Integrative Research in Computer Science
2025, vol. 1, no. 1

KOMBINASI NAÏVE BAYES DAN CHI-SQUARE UNTUK IDENTIFIKASI SMS PENIPUAN

Aditya Priadi Pradana, Arry Maulana Syarif,
Ika Novita Dewi*, Candra Irawan
Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Indonesia

*Penulis Korespondensi: ikadewi@dsn.dinus.ac.id

Abstrak: Ancaman kejahatan siber seperti SMS penipuan telah menjadi masalah serius yang berpotensi mengakibatkan kerugian finansial dan pencurian data pribadi. Penelitian ini bertujuan untuk merancang dan membangun sebuah sistem deteksi yang efektif guna mengklasifikasikan SMS penipuan secara akurat dengan memanfaatkan pendekatan machine learning (ML). Pendekatan yang digunakan adalah penerapan algoritma klasifikasi Naïve Bayes, sebuah metode probabilistik yang dikenal efisien untuk analisis teks. Proses penelitian diawali dengan pengumpulan dataset SMS yang relevan, diikuti oleh tahap pra-pemrosesan data yang komprehensif, mencakup case folding untuk menyeragamkan teks, normalisasi untuk standardisasi kata, stopwords removal untuk eliminasi kata-kata umum, serta stemming untuk mengubah kata ke bentuk dasarnya. Selanjutnya, fitur-fitur teks diekstraksi dan dibobot menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF), dan fitur yang paling signifikan diseleksi menggunakan Chi-Square untuk meningkatkan fokus model. Hasil pengujian dan evaluasi, yang didasarkan pada confusion matrix, menunjukkan performa model yang sangat baik, dengan keberhasilan mencapai tingkat akurasi sebesar 93%. Lebih lanjut, model ini juga menunjukkan keseimbangan yang kuat antara presisi (93%), recall (93%), dan F1-Score (93%). Capaian ini menegaskan bahwa model Naïve Bayes merupakan solusi yang andal dan valid untuk mengembangkan sistem perlindungan pengguna yang efektif terhadap ancaman SMS penipuan.

Kata Kunci: Naïve Bayes, TF-IDF, Chi-Square, Klasifikasi, SMS Penipuan

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah menjadikan pesan singkat (*Short Message Service*/SMS) sebagai salah satu media komunikasi yang paling banyak digunakan. Namun, di sisi lain, hal ini juga memicu maraknya penyalahgunaan SMS untuk tindak penipuan. SMS penipuan umumnya dirancang untuk mengecoh penerima dengan menyamar sebagai pesan resmi dari institusi terpercaya seperti bank, e-commerce, atau instansi pemerintah. Pelaku memanfaatkan teknik social engineering dengan menciptakan rasa urgensi, iming-iming hadiah,

Cara Mengutip:
Pradana, A. P., Syarif, A. M., Dewi, I. N., & Irawan, C. (2025). Kombinasi naive bayes dan chi-square untuk identifikasi sms penipuan. IRCS: Integrative Research in Computer Science, 1(1), 1-22.

atau ancaman agar korban secara tidak sadar memberikan data pribadi atau melakukan transfer dana. Data dari Kementerian Kominfo, tercatat 958 kasus dalam kurun waktu Agustus hingga November 2023 [1]. Tingginya angka ini mengindikasikan perlunya solusi deteksi yang lebih efektif untuk melindungi masyarakat dari ancaman tersebut.

SMS penipuan adalah salah satu modus kejahatan siber yang dilakukan dengan mengirimkan pesan kepada seseorang atau banyak orang dengan tujuan mengecoh penerima agar memberikan informasi pribadi, melakukan tindakan tertentu yang menguntungkan pelaku, atau bahkan mentransfer uang [2]. Tindakan ini memanfaatkan kelengahan, rasa ingin tahu, atau ketakutan korban. Modus-modus penipuan ini sering kali dirancang agar terlihat seperti pesan resmi atau penting, sehingga penerima merasa perlu untuk segera merespons. Contoh umum dari SMS penipuan adalah meminta penerima untuk menghubungi nomor tertentu atau mentransfer sejumlah uang untuk "biaya administrasi" atau "pajak hadiah." Pelaku juga dapat menyertakan tautan atau nomor yang harus dihubungi, yang sebenarnya adalah bagian dari strategi untuk menjebak korban.

Dalam beberapa kasus, pelaku menggunakan nama perusahaan besar atau lembaga pemerintah untuk meminta informasi pribadi, seperti nomor kartu kredit, PIN, atau bahkan kode OTP (*One-Time Password*), dengan alasan adanya masalah pada akun atau perlunya pembaruan data. Bentuk lain dari SMS penipuan adalah pesan yang mengancam, seperti peringatan pemblokiran akun jika tidak segera direspon. Ancaman ini bertujuan untuk membuat korban merasa terdesak dan akhirnya mengikuti instruksi pelaku tanpa berpikir panjang. Teknik-teknik penipuan ini terus berkembang, dan pelaku semakin cerdas dalam merancang isi pesan agar terlihat meyakinkan, sehingga masyarakat perlu selalu waspada terhadap pesan-pesan mencurigakan yang berpotensi merugikan. SMS penipuan bekerja dengan cara memanfaatkan kelengahan, ketakutan, atau rasa ingin tahu korban melalui pesan singkat yang disusun agar terlihat mendesak [3].

Sistem deteksi SMS penipuan yang diusulkan bertujuan untuk membantu pengguna mengidentifikasi pesan penipuan dengan menggunakan metode Naive Bayes, Algoritme Naive Bayes dipilih karena kemampuannya yang efisien dan sederhana dalam menangani klasifikasi teks. Naive Bayes adalah algoritme probabilistik yang menghitung kemungkinan suatu pesan termasuk dalam kategori tertentu berdasarkan pola kata yang muncul dalam pesan tersebut [4]. Dalam kasus deteksi SMS penipuan, Naive Bayes sangat cocok karena SMS penipuan sering mengandung kata-kata yang spesifik dan berbeda dari pesan yang sah. Naive Bayes juga tidak memerlukan banyak waktu untuk pelatihan dan memiliki kinerja yang cepat dalam memproses data teks dalam jumlah besar. Selain itu algoritme Naive Bayes mampu memberikan akurasi yang baik untuk masalah klasifikasi teks dengan data yang terstruktur seperti pesan SMS, sehingga menjadikannya pilihan yang ideal untuk membangun sistem deteksi SMS penipuan yang akurat.

Penelitian ini bertujuan mengembangkan sistem deteksi SMS penipuan menggunakan algoritma Naïve Bayes yang diperkuat dengan seleksi fitur Chi-Square. Naïve Bayes dipilih karena kemampuannya dalam menangani klasifikasi teks secara efisien, bahkan dengan data terbatas, sementara Chi-Square digunakan untuk meningkatkan akurasi dengan menyaring fitur-fitur yang kurang relevan. Pendekatan ini diharapkan dapat mengidentifikasi pola linguistik khas dalam SMS penipuan, sehingga mampu membedakannya dari pesan biasa. Studi ini berfokus pada analisis dataset SMS penipuan tahun 2020 dengan klasifikasi biner (penipuan atau bukan) untuk memastikan model bekerja secara optimal.

Dari segi kontribusi, penelitian ini tidak hanya memperkaya literatur di bidang machine learning dan keamanan siber, tetapi juga memberikan solusi praktis bagi masyarakat dalam menghindari potensi penipuan. Dengan adanya sistem deteksi otomatis, diharapkan tingkat kewaspadaan pengguna terhadap SMS mencurigakan dapat meningkat, sehingga mengurangi risiko kerugian finansial dan psikologis. Temuan dari penelitian ini juga dapat menjadi landasan bagi pengembangan metode deteksi penipuan yang lebih canggih di masa depan.

2. TINJAUAN PUSTAKA

Data mining adalah proses analisis data yang bertujuan untuk menemukan pola, tren, dan informasi berharga dari kumpulan data besar. Data mining dapat didefinisikan sebagai proses menemukan pola yang berguna dari data besar [5]. Proses ini melibatkan penggunaan teknik statistik, algoritma pembelajaran mesin, dan basis data untuk mengubah data mentah menjadi informasi yang dapat digunakan untuk pengambilan keputusan yang lebih baik. Data mining sering kali dianggap sebagai bagian dari ilmu data yang lebih luas, yang mencakup pengumpulan, penyimpanan, dan analisis data. Data mining berfungsi untuk memfasilitasi pekerjaan dengan data yang banyak. Data mining dianggap sebagai bidang penelitian yang penting dan digunakan di berbagai bidang seperti deteksi penipuan, perbankan keuangan, pendidikan, kesehatan, pertanian, industri, dan yang lainnya [6].

