wibisono, 2005)

$P(a_1, a_2 \dots a_n)$ nilainya konstan untuk semua v_j sehingga persamaan ini dapat ditulis sebagai berikut:

$$v_{MAP} = \arg \max_{V_j \in V} P(a_1, a_2 \dots a_n | v_j)P(v_j) \dots \dots \dots (4)$$

(Wibisono, 2005)

Tingkat kesulitan menghitung $P(a_1, a_2 \dots a_n | v_j)$ menjadi tinggi karena jumlah term $P(a_1, a_2 \dots a_n | v_j)$ bisa jadi akan sangat besar. Ini disebabkan jumlah term tersebut sama dengan jumlah semua kombinasi posisi kata dikali dengan jumlah kategori. Naive Bayes Classifier menyederhanakan hal ini dengan mengasumsikan bahwa di dalam setiap kategori, setiap kata independen satu sama lain. Dengan kata lain:

$$P(a_1, a_2 \dots a_n | v_j) = \prod_i P(a_i | v_j) \dots \dots \dots (5)$$

(Wibisono, 2005)

Substitusi rumus 4 dan rumus 5 akan menghasilkan :

$$v_{MAP} = \arg \max_{V_j \in V} P(v_j) \prod_i P(a_i | v_j) \dots \dots \dots (6)$$

(Wibisono, 2005)

$P(v_j)$ dan probabilitas kata w_k untuk setiap kategori $P(w_k | v_i)$ dihitung pada saat pelatihan.

$$P(v_j) = \frac{| docs_j |}{| Contoh |} \dots \dots \dots (7)$$

(Wibisono, 2005)

$$P(w_k | v_j) = \frac{n_k + 1}{n + | kosakata |} \dots \dots \dots (8)$$

(Wibisono, 2005)

Dimana $| docs_j |$ adalah jumlah kata pada kategori j dan $| Contoh |$ adalah jumlah dokumen yang digunakan dalam pelatihan. Sedangkan n_k adalah jumlah kemunculan kata w_k pada kategori v_j , n adalah jumlah semua kata pada kategori v_j dan $| kosakata |$ adalah jumlah kata yang unik (*distinct*) pada semua data latihan.

Algoritma TF-IDF

Jiawei (2007), Algoritma TF-IDF merupakan metode pembobotan dalam bentuk sebuah metode yang merupakan integrasi antar *term frequency (tf)*, dan *inverse document frequency (idf)*. Fungsi metode ini adalah untuk mencari representasi nilai dari tiap-tiap kata dari suatu kumpulan data latih (*Training Set*). Adapun formula yang digunakan untuk menghitung bobot (w) masing-masing dokumen terhadap kata kunci adalah :

$$W_{d,t} = tf_{d,t} * IDF_t \dots \dots \dots (9)$$

(Milkha Harlian, 2006)

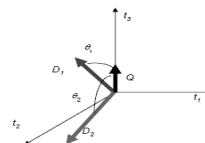
Dimana :

- $tf_{d,t}$ (*term frequency*): merupakan banyaknya kemunculan term ke t pada dokumen ke d .
- df : merupakan banyaknya dokumen dimana *term* ke- t muncul didalamnya (df_t)
- idf (*inverse document frequency*) : merupakan ukuran deskriminankemunculan term ke- t dalam koleksi $idf_t = \log_{10} (N / df_t)$, dimana N adalah banyaknya dokumen.

Algoritma Vector Space Model

Menurut Milkha Harlian Ch (2006), algoritma *Vector Space Model* merupakan metode yang memungkinkan untuk melakukan pemeringkatan dokumen. *Vector Space Model* digunakan untuk mengurutkan (*sorting*) nilai bobot w (hasil perhitungan TF-IDF) jika memiliki nilai yang sama.

Ide awal dari metode ini adalah dengan menghitung nilai cosinus sudut dari dua vector, yaitu W dari tiap dokumen dan W dari kata kunci.



