

868-Article_Text-1297-1-10- 20201109_2.pdf

by - -

Submission date: 19-Mar-2024 10:25AM (UTC+0700)

Submission ID: 2324412944

File name: 868-Article_Text-1297-1-10-20201109_2.pdf (659.7K)

Word count: 3315

Character count: 20765

Teks Mining untuk Klasifikasi Keluhan Masyarakat Menggunakan Algoritma Naive Bayes

Dyah Ariyanti, Kumia Iswardani.

Universitas Panca Marga Probolinggo
Jl. Yos Sudarso 107 Pabean Dringu Probolinggo 67271
E-mail : diyantiku@gmail.com, kurnia.iswardani@gmail.com

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. 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 menghasilkan tingkat akurasi mencapai 95%, sehingga dapat mengklasifikasikan keluhan masyarakat tiap-tiap instansi di pemerintah Kota Probolinggo.

Kata kunci : Klasifikasi, Teks Mining, Naive Bayes

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, steaming 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 .

Keyword : Classification, Text Mining, Naive Bayes

1. PENDAHULUAN

8
Tantangan besar yang dihadapi oleh pemerintah, khususnya pemerintah daerah saat ini adalah bagaimana menampilkan aparatur yang profesional, memiliki etos kerja yang tinggi, keunggulan kompetitif dan kemampuan memegang teguh etika birokrasi

dalam menjalankan tugas dan fungsinya, sehingga dapat memenuhi aspirasi masyarakat.

Jika sebuah lembaga atau instansi banyak dikeluhkan masyarakat, berarti institusi tersebut diperhatikan oleh

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 mengorganisir dan mengklasifikasi keluhan masyarakat secara otomatis.

Aparatur pemerintah daerah di seluruh Kabupaten / Kota saat ini sedang berlomba-lomba untuk mewujudkan daerahnya sebagai kota yang berbasis "SMART CITY", adanya informasi keterbukaan kepada seluruh lapisan masyarakat. Informasi yang diberikan tentunya tidak terbatas dan terbuka untuk segala aspek, baik di bidang ekonomi, sosial, budaya dan lainnya. Dengan demikian, diharapkan masyarakat aktif berpartisipasi untuk menyampaikan kritik dan saran kepada Pemerintah, sehingga apa yang diadukan / dikeluhkan oleh masyarakat tersebut cepat teratasi, karena teknologi-teknologi yang telah disiapkan oleh Pemerintah daerah tersebut.

Seiring perkembangan teknologi dan ilmu pengetahuan tersebut, maka keluhan yang disampaikan semakin berkembang dan menjadi semakin luas, tidak hanya melalui catatan dan telepon tapi juga melalui media internet. Data yang didapatkan / diperoleh semakin banyak dan beragam. Hal ini sangat mempengaruhi efisiensi dan kecepatan jawaban seperti yang diharapkan oleh masyarakat khususnya masyarakat Kota Probolinggo. Oleh karena itu dibutuhkan suatu sistem sebagai alternatif lain untuk memperoleh jawaban yang akurat dan tidak lagi membutuhkan waktu yang lama dalam merespon.

Klasifikasi dokumen teks 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*.

Dengan peran serta dari masyarakat Kota Probolinggo yang semakin kritis, penulis mempunyai inisiatif untuk mengklasifikasikan setiap keluhan masyarakat baik dari data manual (keluhan langsung via telepon, via sms atau datang ke kantor) atau data yang berada di media sosial sehingga dapat memudahkan pengklasifikasian data ke dalam masing-masing Satuan Kerja Pemerintah Daerah (SKPD) Kota Probolinggo. Diharapkan dengan dukungan dari *algoritma naïve bayes* tersebut, sistem yang akan dibuat dapat memilahkan setiap keluhan masyarakat untuk SKPD sehingga dapat memberikan respon dengan cepat dan akurat.

