

IMPLEMENTASI ALGORITMA NAÏVE BAYES PADA ANALISIS SENTIMEN LARANGAN PENJUALAN ROKOK BATANGAN DI MEDIA TWITTER

1st Ai Novi Yanti
Universitas Buana Perjuangan Karawang
Karawang, Indonesia
IF19.aiyanti@mhs.ubpkarawang.ac.id
082121960371

2nd Deden Wahiddin
Universitas Buana Perjuangan Karawang
Karawang, Indonesia
deden.wahiddin@ubpkarawang.ac.id

3rd Rahmat
Universitas Buana Perjuangan Karawang
Karawang, Indonesia
rahmat@ubpkarawang.ac.id

Abstract— Merokok adalah kebiasaan yang tak terpisahkan dari kelompok perokok di Indonesia. Menurut laporan Kementerian Kesehatan, sekitar 33,8% dari total penduduk Indonesia adalah perokok, artinya satu dari tiga orang termasuk perokok. Selain dampak kesehatan, sektor ekonomi juga dipengaruhi karena industri tembakau berkontribusi signifikan pada APBN tahun 2020, mencapai 10,11%. Oleh karena itu, pada tahun 2023, pemerintah berencana melarang penjualan rokok batangan secara perorangan. Hal ini berarti perokok harus membeli rokok dalam kemasan. Kebijakan ini didasarkan pada Keputusan Presiden (Keppres) yang melarang penjualan rokok batangan, yaitu Keppres No. 25 Tahun 2022, sebagai bagian dari Program Penyusunan Peraturan Pemerintah Tahun 2023. Tujuan utama kebijakan ini adalah melindungi anak-anak agar tidak dapat membeli rokok batangan, yang lebih sering dibeli oleh mereka. Oleh karena itu, sebuah penelitian dilakukan untuk mengeksplorasi sentimen masyarakat terkait larangan penjualan rokok pada tahun 2023, khususnya di platform media sosial seperti Twitter. Metode yang digunakan dalam penelitian ini adalah Naive Bayes untuk menganalisis sentimen. Proses melibatkan beberapa tahap, termasuk pembersihan data, konversi huruf kecil, pemisahan kata (tokenizing), normalisasi, penghapusan kata-kata umum (stopword), dan penyederhanaan kata (stemming). Setelah tahap pra-pemrosesan, data diberi label dan diklasifikasikan menggunakan metode Naive Bayes. Hasil kategorisasi tweet cenderung bersifat positif atau negatif. Dari penelitian ini, tingkat akurasi yang dicapai adalah 73%, presisi sebesar 84%, dan recall mencapai 69%.

Kata kunci — Rokok, Twitter, Analisis Sentimen, Algoritma Naïve Bayes

I. PENDAHULUAN

Rencana pemerintah pada Tahun 2023 adalah melarang penjualan rokok batangan, sehingga masyarakat yang merokok akan diwajibkan membeli rokok dalam kemasan. Laura et al (2023) mengungkapkan bahwa merokok adalah aspek yang tidak dapat terpisahkan dari masyarakat Indonesia. Kementerian Kesehatan melaporkan bahwa sekitar 33,8% dari total populasi Indonesia adalah perokok, mengidentifikasi bahwa satu dari tiga orang termasuk perokok. Selain implikasi Kesehatan, sektor ekonomi juga terpengaruh oleh industri tembakau yang memberikan kontribusi signifikan pada APBN tahun 2020, yakni sebesar 10,11%. Penetapan aturan ini mengacu pada Keputusan Presiden (Keppres) mengenai larangan penjualan rokok batangan No. 25 Tahun 2022, yang terkait dengan Program Penyusunan Peraturan pemerintah Tahun 2023. Maksud dari kebijakan ini adalah untuk mencegah pembelian rokok batangan oleh anak di bawah umur, mengingat rokok batangan cenderung banyak dibeli oleh anak-anak.

Larangan penjualan rokok batangan saat ini menjadi topik perbincangan di platform Twitter. Dalam kalangan pengguna Twitter yang secara aktif berbagi pandangan mereka, informasi ini memiliki potensi untuk mengungkap berbagai fakta dan opini. Namun, pemanfaatan informasi tersebut memerlukan analisis yang cermat agar data yang dihasilkan memiliki relevansi bagi berbagai pihak dalam mendukung penilaian dan keputusan tertentu. Salah satu pendekatan yang dapat digunakan adalah Algoritma Naive Bayes, yang berguna dalam menganalisis sentimen. Dengan demikian, diperlukan analisis sentimen, yang merupakan disiplin dalam bidang komputasi yang bertujuan untuk mengubah data mentah yang tak terstruktur menjadi data yang terstruktur. Secara singkat, data tak terstruktur mencakup berbagai bentuk seperti bahasa asing, bahasa daerah, bahasa gaul, emotikon, URL web, dan alamat email. Data-data semacam itu dapat dianalisis melalui analisis sentimen untuk menghasilkan informasi yang berharga dan dapat digunakan dalam penyelidikan lebih lanjut [1].

