

Klasifikasi Penanganan Keluhan Masyarakat Kota Probolinggo Menggunakan Algoritma Naive Bayes

Dyah Ariyanti¹, Kurnia Iswardani², Silvia Rafidah³

^{1,2,3}Universitas Panca Marga Probolinggo

¹Jl. Yos Sudarso 107 Pabean Dringu Probolinggo 67271

Telp : (0335) 422715, 427923, Fax : (0335) 427923

diyantiku@gmail.com

Abstract

Handling public complaints of Probolinggo city, whose known as "Laporo Rek", requires more time to provide the report to the relevant office. Its caused by the administrators sometimes doesn't know where the Public complaints to addressed. Using naive bayes algorithm in Text Mining for Public Complaints of Probolinggo city can help the administrators to work more effective and efficient. The processing data of Public Complaints of Probolinggo city through several stage of text mining which are token, filter, stemming and analyzing. After completed the stage, the data will be classified using naive bayes algorithm. The naive bayes algorithm calculation will view the result of each data class of Public Complaints of Probolinggo city, which is entered by phone, text message complaints. The research using this method has resulted accuray 95%, it's means each public complaints of Probolinggo city can be classified by each government agency in Probolinggo .

Keywords: Klasifikasi, Text mining, Naive Bayes

Abstrak

Penanganan keluhan/ pengaduan masyarakat Kota Probolinggo yang terkenal dengan "Laporo Rek" masih memerlukan waktu yang lama untuk memberikan laporan tentang keluhan masyarakat kepada dinas terkait dikarenakan admin pada dinas terkait terkadang tidak paham harus kemana keluhan masyarakat tersebut diberikan. Dengan adanya penelitian tentang teks mining untuk klasifikasi keluhan masyarakat dengan algoritma naive bayes dapat membantu admin dalam memecahkan persoalan diatas, sehingga waktu untuk menjawab keluhan masyarakat tersebut menjadi lebih efektif dan efisien. Pengolahan data keluhan masyarakat ini melalui beberapa tahapan teks mining yaitu token, filter, stemming dan analyzing. Setelah melalui tahapan praproses, data tersebut akan dilakukan klasifikasi menggunakan algoritma Naive Bayes, hasil perhitungan tersebut yang nantinya akan menunjukkan hasil kelas dari setiap data keluhan masyarakat yang masuk baik melalui telepon, sms. Penelitian ini menghasikan tingkat akurasi mencapai 95%, sehingga dapat mengklasifikasikan keluhan masyarakat tiap-tiap instansi di pemerintah Kota Probolinggo.

Kata kunci: Klasifikasi, Teks Mining, Naive Bayes

1. PENDAHULUAN

Tantangan besar yang dihadapi oleh pemerintah, khususnya pemerintah daerah saat ini adalah bagaimana menampilkan aparatur yang professional, memiliki etos kerja yang tinggi, keunggulan kompetitif dan kemampuan memegang teguh etika birokrasi dalam menjalankan tugas dan fungsinya, sehingga dapat memenuhi aspirasi masyarakat. Apabila sebuah lembaga atau instansi tidak banyak dikeluhkan masyarakat, bukan berarti institusi tersebut diperhatikan masyarakat. Jadi jangan menganggap tidak

ada keluhan pada instansi, bukan berarti instansi tersebut sukses, bisa jadi justru gagal total sehingga masyarakat tidak memberikan perhatian. Sebagai konsekuensi, sangatlah penting untuk bisa mendokumentasikan dan mengklasifikasi keluhan masyarakat secara sistematis. Seiring perkembangan teknologi dan ilmu pengetahuan, maka keluhan yang disampaikan semakin berkembang dan menjadi semakin luas, tidak hanya melalui catatan dan telepon tapi juga melalui media internet. Sehingga data yang diperoleh semakin banyak dan beragam. Hal ini sangat mempengaruhi efisiensi dan kecepatan jawaban seperti yang diharapkan oleh masyarakat Kota Probolinggo.