Tujuan dan Manfaat Penelitian

Berdasarkan latar belakang masalah tersebut di atas, maka tujuan dari penelitian ini adalah :

1. Membuat sebuah sistem pengelompokkan kategori / keyword data keluhan masyarakat berdasarkan seluruh keluhan masyarakat yang tersedia pada Dinas Komunikasi dan Informatika dan Radio Suara Kota Probolinggo dengan menerapkan Algoritma Naïve Bayes.
2. Dengan menerapkan Algoritma Naïve Bayes dapat melakukan proses pengklasifikasian text dari berbagai media yang tersedia sehingga nantinya dapat menghasilkan informasi secara tepat dan akurat.

2. METODOLOGI

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.

Menurut Han and Kamber (2000), Algoritma *naïve bayes classifier*, pada tiap-tiap kalimat keluhan masyarakat di representasikan 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 (2) 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 :

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

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

Keterangan :

$|\text{docs}|$ = 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.

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.

3. LANDASAN TEORI

Keluhan Masyarakat

Pada saat ini kinerja pelayanan publik sudah menjadi ukuran kinerja pemerintah daerah, terutama kepala daerahnya. Dalam berbagai kesempatan ketidak-puasan masyarakat atas kinerja manajemen pelayanan publik ini kian banyak diungkapkan oleh masyarakat secara terbuka. Masyarakat menuntut penyelenggaraan manajemen pelayanan lebih responsif atas kebutuhan masyarakat dan penyelenggaraan manajemen pelayanan publik yang transparan, partisipatif dan akuntabel.

Berbagai upaya telah dilakukan oleh pemerintah dan para stakeholder dalam bidang pelayanan publik untuk dapat memperbaiki kondisi masih rendahnya kualitas pelayanan publik sebagaimana telah diuraikan di depan. Usaha-usaha yang telah dilakukan tersebut, misalnya: peningkatan kualitas sumber daya manusia birokrasi, perbaikan fasilitas pelayanan publik, perbaikan sistem dan prosedur pelayanan dan lain sebagainya.

Dengan meningkatnya tuntutan masyarakat atas tata kelola pemerintahan yang baik dan benar (*good governance*), maka pemberian prioritas atas kegiatan pelayanan menjadi suatu keharusan. Salah satu sektor yang perlu mendapat prioritas adalah manajemen pengaduan masyarakat yang ada di sektor publik. Untuk itu, pemerintah memberikan prioritas utama dibidang pelayanan atas pengaduan yang berasal dari masyarakat. Pelayanan yang baik, efisien dan efektif akan memberikan harapan akan terpenuhinya rasa keadilan di masyarakat serta terminya pengelolaan keuangan negara yang transparan dan terarah.

Pada prinsipnya pelayanan pengaduan masyarakat kepada pemerintah diupayakan agar mempermudah masyarakat yang akan menyampaikan pengaduannya, antara lain dengan menyediakan layanan hotline, faksimili dan situs web.

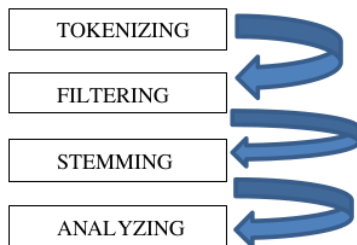
Teks Mining

Teks mining adalah salah satu bidang khusus dari *data mining*. Teks mining dapat didefinisikan sebagai suatu proses untuk menggali informasi dimana seorang user berinteraksi dengan sekumpulan dokumen menggunakan tools analisis yang merupakan komponen-komponen dalam *data mining* yang salah satunya adalah kategorisasi.

Dalam memberikan solusi, teks mining mengadopsi dan mengembangkan banyak teknik dari bidang lain, seperti *data mining, information retrieval, statistic and matematik, machine learning, linguistic, natural language processing, and visualization*.

Tujuan dari *teks mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Jadi, sumber data yang digunakan pada *teks mining* adalah kumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur. Adapun tugas khusus dari teks mining antara lain yaitu pengkategorisasian teks dan pengelompokan teks.

Teks yang akan dilakukan proses *teks mining*, pada umumnya memiliki beberapa karakteristik diantaranya adalah memiliki dimensi yang tinggi, terdapat *noise* pada data, dan terdapat struktur teks yang tidak baik. Cara yang digunakan dalam mempelajari suatu data teks, dengan terlebih dahulu menentukan fitur-fitur yang mewakili setiap kata untuk setiap fitur yang ada pada dokumen. Sebelum menentukan fitur – fitur yang mewakili, diperlukan tahap praproses yang dilakukan secara umum dalam teks mining, yaitu *case folding, tokenizing, filtering, stemming, tagging dan analyzing*.