Telah banyak dilakukan penelitian sebelumnya mengenai analisis sentimen. Salah satunya penelitian oleh Aziz and Fitri (2021) mengenai Analisis Sentimen Terhadap Kebijakan Pemerintah Tentang Larangan Mudik Hari Raya Idulfitri di Indonesia Tahun 2021 menggunakan Metode Naïve Bayes dengan hasil tingkat akurasi mencapai 56,52% [2]. Penelitian selanjutnya menurut Sulastri (2020) mengenai Klasifikasi *Naïve Bayes* dalam Analisis Sentimen terhadap Penolakan Dibukanya Larangan Ekspor Benih Lobster. Penelitian ini menghasilkan nilai evaluasi tertinggi di mana accuracy adalah 72.50%; precision adalah 88.64%; recall adalah 62.42%, dan AUC adalah 86.50% [3]. Penelitian menggunakan algoritma naïve bayes selanjutnya Aditya, Supriyati, and Listyorini (2022) mengenai Analisis Sentimen Pengguna Twitter Pada Rokok Elektrik (Vape) Di Indonesia Menggunakan Metode Naïve Bayes. Hasil Tingkat akurasi mencapai 77.5%, dengan jumlah polarity negatif sebesar 11.7%, jumlah polarity netral sebesar 77.3%, dan jumlah positif sebesar 11% [4].

Penelitian ini menganalisis sentimen data menggunakan pendekatan metode Naive Bayes dengan tujuan mengidentifikasi pendapat yang mencerminkan sentimen positif atau negatif. Fokus dari penelitian ini adalah untuk mengevaluasi tingkat kepuasan masyarakat terhadap tindakan pemerintah terkait larangan penjualan rokok yang akan diterapkan pada tahun 2023. Keputusan untuk memanfaatkan algoritma Naive Bayes diambil karena penelitian sebelumnya menghasilkan tingkat akurasi yang rendah. Oleh karena itu, penelitian ini dirancang untuk meningkatkan tingkat akurasi dan mencapai nilai yang lebih tinggi dalam analisis sentimen.

II. TINJAUAN PUSTAKA

A. Analisis Sentimen

Analisis sentimen adalah metode yang digunakan untuk mengekstraksi informasi dari suatu pendapat, yang kemudian diubah menjadi data yang mencerminkan sentiment dari pendapat masyarakat. Prinsip kerja sentimen melibatkan pengelompokan teks ke dalam kalimat atau dokumen, lalu menentukan pandangan yang terdapat dalam dokumen tersebut untuk selanjutnya dianalisis untuk menentukan apakah sentimennya positif atau negatif [5].

B. Kebijakan Larangan Penjualan Rokok Batangan

Pemerintah berencana melarang penjualan rokok batangan atau ketengan pada tahun 2023. Hal ini berarti bahwa orang-orang diharuskan membeli rokok dalam bentuk perbungkus. Rencana ini dijalankan melalui Keputusan Presiden (Keppres) No. 25 Tahun 2022 tentang Rangkaian Langkah Penyusunan Peraturan Pemerintah Nomor 109 Tahun 2012 yang terkait dengan Pengamanan Bahan yang mengandung Zat Adiktif, khususnya produk Tembakau bagi Kesehatan, Keputusan ini ditandatangani oleh Presiden pada 23 Desember 2022. Tujuan aturan ini adalah untuk mencegah penjualan rokok batangan kepada anak di bawah umur, mengingat rokok batangan banyak dibeli oleh anak-anak. Larangan ini diterapkan sebagai langkah preventif untuk mengurangi peredaran rokok di kalangan anak di bawah umur.

C. Twitter

[6] secara spesifik, Twitter adalah sebuah platform media sosial yang sederhana yang memberikan pengguna kemampuan untuk berbagi pesan singkat atau yang dikenal sebagai "tweet" dalam berbagai bentuk, termasuk teks, gambar, audio, video, dan format lainnya yang serupa. Di Indonesia, Twitter merupakan aplikasi yang populer dan banyak digunakan oleh masyarakat, terutama untuk menyuarakan pendapat mereka. Akun Twitter dengan nama "autobase" bisa menjadi sumber berharga untuk memperoleh pemahaman tentang sentimen yang ada di kalangan masyarakat.

III. METODE

A. Text Mining

Text mining merupakan sebuah proses analisis teks atau informasi yang diperoleh dari berbagai sumber atau dokumen. Data ini dapat berbentuk teks terstruktur maupun tidak terstruktur, yang kemudian diproses melalui text mining untuk menghasilkan informasi yang berguna. Beberapa aplikasi dari text mining mencakup klasifikasi, pengumpulan teks, analisis sentiment, ringkasan dokumen, dan penemuan hubungan antar entitas[7].

