
ANALISIS SENTIMEN MASYARAKAT TENTANG DAMPAK PERTAMBANGAN DI PULAU TALIABU PROVINSI MALUKU UTARA MELALUI MEDIA SOSIAL MENGGUNAKAN METODE K-MEANS DAN NAIVE BAYES

Kori Bantu¹⁾, Marline S. Paendong²⁾, Dodisutarma Lapihu³⁾, Eliasta Ketaren⁴⁾

Program Studi Sistem Informasi, Jurusan Matematika

Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Sam Ratulangi

Jl. Kampus Bahu, Manado 95115, Indonesia

email: koribantu@student.unsrat.ac.id¹⁾, marline.paendong@unsrat.ac.id²⁾, dodisutarma@unsrat.ac.id³⁾, eliasketaren@unsrat.ac.id⁴⁾

Abstrak

Penelitian ini bertujuan untuk menganalisis opini masyarakat terkait isu pertambangan di Pulau Taliabu yang terekam melalui media sosial. Data yang digunakan berupa komentar pengguna dari Facebook, X (Twitter), TikTok, dan Instagram yang kemudian diolah melalui tahap praproses teks sebelum dilakukan pengelompokan menggunakan algoritma *K-Means Clustering*. Hasil analisis menunjukkan bahwa opini masyarakat terbagi ke dalam tiga sentimen utama, yaitu positif, netral, dan negatif. Sentimen negatif menjadi klaster yang paling dominan dengan jumlah 750 data, sedangkan klaster netral berjumlah 142 data yang umumnya berfokus pada diskusi isu pembangunan dan sumber daya alam, dan klaster positif berjumlah 28 data yang berisi apresiasi maupun dukungan terhadap figur publik. Faktor-faktor yang memengaruhi sentimen masyarakat meliputi isu pembangunan infrastruktur, kinerja pemerintah, serta dampak sosial-ekonomi dari aktivitas pertambangan. Dilakukan klasifikasi menggunakan algoritma *Multinomial Naive Bayes* yang berhasil memprediksi sentimen dengan tingkat akurasi sangat baik, yakni mencapai 89,77%. Temuan ini memberikan gambaran yang lebih jelas mengenai persepsi publik terhadap aktivitas pertambangan di Pulau Taliabu.

Kata Kunci: Analisis Sentimen, K-Means Clustering, Media Sosial, Naive Bayes, Pertambangan Pulau Taliabu.

1. Pendahuluan

Provinsi Maluku Utara merupakan daerah yang kaya akan sumber daya alam termasuk bahan galian pertambangan yang mempunyai peran penting dalam perekonomian nasional. Salah satu daerah di Maluku Utara yang kaya akan bahan galian pertambangan seperti bijih besi adalah Pulau Taliabu [1], [2]. Kehadiran industri pertambangan menarik minat investor dan diharapkan dapat meningkatkan Pendapatan Asli Daerah (PAD) serta menciptakan lapangan kerja [3].

Namun, eksplorasi sumber daya tambang mengandung paradigma *double-edged sword*. Di satu sisi memberikan dampak ekonomi positif, namun di sisi lain berisiko menimbulkan degradasi lingkungan seperti deforestasi, pencemaran air, dan hilangnya keanekaragaman hayati [4]. Hal ini memicu beragam reaksi masyarakat yang disampaikan melalui media sosial seperti Facebook, X (Twitter), Instagram, dan TikTok.

Analisis sentimen diperlukan untuk memahami kecenderungan opini publik, apakah bersifat positif, negatif, atau netral terhadap dampak pertambangan tersebut [5]. Penelitian terdahulu menunjukkan bahwa kombinasi metode *clustering* dan *classification* dapat memberikan wawasan mendalam mengenai opini publik [6]. Oleh karena itu, penelitian ini menerapkan metode *K-Means Clustering* untuk mengelompokkan data tanpa label dan *Multinomial Naive Bayes* untuk mengklasifikasikan sentimen masyarakat dengan tujuan memberikan informasi bagi pemerintah dan pemangku kepentingan dalam pengambilan keputusan.

2. Landasan Teori

Analisis Sentimen

Analisis sentimen adalah cabang *data mining* yang mempelajari cara komputer mengevaluasi suasana hati, emosi, dan opini yang diekspresikan lewat teks [7]. Tujuannya adalah mengekstrak atribut objek yang dikomentari dan menentukan polaritas komentar tersebut (positif, negatif, atau netral) [8].

Media Sosial dan Text Mining

Media sosial merupakan alat komunikasi dan kolaborasi yang memungkinkan interaksi masif, menjadikannya sumber data yang kaya untuk *text mining* [9]. *Text mining* sendiri adalah proses mengekstrak informasi pola menarik dari data tekstual tidak terstruktur [10].

K-Means Clustering

K-Means adalah teknik pengelompokan data non-hierarki yang mempartisi data ke dalam k klaster/kelompok.

Data dengan karakteristik serupa dikumpulkan dalam satu klaster dan data dengan karakteristik berbeda dipisahkan ke klaster lain [11]. Pusat klaster baru dihitung menggunakan persamaan (1):

$$\mu^k = \frac{1}{N^k} \sum_{q=1}^{N^k} x^q$$

Dimana:

- μ^k = titik centroid dari cluster ke-K
- N^k = banyaknya data pada cluster ke-K
- x^q = data ke-q pada cluster ke-K

Naive Bayes Classifier

Naive Bayes adalah metode klasifikasi probabilistik berdasarkan teorema Bayes dengan asumsi independensi antar fitur [12]. Rumus umum teorema Bayes ditunjukkan pada persamaan (2):

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Keterangan :

- X = Sampel data yang memiliki class (label) yang tidak diketahui.
- C = Hipotesis bahwa X adalah data class (label)
- $P(C)$ = Probabilitas hipotesis C
- $P(X)$ = Peluang dari data sampel yang diamati (probabilitas C)
- $P(X|C)$ = Probabilitas berdasarkan kondisi pada hipotesis

3. Metode Penelitian

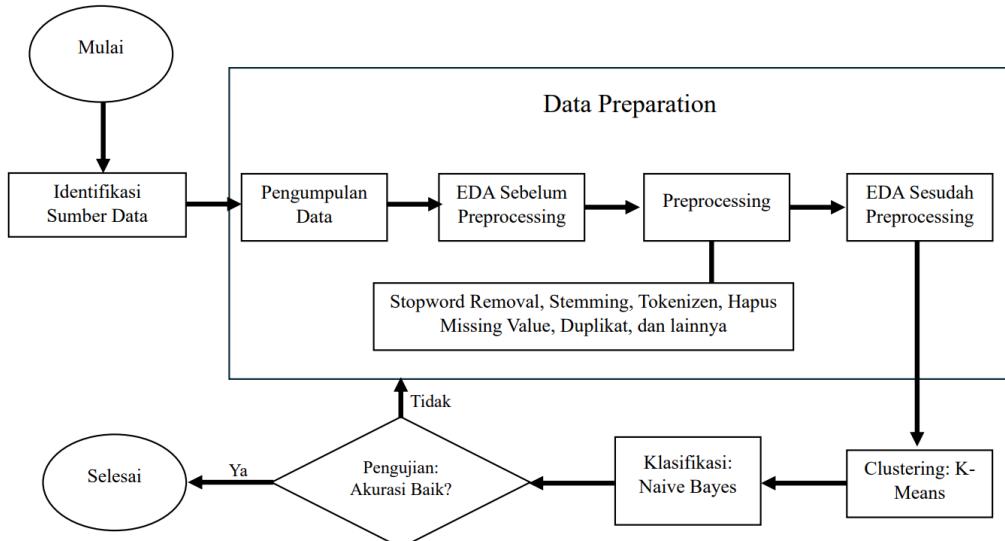
Penelitian ini menggunakan pendekatan kuantitatif dengan tahapan sistematis mulai dari pengumpulan data hingga evaluasi model.

