JURNAL KILAT
Vol. 7, No. 2, September 2018, P-ISSN 2089-1245, E-ISSN 2655-4925

KLASIFIKASI PESAN GANGGUAN PELANGGAN MENGGUNAKAN METODE *NAIVE BAYES CLASSIFIER*

[1]Haryono, [2]Pritasari Palupiningsih, [3]Yessy Asri, [4]Andi Nikma Sri Handayani
[2,3,4]Majoring Informatics Engineering, Sekolah Tinggi Teknik PLN, Jl. Lingkar Luar Duri Kosambi, Cengkareng, Jakarta barat, Indonesia 11750
[1]Majoring Informatics Engineering, STMIK Bani Saleh Bekasi, Jl. Mayor M. Hasibuan No 68, Margahayu, Bekasi Timur, Kota Bekasi, Jawa Barat, Indonesi 17113
[1]haryono.siradmk@gmail.com, [2]prita_palupi@gmail.com [3]yesfar2@gmail.com

ABSTRACT

The application of customer disturbance message classifiers is made because of the process of reporting the interruption by the customer must be done by selection of data disorders by one by the admin to be able to follow-up from the existing customer reports. Naive Bayes is one of machine learning methods that uses probability calculations where the algorithm takes advantage of probability and statistical methods that predict future probabilities based on past experience. The application of the naive bayes classifier method with text mining as the initial data processor of the disorder messaging application can be concluded that this study yields an accuracy of probability values of 95 percent and proves that the Naive Bayes method can be used to help classify interference messages sent by customers.

Keywords: *Distrubance message, Text Mining, Naive Bayes Classifier*

ABSTRAK

Aplikasi pengklasifikasi pesan gangguan pelanggan dibuat karena pada proses pengaduan gangguan oleh pelanggan haruslah dilakukan penyeleksian data gangguan satu persatu oleh admin untuk bisa melakukan tindak lanjut dari laporan pelanggan yang ada. Naive bayes merupakan salah satu metode machine learning yang menggunakan perhitungan probabilitas dimana algoritma ini memanfaatkan metode probabilitas dan statistik yang memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Penerapan metode naive bayes classifier dengan text mining sebagai pemroses data awal dari aplikasi pengklasifikasian pesan gangguan dapat disimpulkan bahwa penelitian ini menghasilkan akurasi dari nilai probabilitas sebesar 95% dan membuktikan bahwa metode Naive bayes dapat digunakan untuk membantu mengklasifikasikan pesan gangguan yang dikirimkan oleh pelanggan.

Kata Kunci: *Pesan gangguan, Text mining, Naive bayes classifier*

1. PENDAHULUAN

Saat ini, untuk penyampaian pengaduan baik berupa keluhan maupun gangguan, PLN (Persero) menyediakan berbagai macam cara yaitu dapat langsung melalui *costumer service* yang ada disetiap kantor PLN, sosial media, maupun *Call Center* 123 yang khusus menangani keluhan maupun gangguan dari pelanggan. Adapun alur dari pengaduan *Call Center* 123 adalah ketika pelanggan telah melakukan pengaduan, maka pada bagian *Call Center* 123 akan menyampaikan laporan gangguan tersebut berupa teks ke PLN area pelayanan Pelanggan. Bagian *comment center* yang bertempat di kantor distribusi pusat akan meneruskan pesan pengaduan ke setiap area-area menggunakan aplikasi AP2T (aplikasi pelayanan pelanggan terpusat). Namun pada kasus pengiriman pesan pengaduan, tiap admin AP2T disetiap bagian bersangkutan tidak dapat langsung menerima pesan gangguan perbagian, namun harus melakukan proses *filtering* manual dengan cara membuka dan membaca satu per satu pesan pengaduan tersebut agar mengetahui apakah pesan itu ditujukan untuk bagiannya atau tidak. Tentu saja hal ini akan membutuhkan waktu yang lebih banyak, mengingat setiap hari terdapat sangat banyak pesan pelanggan yang diterima dalam suatu daerah, sedangkan penanganan pengaduan pelanggan haruslah dilakukan dengan waktu yang cepat untuk mengefesienkan pekerjaan dan membantu memenuhi target untuk penanganan cepat pengaduan pelanggan.

