

3_IMPLEMENTASI NAÏVE BAYES UNTUK KLASIFIKASI VENTING NEGATIF FEELINGS

by Herbertus Yulianton

Submission date: 10-Apr-2023 11:55PM (UTC+0700)

Submission ID: 2060699973

File name: NTASI_NA_VE_BAYES_UNTUK_KLASIFIKASI_VENTING_NEGATIF_FEELINGS.pdf (410.74K)

Word count: 2050

Character count: 12770

Implementasi Naïve Bayes untuk Klasifikasi Venting Negatif Feelings terhadap Penyedia Jasa Komunikasi pada Media Sosial Twitter

Heribertus Yulianton¹, Felix Andreas Sutanto², Kristophorus Hadiono³

^{1,2,3}Fakultas Teknologi Informasi, Universitas Stikubank

e-mail: ¹heri@unisbank.ac.id

Abstrak

Kesan masyarakat terhadap suatu penyedia jasa layanan komunikasi dapat dianalisa melalui respon mereka. Salah satu media yang dapat digunakan untuk mendapatkan kesan tersebut adalah media sosial twitter. Seiring dengan perkembangan Internet, saat ini WoM telah berkembang menjadi electronic word-of-mouth. Electronic Word of Mouth (e-WoM) communication merujuk pada pernyataan positif atau negatif dari pelanggan potensial, pelanggan aktual atau mantan pelanggan mengenai suatu produk atau perusahaan via internet. Salah satu motif dalam e-WoM adalah Venting Negative Feelings, yaitu keinginan mengungkapkan ketidakpuasan konsumen terhadap produk atau perusahaan.

Penelitian ini akan menganalisa percakapan tentang venting negative feelings pada media sosial twitter. Untuk mendapatkan data dari twitter akan menggunakan bahasa R yang telah menyediakan library untuk mengaksesnya. Hasil penelitian ini berupa data motif e-wom Venting Negative Feelings yang dapat digunakan sebagai pendukung keputusan pengguna internet dalam memilih penyedia jasa layanan internet yang baik.

Kata Kunci: venting negative feelings, e-wom, twitter, provider, internet, sosial media.

1. PENDAHULUAN

Dalam dunia bisnis, Word of Mouth (WoM) adalah media yang paling kuat dalam mengomunikasikan produk atau jasa kepada konsumen. Word of Mouth adalah tindakan yang dilakukan konsumen untuk memberikan informasi kepada konsumen lain secara non komersial mengenai merek, produk maupun jasa [1]. Konsumen yang telah memiliki pengalaman unik tentang produk, jasa, dan merek dari perusahaan tertentu, cenderung akan memasukkan produk, jasa, dan merek tersebut ke agenda percakapan. Mereka secara sadar atau tanpa sadar mengungkapkannya kepada orang lain secara lisan (word of mouth) dalam berbagai kesempatan.

Menurut Hennig-Thurau [2], electronic Word of Mouth (e-WoM) communication merujuk pada pernyataan positif atau negatif dari pelanggan potensial, pelanggan aktual atau mantan pelanggan mengenai suatu produk atau perusahaan via internet. Salah satu motif dalam e-WoM adalah Venting Negative Feelings, yaitu keinginan mengungkapkan ketidakpuasan konsumen terhadap produk atau perusahaan.

Penelitian ini akan menganalisa percakapan tentang venting negative feelings pada media sosial twitter. Media sosial Twitter dipilih karena media ini adalah salah satu yang paling banyak digunakan sebagai media e-

WoM selain Facebook, MySpace, Friendster, Foursquare, dll. Orang-orang yang bergabung dalam salah satu komunitas jaringan sosial tersebut saling berbagi pengalaman dan pengetahuan mengenai berbagai macam hal. Penerapan EWOM pada media sosial Foursquare mengungkapkan adanya perbedaan tipe pengguna yang mempengaruhi mereka dalam memberikan suatu pernyataan mengenai suatu produk atau jasa. Untuk memanfaatkan kekuatan dari electronic Word of Mouth (e-WoM), perusahaan harus terlebih dahulu mengidentifikasi dan mengerti pengguna web secara efektif untuk menyebarkan pendapat mereka. [3]. Untuk mendapatkan data dari twitter akan menggunakan bahasa R yang telah menyediakan library untuk mengaksesnya. Media sosial Twitter dipilih karena media ini memungkinkan penggunaannya untuk memberikan komentar atau tanggapan tentang suatu produk secara bebas baik itu mendukung atau mengkritiknya.

2. METODE PENELITIAN

Penelitian ini memilih IndiHome sebagai obyek penelitian. IndiHome dipandang sebagai penyedia jasa komunikasi atau internet yang paling banyak digunakan oleh masyarakat. Sosial media

Twitter dipilih karena media ini adalah salah satu yang banyak penggunanya. Selain itu Twitter adalah media sosial yang berbasis teks, sehingga memungkinkan eksplorasi yang lebih luas tentang suatu topik.

