ANALISIS SENTIMEN RENCANA PENDIRIAN PABRIK APPLE.INC DI BANDUNG MENGGUNAKAN ALGORITMA NAIVE BAYES

Heldi Akbar Kosasih, Ade Yuliana
Teknik Informatika, Politeknik TEDC Bandung
Jl. Pesantren km 2, Cibabat, Kota Cimahi, Jawa Barat, Indonesia
heldiakbark@student.poltektedc.ac.ic

ABSTRAK

Perkembangan teknologi digital telah mendorong peningkatan penggunaan smartphone seperti iPhone di Indonesia. Rencana pendirian pabrik Apple di Bandung memicu beragam respons masyarakat yang terekam melalui media sosial, khususnya platform X (Twitter). Variasi opini ini menunjukkan pentingnya pemetaan sentimen publik secara sistematis. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap isu tersebut dengan memanfaatkan algoritma Naïve Bayes dan metode ekstraksi fitur Term Frequency–Inverse Document Frequency (TF-IDF). Data diperoleh melalui scraping komentar Twitter, dilanjutkan dengan tahap preprocessing dan klasifikasi ke dalam tiga kategori: positif, netral, dan negatif. Penelitian ini mengikuti pendekatan CRISP-DM untuk memastikan alur analisis berjalan terstruktur. Hasil klasifikasi menunjukkan bahwa model cukup efektif dalam mengenali sentimen positif dan netral dengan F1-Score sebesar 0.66, namun belum mampu mengklasifikasikan sentimen negatif dengan baik (F1-Score = 0), serta menghasilkan akurasi sebesar 61.92%. Kelemahan ini diduga disebabkan oleh ketidakseimbangan distribusi data. Dengan hasil tersebut, model ini dapat dimanfaatkan sebagai alat bantu analisis opini publik, namun perlu perbaikan lebih lanjut terutama dalam representasi data negatif.

Kata kunci: *Analisis Sentimen Twitter, Pabrik Apple Bandung, Naïve Bayes, CRISP-DM, TF-IDF*

1. PENDAHULUAN

Perkembangan teknologi digital telah mendorong perubahan signifikan dalam berbagai aspek kehidupan manusia. Salah satu dampaknya terlihat pada meningkatnya penggunaan perangkat *smartphone*, yang menjadi bagian tak terpisahkan dari aktivitas sehari-hari. Peningkatan ini didorong oleh kemajuan internet yang telah menjadi katalis utama dalam memperluas adopsi teknologi digital di seluruh dunia [1]**.**

Smartphone sendiri merupakan perangkat pintar dengan kemampuan tinggi yang mendekati komputer, sehingga sering disebut sebagai komputer mini. Perangkat ini tidak hanya berfungsi sebagai alat komunikasi, tetapi juga dapat menjalankan aplikasi, mengakses internet, hingga mendukung pekerjaan dan hiburan pengguna [2]. Di Indonesia, dominasi pasar *smartphone* diduduki oleh produk-produk dari Apple Inc., seperti *iPhone, iPad,* dan *Apple Watch*, yang dikenal dengan kualitas, inovasi, dan sistem ekosistem tertutupnya yang khas [3].

Apple Inc., sebagai salah satu perusahaan teknologi terbesar di dunia, terus memperluas jangkauan produksinya secara global. Untuk menekan biaya produksi, Apple Inc mulai mengalihkan sebagian operasionalnya ke negara-negara berkembang seperti Vietnam, India, dan China. Strategi ini juga menjadi peluang bagi negara lain, termasuk Indonesia, untuk menjadi mitra produksi Apple [4]. Wacana investasi Apple di Indonesia mencuat setelah kunjungan CEO Apple, Tim Cook, ke Jakarta dalam rangka membahas potensi pendirian fasilitas produksi di Bandung. Investasi ini diyakini akan memberikan dampak positif seperti penciptaan lapangan kerja, *transfer* teknologi, peningkatan kualitas SDM, dan pertumbuhan industri lokal [4].

Meskipun begitu, rencana pendirian pabrik Apple di Bandung menimbulkan berbagai respons dari masyarakat yang terekam dalam ruang diskusi digital seperti media sosial. *Platform* X (sebelumnya Twitter) menjadi salah satu media yang digunakan masyarakat untuk mengekspresikan opini publik, baik dalam bentuk dukungan maupun kekhawatiran. Variasi sentimen publik ini mencerminkan adanya dinamika sosial yang penting untuk dipetakan [5].

Untuk memahami sentimen publik secara sistematis, diperlukan pendekatan berbasis kecerdasan buatan. Salah satu metode yang relevan adalah *Naïve Bayes Classifier*, yaitu algoritma klasifikasi probabilistik yang umum digunakan dalam analisis teks. Algoritma ini mampu memproses data dalam jumlah besar secara cepat dan efisien, serta memberikan hasil klasifikasi yang akurat meskipun dengan asumsi sederhana bahwa setiap fitur bersifat independen [6]. Dalam proses ekstraksi fitur teks, digunakan metode *Term Frequency – Inverse Document Frequency* (TF-IDF). *Term Frequency* (TF) mengukur seberapa sering suatu kata muncul dalam teks atau dokumen, mencerminkan tingkat kepentingannya. Sementara itu, *Inverse Document Frequency* (IDF) berfungsi sebagai teknik pembobotan yang mengevaluasi seberapa jarang sebuah kata muncul dalam kumpulan dokumen, sehingga membantu mengidentifikasi kata-kata yang lebih informatif [7].