Data mining mencakup berbagai teknik yang digunakan untuk menganalisis data. Beberapa teknik utama meliputi: klasifikasi, regresi, clustering, asosiasi dan deteksi anomali. ML adalah cabang dari kecerdasan buatan (*Artificial Intelligence*/AI) yang berfokus pada pengembangan algoritma dan model yang memungkinkan komputer untuk belajar dari data dan membuat prediksi atau keputusan tanpa perlu diprogram secara eksplisit [7]. Proses ini melibatkan penggunaan teknik statistik dan matematis untuk menganalisis data, mengenali pola, dan menggeneralisasi informasi dari data yang telah dipelajari. ML dibagi menjadi beberapa kategori utama, termasuk supervised learning, di mana model dilatih menggunakan data yang sudah diberi label, unsupervised learning, yang melibatkan analisis data tanpa label untuk menemukan struktur atau pola yang mendasarinya, dan

reinforcement learning, di mana agen belajar melalui interaksi dengan lingkungan dan mendapatkan umpan balik dalam bentuk *reward* atau *punishment*.

Aplikasi ML sangat luas, mencakup berbagai bidang seperti kesehatan, di mana ia digunakan untuk diagnosis penyakit dan analisis citra medis, keuangan, untuk deteksi penipuan dan analisis risiko, serta pemasaran, untuk segmentasi pelanggan dan rekomendasi produk. Meskipun memiliki potensi besar, ML juga menghadapi tantangan, seperti overfitting, di mana model terlalu kompleks dan tidak dapat generalisasi dengan baik, serta masalah interpretabilitas, di mana model yang kompleks sulit untuk dipahami dan dijelaskan. Dengan kemajuan teknologi dan peningkatan ketersediaan data, machine learning terus berkembang dan menjadi alat yang sangat berharga dalam pengambilan keputusan berbasis data di berbagai sektor. Secara umum, metode pada machine learning dibagi menjadi empat tipe berdasarkan cara pembelajarannya , yaitu: supervised learning, unsupervised learning, semi-supervised learning dan reinforcement learning. ML menawarkan berbagai macam algoritma yang dapat digunakan untuk menyelesaikan berbagai tugas, salah satunya adalah klasifikasi. Salah satu algoritma klasifikasi yang populer dan mudah digunakan adalah algoritma Naïve Bayes. Algoritma ini didasarkan pada teorema Bayes, yang merupakan rumus probabilitas yang digunakan untuk memperbarui keyakinan kita tentang suatu peristiwa berdasarkan bukti baru.

Berbagai penelitian telah dilakukan untuk mengembangkan sistem deteksi SMS spam dan penipuan menggunakan pendekatan machine learning. Efektivitas algoritma Naïve Bayes dalam mengklasifikasikan SMS spam, dengan hasil menunjukkan performa yang sangat memuaskan [9]. Studi ini menekankan keunggulan Naïve Bayes dalam pemrosesan teks melalui tahapan tokenisasi, stemming, dan pembobotan frekuensi kata. Temuan serupa dilaporkan oleh [2, 10], di mana Naïve Bayes menunjukkan kemampuan klasifikasi yang stabil dan konsisten, didukung oleh kurva ROC yang menunjukkan hasil yang sangat baik. Beberapa peneliti membandingkan kinerja algoritma alternatif. [11] menguji K-Nearest Neighbor (KNN) dengan pembobotan TF-IDF dan cosine similarity untuk SMS berbahasa Indonesia, sementara [12] menemukan Multinomial Naïve Bayes (MNB) memberikan hasil yang lebih baik dibandingkan dengan Support Vector Machine (SVM) dan Random Forest, dengan waktu pemrosesan yang relatif singkat. [13] juga menyoroti kelebihan Naïve Bayes atas SVM dalam konteks karakteristik linguistik pesan. Di sisi lain, regresi logistik mampu memberikan performa yang mengungguli Naïve Bayes pada dataset tertentu, meski memerlukan pra-pemrosesan yang lebih intensif seperti case folding dan stemming [14].

Penelitian-penelitian terdahulu tersebut mengidentifikasi beberapa tantangan krusial: 1) kebutuhan adaptasi model untuk bahasa Indonesia dengan kompleksitas morfologisnya, 2) pentingnya seleksi fitur untuk meningkatkan kinerja algoritma, dan 3) dinamika pola SMS penipuan yang terus

berevolusi. Studi ini berupaya mengisi gap tersebut dengan mengintegrasikan seleksi fitur Chi-Square ke dalam Naïve Bayes, sebuah pendekatan yang belum banyak dieksplorasi dalam konteks SMS berbahasa Indonesia. Temuan dari tinjauan literatur memperkuat dasar teoritis bahwa kombinasi metode probabilistik dan statistik ini berpotensi menghasilkan sistem deteksi yang lebih responsif terhadap modus penipuan terkini.

Short Message Service (SMS) adalah layanan komunikasi yang memungkinkan pengiriman pesan singkat melalui jaringan telekomunikasi. Dalam konteks teknologi informasi, SMS berfungsi sebagai alat yang efisien untuk mengirimkan informasi secara cepat dan langsung kepada pengguna. SMS dapat digunakan dalam berbagai aplikasi, termasuk layanan informasi publik, notifikasi, dan sistem pengamanan rumah otomatis. Deteksi SMS penipuan merupakan proses penting dalam mengidentifikasi pesan teks yang berpotensi menipu, terutama dengan meningkatnya penggunaan ponsel dan SMS sebagai alat komunikasi [2]. Penipuan melalui SMS, yang sering disebut sebagai *smishing*, melibatkan pengiriman pesan yang tampaknya berasal dari sumber tepercaya dengan tujuan mencuri informasi pribadi atau finansial.

Berbagai metode digunakan untuk mendeteksi SMS penipuan, termasuk analisis teks yang memanfaatkan teknik pemrosesan bahasa alami (*Natural Language Processing*/NLP) untuk menganalisis konten pesan. Dalam pendekatan ini, pengenalan pola, pra-pemrosesan data, dan ekstraksi fitur digunakan untuk mengidentifikasi kata-kata atau frasa yang sering muncul dalam SMS penipuan. Selain itu, algoritme pembelajaran mesin Naïve Bayes dilatih menggunakan dataset SMS yang telah diklasifikasikan sebagai penipuan atau bukan, sehingga model ini dapat memprediksi apakah SMS baru adalah penipuan. Selain itu, sistem berbasis aturan dikembangkan berdasarkan karakteristik umum SMS penipuan, seperti penggunaan kata-kata tertentu, format pesan, atau pengirim yang tidak dikenal. Tantangan dalam deteksi SMS penipuan meliputi variasi bahasa yang digunakan oleh penipu untuk menghindari deteksi, risiko false positives yang dapat mengganggu komunikasi, serta isu privasi dan keamanan data pengguna. Dengan demikian, deteksi SMS penipuan adalah bidang yang terus berkembang, dan penelitian serta pengembangan dalam area ini sangat penting untuk melindungi pengguna dari kerugian finansial dan pencurian identitas.

Di Indonesia, SMS penipuan diatur dalam beberapa undang-undang. Salah satunya adalah Undang-Undang No. 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (ITE), yang mengatur tentang penyalahgunaan informasi elektronik, termasuk penipuan melalui SMS. Pasal 28 ayat (1) UU ITE menyatakan bahwa setiap orang dilarang melakukan tindakan yang merugikan orang lain dengan cara menyebarkan informasi yang tidak benar [15]. Selain itu, Undang-Undang No. 19 Tahun 2016 yang merupakan perubahan dari UU ITE juga memperkuat sanksi bagi pelanggar yang

melakukan penipuan melalui media elektronik. Dengan adanya regulasi ini, diharapkan dapat memberikan perlindungan hukum bagi masyarakat dari praktik penipuan yang merugikan.

Preprocessing merupakan tahapan sebelum proses pengklasifikasian yang diperlukan untuk membersihkan, menghilangkan, mengubah sumber data, baik itu berupa karakter non alfabet maupun kata kata yang tidak diperlukan [16]. Hal ini bertujuan agar data yang digunakan lebih optimal ketika digunakan pada proses pengklasifikasiannya. Tahapan preprocessing setiap kasus dapat berbeda-beda. Tahapan ini mencakup berbagai aktivitas seperti pembersihan data, normalisasi, dan transformasi data. Pembersihan data melibatkan penghapusan duplikasi dan penanganan missing values, sedangkan normalisasi dan transformasi memastikan bahwa data berada dalam format yang konsisten dan sesuai dengan persyaratan model. Proses preprocessing penting untuk memastikan kualitas dan integritas data, sehingga analisis selanjutnya dapat dilakukan dengan lebih akurat dan efisien.

