

Analisis Sentiment Masyarakat Terhadap penyebaran Starlink di Indonesia Menggunakan Algoritma Naive Bayes

Bebin Paula^{*1}, Muhammad Fawzan², Hafidz Irsyad³

^{1,2,3}Program Studi Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang

E-mail: ^{*1}Bebinpaula11@mhs.mdp.ac.id, ²Fawzan_mhd@mhs.mdp.ac.id, ³hafzirsyad@mdp.ac.id

Abstrak

Penelitian ini bermaksud menganalisis sentimen masyarakat terhadap penyebaran Starlink di Indonesia dengan menggunakan algoritma Naive Bayes. Data dikumpulkan dari komentar YouTube yang kemudian diberi label sebagai positif atau negatif. Setelah melalui tahap preprocessing, dilakukan analisis sentimen untuk menentukan kecenderungan komentar kepada Starlink. Algoritma Naive Bayes digunakan untuk klasifikasi data tersebut. Hasil penelitian menunjukkan bahwa mayoritas komentar bersifat positif terhadap Starlink, terutama dari daerah terpencil yang sulit mendapatkan akses internet. Selain itu, jurnal ini juga membahas dampak Proyek Satelit Starlink terhadap penyedia layanan internet, perkembangan Internet of Things di sektor energi, serta efisiensi algoritma Naive Bayes dalam klasifikasi buah. Jurnal ini turut menyajikan perbandingan algoritma machine learning untuk analisis sentimen dan tantangan dalam text mining. Kemampuan kerja Naive Bayes menunjukkan akurasi 64% dan Hasil ini dipengaruhi oleh ketidakseimbangan dataset, dengan 946 sentimen negatif dan 543 sentimen positif dari total 1489 data. Kesimpulannya ini cukup jujur dan realistis. Namun, jurnal ini dapat ditingkatkan dengan menyertakan rekomendasi untuk penelitian di masa mendatang, seperti penerapan algoritma berbeda, perbaikan teknik pra-pemrosesan, atau pengumpulan data dari berbagai sumber yang lebih beragam.

Kata Kunci— Analisis Sentimen, Naive Bayes, Sentimen Masyarakat, Starlink, Text Mining

1. PENDAHULUAN

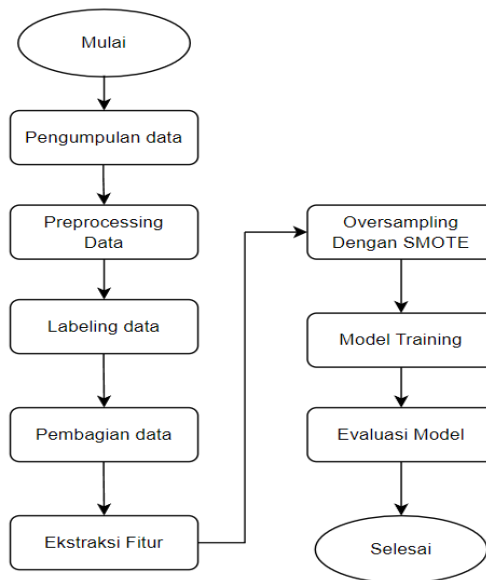
Kemajuan teknologi yang pesat sedang mengubah dunia. Sebagai contoh, kini banyak orang bisa mengakses internet melalui ponsel pintar mereka. Secara keseluruhan, "Industri 4.0" sedang mengubah ekonomi dan masyarakat, dengan model penerapan internal dan eksternal yang berkembang semakin cepat. Banyak negara memiliki posisi yang baik untuk mendukung revolusi teknologi.[1]

Globalisasi telah membawa kita ke tahap perkembangan baru yang signifikan berkat kemajuan dalam teknologi informasi dan komunikasi (TIK), Internet, dan komunikasi seluler. Teknologi utama dalam era globalisasi saat ini adalah komputer dan TIK yang baru dikembangkan, yang menghubungkan seluruh dunia dalam satu sistem komunikasi dan menciptakan ruang keuangan dan informasi yang terintegrasi. Internet menjadi bagian penting dari kehidupan sehari-hari. Sebagai salah satu kemajuan teknologi paling cepat berkembang dalam sejarah manusia, Internet telah mengubah cara kita mencari informasi, mengonsumsi media, dan banyak aspek kehidupan lainnya hanya dalam beberapa dekade.[2]