Gambar 1. Sudut Cosinus dari Dua Vektor

Rumus :

$$\text{sim}(d_j, q) = \frac{\|\vec{d}_j \cdot \vec{q}\|}{\|\vec{d}_j\| \times \|\vec{q}\|} = \frac{\sum_{i=1}^l (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^l w_{ij}^2 \cdot \sum_{i=1}^l w_{iq}^2}} \dots\dots\dots 10$$

(Milkha Harlian, 2006)

Dimana :

$$\|\vec{q}\| = \sqrt{q \cdot q} \text{ dan } \|\vec{d}_1\| = \sqrt{d_1 \cdot d_1}$$

Sedangkan perhitungan *Vector Space Model* untuk proses pengurutannya :

$$\text{Cosine (Di)} = \frac{\text{SUM (kk dot Di)}}{[\text{sqrt (kk)}] \cdot [\text{sqrt(Di)}]} \dots\dots\dots 11$$

(Milkha Harlian, 2006)

Dimana :

- Di = dokumen ke - i

$$\text{Sum (kk dot Di)} = \sum_{j=1}^n k k_j D_{i,j} \dots\dots\dots 12$$

(Milkha Harlian, 2006)

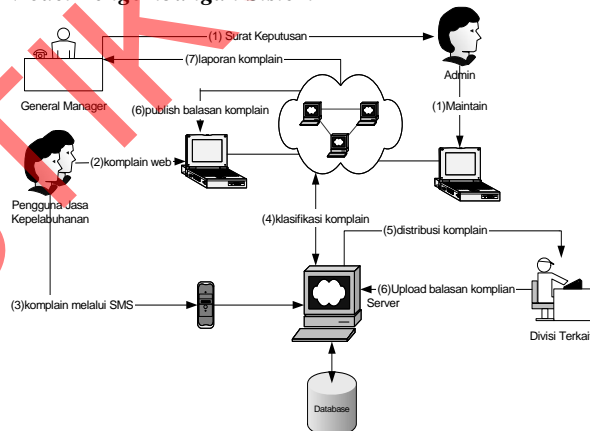
$$\text{Sqrt(kk)} = \text{Sqrt}(\sum_{j=1}^n k k_j^2) \dots\dots\dots 13$$

(Milkha Harlian, 2006)

$$\text{Sqrt(Di)} = \text{Sqrt}(\sum_{j=1}^n D_{i,j}^2) \dots\dots\dots 14$$

(Milkha Harlian, 2006)

METODE
Model Pengembangan Sistem



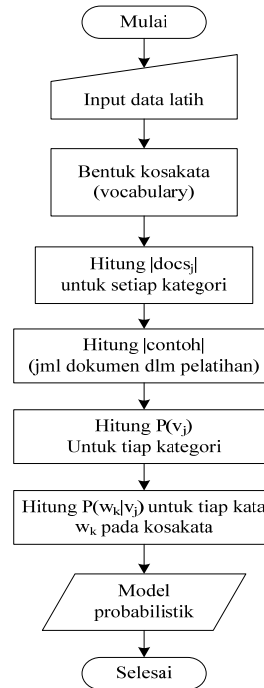
Gambar 2. Model Sistem Informasi Pengklasifikasian Komplain

Gambar 2 merupakan gambar dari model pengembangan sistem pengklasifikasikan komplain layanan operasional pengguna jasa kepelabuhanan, dimana terdapat beberapa orang maupun organisasi yang terlibat didalamnya seperti general manager (GM), pengguna jasa kepelabuhanan, divisi terkait, dan administrator. Pada gambar tersebut dapat pengguna jasa dapat mengirimkan komplain layanan operasional melalui web maupun handphone via SMS. Komplain yang masuk kemudian akan diklasifikasikan kategorinya dan akan dikirimkan ke divisi terkait di kantor Pelabuhan Cabang Tanjung Perak.

