

Implementasi Algoritma *Naive Bayes Classifier* dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan *Call Center 110*

Fitri Handayani dan Feddy Setio Pribadi

Jurusan Teknik Elektro, Fakultas Teknik, Universitas Negeri Semarang, Indonesia
v3handai@gmail.com

Abstrak— Kepolisian Negara Republik Indonesia merupakan salah satu lembaga yang menyediakan layanan pelaporan dan pengaduan melalui call center 110 dengan mengirimkan teks singkat kepada staff terdekat dengan Tkp, kasus yang dilaporkan atau diadukan akan segera ditinjau dan ditindak lanjut. Banyaknya data yang masuk ke dalam sistem dengan pengklasifikasian kasus secara manual membuat petugas kesulitan dalam pengolahan data, oleh karena itu dibutuhkan sistem klasifikasi teks pelaporan dan pengaduan melalui layanan 110. Pada penelitian ini metode yang digunakan dalam pengklasifikasi teks otomatis menggunakan metode Naive bayes Classifier. Metode Naive Bayes Classifiers yaitu salah satu metode klasifikasi teks berdasarkan probabilitas kata kunci dalam membandingkan dokumen latih dan dokumen uji. Keduanya dibandingkan melalui beberapa tahap persamaan, yang akhirnya diperoleh hasil probabilitas tertinggi yang ditetapkan sebagai kategori dokumen baru. Hasil penelitian yang dilakukan oleh peneliti yaitu pengklasifikasian teks otomatis pelaporan dan pengaduan masyarakat dengan menggunakan metode Naive bayes Classifiers menghasilkan rata-rata akurasi yang tinggi, yaitu recall 93%, precision 90 %, dan f-measure 92%.

Kata kunci— *Mining, Call center 110, Naive Bayes classifier, Kepolisian Negara Republik Indonesia*

I. PENDAHULUAN

Dewasa ini dunia teknologi informasi dan komunikasi berkembang sangat pesat, ini menyebabkan bertambahnya data atau file elektronik yang merupakan hasil pertukaran informasi *user* di seluruh dunia. Pertambahan data yang sangat cepat menyebabkan penumpukan data yang belum termanfaatkan yang dianggap sebagai *spam*, padahal dari sekumpulan data tersebut dapat diperoleh informasi yang bermanfaat untuk digali. Salah satu ilmu yang menangani hal ini yakni *data mining*. Data mining memiliki beberapa cabang ilmu yang dapat menggali data sesuai dengan konteks data yang ada, salah satunya adalah *Teks mining*. *Text mining* adalah proses menemukan informasi dalam koleksi dokumen, dan mengidentifikasi secara otomatis pola yang terbentuk, dan berhubungan dengan informasi yang didapat dari kumpulan data yang tidak terstruktur [1]. Teks mining menggali data tekstual yakni berupa data digital yang terdiri dari term-term membentuk suatu dokumen. Layanan 110 Polri merupakan salah satu layanan yang disediakan oleh Kepolisian Negara Republik Indonesia untuk membantu permasalahan menyangkut ketertiban dan keamanan masyarakat melalui jaringan telekomunikasi yang berpusat di Markas Besar Kepolisian Negara Republik Indonesia yang berada di Jakarta selatan. Layanan ini telah diintegrasikan dengan seluruh jajaran polres/polresta, polair (Polisi Air), piket Reskrim se-indonesia untuk membantu menangani pengaduan dan pelaporan masyarakat setiap daerah. Mekanisme pelayanan yakni

pertama pelapor menekan tombol darurat pada handphone maupun telepon rumah dengan nomor darurat 110 kemudian akan tersambung dengan petugas pusat kemudian petugas pusat mencatat pelaporan serta mengklasifikasikan laporan dan aduan sesuai dengan *locus delicti* atau TKP (tempat kejadian perkara) kemudian petugas mengirimkan teks ke jajaran petugas di daerah terdekat dengan TKP. Petugas Daerah mengklasifikasikan kasus dan mencatatnya pada buku pelaporan dan pengaduan masyarakat secara manual. Kegiatan ini menyulitkan dan membuang waktu petugas dalam klasifikasi pelaporan dan pencarian informasi ketika laporan dibutuhkan untuk keperluan administrasi maupun keperluan lainnya. Klasifikasi Teks Otomatis merupakan salah satu solusi untuk menangani masalah ini. Klasifikasi teks merupakan proses menemukan kesamaan dalam dokumen, corpus, maupun kelompok-kelompok dari dokumen yang telah dilabeli sebelumnya (*supervised learning*), berdasarkan topik, tema yang ditunjukkan oleh koleksi dokumen [2]. Salah satu metode klasifikasi yang dikenal simple akan tetapi memiliki akurasi yang cukup tinggi yakni naive bayes classifiers. Tujuan Penelitian ini yakni mengimplementasikan metode NBC kedalam sistem klasifikasi teks otomatis pengaduan dan pelaporan masyarakat melalui layanan 110 dan mengetahui tingkat akurasi sistem.