Case folding merupakan suatu tahapan proses untuk mengubah kata menjadi bentuk yang sama menggunakan python string lower method. Tujuan dari case folding adalah mengembalikan semua kata kedalam bentuk huruf kecil semua supaya data text yang diproses semuanya dalam kondisi bentuk yang sama. Selain itu, case folding ini bertujuan untuk membersihkan teks sebelum analisis lebih lanjut, seperti pengklasifikasian atau analisis sentimen, dengan menghilangkan elemen yang tidak relevan dan memastikan konsistensi format. Normalisasi data adalah teknik preprocessing yang sangat penting, terutama ketika bekerja dengan fitur numerik sebelum menerapkan algoritma klasifikasi atau clustering. Proses ini bertujuan untuk mengatur nilai-nilai fitur ke dalam rentang tertentu, sehingga fitur dengan skala yang lebih kecil tidak terabaikan oleh fitur dengan skala yang lebih besar [17]. Alasan pentingnya proses normalisasi adalah untuk menghindari beberapa fitur yang dipertimbangkan menyembunyikan 19 efek dari fitur lainnya, terutama ketika fitur memiliki rentang yang berbeda-beda.

Di sisi lain, pemilihan teknik normalisasi dan rentang normalisasi (interval) dianggap sebagai langkah penting selama tahap preprocessing, karena “perubahan” yang mempengaruhi data yang dipertimbangkan dan akibatnya hasil dari algoritma machine learning yang akan diterapkan setelah preprocessing. Stopwords Removal adalah proses penghilangan kata-kata yang disebut stopwords dari teks dalam NLP. Stopwords adalah kata-kata yang sering muncul dalam teks namun tidak memiliki makna atau informasi yang signifikan dalam analisis teks atau NLP [18]. Contoh stopwords bahasa Indonesia antara lain “yang”, “di”, “ke”, dan lain-lain. Penghilangan stopwords merupakan langkah penting dalam preprocessing data teks, karena dapat meningkatkan kualitas analisis dengan mengurangi teks yang tidak perlu dalam data. Dengan menghilangkan stopwords, fokus analisis bisa lebih tertuju pada kata kata kunci yang lebih bermakna dan relevan untuk tujuan

analisis yang diinginkan. Stemming adalah proses dalam NLP yang bertujuan untuk mengubah kata-kata ke bentuk dasarnya (stem). Proses ini penting dalam analisis teks karena membantu mengurangi variasi kata yang berbeda menjadi satu bentuk yang sama, sehingga meningkatkan konsistensi dan akurasi dalam analisis. Untuk mengubah kata-kata menjadi kata dasar dibutuhkan beberapa aturan yang disebut dalam morfologi. Sebagaimana dinyatakan oleh [19], berikut ini aturan morfologi yang harus diketahui sebelum melakukan proses stemming: 1) Afiks adalah morfem yang ditambahkan pada kata dasar untuk membentuk kata baru atau mengubah makna kata tersebut. Afiks dapat berupa awalan, akhiran, atau sisipan, dan berfungsi untuk memperkaya kosakata serta memberikan nuansa makna yang lebih dalam. Afiks tidak dapat berdiri sendiri dan selalu melekat pada kata dasar; 2) Prefiks adalah jenis afiks yang ditambahkan di depan kata dasar. Prefiks sering digunakan untuk mengubah makna kata dasar, misalnya "me-" dalam "membaca" (dari kata dasar "baca") yang menunjukkan tindakan; 3) Sufiks adalah afiks yang ditambahkan di belakang kata dasar. Sufiks juga berfungsi untuk mengubah makna kata dasar, contohnya "-kan" dalam "makanan" (dari kata dasar "makan") yang mengubah kata kerja menjadi kata benda; 4) Konfiks adalah kombinasi dari prefiks dan sufiks yang ditambahkan pada kata dasar. Konfiks biasanya digunakan untuk membentuk kata baru yang memiliki makna tertentu. Berikut contoh dari konfiks "pe-" dan "-an" dalam "penulisan" (dari kata dasar "tulis") yang menunjukkan proses atau hasil dari tindakan menulis; 5) Infiks adalah afiks yang disisipkan di tengah kata dasar. Infiks jarang digunakan dalam bahasa Indonesia, tetapi ada beberapa contoh dalam bahasa lain. Berikut adalah contoh infiks dalam bahasa Indonesia "s" dalam "belis" (dari kata dasar "beli"), yang menunjukkan bentuk plural atau intensifikasi. Pemahaman tentang afiks, prefiks, sufiks, konfiks, dan infiks sangat penting dalam analisis morfologi bahasa. Proses ini memungkinkan kita untuk mengurangi variasi kata dan meningkatkan konsistensi dalam analisis data teks.

Term Frequency-Inverse Document Frequency (TF-IDF) adalah sebuah teknik yang digunakan dalam pengolahan bahasa alami dan informasi retrieval untuk menilai seberapa penting sebuah kata dalam sebuah dokumen relatif terhadap kumpulan dokumen (corpus). Metode ini menggabungkan dua komponen utama: frekuensi istilah (Term Frequency, TF) dan frekuensi dokumen terbalik (Inverse Document Frequency, IDF). TF-IDF sering digunakan dalam berbagai aplikasi, termasuk pencarian informasi, pengelompokan dokumen, dan analisis teks. Rumus TF-IDF terdiri dari dua bagian. Yang pertama adalah Term Frequency (TF) untuk mengukur seberapa sering sebuah kata muncul dalam dokumen. TF dihitung dengan rumus:

$$TF(t, d) = \frac{Jumlah\ kemunculan\ kata\ t\ dalam\ dokumen\ d}{Total\ jumlah\ kata\ dalam\ dokumen\ d} \quad (1)$$

Selanjutnya adalah Inverse Document Frequency (IDF) untuk mengukur seberapa penting sebuah kata dalam seluruh kumpulan dokumen. IDF dihitung dengan rumus:

$$TF(t, D) = log \frac{Jumlah\ kemunculan\ kata\ t\ dalam\ dokumen\ d}{Total\ jumlah\ kata\ dalam\ dokumen\ d} \quad (2)$$

Dengan menggabungkan kedua komponen tersebut, nilai TF-IDF untuk sebuah kata t dalam dokumen d dapat dihitung sebagai:

$$TD\text{-}IDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (3)$$

Secara keseluruhan, pembobotan TF-IDF adalah alat yang efektif untuk menilai pentingnya kata dalam konteks dokumen dan koleksi dokumen. Dengan menggunakan rumus ini, kita dapat mengidentifikasi kata-kata kunci yang relevan, yang sangat berguna dalam berbagai aplikasi seperti pencarian informasi, klasifikasi teks, dan pengelompokan dokumen.

Pemilihan fitur menggunakan Chi-Square (χ^2) adalah metode statistik yang digunakan untuk memilih fitur yang paling relevan dalam dataset berdasarkan hubungan antara variabel independen dan variabel dependen [18]. Dalam konteks ini, Chi-Square mengukur seberapa besar perbedaan antara frekuensi yang diamati dan frekuensi yang diharapkan jika tidak ada hubungan antara fitur dan target variabel. Metode ini biasanya digunakan untuk masalah klasifikasi dengan variabel kategorikal, di mana setiap fitur diuji untuk melihat apakah ada ketergantungan yang signifikan terhadap kelas target. Fitur dengan nilai Chi-Square yang tinggi menunjukkan adanya hubungan yang kuat dengan target dan dianggap lebih relevan untuk dimasukkan dalam model. Sebaliknya, fitur dengan nilai Chi-Square yang rendah cenderung tidak memberikan informasi yang signifikan dan dapat dihapus. Proses ini membantu mengurangi kompleksitas model dan meningkatkan kinerjanya dengan hanya mempertahankan fitur yang paling informatif.

Rumus untuk menghitung nilai Chi-Square ($\chi 2$) dalam pemilihan fitur adalah sebagai berikut:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (4)$$

dengan:

O_i = Frekuensi yang diamati (Observed Frequency) untuk kategori i.