Text Mining dapat diartikan sebagai suatu proses penemuan infomasi baru yang berguna dari sekumpulan dokumen yang sebelumnya belum ditemukan dengan melakukan proses penganalisaan data dalam jumlah yang besar, untuk kasus dokumen yang tidak terstuktur (*unstructured text)* penggunaan *text mining* sangatlah diperlukan. Pada *text mining* terbagi kedalam beberapa kelompok yaitu deskripsi, estimasi, prediksi, klasifikasi dan pengklusteran. Pada kelompok klasifikasi sendiri dapat digunakan dalam mengklasifikasikan pesan tidak terstuktur, seperti contohnya pesan dari laporan pengaduan pelanggan.

Pada proses pengklasifikasian digunakan

Metode *naive bayes classifier*, dimana pada metode ini memiliki beberapa kelebihan berupa kesederhanaan dalam komputasi, akurasi yang tinggi dan kecepatan dalam memproses *database* besar. Pada PLN khususnya PLN Distribusi area Cengkareng, pesan gangguan pelanggan biasanya ditujukan untuk 3 bidang yang paling sering menerima pesan gangguan dari pelanggan yaitu bagian Jaringan, bagian konstruksi, dan bagian Transaksi Energi sehingga dari hal tersebut maka data akan diklasifikasikan berdasar kategori pesan pengaduan pelanggan menjadi 3 kelas yaitu bidang Jaringan (B1), bidang Konstruksi (B2), dan bidang Transaksi Energi (B3).

1.1 Literatur Review

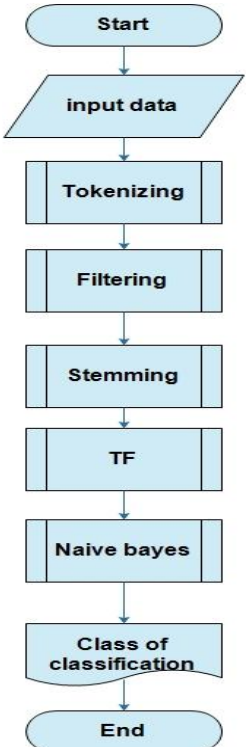
Hamzah A[1] mengklasifikasikan text dengan menggunakan metode naive bayes classifier untuk mengelompokan text berita dan text akademis dengan 1000 dokumen berita dan 450 dokumen abstrak akademis. Hasilnya pengklasifikasi dokumen berita memiliki akurasi lebih tinffi dibandingkan dengan dokumen akademik. Sehingga didapatkan kesimpulan bahwa semakin banyak data latih maka semakin besar akurasi.

Efendi, R et.al[2] mengimplementasikan *naive bayes classifier* untuk mengklasifikasi dokumen berbahasa Indonesia. Dimana pada tahap awal yaitu memproses data latih untuk membentuk nilai probabilitas kata kemudian menjalankan modul klasifikasi pada data uji. Hasil berdasarkan tingkat akurasi menunjukkan sebanyak 86,67 % .

Wongso, R et.al[3] melakukan klasifikasi text artikel dalam bahasa Indonesia dengan menentukan kombinasi metode klasifikasi dengan hasil terbaik. Pengolahan data dengan *text mining* dan kemudian melakukan klasifikasi dengan beberapa metode klasifikasi. Hasilnya ditemukan bahwa kombinasi TFIDF dan Multinomial Naive bayes memberikan hasil terbaik sebesar 98,4%.

Vijayarani, S et.al[6] mengimplementsikan *data mining* untuk memprediksi penyakit ginjal. Klasifikasi dibagi menjadi empat kelas dengan membandingkan metode klasifikasi *naive bayes* dan *space vector mechine* (SVM). Hasilnya disimpulkan untuk tingkat klasifikasi , SVM memberikan hasil terbaik, sedangkan Naive bayes memberikan waktu paling minimum untuk waktu eksekusi.

