BAB II

TINJAUAN PUSTAKA

2.1. LandasanTeori

2.1.1 Analisis Sentimen

Analisis sentimen adalah proses yang memanfaatkan analisis teks untuk mengumpulkan data dari berbagai sumber online dan platform media sosial. Tujuannya adalah untuk mengenali serta mengekstrak pendapat yang terdapat dalam teks. Proses ini memiliki kemampuan untuk mengubah data yang tadinya tidak terstruktur menjadi data yang lebih teratur dan terstruktur.(Lestari & Saepudin, 2021).

Analisis sentimen bertujuan untuk mengelompokkan teks ke dalam kalimat atau dokumen, lalu menilai apakah pendapat yang terungkap dalam teks tersebut cenderung positif, negatif, atau netral. Selain itu, analisis sentimen juga mampu mengidentifikasi ekspresi emosional seperti kesedihan, kebahagiaan, atau kemarahan. (Lestari & Saepudin, 2021).

Analisis sentimen dapat dilakukan menggunakan metode lexicon-based atau machine learning-based. Dalam metode *lexicon-based*, polaritas kata dianalisis berdasarkan nilai yang tercantum dalam kamus yang telah ditetapkan sebelumnya. Sementara dalam pendekatan *Machine Le*arning, algoritma pembelajaran mesin digunakan untuk menganalisis sentimen pada data yang telah diberi label. Setelahnya, data diuji untuk menilai seberapa akurat prediksi sentimen yang dihasilkan. Metode ini juga bisa memanfaatkan pendekatan *bag-of-words*, yang

mengelompokkan kata ke dalam token yang sesuai dengan konteks penelitian. (Wahyudi, 2021)

Dalam penelitian analisis sentimen, pendekatan machine learning digunakan. Pendekatan ini memanfaatkan algoritma *Machine Learning* untuk menganalisis sentimen dengan menggunakan data latih yang telah dilabeli. Langkah berikutnya adalah menguji data untuk menilai seberapa tepat prediksi sentimen yang dihasilkan oleh algoritma machine learning tersebut. Tujuan akhir dari penelitian ini adalah untuk mengevaluasi tingkat akurasi dari prediksi sentimen yang dibuat oleh algoritma *Machine Learning*.

2.1.2 *Blockchain*

Blockchain adalah suatu inovasi teknologi yang memungkinkan pencatatan transaksi keuangan atau data lainnya secara digital melalui jaringan yang terdistribusi secara merata. Dalam sistem *Blockchain*, setiap transaksi atau data yang dimasukkan ke dalam blok dienkripsi dan dihubungkan secara kriptografis ke blok sebelumnya, membentuk rantai blok yang tidak dapat diubah. (Nakamoto, 2009). *Blockchain* memanfaatkan teknologi enkripsi untuk mengamankan dan memvalidasi setiap transaksi yang tercatat dalam jaringannya. Tiap transaksi disimpan dalam blok individual dan terhubung secara berkesinambungan dengan blok-blok sebelumnya, membentuk sebuah rantai yang tidak dapat diubah. Dengan demikian, *Blockchain* menciptakan suatu sistem yang transparan, terukur, dan kebal terhadap pemalsuan karena blok-blok tersebut tidak dapat dimodifikasi atau dihapus dengan cara apapun (Mabruroh dkk., 2021).

Setiap transaksi dalam *Blockchain* tercatat secara permanen di *Ledger,* yang merupakan database publik tempat sistem blockchain disimpan. Jaringan *Blockchain* biasanya diatur oleh serangkaian *Node* yang saling berkomunikasi dan menggunakan protokol tertentu untuk memverifikasi blok-blok baru (Arief & Sundara, 2017). Transaksi merujuk pada aktivitas yang tercatat di dalam blok pada teknologi *Blockchain*. Bentuk transaksi ini bervariasi, seperti pengiriman mata uang digital dari satu alamat *ke* alamat lain dalam konteks *Bitcoin*, atau pelaksanaan operasi dalam kontrak pintar pada *Ethereum*. Setiap transaksi akan diperiksa oleh *node-node* dalam jaringan *Blockchain* sebelum akhirnya dimasukkan ke dalam blok (Nakamoto, 2009).

Blockchain adalah serangkaian node yang berkomunikasi dan menggunakan protokol tertentu untuk memverifikasi blok-blok baru. Transaksi dalam *Blockchain* mencatat aktivitas di *dalam* blok, seperti pengiriman mata uang digital atau pelaksanaan operasi kontrak pintar. Setiap transaksi diperiksa oleh *node-n*ode dalam jaringan sebelum dimasukkan ke dalam blok. Teknologi konsensus memungkinkan semua node dalam jaringan mencapai kesepakatan tentang kevalidan data dalam *Blockchain*, memastikan setiap node setuju tentang status terkini *Blockchain,* yang beroperasi secara terdesentralisasi.

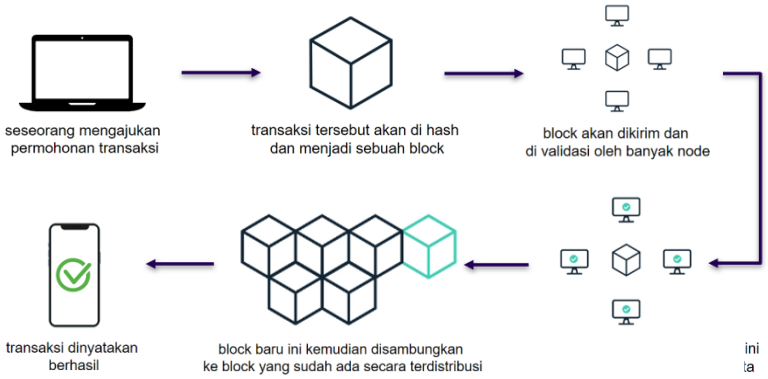
Ada beberapa algoritma konsensus yang berbeda yang digunakan dalam berbagai platform *Blockchain*. Di antara algoritma-algoritma ini, ada dua yang paling umum:

a. *Proof of Work* (PoW): digunakan oleh Bitcoin dan beberapa *Blockchain* lainnya, di mana penambang (*Miners*) harus memecahkan teka-teki

matematika yang rumit untuk menambahkan blok baru ke *Blockchain*. Penambang pertama yang berhasil menyelesaikan teka-teki ini akan diberi hadiah dalam bentuk kripto (Nakamoto, 2009).

b. *Proof of Stake* (PoS): Digunakan oleh *Blockchain* seperti *Ethereum* (yang berencana untukb eralih dari PoW ke PoS), *Cardano*, dan banyaklagi. Di PoS, node yang dipilih untuk mengonfirmasi blok berdasarkan sejumlah koin yang mereka staking (menyimpan dalam dompet mereka). Semakin banyak koin yang dipertaruhkan, semakin besar kemungkinan mereka dipilih untuk memvalidasi transaksi (Panda & Satapathy, 2021).