E_i = Frekuensi yang diharapkan (Expected Frequency) untuk kategori *i*, yang dihitung dengan rumus:

$$E_i = \frac{(jumlah\ baris\ i) x\ (jumlah\ kolom\ j)}{jumlah\ total} \quad (5)$$

Proses ini dihitung untuk setiap kategori dalam tabel kontingensi antara fitur dan target variabel. Setelah menghitung nilai Chi-Square untuk setiap fitur, fitur dengan nilai Chi-Square yang lebih tinggi dianggap memiliki hubungan yang lebih kuat dengan target, sehingga lebih relevan untuk

dimasukkan dalam model. Sebaliknya, fitur dengan nilai Chi-Square yang rendah dianggap kurang berpengaruh dan bisa dihapus dalam proses pemilihan fitur.

Naïve Bayes adalah sebuah algoritme klasifikasi yang digunakan dalam pembelajaran mesin. Algoritme ini didasarkan pada Teorema Bayes dan diasumsikan bahwa setiap pasangan fitur dalam data pelatihan saling independent [20]. Algoritme Naïve Bayes memproses data pelatihan untuk menghasilkan model klasifikasi, yang dapat digunakan untuk memprediksi label kelas yang sesuai untuk data yang belum dilihat sebelumnya. Dalam algoritma Naïve Bayes, kelas target dan setiap fitur dalam data diberikan probabilitas masing-masing. Kemudian, probabilitas setiap fitur diberikan kondisi kelas target, dan probabilitas kelas target diberikan fitur-fitur yang ada.

Dalam pengujian, algoritma Naïve Bayes digunakan untuk memprediksi kelas target dari data yang belum dikenal dengan memperhitungkan probabilitas yang dihitung selama pelatihan. Naïve Bayes merupakan algoritme untuk perhitungan probabilitas bersyarat (posterior), yaitu perhitungan peluang suatu kejadian *X* bila diketahui kejadian H terjadi yang dinotasikan dengan *P(X|H)*. Naïve Bayes berasumsi bahwa efek suatu nilai variabel di sebuah kelas yang ditentukan adalah tidak terkait pada nilai-nilai variabel lain. Algoritme Naïve Bayes memungkinkan secara cepat membuat model yang mempunyai kemampuan untuk prediksi. Dasar dari *Naïve Bayes* yang dipakai dalam pemrograman adalah rumus Bayes seperti pada rumus berikut:

$$P(A|B) = (P(B|A) \times P(A))/P(B) \quad (6)$$

dengan:

P(A|B): Peluang kejadian A terjadi mengingat kejadian B telah terjadi. Ini dikenal sebagai *posterior probability*.

P(B|A): Peluang kejadian B terjadi mengingat kejadian A telah terjadi. Ini dikenal sebagai *likelihood*.

P(A): Peluang terjadinya kejadian A tanpa memperhatikan B. Ini dikenal sebagai *prior probability*.

P(B): Peluang terjadinya kejadian B tanpa memperhatikan A. Ini dikenal sebagai *marginal likelihood* atau *evidence*.

Pada pengaplikasiannya, rumus (6) menjadi:

$$P(Ci|D) = (P(D|Ci) \times P(Ci)) / P(D) \quad (7)$$

dengan:

P(Ci|D): Peluang kelas Ci diberikan dokumen D (probabilitas bahwa dokumen D termasuk dalam kelas Ci).

P(D|Ci): Peluang dokumen D diberikan kelas Ci (probabilitas bahwa dokumen D akan muncul jika diketahui kelasnya adalah Ci).

P(Ci): Peluang awal dari kelas Ci (probabilitas bahwa kelas Ci terjadi tanpa informasi tambahan).

P(D): Peluang dokumen D tanpa memperhatikan kelasnya (sering dianggap konstan di seluruh kelas untuk tujuan klasifikasi).

Naïve Bayes atau bisa disebut sebagai Multinomial Naïve Bayes merupakan model penyederhanaan dari Metode Bayes yang cocok dalam pengklasifikasian teks atau dokumen. Berikut adalah rumus yang merupakan penyederhanaan dari Metode Bayes.

$$V_{MAP} = \arg\max P(Vj \mid a1, a2, ...an) \quad (8)$$

dengan:

VMAP: Kelas yang memiliki probabilitas maksimum (*Maximum A Posteriori*), yaitu kelas yang paling mungkin diberikan bukti yang ada.

Vj: Salah satu kelas yang ada di antara semua kelas yang mungkin V.

a1, a2, ..., an: Atribut atau fitur yang diamati dalam dokumen (kata- kata dalam teks).

Berdasarkan rumus (7), maka rumus (6) dapat ditulis ulang menjadi:

$$V\_MAP = argmax\ (Vj\ eV) \frac{P(a1,a2,\ldots,an\ |(Vj))}{P(a1,a2,\ldots,an)} \quad (9)$$

dengan:

arg max: Operator yang mencari nilai Vj yang memaksimalkan fungsi yang ada.

Vj eV: Vj adalah salah satu kelas dalam himpunan semua kelas V.

P(a1,a2,...,an | Vj): Peluang semua fitur yang diamati (a1, a2, ..., an) muncul dalam dokumen yang termasuk dalam kelas Vj.

P(Vj): Probabilitas awal dari kelas Vj.

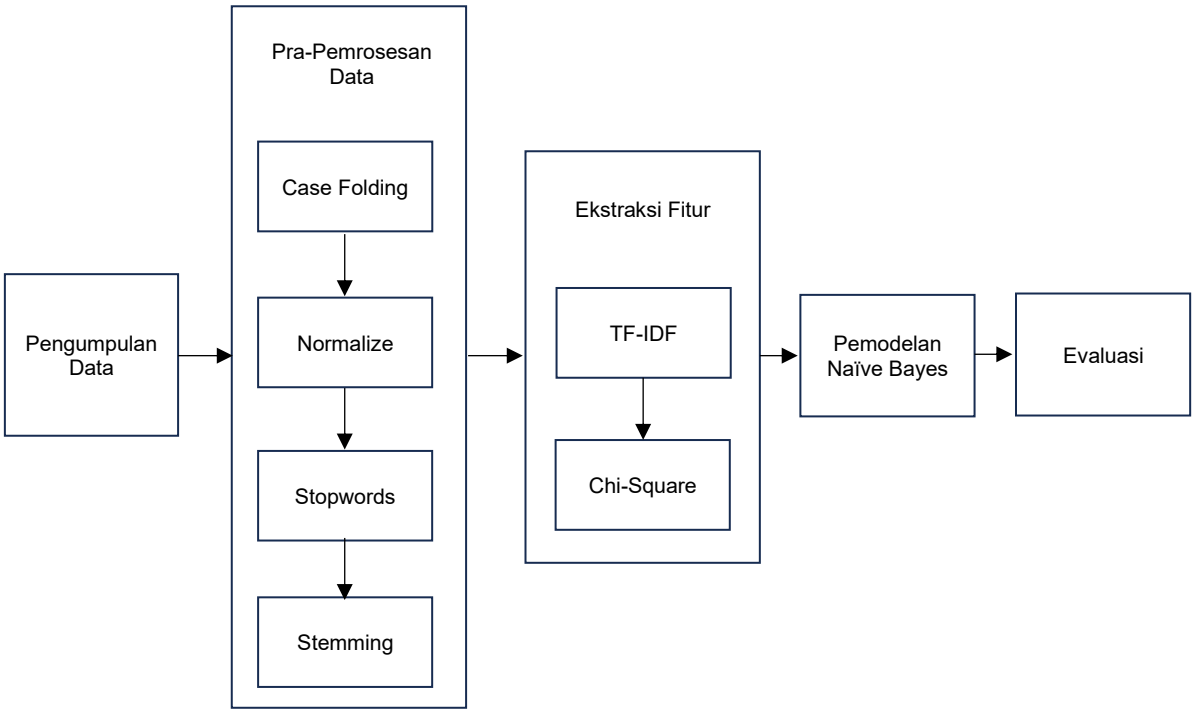
P(a1,a2,...,an): Peluang munculnya semua fitur tanpa memperhatikan kelasnya (konstanta).

Naïve Bayes mengasumsikan bahwa semua fitur (misalnya, kata- kata dalam dokumen) independen satu sama lain, yang menyederhanakan perhitungan probabilitas. Model Naïve Bayes ini digunakan untuk menentukan kelas mana yang paling mungkin untuk sebuah dokumen berdasarkan fitur yang ada. Kelemahan metode Naïve Bayes ini sendiri yaitu adanya asumsi atau dengan kata lain kondisi kelas saling bebas, sehingga kurang akurat, sedangkan pada prakteknya, beberapa kondisi biasanya saling berpengaruh satu sama lain. Pada kenyataannya, asumsi ini tidak selalu benar, dan dapat menyebabkan algoritme Naïve Bayes kurang akurat. Hal ini karena beberapa fitur dalam data point

mungkin saling terkait satu sama lain. Oleh karena itu, penting untuk mengevaluasi kinerja algoritme Naïve Bayes untuk memastikan bahwa algoritme tersebut bekerja dengan baik dan memberikan hasil yang akurat.