Langkah pertama yang akan dilakukan adalah menentukan topik percakapan apa saja yang akan dicari dalam media sosial Twitter. Pada penelitian ini topik ditentukan berdasarkan motif e-WOM yang dianggap sesuai dengan kondisi pada media sosial, yaitu Venting Negative Feelings. Langkah kedua adalah mencari tweet yang sesuai dengan topik yang telah ditentukan. Untuk mendapatkan data tersebut digunakan bahasa R yang dilengkapi dengan library untuk mengakses media sosial Twitter. Langkah ketiga adalah melakukan klasifikasi terhadap data tersebut sesuai dengan motif e-WOM yang telah ditentukan sebelumnya. Klasifikasi dilakukan dengan cara membandingkan kalimat tweet dengan daftar kata yang telah ditentukan dalam motif e-WOM. Jika tweet yang didapat sesuai dengan daftar kata yang ada pada motif, maka bisa disimpulkan bahwa tweet tersebut merupakan bagian dari Venting Negative Feelings. Langkah-langkah penelitian dapat ditunjukkan seperti gambar 1. Untuk mengimplementasikan langkah dalam suatu algoritma program, dimulai dengan pengumpulan data, kemudian dilanjutkan dengan persiapan data, training model dan langkah yang terakhir adalah prediksi.



Gambar 1. Langkah Penelitian

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk membantu perusahaan penyedia jasa komunikasi/internet dalam mendapatkan respon dari pengguna jasanya secara acak. Respon dari pengguna didapatkan dari percakapan yang ada di media sosial Twitter. Seperti yang telah disebutkan sebelumnya bahwa media sosial Twitter mempunyai basis percakapan berupa teks sehingga memungkinkan eksplorasi yang lebih luas tentang suatu topik, khususnya yang dikategorikan sebagai venting negative feelings.

3.1 Pengumpulan Data

Data mengenai kesan pengguna terhadap IndiHome didapat dari media sosial Twitter. Sebagai sampel data, dilakukan pencarian data pada tanggal 2 – 13 Juli 2018. Diambil secara acak sebanyak 500 tweet dan data disimpan dalam bentuk Microsoft Excel untuk kemudahan penyajian informasi. Untuk pengolahan digunakan bahasa R yang telah menyediakan library untuk mendapatkan data dari Twitter yaitu library twitterR. Penggunaannya seperti pada gambar 2.

```

Library (twitterR) api_key<- "kode api" api_secret<- "kode api secret" access_token<- "kode token"
access_token_secret<- "kode token secret"
setup_twitter_oauth(api_key,api_secret,access_token,access_token_secret)

#Pencarian Data
indihome = searchTwitter("indihome", n=500)
indihomedf <- twListToDF(indihome)
write.xlsx(indihomedf, file = "data1.xlsx", sheetName="indihome", append=TRUE)
  
```

Gambar 2. Pencarian Data Twitter

Untuk mendapatkan data komentar pengguna twitter, pada penelitian ini dilakukan dalam dua langkah. Langkah pertama adalah mendapatkan kode-kode untuk mengakses data dari twitter. Untuk mendapatkannya diharuskan memiliki account twitter terlebih dahulu, kemudian mengambil kode-kode tersebut melalui web twitter developer. Langkah kedua adalah menggunakan fungsi

searchTwitter() untuk mencari data yang diinginkan. Pencarian dilakukan dengan metode pencarian tweet yang mengandung kata indihome.

Data yang diperoleh pada saat pencarian bertipe list pada bahasa R, untuk dapat disimpan kedalam Excel, format list harus diubah terlebih dahulu dengan format data frame, setelah itu disimpan ke file excel. Library yang digunakan untuk menyimpan data ke excel adalah library xlsx. Hasil pencarian data dapat dilihat seperti pada gambar 3.

ID	Text	User	Location	Created At	Retweet Count	Reply Count
1	indihome selamat sore indihome...	indihome		2017-07-07 08:00:00	0	0
2	indihome selamat sore indihome...	indihome		2017-07-07 08:00:00	0	0
3	indihome selamat sore indihome...	indihome		2017-07-07 08:00:00	0	0

Gambar 3. Hasil Pencarian Data

3.2 Ekstraksi Data

Pada saat pencarian, library twitterR mengambil data tidak hanya komentar saja, melainkan juga nama pengguna, reply, retweet, tanggal pembuatan, lokasi dan sebagainya. Sedangkan hal yang diperlukan sebenarnya hanyalah field text saja. Oleh karena itu, pada penelitian ini hal pertama yang dilakukan pada tahap ini adalah mengambil text saja. Pada data excel yang didapat dari pencarian, hanya diambil kolom yang menunjukkan percakapannya saja. Data dapat dilihat seperti pada gambar 4.