Sumber Data

Data dikumpulkan dari empat platform media sosial: Facebook, X (Twitter), TikTok, dan Instagram. Pengumpulan data menggunakan teknik scraping (Apify) untuk X, TikTok, dan Instagram, serta pengumpulan manual untuk Facebook. Kata kunci yang digunakan meliputi "Tambang Pulau Taliabu", "dampak tambang", dan tagar relevan seperti #savetaliabu.

Tahapan Penelitian

Alur penelitian digambarkan dalam flowchart berikut:

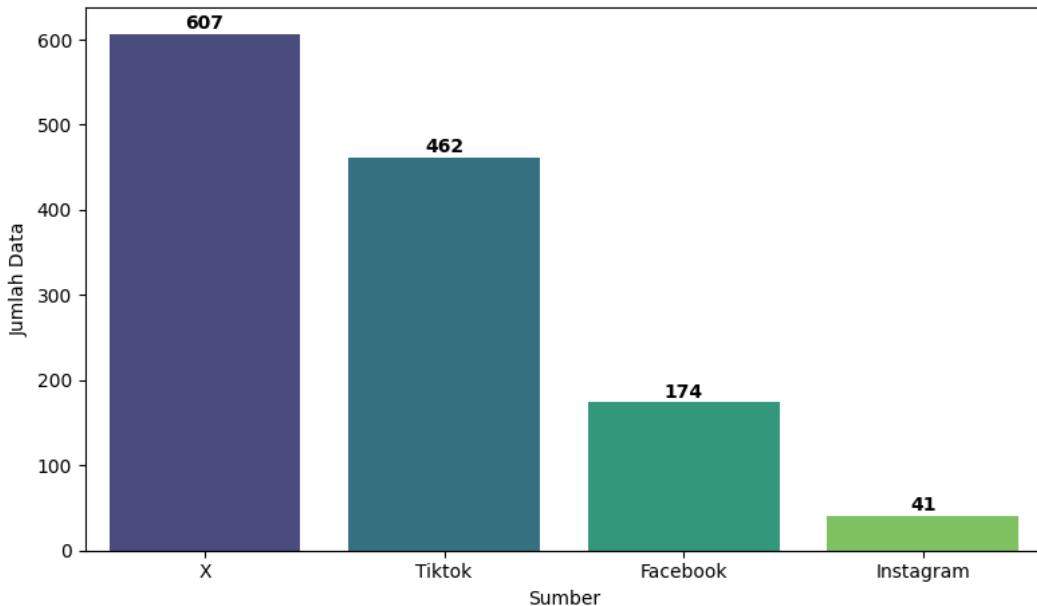


Gambar 1. Flowchart Tahapan Penelitian (Mulai -> Pengumpulan Data -> Preprocessing -> K-Means Clustering -> Naive Bayes Classification -> Evaluasi -> Selesai)

Proses *preprocessing* meliputi pembersihan data (hapus duplikat, URL, simbol), *case folding*, tokenisasi, penghapusan *stopwords*, dan *stemming*. Setelah data bersih, dilakukan klasterisasi menggunakan *K-Means* untuk melabeli data secara otomatis, kemudian diklasifikasikan menggunakan *Multinomial Naive Bayes*. Evaluasi model menggunakan *Confusion Matrix* untuk menghitung Akurasi, Presisi, *Recall*, dan *F1-Score* [13].

4. Hasil Penelitian Distribusi Data

Total data mentah yang dikumpulkan sebanyak 1.284 data. Distribusi data per platform ditunjukkan pada Gambar 2. Platform X (Twitter) menjadi kontributor data terbanyak.



Gambar 2. Grafik Distribusi Jumlah Data Per Sumber (Twitter mendominasi, diikuti TikTok, Instagram, dan Facebook)

Setelah melalui proses *preprocessing* (pembersihan duplikat dan *cleaning*), jumlah data bersih yang digunakan untuk analisis adalah 921 data.

