

ANALISIS SENTIMEN MASYARAKAT TENTANG TAMBANG DI INDONESIA PADA TWITTER MENGGUNAKAN DATA MINING

Norlaila*¹⁾, Wing Wahyu Winarno²⁾, Emha Taufiq Luthfi³⁾

1. Universitas Amikom Yogyakarta, Indonesia
2. Universitas Amikom Yogyakarta, Indonesia
3. Universitas Amikom Yogyakarta, Indonesia

Article Info

Kata Kunci: Data Mining, K-Means, Naïve Bayes, Sentimen Analisis, Tambang Emas, Twitter

Keywords: Data Mining, Gold Mining, K-Means, Naïve Bayes Classifier, Sentiment Analysis, Twitter

Article history:

Received 2 June 2024

Revised 20 July 2024

Accepted 4 August 2024

Available online 1 September 2024

DOI :

<https://doi.org/10.29100/jipi.v9i3.5402>

* Corresponding author.

Norlaila

E-mail address:

norlaila@students.amikom.ac.id

ABSTRAK

Penelitian ini menyelidiki sentimen masyarakat terhadap pertambangan emas di Indonesia melalui Twitter, menggunakan clustering K-Means dan Naïve Bayes Classifier untuk analisis sentimen. Mengingat pertambangan batu bara menjadi isu yang kontroversial, mengukur opini publik sangatlah penting untuk memahami dampak sosial dan mendorong dialog antar pemangku kepentingan. Penelitian ini melibatkan pengumpulan data Twitter, diikuti dengan pra-pemrosesan untuk mempersiapkan analisis. Algoritme K-Means mengidentifikasi tiga kelompok sentimen: netral (1561 tweet), positif (202 tweet), dan negatif (631 tweet). Selanjutnya, Pengklasifikasi Naïve Bayes, yang diterapkan pada set pelatihan yang terdiri dari 1.348 tweet dan set pengujian yang terdiri dari 1.046 tweet, selanjutnya mengkatégorikan sentimen menjadi 324 tweet negatif, 40 netral, dan 682 tweet positif. Metodologi tersebut mencapai akurasi gabungan sebesar 99%, yang keefektifannya dalam mengekstraksi dan menganalisis sentimen publik dari data media sosial. Wawasan-wawasan ini adalah sangat berharga bagi para pembuat kebijakan, pencinta lingkungan, dan perusahaan, memandu keputusan yang lebih tepat dan mempromosikan praktik-praktik berkelanjutan di sektor di sektor pertambangan emas. Penelitian ini menggarisbawahi potensi penambangan data data dalam menangkap nuansa opini publik tentang isu-isu lingkungan dan ekonomi, yang berkontribusi pada pengelolaan sumber daya yang lebih inklusif dan bertanggung jawab.

ABSTRACT

This study investigates public sentiment on gold mining in Indonesia through Twitter, employing K-Means clustering and Naïve Bayes Classifier for sentiment analysis. With gold mining being a contentious issue, gauging public opinion is crucial for understanding societal impacts and fostering dialogue among stakeholders. The research involved collecting Twitter data, followed by pre-processing to prepare for analysis. The K-Means algorithm identified three sentiment groups: neutral (1561 tweets), positive (202 tweets), and negative (631 tweets). Subsequently, the Naïve Bayes Classifier, applied to a training set of 1348 tweets and a test set of 1046 tweets, further categorized sentiments into 324 negative, 40 neutral, and 682 positive tweets. The methodologies achieved a combined accuracy of 99%, demonstrating their effectiveness in extracting and analyzing public sentiment from social media data. These insights are invaluable for policymakers, environmentalists, and corporations, guiding more informed decisions and promoting sustainable practices in the gold mining sector. This research underscores the potential of data mining techniques in capturing the nuances of public opinion on environmental and economic issues, contributing to more inclusive and responsible resource management.

I. PENDAHULUAN

INDONESIA, negara yang kaya akan sumber daya alam, telah lama memikat perhatian dunia. Keberadaan sumber daya alam yang melimpah, termasuk yang terbarukan dan tidak terbarukan, menjadi salah satu alasan negara ini dijajah di masa lalu [1], [2]. Emas, sebagai salah satu sumber daya alam yang berlimpah di Indonesia, dinilai sangat prospektif dari sudut pandang ekonomi karena nilai jualnya yang terus meningkat dan kemudahan dalam

transaksi jual beli [3]. Data dari Kementerian Energi dan Sumber Daya Mineral menunjukkan bahwa cadangan emas global mencapai 50.300 ton, dengan Indonesia termasuk dalam lima besar negara pemilik cadangan emas terbesar di dunia, mencapai 5% atau sekitar 2.600 ton, dengan Papua sebagai pulau dengan cadangan terbesar di negara ini (52%) [4]

Aktivitas penambangan emas yang diizinkan pemerintah tentunya bertujuan untuk kesejahteraan bersama, meskipun terdapat hambatan sosial yang muncul, seperti yang terjadi di tambang emas Tumpang Pitu di Pesanggaran, Banyuwangi, dan di Silo, Jember [5], [6]. Konflik serupa juga terjadi di area penambangan PT. Freeport Indonesia, yang di satu sisi memberikan lapangan kerja luas namun di sisi lain menimbulkan kerusakan lingkungan [1], [7]. Dampak negatif ini mencakup kerusakan fisik lingkungan, polusi suara, pencemaran air dan tanah, serta penurunan produktivitas lahan pertanian dan keanekaragaman hayati [2], [8]

Di era digital, opini masyarakat terkait perusahaan atau isu tertentu kini mudah ditemukan di media sosial, termasuk Twitter, yang menjadi platform penting dalam interaksi sosial dan analisis sentimen [9], [10], [11]. Teknologi informasi, khususnya analisis sentimen, memungkinkan perusahaan untuk menggali pandangan masyarakat dan mengidentifikasi area perbaikan untuk meyakinkan investor atau stakeholder lainnya [12], [13].

