Identifikasi Pola Penipuan Pesan Singkat

Syaikha Amirah Zikrina
S1 Sains Data
Telkom University
Bandung, Indonesia
syaikhaaz@student.telkomuniversity.ac.id

Syifa Putri Fadhilla
S1 Sains Data
Telkom University
Bandung, Indonesia
syifaputrifadhilla@student.telkomuniversity.ac.id

Syifa Salsabila
S1 Sains Data
Telkom University
Bandung, Indonesia
syifasalsabilaa@student.telkomuniversity.ac.id

***Abstract*—Sistem ini menyajikan pengembangan dan implementasi untuk mengidentifikasi pola penipuan dalam pesan singkat menggunakan algoritma *Logistic Regression.* Sistem ini bertujuan untuk mengklasifikasikan pesan SMS sebagai penipuan atau non-penipuan dengan akurasi tinggi. Dataset terdiri dari berbagai sampel pesan singkat yang dikategorikan ke dalam penipuan dan non-penipuan. Metodologi mencakup pra-pemrosesan, ekstraksi fitur, seleksi fitur, dan klasifikasi menggunakan *Logistik Regression*. Hasil menunjukkan efektivitas sistem dengan akurasi 90% dan F1 Score sebesar 0.89.**

***Keywords*—deteksi penipuan, SMS, penambangan teks, regresi logistik, klasifikasi teks**

I. PENDAHULUAN

A. Latar Belakang

Perkembangan teknologi informasi telah memperluas jangkauan dan kemudahan komunikasi, termasuk dalam hal pengiriman pesan teks melalui SMS. Namun, bersamaan dengan manfaatnya, teknologi ini juga membuka pintu bagi penipuan dan eksploitasi. Penjahat cyber semakin cerdik dalam menciptakan skema penipuan yang menggunakan pesan teks untuk mencuri informasi pribadi, meminta pembayaran yang tidak sah, atau menawarkan produk atau layanan palsu. Khususnya, penipuan melalui SMS sering kali menargetkan pengguna yang kurang waspada atau tidak berpengalaman dalam mengenali tanda-tanda penipuan. Oleh karena itu, perlunya sebuah sistem pendeteksi SMS penipuan yang handal menjadi semakin mendesak.

Dengan memahami latar belakang ini, pengembangan solusi yang mampu membedakan antara pesan yang sah dan mencurigakan menjadi sangat penting untuk melindungi pengguna dari potensi kerugian finansial dan pelanggaran privasi. Analisis teks merupakan pilihan tepat untuk membangun sistem ini. Pesan penipuan SMS memiliki karakteristik teks unik yang dapat diidentifikasi dengan analisis teks, seperti penggunaan kata "gratis", "menangkan", "segera", "darurat", dan "tautan", serta kesalahan tata bahasa dan ejaan. Pada penelitian kali ini, kami menggunakan metode klasifikasi Naïve Bayes yang merupakan metode populer dan menjadi pilihan dalam pengklasifikasian, terutama dalam penyaringan spam. Selain itu, Naïve Bayes classifier sangat tepat untuk digunakan dalam pengklasifikasian pesan teks pada SMS inbox karena memberikan hasil filtrasi yang cukup akurat untuk menyaring SMS yang masuk. Keuntungan utama dari penggunaan metode Naïve Bayes adalah kemampuannya untuk melakukan estimasi parameter dengan menggunakan jumlah data latih yang relatif kecil. Dalam banyak situasi dunia nyata yang kompleks, Naïve Bayes sering kali dapat memberikan kinerja yang lebih baik daripada yang diharapkan [3].

B. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan,maka berikut adalah rumusan masalah dalam penelitian ini, sebagai berikut:

1. Bagaimana mengembangkan sistem yang mampu mengenali pola bahasa penipuan secara otomatis?
2. Bagaimana memfilter pesan SMS penipuan dengan akurat dan efisien?
3. Bagaimana memastikan akurasi identifikasi pesan penipuan agar meminimalkan kesalahan klasifikasi?
4. Bagaimana melindungi masyarakat dari potensi kerugian akibat penipuan melalui pesan SMS?