Hasil Clustering K-Means

Metode *Elbow* digunakan untuk menentukan jumlah klaster optimal, yang menghasilkan nilai $k=3\$$. Hasil klasterisasi membagi data menjadi tiga sentimen:

- Klaster 0 (Negatif):** Dominan dengan 750 data. Kata kunci: tambang, rusak, oligarki.
- Klaster 1 (Positif):** 28 data. Kata kunci: semangat, salut, dukung.
- Klaster 2 (Netral):** 142 data. Kata kunci: pembangunan, jalan, wilayah.

Tabel 1 menampilkan contoh data hasil pelabelan sentimen berdasarkan klaster.

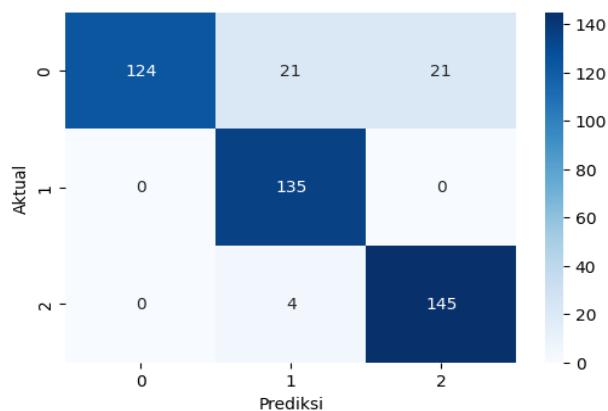
Tabel 1. Contoh Komentar Berdasarkan Klaster Sentimen

Sentimen	Contoh Komentar (Telah diproses)
Negatif	"hancur sudah itu lama lama hutan habis babat habis sda"
Positif	"semangat abang mislan wakil rakyat perjuangkan hak taliabu"
Netral	"taliabu daerah indah butuh perubahan jalan pendidikan kesehatan"

Hasil Klasifikasi Naive Bayes

Karena ketidakseimbangan data (*imbalanced data*) dimana sentimen negatif sangat dominan, dilakukan teknik *Oversampling* sebelum pelatihan model. Data dibagi menjadi 80% data latih dan 20% data uji.

Hasil pengujian model *Multinomial Naive Bayes* divisualisasikan menggunakan *Confusion Matrix* pada Gambar 3.

**Gambar 3.** Confusion Matrix Hasil Pengujian Naive Bayes

Berdasarkan *confusion matrix*, kinerja model dievaluasi dan dirangkum dalam Tabel 2.

Tabel 2. Metrik Evaluasi Hasil Pengujian Multinomial Naive Bayes

Kelas	Precision	Recall	F1-Score
0 (Netral)	1.00	0.75	0.86
1 (Positif)	0.84	1.00	0.92
2 (Negatif)	0.87	0.97	0.92
Akurasi Total	89.77%		

Model menghasilkan akurasi keseluruhan sebesar 89,77%. Nilai *F1-Score* yang tinggi (>0.85) pada semua kelas menunjukkan bahwa model mampu mengklasifikasikan sentimen dengan baik meskipun pada data yang awalnya tidak seimbang [14], [15]. Sentimen negatif yang dominan mencerminkan kekhawatiran masyarakat terhadap dampak lingkungan dan sosial akibat pertambangan.

5. Kesimpulan

Berdasarkan analisis yang dilakukan, sentimen masyarakat terhadap dampak pertambangan di Pulau Taliabu cenderung negatif (dominan 750 data), diikuti sentimen netral (142 data), dan positif (28 data). Isu utama yang diangkat meliputi kerusakan lingkungan dan dampak sosial. Implementasi metode *K-Means* efektif untuk pelabelan otomatis, dan *Multinomial Naive Bayes* terbukti handal dalam klasifikasi dengan akurasi 89,77%.