Penelitian ini akan menggunakan metode K-Means dan Naïve Bayes Classifier dalam analisis sentimen terkait tambang emas di Indonesia berdasarkan data Twitter. Kedua metode ini dipilih karena efektivitasnya dalam mengklasifikasikan data teks ke dalam kategori sentimen yang berbeda dan telah terbukti memberikan akurasi yang tinggi dalam penelitian sebelumnya [14], [15]. Penelitian ini bertujuan untuk mencakup ekstraksi dan klasifikasi sentimen menjadi positif, negatif, dan netral, evaluasi akurasi dan efisiensi kedua metode, serta analisis kecocokan mereka untuk data media sosial tentang isu penambangan emas. Melalui penelitian ini, diharapkan dapat memberikan pemahaman yang lebih mendalam tentang persepsi publik terhadap penambangan emas dan menawarkan insight tentang pendekatan analisis sentimen yang paling efektif. Penelitian ini tidak hanya berkontribusi pada pemahaman opini publik terhadap isu penambangan emas namun juga pada metodologi analisis sentimen untuk data sosial media, selain itu juga memberikan wawasan praktis bagi pemangku kepentingan di bidang pertambangan, kebijakan publik, dan konservasi lingkungan tentang cara terbaik untuk memahami dan menavigasi opini publik mengenai isu penambangan emas di Indonesia. Ini membuka peluang untuk dialog yang lebih informasi dan bertanggung jawab antara perusahaan penambangan, pemerintah, dan masyarakat.

Penelitian-penelitian yang dilakukan dalam bidang analisis sentimen menunjukkan pemanfaatan berbagai metode dalam menggali informasi dari data sosial media untuk berbagai tujuan. Pemanfaatan analisis sentimen pada penelitian [16] untuk memprediksi produk di masa akan datang berdasarkan hasil review dari pelanggan. Selain itu, analisis sentimen menggunakan metode *backpropagation* pada tweet mahasiswa [17], ditemukan bahwa mayoritas emosi yang diungkapkan cenderung netral (53,33%), diikuti oleh emosi positif (33,33%) dan emosi negatif (13,33%). Ini menunjukkan bahwa walaupun ada pandangan positif dan negatif, mahasiswa cenderung mengekspresikan sentimen yang lebih netral dalam konteks yang dianalisis. Analisis sentimen juga telah diterapkan di berbagai domain, dan salah satu domain yang paling populer adalah media sosial [18]. Twitter merupakan salah satu platform media sosial terpopuler, memiliki pengguna dari berbagai daerah dengan beragam budaya dan bahasa. Hal ini dapat memberikan informasi berharga untuk data yang beragam dan berjumlah besar untuk digunakan guna meningkatkan pengambilan keputusan [19].

Dalam konteks pemasaran, penelitian oleh [20] menggunakan algoritma K-Means *Clustering* untuk mengkategorikan produk busana muslim berdasarkan tingkat penjualannya. Hasilnya membantu dalam mengoptimalkan persediaan barang dan strategi promosi dengan mengidentifikasi produk yang kurang laris, cukup laris, dan sangat laris. Ini menunjukkan efektivitas K-Means dalam membantu pengambilan keputusan di sektor ritel.

Penelitian oleh [21] terkait analisis sentimen Pilpres 2019 menggunakan K-Means dan *Naïve Bayes Classifier* menunjukkan bahwa kombinasi kedua metode tersebut mampu menghasilkan akurasi tinggi (93,35%) dalam menganalisis sentimen masyarakat terhadap dua kubu politik, menekankan potensi analisis sentimen dalam konteks politik untuk memahami preferensi dan pandangan masyarakat.

Penelitian yang dilakukan oleh [22] menyebutkan bahwa metode Naïve Bayes memiliki tingkat akurasi 83,43% dalam melakukan analisis sentimen menggunakan informasi dari media sosial Twitter.

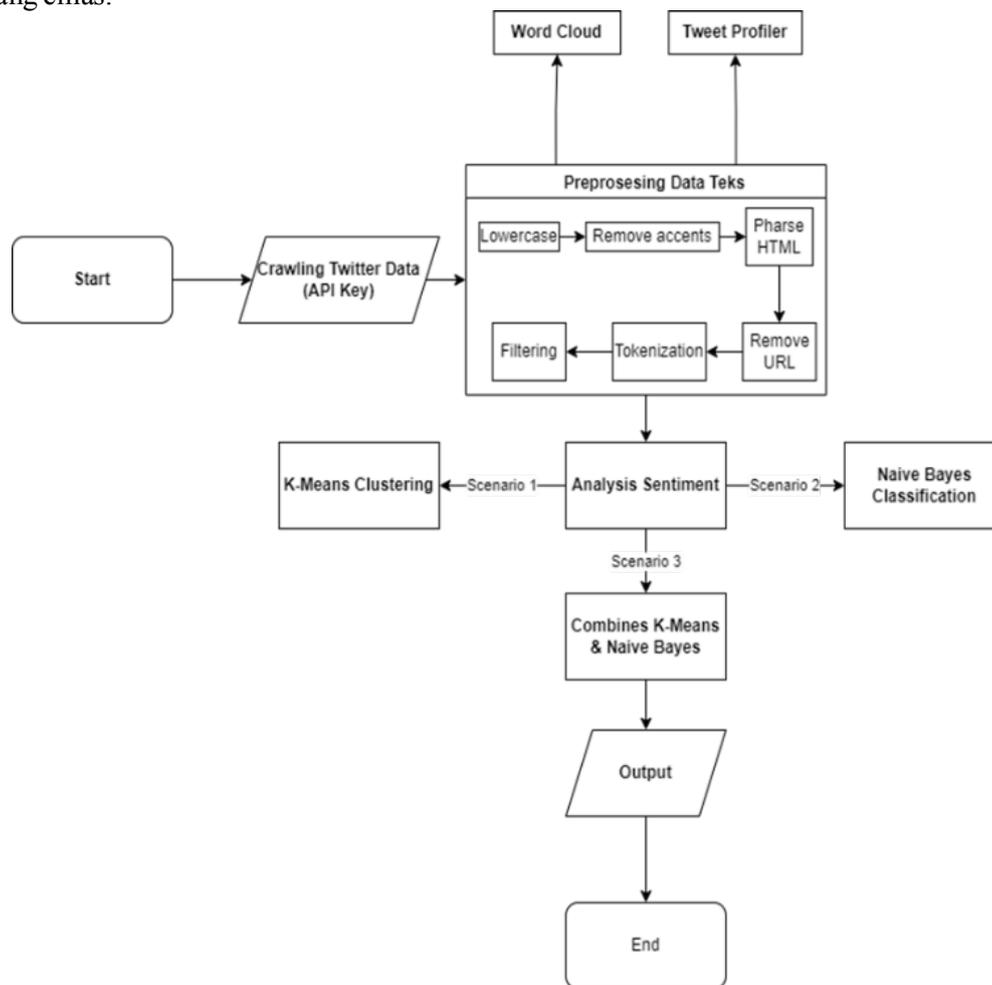
Pada penelitian [23] mengimplementasikan K-Means *Clustering* untuk mengeksplorasi komentar pada fanpage Shopee Indonesia, menemukan bahwa komentar terbanyak berkaitan dengan kegiatan promosi seperti kuis hadiah. Hal ini menunjukkan bahwa interaksi di media sosial sangat dipengaruhi oleh konten promosi, memberikan insight bagi strategi pemasaran digital.