Gambar 1. Proses teks mining

Praproses

Dengan membandingkan dari penelitian terdahulu menurut Triawati (2009), algoritma yang digunakan pada teks mining, biasanya tidak hanya melakukan perhitungan hanya pada dokumen, tetapi pada juga feature. Berikut ini beberapa macam *feature* yang sering digunakan :

1. Character

Merupakan komponen individual, bisa huruf, angka, karakter spesial dan spasi, merupakan block pembangun pada level paling tinggi pembentuk semantik feature, seperti kata *term* dan *concept*. Pada umumnya, representasi dasar karakter ini jarang digunakan pada beberapa teknik pemrosesan teks.

2. Words

Merupakan single word dan frasa multiword yang terpilih secara langsung dari corpus. Representasi term-based dari dokumen tersusun dari subset term dalam dokumen.

3. Concept

Merupakan feature yang di-generate dari sebuah dokumen secara manual, rule-based, atau metodologi lain. Pada tugas akhir ini, concept di-generate dari argument atau verb yang sudah diberi label pada suatu dokumen.

Klasifikasi

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu : pertama, pembangunan model sebagai prototype untuk disimpan sebagai memori dan kedua, penggunaan model tersebut untuk melakukan pengenalan / klasifikasi / prediksi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang mudah disimpan. Contoh aplikasi yang sering ditemui adalah pengklasifikasian jenis hewan yang mempunyai sejumlah atribut.

Menurut Suyanto (2019), model dalam klasifikasi mempunyai arti dimana ada suatu model yang menerima masukan, kemudian mampu melakukan pemikiran

terhadap masukan tersebut dan memberikan jawaban sebagai keluaran dari hasil pemikirannya.

Setiap algoritma mempunyai kelebihan dan kekurangan, tetapi semua algoritma berprinsip sama, yaitu melakukan suatu pelatihan, model dapat memetakan setiap vector masukan ke label kelas keluaran dengan benar.

Pada penelitian ini, untuk klasifikasi proses dalam menghitung akurasi digunakan formula :

$$\text{Akurasi} = \frac{\text{jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} = \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad (6)$$

Untuk menghitung laju error (kesalahan prediksi) digunakan formula :

$$\text{Laju error} = \frac{\text{jumlah data yang diprediksi secara salah}}{\text{jumlah prediksi yang dilakukan}} = \frac{f_{10} + f_{01}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad (7)$$

Semua algoritma klasifikasi berusaha membentuk model yang mempunyai akurasi tinggi atau (laju error yang rendah). Umumnya model yang dibangun memprediksi dengan benar pada semua data yang menjadi data latihnya, tetapi ketika model berhadapan dengan data uji, barulah kinerja model dari sebuah algoritma klasifikasi ditentukan

Naïve Bayes

Teorema Bayes menerangkan hubungan antara probabilitas terjadinya peristiwa A dengan syarat peristiwa B telah terjadi dan probabilitas terjadinya peristiwa B dengan syarat peristiwa A telah terjadi. Teorema ini didasarkan pada prinsip bahwa tambahan informasi dapat memperbaiki probabilitas. Teorema Bayes ini bermanfaat untuk mengubah atau memutakhirkan (meng-update) probabilitas yang dihitung dengan tersedianya data dan informasi tambahan.