B. Naïve Bayes

Algoritma *Naïve Bayes* adalah teknik pembelajaran mesin yang berdasarkan teorema Bayes dan digunakan untuk membuat prediksi berdasarkan data historis. Ini merupakan metode kategorisasi sederhana namun presisinya tinggi. Metode ini menghitung probabilitas langsung dengan mudah, cocok untuk data besar. Klasifikasi *Naïve Bayes* masuk dalam pembelajaran terawasi karena memerlukan data pelatihan untuk membangun model. Dalam praktiknya, algoritma ini mengkategorikan data berdasarkan kesamaannya dengan cara yang paling mirip. Langkah-langkahnya termasuk membangun model dari data pelatihan, memvalidasi model dengan data uji, dan mengukur akurasi berdasarkan perbandingan hasil kontras antara klasifikasi sebenarnya dan prediksi model. Evaluasi ini memberikan penilaian kualitas model, mengidentifikasi apakah kinerjanya lemah, sedang, baik, atau sangat baik.

Menurut Rahayu, Fauzi, and Rahmat (2022) Kinerja *Naïve Bayes* menonjol dengan waktu klasifikasi yang sangat cepat, sehingga dapat meningkatkan kecepatan dalam memproses sistem analisis sentiment [6]. Berikut adalah rumus untuk menghitung Naïve Bayes:

[5] persamaan algoritma Naïve Bayes dapat dilihat pada (1)

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \dots (1)$$

Dimana:

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis

H : Merupakan sebuah class data kosong / belum diketahui

H : Hipotesis atau spekulasi data

$P(H|X)$: Merupakan hipotesis kemungkinan spekulasi H berdasarkan kondisi X (posterior probabilitas)

$P(H)$: Merupakan hipotesis kemungkinan spekulasi H (prior probabilitas)

$P(X)$: Probabilitas / kemungkinan X.

C. Confusion Matrix

Confusion matrix adalah metode yang digunakan untuk menilai kinerja model klasifikasi yang digunakan untuk memprediksi apakah suatu objek diklasifikasikan dengan benar atau salah. Matriks ini membandingkan hasil prediksi dengan kelas asli dari objek tersebut, yang berisi informasi mengenai nilai aktual dan prediksi dari klasifikasi. Setelah tweet berhasil diklasifikasikan oleh sistem, diperlukan ukuran yang akan mengindikasikan sejauh mana klasifikasi yang dilakukan oleh sistem adalah valid atau akurat.

Tabel 1 Confusion Matrix

Manual	Sistem	
	Positif	Negatif
Positif	True Positive (TP)	False Negative (FN)
Negatif	False Positive (FP)	True Negative (TN)

Keterangan :

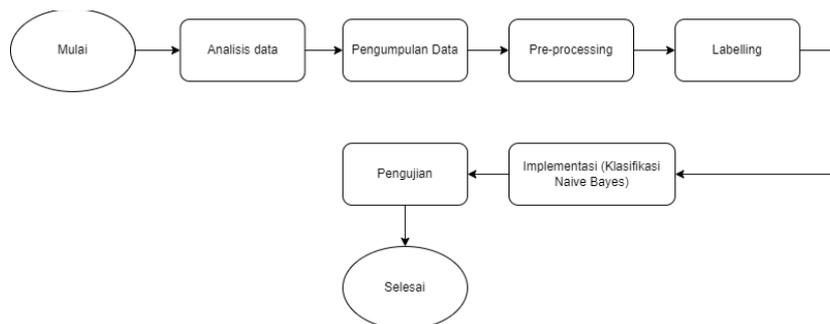
- TP = Jumlah data positif yang terklasifikasi dengan benar oleh sistem.
- TN = Jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
- FN = jumlah data negatif namun terklasifikasi salah oleh sistem.
- FP = jumlah data positif namun terklasifikasi salah oleh sistem.

D. TF-IDF

Tf-idf adalah teknik yang menggabungkan dua komponen: Term Frequency (TF) dan Inverse Document Frequency (IDF). Komponen Frekuensi Kata (TF) mengukur berapa kali suatu kata muncul dalam suatu dokumen, sementara komponen Inverse Document Frequency (IDF) mengukur sejauh mana kata tersebut jarang muncul di seluruh dokumen dalam kumpulan data.

IV. HASIL DAN PEMBAHASAN

Penelitian ini dimulai dengan analisis data lalu dilakukan pengumpulan data, kemudian setelah pengumpulan data dilakukan pre-processing, kemudian data yang sudah di *pre-processing* dilakukan pelabelan data, selanjutnya diimplementasikan menggunakan algoritma *Naïve Bayes*, kemudian data yang sudah diimplementasikan dilakukan pengujian dengan menggunakan *confusion matrix*.



Gambar 1 Proses Metode Penelitian

A. Analisis Data

Penelitian ini dilakukan analisis data untuk mengidentifikasi isu yang muncul di platform media sosial *Twitter*, yakni pandangan masyarakat terkait larangan penjualan rokok batangan. Analisis tersebut bertujuan untuk mengungkapkan dan mengumpulkan sentimen positif dan negatif. Hasil analisis akan di gambarkan sebagai berikut.