Secara keseluruhan, penelitian-penelitian ini menegaskan potensi metode data mining seperti K-Means dan Naïve Bayes dalam menggali wawasan dari data besar, baik untuk keperluan akademis, bisnis, maupun politik.

II. METODE PENELITIAN

Penelitian ini dilakukan untuk mengetahui sentimen masyarakat di sosial media Twitter dengan API Twitter

terhadap tambang emas di Indonesia dengan data sebanyak 5000 cuitan, kemudian dilakukan pengujian tingkat akurasi metode K-Means dan Naïve Bayes Classifier pada analisis sentimen tersebut dengan kata kunci tambang emas dan tambang emas.



Gambar 1 Tahapan Penelitian

A. Crawling Twitter (API Keys)

Data yang digunakan pada penelitian ini adalah data yang didapatkan di platform Twitter dengan menggunakan API Key Twitter, kemudian kata kunci yang cari dimasukkan pada tahapan ini seperti kata emas, tambang, dan tambang emas. Data awal tweet yang akan dipakai adalah 5000 data terbaru yang ada.

B. Preprocessing Data Teks

Pada bagian ini akan dilakukan beberapa tahapan, yaitu menjadikan semua *tweet* menjadi huruf kecil, menghilangkan *link*, menghilangkan simbol-simbol, Tokenisasi, *Filtering* berupa *stopword* dengan kamus kata yang telah disesuaikan.

C. World Cloud

Word cloud adalah salah satu algoritma *Machine Learning* [24]. *Word Cloud* merupakan salah satu metode untuk memvisualisasikan data teks secara visual. *Word Cloud* populer dalam *text mining* karena mudah dipahami. Dengan menggunakan *word cloud*, gambaran frekuensi kata-kata dapat ditampilkan dalam bentuk yang menarik namun tetap informatif [25].

D. Tweet Profiler

Tweet Profiler merupakan salah satu widget pada aplikasi data mining yang bertujuan untuk mengambil data tentang sentimen dari server untuk setiap *tweet* (ataupun dokumen) yang diberikan [26].

E. Analisis Sentimen

Analisis sentimen pada penelitian ini merupakan salah satu tahap untuk melihat opini yang didapat dari data kicauan Twitter yang telah dikumpulkan sebelumnya. Proses analisis menggunakan algoritma vader untuk mengklasifikasi popularity/class sentimen yaitu positif, negatif, dan neutral. Vader akan mengkategorikan, dan

berikan skor text bersumber pada nilai tiap kata–kata yang terdaftar dalam lexicon vader. Hasil akhir dari evaluasi merupakan skor total ialah compound. Skor total inilah yang hendak direkapitulasi serta dibanding hasilnya [26].

F. Implementasi Algoritma K-Means dan Naïve Bayes

Pada bagian ini peneliti akan melakukan implementasi Algoritma K-Means dan Naïve Bayes. Penelitian akan melakukan klusterisasi terlebih dahulu menggunakan K-Means, kemudian melakukan klasifikasi secara terpisah dengan Algoritma Naïve Bayes. Setelah melihat hasil kedua algoritma tersebut, peneliti akan menggabungkan dua algoritma K-Means dan Naïve Bayes. Karena berdasarkan dari penelitian [14] menyebutkan bahwa menggabungkan dua algoritma ini akan meningkatkan akurasi hasil yang didapat. Penggabungan 2 metode ini akan dilakukan dengan menggunakan aplikasi Orange Data Mining yang memiliki dasar dari bahasa pemrograman *python*.