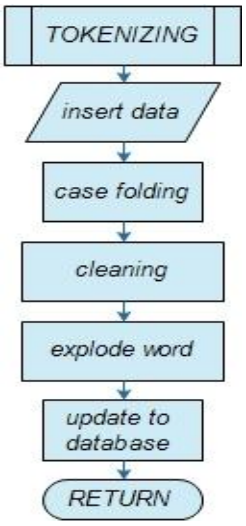
2. METODOLOGI PENELITIAN



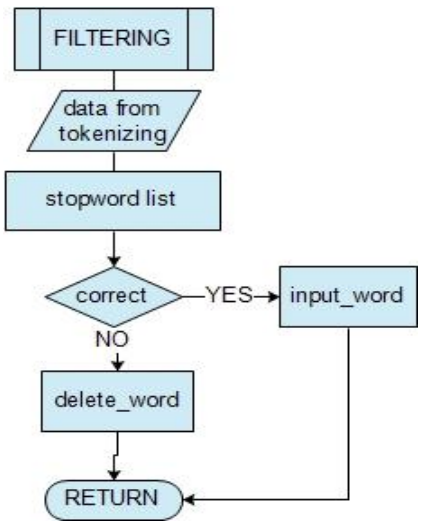
Gambar 1. Flowchart Metode

2.1 Reprocessing Data

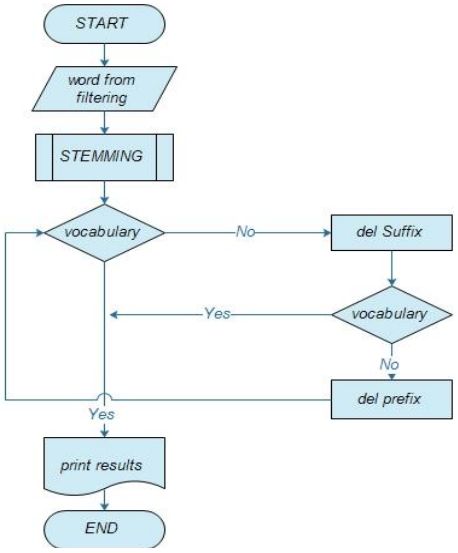
Adapun tahapan dari *reprocessing data* yaitu, Tahap *tokenizing, filtering, dan stemming*. Data yang diproses terlebih dahulu adalah data latih yang nantinya akan digunakan sebagai parameter klasifikasi. Tiap proses dari reprocessing data dapat dilihat dari *flowchart* dibawah ini:



Gambar 2. Tahap Tokenizing



Gambar 3. Tahap *Filtering*



Gambar 4. Tahap *Stemming*

Tahap diatas akan menghasilkan kata dasar yang selanjutnya akan diberikan bobot berdasarkan jumlah kemunculan dari kata tersebut .

2.2 *Term Frequency* (TF)

pemrosesan menghitung *Term Frequency (*TF), yakni menghitung frekuensi kemunculan kata dari setiap kata (token). Untuk dokumen tunggal tiap kata dianggap sebagai sebuah dokumen.

Tabel 1. Hasil penggabungan bobot dokumen tiap kelas

	Class	Mcb	App	Percik	ptl	Padam	persil	ganggu
TOTAL KATA	B1	1	1	1	2	2	1	1
	B2	0	0	0	2	2	0	2
	B3	1	4	0	1	2	0	0

Listrik	gardu	Dengar	ledak	Tiang	miring	Kabel	bawah	tanah
1	1	1	1	0	0	0	0	0
3	0	0	1	1	1	1	1	1
0	0	0	0	0	0	0	0	0

display	Tera	Error	Id	Rusak	blank	Tidak	input	token
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1