1. Sentimen positif merujuk pada Tindakan atau respons yang terjadi tanpa menghadapi penolakan atau memberikan dukungan.
2. Sentimen negatif menggambarkan Tindakan atau respons yang terjadi dalam situasi penolakan dan ditandai dengan sikap yang bersifat negatif atau tidak mendukung.

B. Pengumpulan Data

Penelitian ini dimulai dengan mengumpulkan data melalui platform media sosial *Twitter*. Data yang relevan dengan opini masyarakat tentang larangan penjualan rokok batangan dengan menggunakan metode algoritma *Naïve Bayes*. Proses pengumpulan data dari *Twitter* dijalankan melalui *API*, dan hasilnya disimpan dalam format csv.

1. Cleaning Data

Cleaning data adalah proses penghapusan data yang berupa data yang duplikat dan teks yang tidak mengandung makna dari hasil pengumpulan data.

Tabel 2 Penghapusan Data Duplikat

No	Data Duplikat
87	Larangan penjualan rokok eceran/batangan merugikan kelangsungan hidup jutaan pedagang kecil/asongan. #BatalkanRevisiPP109 Rokok Bukan Untuk Anak https://t.co/kis5rwVzua

Tabel 5 Hasil *Case Folding*

No	Sebelum	Sesudah
1	"aturan baru tata niaga rokok". beli bungkusannya, bawa foto copy npwp. beli batangan, bawa sktm dari desa setempat.	aturan baru tata niaga rokok beli bungkusannya bawa foto copy npwp beli batangan bawa sktm dari desa setempat
2	"di beberapa negara justru sudah dilarang. kita kan masih, tapi untuk yang batangan, tidak ya," jelas @jokowi #rokok #larangan #kerentanparokok https://t.co/iz9zpqxmgv	di beberapa negara justru sudah dilarang kita kan masih tapi untuk yang batangan tidak ya jelas
3	"pengawasannya akan terus dilakukan, karena itu sudah menjadi perintah undang-undang. jadi kita harus kerjakan," ujar wapres maruf amin. selain soal larangan penjualan rokok batangan, ada pula perubahan aturan. apa saja?	pengawasannya akan terus dilakukan karena itu sudah menjadi perintah undang-undang jadi kita harus kerjakan ujar wapres maruf amin selain soal larangan penjualan rokok batangan ada pula perubahan aturan apa saja
4	"perokok yang telah kecanduan akan tetap membeli rokok meskipun tidak dapat lagi membeli secara batangan"	perokok yang telah kecanduan akan tetap membeli rokok meskipun tidak dapat lagi membeli secara batangan
...
917	yg paling asyik nih, saat nyala batangan rokoknya ditaruh asbak, asapnya menebarkan aroma wangi yg lembut. kayak aroma parfum wangi apel itu. krasa segar seperti mengandung mint. secara keseluruhan rasanya halus.	yg paling asyik nih saat nyala batangan rokoknya ditaruh asbak asapnya menebarkan aroma wangi yg lembut kayak aroma parfum wangi apel itu krasa segar seperti mengandung mint secara keseluruhan rasanya halus
918	ylki dukung wacana larangan penjualan rokok batangan https://t.co/bcys3f3k2x	ylki dukung wacana larangan penjualan rokok batangan

2. Tokenizing

Pada proses ini adalah proses pemecahan string (kalimat yang dipecah menjadi perkata) teks setelah proses *Case Folding* selesai kalimat dipecah menjadi perkata.

Tabel 6 Hasil *Tokenizing*

No	Case Folding	Tokenizing
1	aturan baru tata niaga rokok beli bungkusannya bawa foto copy npwp beli batangan bawa sktm dari desa setempat	['aturan', 'baru', 'tata', 'niaga', 'rokok', 'beli', 'bungkusannya', 'foto', 'copy', 'npwp', 'beli', 'batangan', 'sktm', 'dari', 'desa', 'setempat']
2	di beberapa negara justru sudah dilarang kita kan masih tapi untuk yang batangan tidak ya jelas	['di', 'beberapa', 'negara', 'justru', 'sudah', 'dilarang', 'kita', 'kan', 'masih', 'tapi', 'untuk', 'yang', 'batangan', 'tidak', 'ya', 'jelas']
3	pengawasannya akan terus dilakukan karena itu sudah menjadi perintah undang-undang jadi kita harus kerjakan ujar wapres maruf amin selain soal larangan penjualan rokok batangan ada pula perubahan aturan apa saja	['pengawasannya', 'akan', 'terus', 'dilakukan', 'karena', 'itu', 'sudah', 'menjadi', 'perintah', 'undang-undang', 'jadi', 'kita', 'harus', 'kerjakan', 'ujar', 'wapres', 'maruf', 'amin', 'selain', 'soal', 'larangan', 'penjualan', 'rokok', 'batangan', 'ada', 'pula', 'perubahan', 'aturan', 'apa', 'saja']
4	perokok yang telah kecanduan akan tetap membeli rokok meskipun tidak dapat lagi membeli secara batangan	['perokok', 'yang', 'telah', 'kecanduan', 'akan', 'tetap', 'membeli', 'rokok', 'meskipun', 'tidak', 'dapat', 'lagi', 'membeli', 'secara', 'batangan']
...
917	yg paling asyik nih saat nyala batangan rokoknya ditaruh asbak asapnya menebarkan aroma wangi	['yg', 'paling', 'asyik', 'nih', 'saat', 'nyala', 'batangan', 'rokoknya', 'ditaruh', 'asbak', 'asapnya', 'menebarkan']