3. METODOLOGI PENELITIAN

Tahap penelitian ini dimulai dengan pengumpulan data yang relevan dan representatif, diikuti oleh serangkaian langkah pra-pemrosesan yang bertujuan untuk membersihkan dan menyiapkan data agar siap untuk analisis lebih lanjut. Setelah data diproses, fitur-fitur penting diekstraksi menggunakan teknik pembobotan yang sesuai, yang kemudian digunakan dalam algoritme klasifikasi. Akhirnya, model yang dihasilkan dievaluasi menggunakan berbagai metrik untuk memastikan kinerjanya dalam mengidentifikasi pesan penipuan secara akurat. Ilustrasi tahapan penelitian dapat dilihat dalam Gambar 1.



Gambar 1. Diagram Alur Sistem

3.1 Pengumpulan Data

Pada tahap ini dataset yang akan digunakan dalam penelitian ini bersumber dari repositori GitHub dengan judul “klasifikasi spam sms” [21]. Dataset berisi 1.145 teks SMS yang terbagi ke dalam 2 kelas, yaitu penipuan dan bukan penipuan. Tabel 1 memperlihatkan contoh 6 record dataset yang diantaranya 3 record untuk kelas penipuan diberi label 1 dan 3 record untuk kelas bukan penipuan diberi label 0.

Tabel 1. Sampel dataset pesan teks

id	teks	label
3	2016-08-07 11:29:47.Plg Yth, sisa kuota Flash Anda 7160KB. Download MyTelkomsel apps di http://tsel.me/tsel utk cek kuota&beli paket Flash atau hub *363#	0
16	Anda sedang menikmati Paket Reguler dgn sisa kuota 209715 KB.Dapatkan beragam Paket INTERNET HEMAT Tri di 1233#	0
271	Bpk/Ibu Mengenai Rekening Anda Terpilih Sebagai Pemenang Cek 35jt Dri BNI U/Info klik www.promobni46.tk Kode Cek 03299757 Hub.085288991559	1
317	Ini bpk pnjam hp orng, tlng bliin bpk pls 50rb, krm ke nomr brunya bpk=081215008228, bpk lagi ada msalh di kantor polisi, jgn tlp/sms nnti bpk yg tlp, penting	1

3.2 Pra-Pemrosesan Data

Pra-pemrosesan data adalah langkah penting dalam pengembangan sistem klasifikasi SMS untuk mendeteksi penipuan. Proses ini bertujuan untuk membersihkan dan menyiapkan data agar siap untuk analisis lebih lanjut [22]. Dengan melakukan pra-pemrosesan, kualitas data dapat ditingkatkan, dan memastikan bahwa model yang dibangun dapat belajar dengan efektif.

3.2.1 Case Folding

Case folding adalah teknik yang digunakan dalam pengolahan bahasa alami atau NLP untuk mengubah semua huruf dalam teks menjadi huruf kecil. Proses ini bertujuan untuk menghilangkan perbedaan antara huruf besar dan kecil, sehingga memudahkan analisis dan pemrosesan teks. Misalnya, kata "Penipuan" dan "penipuan" akan dianggap sebagai dua entitas yang berbeda jika tidak dilakukan case folding. Dengan menyamakan semua huruf, kita dapat memastikan bahwa analisis dilakukan secara konsisten. Proses ini sangat penting dalam pengolahan bahasa alami (NLP) karena membantu mengurangi kompleksitas data dan meningkatkan akurasi model. Contoh case folding dapat dilihat di Tabel 2.

3.2.2 Normalize

Normalize adalah proses yang krusial dalam pengolahan bahasa alami yang bertujuan untuk mengubah teks menjadi bentuk yang lebih standar dan konsisten. Proses ini mencakup penggantian istilah yang disingkat atau tidak baku dengan bentuk yang lebih umum dan dapat diterima. Misalnya, dalam konteks pesan teks, istilah seperti "yg" bisa diganti dengan "yang", sehingga menghasilkan teks yang lebih formal dan seragam. Selain itu, normalisasi juga penting untuk menangani variasi dalam penulisan, seperti mengubah semua huruf menjadi huruf kecil atau huruf besar untuk menghilangkan perbedaan antara penulisan yang berbeda. Dengan melakukan

normalisasi ini, kita dapat mengurangi jumlah noise dalam data dan memastikan bahwa analisis yang dilakukan dapat menghasilkan hasil yang lebih akurat dan konsisten. Contoh proses normalize dapat dilihat di Tabel 3.

Tabel 2. Contoh case folding

id	Raw Data	Case Folding
3	2016-08-07 11:29:47.Plg Yth, sisa kuota Flash Anda 7160KB. Download MyTelkomsel apps di http://tsel.me/tsel utk cek kuota&beli paket Flash atau hub *363#	plg yth sisa kuota flash anda kb download mytelkomsel apps di utk cek kuotabeli paket flash atau hub
16	Anda sedang menikmati Paket Reguler dgn sisa kuota 209715 KB.Dapatkan beragam Paket INTERNET HEMAT Tri di 1233#	anda sedang menikmati paket reguler dgn sisa kuota kbdapatkan beragam paket internet hemat tri di
271	Bpk/Ibu Mengenai Rekening Anda Terpilih Sebagai Pemenang Cek 35jt Dri BNI U/Info klik www.promobni46.tk Kode Cek 03299757 Hub.085288991559	bpkibu mengenai rekening anda terpilih sebagai pemenang cek jt dri bni uinfo klik kode cek hub
317	Ini bpk pnjam hp orng, tlng bliin bpk pls 50rb, krm ke nomr brunya bpk=081215008228, bpk lagi ada msalh di kantor polisi, jgn tlp/sms nnti bpk yg tlp, penting	ini bpk pnjam hp orng tlng bliin bpk pls rb krm ke nomr brunya bpk bpk lagi ada msalh di kantor polisi jgn tlpsms nnti bpk yg tlp penting

Tabel 3. Contoh normalize

id	Raw Data	Normalize
3	2016-08-07 11:29:47.Plg Yth, sisa kuota Flash Anda 7160KB. Download MyTelkomsel apps di http://tsel.me/tsel utk cek kuota&beli paket Flash atau hub *363#	pelanggan yang terhormat sisa kuota flash anda kb download mytelkomsel apps di untuk cek kuotabeli paket flash atau hub
16	Anda sedang menikmati Paket Reguler dgn sisa kuota 209715 KB.Dapatkan beragam Paket INTERNET HEMAT Tri di 1233#	anda sedang menikmati paket reguler dengan sisa kuota kbdapatkan beragam paket internet hemat tri di
271	Bpk/Ibu Mengenai Rekening Anda Terpilih Sebagai Pemenang Cek 35jt Dri BNI U/Info klik www.promobni46.tk Kode Cek 03299757 Hub.085288991559	bpkibu mengenai rekening anda terpilih sebagai pemenang cek jt dri bni uinfo klik kode cek hub
317	Ini bpk pnjam hp orng, tlng bliin bpk pls 50rb, krm ke nomr brunya bpk=081215008228, bpk lagi ada msalh di kantor polisi, jgn tlp/sms nnti bpk yg tlp, penting	ini bapak pinjam hp orang tolong bliin bapak pulsa rb krm ke nomr brunya bapak bapak lagi ada msalh di kantor polisi jgn tlpsms nnti bapak yang telepon penting

3.2.3 Stopword

Stopwords adalah langkah yang dilakukan untuk menghilangkan kata-kata umum yang tidak memberikan informasi penting dalam analisis teks. Kata-kata seperti "dan", "atau", dan "adalah" sering kali tidak memberikan kontribusi signifikan terhadap makna keseluruhan teks. Dengan menghapus stopwords, kita dapat fokus pada kata-kata yang lebih bermakna dan relevan. Proses ini membantu dalam menyederhanakan data dan meningkatkan efisiensi analisis, sehingga model dapat lebih mudah mengenali pola yang ada dalam data. Dengan demikian, penghapusan stopwords berkontribusi pada peningkatan akurasi dan efektivitas model yang dibangun. Contoh stopword terdapat dalam Tabel 4.