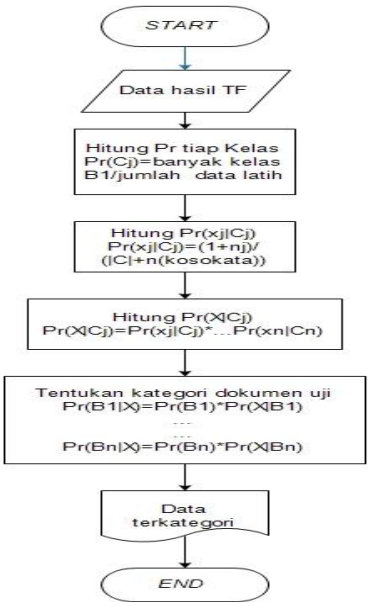
migrasi	**TOTAL**
0	13
0	15
1	18

Sehingga dapat diperoleh total dari setiap token pada keseluruhan data latih yaitu:

(Kosakata) = |B1|+|B2|+|B3| (3.1)
= 13 + 15 + 18
= 46

2.3 Naive bayes classifier

Naive Bayes memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Pada proses *training*, dokumen input telah diketahui kelasnya. Proses *training* berguna untuk membentuk pengetahuan berupa nilai probabilitas kata. Perlu diketahui bahwa pada proses *training*, tidak dijalankan modul klasifikasi, tetapi hanya menghasilkan dokumen yang mengandung kata untuk mengkarakteristik suatu kelas. Berikut adalah flowchart proses naive bayes:



Gambar 5. *Flowchart naive bayes*

3.3.1 Menghitung probabilitas untuk setiap kelas :

- Pr(B1) = $\frac{banyaknya\ kelas\ B1}{jumlah\ data\ training}$ (3.2)
  $= \frac{2}{6}$
  = 0,333
- Pr(B2) = $\frac{banyaknya\ kelas\ B2}{jumlah\ data\ training}$
  $= \frac{2}{6}$
  = 0,333
- Pr(B3) = $\frac{banyaknya\ kelas\ B3}{jumlah\ data\ training}$
  $= \frac{2}{6}$
  = 0,333

3.3.2 Menghitung probabilitas setiap kata (token) pada setiap kelas

$Pr(w_i|c_j) = \frac{n_i+1}{|C|+(kosokata)}$ (3.3)

- Term “mcb”
  Pr(mcb|B1)= $\frac{1+1}{13+46}$ = 0,034
  Pr(mcb|B2) = $\frac{1+0}{15+46}$ = 0,016
  Pr(mcb|B3) = $\frac{1+1}{18+46}$ = 0,031
- Term “app”
  Pr(app|B1) = $\frac{1+1}{13+46}$ = 0,034
  Pr(app|B2) = $\frac{1+0}{15+46}$ = 0,016
  Pr(app|B3) = $\frac{1+4}{18+46}$ = 0,078
- Term “percik”
  Pr(percik|B1) = $\frac{1+1}{13+46}$ = 0,034
  Pr(percik|B2) = $\frac{1+0}{15+46}$ = 0,016

Pr(percik|B3) = $\frac{1+0}{18+46}$ = 0,016

- Term "ptl"

Pr(ptl|B1) = $\frac{1+2}{13+46}$ = 0,051

Pr(ptl|B2) = $\frac{1+2}{15+46}$ = 0,049

Pr(ptl|B3) = $\frac{1+1}{18+46}$ = 0,031

- Term "padam"

Pr(padam|B1) = $\frac{1+2}{13+46}$ = 0,051

Pr(padam|B2) = $\frac{1+2}{15+46}$ = 0,049

Pr(padam|B3) = $\frac{1+2}{18+46}$ = 0,047

...