C. Tujuan

Dalam mengembangkan sistem pendeteksi SMS penipuan yang handal dan akurat

untuk melindungi pengguna dari potensi kerugian finansial dan pelanggaran privasi [4],

dengan cara:

1. Membedakan pesan SMS yang sah dan mencurigakan
2. Memberikan peringatan kepada pengguna
3. Analisis pesan mencurigakan

II. KAJIAN TEORI

A. Penelitian Terkait

Berdasarkan tinjauan pustaka yang telah dilakukan, terdapat peneliti terdahulu yang telah menganalisis mengenai identifikasi sms penipuan seperti yang telah dilakukan Ferin dkk. Penelitian tersebut melakukan klasifikasi sms spam dengan menggunakan logistic regression dan penelitian berhasil membuktikan model tersebut mampu mencapai akurasi yang baik yaitu sebesar 95%. Selain itu pada penelitian lain yang memiliki topik relevan dengan topik ini yaitu seperti yang dilakukan oleh Sravya dkk [Sravya]

B. Klasifikasi Teks

Dataset yang digunakan dalam studi ini terdiri dari pesan SMS yang diberi label sebagai penipuan atau non-penipuan. Pesan-pesan ini dikumpulkan dari berbagai sumber untuk memastikan sampel yang beragam dan representatif. Dataset

kemudian dipra-pemrosesan untuk menghilangkan kebisingan dan informasi yang tidak relevan.

C. Logistic Regression

Logistic Regression (LR) adalah teknik statistika yang digunakan untuk memodelkan hubungan antara satu atau beberapa variabel independen (predictor) dengan variabel dependen biner (respon). Tujuan LR adalah untuk memprediksi kemungkinan bahwa variabel-variabel independen akan menentukan kemungkinan terjadinya suatu kejadian. Pada dasarnya, LR mencari hubungan antara variabel input, seperti fitur data, dan kemungkinan bahwa output dalam kasus klasifikasi biner adalah kelas tertentu [1]

Beberapa konsep yang relevan dalam regresi logistik meliputi:

1. Variabel dependen biner: Variabel dependen dalam regresi logistik adalah variabel biner yang menggambarkan hasil atau kategori yang mungkin.
2. Koefisien regresi: Koefisien regresi dalam regresi logistik mengindikasikan pengaruh variabel prediktor terhadap peluang kejadian sukses. Koefisien positif menunjukkan hubungan positif, sedangkan sebaliknya.
3. Odds ratio: Odds ratio menggambarkan perubahan dalam peluang kejadian sukses sebagai hasil dari perubahan satu unit dalam variabel prediktor. Odds ratio yang lebih besar dari 1 menunjukkan peningkatan peluang, sementara yang lebih kecil dari 1 menunjukkan penurunan peluang.
4. Evaluasi model: Akurasi,Presisi,Recall dan area bawah kurva ROC (AUC-ROC)[2].

D. Term Frequency – Inverse Document Frequency (TF-IDF)

Term Frequency—Inverse Document Frequency (TF-IDF) adalah sebuah algoritma yang bermanfaat untuk mengevaluasi bobot relatif dari semua kata yang sering digunakan. Setelah proses preprocessing, pembobotan kata adalah tahap yang sangat penting [3]. Skor TF-IDF yang dihasilkan untuk setiap frasa dalam setiap tweet diintegrasikan ke dalam vektor fitur yang mewakili tweet selama proses analisis sentimen pembelajaran mesin. Teknik ini umum digunakan untuk mengevaluasi pentingnya frasa dalam teks dan kelangkaan relatifnya dalam korpus [4].
Menurut penelitian sebelum nya, TF-IDF bekerja dengan melibatkan perkalian antara Term Frequency (TF) dengan Inverse Document Frequency (IDF). TF memiliki tujuan untuk menunjukkan jumlah kemunculan sebuah kata pada suatu tweet. IDF memiliki tujuan untuk menghitung frekuensi kemunculan suatu kata pada seluruh tweet (Brownlee, 2019). Semakin tinggi nilai dari TF-IDF makan semakin jaran kemunculan suatu kata tersebut dalam sekumpulan tweet. Rumus yang digunakan dalam memperoleh nilai TF-IDF dituliskan dengan persamaan berikut:

$$w_{ij} = tf_{ij} \times idf$$

$$idf = \log\left(\frac{N}{df_i}\right)$$

dimana wij merupakan bobot dari kata i pada tweet ke-j, N merupakan jumlah seluruh tweet, ij tf merupakan jumlah munculnya kata i pada tweet ke-j, dan i df adalah banyaknya tweet yang mengandung kata i [5].