Row	Text
515	indihome selamat sore indihome...
516	indihome selamat sore indihome...
517	indihome selamat sore indihome...
518	indihome selamat sore indihome...
519	indihome selamat sore indihome...
520	indihome selamat sore indihome...
521	indihome selamat sore indihome...
522	indihome selamat sore indihome...
523	indihome selamat sore indihome...
524	indihome selamat sore indihome...
525	indihome selamat sore indihome...
526	indihome selamat sore indihome...
527	indihome selamat sore indihome...
528	indihome selamat sore indihome...
529	indihome selamat sore indihome...
530	indihome selamat sore indihome...
531	indihome selamat sore indihome...
532	indihome selamat sore indihome...
533	indihome selamat sore indihome...
534	indihome selamat sore indihome...
535	indihome selamat sore indihome...
536	indihome selamat sore indihome...
537	indihome selamat sore indihome...
538	indihome selamat sore indihome...
539	indihome selamat sore indihome...
540	indihome selamat sore indihome...

Gambar 4. Hasil Text

Ekstraksi fitur diperlukan agar kata-kata yang akan diprediksi sentimennya efektif. Dalam bahasa indonesia juga dikenal kata-kata yang tidak memiliki makna atau tidak penting yang umumnya disebut stopword. Penghilangan stopword dilakukan berbasis daftar stopword yang dibuat Tala dengan penambahan kata yang tidak baku dalam bahasa indonesia yang juga tidak penting. Contoh daftar kata yang dimasukkan dalam daftar stopword seperti: ada, adalah, adanya, adapun, agak, agaknya, agar, akan dan sebagainya.

3.3 Persiapan Data

Persiapan data dilakukan dengan cara membuat data yang akan digunakan untuk training dan testing. Persiapan data dilakukan dengan memilih sample sebanyak 20 tweet dengan mengkategorikan motif negatif serta memberikan hasil apakah kalimat tweet tersebut merupakan motif e-WOM Venting Negative Feelings dengan memberi nilai '1' atau '0'. Pada penelitian ini, venting negative feelings yang dipilih adalah kecewa (m1) dan marah(m2). Hasil dari langkah ini dapat dilihat seperti gambar 5.

1	x	m1	m2
2	sobat indihome paket ramadhan lho ju	0	0
3	alesadewi pake speedy cari ganti smpt	1	0
4	alvionitarani speedytrs indihome spee	1	0
5	saksikan aksi udin uda kompetisi indiho	0	0
6	amaliasugianto selviselv indihome ihhh	1	0
7	rudiantorudy telkomcare selamat siang t	0	0
8	dan klo provider yg masuk jogja suam	0	0
9	alesadewi duhh ga ganti indihome deh k	1	0
10	telkomcare mohon info daftar harga ind	0	0
11	akhirnya si temennya bojoku telepon k	0	0
12	telkompromo untung indihome ada beb	0	0
13	saya suka video youtube cara bobol indi	0	1
14	kak nomor telepon memanggil teknisi r	0	0
15	rayindra karna ultah telkom kl yaa chann	0	0
16	kini dgn hybrid box tambahan hdk sobat	0	0
17	hallo warga jatiasih jika keluhan info pe	1	0
18	halo sobat indihome kini menikmati ta	0	1
19	telkonsel males banget nonton hp kuota	1	0
20	indihome segera masuk karimun telkom	0	0

Gambar 5. Persiapan Data

3.4 Training Model

Data training disusun berdasarkan hasil persiapan data. Pada tahap ini peneliti melakukan analisa terhadap tweet untuk memberikan hasil apakah kalimat tweet tersebut merupakan e-WOM atau bukan. Penentuannya e-WOM atau bukan dengan memberi nilai ‘Y’ untuk kondisi benar dan ‘T’ untuk kondisi yang salah. Hasil dari langkah ini dapat dilihat seperti tabel 1.

Tabel 1. Data Training

Tweet	m1	m2	Hasil
sobat indihome paket ramadhan lho juli segera berlangganan lsg	0	0	T
alesadewi pake speedy cari ganti smpt ganti indihome tp liat gini jd ga hehe	1	0	Y
Indihome lemot disini jaringan tdk stabil	1	0	T
amaliasugianto selviselv indihome ihhh malesin bgt	0	1	Y
alesadewi duhh ga ganti indihome deh kalo gitu	1	0	Y
saya warga jatiasih kecewa dengan pemasangan indihome	1	0	Y
males banget nonton hp kuota terbuang yg indihome aja gak ketonton	0	1	Y
Bikin marah download dikit nunggu lama	0	1	Y