Dst

- Term "migrasi"

Pr(token|B1) = $\frac{1+0}{13+46}$ = 0,017

Pr(token|B2) = $\frac{1+0}{15+46}$ = 0,016

Pr(token|B3) = $\frac{1+1}{18+46}$ = 0,031

Hasilnya dapat dilihat dari tabel dibawah ini :

Tabel 2. Hasil hitung probabilitas setiap kata (token) pada setiap kelas

Menghitung probabilitas setiap kata (token) pada setiap kelas	Mcb	0,034	0,016	0,031
	App	0,034	0,016	0,078
	Percik	0,034	0,016	0,016
	Ptl	0,051	0,049	0,031
	Padam	0,051	0,049	0,047
	Persil	0,034	0,016	0,016
	Ganggu	0,034	0,049	0,016
	Listrik	0,034	0,066	0,016
	Gardu	0,034	0,016	0,016
	Dengar	0,034	0,016	0,016
	Ledak	0,034	0,033	0,016
	Tiang	0,017	0,033	0,016
	Miring	0,017	0,033	0,016
	Kabel	0,017	0,033	0,016
	Bawah	0,017	0,033	0,016
	Tanah	0,017	0,033	0,016
	Display	0,017	0,016	0,031
	Tera	0,017	0,016	0,031
	Error	0,017	0,016	0,031
	Id	0,017	0,016	0,031
	Rusak	0,017	0,016	0,031
	blank	0,017	0,016	0,031
	tidak	0,017	0,016	0,031
	input	0,017	0,016	0,031
	token	0,017	0,016	0,031
	migrasi	0,017	0,016	0,031

3.3.3 Perkalian nilai *Class-Conditional probability* (Probabilitas pada setiap kelas pada dokumen uji)

Setelah melakukan proses preprocessing yang sama dengan data uji, didapatkan hasil dari data uji adalah sebagai berikut:

Tabel 3. Tabel hasil *preprocessing* dokumen uji

Dok	Isi Dokumen	Hasil *Tokenizing*	Hasil *Filtering*	Hasil Stemm*ing*
Dokumen Uji	MOHON PERIKSA APP. DISPLAY APP TERTERA PERIKSA (PTL TIDAK PADAM) SEJAK HARI INI . DENGAN ID METER : 5670XXXX . INFO PELANGGAN PERNAH DILAKUKAN PERGANTIAN APP SEJAK 1 BULAN YANG LALU DAN SAAT INI TIDAK DAPAT MELAKUKAN PENGINPUTAN TOKEN.	mohon	App	App
		periksa	display	Display
		app	App	App
		display	tertera	Tera
		app	Ptl	Ptl
		tertera	tidak	Tidak
		periksa	padam	Padam
		ptl	Id	Id
		tidak	ganti	Ganti
		padam	App	App
		sejak	Tdak	Tdak
		hari	input	Input
		ini	token	Token
		dengan		
		id		
		meter		
		5670XXXX		
		Info		
		Pelanggan		
		pernah		
		Dilakukan		
		Pergantian		
		App		
		Sejak		
		1		
		Bulan		
		Yang		
		Lalu		
		Dan		
		Saat		
		Ini		
		Tidak		
		Dapat		
		Melakukan		
		Penginputan		
		Token		

Selanjutnya adalah melakukan perkalian nilai pada probabilitas setiap kelas uji dengan persamaan

$$\Pr(X|c_j) = \Pi_{i=1}^{n} \Pr(w_i|c_j) \qquad (3.4)$$

- Probabilitas dokumen uji terhadap kelas B1
  Pr(X|B1)=Pr(app|B1)*Pr(display|B1)*Pr(app|B1)*Pr(tera|B1)*Pr(ptl|B1)*Pr(tidak|B1)*Pr(padam|B1)*Pr(id|B1)*Pr(ganti|B1)*Pr(app|B1)*Pr(tidak|B1)*Pr(input|B1)*Pr(token|B1)
  = 0,034*0,017*0,034*0,017*0,051*0,0
  51*0,017*0,051*0,017*0,034*0,017*
  0,017*0,017
  = $4*10^{-20}$

- Probabilitas dokume uji terhadap kelas B2
  Pr(X|B2)=Pr(app|B2)*Pr(display|B2)*Pr
  (app|B2)*Pr(tera|B2)*Pr(ptl|B2)*Pr(tidak|B2)*Pr(padam|B2)*Pr(id|B2)*Pr(ganti|B2)
  *Pr(app|B2)*Pr(tidak|B2)*Pr(input|B2)*Pr
  (token|B2)
  =0,016*0,016*0,016*0,016*0,049*0,016*0,049*0,016*0,016*0,016*0,016*0,016
  =$3*10^{-21}$

