

Analisis Sentimen Masyarakat Indonesia Terhadap Kota Jakarta, Surabaya dan Medan Menggunakan Algoritma Naïve Bayes

Andi Rukmana¹, Ferdi Kuswandi², Adiyanto³

Dosen Tetap, Universitas Insan Pembangunan Indonesia

rukmana.andy@gmail.com¹, Ferdikuswandi@gmail.com², adiet031170@gmail.com³

Abstract

The theory of sentiment analysis is interesting to research, this theory aims to group the polarity of a text in a sentence, feature/aspect, or document and then determine whether the opinion contained in the sentence, feature/aspect, or document has positive or negative sentiment. . Sentiment analysis helps in understanding how people respond to certain issues related to these cities (Jakarta, Surabaya, Medan). This research uses the Naïve Bayes algorithm by conducting sentiment analysis on cities such as Jakarta, Surabaya and Medan, we can gain valuable insights to support better decision making in various aspects of city development and management. The results of this research provide an accuracy value of 77% for the city of Jakarta, 76% for the city of Surabaya, 75% for the city of Medan. These results also show that the Naïve Bayes model has quite good performance in carrying out sentiment analysis.

Keywords: Indonesia, Sentimen, Naïve Bayes, Twitter

PENDAHULUAN

Dengan kemajuan ilmu pengetahuan dan teknologi (IPTEK) yang terus berkembang pesat hingga saat ini, hampir setiap aktivitas dan aktivitas manusia sangat bergantung pada teknologi. Teknologi informasi memegang peranan penting dalam banyak aspek kehidupan manusia karena dapat menghubungkan dan memberikan banyak informasi berbeda melalui web. Menurut Arifin dan Krisnadita (2017:76) dalam ¹, Website adalah salah satu aplikasi yang berisikan dokumen-dokumen multimedia (teks, gambar, suara, animasi, video) di dalamnya menggunakan protocol HTTP (Hypertext Transfer Protocol) dan untuk mengaksesnya menggunakan perangkat yang disebut browser. Tujuan pembuatan website pada saat itu adalah untuk memfasilitasi pertukaran dan pemutakhiran informasi dengan sesama peneliti di tempat kerjanya.

Jakarta adalah Ibu Kota Republik Indonesia. Provinsi DKI Jakarta terbagi menjadi lima wilayah Kota Administrasi dan satu Kabupaten Administrasi dengan luas keseluruhan wilayah 662,33 km². Sebagai Ibu Kota Negara, Jakarta menjadi pusat dari perniagaan dan pemerintahan nasional. Kota Jakarta merupakan pusat dari wilayah

aglomerasi Jabodetabekpunjur (Jakarta, Bogor, Depok, Tangerang, Bekasi, Puncak, dan Cianjur) sesuai dengan Peraturan Presiden Nomor 60 Tahun 2020 tentang Rencana Tata Ruang Wilayah Jabodetabekpunjur. Hingga tahun 2030, Jakarta akan mengoptimalkan pembangunan transportasi publik yang semakin memudahkan perpindahan tiap warganya. Melalui pengembangan Bus Rapid Transit (BRT) TransJakarta serta KRL loop line Jabodetabek yang akan didukung oleh penyediaan moda transportasi baru seperti MRT Jakarta, LRT Jakarta, dan LRT Jabodebek.

Surabaya adalah ibu kota Provinsi Jawa Timur yang menjadi pusat pemerintahan dan perekonomian sekaligus kota terbesar di provinsi tersebut. Surabaya juga merupakan sebuah kota yang terletak di Provinsi Jawa Timur, Indonesia. Surabaya merupakan kota terbesar kedua di Indonesia setelah Kota Jakarta. Menurut Bappenas, Kota Surabaya adalah salah satu dari empat kota pusat pertumbuhan utama di Indonesia, bersama dengan Medan, Jakarta, dan Makassar. Surabaya juga merupakan salah satu kota terpenting dalam menopang perekonomian Indonesia. Sebagian besar penduduknya bergerak dalam bidang jasa, industri, dan perdagangan. Surabaya adalah pusat

perdagangan yang mengalami perkembangan pesat. Industri-industri utamanya antara lain galangan kapal, alat-alat berat, pengolahan makanan dan agrikultur, elektronik, perabotan rumah tangga, serta kerajinan tangan.