E. Chi-Square

Metode Chi Square Feature Selection digunakan untuk memilih fitur yang paling relevan dalam klasifikasi teks untuk menentukan apakah sebuah buku termasuk dalam kategori komik atau bukan. Metode ini didasarkan pada uji statistika Chi Square untuk mengevaluasi keterkaitan antara atribut (seperti periode penerbitan, materi, dan fisik) dengan klasifikasi buku sebagai komik atau bukan komik [6].

III. METODE PENELITIAN

A. Dataset

Pada penelitian penelitian digunakanlah dataset SMS Spam berbahasa Indonesia yang diperoleh dari link github.com/kmkurn/id-nlp-resource. Dataset tersebut memiliki kolom teks yang berisi teks dari pesan SMS dan kolom label yang berisi label dari kelas pesan SMS. Data dalam dataset tersebut dapat digunakan untuk melakukan pengujian dan evaluasi sistem pengklasifikasi SMS spam. Tahap pra-pemrosesan dataset akan melibatkan langkah-langkah untuk membersihkan teks dari karakter khusus, mengubahnya menjadi huruf kecil, serta menghilangkan unsur-unsur yang tidak relevan atau mengganggu.

B. Preprocessing

Data yang telah didapatkan kemudian diolah melalui tahap *preprocessing* atau persiapan data. Tahap ini bertujuan untuk membersihkan dan meningkatkan 18 kualitas data agar siap digunakan dalam analisis selanjutnya.Dengan data yang bersih dan berkualitas tinggi, hasil analisis akan menjadi lebih akurat, efisien, dan mudah diinterpretasikan. Adapun tahapan yang dilakukan di antaranya:

1. *Data Cleaning*

   Pada tahap ini dilakukan pembersihan data terkait tanda baca, simbol, angka, emoji yang tidak diperlukan dalam analisis.

2. *Tokenization*

   Tahap ini merupakan proses pemisahan kalimat menjadi token token atau kumpulan kata yang dipisah oleh spasi menjadi kata kata secara individu.

3. Normalisasi Kata

   Pada normalisasi akan dilakukan pengubahan kata singkatan dan kata tidak baku menjadi bentuk lengkap dan bakunya.

4. *Case Folding*

   Pada tahap ini seluruh kata dalam data diubah menjadi huruf kecil, sehingga data menjadi seragam dan mudah untuk dibandingkan serta mempercepat proses analisis.

5. *Stopword removal*

   Pada tahap ini merupakan proses penghapusan kata yang dianggap tidak memiliki pengaruh penting serta makna yang tidak spesifik.

6. *Stemming*

   Pada tahap ini dilakukan proses pengubahan suatu kata yang memiliki imbuhan menjadi kata dasar.

Berikut ini merupakan contoh pada preprocessing data dengan data asli "2016-07-08 11:47:11.Plg Yth, sisa kuota Flash Anda 478 KB. Download MyTelkomsel apps"

TABLE I. CONTOH PREPROCESSING

Cleaning	Plg Yth sisa kuota Flash Anda KB Download MyTelkomsel apps
Tokenisasi	"Plg", "Yth", "sisa", "kuota", "Flash", "Anda", "KB", "Download", "MyTelkomsel", "apps"
Normalisasi Kata	"Pelanggan", "Yang", "Terhormat", "sisa", "kuota", "Flash", "Anda", "Kilobyte", "Download", "MyTelkomsel", "aplikasi"
Case Folding	"pelanggan", "yang", "terhormat", "sisa", "kuota", "flash", "anda", "kilobyte", "download", "mytelkomsel", "aplikasi"
Stopword removal	"pelanggan", "terhormat", "sisa", "kuota", "flash", "kilobyte", "download", "mytelkomsel", "aplikasi"
Stemming	"pelanggan", "terhormat", "sisa", "kuota", "flash", "kilobyte", "download", "mytelkomsel", "aplikasi"

