

ANALISIS SENTIMEN PERSEPSI MASYARAKAT TERHADAP UNHAN RI MELALUI MEDIA SOSIAL

Diterima Redaksi: 18 September 2024; Revisi Akhir: 30 September 2024; Diterbitkan Online: 15 November 2024

Nadiza Lediwara¹⁾, Sembada Denrineksa Bimorogo²⁾, Aulia Khamas Heikmakhtiar³⁾,
Ido Jaya Gainal⁴⁾, Sufi Naylil Karomah⁵⁾, Khazali Fahmi⁶⁾

^{1, 2, 3),4),5)} Informatika, Fakultas Sains dan Teknologi Pertahanan, Universitas Pertahanan Republik Indonesia

⁶⁾ Fisika, Fakultas MIPA Militer, Universitas Pertahanan Republik Indonesia

^{1, 2, 3),4),5),6)} Kawasan Indonesia Peace and Security Center (IPSC), Sentul, Bogor, Jawa Barat

e-mail: nadizalediwara@gmail.com¹⁾, denri5693@gmail.com²⁾, auliakhphd@gmail.com³⁾,
gainalidojaya@gmail.com⁴⁾, karomahnaylil35@gmail.com⁵⁾, fahmikhazali1@gmail.com⁶⁾

Abstrak: Penelitian ini bertujuan untuk mengetahui opini publik mengenai Universitas Pertahanan Republik Indonesia (UNHAN RI). Sebagai universitas yang relatif masih baru, tanggapan mengenai opini publik ini perlu dilakukan agar pihak universitas dapat memberikan layanan dan program pendidikan yang lebih baik. Untuk melihat tanggapan publik terhadap UNHAN RI digunakan suatu metode Machine Learning yaitu Naïve Bayes. Pemodelan Naïve Bayes ini dapat membantu untuk mengklasifikasikan analisis sentimen publik terhadap UNHAN RI. Hasil klasifikasi dibagi dalam tiga kelas yaitu positif, negatif, dan netral. Hasil klasifikasi terbesar diperoleh pada kelas positif sebesar 60.8%, kelas netral sebesar 33.3%, dan kelas negatif sebesar 5.9%. Pada tahapan evaluasi, peneliti menggunakan Confusion Matrix dengan tiga parameter yaitu akurasi, precision, dan recall. Skor precision didapat sebesar 83,60% sedangkan skor recall adalah 78,40%. Dalam hal akurasi diperoleh nilai sebesar 60.78%. Hal ini berarti bahwa model mampu mengklasifikasikan 60.78% dari dokumen dengan benar, skor ini mengartikan bahwa mayoritas pandangan publik terhadap UNHAN RI adalah positif.

Kata Kunci— Sentimen Analisis, UNHAN RI, Naïve Bayes

Abstract: This study aims to determine public opinion regarding Universitas Pertahanan Republik Indonesia (UNHAN RI). As a new university, responses regarding public opinion need to be carried out so that the university can provide better services and educational programs. To see the public response to UNHAN RI, a Machine Learning method is used, namely Naïve Bayes. This Naïve Bayes modeling can help to classify public sentiment analysis towards UNHAN RI. The classification results are divided into three classes, namely positive, negative, and neutral. The largest classification results were obtained in positive class of 60.8%, neutral class of 33.3%, and negative class of 5.9%. The evaluation stage, researchers used Confusion Matrix with three parameters, namely accuracy, precision, and recall. The precision score was 83.60% while the recall score was 78.40%. In terms of accuracy, a value of 60.78%. This means that the model is able to classify 60.78% of the documents correctly, this score means that the majority of public views on UNHAN RI are positive.

Keywords— Sentiment Analysis, UNHAN RI, Naïve Bayes

I. PENDAHULUAN

UNIVERSITAS Pertahanan Republik Indonesia (UNHAN RI) merupakan institusi perguruan tinggi dalam negeri yang memiliki fokus di bidang pertahanan dan keamanan. Kampus ini memiliki kurikulum berbasis akademik pada umumnya dan kemiliteran. Sebagai universitas yang relatif baru dan unik, UNHAN RI sering menjadi topik pembicaraan di berbagai platform media sosial, khususnya Twitter. Mengingat pentingnya memahami persepsi publik, analisis tanggapan umum terhadap UNHAN RI dapat memberikan wawasan yang berharga bagi pihak universitas untuk memperbaiki dan mengembangkan layanan serta program pendidikan [1].

Dalam hal pengetahuan terhadap opini atau persepsi publik, diperlukan suatu penerapan dari algoritma Machine Learning. Salah satu manfaat dari algoritma Machine Learning adalah dapat digunakan untuk memberikan pengetahuan dalam mengetahui persepsi publik [2]. Sentimen analisis

adalah proses analisis teks yang digunakan untuk mendapatkan informasi data dari beragam sumber data melalui media internet. Selain mendapatkan data yang bersumber dari internet, sentimen analisis ini juga dapat berkerja pada bermacam-macam media sosial [3]. Tujuan dari sentimen analisis ini agar dapat diperolehnya opini dari pengguna ataupun publik yang terdapat pada bermacam platform digital [4]. Data dikumpulkan dari Twitter menggunakan teknik *crawling data*. Proses *crawling* ini memungkinkan untuk mengumpulkan berbagai tweet yang menyebutkan Unhan RI. Setelah data terkumpul, selanjutnya diolah menggunakan perangkat lunak.

