

Sentiment Analysis of Police Performance On Twitter Users Using Naïve Bayes Method

Muhammad Hidayatullah^{1*}, Syariful Alam², Irsan Jaelani³

¹²³Teknik Informatika, Sekolah Tinggi Teknologi Wastukencana, Jl. Cikopak No.53, Babakancikao,
Purwakarta, Jawa Barat 41151, Indonesia

*Penulis koresponden, e-mail : muhammadhidayatullah44@wastukencana.ac.id, irsan@wastukencana.ac.id,
syarifulalam@wastukencana.ac.id.

Abstract: *After the case of alleged child rape in East Luwu which was stopped went viral in the aftermath of many other cases of sexual violence that were considered by the police to be inconsistent with procedures. After the hashtag #PercumaLaorPolisi appeared #PolriSesuaiProsedur hashtag became a trending topic on Twitter. This study discusses the sentiment of police performance on twitter users, aiming to measure how much sentiment the performance of the police according to twitter citizens who earned. The topic of this study is a mining text that uses the naïve bayes method. Text mining is a computer-based algorithmic technique/approach to gaining new knowledge hidden from a set of texts. The data from crwalling on twitter were analyzed using naive bayes which is a method for analyzing. Naive Bayes' algorithm is very effective in classification or classification problems. This algorithm works based on existing probabilities to determine the probability of the future. The steps in the Naïve Bayes method are preprocessing which includes transformation, tokenization and filtering processes. It is followed by the weighting of words such as TF-IDF and ends with classification and evaluation. As a result of this study, according to tweet data processed using the orange application and confusion matrix calculations, the police performance sentiment entered the neutral classification of 75.8%, negative 58.1% and positive 39.5% in the last order, as well as the resulting model at an accuracy value of 0.929, precision 0.933, recall 0.923, and f-measure 0.954*

Keywords: *Text Mining; Twitter; Police Performance; Naïve Bayes.*

Abstrak: *Setelah kasus dugaan pemerkosaan anak di Luwu Timur yang dihentikan viral seba ga i buntut banyaknya kasus kekerasan seksual lain yang dianggap kinerja polisi tidak sesuai prosedur. Usai ta ga r #PercumaLaorPolisi muncul ta ga r #PolriSesuaiProsedur menjadi trending topik di Twitter. Penelitian ini membahas tentang sentimen kinerja polisi pada pengguna twitter, bertujuan untuk mengukur berapa sentiment kinerja polisi menurut wa rga twitter yang mengha silkan. Topik penelitian ini merupakan teks mining yang menggunakan metode naïve bayes. Teks mining ada lah teknik/pendekatan algoritmik berbasis komputer untuk mendapatkan suatu pengetahuan ba ru yang tersembunyi dari sekumpulan teks. Data hasil crwalling di twitter diana lisis menggunakan na ive bayes yang merupakan suatu metode untuk mengana lisis. Algoritma Na ive Bayes sangat efektif da lam permasa lahan kla sifikasi atau penggolongan. Algoritma ini bekerja berdas rkan probabilitas yang sudah ada untuk menentukan probabilita s yang akan datang. Langkah-langkah da lam metode Naïve Bayes yaitu preprocessing yang meliputi proses transformation, tokenization dan filtering. Dilanjut dengan pembobotan kata seperti TF-IDF dan diakhiri dengan kla sifikasi serta evalua si. Hasil dari penelitia n ini, menurut data tweet yang diolah menggunakan aplikasi orange dan perhitungan confusion matrix bahwa sentiment kinerja polisi masuk ke kla sifikasi netra l 75,8 %, negative 58.1% dan positif 39,5% di urutan terakhir, serta model yang dihasilkan pada nilai akura si 0.929, precision 0.933, recall 0.923, dan f-measure 0.954.*