Revolusi industri keempat diperkirakan akan digerakkan oleh berbagai teknologi canggih, seperti jaringan komunikasi 5G, robot cerdas, dan Internet of Things (IoT). IoT memungkinkan konektivitas tanpa batas antara berbagai objek, individu, proses, dan data. Dengan demikian, IoT dapat membantu meningkatkan efisiensi berbagai operasi dengan membuatnya lebih dapat diamati dan diukur. Pengumpulan dan analisis data yang melimpah melalui IoT dapat memperkuat kemampuan dalam pengukuran dan skala operasional. Potensi peningkatan kualitas hidup melalui IoT dapat dilihat dalam berbagai bidang seperti kesehatan, kota pintar, industri bangunan dan konstruksi, pertanian, manajemen air, dan sektor energi. Hal ini tercapai dengan memberikan alat kepada pengguna untuk membuat keputusan yang lebih baik dan mengotomatisasi keputusan secara real-time.[3]

2. METODE PENELITIAN

Metode Penelitian (dapat mencakup analisis, arsitektur, metode yang digunakan untuk menyelesaikan masalah, dan implementasi), dalam bagian ini, penulis dapat menjelaskan bagaimana penelitian tersebut akan dilaksanakan.



Gambar 1. Tahapan penelitian

2.1. Naive Bayes

Naive Bayes adalah teknik yang memanfaatkan *Bayesian classification* untuk mengklasifikasikan data, dengan asumsi bahwa setiap atribut atau variabel independen. Metode ini diperkenalkan oleh Thomas Bayes, seorang ilmuwan Inggris, yang memungkinkan prediksi probabilitas keanggotaan dalam sebuah kelas berdasarkan pengalaman di masa lalu.[4] Keunggulan *Naive Bayes* terletak pada kemampuannya untuk melakukan klasifikasi dengan data pelatihan yang relatif kecil, sementara data uji digunakan untuk mengevaluasi keakuratan prediksi dan keputusan yang dihasilkan.[5]

2.2. *Analisis Sentiment*

Analisis sentimen adalah proses komputasi yang menilai sentimen, opini, dan emosi yang terkait dengan suatu objek dan diungkapkan dalam teks. Tujuan dari proses ini adalah untuk mengekstraksi atribut dan komponen yang terdapat dalam teks serta menentukan apakah komentar tersebut memiliki nada negatif atau positif..[6]

2.3. *Youtube*

YouTube adalah platform internasional utama untuk distribusi media yang didukung oleh penyebaran sosial. Berdasarkan data publik, setiap menit lebih dari 48 jam konten video diunggah dan setiap hari ada 3 miliar penayangan. Untuk meningkatkan pengalaman menonton dan berbagi konten, YouTube terintegrasi dengan lancar dengan jejaring sosial utama seperti Facebook, Twitter, dan Google+ untuk mempermudah penyebaran di luar situs. Bahkan, 12 juta pengguna telah menghubungkan akun YouTube mereka dengan setidaknya satu jejaring sosial untuk berbagi otomatis, dan lebih dari 150 tahun konten YouTube ditonton di Facebook setiap harinya.[7]

2.4 *Text Mining*

Text Mining, disebut juga penambangan data teks atau penemuan pengetahuan dari database, merujuk pada proses menemukan pola menarik dan berharga dari dokumen teks. Banyak yang melihatnya sebagai evolusi penting dalam penemuan pengetahuan, dengan nilai komersial yang tinggi. Saat ini, lebih dari sepuluh perusahaan teknologi utama menawarkan produk untuk text mining, menunjukkan pertumbuhan pesat dalam bidang ini. Apakah penambangan teks telah berkembang menjadi bidang yang matang? Artikel ini akan membahas hal tersebut.[8]

3. HASIL DAN PEMBAHASAN

3.1. *Pembahasan*

Pembahasan hasil dari penelitian ini, akan menyampaikan beberapa hal yakni terkait cara kami mengumpulkan dataset yang digunakan untuk pengujian, kemudian menunjukkan cara yang dilakukan sebelum menuju proses pengujian dan menunjukkan hasil yang kami dapatkan setelah klasifikasi.