II. TINJAUAN PUSTAKA

Penelitian tentang klasifikasi teks menggunakan metode naive bayes classifier sudah banyak dilakukan sebelumnya diantaranya adalah Dyta Anggraeni dari Fakultas Ilmu Komputer Universitas Indonesia dalam penelitiannya yang berjudul “Klasifikasi topik berbasis *Naive Bayes* dan *Maximum Entropy* pada Artikel Media Massa dan Abstrak Tulisan Ilmiah”, hasil penelitiannya menyatakan bahwa nilai akurasi tertinggi didapat pada saat menggunakan metode Naive Bayes dengan informasi fitur *frequency – normalized* yaitu 95,73%. Penelitian dengan menggunakan metode klasifikasi teks yang sama dilakukan oleh Mohammed EK, dan Tajje-Eddine R dari Universitas Alakhawyn, Maroko dalam penelitiannya yang berjudul “Pengkategorian dokumen berbahasa Arab menggunakan Algoritma Naive Bayes”, dan hasil penelitiannya menyatakan bahwa dengan data set 300 dan jumlah 2000 term dan 5 kategori dalam pengklasifikasiannya didapatkan rata-rata akurasi 68,78% *recall*, 92,8 % *precision*, 90% *F-measure* [3].

A. Klasifikasi Teks

Klasifikasi teks adalah proses klasifikasi teks berdasarkan kata, frase maupun kombinasinya dengan baik untuk menentukan kategori yang telah ditetapkan sebelumnya. Pengolahan klasifikasi teks melibatkan dua proses utama, yakni pertama ekstraksi fitur yang menjadi kata kunci yang efektif dalam tahap pelatihan atau *training* dan kemudian proses kedua yakni klasifikasi dokumen setelah melalui tahap uji atau *testing* [4].

B. Teks Preprocessing

Persiapan Dokumen teks (*Text Preprocessing*) merupakan tahapan proses yang sangat dalam menentukan kualitas tahap selanjutnya yakni tahap klasifikasi. Tahap ini penting dalam memilih kata kunci yang signifikan dan membuang kata-kata yang tidak berkontribusi dalam membedakan antar dokumen. Tahap *Preprocessing* menkonversi data tekstual dalam struktur data mining. Berikut langkah-langkah *text preprocessing* akan digambarkan pada Gambar 1.

1) Extraction

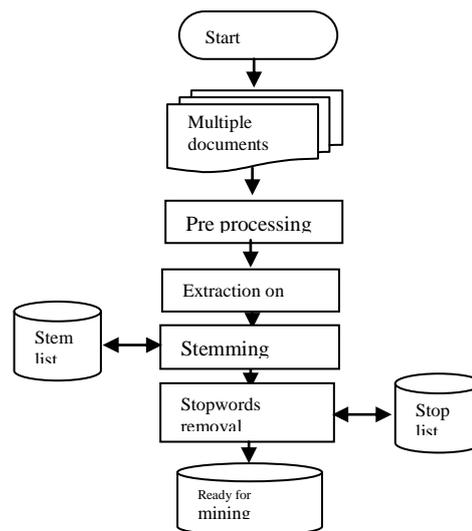
Metode ini digunakan untuk *tokenize* atau memisah isi kalimat menjadi kata-kata individu atau biasa disebut dengan term.

2) Stopword removal

Metode ini merupakan cara menghilangkan kata-kata yang tidak memiliki kontribusi dalam pengklasifikasian teks. Kata – kata yang akan dihilangkan telah disimpan pada ruang stop list.

3) Stemming

Metode ini merupakan pemisahan antar imbuhan baik awalan, akhiran dengan kata dasar. Seluruh kata-kata yang akan dihilangkan telah disimpan pada ruang stem list, data akan dicocokkan dengan stem list. Data yang sama dengan *stem list* akan dianggap *noisy* dan hilang. Kata dasar akan disimpan didalam *database*.