Dalam melakukan proses sentimen analisis, salah satu model *Machine Learning* yang digunakan adalah Naïve Bayes. Metode Naive Bayes dipilih karena kesederhanaannya dan kemampuannya untuk menangani data teks dengan efektif. Dengan menggunakan metode ini, kita dapat mengklasifikasikan tanggapan menjadi kategori positif, negatif, atau netral. Kajian dari hasil analisis yang didapatkan, hal ini dapat memberikan pandangan masyarakat yang lebih jelas tentang bagaimana masyarakat ataupun publik memandang UNHAN RI, serta memberikan masukan yang konstruktif bagi peningkatan kualitas di universitas tersebut [5].

Selain memahami persepsi publik melalui media sosial, sentimen analisis juga dapat membantu UNHAN RI dalam menyusun strategi komunikasi yang lebih efektif. Dengan mengetahui sentimen umum terhadap universitas, pihak manajemen dapat menanggapi isu-isu yang berkembang secara lebih tepat dan proaktif [6]. Hal ini tidak hanya membantu dalam menjaga citra positif universitas, tetapi juga membangun hubungan yang lebih baik dengan mahasiswa, alumni, dan masyarakat luas.

Lebih jauh lagi, analisis tanggapan ini juga dapat menjadi dasar bagi pengembangan kurikulum dan program pendidikan yang dapat menjawab kebutuhan masyarakat dan industri pertahanan. Melalui pemahaman yang mendalam tentang opini publik, UNHAN RI dapat menyesuaikan pendekatan pendidikan agar lebih sesuai dengan harapan dan tuntutan zaman. Hal ini akan membantu UNHAN RI dalam mencetak lulusan yang tidak hanya kompeten dalam bidang pertahanan, tetapi juga siap menghadapi tantangan global.

Secara keseluruhan, analisis ini bertujuan untuk dapat memberikan gambaran yang komprehensif mengenai apa saja dan bagaimana UNHAN RI dipersepsikan oleh masyarakat melalui media sosial. Dengan hasil yang diperoleh, diharapkan UNHAN RI dapat terus berinovasi dan berkembang, memperkuat perannya sebagai lembaga pendidikan tinggi yang unggul dalam bidang pertahanan dan keamanan, serta memberikan hal berupa sumbangsih nyata dalam hal kemajuan bangsa dan negara.

II. TINJAUAN PUSTAKA

A. Sentimen Analisis

Sentimen analisis adalah proses analisis teks yang diperoleh melalui beragam sumber data seperti internet dan beragam platform sosial media [7]. Kegiatan sentimen analisis ini adalah untuk memperoleh opini dari pengguna dari beragam platform seperti sosial media. Sentimen analisis ini merupakan salah satu bidang dari banyaknya bidang *Natural Language Processing* (NLP) [8]. Dengan bantuan sentimen analisis, data yang diperoleh sebelumnya belum menghasilkan informasi dapat diubah menjadi data yang lebih terstruktur dan menghasilkan informasi dan pengetahuan yang berguna.

Pada tahapan pengumpulan data, peneliti menggunakan Tweet Harvest. Tweet Harvest adalah *command-line tool* yang dapat digunakan untuk mengambil tweet berdasarkan kata kunci dan waktu tertentu [9].

B. Rapid Miner

Rapid Miner merupakan *software* yang dapat mengolah data mining seperti analisis, proses dan visualisasi [10]. *Software* ini merupakan *software open source*. Untuk pemrosesan sentimen analisis, dari pembersihan data hingga visualisi dapat dikerjakan dengan bantuan *software* ini.

Dataset yang akan diambil pada penelitian ini adalah berkaitan dengan *keyword* seputaran UNHAN RI. Data tersebut nantinya akan diproses dari proses pembersihan hingga visualisasi di *software* Rapid Miner.

C. Naïve Bayes

Algoritma Naïve bayes adalah satu metode klasifikasi dengan menggunakan teori probabilitas yang ditemukan oleh Thomas Bayes [11]. Model ini dapat memprediksi dengan teori peluang berdasarkan pengalaman lampau. Teori ini dikenal dengan Teorema Bayes.

Tahapan dalam pengerjaan Naïve Bayes adalah dimulai dengan menghitung peluang dari keseluruhan jumlah kelas [12]. Proses selanjutnya yaitu dengan menghitung peluang yang lebih spesifik dari jumlah kasus pada setiap kelas. Hasil dari perbandingan keseluruhan isi kelas ini dikalikan dengan keseluruhan jumlah kelas. Tahapan terakhirnya yaitu membandingkan hasil setiap kelasnya.

D. CRISP-DM

Cross Industry Standard Process for Data Mining (CRISP-DM) merupakan tahapan standar dalam suatu pengerjaan data mining [13]. Isi dari CRISP-DM adalah Langkah-langkah yang digunakan untuk mengolah data dari data mentah menjadi suatu informasi dan pengetahuan yang dapat digunakan untuk pengambilan keputusan.

Tahapan awal dari CRIPS-DM ini adalah *Business Understanding*. Tahapan ini merupakan tahapan untuk menentukan tujuan penelitian atau project yang ingin dicapai [14]. Proses ini juga menentukan ruang lingkup apa saja yang akan menjadi tujuan dari penelitian atau suatu project.