Pemanfaatan teknologi *Blockchain* memiliki potensi untuk meningkatkan efisiensi dan keamanan sistem yang sudah ada, seperti mengurangi risiko kecurangan, meningkatkan tingkat keterbukaan, dan menurunkan biaya transaksi. Penggunaan *Blockchain* tidak terbatas pada sektor-sektor tertentu, melainkan dapat diterapkan dalam berbagai bidang seperti keuangan, manajemen rantai pasokan, pemilihan umum, layanan kesehatan, dan lainnya. Keunggulan utama dari *Blockchain* meliputi tingkat keamanan yang tinggi, transparansi, dan keotentikan yang terjamin, menjadikannya sebagai sistem yang dapat diandalkan untuk pelaksanaan transaksi secara daring.



Gambar 2.1 Cara Kerja *Blockchain*

Pada gambar 2.1 merupakan cara kerja *Blockchain* adalah sebagai berikut:

1. Sebuah transaksi terjadi antara dua pihak, misalnya A mengirimkan uang kepada B.
2. Transaksi tersebut kemudian diverifikasi oleh sejumlah *node* (komputer) di dalam jaringan *Blockchain*.
3. Setelah diverifikasi, transaksi dimasukkan ke dalam *Blockchain* sebagai blok baru. Blok baru ini terhubung dengan blok-blok sebelumnya, membentuk satu rantai (chain). Pada beberapa *Blockchain*, seperti *Bitcoin*, langkah ini melibatkan miners yang menggunakan algoritma *Proof of Work*. Para *miners* memecahkan teka-teki *kriptografi* kompleks untuk menemukan *nonce* yang, ketika digabungkan dengan data blok, menghasilkan hash blok yang memenuhi kriteria tertentu.
4. Setiap blok dalam *Blockchain* menyimpan informasi tentang transaksi, tanggal dan waktu transaksi, serta hash (kode unik) dari blok sebelumnya. Hash ini digunakan untuk memastikan integritas data dalam blok.
5. Setelah blok baru ditambahkan ke dalam *Blockchain*, transaksi tersebut tidak dapat diubah atau dibatalkan lagi.

Dengan cara kerja yang demikian, *Blockchain* menjadi sistem yang dapat diandalkan dalam hal *keamanan* dan validasi transaksi. Kemampuan *Blockchain* untuk memberikan tingkat transparansi yang tinggi memungkinkan setiap individu yang terhubung ke jaringannya untuk melihat semua transaksi yang terjadi.

2.1.3 *Cryptocurrency*

Cryptocurrency adalah bentuk mata uang digital yang menggunakan teknologi kriptografi untuk menjalankan transaksi. Berbeda dengan mata uang tradisional, *Cryptocurrency* mengandalkan teknologi *Blockchain*, sebuah sistem *Ledger* Terdistribusi *(DLT*) yang mencatat dan mengamankan setiap transaksi dalam serangkaian blok terhubung.(AlMeasam dkk., 2023).

Bitcoin (BTC), *Ethereum* (ETH), *Ripple* (XRP), *Litecoin* (LTC), dan *Solana* (SOL) adalah beberapa contoh *Cryptocurrency* populer saat ini. *Cryptocurrency* telah menjadi populer karena potensinya sebagai bentuk investasi, alat pembayaran, dan teknologi yang dapat mengubah industri keuangan dan banyak sektor lainnya (Hasan & Salah, 2018).

2.1.4 *Solana*

Solana adalah sebuah platform *Blockchain* yang dirancang untuk mendukung aplikasi dan proyek terdesentralisasi. *Solana* menggunakan teknologi yang disebut *Proof-of-History* (PoH) untuk mencatat transaksi secara kronologis sebelum memasukkannya ke dalam blockchain utama. *Proof-of-History* membantu meningkatkan efisiensi dan kecepatan konsensus dalam jaringan *Solana*.

Selain itu, Solana juga menggunakan *Proof-of-Stake* (PoS) sebagai mekanisme konsensus utamanya. PoS memungkinkan pemegang token Solana untuk berpartisipasi dalam validasi transaksi dan pembuatan blok baru. *Solana* dikenal karena kemampuannya untuk menangani jumlah transaksi yang tinggi dengan biaya yang rendah dan kecepatan yang tinggi.

Token *native Solana* disebut SOL, dan *platform* ini telah menjadi basis untuk berbagai macam proyek terdesentralisasi, termasuk aplikasi keuangan, pasar prediksi, dan permainan terdesentralisasi. *Solana* juga menciptakan ekosistem pengembangan yang aktif dan terus berkembang.

2.1.5 NFT

NFT, singkatan dari *Non-Fungible Token*, adalah aset digital yang unik dan diverifikasi melalui teknologi *Blockchain*. Berbeda dengan C*ryptocurrency* seperti *Bitcoin* atau *Ethereum* yang dapat saling dipertukarkan, NFT memiliki karakteristik tunggal yang membuatnya tidak dapat dipertukarkan langsung dengan yang lain.(Mahardika, 2022).

NFT dapat berbentuk beragam digital, termasuk karya seni digital, musik, item dalam permainan video, dan berbagai jenis barang lainnya. Dengan menggunakan teknologi *Blockchain* yang mencatat setiap transaksi, pemilik NFT memiliki kemampuan untuk memverifikasi keaslian dan karakteristik unik dari aset digital tersebut. Mereka juga dapat melacak sejarah kepemilikan yang terkait dengan NFT tersebut.

2.1.6 X

X, sebagai platform media sosial yang terus berkembang, telah menjadi tempat yang ideal bagi pengguna untuk berkomunikasi secara aktif dalam kehidupan sehari-hari. Kecepatan dan kenyamanannya membuatnya menjadi saluran komunikasi publik yang efisien. Selain itu, X juga digunakan untuk melaporkan masalah teknis dan menyebarkan informasi terkait bencana alam yang sedang atau akan terjadi.(Herdhianto, 2020).

X, sebuah *platform* media sosial yang dikenal karena *microblogging*, memungkinkan pengguna untuk berbagi berita real-time melalui pesan singkat yang disebut *"tweet"*. Dengan batasan pesan hanya 140 karakter, X diciptakan untuk memfasilitasi berbagi pengalaman tanpa batasan. Fitur-fitur seperti hashtag dan tren memungkinkan pengguna mengetahui topik populer dengan mudah (Nurrun Muchammad Shiddieqy dkk., 2016).