Gambar 1. Tahap teks preprocessing

C. Metode Naive Bayes Classifier

Pengklasifikasi bayes merupakan salah satu pengklasifikasi statistik, dimana pengklasifikasi ini dapat memprediksi probabilitas keanggotaan kelas suatu data tuple yang akan masuk ke dalam kelas tertentu, sesuai dengan perhitungan probabilitas. Pengklasifikasi Bayes didasari oleh teorema bayes yang ditemukan oleh Thomas Bayes pada abad ke-18. Dalam studi perbandingan algoritma klasifikasi telah ditemukan simple bayesian atau yang biasa dikenal dengan *Naive Bayes classifier*. *Naive Bayes classifier* menunjukkan akurasi dan kecepatan yang tinggi bila diterapkan pada database yang besar [5]. Metode ini sering digunakan dalam menyelesaikan masalah dalam bidang mesin pembelajaran karena metode ini dikenal memiliki tingkat akurasi yang tinggi dengan perhitungan sederhana [6].

Teorema bayes merupakan dasar aturan dari *naive bayes classifier* berikut teorema bayes akan disajikan pada persamaan (1).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Dimana X merupakan data *tuple* hasil pengujian dari suatu set data yang telah ditentukan masuk ke ke dalam kelas tertentu. H merupakan suatu hipotesis yang akan menentukan X masuk ke dalam kelas C. P(H|X) merupakan peluang atau probabilitas X yang merupakan data *tuple* atau bukti yang diperoleh pada saat observasi masuk ke dalam kelas C, dengan kata lain mencari probabilitas X dimiliki oleh kelas C. P(H|X) merupakan probabilitas posterior, H dikondisikan pada X. Sebaliknya P(H) merupakan probabilitas prior, atau probabilitas sebelumnya. Kemudian P(X|H) merupakan probabilitas posterior dimana X dikondisikan pada H. Sedangkan P(X) merupakan probabilitas sebelumnya dari X [5].

Dengan aturan Bayes maka penelitian ini akan mengimplementasikan aturan bayes dengan studi kasus tertentu oleh karena itu aturan bayes dapat dinyatakan :

$$P(c_j|X) = \frac{p(X|c_j)p(c_j)}{p(X)} \quad (2)$$

Dimana c_j adalah kategori teks yang akan diklasifikasikan, dan $p(c_j)$ merupakan probabilitas prior dari kategori teks c_j . Sedangkan d merupakan dokumen teks yang direpresentasikan sebagai himpunan kata (W_1, W_2, \dots, W_n), dimana W_1 adalah kata pertama, W_2 adalah kata kedua dan seterusnya.

Pada saat proses pengklasifikasian dokumen teks, maka pendekatan Bayes akan memilih kategori yang memiliki probabilitas paling tinggi (C_{MAP}) yaitu :

$$C_{MAP} = \operatorname{argmax} \frac{p(c_j)p(X|c_j)}{p(X)} \quad (3)$$

Nilai $p(X)$ dapat diabaikan karena nilainya adalah konstan untuk semua c_j , sehingga persamaan (3) dapat dituliskan :

$$C_{MAX} = \operatorname{argmax} p(c_j) p(X|c_j) \quad (4)$$

Probabilitas $p(c_j)$ dapat diestimasi dengan menghitung jumlah dokumen training pada setiap kategori c_j . Sedangkan untuk menghitung distribusi $p(X|c_j)$ akan sulit karena jumlah term menjadi sangat besar. Hal ini disebabkan jumlah term tersebut sama dengan jumlah semua kombinasi posisi kata dikalikan dengan jumlah kategori yang akan diklasifikasikan.

Dengan pendekatan *Naïve Bayes* yang mengasumsikan bahwa setiap kata dalam setiap kategori adalah tidak bergantung satu sama lain, maka perhitungan dapat lebih disederhanakan dan dapat dituliskan sebagai berikut:

$$p(X|c_j) = \prod_{i=1}^n p(w_i|c_j) \quad (5)$$

Dengan menggunakan persamaan (2), maka persamaan (5) dapat dituliskan menjadi :

$$C_{MAP} \operatorname{argmax} p(c_j) \prod_{i=1}^n p(w_i|c_j) \quad (6)$$

Nilai $p(c_j)$ dan $p(w_i|c_j)$ dihitung pada saat proses pelatihan dimana persamaannya adalah sebagai berikut :

