JURNAL SISTEM INFORMASI TGD
Volume 4 Nomor 1, Januari 2025, Hal 55-64
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

Model Text Embedding dan TF-IDF+Ngram untuk Meningkatkan Kinerja Algoritma Binary Classifier pada Klasifikasi SMS Palsu

Sutriawan[1], Siti Mutmainnah[2], Teguh Ansyor Lorosae[3], Sahrul Ramadhan[4]
[1,2,3,4] Ilmu Komputer, Universitas Muhammadiyah Bima
Email: [1]sutriawan@umbima.ac.id[*], [2]siti.mutmainnah.id19@gmail.com, [3]ansyorlorosae95@gmail.com, [4]sahrulramadhanbinaswan@gmail.com
Email Penulis Korespondensi: sutriawan@umbima.ac.id

Abstrak
Seiring meningkatnya penggunaan SMS, deteksi SMS palsu (spam) menjadi tantangan dalam menjaga keamanan komunikasi. Algoritma klasifikasi berbasis teks, seperti Naive Bayes, Logistic Regression, dan Random Forest, memiliki performa yang bervariasi tergantung pada representasi fitur teks yang digunakan. Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma binary classifier dalam klasifikasi SMS palsu menggunakan representasi fitur TF-IDF, TF-IDF + Ngram, dan Word2Vec. Algoritma yang diuji meliputi Naive Bayes, Logistic Regression, Random Forest, dan Decision Tree, dengan metrik akurasi, precision, recall, dan F1-score sebagai evaluasi. Hasil penelitian menunjukkan bahwa Naive Bayes dengan TF-IDF mencapai akurasi 91.26%, sementara Random Forest dengan Word2Vec memperoleh akurasi 89.08%. Logistic Regression dengan TF-IDF + Ngram menunjukkan hasil lebih rendah. Temuan ini menegaskan pentingnya pemilihan representasi fitur yang tepat untuk meningkatkan akurasi klasifikasi SMS palsu.
Kata Kunci : Klasifikasi SMS, Naive Bayes, TF-IDF, Word2Vec, algoritma binary classifier.

Abstract
As the use of SMS increases, detection of fake SMS (spam) becomes a challenge in maintaining communication security. Text-based classification algorithms, such as Naive Bayes, logistic regression, and random forest, have varying performance depending on the text feature representation used. This study aims to evaluate the performance of binary classifier algorithms in fake SMS classification using TF-IDF, TF-IDF + N-gram, and Word2Vec feature representations. The tested algorithms include Naive Bayes, logistic regression, random forest, and decision tree, with accuracy, precision, recall, and F1-score metrics as evaluation. The results showed that Naive Bayes with TF-IDF achieved 91.26% accuracy, while Random Forest with Word2Vec obtained 89.08% accuracy. Logistic regression with TF-IDF + N-gram showed lower results. These findings emphasize the importance of selecting the right feature representation to improve the accuracy of fake SMS classification.
Keyword : *SMS classification, Naive Bayes, TF-IDF, Word2Vec, binary classifier algorithm.*

1. PENDAHULUAN

Penipuan melalui pesan singkat (SMS), yang dikenal sebagai SMS palsu, telah menjadi salah satu ancaman utama di era digital. Jumlah serangan yang menggunakan SMS untuk menipu korban melalui penawaran palsu, phishing, dan spamming terus meningkat. Masalah ini berdampak pada banyak layanan digital dan aplikasi mobile, sehingga diperlukan sistem deteksi yang efektif untuk mengidentifikasi dan mencegah SMS palsu. Oleh karena itu, klasifikasi SMS palsu menjadi topik yang penting dalam penelitian keamanan siber dan pengolahan bahasa alami (Natural Language Processing/NLP) [1]. SMS palsu, atau yang sering disebut SMS spam atau SMS penipuan, merupakan salah satu bentuk kejahatan siber yang banyak terjadi melalui platform pesan singkat. Pesan ini sering kali berisi informasi yang tidak diinginkan atau bahkan berbahaya, seperti penawaran palsu, tautan phishing, atau permintaan informasi pribadi. Pesan-pesan semacam itu tidak hanya mengganggu pengguna, tetapi juga dapat menyebabkan kerugian finansial yang signifikan bagi korban [2]. Fenomena SMS palsu semakin meningkat seiring dengan kemajuan teknologi komunikasi dan tingginya adopsi penggunaan ponsel. Dalam banyak kasus, pelaku kejahatan menggunakan teknik manipulasi psikologis untuk mendorong penerima mengikuti instruksi dalam pesan, seperti mengklik tautan berbahaya atau memberikan informasi pribadi yang kemudian dapat disalahgunakan. Akibatnya, SMS palsu menjadi ancaman serius, tidak hanya bagi individu, tetapi juga bagi perusahaan dan lembaga keuangan. Ancaman ini dapat menyebabkan kerugian berupa kebocoran data atau penyalahgunaan informasi pelanggan [1]. Sebagai respons terhadap ancaman tersebut, berbagai teknik telah dikembangkan untuk mengidentifikasi dan menyaring SMS palsu. Salah satu pendekatan yang populer adalah menggunakan algoritma klasifikasi dalam machine learning, yang memungkinkan deteksi otomatis terhadap pesan-pesan mencurigakan. Salah satu algoritma yang umum digunakan untuk tugas ini adalah binary classifier, yang mampu membedakan antara SMS palsu dan SMS sah berdasarkan karakteristik teks yang ada dalam pesan. Namun, untuk mencapai akurasi yang tinggi dalam mendeteksi SMS palsu, representasi teks yang digunakan oleh model klasifikasi sangatlah penting. Oleh karena itu, pemilihan model embedding yang tepat menjadi faktor kunci dalam meningkatkan kinerja algoritma binary classifier [2], [3]. Seiring meningkatnya volume SMS palsu yang beredar, penting untuk mengembangkan sistem deteksi yang lebih efektif dan efisien. Sistem ini harus mampu mengenali pola-pola umum dalam SMS palsu serta beradaptasi dengan berbagai teknik dan bahasa yang digunakan oleh pelaku kejahatan [4]. Dalam konteks