Oleh karena itu dibutuhkan suatu sistem sebagai alternatif lain untuk memperoleh jawaban yang akurat dan tidak lagi membutuhkan waktu yang lama. Klasifikasi dokumen teks[8] adalah permasalahan yang mendasar dan penting. Didalam dokumen teks, tulisan yang terkandung adalah bahasa alami manusia, yang merupakan bahasa dengan struktur yang kompleks dan jumlah kata yang sangat banyak. Permasalahan ini merupakan masalah yang cukup kompleks dikarenakan penggunaan bahasa alami tersebut. Salah satu algoritma yang dapat digunakan untuk mengklasifikasikan dokumen dimulai dari pengelolaan dokumen teks menggunakan text mining yaitu dengan Algoritma Naïve Bayes.

Tujuan dari penelitian ini adalah untuk menganalisis data keluhan masyarakat yang masuk ke "Laporo Rek" dengan algoritma Naïve Bayes, sehingga nantinya dapat dibuatkan sebuah aplikasi yang dapat membantu pelayanan dalam penyampaian keluhan masyarakat lebih cepat dan tepat. Dengan peran serta dari masyarakat Kota Probolinggo yang semakin kritis, penelitian ini diharapkan dapat mengelompokkan setiap keluhan masyarakat dari data manual (keluhan langsung via telepon, via sms atau datang ke kantor) sehingga dapat memudahkan pengklasifikasian data ke dalam masing-masing Satuan Kerja Pemerintah Daerah (SKPD) Kota Probolinggo. Diharapkan dengan dukungan dari algoritma Naïve Bayes tersebut, dapat memilahkan setiap keluhan masyarakat untuk SKPD sehingga dapat memberikan respon dengan cepat dan akurat.

Oleh karena itu dibutuhkan suatu sistem sebagai alternatif lain untuk memperoleh jawaban yang akurat dan tidak lagi membutuhkan waktu yang lama dalam merespon. Merujuk pada penelitian terdahulu[7], salah satu algoritma yang dapat digunakan untuk mengklasifikasikan dokumen dimulai dari pengelolaan dokumen teks menggunakan text mining. Dalam proses penentuan kelas untuk Dinas / Badan / Perusahaan dapat menggunakan algoritma Naïve Bayes (NB). Pada kasus ini menggunakan sample data training sebanyak 160 data keluhan masyarakat dan data testing sebanyak 40 data keluhan masyarakat.

Batasan masalah dalam penelitian ini adalah sistem yang akan dibangun merupakan sistem pengklasifikasian mengenai keluhan masyarakat Kota Probolinggo, serta algoritma yang akan diimplementasikan dengan proses pengklasifikasian menggunakan Naïve Bayes.



2. METODELOGI PENELITIAN

Penelitian ini menggunakan metodologi penelitian yaitu klasifikasi dengan Algoritma Naïve Bayes. Data yang digunakan untuk uji coba adalah data dari keluhan masyarakat Kota Probolinggo. Naïve Bayes Classifier digunakan untuk memperkirakan kejadian bersyarat munculnya suatu kelas dengan asumsi bahwa variabel-variabel bersifat independen. Naïve Bayes memiliki kinerja yang sederhana, sehingga menunjukkan akurasi dan kecepatan tinggi ketika diaplikasikan pada database yang besar.

Algoritma *naïve bayes classifier*[5], tiap-tiap kalimat keluhan masyarakat direpresentasikan dengan pasangan atribut “ $x_1, x_2, x_3, \dots, x_n$ ” dimana x_1 adalah kata pertama, x_2 adalah kata kedua dan seterusnya, serta V adalah himpunan kategori. Pada saat proses klasifikasi algoritma akan mencari nilai probabilitas yang tertinggi dari semua kategori dokumen yang diujikan (V_{MAP}), dimana persamaannya adalah sebagai berikut :

$$v_{MAP} = \underset{v \in V}{\operatorname{arg\,max}} \left(\frac{P(x_1, x_2, x_3, \dots, x_n | v_j) P(v_j)}{P(x_1, x_2, x_3, \dots, x_n)} \right) \quad (1)$$