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap rencana pendirian pabrik Apple di Bandung dengan memanfaatkan

algoritma *Naïve Bayes*. Penelitian ini mengikuti pendekatan CRISP-DM (*Cross Industry Standard Process for Data Mining*) terdiri dari enam tahapan utama yang saling berulang yaitu, *businiess understanding*, *data understanding*, *preparation*, *modelling*, *evaluastion*, dan *deployment*. untuk memastikan proses analisis dilakukan secara sistematis dan terstruktur [8]. Fokus diarahkan pada komentar publik yang diambil dari platform X, guna memperoleh pemetaan sentimen masyarakat secara kuantitatif. Inovasi dari penelitian ini terletak pada pemanfaatan data sosial media *real-time* untuk mendukung pengambilan kebijakan strategis dalam bidang investasi dan teknologi.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian dengan judul "Analisis Strategi Pemasaran Perusahaan Apple Dalam Meningkatkan Pemasaran Global" Penelitian ini mengkaji strategi branding Apple melalui keunikan desain, fitur, dan sistem teknologinya dengan pendekatan deskriptif. Hasilnya, Apple sukses membangun citra sebagai merek teknologi premium yang meningkatkan loyalitas konsumen melalui pengalaman pengguna yang eksklusif dan sistem tertutup [3].

Penelitian dengan judul "Risiko Apple Dalam Rencana Investasi Di Indonesia". Penelitian ini menganalisis risiko investasi Apple di Indonesia dengan pendekatan deskriptif-analitis. Hasil menunjukkan bahwa Apple memilih investasi berbasis inovasi dibandingkan mendirikan pabrik, sebagai respons terhadap risiko seperti keamanan dan regulasi, serta sebagai masukan bagi kebijakan investasi nasional [4].

penelitian dengan judul "Analisis Sentimen Opini Publik Terkait Judi *Online* Menggunakan Algoritma *Naïve Bayes* Dan *Support Vector Mechine*". Penelitian ini menganalisis sentimen publik terkait judi online menggunakan algoritma Naïve Bayes dan SVM. Hasil menunjukkan bahwa SVM lebih unggul dengan akurasi 98%, dibandingkan Naïve Bayes yang mencapai 93%, sehingga lebih efektif dalam mengklasifikasikan opini publik [5].

Penelitian dengan judul "Analisis Sentimen Isu Kecurangan Pemilu 2024 Berdasarkan Opini Pada Media Sosial Twitter Menggunakan Metode CRISP - DM Dengan Algoritma *Naïve Bayes Classifier*". Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap dugaan kecurangan Pemilu 2024 di Twitter. Metode yang digunakan adalah algoritma Naïve Bayes dengan klasifikasi sentimen positif, negatif, dan netral. Dari 1810 tweet yang dianalisis, model menghasilkan akurasi 66.85%, precision 68.05%, recall 56.11%, dan F1-score 61.56% [9].

2.2. Analisis Sentimen

Analisis sentiment berfokus pada penggalian opini dari teks yang tidak terstruktur guna mengetahui persepsi publik terhadap suatu topik, produk, atau layanan. Hasil analisis ini kemudian dikategorikan ke dalam sentimen positif, negatif, atau netral, dan dapat dimanfaatkan sebagai dasar dalam proses pengambilan keputusan [8].

2.3. X (*twitter*)

X, yang sebelumnya dikenal luas sebagai *Twitter*, merupakan *platform microblogging* dan media sosial daring yang memungkinkan penggunanya mengirim serta membaca pesan singkat yang dikenal sebagai "*tweet*", dengan batasan karakter hingga 140 huruf pada awal kemunculannya. X berkembang pesat menjadi salah satu situs web paling populer di dunia dan sering dijuluki sebagai layanan pesan singkat berbasis internet [6].

2.4. Data Mining

Data Mining didefinisikan sebagai satu set teknik yang digunakan secara otomatis untuk mengeksplorasi secara menyeluruh dan membawa ke permukaan relasi-relasi yang kompleks pada set data yang sangat besar. Set data yang dimaksud di sini adalah set data yang berbentuk tabulasi, seperti banyak diimplementasikan dalam teknologi manajemen basis data relasional. Akan tetapi, teknik-teknik Data Mining dapat juga diaplikasikan pada representasi data yang lain, seperti domain data spatial, berbasis teks, dan multimedia (citra) [5].

2.5. Text Mining

Didalam teks mining, sistem klasifikasi teks merupakan proses pengkajian dan evaluasi yang terpantau. Sistem klasifikasi teks merupakan sebuah proses penentuan untuk pemisahandan pengelompokan teks secara langsung atau otomatis, sesuai dengan klasifikasi konten teks yang di bawah yang diberikansistem [10].

2.6. Klassifikasi

Klasifikasi merupakan suatu teknik menemukan suatu pola yang mampu memisahkan kelas data yang satu dengan yang lainnya untuk menentukan objek yang masuk dengan kategori tertentu dengan melihat kelakuan dan atribut dari kelompok yang telah di definisikan. Teknik ini mampu mengklarifikasi data baru dengan menggunakan hasilnya untuk memberikan sejumlah aturan [11].

Klasifikasi sentimen merupakan proses identifikasi serta pengelompokan opini dalam teks ke dalam kategori seperti positif, negatif, atau netral. Metode ini memanfaatkan algoritma pembelajaran mesin serta teknik pemrosesan bahasa alami untuk secara otomatis menentukan kategori sentimen berdasarkan konteks dan kata kata yang terkandung dalam teks [8].

2.7. Algoritma Naïve Bayes

Salah satu tujuan utama Data Mining adalah mengklasifikasikan data ke dalam kategori yang telah ditentukan. Algoritma Naïve Bayes, yang berbasis

pada teorema Bayes, merupakan salah satu metode yang populer untuk melakukan klasifikasi. Metode ini bekerja dengan menghitung probabilitas suatu kelas berdasarkan fitur-fitur yang dimiliki oleh data tersebut. Dengan kata lain, Naïve Bayes menggunakan data historis untuk membuat prediksi tentang data baru [6].

Naïve Bayes classifiers merupakan metode pengklasifikasian berdasarkan aturan bayes berdasarkan rumus berikut:

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

Dimana:

- P(A|B) adalah probabilitas bahwa suatu opini termasuk dalam kelas A (misalnya, positif) setelah data (B) diamati.
- P(B|A) adalah probabilitas bahwa data B muncul jika opini tersebut positif.
- P(A) adalah probabilitas awal opini positif tanpa mengandalkan data.
- P(B) adalah probabilitas bahwa data B muncul.