Penerapan Algoritma

a. Proses Pelatihan

Proses pelatihan dimulai dengan pembentukan kosakata (*vocabulary*) dimana kosakata adalah himpunan semua kata unik dari data latih (Masayu, 2007 dan Vincent, 2005). Dari himpunan tersebut masing-masing kata akan dihitung jumlahnya untuk mendapatkan nilai-nilai probabilitas yaitu $P(v_j)$ dan $P(w_k|v_j)$, dimana $P(w_k|v_j)$ merupakan probabilitas kemunculan kata w_k pada tiap kategori v_j . Untuk lebih jelasnya proses pelatihan dapat dilihat pada gambar 3 berikut ini :

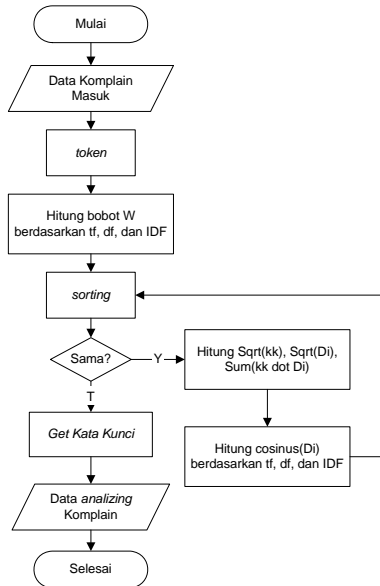


Gambar 3. Flowchart Proses Pelatihan

b. Proses Analizing

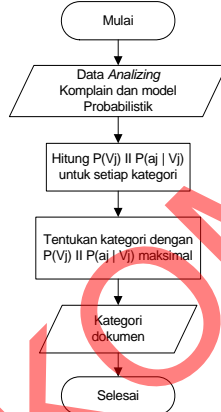
Proses *analizing* bertujuan untuk mencari kemiripan kalimat dengan kata kunci *term* terpilih dari masing-masing kategori yang ada sehingga kata kunci yang mewakili (*representation*) kalimat tersebut dapat diketahui. Dalam proses *analizing* ini akan diterapkan algoritma TF-IDF untuk pembobotan nilai (w) dan

algoritma *Vector Space Model (Cosine Similarity)* yang digunakan untuk menunjukkan kemiripan antar dokumen. Proses *analyzing* dapat dilihat pada gambar 4 berikut ini :



Gambar 4. Flowchart Proses Analyzing

c. Proses Klasifikasi



Gambar 5. Flowchart Proses Analyzing

Data komplain yang dihasilkan dari proses *Analyzing* dan model probabilistik pada tahap pelatihan akan dijadikan *input* untuk proses selanjutnya yaitu proses klasifikasi dimana di dalamnya akan dilakukan perhitungan v_{MAP} untuk setiap kategori dengan rumus 6. Setelah didapatkan hasil perhitungan dari masing-masing kategori tersebut selanjutnya dipilih kategori dengan v_{MAP} yang maksimal dan digunakan sebagai acuan untuk melakukan klasifikasi terhadap data komplain. Komplain akan diklasifikasikan sesuai dengan kategori yang memiliki nilai v_{MAP} yang maksimal.

HASIL DAN PEMBAHASAN

Pada tahap ini diuraikan hasil dan pembahasan penelitian terhadap perangkat sistem pengklasifikasian komplain layanan operasional jasa kepelabuhanan menggunakan *Naive Bayes Classifier*. Proses evaluasi dari aplikasi *complaint center* dengan menggunakan algoritma *Naive Bayes Classifier* ini dilakukan dengan beberapa tahap, yaitu :