Untuk $P(x_1, x_2, x_3, \dots, x_n)$ nilainya konstan untuk semua kategori (V_j) sehingga persamaan dapat ditulis sebagai berikut :

$$v_{MAP} = \underset{v \in V}{\operatorname{arg\,max}} (P(x_1, x_2, x_3, \dots, x_n | v_j) P(v_j)) \quad (2)$$

Persamaan diatas disederhanakan menjadi sebagai berikut :

$$v_{MAP} = \underset{v \in V}{\operatorname{arg\,max}} \prod_{i=1}^n (P x_i | V_j) P(v_j) \quad (3)$$

Keterangan :

V_j = Kategori keluhan masyarakat,

$(P x_i | V_j)$ = Probabilitas X_i pada kategori V_j , $P(v_j)$ = Probabilitas dari V_j .

$P(v_j)$ dan $(P x_i | V_j)$ dihitung pada saat pelatihan, dimana persamaannya adalah sebagai berikut :

$$P(v_j) = \frac{|\text{docs } j|}{|\text{contoh}|} \quad (4)$$

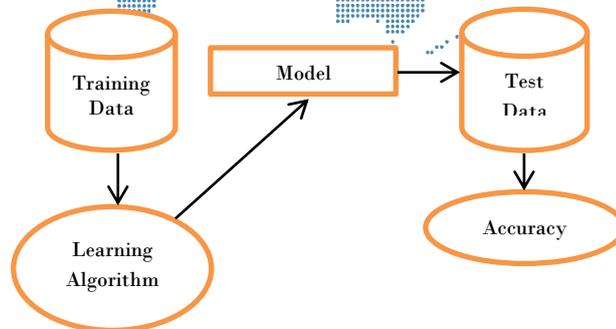
$$P(x_i | v_j) = \frac{n_k + 1}{n + |\text{kosakata}|} \quad (5)$$

Keterangan :

$|\text{docs } j|$ = jumlah dokumen setiap kategori j , $|\text{contoh}|$ = jumlah dokumen dari semua kategori, n_k = jumlah frekuensi kemunculan setiap kata, n = jumlah frekuensi kemunculan kata dari setiap kategori, $|\text{kosakata}|$ = jumlah semua kata dari semua kategori.

Dalam penelitian ini yang menjadi data uji adalah data dokumen keluhan masyarakat. Ada dua tahap pada klasifikasi dokumen. Tahap pertama adalah pelatihan terhadap dokumen yang sudah diketahui kategorinya dan hasil kelasnya atau dinas penerimanya. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya

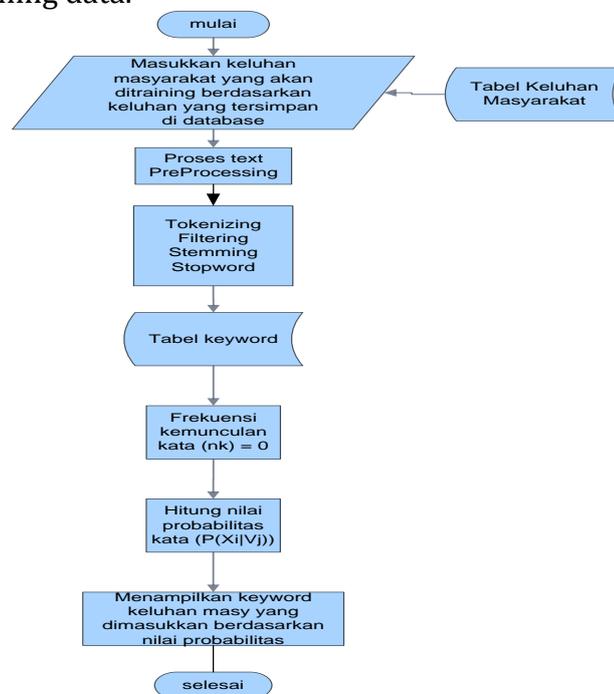
dan belum diketahui kelasnya atau dinas penerimanya. Berikut ini adalah ilustrasi sistem yang menunjukkan alur kerja metodologi penelitian:



Gambar 1. Alur Metodologi Penelitian

Sistem klasifikasi keluhan masyarakat pada pemerintah Kota Probolinggo membutuhkan beberapa data dalam suatu proses. Dalam sebulan data yang akan digunakan untuk training sebanyak 20 (dua puluh) keluhan, dengan jangka waktu periode 8 (delapan) bulan. Data-data yang dibutuhkan tersebut bisa diperoleh dari SMS dan Telepon.