3.2 *Pengumpulan Data*

Data ini dikumpulkan dari komentar dari youtube dengan memanfaatkan netylic untuk menelusuri data komentar youtube dari keyword starlink. Data ini dikumpulkan pada tanggal 29 mei 2024. Dataset berisi Kumpulan komentar youtube dengan berbagai pendapat tentang starlink di Indonesia. Hasil dari pengumpulan data dapat dilihat pada table 1.

Table 1: hasil dataset yang dikumpulkan

Author	Description
@yusufmalikulm	Kalau untuk kota kurang cocok. Ada kabel fiber yang lebih kenceng dan murah. Kalau untuk tempat terpencil cocok nih. Apalagi pekerja remote, harga dan biaya starlink bukan jadi masalah. Bayangin aja bisa kerja di pulau idaman dengan internet kenceng.
@yosefd1308	Hebat sih Starlink, tapi mungkin masih harus di test keawetan alatnya jika digunakan di lokasi yang listriknya sering byarpet . Karena biasanya di lokasi terpencil walau ada listrik terkadang arusnya kurang stabil mati hidup..mati hidup, sering merusak peralatan yang menggunakan listrik. Durabilitas alatnya yang masih harus diuji, otherwise opsi starlink ini bagus banget buat daerah terpencil yang susah dapat akses internet yang mumpuni.
@seving_gaol	Starlink akan sangat menguntungkan RT/RW-net ... kecepatan 200Mbps, Invest modem 8 jt, bulanan 800rb - di bagi 20 rumah; langganan Rp. 50.000/bulan atau voucheran Rp. 5000/hari ... balik modal 1 tahun :) ...

Setelah data telah berhasil dikumpulkan, data akan diberi label antara positif dan negatif untuk menjadi pembandingan saat melakukan pengujian. Pemberian data biasanya dilakukan secara manual untuk hasil yang terbaik, namun pada penelitian ini menggunakan *auto-label* transform dan di ketahui jumlah data yang dilabel 946 sebagai negative dan 543 sebagai positif.

A. Preprocessing Data

Setelah data dikumpulkan, *dataset* akan melewati proses *preprocessing* untuk agar dapat di terima oleh mesin saat melakukan pengujian. Data komentar yang diambil dari media sosial. *Dataset* youtube saat melakukan pengumpulan data bentuknya sulit dibaca mesin karena proses *scrapping* mengambil semua *variable* maka diperlukan melakukan *preprocessing* data. Namun sebelum menuju ke tahap *preprocessing* perlu melewati tahap *cleaning*. Tahap ini merupakan tahap untuk menghapus segala yang tidak di gunakan dalam proses pengujian seperti menghapus emot ikon atau simbol unik.

Sebelum memulai proses data mining, perlu dilakukan pembersihan data pada data yang menjadi fokus KDD. Pembersihan data meliputi penghapusan duplikasi data, pemeriksaan inkonsistensi data, serta perbaikan kesalahan pada data, termasuk kesalahan tipografi. Selain itu, dilakukan pula proses enrichment, yaitu proses "memperkaya" data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal[9][10][11][12].

Table 2: Stopword

Author	Description	after
@yusufmalikulm	Kalau untuk kota kurang cocok. Ada kabel fiber yang lebih kenceng dan murah. Kalau untuk tempat terpencil cocok nih. Apalagi pekerja remote, harga dan biaya starlink bukan jadi masalah. Bayangin aja bisa kerja di pulau idaman dengan internet kenceng.	kalau kota kurang cocok kabel fiber lebih kenceng murah kalau tempat pencil cocok nih kerja remote harga biaya starlink bukan jadi masalah bayangin aja