$$p(c_j) = \frac{|\text{docs } j|}{|\text{contoh}|} \quad (7)$$

$$p(w_i|c_j) = \frac{1+n_i}{|C|+n(\text{kosakata})} \quad (8)$$

- $p(w_i|c_j)$ = probabilitas kata w_i pada kategori c_j
- $|\text{docs } j|$ = jumlah dokumen pada kategori j
- $|\text{contoh}|$ = jumlah seluruh dokumen sampel yang digunakan dalam proses training
- n_i = frekuensi kemunculan kata w_i pada kategori c_j
- $|C|$ = jumlah semua kata pada kategori c_j
- $n(\text{kosakata})$ = jumlah kata yang unik pada semua data training.

D. Performance Measure

Performance measure merupakan tahapan terakhir dalam klasifikasi teks. Pada tahap ini akan mengevaluasi hasil percobaan, membandingkan dan menganalisis terhadap kinerja klasifikasi teks. Banyak cara pengukuran yang telah digunakan, seperti *Precision* dan *Recall*, error, akurasi, dan lainnya. Pada penelitian ini metode evaluasi yang digunakan adalah *Recall*, *Precision*, dan *F-measure*.

Pengevaluasian hasil klasifikasi dokumen pada tiap kategori ditunjukkan pada Tabel 1.

TABEL I. TABEL CONFUSION MATRIX

Category C_j	Expert judgements		
	YES	TP	NO
Classifier judgements	YES	TP	FN
	NO	FP	TN

Tabel 1 menunjukkan bahwa hasil klasifikasi benar sesuai dengan keputusan sistem (TP) dan Klasifikasi salah menurut sistem (FP). Sedangkan dokumen yang tidak termasuk dalam hasil klasifikasi kategori, adakalanya memang bukan anggota dari kategori itu (TN) dan adakalanya ternyata seharusnya menjadi anggota kategori tersebut (FN).

Keempat parameter di atas digunakan untuk menghitung 3 metode evaluasi yakni :

- 1) *Recall*, yaitu perbandingan jumlah dokumen yang relevan terkenal dengan jumlah seluruh dokumen relevan. *Recall* memiliki rumusan sebagai berikut :

$$Recall = \frac{TP}{TP+FP} \quad (9)$$

- 2) *Precision* yaitu perbandingan jumlah dokumen yang relevan terkenal dengan jumlah dokumen yang terkenal. *Precision* memiliki persamaan sebagai berikut :

$$Precision = \frac{TP}{TP+FN} \quad (10)$$

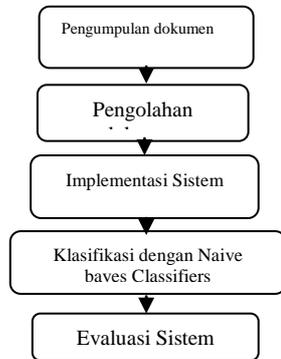
- 3) *F - measure* merupakan nilai yang mewakili seluruh kinerja sistem yang merupakan penggabungan nilai *Recall* dan *Precision*. *F-measure* memiliki persamaan sebagai berikut :

$$F - Measure = \frac{2PR}{P+R} \quad (11)$$

Nilai *recall*, *precision* dan *F-measure* dinyatakan dalam persen. Semakin tinggi persentase ketiga nilai tersebut menunjukkan semakin baik kinerja sistem klasifikasi teks otomatis.

III. METODE

Penelitian ini dilaksanakan melalui beberapa tahap yaitu pengumpulan data, pengolahan data, implementasi sistem, klasifikasi teks dengan metode *naive bayes classifier*, evaluasi klasifikasi dan melakukan perbandingan antara sistem dan manual. Tahap-tahap yang akan dilakukan dalam penelitian ini diilustrasikan pada Gambar 2.



Gambar 2. Metode penelitian

A. Pengumpulan Dokumen

Dokumen yang digunakan dalam penelitian ini yakni dokumen catatan laporan dan aduan masyarakat selama 2 tahun 2013-2014 yang diperoleh dari hasil observasi peneliti di SPKT Polrestabes Semarang .