Tahapan kedua dari CRIPS-DM adalah *Data Understanding*. Pemahaman terhadap data apa saja yang diperlukan wajib dilakukan. Pada tahapan ini, peneliti menentukan data-data apa saja yang diperlukan untuk mencapai tujuan bisnis [15]. Sehingga jelas cakupan dari ruang lingkup data yang harus diperoleh.

Tahapan ketiga yaitu *Data Preparation*. Penyiapan data perlu dilakukan agar menghasilkan data-data yang berkualitas untuk proses selanjutnya. Pada tahapan penyiapan data, hal yang dapat dilakukan seperti pembersihan data, pemilihan fitur data, dan pembagian data [16].

Tahapan keempat yaitu *Modelling*. Data yang telah disiapkan seperti proses pembersihan data maka data tersebut akan diterapkan ke dalam suatu algoritma atau model. Model yang digunakan pada tahapan ini disesuaikan dengan tujuan penelitian. Tahap pemodelan ini sebagai alat bantu untuk menyelesaikan suatu proses data mining [17].

Tahapan kelima yaitu *Evaluation*. Fungsi tahap ini adalah untuk mengukur performa dari model yang telah digunakan. Selain itu dapat juga menjadi tolak ukur seberapa baik model yang telah diterapkan untuk mencapai tujuan bisnis [18]. Jika skor evaluasi terbilang cukup rendah, maka langkah yang diambil bisa mengulang lagi dari tahapan awal CRIPS-DM.

Tahapan keenam atau tahapan terakhir dari CRIPS-DM adalah *Deployment*. Proses ini adalah tahapan untuk menampilkan hasil kinerja model sehingga dapat mudah dilihat oleh orang lain. Tahap ini juga terkadang memerlukan kemampuan dalam mendesain antar muka suatu tampilan [19], sehingga informasi ataupun pengetahuan yang didapat dari proses awal hingga akhir dapat ditampilkan ke pembaca.

III. METODE PENELITIAN

Untuk mendapatkan gambaran yang jelas tentang persepsi masyarakat Indonesia terhadap Universitas Pertahanan Republik Indonesia (UNHAN RI), dilakukanlah serangkaian tahapan proses yang sistematis. Tahap awal proses penelitian yaitu Pengumpulan data. Pada tahap ini, proses yang dilakukan adalah mencari komentar dari Twitter yang berkaitan dengan UNHAN RI. Adapun *software* yang digunakan adalah Python versi 3.10 untuk *scraping* data dan Rapid Miner 10.1.003 untuk pemrosesan dari pembersihan data hingga ke tahapan visualisasi. Untuk *hardware* yang digunakan adalah Laptop Lenovo IdeaPad 5.14.0. Teknik pengumpulan data yang digunakan menggunakan teknik pengumpulan data secara otomatis (*crawling*) dengan Metode Harvest. Teknik ini memungkinkan untuk mengumpulkan data dalam jumlah besar dari berbagai sumber di Twitter.

Tahapan kedua yang dilakukan yaitu Pra-pemrosesan Data (*Preprocessing*). Pada tahap ini, dilakukan proses pembersihan data yang telah dikumpulkan untuk memastikan kualitas data yang baik sebelum dilakukan analisis lebih lanjut. Tahapan preprocessing terdiri dari empat langkah utama yaitu:

A. *Cleaning*

Proses ini adalah menghapus karakter dan simbol yang tidak diperlukan, seperti tanda baca, angka, dan simbol khusus yang tidak relevan. Adapun simbol khusus yang tidak relevan dapat dilihat pada Tabel 1.

Tabel 1. *Cleaning Simbol*

Simbol	Contoh
Tanda Baca	.,''"/''-
Angka	1234567890
Simbol Lain	@#&RT

B. *Remove Stopwords*

Proses ini berguna untuk menghilangkan kata-kata umum yang tidak memiliki makna penting dalam analisis, seperti "dan", "yang", "di", dan lain-lain. Kata-kata ini disebut *stopwords*.

C. *Tokenization*

Proses ini memecahkan teks menjadi unit-unit kata atau token individu. Misalnya, kalimat "UNHAN RI adalah universitas" akan dipecah menjadi ["UNHAN", "RI", "adalah", "universitas"].

D. *Stemming*

Proses ini mengubah kata-kata ke bentuk dasarnya. Misalnya, kata "berlari", "lari", dan "pelari" akan dikembalikan ke bentuk dasar "lari".

Setelah data dibersihkan dan diproses, langkah berikutnya adalah ekstraksi fitur. Pada tahap ini, data akan diubah menjadi teks yang dapat digunakan oleh algoritma *machine learning*. Salah satu metode yang umum digunakan adalah representasi *Bag of Words* (BoW) atau *Term Frequency-Inverse Document Frequency* (TF-IDF). Proses ini melibatkan pembuatan vektor fitur yang merepresentasikan frekuensi atau pentingnya kata-kata dalam teks, sehingga model dapat dengan mudah memproses dan menganalisis data tersebut.

Tahapan berikutnya yaitu pelabelan. Data yang telah dibersihkan dan diekstraksi, dilakukan pelabelan manual pada setiap komentar untuk menentukan sentimen yang terkandung di dalamnya, apakah itu positif, negatif, atau netral. Pelabelan ini penting untuk membangun dataset yang akan digunakan dalam proses analisis selanjutnya.