Fitur *streaming* API X memfasilitasi pengumpulan data dengan *latency* rendah melalui akses ke stream global. Terdapat berbagai tipe *endpoint* yang didukung oleh X, memungkinkan pengembang untuk mengakses data secara efisien. seperti:

1. *PublicStreams*: adalah jenis aliran data X yang memberikan akses ke data publik yang tersedia. Endpoint ini memungkinkan pengguna untuk mencari pengguna tertentu atau topik tertentu, serta melakukan analisis data dengan mudah.
2. *User Streams* adalah Jenis aliran yang terkait dengan satu pengguna, memberikan akses penuh terhadap semua data dan informasi mengenai

aktivitas pengguna tersebut, termasuk keputusan-keputusan yang diambilnya.

3. *Site Streams* adalah streams yang digunakan untuk melakukan pencarian data khusus untuk menemukan semua informasi tentang banyak pengguna. *Endpoint* ini mengharuskan pengembang untuk terhubung ke *X* dengan menggunakan autentikasi banyak pengguna.

2.1.7 *Text Mining*

Text mining, juga dikenal sebagai penambangan teks, adalah proses penggalian informasi berguna, bermakna, dan relevan dari sejumlah besar data teks terstruktur atau semi-terstruktur yang berupa tulisan, dokumen, atau teks dalam bentuk *clustering* dan klasifikasi. Ini adalah bagian dari data mining, di mana data atau teks dan semua dokumen diproses dalam jumlah yang sangat besar. Oleh karena itu, tujuan penambangan teks adalah untuk menemukan informasi berharga yang tersembunyi di balik banyak teks dan dokumen (Firdaus & Firdaus, 2021).

Textmining dapat menyelesaikan masalah seperti pemrosesan, pengorganisasian/pengelompokkan, dan analisis data yang tidak terstruktur dalam jumlah besar. Dalam penelitian ini, data yang digunakan diambil dari *tweet* media sosial X. *Text mining* memanfaatkan dan mengembangkan berbagai teknik dari bidang lain, seperti data mining, pengumpulan informasi, statistik dan matematik, pembelajaran mesin, linguistik, pengolahan bahasa alami, dan visualisasi. Tujuan penelitian *Text mining* adalah untuk mengekstrak dan menyimpan teks, melakukan *preprocessing* teks, mengumpulkan data statistik, dan mengindeks dan menganalisis sentimen(Nurhuda dkk., 2013).

2.1.8 *Text Preprocessing*

Preprocessing teks adalah langkah awal dalam pengolahan data teks yang bertujuan untuk menyiapkan data tersebut agar siap untuk diproses lebih lanjut. Proses ini mencakup berbagai tahapan, seperti membersihkan data teks dari *noise*, membagi teks menjadi token, dan memilih token yang relevan untuk analisis. Tujuan dari *preprocessing* teks adalah untuk menyajikan data teks dalam bentuk yang bersih, terstruktur, dan siap untuk dianalisis lebih lanjut, terutama dalam konteks analisis sentimen. Hal ini diperlukan karena teks yang berasal dari media sosial seringkali tidak formal dan memiliki struktur yang tidak teratur dengan beragam suara yang berbeda.(Rozi & Sulistyawati, 2019).

Proses pengolahan teks melibatkan transformasi data teks yang semula tidak terstruktur menjadi format yang lebih teratur dan terstruktur. Tujuannya adalah untuk menghasilkan data teks yang lebih teratur, bersih, dan tersusun dengan baik sehingga siap untuk tahapan analisis lebih lanjut.(Ramadhanty, 2021) diantaranya:

1. *Case folding* (pengolahanteks)

   Proses ini mencakup menghilangkan tanda baca yang tidak perlu, angka, dan huruf besar, atau "kapital", menjadi huruf kecil.

2. *Tokenization*

   Adalah proses membagi teks atau kalimat menjadi *token-token* yang lebih kecil, seperti kata per kata atau frasa.

3. *Filtering*

Proses ini melibatkan pemilihan token yang relevan dan penghapusan token yang tidak relevan, seperti kata-kata henti atau kata-kata yang tidak relevan untuk teks.

4. *Stemming*

   Ini adalah langkah pertama, mengubah kata imbuhan menjadi kata dasar, seperti mengubah "jogging" menjadi "jog" atau "stemmer" menjadi "stem".

5. *Labelling*

   Adalah langkah yang dilakukan dengan memberikan set karakter atau tanda yang akan digunakan untuk mengidentifikasi variabel atau komponen dari teks atau data. Label diberi kepada token dengan kata penguat atau kata negasi..

2.1.9 Klasifikasi

Dalam proses pengolahan data, klasifikasi merupakan suatu metode yang digunakan untuk mengkaji data dengan tujuan meramalkan nilai dari beberapa atribut tertentu. Algoritma klasifikasi menghasilkan sejumlah aturan yang disebut model, yang berfungsi sebagai panduan untuk mengidentifikasi kelas atau kategori yang diinginkan dari data yang akan diprediksi. Aturan-aturan ini membantu dalam menetapkan pola-pola yang dapat ditemukan dalam data dan mengarahkan ke dalam penentuan kelas yang sesuai dengan prediksi yang dibuat.(Wahyuningsih & Utari, 2018). Salah satu metode yang efektif dalam analisis data adalah klasifikasi. Proses klasifikasi melibatkan tahap pelatihan yang terawasi, dimana data pelatihan, seperti pengamatan atau pengukuran, dilengkapi dengan label yang menandakan kelas dari setiap pengamatan tersebut. Tujuan utama dari klasifikasi adalah untuk

membangun sebuah model berdasarkan data pelatihan beserta label kelas yang terkait, sehingga model tersebut dapat digunakan untuk mengklasifikasikan data baru yang belum pernah dilihat sebelumnya. Dengan penerapan teknik klasifikasi ini, berbagai bidang dapat memanfaatkannya dengan luas, mulai dari deteksi kecurangan dalam bidang pemasaran, prediksi kinerja dalam industri, hingga diagnosis medis yang akurat.(Taufik, 2018).