Didalam proses training data, data keluhan masyarakat yang didapatkan dari sms dan telepon terlebih dahulu perlu diolah dengan proses ekstraksi data yang disebut praproses (*pre-processing*). Berikut ini gambar dari proses training data.



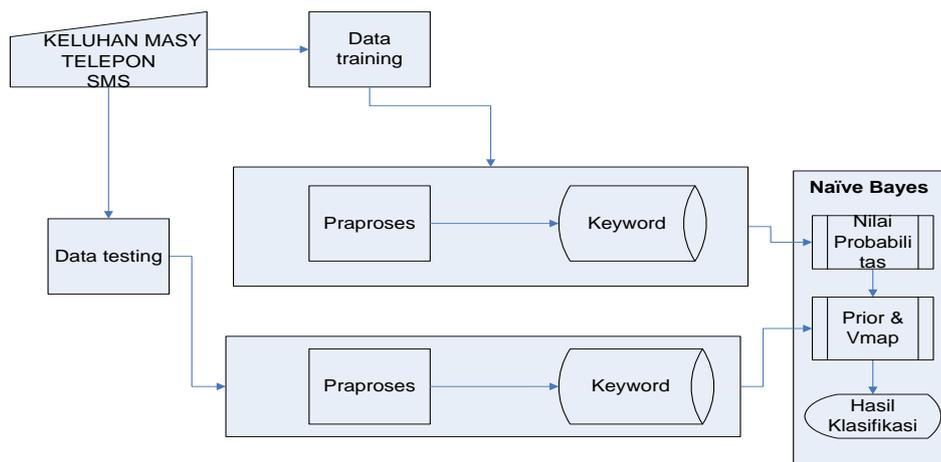
Gambar 2. Proses Training Data

Setelah melakukan praproses, maka tahap berikutnya adalah menghitung jumlah kata pada setiap kelas yang telah ditentukan. Sesuai dengan penjelasan rumus Algoritma Bayes Classifier, kemudian perlu dilakukan mencari nilai V_{map} sehingga diketahui nilai terbesar dari kelas-kelas yang telah ditentukan.

Maka dari itu, kontribusi yang diharapkan dari penelitian ini adalah sistem mampu memberikan hasil pengelompokan Dinas / Badan / Perusahaan secara efektif dan efisien. Sehingga penyampaian keluhan masyarakat ke tiap-tiap Instansi Pemerintah Kota Probolinggo cepat untuk direspon dan tentunya masyarakat Kota Probolinggo sesegera mungkin untuk memperoleh jawaban yang akurat dan tidak lagi membutuhkan waktu yang lama. Keberhasilan klasifikasi untuk keluhan masyarakat ini yang menggunakan algoritma Naïve Bayes, kemudian ditentukan dengan menghitung nilai akurasi.

3. HASIL DAN PEMBAHASAN

Bagian ini akan menjelaskan beberapa proses sistem yang akan dijalankan pada aplikasi penelitian ini. Berikut ini alur cara kerja aplikasi.



