

Anlisa Terpilihnya Tri Rismaharini sebagai Menteri Sosial dengan Pendekatan Algoritma Naïve Bayes

Muhamad Fatchan^{1,*}, Hari Sugeng²,

^{1,*}Teknik Informatika, Universitas Pelita Bangsa
¹fatchan@pelitabangsa.ac.id ²harisugeng23@gmail.com

ABSTRAK

Analisis sentimen merupakan penelitian dari suatu pendapat, atau opini dan luapan perasaan yang di ekspresikan dalam bentuk tekstual. Twitter merupakan sosial media yang memiliki jumlah pengguna yang banyak khususnya di Indonesia twitter sering digunakan untuk meluapkan perasaan maupun tanggapan opini berupa tanggapan positif maupun negatif. Tujuan dari penelitian ini adalah untuk menentukan sentimen masyarakat terhadap terpilihnya Bu Tri Rismaharini menjadi Menteri Sosial pada twitter serta mengetahui tingkat akurasi metode dalam mengklasifikasikan sentimen positif dan sentimen negatif. Metode yang digunakan dalam penelitian ini yaitu Algoritma Naïve Bayes dan pemrosesan data menggunakan aplikasi Rstudio. Hasil dari penelitian ini yaitu mengetahui tingkat akurasi metode Naïve Bayes dalam mengklasifikasikan data sentimen positif dan negatif dengan membagi data penelitian menjadi 3 data testing yaitu 30%, 25%, dan 20% dan jumlah akurasi tertinggi yang ada pada penelitian ini adalah data testing 30% dan data training 70% yaitu dengan tingkat akurasi 90,3%. Dari penelitian ini bisa diketahui bahwa menggunakan aplikasi Rstudio dan metode Naïve Bayes memiliki tingkat akurasi yang baik dalam menentukan sentiment analisis.

Kata kunci: TNaïve Bayes, Analisis sentiment, RStudio, Mensos Risma.

ABSTRACT

Sentiment analysis is a study of an opinion, or opinions and overflow of feelings that are expressed in textual form. Twitter is social media that has many users, especially in Indonesia. Twitter is often used to express feelings and opinion responses in the form of positive or negative responses. The purpose of this research is to determine public sentiment towards the election of Mrs. Tri Rismaharini become the Minister of Social Affairs on twitter and know the level of accuracy of the method in classifying positive sentiment and negative sentiment. Which method used in this research is the Naïve Bayes Algorithm and data processing using the Rstudio application. The results of this study are knowing the level of the accuracy of the Naïve Bayes method in classifying positive sentiment data and negative by dividing the research data into 3 testing data, namely 30%, 25%, and 20% and the highest amount of accuracy in this study is data testing 30% and training data 70% with an accuracy rate of 90.3%. From In this research, using the Rstudio application and the Naïve Bayes has a good level of accuracy in determining sentiment analysis.

Keyword: Naïve Bayes, Sentiment Analysis, RStudio, Social Minister Risma

PENDAHULUAN

Kementerian Sosial Republik Indonesia (Kemensos) adalah kementerian yang mempunyai tugas menyelenggarakan dan membidangi urusan dalam negeri di dalam pemerintahan. Kementerian Sosial dipimpin oleh seorang Menteri Sosial (Mensos) yang sejak tanggal 23 Desember 2020 dijabat oleh Tri Rismaharini.[1]

Mempin koKota Surabaya menjadi kota yang terbaik partisipasinya se-Asia Pasifik pada tahun 2012 versi Citynet atas keberhasilan pemerintah kota dan partisipasi rakyat dalam mengelola lingkungan.[2]

analisis sentimen masyarakat terhadap calon presiden dan wakil presiden Indonesia 2014 yang diungkapkan melalui jejaring sosial Twitter menggunakan naïve bayes [3]

Dari hasil penelitian yang dilakukan menggunakan naïve bayes adalah melakukan text mining pada komentar-komentar yang terkait postingan mengenai efektifitas ganjil genap di tol bekasi pada Twitter, Instagram, Youtube dan Facebook. Ada beberapa tahapan untuk melakukan text minning antara lain selection, preprocessing, transformation, datamining dan evaluation [4]