Tahapan berikutnya yaitu pemodelan dengan Algoritma Naive Bayes. Pada tahap ini, penerapan algoritma Naive Bayes untuk klasifikasi sentimen. Naive Bayes adalah algoritma yang efektif dan sederhana untuk tugas klasifikasi teks. Dengan menggunakan data yang telah diberi label dan fitur yang telah diekstraksi, model Naive Bayes akan melatih untuk mengenali pola dalam data dan membuat prediksi sentimen untuk komentar baru. Algoritma ini bekerja dengan menghitung probabilitas dari setiap kelas (positif, negatif, netral) berdasarkan fitur yang ada, kemudian memilih kelas dengan probabilitas tertinggi sebagai hasil prediksi.

Tahapan terakhir yaitu evaluasi. Tahapan ini untuk melihat sudah sebaik apa model yang dipilih untuk memenuhi tujuan penelitian. Pada tahapan evaluasi ini menggunakan Confusion Matrix untuk melihat performa seperti akurasi.

IV. HASIL DAN PEMBAHASAN

Pengumpulan tahap awal dataset dilakukan dengan menggunakan Tweet Harvest. Dari Tweet Harvest didapat beberapa data yang berhubungan dengan kata kunci yang telah diinputkan. Tampilan untuk pengambilan dataset dapat dilihat pada Gambar 1.

Pengambilan dataset dilakukan beberapa hari. Hal ini dikarenakan pengambilan jumlah record dataset dibatasi oleh Tweet Harvest. Per hari pengambilan data tersebut hanya dilakukan sekitar 500 record data. Jadi dibutuhkan pengambilan data beberapa hari agar jumlah dataset yang terambil cukup banyak. Contoh dataset dapat dilihat pada Gambar 2.

```

[ ]
twitter_auth_token = "4907891bF20F4R267aed081809218atceffa7c"

[ ] [!pip install pandas

[!sudo apt-get update
[!sudo apt-get install -y ca-certificates curl gnupg
[!sudo mkdir -p /etc/apt/keyrings
[!curl -fsSL https://deb.nodesource.com/gpgkey/gpgkey.gpg | sudo gpg --dearmor -o /etc/apt/keyrings/nodesource.gpg
[!NODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/node_$NODE_MAJOR.x nodistro main" | sudo tee /etc/apt/sources.list.d/nodesource.list
[!sudo apt-get update
[!sudo apt-get install nodejs -y
[!node -v

Tampilkan output tersembunyi

[ ]
filename = "dataset.csv"
search_keyword = "Unhan"
limit = 100

[!jq -y --raw-output --slurpfile <n "filename" --tab "latest" -l (limit) --token {twitter_auth_token}
    
```

Gambar 1. Pengambilan dataset

conversation_id_str	created_at	favorite_count	full_text	id_str
17907241828311698	Fri May 17 11:25:23 +0000 2024	0	@kenapagituyakk @tanyarlfe Unhan gratis bang	1791429807806496973
17907241828311698	Thu May 16 15:45:51 +0000 2024	0	@tanyarlfe Unhan gratis	1791132964379930812
17907241828311698	Thu May 16 08:50:20 +0000 2024	0	@lafairaa @130moodtrawbel @tanyarlfe Unhan itu full gratis	1791028396669256070

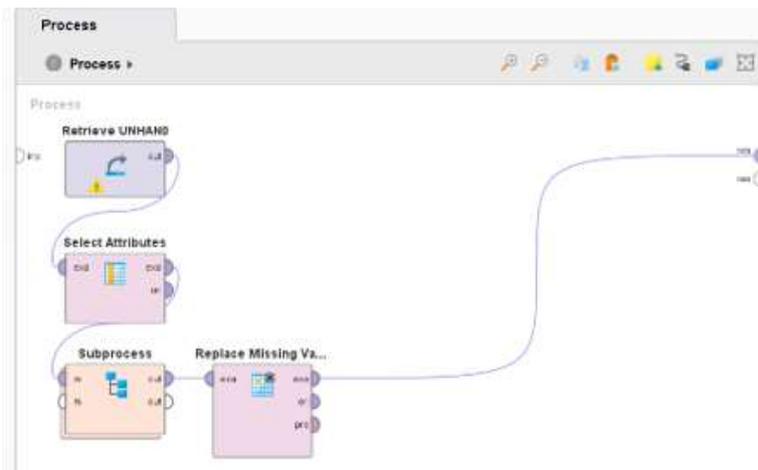
Gambar 2. Contoh dataset

Setelah pengambilan record data dirasa cukup maka proses yang dilakukan adalah pembersihan data. Proses ini dilakukan agar pada saat tahap pemodelan, data sudah cukup bersih dan diharapkan dapat memberikan suatu kesimpulan yang baik [20]. Tahapan ini dikenal juga dengan *preprocessing*. Bentuk dataset awal yang diambil dapat dilihat pada Tabel 2.

Tabel 2. Contoh dataset awal

conversation_id_str	created_at	favorite_count	full_text	id_str
17907241828311698 45	Fri May 17 11:25:23 +0000 2024	0	@kenapagituyakk @tanyarlfe Unhan gratis bang	1791429807806496973
17907241828311698 45	Thu May 16 15:45:51 +0000 2024	0	@tanyarlfe Unhan gratis	1791132964379930812
17907241828311698 45	Thu May 16 08:50:20 +0000 2024	0	@lafairaa @130moodtrawbel @tanyarlfe Unhan itu full gratis	1791028396669256070

Pada tahap preprocessing, data diproses dengan bantuan software Rapid Miner. Proses yang dilakukan diantaranya *cleaning*, *remove data*, dan *tokenizing*. Proses cleaning bertujuan untuk membersihkan tanggapan dari hal yang tidak diperlukan contohnya seperti mengganti record data yang kosong dengan fitur "*Replace Missing Values*". Hal tersebut dapat dilihat pada tampilan Gambar 3.