Tabel 4. Contoh stopword

id	Raw Data	Stopword
3	2016-08-07 11:29:47.Plg Yth, sisa kuota Flash Anda 7160KB. Download MyTelkomsel apps di http://tsel.me/tsel utk cek kuota&beli paket Flash atau hub *363#	plg yth sisa kuota flash anda kb download mytelkomsel apps di utk cek kuotabeli paket flash atau hub
16	Anda sedang menikmati Paket Reguler dgn sisa kuota 209715 KB.Dapatkan beragam Paket INTERNET HEMAT Tri di 1233#	anda sedang menikmati paket reguler dgn sisa kuota kbdapatkan beragam paket internet hemat tri di
271	Bpk/Ibu Mengenai Rekening Anda Terpilih Sebagai Pemenang Cek 35jt Dri BNI U/Info klik www.promobni46.tk Kode Cek 03299757 Hub.085288991559	bpkibu mengenai rekening anda terpilih sebagai pemenang cek jt dri bni uinfo klik kode cek hub
317	Ini bpk pnjam hp orng, tlng bliin bpk pls 50rb, krm ke nomr brunya bpk=081215008228, bpk lagi ada msalh di kantor polisi, jgn tlp/sms nnti bpk yg tlp, penting	ini bpk pnjam hp orng tlng bliin bpk pls krm ke nomr brunya bpk bpk lagi ada msalh di kantor polisi jgn tlpsms nnti bpk yg tlp penting

3.2.4 Stemming

Stemming adalah proses mengubah kata-kata ke bentuk dasarnya (stem) untuk mengurangi variasi kata yang sama. Misalnya, kata "berlari" dan "lari" akan diubah menjadi "lari". Proses ini membantu dalam menyederhanakan analisis dan memastikan bahwa variasi kata yang sama diakui sebagai entitas yang sama. Dengan melakukan stemming, kita dapat mengurangi kompleksitas data dan meningkatkan kemampuan model dalam mengenali pola yang ada. Hal ini sangat penting dalam pengolahan bahasa alami, di mana kata-kata sering kali memiliki bentuk yang berbeda tetapi makna yang sama. Dengan demikian, stemming berkontribusi pada peningkatan akurasi dan efektivitas model yang dibangun. Contoh stemming terdapat dalam Tabel 5.

Tabel 5. Contoh stemming

id	Raw Data	Stemming
3	2016-08-07 11:29:47.Plg Yth, sisa kuota Flash Anda 7160KB. Download MyTelkomsel apps di http://tsel.me/tsel utk cek kuota&beli paket Flash atau hub *363#	plg yth sisa kuota flash anda kb download mytelkomsel apps di utk cek kuotabeli paket flash atau hub
16	Anda sedang menikmati Paket Reguler dgn sisa kuota 209715 KB.Dapatkan beragam Paket INTERNET HEMAT Tri di 1233#	anda sedang nikmat paket reguler dgn sisa kuota kbdapatkan agam paket internet hemat tri di
271	Bpk/Ibu Mengenai Rekening Anda Terpilih Sebagai Pemenang Cek 35jt Dri BNI U/Info klik www.promobni46.tk Kode Cek 03299757 Hub.085288991559	bpkibu kena rekening anda pilih bagai menang cek jt dri bni uinfo klik kode cek hub
317	Ini bpk pnjam hp orng, tlng bliin bpk pls 50rb, krm ke nomr brunya bpk=081215008228, bpk lagi ada msalh di kantor polisi, jgn tlp/sms nnti bpk yg tlp, penting	ini bpk pnjam hp orng tlng bliin bpk pls krm ke nomr brunya bpk bpk lagi ada msalh di kantor polisi jgn tlpsms nnti bpk yg tlp penting

3.3 Ekstraksi Fitur

Ekstraksi fitur adalah tahap penting dalam NLP yang bertujuan untuk mengubah teks yang telah dipra-pemrosesan menjadi representasi numerik yang dapat digunakan oleh model pembelajaran mesin. Proses ini memungkinkan model untuk mengenali pola dan membuat prediksi berdasarkan data yang ada. Dalam konteks deteksi SMS penipuan, ekstraksi fitur membantu mengidentifikasi karakteristik linguistik khusus seperti kata kunci, pola kalimat, atau frekuensi istilah tertentu yang sering muncul dalam pesan penipuan. Teknik-teknik seperti TF-IDF dan word embedding dapat digunakan untuk menangkap makna semantik dan statistik dari teks secara efektif. Dengan representasi numerik yang tepat, model machine learning dapat membedakan antara pesan legit dan penipuan dengan akurasi yang lebih tinggi.

3.3.1 Pembobotan TF-IDF

Dalam tahap ini, teknik TF-IDF digunakan untuk mengubah teks yang telah dipra-pemrosesan menjadi representasi numerik. Gambar 2 memperlihatkan format tabular dari 3.563 data fitur yang mewakili kata-kata unik (fitur) yang diekstrak dari corpus teks dan 1.143 baris yang mewakili dokumen atau entitas teks dalam dataset. Mayoritas nilai adalah 0.0, yang menunjukkan bahwa sebagian besar kata tidak muncul dalam dokumen tertentu. Ini adalah karakteristik umum data teks yang direpresentasikan dalam bentuk matriks TF-IDF. Kesimpulannya, data ini merupakan hasil transformasi teks menjadi representasi numerik berbasis TF-IDF yang siap digunakan untuk berbagai keperluan analisis dan model pembelajaran mesin.

	aa	aamiiiin	aamiin	ab	abadi	abai	abbee	abdul	acara	acaratks	...	yudisium	yuk	yuks	yuni	yunit	zalora	zarkasi	zjt	zona	ztkm
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1138	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1139	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1140	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1141	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1142	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

1143 rows × 3563 columns

Gambar 2. Pembobotan TF/IDF

3.3.2 Chi-Square

Setelah dilakukan ekstraksi fitur dengan pembobotan TF-IDF, langkah selanjutnya adalah melakukan seleksi data fitur menggunakan metode Chi-Square. Chi-Square adalah metode statistik yang digunakan untuk mengukur hubungan antara kata-kata (fitur) dengan label target dalam dataset. Seleksi fitur ini bertujuan untuk mengidentifikasi kata-kata yang paling relevan dalam membedakan antara kelas target, seperti SMS penipuan dan bukan penipuan. Fitur yang memiliki nilai Chi-Square tinggi menunjukkan bahwa kata-kata tersebut memiliki pengaruh yang kuat dalam membedakan antara SMS penipuan dan bukan penipuan. Sebaliknya, fitur dengan nilai Chi-Square rendah dianggap kurang relevan dan dapat diabaikan. Dengan cara ini, seleksi fitur membantu mengurangi jumlah kata yang digunakan dalam model, sehingga meningkatkan efisiensi dan akurasi.

	aa	aamiiiin	aamiin	ab	abadi	abai	abbee	abdul	acaratks	ada	...	yudisium	yuk	yuks	yuni	yunit	zalora	zarkasi	zjt	zona	ztkm
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1138	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1139	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.185855	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1140	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1141	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1142	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

1143 rows × 3000 columns

Gambar 3. Seleksi fitur dengan Chi-Square

Gambar 3 adalah hasil seleksi data fitur menggunakan metode Chi-Square pada dataset teks yang telah diekstraksi menggunakan TF-IDF. Dataset ini memiliki 3.000 kolom yang merupakan fitur yang telah dipilih melalui proses seleksi menggunakan Chi-Square dan 1.143 baris yang masing-

masing mewakili dokumen dalam corpus. Setelah proses seleksi fitur selesai, fitur-fitur yang terpilih akan digunakan dalam tahap modeling.