Evaluasi Tahap Pelatihan

Tahap pelatihan bertujuan untuk mencari nilai $P(w_k | v_j)$ dan nilai $P(V_j)$ (Muhamad Rachli, 2007). Data latih yang digunakan adalah data komplain layanan operasional yang telah diklasifikasikan jenis kategori dan divisinya. Data latih komplain yang telah diklasifikasikan sebelumnya, akan dibersihkan dari noiseword dan disimpan ke dalam file vocab kata latih. Nilai-nilai probabilistik yang dihasilkan dari proses pelatihan tersebut adalah $P(w_k | v_j)$ dan nilai $P(V_j)$. Pada aplikasi nilai probabilistik $P(w_k | v_j)$ dapat ditunjukkan pada gambar 6, sebagai berikut :

Daftar $P(w_k | v_j)$ masing-masing kata per-kategori :

No.	Kategori	Kata	$P(w_k v_j)$
2	KELUHAN	kecewa	(0,438)
3	SARAN	kecewa	(0,091)
4	KEBERATAN	kecewa	(0,067)
5	KELUHAN	keluhan	(0,313)
6	SARAN	keluhan	(0,091)
7	KEBERATAN	keluhan	(0,067)
11	KELUHAN	saran	(0,063)
12	SARAN	saran	(0,364)
13	KEBERATAN	saran	(0,067)
14	KELUHAN	mohon	(0,063)
15	SARAN	mohon	(0,273)
16	KEBERATAN	mohon	(0,4)
17	KELUHAN	koreksi	(0,063)
18	SARAN	koreksi	(0,091)
19	KEBERATAN	koreksi	(0,267)
20	KELUHAN	kesalahan	(0,063)
21	SARAN	kesalahan	(0,091)
22	KEBERATAN	kesalahan	(0,133)

Gambar 6. nilai probabilistik $P(w_k | v_j)$

Sedangkan untuk $P(V_j)$ dapat ditunjukkan pada gambar 7, sebagai berikut :

Probabilitas masing-masing kategori :

No.	Kategori	$P(V_j)$
1	KELUHAN	(0,556)
2	SARAN	(0,278)
3	KEBERATAN	(0,5)

Gambar 7. nilai probabilistik $P(V_j)$

Evaluasi Tahap Uji Coba

a. Uji Coba Analyzing

Sebelum dilakukan klasifikasi menggunakan algoritma *Naive Bayes Classifier*, kalimat komplain layanan operasional akan dianalisa menggunakan algoritma TF-IDF dan *Vector Space Model* sehingga

menghasilkan dokumen analisa yang merupakan representasi / maksud dari kalimat komplain layanan operasional pengguna jasa kepelabuhanan. Dari dokumen analisa kalimat komplain layanan operasional pengguna jasa kepelabuhanan akan menjadi *input* untuk proses klasifikasi pada algoritma *Naïve Bayes Classifier*.

Misal diasumsikan masing-masing terdapat 3 dokumen yang ada proses *analizing* kategori, yaitu dokumen 1 yang merupakan kumpulan kata kunci dari kategori 'keluhan', dokumen 2 merupakan kumpulan kata kunci dari kategori 'saran', dan dokumen 3 merupakan kumpulan kata kunci dari kategori 'keberatan'.

Dokumen 1 : kecewa, keluhan

Dokumen 2 : tolong, usul, supaya

Dokumen 3 : keberatan, permohonan, peninjauan.

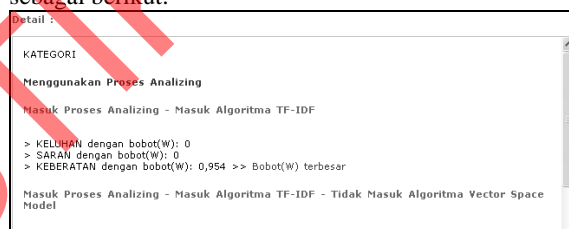
Setelah kata kunci masing-masing kategori didefinisikan, proses selanjutnya adalah menemukan kata pada kalimat komplain yang ada pada masing-masing dokumen kategori. contoh kalimat komplain yang akan di analisa, sebagai berikut:

Bersama ini kami sampaikan permohonan peninjauan ulang dan perubahan untuk nota penjualan jasa kepelabuhanan no. 010.001.06.00001561 untuk kegiatan kapal kami M/V WAN HAI 206 VOY.N211, sehubungan dengan uang pandu sebesar USD.105.72. Terima kasih atas kerja samanya yang baik

Pada kalimat tersebut ditemukan kata yang terdapat pada salah satu dokumen kategori, sebagai berikut:

Kata yang ditemukan : permohonan, peninjauan.

