JOURNAL SHIFT
VOL 2 NO 2 (2022)
ISSN 2776-8961 (ONLINE)
2808-3385 (PRINT)

Analisis Sentimen Terhadap Jasa Ekspedisi Pos Indonesia Pada Sosial Media Twitter Menggunakan *Naïve Bayes Classifier*

Muhammad Nur Akbar[1*], Darmatasia[2], Yulia Ardana[3]
[1,2,3]Jurusan Teknik Informatika, Universitas Islam Negeri Alauddin Makassar
[1]muhammad.akbar@uin-alauddin.ac.id, [2]darmatasia@uin-alauddin.ac.id, [3]60200117043@uin-alauddin.ac.id

Informasi Artikel

Article historys:

Diterima Juni 19, 2022
Disetujui Juni 29, 2022
Dipublikasi Juni 30, 2022

Kata Kunci:

Sentiment Analysis
Twitter
Naïve Bayes
Text Mining

ABSTRACT

Nowdays, the rapid growth of information technology positively impacts companies engaged in industry, sales, and services, especially e-commerce. The increase in the number of transactions in various e-commerce impacts the increase in the use of expedition services. Pos Indonesia is the oldest expedition service provider in Indonesia and is required to be able to innovate in providing the best service for its customers. The importance of customers for a company depends on how the company builds customer relationships. A strong company will have good customer relations. Many customers have expressed their opinions regarding Pos Indonesia through Twitter. In this study, text mining techniques are used, namely sentiment analysis which helps analyze opinions, sentiments, evaluations, assessments, attitudes, and public emotions towards Pos Indonesia services. Naïve Bayes Classifier was chosen because it is simple, fast, and has high accuracy. The Naïve Bayes Classifier has successfully classified positive and negative sentiments on 100 tweets from Pos Indonesia customers with an accuracy of 87%.

Koresponden Author:

Muhammad Nur Akbar,
Jurusan Teknik Informatika,
Universitas Islam Negeri Alauddin Makassar,
Jl. H.M. Yasin Limpo No. 36 Samata, Kab Gowa, Sulawesi Selatan, Indonesia.
Email: muhammad.akbar@uin-alauddin.ac.id

1. PENDAHULUAN

Kemajuan teknologi informasi dan internet berkembang begitu pesat di segala bidang kehidupan. Membuat suatu organisasi memiliki ketergantungan terhadap teknologi informasi[12]. Kemudahan serta fitur yang lengkap menjadi alasan sebuah organisasi tidak mungkin tanpa menggunakan suatu teknologi informasi untuk menyelesaikan tugas-tugasnya. Demikian pula dengan organisasi yang bergerak dalam bidang jasa ekspedisi barang, teknologi informasi memiliki perananan penting. Setiap barang yang akan dikirim harus dikontrol dengan berbasis teknologi

ISSN 2776-8961 (ONLINE)
2808-3385 (PRINT)

informasi, mulai dari barang diserahkan pengirim ke petugas ekspedisi, hingga barang sampai ke alamat tujuan. Di sisi lain teknologi internet juga ikut mendorong jumlah transaksi penjualan online pada berbagai *e-commerce* berdampak pada lonjakan penggunaan jasa ekpedisi. Selain digunakan untuk mempermudah menyelesaikan tugas, teknologi informasi dan internet juga dapat dimanfaatkan dalam menambah daya saing yaitu meningkatkan jumlah pelanggan.

Pentingnya pelanggan bagi sebuah perusahaan tergantung kepada bagaimana perusahaan menjalin hubungan dengan pelanggan. Perusahaan yang kuat akan memiliki relasi pelanggan yang baik[9]. Banyak pelanggan yang menuangkan opini mereka terkait jasa ekspedisi barang yang ada di Indonesia melalui media sosial. Media sosial adalah sebuah layanan yang memfasilitasi dalam pertukaran informasi dan topik secara berkelanjutan dengan cakupan yang luas. Salah satu media sosial yang popular di masyarakat adalah twitter.