Gambar 3. Arsitektur Sistem

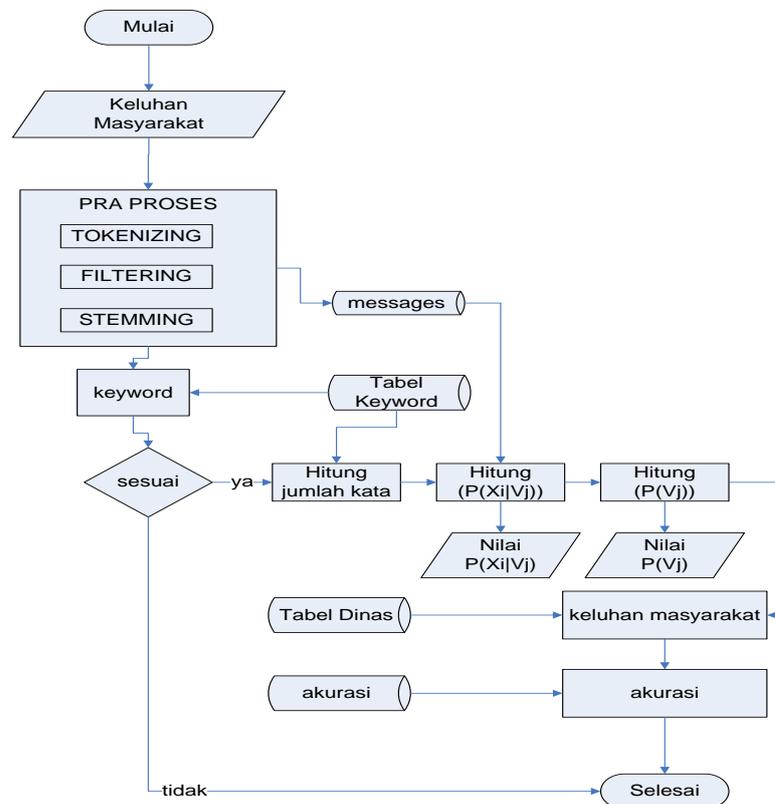
Arsitektur Sistem diatas menjelaskan bahwa data keluhan masyarakat akan diproses dengan dua tahap yaitu proses data training dan data testing. Data keluhan masyarakat diambil melalui sms dan telepon. Pada proses data training dan data testing, selanjutnya dilakukan proses text mining (praproses). Pada praproses ini kemudian ditentukan kata-kata baku yang termasuk dalam kamus. Kata yang terpilih merupakan kata dasar (tidak berimbuhan) dan bukan termasuk di dalam stopword. Kata yang terpilih ini dapat dilihat di tabel 2. Dari hasil keyword pada tabel 2 tersebut yang berikutnya akan menentukan proses perhitungan Naïve Bayes.

Sedangkan struktur sistem dalam proses testing akan dijelaskan pada gambar 4. Gambar tersebut secara menyeluruh menjelaskan bagaimana model klasifikasi yang akan dibangun, baik secara manual ataupun secara

otomatis. Prosesnya dimulai dengan melakukan input data keluhan masyarakat. Data tersebut kemudian diproses untuk menentukan apakah kalimat yang digunakan sudah baku atau belum baku.

Hasilnya kemudian diolah ke dalam pra proses melalui beberapa cara yaitu tokenizing, filtering dan stemming. Setelah itu, hasil dari pra proses menemukan kata baku dan membandingkan kata hasil stemming dengan kata yang terdapat pada kamus (keyword). Apabila kata baku sesuai dengan nilai probabilitas keyword yang sudah dihitung, maka selanjutnya dilakukan proses perhitungan Naïve Bayes (Vmap) yang menentukan apakah data keluhan masyarakat tersebut masuk dalam hasil klasifikasi dinas yang sudah tersimpan. Sedangkan apabila kata baku tidak sesuai maka proses perhitungan Naïve Bayes diakhiri.

Berikut ini adalah gambar alur diagram dari proses testing untuk menentukan klasifikasi menggunakan algoritma naïve bayes.



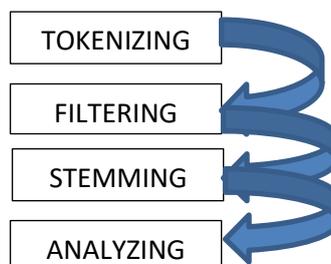
Gambar 4. Proses Testing

Proses awal pertama kali yang dilakukan pada penelitian ini adalah melakukan testing data keluhan masyarakat yang sudah dikelompokkan secara manual oleh Dinas Komunikasi Informatika Kota Probolinggo. Data yang diinputkan sebanyak 160 (seratus enam puluh) komentar, yang dipilih dari 10 (sepuluh) Dinas yang telah ditentukan. Dapat dilihat pada tabel 1 berikut ini.