Klasifikasi merupakan metode yang umum digunakan dalam memprediksi kelas tertentu dari suatu dataset dengan cara mengelompokkan data berdasarkan informasi yang terdapat dalam set pelatihan, beserta dengan nilai-nilai atau label kelas yang sudah diberikan. Dalam dunia klasifikasi, terdapat lima kategori utama yang berbeda berdasarkan konsep matematika yang mendasarinya. Kategori-kategori tersebut mencakup pendekatan statistik, metode berbasis jarak, teknik berbasis pohon keputusan, pendekatan berbasis jaringan saraf, dan strategi berbasis aturan. Dengan memahami perbedaan serta karakteristik dari masing-masing kategori klasifikasi ini, para peneliti dan praktisi data dapat memilih pendekatan yang paling sesuai dengan kebutuhan dan sifat data yang mereka hadapi.(Tangkelayuk, 2022). Proses pencarian model atau fungsi yang mampu memisahkan atau mengklasifikasikan konsep atau kelas dalam suatu dataset dikenal sebagai klasifikasi. Tujuan utama dari klasifikasi adalah untuk mengidentifikasi dan menetapkan kelas yang sesuai terhadap objek atau data yang belum diketahui sebelumnya. Dengan kata lain, klasifikasi berfungsi sebagai mekanisme untuk memahami pola atau karakteristik yang membedakan antara berbagai kelas dalam dataset, sehingga dapat digunakan untuk membuat prediksi atau penentuan terhadap

data baru yang masuk. Dengan adanya proses klasifikasi yang efektif, kita dapat menghasilkan model yang mampu memahami dan mengklasifikasikan data dengan akurasi yang tinggi, sehingga dapat diterapkan dalam berbagai konteks seperti dalam analisis data, pembelajaran mesin, dan pengambilan keputusan.(Wijaya & Santoso, 2016). Klasifikasi dilakukan melalui dua langkah:

1. Proses *training*

   Selama proses ini, set pelatihan yang sudah diketahui labelnya digunakan untuk membangun model.

2. Proses *testing*

   Umumnya digunakan set data validasi untuk menguji keakuratan model yang telah dikembangkan selama fase pelatihan, dengan maksud untuk menilai tingkat akurasi model.

2.1.10 TF-IDF

Metode TF-IDF merupakan suatu teknik yang menghitung bobot untuk setiap kata yang sering muncul dalam kumpulan data yang dianalisis. Pendekatan ini dianggap efisien dan sederhana dalam penggunaannya, sementara juga mampu memberikan hasil yang akurat. Dalam metode ini, dilakukan pengukuran terhadap Frekuensi Kemunculan Kata (TF) dan Frekuensi Dokumen Terbalik (IDF) untuk setiap kata dalam keseluruhan koleksi dokumen. Nilai-nilai ini kemudian digunakan untuk menentukan signifikansi relatif dari masing-masing kata dalam dokumen, membantu dalam mengidentifikasi kata-kata kunci yang memiliki kontribusi paling besar terhadap makna atau tema yang terdapat dalam teks tersebut. (Santoso, 2021), dengan rumus:

tf = 0.5 + 0.5 * (ft,d)/max$_{f_{0i}}$(ft,d)…… **(2.1)**

Frekuensi dokumen merujuk pada jumlah dokumen yang memuat istilah tersebut, bukan sekadar seberapa sering istilah tersebut muncul. Seperti yang dinyatakan dalam rumus sebelumnya, term frequency (frekuensi istilah) dari sebuah dokumen digunakan untuk perhitungan. Inversely, nilai invers dokumen frekuensi (idf) diperoleh dari perhitungan berikut ini:

$$idf = \log\left(\frac{n}{df}\right)$$ ……..……...(**2.2)**

$$w = tf * idf$$ .....**(2.3)**

Keterangan :

d = dokumen ke-d

t = kata kunci ke-t dari kata kunci

w = bobot dokumen ke-d terhadap kaya ke-t

tf = jumlah kata dalam dokumen yang dicari

idf = Inversed Document Frequency

ft,d = frekuensi kata pada d

df = banyak dokumen yang berisi kata pencarian

2.1.11 *Wordcloud*

Wordcloud adalah sebuah alat atau aplikasi yang memungkinkan pengguna untuk menghasilkan visualisasi yang menarik dari teks dengan menyoroti keberadaan kata-kata yang signifikan dalam teks tersebut. Dengan menggunakan *Wordcloud*, pengguna dapat melihat dengan jelas seberapa sering kata-kata tertentu muncul dalam teks, karena kata-kata yang lebih sering muncul akan ditampilkan lebih besar dan lebih menonjol dalam visualisasi.(Sodik & Kharisudin, 2021). Salah

satu teknik visualisasi data yang populer adalah *Wordcloud*. Dalam visualisasi ini, ukuran *font* (huruf) kata-kata berkorelasi dengan frekuensi kemunculannya dalam sebuah set dokumen; semakin sering topik muncul, semakin besar fontnya (Maarif, 2018).

Gambar 2.2 *Wordcloud*

2.1.12 *Support Vector Machine*

Dalam evaluasi akurasi klasifikasi, *Support Vector Machine* menggunakan sekelompok calon untuk membandingkan parameter standar yang berupa nilai diskrit dengan nilai yang telah ditetapkan. Pada tahun 1992, Vapnik, Boser, dan Guyon melakukan pengembangan terhadap *Support Vector Machine* (SVM).(Riadi dkk., 2019). SVM, singkatan dari *Support Vector Machine,* merupakan algoritma yang memberikan label pada objek berdasarkan data yang tersedia. SVM adalah konsep matematika dan algoritma yang bertujuan untuk memaksimalkan suatu fungsi matematika terkait dengan sekumpulan data tertentu. Meskipun begitu, prinsip dasar dari algoritma SVM bisa dijelaskan tanpa harus merujuk pada rumus matematika. (Noble, 2006).

Dalam tahap pelatihan, *Support Vector Machine* (SVM) diajarkan untuk menghitung dan memperbarui bobot dan bias sesuai dengan rumus yang diberikan.

Hasil akhir dari pelatihan adalah menemukan persamaan garis yang analog dengan persamaan matematika tertentu (Nafi'iyah, 2020).

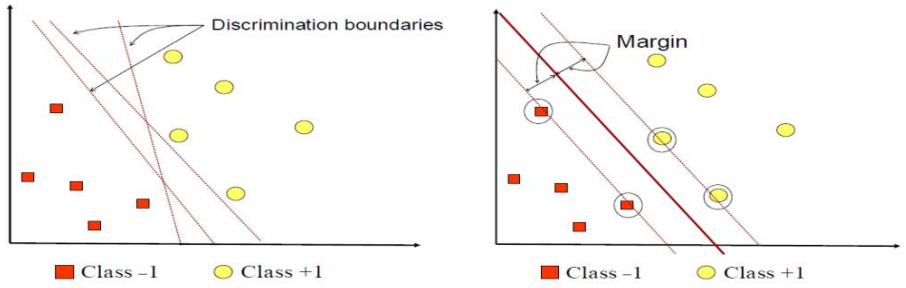
$$W^T X_i + b \geq 1 \quad \textbf{(2.4)}$$

$$W^T X_i + b \geq -1 \quad \textbf{(2.5)}$$

Di mana variabel input adalah persamaan xi, hasil pelatihan dari persamaan sebelumnya adalah persamaan.

$$F_{w,b}(x) = sgn(W^T + b) \quad \textbf{(2.6)}$$

Mencari hyperplane optimal yang berfungsi sebagai pemisah antara dua kelas dalam ruang input adalah cara yang bisa digunakan untuk menjelaskan konsep SVM secara ringkas. (Drajana, 2017).