- Probabilitas dokumen uji terhadap kelas B3 =
  Pr(X|B1)=Pr(app|B3)*Pr(display|B3)*Pr
  (app|B3)*Pr(tera|B3)*Pr(ptl|B3)*Pr(tidak|B3)*Pr(padam|B3)*Pr(id|B3)*Pr(ganti|B3)*Pr(app|B3)*Pr(tidak|B3)*Pr(input|B3)*Pr
  (token|B3)
  = 0,078*0,031*0,078*0,031*0,049*0,031 *0,047*0,031*0,078*0,031*0,031*0,031
  = $2*10^{-17}$
  Tentukan kategori dokumen data *testing*

C_{MAP} = argmax $p(c_j)$ $\Pi_{i=1}^{n} p(w_i|c_j)$ (3.5)

- Pr(B1|X) = Pr(B1) * Pr(X|B1)
  = 0,333 * $4*10^{-20}$
  = $1x10^{-20}$
- P(B2|X) = Pr(B2) * Pr(X|B2)
  = 0,333* $3*10^{-21}$
  =$1x10^{-21}$
- Pr(B3|X) = Pr(B3) * Pr(X|B3)
  = 0,333 * $2*10^{-17}$
  = $7x10^{-18}$
- C_{MAP} = argmax $p(c_j)$ $\Pi_{i=1}^{n} p(w_i|c_j)$= P(P1|X)
  = $7x10^{-18}$

Berdasarkan proses klasifikasi teks tingkat pelanggaran pelanggan dengan menggunakan metode Naïve Bayes diperoleh bahwa dokumen uji termasuk ke dalam kelas **B3 (Bidang Transaksi Energi).**

3. HASIL DAN PEMBAHASAN

Data yang digunakan menggunakan 100 data *training* dan 20 data *testing*, setelah melalui *reprocessing data*, proses perhitungan salah satu data *testing* dapat dilihat sebagai berikut:

Tabel 4. Perhitungan data *testing* untuk 1 data.

Proses / Langkah Perhitungan Naive Bayes Pada Data Uji		**B1**	**B2**	**B3**
Menghitung probabilitas untuk setiap kelas	P(Vj)= jumlah seluruh kata kelas jumlah seluruh data latih	0,3231	0,3068	0,370
Hitung ProbabilItas setiap kata (token) pada setiap kelas terhadap data uji	Tiang	0,0182	0,0952	0,004
	Listrik	0,0913	0,1428	0,004
	Percik	0,0365	0,0048	0,004
	Api	0,0411	0,0048	0,004
	Ptl	0,1050	0,1	0,028
	Padam	0,1689	0,1	0,069
	Ganggu	0,0913	0,1190	0,004
	Listrik	0,0913	0,1428	0,004
Class-conditional probability	**Dok Uji**	4,E-10	5,E-11	9,E-18
Menentukan kategori dari dokumen uji		1,E-10	2,E-11	3,E-18
Nilai tertinggi		**V**		

Perhitungan dengan 19 data lainnnya dilakukan dengan menggunakan langkah yang sama. Setelah itu maka semua klasifikasi kelas dapat diketahui.