Kota Medan merupakan ibu kota provinsi Sumatera Utara, Indonesia. Kota Medan adalah kota terbesar posisi ketiga di Indonesia setelah kota Jakarta dan kota Surabaya, serta kota terbesar di luar pulau Jawa. Kota Medan memiliki jalan masuk dari inti kota menuju ke pelabuhan kapal dan bandar udara yang sudah difasilitasi oleh jalan besar tol dan stasiun kereta api, kota ini juga memiliki beraneka ragam pariwisata begitu juga ragam budaya dan adat sosial nya. Selain salah satu dari kota terbesar di Negara Indonesia, Medan merupakan kota memiliki berbagai entertainment dan hiburan, begitu juga ragam budaya untuk dijadikan sebagai destinasi dari pariwisata nya dalam kota tersebut. Oleh karena itu maka kota Medan layak dijadikan bahan perbincangan bagi masyarakat Indonesia sendiri maupun Internasional.²

Teori mengenai analisis sentimen menarik untuk diteliti, teori ini bertujuan untuk pengelompokan polaritas pada sebuah teks yang ada di dalam kalimat, fitur/aspek, atau dokumen kemudian menetapkan apakah pendapat yang terdapat dalam kalimat, fitur/aspek, atau dokumen tersebut memiliki sentimen positif atau negatif ³. Analisis sentimen membantu dalam memahami bagaimana masyarakat merespons terhadap isu-isu tertentu yang berkaitan dengan kota-kota tersebut. Ini bisa mencakup infrastruktur, transportasi, layanan publik, keamanan, dan berbagai aspek lain dari kehidupan kota. Dengan memahami sentimen masyarakat, pemerintah dapat mengembangkan kebijakan yang lebih responsif dan sesuai dengan kebutuhan serta aspirasi penduduk setempat. Misalnya, jika banyak penduduk merasa frustrasi dengan kemacetan lalu lintas di Jakarta, pemerintah bisa memprioritaskan proyek-proyek transportasi yang dapat mengatasi masalah tersebut.

Algoritma Naive Bayes adalah salah satu algoritma klasifikasi yang populer dalam pembelajaran mesin. Meskipun sederhana, algoritma ini sering kali memberikan hasil yang baik dalam banyak kasus, terutama pada

dataset yang cukup besar. Ini sangat cocok untuk analisis sentimen.

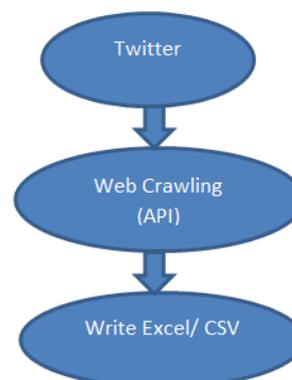
Tujuan Penelitian ini adalah membandingkan sentimen masyarakat di berbagai kota, maka akan didapatkan wawasan tentang perbedaan dan kesamaan antara mereka. Hal ini dapat membantu dalam memahami tantangan dan peluang yang dihadapi oleh masing-masing kota dalam hal pembangunan dan pengelolaan kota. Dengan melakukan analisis sentimen terhadap kota-kota seperti Jakarta, Surabaya, dan Medan, kita dapat mendapatkan wawasan yang berharga untuk mendukung pengambilan keputusan yang lebih baik dalam berbagai aspek pembangunan dan pengelolaan kota.

METODE PENELITIAN

Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan yaitu metode Web Crawling dimana data tersebut bersumber dari sosial media Twitter. Crawling merupakan proses penjelajahan suatu web untuk mendapatkan data atau informasi dari web tersebut ⁴.

Dalam teknik penambangan web, data dikumpulkan dari halaman yang berisi pencari sumber daya seragam (URL) menggunakan antarmuka pemrograman aplikasi (API) untuk mengambil kumpulan data besar. Twitter merupakan salah satu media sosial yang menyediakan API bagi penggunaanya untuk mendapatkan datanya. Data yang biasanya digunakan dari sosial media twitter yaitu berupa teks postingan atau biasanya disebut dengan cuitan (tweets)



Gambar 1. Alur Pengumpulan Data

Tahap Pengumpulan Data

Tahap ini digunakan untuk memastikan bahwa data sudah dalam kondisi yang optimal. Data yang diperoleh belum tentu dalam kondisi baik, oleh karena itu diperlukan tahap data preprocessing. Beberapa langkah yang digunakan antara lain:

