

Implementasi Teori Naive Bayes dalam Klasifikasi Ujaran Kebencian di Facebook

Willianto¹, Izmy Alwiah Musdar², Junaedy³, Husni Angriani⁴

^{1,2,3,4}Program Studi Informatika, STMIK KHARISMA Makassar

Jl. Baji Ateka No. 20, Makassar, Indonesia 90134

e-mail: ¹willianto_16@kharisma.ac.id, ²izmyalwiah@kharisma.ac.id, ³junaedy@kharisma.ac.id,

⁴husniangriani@kharisma.ac.id

Submitted Date: August 15th, 2021

Revised Date: January 10th, 2022

Reviewed Date: January 05th, 2022

Accepted Date: January 31st, 2022

Abstract

Hate Speech can be orally or in writing which is expressed intentionally by someone for the purpose of spreading and leading to hatred between groups of people. The phenomenon of Hate Speech has become a hot topic. This is motivated by netizens who often express Hate Speech either in the comments column or in their personal status on social media. The impact of this phenomenon is the emergence of hatred in society which can lead to conflict. The purpose of this study is to implement the Naïve Bayes Theory in the classification of Hate Speech on Facebook. In this study Naïve Bayes is used as a Classifier. Naïve Bayes method is applied to find the probability of words in documents would be categorized as hate speech or not hate speech. This Classifier is implemented using Python programming language. In the Classifier design stage, 500 data are collected randomly on Facebook. Data is divided by 80% - 20% , 400 text data for training and 100 text data for testing. The accuracy for hate speech classification in this study is 83%. These results are obtained from Classifier evaluations using test data where the Classifier correctly labels 83 out of 100 test data.

Keywords: Hate Speech Classification; Naïve Bayes.

Abstrak

Ujaran Kebencian dapat berbentuk ucapan atau tulisan yang dengan sengaja dilontarkan oleh seseorang dengan tujuan menyebarkan dan menimbulkan kebencian antar kelompok orang. Fenomena Hate Speech atau ujaran kebencian menjadi perbincangan hangat. Hal ini dipicu oleh banyaknya warganet yang sering melontarkan ujaran kebencian melalui komentar atau status pribadi di media sosial. Dampak dari fenomena ini yaitu timbulnya kebencian dalam masyarakat yang dapat mengakibatkan konflik. Penelitian ini bertujuan untuk mengimplementasi Teori Naïve Bayes dalam klasifikasi ujaran kebencian di Facebook. Penelitian ini menggunakan metode Naïve Bayes sebagai *Classifier*. Metode Naïve Bayes digunakan untuk menghitung nilai probabilitas kata-kata dalam dokumen teks termasuk dalam ujaran kebencian atau bukan ujaran kebencian. *Classifier* ini dibuat dengan bahasa pemrograman Python. Pada tahap perancangan *Classifier*, 500 data dikumpulkan secara acak pada sosial media Facebook. Data dibagi menggunakan rasio 80% - 20%, 400 data teks sabagi data latih dan 100 data digunakan sebagai data uji. Penelitian ini menghasilkan tingkat akurasi klasifikasi teks ujaran kebencian sebesar 83%. Hasil ini diperoleh dari evaluasi *Classifier* menggunakan data uji dimana *Classifier* memberikan label dengan tepat sebanyak 83 dari 100 data uji.

Kata Kunci: Ujaran Kebencian; Klasifikasi Teks; Naïve Bayes.

1. Pendahuluan

Kemajuan dunia informasi dan teknologi saat ini sangat pesat sejak dimulainya era millennium. Internet yang menyediakan berbagai

layanan yang berguna bagi kehidupan manusia merupakan salah satu produk kemajuan bidang teknologi dan informasi. Layanan tersebut dapat dimanfaatkan untuk akses informasi, media

komunikasi dan masih banyak lagi. Internet juga dapat digunakan oleh siapapun dari berbagai kalangan usia mulai dari anak-anak hingga orang dewasa. Selain itu, kemudahan mengakses internet membuat internet telah digunakan oleh hampir semua masyarakat Indonesia. Kemudahan – kemudahan tersebut tentunya memanjakan para penggunanya namun ada juga beberapa dampak negatif yang ditimbulkan dari kemudahan tersebut, salah satunya adalah ujaran kebencian.