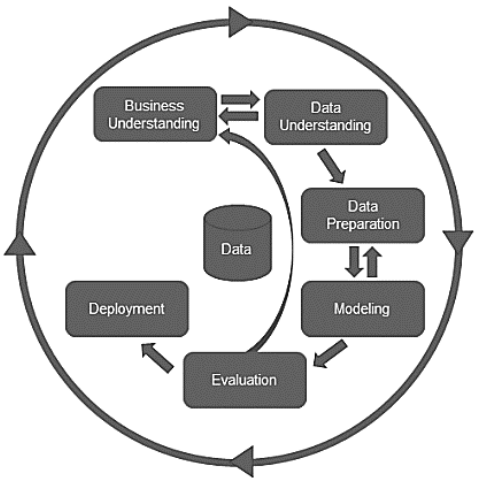
Twitter adalah media sosial yang bisa menghubungkan banyak orang dengan berbagai topik dari seluruh dunia. Dengan menggunakan twitter, masyarakat dapat memberikan pendapat mereka tentang apapun yang terjadi secara langsung[3]. Penelitian ini mencoba memanfaatkan twitter dengan menganalisis tweet berbahasa Indonesia yang membicarakan tentang jasa ekspedisi barang di Indonesia. Jasa ekspedisi barang yang dianalisis merupakan salah satu jasa ekspedisi tertua di Indonesia yaitu Pos Indonsia. Nantinya penelitian ini akan menghasilkan ulasan dengan kategori kelas positif yang mengandung kata-kata yang baik dan mendukung, atau kategori kelas negatif yang mengandung kata-kata yang buruk. Dari ulasan tersebut diharapkan dapat membantu pemegang kebijakan dalam mengevaluasi kinerja perusahaan melalui tanggapan pelanggan.

Pada penelitian analisis sentimen ini, dibahas bagaimana mengimplementasikan analisis sentimen untuk menentukan kecenderungan pandangan publik terhadap jasa ekspedisi barang dengan menggunakan metode klasifikasi *Naïve Bayes Classifier*. Naïve Bayes merupakan metode pembelajaran mesin yang memiliki model dalam membentuk probalilitas dan peluang. Maka dari itu, Naïve Bayes akan menghitung probabilitas kemunculan kata yang mempresentasikan komentar berdasarkan kelas positif maupun negatif. Berdasarkan hasil klasifikasi dilakukan pengujian untuk mendapatkan nilai akurasi dari hasil prediksi analisis sentimen dengan metode naïve bayes[13].

2. METODE PENELITIAN

Penelitian ini menggunakan metode *Cross Industry Standard Process for Data Mining* atau yang biasa disingkat menjadi CRISP-DM. CRISP-DM dianggap teknologi yang netral, industri independen dan merupakan standar de-facto untuk *data mining* [2].

2.1. Model CRISP-DM



Gambar 1. Model CRISP-DM

Proses data mining berdasarkan CRISP-DM terdiri dari 6 fase. Yaitu:

ISSN 2776-8961 (ONLINE)
2808-3385 (PRINT)

1. *Business Understanding*
   Merupakan pemahaman tentang substansi dari kegiatan *data mining* yang akan dilakukan, kebutuhan dari perspektif bisnis. Kegiatannya antara lain: menentukan sasaran atau tujuan bisnis, memahami situasi bisnis, menentukan tujuan *data mining* dan membuat perencanaan strategi serta jadwal penelitian.
2. *Data Understanding*
   Fase mengumpulkan data awal, mempelajari data untuk bisa mengenal data yang akan dipakai, mengidentifikasikan masalah yang berkaitan dengan kualitas data, mendeteksi subset yang menarik dari data untuk membuat hipotesa awal.
3. *Data Preparation*
   Fase yang padat karya. Aktivitas yang dilakukan antara lain memilih *table* dan *field* yang akan ditransformasikan ke dalam database baru untuk bahan *data mining* (set data mentah).
4. *Modeling*
   Fase menentukan tehnik *data mining* yang digunakan, menentukan *tools data mining*, teknik *data mining*, algoritma *data mining*, menentukan parameter dengan nilai yang optimal.
5. *Evaluation*
   Fase interpretasi terhadap hasil *data mining* yang ditunjukan dalam proses pemodelan pada fase sebelumnya. Evaluasi dilakukan secara mendalam dengan tujuan menyesuaikan model yang didapat agar sesuai dengan sasaran yang ingin dicapai dalam fase pertama.
6. *Deployment*
   Fase penyusunan laporan atau presentasi dari pengetahuan yang didapat dari evaluasi pada proses *data mining*[14].