(a) (b)

Gambar 2. 3 *SVM* berusaha Menemukan *Hyperplane* Terbaik

Dalam Gambar 2a, pola-pola yang termasuk dalam kategori +1 dan –1 direpresentasikan dengan warna merah, sementara pola dari kategori +1 ditampilkan dalam warna kuning. Untuk menyelesaikan masalah klasifikasi ini, langkah pertama adalah mencari sebuah garis (*hyperplane*) yang mampu memisahkan kedua kelompok tersebut. Dalam Gambar 2a, beberapa garis pemisah alternatif, yang juga dikenal sebagai batas keputusan, disajikan. *Hyperplane* terbaik yang dapat memisahkan kedua kelompok ini ditemukan dengan mengukur margin *Hyperplane* dan menentukan titik maksimumnya. Margin merupakan jarak antara

Hyperplane dengan pola terdekat dari setiap kelas. Gambar 2b menampilkan garis solid yang mewakili hyperplane terbaik, yang terletak tepat di tengah-tengah kedua kelas, sementara lingkaran hitam menunjukkan titik-titik merah dan kuning yang disebut sebagai *Support Vector*. Proses pembelajaran SVM bergantung pada penempatan *Hyperplane* ini untuk memisahkan dengan baik kedua kelas dalam ruang fitur yang digunakan.

Dengan menerapkan pendekatan kernel, SVM memiliki kemampuan untuk mengatasi data yang bersifat non-linier dengan menggunakan ciri-ciri asli dari dataset. Fungsi kernel ini bertugas untuk mengubah ruang fitur dari himpunan data yang awalnya memiliki dimensi yang lebih rendah menjadi ruang fitur yang memiliki dimensi yang lebih tinggi, sehingga memungkinkan SVM untuk melakukan pemisahan yang lebih baik antara kelas-kelas yang kompleks dalam data non-linier.(Rizki dkk., 2021). Ada beberapafungsi*kernel*, diantaranya :

1. *Kernel Linear*

$$K(x, y) = x.y \quad \text{..... (2.7)}$$

2. *Kernel Polynomial*

$$K(x, y) = (x.y)^d \quad \text{..... (2.8)}$$

3. *Kernel Radial Basis Function (RBF)*

$$K(x_i, x_j) = \exp(-y|x_i, x_j|)^2 \quad \text{..... (2.9)}$$

4. *Kernel Sigmoid*

$$K(x_i, x_d) = \tan y\, X_i^T X^j + r \quad \text{..... (2.10)}$$

2.1.13 *Random Forest*

Metode Random Forest adalah salah satu teknik machine learning yang digunakan untuk klasifikasi dan regresi. Metode ini berdasarkan konsep agregasi (ensemble) dari beberapa pohon keputusan (decision tree) yang dikombinasikan untuk menghasilkan prediksi yang lebih akurat dan stabil. Setiap pohon keputusan pada Random Forest dibangun secara acak dan independen, dengan cara melakukan sampling acak pada data pelatihan dan variabel fitur yang digunakan dalam pembentukan pohon (Primajaya & Sari, 2018).

Pendekatan Random Forest yang diusulkan oleh Breiman adalah algoritma pembelajaran mesin dengan banyak pohon keputusan. Random forest adalah kombinasi dari metode Bagging dan Random Sub spaces (Apriliah dkk., 2021). Kombinasi Bagging dan Random Subspaces membuat Random Forest menjadi algoritma yang unggul dalam berbagai tugas regresi dan klasifikasi. Dengan memanfaatkan banyak pohon keputusan yang bekerja secara independen, Random Forest dapat mengurangi overfitting dan meningkatkan akurasi prediksi pada data yang belum pernah dilihat sebelumnya.

Metode Random Forest adalah metode pelatihan berbasis ensemble learning yang menggunakan algoritma Decision Tree. Kemudian, prediksi dibuat dari beberapa model Decision Tree menggunakan data uji yang sama di setiap model. Hasil prediksi kemudian dibuat melalui proses majority voting menggunakan metode modus dan akhirnya kelas dari modus tersebut menjadi kelas prediksi. Rumus untuk menghitung Entropy dan untuk menghitung Gain pada metode Random Forest (Sejati dkk., 2022).

$$Entropy\ (E) = -\sum_{i=1}^{n} p_i\ log_2\ p_i \quad \textbf{(2.11)}$$

Keterangan:

E : ruang (data) sampel yang digunakan untuk training

n : banyaknya partisi pada E

p_i : peluang terpilihnya contoh secara acak di kelas i.

Untuk rumus persamaan nilai Information Gain, dapat menggunakan rumus:

$$Gain = Entropy_{parent} - Entropy_{parent} \quad \textbf{(2.12)}$$

Keterangan

E_{parent} adalah entropi dari node induk

$E_{children}$ adalah entropi rata-rata dari node anak.

2.1.14 Evaluasi Klasifikasi

Setelah tahap pengolahan data dan klasifikasi teks, langkah terakhir adalah evaluasi. Di dalam tahap ini, hasil eksperimen akan diperiksa, dibandingkan, dan dianalisis berdasarkan kinerja klasifikasi yang telah dilakukan. Evaluasi umumnya melibatkan pertimbangan terhadap hasil implementasi metode yang digunakan dalam analisis sentimen, dan sering kali menggunakan *Confusion Matrix* sebagai alat untuk menganalisis performa *Confusion Matrix*.(Rizki dkk., 2021). Tabel yang mengklasifikasikan jumlah data uji yang benar dan salah dikenal sebagai *Confusion Matrix* (Normawati & Prayogi, 2021). Tabel 2.1 menunjukkan contoh *Confusion Matrix* untuk klasifikasi *Biner*.

Tabel 2.1 *Confusion Matrix*

Predicted Class	***True Class***	
	Positif	*Negatif*
Positif	*True Positif (Tp)*	*False Positif (Fp)*
Negatif	*False Negatif (Fn)*	*True Negatif (Tn)*

Untuk menghitung tiga metode evaluasi, yaitu:

1. *Recall*, yang mencerminkan perbandingan antara jumlah dokumen yang relevan yang berhasil diidentifikasi dengan total jumlah dokumen yang sebenarnya relevan, dihitung menggunakan rumus persamaan.

$$\frac{TP}{TP+FN} * 100\% \quad \textbf{(2.11)}$$

2. *Precision*yang menunjukkan perbandingan antara jumlah dokumen yang relevan yang berhasil diidentifikasi dengan total jumlah dokumen yang diidentifikasi, dihitung menggunakan rumus persamaan.

$$\frac{TP}{TP+FP} * 100\% \quad \textbf{(2.12)}$$

3. *Accuracy*, yang merupakan kemampuan sistem untuk mengklasifikasikan data dengan akurat, diukur menggunakan rumus persamaan tertentu.

$$\frac{TP+TN}{TN+TP+FN+FP} * 100\% \quad \textbf{(2.13)}$$

4. *f1-Scores,* adalah sebuah metrik kinerja yang menggabungkan presisi dan *recall,* yang digunakan untuk menilai kualitas secara keseluruhan dari suatu model klasifikasi.

$$2 * \frac{Presisi*Recal}{Presisi+recal} * 100\% \quad \textbf{(2.14)}$$

*Recall, precision, dan accuracy*dinyatakan dalam bentuk nilai *f1-scores,* yang diekspresikan dalam persentase. Tingkat persentase untuk ketiga metrik ini mengindikasikan bahwa kinerja sistem klasifikasi teks otomatis meningkat.