3.4 Pemodelan

Dalam penelitian ini, algoritme yang digunakan adalah Naïve Bayes, yang merupakan metode populer dalam klasifikasi teks. Algoritme ini bekerja dengan menghitung probabilitas suatu teks termasuk dalam kategori tertentu berdasarkan fitur yang telah diekstraksi dan diseleksi. Kode berikut, yang menggunakan bahasa pemrograman Python, merupakan contoh cara melatih model klasifikasi menggunakan algoritma Naïve Bayes. Dimulai dengan memilih fitur-fitur terbaik dari dataset yang ada, kemudian data dipersiapkan dengan memisahkan antara fitur (input) dan label (output). Selanjutnya, dataset dibagi menjadi dua bagian: data latih (80%) yang digunakan untuk melatih model, dan data uji (20%) yang digunakan untuk menguji performa model setelah dilatih.

```
selected_x = x_kbest_features
selected_x

# import library
import random
from sklearn.model_selection import train_test_split

# import algoritma Naïve Bayes
from sklearn.naive_bayes import MultinomialNB

x = selected_x
y = df.label

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2,
random_state=0)


# proses training menggunakan Naïve Bayes
text_algorithm = MultinomialNB()

model = text_algorithm.fit(x_train, y_train)
```

3.5 Evaluasi

Proses evaluasi melibatkan penggunaan berbagai metrik untuk menilai kinerja model. Dengan menganalisis hasil evaluasi, kita dapat mengidentifikasi kekuatan dan kelemahan model, serta melakukan perbaikan yang diperlukan untuk meningkatkan akurasi dan efektivitasnya. Salah satu alat evaluasi yang digunakan adalah confusion matrix. Confusion matrix memberikan gambaran yang mendetail mengenai performa model, termasuk jumlah prediksi yang benar (baik positif maupun negatif) serta jumlah prediksi yang salah (false positive dan false negative). Dengan menggunakan confusion matrix, dapat dihitung berbagai metrik evaluasi seperti akurasi, presisi,

recall, dan skor F1, yang semuanya memberikan wawasan komprehensif tentang seberapa efektif model dalam mendeteksi SMS penipuan atau bukan penipuan.

4. HASIL DAN PEMBAHASAN

Pemeringkatan fitur berdasarkan nilai Chi-Square untuk seleksi fitur dalam klasifikasi teks dapat dilihat dalam Gambar 5. Fitur dengan nilai tinggi, seperti “paket,” “hadiah,” dan “kuota,” berkontribusi signifikan dalam membedakan kategori, sedangkan fitur dengan nilai rendah kurang relevan. Seleksi fitur ini membantu meningkatkan akurasi dan efisiensi model klasifikasi.

	Nilai	Fitur
2337	44.191916	paket
1145	42.656336	hadiah
1749	39.717546	kuota
2433	37.689749	pin
1663	31.976164	klik
...	...	...
1708	0.035780	kopi
939	0.032327	fb
652	0.032191	daftar
1922	0.025060	maksimal
3428	0.003971	via

3563 rows × 2 columns

Gambar 5. Urutan nilai fitur terbaik

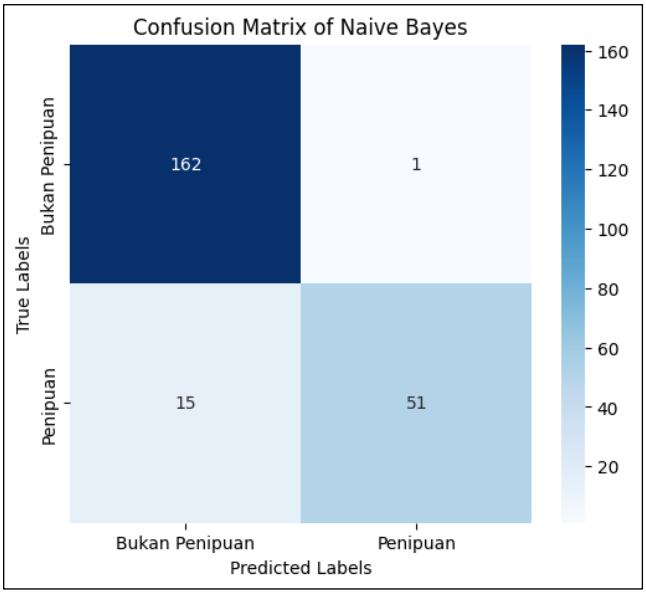
Hasil evaluasi kinerja dari model klasifikasi yang diuji menggunakan beberapa metrik seperti terlihat dalam Gambar 6, yaitu precision, recall, f1-score, dan support untuk dua kelas (0 dan 1). Kelas 0 merupakan kategori bukan penipuan, sementara kelas 1 merupakan kategori data penipuan. Precision mengukur proporsi prediksi positif yang benar, dan model menunjukkan hasil yang sangat baik dengan nilai precision 0.92 untuk kelas 0 dan 0.98 untuk kelas 1. Recall, yang mengukur proporsi data positif yang berhasil diprediksi dengan benar, menunjukkan nilai yang tinggi untuk kelas 0 (0.99), namun lebih rendah untuk kelas 1 (0.77), yang berarti model lebih efektif dalam mengidentifikasi kelas 0. F1-score, yang merupakan rata-rata harmonis dari precision dan recall, memberikan gambaran keseimbangan antara keduanya; untuk kelas 0, nilai F1 mencapai 0.95, sementara untuk kelas 1, nilainya 0.86. Support mengindikasikan jumlah data pada masing-masing kelas, dengan 163 sampel untuk kelas 0 dan 66 sampel untuk kelas 1. Akurasi keseluruhan model adalah 93%, yang berarti model dapat memprediksi dengan benar 93% dari total data. Selain itu,

terdapat dua jenis rata-rata: macro average dan weighted average. Macro average menghitung rata-rata dari precision, recall, dan F1-score tanpa memperhatikan proporsi kelas, menghasilkan nilai rata-rata yang lebih seimbang, sedangkan weighted average memperhitungkan jumlah sampel pada setiap kelas, memberikan hasil yang lebih representatif terhadap distribusi kelas.

	precision	recall	f1-score	support
0	0.92	0.99	0.95	163
1	0.98	0.77	0.86	66
accuracy			0.93	229
macro avg	0.95	0.88	0.91	229
weighted avg	0.93	0.93	0.93	229

Gambar 6. Nilai akurasi, precision, recall, dan f-1 score

Gambar 7 merupakan confusion matrix dari model klasifikasi Naïve Bayes, yang mengevaluasi kinerja model dalam mengklasifikasikan dua kelas: "Bukan Penipuan" dan "Penipuan." Dalam matriks ini, terdapat 162 true negative (TN) yang menunjukkan bahwa model berhasil mengidentifikasi 162 contoh sebagai "Bukan Penipuan" dengan benar. Hanya terdapat 1 false positive (FP), yaitu satu contoh sebenarnya adalah "Bukan Penipuan" tetapi diprediksi sebagai "Penipuan." Di sisi lain, terdapat 15 false negative (FN), yang berarti model gagal mengenali 15 contoh "Penipuan," sedangkan 51 true positive (TP) berhasil diidentifikasi.



Gambar 7. Confusion matrix

Hasil analisis menunjukkan bahwa pemeringkatan fitur menggunakan Chi-Square berhasil mengidentifikasi kata-kata kunci seperti “paket,” “hadiah,” dan “kuota” sebagai fitur paling signifikan dalam membedakan SMS penipuan dari pesan legit. Seleksi fitur ini terbukti meningkatkan performa model secara keseluruhan dengan mempertahankan hanya fitur-fitur yang benar-benar informatif, sekaligus mengurangi dimensi data yang perlu diproses. Temuan ini konsisten dengan karakteristik umum SMS penipuan di Indonesia yang sering memanfaatkan iming-iming hadiah atau kuota internet sebagai umpan. Evaluasi model menggunakan berbagai metrik menunjukkan performa yang mengesankan dengan akurasi keseluruhan mencapai 93%. Model ini menunjukkan precision yang sangat tinggi (0.98) untuk kelas penipuan, artinya hampir semua pesan yang diklasifikasikan sebagai penipuan memang benar-benar penipuan. Namun, recall yang lebih rendah untuk kelas penipuan (0.77) mengindikasikan bahwa model masih melewatkan beberapa kasus penipuan, mungkin karena variasi pola teks yang belum sepenuhnya tercakup dalam data pelatihan.

Confusion matrix mengungkapkan bahwa model sangat baik dalam mengidentifikasi pesan legit (162 TN dari 163) namun memiliki keterbatasan dalam mendeteksi semua kasus penipuan (51 TP dari 66). Hal ini menimbulkan trade-off antara keamanan (mengurangi false negative) dan kenyamanan pengguna (menghindari false positive). Dalam konteks deteksi penipuan, false negative yang memungkinkan pesan penipuan lolos mungkin lebih berisiko daripada false positive yang hanya akan mengganggu pengguna. Implementasi pipeline pemrosesan teks yang terdiri dari case folding, normalisasi, stopwords removal, dan stemming terbukti efektif dalam mempersiapkan data untuk klasifikasi. Kombinasi TF-IDF untuk ekstraksi fitur dan Chi-Square untuk seleksi fitur menghasilkan representasi teks yang optimal untuk algoritma Naïve Bayes. Hasil ini memperkuat bukti bahwa pendekatan tradisional dalam pemrosesan bahasa alami tetap relevan untuk tugas klasifikasi teks tertentu, terutama dengan dataset berukuran sedang.