C. *Pembobotan Kata (TF-IDF)*

Setelah melakukan tahapan preprocessing maka proses selanjutnya yaitu pembobotan kalimat dengan menggunakan metode TF-IDF (Term Frequency -Inverse Document Frequency). TF-IDF merupakan proses pembobotan kata dalam analisis sentimen dan text mining. Tujuannya pembobotan ini adalah untuk mengoptimalkan kemampuan analisis dengan mengidentifikasi kata-kata yang paling penting dalam dokumen. Metode pembobotan ini dilakukan melalui perhitungan bobot kata berdasarkan dua faktor yaitu pertama, bobot kata dihitung berdasarkan frekuensi kemunculan kata dalam dokumen(TF). Nilainya akan semakin tinggi seiring dengan frekuensi kemunculan kata tersebut yang semakin sering. Bentuk TF terdapat pada persamaan ()

$$TF = \frac{jumlah\ kemunculan\ term\ pada\ suatu\ dokumen}{jumlah\ seluruh\ term\ dalam\ dokumen}$$

Kedua, bobot kata dihitung berdasarkan kebalikan frekuensi dokumen (IDF). Semakin jarang kata muncul di seluruh dokumen, semakin tinggi nilainya [7].

$$IDF = log\frac{jumlah\ seluruh\ dokumen}{jumlah\ dokumen\ suatu\ term\ muncul}$$

Nilai TF-IDF dihitung dengan mengalikan bobot frekuensi kemunculan kata (TF) dengan bobot kebalikan frekuensi dokumen (IDF) [7].

$$\text{TF} - \text{IDF} = TF \times IDF$$

Kata-kata dengan nilai TF-IDF tinggi dianggap lebih penting karena menunjukkan bahwa kata tersebut sering muncul dalam dokumen tertentu, tetapi jarang muncul di seluruh dokumen

D. *Seleksi Fitur (Chi-Square)*

Proses pemilihan fitur yang relevan dalam data dikenal sebagai seleksi fitur. Tujuan seleksi fitur adalah untuk menghilangkan fitur-fitur yang tidak relevan dan mengurangi dimensi data, sehingga algoritma klasifikasi dapat bekerja lebih baik dalam menganalisa dan memprediksi data. Untuk mengetahui apakah fitur dan kelas memiliki ketergantungan yang kuat satu sama lain, nilai chi-kuadrat dan signifikansi dihitung dengan metode statistik chi-kuadrat [6].

E. Klasifikasi (*Logistic Regression*)

Logistic Regression merupakan metode klasifikasi probabilistik yang memanfaatkan pembelajaran mesin terawasi. Pengklasifikasi ini membutuhkan input dan output untuk membedakan antara berbagai kelas. Logistic Regression digunakan untuk mengklasifikasikan observasi ke dalam salah satu dari dua kelas atau kedalam salah satu dari banyak kelas. [1]

$$P(Y|X) = \frac{\exp\left[\beta_0 + \sum_i \beta_i X_i\right]}{1 + \exp\left[\beta_0 + \sum_i \beta_i X_i\right]}$$

Keterangan:
P(Y | X) = Peluang kelas Y pada observasi X
β0 = Bias
βi = Vektor bobot
Xi = Variabel pengamatan

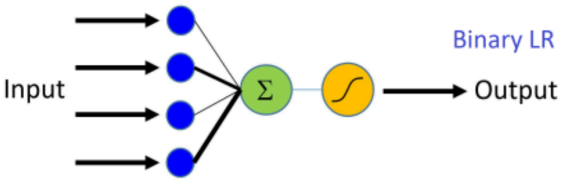

Fig. 1. Binary Logistic Regression

Binary Logistic Regression adalah metode untuk memprediksi salah satu dari dua kelas yang mungkin. Model ini menggabungkan beberapa fitur input dengan bobot tertentu untuk menghasilkan kombinasi linear dari input tersebut. Hasil dari kombinasi linear ini kemudian dilewatkan melalui fungsi aktivasi sigmoid.