Definisi dari ujaran kebencian dapat diketahui dari jenis perbuatan ujaran kebencian yang tertuang dalam Surat Edaran Kapolri No SE/X/06/2015. Berdasarkan surat tersebut perbuatan yang tergolong dalam ujaran kebencian (Hate Speech) adalah penghinaan, pencemaran nama baik, penistaan, perbuatan tidak menyenangkan, memprovokasi, menghasut dan menyebarkan berita bohong baik secara langsung di muka umum maupun lewat sosial media.

Saat ini, fenomena Hate Speech atau ujaran kebencian menjadi perbincangan hangat setelah Kepolisian Republik Indonesia mengeluarkan pernyataan bahwa penyebar ujaran kebencian di media sosial dapat dikenakan sanksi pidana. Hal ini dilatarbelakangi oleh banyaknya warganet di media sosial yang mengeluarkan ujaran kebencian baik di kolom komentar maupun lewat status pribadi. Para penyebar ujaran kebencian sering membawa isu-isu sensitif seperti suku, agama, ras, warna kulit, gender, dan kaum difabel. Dampak dari fenomena ini tentu saja akan berakibat buruk bagi banyak orang terkhusus pengguna media sosial karena mendapatkan informasi yang keliru dapat menimbulkan kebencian dalam masyarakat dan dapat mengakibatkan konflik. Konflik dan kebencian dalam masyarakat dapat berpengaruh negatif pada kinerja bidang lainnya seperti bidang birokrasi pemerintah dan lain-lain.

Pendeteksian ujaran kebencian dari data berbasis teks dapat dilakukan dengan pendekatan analisis sentiment. Analisis sentimen merupakan bidang kajian yang dapat digunakan untuk menganalisa sentiment dari dokumen teks. Salah satu jenis dari analisis sentiment disebut dengan klasifikasi sentiment atau penentuan polaritas dokumen teks. Polaritas umumnya dibagi atas positif atau negative (Cortis & Davis, 2021). Pada klasifikasi ujaran kebencian maka polaritas dokumen teks yang digunakan adalah dokumen mengandung ujaran kebencian atau tidak mengandung ujaran kebencian.

Penentuan polaritas dari dokumen teks dapat dilakukan menggunakan algoritma machine

learning. Algoritma yang dapat digunakan adalah yang tergolong dalam supervised learning. Algoritma jenis ini membutuhkan data latih, yaitu data teks yang sudah dilabeli sebelumnya, kemudian algoritma akan mengklasifikasi teks berdasarkan dari data latih. Naïve Bayes, Nearest Neighbour, dan Decision Tree merupakan contoh algoritma supervised learning. Selain menggunakan pendekatan algoritma machine learning, juga dapat digunakan metode lexicon-based (Khoo & Johnkhan, 2018).

Naïve Bayes cocok digunakan untuk menentukan polaritas dokumen. Metode ini termasuk metode yang sederhana namun memiliki performansi yang baik dalam klasifikasi sentiment, serta metode ini juga metode ke-2 terbanyak yang digunakan dalam analisis sentiment (Lighthart et al., 2021). Klasifikasi dokumen dengan Naïve Bayes dilakukan dengan menentukan nilai probabilitas kemiripan antara dokumen yang telah memiliki label dengan dokumen baru yang belum memiliki label dengan melakukan perhitungan statistik.

Beberapa penelitian klasifikasi sentimen dengan metode Naïve Bayes telah dilakukan. Sumber data teks yang digunakan juga beragam seperti Twitter (Suryono & Taufiq Luthfi, 2018), Facebook (Elouardighi et al., 2017), dan data teks ulasan pariwisata dari Tripadvisor (Fanissa et al., 2018), dan blog (Akella & Akella, 2018). Tingkat akurasi analisis sentimen menggunakan Naïve Bayes cukup memuaskan bahkan memiliki akurasi yang lebih baik dibandingkan dengan lexicon-based untuk analisis sentiment teks berbahasa Indonesia (Ahmad et al., 2019).