ini, penelitian ini bertujuan untuk mengeksplorasi bagaimana pemilihan model embedding yang tepat dapat meningkatkan akurasi deteksi SMS palsu menggunakan binary classifier .

Dalam konteks klasifikasi, pendekatan algoritma binary classifier mampu memisahkan pesan SMS menjadi dua kategori yaitu palsu atau bukan. Namun, tantangan utama dalam membangun model klasifikasi yang akurat terletak pada representasi teks yang digunakan sebagai input oleh algoritma. Representasi teks yang buruk dapat mengurangi kinerja model, menyebabkan kesalahan klasifikasi yang lebih tinggi. Oleh karena itu, penting untuk memilih model embedding yang tepat untuk menghasilkan representasi teks yang lebih baik dan meningkatkan kinerja classifier [5].

Permasalahannya adalah Penggunaan algoritma machine learning, khususnya binary classifier, telah terbukti efektif dalam mengklasifikasikan SMS palsu, namun akurasi deteksi yang tinggi sangat bergantung pada representasi teks yang digunakan. Meskipun teknik embedding seperti Word2Vec dan TF-IDF+N-gram telah banyak digunakan dalam berbagai aplikasi NLP, permasalahan ini belum dapat diatasi dengan baik terutama mengenai metode representasi teks mana yang lebih efektif dalam meningkatkan kinerja binary classifier pada klasifikasi SMS palsu. Oleh karena itu, penting untuk mengeksplorasi dan membandingkan pengaruh teknik embedding yang berbeda, seperti Word2Vec dan TF-IDF+N-gram, untuk menentukan model terbaik yang dapat meningkatkan akurasi dan efisiensi algoritma binary classifier dalam mendeteksi SMS palsu.

Model embedding memainkan peran penting dalam mengubah data teks menjadi vektor numerik yang dapat digunakan oleh algoritma machine learning [6]. Teknik embedding yang umum digunakan meliputi Word2Vec dan TF-IDF+N-gram [7]. Word2Vec adalah teknik berbasis neural network yang menghasilkan representasi kata dalam bentuk vektor, mempelajari hubungan semantik antar kata berdasarkan konteks dalam teks melalui pendekatan seperti Continuous Bag of Words (CBOW) dan Skip-gram [8]. Keunggulan Word2Vec terletak pada kemampuannya untuk menangkap makna semantik kata-kata, sehingga sangat cocok untuk analisis teks yang membutuhkan pemahaman konteks. TF-IDF mengukur seberapa penting suatu kata dalam dokumen relatif terhadap koleksi dokumen lainnya, dan ketika dikombinasikan dengan N-gram, TF-IDF dapat menangkap pola urutan kata, seperti bigram atau trigram, yang berguna untuk menangani data teks dengan struktur dan pola tertentu, seperti SMS palsu [9]–[11]. Kombinasi kedua teknik ini, dengan Word2Vec yang unggul dalam menangkap makna semantik dan TF-IDF+N-gram yang efektif untuk mengidentifikasi pola frekuensi kata spesifik, memberikan fleksibilitas dalam meningkatkan akurasi klasifikasi SMS palsu.