Dengan menerapkan Teorema Bayes ini, dapat mengklasifikasikan sentimen opini publik berdasarkan data yang tersedia dari platform X (*Twitter*), dan menentukan apakah opini tersebut cenderung positif, negatif, atau netral. Pengelompokan data dengan jumlah yang besar adalah salah satu tugas dari Algoritma Naïve Bayes, juga banyak digunakan untuk dan memprediksi kondisi yang akan terjadi [12].

2.8. CRISP – DM

Penelitian ini mengacu pada metode CRISP-DM (*Cross Industry Standard Process for Data Mining*) terdiri dari enam tahapan yang saling berulang, yaitu *Business Understanding* (pemahaman bisnis), *Data Understanding* (pemahaman data), *Data Preparation* (persiapan data), *Modeling* (pemodelan), *Evaluation* (evaluasi), dan *Deployment* (penerapan) [8].

2.9. TF - IDF

Dalam penelitian ini, digunakan model *Term Frequency-Inverse Document Frequency* (TF-IDF). *Term Frequency* (TF) mengukur seberapa sering suatu kata muncul dalam teks atau dokumen, mencerminkan tingkat kepentingannya. Sementara itu, *Inverse Document Frequency* (IDF) berfungsi sebagai teknik pembobotan yang mengevaluasi seberapa jarang sebuah kata muncul dalam kumpulan dokumen, sehingga membantu mengidentifikasi kata-kata yang lebih informatif [7].

2.10. *Lexicon Based*

Metode berbasis leksikon dalam analisis sentimen menggunakan kamus yang telah dikurasi secara manual, di mana setiap kata atau frasa diberi label (positif, negatif, atau netral) guna mendeteksi sentimen dalam kalimat [13]. Kamus ini juga membantu penentuan bobot kata untuk klasifikasi otomatis sebelum pelatihan model [14].

Beberapa jenis kamus yang digunakan dalam pendekatan leksikon meliputi kamus positif yang berisi kata-kata bernuansa positif, kamus negatif untuk mendeteksi kata bermakna negatif, kamus negasi yang mengenali kata-kata penyangkal yang dapat membalikkan makna sentimen, kamus kata dasar dan KBBI yang digunakan dalam proses *stemming* untuk mengubah kata berimbuhan menjadi bentuk dasar, serta kamus *stopwords* yang memuat kata-kata umum tanpa makna signifikan dan dihapus untuk meningkatkan efisiensi klasifikasi.

2.11. *Web scraping*

Web Scraping adalah proses otomatisasi untuk mengakses dan mengumpulkan informasi dari situs web berdasarkan kata kunci tertentu. Data yang diambil, biasanya berupa teks, kemudian disimpan dan diindeks untuk memudahkan pencarian, pengelompokan, dan analisis informasi dari berbagai sumber daring [15].

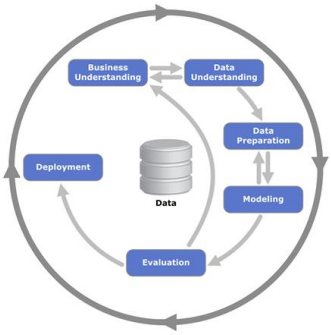
2.12. *Google Collab*

Google Collab, atau *Google Colaboratory*, adalah *platform* versi *Jupyter Notebook* berbasis *cloud* yang dapat diakses melalui *browser* seperti *Google Chrome* dan *Mozilla Firefox*. Pengguna *Google Colab* dapat menjalankan kode *Python* tanpa perlu melakukan instalasi atau konfigurasi tambahan, karena seluruh pengaturan dan penyesuaian sistem dilakukan di *cloud* [16].

3. METODE PENELITIAN

Penelitian ini menggunakan analisis deskriptif kualitatif, untuk menganalisis, menggambarkan, dan meringkas opini publik dari berbagai data yang dikumpulkan pada *platform* X yang hasil pengamatannya akan menentukan apakah nada emosional opini tersebut positif, negatif, dan netral. Dalam konteks analisis sentimen, penelitian ini bertujuan untuk memahami sikap atau perasaan yang terkandung dalam suatu teks, komentar di media sosial, atau artikel berita.

Dalam proses klasifikasi opini masyarakat dalam *text mining* pada penelitian ini disertai dengan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang disesuaikan dengan tahapan sebagai berikut [9].



Gambar 1. Tahapan metode CRISP- DM

3.1. *Business Understanding*

Tahap ini merumuskan tujuan penelitian untuk memahami sentimen masyarakat terhadap pendirian pabrik Apple di Bandung berdasarkan data Twitter. Penelitian ini menganalisis distribusi sentimen (positif, negatif, dan netral), mengidentifikasi faktor utama yang mempengaruhi sentimen tersebut, serta mengamati tren perubahan sentimen dari April 2024 hingga Februari 2025. Hasil penelitian ini dapat dimanfaatkan oleh Apple, pemerintah, dan pelaku industri teknologi untuk mengukur tingkat penerimaan masyarakat tentang dampak positif dalam perluasan lapangan kerja pada masyarakat setempat.

3.2. *Data Understanding*

Pada tahap ini, merupakan pengumpulan dataset atau data yang digunakan berupa kumpulan tweet dan komentar dari pengguna Twitter yang mengandung kata kunci tertentu '*Apple Investasi', 'Pabrik Apple di Bandung', 'Apple Bandung', 'Apple Investasi Indonesia'.*

3.3. *Data Preparation*

Setelah memahami data, langkah selanjutnya adalah mempersiapkan data untuk dianalisis. Proses ini melibatkan beberapa tahapan, seperti *Cleaning, Normalisasi, Tokenized, Stopword, Stemming dan Labeling.*

3.3.1. *Cleaning*

Pada tahap *Cleaning*, data teks dibersihkan dari karakter yang tidak diperlukan seperti tanda baca, angka, emoji, URL, dan simbol khusus lainnya. Proses ini bertujuan untuk memastikan teks lebih bersih dan mudah diproses oleh model, sehingga hanya informasi relevan yang digunakan dalam analisis.