		kerja pulau idam internet kenceng
@yosefd1308	Hebat sih Starlink, tapi mungkin masih harus di test keawetan alatnya jika digunakan di lokasi yang listriknya sering byarpet . Karena biasanya di lokasi terpencil walau ada listrik terkadang arusnya kurang stabil mati hidup..mati hidup, sering merusak peralatan yang menggunakan listrik. Durabilitas alatnya yang masih harus diuji, otherwise opsi starlink ini bagus banget buat daerah terpencil yang susah dapat akses internet yang mumpuni.	hebat sih starlink mungkin test awet alat guna lokasi listrik sering byarpet biasa lokasi pencil listrik terkadang arus kurang stabil mati hidup mati hidup sering rusak alat guna listrik durabilitas alat uji otherwise opsi starlink bagus banget buat daerah pencil susah akses internet mumpuni
@admin9964	Tingkat Kualitas Link, Pertama kabel FO, Kedua Via Radio dan Ketiga Vsat/Satelit (Starlink atau lainnya), karena kualitas Vsat tergantung lokasi, jika banyak Pohon, gedung2 tinggi, Vsat kurang direkomendasikan, karena akan terhalang dan mempengaruhi kualitas sinyal dan Vsat Kalau Hujan kualitas sinyanya menurun	tingkat kualitas link pertama kabel fo dua via radio tiga vsat satelit starlink lai kualitas vsat gantung lokasi banyak pohon gedung2 tinggi vsat kurang rekomendasi halang pengaruh kualitas sinyal vsat kalau hujan kualitas sinyanya turun
@seving_gaol	Starlink akan sangat menguntungkan RT/RW-net ... kecepatan 200Mbps, Invest modem 8 jt, bulanan 800rb - di bagi 20 rumah; langganan Rp. 50.000/bulan atau voucheran Rp. 5000/hari ... balik modal 1 tahun :) ...	starlink sangat untung rt rw-net cepat 200mbps invest modem 8 jt bulan 800rb - 20 rumah langgan rp 50 000 bulan voucheran rp 5000 hari balik modal 1 tahun

B. Klasifikasi *Naive Bayes*

Sebelum melakukan klasifikasi data dibagi dahulu menjadi 2, data latih dan data uji. Data perlu di bagi untuk menemukan perbandingan mana yang memiliki kemampuan yang terbaik. Dilihat pada table 3.

Laporan Klasifikasi :

	precision	recall	f1-score	support
negative	0.70	0.75	0.72	142
positive	0.51	0.45	0.48	82
accuracy			0.64	224
macro avg	0.60	0.60	0.60	224
weighted avg	0.63	0.64	0.63	224

Gambar 2. Laporan Akurasi Evaluasi Model

- BAYES YANG EFISIEN UNTUK KLASIFIKASI BUAH,” vol. 1, no. 2, pp. 236–248, 2024.
- [5] N. Asmiati and Fatmawati, “Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Pengaruh Negatif Game Online Bagi Remaja Milenial,” *JTIM J. Teknol. Inf. dan Multimed.*, vol. 2, no. 3, pp. 141–149, 2020, doi: 10.35746/jtim.v2i3.102.
- [6] C. A. A. Soemedhy, N. Trivetisia, N. A. Winanti, D. P. Martiyaningsih, T. W. Utami, and S. Sudianto, “Analisis Komparasi Algoritma Machine Learning untuk Sentiment Analysis (Studi Kasus: Komentar YouTube ‘Kekerasan Seksual’),” *J. Inform. J. Pengemb. IT*, vol. 7, no. 2, pp. 80–84, 2022, doi: 10.30591/jpit.v7i2.3547.
- [7] Z. Wattenhofer, M., Wattenhofer, R., & Zhu, “The YouTube Social,” *Int. Conf. Weblogs Soc. Media*, no. January 2012, pp. 354–361, 2012.
- [8] A. Tan, “Text Mining: The state of the art and the challenges Concept-based,” *Proc. PAKDD 1999 Work.*, no. November 2000, pp. 65–70, 2011.
- [9] H. Sulistiani and A. A. Aldino, “Decision Tree C4.5 Algorithm for Tuition Aid Grant Program Classification (Case Study: Department of Information System, Universitas Teknokrat Indonesia),” *EduTic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 40–50, 2020, doi: 10.21107/edutic.v7i1.8849.
- [10] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [11] D. Alita, “Multiclass SVM Algorithm for Sarcasm Text in Twitter,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 1, pp. 118–128, 2021, doi: 10.35957/jatisi.v8i1.646.
- [12] A. D. Cahyo, “METODE NAIVE BAYES UNTUK KLASIFIKASI MASA STUDI,” vol. 3, no. 4, 2023.