Gambar 3. Proses cleaning data

Setelah proses *cleaning data* juga menggunakan bantuan *software* Rapid Miner. Proses yang dilakukan diantaranya adalah *remove data*. Adapun yang dilakukan pada proses ini adalah menghapus hastag RT (*Retweet*), “#”, alamat URL, *mention*, data yang terduplikasi, dan simbol (seperti tanda baca). Dalam hal ini bukan data yang dihapus, melainkan yang dihapus adalah sesuatu yang dirasa tidak terlalu penting untuk proses selanjutnya. Kumpulan proses ini dapat dilihat pada Gambar 4.



Gambar 4. Prapocessing tahap kedua

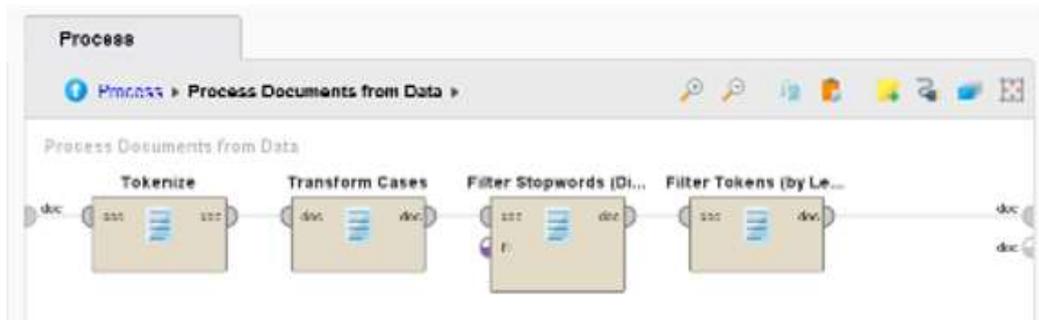
Dari proses yang telah dilakukan, maka didapatlah hasil data yang telah dibersihkan. Untuk contoh data sebelum dan sesudah dilakukan proses pembersihan dapat dilihat pada Tabel 3.

Tabel 3. Contoh dataset awal

Sebelum Dibersihkan	Sesudah Dibersihkan
@Unhan_RI Prabowo mencontohkan skema biaya kuliah gratis yang telah diterapkan oleh Universitas Pertahanan (Unhan)	Prabowo mencontohkan skema biaya kuliah gratis yang telah diterapkan oleh Universitas Pertahanan Unhan
Buktinya adalah UNHAN tapi bukan bagi calon mahasiswa dgn IQ 78 ke bawah yah. Prabowo harusnya kuliah gratis . https://t.co/oITVzKaDBh	Buktinya adalah UNHAN tapi bukan bagi calon mahasiswa dgn IQ 78 ke bawah yah. Prabowo harusnya kuliah gratis .
@paipiapaipia Unhan lulus jd perwira TNI	Unhan lulus jd perwira TNI

Tahapan berikutnya setelah tahap *remove data* adalah tahapan *tokenizing*. Proses *tokenizing* juga menggunakan bantuan *software* Rapid Miner. Proses ini dilakukan untuk memisahkan kata per kata pada akhirnya. Proses ini dilakukan sebagai pembelajaran mesin agar mesin dapat memahami bahasa manusia yang memiliki kompleksitas yang tinggi dan beragam. Dalam tokenisasi ini, mesin dapat menghitung statistik seperti frekuensi kemunculan kata-kata. Tahap tokenisasi ini dapat dilihat pada Gambar 5. Tahapan tokenisasi ini terdiri dari *Transform Cases*, *Filter*

Stopwords, dan *Filter Tokens*. *Transform Cases* adalah proses mentransformasi huruf menjadi huruf kecil (*lower case*). *Filter Stopwords* adalah proses untuk meniadakan *stopwords* yang terkandung dalam suatu kata. Lalu *Filter Tokens* adalah memisahkan setiap kata per kata agar menjadi suatu kata tunggal.



Gambar 5. Praprocessing tokenizing

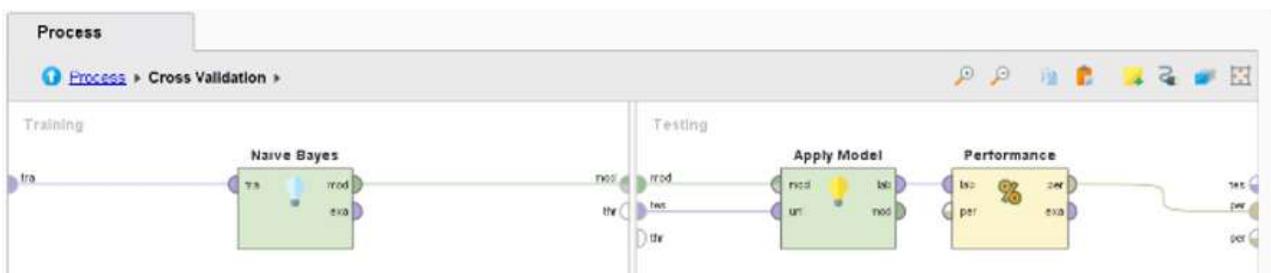
Setelah proses tokenisasi ini dilakukan maka mesin akan memberikan bobot penilaian sehingga pada saat visualisasi dapat direkap bagaimana tanggapan masyarakat terhadap UNHAN RI. Proses tersebut dapat dilihat pada Gambar 6.