3.3.2. *Normalize*

Normalisasi bertujuan untuk mengubah kata-kata yang tidak baku menjadi bentuk standar atau bakunya. Hal ini penting karena dalam media sosial banyak ditemukan kata-kata dengan ejaan tidak formal, seperti singkatan atau bahasa gaul. Contohnya, kata "gak", "ga", "nggak" akan dinormalisasi menjadi "tidak", atau "ak", "gw", "gue" menjadi "saya".

3.3.3. *Tokenizing*

Selanjutnya, *Tokenized* adalah proses memecah teks menjadi unit kata atau token. Setiap kalimat atau teks yang diambil dari X diubah menjadi daftar kata yang lebih kecil agar dapat dianalisis secara lebih spesifik. Misalnya, kalimat "Apple membangun pabrik di Bandung" akan diubah menjadi ["Apple", "membangun", "pabrik", "di", "Bandung"]. Tokenisasi membantu dalam memahami struktur kalimat dan mempermudah analisis lebih lanjut.

3.3.4. *Stopword*

Proses *Stopword Removal* bertujuan untuk menghapus kata-kata umum yang sering muncul tetapi tidak memiliki pengaruh signifikan terhadap sentimen, seperti "di", "ke", "yang", dan "dan". Kata-kata ini umumnya tidak memiliki makna spesifik dalam analisis sentimen, sehingga penghapusannya dapat meningkatkan efisiensi model dengan hanya mempertahankan kata-kata yang benar-benar penting dalam menentukan sentimen.

3.3.5. *Stemming*

Stemming adalah proses mengubah kata ke bentuk dasarnya atau akar katanya. Misalnya, kata "membangun", "dibangun", dan "pembangunan" akan diubah menjadi "bangun". Proses ini membantu mengurangi kompleksitas data dengan memastikan bahwa berbagai bentuk kata dari akar yang sama dianggap sebagai satu entitas, sehingga model dapat lebih konsisten dalam menganalisis sentimen tanpa terganggu oleh variasi bentuk kata.

3.3.6. *Labelling*

Proses *labeling* dilakukan dengan pendekatan semi-otomatis, yaitu menggabungkan metode *lexicon-based labeling* (menggunakan kamus sentimen). Sentimen dikategorikan menjadi positif, negatif, atau netral. Pada tahap ini, algoritma Naive Bayes diterapkan untuk menganalisis data. Model dilatih menggunakan data training untuk mengenali pola-pola yang berkaitan dengan sentimen tertentu. Setelah pelatihan, model digunakan untuk mengklasifikasikan *tweet* berdasarkan sentiment.

3.4. *Modelling*

Langkah selanjutnya adalah teknik pemodelan, yang kemudian diikuti dengan penyusunan skenario pengujian serta validasi kualitas model. Berdasarkan pertimbangan tersebut, penelitian ini mengusulkan penggunaan metode klasifikasi dengan algoritma Naïve Bayes.

3.5. *Evaluation*

Tahap evaluasi dilakukan untuk menilai apakah model yang dibangun sudah sesuai dengan tujuan penelitian. Evaluasi dilakukan dengan membandingkan hasil prediksi model dengan data aktual menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-Score*. Salah satu indikator utama adalah *True Positive* (TP), yaitu data yang diprediksi dan benar-benar termasuk kategori positif.

3.6. *Deployment*

Tahap ini merupakan tahap akhir dari metode CRISP-DM, yaitu pembangunan aplikasi atau sistem berdasarkan model dan data yang telah diproses pada tahapan sebelumnya. Pada tahap ini, hasil dari seluruh proses dianalisis dan diimplementasikan ke dalam bentuk aplikasi yang sesuai dengan kebutuhan.

4. HASIL DAN PEMBAHASAN

4.1. *Crawling Data*

Langkah yang akan dilakukan oleh peneliti untuk melakukan *crawling data* atau pengambilan dataset adalah melakukan *login* atau masuk akun *google* untuk mengakses fitur-fitur yang disediakan oleh *google* seperti X atau X *(Twitter)* API. Kemudian, peneliti dapat menggunakannya untuk proses *crawling* data dengan menggunakan Bahasa pemrograman *python*. Komentar yang didapat berjumlah 1.810. Komentar-komentar tersebut disimpan didalam file bernama *apple_new.csv*. Tabel 1 menunjukan dataset yang diambil dari komentar X.

Table 1. *Dataset* komentar X

No	Komentar (*Full_Text)*	*Username*
1	@Strategi_Bisnis Perusahaan sebesar Apple cuma mampu invest 157 miliar? Itu mah cuma seharga rumah di Pondok Indah gak sih	Satuterkasih
2	Penghinaan begini kok dibilang lumayan mas mas. Bargaining sama perusahaan kalah dan lemah begini	pendekarads
3	Nah bener lumayan daripada lu manyun	vicko_twitt
…	…	…
1809	@dagelanx Baru tau aku tentang informasi terkait apple bakal bangun pabrik di Bandung	gincu
1810	Yg bener nih infonya min akurat kah?	chichiochi_

4.2. *Data Preparasi*

Pada tahap ini, data yang telah dikumpulkan dibersihkan dan diproses untuk memastikan kualitas data sebelum analisis sentimen dilakukan. Hasil dari tahapan *pre-processing*.

Tahapan pertama yaitu *cleaning* menghapus simbol, tanda baca, mention (seperti *@username*), serta format yang tidak penting dari data mentah (contohnya *tweet*). Tujuannya agar hanya teks penting yang tersisa untuk dianalisis.

Table 2. Hasil *Cleaning*

Sebelum	Sesudah
@Strategi_Bisnis Perusahaan sebesar Apple cuma mampu invest 157 miliar? Itu mah cuma seharga rumah di Pondok Indah gak sih	perusahaan sebesar apple cuma mampu invest 157 miliar itu mah cuma seharga rumah di pondok indah gak sih
@Strategi_Bisnis Penghinaan begini kok dibilang lumayan mas mas. Bargaining sama perusahaan kalah dan lemah begini	penghinaan begini kok dibilang lumayan mas mas. bargaining sama perusahaan kalah dan lemah begini
Nah bener lumayan daripada lu manyun	nah bener lumayan daripada lu manyun
…	…
@dagelanx Baru tau aku tentang informasi terkait apple bakal bangun pabrik di Bandung	baru tau aku tentang informasi terkait apple bakal bangun pabrik di bandung
Yg bener nih infonya min akurat kah?	yg bener nih infonya min akurat kah _

Proses selanjutnya normalisasi mengubah kata tidak baku atau bahasa gaul menjadi bentuk baku. Contohnya: "cuma" menjadi "cuman", "lu" menjadi "kamu", dan sebagainya. Proses ini penting agar analisis teks dapat dilakukan secara konsisten.

Table 3. Hasil *Normalize*

Sebelum	Sesudah
perusahaan sebesar apple cuma mampu invest 157 miliar itu mah cuma seharga rumah di pondok indah gak sih	perusahaan sebesar apple cuman mampu investasi 157 miliar itu mah cuman seharga rumah di pondok indah tidak sih
penghinaan begini kok dibilang lumayan mas mas. bargaining sama perusahaan kalah dan lemah begini	penghinaan begini kenapa bilang lumayan mas. bargaining sama perusahaan kalah dan lemah begini
nah bener lumayan daripada lu manyun	nah bener lumayan daripada kamu manyun
…	…
baru tau aku tentang informasi terkait apple bakal bangun pabrik di bandung	baru tau aku tentang informasi terkait apple akan bangun pabrik di bandung
yg bener nih infonya min akurat kah _	yang bener ini informasi admin apakah akurat

Selanjutnya yaitu Tokenisasi untuk memecah kalimat menjadi kata-kata per satuan terkecil (token). Misalnya kalimat diubah menjadi daftar kata. Proses ini akan mempermudah proses selanjutnya seperti penghapusan kata tidak penting atau pencocokan kata.

Table 4. Hasil *Tokenizing*

Sebelum	Sesudah
perusahaan sebesar apple cuman mampu investasi 157 miliar itu mah cuman seharga rumah di pondok indah tidak sih	[‘perusahaan’, ‘sebesar’, ‘apple’ ‘cuman’, ‘mampu’, ‘investasi’, ‘157 miliar’, ‘itu’, ‘mah’, ‘cuman’, ‘seharga’, ‘rumah’, ‘di’, ‘pondok’, ‘indah’, ‘tidak’, ‘sih’]
penghinaan begini kenapa bilang lumayan mas. bargaining sama perusahaan kalah dan lemah begini	[‘penghinaan’, ‘begini’, ‘kenapa’, ‘bilang’, ‘lumayan’, ‘mas’, ‘bargaining’, ‘sama’, ‘perusahaan’, ‘kalah’, ‘dan’, ‘lemah’, ‘begini’]
nah bener lumayan daripada kamu manyun	[‘nah’, ‘bener’, ‘lumayan’, ‘daripada’, ‘kamu’, ‘manyun’]
…	…
baru tau aku tentang informasi terkait apple akan bangun pabrik di bandung	[‘baru’, ‘tau’, ‘aku’, ‘tentang’, ‘informasi’, ‘terkait’, ‘apple’, ‘bakal’, ‘bangun’, ‘pabrik’, ‘di’, ‘bandung’]
yang bener ini informasi admin apakah akurat	[‘yang’, ‘bener’, ‘nih’, ‘informasi’, ‘kak’, ‘akurat’, ‘kah’]

Setelah itu *stopword removal* menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis, seperti "di", "itu", "dan", "sih". Hal ini dilakukan agar hanya kata-kata yang relevan dengan sentimen atau makna yang dianalisis.

Table 5. Hasil *Stopword Removal*

Sebelum	Sesudah
[‘perusahaan’, ‘sebesar’, ‘apple’ ‘cuman’, ‘mampu’, ‘investasi’, ‘157 miliar’, ‘itu’, ‘mah’, ‘cuman’, ‘seharga’, ‘rumah’, ‘di’, ‘pondok’, ‘indah’, ‘tidak’, ‘sih’]	[‘perusahaan’, ‘sebesar’, ‘apple’ ‘cuman’, ‘mampu’, ‘investasi’, ‘157 miliar’, ‘cuman’, ‘seharga’, ‘rumah’, ‘pondok’, ‘indah’]
[‘penghinaan’, ‘begini’, ‘kenapa’, ‘bilang’, ‘lumayan’, ‘mas’, ‘bargaining’, ‘sama’, ‘perusahaan’, ‘kalah’, ‘dan’, ‘lemah’, ‘begini’]	[‘penghinaan’, ‘kenapa’, ‘bilang’, ‘lumayan’, ‘sama’,‘perusahaan’, ‘kalah’, ‘lemah’]
[‘nah’, ‘bener’, ‘lumayan’, ‘daripada’, ‘kamu’, ‘manyun’]	[‘bener’, ‘lumayan’, ‘daripada’, ‘kamu’, ‘manyun’]
…	
[‘baru’, ‘tau’, ‘aku’, ‘tentang’, ‘informasi’, ‘terkait’, ‘apple’, ‘bakal’, ‘bangun’, ‘pabrik’, ‘di’, ‘bandung’]	[‘baru’, ‘tau’, ‘aku’, ‘informasi’, ‘terkait’, ‘apple’, ‘bakal’, ‘bangun’, ‘pabrik’, ‘bandung’]
[‘yang’, ‘bener’, ‘nih’, ‘informasi’, ‘kak’, ‘akurat’, ‘kah’]	[‘bener’,‘informasi’, ‘akurat’]

Kemudian proses *stemming* mengubah kata ke bentuk dasarnya. Contoh: "penghinaan" menjadi "hina", "investasi" tetap "investasi", atau "seharga" menjadi "harga". Ini penting untuk menyamakan makna dari variasi kata yang berbeda.

Table 6. Hasil *Stemming*

Sebelum	Sesudah
[‘perusahaan’, ‘sebesar’, ‘apple’ ‘cuman’, ‘mampu’, ‘investasi’, ‘157 miliar’, ‘cuman’, ‘seharga’, ‘rumah’, ‘pondok’, ‘indah’]	[‘perusahaan’, ‘besar, ‘apple’ ‘cuman’, ‘mampu’, ‘investasi’, ‘157 miliar’, ‘cuman’, ‘harga, ‘rumah’, ‘pondok’, ‘indah’]
[‘penghinaan’, ‘kenapa’, ‘bilang’, ‘lumayan’, ‘sama’, ‘perusahaan’, ‘kalah’, ‘lemah’]	[‘hina’, ‘kenapa’, ‘bilang’, ‘lumayan’ ,‘perusahaan’, ‘kalah’, ‘lemah’]
[‘bener’, ‘lumayan’, ‘daripada’, ‘kamu’, ‘manyun’]	[‘bener’, ‘lumayan’, ‘daripada’, ‘kamu’, ‘manyun’]
…	…
[‘baru’, ‘tau’, ‘aku’, ‘informasi’, ‘terkait’, ‘apple’, ‘bakal’, ‘bangun’, ‘pabrik’, ‘bandung’]	[‘baru’, ‘tau’, ‘aku’, ‘informasi’, ‘terkait’, ‘apple’, ‘bakal’, ‘bangun’, ‘pabrik’, ‘bandung’]
[‘bener’, ‘informasi’, ‘akurat’]	[‘bener’,‘informasi’, ‘akurat’]

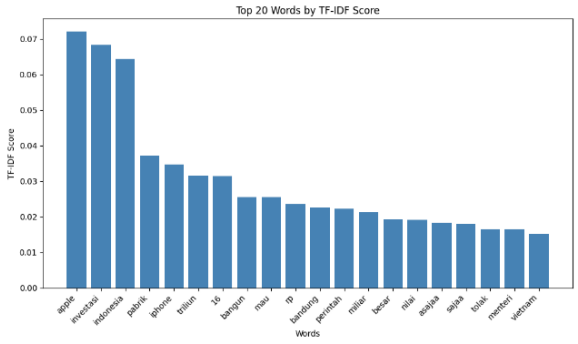
Setelah semua proses selesai, tiap data diberi label sentimen: Positif, Negatif, atau Netral. Proses ini merupakan hasil akhir dari proses data preparasi sebelum digunakan dalam pelatihan model analisis sentiment.

Table 7. Hasil *Labelling*

Full_Text	**Hasil label**
perusahaan besar apple cuman mampu investasi 157 miliar cuman harga rumah pondok indah	Negatif
hina kenapa bilang lumayan perusahaan kalah lemah	Negatif
bener lumayan daripada kamu manyun	Positif
…	…
baru tau aku informasi terkait apple bakal bangun pabrik bandung	Netral
bener informasi akurat	Netral

4.3. TF – IDF

Visualisasi TF-IDF menunjukkan 20 kata teratas yang memiliki skor tertinggi dalam kumpulan data. Kata “apple”, “investasi”, dan “indonesia” memiliki skor TF-IDF paling tinggi, yang berarti kata-kata tersebut paling menonjol dan relevan dalam dokumen. Tingginya skor TF-IDF menunjukkan bahwa kata tersebut sering muncul di dokumen tertentu namun jarang muncul secara umum, sehingga dianggap penting dalam membedakan konteks setiap dokumen. Hasil ini menunjukkan bahwa topik sentimen yang dianalisis berkaitan erat dengan isu ekonomi dan kebijakan industri.



Gambar 2. Visualisasi TF – IDF

4.4. Hasil Klasifikasi Algoritma Naïve Bayes

Setelah proses transformasi data menggunakan TF-IDF selesai dilakukan, data kemudian digunakan untuk pelatihan dan pengujian model klasifikasi menggunakan algoritma Naïve Bayes. Dataset yang berjumlah total 1.810 data dibagi menjadi dua bagian, yaitu 80% data untuk pelatihan (*training*) sebanyak 1.448 data, dan 20% untuk pengujian (*testing*) sebanyak 362 data. Pembagian ini dilakukan dengan menggunakan fungsi *train_test_split* dari *library Scikit-Learn* dengan parameter *test_size* = 0.2.

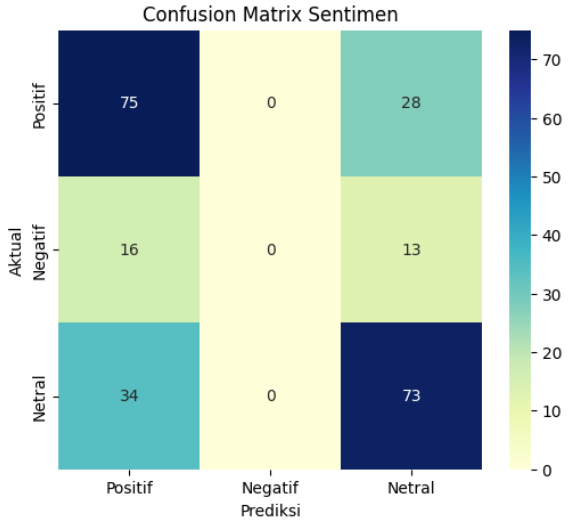
Proses pelatihan dilakukan dengan menerapkan *Multinomial* Naïve Bayes pada hasil vektorisasi TF-IDF dari *data train*. Setelah model dilatih, dilakukan

pengujian terhadap data uji untuk melihat performa model. Berdasarkan hasil evaluasi, model menghasilkan akurasi sebesar 61,92%.
Selain itu, model menunjukkan performa yang cukup baik dalam mengklasifikasikan sentimen positif dan netral dengan nilai F1-Score masing-masing sebesar 0.66, namun belum dapat mengenali kelas negatif dengan baik (F1-Score: 0.00). Hal ini juga terlihat pada *confusion matrix*, di mana semua data negatif salah diklasifikasikan.

```
Confusion Matrix:
[[ 0 13 16]
 [ 0 73 34]
 [ 0 28 75]]

Classification Report:
              precision    recall  f1-score   support

     Negatif       0.00      0.00      0.00        29
      Netral       0.64      0.68      0.66       107
     Positif       0.60      0.73      0.66       103

    accuracy                           0.62       239
   macro avg       0.41      0.47      0.44       239
weighted avg       0.55      0.62      0.58       239


Akurasi: 0.6192
```