Tabel 1. Hasil Klasifikasi

No	Instansi
1.	DPU = DINAS PEKERJAAN UMUM
2.	PDAM = PERUSAHAAN DAERAH AIR MINUM
3.	DISHUB = DINAS PERHUBUNGAN
4.	DINSOS = DINAS SOSIAL
5.	POLRESTA = KEPOLISIAN KOTA PROBOLINGGO
6.	PLN = PERUSAHAAN LISTRIK NEGARA
7.	DISPENDUK CAPIL = DINAS KEPENDUDUKAN DAN CATATAN SIPIL
8.	DISPENDIK = DINAS PENDIDIKAN
9.	BLH = BADAN LINGKUNGAN HIDUP
10.	DINKES = DINAS KESEHATAN

Data yang dimasukkan tersebut selanjutnya dilakukan proses teks mining. Berikut ini adalah gambar dari proses teks mining.



Gambar 5. Proses Teks Mining

Dari hasil akhir proses stemming kemudian selanjutnya komentar tersebut dibuatkan daftar stop list. Kata yang tidak termasuk dalam stop list, akhirnya disimpan dalam database keyword yang telah disiapkan. Apabila daftar keyword tersebut termasuk kata umum yang sering digunakan pada semua dinas, maka bisa dipilih untuk dihapus (dimasukkan dalam daftar stop list). Sehingga akan terlihat jumlah kata yang tersimpan, selanjutnya akan menentukan dalam perhitungan probabilitas.

Beberapa kata/kategori menjadi poin utama dalam pengelompokkan data keluhan masyarakat. Tabel 2 berikut ini merupakan beberapa kategori kata yang digunakan pada tiap-tiap dinas.

Tabel 2. Data Kategori/Keyword

No	Dinas	Kategori
		Keyword
1	Dinas Pekerjaan Umum (DPU)	Aspal, Buang, Hancur, Hujan, Kendaraan,
		Kotor, Lubang, Macet, Niaga, Pasar,
		Pasir, Realisasi, Rusak, Rusunawa, Truk,
		Sungai, Trotoar,
2	Perusahaan Daerah Air Minum	Air, Alir, Arah, Bau, Bayar, Biaya, Mati,
		Bocor, Gali, Keruh, Loket, Meter, Pasir,

