JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 2, April 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i2.8528
Hal 108-117
http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Analisis Sentimen Pada Komentar Mengenai Kartu Indonesia Pintar Menggunakan Metode Naïve Bayes

Dava Sindy, Sriani*

Fakultas Sains dan Teknologi, Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
Email: [1]davasindyy@gmail.com, [2,8]sriani@uinsu.ac.id
Email Penulis Korespondensi: sriani@uinsu.ac.id
Submitted **30-03-2025;** Accepted **28-04-2025;** Published **30-04-2025**

Abstrak

Kartu Indonesia Pintar (KIP) adalah program pemerintah Indonesia yang bertujuan untuk memberikan bantuan pendidikan kepada siswa dari keluarga kurang mampu. Program ini bertujuan untuk meningkatkan akses dan kesempatan belajar bagi anak-anak agar dapat menyelesaikan pendidikan hingga jenjang menengah atau kejuruan. Dalam era digital, opini publik mengenai kebijakan pemerintah, termasuk KIP, sering kali disampaikan melalui media sosial seperti X. Analisis sentimen adalah teknik dalam pemrosesan bahasa alami (NLP) yang digunakan untuk mengidentifikasi, mengekstrak, dan mengklasifikasikan opini dari teks. Salah satu algoritma yang umum digunakan dalam analisis sentimen adalah Naïve Bayes, yang bekerja berdasarkan Teorema Bayes dengan asumsi independensi antar fitur. Algoritma ini efektif dalam klasifikasi teks karena kesederhanaannya dan kemampuannya dalam menangani data dengan jumlah besar. Dengan menggunakan algoritma Naïve Bayes, analisis sentimen terhadap program KIP dapat memberikan wawasan mendalam mengenai respons masyarakat. Hasil analisis penelitian menunjukkan bahwa dari 656 data komentar yang dianalisis, diperoleh akurasi kualifikasi sebesar 88,6% dengan 58,2% komentar bersentimen positif dan 41,8% bersentimen negatif. Hasil analisis ini dapat membantu pemerintah dalam mengevaluasi kebijakan, memahami persepsi publik, serta mengoptimalkan implementasi program tepat sasaran.

Kata Kunci: Sentimen; Kartu Indonesia Pintar; X; Naïve Bayes; pemrosesan teks

Abstract

The Indonesia Smart Card (Kartu Indonesia Pintar – KIP) is a government program aimed at providing educational assistance to students from underprivileged families. This program seeks to improve access to education and increase learning opportunities for children, enabling them to complete their education up to the secondary or vocational level. In the digital era, public opinion regarding government policies, including KIP, is often expressed through social media platforms such as X. Sentiment analysis is a technique in natural language processing (NLP) used to identify, extract, and classify opinions from text. One of the commonly used algorithms for sentiment analysis is Naïve Bayes, which operates based on Bayes' Theorem with the assumption of feature independence. This algorithm is effective for text classification due to its simplicity and its ability to handle large datasets. By using the Naïve Bayes algorithm, sentiment analysis on the KIP program can provide deep insights into public responses. The research results show that out of 656 collected comments, classification accuracy reached **88.6%**, with **58.2%** of comments being positive and **41.8%** negative. These findings can assist the government in evaluating policies, understanding public perceptions, and optimizing program implementation to ensure it effectively reaches the intended beneficiaries.

Keywords: Sentiment; Smart Indonesia Card; X; Naïve Bayes; text preprocessing

1. PENDAHULUAN

Pendidikan di Indonesia selalu menghadapi masalah kemiskinan, dan kinerja siswa dievaluasi pada setiap akhir semester untuk mengetahui hasil pembelajaran yang dicapai. Kemiskinan inilah yang menjadi alasan masyarakat mengajukan KIP (Kartu Indonesia Pintar) pada lembaga pendidikan.[1] Upaya pemerintah untuk memberikan kesempatan seluas-luasnya kepada masyarakat agar memperoleh layanan pendidikan yaitu salah satunya melalui program Kartu Indonesia Pintar. Program tersebut diharapkan dapat membangun generasi yang unggul dan masyarakat generasi muda mendapatkan pendidikan yang layak.[2] Kebijakan Kartu Indonesia Pintar merupakan program pemerintah yang diluncurkan untuk mengatasi masalah yang terjadi karena masih banyak ditemukan kasus siswa yang masih usia sekolah namun memutuskan berhenti karena kesulitan biaya.[3]

Analisis sentimen adalah riset komputasional dari opini, sentimen dan emosi yang diekpresikan secara tekstual. Untuk melakukan analisis sentimen ada beberapa algoritme yang dapat digunakan salah satunya adalah algoritme *Naive Bayes. Naïve Bayes Classifier* merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yaitu memprediksi peluang berdasarkan pengalaman di masa sebelumnya (Teorema Bayes) dengan ciri utamanya adalah asumsi yang sangat kuat (naif) akan ketergantungan dari masing-masing kondisi atau kejadian.[4]

Dalam era digital dan perkembangan media sosial, X telah menjadi salah satu *platform* utama bagi masyarakat untuk berbagi pemikiran, pendapat, dan pengalaman mereka secara real-time. Dengan jutaan pengguna aktif yang terlibat dalam percakapan yang beragam, X menjadi sumber data yang kaya untuk menganalisis sentimen masyarakat terhadap berbagai topik termasuk program-program pemerintah seperti KIP (Kartu Indonesia Pintar). Penelitian dengan menggunakan topik Kartu Indonesia Pintar sudah dilakukan oleh Guterres dengan judul penelitian Analisis Sentimen Netizen *Twitter* Mengenai Beasiswa Kip Kuliah Di Indonesia Menggunakan Vader menggunakan data sebanyak 10.012 data yang diperoleh dari *Twitter* dengan hasil sentimen Positif sebesar 20% atau 2063 ulasan, sentimen Negatif sebesar 4% atau 372 ulasan dengan sentimen Netral 76% atau 7577 ulasan. Sentimen Netral menunjukkan hasil presentase yang paling banyak yang artinya bahwa pendapat netizen mengenai beasiswa KIP-Kuliah adalah biasa-biasa saja.[5]

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 2, April 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i2.8528
Hal 108-117
http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Penelitian lainnya yang di lakukan oleh Ginting et al., dengan judul Analisis Sentimen Berdasarkan Opini Dari *Twitter* Terhadap Pelaku Penyalahgunaan KIP K menggunakan data sebanyak 1159 data *tweet* yang setelah dilakukan proses eliminasi yang tidak berkaitan dengan isu yang dibahas menghasilkan hanya sekitar 42 data yang berkaitan dengan isu. Data yang digunakan kemudian dilakukan pemrosesan data menggunakan data latih sebanyak 30% dan 70% data uji sehingga total data uji adalah 30, dengan hasil pemberian label Positif sebanyak 10 data, Negatif sebanyak 15 data, dan Netral sebanyak 5 data.[6]

Penelitian ini memiliki GAP atau beberapa perbedaan signifikan dibandingkan dengan penelitian sebelumnya. Salah satu perbedaan utama terletak pada jumlah dan kualitas data yang digunakan. Penelitian Ginting et al. (2024) hanya menggunakan 42 data setelah eliminasi, yang relatif kecil dan kurang representatif untuk menggambarkan opini publik secara menyeluruh. Sebaliknya, penelitian ini menggunakan 656 data komentar dari media sosial X, yang memberikan cakupan lebih luas dan hasil yang lebih dapat diandalkan. Dari segi metode, beberapa penelitian sebelumnya menerapkan algoritma seperti *Vader, Random Forest, Support Vector Machine,* dan *IndoBERT* yang meskipun canggih, memiliki tingkat kompleksitas yang tinggi. Penelitian ini memilih algoritma Naïve Bayes yang lebih sederhana namun tetap efektif, terbukti dengan pencapaian akurasi sebesar 88,6%. Selain itu, fokus penelitian sebelumnya umumnya terbatas pada beasiswa KIP-Kuliah atau isu penyalahgunaan program, sedangkan penelitian ini menganalisis sentimen masyarakat secara umum terhadap program Kartu Indonesia Pintar (KIP). Dari sisi evaluasi, penelitian ini juga menambahkan pengukuran akurasi menggunakan *confusion matrix,* yang belum banyak dibahas pada penelitian-penelitian terdahulu.

Penelitian ini menunjukkan bahwa Program KIP telah berhasil meningkatkan akses pendidikan di Indonesia, terutama bagi laki-laki dan perempuan. Hasil penelitian ini memberikan indikasi bahwa program KIP telah efektif dalam meningkatkan akses pendidikan dan mengurangi anak putus sekolah di Indonesia.[7]

Penelitian ini bertujuan untuk menganalisis pengaruh dari adanya Program Indonesia Pintar dalam upaya pemerataan pendidikan di Indonesia. Data yang digunakan dalam penelitian ini didapat melalui studi literatur dengan mencari, mengumpulkan, dan membaca berbagai jurnal ataupun artikel tentang Program Indonesia Pintar yang mempunyai dampak pada pemerataan pendidikan di Indonesia. Dalam pengimplementasiannya, Program Indonesia Pintar sendiri memang masih terdapat berbagai hambatan yang menjadi salah satu faktor penghambat dalam menekan turunnya angka putus sekolah di Indonesia. Seperti kurang akuratnya penentuan calon peserta didik penerima Program Indonesia Pintar, masih terjadi keterlambatan dalam pencairan dana, sosialisasi program ini yang masih kurang, lamanya proses dalam memverifikasi kepemilikan kartu, dan tingkat kesadaran para wali murid terhadap peruntukkan bantuan Program Indonesia Pintar yang masih kurang. Dengan begitu, upaya pemerintah untuk menekan angka putus sekolah kurang optimal yang berdampak pada pemerataan pendidikan di Indonesia. Sehingga pengimplementasian Program Indonesia Pintar perlu didukung oleh beberapa faktor-faktor yang mendorong agar program ini dapat berjalan dengan lebih efektif. [8]

Penelitian ini bertujuan untuk menganalisis sentimen dari komentar *netizen Twitter* terhadap kesehatan mental masyarakat Indonesia dengan menggunakan metode Naïve Bayes dengan data yang diperoleh menggunakan *tools RapidMiner*. Data yang digunakan mulai dari tanggal 01 November 2021 sampai 30 November 2021 dengan jumlah data yang diperoleh adalah sebanyak 2369 data. Dari data tersebut di peroleh sentimen negatif sebanyak 1204 data atau 50,8%, positif sebanyak 1068 data atau 45,1%, dan netral sebanyak 97 data atau 4,1%. Sehingga disimpulkan dari hasil penelitian siketahui bahwa komentar pengguna media sosial *Twitter* terhadap kesehatan mental adalah negatif dengan klasifikasi negatif sebesar 50,8%. [9]

Pada penelitian yang dilakukan oleh Ginting et al., ini terdapat keterbatasan data pada beberapa penelitian sejenis yaitu kurangnya data yang memadai untuk dianalisis sebagai sample, adanya penggunaan metode yang berbeda pada beberapa penelitian sejenis sebelumnya dengan penelitian yang sedang dilakukan, minimnya informasi tambahan yang dapat memberikan wawasan lebih dalam, ketidakkonsistenan hasil, adanya hasil analisis yang tidak konsisten atau bertentangan satu sama lain, yang mengakibatkan tidak adanya hasil akhir yang jelas dari penelitian tersebut. Untuk mengatasi kekurangan ini, penelitian lebih lanjut untuk mengembangkan hasil menjadi lebih baik dengan penambahan jumlah *sample* data yang akan digunakan.

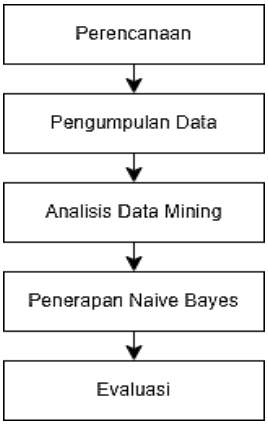
2. METODOLOGI PENELITIAN

2.1 Kerangka Penelitian

Penelitian ini akan menerapkan algoritma *Naïve Bayes* untuk menganalisis sentimen tentang Kartu Indonesia Pintar di media sosial X.

Pada Gambar 1 dibawah ini adalah tahapan yang dilalui dalam penerapan algoritma *Naïve Bayes* tersebut :

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 2, April 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i2.8528
Hal 108-117
http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 1 Kerangka Penelitian

Kerangka penelitian pada gambar 1 diatas ini menyusun urutan alur penelitian secara sistematis untuk mencapai tujuan yang diterapkan, sehingga dapat menjadi pedoman dalam pemecahan masalah yang akan dihadapi. Penelitian ini menggunakan metode kuantitatif, yang melibatkan pengumpulan data terukur melalui teknik statistik, matematika, atau komputasi untuk menyelidiki fenomena. Metode ini umum digunakan dalam Sains maupun matematis.

2.2 Perencanaan

Perencanaan penelitian tidak lain adalah gambaran tentang proses penelitian yang akan dilakukan untuk dapat memecahkan suatu permasalahan. Perencanaan dalam arti lain sama dengan *planning* sering kali rancu dengan rencana dalam arti desain penelitian. Perencanaan atau *planning* itu seharusnya dibuat sebelum melakukan penelitian di lapangan. Perencanaan penelitian pada umumnya berisi komponen-komponen penelitian yang secara komprehensif menggambarkan urutan tindakan yang harus dilakukan untuk mencapai tujuan penelitian.[10]

Langkah perencanaan yang dilakukan untuk menyusun penelitian, termasuk menentukan isu yang akan diteliti. Isu yang dipilih ialah mengenai analisis sentimen pada komentar mengenai Kartu Indonesia Pintar menggunakan metode *Naïve Bayes*.

2.3 Pengumpulan Data

Dalam melakukan penelitian, peneliti bisa menggunakan berbagai jenis teknik pengumpulan data, tergantung teknik mana yang sesuai dengan jenis penelitian dan juga pencarian sumber datanya, tergantung teknik mana yang sesuai dengan jenis penelitian dan juga pencarian sumber datanya. Teknik pengumpulan data harus dilakukan agar proses pengumpulan datanya baik dan peneliti juga harus melakukan berbagai tata cara dan juga prosedur yang berlaku agar pengumpulan data tersebut dapat menghasilkan data yang lebih valid dan akurat. [11]

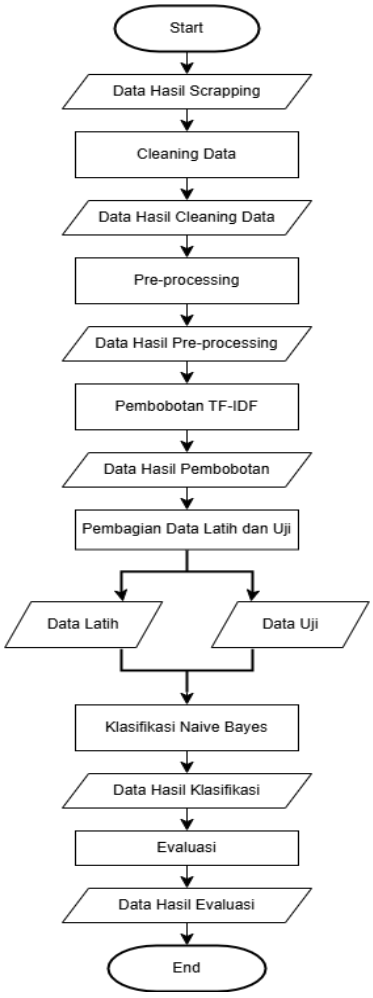
Proses pengumpulan data mengenai Kartu Indonesia Pintar dari aplikasi X dilakukan dengan cara *Web Scrapping* dengan menggunakan *tools Google Colaboratory* dan bahasa pemrograman Python dengan menggunakan kata kunci kartu indonesia pintar. Sehingga data yang telah didapatkan kemudian diproses melalui *preprocessing text* yang terdiri dari beberapa tahapan yaitu *case folding, tokenizing, filtering* dan *stemming*.[12]

2.4 Analisis *Data Mining*

Beberapa definisi awal dari *data mining* menyertakan fokus pada proses otomatisasi. Berry dan Linoff dalam buku *Data Mining Technique for Marketing, Sales, and Customers Support* mendefinisikan *data mining* sebagai suatu proses eksplorasi dan analisis secara otomatis maupun semi-otomatis terhadap data dalam jumlah besar dengan tujuann menemukan pola atau aturan yang berarti.[13]

Analisis *data mining* dilakukan berdasarkan data yang dikumpulkan selama proses pengumpulan data. Pada Gambar 2 dibawah ini adalah diagram alur yang menunjukkan setiap proses analisis sehingga penelitian dapat disusun dengan baik. Diagram ini memberikan gambaran jelas tentang struktur penelitian yang dilakukan, seperti yang terlihat pada gambar 2.2 dibawah ini :

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 2, April 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i2.8528
Hal 108-117
http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 2. *Flowchart* Klasifikasi Naive Bayes

Pada *flowchart* gambar 2.2 diatas menjelaskan bahwa proses dimulai dengan menganalisis kebutuhan data yang menjadi dasar untuk melakukan analisis sentimen pada komentar mengenai Kartu Indonesia Pintar menggunakan metode *Naïve Bayes*. Kemudian dilanjutkan dengan proses pengumpulan data yang dilakukan dengan cara *web scrapping*, kemudian dilakukan pelabelan data yang dimaksudkan untuk menentukan kategori tertentu untuk mengindikasi sentimen yang terkandung dalam teks tersebut dan lanjut ke tahap *preprocessing text* untuk membersihkan, mempersiapkan dan pengorganisasian data mentah dalam memastikan data yang di analisis telah tepat, relevan, dan siap untuk digunakan, dalam *preprocessing text* terdapat tahap lainnya yaitu *case folding* untuk menjadikannya huruf kecil untuk konsistensi, lalu *cleansing* untuk menghapus komponen lain selain kata dalam teks, lalu tokenisasi untuk memisahkan karakter, lalu *stopword removal* untuk menghapus kata yang kurang memberikan informasi, dan proses terakhir dari *preprocessing text* ialah *stemming* mengubah kata berimbuhan menjadi kata dasar.

Proses yang selanjutnya dilakukan ialah pembobotan TF-IDF sehingga meningkatkan kemampuan analisis sentimen, kemudian membagi data latih dan uji dengan alokasi 20% dari total dataset sebagai data uji dan alokasi 80% dari total dataset sebagai data pelatihan. Kemudian tahapan klasifikasi *Naïve Bayes* dan diakhiri dengan evaluasi untuk menilai kinerja analisis yang telah dilakukan dalam proses validasi.

2.5 Penerapan Naïve Bayes

Naïve Bayes merupakan suatu algoritma yang dapat mengklasifikasikan suatu variabel tertentu dengan menggunakan metode probabilitas dan statistik.[14] Algoritma klasifikasi ini dapat memprediksi probabilitas keanggotaan suatu kelas pada suatu objek dengan ciri-ciri yang termasuk dalam berkelompok atau kelas didasarkan pada teorema probabilitas Bayes.[15]

Kemudian setelah dataset selesai melalui tahap preprocessing dan ekstraksi fitur, proses berikutnya adalah pembelajaran dengan menggunakan metode klasifikasi *Naïve Bayes.* Setelah data dilatih, langkah selanjutnya adalah melakukan pengujian dengan data uji untuk mengevaluasi akurasi yang telah dilakukan.

Berikut ini merupakan langkah-langkah yang dilakukan dalam proses klasifikasi Naïve Bayes :

a. Menghitung probabilitas awal (*Prior Probability).*
   Tentukan probabilitas awal untuk setiap kelas berdasarkan proporsi data latih.

e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i2.8528
Hal 108-117
http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

$$P(Positif) = \frac{jumlah\ data\ positif}{total\ data} \quad (1)$$

b. Menghitung Probabilitas Kondisional (*Likelihood*).
Untuk setiap fitur, hitung probabilitas kemunculannya dalam setiap kelas.

$$P(kata = \text{kartu}|kelas = spam) \quad (2)$$

c. Menerapkan Teorema Bayes
Hitung probabilitas posterior untuk setiap kelas menggunakan rumus :

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

Keterangan :
X = Data dengan class yang belum diketahui
H = Hipotesis data X merupakan suatu *class* spesifik
$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi X (parteriori probabilitas)
$P(H)$ = Probabilitas hipotesis H (prior probabilitas)
$P(X|H)$ = Probabilitas X berdasarkan kondisi pada hipotesis H
$P(X)$ = Probabilitas dari X. [16].

d. Menentukan Kelas dengan Probabilitas Tertinggi
Pilih kelas dengan nilai probabilitas posterior tertinggi sebagai hasil prediksi.

Hasil klasifikasi ini kemudian akan dievaluasi menggunakan *convusion matrix* untuk melihat tingkat keakuratan pengelompokkan dataset. Evaluasi dilakukan untuk menilai kinerja analisis yang telah dilakukan dalam proses validasi dan pengujian menggunakan metode *Naive Bayes* yaitu *convusion matrix* dan klasifikasi hasil pengujian secara manual. Dalam penelitian ini, terdapat dua kelas yang dianalisis dalam *convusion matrix*, yaitu kelas positif dan kelas negatif.

3. HASIL DAN PEMBAHASAN

3.1 Pembahasan

Pada penelitian ini dilakukan analisis yang bertujuan untuk menilai sejauh mana sistem yang dikembangkan mampu membangun model prediksi dan klasifikasi berdasarkan pembelajaran dari data latih. Evaluasi ini bertujuan untuk mengukur kemampuan sistem dalam menganalisis sentimen opini masyarakat terkait kartu indonesia pintar pada laman media sosial X. *Google Colaboratory* digunakan sebagai *platform* pemrosesan data. Dataset yang diperoleh melalui teknik *crawling* disimpan dalam format .csv agar lebih mudah dianalisis.

Proses *preprocessing* teks melalui beberapa tahapan agar data siap untuk tahap pengolahan berikutnya. *Preprocessing* teks melibatkan lima langkah utama, yaitu normalisasi, *case folding*, tokenisasi dan *stemming*. Proses ini memungkinkan model untuk menentukan kategori teks baru secara efisien dan efektif, meskipun algoritma ini membuat asumsi independensi antar fitur. Dengan memahami setiap langkah dalam implementasi ini, kita dapat mengoptimalkan penggunaan Naïve Bayes pada berbagai tugas klasifikasi teks.[17]

Pada Gambar 3 berikut ini adalah hasil *output* dai *preprocessing* teks yang dilakukan dengan *Phyton* :

1 to 25 of 656 entries Filter

index	Original Text	Cleaning	Normalization	Stopword Removal	Stemming
0	penjabat pj gubernur dki jakarta heru budi hartono mencabut kartu jakarta pintar kjp pelssajaar ketahuan merkk ngevpe mengonsumsi rkk elektrik menekan antidaka konsumsi rkk indonesia adikadik mendapatidakan merkk merkk itu	penjabat pj gubernur dki jakarta heru budi hartono mencabut kartu jakarta pintar kjp pelssajaar ketahuan merkk ngevpe mengonsumsi rkk elektrik menekan antidaka konsumsi rkk indonesia adikadik mendapatidakan merkk merkk itu	penjabat pj gubernur dki jakarta heru budi hartono mencabut kartu jakarta pintar kjp pelssajaar ketahuan merkk ngevpe mengonsumsi rkk elektrik menekan antidaka konsumsi rkk indonesia adikadik mendapatidakan merkk merkk itu	penjabat pj gubernur dki jakarta heru budi hartono mencabut kartu jakarta pintar kjp pelssajaar ketahuan merkk ngevpe mengonsumsi rkk elektrik menekan antidaka konsumsi rkk indonesia adikadik mendapatidakan merkk merkk itu	jabat pj gubernur dki jakarta heru budi hartono cabut kartu jakarta pintar kjp pelssajaar tahu merkk ngevpe konsumsi rkk elektrik tekan antidaka konsumsi rkk indonesia adikadik mendapatidakan merkk merkk itu
1	sepertinya si kehabisan supplemen kartu indonesia pintar junjuntidaknnya tolol sekali tweet id	sepertinya si kehabisan supplemen kartu indonesia pintar junjuntidaknnya tolol sekali tweet id	sepertinya si kehabisan supplemen kartu indonesia pintar junjuntidaknnya tolol sekali tweet id	sepertinya si kehabisan supplemen kartu indonesia pintar junjuntidaknnya tolol sekali tweet id	seperti si habis supplemen kartu indonesia pintar junjuntidaknnya tolol sekali tweet id
2	beasiswa kip kuliah 2024 halo vets berikut informasi mengenai beasiswa kartu indonesia pintar kip kuliah tahun 2024 beasiswa diikuti mahasiswa antidakatan 2024 sayaarat ketentuan dapat dilihat postintidakn atas	beasiswa kip kuliah 2024 halo vets berikut informasi mengenai beasiswa kartu indonesia pintar kip kuliah tahun 2024 beasiswa diikuti mahasiswa antidakatan 2024 sayaarat ketentuan dapat dilihat postintidakn atas	beasiswa kip kuliah 2024 halo vets berikut informasi mengenai beasiswa kartu indonesia pintar kip kuliah tahun 2024 beasiswa diikuti mahasiswa antidakatan 2024 sayaarat ketentuan dapat dilihat postintidakn atas	beasiswa kip kuliah 2024 halo vets berikut informasi mengenai beasiswa kartu indonesia pintar kip kuliah tahun 2024 beasiswa diikuti mahasiswa antidakatan 2024 sayaarat ketentuan dapat dilihat postintidakn atas	beasiswa kip kuliah 2024 halo vets ikut informasi kena beasiswa kartu indonesia pintar kip kuliah tahun 2024 beasiswa ikut mahasiswa antidakatan 2024 sayaarat tentu dapat lihat postintidakn atas
3	penjabat pj gubernur dki jakarta heru budi hartono mentidakncam mencabut kartu jakarta pintar kjp kartu jakarta mahasiswa unggul kjmu milik penerima diketahui kecanduan judi online judol	penjabat pj gubernur dki jakarta heru budi hartono mentidakncam mencabut kartu jakarta pintar kjp kartu jakarta mahasiswa unggul kjmu milik penerima diketahui kecanduan judi online judol	penjabat pj gubernur dki jakarta heru budi hartono mentidakncam mencabut kartu jakarta pintar kjp kartu jakarta mahasiswa unggul kjmu milik penerima diketahui kecanduan judi online judol	penjabat pj gubernur dki jakarta heru budi hartono mentidakncam mencabut kartu jakarta pintar kjp kartu jakarta mahasiswa unggul kjmu milik penerima diketahui kecanduan judi online judol	jabat pj gubernur dki jakarta heru budi hartono mentidakncam cabut kartu jakarta pintar kjp kartu jakarta mahasiswa unggul kjmu milik terima tahu candu judi online judol
4	pemulihan sistem layanan kartu indonesia pintar kuliah kip kuliah selesai sistem kip kuliah ssudah diakses yuk simak dentidakn saksama informasi atas mengetahui lantidakah apa harus kamu lakukan selanjutnya	pemulihan sistem layanan kartu indonesia pintar kuliah kip kuliah selesai sistem kip kuliah ssudah diakses yuk simak dentidakn saksama informasi atas mengetahui lantidakah apa harus kamu lakukan selanjutnya	pemulihan sistem layanan kartu indonesia pintar kuliah kip kuliah selesai sistem kip kuliah ssudah diakses yuk simak dentidakn saksama informasi atas mengetahui lantidakah apa harus kamu lakukan selanjutnya	pemulihan sistem layanan kartu indonesia pintar kuliah kip kuliah selesai sistem kip kuliah ssudah diakses yuk simak dentidakn saksama informasi atas mengetahui lantidakah apa harus kamu lakukan selanjutnya	pulih sistem layan kartu indonesia pintar kuliah kip kuliah selesai sistem kip kuliah ssudah akses yuk simak dentidakn saksama informasi atas tahu lantidakah apa harus kamu laku lanjut
5	keren sekali kartu indonesia pintar ntidaksih harapan buat anakanak sekamuruh indonesia semotidak semakin banyak sukses berkat bantuan	keren sekali kartu indonesia pintar ntidaksih harapan buat anakanak sekamuruh indonesia semotidak semakin banyak sukses berkat bantuan	keren sekali kartu indonesia pintar ntidaksih harapan buat anakanak sekamuruh indonesia semotidak semakin banyak sukses berkat bantuan	keren sekali kartu indonesia pintar ntidaksih harapan buat anakanak sekamuruh indonesia semotidak semakin banyak sukses berkat bantuan	keren sekali kartu indonesia pintar ntidaksih harap buat anakanak sekamuruh indonesia semotidak makin banyak sukses berkat bantu
6	penjabat pj gubernur dki jakarta heru budi hartono mentidakncam mencabut kartu jakarta pintar kjp kartu jakarta mahasiswa unggul kjmu milik penerima diketahui kecanduan judi online judol	penjabat pj gubernur dki jakarta heru budi hartono mentidakncam mencabut kartu jakarta pintar kjp kartu jakarta mahasiswa unggul kjmu milik penerima diketahui kecanduan judi online judol	penjabat pj gubernur dki jakarta heru budi hartono mentidakncam mencabut kartu jakarta pintar kjp kartu jakarta mahasiswa unggul kjmu milik penerima diketahui kecanduan judi online judol	penjabat pj gubernur dki jakarta heru budi hartono mentidakncam mencabut kartu jakarta pintar kjp kartu jakarta mahasiswa unggul kjmu milik penerima diketahui kecanduan judi online judol	jabat pj gubernur dki jakarta heru budi hartono mentidakncam cabut kartu jakarta pintar kjp kartu jakarta mahasiswa unggul kjmu milik terima tahu candu judi online judol

✓ 0 d selesai pada 02.47

Gambar 3. *Preprocessing* Teks

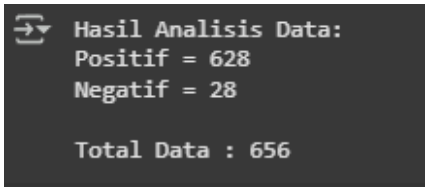
Pada gambar 3 ditampilkan hasil dari *preprocessing* data akan digunakan untuk melakukan pelabelan. Langkah-langkah yang dilakukan dalam proses *preprocessing text* adalah normalisasi (mengubah kata-kata tidak baku menjadi

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 2, April 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i2.8528
Hal 108-117
http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

baku), *case folding* (mengubah semua teks menjadi huruf kecil), tokenisasi (memisahkan kalimat menjadi kata-kata), *stemming* (mengubah kata ke bentuk dasar). Penilaian skor dokumen dilakukan dengan merujuk pada *wordlist* dalam kamus *lexicon* yang telah diberikan bobot. Skor diperoleh dengan menjumlahkan nilai bobot setiap kata dalam dokumen teks ulasan, yang kemudian menghasilkan tiga kelas yaitu positif, netral, dan negatif. Dalam penelitian ini, hanya terdapat dua kelas yang digunakan yaitu positif dan negatif.

Pelabelan data merupakan proses memberikan label positif dan negatif pada data teks. Proses ini dapat dilakukan secara manual dan otomatis, yaitu dengan bantuan *machine learning.* Pelabelan data secara manual tidak efektif karena sangat bersifat subjektif. Pelabelan otomatis dapat menggunakan *sentiment lexicon dataset* atau kamus emosi, cara ini baik digunakan untuk jumlah teks yang besar sehingga dapat meminimalkan kesalahan pelabelan dokumen. Pelabelan dilakukan untuk membentuk data latih, dan dilakukan untuk membentuk data latih, dan dilakukan juga pelabelan pada data uji untuk menghitung akurasi dari hasil sentimen yang di dapat secara otomatis.[18]

Pada gambar 4 ditampilkan hasil dari pemrosesan *labelling data* yang dilakukan menggunakan *google colaboratory* :

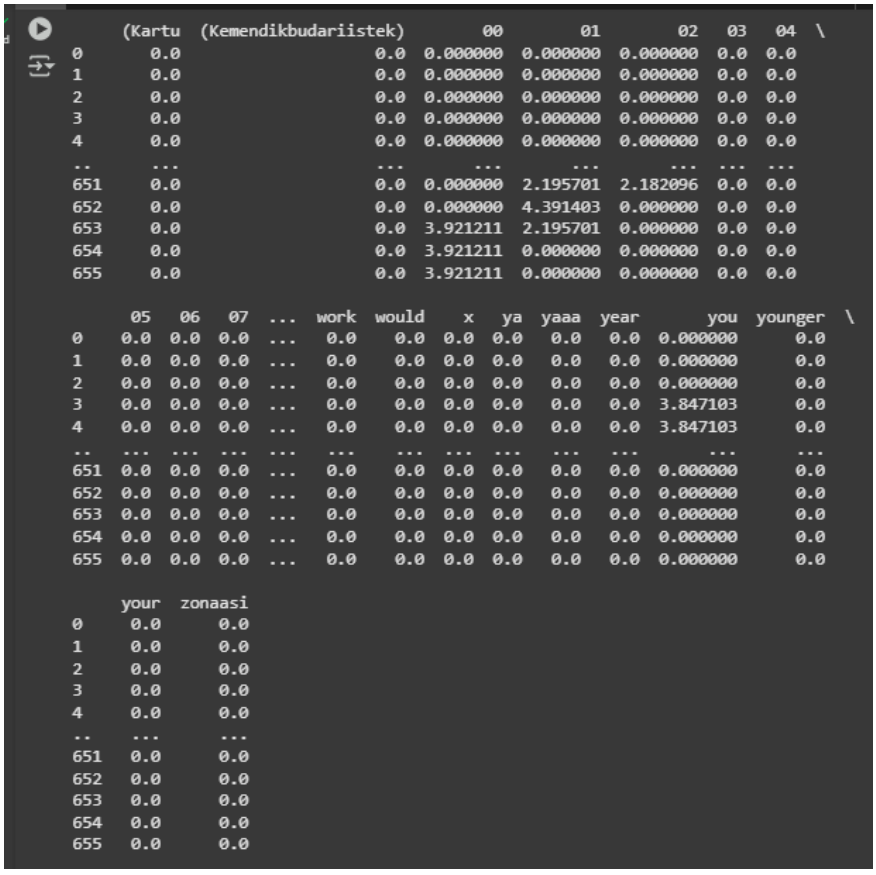


Gambar 4. *Labeling* Data

Berdasarkan gambar 4 diperoleh komentar positif sebanyak 628 data dan komentar negatif sebanyak 28 data dengan jumlah total komentar setelah proses *preprocessing text* sebanyak 656 data.

Term Frequency-Inverse Document Frequency (TF-IDF) adalah proses mengubah kata menjadi nilai numerik atau *vector* yang biasa disebut pembobotan. Tujuan dari TF-IDF adalah untuk menghitung pentingnya sebuah kata dalam sebuah dokumen dengan memberikan bobot pada setiap kata dalam setiap dokumen untuk mengidentifikasi dan menghitung jumlah kemunculan kata tersebut.[19] Perhitungan pembobotan kata (*term*) menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dilakukan secara otomatis pada seluruh kata dalam data ulasan aplikasi X. Proses ini dilakukan dengan memanfaatkan *Python* dan menggunakan *syntax TfidfVectorizer* dari *sklearn.feature_extraction.*

Pada gambar 5 menampilkan hasil dari perhitungan *output* pembobotan TF-IDF yang dapat dilihat seperti berikut:



	(Kartu	(Kemendikbudariistek)	00	01	02	03	04	\
0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	
1	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	
2	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	
3	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	
4	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	
..	...	...	...	...	...	...	...	
651	0.0	0.0	0.000000	2.195701	2.182096	0.0	0.0	
652	0.0	0.0	0.000000	4.391403	0.000000	0.0	0.0	
653	0.0	0.0	3.921211	2.195701	0.000000	0.0	0.0	
654	0.0	0.0	3.921211	0.000000	0.000000	0.0	0.0	
655	0.0	0.0	3.921211	0.000000	0.000000	0.0	0.0	

	05	06	07	...	work	would	x	ya	yaaa	year	you	younger	\
0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	
1	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	
2	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	
3	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	3.847103	0.0	
4	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	3.847103	0.0	
..	...	...	...	...	...	...	...	...	...	...	...	...	
651	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	
652	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	
653	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	
654	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	
655	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	

	your	zonaasi
0	0.0	0.0
1	0.0	0.0
2	0.0	0.0
3	0.0	0.0
4	0.0	0.0
..	...	...
651	0.0	0.0
652	0.0	0.0
653	0.0	0.0
654	0.0	0.0
655	0.0	0.0

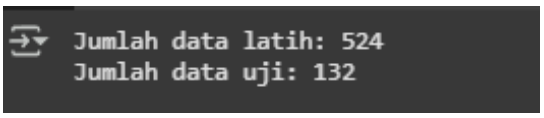
Gambar 5. *Output* TF-IDF

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 2, April 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i2.8528
Hal 108-117
http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Berdasarkan hasil yang ditampilkan pada Gambar 3.3, diperoleh nilai bobot untuk 847 kata (term) yang ada pada setiap dokumen, mulai dari dokumen 0 (d0) hingga dokumen 655 (d655).

Setelah memberikan nilai dan bobot pada setiap kata dalam teks ulasan, data kemudian dibagi menjadi dua bagian, yaitu data latih dan data uji. Berdasarkan Prinsip Pareto, perbandingan yang paling umum digunakan adalah 80:20 untuk membagi data latih dan data uji. Oleh karena itu, dalam penelitian ini, perbandingan data latih dan data uji yang digunakan adalah 80:20. Pembagian data tersebut dilakukan dengan menggunakan Python dan fungsi *train_test_split* dari *sklearn.model_selection*.

Pada gambar 6 menampilkan hasil pembagian data sebagai berikut :



Gambar 6. *Split* Data

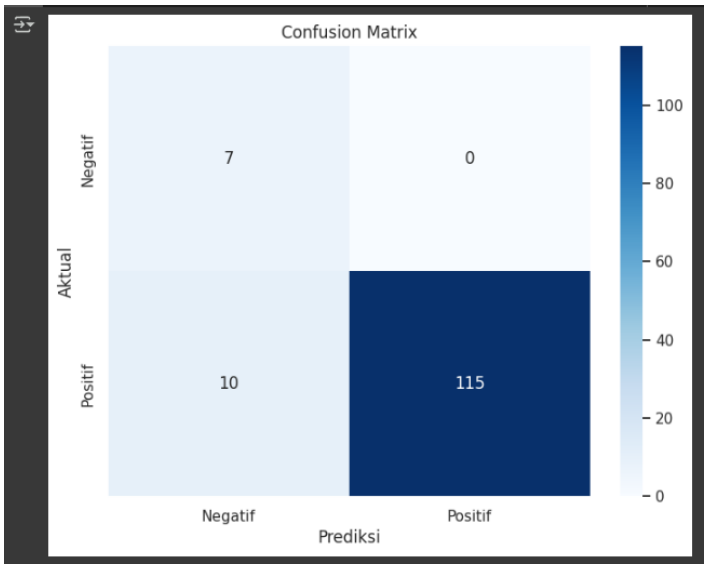
Pada gambar 3.4 dapat dilihat bahwa pembagian data (*split data*) dilakukan dengan perbandingan 80% data latih dan 20% data uji, yang kemudian menghasilkan pembagian data sebanyak 524 data sebagai data latih dan 132 data sebagai data uji dari total keseluruhan data sebanyak 656 data.

Metode *Naïve Bayes* diterapkan dalam proses klasifikasi, dengan hasil yang ditampilkan dalam bentuk *confusion matrix*. Matriks ini menggambarkan jumlah data uji yang berhasil diprediksi kedalam kategori tertentu, menggunakan format 2x2 untuk merepresentasikan dua kelas sentime yaitu, positif dan negatif. Dari hasil *confusion matrix*, tingkat akurasi presisi, *recall*, dan *F1-score* sistem dapat dievaluasi.

3.2 Hasil Pengujian

Hasil evaluasi dengan metode *Naïve Bayes* disajikan dalam bentuk *confusion matrix*, yang menunjukkan perbandingan hasil klasifikasi yang diperoleh dengan menggunakan metode pembobotan TF-IDF. Matriks ini memiliki format 2x2 dan digunakan untuk menghitung nilai akurasi, presisi, *recall*, serta *F1-score*. Nilai-nilai tersebut membantu dalam mengevaluasi performa sistem dalam analisis sentimen dan klasifikasi menggunakan pendekatan *Naive Bayes*.

Pada gambar 7 dapat dilihat hasil dari *confusion matrix* dari pemrosesan yang dilakukan menggunakan *google colaboratory :*



Gambar 7. *Confusion Matrix*

Pada gambar 7 memunculkan hasil kesuluruhan perhitungan secara otomatis yang dilakukan dengan menggunakan *google colaboratory yang* menunjukkan bahwa metode *Naïve Bayes* dengan pembobotan TF-IDF menghasilkan akurasi sebesar 0.9242 atau 92.42%. Selain itu, nilai presisi yang diperoleh adalah 1 atau 100%, dengan *recall* sebesar 1 atau 100% dan *f1-score* mencapai 1 atau 100%. Dan, dibawah ini merupakan perhitungan manual dari *confusion matrix* yang menunjukkan bahwa data yang dihasilkan secara otomatis dengan data yang dihitung secara manual memiliki hasil yang sama atau setara.

$$(1)\ Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} = \frac{115 + 7}{(115 + 7 + 0 + 10)} = 0.9242424242$$

$$(2) Precision\ Positif = \frac{TP}{(TP + FP)} = \frac{115}{(115 + 0)} = 1.0$$

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 2, April 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i2.8528
Hal 108-117
http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

$$Precision\ Negatif = \frac{TN}{(TN+FN)} = \frac{7}{(7+10)} = 0.4117647059$$

$$Weighted\ Average = \frac{(Precision\_Positif) \times (Number\_Positif) + (Precision\_Negatif) \times (Number\_Negatif)}{(Total\_Number)}$$

$$= \frac{1 \times 125 + 0.4117647059 \times 17}{(132)} = 1$$

$$(3) Recall\ Positif = \frac{TP}{(TP+FN)} = \frac{115}{(115+10)} = 0.92$$

$$Recall\ Negatif = \frac{TN}{(TN+FP)} = \frac{7}{(7+0)} = 1$$

$$Weighted\ Average = \frac{(Recall\_Positif) \times (Number\_Positif) + (Recall\_Negatif) \times (Number\_Negatif)}{(Total\_Number)} = \frac{0.92 \times 125 + 1 \times 17}{(132)}$$

$$= 1$$

$$(4)\ F - Measure = \frac{2\ (recall\ \times precision)}{(recall + precision)} = \frac{2(1\ .\times 1)}{(1\ +1)} = 1$$

Pada gambar 8 dibawah ini, menunjukkan hasil akurasi yang dilakukan menggunakan *google colaboratory :*

```
Akurasi model naive bayes : 0.9242424242424242
Precision: 0.9688057040998217
Recall: 0.9242424242424242
F1-score: 0.9384469696969697
Confusion Matrix:
[[  7   0]
 [ 10 115]]
              precision    recall  f1-score   support

     Negatif       0.41      1.00      0.58         7
     Positif       1.00      0.92      0.96       125

    accuracy                           0.92       132
   macro avg       0.71      0.96      0.77       132
weighted avg       0.97      0.92      0.94       132
```

Gambar 8. Hasil Akurasi

Dari gambar 8 diatas, hasil perhitungan model *Naive Bayes* dengan pembobotan TF-IDF menunjukkan performa yang sangat baik dalam mengklasifikasikan sentimen tweet mengenai Kartu Indonesia Pintar, dengan akurasi, presisi, *recall*, dan *F1-score* yang tinggi. Model sangat handal dalam mengidentifikasi *tweet* positif, dengan presisi yang mendekati sempurna.

Meskipun *recall*-nya sedikit lebih rendah daripada presisi, model tetap berhasil menemukan sebagian besar *tweet* positif. Area yang perlu diperhatikan adalah *False Negative* (FN), di mana model salah memprediksi 10 *tweet* positif sebagai negatif. Ini mungkin perlu dianalisis lebih lanjut untuk meningkatkan performa model. Secara keseluruhan, model ini dapat dianggap sangat efektif untuk tugas klasifikasi sentimen ini. Namun, selalu ada ruang untuk peningkatan dan analisis lebih lanjut untuk menyempurnakan model.

Gambar 9 merupakan hasil dari proses visualisasi yang dilakukan menggunakan *google colaboratory* secara otomatis :

Gambar 9. Hasil Visualisasi

Pada gambar 9 menampilkan hasil dari proses visualisasi yang dilakukan menggunakan *google colaboratory* secara otomatis yang menunjukkan bahwa semakin besar ukuran kata tersebut berarti jumlah frekuensi kata tersebut

banyak muncul di dalam dataset yang digunakan, dan sebaliknya semakin kecil ukuran kata tersebut maka semakin sedikit pula frekuensi kata tersebut muncul di dalam dataset. Dari hasil visualisasi diatas dapat dilihat bahwa kata "kartu" banyak muncul di dalam dataset di tandai dengan ukurannya yang besar, sedangkan kata yang paling sedikit muncul dalam dataset ialah kata "ka" di tandai dengan ukurannya yanbg paling kecil diantara seluruh kata yang divisualisasikan.

4. KESIMPULAN

Penelitian mengenai analisis sentimen terhadap komentar aplikasi X tentang Kartu Indonesia Pintar dilakukan dengan mengumpulkan 662 komentar menggunakan kata kunci "kartu indonesia pintar". Data yang diperoleh kemudian diproses melalui tahap *preprocessing*, yang mencakup pembersihan teks (*cleaning*), normalisasi huruf (*case folding*), pemisahan kata (*tokenizing*), penghapusan kata umum (*stopword removal*), *stemming*, dan normalisasi. Proses ini bertujuan untuk menyiapkan data sebelum dilakukan pelabelan menggunakan metode TF-IDF. Setelah proses pelabelan selesai, data diklasifikasikan menggunakan metode *Naïve Bayes*. Hasil analisis menunjukkan bahwa dari 662 komentar, sebanyak 628 termasuk dalam kategori sentimen positif, sementara 28 memiliki sentimen negatif. Dominasi komentar positif ini disebabkan oleh banyaknya pengguna yang memberikan respon baik terhadap Kartu Indonesia Pintar.
Evaluasi performa klasifikasi menggunakan *confusion matrix* menunjukkan bahwa metode *Naïve Bayes* dengan pembobotan TF-IDF menghasilkan akurasi sebesar 0.9242 atau 92.42%. Selain itu, nilai presisi yang diperoleh adalah 1 atau 100%, dengan *recall* sebesar 1 atau 100% dan *f1-score* mencapai 1 atau 100%.

REFERENCES

[1] N. E. Rohaeni and O. Saryono, "Implementasi Kebijakan Program Indonesia Pintar (PIP) Melalui Kartu Indonesia Pintar (KIP) dalam Upaya Pemerataan Pendidikan," *J. Educ. Manag. Adm. Rev.*, vol. 2, no. 1, pp. 193–204, 2018, [Online]. Available: https://jurnal.unigal.ac.id/ijemar/article/view/1824

[2] J. Jumanah and H. Rosita, "EVALUASI PROGRAM INDONESIA PINTAR DALAM UPAYA PEMERATAAN PENDIDIKAN," *\Indonesian J. Soc. Polit. Sci.*, vol. 4, no. 1, p. 54, 2023, [Online]. Available: https://www.researchgate.net/publication/370406946_EVALUASI_PROGRAM_INDONESIA_PINTAR_DALAM_UPAYA_PEMERATAAN_PENDIDIKAN

[3] M. Sari, S. Musdalifah, and E. A. Asfar, "Implementasi Kebijakan Kartu Indonesia Pintar Dalam Upaya Pemerataan Pendidikan di MTsN 1 Watampone," *MAPPESONA J. Manaj. Pendidik. Islam*, vol. 3, no. 1, p. 44, 2021, doi: 10.35673/ajmpi.v8i1.422.

[4] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *J. Inovtek Polbeng - Seri Inform.*, vol. 3, no. 1, p. 51, 2018, [Online]. Available: https://media.neliti.com/media/publications/256208-implementasi-algoritma-naive-bayes-terha-8b9cc418.pdf

[5] A. Nasir, et, "ANALISIS SENTIMEN NETIZEN TWITTER MENGENAI BEASISWA KIP KULIAH DI INDONESIA MENGGUNAKAN VADER," vol. 9, no. 981, pp. 356–363, 2023, [Online]. Available: http://repository.unwira.ac.id/13329/

[6] J. S. B. Ginting, M. Irawan, and M. G. Fadhliphya, "ANALISIS SENTIMEN BERDASARKAN OPINI DARI TWITTER TERHADAP PELAKU PENYALAHGUNAAN KIP K," *Etika Teknol. Inf.*, pp. 1–5, 2024, [Online]. Available: https://www.researchgate.net/publication/381829493_ANALISIS_SENTIMEN_BERDASARKAN_OPINI_DARI_TWITTER_TERHADAP_PELAKU_PENYALAHGUNAAN_KIP_K.

[7] N. Sufni, "Analisis Keberhasilan Program Kartu Indonesia Pintar (KIP) dalam Meningkatkan Akses Pendidikan di Indonesia," *BENEFIT J. Business, Econimics, Financ.*, vol. 2, no. 2, pp. 38–45, 2024, [Online]. Available: https://publikasi.abidan.org/index.php/benefit/article/download/393/301/1444

[8] R. Achmad and T. W. A., "Evaluasi Program Indonesia Pintar Dalam Upaya Pemerataan Pendidikan di Indonesia," *J. Adminitrasi Publik*, p. 2, 2022, [Online]. Available: https://journal.uta45jakarta.ac.id/index.php/admpublik/article/view/6042

[9] K. Yan, D. Arisandi, and Toni, "ANALISIS SENTIMEN KOMENTAR NETIZEN TWITTER TERHADAP KESEHATAN MENTAL MASYARAKAT INDONESIA," *J. Ilmu Komput. dan Sist. Inf.*, vol. 9, no. 2, pp. 5–6, 2021, [Online]. Available: https://journal.untar.ac.id/index.php/jiksi/article/view/17865/9870

[10] H. M. Sukardi, *Metodologi Penelitian Pendidikan: Kompetensi dan praktiknya, Edisi Revisi*, 1st ed. Jakarta: Bumi Aksara, 2021. [Online]. Available: https://www.google.co.id/books/edition/Metodologi_Penelitian_Pendidikan/gJo_EAAAQBAJ?hl=id&gbpv=1&dq=perencanaan+penelitian&pg=PA299&printsec=frontcover

[11] S. Y. Kusumastuti, A. F. Anggraeni, A. Rustam, D. E. Desi, and B. Waseso, *Metodologi Penelitian (Pendekatan Kualitatif dan Kuantitatif)*, 1st ed. Jambi: PT. Sonpedia Publishing Indonesia, 2025. [Online]. Available: https://www.google.co.id/books/edition/Metodologi_Penelitian_Pendekatan_Kualita/0DxHEQAAQBAJ?hl=id&gbpv=1&dq=mengumpulkan+data&pg=PA93&printsec=frontcover

[12] Y. Asri and M. Fajri, *Implementasi Lexicon Vader dan Naive Bayes Pada Aplikasi PLN Mobile*, 1st ed. Sidoarjo: Uwais Inspirasi Indonesia, 2019. [Online]. Available: https://www.google.co.id/books/edition/Implementasi_Lexicon_Vader_dan_Naïve_Ba/CqPtEAAAQBAJ?hl=id&gbpv=1&dq=preprocessing+text+naive+bayes&pg=PA46&printsec=frontcover

[13] Kusrini and E. T. Luthfi, *Algoritma Data Mining*, 1st ed. Yogyakarta: C.V ANDI OFFSET, 2009. [Online]. Available: https://www.google.co.id/books/edition/Algoritma_Data_Mining/-Ojclag73O8C?hl=id&gbpv=1&dq=analisis+data+mining&pg=PA150&printsec=frontcover

[14] E. A. Novia, W. I. Rahayu, and C. Prianto, *Sistem Perbandingan Algoritma K-Means dan Naive Bayes Untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan*, 1st ed. Bandung: Kreatif Industri Nusantara, 2020. [Online]. Available:

https://www.google.co.id/books/edition/SISTEM_PERBANDINGAN_ALGORITMA_K_MEANS_DA/MND9DwAAQBAJ?hl=id&gbpv=1&dq=NAIVE+BAYES&pg=PA37&printsec=frontcover

[15] R. T. Yunardi and N. Z. Dina, *Data Mining dan Machine Learning dengan Orange3 Tutorial dan Aplikasinya*, 1st ed. Surabaya: Airlangga University Press, 2022. [Online]. Available: https://www.google.co.id/books/edition/DATA_MINING_dan_MACHINE_LEARNING_dengan/hplvEAAAQBAJ?hl=id&gbpv=1&dq=naive+bayes+classifier&pg=PA69&printsec=frontcover

[16] Sriani, Suhardi, and L. S. Yardha, "Analisis Sentimen Pengguna Aplikasi Monile JKN Menggunakan Algoritma Naive Bayes Classifier Dan C4.5," *J. Sci. Soc. Res.*, vol. 7, no. 2, pp. 555–563, 2024, [Online]. Available: https://jurnal.goretanpena.com/index.php/JSSR/article/view/1873/1181

[17] C. R. A. Widiawati and A. M. Wahid, *Pemrosesan Bahasa Alami Konsep, Algoritma & Implementasi*, 1st ed. Purwokerto: CV. ZT CORPORA, 2025. [Online]. Available: https://www.google.co.id/books/edition/PEMROSESAN_BAHASA_ALAMI_Konsep_Algoritma/bKlOEQAAQBAJ?hl=id&gbpv=1&dq=teks+preprocessing+naive+bayes&pg=PA157&printsec=frontcover

[18] Solimun, A. A. R. Fernandes, Nurjannah, E. G. Erwinda, R. Hardianti, and L. H. Y. Arini, *Metodologi Penelitian: Variabel Mining Berbasis Big Data Dalam Pemodelan Sistem untuk Mengungkap Research Novelty*, 1st ed. Malang: UB Press, 2023. [Online]. Available: https://www.google.co.id/books/edition/Metodologi_Penelitian/GfXaEAAAQBAJ?hl=id&gbpv=1&dq=pelabelan+data+adalah&pg=PA34&printsec=frontcover

[19] T. C. S. Rani and E. Sahputra, *Penerapan Machine Learning Terhadap Analisis Sentimen Masyarakat Dalam Pemilihan Presiden Indonesia 2024*, 1st ed. Bengkulu: Penerbit NEM, 2024. [Online]. Available: https://www.google.co.id/books/edition/Penerapan_Machine_Learning_terhadap_Anal/KRALEQAAQBAJ?hl=id&gbpv=0