Tabel 5. Perbandingan Kelas Sebenarnya Dan Kelas Prediksi

	Kelas Sebenarnya	**Kelas Prediksi**
D1	B1	B1
D2	B1	B1
D3	B2	B2
D4	B2	B2
D5	B3	B3
D6	B3	B3
D7	B3	B3
D8	B1	B1
D9	B3	B3
D10	B1	B2
D11	B1	B1
D12	B1	B1
D13	B1	B1
D14	B1	B1
D15	B3	B3
D16	B3	B3
D17	B2	B2
D18	B2	B2
D19	B2	B2
D20	B1	B1

Perhitungan akurasi dengan confesion matrix

Tabel 6. Perhitungan akurasi dengan *confesion matrix*

		Kelas Prediksi		
		B1	B2	B3
Kelas sebenarnya	B1	8	1	0
	B2	0	5	0
	B3	0	0	6

$$\text{Akurasi} = \frac{TP+TN}{TP+FN+FP+TN} x100 \qquad (4.1)$$

$$\text{Akurasi} = \frac{(8+5+6)}{(8+1+0+0+5+0+0+6} x100\%$$

$$\text{Akurasi} = \frac{19}{20} x100\%$$

$$\text{Akurasi} = 95\%$$

Penerapan metode *naive bayes classifier* dengan reprocessing data sebagai langkah awal pengolahan data yang kemudian diberikan nilai, dimana semakin sering suatu kata muncul maka bobot kata pada kalimat akan semakin berpengaruh pada proses klasifikasi. Setelah itu dilakukan pengklasifikasian dengan menerapkan metode *naive bayes classifier* dan pengujian dengan metode *confession matrix.*

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan uraian pada bab-bab sebelumnya maka penulis dapat mengambil kesimpulan bahwa pengklasifikasian pesan gangguan dengan pengujian menggunakan 100 data latih dan 20 dokumen uji dapat disimpulkan bahwa metode *naive bayes classifier* dengan *text mining* sebagai pemroses data awal dari aplikasi pengklasifikasian pesan gangguan ini menghasilkan akurasi dari nilai probabilitas sebesar 95%, sedangkan kesalahan sebesar 5% dapat disebabkan karena banyaknya kata- kata yang mirip pada setiap kelas sehingga memiliki bobot kata yang berdekatan. Dengan akurasi yang tinggi membuktikan bahwa metode *Naive bayes* dapat digunakan untuk membantu mengklasifikasikan pesan gangguan yang dikirimkan oleh pelanggan.

Dari hasil percobaan yang telah dilakukan, maka didapatkan bahwa semakin banyak data latih yang digunakan, maka semakin tinggi akurasi dalam melakukan pengklasifikasian pesan gangguan.

4.2 Saran

Beberapa kemungkinan pengembangan lebih lanjut yang dapat dilakukan pada penelitian ini dapat berupa penambahan fitur-fitur khusus pada proses *tokenizing* dan *filtering* sehingga hasil klasifikasi dapat lebih optimal lagi dengan pemrosesan kata yang lebih akurat.

REFERENSI

[1] Hamzah, A. (2012). Klasifikasi Teks Dengan Naïve Bayes Classifier (NBC) Untuk Pengelompokan Teks Berita Dan Abstract Akademis. *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III*, (2011), 269–277. https://doi.org/1979-911X

[2] Efendi, R., Malik, R. F., & U, J. M. (2012). Klasifikasi Dokumen Berbahasa Indonesia Menggunakan Naive Bayes Classifier. *Research in Computer Science and Applications*, *I*(I), 7–13.

[3] Wongso, R., Luwinda, F. A., Trisnajaya, B. C., Rusli, O. (2017). *ScienceDirect ScienceDirect News Article Text Classification in Indonesian Language News Article Text Classification in Indonesian Language*. *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2017.10.039

[4] Kurniawan, B., Effendi, S., & Sitompul, O. S. (2012). Klasifikasi Konten Berita Dengan Metode Text Mining. *Jurnal Dunia Teknologi Informasi*, *1*(1), 14–19.

[5] Tang, B., He, H., Baggenstoss, P. M., & Kay, S. (2016). A Bayesian Classification Approach Using Class-Specific Features for Text Categorization. *IEEE Transactions on Knowledge and Data Engineering*, *28*(6), 1602–1606.

[6] S, Vijayari., & S, D. (2015). Data Mining Classification Algorithms for Kidney Disease Prediction. *International Journal on Cybernetics & Informatics*, *4*(4), 13–25. https://doi.org/10.5121/ijci.2015.4402