Metode klasifikasi naïve bayes cukup baik digunakan untuk memprediksi sentimen masyarakat terhadap sistem sekat sekolah karena memiliki akurasi yang cukup tinggi yaitu 0,8079.[5]

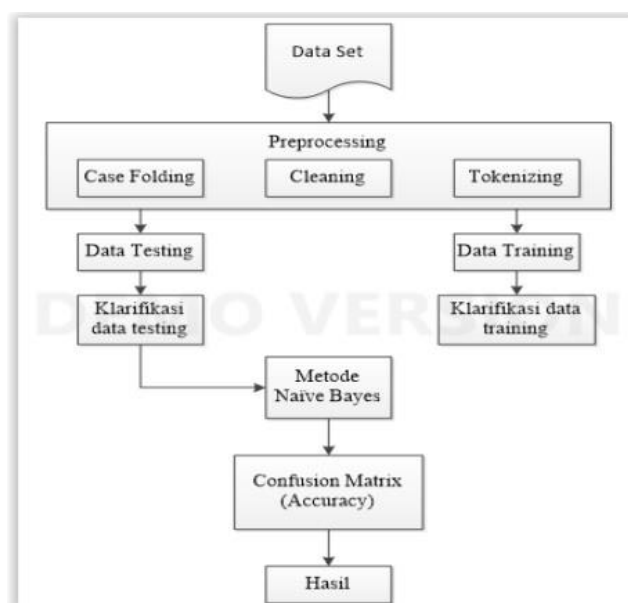
Dari hasil penelitian yang dilakukan dilakukan analisis sentimen terhadap tweet dengan mengklasifikasikan kedalam kategori positif atau negatif menggunakan algoritma Naïve Bayes Classifier. Dari hasil klasifikasi tersebut kita dapat memprioritaskan bantuan dengan mengkategorikan setiap hasil klasifikasi [6]

memperoleh data review film yang bersumber dari komentar film di Twitter. Jumlah opini yang diposting di Twitter perlu dikategorikan menurut sentimennya, sehingga mudah untuk membuat opini tersebut cenderung berupa film, baik opini tersebut cenderung positif maupun negatif. Algoritma yang akan digunakan dalam penelitian ini adalah algoritma Naive Bayes.[7]

Tugas dasar analisis sentimen adalah mengelompokkan teks yang ada ke dalam kalimat atau kemudian dokumen menentukan opini yang mana dinyatakan dalam kalimat atau dokumen apakah itu positif, negatif atau netral dalam beberapa penelitian terutama pada bagian komentar sebelum produknya, tentukan dulu elemen produknya produk didiskusikan sebelum memulai proses opinion minning.[9]

METODE PENELITIAN

Menganalisis sentimen untuk mengetahui sentimen masyarakat terhadap terpilihnya Tri Rismaharini menjadi Menteri Sosial dengan menggunakan metode Naïve Bayes, pada metode ini terdapat dua pemrosesan yaitu pemrosesan data training dan pemrosesan data testing. Langkah awal yang harus dilakukan adalah proses preprocessing dataset. Proses ini bertujuan untuk membersihkan data dan menormalisasi dataset dengan tujuan agar mempermudah melakukan pelabelan sentimen.



Gambar 1 Tahapan Penelitian

Pada metode ini terdapat dua pemrosesan yaitu pemrosesan data training dan pemrosesan data testing. Langkah awal yang harus dilakukan adalah proses preprocessing dataset. Proses ini bertujuan untuk membersihkan data dan menormalisasi dataset dengan tujuan agar mempermudah melakukan pelabelan sentimen. Setelah dilakukan labeling sentimen pada dataset dengan memberikan sentiment positif dan negatif kemudian data dibagi menjadi data training dan data testing. Langkah berikutnya yaitu melakukan model klasifikasi Naïve Bayes terhadap data training, selanjutnya melakukan perhitungan prediksi terhadap data testing sehingga hasilnya akan muncul tabel prediksi confusion matrix dari pengujian data testing dan hasil akurasi metode Naïve Bayes.