2.1.15 *Python*

Python adalah bahasa pemrograman interpretatif dan serbaguna yang menekankan kejelasan kode untuk memudahkan pembacaan dan pemahaman. Keanekaragaman fiturnya membuatnya cocok untuk berbagai keperluan seperti pengembangan web, analisis data, kecerdasan buatan, dan pemrosesan teks. Keunggulan lainnya adalah produktivitas dan efisiensinya berkat pustaka standar yang kaya fitur. (Syahrudin & Kurniawan, 2018).

2.1.16 *Library Tweet Harvest*

Tweet Harvest adalah praktik mengumpulkan atau mengekstrak *tweet* dari platform media sosial seperti *Twitter* untuk berbagai tujuan, seperti analisis sentimen, riset pasar, atau pemantauan tren. Ini melibatkan penggunaan alat atau perangkat lunak khusus yang memungkinkan pengguna untuk menarik *tweet* yang relevan berdasarkan kata kunci, tanggal, lokasi, atau parameter lainnya. Praktik ini dapat memberikan wawasan berharga tentang opini, kecenderungan, atau pola perilaku di platform media sosial tersebut. Namun, penting untuk memperhatikan privasi pengguna dan kebijakan platform saat melakukan aktivitas semacam ini. Berikut proses crawling tweet pmenggunakan tweet harvest:

1. Siapkan auth token akun *twitter*.
2. *Install library Tweet Harvest*.
3. Ganti *auth token* dengan *auth* token akun *twitter*.
4. Buka terminal atau *command prompt*.

5. Navigasikan ke direktori tempat Anda menyimpan file tweet_harvest.py.
6. Jalankan script dengan perintah python tweet_harvest.py.
7. Tungguhingga proses *crawling* selesai,

2.1.17 *Pandas Library*

Pandas adalah pustaka atau library Python yang populer digunakan untuk manipulasi dan analisis data. Ini menyediakan struktur data dan alat untuk bekerja dengan data terstruktur, terutama dalam bentuk tabel seperti yang sering ditemui dalam spreadsheet dan database.

Berikut beberapa fitur utama Pandas:

1. *DataFrame*: Ini adalah struktur data utama dalam Pandas. *DataFrame* adalah representasi dua dimensi dari data yang dapat berupa berbagai jenis, termasuk numerik, string, atau bahkan objek *Python* lainnya. Setiap kolom dalam *DataFrame* dapat memiliki tipe data yang berbeda.
2. Series: Ini adalah struktur data satu dimensi yang mirip dengan *array* atau list, tetapi lebih kuat dan fleksibel. Series sering digunakan untuk merepresentasikan satu kolom dari *DataFrame*.
3. Membaca dan Menulis Data: Pandas memiliki kemampuan yang kuat untuk membaca dan menulis data dari berbagai sumber seperti CSV, *Excel*, SQL *databases*, dan lebih banyak lagi.
4. Manipulasi Data: Pandas menyediakan berbagai fungsi untuk membersihkan, memanipulasi, dan mentransformasi data. Ini

termasuk penghapusan baris dan kolom, penggabungan dan penyatuan data, pengurutan, pengindeksan ulang, dan banyak lagi.

5. Pengindeksan dan Pemilihan: Anda dapat mengakses, memfilter, dan memanipulasi data menggunakan indeks dan label kolom dengan sangat mudah menggunakan Pandas.
6. Operasi Data: Pandas mendukung operasi matematika dan statistik pada data, seperti agregasi, penghitungan statistik deskriptif, dan pemodelan data.
7. Visualisasi Data: Meskipun Pandas tidak memiliki kemampuan visualisasi bawaan seperti *Matplotlib* atau *Seaborn*, namun sangat mudah berintegrasi dengan pustaka visualisasi ini untuk membuat grafik dan visualisasi data.
8. Penanganan *Missing* Data: *Pandas* memiliki fungsi untuk menangani nilai yang hilang (*missing values*) dalam data, seperti mengisi nilai yang hilang atau menghapus baris atau kolom dengan nilai yang hilang.

2.2. Kajian Penelitian

Tabel 2.2 Kajian Penelitian

No	Judul Artikel Ilmiah	Penulis (Tahun)	Nama Jurnal Penerbit	Hasil Penelitian
1.	Aplikasi Klasifikasi Sentimen Pada Ulasan *Smartphone* Di Situs Jual Beli *Online* Berbasis Web Menggunakan *Naive Bayes* Dengan Tf-Idf	Risky Kararisma, Sri Lestanti, M Taofik Chulkamdi (2022)	JATI (Jurnal Mahasiswa Teknik Informatika)	Proses pengembangan aplikasi klasifikasi web ini mengikuti lima tahap Siklus Hidup Pengembangan Perangkat Lunak (SDLC) dan menggunakan *Framework Laravel* sebagai basisnya. Ulasan dikumpulkan dan diberi bobot dengan metode *TF-IDF*, lalu disimpan dalam basis data. Hasil pengujian menunjukkan tingkat akurasi mencapai 93%, dengan nilai precision dan recall di atas 85%, menunjukkan kinerja yang baik dari algoritma *Naive Bayes*. Kesimpulan dari penelitian ini menegaskan bahwa aplikasi klasifikasi web menggunakan algoritma *Naive Bayes* dapat secara otomatis menganalisis sentimen dan membantu dalam evaluasi kepuasan pelanggan.
2.	Analisis Sentimen Pengguna Kereta Api Indonesia Melalui Sosial Media X Dengan Algoritma *Naïve Bayes Classifier*	M Azahri, N Sulistiyowati, M Jajuli (2023)	JATI (Jurnal Mahasiswa Teknik	Penelitian ini menggunakan algoritma *Naïve Bayes Classifier* dan metode *Knowledge Discovery in Databases* (KDD) untuk mengkategorikan sentimen pengguna layanan transportasi Kereta Api Indonesia. Hasilnya menunjukkan tingkat akurasi klasifikasi sentimen sebesar 92.15%, dengan mayoritas sentimen *negatif* terkait kenaikan harga tiket, kondisi kursi, dan fasilitas lainnya. Studi ini dapat memberikan panduan kepada Kereta Api Indonesia untuk meningkatkan kualitas layanan berdasarkan umpan balik pengguna. Dengan menganalisis sentimen, perusahaan dapat merespons keluhan pengguna secara lebih efektif dan meningkatkan kepuasan pelanggan.
3.	Analisis Sentimen Tanggapan Masyarakat Terhadap Pembelajaran *Daring* di Masa Pandemi COVID-19 Melalui Sosial Media X Menggunakan Klasifikasi *Naïve Bayes*	Z Zulfachmi, NM Kautsar, ZA Ramadhan(2023)	*Prosiding* Seminar Implementasi Teknologi Informasi dan Komunikasi	Penelitian ini bertujuan untuk mengevaluasi pandangan masyarakat terhadap pembelajaran daring selama masa pandemi COVID-19 menggunakan metode analisis sentimen *Naive Bayes*. Hasil analisis data dari *platform* X menunjukkan bahwa mayoritas masyarakat menunjukkan sentimen *negatif* terhadap pembelajaran daring, mencapai 45,33%. Meskipun demikian, terdapat juga 28,67% sentimen *positif*, sementara 26% bersifat *netral*. Algoritma *Naive Bayes* mencapai tingkat akurasi sebesar 80,67%, menandakan efektivitasnya dalam mengevaluasi respons masyarakat terhadap pembelajaran daring selama pandemi. Penelitian ini memberikan wawasan yang