3.5 Prediksi

Klasifikasi adalah salah satu teknik machine learning. Teknik ini termasuk ke dalam tipe supervised learning. Istilah klasifikasi didapat dari tujuan utama teknik ini untuk memprediksi sebuah kategori dari input data. Dalam Machine Learning, algoritma Naïve Bayes termasuk tipe Supervised Learning. Pembelajaran pada tipe ini telah diketahui apa yang akan diprediksi atau target variabelnya. Model prediksi digunakan untuk melakukan proses prediksi target variabel (target feature) berdasarkan feature-feature lain pada suatu dataset [4]. Cara kerja algoritma naïve bayes untuk melakukan klasifikasi adalah dengan cara menghitung peluang. Perhitungan peluang berdasarkan rumus Bayesian Rule sebagai berikut:

(1) Implementasi algoritma Naïve Bayes pada lingkungan R dapat menggunakan fungsi naiveBayes() dari package e1071. Jika pada lingkungan R yang digunakan belum memiliki package ini maka package ini perlu diinstall dengan fungsi sebagai berikut:

```
install.packages("e1071")
model = naiveBayes(formula, training_data)
predict(model, testing_data)
```

Keterangan:

- formula, formula untuk menentukan target variable.
- training_data, obyek data training.
- model adalah obyek output dari fungsi naiveBayes().
- testing_data adalah obyek data testing.

Berikut ini adalah implementasi dari Naïve Bayes untuk memprediksi Venting Negative Feelings dari data yang telah diambil dari media sosial Twitter:

```
#proses 1
#membaca data training & data testing
data.training = read.csv("train.1.csv")
data.test = read.csv("test.1.csv")

#proses 2
#membuat model
model = naiveBayes(Benci~, data.training)

#proses 3
#melakukan prediksi
predict_result = predict(model, data.test[,4])
#menampilkan hasil prediksi
comparation_result = cbind(prediction = as.character(predict_result), actual =
as.character(data.test[,4]))
print(comparation_result)
```

Untuk mengetahui ketepatan hasil prediksi maka nilai di atas dapat dibandingkan dengan nilai class sebenarnya dari setiap instance. Berikut adalah cara untuk membandingkan kedua nilai tersebut:

```
comparation_result = cbind(prediction = as.character(model$fitted.values), actual
=as.character(data.test[,5]))
comparation_result
```

Hasil prediksi adalah sebagai berikut:

actual prediction

```
[1,] "T" "T"
[2,] "Y" "Y"
[3,] "Y" "T"
[4,] "T" "T"
[5,] "Y" "Y"
[6,] "T" "T"
[7,] "Y" "T"
[8,] "Y" "Y"
[9,] "T" "Y"
[10,] "Y" "Y"
[11,] "T" "Y"
[12,] "Y" "Y"
[13,] "T" "T"
[14,] "T" "T"
[15,] "Y" "Y"
[16,] "Y" "Y"
[17,] "Y" "Y"
[18,] "Y" "Y"
[19,] "T" "Y"
[20,] "Y" "T"
```

Dari 20 data training dan data test 20 data juga, diperoleh hasil prediksi yang salah sebanyak 6 data. Sehingga prosentasi keberhasilan adalah 14/20 yaitu 70%.

4. KESIMPULAN

Kesimpulan dari penelitian ini adalah sebagai berikut:

- a. Opini pengguna jasa layanan internet dapat diperoleh dari media sosial twitter.
- b. Penentuan Venting Negative Feelings dapat memanfaatkan daftar kata yang telah ditentukan sebelumnya (kamus).
- c. Variasi kamus data dapat menentukan keakuratan hasil prediksi.

DAFTAR PUSTAKA

- [1] Hasan, A, 2010, Marketing dari Mulut ke Mulut. MedPress, Yogyakarta.
- [2] Hennig-Thurau, T., Gwinner, K. P., Walsh, G. and Gremler, D.D. 2004. Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*. Vol. 18 No. 1 2004: 38-52.
- [3] Rita, Karyana Hutomo dan Natalia, 2013, Electronic Word Of Mouth (E-Wom) Foursquare: The New
- [4] Social Media, *Binus Business Review* Vol. 4 No. 2 November 2013: 711-724.
- [5] Faisal, M Reza, 2016, Data Science Klasifikasi Dengan Bahasa Pemrograman R, INDC, Banjarmasin.

3_IMPLEMENTASI NAÏVE BAYES UNTUK KLASIFIKASI VENTING NEGATIF FEELINGS

ORIGINALITY REPORT

21 %
SIMILARITY INDEX

18 %
INTERNET SOURCES

2 %
PUBLICATIONS

17 %
STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

9%
★ 123dok.com
Internet Source

Exclude quotes On
Exclude bibliography On

Exclude matches < 2%