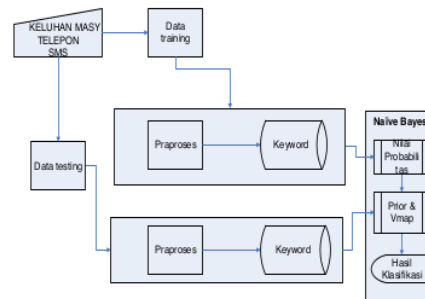
Menurut Niloy (2018), bahwa kaitan antara Naïve Bayes dengan klasifikasi, korelasi hipotesis, dan bukti dengan klasifikasi adalah bahwa hipotesis dalam teorema Bayes merupakan label kelas yang

menjadi target pemetaan dalam klasifikasi, sedangkan bukti merupakan fitur-fitur yang menjadi masukan dalam model klasifikasi. Jika X adalah vector masukan yang berisi fitur dan Y adalah label kelas, Naïve Bayes dituliskan dengan P(Y|X). Notasi tersebut berarti probabilitas label kelas Y didapatkan setelah fitur-fitur X diamati. Notasi ini disebut juga probabilitas akhir (posterior probability) untuk Y, sedangkan P(Y) disebut probabilitas awal (prior probability) Y.

Naive bayesian klasifikasi adalah suatu klasifikasi yang berpeluang sederhana berdasarkan aplikasi teorema Bayes dengan asumsi antar variabel penjelas saling bebas (independen). Dalam hal ini, diasumsikan bahwa kehadiran atau ketiadaan dari suatu kejadian tertentu dari suatu kelompok tidak berhubungan dengan kehadiran atau ketiadaan dari kejadian lainnya.

4. HASIL DAN PEMBAHASAN

Pada bagan dibawah ini akan menjelaskan beberapa proses sistem yang akan dijalankan pada aplikasi penelitian ini. Dan akan diberi penjelasan cara kerja pada aplikasi tersebut.



Gambar 2. Cara kerja aplikasi

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

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

Contoh Testing Keluhan Masyarakat :

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

Token

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

Hasil Stemming – Stop Word

gelisah (1), padam (1), pln (1)

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

Tabel 1. Hasil 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
POLISI	0.00386	0.00386	0.1	0.000001490735

Karena P(“PLN”|”KeluhanMasyarakat”) mempunyai nilai yang lebih besar dibandingkan dengan 9 (sembilan) kelas lainnya antara lain :
P(“DPU”|”KeluhanMasyarakat”),
P(“DISHUB”|”KeluhanMasyarakat”),
P(“DINSOS”|”KeluhanMasyarakat”),
P(“DINKES”|”KeluhanMasyarakat”),
P(“PDAM”|”KeluhanMasyarakat”),

P(“CAPIL”|”KeluhanMasyarakat”),
P(“DISDIK”|”KeluhanMasyarakat”),
P(“BLH”|”KeluhanMasyarakat”),
P(“POLISI”|”KeluhanMasyarakat”)
Sehingga hasil data training pada tabel. 1 masuk pada kelas PLN.

Berikut ini segmen program Menghitung Naïve Bayes.

```

1 : baris = 0
2 : VSNB.Clear
3 : For kolom = 1 To GridNB.Cols - 1
4 : nilaiawal = 0
5 : nilaiakhir = 1
6 : For j = 1 To GridNB.rows - 1
7 : For i = 1 To VSDocMuncul.rows - 1
8 : If VSDocMuncul.TextMatrix(i, 1) = GridNB.TextMatrix(j, 0) Then
9 : baris = baris + 1
10 : VSNB.rows = baris + 1
11 : VSNB.TextMatrix(baris, 0) = baris
12 : VSNB.TextMatrix(baris, 1) = "P (" & GridNB.TextMatrix(0, kolom) & " | " & GridNB.TextMatrix(j, 0) & " ) = " & GridNB.TextMatrix(j, kolom) & ""
13 : VSNB.TextMatrix(baris, 2) = GridNB.TextMatrix(j, kolom) 'hasil
14 : VSNB.TextMatrix(baris, 3) = GridNB.TextMatrix(0, kolom) 'dinas
15 : VSNB.TextMatrix(baris, 4) = GridNB.TextMatrix(j, 0) 'keyword
16 : nilaiawal = Cdbl(GridNB.TextMatrix(j, kolom))
17 : nilaiakhir = nilaiakhir * nilaiawal
18 : End If
19 : Next i
20 : Next j
21 : nilaiawal = GridDinas.rows - 1
22 : For i = 1 To GridDinas.rows - 1
23 : If GridDinas.TextMatrix(i, 2) = GridNB.TextMatrix(0, kolom) Then
24 : nilaiakhir = (1 / nilaiawal) * nilaiakhir '( 1/totaldinas) * nilai probabilitas per kosakata
25 : GridDinas.TextMatrix(i, 9) = Cdbl(Format(nilaiakhir, "##.#####"))
26 : End If
27 : Next i