	(PDAM)	Pasang, PDAM, Pipa, Rusak, Rusunawa, Salur, Selokan
3	Dinas Perhubungan (DISHUB)	Anak, Arah, Arus, Becak, Celaka, Macet, Dishub, Ganggu, Hijau, Karyawan, Mati, Kendara, Korban, Kuning, Lalin, Maut, Merah, Meter, Niaga, Nyala, Padat, Pos, Parkir, Pasar, Polisi, Rawan, Sekolah, Sepeda, Siswa, Traffic, Trafic, Truk,
4	Dinas Sosial (DINSOS)	Anak, Asuh, Bantuan, Beras, Blsm, Blt, Cacat, Dagang, Dana, Dewan, Ekonomi, Hilang, Jahit, Kartu, Lumpuh, Mesin, Pos Masyarakat, Modal, Netra, Pasar, Pindah, SD, Sekolah, Sembako, Sepeda, SMA, SMP, Taman, Tetangga, Tuna, Tunanetra, Umur,
5	Dinas Kesehatan (DINKES)	Bayar, Becak, Bpjs, Dokter, Gratis, KTP, Jamkesmas, Kartu, Karyawan, Katarak, Klinik, Korban, Malaria, Mata, Mati, Pasien, Periksa, Poliklinik, Puskesmas, Sakit, Sehat, Selaput, Serangga, Uang, Wabah
6	Perusahaan Listrik Negara (PLN)	Biaya, Daya, Denda, Ganggu, Gardu, Kabel, Listrik, Mati, Meter, Nyala, Pasar, Oknum, Pasang, PLN, Segel, Stabil, Tiang, Tegang, Teknisi, Turun, Voltase, Watt
7	Dinas Kependudukan dan Catatan Sipil (DISPENDUK CAPIL)	Akta, Anak, Angkat, Berkas, Biaya, KTP, Capil, Denda, E, Gratis, Hilang, KK, Lahir, Masa, Negeri, Pangkal, Sertifikat, Pindah, Reguler, Rekam, Rusak, Umur
8	Dinas Pendidikan (DISPENDIK)	Ajar, Anak, Bantuan, Bayar, Beras, Bos, Biaya, Dana, Didik, Guru, Kartu, Kelas, Komite, Magang, Modal, Pramuka, SD, Pungli, Ragam, Realisasi, Sertifikasi, Uang, Sekolah, Siswa, SMA, SMK, Swasta,
9	Badan Lingkungan Hidup (BLH)	Air, Angin, Bak, Banjir, Bau, Bayar, PJU, Bersih, Busuk, Celaka, Gelap, Genang, Hijau, Kering, Kotoran, Mati, Nyala, Pos, Padam, Pangkas, Pasar, Pohon, Polisi, Ranting, Roboh, Sampah, Selokan, Sengon, Sungai, Taman, Tebang, Terang
10	Polisi Resort Kota (POLRESTA)	Aman, Anak, Arah, Area, Arena, Judi, Bandar, Biaya, Gerebek, Hilang, Indikasi, Kendara, Korban, Kriminal, Masa, Miras, Minum, Obat, Oknum, Polisi, Pos, Punk, Rampas, Rawan, Razia, Sekolah, Sepeda, Sim, SMP, Sopir, Tabrak, Tilang, Traffic, Turun, Uang, Umur,

Pada perhitungan Naïve Bayes, proses yang pertama kali dilakukan dalam tahap perhitungan adalah pembetulan data training, kemudian menghitung nilai kata yang muncul pada tiap kelas.

Berikut ini merupakan salah satu contoh bentuk uji coba, untuk membuktikan teks mining klasifikasi keluhan masyarakat dengan menggunakan algoritma NB.

3.1. Contoh Testing Keluhan Masyarakat :

Perumahan sti gelisah masalah lampu padam hampir tiap malam ada apa PLN ?

Prose Token

perumahan // sti // gelisah // masalah // lampu // padam // hampir // tiap // malam // ada // apa // pln

Hasil Stemming – Stop Word : gelisah (1), padam (1), pln (1)

Beberapa kata pada keluhan masyarakat tersebut tidak muncul pada hasil stemming, akan dijelaskan berikut ini :

- a. Kata “sti” tidak muncul karena tidak terdaftar dalam kamus bahasa indonesia
- b. Kata “masalah”, “hampir”, “tiap”, “malam”, “ada”, “apa”, tidak muncul karena kata-kata tersebut masuk dalam stopwords.

Selanjutnya diambil perhitungan nilai probabilitas yang sesuai dengan keyword yang termasuk pada contoh tersebut. Dan kemudian dilakukan perhitungan klasifikasi, yang hasilnya dijabarkan tabel 3 berikut ini :

Tabel 3. Perhitungan Klasifikasi

Kelas	pln	padam	P(Ci)	VMAP
DPU	0.00446	0.00446	0.1	0.000001992985
PDAM	0.00412	0.00412	0.1	0.000001693509
DISHUB	0.00382	0.00382	0.1	0.000001456792
DINSOS	0.00388	0.00388	0.1	0.000001502314
DINKES	0.00403	0.00403	0.1	0.000001625911
PLN	0.03543	0.01574	0.1	0.000055800112
CAPIL	0.00376	0.00376	0.1	0.000001413308
DISPENDIK	0.00364	0.00364	0.1	0.000001322314
BLH	0.00411	0.01646	0.1	0.000006774035
POLRESTA	0.00386	0.00386	0.1	0.000001490735