Kata kunci: *Teks Mining; Twitter; Kinerja Polisi; Naïve Bayes.*

1. PENDAHULUAN

Kepolisian Negara Republik Indonesia merupakan instrumen negara yang berperan dalam memelihara keamanan dan ketertiban masyarakat, menegakkan hukum serta memberikan perlindungan, pengayoman dan pelayanan kepada masyarakat dalam rangka terpeliharanya keamanan dalam negeri sesuai dengan pasal 5 UU No.2 Tahun 2002, kepolisian memiliki kewenangan dalam menyelesaikan permasalahan yang terjadi baik masalah pidana maupun masalah perdata. Tujuan dibentuknya kepolisian menurut undang-undang dasar negara 1945 pasal 4 UU No.2 Tahun 2002, adalah untuk mewujudkan keamanan dan ketertiban masyarakat, tertib dan tegaknya hukum, terselenggaranya perlindungan, pengayoman dan pelayanan masyarakat, serta terbinanya ketentraman masyarakat dengan menjunjung tinggi hak asasi manusia (Suwondo and Polri n.d.). Terkait tugas dan tanggung jawab yang diemban oleh kepolisian, kinerja kepolisian merupakan bagian yang mendapat perhatian dari masyarakat. Beberapa waktu belakangan, kepolisian mendapat sorotan publik terkait dengan kasus cicak vs buaya. Isu mengenai kinerja kepolisian tidak hanya pada kasus cicak vs buaya, sepanjang tahun 2009-2010, akuntabilitas kepolisian dimata publik juga semakin menurun. Indikator yang muncul tidak lain berasal dari internal kepolisian sendiri, diantaranya terjadinya dugaan tindakan kekerasan dan penyalahgunaan kekuasaan (<http://www.kontras.org/index>).

Dari jawapos.com bahwa Komisi Nasional Hak Asasi Manusia (Komnas HAM) menanggapi dugaan represivitas aparat kepolisian terhadap warga Desa Wadas, Purworejo, Jawa Tengah. Komnas HAM secara tegas mengecam tindakan kekerasan yang dilakukan oleh aparat kepolisian kepada warga termasuk pendamping hukum warga Wadas yang menolak desanya dijadikan lokasi penambangan quarry, untuk pembangunan Bendungan Bener. Pada 2019, Ketua Yayasan Lembaga Bantuan Huku Indonesia (YLBHI) Asfinawati mengatakan tindakan kekerasan yang dilakukan aparat kepolisian saat demonstrasi sebanyak 68 kasus. Penangkapan sewenang-wenang 3.539 korban kemudian, penahanan sewenang-wenang 3.539 korban dan penyiksaan sebanyak 474 korban (Saputra, Penelitian, and Polri n.d.). Tidak hanya kasus diatas, tagar #PercumaLaporPolisi sempat trending di Twitter setelah kasus dugaan pemerkosaan anak di Luwu Timur yang dihentikan viral. Tagar itu juga muncul sebagai buntut banyaknya kasus kekerasan seksual lain yang dianggap diabaikan polisi (news.detik.com). Usai tagar #PercumaLaorPolisi trending topik, muncul tagar baru yang berlawanan. Polisi memasang tagar #PolriSesuaiProsedur dan menjadi trending topik di Twitter. Berdasarkan pantauan di Twitter, Senin (11/10/2021) pukul 18.40 WIB, terlihat tagar #PolriSesuaiProsedur bertengger di jajaran 'Trending in Indonesia' (news.detik.com).

Berdasarkan permasalahan tersebut dengan perkembangan akses teknologi informasi yang semakin pesat dan mudah dijangkau pada era globalisasi saat ini dapat memberikan kemudahan untuk berkomunikasi lebih efektif dan efisien, contohnya adalah internet. Seperti halnya yang telah

diungkap oleh kementerian komunikasi dan informatika (Kemenkominfo, 2022) bahwa penggunaan internet di Indonesia sebanyak 202,35 juta pengguna atau 76,8% sedangkan untuk pengguna twitter di Indonesia mencapai 18,4 juta orang atau setara 6,6% dari total populasi pada tahun 2022 (suara.com). Berdasarkan hasil penjelasan diatas berapa sentimen kinerja polisi pada pengguna twitter? maka penelitian ini dibuat dengan judul “Analisis sentimen kinerja polisi pada pengguna twitter menggunakan metode naive bayes”

2. DASAR TEORI

2.1. Twitter

Twitter adalah layanan jejaring sosial yang membantu penggunaannya mengirim dan membaca pesan berbasis teks hingga 140 karakter. Tingginya popularitas twitter menyebabkan layanan ini telah dimanfaatkan untuk berbagai keperluan dalam berbagai aspek, misalnya sebagai sarana protes, kampanye politik sarana pembelajaran dan sebagai media komunikasi darurat. Twitter mengizinkan akses ke bagian dari layanan melalui API untuk memungkinkan orang-orang membangun perangkat lunak yang terintegrasi dengan Twitter seperti solusi yang membantu sebuah perusahaan menjawab umpan balik pelanggan di Twitter. Data Twitter berbeda dari data yang dibagi oleh kebanyakan platform sosial lain karena data tersebut mencerminkan informasi yang dipilih pengguna untuk dibagikan ke publik. Platform API twitter menyediakan akses luas ke data Twitter publik yang telah dipilih pengguna untuk dibagikan ke dunia. Twitter juga mendukung API yang memungkinkan pengguna mengelola informasi Twitter pengguna yang non-publik dan memberikan informasi ini ke pengembang yang telah diizinkan pengguna untuk melakukannya (Duei Putri et al. 2022).

2.2. Teks Mining

Teks mining adalah sebuah teknik/pendekatan algoritmik berbasis komputer untuk mendapatkan suatu pengetahuan baru yang tersembunyi dari sekumpulan teks. Teks mining merupakan bagian dari keilmuan information retrieval (temu balik informasi) yang bekerja pada data bertipe teks yang cenderung tidak terstruktur. Pada dasarnya mekanisme kerja algoritma-algoritma teks mining memiliki kemiripan dengan algoritma-algoritma data mining secara umum. Perbedaan pokok dari teks mining dan data mining adalah pada tipe data yang menjadi objek kerjanya. Teks mining adalah ekstraksi informasi dari data sumber yang belum terstruktur yang mengacu pada teknik penambangan data untuk menganalisis dan diolah melalui tahap preprocessing yang didalamnya merupakan proses case folding, tokenization, filtering, dan stop (Nurhuda, Sihwi, and Doewes 2013).

2.3. Naïve Bayes

Naïve Bayes adalah metode calssifier berdasarkan probabilitas dan teorma bayesian dengan asumsi bahwa setiap variabel X bersifat bebas (Buntoro, 2017). Naïve Bayes merupakan metode statistik untuk memprediksi probabilitas suatu atribut dalam suatu kelas, dalam proses klasifikasinya menganggap bahwa kemunculan suatu kata pada suatu kalimat tidak dipengaruhi oleh kemunculan kata-kata yang lain pada kalimat tersebut, namun pada kenyataannya kemunculan suatu kata dipengaruhi oleh kata-kata lainnya, atau disebut dengan naïve (Susanti, 2016). Naïve Bayes adalah metode dalam pembelajaran mesin, dimana pada proses klasifikasi terjadi dua tahap yaitutahap penelitan dan tahap klasifikasi. Tahap penelitian merupakan proses terhadap data sampel untuk menemukan term-term yang ada dalam suatu kelas dokumen, sedangkan tahap klasifikasi yaitu tahap penentuan kelas suatu dokumen dalam data uji berdasarkan data sampel, pengklasifikasian dokumen dalam data uji ditentukan oleh banyaknya probabilitas term-term yang muncul pada dokumen tersebut (Renny et al. n.d.)

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \quad (1)$$

Dimana:

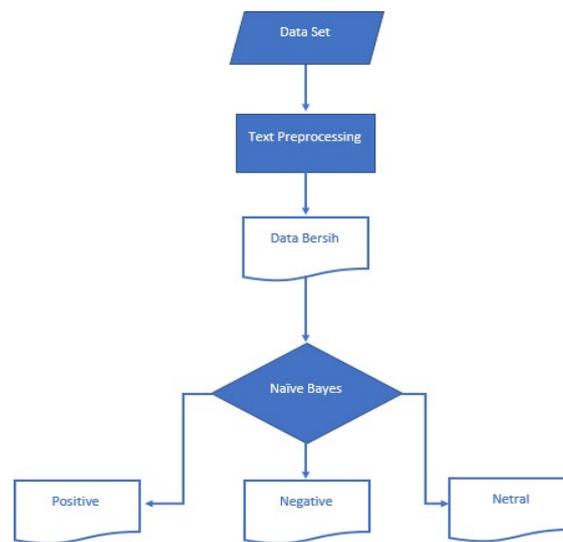
$P(A|B)$ = Probabilitas posterior dari A pada kondisi B (posterior probabilitas).

$P(B|A)$ = Probabilitas posterios dari B pada kondisi A (likelihood).

$P(A)$ = Probabilitas prior dari A (class prior probabilitias).

$P(B)$ = Probabilitas prior dari B (predictor prior probability).

Proses untuk menghitung probabilitas kelas suatu data dimulai dengan menentukan likelihood berdasarkan dataset yang digunakan, menggunakan metode yang sesuai dengan bentuk dari data yang digunakan. Likelihood yang diperoleh akan dikalikan dengan probabilitas dengan masing-masing kelas. Hasil dari proses tersebut akan digunakan sebagai acuan untuk mengklasifikasikan data baru. Pada praktikny, seringkali $P(B)$ dihiraukan, karena nilai $P(B)$ selalu tetap (Widhi Saputro and Wulan Sari 2019).



Gambar 2. 1. Flowchart Naive Bayes

2.4. Confusion Matrix

Merupakan tabel yang menggambarkan performa dari model atau algoritma secara spesifik. Setiap baris dari matrix merepresentasikan kelas aktual dari data, dan setiap kolom merepresentasikan kelas prediksi dari data (atau sebaliknya). Matrix tersebut dijelaskan pada tabel 2.1. (Nurhuda et al. 2013).

Tabel 2. 1. Confusion Matrix

	Predicted Negative	Predicted Positive
Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

1. True Positive = Berarti seberapa banyak data yang aktual kelasnya positif, dan model juga memprediksi positif
2. True Negative = Berarti seberapa banyak data yang aktual kelasnya negatif, dan model memprediksi negatif.
3. False Positive = Berarti seberapa banyak data yang aktual kelasnya negative, namun model memprediksi positif.
4. False Negative = Berarti seberapa banyak data aktual yang positif, namun model memprediksi negatif.

Melalui 4 data tersebut, dapat diperoleh data-data lain yang sangat berguna untuk mengukur performa sebuah model, diantaranya:

1. *Accuaracy* = Total keseluruhan seberapa sering model benar mengklasifikasi. Formula accuary dapat ditulis menggunakan persamaan 2.

$$\frac{TP + TN}{Total} \quad (2)$$

2. *Precision* = Ketika model memprediksi positif, seberapa sering prediksi itu benar. Formula precision dapat ditulis menggunakan persamaan 3.

$$\frac{TP}{FP + TP} \quad (3)$$

3. *Recall* (Sensintivity / True Positive Rate) = Ketika kelas aktualnya positif, seberapa sering model memprediksi positif. Formula recaldapat ditulis menggunakan persamaan 4.

$$\frac{TP}{FN + TP} \quad (4)$$

4. *F1-Score* = Merupakan rata-rata harmonik dari Precision dan Recall. Formula F1-Score dapat ditulis menggunakan persamaan 5.

$$2 * \frac{precision * recall}{precision + recall} \quad (5)$$

3. METODE PENELITIAN

3.1. Studi Literatur Dan Pemahaman

Pada tahap ini, peneliti mencari dan mempelajari referensi berupa textbook, artikel ilmiah maupun jurnal yang berkaitan dengan penelitian. Topik yang akan dibahas antara lain: analisis sentimen kinerja polisi pada pengguna twitter menggunakan metode *Naïve Bayes*.