```



```

28 : Next kolom
29 : nilaiawal =
    GridDinas.TextMatrix(1, 9)
    'catch nilai dinas pertama
30 : nilaiakhir = GridDinas.rows - 1
    'catch maksimal indeks grid
    dinas
31 : For i = 1 To GridDinas.rows - 1
32 : If nilaiawal =
    Val(GridDinas.TextMatrix(i, 9))
    Then
33 : kalimat =
    GridDinas.TextMatrix(i, 2)
34 : xkddinas =
    GridDinas.TextMatrix(i, 8)
35 : ElseIf
    Val(GridDinas.TextMatrix(i, 9))
    > nilaiawal Then
36 : nilaiawal =
    Val(GridDinas.TextMatrix(i, 9))
37 : kalimat =
    GridDinas.TextMatrix(i, 2)
38 : xkddinas =
    GridDinas.TextMatrix(i, 8)
39 : End If
40: Next i

```

5. KESIMPULAN

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

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 %.

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.

DAFTAR PUSTAKA

Annur, Haditsah. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naïve Bayes. ILKOM Jurnal Ilmiah. Vol.10 No.02. ISSN 2548-7779

Albert, J. (2009). Bayesian Computation with R, Springer NY.

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

Fadillah, AP, Hardiyana, Bela. (2018). Penerapan Naïve Bayes Cllasifier untuk Pemilihan Konsentrasi Mata Kuliah. JATI (Jurnal Teknologi dan Informasi). Vol.08 No.02. ISSN 2655-6839

Han, J., and Kamber M. (2000). Data Mining, Concept and Techniques, New York, NY : Mogan Kaufman

Kim. J. and Thoma GR. (2008). Extracting Bibliographic Information from Biomedical Online Articles, International Conference On Data Mining. Las Vegas, Nevada, USA.

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.

Rivki, M.and Bachtiar, A.M., (2017). Implementasi Algoritma K-Nearest Neighbor dalam Pengklasifikasian Follower Twitter Menggunakan Bahasa Indonesia. Jurnal Sistem Informasi, 13(1).

Suyanto, Dr. (2019). Data Mining untuk Klasifikasi dan Klasterisasi Data (Edisi Revisi). Informatika Bandung.

Triawati Candra. (2009). Metode Pembobotan Statistical Concept Based untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia. SNATI STT Telkom: Bandung.

ORIGINALITY REPORT

18%

SIMILARITY INDEX

0%

INTERNET SOURCES

0%

PUBLICATIONS

18%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Tidar Student Paper	3%
2	Submitted to Forum Perpustakaan Perguruan Tinggi Indonesia Jawa Timur Student Paper	2%
3	Submitted to Universitas Amikom Student Paper	2%
4	Submitted to Universitas Mercu Buana Student Paper	1%
5	Submitted to Universitas Indonesia Student Paper	1%
6	Submitted to Institut Pertanian Bogor Student Paper	1%
7	Submitted to UPN Veteran Yogyakarta Student Paper	1%
8	Submitted to LL DIKTI IX Turnitin Consortium Part IV Student Paper	1%
9	Submitted to STT PLN	

Student Paper

1 %

10

Submitted to Universitas Mulawarman

Student Paper

1 %

11

Submitted to Universitas Jember

Student Paper

1 %

12

Submitted to Universitas Nahdlatul Ulama
Surabaya

Student Paper

1 %

13

Submitted to Universitas Islam Negeri
Sumatera Utara

Student Paper

1 %

14

Submitted to Universitas Bina Sarana
Informatika

Student Paper

<1 %

15

Submitted to Universitas Jenderal Achmad
Yani

Student Paper

<1 %

16

Submitted to Universitas Merdeka Malang

Student Paper

<1 %

17

Submitted to Universitas Kristen Satya
Wacana

Student Paper

<1 %

18

Submitted to Universitas Sebelas Maret

Student Paper

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography On