ANALISIS SENTIMEN INDIHOME DAN STARLINK MENGGUNAKAN METODE NAÏVE BAYES BAGI PT. SARIMELATI KENCANA, TBK DALAM PENGAMBILAN KEPUTUSAN

Dhea Gitawijaya[1], Arif Siswandi[2], Irfan Afriantoro[3]
[1]Program Studi Teknik, Informatika, Universitas Pelita Bangsa, Bekasi, Indonesia
e-mail: [1]dheagitawijaya@mhs.pelitabangsa.ac.id, [2]arif.siswandi@pelitabangsa.ac.id, [3]irfanafriantoro@pelitabangsa.ac.id

Abstrak

Perkembangan teknologi layanan internet di Indonesia memengaruhi berbagai sektor industri, termasuk sektor makanan di PT Sarimelati Kencana Tbk. Perusahaan awalnya menggunakan layanan IndiHome, namun menghadapi tantangan seperti tingginya biaya dan ketidakstabilan jaringan yang berdampak pada efisiensi operasional. Sebagai alternatif, layanan Starlink hadir dengan jaringan satelit yang menjangkau wilayah luas dan mendapat respons publik yang lebih positif. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap IndiHome dan Starlink menggunakan metode Naïve Bayes yang dievaluasi melalui K-Fold Cross Validation. Data dikumpulkan dari media sosial X (sebelumnya Twitter) dengan total 14.200 tweet (7100 tweet untuk IndiHome dan 7100 untuk Starlink), yang kemudian diproses melalui tahapan preprocessing seperti cleansing, tokenizing, stopword removal, dan stemming. Hasil pengujian menunjukkan bahwa metode Naïve Bayes memberikan akurasi terbaik pada data Starlink sebesar 83,74%, sedangkan pada IndiHome sebesar 73,31%. Analisis preferensi menunjukkan bahwa sentimen positif terhadap IndiHome sebesar 34%, sedangkan Starlink mencapai 47%.Temuan ini menunjukkan bahwa Starlink lebih unggul dalam persepsi publik dibandingkan IndiHome. Oleh karena itu, hasil penelitian ini dapat menjadi dasar pertimbangan strategis bagi PT Sarimelati Kencana Tbk. dalam memilih penyedia layanan internet yang lebih stabil, efisien, dan mendukung kelancaran operasional perusahaan.

Kata Kunci: Naïve Bayes, Analisis Sentimen, IndiHome, Starlink, Media Sosial X, Pengambilan Keputusan.

Abstract

The development of internet service technology in Indonesia has impacted various industrial sectors, including the food sector at PT Sarimelati Kencana Tbk. The company initially used the IndiHome service but faced challenges such as high costs and unstable network connections, which affected operational efficiency. As an alternative, Starlink offers satellite-based internet with broader coverage and has received more positive public feedback. This study aims to analyze public sentiment toward IndiHome and Starlink using the Naïve Bayes method, evaluated through K-Fold Cross Validation. Data were collected from social media platform X (formerly Twitter), totaling 14,200 tweets (7,100 for IndiHome and 7,100 for Starlink), and processed through preprocessing stages including cleansing, tokenizing, stopword removal, and stemming. The testing results show that the Naïve Bayes method achieved the best accuracy for Starlink data at 83.74%, while for IndiHome it was 73.31%. Sentiment analysis results indicate that 34% of tweets about IndiHome were positive, while Starlink received 47% positive sentiment. These findings suggest that Starlink is perceived more favorably by the public compared to IndiHome. Therefore, the results of this study can serve as a strategic consideration for PT Sarimelati Kencana Tbk. in selecting a more stable and efficient internet service provider to support the company's operational continuity.

Keywords: *Naïve Bayes, Sentiment Analysis, IndiHome, Starlink, Social Media X, Decision Making.*

1. PENDAHULUAN

Perkembangan teknologi informasi di Indonesia dalam beberapa tahun terakhir telah membawa perubahan signifikan, khususnya dalam kebutuhan akses internet yang kini menjadi kebutuhan utama masyarakat. Berdasarkan laporan Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) tahun 2024, pengguna internet di Indonesia mencapai 221,5 juta orang atau 79,5% dari total populasi. Internet digunakan untuk berbagai aktivitas, mulai dari informasi, pembelajaran, hingga interaksi sosial[1].

Penyedia layanan internet di Indonesia sangat beragam, mulai dari penyedia besar seperti IndiHome, MNC Play, Biznet, XL Fiber, First Media, MyRepublic, Orbit Telkomsel, hingga penyedia kecil seperti RT RW net. IndiHome, sebagai salah satu penyedia internet terbesar di Indonesia, menawarkan jaringan fiber optic yang menjangkau berbagai wilayah di Indonesia. Layanan ini menjadi pilihan utama bagi banyak pengguna karena keandalannya dalam mendukung kebutuhan internet rumah tangga maupun bisnis. Namun, semakin tingginya biaya layanan IndiHome telah menyebabkan perusahaan, termasuk PT Sarimelati Kencana Tbk., menghadapi tantangan dalam mengelola efisiensi biaya operasional. Selain itu, menurunnya kestabilan layanan IndiHome menyebabkan sentimen negatif masyarakat terhadap IndiHome dan menghambat produktivitas dan kelancaran operasional bisnis perusahaan, terutama dalam mendukung kegiatan yang sangat bergantung pada konektivitas internet yang turut memengaruhi keputusan strategis perusahaan[2].

Untuk mengatasi tantangan tersebut, PT Sarimelati Kencana Tbk. mulai mempertimbangkan alternatif lain, salah satunya adalah Starlink. Starlink adalah layanan internet berbasis satelit yang dikembangkan oleh SpaceX. Starlink telah menjadi perhatian besar di Indonesia dengan sentimen positif yang dominan di kalangan pengguna karena memiliki biaya layanan yang terjangkau dan dirancang untuk menjangkau wilayah terpencil yang sulit dijangkau oleh jaringan fiber optic, menjadikannya pilihan menarik bagi perusahaan yang membutuhkan akses internet yang cepat dan stabil[3].

Menurut riset We Are Social pada tahun 2022, penggunaan media sosial di Indonesia mencakup 191,4 juta pengguna dari total populasi sebesar 277,7 juta jiwa. Platform media sosial X sendiri memiliki 383 juta tweet per hari, dengan lebih dari 29,5 juta pengguna di Indonesia, menjadikannya platform yang banyak digunakan untuk menyuarakan opini terkait layanan internet seperti IndiHome dan Starlink. Analisis sentimen terhadap opini masyarakat di media sosial menjadi salah satu cara untuk memahami persepsi publik terhadap kedua layanan ini[4].

Analisis sentimen terhadap opini masyarakat di media sosial menjadi penting untuk mendapatkan wawasan mengenai persepsi publik terhadap kedua layanan ini. Metode analisis sentimen Naïve Bayes, dipilih oleh penulis karena kemampuannya dalam mengolah data besar dan kompleks, serta menghasilkan prediksi yang akurat. Dengan menggunakan metode ini, penelitian ini bertujuan untuk membandingkan sentimen masyarakat terhadap IndiHome dan Starlink, serta memberikan alternatif strategis berbasis data bagi PT Sarimelati Kencana Tbk., dalam memilih layanan internet yang mendukung efisiensi operasional dan daya saing perusahaan di pasar yang semakin kompetitif[5].

Penelitian ini dilatarbelakangi oleh kebutuhan untuk memahami persepsi publik terhadap layanan internet yang digunakan oleh perusahaan, khususnya dalam konteks efisiensi dan kestabilan layanan. Oleh karena itu, rumusan masalah dalam penelitian ini mencakup beberapa hal penting. Pertama, bagaimana sentimen yang muncul di media sosial terhadap biaya layanan dan kestabilan IndiHome. Kedua, bagaimana sentimen yang muncul terhadap biaya layanan dan kestabilan Starlink di media sosial. Ketiga, seberapa besar akurasi yang dapat diperoleh dalam analisis sentimen terhadap IndiHome dan Starlink dengan menggunakan metode Naïve Bayes. Keempat, bagaimana data sentimen tersebut dapat memberikan alternatif strategis bagi PT Sarimelati Kencana Tbk. dalam memilih penyedia layanan internet yang lebih efisien dan mendukung kelancaran operasional perusahaan.

2. TINJAUAN PUSTAKA

Penelitian pertama yang dilakukan oleh Alfandi Safira dan Firman Noor Hasan berjudul "Analisis Sentimen Terhadap Penggunaan Paylater Menggunakan Naïve Bayes Classifier dan TextBlob" bertujuan untuk mengetahui pandangan masyarakat terhadap layanan paylater yang kian populer seiring meningkatnya tren belanja online, namun juga berpotensi memicu perilaku negatif seperti impulse buying. Penelitian ini menggunakan metode analisis sentimen Naïve Bayes Classifier dan TextBlob dengan bantuan bahasa pemrograman Python, berdasarkan 405 data tweet dari Twitter yang diklasifikasikan ke dalam tiga kategori sentimen: positif, negatif, dan netral. Hasil analisis menunjukkan bahwa sentimen negatif mendominasi baik pada metode Naïve Bayes (70,62%) maupun TextBlob (55,8%), dan dari hasil pengujian akurasi menggunakan confusion matrix, metode Naïve Bayes Classifier terbukti lebih andal dengan akurasi 91% dibandingkan TextBlob yang hanya mencapai 61%. Dengan demikian, penelitian ini menyimpulkan bahwa sebagian besar masyarakat memiliki

persepsi negatif terhadap penggunaan paylater dan bahwa Naïve Bayes lebih efektif dalam klasifikasi sentimen untuk topik ini[6].

Penelitian kedua yang dilakukan oleh Aji Susanto dan Iskandar Agung Dzulkarnain berjudul "Analisis Sentimen Data Twitter Topik Ekonomi dan Industri dengan Metode Random Forest dan Naïve Bayes" bertujuan untuk menganalisis dan membandingkan performa algoritma Random Forest dan Naïve Bayes dalam mengklasifikasikan sentimen pada data Twitter yang membahas isu ekonomi dan industri. Twitter dipilih karena dianggap mampu mencerminkan pandangan publik yang relevan bagi pengambilan keputusan strategis. Sebanyak 11.833 data dikumpulkan dan dibagi menjadi data training (70%) dan data testing (30%) untuk kemudian dianalisis menggunakan kedua algoritma tersebut. Hasil klasifikasi menunjukkan distribusi sentimen sebesar 28,52% positif, 31,44% negatif, dan 40,04% netral, dengan Naïve Bayes menghasilkan akurasi tertinggi sebesar 71,89%. Penelitian ini menyimpulkan bahwa meskipun kedua metode memiliki keunggulan masing-masing, Naïve Bayes lebih efektif dalam klasifikasi sentimen pada konteks ekonomi dan industri[7].

Penelitian yang dilakukan oleh Aprilia Wandani, Fauziah, dan Andrianingsih berjudul *"Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Naïve Bayes, dan Random Forest"* bertujuan untuk menganalisis opini pengguna Twitter terhadap event Flash Sale yang diselenggarakan oleh platform e-commerce, sebuah strategi penjualan yang menawarkan harga rendah dalam waktu terbatas untuk mendorong minat beli. Twitter dipilih sebagai sumber data karena dianggap sebagai platform yang representatif untuk menangkap opini publik. Penelitian ini menggunakan tiga algoritma klasifikasi, yaitu Naïve Bayes, K-Nearest Neighbour (K-NN), dan Random Forest, dengan dua kata kunci utama: "flash sale" dan "flash sale shopee." Hasil analisis menunjukkan bahwa algoritma Naïve Bayes consistently memberikan akurasi tertinggi, yaitu 83,53% untuk kata kunci "flash sale" dan 81,48% untuk "flash sale shopee," dibandingkan K-NN dan Random Forest. Oleh karena itu, penelitian ini menyimpulkan bahwa Naïve Bayes merupakan metode yang paling direkomendasikan untuk analisis sentimen pada topik Flash Sale di Twitter karena memberikan hasil akurasi yang lebih tinggi dan stabil[8].

3. METODOLOGI PENELITIAN

3.1 Analisis Sentimen

Analisis sentimen adalah proses untuk memahami, mengekstrak, dan mengolah data teks secara otomatis guna memperoleh informasi mengenai sentimen yang terkandung dalam suatu opini. Proses ini bertujuan untuk menganalisis kecenderungan opini terhadap suatu objek yang kemudian dikategorikan ke dalam sentimen positif, negatif, atau netral. Dalam konteks penelitian ini, perhitungan Preference Value digunakan untuk mengidentifikasi tingkat sentimen positif pengguna Twitter terhadap peluncuran dan penggunaan IndiHome serta Starlink. Preference Value merupakan ukuran yang digunakan untuk menilai tingkat respons positif dari pengguna Twitter. Nilai Preference Value yang lebih tinggi menunjukkan tingkat respons positif yang lebih besar dari pengguna terhadap kedua layanan tersebut[9]. Rumus untuk menghitung Preference Value ditunjukan pada persamaan (1)

$$Preference\ value = \frac{sp}{sp+sn} * akurasi \quad (1)$$

3.2 Metode Penelitian dan Pengumpulan Data

Penelitian ini menggunakan data berjenis kualitatif, berupa teks yang diperoleh melalui proses pengumpulan data (crawling) menggunakan API dari platform media sosial X (sebelumnya dikenal sebagai Twitter). Crawling adalah metode yang digunakan untuk mengumpulkan informasi yang ada di internet. Ini beroperasi secara otomatis dan mengumpulkan data berdasarkan kata kunci yang diberikan pengguna(ref crawling). Penelitian ini memanfaatkan tiga jenis sumber data utama yang digunakan dalam proses analisis sentimen. Sumber data tersebut meliputi data dari media sosial X (sebelumnya Twitter), daftar stopword, dan kumpulan kata dasar. adalah platform jejaring sosial dan mikroblog yang memungkinkan pengguna untuk mengirim dan membaca pesan pendek hingga 280 karakter, yang dikenal sebagai tweet. Platform ini berfungsi sebagai alat untuk menghubungkan pengguna satu sama lain serta berbagi informasi secara cepat dan luas[10].

3.3 Metode Pengolahan Data

Data mining adalah proses pengolahan data dalam skala besar yang bertujuan untuk mengekstrak pola atau informasi berharga yang tersembunyi di dalamnya[11]. Penelitian ini mengikuti lima tahapan utama dalam proses Knowledge Discovery in Databases (KDD), dimulai dari tahap seleksi, di mana data operasional yang relevan dipilih dan dipisahkan dari basis data utama untuk dianalisis lebih lanjut. Selanjutnya, pada tahap pra-pengolahan (preprocessing), data dibersihkan dari elemen yang tidak konsisten atau nilai yang hilang guna memastikan integritas dan kesiapan dataset. Setelah itu, dilakukan transformasi, yaitu mengubah data ke dalam format yang sesuai, seperti mengonversi data nominal menjadi numerik agar dapat diproses secara algoritmik. Tahap inti, yaitu data mining, dilakukan untuk mengeksplorasi dan menemukan pola dalam data menggunakan metode tertentu, seperti Naïve Bayes dalam analisis sentimen. Terakhir, pada tahap evaluasi, hasil yang diperoleh dianalisis dan diinterpretasikan untuk memastikan kesesuaiannya dengan tujuan analisis serta memberikan wawasan yang dapat dipahami oleh pihak terkait[12].

3.4 Metode Analisis

Pada peneltiian ini Naïve Bayes Classifier dipilih sebagai metode yang digunakan untuk mengolah data setelah proses crawling, Naïve Bayes juga merupakan sebuah algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat. Naïve Bayes classifier mengasumsikan bahwa ada atau tidaknya suatu fitur tertentu pada sebuah kelas tidak akan mempengaruhi keberadaan fitur yang lain[13]. Konsep dasar yang digunakan pada Naïve Bayes classifier adalah Teorema Bayes yang dinyatakan pertama kali oleh Thomas Bayes[14] Adapun rumus teorema naive bayes ditunjukan pada persamaan (2)

$$P(c) = \frac{Nc}{Ndoc} \tag{2}$$

Keterangan :
Nc = Jumlah data training dalam sebuah dokumen dengan kelas c.
Ndoc = Jumlah total dokumen dalam data training.

3.5 K-Fold Cross Validation

K-Fold Cross Validation merupakan salah satu metode yang digunakan untuk membagi dataset menjadi data pelatihan (training data) dan data pengujian (test data). Teknik ini sering digunakan oleh para peneliti karena mampu mengurangi bias yang mungkin terjadi selama proses pengambilan sample[12].

3.6 Rapid Miner

RapidMiner adalah perangkat lunak yang digunakan untuk mempelajari data mining dalam proses analisis data, pemrosesan data, serta pengembangan model prediktif. Perangkat lunak ini bersifat open-source dan banyak digunakan karena fleksibilitas dan kemampuannya dalam menangani berbagai tugas terkait analisis data[15].

4. Hasil dan Pembahasan

4.1 Pengumpulan Data

Data dikumpulkan melalui proses crawling menggunakan alat Twitter-Harvest. Data yang diambil berupa tweet berbahasa Indonesia yang diperoleh dari pencarian menggunakan kata kunci IndiHome dan Starlink serta dari balasan thread yang dibuat oleh konten kreator yang telah menggunakan layanan IndiHome dan Starlink.

Data hasil crawling dapat dilihat pada tabel 1 dibawah ini :

Table 1 Data dari hasil proses crawling

Data Crawling Indihome	Data Crawling Starlink
b'RT @gadgetfreak_ID: SEKALI doang nyambuuuung ke STARLINK!!! semua alat2 GUA langsung LANCARRRRR JAYAAAA gaaaada hambatan samaa sekaliii #internetcepat #starlink'	b'INDIHOME = koneksi CEPATTTTTTTT hidup berasa lengkap cuy #indihome #wificepat'

b'@milenialtech Upload konten panjang SEKARANG tuh kyk upload foto doangggg makasih bgt STARLINK!!! #starlink #internetcepat #gasspoll'	b'udah restart modem 5x, tetep aja GABISAAA buka web kayaknya INDIHOME emg sering gini'
b'gila sih ini STARLINK... semua device gua konek bareng nonton YouTube HD lancar tanpa buffering #puas #starlink #ngebut'	b'indihome bikin semua device di rumah konek bareng streaming lancar tanpa ngelag #puas #indihome'
b'starlink ngajarin gua sabarrrrr sinyal ilang nyambung lagi ilang lagi nyambung lagiiii '	b'INDIHOME sinyalnya PALING waras dari semua provider yg pernah gua coba#internet'
b'pake starlink beneran bikin gua makin deket sm kata resign tiap zoom meeting auto reconnect wkwk'	b'kerjaan kantor kelar terus karena INDIHOME stabil bgt #thankyouindihome'

Tabel tersebut menampilkan hasil data crawling dari media sosial terkait Indihome dan Starlink. Menariknya, meskipun pencarian berdasarkan kata kunci "Indihome", banyak komentar justru membahas keunggulan Starlink, seperti koneksi cepat, stabil, tanpa buffering, dan cocok untuk aktivitas berat seperti upload dan Zoom. Namun, ada satu keluhan terkait sinyal yang tidak stabil. Sebaliknya, pada pencarian dengan kata kunci "Starlink", banyak komentar membicarakan Indihome, dengan sentimen campuran. Beberapa memuji koneksi cepat dan stabil, namun ada juga yang mengeluhkan masalah teknis seperti modem sering di-restart dan akses web yang bermasalah. Hal ini mencerminkan perbandingan aktif antara kedua layanan di kalangan pengguna.

4.2 Preprocessing Data

Preprocessing adalah tahap awal dalam pengolahan data untuk memastikan data yang digunakan bersih dan siap untuk dianalisis lebih lanjut. Proses ini dilakukan secara manual oleh peneliti. Tahapan-tahapan dalam proses preprocessing meliputi seleksi data, penghapusan duplikat (remove duplicate), pembersihan data, normalisasi data, penentuan kelas kategori sentimen, pembagian data set, dan terakhir pemrosesan data. Pemrosesan data itu sendiri meliputi transform cases, tokenize, filter stopwords, filter token bylenght, stemming dan generate n-gram. Hasil preprocessing data dapat dilihat pada tabel 2 dibawah ini.

Table 2 Hasil Preprocessing Data

Data Mining Indihome	Sentimen	Data Mining Starlink	Sentimen
sekali doang sambung starlink alat langsung lancar jaya gak hambat	Positif	indihome koneksi cepat hidup rasa lengkap	Positif
upload konten panjang upload foto doang makasih starlink	Positif	restart modem bisa buka kaya indihome sering gini	Negatif
gila starlink device konek bareng nonton youtube lancar	Positif	indihome buat device rumah konek bareng stream lancar	Positif
starlink ajar sabar sinyal hilang sambung hilang sambung lagi	Negatif	indihome sinyal paling waras provider pernah coba	Positif
starlink benar buat makin dekat resign zoom meeting auto reconnect wkwk	Negatif	kerja kantor kelar indihome stabil	Positif

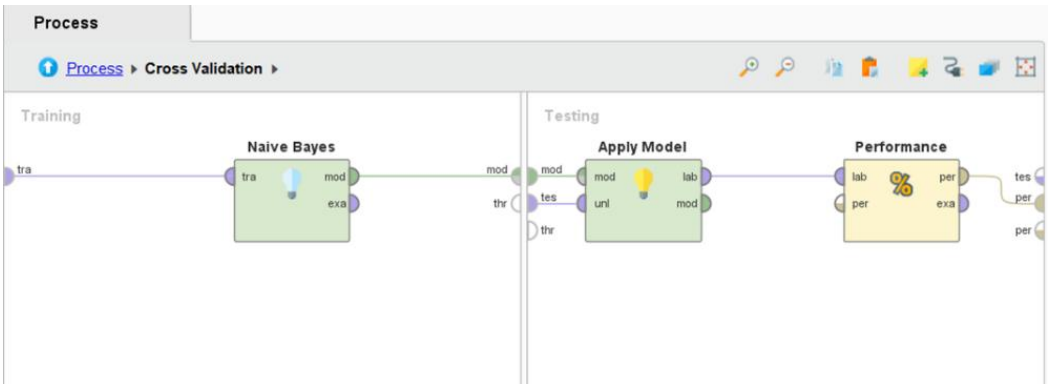
Tabel di atas menunjukkan hasil akhir dari proses data mining terhadap opini pengguna terkait layanan internet Indihome dan Starlink, yang telah dikelompokkan berdasarkan sentimen positif dan negatif. Setiap kalimat pada tabel merupakan hasil dari proses preprocessing yang dilakukan terhadap data mentah hasil crawling media sosial. Pada kolom Data Mining Indihome dan Data Mining Starlink ditampilkan pernyataan-pernyataan pengguna yang telah dibersihkan dan dinormalisasi, sementara kolom Sentimen menunjukkan hasil klasifikasi dari masing-masing pernyataan tersebut. Sebagai contoh, kalimat "sekali doang sambung starlink alat langsung lancar jaya gak hambat" diklasifikasikan sebagai sentimen positif karena menunjukkan kepuasan pengguna terhadap layanan Starlink, sementara kalimat "starlink ajar sabar sinyal hilang sambung hilang sambung lagi" dikategorikan sebagai sentimen negatif karena berisi keluhan.

Data dalam tabel ini telah melalui tahapan preprocessing yang terdiri dari: seleksi data, penghapusan duplikat (remove duplicate), pembersihan data, normalisasi data, penentuan kelas kategori

sentimen, pembagian dataset, serta pemrosesan data. Tahap pemrosesan data itu sendiri mencakup transform cases (mengubah semua huruf menjadi huruf kecil), tokenize (memecah kalimat menjadi kata-kata), filter stopwords (menghapus kata umum yang tidak bermakna signifikan), filter token by length (menghapus kata yang terlalu pendek), stemming (mengubah kata ke bentuk dasar), dan generate n-gram (menggabungkan kata dalam pola tertentu untuk meningkatkan akurasi analisis). Dengan rangkaian proses tersebut, data menjadi lebih bersih, terstruktur, dan siap digunakan untuk analisis sentimen maupun pemodelan machine learning lanjutan.