Langkah berikutnya yaitu menerapkan metode yang diusulkan yakni algoritma Naïve Bayes. Agar dapat melakukan perhitungan algoritma Naïve Bayes menggunakan software RStudio, maka penulis perlu membagi dataset yang sudah diberi sentimen menjadi dua bagian yaitu data training dan data testing, Pada pengujian ini, penulis membagi pengujian dalam tiga perhitungan yaitu:

Tabel 1: Data Training dan Data Testing

No	Data Training	Data Testing
1	70%	30%
2	75%	25%
3	80%	20%

1. Pengujian data Testing 30%

Tabel 2: Hasil Prediksi Testing 30%

Prediksi	Aktual Kebenaran	
	Negative	Positive
Negative	173	0
Positive	29	98

2. Pengujian Data Testing 25%

Tabel 3: Prediksi Data Testing 25%

Prediksi	Aktual Kebenaran	
	Negative	Positive
Negative	144	0
Positive	20	80

3. Pengujian Data Testing 20%

Table 4: Hasil Prediki Testing 20%

Prediksi	Aktual Kebenaran	
	Negative	Positive
Negative	115	0
Positive	21	64

Dalam penerapan algoritma Naïve Bayes pada software RStudio, ada beberapa package yang akan diperlukan untuk melakukan proses perhitungan algoritma Naïve Bayes di antaranya adalah package “e1071”, package “caret”, dan

package “caTools”. Package “e1071” merupakan package yang digunakan untuk menjalankan metode Naïve Bayes pada RStudio, package “caret” merupakan package yang digunakan untuk melihat apakah metode Naïve Bayes ini baik atau tidak untuk digunakan dilihat dari akurasi, sensitifitas dan spesifisitas, dan package “caTools”

HASIL DAN PEMBAHASAN

Dalam mengklasifikasikan kalimat sentimen positif dan negatif informasi sentiment masyarakat terhadap terpilihnya Tri Rismaharini menjadi Menteri Sosial yang didapat dari media sosial yaitu Twitter Setelah melakukan analisa dan pengujian data yang dibahas pada bab sebelumnya, maka penulis akan memaparkan hasil pengolahan dan pengujian data, proses crawling data pada media sosial twitter dengan menggunakan software RStudio untuk memperoleh data yang digunakan sebagai bahan penelitian. Crawling data menggunakan kata kunci “mentos risma” dengan jumlah data tweet sebanyak 3000 data.

	user_id	status_id	created_at	screen_name	text
1	1051199703503622145	1352206739672686594	2021-01-21 10:49:44	Innayaputri72	Mentos risma antar 15 g
2	18129942	1352201031065640960	2021-01-21 10:27:03	tempodotco	Sebanyak 15 mantan pe
3	18129942	1352161896242114561	2021-01-21 07:51:32	tempodotco	Menteri Sosial Tri Risma
4	18129942	1351221950136410122	2021-01-18 17:36:32	tempodotco	Mentos Risma: Erupsi Gi
5	18129942	1349966320222105600	2021-01-15 06:27:06	tempodotco	Kementerian Sosial telah
6	18129942	1349938084628975616	2021-01-15 04:34:54	tempodotco	Joko Widodo dikabarkan
7	18129942	1350345175638237185	2021-01-16 07:32:32	tempodotco	Potensi Gempa Susulan,
8	18129942	1351313207529361408	2021-01-18 23:39:09	tempodotco	Menteri Sosial Tri Risma
9	748455664968568833	1352200166946832385	2021-01-21 10:23:37	suarapubliknews	Mentos Risma Antarkan
10	780926929	1352197858892500993	2021-01-21 10:14:26	FaqihSyahrir	@republikaonline Pura-j
11	80885324	1352192226567327744	2021-01-21 09:52:03	tribuntimur	Viral foto Mentos Risma
12	799023207294504960	1352188321946976256	2021-01-21 09:36:33	LOVE_AG4EVER	Kala Mentos Risma di nj
13	799023207294504960	1351391249471717377	2021-01-19 04:49:16	LOVE_AG4EVER	Mentos Risma Lari saat i
14	2188143474	1352149204504768513	2021-01-21 07:01:06	Tribun_Jakarta	Mentos Risma Antarkan
15	2188143474	1352186651976507395	2021-01-21 09:29:54	Tribun_Jakarta	Antar Pemulung Bekerja
16	168469708	1352181695852785664	2021-01-21 09:10:13	IDCorner	Mentos Risma Antarkan
17	1856666713	1352177080226594818	2021-01-21 08:51:52	halakhitaT08A	terimakasih Ibu MenSos,
18	1667231432	1352176637027225600	2021-01-21 08:50:07	nawacitadotco	Mentos Risma Antar 15
19	3158354773	1352176634615496704	2021-01-21 08:50:06	nawacitamedia	Mentos Risma Antar 15
20	3158354773	1351694089490731014	2021-01-20 00:52:38	nawacitamedia	Mentos Risma Ikut Bung