Beberapa tahun terakhir penelitian tetang klasifikasi SMS palsu telah mengalami perkembangan yang cukup signifikan seperti penelitian ini mengusulkan sebuah metode yang menggunakan embeddings dan TF-IDF untuk meningkatkan akurasi penyaringan spam SMS, dan percobaan pada dataset nyata menunjukkan keefektifan pendekatan yang diusulkan. Namun, permasalahannya adalah Data yang tidak seimbang dan semantik yang tidak jelas dalam pesan teks pendek, yang dapat membuat penyaringan spam SMS menjadi sulit sehingga tantangan ini menjadi hal yang sangat penting untuk mencapai akurasi yang tinggi dalam penyaringan spam SMS, bahkan dengan berbagai metode yang ada, disisilain, Kurangnya rincian tentang dataset nyata yang digunakan dan parameter evaluasi, yang dapat membatasi kemampuan untuk menilai generalisasi terhadap pendekatan yang diusulkan [12]. Analisis komparatif dari berbagai teknik penyisipan kata untuk deteksi spam SMS, mengevaluasi kinerja mereka menggunakan lima pengklasifikasi pembelajaran mesin yang berbeda dengan melakukan perbandingan teknik word embedding, termasuk N-gram, BOW, dan TF-IDF, untuk deteksi spam SMS menggunakan pengklasifikasi pembelajaran mesin [13]. Menggunakan pendekatan pembelajaran mesin yang diawasi menggunakan Multi-Layer Perceptron (MLP) untuk mengklasifikasikan artikel berita untuk mendeteksi artikel berita palsu dan membedakannya dari artikel berita yang valid, melalui pendekatan klasifikasi teks biner yang mencapai presisi dan recall yang tinggi [14]. Penelitian ini ini mengusulkan penggunaan penyematan teks dan model TF-IDF+Ngram untuk meningkatkan kinerja algoritme pengklasifikasi biner untuk mendeteksi spam dalam pesan teks pendek (SMS). Model-model ini menganalisis teks dengan mengubah pesan teks diskrit menjadi bentuk vektor numerik yang kontinu. Sebuah vektor mewakili setiap kata dalam teks, dan nilai numerik dari dimensi sebuah kata didasarkan pada konteks kata tersebut [15]. Menggunakan beberapa model klasifikasi pembelajaran mesin untuk mendeteksi spam SMS dari dataset yang terdiri dari hampir 6.000 pesan, menggunakan kombinasi fitur TF-IDF dan Count Vectorization, dan kemudian membandingkan keakuratan model yang berbeda untuk menentukan model yang paling akurat untuk mendeteksi spam.[16]. Penelitian ini menggunakan model pembelajaran penyematan teks dan ansambel berbasis Transformer mencapai kinerja yang canggih dalam pendeteksian spam SMS yang menggunakan teknik penyematan teks berdasarkan GPT-3 Transformer dan strategi Ensemble Learning dengan Weighted Voting untuk mencapai kinerja yang canggih dengan akurasi 99,91%. Walaupun akurasi yang dihasilkan tinggi, tetapi dataset yang digunakan relatif kecil, dengan hanya 5.574 pesan berbahasa Inggris, dan perlu diperluas dengan data pelatihan yang lebih beragam dan lebih besar. Metode Ensemble Learning yang digunakan cukup kompleks, dengan empat algoritma klasifikasi yang berbeda, yang dapat mempengaruhi performa sistem dengan meningkatkan waktu komputasi dan penggunaan memori. Model ini dapat menghadapi masalah kinerja jika diperluas untuk mendukung pesan yang lebih besar, misalnya email atau digunakan pada perangkat dengan perangkat keras yang terbatas [17]. Melakukan evaluasi dan membandingkan kinerja dua pendekatan ekstraksi fitur BoW dan TF-IDF dengan N-gram dan tiga pengklasifikasi pembelajaran mesin konvensional SVM, NB, dan DT untuk tugas mendeteksi berita palsu, dan membandingkan

JURNAL SISTEM INFORMASI TGD
Volume 4 Nomor 1, Januari 2025, Hal 55-64
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

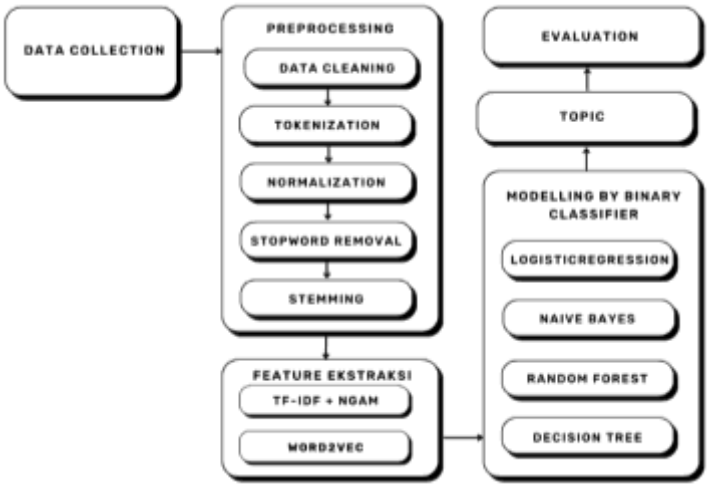
kinerjanya dengan model transformasi BERT yang telah disetel dengan baik. - Metode prapemrosesan dan ekstraksi fitur yang digunakan masih terbatas, Karena dalam mendeteksi berita palsu alami yang sulit dilakukan oleh manusia merupakan keterbatasan dari pendekatan yang yang diusulkan. Meskipun pendekatan yang diusulkan berkinerja baik, namun, masih ada ruang untuk pengembangan dan membandingkannya dengan hasil yang sudah ada dan model BERT yang digunakan [18]. Mengevaluasi dan membandingkan akurasi dari enam model pembelajaran mesin yang berbeda (Random Forest, Gradient Boosting, Extra Trees, Regresi Logistik, Support Vector Machine, dan Multinomial Naive Bayes) untuk klasifikasi spam SMS, dengan menggunakan berbagai metode pra-pemrosesan seperti lemmatisasi, stemming, dan ekstraksi TF-IDF [19]. model pembelajaran mesin menggunakan TF-IDF dan Stochastic Gradient Descent Classifier dikembangkan untuk mendeteksi spam SMS Indonesia dengan akurasi 97% [20]. Mengusulkan model klasifikasi sentimen teks menggunakan TF-IDF dan Next Word Negation (NWN), dan mereka menemukan bahwa model dengan algoritma Linear Support Vector Machine (LSVM) ini mencapai akurasi yang jauh lebih tinggi daripada metode sebelumnya [21]. Penelitian ini mengusulkan pendekatan hibrida menggunakan TF-IDF dan N-Grams untuk ekstraksi fitur guna mendeteksi berita palsu pada data Twitter secara akurat menggunakan teknik pembelajaran mesin. Namun didalam penelitian ini mengakui bahwa mendeteksi berita palsu adalah tantangan yang sedang berlangsung dengan kesulitan besar yang belum sepenuhnya terselesaikan oleh studi saat ini. Selain itu, bertujuan untuk membangun model untuk mendeteksi berita palsu, menyiratkan bahwa model tersebut belum sepenuhnya dikembangkan atau divalidasi. Sehingga perlu mengeksplorasi teknik ekstraksi fitur dan klasifikasi yang berbeda, menyiratkan bahwa pendekatan terbaik belum diidentifikasi secara pasti [21].