4.3 Penerapan Algoritma

Dalam penelitian ini, Naïve Bayes digunakan untuk menganalisis sentimen. Penerapan Naïve Bayes dilakukan menggunakan operator cross-validation dalam aplikasi RapidMiner. Pada tahap pelatihan (training), Naïve Bayes diterapkan dengan konfigurasi parameter tertentu. Sementara itu, pada tahap pengujian (testing), digunakan operator apply model untuk menerapkan model yang telah dilatih, serta operator performance untuk menghitung akurasi dan performa model dalam klasifikasi data. Berikut dapat dilihat untuk proses penerapan algoritma pada gambar 1.



Gambar 1 Penerapan Algoritma

Gambar diatas menunjukkan proses evaluasi model klasifikasi menggunakan algoritma Naïve Bayes dalam platform RapidMiner dengan metode Cross Validation. Proses ini terdiri dari dua tahap utama, yaitu tahap pelatihan (training) dan tahap pengujian (testing). Pada tahap pelatihan, data yang telah dibagi oleh skema Cross Validation dimasukkan ke dalam operator Naïve Bayes untuk membangun model. Model yang dihasilkan selanjutnya digunakan pada tahap pengujian, di mana operator Apply Model menerapkan model tersebut pada data uji untuk menghasilkan prediksi.

Hasil prediksi ini kemudian dibandingkan dengan label asli menggunakan operator Performance, yang akan menghitung dan menampilkan metrik evaluasi seperti akurasi, presisi, dan recall. Proses ini diulang secara otomatis sesuai jumlah lipatan (fold) pada Cross Validation, sehingga memungkinkan evaluasi performa model yang lebih objektif dan menghindari overfitting. Diagram ini merepresentasikan alur kerja umum dalam membangun dan menguji model machine learning secara valid dan sistematis.

4.4 Perhitungan Naïve Bayes

Perhitungan Naïve Bayes dalam penelitian ini dilakukan secara manual berdasarkan data tweet yang telah melalui tahapan preprocessing. Contoh data yang digunakan berasal dari tweet pengguna terhadap layanan IndiHome dan Starlink, masing-masing diklasifikasikan ke dalam kelas positif dan negatif. Tahap awal adalah perhitungan prior probability, yaitu probabilitas awal dari masing-masing kelas. Berdasarkan data sampel, terdapat dua tweet positif dan dua tweet negatif, sehingga dapat dilihat pada perhitungan (3):

$$P(c) = \frac{Nc}{Ndoc}$$

$$P(Positif) = \frac{2}{4} = 0{,}5$$

$$P(Negatif) = \frac{2}{4} = 0{,}5 \quad (3)$$

Keterangan :

Nc = Jumlah data training dalam sebuah dokumen dengan kelas c.

Ndoc = Jumlah total dokumen dalam data training.

Selanjutnya dilakukan likelihood calculation untuk menghitung probabilitas kemunculan kata dalam setiap kelas. Kata-kata dari tweet diuji terhadap data training menggunakan rumus (4):

$$P(wi|c) = \frac{count(wi,c)+1}{\sum \text{count(w,c)}+|V|} \quad (4)$$

Keterangan :

$P(w_i|c)$ = Probabilitas kata w_i muncul dalam kelas c

$Count(w_i,c)$ = Jumlah kemunculan kata w_i pada seluruh dokumen dalam kelas c

$\sum count(w,c)$ = Jumlah total seluruh kata dalam kelas c

$|V|$ = Jumlah data unik dalam keseluruhan korpus data

+1 pada pembilang = Digunakan untuk mencegah probabilitas nol jika sebuah kata tidak muncul dalam kelas tersebut.

Sebagai contoh, untuk tweet uji "Dulu sampai sekarang mantap" dari IndiHome, semua kata tersebut ditemukan di kelas positif dengan masing-masing frekuensi 1, dan tidak ditemukan di kelas negatif. Maka perhitungan tersebut dapat dilihat pada rumus (5):

$$P(wi|c) = \frac{count(wi,c)+1}{\sum \text{count(w,c)}+|V|} \quad (5)$$

$$P(wi|Positif) = \frac{1+1}{32+4} x4 = \frac{2}{36} x4 = 0.0000095$$

$$P(wi|Negatif) = \frac{0+1}{32+4} x4 = \frac{1}{36} x4 = 0.00000059$$

Probabilitas akhir dihitung dengan mengalikan prior dan likelihood, dengan rumus perhitungan (6) dibawah ini :

$$P(X|c) = P(c) x\, P(wi|c) \quad (6)$$

$$P(X|Positif) = 0{,}5\ x\ 0.0000095 = 0.00000476$$

$$P(w|Negatif) = 0{,}5\ x\ 0.00000059 = 0.00000029$$

Keterangan :

P(X|c) = Probabilitas (posterior) bahwa dokumen atau tweet X termasuk dalam kelas c

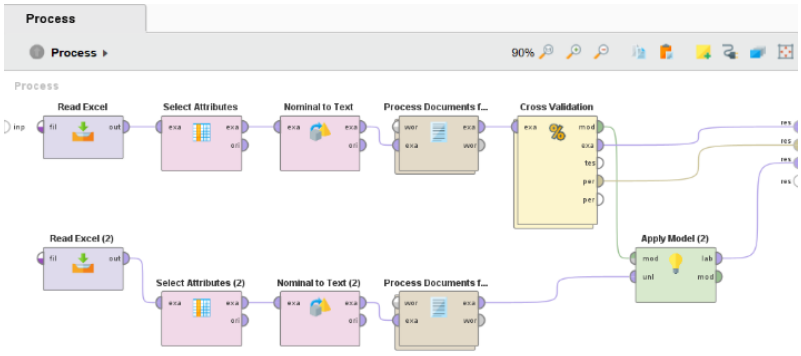
P(c) = Probabilitas kelas awal atau prior dari kelas c

$P(w_i|c)$ = Probabilitas kemunculan kata w_i dalam kelas c (likelihood)

Karena P(X|Positif) lebih besar dari P(X|Negatif), maka tweet diklasifikasikan sebagai sentimen positif. Proses serupa dilakukan untuk tweet uji Starlink. Hasil perhitungan menunjukkan bahwa tweet “Terima kasih Elon Musk dari warga pedalaman” diklasifikasikan sebagai positif, karena kata-kata seperti mudah, dipakai, internetnya, cepat memiliki nilai likelihood tinggi di kelas positif Starlink. Hasil ini membuktikan bahwa algoritma Naïve Bayes mampu mengklasifikasikan sentimen secara akurat berdasarkan probabilitas kemunculan kata dari masing-masing kelas.