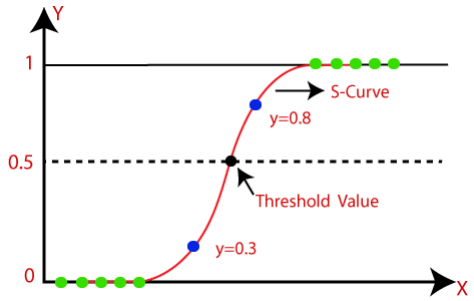

Fig. 2. Kurva Sigmoid

Dalam *Binary Logistic Regression*, fungsi sigmoid digunakan untuk mengubah hasil kombinasi linear dari input menjadi nilai probabilitas antara 0 dan 1. Probabilitas ini digunakan untuk menentukan kelas output, biasanya dengan menetapkan ambang batas pada 0.5. Jika nilai probabilitas di atas 0.5, input diklasifikasikan sebagai kelas 1, dan jika di bawah 0.5, diklasifikasikan sebagai kelas 0.

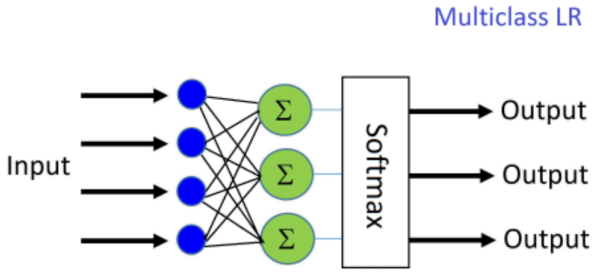

Fig. 3. Multiclass Logistic regression

Multiclass Logistic Regression digunakan untuk memprediksi salah satu dari beberapa kelas yang mungkin. Model ini juga menggabungkan beberapa fitur input dengan bobot tertentu, tetapi menghasilkan kombinasi linear terpisah untuk setiap kelas yang mungkin. Hasil dari kombinasi linear ini kemudian dilewatkan melalui fungsi softmax, yang mengubah nilai-nilai tersebut menjadi probabilitas untuk setiap kelas. Probabilitas total untuk semua kelas dijamin berjumlah 1. Prediksi akhir dibuat berdasarkan kelas dengan probabilitas tertinggi, sehingga model ini mampu menangani masalah klasifikasi dengan lebih dari dua kategori.

IV. Hasil dan Pembahasan

A. Hasil Eksperimen

Eksperimen ini mengembangkan aplikasi berbasis Streamlit untuk mendeteksi pesan singkat yang mungkin termasuk dalam kategori penipuan, promosi, atau spam. Aplikasi ini menggunakan model klasifikasi teks yang telah dilatih sebelumnya dan diimplementasikan menggunakan `pickle` untuk membuat model deteksi pesan dan fitur terpilih dari vectorizer TF-IDF.

Pada halaman utama aplikasi, pengguna disambut dengan penjelasan singkat tentang tujuan aplikasi dan beberapa metrik menarik terkait penipuan pesan singkat, seperti total kerugian akibat penipuan dan persentase penerima pesan penipuan. Aplikasi ini juga menampilkan grafik horizontal yang menunjukkan berbagai modus penipuan digital beserta persentasenya, dengan warna latar belakang dan teks yang telah disesuaikan agar kontras dengan latar belakang aplikasi.

Selanjutnya, aplikasi menampilkan dua word cloud yang menunjukkan kata-kata yang sering muncul pada pesan dengan label "Fraud" dan "Promo", yang membantu pengguna untuk memahami pola umum dalam pesan-pesan tersebut.

Pada halaman "Deteksi Pesan", pengguna dapat memasukkan pesan teks yang ingin mereka periksa. Setelah menekan tombol "Hasil Deteksi", aplikasi menggunakan model yang dimuat untuk mengklasifikasikan pesan tersebut menjadi salah satu dari empat kategori: SMS Normal, SMS Fraud, SMS Promo, atau SMS Spam. Hasil deteksi ditampilkan dengan pesan peringatan yang sesuai, memberikan informasi tambahan kepada pengguna tentang cara merespons pesan tersebut.