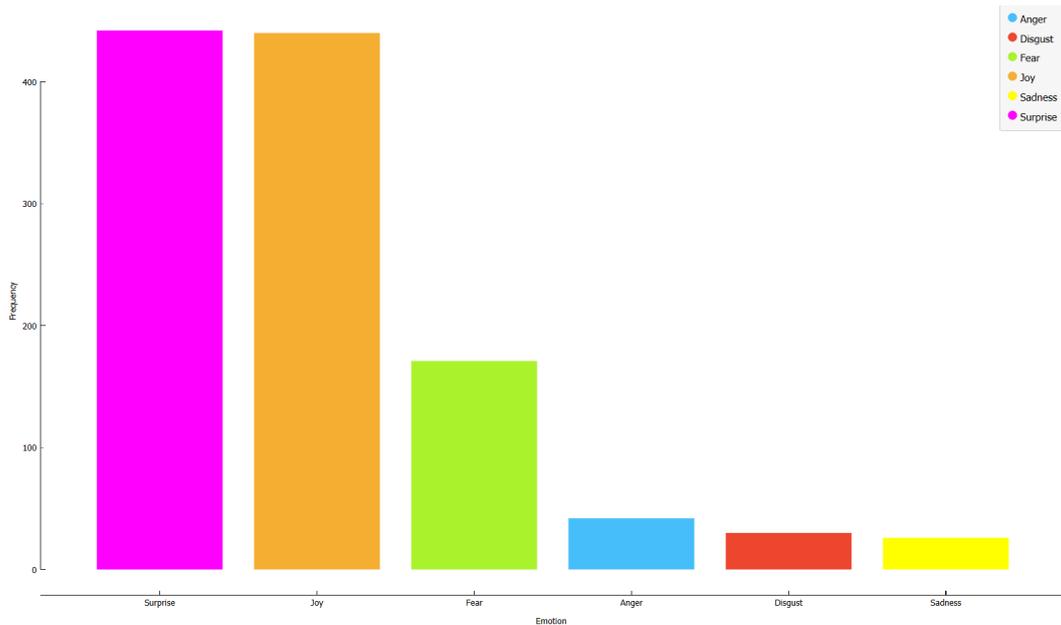
III. HASIL DAN PEMBAHASAN

Implementasi dilakukan secara bertahap. Pertama akan dilakukan dengan menggunakan Algoritma K-Means Clustering kemudian menggunakan klasifikasi Naïve Bayes. Berikut contoh data yang dikumpulkan melalui Twitter pada penelitian ini:

TABEL 1
 CONTOH DATA TWITTER

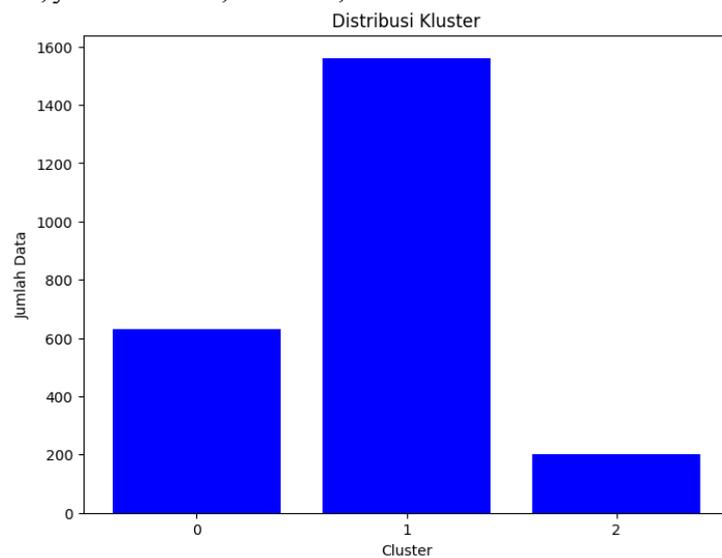
tweet_text	tweet_date	user_name
RT @jtdtodayco: Rugikan Negara Sebesar Rp 5,7 triliun, Eks Relawan Jokowi Ditetapkan Sebagai Tersangka Tambang Nikel Ilegal di Konawe Utaraâ€¦ Kapolsek Lamala Hadiri Rapat Penghentian Sementara Aktivitas Tambang PT. ATN â€¢ â€¢ â€¢ #polri #humaspoli #divisihumaspoli #bidhumaspoldasulteng #poldasulteng #polresbanggai https://t.co/vzFUJshdtd	Thu Jul 20 06:11:03+0000 2023	Kyai Bonorejo
RT @convomf: ðŸ™ GAES INI WARNA EMAS PUTIH ATAU HITAM BIRU? https://t.co/uAVIEGEMPC â€¢Boleh gadai aja kat Arrahnuâ€¢ kata mak lagi. Sebab itu saya sangat percaya kepada emas fizikal gais. Mak saya bukti hidup ðŸ™	Thu Jul 20 06:10:41+0000 2023 Thu Jul 20 06:10:14+0000 2023	Polres Banggai mihyaðŸ™ LAGI LIMIT
#NotaEmasSyukorHashim â€¢Boleh gadai aja kat Arrahnuâ€¢ kata mak lagi. Sebab itu saya sangat percaya kepada emas fizikal gais. Mak saya bukti hidup ðŸ™ Emas dah beli 33 tahun dia masih simpan rapat-rapat gais. Takde tanda-tanda nak kena jual gais ðŸ™	Thu Jul 20 06:10:10+0000 2023	fairuz_imran
#NotaEmasSyukorHashim RT @HMMurah3: Mantan Relawan Jokowi Jadi Tersangka Korupsi Tambang Nikel Ilegal Senilai Rp5,7 Triliun https://t.co/EqgmVHEXPL Menko Perekonomian Airlangga Hartarto membahas soal target Indonesia Emas 2045. Dia menyampaikan pertumbuhan ekonomi harus didorong di kisaran 6-7% #AirlanggaHartarto #IndonesiaEmas2045 https://t.co/wvnVcRgt8t	Thu Jul 20 06:09:13+0000 2023 Thu Jul 20 06:09:04+0000 2023	fairuz_imran YP20
Seorang penyelia syarikat didakwa di Mahkamah Majistret, di sini, hari ini kerana menyamar sebagai kontraktor Majlis Perbandaran Ampang Jaya (MPAJ) untuk menipu seorang pesara warga emas berhubung penurapan jalan. #KwikNews #Berita https://t.co/LPrHpYEMlW	Thu Jul 20 06:07:37+0000 2023 Thu Jul 20 06:07:34+0000 2023	Katadata.co.id KwikNews

Tahapan kedua setelah pengumpulan data adalah *Preprocessing* Data Teks untuk dilakukan normalisasi, filterisasi dan proses pembersihan data lainnya yang dapat digunakan untuk proses selanjutnya, berikut contoh hasil dari

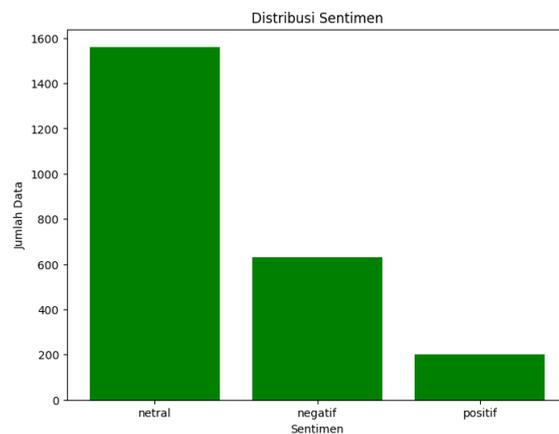


Gambar 3 Tweet Profiler

Untuk proses implementasi algoritma K-Means Clustering menggunakan tools Google Collabs dengan bahasa python. Berikut hasil dari klasterisasi dari algoritma K-Means Clustering dalam bentuk grafik gambar 2 di bawah ini yang terbagi menjadi 3 klaster, yaitu klaster 0, klaster 1, dan klaster 2.



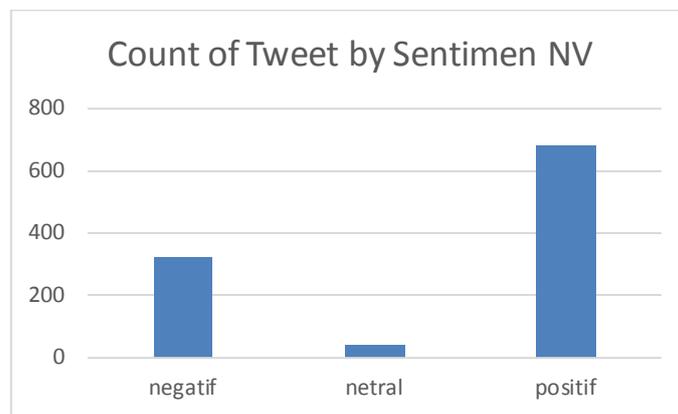
Gambar 4 Hasil K-Means Clustering



Gambar 5 Hasil K-Means Clustering berdasarkan Sentimen Analisis

Pada K-Means Clustering dilakukan beberapa pengujian, yaitu pengujian menggunakan *Silhouette Score*, *Inertia*, *Davies-Bouldin Index* dan pengujian dengan *Calinski-Harabasz Index*. Pengujian klustering dengan menggunakan metrik *Silhouette Score*, *Inertia*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index* menunjukkan hasil yang relatif baik dalam analisis klustering. *Silhouette Score* sebesar 0.23663621791951647 mengindikasikan bahwa klustering telah berhasil membentuk kelompok-kelompok dengan kesamaan yang cukup tinggi di dalam kluster dan perbedaan yang signifikan antara kluster, meski masih terdapat ruang untuk peningkatan, karena berdasarkan dari penelitian yang dilakukan oleh [27] menyebutkan bahwa rentang nilai indeks silhouette adalah -1 hingga 1. Nilai di bawah 0 menunjukkan penempatan objek yang tidak tepat, nilai di atas 0 menunjukkan penempatan yang tepat, dan nilai 0 menandakan objek berada di perbatasan antara dua kluster. *Inertia* dengan nilai 1706,1048684048876 menunjukkan kluster yang cukup padat dan mungkin optimal. Sementara itu, *Davies-Bouldin Index* sebesar 1,5529956947272685 yang mana hal ini sesuai dengan pernyataan pada penelitian [27] yang menyatakan bahwa jumlah kluster optimal ditunjukkan dengan nilai indeks *Davies Bouldin* yang lebih rendah. *Calinski-Harabasz Index* dengan nilai 352,93 yang mana nilai ini sudah tinggi, berdasarkan penelitian yang dilakukan oleh [28] yang menyatakan bahwa nilai CH yang lebih tinggi menunjukkan jumlah kluster yang lebih optimal. Keseluruhan hasil ini memberikan indikasi positif tentang efektivitas klustering yang dihasilkan oleh model dalam konteks analisis yang ditinjau.

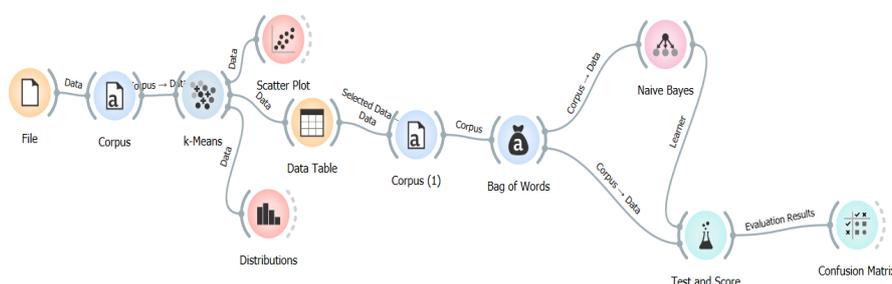
Implementasi Naïve Bayes data akan dibagi menjadi 2 bagian, yaitu data latih dan data uji. Pada bagian data latih, akan diisi terlebih dahulu sentimen pada data tweet yang sudah ada, kemudian pada data uji akan dilakukan implementasi Naïve Bayes untuk melihat hasil dari penggunaannya algoritma tersebut. Implementasi Naïve Bayes dilakukan menggunakan aplikasi Orange Data Mining.