Secara keseluruhan, temuan penelitian ini mendemonstrasikan bahwa model Naïve Bayes dengan optimasi seleksi fitur dapat menjadi solusi praktis dan efisien untuk deteksi SMS penipuan. Meskipun memiliki beberapa keterbatasan, model ini menawarkan akurasi yang memadai untuk implementasi nyata, terutama jika dikombinasikan dengan mekanisme verifikasi tambahan. Keberhasilan ini membuka peluang untuk pengembangan sistem deteksi yang lebih canggih sekaligus menyoroti pentingnya kolaborasi multidisiplin dalam memerangi kejahatan siber berbasis teks.

5. PENUTUP

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan, dapat disimpulkan bahwa algoritma Naïve Bayes terbukti sangat efektif dan andal dalam mendeteksi SMS penipuan. Melalui

serangkaian tahapan yang sistematis, mulai dari pra-pemrosesan data (meliputi case folding, normalisasi, stopwords removal, dan stemming), ekstraksi fitur menggunakan TF-IDF, hingga seleksi fitur dengan Chi-Square, model yang dikembangkan berhasil mencapai tingkat akurasi sebesar 93%. Kinerja model ini juga diperkuat dengan perolehan nilai presisi, recall, dan F1-score yang seimbang, yaitu sebesar 93%, yang mengindikasikan kemampuan tinggi dalam mengklasifikasikan SMS penipuan dan bukan penipuan secara akurat serta memperkuat validitas hasil analisis. Keberhasilan ini menunjukkan bahwa pendekatan klasifikasi teks berbasis Naïve Bayes merupakan solusi yang kuat untuk membedakan antara pesan yang sah dan yang berpotensi menipu, sehingga dapat membantu melindungi pengguna dari ancaman kejahatan siber.

Menindaklanjuti keberhasilan model Naïve Bayes ini, beberapa saran diajukan untuk pengembangan di masa depan. Pertama, disarankan agar model ini diimplementasikan ke dalam aplikasi mobile atau sistem peringatan SMS real-time yang dapat memberikan notifikasi langsung kepada pengguna mengenai potensi adanya penipuan. Kedua, penelitian lebih lanjut perlu dilakukan dengan menggunakan dataset yang lebih besar dan beragam, mencakup modus-modus penipuan yang lebih baru, untuk memastikan model tetap relevan dan akurat. Eksplorasi algoritma lain seperti ensemble methods atau deep learning juga dapat dipertimbangkan untuk lebih meningkatkan kinerja deteksi. Terakhir, kolaborasi dengan lembaga pemerintah dan penyedia layanan telekomunikasi menjadi krusial untuk menyebarluaskan informasi dan meningkatkan upaya pencegahan, sehingga dapat mengurangi jumlah korban penipuan dan meningkatkan keamanan komunikasi digital secara menyeluruh.

REFERENSI

[1] Rosmayati. (Nov 2023). Kominfo: Ada 958 kasus penipuan berkedok sms. Bloomberg Technoz. https://www.bloombergtechnoz.com/detail-news/21101/kominfo-ada-958-kasus-penipuan-berkedok-sms.

[2] Sofyan, M. A., Rahaningsih, N., & Dana, R. D. (2024). Deteksi sms spam berbahasa indonesia menggunakan algoritma support vector machine. JATI: Jurnal Mahasiswa Teknik Informatika, 8(3), 3071-3079.

[3] Rosyidi, M. I. U., & Rochmawati, N. (2024). Implementasi ensemble learning adaboost pada algoritma klasifikasi decision tree dan svm untuk klasifikasi sms berbahasa Indonesia. JIEET: Journal of Information Engineering and Educational Technology, 8(1), 7-13.

[4] Alvares, J., & Saputro, U. A. (2023). Klasifikasi short message service spam menggunakan algoritma naïve bayes classifier. Smart Comp: Jurnalnya Orang Pintar Komputer, 12(4), 885-893.

[5] Liliana, L., Hartono, H., & Bernanda, D. Y. (2020). Integrasi data mining dan online analytical processing (olap) pada data performa siswa. Jurnal Sisfokom: Sistem Informasi dan Komputer, 9(3), 400-406.

[6] Maryoosh, A. A., & Hussein, E. M. (2022). A review: Data mining techniques and its applications. International Journal of Computer Science and Mobile Applications, 10(3), 1-14.

[7] Darmawan, I. P. E., Djuri, P. A., & Rhamadhani, R. F. (2024). Implementasi artificial intelligence dalam dunia auditing: sebuah peluang atau tantangan baru. JAIM: Jurnal Akuntansi Manado, 5(3), 675-683.

[8] Rajeswari, P., Sathishkumar, V. E., Anilkumar, C., Thilakaveni, P., & Moorthy, U. (2023). Big data analytics and implementation challenges of machine learning in big data. Applied and Computational Engineering, 233-238.

[9] Wahid, A., Baharulloh, M., Kahfiansyah, R., Abrilianto, T., Saifudin, A., & Mulyati, S. (2021). Identifikasi sms spam menggunakan metode naive bayes. Jurnal Informatika Universitas Pamulang, 6(3), 536-539.

[10] Azzahra, F. N., Rohana, T., Rahmat, R., & Juwita, A. R. (2024). Penerapan metode naive bayes dalam klasifikasi spam sms menggunakan fitur teks untuk mengatasi ancaman pada pengguna. Journal of Information System Research (JOSH), 5(3), 873-880.

[11] Putera, A. W., S., & Lestari, Y. D. (2023). Klasifikasi SMS Spam Menggunakan Algoritma K-Nearest Neighbour. Jurnal Ilmu Komputer Dan Sistem Komputer Terapan, 5(1), 43-55.

[12] Herwanto, H., Chusna, N. L., & Arif, M. S. (2021). Klasifikasi sms spam berbahasa indonesia menggunakan algoritma multinomial naïve bayes. Jurnal Media Informatika Budidarma, 5(4), 1316.

[13] Dwiyansaputra, R., Nugraha, G. S., Bimantoro, F., & Aranta, A. (2021). Deteksi sms spam berbahasa Indonesia menggunakan tf-idf dan stochastic gradient descent classifier. JTIKA: Jurnal Teknologi Informasi, Komputer, dan Aplikasinya, 3(2), 200-207.

[14] Reviantika, F., Azhar, Y., & Marthasari, G. I. (2021). Analisis klasifikasi sms spam menggunakan logistic regression. Jurnal Repositor, 3(4), 387-392.

[15] Kesuma, I. G. M. J., Widiati, I. A. P., & Sugiartha, I. N. G. (2020). Penegakan hukum terhadap penipuan melalui media elektronik. Jurnal Preferensi Hukum, 1(2), 72-77.

[16] Astuti, A. P., Alam, S., & Jaelani, I. (2022). Komparasi algoritma support vector machine dengan naive bayes untuk analisis sentimen pada aplikasi brimo. Jurnal Bangkit Indonesia, 11(2), 1-6.

[17] Alshdaifat, E. A., Alshdaifat, D. A., Alsarhan, A., Hussein, F., & El-Salhi, S. M. D. F. S. (2021). The effect of preprocessing techniques, applied to numeric features, on classification algorithms' performance. Data, 6(2), 11.

[18] Ma'rifah, H., Wibawa, A. P., & Akbar, M. I. (2020). Klasifikasi artikel ilmiah dengan berbagai skenario preprocessing. Sains, Aplikasi, Komputasi dan Teknologi Informasi, 2(2), 70-78.

[19] Ramadhanti, F., Wibisono, Y., & Sukamto, R. A. (2019). Analisis morfologi untuk menangani out-of-vocabulary words pada part-of-speech tagger bahasa Indonesia menggunakan hidden markov model. JLK: Jurnal Linguistik Komputasional, 2(1), 6-12.

[20] Duha, T., Laia, M., Huda, A. K., & Jasuma, A. (2023). Klasifikasi data gempa bumi di pulau sumatera menggunakan algoritma naïve bayes. Jurnal Informatika, 2(1), 23-27.

[21] Ksnugroho. Klasifikasi-spam-sms. (2019). https://github.com/ksnugroho/klasifikasi-spam-sms.

[22] Widyanto, A. (2023). Pengaruh Keseimbangan Data terhadap Akurasi Model Support Vector Machine pada Data Set Donor Darah. Jurnal Teknologi Terpadu, 9(2), 79-88.