Eksperimen ini menunjukkan bagaimana penggunaan model klasifikasi teks dan visualisasi data dapat membantu dalam mengidentifikasi dan memahami ancaman dari pesan penipuan dan promosi yang sering diterima oleh pengguna. Aplikasi ini juga mengilustrasikan bagaimana Streamlit dapat digunakan untuk membuat antarmuka pengguna yang interaktif dan informatif.

B. Pembahasan

1. Visualisasi Data
   - Grafik batang label

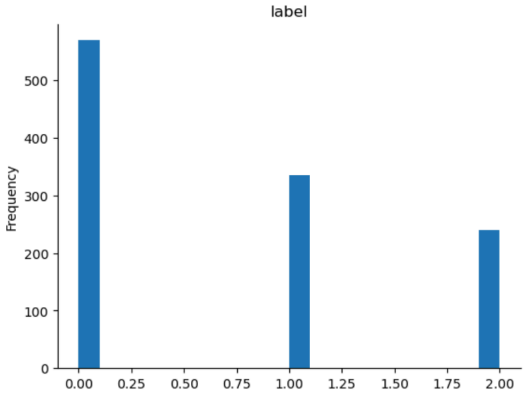

Fig. 4. Grafik batang label

Grafik batang diatas menggambarkan jumlah pesan yang telah diberi label berdasarkan frekuensi kedatangannya. Grafik ini memiliki sumbu sebagai berikut:

- **Sumbu X:**Jumlah label
- **Sumbu Y:** Frekuensi

Label 0 (Normal) Label ini kemungkinan menunjukkan bahwa titik data mewakili aktivitas promosi normal atau asli tanpa perilaku penipuan.

Label 1 (Penipuan) Label ini menunjukkan bahwa titik data terkait dengan aktivitas promosi

penipuan, yang menyiratkan bahwa promosi tersebut mungkin menipu, menyesatkan, atau ditujukan untuk keuntungan ilegal.

Label 2 (Promosi) Label ini dapat menunjukkan bahwa titik data mewakili aktivitas promosi, tetapi tidak jelas apakah promosinya curang atau tidak. Mungkin diperlukan penyelidikan lebih lanjut untuk menentukan keabsahannya.

Grafik menunjukkan adanya korelasi positif yang kuat antara frekuensi kedatangan pesan dan jumlah label yang ditetapkan untuk setiap pesan. Artinya, pesan yang datang lebih sering lebih mungkin diberi label daripada pesan yang jarang datang.

- Word Cloud

Fig. 5. Word Cloud Fraud dan Promosi

Untuk memberikan gambaran lebih lanjut, di dalam aplikasi menampilkan word cloud yang menggambarkan kata-kata yang sering muncul dalam pesan fraud dan promo. WordCloud ini membantu pengguna untuk secara intuitif melihat pola kata-kata yang mencolok dalam setiap kategori pesan fraud atau pesan promosi.

- Grafik Batang Modus Penipuan

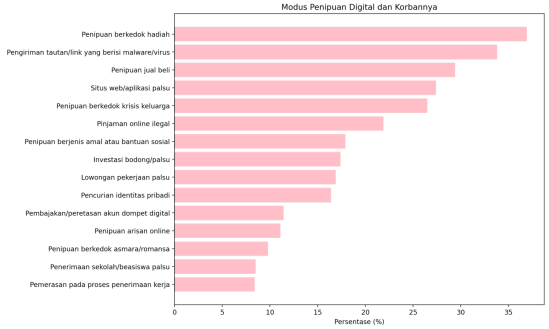

Fig. 6. grafik Batang Modus Penipuan

Grafik diatas menunjukkan bahwa modus penipuan digital yang paling banyak menelan korban adalah penipuan berkedok hadiah, dengan persentase 36,9%. Diikuti oleh pengiriman tautan/link yang berisi malware atau virus dengan persentase 33,8%, penipuan jual beli dengan persentase 29,4%, situs web/aplikasi palsu dengan persentase 27,4%, dan penipuan berkedok krisis keluarga dengan persentase 26,5%. Hal ini menunjukkan bahwa modus penipuan digital yang paling sering terjadi adalah modus yang memanfaatkan psikologis korban, seperti penipuan berkedok hadiah dan penipuan berkedok krisis keluarga. Modus penipuan digital lainnya,seperti pengiriman tautan/link yang berisi malware/virus dan situs web/aplikasi palsu, memanfaatkan kelemahan teknologi yang dimiliki korban.