Dokumen telah diklasifikasikan manual oleh petugas polrestabes semarang menurut aturan yang telah diatur dalam Peraturan Kapolri nomor 7 dan 9. Dokumen Laporan dan Aduan Masyarakat tersebut diklasifikasikan ke dalam tiga kategori yaitu :

- 1) Kategori Informasi yang berisi laporan dan aduan yang bersifat informatif
- 2) Kategori Kamtibmas yang berisi laporan dan aduan yang bersifat mengganggu ketertiban dan keamanan masyarakat.
- 3) Kategori Tindak Pidana yang berisi laporan dan aduan yang bersifat menjeruk unsur tindak pidana.

Sebanyak 113 dokumen yang diperoleh dari observasi dipolrestabes semarang dari seluruh dokumen akan dibagi menjadi 70% sebagai dokumen latih dan 30% sebagai dokumen uji.

B. Implementasi Sistem

Implementasi sistem menggunakan bahasa pemrograman PHP dan MySQL sebagai tempat penyimpanan data.

1) Teks *Preprocessing*

Sebelum melakukan tahap klasifikasi dilakukan tahap praproses teks yakni proses reduksi teks yang bertujuan membuang kata-kata atau term-term yang tidak memiliki kontribusi atau bobot yang mempengaruhi tahap selanjutnya. Praproses teks diantaranya yakni *extraction* teks, *stemming*, *stopword removal* akan menghasilkan *query* yang akan digunakan pada tahap selanjutnya.

2) Klasifikasi dengan *Naive bayes classifier*

Dalam metode *Naive bayes classifier* dilakukan proses pengklasifikasian teks berdasarkan data latih yang sebelumnya telah ada didalam database sistem. Penelitian terkait *naive bayes classifiers* melalui beberapa tahap

yakni *prior probability*, *conditional probability*, pemilihan kategori akan ditentukan melalui nilai maksimum kategori terpilih [5].

3) Evaluasi Klasifikasi

Evaluasi dilakukan dengan cara membandingkan antara hasil klasifikasi *naive bayes classifier* dan hasil klasifikasi manual oleh petugas spkt polrestabes semarang serta dihitung akurasi dengan menggunakan *confussion matrix*.

IV. HASIL DAN PEMBAHASAN

Hasil Uji coba yang dilaksanakan oleh peneliti akan dibahas secara jelas pada bab ini. Uji Coba yang dilakukan pada 33 dokumen mengalami seleksi kembali dan peneliti menemukan kesalahan pada hasil klasifikasi oleh petugas karena ketidakteelitian petugas dalam menulis kategori laporan. Dari 33 dokumen yang diujikan terdapat 4 dokumen yang tidak terkategori dengan benar seacara manual. Oleh karena itu dokumen uji yang semula berjumlah 33 dokumen menjadi 29 dokumen, 10 dokumen akan dibahas secara lebih rinci untuk dibandingkan antara hasil klasifikasi manual dan hasil klasifikasi sistem. Hasil Pengujian akan disajikan pada Tabel 2 dan Tabel 3.

TABEL II. HASIL KLASIFIKASI MANUAL

NO	Isi Teks	Kategori
1	Dokumen 1	Kamtibmas
2	Dokumen 2	Informasi
3	Dokumen 3	Informasi
4	Dokumen 4	Informasi
5	Dokumen 5	Informasi
6	Dokumen 6	Informasi
7	Dokumen 7	Tindak Pidana
8	Dokumen 8	Informasi
9	Dokumen 9	Informasi
10	Dokumen 10	Kamtibmas

TABEL III. HASIL KLASIFIKASI SISTEM

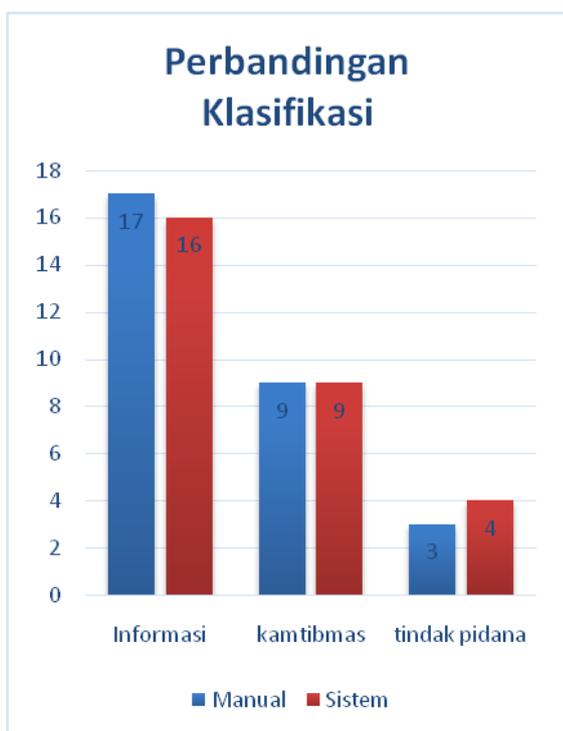
NO	Isi Teks	Kategori
1	Dokumen 1	Kamtibmas
2	Dokumen 2	Informasi
3	Dokumen 3	Tindakpidana
4	Dokumen 4	Informasi
5	Dokumen 5	Informasi
6	Dokumen 6	Informasi
7	Dokumen 7	Tindak Pidana
8	Dokumen 8	Tindakpidana
9	Dokumen 9	Informasi
10	Dokumen 10	Kamtibmas