	yg lembut kayak aroma parfum wangi apel itu krasa segar seperti mengandung mint secara keseluruhan rasanya halus	'aroma', 'wangi', 'yg', 'lembut', 'kayak', 'aroma', 'parfum', 'wangi', 'apel', 'itu', 'krasa', 'segar', 'seperti', 'mengandung', 'mint', 'secara', 'keseluruhan', 'rasanya', 'halus']
918	ylki dukung wacana larangan penjualan rokok batangan	['ylki', 'dukung', 'wacana', 'larangan', 'penjualan', 'rokok', 'batangan']

3. *Normalize*

Proses ini dilakukan untuk merubah kata yang tidak baku menjadi kata baku.

Tabel 7 Hasil *Normalize*

No	Tokenizing	Normalize
1	['aturan', 'baru', 'tata', 'niaga', 'rokok', 'beli', 'bungkusanbawa', 'foto', 'copy', 'npwp', 'beli', 'batanganbawa', 'sktm', 'dari', 'desa', 'setempat']	['aturan', 'baru', 'tata', 'niaga', 'rokok', 'beli', 'bungkusanbawa', 'foto', 'copy', 'npwp', 'beli', 'batanganbawa', 'sktm', 'dari', 'desa', 'setempat']
2	['di', 'beberapa', 'negara', 'justru', 'sudah', 'dilarang', 'kita', 'kan', 'masih', 'tapi', 'untuk', 'yang', 'batangan', 'tidak', 'ya', 'jelas']	['di', 'beberapa', 'negara', 'justru', 'sudah', 'dilarang', 'kita', 'kan', 'masih', 'tapi', 'untuk', 'yang', 'batangan', 'tidak', 'ya', 'jelas']
3	['pengawasannya', 'akan', 'terus', 'dilakukan', 'karena', 'itu', 'sudah', 'menjadi', 'perintah', 'undangundang', 'jadi', 'kita', 'harus', 'kerjakan', 'ujar', 'wapres', 'maruf', 'amin', 'selain', 'soal', 'larangan', 'penjualan', 'rokok', 'batangan', 'ada', 'pula', 'perubahan', 'aturan', 'apa', 'saja']	['pengawasannya', 'akan', 'terus', 'dilakukan', 'karena', 'itu', 'sudah', 'menjadi', 'perintah', 'undangundang', 'jadi', 'kita', 'harus', 'kerjakan', 'ujar', 'wapres', 'maruf', 'semoga', 'selain', 'soal', 'larangan', 'penjualan', 'rokok', 'batangan', 'ada', 'pula', 'perubahan', 'aturan', 'apa', 'Hanya']
4	['perokok', 'yang', 'telah', 'kecanduan', 'akan', 'tetap', 'membeli', 'rokok', 'meskipun', 'tidak', 'dapat', 'lagi', 'membeli', 'secara', 'batangan']	['perokok', 'yang', 'telah', 'kecanduan', 'akan', 'tetap', 'membeli', 'rokok', 'meskipun', 'tidak', 'dapat', 'lagi', 'membeli', 'secara', 'batangan']
...
917	['yg', 'paling', 'asyik', 'nih', 'saat', 'nyala', 'batangan', 'rokoknya', 'ditaruh', 'asbak', 'asapnya', 'menebarkan', 'aroma', 'wangi', 'yg', 'lembut', 'kayak', 'aroma', 'parfum', 'wangi', 'apel', 'itu', 'krasa', 'segar', 'seperti', 'mengandung', 'mint', 'secara', 'keseluruhan', 'rasanya', 'halus']	['Yang', 'paling', 'asyik', 'ini', 'saat', 'nyala', 'batangan', 'rokoknya', 'ditaruh', 'asbak', 'asapnya', 'menebarkan', 'aroma', 'wangi', 'Yang', 'lembut', 'kayak', 'aroma', 'parfum', 'wangi', 'apel', 'itu', 'krasa', 'segar', 'seperti', 'mengandung', 'mint', 'secara', 'keseluruhan', 'rasanya', 'halus']
918	['ylki', 'dukung', 'wacana', 'larangan', 'penjualan', 'rokok', 'batangan']	['ylki', 'dukung', 'wacana', 'larangan', 'penjualan', 'rokok', 'batangan']

4. *Stopword*

Dalam tahap ini, dilakukan eliminasi kata-kata bantu yang tidak relevan, yang dikenal sebagai “stopword”.