				penting mengenai dinamika dan tantangan pembelajaran daring di Indonesia.
4.	Analisis Sentimen Pengguna Aplikasi E-*Wallet* Dana Melalui Postingan di Media Sosial X Menggunakan *Naïve Bayes*	O Hondro (2023)	KETIK: Jurnal Informatika	Penelitian ini mengeksplorasi pendapat pengguna tentang aplikasi DANA di X menggunakan metode *Naïve Bayes*. Meskipun DANA memiliki popularitas tinggi dengan lebih dari 10 juta unduhan, hasil analisis sentimen menunjukkan variasi opini *positif* dan *negatif* dari pengguna. Pengguna cenderung berbagi pendapat mereka secara terbuka di *platform* ini. Pendekatan penelitian menggunakan *Knowledge Discovery* in *Databases* (KDD), meliputi langkah-langkah seleksi data, *preprocessing*, transformasi, data *mining*, dan evaluasi. Algoritma *Naïve Bayes* dengan *kernel linear* memberikan akurasi tertinggi sebesar 98.7% dalam mengklasifikasikan sentimen aplikasi DANA di X, menegaskan keefektifan metode ini.
5.	AnalisisSentimenXTerhadap *Bitcoin* dan *Cryptocurrency* Berbasis *Python TextBlob*	R Parlika, SI Pradika, AM Hakim(2020)	JurnalIlmiahTeknologi	Penelitian ini menganalisis pendapat pengguna X tentang *Bitcoin* menggunakan analisis sentimen dengan *Python* dan teknik *Word Cloud*. Dengan mengumpulkan data tweet melalui API X, studi ini mengkategorikan sentimen menjadi *positif* (41,3%), netral (44,9%), dan *negatif* (13,7%). Visualisasi dengan *Word Cloud* menampilkan kata-kata kunci seperti "*blockchain*," "harga," dan "beli," yang mencerminkan minat positif dan kekhawatiran terhadap pasar *Bitcoin*. Temuan menunjukkan bahwa *Bitcoin* mendapat pengakuan *positif*, tetapi juga ada sentimen negatif. Penelitianinimemberikanwawasanpentingtentangpersepsipengguna X terhadap *Bitcoin*.
6	Model Klasifikasi *Multinomial Naive Bayes* UntukAnalisis *Sentiment* Terkait *Non-Fungible Token*	RYL Lesmana, R Andarsyah (2022)	Jurnal Teknik Informatika	Penelitian ini menunjukkan bahwa Model Klasifikasi *Multinomial Naïve Bayes* berhasil menganalisis sentimen pengguna X terhadap *Non-Fungible Token* (NFT). Dari 7060 tweet dengan tagar #NFT, hasil analisis sentiment menggunakan metode Vader menunjukkan 3840 tweet dengan sentimen *negatif* dan 3220 tweet dengan sentimen *positif*. Setelah tahap klasifikasi dengan metode *Multinomial Naïve Bayes*, model mencapai tingkat akurasi 84%, dengan precision 84%, recall 81%, dan f1-score 83%. Hasil dari *Confusion Matrix* menunjukkan kinerja model yang baik dengan *True Positive* (830), *True Negative* (661), *False Positive* (122), dan *False Negative* (152). Temuan ini memberikan wawasan penting tentang persepsi publik terhadap NFT di X.

7.	Analisis Sentimen Masyarakat Indonesia Terhadap *Non Fungible Token* (Nft) Menggunakan Algoritma *Naïve Bayes Classifier*	DT Mahardika (2022)	repository.nusaputra.ac.id	Penelitian ini memberikan kontribusi penting dalam memahami persepsi masyarakat Indonesia terhadap *Non-Fungible Token* (NFT) melaluianalisissentimen data darikomentar *YouTube*. Metode *Naïve Bayes Classifier* digunakan untuk mengklasifikasikan sentimen menjadi tiga kelas: *Positif*, *Negatif*, dan *Netral*. Hasilnya menunjukkan mayoritas masyarakat Indonesia menyambut NFT dengan pandangan positif (64%), menunjukkan antusiasme yang tinggi. Meskipun informasinya masih terbatas, penelitian ini memberikan wawasan tentang potensi NFT sebagai alat pemasaran bagi seniman, sambil juga menyoroti risiko kejahatan di dunia digital. Kesimpulan dari penelitian ini dapat menjadi panduan bagi masyarakat, peneliti, dan pengembangan lebih lanjut dalam studi serupa.
8	Implementasi Algoritma K-*Means* Dalam Pembelian NFT Pada Jaringan *Solana*	MRR Harahap, T Arifin (2023)	Jurnal Nasional Komputasi dan Teknologi Informasi	Penelitian ini menekankan analisis NFT di platform Solana dengan menggunakan metode *Clustering* menggunakan algoritma K-*Means*. Dari dataset *Magic Eden*, hasil pengelompokan data mengidentifikasi lima kelompok dengan karakteristik yang berbeda. Kelompok 0, dengan rating 1, terdiri dari 389 item seperti *Nekoverse* dan *Carton Kids*. Kelompok 1 (rating 3) mencakup tiga item seperti *Solana Monkey Business*. Kelompok 2 (rating 4) memiliki dua item seperti Degods. Kelompok 3 (rating 5) hanya terdiri dari satu item, yaitu Okay Bears. Sedangkan Kelompok 4 (rating 2.3) terdiri dari 24 item dengan variasi seperti *Thugbirdz* dan Primates. Temuan ini memberikan wawasan berharga bagi investor NFT untuk membuat keputusan yang lebih terinformasi.
9	*Solana: A new architecture for a high performance blockchain v0. 8.13*	A Yakovenko (2018)	*WhitepaperGithub*	Penelitian Yakovenko pada tahun 2018 menghasilkan kontribusi berupa pengembangan arsitektur *Blockchain* baru, dikenal sebagai *Solana*. Arsitektur ini, terutama dalam versi 0.8.13, menonjolkan kinerja tinggi dalam teknologi *Blockchain*. Dari whitepaper yang dihasilkan, penelitian tersebut memberikan detail tentang desain *Solana*, yang sangat berarti dalam evolusi teknologi *Blockchain* yang lebih maju. Dengan fokus pada peningkatan kinerja, *Solana* menawarkan solusi untuk meningkatkan efisiensi dan skalabilitas dalam transaksi *Blockchain*, membuka pintu bagi inovasi dalam pengembangan aplikasi terdesentralisasi. Kontribusi dari penelitian ini menjadi elemen penting dalam