Kata yang ditemukan selanjutnya akan disebut "Kata Kunci". Pada proses *analizing* akan dicari kedekatan kalimat komplain dengan dokumen yang merupakan kata kunci masing-masing kategori untuk kedekatan kata kunci dengan masing-masing dokumen pada kategori tertentu. Dari proses *analizing* dihasilkan analisa berupa dokumen keberatan. Pada gambar 8 merupakan gambar hasil proses analisa, sebagai berikut:



Gambar 8. Hasil Proses *Analizing* Kalimat

b.Uji Coba Klasifikasi

Pada proses klasifikasi langkah yang dilakukan adalah menghitung nilai V_{MAP} masing-masing kategori maupun divisi. Pada proses klasifikasi akan dihasilkan label kategori yang paling tinggi

probabilitasnya (V_{MAP}), sehingga komplain yang masuk dapat diketahui kategorinya dan akan dikirimkan ke divisi terkait PT.(Persero) Pelabuhan Indonesia III Cabang Tanjung Perak Surabaya.

Nilai V_{MAP} masing-masing kategori akan dihitung menggunakan algoritma *naïve bayes classifier* dengan rumus 6. Dengan menggunakan contoh kalimat seperti pada proses *analizing*, maka data komplain akan diklasifikasikan berdasarkan kategorinya, sebagai berikut :

Keluhan

$P(v_j)$ kategori keluhan = 0,917

$P(w_k | v_j)$ masing-masing kata pada kategori keluhan :

$P(\text{"permohonan"} | \text{"keluhan"}) = 0,02857$

$P(\text{"peninjauan"} | \text{"keluhan"}) = 0,02857$

$P(\text{"ulang"} | \text{"keluhan"}) = 0,02857$

$P(\text{"baik"} | \text{"keluhan"}) = 0,02857$

Saran

$P(v_j)$ kategori saran = 0,5

$P(w_k | v_j)$ masing-masing kata pada kategori saran:

$P(\text{"permohonan"} | \text{"saran"}) = 0,033333$

$P(\text{"peninjauan"} | \text{"saran"}) = 0,033333$

$P(\text{"ulang"} | \text{"saran"}) = 0,033333$

$P(\text{"baik"} | \text{"saran"}) = 0,066667$

Keberatan

$P(v_j)$ kategori saran = 1,25

$P(w_k | v_j)$ masing-masing kata pada kategori keberatan

:

$P(\text{"permohonan"} | \text{"keberatan"}) = 0,0769231$

$P(\text{"peninjauan"} | \text{"keberatan"}) = 0,0769231$

$P(\text{"ulang"} | \text{"keberatan"}) = 0,0769231$

$P(\text{"baik"} | \text{"keberatan"}) = 0,0256410$

Sehingga perhitungan V_{MAP} dengan rumus 7 untuk masing-masing kategori diperoleh :

V_{MAP} untuk kategori keluhan = 0,104762

V_{MAP} untuk kategori saran = 0,083333

V_{MAP} untuk kategori keberatan = 0,3205128

Hasil proses klasifikasi kategori menunjukkan bahwa kategori yang dipilih dari proses klasifikasi untuk data komplain yang masuk adalah kategori keberatan dengan nilai V_{MAP} yang maksimal sebesar 0,3205128.