Penelitian ini bertujuan untuk mengeksplorasi dan membandingkan pengaruh teknik embedding, khususnya Word2Vec dan TF-IDF+N-gram, terhadap peningkatan kinerja algoritma binary classifier dalam klasifikasi SMS palsu. Dalam konteks ini, binary classifier digunakan untuk membedakan antara SMS palsu dan SMS yang sah, dengan teknik embedding berfungsi untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh model. Penelitian ini akan menganalisis bagaimana kedua teknik embedding tersebut mempengaruhi akurasi deteksi, kecepatan komputasi, dan kemampuan model dalam menangkap pola-pola penting yang membedakan SMS palsu dari pesan yang sah. Dengan membandingkan kinerja Word2Vec dan TF-IDF+N-gram, penelitian ini bertujuan untuk mengidentifikasi teknik embedding yang paling optimal dalam meningkatkan efektivitas binary classifier, sehingga dapat menghasilkan sistem deteksi SMS palsu yang lebih akurat dan efisien. Capaian akhir yang diharapkan adalah rekomendasi teknik embedding terbaik yang dapat diterapkan dalam aplikasi deteksi SMS palsu berbasis machine learning, serta kontribusi terhadap pengembangan metode deteksi yang lebih baik.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Metode penelitian yang digunakan meliputi serangkaian tahapan sistematis mulai dari pengumpulan data hingga evaluasi model untuk memastikan keakuratan dan relevansi hasil yang diperoleh. Tahapan yang diusulkan dalam penelitian ini mencakup Data Collection, Preprocessing, Feature Extraction, Modeling, dan Evaluation dalam menghasilkan kinerja model yang lebih baik. Gambar 1 Menunjukkan usulan metode penelitian.



Gambar 1. Usulan Metode Penelitian

2.2 Data Collection

Tahap pengumpulan data merupakan langkah awal dalam penelitian ini, di mana data yang digunakan memiliki karakteristik berupa teks yang relevan dengan permasalahan penelitian. Dataset yang digunakan bersifat publik dan diambil dari GitHub Repository, yang menyediakan data terbuka untuk penelitian *Sentiment Analysis* dalam bahasa

JURNAL SISTEM INFORMASI TGD
Volume 4 Nomor 1, Januari 2025, Hal 55-64
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

Indonesia. Sumber dataset ini dapat diakses melalui URL berikut: https://github.com/Andikazidanef15/Sentiment-Analysis-on-Indonesian-SMS-Dataset/blob/main/dataset_sms_spam_v1.csv, Dataset ini berisi data pesan SMS berbahasa Indonesia yang diklasifikasikan sebagai spam dan non-spam. Data tersebut dipilih karena sesuai dengan tujuan penelitian dalam membangun model klasifikasi teks berbasis machine learning menggunakan pendekatan Natural Language Processing (NLP) [2]. Selain itu, penggunaan dataset publik dari GitHub memastikan transparansi, aksesibilitas, serta potensi replikasi untuk penelitian serupa di masa mendatang.

2.3 Text Preprocessing

Tahap preprocessing bertujuan untuk mempersiapkan data teks sebelum memasuki proses pemodelan agar lebih bersih dan terstruktur. Proses ini diawali dengan data cleaning, yaitu menghapus karakter-karakter yang tidak relevan, seperti tanda baca, angka, atau simbol khusus yang tidak memiliki makna dalam analisis [22]. Selanjutnya, dilakukan tokenization untuk memecah teks menjadi kata-kata atau token yang lebih kecil [23]. Proses normalization diterapkan untuk menyamakan format kata, seperti mengubah huruf kapital menjadi huruf kecil dan menyederhanakan kata-kata tidak baku [24]. Setelah itu, stopword removal digunakan untuk menghilangkan kata-kata umum yang tidak memiliki pengaruh signifikan terhadap makna teks, seperti "yang", "dan", atau "di" [20]. Tahap terakhir adalah stemming, yang berfungsi untuk mengubah kata berimbuhan menjadi bentuk dasar, sehingga kata-kata seperti "makanlah" atau "dimakan" dikonversi menjadi "makan" [25]. Hasil dari proses preprocessing ini adalah data teks yang lebih terstruktur, konsisten, dan siap untuk tahap ekstraksi fitur.