Gambar 6 Hasil Implementasi Naïve Bayes menggunakan Orange Data Mining

Dari pengujian model Naïve Bayes menggunakan Cross Validation dengan 5 fold, diperoleh nilai AUC sebesar 0,716, yang menandakan kemampuan model yang cukup baik dalam membedakan antara kelas positif dan negatif, dengan AUC mendekati 1 menunjukkan performa yang lebih baik. Nilai sebesar 0,577 menunjukkan akurasi model dalam membuat prediksi yang benar sebesar 57,7%. F1-Score sebesar 0,567 mencerminkan keseimbangan antara precision dan recall, menunjukkan model cukup efektif dalam mengelola trade-off antara kedua metrik tersebut. Precision sebesar 0,574 menunjukkan tingkat keberhasilan model dalam mengidentifikasi prediksi positif yang akurat, sementara recall sebesar 0,577 menunjukkan kemampuan model dalam mengidentifikasi kasus positif secara keseluruhan. MCC sebesar 0,356 menunjukkan korelasi yang moderat antara prediksi dan nilai sebenarnya, dengan nilai +1 merepresentasikan prediksi sempurna. Kesimpulannya, model menunjukkan kinerja yang memadai dalam klasifikasi, dengan ruang untuk peningkatan dalam beberapa aspek untuk mencapai prediksi yang lebih akurat dan reliabel.

Pada Implementasi dari Algoritma K-Means Clustering dengan Naïve Bayes dengan menggunakan Orange Data Mining.



Gambar 7 Model K-Means dan Naïve Bayes menggunakan Orange Data Mining

Model diatas dijalankan secara bertahap, yang mana akan dilakukan klusterisasi dari metode K-Means, kemudian dilakukan klasifikasi dengan menggunakan Naïve Bayes. Berdasarkan model pada gambar 5 tersebut, didapatkan data diambil sample sebanyak 80% dari total data yang ada dengan 20 kali perulangan, sehingga menghasilkan AUC bernilai 1,000 yang berarti model sudah sangat baik karena berhasil memisahkan kelas data. Kemudian untuk nilai CA atau tingkat akurasi yang menjadi 0,99 atau 99%. Nilai F1 0,99 memiliki trade-off yang baik antara precision dan recall. Prec juga menunjukkan angka 0,99 yang berarti Precision mengukur persentase prediksi positif yang benar dari semua prediksi positif. Untuk nilai Recall 0,99 menunjukkan bahwa model hampir selalu berhasil menemukan data yang seharusnya positif. Selain itu untuk nilai MCC 0,98 menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam memprediksi sentimen dengan mempertimbangkan semua aspek. Kedua metode ini, ketika digabungkan, memberikan kerangka kerja yang kuat untuk analisis sentimen, memungkinkan penelitian ini untuk mencapai pemahaman yang mendalam dan akurat tentang sentimen masyarakat terhadap tambang emas di Indonesia. Dengan demikian, integrasi metode ini tidak hanya meningkatkan akurasi analisis tetapi juga memberikan wawasan komprehensif tentang dinamika sentimen masyarakat.

Selain itu dengan variasi jumlah data yang dicoba pada model terlihat ada beberapa perbedaan, berikut tabel variasi data yang dicoba untuk mengetahui kinerja algoritma pada gambar 7.

TABEL 3
VARIASI DATA UNTUK MENGUJI ALGORITMA PADA MODEL

No.	Jumlah Data	AUC	CA	F1	Prec	Recall	MCC
1	500	1.000	0.888	0.914	0.966	0.888	0.518
2	1200	1.000	1.000	1.000	1.000	1.000	1.000
3	1819	1.000	0.992	0.993	0.993	0.992	0.981
4	2394	1.000	0.990	0.990	0.990	0.990	0.980

Tabel 3 menyajikan hasil evaluasi model dengan menggunakan berbagai metrik performa terhadap empat kumpulan data berbeda yang memiliki jumlah data yang beragam, yaitu 500, 1200, 1819, dan 2394. Berdasarkan nilai AUC (Area Under the Curve) yang konsisten mencapai 1.000 untuk semua kumpulan data, menunjukkan bahwa model memiliki kemampuan sempurna dalam membedakan antara kelas positif dan negatif tanpa kesalahan. Ini adalah indikator yang sangat kuat dari performa model yang luar biasa dalam konteks klasifikasi.

Pada kumpulan data dengan 500 entri, model mencatatkan Classification Accuracy (CA) sebesar 0.888, F1-Score sebesar 0.914, Precision (Prec) 0.966, Recall 0.888, dan Matthews Correlation Coefficient (MCC) sebesar 0.518. Meskipun nilai AUC sempurna, metrik lainnya menunjukkan ruang untuk peningkatan, khususnya nilai MCC yang relatif lebih rendah dibandingkan dengan kumpulan data lainnya. Ini mengindikasikan bahwa pada kumpulan data yang lebih kecil, model mungkin mengalami kesulitan dalam memprediksi kelas dengan tingkat kepercayaan yang tinggi.

Untuk kumpulan data dengan 1200 entri, semua metrik evaluasi (CA, F1, Prec, Recall, dan MCC) mencapai nilai sempurna 1.000, menunjukkan bahwa model tidak hanya sempurna dalam memisahkan kelas tapi juga memiliki akurasi, presisi, sensitivitas, dan korelasi prediksi terhadap nilai sebenarnya yang sangat tinggi. Hal ini menunjukkan peningkatan signifikan dibandingkan dengan kumpulan data yang lebih kecil.

Saat jumlah data meningkat menjadi 1819 dan 2394, model tetap menunjukkan performa yang sangat baik dengan nilai CA, F1, Prec, dan Recall yang sangat tinggi, yaitu sekitar 0.990 hingga 0.993, dan nilai MCC yang juga tinggi, yaitu 0.981 dan 0.980. Ini menunjukkan bahwa model sangat konsisten dalam performanya, bahkan ketika jumlah data bertambah, menunjukkan skalabilitas yang baik.