- 1) Case Folding Digunakan untuk menutup huruf besar/kecil. Setiap kata dalam kalimat tweet diubah menjadi huruf kecil semua. Sebagai contoh, terdapat kata "KUNCI" dan mencoba memasukkan "KunCI" atau "KuNci" akan tetap mendapatkan hasil pencarian yang sama dengan "kunci".
- 2) Cleaning Digunakan untuk menghapus tanda baca, angka, simbol, tautan URL, dan nama pengguna dari teks. Di dalam data yang berasal dari twitter sendiri tentu saja banyak terdapat noise atau datanya masih kotor karena terdapat karakter tidak penting untuk proses klasifikasi seperti tanda baca, mention, hashtag, angka, username, alamat URL, dan retweet text)
- 3) Filter Stopword Pada langkah ini data teks akan diperiksa, kemudian akan dihilangkan kata yang dianggap tidak terlalu penting. Sebagai contoh penggunaan kata depan, kata ganti, kata sambung, ataupun kata yang tidak memiliki makna yang tidak dibutuhkan^{5,6}.
- 4) Tokenize Langkah ini bertujuan untuk memotong kata di dalam setiap teks dan mengubahnya menjadi huruf kecil. Sehingga hanya huruf yang dapat diterima pada langkah ini, sedangkan karakter khusus dan tanda baca akan dihilangkan. Hasil dari pemrosesan pada langkah tokenization merupakan kata-kata penyusun kalimat yang sudah dipisahkan dari tanda baca.

HASIL DAN PEMBAHASAN

Pengumpulan Data (Web Crawling)

Tabel 1. Contoh Pengumpulan Data

NO	TEXT
1	@\Tingkatkan Tugas dan Fungsi Kepegawaian Rutan Perempuan Kelas IIA Medan ikuti Sosialisasi oleh Kanwil Kemenkumham Sumut secara virtual Medan 30 Januari 2024. Baca berita selengkapnya di: https://t.co/tKahQQ7dDW https://t.co/09BVkLdMNb

2	Kebun binatang harusnya tempat yg aman dan nyaman bagi hewan. Jika hal paling dasar yaitu makanan yg cukup tidak bisa diberikan hingga menyebabkan kematian hal ini merupakan bukti bahwa pemkot medan sebagai pihak yang bertanggungjawab telah gagal menjalankan tugasnya
3	Ga tau la nta aku yang bodoh nta gimana lagi lagi ditipu Hati hati guys sama manusia ini hidup lagi susah susah cari uang dia malah nipu #penipu #MEDAN #tnipenipu #aparapenipu https://t.co/xP7OzLm895

Data Processing

Tujuan dari tahapan data processing untuk mempermudah dalam analisis data pada tahap selanjutnya. Pada tahap ini, dilakukan pembersihan dan persiapan data sehingga data yang akan dianalisis menjadi lebih teratur dan sudah dalam kondisi siap.

- 1) Case Folding
 Pada tahap ini seluruh teks yang terdapat huruf besar (kapital) akan diubah menjadi huruf kecil semua (lower case). Sebagai contoh, hasil dari tahapan case folding dapat dilihat pada Tabel 2.

Tabel 2. Contoh Case Folding

TEXT	HASIL CASE FOLDING
Tingkatkan Tugas dan Fungsi Kepegawaian Rutan Perempuan Kelas IIA Medan ikuti Sosialisasi oleh Kanwil Kemenkumham Sumut secara virtual Medan 30 Januari 2024 Baca berita selengkapnya di	tingkatkan tugas dan fungsi kepegawaian rutan perempuan kelas iia medan ikuti sosialisasi oleh kanwil kemenkumham sumut secara virtual medan 30 januari 2024 baca berita selengkapnya di

- 2) *Cleaning*
 Pada tahap ini dilakukan penghapusan semua karakter yang bukan huruf, contohnya yaitu angka dan simbol yang ada dalam teks tweet.

Tabel 3. Contoh Cleaning

TEXT HASIL CASE FOLDING	HASIL CLEANING
@\Tingkatkan Tugas dan Fungsi Kepegawaian Rutan Perempuan Kelas IIA Medan ikuti Sosialisasi oleh Kanwil Kemenkumham Sumut secara virtual Medan 30 Januari 2024 Baca berita selengkapnya di : https://t.co/tKahQQ7dD W https://t.co/09BVkLdMN b	Tingkatkan Tugas dan Fungsi Kepegawaian Rutan Perempuan Kelas IIA Medan ikuti Sosialisasi oleh Kanwil Kemenkumham Sumut secara virtual Medan 30 Januari 2024 Baca berita selengkapnya di

3) *Filter Stopword*

Pada tahap filter stopwords ini, dilakukan penghapusan kata yang dianggap tidak terlalu penting untuk proses selanjutnya seperti kata penghubung, waktu, dan kata lain yang tidak dibutuhkan. Sebagai contoh, hasil dari proses filter stopwords dapat dilihat pada Tabel 4.

Tabel 4. Contoh Filter Stopword

TEXT HASIL CLEANING	HASIL FILTER STOPWORD
tingkatkan tugas dan fungsi kepegawaian rutan perempuan kelas iia medan ikuti sosialisasi oleh kanwil kemenkumham sumut secara virtual medan 30 januari 2024 baca berita selengkapnya di	tingkatkan tugas fungsi kepegawaian rutan perempuan kelas iia medan ikuti sosialisasi kanwil kemenkumham sumut virtual medan 30 januari 2024 baca berita selengkapnya

4) *Tokenize*

Tujuan dari tahap tokenize yaitu memisahkan kata-kata pada teks tweet yang kemudian dipisahkan dengan spasi

berdasarkan kata yang menyusunnya menjadi kata tunggal. Proses ini sangat penting untuk dilakukan, sebagai contoh terdapat kata-kata yang seharusnya terpisah akan tetapi kata-kata tersebut masih tergabung. Contoh hasil proses tokenize dapat dilihat pada Tabel 5.

Tabel 5. Contoh Tokenize

TEXT HASIL FILTER STOPWORD	HASIL TOKENIZE
tingkat tugas fungsi pegawai rutan perempuan kelas iia medan ikut sosialisasi kanwil kemenkumham sumut virtual medan 30 januari 2024 baca berita lengkap	['tingkat', 'tugas', 'fungsi', 'pegawai', 'rutan', 'perempuan', 'kelas', 'iia', 'medan', 'ikut', 'sosialisasi', 'kanwil', 'kemenkumham', 'sumut', 'virtual', 'medan', '30', 'januari', '2024', 'baca', 'berita', 'lengkap']

Klasifikasi Teks Hasil Preprocessing

Pada tahap ini seluruh komentar pengguna yang telah dilakukan perbaikan sehingga dapat dipahami arah sentimen pengguna akan diklasifikasi sesuai dengan kelas positif atau negatif, dapat dilihat pada Tabel 6

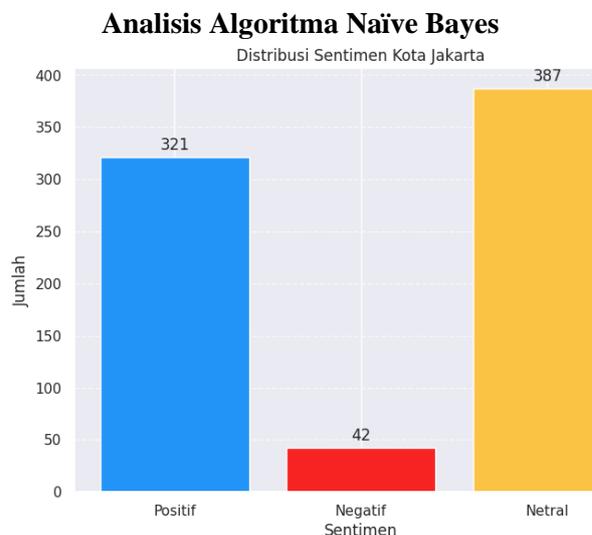
Tabel 6. Klasifikasi Hasil Text

NO	TEXT	LABEL
1	tingkat tugas fungsi pegawai rutan perempuan kelas iia medan ikut sosialisasi kanwil kemenkumham sumut virtual medan 30 januari 2024 baca berita lengkap	Positif
2	kebun binatang harus tempat aman nyaman hewan paling dasar makan cukup bisa beri hingga sebab mati hal ini rupa bukti pemkot medan pihak bertanggungjawab gagal jalan tugas	Positif
3	ga tau la ntah aku bodoh ntah gimana tipu hati hati	Negatif