2.4 Feature Extraction

Pada tahap Feature Extraction, teks yang telah melalui proses preprocessing diubah menjadi representasi numerik yang dapat diproses oleh algoritma machine learning. Dalam penelitian ini, dua teknik utama digunakan untuk ekstraksi fitur, yaitu TF-IDF (Term Frequency-Inverse Document Frequency) dan Word2Vec [26]. Teknik TF-IDF menghitung bobot kata berdasarkan frekuensi kemunculannya dalam suatu dokumen dibandingkan dengan kemunculannya di seluruh korpus, sehingga memberikan bobot lebih pada kata-kata yang unik dan relevan [7]. Sementara itu, Word2Vec digunakan untuk menghasilkan representasi vektor kata yang lebih mendalam, di mana kata-kata dengan konteks yang serupa akan memiliki vektor yang lebih dekat satu sama lain dalam ruang vektor. Kedua teknik ini memungkinkan model untuk menangkap pola-pola penting dalam teks dan mengonversi informasi yang bersifat kualitatif menjadi bentuk yang bisa diproses lebih lanjut oleh algoritma pembelajaran mesin [27].

2.5 Modelling by Binary Classifier

Pada tahap Modeling, sejumlah algoritma machine learning diterapkan untuk membangun model klasifikasi teks berdasarkan fitur yang telah diekstraksi. Dalam penelitian ini, empat algoritma yang berbeda digunakan, yaitu Logistic Regression [28], Naive Bayes [29], Random Forest [30], dan Decision Tree [31]. Logistic Regression diterapkan untuk menguji hubungan antara fitur yang diekstraksi dengan probabilitas suatu kelas, sedangkan Naive Bayes digunakan untuk klasifikasi berbasis probabilitas dengan asumsi independensi antar fitur. Random Forest, yang merupakan ensemble method, digunakan untuk meningkatkan akurasi dengan menggabungkan beberapa pohon keputusan, sedangkan Decision Tree dipilih karena kemampuannya dalam menghasilkan model yang interpretable berdasarkan pemisahan data menggunakan fitur terbaik. Keempat algoritma ini diuji untuk menentukan kinerja masing-masing dalam mengklasifikasikan pesan SMS sebagai spam atau non-spam, dan hasilnya akan dibandingkan berdasarkan metrik evaluasi yang relevan untuk memilih model yang paling efektif.

2.6 Evaluation

Pada tahap Evaluation, kinerja model klasifikasi diuji dan dianalisis menggunakan metrik evaluasi untuk menilai hasil klasifikasikan data teks SMS palsu. Metode yang digunakan adalah Confusion Matrix, yang memberikan gambaran visual mengenai performa model dengan membandingkan prediksi model dengan nilai sebenarnya.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Confusion Matrix menghasilkan empat komponen utama yaitu True Positives (TP), True Negatives (TN), False Positives (FP), dan False Negatives (FN), yang digunakan untuk menghitung metrik lainnya seperti *accuracy*, *precision*,

JURNAL SISTEM INFORMASI TGD
Volume 4 Nomor 1, Januari 2025, Hal 55-64
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

recall, dan F1-score. *Accuracy* mengukur seberapa banyak prediksi yang benar, sementara precision dan recall memberikan informasi lebih rinci mengenai ketepatan dan kemampuan model dalam mendeteksi kelas tertentu, seperti spam. F1-score menjadi indikator keseimbangan antara precision dan recall. Dengan menggunakan Confusion Matrix dan metrik terkait, evaluasi ini memungkinkan untuk menganalisis kekuatan dan kelemahan masing-masing model, serta memilih model yang paling optimal untuk tugas klasifikasi.

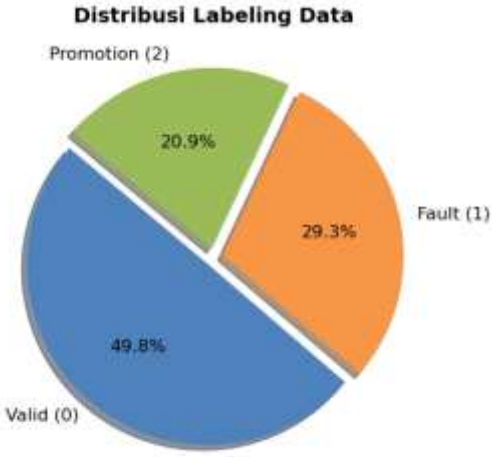
3. HASIL DAN PEMBAHASAN

Pada tahapan ini, hasil penelitian yang telah dilakukan melalui eksperimen terhadap model yang diusulkan akan dipresentasikan. Eksperimen bertujuan untuk menguji efektivitas dan kinerja model yang dikembangkan dalam mencapai tujuan yang ditetapkan, baik dari segi akurasi, precision recall dan f-score. Pada tahapan ini juga dijelaskan pembahasan mengenai temuan-temuan penting yang dihasilkan dari eksperimen ini, mengidentifikasi kekuatan dan kelemahan model yang diuji. Penelitian ini menggunakan model machine learning dengan beberapa algoritma Binary Classifier, yaitu Logistic Regression, Naive Bayes, Random Forest, dan Decision Tree. Kinerja model dievaluasi berdasarkan metrik akurasi (accuracy), presisi (precision), recall, dan F1-score.