Pada aplikasi proses klasifikasi kategori dapat ditunjukkan pada :

- Detail $P(w_k | v_j)$ masing-masing kata pada kategori tertentu gambar 9, sebagai berikut :

KATEGORI

Menggunakan Proses Analizing

$P(\text{permohonan} \mid \text{KELUHAN}) = 0,0285714285714286$
 $P(\text{peninjauan} \mid \text{KELUHAN}) = 0,0285714285714286$
 $P(\text{ulang} \mid \text{KELUHAN}) = 0,0285714285714286$
 $P(\text{baik} \mid \text{KELUHAN}) = 0,0285714285714286$
 $P(\text{permohonan} \mid \text{SARAN}) = 0,0333333333333333$
 $P(\text{peninjauan} \mid \text{SARAN}) = 0,0333333333333333$
 $P(\text{ulang} \mid \text{SARAN}) = 0,0333333333333333$
 $P(\text{baik} \mid \text{SARAN}) = 0,0666666666666667$
 $P(\text{permohonan} \mid \text{KEBERATAN}) = 0,0769230769230769$
 $P(\text{peninjauan} \mid \text{KEBERATAN}) = 0,0769230769230769$
 $P(\text{ulang} \mid \text{KEBERATAN}) = 0,0769230769230769$
 $P(\text{baik} \mid \text{KEBERATAN}) = 0,0256410256410256$

Gambar 9 Detail $P(w_k \mid v_j)$ Klasifikasi Kategori

Sedangkan hasil nilai V_{MAP} pada aplikasi dapat dilihat pada gambar 10, sebagai berikut :

Detail Perhitungan Kategori
$V_{MAP}(\text{KELUHAN}) = 0,104761904761905$ $V_{MAP}(\text{SARAN}) = 0,0833333333333333$ $V_{MAP}(\text{KEBERATAN}) = 0,32051282051282 >> V_{MAP} \text{ Maksimal}$

Gambar 10 Detail V_{MAP} Klasifikasi Kategori

Sedangkan untuk klasifikasi berdasarkan divisi terkait dapat ditunjukkan sebagai berikut:

Terminal

$P(v_j \text{ divisi terminal}) = 0,8$

$P(w_k \mid v_j)$ masing-masing kata pada divisi terminal :

$P(\text{"penjualan"} \mid \text{terminal}) = 0,125$

$P(\text{"pandu"} \mid \text{terminal}) = 0,0625$

Pelayanan Kapal

$P(v_j \text{ divisi pelayanan kapal}) = 0,8$

$P(w_k \mid v_j)$ masing-masing kata pada divisi pelayanan kapal :

$P(\text{"penjualan"} \mid \text{pelayanan kapal}) = 0,0625$

$P(\text{"pandu"} \mid \text{pelayanan kapal}) = 0,375$

Sehingga perhitungan V_{map} dengan rumus 6 untuk masing-masing divisi diperoleh :

V_{MAP} untuk divisi terminal = 0,15

V_{MAP} untuk divisi pelayanan kapal = 0,35

Hasil proses klasifikasi divisi menunjukkan bahwa divisi yang dipilih dari proses klasifikasi untuk data komplain yang masuk adalah divisi pelayanan kapal dengan nilai V_{MAP} yang maksimal sebesar 0,35.

Pada aplikasi proses klasifikasi divisi dapat ditunjukkan pada :

- Detail $P(w_k \mid v_j)$ masing-masing kata pada divisi tertentu gambar 11, sebagai berikut :

DIVISI
Menggunakan Proses Analizing
$P(\text{penjualan} \mid \text{Pelayanan Kapal}) = 0,0625$ $P(\text{pandu} \mid \text{Pelayanan Kapal}) = 0,375$ $P(\text{penjualan} \mid \text{Terminal}) = 0,125$ $P(\text{pandu} \mid \text{Terminal}) = 0,0625$