				mendorong perkembangan teknologi *Blockchain* ke arah yang lebih efisien dan inovatif.
10.	*Can solana be the solution to the blockchain scalability problem?*	GA Pierro, R Tonelli (2022)	2022 IEEE International Conference on Communications	Penelitian ini mengeksplorasi platform blockchain Solana sebagai solusi untuk masalah skalabilitas yang umum terjadi dalam teknologi *Blockchain*. Diluncurkan pada April 2018, Solana bertujuan untuk meningkatkan kemampuan skalabilitasnya tanpa mengorbankan aspek desentralisasi dan keamanan. Data dari *Blockchain Solana* dikumpulkan selama dua bulan untuk memverifikasi beberapa karakteristiknya, termasuk *throughput* transaksi. Hasil analisis menunjukkan bahwa rata-rata *throughput* transaksi *Solana* mencapai 2812 TPS, dengan biaya pengguna untuk mengkonfirmasi transaksi jauh lebih rendah daripada *blockchain* lain yang serupa. Penelitian ini memberikan pemahaman lebih dalam tentang mekanisme Solana, yang didesain untuk mengatasi masalah skalabilitas tanpa mengorbankan aspek desentralisasi dan keamanan. Data penelitian ini tersedia secara publik di repositori *GitHub*.

Penelitian ini dilakukan dengan merujuk pada hasil-hasil penelitian terdahulu sebagai bahan perbandingan dan kajian. Hasil-hasil penelitian sebelumnya yang menjadi perbandingan tidak lepas dari topik penelitian, yaitu algoritma *Naïve Bayes Classifier*. Berikut adalah beberapa hasil penelitian terdahulu yang menjadi acuan dalam penelitian ini.

Penelitian yang dilakukan oleh M Azahri, N Sulistiyowati, M Jajuli pada tahun 2023 dengan judul Analisis Sentimen Pengguna Kereta Api Indonesia Melalui Sosial Media X Dengan Algoritma *Naïve Bayes Classifier* ini mengaplikasikan algoritma *Naïve Bayes Classifier* untuk menganalisis sentimen pengguna layanan Kereta Api Indonesia melalui data yang diperoleh dari media sosial X, khususnya Twitter. Dengan menggunakan metode KDD (*Knowledge*

Discovery in Database), penelitian ini melalui tahapan Data *Selection*, *preprocessing*, Data *transformation*, Data *Mining*, dan *Evaluation* untuk meningkatkan kualitas data dan hasil analisis. Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes Classifier* memiliki tingkatan kurasi yang tinggi, mencapai 92.15%. Melalui analisis sentimen, penelitian ini dapat mengekstrak informasi dari opini pengguna dan mengklasifikasikannya menjadi kategori *positif* dan *negatif* terhadap layanan yang diberikan oleh Kereta Api Indonesia. Penelitian ini juga memberikan wawasan tentang perbedaan pendapat pengguna terkait berbagai aspek layanan, seperti fasilitas kereta dan kenyamanan perjalanan. Metode analisis yang digunakan didukung oleh tools *Python*, *Scikit-learn*, dan *Jupyter Notebook*, serta memperhatikan konsep-konsep seperti TF-IDF dan *Confusion Matrix*.

Kebaruan Kelebihan penelitian M.Azahri (2023) tentang analisis *sentimen* pengguna layanan Kereta Api Indonesia menggunakan algoritma *Naïve Bayes Classifier* dan data dari *Twitter*. Penelitian ini berhasil mencapai tingkat akurasi sebesar 92.15% dan memberikan wawasan tentang opini pengguna terkait layanan kereta api, termasuk aspek fasilitas dan kenyamanan perjalanan. Di sisi lain, penelitian Anda tentang analisis sentimen NFT pada *Blockchain Solana* menggunakan metode klasifikasi *Support Vector Machine* dan data dari media sosial yang belum dijelaskan. Dengan fokus pada teknologi *Blockchain* dan aset digital NFT, penelitian Anda membawa perspektif baru dalam analisis sentimen di domain yang berbeda.

Persamaan dengan penelitian "Analisis Sentimen NFT Pada *Blockchain Solana* Melalui Media Sosial X Menggunakan Metode Klasifikasi *Support Vector Machine*":

a. Kedua penelitian bertujuan untuk menganalisis sentimen pengguna terhadap suatu subjek tertentu (layanan Kereta Api Indonesia dan NFT di *Blockchain Solana*) melalui data yang diperoleh dari media sosial X.
b. Kedua penelitian menggunakan metode klasifikasi untuk mengklasifikasikan sentimen dari data yang dikumpulkan. Salah satunya menggunakan algoritma *Naïve Bayes Classifier*, sedangkan yang lain menggunakan metode klasifikasi *Support Vector Machine* (SVM).
c. Kedua penelitian memanfaatkan tools dan bahasa pemrograman untuk melakukan analisis data. Salah satunya menggunakan *Python*, *Scikit-Learn*, dan *Jupyter Notebook*, dan yang lain mungkin akan menggunakan alat dan bahasa pemrograman yang serupa.
d. Kedua penelitian memiliki proses analisis data yang melibatkan tahapan preprocessing dan evaluasi, serta memperhatikan konsep-konsep yang relevan dalam analisis sentimen seperti TF-IDF dan *Confusion Matrix*.
e. Meskipun kedua penelitian memiliki persamaan dalam tujuan dan metodologi analisis sentimen, subjek yang dianalisis dan algoritma klasifikasi yang digunakan berbeda.