4.5 Pengujian

Proses Pengujuan dilakukan dengan mengikuti alur yang telah ditentukan yaitu dengan metode klasifikasi Naïve Bayes. Dalam proses ini peneliti menggunakan RapidMiner untuk melakukan pengujian dengan merangkai beberapa proses yaitu read excel, Select Attribut, Nominal to text, Process Document, Cross Validation dan Apply Model. Proses ini juga akan menggunakan K-Folds validation. Dimana nantinya akan dilakukan perubahan nilai fols pada operator Cross Validation. Proses tersebut dapat dilihat pada gambar 2 dibawah ini :



Gambar 2 Proses Pengujian

Gambar di atas dapat dijelaskan sebagai berikut bahwa Operator read excel nantinya akan digunakan untuk membaca data training dan read excel (2) nantinya akan digunakan untuk membaca data testing, select attribute akan digunakan untuk memilih subset attribute, nominal to text akan digunakan untuk mengubah jenis atribut yang dipilih untuk teks, process document from data akan digunakan untuk proses preprocessing, cross validation akan digunakan untuk memberikan pelatihan, penerapan algoritma dan perhitungan performa data testing dan apply model digunakan untuk menerapkan klasifikasi.

4.6 Hasil Pengujian

Secara keseluruhan, hasil pengujian menunjukkan bahwa algoritma Naïve Bayes lebih efektif dalam menganalisis sentimen terhadap layanan Starlink dibandingkan IndiHome, baik dari sisi akurasi maupun keseimbangan antara precision dan recall. Hal ini memperkuat temuan bahwa karakteristik teks sangat memengaruhi performa model. Bahasa yang lebih langsung dan tidak ambigu memberikan hasil klasifikasi yang lebih baik. Oleh karena itu, dalam kasus layanan digital yang sering mendapat opini sarkastik seperti IndiHome, penggunaan metode tambahan seperti sentiment scoring atau pengayaan data (data augmentation) dapat dipertimbangkan untuk meningkatkan hasil klasifikasi. Dari hasil pengujian yang telah dilakukan menggunakan algoritma Naïve Bayes dan dilakukanya pengujian menggunakan teknik Cross Validation dengan nilai K-Folds 2 sampai dengan 10 penulis mendapatkan hasil pengujian seperti pada tabel berikut :

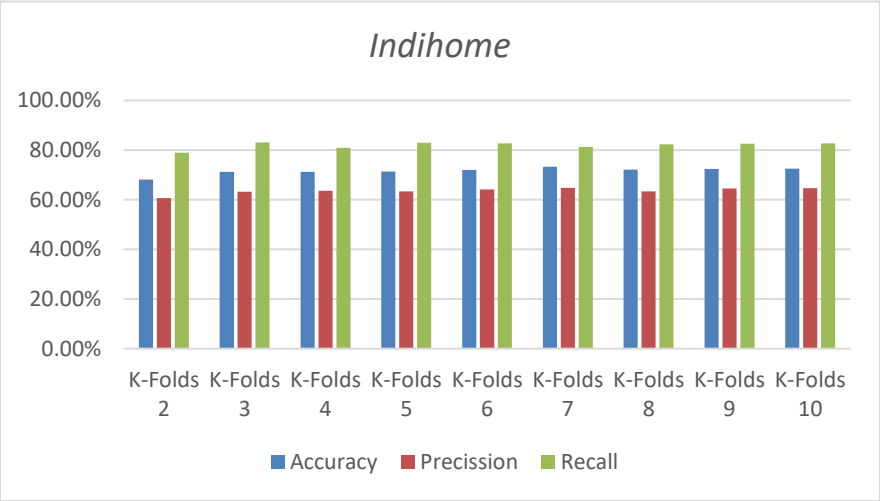
Table 3 Akurasi Hasil Pengujian

Rekap Data *Indihome*			
Dataset	***Accuracy***	***Precission***	***Recall***
K-Folds 2	68,19%	60,64%	79,00%
K-Folds 3	71,25%	63,20%	83,05%
K-Folds 4	71,24%	63,62%	80,91%
K-Folds 5	71,41%	63,40%	82,91%
K-Folds 6	71,99%	64,08%	82,73%
K-Folds 7	73,31%	64,77%	81,32%
K-Folds 8	72,17%	63,37%	82,36%
K-Folds 9	72,39%	64,58%	82,55%
K-Folds 10	72,49%	64,64%	82,77%

Dari tabel diatas maka dapat disimpulkan bahwa nilai Accuracy , Precission dan recall tertinggi terdapat pada pengujian K-Folds (Indihome) dengan nilai 7 yang mana mendapatkan hasil paling tinggi dengan nilai Accuracy 73.31% Precission 64.77% dan Recall sebanyak 81.32%.

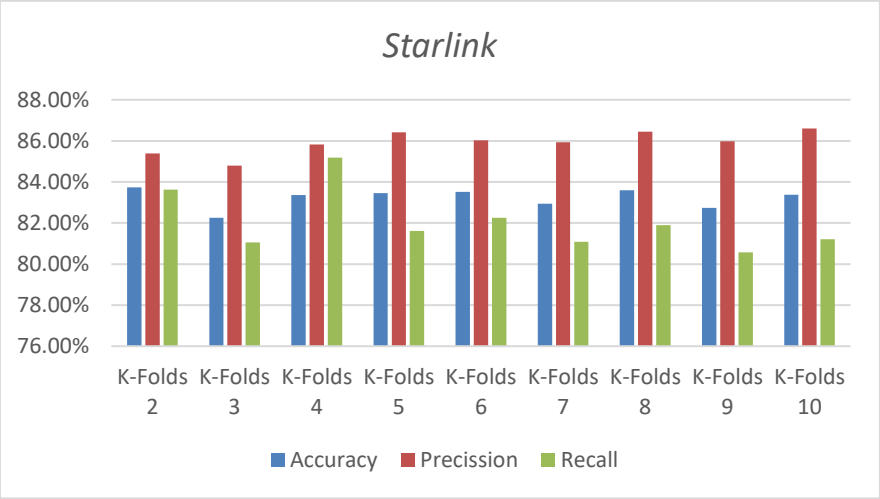
Berikut ini ditampilkan dua grafik yang menggambarkan hasil evaluasi performa model klasifikasi sentimen terhadap dua penyedia layanan internet, yaitu Indihome dan Starlink. Grafik-grafik tersebut merupakan hasil dari pengujian model menggunakan metode K-Fold Cross Validation dengan variasi jumlah lipatan mulai dari 2 hingga 10. Masing-masing grafik menampilkan tiga metrik evaluasi utama, yaitu Accuracy (akurasi), Precision (presisi), dan Recall, yang digunakan untuk menilai sejauh mana model mampu mengklasifikasikan data dengan benar. Hasil ini penting untuk memberikan