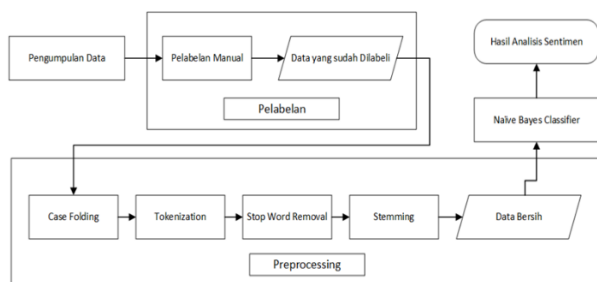
Pada penelitian ini dilakukan pengklasifikasian dokumen teks yang mengandung ujaran kebencian atau tidak mengandung ujaran kebencian. Dokumen teks yang digunakan adalah dokumen berbahasa Indonesia dan diperoleh dari media sosial Facebook. Penelitian ini menerapkan metode Naïve Bayes untuk klasifikasi teks dikarenakan metode ini memiliki tingkat akurasi yang baik dalam penentuan polaritas dokumen teks dari berbagai sumber data dan dokumen teks berbahasa Indonesia.

2. Metode Penelitian

Penelitian ini dilakukan melalui beberapa tahapan. Penelitian dimulai dengan pengumpulan data, kemudian setiap data tersebut diberi label secara manual yaitu ujaran kebencian atau bukan ujaran kebencian. Setelah data diberi label, data masuk tahap *Preprocessing* yaitu mengubah semua katakter menjadi huruf kecil (Case Folding),

membagi teks menjadi potongan kata (*Tokenization*), menghapus kata yang tidak memiliki makna (*Stop Word Removal*), dan mengubah kata menjadi kata bakunya (*Stemming*). Setelah data melewati proses *Preprocessing*, data akan masuk ke *Classifier* untuk diproses dan hasilnya adalah klasifikasi apakah teks tersebut ujaran kebencian atau bukan ujaran kebencian. Gambar 1 menampilkan tahapan metode dari penelitian ini.

Pengumpulan data dilakukan dengan mengambil data dari komentar-komentar ataupun tulisan pengguna di media sosial Facebook. Data yang dikumpulkan terdiri $\pm 50\%$ Positif ujaran kebencian dan $\pm 50\%$ bukan ujaran kebencian. Proses penentuan label data dilakukan secara manual. Data dikumpulkan tanpa menggunakan web crawler tapi disalin dari sosial media dan disimpan ke dalam file dengan format csv. Data diambil secara acak tanpa kriteria kriteria tertentu agar diperoleh data yang lebih bervariasi.



Gambar 1. Metode Klasifikasi Ujaran Kebencian

Data yang telah dilakukan selanjutnya diproses dengan beberapa tahapan *preprocessing*. *Preprocessing* diawali dengan mengubah seluruh karakter menjadi huruf kecil, proses ini disebut *Case Folding*. Selanjutnya tiap dokumen teks dipisah perkata, tahap ini disebut *Tokenization*. Setelah *Tokenization*, Masuk tahap *Stop Word Removal* yaitu setiap kata yang tidak memiliki makna atau arti penting akan dihapus. Selanjutnya Setiap kata diubah menjadi kata baku, tahap ini disebut tahap *Stemming*. Proses *Preprocessing* dilakukan menggunakan Bahasa Pemrograman Python dengan memanfaatkan Library NLTK (Bird et al., 2009). Setelah *Preprocessing* dilakukan, data-data akan diberi label polaritas yaitu Ujaran Kebencian atau bukan Ujaran kebencian. Data set yang digunakan dibagi menjadi dua yaitu Data set untuk melatih *Classifier* (Training Data) dan data set untuk evaluasi *Classifier* (Test Data).