Gambar 1: Hasil Crawling Data

Preprocessing dilakukan sebelum proses pemberian sentimen terhadap data. Preprocessing adalah proses untuk membersihkan dataset, menghilangkan variabel yang tidak digunakan, merubah huruf kapital menjadi huruf kecil, menghilangkan tanda #, @, //, URL, username, nomor serta menghapus spasi yang berlebihan.

No	Description
1	mendingan sih bu risma tuker jabatan aja sama pak anies b...
2	seharusnya pengamat politik menyoroti kinerja gubernur s...
3	rupanya belum koordinasi dengan dinsos dki kalau begini ni...
4	sudahlah pak yg bener kerja ngga didoakan tp hoax digada...
5	risma sbnrnya mensos atau mau nyalonin gubernur dki dia ...
6	risma kamu itu mensos bukan walikota tau dirilah adoh2
7	dari postingan ini dari berita2 yang ada di media tentang b...
8	anda itu cuman otaknya d pantat mangnya bu risma sudah ...
9	betul bru sj bu risma jd mensos mereka omong tdk pakai o...
10	manusia goblok anjing omong tdk pakai otak bego bu rism...
11	manusia tolok goblok bego anjing puki mai emangnya bu ris...
12	bego d pelihara emangnya bu risma sdh berapa bulan n be...
13	gubrakkk u0001f600 wow oh oh kamu ketahuan blusukan ...
14	iya bang kalo kaya gitu saran gue jadi mensos aja si bang g...
15	respon mensos risma dituding settingan temul tunawisma ...
16	kio saya jd karyawan mensos di era bu risma pasti bahagia ...
17	3 masalah besar ini perlu diselesaikan mensos risma daripad...
18	blusukannya dituding settingan mensos risma akhirnya buk...
19	dan kata mensos beliau juga sekarang bekerja di perusaha...
20	emangnya bu risma blusukan baru setelah jd mensos atau ...

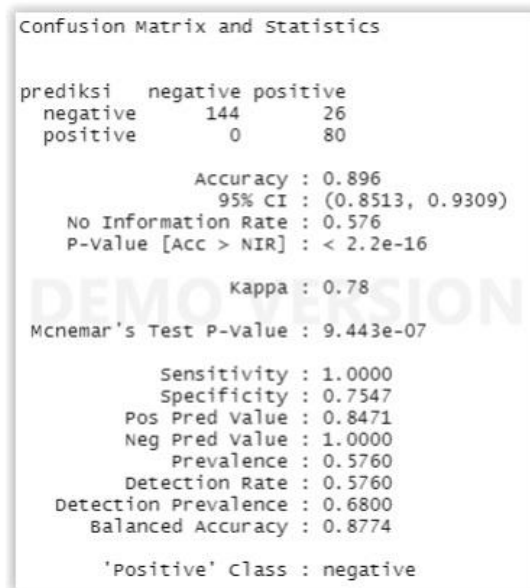
Gambar 2: Hasil Preprocessing Data

Dalam melakukan pengujian pada dataset ini, penulis membagi data menjadi dua yaitu data training dan data testing. Pada penelitian ini dilakukan tiga kali pengujian data dengan rasio yang berbeda-beda, yaitu sebagai berData Training 70% dan Data Testing 30% Pada pengujian ini akan dihitung akurasi terhadap data testing untuk mengetahui seberapa akurat hasil pengujian data tersebut. Berikut ini hasil pengujian Algoritma Naïve Bayes dengan data testing 30% Berikut ini nilai prediksi pada pengujian data testing 30%.