Kesimpulannya, analisis dari tabel 3 menunjukkan bahwa performa model secara keseluruhan sangat mengesankan, khususnya pada dataset dengan jumlah data yang lebih besar. Model ini memiliki kemampuan luar biasa dalam klasifikasi sentimen, dengan performa yang meningkat secara signifikan ketika jumlah data yang diolah meningkat, yang ditunjukkan oleh peningkatan nilai MCC. Ini menunjukkan pentingnya jumlah data dalam meningkatkan kualitas prediksi model.

IV. KESIMPULAN

Berdasarkan hasil yang dilakukan, penggunaan teknik K-Means Clustering dalam pengelompokan sentimen dari data Twitter berhasil mengidentifikasi tiga kluster yang signifikan, yakni netral, positif, dan negatif, dengan distribusi yang beragam antar kluster. Kluster netral menonjol dengan jumlah data terbanyak sebanyak 1561, diikuti oleh kluster negatif dengan 631 data, dan kluster positif yang terdiri dari 202 data. Selanjutnya, penerapan algoritma Naïve Bayes Classifier pada data yang sama, dengan membagi data menjadi 1348 tweet untuk data latih dan 1046 tweet untuk data uji, menghasilkan klasifikasi sentimen yang lebih detail dengan jumlah data uji negatif sebanyak 324, netral 40, dan positif 682. Integrasi kedua algoritma ini, K-Means Clustering dan Naïve Bayes Classifier, dalam analisis sentimen menunjukkan peningkatan signifikan dalam akurasi klasifikasi, mencapai 99%. Temuan ini menegaskan efektivitas gabungan kedua metode dalam analisis sentimen, dimana K-Means Clustering efektif dalam

pengelompokan data berdasarkan kesamaan karakteristik sentimen, dan Naïve Bayes Classifier memperkuat hasil klasifikasi dengan akurasi yang tinggi. Hal ini menunjukkan potensi besar penggunaan metode hibrid dalam analisis sentimen untuk data sosial media, khususnya dalam konteks pengelolaan dan pemahaman opini publik secara lebih akurat dan efisien.