Data teks yang telah melewati *preprocessing* dan penentuan label digunakan untuk melatih

metode klasifikasi yang digunakan. Implementasi metode klasifikasi, Naive Bayes, menggunakan Bahasa pemrograman Python dengan menggunakan library Scikit-learn. Cara kerja metode klasifikasi adalah setiap kata pada data berlabel Ujaran kebencian maupun bukan ujaran kebencian akan diberikan bobot dan frekuensi munculnya kata tersebut akan dihitung. Setelah metode klasifikasi dilatih, maka metode dapat digunakan untuk menentukan data teks yang belum berlabel dengan cara melakukan *Preprocessing* pada teks yang akan diklasifikasi dan menghitung probabilitas tiap katanya menggunakan rumus Naive Bayes. Rumus Naive Bayes ditunjukkan pada Persamaan (1).

$$V_{MAP} = \arg \max_{v_j \in V} \frac{P(a_1, a_2, \dots, a_n | v_j) P(v_j)}{P(a_1, a_2, \dots, a_n)} \quad (1)$$

Keterangan :

$P(a_1, a_2, \dots, a_n | v_j)$ = Probabilitas kata-kata dalam sebuah dokumen terhadap polaritas v_j

$P(v_j)$ = Probabilitas v_j pada dataset

$P(a_1, a_2, \dots, a_n)$ = Probabilitas kata-kata dalam dokumen terhadap seluruh kata pada dataset

Pada rumus Naive Bayes pada Persamaan (1), setiap probabilitas kondisional atau setiap probabilitas suatu kata dikalikan. Namun, Perkalian nilai desimal yang terlalu besar akan membuat nilai terlalu banyak angka di belakang koma (.) (floating point underflow) sehingga data tidak dapat ditampung dalam computer (Rahmad & Pribadi, 2015). Untuk permasalahan ini dapat digunakan penambahan logaritma dalam perhitungan, dimana $\log(xy) = \log(x) + \log(y)$, Sehingga persamaan Naive Bayes yang digunakan untuk klasifikasi pada penelitian ini adalah Persamaan (2) :

$$V_{MAP} = \arg \max_{v_j \in V} \log P(v_j) + \sum \log P(a_i | v_j) \quad (2)$$

Tahap terakhir dari penelitian ini adalah evaluasi hasil klasifikasi. Proses evaluasi dilakukan dengan menghitung akurasi dari hasil klasifikasi data uji. Akurasi diperoleh dengan membandingkan label polaritas yang dihasilkan dari metode klasifikasi dengan label yang diberikan secara manual.

3. Hasil dan Pembahasan

3.1 Analisis Data

Data pada penelitian ini menggunakan data teks yang dikumpulkan dari facebook. Data terdiri dari 500 data teks dimana 400 data merupakan data latih dan 100 data lainnya adalah data uji. Data latih terdiri dari 200 data teks yang termasuk dalam ujaran kebencian dan 200 yang bukan. Sedangkan untuk data uji juga digunakan perbandingan yang sama, yaitu 50 untuk data teks ujaran kebencian dan 50 bukan merupakan ujaran kebencian.

Seluruh data teks diberikan label, Ujaran kebencian atau bukan, secara manual untuk keperluan pembelajaran dan pengujian. Data teks yang dikumpulkan disimpan dalam format file csv. Data disimpan dalam dua file berbeda, yaitu satu file berisi data teks yang termasuk ujaran kebencian dan file lainnya berisi teks yang bukan merupakan ujaran kebencian. Gambar 2 menunjukkan tampilan data teks dalam file.

1	ini demok	lasi	terserah	lhoe	pilih	capres	nmr	brp	sesuai	hati	nurani
2	an**ng										
3	Pak emang	bisa	kalo	org	kafir	mendirikan	negra	khilafah?	Mohon	pencerahanya.	
4	Yakin polri	netral???	GUE	GAK	YAKIN	1000	%				
5	MONSTER	KALIAN	YANG	KALIAN	PUJA	PUJI	YA	INI	!!!		

Gambar 2 Data Mentah Ujaran Kebencian

3.2 Preprocessing

Proses *preprocessing* yang pertama dilakukan adalah *casefolding*. Potongan hasil *casefolding* ditunjukkan pada Gambar 3. Terlihat bahwa semua huruf telah berubah menjadi huruf kecil.