2.2. *Text Mining*

Text mining dapat diartikan sebagai penemuan informasi yang baru dan tidak diketahui sebelumnya oleh komputer, dengan secara otomatis mengekstrak informasi dari sumber- sumber yang berbeda. Kunci dari proses ini adalah menggabungkan informasi yang berhasil diekstraksi dari berbagai sumber[7]. Sedangkan menurut Harlian-Milkha[5] *text mining* memiliki definisi menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen[6].

Tipe pekerjaan *text mining* meliputi kategorisasi*, text clustering*, ekstraksi konsep/entitas, analisis sentimen, *document summarization*, dan *entity-relation modeling* (yaitu, hubungan pembelajaran antara entitas). Sumber data yang digunakan pada *text mining* adalah kumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur. Tujuan dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen[13].

Dengan *text mining*, tugas-tugas yang berhubungan dengan penganalisaan teks dengan jumlah yang besar, penemuan pola, serta penggalian informasi yang mungkin berguna dari suatu teks dapat dilakukan[6]. Sebagai bentuk aplikasi dari *text mining*, analisis sentimen jasa ekspedisi barang menggunakan ulasan masyarakat pada media sosial twitter sebagai sumber data.

2.3 Twitter

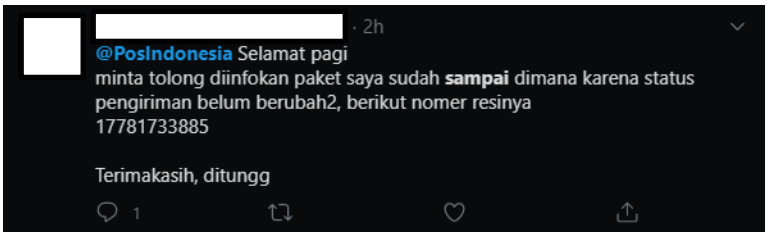
Twitter adalah layanan microblogging yang dirilis secara resmi pada 13 Juli 2006[8]. Aktifitas utama twitter adalah mem-posting sesuatu yang pendek (tweet) melalui web atau mobile. Panjang maksimal dari tweet adalah 140 karakter, kira-kira seperti panjang karakter dari judul koran. Twitter menjadi sumber yang hampir tak terbatas yang digunakan pada text classification sebab

ISSN 2776-8961 (ONLINE)
2808-3385 (PRINT)

memiliki karakteristik yang sangat beragam. Pesan pada twitter memiliki banyak attribute yang unik, yang membedakan dari media sosial lainnya :

- Twitter memiliki maksimal panjang karakter yaitu 140 karakter.
- Twitter menyediakan data yang bisa diakses secara bebas dengan menggunakan Twitter API, mempermudah saat proses pengumpulan tweets dalam jumlah yang sangat banyak.
- Pengguna twitter mem-posting pesan melalui banyak media yang berbeda. Frekuensi dari salah ejaan, bahasa gaul dan singkatan lebih tinggi daripada media sosial lainnya.
- Pengguna twitter mengirim pesan singkat tentang berbagai topik yang disesuaikan dengan topik tertentu dan itu berlaku secara global.

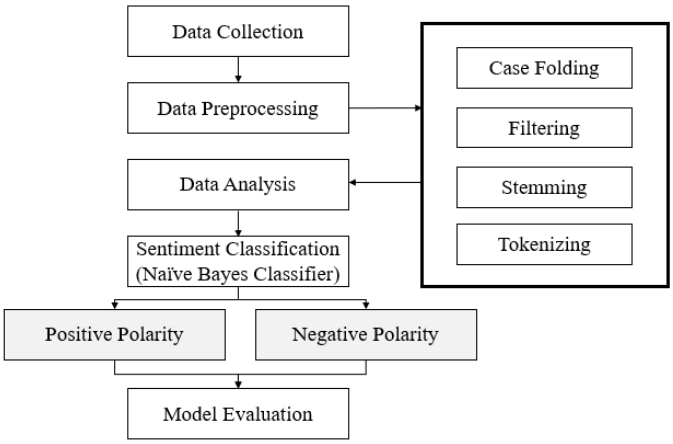
Selama beberapa tahun terakhir, twitter menjadi sangat populer. Jumlah pengguna twitter telah naik menjadi 190 juta dan jumlah tweet yang dipublikasikan di twitter setiap hari adalah lebih dari 65 juta[10].