Tabel 8 Hasil *Stopword*

No	Normalize	Stopword
1	['aturan', 'baru', 'tata', 'niaga', 'rokok', 'beli', 'bungkusanbawa', 'foto', 'copy', 'npwp', 'beli', 'batanganbawa', 'sktm', 'dari', 'desa', 'setempat']	['aturan', 'tata', 'niaga', 'rokok', 'beli', 'bungkusanbawa', 'foto', 'copy', 'npwp', 'beli', 'batanganbawa', 'sktm', 'desa']
2	['di', 'beberapa', 'negara', 'justru', 'sudah', 'dilarang', 'kita', 'kan', 'masih', 'tapi', 'untuk', 'yang', 'batangan', 'tidak', 'ya', 'jelas']	['negara', 'dilarang', 'batangan']
3	['pengawasannya', 'akan', 'terus', 'dilakukan', 'karena', 'itu', 'sudah', 'menjadi', 'perintah', 'undangundang', 'jadi', 'kita', 'harus', 'kerjakan', 'ujar', 'wapres', 'maruf', 'amin', 'selain', 'soal', 'larangan', 'penjualan', 'rokok', 'batangan', 'ada', 'pula', 'perubahan', 'aturan', 'apa', 'Hanya']	['pengawasannya', 'perintah', 'undangundang', 'kerjakan', 'wapres', 'maruf', 'semoga', 'larangan', 'penjualan', 'rokok', 'batangan', 'perubahan', 'aturan', 'Hanya']

4	['perokok', 'yang', 'telah', 'kecanduan', 'akan', 'tetap', 'membeli', 'rokok', 'meskipun', 'tidak', 'dapat', 'lagi', 'membeli', 'secara', 'batangan']	['perokok', 'kecanduan', 'membeli', 'rokok', 'membeli', 'batangan']
...
917	['Yang ', 'paling', 'asyik', 'ini', 'saat', 'nyala', 'batangan', 'rokoknya', 'ditaruh', 'asbak', 'asapnya', 'menebarkan', 'aroma', 'wangi', 'Yang ', 'lembut', 'kayak', 'aroma', 'parfum', 'wangi', 'apel', 'itu', 'krasa', 'segar', 'seperti', 'mengandung', 'mint', 'secara', 'keseluruhan', 'rasanya', 'halus']	['Yang ', 'asyik', 'nyala', 'batangan', 'rokoknya', 'ditaruh', 'asbak', 'asapnya', 'menebarkan', 'aroma', 'wangi', 'Yang ', 'lembut', 'kayak', 'aroma', 'parfum', 'wangi', 'apel', 'krasa', 'segar', 'mengandung', 'mint', 'halus']
918	['ylki', 'dukung', 'wacana', 'larangan', 'penjualan', 'rokok', 'batangan']	['ylki', 'dukung', 'wacana', 'larangan', 'penjualan', 'rokok', 'batangan']

5. *Stemming*

Proses ini adalah Langkah untuk mengubah kata-kata menjadi bentuk dasarnya, dengan menghapus imbuhan (awalan dan akhiran) dari kata-kata yang telah dihasilkan dari proses *stopword*. Proses ini dilakukan dengan menggunakan library sastrawi, karena Bahasa yang diuji adalah Bahasa Indonesia.

Tabel 9 Hasil *Stemming*

No	<i>Stopword</i>	<i>Stemming</i>
1	['aturan', 'tata', 'niaga', 'rokok', 'beli', 'bungkusanbawa', 'foto', 'copy', 'npwp', 'beli', 'batanganbawa', 'sktm', 'desa']	['atur', 'tata', 'niaga', 'rokok', 'beli', 'bungkusanbawa', 'foto', 'copy', 'npwp', 'beli', 'batanganbawa', 'sktm', 'desa']
2	['negara', 'dilarang', 'batangan']	['negara', 'larang', 'batang']
3	['pengawasannya', 'perintah', 'undangundang', 'kerjakan', 'wapres', 'maruf', 'semoga', 'larangan', 'penjualan', 'rokok', 'batangan', 'perubahan', 'aturan', 'Hanya']	['awas', 'perintah', 'undangundang', 'kerja', 'wapres', 'maruf', 'moga', 'larang', 'jual', 'rokok', 'batang', 'ubah', 'atur', 'hanya']
4	['perokok', 'kecanduan', 'membeli', 'rokok', 'membeli', 'batangan']	['rokok', 'candu', 'beli', 'rokok', 'beli', 'batang']
...
917	['Yang ', 'asyik', 'nyala', 'batangan', 'rokoknya', 'ditaruh', 'asbak', 'asapnya', 'menebarkan', 'aroma', 'wangi', 'Yang ', 'lembut', 'kayak', 'aroma', 'parfum', 'wangi', 'apel', 'krasa', 'segar', 'mengandung', 'mint', 'halus']	['yang', 'asyik', 'nyala', 'batang', 'rokok', 'taruh', 'asbak', 'asap', 'tebar', 'aroma', 'wangi', 'yang', 'lembut', 'kayak', 'aroma', 'parfum', 'wangi', 'apel', 'krasa', 'segar', 'kandung', 'mint', 'halus']
918	['ylki', 'dukung', 'wacana', 'larangan', 'penjualan', 'rokok', 'batangan']	['ylki', 'dukung', 'wacana', 'larang', 'jual', 'rokok', 'batang']

D. *Labelling*

Tahap *labelling* merupakan Langkah pelabelan data yang dilakukan setelah proses preprocessing. Dalam penelitian ini, pelabelan dilakukan dengan membagi data menjadi 2 kategori berdasarkan kalimat opini masyarakat, yakni sentimen positif dan sentimen negatif. Proses pelabelan ini dilakukan secara manual oleh Pionika Linda Sari, S.Pd. (Lulusan Sarjana Sastra Indonesia), dan data hasil pelabelan disimpan dalam format Excel.