3.2. Pengumpulan data

Dataset yang akan digunakan dalam penelitian adalah *public timeline tweet* bahasa Indonesia yang merupakan pencarian berdasarkan tagar (*#percumalaporpholisi*, *#polisi* dan *#polri*). Dataset didapatkan dengan cara *crawling data* memanfaatkan API dari twitter menggunakan *software* Orange 3.32, kemudian disimpan dalam bentuk database. Pengambilan data di twitter dilakukan pada tanggal 30 Maret 2022, dengan jumlah 2059, dataset dan menjadi 1734, setelah melewati proses preprofcessing.

3.3. Implementasi

Implementasi penelitian ini dilakukan dengan langkah kerja seperti pada gambar 3.1.

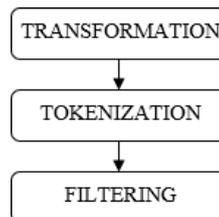


Gambar 3. 1. Diagram alur kerja penelitian

Proses awal pada tahap implementasi adalah dataset yang akan digunakan baik *testing* maupun *training*. Pengumpulan data set dengan cara memanfaatkan API dari pihak Twitter melalui teknik *crawling* data. Hasil dari *crawling* data kemudian disimpan kedalam database yang kemudian dilakukan *preprocessing* data untuk menjadikan sederhana dimensi dari dataset.

3.4. Preprocessing Data

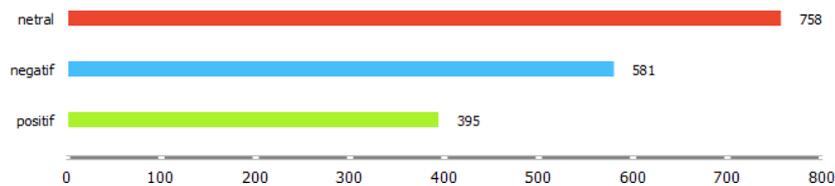
Sebelum dataset siap digunakan maka akan terlebih dahulu dilakukan tahap *preprocessing* data sehingga dataset telah bersih dan siap digunakan dalam proses selanjutnya. Berikut adalah diagram alur mengenai *preprocessing* data pada gambar 3.2.



Gambar 3. 2. Preprocessing Data

Terlihat dari diagram diatas bahwa tahap pertama preprocessing data adalah transformation. Yaitu mengubah semua huruf kapital menjadi huruf kecil atau lowercase. Kemudian dilakukan tokenization berupa memecah teks menjadi komponen yang lebih kecil, yaitu pemecahan berdasarkan perkata. Hal ini bisa dilakukan dengan menandai karakter spasi sebagai pembatas. Selanjutnya dataset akan dilakukan *filtering* berupa penghapusan semua karakter selain string serta penghapusan beberapa karakteristik dari data twitter, misalnya *@username*, *#hashtag*, *http:URL*, dan "RT" atau kata yang menandakan kalau itu perulangan tweet. Dalam *filtering* ini juga dilakukan penghapusan terhadap *stopword*. Hal ini bermanfaat untuk mengurangi *load* atau *performance* saat melakukan training maupun testing dataset.

Berdasarkan gambar 4.4, terdapat kata polisi, polri dan aman yang sering muncul pada tweet pengguna twitter. Adapun visualisasi data berdasarkan kategori dari keseluruhan tweet dalam bentuk box plot yang dapat dilihat pada gambar 4.5, sebagai berikut:



Gambar 4. 5. Box Plot Kategori Sentimen

Gambar 4.5, menunjukkan tweet yang berkategori netral sebanyak 758 tweet atau 75,8%, tweet yang berkategori negatif sebanyak 581 tweet atau 58,1% dan yang berkategori positif sebanyak 395 tweet atau 39,5%. Dari keterangan tersebut, tweet yang berkategori netral lebih banyak dibandingkan dengan tweet yang berkategori negatif dan positif.

4.3. Evaluasi

Evaluasi pada penelitian ini menggunakan Confusion matrix adalah tool yang digunakan sebagai evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah. Sebuah matrix yang digunakan untuk membandingkan hasil klasifikasi dengan data asli. Pada tahap ini dilakukan pengujian dan evaluasi menggunakan confusion matrix merepresentasikan prediksi dan kondisi sebenarnya (aktual) dari data yang dihasilkan pada setiap produk oleh algoritma. Berdasarkan confusion matrix, penulis bisa menentukan akurasi, precision, recall, dan f-measure. Adapun tabel confusion matrix bisa dilihat pada tabel 4.1.