Gambar 2. Contoh *tweet* ulasan pelanggan Pos Indonesia

2.4 Analisis Sentimen

Lebih spesifik pada penelitian ini langkah-langkah proses dalam melakukan analisis sentimen dapat dilihat pada gambar berikut.



Gambar 3. Alur proses analisis sentimen menggunakan *Naïve Bayes Classifier*

Analisis sentimen disebut juga *opinion mining*, adalah bidang ilmu yang menganalisa pendapat, sentimen, evaluasi, penilaian, sikap dan emosi publik terhadap entitas seperti produk, jasa, organisasi, individu, masalah, peristiwa, topik, dan atribut mereka. Analisis sentimen berfokus pada opini- opini yang mengekspresikan atau mengungkapkan sentimen positif atau negatif[13].

1. Proses Analisis Sentimen

ISSN 2776-8961 (ONLINE)
2808-3385 (PRINT)

Menurut Buntoro, Adji dan Purnamasari sebuah tweet memiliki kandungan sentimen positif dan negatif. Analisis sentimen dapat digunakan untuk mengidentifikasi kandungan sentimen pada tweet[11]. Parameter sentimen yang digunakan pada sistem analisis sentimen ini yaitu positif dan negatif. Tahapan sistem analisis sentimen yang dibangun yaitu pengumpulan data, preprocessing, pembobotan fitur, dan klasifikasi class dengan menggunakan metode naïve bayes. Uji coba sistem dilakukan dengan jumlah data 100 tweets, untuk mengetahui seberapa baik sistem yang telah dibangun. Pada proses klasifikasi, sistem menghitung nilai sentimen setiap tweet dan mencocokkan dengan 2 parameter sentimen (positif/negatif).

2. *Crawling Data*

Crawling data di twitter adalah sebuah proses untuk mengambil atau mengunduh data dari server twitter dengan melakukan *connecting* ke *Application Programming Interface* (API) twitter untuk mendapatkan *token access* yang menjadi syarat untuk mengumpulkan dataset twitter.
Crawling data ini dilakukan untuk mengambil data dari twitter dimana data tersebut dibutuhkan untuk penelitian ini. Cara melakukan crawling data ialah dengan membuat program dengan memasukkan kata kunci untuk mencari tweet sesuai yang kita inginkan. Misalnya "@posindonesia", maka program akan mengambil tweet yang mention ke akun Pos Indonesia. Kumpulan tweet tersebut merupakan data yang akan dianalisis nantinya.

Tabel 1. Contoh Dokumen Hasil *Crawling*

Tweet
{"created_at": "Sat May 23 23:18:26 +0000 2020", "id": 1264334904663306240, "id_str": "1264334904663306240", "text": "@PosIndonesia Udah dibls dari kemarin kok kak. Makasih y", "truncated": false, "entities": …..}
{"created_at": "Sat May 23 10:59:25 +0000 2020", "id": 1264148925226893312, "id_str": "1264148925226893312", "text": "@PosIndonesia Kapok pake \n@PosIndonesia \n dari tanggal 16 mei sampe sekarang belum sampe...", "truncated": false, "entities":…..}

3. *Preprocessing*

Preprocessing dilakukan untuk menghindari data yang kurang sempurna, gangguan pada data, dan data-data yang tidak konsisten[1]. Berikut adalah tahap dari text preprocessing:

- *Case Folding*: Dilakukan proses konversi teks menjadi bentuk standar, biasanya diubah menjadi lower case pada dokumen. Selain itu, juga dilakukan penghapusan seluruh angka dan karakter selain huruf.
- *Filtering*: Dilakukan dengan menggunakan algoritma stopword removal. Stopword removal digunakan untuk membuang kata-kata yang sering muncul, bersifat umum, dan kurang menunjukan relevansinya dengan teks[13].
- *Stemming*: Proses mencari akar kata dan menghilangkan imbuhan pada kata[13]. Bertujuan untuk mengurangi variasi kata yang memiliki kata dasar sama.
- *Tokenizing*: Proses pemotongan string input berdasarkan tiap kata penyusunnya dan memecah sekumpulan karakter dalam suatu teks ke dalam suatu kata. Pemisah dapat berupa spasi, enter, dan tabulasi[15].