3.1 Hasil Eksperimen

3.1.1 Distribusi Label dataset.

Penelitian ini menggunakan dataset dari 1143 pesan SMS yang diklasifikasikan ke dalam tiga label: “Valid”, “Fault” dan “promotion”. Distribusi jumlah data untuk setiap label setelah penggabungan ditunjukkan pada Gambar 2.



Gambar 2. Distribusi Label Data

Gambar 2. Menjelaskan distribusi tiga kategori data: "Valid (0)" sebesar 49.8%, menunjukkan bahwa hampir setengah dari data terverifikasi dan memenuhi kriteria; "Fault (1)" sebesar 29.3%, mencerminkan sekitar sepertiga data mengalami kesalahan atau masalah; dan "Promotion (2)" sebesar 20.9%, yang menunjukkan bahwa sekitar 1/5 dari data berkaitan dengan kegiatan promosi. Secara keseluruhan, grafik ini menunjukkan bahwa mayoritas data adalah data valid, diikuti dengan data bermasalah dan data terkait promosi.

3.1.2 Pengujian Model Binary Classifier Berdasarkan Representasi Fitur

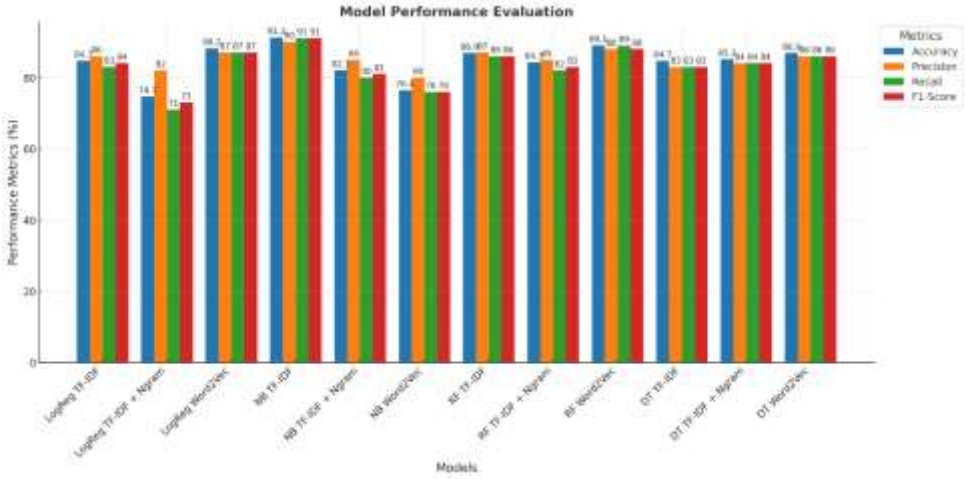
Eksperimen klasifikasi SMS Spam dilakukan untuk dengan menggunakan beberapa algoritma binary classifier, yaitu Logistic Regression, Naive Bayes (NB), Random Forest, dan Decision Tree. Tujuan utama eksperimen adalah untuk mengukur kemampuan masing-masing algoritma berdasarkan representasi fitur ekstraksi yang digunakan dalam mendeteksi kelas ambigu (label 1) dan tidak ambigu (label 0) pada dataset BBC News. Pengukuran performa dilakukan berdasarkan precision, recall, f1-score, dan confusion matrix untuk memberikan analisis menyeluruh terhadap hasil klasifikasi. Tabel 1 menunjukkan performa dari komparasi algoritma Binary Classifier.

Tabel 1. Evaluasi Performa Model dengan Confusion Matrix

Algoritma	Akurasi	Precision	Recall	F1-Score
Logistic Regression TF-IDF	84.71%	86%	83%	84%
Logistic Regression TF-IDF + Ngram	74.67%	82%	71%	73%
Logistic Regression Word2Vec	88.20%	87%	87%	87%
Naive Bayes TF-IDF	91.26%	90%	91%	91%
Naive Bayes TF-IDF + Ngram	82.09%	85%	80%	81%

JURNAL SISTEM INFORMASI TGD
Volume 4 Nomor 1, Januari 2025, Hal 55-64
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

Naive Bayes Word2Vec	76.41%	80%	76%	76%
Random Forest TF-IDF	86.89%	87%	86%	86%
Random Forest TF-IDF + Ngram	84.27%	85%	82%	83%
Random Forest Word2Vec	89.08%	88%	89%	88%
Decision Tree TF-IDF	84.71%	83%	83%	83%
Decision Tree TF-IDF + Ngram	85.15%	84%	84%	84%
Decision Tree Word2Vec	86.89%	86%	86%	86%

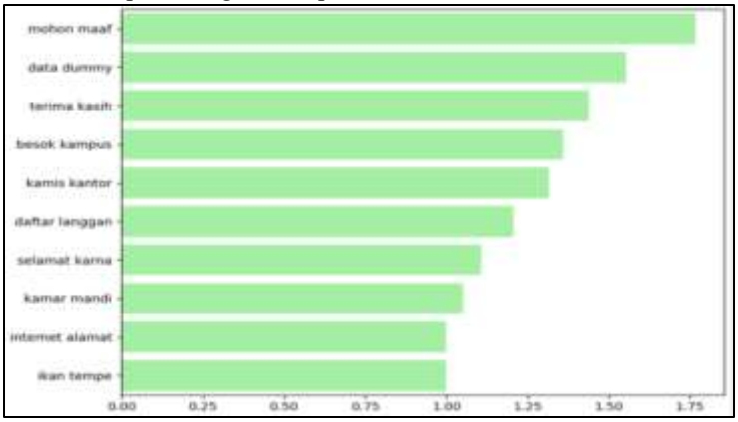


Gambar 4. Evaluasi kinerja model berdasarkan representasi fitur.