No	<i>Tweet</i>	Sentimen
1	['atur', 'tata', 'niaga', 'rokok', 'beli', 'bungkusanbawa', 'foto', 'copy', 'npwp', 'beli', 'batanganbawa', 'sktm', 'desa']	Positif
2	['negara', 'larang', 'batang']	Negatif

...
917	['yang', 'asyik', 'nyala', 'batang', 'rokok', 'taruh', 'asbak', 'asap', 'tebar', 'aroma', 'wangi', 'yang', 'lembut', 'kayak', 'aroma', 'parfum', 'wangi', 'apel', 'krasa', 'segar', 'kandung', 'mint', 'halus']	Negatif
918	['ylki', 'dukung', 'wacana', 'larang', 'jual', 'rokok', 'batang']	Positif

E. Implementasi

1. TF-IDF

TF-IDF dilakukan setelah tahapan *pre-processing* dan pelabelan data kemudian akan dilakukan pembobotan data menggunakan TF-IDF berikut tahapan-tahapan TF-IDF.

- 1) Hitung Term Frequency (TF) total kemunculan istilah dalam dokumen di bagi dengan jumlah total kata dalam dokumen.
- 2) Hitung frekuensi dokumen terbalik (IDF) logaritma dari jumlah total dokumen dibagi dengan jumlah dokumen yang berisi term.
- 3) Hitung TF-IDF hasil perkalian TF dan IDF. Nilai menunjukkan pentingnya istilah dalam dokumen.
- 4) Normalisasi TF-IDF pastikan nilai TF-IDF berada pada skala yang sama sehingga dapat dibandingkan di seluruh dokumen.

2. Pembagian Data

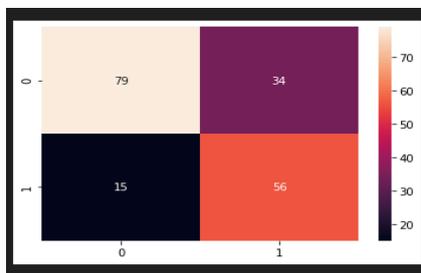
Tahapan ini dilakukan setelah melakukan tahapan TF-IDF. Pada proses pembagian data akan dibagi menjadi dua yaitu data latih dan data uji secara acak.

3. Klasifikasi Naïve Bayes

Setelah menyelesaikan Langkah-langkah TF-IDF, data akan dijalankan melalui algoritma *Naïve Bayes*. Implementasi Algoritma *Naïve Bayes* dimulai dengan menyiapkan data yang telah dikumpulkan dari media sosial *Twitter*. Dalam penelitian ini, data tersebut dikelompokkan menjadi dua kategori, yaitu sentimen positif dan sentimen negatif.

F. Pengujian

Pada penelitian ini, dilakukan pengujian untuk menilai efektivitas algoritma yang diterapkan. Proses pengujian dimulai dengan menggunakan matriks konfusi untuk mengukur tingkat akurasi algoritma. Langkah selanjutnya dilakukan pengujian algoritma menggunakan metode *Confusion Matrix* untuk meraih gambaran lebih lengkap mengenai kinerja algoritma tersebut.



Gambar 2 Hasil Confusion Matrix

Dalam Gambar 2 terlihat bahwa nilai 0 melambangkan kategori negatif, nilai 1 merupakan kategori positif. Setelah memahami struktur dari *confusion matrix* yang dijelaskan di atas, Langkah berikutnya adalah menyusun tabel *confusion matrix* untuk lebih mudah memahami performa algoritma Naïve Bayes terhadap data. Adapun tabel *confusion matrix* seperti tabel 2.

Tabel 10 Hasil *Confusion Matrix*

Manual	Sistem	
	Positif	Negatif
Positif	79	34
Negatif	15	56

$$\begin{aligned}
 akurasi &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
 &= \frac{79+56}{184} \times 100\% \\
 &= \frac{135}{184} \times 100\%
 \end{aligned}$$

$$\begin{aligned}
 &= 73\% \\
 \text{Precision} &= \frac{TP}{FP + TP} \times 100\% \\
 &= \frac{79}{15+79} \times 100\% \\
 &= \frac{79}{94} \times 100\% \\
 &= 84\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{FN + TP} \times 100\% \\
 &= \frac{79}{34+79} \times 100\% \\
 &= \frac{79}{113} \times 100\% \\
 &= 69\%
 \end{aligned}$$