prediksi	negative	positive
negative	144	26
positive	0	80

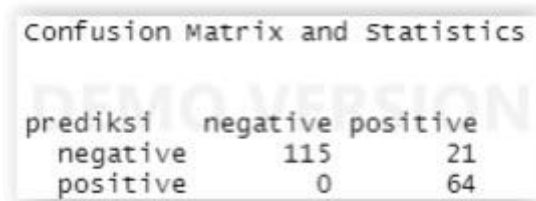
Gambar 3: Predikis Data Testing 25%

Prediksi dari data sentimen negative yang terdeteksi negative adalah 144 dan yang salah adalah 0, sedangkan data sentimen positive yang terdeteksi positive adalah 80 dan yang salah 26.



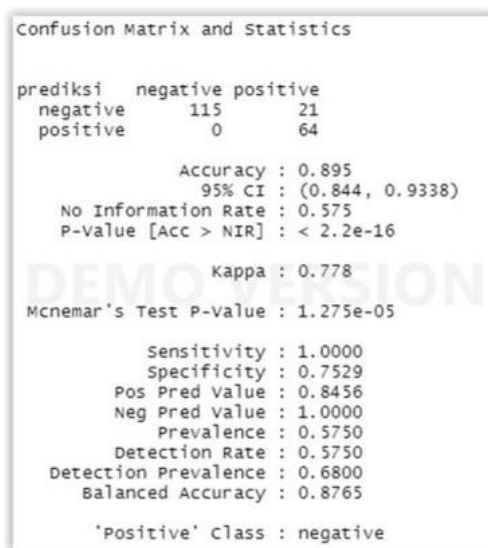
Gambar 4 : Akurasi Data Testing 25%

Dengan cara yang sama seperti data testing 30% untuk menentukan akurasi yang kedua dengan data testing 25%. Berikut adalah hasilnya: Prediksi dari data sentimen negative yang terdeteksi negative adalah 115 dan yang salah adalah 0, sedangkan data sentimen positive yang terdeteksi positive adalah 64 dan yang salah 21



Gambar 5: Predikis Data Testing 20 %

Hasil perhitungan dengan menggunakan algorithmna Naïve Bayes



Gambar 6 : Akurasi Data Testing 20%

Dari hasil perhitungan tiga sampel pembagian data, kemudian hasil dari sistem RStudio diperoleh hasil prediksi nilai accuracy sebagai berikut:

Tabel 5 Akurasi Hasil Prediksi

Data Training	Data Testing	Accuracy
70%	30%	90.33%
75%	25%	89.60%
80%	20%	89.50%

Hasil dari dataset Menteri Sosial Risma menyatakan bahwa tingkat akurasi dari Naïve Bayes menggunakan software RStudio dengan tiga pengujian data adalah 90.33%, 89.60% dan 89.50%. Dapat disimpulkan bahwa presentase hasil tertinggi dari tiga pengujian yang dilakukan adalah 90.33% dari pembagian data testing 30% dan data training 70%.

Setelah mendapatkan hasil akurasi pada pengujian Naïve Bayes, kemudian dilakukan proses pembentukan wordcloud. Wordcloud dilakukan untuk mengetahui frekuensi setiap kata. Untuk menentukan wordcloud, maka dataset yang semula dalam bentuk file excel harus dirubah menjadi bentuk file word, kemudian setiap kalimat dibuat menjadi perkata. Dalam hasil wordcloud, untuk tulisan yang paling besar menandakan bahwa kata tersebut yang paling sering muncul pada teks, begitupun pada ukuran yang lebih kecil menandakan kata tersebut lebih sedikit muncul dalam teks. Pada hasil wordcloud dalam penelitian ini, kata yang paling sering muncul adalah Mensos kemudian risma dan seterusnya hingga kata yang paling sedikit muncul. Disini diambil minimal kata yang dimunculkan pada wordcloud adalah 100 pengulangan kata. Berikut adalah hasil wordcloud pada penelitian ini.