1	ini demok	lasi	terserah	lhoe	pilih	capres	nmr	brp	sesuai	hati	nurani
2	an**ng										
3	Pak emang	bisa	kalo	org	kafir	mendirikan	negra	khilafah?	Mohon	pencerahanya.	
4	Yakin polri	netral???	GUE	GAK	YAKIN	1000	%				
5	monster	kalian	yang	kalian	puja	puji	ya	ini	!!!		

Gambar 3 Hasil Casefolding

Proses *preprocessing* berikutnya adalah tokenizing. Data teks hasil *casefolding* yang berupa kalimat dipecah menjadi perkata. Hasil *tokenization* ditunjukkan pada Gambar 4. Terlihat bahwa seluruh setiap baris dokumen teks berisi daftar kata pada dokumen teks bukan lagi kalimat.

1	['ini', 'demokrasi', 'terserah', 'lhoe', 'pilih', 'capres', 'nmr', 'brp', 'sesuai', 'hati', 'nurani', 'u', 'yg', 'kau', 'anggap', 'baik', 'kl', 'i']
2	['akan', 'setia', 'sama', 'tuannya', 'dan', 'akan', 'selalu', 'menggonggong', 'atau', 'menggigit', 'bila', 'dirasa', 'tuannya', 'dalam', 'bahaya']
3	['pak', 'emang', 'bisa', 'kalo', 'org', 'kafir', 'mendirikan', 'negra', 'khilafah', 'mohon', 'pencerahanya']
4	['yakin', 'polri', 'netral', 'gue', 'gak', 'yakin']
5	['monster', 'kalian', 'yang', 'kalian', 'puja', 'puji', 'ya', 'ini']

Gambar 4 Hasil Tokenization

Hasil *tokenization* selanjutnya melewati proses *Stopwords Removal*. Pada proses ini kata-kata yang dianggap netral dan tidak mengandung polarisasi dihilangkan. Kata-kata ini disebut dengan stopwords, contohnya adalah “dan”, “atau”, “dari”, dst. Proses *Stopwords Removal* dilakukan dengan memanfaatkan library NLTK, dimana pada library ini sudah terdapat daftar stopwords untuk

Bahasa Indonesia. Kata yang ada di dalam daftar stopwords dihilangkan dari data teks. Hasil dari *Stopwords Removal* ditampilkan pada Gambar 5.

1	demokrasi	terserah	lhoe	pilih	capres	nmr	brp	sesuai	hati	nurani
2	setia	tuannya	menggonggong	menggigit	dirasa	tuannya	tuannya	penjahat	pembohor	manusia
3	emang	kalo	org	kafir	mendirikan	negra	khilafah	mohon	pencerahanya	
4	polri	netral	gue	gak						
5	monster	puja	puji	ya						

Gambar 5 Hasil Stopwords Removal

Proses terakhir dari *preprocessing* adalah stemming. Stemming merupakan proses penghilangan awalan dan atau akhiran pada sebuah kata, sehingga kata-kata dengan kata dasar yang sama dianggap sebagai kata yang sama. Stemming pada penelitian ini memanfaatkan library Sastrawi (Robbani, 2016), yaitu library yang menyediakan kamus stemming untuk dokumen teks berbahasa Indonesia. Hasil dari proses stemming ditunjukkan pada Gambar 6. Terlihat dari Gambar 6 terdapat kata yang berhasil diekstrak kata dasarnya seperti “gonggong”, namun masih ada kata lain yang masih berbentuk kata berimbuhan seperti “pencerahannya”.

1	demokrasi	serah	lhoe	pilih	capres	nmr	brp	sesuai	hati	nurani
2	setia	tuan	gonggong	menggigit	rasa	tuan	tuan	jahat	bohong	manusia
3	emang	kalo	org	kafir	diri	negra	khilafah	mohon	pencerahanya	
4	polri	netral	gue	gak						
5	monster	puja	puji	ya						

Gambar 6 Hasil Stemming

Setelah semua data teks melewati tahap *Preprocessing*, maka data teks dibagi atas dua bagian. Bagian pertama merupakan data teks untuk data latih *Classifier* sedangkan bagian lain digunakan sebagai data uji.