6. Daftar Pustaka

- [1] F. A. Harahap, "Penentuan sektor basis di Indonesia," Doctoral dissertation, UIN Syekh Ali Hasan Ahmad Addary Padangsidimpuan, 2022.
- [2] H. B. Aji dan A. Teapon, "Pengaruh batuan induk dan kimia tanah terhadap potensi kesuburan tanah di Kabupaten Kepulauan Sula, Provinsi Maluku Utara," *Jurnal Pengkajian dan Pengembangan Teknologi Pertanian*, vol. 22, no. 3, hal. 343-353, 2019.
- [3] D. Fauzah, N. Nurany, dan F. Firman, "Evaluasi kemajuan tambang triwulan ke IV (Oktober-Desember) 2022 pada tambang bijih besi PT. Bintani Megahindah Kabupaten Pulau Taliabu Provinsi Maluku Utara," *Jurnal Teknologi Sumberdaya Mineral*, vol. 4, no. 2, hal. 45-52, 2023.
- [4] E. Lovenda, "Kajian ekonomi lingkungan: Biaya eksternal dalam aktivitas industri pertambangan," *Circle Archive*, vol. 1, no. 7, 2025.
- [5] R. Yusuf, "Implementasi algoritma Support Vector Machine untuk pengaruh emosi dasar pada sentimen ulasan film," Doctoral dissertation, Universitas Buddhi Dharma, 2023.
- [6] H. Setiawan, E. Utami, dan S. Sudarmawan, "Analisis sentimen Twitter kuliah online pasca Covid-19 menggunakan algoritma Support Vector Machine dan Naive Bayes," *Jurnal Komtika (Komputasi dan Informatika)*, vol. 5, no. 1, hal. 43-51, 2021.
- [7] R. L. Atimi dan E. E. Pratama, "Implementasi model klasifikasi sentimen pada review produk Lazada Indonesia," *Jurnal Sains dan Informatika*, vol. 8, no. 1, hal. 88-96, 2022.
- [8] S. Wulandari dan F. N. Hasan, "Analisis sentimen masyarakat Indonesia terhadap pengalaman belanja thrifting pada media sosial Twitter menggunakan algoritma Naïve Bayes," *Jurnal Media Informatika Budidarma*, vol. 8, no. 2, hal. 768-776, 2024.
- [9] T. Wibisino dan Y. S. Mulyani, "Analisis dampak penggunaan media sosial terhadap prestasi akademik pelajar tingkat Sekolah Menengah Pertama," *Jurnal Ekonomi Manajemen*, vol. 4, no. 1, hal. 1-7, 2018.

- [10] A. F. Firdaus dan W. I. Firdaus, "Text Mining dan pola algoritma dalam penyelesaian masalah informasi: (Sebuah ulasan)," *JUPITER (Jurnal Penelitian Ilmu dan Teknologi Komputer)*, vol. 13, no. 1, hal. 66-78, 2021.
- [11] H. Gunawan dan V. Purwayoga, "Data mining menggunakan algoritma K-Means Clustering untuk mengetahui potensi penyebaran virus corona di Kota Cirebon," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 11, no. 1, hal. 1-8, 2022.
- [12] S. K. Wardani, Y. A. Sari, dan I. Indriati, "Analisis sentimen menggunakan metode Naïve Bayes Classifier terhadap review produk perawatan kulit wajah menggunakan seleksi fitur N-gram dan document frequency thresholding," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 12, hal. 5582-5590, 2021.
- [13] Y. O. Sihombing dan N. V. Situmorang, "Prediksi Sentimen Pada Teks Media Sosial Corporate University Menggunakan RoBERTa," *Prosiding PITNAS Widyaaiswara*, vol. 1, hal. 302-316, 2024.
- [14] N. Norlaila, W. W. Winarno, dan E. T. Luthfi, "Analisis Sentimen Masyarakat Tentang Tambang Di Indonesia Pada Twitter Menggunakan Data Mining," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 3, hal. 1091-1099, 2024.
- [15] T. Y. Pahtoni dan H. Jati, "Analisis sentimen data Twitter terkait ChatGPT menggunakan Orange Data Mining," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 2, hal. 329-336, 2024.