Gambar 3. *Model Training Data*

4.5. Hasil *Confusion Matrix*



Gambar 4. *Confusion Matrix Sentimen*

a. Kelas Positif
TP (True Positive) = 75
FP (False Positive) = 34 (Netral → Positif) + 16 (Negatif → Positif) = 50
FN (False Negative) = 28 (Positif → Netral)

$$Precision = \frac{TP}{TP + FP} = \frac{75}{75 + 50} = \frac{75}{125} = 0{,}60$$

$$Recall = \frac{TP}{TP + FN} = \frac{75}{75 + 28} = \frac{75}{103} = 0{,}728$$

$$F1 - Score = \frac{2\ x\ 0{,}60\ x\ 0{,}728}{0{,}60 + 0{,}728} = 0{,}656$$

b. Kelas Netral
TP = 73
FP = 28 (Positif → Netral) + 13 (Negatif → Netral) = 41
FN = 34 (Netral → Positif)

$$Precision = \frac{73}{73 + 41} = \frac{73}{114} = 0{,}640$$

$$Recall = \frac{73}{73 + 34} = \frac{73}{107} = 0{,}682$$

$$F1 - Score = \frac{2\ x\ 0{,}640\ x\ 0{,}682}{0{,}640 + 0{,}682} = 0{,}660$$

c. Kelas Negatif
TP = 0
FP = 0
FN = 16 (Negatif → Positif) + 13 (Negatif → Netral) = 29

$$Precision = \frac{0}{0 + 0} = 0$$

$$Recall = \frac{0}{0 + 29} = 0$$

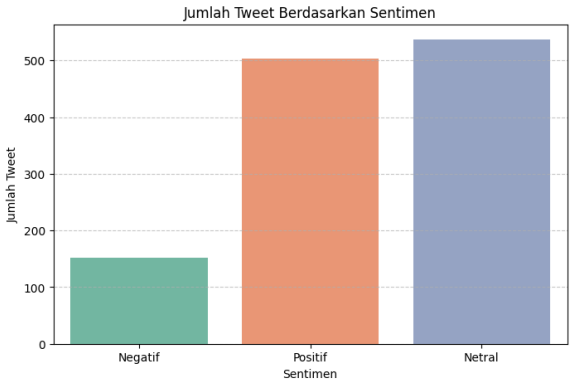
$$F1 - Score = 0$$

True Positives (TP total) = 75 (Positif) + 0 (Negatif) + 73 (Netral) = 148
Total Data = semua elemen:
75 + 0 + 28 + 16 + 0 + 13 + 34 + 0 + 73 = 239

$$\text{Akurasi} = \frac{148}{239} = 0{,}6192 \text{ atau } 61{,}92\ \%$$

Model Naïve Bayes berhasil mengklasifikasikan sentimen Positif dan Netral dengan F1-Score sebesar 0.66, namun gagal mengenali kelas Negatif (*precision*, *recall*, dan *F1-Score* = 0). Hal ini kemungkinan disebabkan oleh jumlah data Negatif yang lebih sedikit atau fitur yang kurang representatif. Dengan akurasi 61.92%, model tergolong cukup baik untuk dua kelas utama, namun perlu perbaikan lebih lanjut pada distribusi data dan representasi fitur untuk meningkatkan performa keseluruhan.

4.6. Visualisasi Sentimen



Gambar 5. Visualisasi Sentimen

Visualisasi sentimen ditampilkan dalam bentuk diagram batang yang menunjukkan jumlah tweet berdasarkan kategori sentimen. Hasilnya

menunjukkan bahwa tweet dengan sentimen positif memiliki jumlah terbanyak, disusul oleh sentimen netral, dan terakhir sentimen negatif. Hal ini mengindikasikan bahwa mayoritas respons publik terhadap topik yang dianalisis bersifat positif.

4.7. Hasil *Wordcloud*

Wordcloud untuk sentimen positif menampilkan kata-kata yang paling sering muncul dalam *tweet* dengan nada positif. Kata yang dominan di antaranya adalah "indonesia", "investasi", "apple", dan "bangun pabrik", "Bandung". Ini menunjukkan adanya dukungan masyarakat terhadap rencana investasi Apple di Indonesia, khususnya di Bandung dalam pembangunan fasilitas produksi.

Gambar 6. *Wordcloud* Positif

Pada *wordcloud* sentimen negatif, kata-kata utama yang muncul adalah "tidak", "buat apa", "mau investasi apa", "masasih", "bukan", dan "masih". Kemunculan kata-kata dengan konotasi negatif ini mencerminkan adanya keraguan atau penolakan terhadap isu yang dibahas, seperti kekhawatiran terhadap realisasi investasi atau skeptisisme terhadap dampaknya.

Gambar 7. *Wordcloud* Negatif

Wordcloud sentimen netral didominasi oleh kata-kata seperti "iphone", "bangun pabrik", "apple", dan "rp triliun". Kata-kata ini bersifat informatif dan tidak menunjukkan emosi tertentu, yang mengindikasikan bahwa *tweet-tweet* dalam kategori ini cenderung menyampaikan informasi atau berita tanpa opini yang kuat.