3.2 Pengujian

Metode Pengujian yang digunakan adalah pengujian menggunakan data uji berjumlah 100 data. Data uji yang telah diberi label sebelumnya, diuji menggunakan *Classifier* yang telah dibuat. Hasil dari klasifikasi tersebut dibandingkan dengan label yang telah diberikan secara manual. Berdasarkan hasil pengujian, didapatkan tingkat akurasi dari *Classifier* yang digunakan. Perhitungan untuk mendapatkan tingkat akurasi dalam persen ditunjukkan pada Persamaan (3).

$$\text{Akurasi} = \frac{\text{Jumlah data uji dengan polaritas yang sesuai}}{\text{Jumlah seluruh data uji}} \times 100\% \quad (3)$$

Beberapa hasil klasifikasi data uji menggunakan Naïve Bayes yang sesuai antara label manual dengan hasil klasifikasi ditunjukkan Pada Tabel 1, sedangkan hasil klasifikasi dokumen teks yang tidak sesuai dengan label manual ditunjukkan

pada Tabel 2. Kolom label berisi polaritas yang diberikan manual kepada dokumen teks sedangkan kolom hasil klasifikasi berisi label yang diberikan oleh *Classifier*. Label untuk dokumen teks yang tidak mengandung ujaran kebencian menggunakan **pos**, sedangkan yang mengandung ujaran kebencian menggunakan **neg**.

Tabel 1. Hasil Klasifikasi Sesuai

No	Data Teks	Label	Hasil Klasifikasi
1	Sebelumnya banyak teman2 dari jasa kontraktor hampir putus asa. Tetapi sekarang mereka tidak sempat pulang ke rumah.	pos	pos
2	Jangan salah menurut informasi dr Sodara gw si Mak ini beneran cakep lho orangnya ?? , gw akan tetap support Mak ??	pos	pos
3	Org macam si cabul buron ini jika hidup di arab, china, rusia atau thailand kira2 butuh nyawa brp ya biar bisa panjang umur? Beruntung sekali dia hidup di jaman jokowi	neg	neg
4	sdhlah, muak lihat kambing satu ini, aparat keamanan kelamaan menumpas HRS dan gerombolannya, bikin negeri ini gonjang-ganjing terus !!!	neg	neg

Tabel 2. Hasil Klasifikasi Tidak Sesuai

No	Data Teks	Label	Hasil Klasifikasi
1	Mak soal dukung mendukung itu sah. Yg penting jangan bikin fitnah menghina berita hoax. I love you MLT	pos	neg
2	Akun mak berdiri sendiri tanpa tongkat siapapun., iye kan mak...??	pos	neg
3	Dari awal dah tau mak.. makanya sering ta rekomendasi ke temen2 biar follow	pos	neg

	akun ini...????		
4	Presiden kau buat becandaan kau hina padahal kau ustad sekarang ada yg becaidain namamu kau marah ternyata balasan dr Alloh dibayar langsung dan cepat	pos	neg
5	Klu nyebut Adidas kan malah mempromosikan karna gk boleh nyebut merek klu pas on air karna gk ikut mensponsori makaya d singkat,itu pandangan gua,gk tau pandangan orang lain kan jenis orang itu ada dua sehat dan kurang sehat termasuk d golongan mana tentunya orang lain yg bisa menerjemahkan nya.	pos	neg
6	tambah sinting aja si porno nih di arab haha.... gak bisa pulang lagi ?? jd gembel deh loe di arab gih... atau jd aktor porno berdaster aja haha	neg	pos
7	Lha yg cabul ini yah yg nyuruh kepung istana ? Mau dong dikepung wanita wkwkwkwkw	neg	Pos
8	mestinya si Cabul ini dikasi kopi biar cpat modyar. onta keturunan yaman coba coba bikin onar di Indonesia. ngak ada tempat kalian di Indonesia	neg	Pos
9	https://www.facebook.com/elfere	neg	pos
10	Penjara aja dah nih habib provokator n cabul	neg	Pos
11	Onta rizek sdh provokasi makar....tangkap dan adili.....siburonan cabul.....	neg	pos
12	Woiiii pulang lo pengecut!!!!!!	neg	Pos
13	Mental pecundang	neg	pos
14	suruh mereka2 bunuh diri masal aja mak,	neg	pos

	hidup buat onar aja ,		
15	Dan akhirnya orang gila jadi presidenpun mulai ditinggal satu persatu koalisinya, biarkan dia edan sendirian, politik itu teman kalo menguntungkan, kalo sudah tidak menguntungkan tinggal aja, salam edan terus capres abadi	neg	pos
16	Pak Djoko aja yang mati duluan! Haus kekuasaan banget sih!	neg	Pos
17	orang tampang bangkai gini kok bisa ya jadi politisi..???	neg	pos