Berdasarkan Tabel 1, dan Gambar 4. Menunjukkan evaluasi performa empat algoritma (Naive Bayes, Logistic Regression, Random Forest, dan Decision Tree) dengan berbagai representasi fitur teks (TF-IDF, TF-IDF + Ngram, dan Word2Vec) berdasarkan empat metrik (Accuracy, Precision, Recall, dan F1-Score). Naive Bayes dengan TF-IDF menunjukkan performa tertinggi di semua metrik, dengan akurasi 91.26%, diikuti oleh nilai Precision, Recall, dan F1-Score yang juga tinggi. Logistic Regression memiliki perbedaan signifikan antar representasi fitur, di mana TF-IDF + Ngram menunjukkan performa terendah, sementara Word2Vec meningkatkan performa secara signifikan. Random Forest dan Decision Tree menunjukkan performa yang cukup baik, dengan Random Forest menggunakan Word2Vec mencatatkan akurasi tertinggi di 89.08%. Kesimpulannya, Naive Bayes TF-IDF unggul, namun Random Forest Word2Vec menjadi alternatif kompetitif, dan penting untuk mempertimbangkan beberapa metrik dalam mengevaluasi performa model klasifikasi teks.

3.2 Hasil Topik Label Klasifikasi

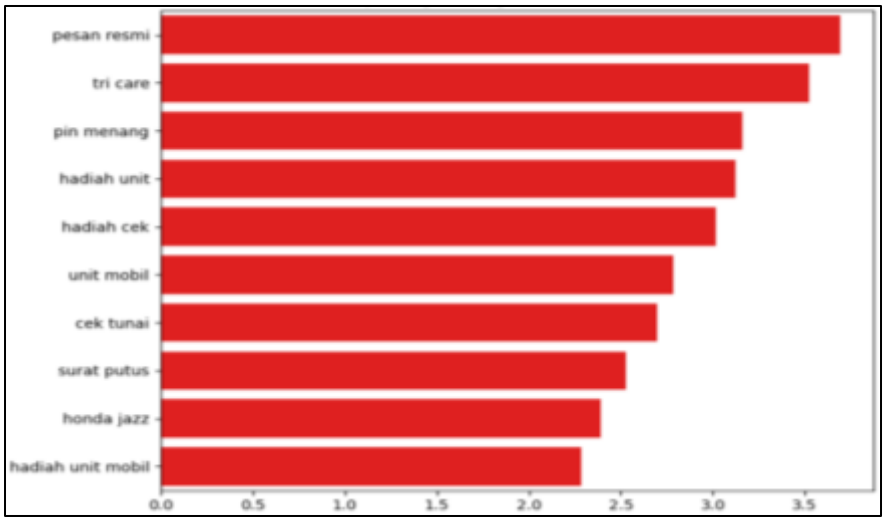
Hasil topik pada masing-masing label klasifikasi pesan SMS yaitu "Valid," "Fault," dan "Promotion," untuk memberikan pemahaman mendalam masing-masing label pesan. Topik masing-masing label di sajikan pada Gambar 5, 6 dan 7. Gambar 5, Topik Label "Valid" mencakup pesan-pesan yang bersifat relevan, informatif, dan sesuai dengan kebutuhan pengguna, seperti "mohon maaf", "besok kampus", "kamis kantor". Gambar 6, Label "Fault" mengacu pada pesan-pesan spam dan penipuan seperti "pin pemenang", "hadiah unit", "cek tunai". Kemudian Gambar 7 menunjukan label "Promotion" meliputi pesan-pesan yang berisi promosi, iklan, atau ajakan komersial terkait produk atau layanan tertentu, seperti "beli paket", "bonus pulsa", "gratis nelpon".



Gambar 5. Topik Label Valid

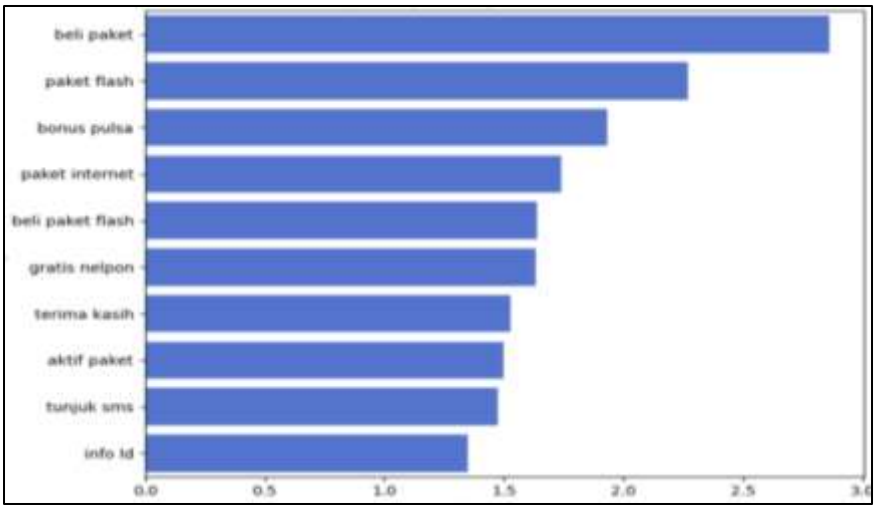
JURNAL SISTEM INFORMASI TGD
Volume 4 Nomor 1, Januari 2025, Hal 55-64
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