A. Hasil Klasifikasi Manual

Tabel 2 merupakan hasil klasifikasi manual yang dilakukan oleh petugas berdasarkan peraturan yang mengatur kategorisasi dalam layanan 110. Kasus yang berupa teks singkat akan diklasifikasikan berdasarkan golongan kasus kejahatan yang telah dijelaskan pada bab sebelumnya. Kategori – kategori yang terdapat dalam sistem yakni meliputi Tindak pidana, Informasi, dan Kamtibmas (Keamanan ketertiban masyarakat). Tindak pidana meliputi kasus-kasus tergolong kejahatan I, sedangkan kategori Informasi meliputi kejadian-kejadian yang mengandung makna memberi informasi kepada petugas, dan kategori Kamtibmas meliputi kejadian / kasus yang meresahkan msasyarakat dan perlu adanya pembinaan terhadap pelaku kejahatan.

B. Hasil Klasifikasi Sistem

Hasil klasifikasi sistem yang direpresentasikan melalui tabel 3 diperoleh informasi bahwa dari 10 dokumen uji didapatkan 5 dokumen berkategori informasi, 3 dokumen berkategori kamtibmas, 2 dokumen berkategori tindak pidana. Hasil sistem tersebut diperoleh melalui beberapa tahap yakni tahap *text preprocessing*, pembobotan dan klasifikasi teks dengan menggunakan Algoritma naïve bayes yang melalui beberapa proses . Proses klasifikasi naïve bayes melalui 3 proses yaitu penghitungan peluang *priors*, penghitungan peluang *conditional* , dan pemilihan kategori. Hasil dari pemilihan kategori pada tahap klasifikasi teks menunjukkan hasil nilai maksimum peluang untuk pengkategorian dokumen baru dan terpilih kategori kamtibmas sebagai kategori dokumen *testing* yang masuk ke dalam sistem sebagai dokumen ke 11.



Gambar 3. Grafik perbandingan klasifikasi

C. Evaluasi Sistem

Evaluasi digunakan untuk mengukur kinerja suatu sistem, dalam penelitian ini evaluasi digunakan untuk mengukur keakuratan merode klasifikasi teks. Teknik evaluasi yang biasa digunakan untuk mengukur akurasi metode klasifikasi teks diantaranya *precision* ,*recall*, dan *f-measure*.

Teknik evaluasi hasil klasifikasi teks pada tiap kategori menggunakan standar tabel 1. Tabel 1 berisi setiap kemungkinan hasil pada tiap kategori. Berdasarkan Tabel 1 diperoleh persamaan *precision*, *recall*, dan *F-measure*. Penghitungan akurasi sistem berdasarkan informasi diatas diperoleh hasil yang dapat dilihat pada Tabel 4.

TABEL IV. PENGUKURAN KEAKURATAN KLASIFIKASI TEKS PADA SISTEM DENGAN DATA SEBANYAK 29 DATA TESTING

Measure	Formula	Result	Percent (%)
Recall	$\frac{TP}{TP + FP}$	$\frac{27}{27 + 2}$	93
Precision	$\frac{TP}{TP + FN}$	$\frac{27}{27 + 2}$	93
F-Measure	$\frac{2XPXP}{R + P}$	$\frac{2X90X93}{90 + 93}$	92
Rata – rata akurasi			92,67

Berdasarkan tabel 4 dinyatakan bahwa sistem menunjukkan nilai akurasi yang tinggi, oleh karena itu sistem dapat dinyatakan layak digunakan. Rata-rata akurasi dari beberapa jumlah data *training* dan data *testing* yang berbeda dapat dilihat pada tabel 5.