Gambar 8. *Wordcloud* Netral

5. KESIMPULAN DAN SARAN

Penelitian ini melakukan analisis sentimen terhadap *tweet* terkait pendirian pabrik Apple di Bandung dengan menggunakan algoritma Naïve Bayes dan pendekatan TF-IDF untuk ekstraksi fitur. Berdasarkan hasil klasifikasi, sebagian besar *tweet* mengandung sentimen positif dan netral, sementara sentimen negatif sangat sedikit terdeteksi. Evaluasi model melalui *confusion matrix* menunjukkan bahwa model memiliki akurasi sebesar 61.92%, dengan *F1-Score* tertinggi pada kelas Positif dan Netral (masing-masing 0.66), namun gagal mengklasifikasikan kelas Negatif secara akurat (*F1-Score* = 0). Hal ini menunjukkan bahwa model cukup efektif dalam mengenali opini netral dan positif, tetapi masih lemah dalam mengidentifikasi opini negatif. Untuk penelitian selanjutnya, disarankan agar jumlah data pada setiap kelas diseimbangkan guna menghindari bias model terhadap kelas mayoritas. Selain itu, penggunaan teknik lain seperti oversampling, pemilihan fitur yang lebih baik, atau penerapan algoritma pembelajaran mesin lain seperti *SVM, Random Forest, atau Logistic Regression* dapat dieksplorasi untuk meningkatkan performa klasifikasi, terutama pada kelas dengan distribusi kecil seperti sentimen negatif.

DAFTAR PUSTAKA

[1] M. S. Riswan, H. D. Waloejo, S. Lisyorini, D. A. Bisnis, and U. Diponegoro, "Pengaruh Inovasi Produk Dan Kualitas Produk Terhadap Kepuasan Konsumen Pada Pengguna Smartphone Merek Iphone Apple Di Kota Semarang," *J. Ilmu Adm. Bisnis*, vol. 11, no. 2, pp. 272–280, 2022.

[2] D. Murni, J. Jamna, R. Handican, and S. Solfema, "Pemanfaatan Smartphone dalam Pembelajaran Matematika : Bagaimana Persepsi Mahasiswa?," *J. Cendekia J. Pendidik. Mat.*, vol. 7, no. 1, pp. 590–603, 2023, doi: 10.31004/cendekia.v7i1.2153.

[3] F. Damayanti and K. Rostiana, "Analisis Strategi Pemasaran Perusahaan Apple Dalam Meningkatkan Pemasaran Global," *J. Publ. Ilmu Manaj.*, vol. 3, no. 1, pp. 138–142, 2024, [Online]. Available: https://doi.org/10.55606/jupiman.v3i1.3271

[4] V. Sasmita, "Risiko Apple Dalam Rencana Investasi Di Indonesia," *J. Huk. dan Kewarganegaraan*, vol. 3, no. 11, pp. 1–11,

2024, doi: 10.3783/causa.v2i9.2461.

[5] A. Maulana and A. Yuliana, "Analisis Sentimen Opini Publik Terkait Judi Online Menggunakan Algoritma Naïve Bayes Dan Support Vector Mechine," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 12, no. 3, pp. 3706–3714, 2024.

[6] 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.

[7] O. I. Gifari, M. Adha, F. Freddy, and F. F. S. Durrand, "Film Review Sentiment Analysis Using TF-IDF and Support Vector Machine," *J. Inf. Technol.*, vol. 2, no. 1, pp. 36–40, 2022.

[8] Lady Agustin Fitrana, S. Linawati, N. Herlinawati, R. Sa'adah, and S. Seimahuria, "Analisis Sentimen Pengguna Twitter Terhadap Brand Indosat Menggunakan Metode Naïve Bayes Classifier," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 4291–4297, 2024, doi: 10.36040/jati.v8i3.9866.

[9] A. A. Muttaqin, S. Alam, and M. A. Komara, "Analisis Sentimen Isu Kecurangan Pemilu 2024 Berdasarkan Opini Pada Media Sosial Twitter Menggunakan Metode Crisp-Dm Dengan Algoritma Naïve Bayes Classifier," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 5, pp. 8764–8772, 2024.

[10] A. Firdaus and W. Istalama Firdaus, "Text Mining dan Pola Algoritma dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan)," vol. 13, no. 1, pp. 66–78, 2021.

[11] I. Romli and A. T. Zy, "Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5," *J. Sains Komput. Inform.*, vol. 4, no. 2, pp. 694–702, 2020.

[12] A. M. Taufiqi and A. Nugroho, "Sentimen Pengguna Twitter Mengenai Isu Kebocoran Data Dengan Algoritma Naïve Bayes," *J. Nas. Ilmu Komput.*, vol. 4, no. 1, pp. 1–11, 2023, doi: 10.47747/jurnalnik.v4i1.1091.

[13] O. Manullang, C. Prianto, and N. H. Harani, "Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based Dan Random Forest," *J. Ilm. Inform.*, vol. 11, no. 02, pp. 159–169, 2023, doi: 10.33884/jif.v11i02.7987.

[14] M. F. Naufal Fathoni, E. Yulia Puspaningrum, and A. Nugroho Sihananto, "Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM," *J. Inform. dan Sains Teknol.*, vol. 2, no. 3, pp. 62–76, 2024, doi: 10.62951/modem.v2i3.112.

[15] N. Fadhlullah, S. Setiawansyah, and A. Surahman, "Penerapan Teknologi Web Scraping Sebagai Pengumpulan Data Covid-19 Di Provinsi Lampung," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 3, no. 1, pp. 25–30, 2022, doi: 10.33365/jatika.v3i1.1841.

[16] M. S. S. Nugroho and F. Nurlaila, "Klasifikasi Spesies Burung Dengan Menggunakan Convolutional Neural Network," *J. Ilmu Komput. dan Sci.*, vol. 2, no. 11, pp. 2867–2878, 2023, [Online]. Available: https://journal.mediapublikasi.id/index.php/oktal