gambaran kinerja model secara menyeluruh, terutama dalam konteks perbandingan analisis sentimen terhadap kedua layanan. Grafik tersebut dapat dilihat pada gambar 3 dan 4 dibawah ini



Gambar 3 Diagram Akurasi (Indihome)

Pada grafik pertama digambar 3 yang menampilkan hasil untuk Indihome, terlihat bahwa nilai akurasi berada di kisaran 68% hingga 72% di seluruh variasi K-Folds. Nilai precision sedikit lebih rendah, yaitu sekitar 61% hingga 65%, sedangkan nilai recall justru lebih tinggi dan stabil, mendekati 80%. Hal ini menunjukkan bahwa model cukup baik dalam mengenali seluruh data yang termasuk dalam kelas positif (recall tinggi), meskipun ada kecenderungan untuk memprediksi positif secara kurang tepat (precision lebih rendah).



Gambar 4 Diagram Akurasi (Starlink)

Sementara itu, grafik kedua menunjukkan hasil untuk Starlink, yang memperlihatkan performa yang berbeda. Precision mendominasi dengan nilai tinggi dan stabil, berada di kisaran 84% hingga 87%, menunjukkan bahwa prediksi positif yang dilakukan model hampir seluruhnya benar. Accuracy juga cukup baik, berkisar antara 82% hingga 84%. Namun, recall menunjukkan variasi yang lebih besar dan cenderung lebih rendah dibandingkan precision, berada antara 80% hingga 85%. Ini mengindikasikan bahwa model lebih berhati-hati dalam mengidentifikasi kelas positif, namun sangat akurat saat melakukannya.

Dari kedua grafik tersebut, dapat disimpulkan bahwa model klasifikasi untuk Starlink menunjukkan performa yang lebih kuat dalam precision, sedangkan model untuk Indihome lebih menonjol dalam recall. Perbedaan ini mengindikasikan karakteristik distribusi data sentimen yang berbeda antara kedua layanan, serta kekuatan dan kelemahan model dalam menangani masing-masing data.

4.7 Preferences Value

Preference Value digunakan untuk mengetahui respon positif dari Indihome dan Starlink yang telah diambil dan di proses sebelumnya, jumlah data positif pada data training dan testing Indihome berjumlah 3350 dan jumlah data negatif pada data training dan testing berjumlah 3750. Total dari data tersebut berjumlah 7100 data. Jumlah data positif pada data training dan testing Starlink berjumlah 4000 dan jumlah data negatif pada data training dan testing berjumlah 3100. Total dari data tersebut berjumlah 7100 data. Dan berikut adalah perhitungan preference value (7) dari data tersebut dikalikan dengan nilai akurasi terbaik :

$$Preferences\ Value\ (Indihome) = \frac{3350}{3350+3750} x\ 73{,}31\% = 0{,}3458 \quad (7)$$

Dari hasil perhitungan di atas, dapat disimpulkan bhwa respon positif masyarakat terhadap layanan dapat dilihat dari perhitungan tersebut menunjukan *Indihome* di Indonesia adalah sebesar 34%, lalu perhitungan preference value untuk starlink dapat dilihat pada perhitungan (8) :

$$Preferences\ Value\ (Starlink) = \frac{4000}{4000+3100} x\ 83{,}74\% = 0{,}4717 \quad (8)$$

Dari hasil perhitungan di atas, dapat disimpulkan bahwa respon positif masyarakat terhadap layanan Starlink di Indonesia adalah sebesar 47%

Table 4 Hasil Perhitungan Preference Value

Hasil Perhitungan *Preference Value*				
Layanan	**Sentimen Positif (%)**	**Sentimen Negatif (%)**	**Akurasi Tinggi**	***Preference Value***
Indihome	34%	66%	73,31%	24,93%
Starlink	47%	53%	83,74%	39,39%

Nilai Preference Value menunjukkan bahwa Starlink memiliki tingkat preferensi publik yang lebih tinggi dibandingkan IndiHome, baik dari sisi proporsi sentimen positif maupun akurasi model klasifikasi. Nilai 39,39% menunjukkan bahwa dari keseluruhan data dan performa model, respon positif terhadap Starlink lebih kuat dan lebih dapat dipercaya oleh sistem klasifikasi dibandingkan IndiHome yang hanya memperoleh 24,93%. Dengan demikian, dapat disimpulkan bahwa dari perspektif publik berdasarkan opini di media sosial, Starlink dianggap lebih unggul dalam:

a. Stabilitas jaringan
b. Kemudahan akses di daerah terpencil
c. Tingkat kepuasan awal pelanggan

Sementara itu, IndiHome masih menghadapi persepsi negatif publik yang tinggi, terutama terkait kestabilan jaringan dan biaya layanan. Hasil ini memberikan dasar kuantitatif bagi PT Sarimelati Kencana, Tbk untuk mempertimbangkan Starlink sebagai penyedia layanan internet alternatif.