Karena $P(\text{“PLN”}|\text{“Keluhan_Masyarakat”})$ mempunyai nilai yang lebih besar dibandingkan dengan 9 (sembilan) kelas lainnya antara lain :
 $P(\text{“DPU”}|\text{“Keluhan_Masyarakat”})$,
 $P(\text{“DISHUB”}|\text{“Keluhan_Masyarakat”})$,
 $P(\text{“DINSOS”}|\text{“Keluhan_Masyarakat”})$,
 $P(\text{“DINKES”}|\text{“Keluhan_Masyarakat”})$,
 $P(\text{“PDAM”}|\text{“Keluhan_Masyarakat”})$,
 $P(\text{“CAPIL”}|\text{“Keluhan_Masyarakat”})$,
 $P(\text{“DISDIK”}|\text{“Keluhan_Masyarakat”})$,
 $P(\text{“BLH”}|\text{“Keluhan_Masyarakat”})$,
 $P(\text{“POLRESTA”}|\text{“Keluhan_Masyarakat”})$
 sehingga hasil data training pada tabel. 3 masuk pada kelas “PLN”.

4. SIMPULAN

Dari hasil uji coba pada penelitian ini maka dapat ditarik beberapa kesimpulan sebagai berikut:

- a) Hasil Klasifikasi yang diperoleh dalam penelitian ini mendapatkan hasil yang baik. Uji coba yang pertama dilakukannya uji klasifikasi dengan menggunakan semua proses sampai tahap analyzing data mendapatkan hasil nilai rata-rata akurasi sebesar 95 %.
- b) Hasil terbaik yang diberikan selama uji coba pada penelitian ini adalah melakukan klasifikasi terhadap keluhan masyarakat, dimana data yang dipakai bersumber dari telepon dan sms.
- c) Algoritma Naive Bayes yang digunakan dalam penelitian ini memberikan perhitungan yang maksimal.

UCAPAN TERIMA KASIH

Penelitian ini sepenuhnya dibiayai oleh Direktorat Penelitian dan Pengabdian Masyarakat Pendidikan Tinggi melalui skema Penelitian Dosen Pemula.

DAFTAR PUSTAKA

- [1] Annur, Haditsah. 2018. Klasifikasi Masyarakat Miskin Menggunakan Metode Naïve Bayes. *ILKOM Jurnal Ilmiah*. Vol.10 No.02. ISSN 2548-7779
- [2] Albert, J. 2009, *Bayesian Computation with R*, Springer NY.
- [3] Devita, R.N, Herwanto, H.W., Wibawa A.P. 2018. *Perbandingan Kinerja Metode Naïve Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia*. *JTIK (Jurnal Teknologi Informasi dan Ilmu Komputer)* Vol.05 No.04. ISSN 2528-6579
- [4] Fadillah, AP, Hardiyana, Bela. 2018. Penerapan Naïve Bayes Classifier untuk Pemilihan Konsentrasi Mata Kuliah. *JATI (Jurnal Teknologi dan Informasi)*. Vol.08 No.02. ISSN 2655-6839
- [5] Han, J., and Kamber M. 2000. *Data Mining, Concept and Techniques*, New York, NY : Morgan Kaufman
- [6] Kim. J. and Thoma GR. 2008. Extracting Bibliographic Information from Biomedical Online Articles, *International Conference On Data Mining*. Las Vegas, Nevada, USA.
- [7] Niloy, NH., Mai, Navid, 2018. Naïve Bayes Classifier and Classification Trees for the Predictive Accuracy of Probability of Default Credit Card Clients. *American Journal of Data Mining and Knowledge Discovery*, Vol.03 Issue 1.
- [8] Rivki, M. & Bachtiar, A.M., 2017. Implementasi Algoritma K-Nearest Neighbor dalam Pengklasifikasian Follower Twitter yang Menggunakan Bahasa Indonesia. *Jurnal Sistem Informasi*, 13(1), Pp.31-37.