The screenshot shows a data table with 100 rows and 9 columns. The columns are labeled: "Row No.", "label", "positif", "negatif", "netral", "afirmasi", "skeptis", "dipertanyakan", and "tidak". The "label" column contains values like "Netral" and "Positif". The other columns contain numerical values representing scores for each category.

Row No.	label	positif	negatif	netral	afirmasi	skeptis	dipertanyakan	tidak
80	Netral	0	0	0	0	0	0	0
81	Netral	0	0	0	0	0	0	0
82	Netral	0	0	0	0	0	0	0
83	Netral	0	0	0	0	0	0	0
84	Netral	0	0	0	0	0	0	0
85	Netral	0	0	0	0	0	0	0
86	Netral	0	0	0	0	0	0	0
87	Netral	0	0	0	0	0	0	0
88	Netral	0	0	0	0	0	0	0
89	Netral	0	0	0	0	0	0	0
90	Positif	0	0	0	0	0	0	0
91	Netral	0	0	0	0	0	0	0
92	Netral	0	0	0	0	0	0	0
93	Netral	0	0	0	0	0	0	0
94	Netral	0	0	0	0	0	0	0
95	Netral	0	0	0	0	0	0	0
96	Positif	0	0	0	0	0	0	0
97	Netral	0	0	0	0	0	0	0
98	Netral	0	0	0	0	0	0	0
99	Positif	0	0	0	0	0	0	0
100	Positif	0	0	0	0	0	0	0

Gambar 6. Bobot Penilaian

Setelah semua proses preprocessing selesai, langkah selanjutnya adalah mengimplementasikan algoritma Naive Bayes menggunakan software Rapid Miner untuk menganalisis tanggapan tentang hasil data yang telah diolah. Pada tahap ini, penerapan model Naive Bayes untuk memprediksi kategori dari setiap tanggapan, apakah positif, negatif, atau netral. Hasilnya menunjukkan bahwa sebagian besar tanggapan terhadap data bersifat positif, namun ada juga beberapa tanggapan negatif dan netral. Proses ini dapat membantu untuk memahami sentimen umum masyarakat terhadap Unhan RI dengan lebih baik. Proses tersebut dapat dilihat pada Gambar 7.



Gambar 7. Pemodelan dengan Naive Bayes

Dalam proses pemodelan, pengerjaan yang dilakukan juga sekaligus melihat evaluasi skor yang didapat dari hasil pemodelan. Untuk mengevaluasi pemodelan, metode yang digunakan dengan Confussion Matrix. Nilai akurasi yang didapat algoritma Naïve Bayes ini adalah 60.78%. Bisa diartikan bahwa sejumlah 60.78% model ini dapat mengklasifikasikan data dengan benar. Sedangkan margin error atau kesalahan dalam pengambilan sample sejumlah + - 4.26%. Nilai + - 4.26% menunjukkan akurasi rata-rata untuk semua kelas (positif, negative, dan netral). Akurasi keseluruhan model adalah 60.78%. Hal ini berarti bahwa model mampu mengklasifikasikan 60.78% dari dokumen dengan benar. Hasil dari evaluasi ini dapat dilihat pada Gambar 8.

● Table View ○ Plot View

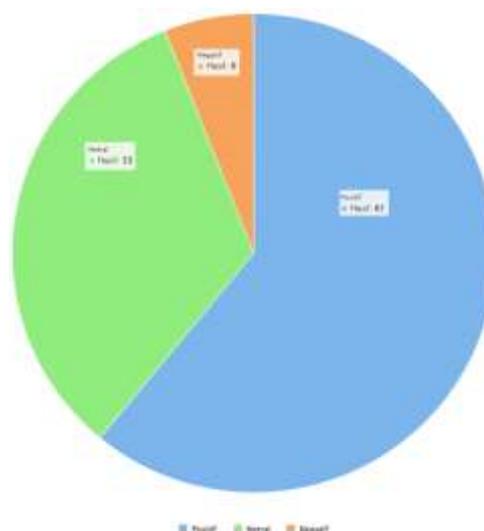
accuracy: 60.78% +/- 4.26% (micro average: 60.78%)

	true Negatif	true Netral	true Positif	class precision
pred. Negatif	150	12	17	83.60%
pred. Netral	170	132	56	36.87%
pred. Positif	34	64	265	73.00%
class recall	42.37%	63.46%	78.40%	

Gambar 8. Tampilan Confussion Matrix

Secara keseluruhan, model Naive Bayes yang digunakan untuk analisis sentimen ini memiliki akurasi moderat. Performa terbaik pada kelas positif dan terburuk pada kelas netral. Hal tersebut memiliki arti yaitu model lebih baik dalam mengidentifikasi sentimen positif dibandingkan dengan sentimen netral atau negatif.