ISSN 2776-8961 (ONLINE)
2808-3385 (PRINT)

Tabel 2. Contoh Hasil *Preprocessing*

Dokumen Awal	@PosIndonesia TOLONG dibantu, tanggal 19 tidak ada pergerakan cek pesan.
Case Folding	tolong di bantu tanggal tidak ada pergerakan cek pesan
Filtering	tolong bantu tanggal ada pergerakan cek pesan
Stemming	tolong bantu tanggal ada gerak cek pesan
Tokenizing	tolong bantu tanggal ada gerak cek pesan

4. *Term Weighting*

TF-IDF adalah metode pembobotan yang mengaitkan antara *term frequency* (TF) dan *inverse document frequensi* (IDF). Langkah awal pada pembobotan TF-IDF adalah menemukan nomor kata yang diketahui sebagai bobot atau *frequency term* di setiap dikumen setelah dilakukan pengalianoleh *inverse deocument frequency*. Adapun rumus untuk menemukan bobot dari kata menggunakan TF-IDF adalah :

a) *TF (Term Frequency): Term Frequency* adalah cara pembobotan *term* (kata) yang paling sederhana. Bobot kata t pada dokumen diberikan dengan :

$$w_{ij} = tf_{ij} . idf \qquad (1)$$

b) *IDF (Inverse Document Frequency):* Jika TF memperhatikan kemunculan kata dalam dokumen, IDF memperhatikan kemunculan kata pada kumpulan dokumen. Faktor IDF pada suatu kata t diberikan oleh :

$$idf = log \frac{N}{df_j} \qquad (2)$$

Dimana *wi j* adalah bobot kata *i* pada dokumen *j,* semantaraN adalah jumlah dokumen, dan *tfi j (term frequency)* adalah jumlah dari kemunculan kata *i* pada dokumen *j*, *dfj (document frequency)* adalah jumlah dokumen *j* yang berisi kata *i*.

5. *Naïve Bayes Classifier*

Merupakan metode klasifikasi statistik yang dapat memprediksi probabilitas keanggotaan kelas, seperti probabilitas bahwa sampel yang diberikan termasuk dalam kelas tertentu. Metode ini yang akan digunakan pada penelitian ini untuk klasifikasi data yang diambil dari twitter dan data tersebut akan diklasifikasikan menjadi kelas positif dan kelas negatif. *Naïve Bayes Classifier* merupakan sebuah metode klasifikasi yang berakar pada teorema bayes, yaitu memprediksi peluang di masa yang akan datang berdasarkan pengalaman di masa sebelumnya. Ciri utama dari naïve bayes classifier ini adalah asumsi yang sangat kuat dari masing- masing kondisi atau kejadian. Metode ini sangat cocok digunakan sebagai pengklasifikasian sentimen pada tugas

ISSN 2776-8961 (ONLINE)
2808-3385 (PRINT)

akhir ini dikarenakan memiliki beberapa kelebihan antara lain, sederhana, cepat, dan berakurasi tinggi[3].

Pada pengujian *Naïve Bayes Classifier* ini menggunakan *library* NLTK. Tujuan pengujian ini adalah untuk dapat secaraotomatis mengklasifikasikan *tweet* sebagai sentimen *tweet* yang positif atau negatif. Pada implementasinya terdapat tiga tahap yaitu membuat daftar tupel tunggal, membuat daftar fiturkata, dan membuat *classifier*.

Naïve bayes classifier menggunakan *prior probability* (yaitu nilai probabilitas yang diyakini benar sebelum melakukan eksperimen) dari setiap label yang merupakan frekuensi masing-masing label pada *training set* dan kontribusi dari masing-masing fitur. Berdasarkan dari ciri alami dari sebuah model probabilitas, klasifikasi *naïve bayes* bisa dibuat lebih efisien dalam bentuk pembelajaran *supervised*.