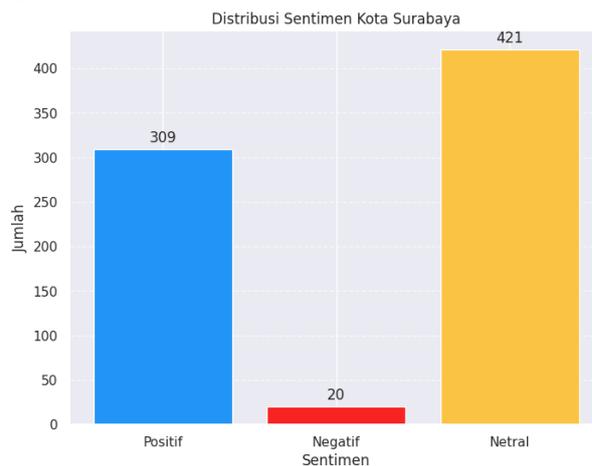
guys sama manusia hidup lagi susah susah cari uang malah nipu

Hasil Akurasi Algoritma Naïve Bayes

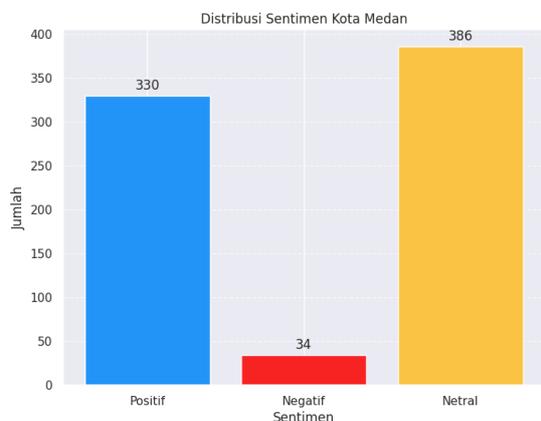
Kota	Akurasi
JAKARTA	77 %
SURABAYA	76%
MEDAN	75%



Gambar 2. Distribusi Sentimen Kota Jakarta



Gambar 3. Distribusi Sentimen Kota Surabaya



Gambar 4. Distribusi Sentimen Kota Medan

KESIMPULAN

Dari hasil pemrosesan data terhadap yang terdapat di twitter terkait dengan kota Jakarta, Surabaya dan Medan maka didapatkan Kota yang paling banyak mendapatkan sentimen positif dan negative adalah Kota Jakarta dimana terdapat 321 komentar positif dan 42 komentar negatif. Hal ini menunjukkan bahwa masyarakat Indonesia yang beranggapan positif terhadap Kota Jakarta. Dari hasil pengujian, algoritma naïve bayes bekerja sangat baik pada kasus penelitian ini. Naïve Bayes tidak memerlukan waktu yang lama dalam proses evaluasi dibanding algoritma yang lain.

Disarankan pada penelitian selanjutnya, kami akan mencoba menggunakan algoritma klasifikasi lain seperti algoritma SVM untuk mendapatkan hasil dengan akurasi yang lebih tinggi.

DAFTAR PUSTAKA

Kesuma Astuti F, Sri Agustina D. Membangun Website MTS Negeri 01 OKU Timur Menggunakan Php dan Mysql. Jik. 2022;13(1):7–14.

Ridwansyah T. Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier. KLIK Kaji Ilm Inform dan Komput. 2022;2(5):178–85.

Ramadhan D, Setiawan EB. Analisis Sentimen Program Acara di SCTV pada Twitter Menggunakan Metode Naive Bayes dan Support Vector Machine. ... TelkomuniversityAcId [Internet]. 2019;6(2):9736–43. Available from: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/10708>

Hanifah R, Nurhasanah IS. Implementasi Web Crawling untuk Mengumpulkan Informasi Wisata Kuliner di Bandar Lampung. *J Teknol Inf dan Ilmu Komput.* 2018;5(5):531.

LING J, N. KENCANA IPE, OKA TB. Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square. *E-Jurnal Mat.* 2014;3(3):92.

Sari R, Hayuningtyas RY. Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website. *Indones J Softw Eng.* 2019;5(2):51–60.