5. KESIMPULAN

Berdasarkan hasil penelitian dan analisis data yang telah dilakukan, dapat disimpulkan bahwa sentimen terhadap layanan IndiHome di media sosial X cenderung lebih negatif. Hal ini terutama disebabkan oleh keluhan terhadap biaya layanan yang tinggi serta kestabilan jaringan yang tidak konsisten. Proporsi sentimen negatif lebih mendominasi dibandingkan sentimen positif, dengan persentase sentimen positif hanya sebesar 34%. Sebaliknya, sentimen terhadap layanan Starlink secara umum menunjukkan persepsi yang lebih positif. Hal ini didorong oleh keunggulan akses internet yang luas serta kestabilan layanan, khususnya di wilayah terpencil. Starlink memperoleh tingkat sentimen positif sebesar 47%, yang lebih tinggi dibandingkan IndiHome, meskipun masih terdapat beberapa keluhan dari pengguna terkait harga langganan dan performa saat kondisi cuaca buruk.

Selain analisis sentimen, penelitian ini juga mengevaluasi kinerja metode klasifikasi Naïve Bayes dalam menganalisis data. Hasilnya menunjukkan bahwa metode ini mampu memberikan performa yang baik dalam klasifikasi sentimen. Pada data IndiHome, akurasi tertinggi yang dicapai adalah 73,31%, sementara pada data Starlink, akurasi tertinggi mencapai 83,74%. Hal ini menunjukkan bahwa Naïve Bayes merupakan metode yang efektif untuk mengolah dan menginterpretasi data sentimen dari media sosial.

Secara keseluruhan, berdasarkan hasil analisis sentimen dan perbandingan akurasi, dapat disimpulkan bahwa data sentimen dari media sosial berpotensi besar untuk dijadikan dasar dalam pengambilan keputusan strategis. Dalam konteks PT Sarimelati Kencana Tbk., informasi ini dapat menjadi bahan pertimbangan penting dalam memilih penyedia layanan internet yang lebih stabil, efisien, dan mendukung kelancaran operasional perusahaan. Berdasarkan hasil penelitian, Starlink menjadi alternatif yang lebih menjanjikan dibandingkan IndiHome.

6. SARAN

Berdasarkan hasil penelitian yang telah diperoleh, terdapat beberapa saran yang dapat dijadikan acuan untuk penelitian selanjutnya maupun implementasi praktis. Bagi PT Sarimelati Kencana Tbk., disarankan untuk mempertimbangkan penggunaan layanan internet Starlink sebagai solusi alternatif, terutama untuk cabang-cabang yang berada di wilayah dengan keterbatasan infrastruktur jaringan. Rekomendasi ini didasarkan pada tingginya sentimen positif dari publik serta hasil analisis sentimen yang menunjukkan bahwa Starlink dinilai lebih stabil dan efisien oleh pengguna media sosial.

Bagi peneliti selanjutnya, penelitian ini masih terbatas pada dua kategori sentimen, yaitu positif dan negatif, serta hanya menggunakan algoritma klasifikasi Naïve Bayes. Oleh karena itu, diharapkan penelitian di masa mendatang dapat mengembangkan pendekatan dengan menambahkan kategori netral atau menggunakan algoritma lain seperti Support Vector Machine (SVM), Random Forest, atau Deep Learning guna memperluas cakupan analisis dan meningkatkan akurasi klasifikasi.

Selanjutnya, bagi akademisi dan praktisi, hasil dari penelitian ini membuktikan bahwa analisis sentimen berbasis media sosial memiliki potensi besar sebagai alat bantu dalam pengambilan keputusan bisnis. Oleh karena itu, pendekatan ini sebaiknya terus dikembangkan dan diintegrasikan ke dalam kerangka business intelligence serta digunakan sebagai alat evaluasi reputasi merek, khususnya di sektor yang sangat bergantung pada kualitas layanan publik.

Terkait dengan data dan sistem analitik, perlu diperhatikan bahwa media sosial merupakan sumber data yang sangat dinamis dan rentan terhadap bias waktu, seperti munculnya tren sesaat. Untuk itu, disarankan agar ke depannya dibangun sistem pemantauan sentimen secara berkala atau real-time monitoring, guna menyediakan wawasan yang lebih akurat, relevan, dan mutakhir sebagai dasar dalam proses pengambilan keputusan perusahaan.

DAFTAR PUSTAKA

[1] Prasetiyo SM, Gustiawan R, Faarhat, Albani FR. Analisis Pertumbuhan Pengguna Internet Di Indonesia. J Bul Ilm Ilmu Komput Dan Multimed 2024;2:65–71.
[2] PADA LAYANAN INDIHOME PT TELKOM 2023.
[3] Dinamika H, Bisnis P. Strategi sinergi revolusioner starlink dan isp indonesia 2023.
[4] Braniwati M, Rahmawan B, Susilo CB, Reynaldo K, Sosial M, Sekolah P, et al. PERANCANGAN MOTION GRAPHIC TENTANG PENGARUH MEDIA SOSIAL BAGI PELAJAR n.d.:63–77.
[5] A MA, Alamsyah M, Arif MF. Analisis Sentimen Twitter Tentang Pinjaman Online di Indonesia Menggunakan Metode Random Forest n.d.
[6] Safira A, Hasan FN, Timur KJ, Classifier NB. PAYLATER MENGGUNAKAN METODE NAIVE BAYES 1978;5:59–70.
[7] Susanto A, Agung Dzulkarnain I, Studi Statistika P, Matematika Dan Ilmu Pengetahuan Alam F. Analisis Sentimen Data Twitter Topik Ekonomi Dan Industri Dengan Metode Naive Bayes Dan Random Forest. J Ilm Wahana Pendidikan, Oktober 2023;2023:59–65.
[8] Wandani A. Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes. J Sains Komput Inform (J-SAKTI 2021;5:651–65.
[9] Mailoa FF. Analisis sentimen data twitter menggunakan metode text mining tentang masalah obesitas di indonesia. J Inf Syst Public Heal 2021;6:44. https://doi.org/10.22146/jisph.44455.
[10] Qalbina GS, Asmara R, Padang UN. A L - D Y n.d.;3:816–24.
[11] Asiva Noor Rachmayani. No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title 2015:6.
[12] Pradana AI, Atina V. Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa 2024:239–48. https://doi.org/10.33364/algoritma/v.21-1.1618.
[13] Siregar AP, Purba DP, Pasaribu JP. Implementasi Algoritma Random Forest Dalam Klasifikasi

Diagnosis Penyakit Stroke 2023;2.
[14] Komputer JS. Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter 2021;5:697–711.
[15] Wicaksana IWS. No Title n.d.