Gambar 11 Detail $P(w_k \mid v_j)$ Klasifikasi Divisi

Sedangkan hasil nilai V_{MAP} pada aplikasi dapat dilihat pada gambar 12, sebagai berikut :

Detail Perhitungan Divisi
$V_{MAP}(\text{Pelayanan Kapal}) = 0,35 >> V_{MAP} \text{ Maksimal}$ $V_{MAP}(\text{Terminal}) = 0,15$

Gambar 12 Detail V_{MAP} Klasifikasi Kategori

SIMPULAN

Dari pendefinisian masalah serta analisa dan pembuatan perangkat lunak pengklasifikasian komplain / *complaint center* dengan menggunakan Naïve Bayes Classifier, maka dapat diambil beberapa kesimpulan sebagai berikut :

1. Pada hasil uji coba perangkat lunak dengan menggunakan data komplain, menunjukkan bahwa prototipe sistem pengklasifikasian komplain layanan operasional jasa kepelabuhanan dapat mengklasifikasikan komplain secara cepat dan tepat sehingga dapat digunakan sebagai acuan dalam pengembangan sistem pengklasifikasian komplain pada PT.(Persero) Pelabuhan Indonesia III Cabang Tanjung Perak Surabaya.
2. *Naïve Bayes Classifier* dapat diterapkan untuk mengklasifikasikan komplain layanan operasional pengguna jasa kepelabuhanan menjadi kategori-kategori yang ditentukan oleh PT. (Persero) Pelabuhan Indonesia III Cabang Tanjung Perak Surabaya.
3. *Naïve Bayes Classifier* dapat diterapkan untuk mengklasifikasikan komplain layanan operasional pengguna jasa kepelabuhanan berdasarkan jenis divisi-divisi yang telah ditentukan oleh PT. (Persero) Pelabuhan Indonesia III Cabang Tanjung Perak Surabaya.
4. Algoritma pembobotan TF-IDF dan *vector space model (cosine similarity)* digunakan untuk menunjukkan kemiripan antar dokumen yang merupakan representasi dari kalimat komplain layanan operasional pengguna jasa kepelabuhanan sehingga dapat digunakan untuk mendukung penerapan algoritma *Naive Bayes Classifier* dalam mengklasifikasikan data komplain layanan operasional yang berupa data teks yang tidak terstruktur dengan tingkat akurasi yang cukup tinggi.

Beberapa saran yang dapat digunakan sebagai ide untuk pengembangan dalam penelitian yang akan datang adalah sebagai berikut :

1. Perlu adanya riset lebih mendalam untuk perbandingan dengan algoritma lain untuk membandingkan hasil klasifikasi sehingga diperoleh hasil yang lebih optimal.
2. Menerapkan atau mengimplementasikan *xml web service* pada aplikasi *complaint center*.
3. Memberikan fasilitas *digital signature* untuk menguji keaslian data komplain yang dikirimkan melalui website *complaint center*.

DAFTAR RUJUKAN

Harlian, Milkha. 2006. *Machine Learning Text Categorization*. University of Texas, Austin.
Jiawei, Micheline K., 2007. *Data Mining: Concepts and Techniques, Second Edition*, University of Illinois, Urbana-Champaign.

Leyla Khodra, Masayu. 2007. *Text Mining Kategorisasi Teks Naïve Bayes*, (Online), (<http://matematika.upi.edu/staff/masayu>, diakses 08 April 2008).

Rachli, Muhamad. 2007. *Email Filtering Menggunakan Naïve Bayesian*, (Online), (<http://matematika.upi.edu/staff/rachli>, diakses 10 April 2008).

Vincent, Kelly P. 2005. *Text Mining Methods for Event Recognition in Stories*, Knowledge Media Institute.

Wibisono, Yudi. 2005. *Klasifikasi Berita Berbahasa Indonesia menggunakan Naïve Bayes Classifier*, (Online), (<http://matematika.upi.edu/staff/yudi>, diakses 08 April 2008).