TABEL V. AKURASI SISTEM

Dokumen latih	Dokumen uji	Akurasi (%)
79	29	82 %
39	17	76 %
20	8	67 %

Berdasarkan hasil rata-rata akurasi yang ditunjukkan pada tabel 5, diperoleh informasi yakni dengan data latih sebanyak 79 dokumen dan data uji sebanyak 33 dokumen dengan dokumen yang relevan sebanyak 29 dokumen diperoleh rata-rata akurasi 82% , dan seterusnya maka dapat disimpulkan bahwa jumlah data *testing* mempengaruhi nilai akurasi, semakin banyak jumlah data testing semakin besar nilai akurasi sistem. Hasil pengujian evaluasi menunjukkan bahwa akurasi sistem dalam pengklasifikasian teks pelaporan dan pengaduan masyarakat melalui 110 dinyatakan baik dan akurat ditunjukkan dengan hasil pengukuran akurasi *precision* 93%, *recall* 93%, *F-measure* 92 % dan nilai rata-rata akurasi untuk 29 data *testing* sebesar 92,67 %.

Proses pengujian yang serupa telah lebih dahulu dilakukan oleh Sandi F., dan Edi W dalam penelitiannya yang berjudul Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification. Pada penelitiannya peneliti berpendapat bahwa penelitian evaluasi hasil uji coba peneliti berpendapat bahwa keakuratan suatu sistem akan dihasilkan lebih baik ketika data yang diuji cobakan dalam jumlah yang banyak, hasil uji coba aplikasi menunjukkan bahwa nilai akurasi terkecil sebesar 72% untuk 100 sampel dan nilai akurasi tertinggi yakni 93,58% untuk 13106 sampel.

V. SIMPULAN

Berdasarkan hasil penelitian yang diperoleh, dapat ditarik beberapa kesimpulan, yaitu pertama, jumlah kategori yang ada mempengaruhi kinerja klasifikasi teks menggunakan metode Naive Bayes. Tingkat kemiripan diantara kategori satu dengan yang lain mempengaruhi tingkat akurasi klasifikasi teks. Jika tingkat kemiripan diantara dua kategori tinggi, maka akan sulit membedakan kedua kategori tersebut sehingga tingkat akurasi klasifikasi teks akan menurun. Kedua, penggunaan stopwords dan stemming dapat meningkatkan tingkat akurasi klasifikasi teks.

Klasifikasi teks pengaduan dan pelaporan masyarakat dengan 79 dokumen latih dan 29 data uji diperoleh 27 dokumen relevan dan 2 dokumen tidak relevan. Klasifikasi

Teks menggunakan metode *Naive Bayes Classifier* mencapai tingkat akurasi untuk *recall* sebesar 93%, *precision* sebesar 93%, *f-measure* sebesar 92%, dan rata-rata akurasi sebesar 92,67%. Akurasi yang cukup tinggi dapat disimpulkan bahwa sistem pengklasifikasian teks pelaporan dan pengaduan masyarakat melalui layanan 110 ini dapat diimplementasikan dengan menggunakan metode *Naive Bayes Classifier*

REFERENSI

- [1] Feldman, R., and J. Sanger. 2007. *The Text Mining Handbook : Advanced Approaches in Analysing Unstructured Data*. New York : Cambridge University Press
- [2] Chakraborty et al. 2013. *Text Mining and Analysis, Practical Methods, Examples and Case Studies Using SAS*. North Carolina : SAS Institute Inc.
- [3] El Kourdi et al. 2004. *Automatic Arabic Document Categorization Based on the Naïve Bayes Algorithm*. Alakhawayn University, Morocco, <http://www.aclweb.org/anthology/W04-1610.pdf>, 2 Mei 2015
- [4] S, Menaka and Radha, Desember 2013. *Text Classification using keyword extraction Technique*. *International Journal of Advanced Research in Computer Science and Software Engineering*, vol 3, Issue 12, ISSN : 2277 128 X, http://www.ijarcsse.com/docs/papers/Volume_3/12_December2013/V3I12-0271.pdf
- [5] Han, J., and M. Kamber. 2006. *Data Mining Concept and Techniques Second Edition*. San Francisco : Elsevier.
- [6] Aggarwal, C., and C. Zhai. 2012. *Mining Text Data Chapter A Survey of Text Classification Algorithms*. London : Kluwer Academics Publisher.