Dari hasil penjumlahan di atas, angka 135 mencerminkan jumlah data yang diklasifikasikan dengan benar atau sesuai setiap kelas, sementara angka 184 mengindikasikan total data yang digunakan dalam uji. Tingkat akurasi mengindikasikan sejauh mana model yang telah dibuat mampu mengklasifikasikan data secara tepat. Akurasi diperoleh dengan membandingkan jumlah klasifikasi yang benar terhadap total data dalam kelompok. Dengan mengetahui nilai precision secara keseluruhan, seseorang dapat menilai sejauh mana sistem mampu mengidentifikasi keakuratan antara informasi yang diharapkan oleh pengguna dan jawaban yang dihasilkan oleh sistem. Dalam konteks ini, tingkat keberhasilan sistem dalam mencari informasi mencapai 73% sesuai dengan hasil penelitian ini.

V. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian terhadap larangan penjualan rokok batangan menggunakan algoritma *Naïve Bayes*, dapat diambil kesimpulan sebagai yaitu, penerapan algoritma *Naïve Bayes* dalam Analisis sentiment terhadap larangan penjualan rokok batangan melibatkan beberapa tahapan, termasuk analisis data, pengumpulan data, *preprocessing*, *labelling*, implementasi, dan pengujian. Kemudian dapat disimpulkan bahwa berdasarkan penelitian ini, larangan penjualan rokok batangan mendapat sentimen negatif. Proses pengujian menggunakan *confusion matrix* dengan algoritma *Naïve Bayes* menghasilkan tingkat akurasi sebesar 73%. Angka akurasi ini dihasilkan dari kumpulan data sebanyak 918 data yang mencakup sentimen positif dan negatif. Dalam pengujian ini, data dibagi menjadi dua bagian, dimana 80% sebagai data latih dan 20% sebagai data uji.

Saran untuk penelitian selanjutnya yaitu menerapkan algoritma klasifikasi lain untuk mengetahui tingkat akurasi yang berbeda seperti *Support Vector Machine*(SVM), *K-Nearest Neighbor*(KNN), dan *Decision Tree*, selain itu, menambahkan lebih banyak data latih dalam analisis sentimen terhadap larangan penjualan rokok batangan bisa membantu menyeimbangkan jumlah data pada setiap kelas.

PENGAKUAN

Naskah ilmiah ini merupakan bagian dari penelitian Tugas Akhir yang dilakukan oleh Ai Novi Yanti dengan judul Analisis Sentimen Larangan Penjualan Rokok Batangan Menggunakan Algoritma *Naïve Bayes*. Penelitian ini dibimbing oleh **Bapak Deden Wahiddin, M.Kom dan Bapak Rahmat, M.Pd.**

DAFTAR PUSTAKA

- [1] A. Laura *et al.*, "MEROKOK," vol. 6, no. 1, pp. 1–28, 2023.
- [2] A. Aziz and I. Fitri, "Analisis Sentimen Terhadap Kebijakan Pemerintah ... Metode *Naïve Bayes* (Abdul Aziz) |842 Teknik Informatika, Fakultas Teknologi Komunikasi dan Informatika," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, p. 7806700, 2021.
- [3] K. Sulastri, "Klasifikasi *Naïve Bayes* pada Analisis Sentimen atas Penolakan Dibukanya Larangan Ekspor Benih Lobster," *KERNEL J. Ris. Inov. Bid. Inform. dan Pendidik. Inform.*, vol. 1, no. 2, pp. 68–75, 2020, doi: 10.31284/j.kernel.2020.v1i2.1501.
- [4] D. R. Aditya, E. Supriyati, and T. Listyorini, "Analisis Sentimen Pengguna Twitter Terhadap Rokok Elektrik (Vape) Di Indonesia Menggunakan Metode *Naïve Bayes*," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 7, no. 1, pp. 43–50, 2022, doi: 10.29100/jipi.v7i1.2145.
- [5] T. D. Ramadhan, D. Wahiddin, and E. E. Awal, "Klasifikasi Sentimen Terhadap Pinjaman Online (Pinjol) Menggunakan Algoritma *Naive Bayes*," *Sci. Student J. Information, Technol. Sci.*, vol. IV, no. 1, pp. 82–87, 2023, [Online]. Available: www.tripadvisor.com
- [6] A. Safira, F. N. Hasan, K. J. Timur, and N. B. Classifier, "PAYLATER MENGGUNAKAN METODE NAIVE BAYES," vol. 5, no. 1, pp. 59–70, 1978.
- [7] Y. Pratama, D. T. Murdiansyah, and K. M. Lhaksmana, "Analisis Sentimen Kendaraan Listrik Pada Media Sosial Twitter Menggunakan Algoritma Logistic Regression dan Principal Component Analysis," vol. 7, pp. 529–535, 2023, doi:

- 10.30865/mib.v7i1.5575.
- [8] A. S. Rahayu, A. Fauzi, and R. Rahmat, "Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, p. 349, 2022, doi: 10.30865/json.v4i2.5398.