Dalam beberapa bentuk praktiknya, parameter untuk perhitungan model *naïve bayes* menggunakan metode *maximum likelihood* atau kemiripan tertinggi. Untuk ranah klasifikasinya yang dihitung adalah P(H|X), yaitu peluang bahwa hipotesa benar (*valid*) untuk data *sample*X yang diamati dapat diterapkan pada persamaan di bawah[4].

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (3)$$

X = Data dengan kelas tidak dikenal
H = Hipotesis data X adalah kelas khusus
P(H|X) = Probabilitas hipotesis H didasarkan pada kondisi X
P(H) = Probabilitas hipotesis H
P(X|H) = Probabilitas hipotesis X didasarkan pada kondisi H
P(X) = Probabilitas X

3. HASIL PENELITIAN DAN PEMBAHASAN

3.1. Hasil *Preprocessing*

Preprocessing dilakukan di Jupyter Notebook dengan menggunakan bahasa pemrograman python. Terdapat 100 data yang akan di analisis dan class sebanyak 2, yaitu 1 (positif) dan 0 (negatif). Berikut merupakan hasil preprocessing dari setiap langkah- langkahnya:

1. Import dataset ke dalam Jupiter Notebook

Tabel 3. Contoh Dataset

No.	*Tweet*	*Class*
1	@PosIndonesia Sudah di balas kemarin kakak. Terima kasih,	1
2	@PosIndonesia itu sudah seminggu sejak pengiriman kakak, isinya makanan kalau rusak di jalan bagaimana??	0
…	…	…
99	Paketku sampai sesuai pesanan @PosIndonesia	1
100	@PosIndonesia saya capek menunggu paket saya..	0

ISSN 2776-8961 (ONLINE)
2808-3385 (PRINT)

2. Menghilangkan *username* akun twitter

Tabel 4. Contoh input output data hasil menghilangkan *username*

Input		Output
@PosIndonesia BAGUS!	→	BAGUS!
@PosIndonesia tidak bergerak resi 17634248840	→	tidak bergerak resi 17634248840

3. Menghilangkan tanda baca, angka, dan karakter spesial

Menghapus seluruh tanda baca (-, ?, !, dan lain-lain), angka (1-0), dan karakter spesial (+, $, ^, dan lain-lain) yang terdapat dalam dokumen.

Tabel 5. Contoh input output data hasil menghilangkan tanda baca, angka, dan karakter spesial

Input		Output
BAGUS!	→	BAGUS
tidak bergerak resi 17634248840	→	tidak bergerak resi

4. *Tokenizing* dan *case folding*

Tabel 6. Contoh input output data hasil *tokenizing* dan *case folding*

Input		Output
BAGUS	→	bagus
tidak bergerak resi	→	tidak bergerak resi

5. *Filtering*

Tabel 7. Contoh input output data hasil *filtering*

Input		Output
bagus	→	bagus
tidak bergerak resi	→	bergerak resi

6. *Stemming*

Tabel 8. Contoh input output data hasil *stemming*

Input		Output
bagus	→	bagus
bergerak resi	→	gerak resi

3.2. Data

Data yang digunakan dibagi menjadi 2, yaitu:

a. Data *training*: Berfungsi untuk melatih algoritma, pembentukan kelas, dan sebagai acuan bagaimana dokumen akan diklasifikasikan. Digunakan 80% dari total data.
b. Data *testing*: Berfungsi untuk mengetahui performa algoritma yang sudah dilatih sebelumnya Ketika menemukan data baru yang sebelumnya belum ditemui. Digunakan 20% dari total data.

ISSN 2776-8961 (ONLINE)
2808-3385 (PRINT)

3.3. Vektorisasi

Vektorisasi bertujuan untuk mengubah teks dalam tweet menjadi angka yang bisa diolah dan dicari polanya pada saat proses klasifikasi. Terdapat dua proses dalam vektorisasi yaitu tokenisasi dan pembobotan setiap kata dengan metode TD/IDF.

Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan suatu metode yang bertujuan memberi bobot hubungan suatu kata (term) terhadap dokumen/tweet. Perhitungan TF-IDF menggunakan library di Sklearn python yaitu TfidfVectorizer().