Tahapan terakhir yaitu memberikan rekapitan terhadap hasil cuitan yang didapat dari proses awal (pengambilan data), pembersihan data, pemodelan, dan evaluasi. Rekapitan ini digunakan sebagai gambaran tanggapan masyarakat terhadap UNHAN RI pada suatu media sosial. Dari hasil visualisasi diatas dapat disimpulkan bahwa kelas netral mempunyai persentase 33.3%. Dari total data, sekitar sepertiga atau 33.3% termasuk dalam kategori netral. Kelas positif mempunyai persentase 60.8%. Lebih dari setengah data, yaitu 60.8%, termasuk dalam kategori positif. Ini adalah kategori dengan jumlah data terbanyak. Hal ini berarti tanggapan masyarakat cukup baik terhadap UNHAN RI. Namun dari sisi tanggapan masyakat ini hendaknya dapat ditingkatkan lagi untuk mencapai skor yang lebih baik. Kelas negatif mempunyai persentase 5.9%. Hanya sebagian kecil data, yaitu 5.9%, termasuk dalam kategori negatif. Ini adalah kategori dengan jumlah data paling sedikit. Tampilan Visualisasi dapat dilihat pada Gambar 9.



Gambar 9. Distribusi hasil cuitan

Berdasarkan Gambar 9, mayoritas cuitan adalah berpandangan positif terhadap UNHAN RI. Hanya sedikit yang saja yang berpandangan positif. Dari hasil akurasi juga didapatkan skor sebesar 60,78% seperti yang terlihat pada Gambar 9. Hasil ini tentunya masih bisa ditingkatkan lagi dengan menambah data yang lebih banyak dan melakukan preprocessing dengan tahapan yang lebih kompleks sehingga menghasilkan nilai performa yang lebih baik. Untuk skor *recall* adalah 78,40%. Skor 78,40% ini mempunyai artinya bahwa model yang digunakan dapat memprediksi dengan baik cuitan positif sebesar 78,40%. Sedangkan untuk skor *precision* didapat sebesar 83,60%. Skor ini mempunyai artinya bahwa model dapat memprediksi dengan baik cuitan yang diklasifikasikan dalam cuitan negatif.

V. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pengujian metode Naïve Bayes untuk mengklasifikasi tanggapan masyarakat terhadap UNHAN RI dengan cuitan berupa bahasa Indonesia menggunakan proses tahapan tahapan CRISP-DM yang terdiri dari *business understanding*, *data understanding*, *data preparation*, *modeling*, dan *evaluation*. Hasil akurasi didapat sebesar 60.78%. Hasil ini tentunya sudah cukup untuk mengklasifikasi cuitan dari publik terhadap UNHAN RI. Hasil ini juga dapat memprediksi data yang benar sebesar 60.78%.

Adapun pengelompokkan terdapat tanggapan masyarakat terbagi dalam tiga kelas. Kelas pertama yaitu positif sebesar 60.8%. Kelas kedua yaitu kelas netral sebesar 33.3% dan kelas terakhir yaitu kelas negatif yaitu sebesar 5.9%. Secara garis besar, opini publik terhadap UNHAN RI adalah positif, namun hal ini harus ditingkatkan agar pelayanan dalam bidang akademik semakin baik.

Sebagai saran pada penelitian ini sebaiknya menguji sentimen analisis ini dengan model algoritma yang lain seperti Random Forest (RF), k-Nearest Neighbor (kNN), dan Support Vector Machine (SVM). Diharapkan dengan mencoba model lain, hasil akurasi akan semakin meningkat.

DAFTAR PUSTAKA

- [1] N. Nikolić, O. Grljević, and A. Kovačević, “Aspect-based sentiment analysis of reviews in the domain of higher education,” *Electronic Library*, vol. 38, no. 1, pp. 44–64, Mar. 2020, doi: 10.1108/EL-06-2019-0140/FULL/XML.
- [2] A. Kusuma and A. Nugroho, “Analisa Sentimen Pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naïve Bayes,” *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 15, no. 2, pp. 137–146, Dec. 2021, doi: 10.32815/JITIKA.V15I2.557.
- [3] R. Saptari, R. Rianto, and A. I. Gufroni, “Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Unit Gawat Darurat Rumah Sakit Umum di Indonesia Menggunakan Seleksi Fitur Information Gain dan Support Vector Machine,” *Joined Journal*, vol. 4, no. 2, pp. 104–110, 2021, doi: <https://doi.org/10.31331/joined.v4i2.1925>.
- [4] R. Mursyid and A. D. Indriyanti, “Perbandingan Akurasi Metode Analisis Sentimen Untuk Evaluasi Opini Pengguna Pada Platform Media Sosial (Studi Kasus: Twitter),” *Journal of Informatics and Computer Science (JINACS)*, vol. 06, pp. 371–383, Jun. 2024, [Online]. Available: <https://ejournal.unesa.ac.id/index.php/jinacs/article/view/61322>
- [5] M. Alfarizi, M. Rizqy, R. I. Ghufroni, D. Fathurahman, R. D. Saputra, and F. Kurniawan, “Analisis Sentimen Persepsi Publik Terhadap Kasus Bullying Siswa Cilacap Menggunakan Pendekatan Machine Learning,” *Journal of Information Technology Ampera*, vol. 4, no. 3, pp. 265–276, Dec. 2023, doi: 10.51519/JOURNALITA.V4I3.436.
- [6] S. Verma, “Sentiment analysis of public services for smart society: Literature review and future research directions,” *Gov Inf Q*, vol. 39, no. 3, p. 101708, Jul. 2022, doi: 10.1016/J.GIQ.2022.101708.
- [7] M. Haikal, B. Prayogi, N. P. Ganatha, M. R. Ferdiansyah, and N. A. R. Rakhmawati, “Analisis Persepsi Masyarakat Dunia Terhadap Virtual Youtuber yang Menawarkan Pengalaman Kekasih,” *E-Link: Jurnal Teknik Elektro dan Informatika*, vol. 19, no. 1, pp. 90–98, May 2024, doi: 10.30587/E-LINK.V19I1.6746.
- [8] H. D. Al Assyam and F. N. Hasan, “Analisis Sentimen Twitter Terhadap Perpindahan Ibu Kota Negara Ke IKN Nusantara Menggunakan Orange Data Mining,” *KLIK: Kajian Ilmiah*