Jumlah dokumen teks yang tidak berhasil diklasifikasikan dengan tepat adalah 17 dari 100 dokumen. Terdapat 5 dokumen teks yang seharusnya bukan ujaran kebencian namun diklasifikasikan sebagai ujaran kebencian dan 12 lainnya adalah dokumen dengan label ujaran kebencian namun diklasifikasikan sebagai bukan ujaran kebencian. Dokumen teks yang tidak berhasil diklasifikasikan dengan benar ditampilkan pada Tabel 2.

Berdasarkan hasil klasifikasi 100 data uji, maka *Classifier* Naïve Bayes berhasil mengklasifikasikan sebanyak 83 data dengan benar dari total 100 data uji. Maka, tingkat akurasi dari *Classifier* Naïve Bayes untuk klasifikasi ujaran kebencian pada dokumen teks sebesar 83%.

$$Akurasi = \frac{83}{100} \times 100 = 83\%$$

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan dapat ditarik kesimpulan bahwa tingkat akurasi klasifikasi ujaran kebencian menggunakan metode Naive Bayes mencapai angka 83%. Hasil tingkat akurasi ini menunjukkan bahwa kinerja *Classifier* Naïve Bayes cukup baik dalam mengklasifikasikan ujaran kebencian.

Referensi

Ahmad, M., Octaviansyah, M. F., & Kardiana, A. (2019). Sentiment Analysis System of Indoneisan Tweets using Lexicon and Naive Bayes Approach. *2019 Fourth International Conference on Informatics and Computing (ICIC)*, 7–11.

- Akella, J. O., & Akella, L. N. Y. (2018). Sentiment Analysis Using Naïve Bayes Algorithm: With Case Study. *Proceedings of the 3rd International Conference on Inventive Computation Technologies, ICICT 2018*. <https://doi.org/10.1109/ICICT43934.2018.9034394>
- Bird, S., Loper, E., & Klein, E. (2009). *Natural Language Processing with Python*. O'Reilly Media, Inc.
- Cortis, K., & Davis, B. (2021). Over a decade of social opinion mining: a systematic review. *Artificial Intelligence Review*, 54(7), 4873–4965. <https://doi.org/10.1007/s10462-021-10030-2>
- Elouardighi, A., Maghfour, M., & Hammia, H. (2017). Collecting and processing arabic facebook comments for sentiment analysis. *International Conference on Model and Data Engineering*, 262–274.
- Fanissa, S., Fauzi, M. A., & Adinugroho, S. (2018). Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(8), 2766–2770.
- Khoo, C. S., & Johnkhan, S. B. (2018). Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. *Journal of Information Science*, 44(4), 491–511. <https://doi.org/10.1177/0165551517703514>
- Lighthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. *Artificial Intelligence Review*, 54(6). <https://doi.org/10.1007/s10462-021-09973-3>
- Rahmad, A. N., & Pribadi, F. S. (2015). Pemilihan Feature Dengan Chi Square Dalam Algoritma Naïve Bayes Untuk Klasifikasi Berita. *Edu Komputika Journal*, 2(1), 13–21. <https://doi.org/10.15294/edukomputika.v2i1.7823>
- Robbani, H. A. (2016). *PySastrawi* (1.2). <https://github.com/har07/PySastrawi>
- Suryono, S., & Taufiq Luthfi, E. (2018). Analisis sentimen pada Twitter dengan menggunakan metode Naïve Bayes Classifier. *Seminar Nasional Geotik*, 9–15. <https://doi.org/10.36802/jnanaloka.2020.v1-no2-81-86>