3.4. Hasil Prediksi

Berikut hasil analisa pengujian sistem secara keseluruhan:

	precision	recall	f1-score	support
0	0.82	1.00	0.90	18
1	1.00	0.67	0.80	12
accuracy			0.87	30
macro avg	0.91	0.83	0.85	30
weighted avg	0.89	0.87	0.86	30

Gambar 4. Performa Hasil Pengujian

Dari hasil pengujian tweet Pos Indonesia menggunakan metode *Naïve Bayes*, diperoleh tingkat akurasi sebesar 0,87 atau 87%.

4. KESIMPULAN

Beberapa kesimpulan yang diperoleh dari penelitian ini, antara lain:

1) Algoritma Naïve Bayes relatif memiliki performa yang baik dalam melakukan klasifikasi sentimen tweet Pos Indonesia dengan tingkat akurasi sebesar 87%
2) Komentar-komentar yang ada di media sosial khususnya Twitter dapat diidentifikasi apakah positif atau negative dengan membuat model klasifikasi.
3) Dengan mengetahui jenis sentimen sebuah komentar, pemangku kebijakan Pos Indonesia diharapkan dapat terbantu dalam membuat keputusan bisnis sebagai bentuk tindak lanjut terhadap *feedback* dari pelanggan. Hal yang baik dapat pertahankan maupun ditingkatkan, sedangkan hal yang dianggap kurang baik dapat segera dibenahi.

DAFTAR PUSTAKA

[1] A. F. Hidayatullah and S. N. Azhari. (2014). Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik pada Twitter. Seminar Nasional Informatika, Agustus 2014.
[2] Azevedo, A. and Santos, M.F. (2008) KDD, SEMMA and CRISP-DM: A Parallel Overview. Proceedings of the IADIS European Conference Data Mining, Amsterdam, 24-26 July 2008, 182-185.
[3] B. M. Pintoko and K. Muslim L. (2018). Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier. e- Proceeding of Engineering, vol. 5, no. 3, Desember 2018.
[4] F. Ratnawati. (2018). Implementasi Algoritma Naïve Bayes Terhadap Analisis Sentimen Opini Film ada Twitter. Jurnal INOVTEK POLBENG - Seri Informatika, vol. 3, no. 1, Juni 2018.
[5] Harlian, Milka. (2006). Machine Learning Text Categorization. Austin : University of Texas.
[6] H. Februariyanti and E. Zuliarso. (2012). Klasifikasi Dokumen Berita Teks Bahasa Indonesia Menggunakan Ontologi. Jurnal Teknologi Informasi DINAMIK, vol. 17, no. 1, Januari 2012.

[7] Marti Hearst. (2003). What Is Text Mining? Retrieved from http://people.ischool.berkeley.edu/~hearst/text-mining.html
[8] Mostafa, M. (2013). More than words: Social networks ‟text mining for consumer brand sentiments"
[9] N. Haqqizar and T. N. Larasyanti. (2019). Analisis Sentimen Terhadap Layanan Provider Telekomunikasi Telkomsel di Twitter Dengan Metode Naïve Bayes. Prosiding TAU SNAR-TEK 2019 Seminar Nasional Rekayasa dan Teknologi, November 2019.
[10] Ravichandran, M., & Kulanthaivel, G. (2014). Twitter Sentiment Mining (TSM) Framework Based Learners Emotional State Classification And Visualization For E-Learning System. Journal of Theoretical and Applied Information Technology , 69 (1), 84-90.
[11] R. Habibi, D. B. Setyohadi and Ernawati. (2016). Analisis Sentimen pada Twitter Mahasiswa Menggunakan Metode Backpropagation. INFORMATIKA, vol. 12, no. 1, April 2016.
[12] R. Raksanagara, Y. H. Chrisnanto and A. I. Hadiana. (2016). Analisis Sentimen Jasa Ekspedisi Barang Menggunakan Metode Naïve Bayes. Prosiding SNST, 2016.
[13] S. Gusriani, K. D. K. Wardhani and M. I. Zul. (2016). Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes (Studi Kasus: Facebook Page Berrybenka).
[14] Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. Journal of Data Warehousing, 5, 13-22.
[15] Utami, P. D. (2018). Analisis Sentimen Review Kosmetik Bahasa Indonesia Menggunakan Algoritma Naïve Bayes.