Tabel 4. 1. Confusion Matrix

	Predicted Negative	Predicted Positive	Predicted Netral
Actual Negatife	539	25	17
Actual Positive	28	694	36
Actual Netral	4	13	378

Accuracy score didapat dari membandingkan value dari label yang terdapat pada dataset dengan value label yang diprediksi. Kemudian terdapat classification report, pada classification report terdapat berbagai macam score yaitu precision, recall, f1-score dan support score. Adapun hasil yang didapatkan sebesar 0,929 atau 92,9% akurasi. Nilai rata-rata dari presisi score adalah 0,933 atau 93,3% sedangkan untuk score recall didapatkan sebesar 0,923 atau 92,3% dan untuk F-Measure

sendiri untuk score yang didapatkan sebesar 0,954 atau 95,4% hal ini membuktikan sistem mampumendapatkan prediksi secara akurat dari total 1734 data yang telah melewati tahap *preprocessing*.

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Kesimpulan yang didapat dari penelitian sentiment kinerja polisi pada pengguna twitter menggunakan metode naïve bayes dengan jumlah data yang telah melewati tahap preprocessing berjumlah 1734 data dan hasil klasifikasi sentimen yaitu positif 39,5%, netral 75,8 %, dan negatif 58.1%. Serta model yang dihasilkan pada nilai akurasi 0.929, precision 0.933, recall 0.923, dan f- measure 0.954. Hasil tersebut membuktikan bahwa algoritma Naive Bayes terbukti cukup baik dalam mengklasifikasi dokumen sentimen kinerja polisi pada pengguna twitter.

5.2. Saran

Saran yang didapatkan dari penelitian Analisis Sentimen Kinerja Polisi Pada Pengguna Twitter Menggunakan Metode Naïve Bayes, adalah (1) Perlu dilakukan penelitian yang serupa dengan menggunakan tools selain orange. (2) Perlu dilakukan penelitian yang serupa dengan menggunakan tools selain orange. (3) Perlu dilakukan penelitian yang sama dengan menggunakan algoritma yang berbeda agar hasil keduanya dapat dibandingkan. (4) Lembaga kepolisian disarankan agar terus berupaya meningkatkan kinerjanya guna meningkatkan sentiment positif pada pengguna twitter.

DAFTAR PUSTAKA

- Duei Putri, Dianati, Gigih Forda Nama, Wahyu Eko Sulistiono Jurusan Teknik Elektro Universitas Lampung, Bandar Lampung Jl Sumantri Brojonegoro No, and Bandar Lampung. 2022.
Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier. Vol. 10.
- Nurhuda, Faishol, Sari Widya Sihwi, and Afrizal Doewes. 2013. "Analisis Sentimen Masyarakat Terhadap Calon Presiden Indonesia 2014 Berdasarkan Opini Dari Twitter Menggunakan Metode Naive Bayes Classifier." 2(2).
- Renny, Supervisor, Pradina Kusumawardani, Irmasari Hafidz, and S. Kom. n.d. *Analysis Of Sentimentality Towards The Performance Of Public Services In Surabaya Based On Commentary Classification In Social Media Using Naïve Bayes Algorithm* Nuke Yulnida Aden Faradhillah NRP 5212 100 123.
- Saputra, Azis, Pusat Penelitian, and Pengembangan Polri. n.d. *Tingkat Kepercayaan Masyarakat Terhadap Kinerja Polri Tahun 2020*.
- Suwondo, Dadang, and Puslitbang Polri. n.d. "Analisis Tingkat Kepercayaan Masyarakat

Terhadap Kinerja Polri.”

Widhi Saputro, Irkham, and Bety Wulan Sari. 2019. “Uji Performa Algoritma Naïve Bayes Untuk Prediksi Masa Studi Mahasiswa Naïve Bayes Algorithm Performance Test for Student StudyPrediction.” Citec Journal 6(1).