2. Confusion Matrix

Confusion matrix (matriks kebingungan) adalah alat yang digunakan dalam machine learning untuk mengevaluasi kinerja model klasifikasi. Matriks ini menyajikan informasi tentang hasil klasifikasi sebenarnya vs. prediksi yang dihasilkan oleh model. Confusion matrix memungkinkan kita untuk melihat dengan jelas bagaimana model kita membuat keputusan dan di mana kesalahan prediksi terjadi.

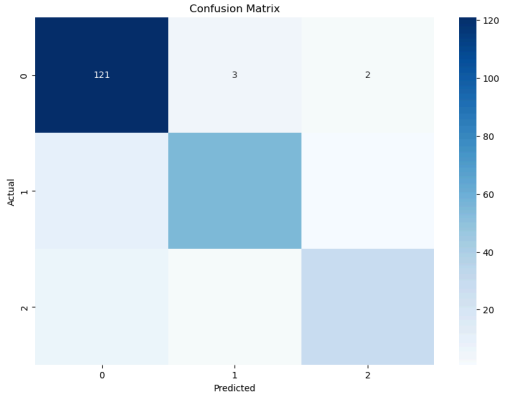

Fig. 7. Confusion Matrix

- Akurasi untuk Kelas 0: Dari total 126 instance yang sebenarnya merupakan kelas 0 (121+3+2), 121 instance diprediksi dengan benar. Ini berarti model cukup baik dalam memprediksi kelas 0.
- Akurasi untuk Kelas 1: Dari total 55 instance yang sebenarnya merupakan kelas 1, semuanya diprediksi dengan benar.
- Akurasi untuk Kelas 2: Dari total 42 instance yang sebenarnya merupakan kelas 2, semuanya diprediksi dengan benar.Dari confusion matrix didapatkan hasil berupa precision, recall, f1-score, dan akurasi.Didapatkan akurasi sebesar 90% dengan rata-rata presisi sebesar 0.90, recall sebesar 0.86, dan F1-Score sebesar 0.88.

	precision	recall	f1-score	support
0	0.88	0.96	0.92	126
1	0.92	0.83	0.87	66
2	0.91	0.78	0.84	37
accuracy			0.90	229
macro avg	0.90	0.86	0.88	229
weighted avg	0.90	0.90	0.89	229

Fig. 8. Evaluasi Kinerja

3. Aplikasi Streamlit

Aplikasi ini memungkinkan pengguna untuk memasukkan teks pesan yang mereka terima dan memberikan hasil deteksi berdasarkan model machine learning yang telah dilatih. Dengan antarmuka yang sederhana dan intuitif, pengguna dapat dengan mudah mengetahui apakah pesan yang mereka terima termasuk dalam kategori fraud, promo, atau normal.

- Halaman Home

Fig. 9. Halaman Home Aplikasi

Pada halaman home terdapat penjelasan mengenai aplikasi kami, visualisasi modus penipuan digital dan korbannya, dan juga visualisasi word cloud.

- Halaman Deteksi Pesan

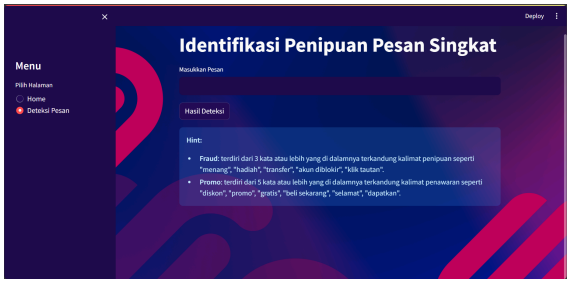

Fig. 10. Halaman Deteksi Pesan Aplikasi

Pada halaman deteksi pesan, pengguna dapat menginputkan pesan yang ingin di deteksi. Selain itu, aplikasi ini juga memberikan panduan (hint) mengenai kata-kata atau frasa yang biasanya terkait dengan penipuan atau promosi, membantu pengguna memahami kriteria yang digunakan dalam proses deteksi.