- Informatika dan Komputer*, vol. 4, no. 1, pp. 341–349, Aug. 2023, doi: 10.30865/KLIK.V4I1.957.
- [9] A. Zuntriana, “#1LIB1REF : Analisis Isi Pesan Advokasi Digital Satu Pustakawan Satu Referensi (One Librarian One Reference) di Twitter,” *Jurnal Al-Ma’arif: Ilmu Perpustakaan dan Informasi Islam*, vol. 3, no. 2, pp. 131–140, 2023, [Online]. Available: <https://rjfahuinib.org/index.php/almaarif/article/view/1284>
- [10] H. Jafari Pavarsi, N. Hariri, M. Alipour-Hafezi, F. Bab Al-Hawaeji, and M. Khademi, “Machine Indexing of Documents in the Field of Information Retrieval Using Text Mining in the RapidMiner Software,” *Iranian Journal of Information Processing and Management*, vol. 35, no. 2, pp. 349–374, Feb. 2020, doi: 10.35050/JIPM010.2020.054.
- [11] R. H. Sukarna and Y. Ansori, “Implementasi Data Mining Menggunakan Metode Naive Bayes dengan Feature Selection untuk Prediksi Kelulusan Mahasiswa Tepat Waktu,” *Jurnal Ilmiah Sains dan Teknologi*, vol. 6, no. 1, pp. 50–61, Feb. 2022, doi: 10.47080/SAINTEK.V6I1.1467.
- [12] K. V. S. Toy, Y. A. Sari, and I. Cholissodin, “Analisis Sentimen Twitter menggunakan Metode Naive Bayes dengan Relevance Frequency Feature Selection (Studi Kasus: Opini Masyarakat mengenai Kebijakan New Normal),” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 11, pp. 5068–5074, Oct. 2021, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/10172>
- [13] R. Mubarak, M. Hanafi, and D. Sasongko, “Komparasi Performa Naive Bayes Gaussian dan K-NN Untuk Prediksi Kelulusan Mahasiswa dengan CRISP-DM,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 6, pp. 2982–2991, Jun. 2024, doi: 10.30865/KLIK.V4I6.1924.
- [14] H. Asyraf and M. E. Prasetya, “Implementasi Metode CRISP DM dan Algoritma Decision Tree Untuk Strategi Produksi Kerajinan Tangan pada UMKM A,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 1, pp. 94–105, Jan. 2024, doi: 10.30865/MIB.V8I1.7050.
- [15] S. A. Perdana, S. F. Florentin, and A. Santoso, “ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN K-MEANS CLUSTERING STUDI KASUS APLIKASI ALFAGIFT,” *Sebatik*, vol. 26, no. 2, pp. 446–457, Dec. 2022, doi: 10.46984/SEBATIK.V26I2.1991.
- [16] H. Nuraliza, O. N. Pratiwi, and F. Hamami, “Analisis Sentimen IMBd Film Review Dataset Menggunakan Support Vector Machine (SVM) dan Seleksi Feature Importance,” *Jurnal Mirai Management*, vol. 7, no. 1, pp. 1–17, Aug. 2022, doi: 10.37531/MIRAI.V7I1.2222.
- [17] A. Pramudiansyah and H. Munte, “Segmentasi Pelanggan Menggunakan Algoritma K-Means Berdasarkan Model Recency Frequency Monetary,” *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, vol. 7, no. 2, pp. 06–19, Sep. 2021, doi: 10.35329/JIIK.V7I2.201.
- [18] Y. A. Singgalen, “Penerapan CRISP-DM dalam Klasifikasi Sentimen dan Analisis Perilaku Pembelian Layanan Akomodasi Hotel Berbasis Algoritma Decision Tree (DT),” *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 5, no. 2, pp. 237–248, Dec. 2023, doi: 10.30865/JSON.V5I2.7081.
- [19] M. Y. Abdilah and S. Suparni, “Rancang Bangun Sistem Informasi Inventory Stock Barang Usaha Fashion Pada Toko Online Mezzaluna Signature Depok,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, pp. 3852–3859, Jan. 2023, doi: 10.36040/JATI.V7I6.7735.
- [20] A. Fitri Ariani, K. Aulia, and A. A. La Ode, “Pengembangan Dashboard Interaktif Menggunakan Looker Studio untuk Visualisasi dan Prediksi Harga Komoditas Cabe di Jawa Timur,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 4, pp. 8067–8074, Aug. 2024, doi: 10.36040/JATI.V8I4.10616.