Gambar 7: Wordcloud

KESIMPULAN

Berdasarkan hasil analisis dan pengujian terhadap analisis sentimen menggunakan algoritma Naïve Bayes, maka dapat disimpulkan bahwa : teknik Algoritma Naïve Bayes dapat digunakan untuk melakukan analisis sentimen terhadap terpilihnya Tri Rismaharini sebagai Menteri sosial berdasarkan pendapat dan opini yang ada pada sosial media twitter. Analisis sentimen terhadap terpilihnya Tri Rismaharini sebagai Menteri Sosial menggunakan metode Naïve Bayes menghasilkan nilai akurasi tertinggi 90.33%, Precision sebesar 77,7% dan Sensitivity atau Recall sebesar 99,9%, pada pengujian yang menggunakan dataset 30% data testing dan 70% data training. sedangkan

nilai akurasi terendah 89.5%, Precision sebesar 75,3% dan Sensitivity atau Recall sebesar 99,9% dengan pengujian menggunakan dataset 20% data testing dan 80% data training. Proses klasifikasi semakin akurat jika data training yang digunakan dalam pengujian berjumlah banyak akan tetapi dapat juga mengurangi keakuratan data jika kata-kata yang terdapat pada tweet mengalami bias atau makna ganda.

DAFTAR PUSTAKA

- [1] Kementerian sosial republik indonesia - wikipedia bahasa indonesia, ensiklopedia bebas.” [online]. available: https://id.wikipedia.org/wiki/kementerian_sosial_republik_indonesia. [accessed: 22-jan-2021].
- [2] Profil tri rismaharini, wali kota surabaya yang jadi menteri sosial.” [online]. available: <https://nasional.kompas.com/read/2020/12/22/15543161/profil-trismaharini-wali-kota-surabaya-yang-jadi-menteri-sosial>. [accessed: 20-jan 2021].
- [3] f. nurhuda and s. w. sihwi, “analisis sentimen masyarakat terhadap calon presiden indonesia 2014 berdasarkan opini dari twitter menggunakan metode naive bayes classifier,” *itsmart j. ilm. teknol. dan inf.*, vol. 2, no. 2, pp. 35–42, 2014.
- [4] h. s. utama, d. rosiyadi, d. aridarma, and b. s. prakoso, “sentimen analisis kebijakan ganjil genap di tol bekasi menggunakan algoritma naive bayes dengan optimalisasi information gain,” *j. pilar nusa mandiri*, vol. 15, no. 2, pp. 247–254, 2019.
- [5] n. n. rosyad, “analisis sentimen publik terhadap sistem zonasi sekolah menggunakan data twitter dengan metode naive bayes classification,” vol. 12, no. 4, pp. 315–322, 2019.
- [6] i. fahrur rozi, a. taufika firdausi, and k. islamiyah, “analisis sentimen pada twitter mengenai pasca bencana menggunakan metode naive bayes dengan fitur n-gram,” *j. inform. polinema*, vol. 6, no. 2, pp. 33–39, 2020.
- [7] f. ratnawati, “implementasi algoritma naive bayes terhadap analisis sentimen opini film pada twitter,” *inovtek polbeng - seri inform.*, vol. 3, no. 1, p. 50, 2018
- [8] m. chalida, “analisis sentimen ujaran kebencian pemilihan presiden 2019 menggunakan algoritma naive bayes (studi kasus: tweet #pilpres2019 di kota jakarta, bandung, semarang, surabaya dan yogyakarta),” *jnanaloka (jurnal open access yayasan lentera dua indones.*, no. 2001, pp. 5–10, 2019