V. Kesimpulan

Penelitian ini berhasil membangun model deteksi pesan penipuan dengan menggunakan teknik NLP dan algoritma Machine Learning. Model yang dikembangkan memiliki akurasi tinggi dan dapat digunakan untuk mendeteksi pesan penipuan secara otomatis. Penelitian ini diharapkan dapat memberikan kontribusi dalam upaya pencegahan penipuan melalui pesan singkat. Sistem ini bekerja dengan cara mengumpulkan data SMS yang diberi label sebagai penipuan atau non-penipuan, memproses data SMS untuk membersihkan dan meningkatkan kualitasnya, memilih fitur yang paling relevan dalam klasifikasi teks dengan menggunakan metode Chi-Square, serta mengklasifikasikan pesan SMS sebagai penipuan atau non-penipuan dengan menggunakan algoritma Logistic Regression.

Hasil penelitian menunjukkan bahwa sistem ini mampu mencapai akurasi sebesar 90% dengan rata-rata presisi sebesar 0.90, recall sebesar 0.86, dan F1-Score sebesar 0.88, yang menunjukkan efektivitas sistem ini dalam mendeteksi SMS penipuan. Sistem ini memiliki potensi untuk diterapkan dalam berbagai aplikasi, seperti aplikasi mobile untuk mendeteksi SMS penipuan, sistem keamanan untuk menyaring pesan SMS spam, dan layanan edukasi untuk meningkatkan kesadaran masyarakat tentang penipuan SMS. Dengan menerapkan sistem ini, diharapkan dapat membantu mengurangi jumlah korban penipuan SMS dan meningkatkan keamanan pengguna internet.

Referensi

[1] R. Klabunde, “Daniel Jurafsky/James H. Martin, Speech and Language Processing,” *Zeitschrift Für Sprachwissenschaft*, vol. 21, no. 1, pp. 134–135, Jan. 2002, doi: 10.1515/zfsw.2002.21.1.134.

[2] M. A. Alrasyid, “Analisis regresi logistik.” https://rstudio-pubs-static.s3.amazonaws.com/1048753_c5900713e4eb425d813f602056db6ebd.html

[3] M. Nurjannah, H. Hamdani, and I. F. Astuti, “PENERAPAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK TEXT MINING,” *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, vol. 8, no. 3, pp. 110–113, Jun. 2016, doi: 10.30872/jim.v8i3.113.

[4] W. Ahmed, N. Semary, K. Amin, and M. A. Hammad, “Sentiment analysis on Twitter using machine learning techniques and TF-IDF feature extraction: a comparative study,” *IJCI International Journal of Computers and Information*, vol. 10, no. 3, pp. 52–57, Nov. 2023, doi: 10.21608/ijci.2023.236052.1128.

[5] E. D. N. Sari and I. Irhamah, “Analisis Sentimen Nasabah pada Layanan Perbankan Menggunakan Metode Regresi Logistik Biner, Naïve Bayes Classifier (NBC), dan Support Vector Machine (SVM),” *Jurnal Sains Dan Seni ITS*, vol. 8, no. 2, Feb. 2020, doi: 10.12962/j23373520.v8i2.44565.

[6] S. Anisah, A. S. Honggowibowo, and A. Pujiastuti, “KLASIFIKASI TEKS MENGGUNAKAN CHI SQUARE FEATURE SELECTION UNTUK MENENTUKAN KOMIK BERDASARKAN PERIODE, MATERI DAN FISIKDENGAN ALGORITMA NAIVEBAYES,” *Compiler*, vol. 5, no. 2, Nov. 2016, doi: 10.28989/compiler.v5i2.171.

[7] Suhasini and N.Vimala, “A hybrid TF-IDF and N-Grams based feature extraction approach for accurate detection of fake news on Twitter data,” *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 6, pp. 5710–5723, Sep. 2021, [Online]. Available: https://turcomat.org/index.php/turkbilmat/article/view/10885