DAFTAR PUSTAKA

- [1] A. Saputra, P. Dita, and H. Micelyn, "DAMPAK BURUK AKTIVITAS PERTAMBANGAN PT. FREEPORT INDONESIA," Yogyakarta, Jun. 2017. Accessed: Nov. 09, 2022. [Online]. Available: https://www.academia.edu/35965184/DAMPAK_BURUK_AKTIVITAS_PERTAMBANGAN_PT_FREEPORT_INDONESIA#:~:text=Dampak%20tersebut%20di%20antaranya%20adalah%20timbulnya,akan%20keberadaan%20Freeport%20di%20Papua.
- [2] L. A. Wowling, A. J. Rorong, and N. N. Plangiten, "DAMPAK KEBIJAKAN PERTAMBANGAN EMAS DI DESA RATATOTOK SELATAN MINAHASA TENGGARA," *Jurnal Administrasi Publik*, vol. 7, no. 106, pp. 1–11, 2021, Accessed: Nov. 09, 2022. [Online]. Available: <https://ejournal.unsrat.ac.id/index.php/JAP/article/view/34885>
- [3] M. I. Ali, *Penataan Infrastruktur dan Penambang Emas Tradisional*, 1st ed., vol. 1. Makassar: Global Research and Consulting Institute (Global-RCI), 2017.
- [4] K. ESDM, "Peluang Investasi Emas-Perak di Indonesia," Jakarta, 2020. Accessed: Nov. 09, 2022. [Online]. Available: <https://www.esdm.go.id/id/booklet/booklet-tambang-emas-perak-2020>
- [5] C. E. Yunita, "KONFLIK TAMBANG EMAS TUMPANGPITU. DESA SUMBER AGUNG. PESANGGARAN, BANYUWANGI, JAWA TIMUR Oleh," *Jurnal Pendidikan Sosiologi*, vol. 7, no. 3, pp. 1–16, 2018, Accessed: Nov. 09, 2022. [Online]. Available: <https://journalstudent.uny.ac.id/ojs/index.php/societas/article/view/12530>
- [6] M. Ramadhan, "MAQASID SYARI'AH DAN LINGKUNGAN HIDUP (Bahtsul Masa'il Sebagai Perlawanan Kaum Santri Terhadap Eksploitasi Pertambangan Emas di Silo Jember)," *ANALYTICA ISLAMICA*, vol. 21, no. 2, pp. 126–136, 2019, Accessed: Nov. 09, 2022. [Online]. Available: <http://jurnal.uinsu.ac.id/index.php/analytica/article/view/7076>
- [7] Y. Pakniany and F. N. Patty, "Conflict Accommodation Ideas Based on Custom Institutions Siram Sopi in Gold Mining Area of Romang Island, Indonesia," *Sodality: Jurnal Sosiologi Pedesaan*, vol. 7, no. 3, pp. 224–235, Dec. 2019, doi: 10.22500/sodality.v7i3.27836.
- [8] A. Y. Emmanuel, C. S. Jerry, and D. A. Dzighodi, "Review of Environmental and Health Impacts of Mining in Ghana," *J Health Pollut*, vol. 8, no. 17, pp. 43–52, Mar. 2018, doi: 10.5696/2156-9614-8-17-43.
- [9] E. Y. Hidayat, R. W. Hardiansyah, and A. Affandy, "Analisis Sentimen Twitter untuk Menilai Opini Terhadap Perusahaan Publik Menggunakan Algoritma Deep Neural Network," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 2, pp. 108–118, Sep. 2021, doi: 10.25077/teknosi.v7i2.2021.108-118.
- [10] T. Liedfray, F. J. Waani, and J. J. Lasut, "Peran Media Sosial Dalam Mempererat Interaksi Antar Keluarga Di Desa Esandom Kecamatan Tombatu Timur Kabupaten Minahasa Tenggara," *Jurnal Ilmiah Society*, vol. 2, no. 1, pp. 1–13, 2022.
- [11] E. A. Marwa and A. B. Kristanto, "Analisis Sentimen Pengungkapan Informasi Manajemen: Text Mining Berbasis Metode VADER," *Owner*, vol. 6, no. 3, pp. 2853–2864, Jul. 2022, doi: 10.33395/owner.v6i3.895.
- [12] A. Alsaedi and M. Z. Khan, "A study on sentiment analysis techniques of Twitter data," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 2, pp. 361–374, 2019, doi: 10.14569/ijacsa.2019.0100248.
- [13] A. S. Neogi, K. A. Garg, R. K. Mishra, and Y. K. Dwivedi, "Sentiment analysis and classification of Indian farmers' protest using twitter data," *International Journal of Information Management Data Insights*, vol. 1, no. 2, Nov. 2021, doi: 10.1016/j.ijime.2021.100019.
- [14] Z. M. Fadhil, "Hybrid of K-means clustering and naive Bayes classifier for predicting performance of an employee," *Periodicals of Engineering and Natural Sciences ISSN*, vol. 9, no. 2, pp. 799–807, 2021, doi: <http://dx.doi.org/10.21533/pen.v9i2.1898>.
- [15] M. I. Zul, F. Yulia, and B. Nurmalasari, "Social Media Sentiment Analysis Using K-Means and Naïve Bayes Algorithm," in *2nd International Conference on Electrical Engineering and Informatics (ICOn EEI 2018)*, Batam: Institute of Electrical and Electronics Engineers Inc., Oct. 2018, pp. 24–29.
- [16] P. Sasikala and L. M. I. Sheela, "Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS," *J Big Data*, vol. 7, no. 1, pp. 1–20, Dec. 2020, doi: 10.1186/s40537-020-00308-7.
- [17] R. Habibi, D. Budiyo Setyohadi, and Ernawati, "ANALISIS SENTIMEN PADA TWITTER MAHASISWA MENGGUNAKAN METODE BACKPROPAGATION," *Informatika*, vol. 12, no. 1, pp. 103–109, 2016, doi: <http://dx.doi.org/10.21460/inf.2016.121.462>.
- [18] A. Lighthart, C. Catal, and B. Tekinerdogan, "Systematic reviews in sentiment analysis: a tertiary study," *Artif Intell Rev*, vol. 54, no. 7, pp. 4997–5053, Oct. 2021, doi: 10.1007/s10462-021-09973-3.
- [19] A. Alrumaih, A. Al-Sabbagh, R. Alsabah, H. Kharufa, and J. Baldwin, "Sentiment Analysis of Comments in Social Media," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 6, pp. 5917–5922, Dec. 2020, doi: 10.11591/ijece.v10i6.pp5917-5922.
- [20] M. Syukron Nawawi, F. Sembiring, and A. Erfina, "Implementasi Algoritma K-Means Clustering Menggunakan Orange Untuk Penentuan Produk Busana Muslim Terlaris," in *Seminar Nasional Teknologi Informasi dan Komunikasi-2021*, Madiun, 2021, pp. 789–797.
- [21] I. Kurniawan and A. Susanto, "Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *Eksplora Informatika*, vol. 9, no. 1, pp. 1–10, Sep. 2019, doi: 10.30864/eksplora.v9i1.237.
- [22] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment Analysis of Social Media Twitter with Case of Anti-LGBT Campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest Algorithm," in *Procedia Computer Science*, Bandung: Information Systems Department School of Industrial Engineering, Jan. 2019, pp. 765–772. doi: 10.1016/j.procs.2019.11.181.
- [23] A. Sentiya and H. Suroyo, "ANALISIS TEXT CLUSTERING AKUN FANPAGE SHOPEE INDONESIA DENGAN KOMENTAR FOLLOWERS MENGGUNAKAN TOOLS ORANGE DATA MINING," in *Bina Darma Conference on Computer Science*, 2019, pp. 1055–1067.
- [24] S. Trihandaru, H. A. Parhusip, B. Susanto, and C. F. R. Putri, "Word Cloud of UKSW Lecturer Research Competence Based on Google Scholar Source," *Khazanah Informatika*, vol. 7, no. 1, pp. 52–59, Apr. 2021, doi: <https://doi.org/10.23917/khif.v7i2.13123>.
- [25] M. G. Pradana, "PENGUNAAN FITUR WORDCLOUD DAN DOCUMENT TERM MATRIX DALAM TEXT MINING," *Jurnal Ilmiah Informatika (JIF)*, vol. 8, no. 1, pp. 38–43, Mar. 2020, doi: <https://doi.org/10.33884/jif.v8i01.1838>.
- [26] S. I. Nurhafida and F. Sembiring, "ANALISIS TEXT CLUSTERING MASYARAKAT DI TWITER MENGENAI MCDONALD'S XSBTS MENGGUNAKAN ORANGE DATA MINING," in *SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika)*, Sukabumi, Aug 2021, pp. 28–35.
- [27] R. Rahmati and A. W. Wijayanto, "ANALISIS CLUSTER DENGAN ALGORITMA K-MEANS, FUZZY C-MEANS DAN HIERARCHICAL CLUSTERING," *JIKO (Jurnal Informatikan dan Komputer)*, vol. 5, no. 2, pp. 73–80, 2021.
- [28] E. Luthfi and A. W. Wijayanto, "Analisis Perbandingan Metode Hierarchical, K-Means, dan K-Medoids Clustering dalam Pengelompokan Indeks Pembangunan Manusia Indonesia," *JEB (Journal of Engineering and Business)*, vol. 17, no. 4, pp. 761–773, 2021, [Online]. Available: <http://journal.feb.unmul.ac.id/index.php/INOVASI>