Gambar 5. Menjelaskan label "Valid" mencakup pesan-pesan yang relevan, informatif, dan sesuai dengan kebutuhan pengguna. Pesan-pesan ini sering kali berisi pemberitahuan penting atau informasi yang membantu dalam aktivitas sehari-hari, seperti jadwal, pengumuman, atau permintaan maaf, dengan contoh frasa seperti "mohon maaf," "besok kampus," dan "kamis kantor." Karakteristik utama dari label ini adalah keandalannya dalam menyampaikan konten yang berguna dan tidak mengganggu, sehingga memberikan nilai tambah bagi penerimanya.



Gambar 6. Topik Label Fault

Gambar 6. Menjelaskan label "Fault" mencakup pesan-pesan yang tidak valid, seperti spam atau penipuan, yang cenderung mengganggu dan tidak relevan bagi penerima. Pesan-pesan ini sering kali memiliki maksud manipulatif atau mencoba menipu pengguna dengan menawarkan hadiah palsu, informasi yang menyesatkan, atau aktivitas phishing. Contoh frasa yang sering muncul dalam kategori ini adalah "pin pemenang," "hadiah unit," dan "cek tunai." Karakteristik utama dari label ini adalah sifatnya yang merugikan atau mengganggu, sehingga penting untuk dikenali dan dihindari.



Gambar 7. Topik Label Promosi

Gambar 7. Menjelaskan label "Promotion" mencakup pesan-pesan berisi promosi, iklan, atau ajakan komersial yang terkait dengan produk atau layanan tertentu. Pesan-pesan ini biasanya menawarkan penawaran khusus, diskon, atau manfaat tambahan yang bertujuan menarik perhatian penerima untuk membeli atau menggunakan layanan. Contoh frasa yang sering ditemukan dalam kategori ini meliputi "beli paket," "bonus pulsa," dan "gratis nelpon." Karakteristik utama dari label ini adalah kontennya yang bersifat promosi dan cenderung berorientasi pada pemasaran.

3.3 Analisis *dan Pembahasan*

Penelitian ini mengeksplorasi pengaruh representasi fitur teks, seperti TF-IDF, TF-IDF + Ngram, dan Text Embedding (Word2Vec), terhadap performa beberapa algoritma binary classifier dalam klasifikasi SMS palsu dan SMS valid. Berdasarkan hasil eksperimen, terlihat bahwa setiap kombinasi algoritma dan representasi fitur memiliki karakteristik performa yang berbeda, yang dapat dianalisis sebagai berikut:

3.3.1 Kinerja Naive Bayes dengan TF-IDF
Naive Bayes menunjukkan kinerja terbaik dengan representasi TF-IDF, menghasilkan akurasi 91.26% serta nilai precision, recall, dan F1-score yang tinggi. Hal ini dapat dijelaskan karena Naive Bayes bekerja optimal dengan fitur berbasis frekuensi seperti TF-IDF, yang mampu memodelkan distribusi probabilitas kata secara sederhana namun efektif. Representasi TF-IDF juga berhasil menangkap informasi penting dari data teks, terutama dalam membedakan SMS palsu dan valid.

3.3.2 Kinerja Text Embedding (Word2Vec)
Word2Vec, yang mampu menangkap konteks semantik kata, memberikan hasil yang kompetitif pada algoritma seperti Random Forest dan Bagging Classifier, dengan akurasi mencapai 89.08% dan F1-score 88%. Meskipun tidak melampaui performa TF-IDF pada Naive Bayes, embedding ini mampu meningkatkan performa algoritma berbasis ensemble dengan menyediakan representasi vektor yang lebih kaya akan makna dan hubungan antar kata. Hal ini menunjukkan potensi Word2Vec dalam menangani dataset yang memerlukan analisis semantik mendalam.

3.3.3 Performa TF-IDF + Ngram
Kombinasi TF-IDF dengan Ngram menghasilkan performa yang beragam di antara algoritma. Sementara Decision Tree menunjukkan hasil stabil, Logistic Regression memiliki performa terendah dengan pendekatan ini. Hal ini menunjukkan bahwa meskipun Ngram dapat menangkap pola kata yang lebih panjang, representasi ini mungkin kurang cocok untuk algoritma yang sensitif terhadap dimensi data tinggi, seperti Logistic Regression.

3.3.4 Evaluasi Berbasis Metrik
Analisis berbasis metrik precision, recall, dan F1-score mengungkapkan bahwa penting untuk tidak hanya mengandalkan akurasi dalam mengevaluasi performa model. Naive Bayes dengan TF-IDF menunjukkan keseimbangan metrik yang baik, mengindikasikan kemampuan yang konsisten dalam mendeteksi kedua kelas (SMS palsu dan valid). Sebaliknya, beberapa algoritma dengan kombinasi fitur tertentu menunjukkan ketidakseimbangan performa pada kelas tertentu, yang dapat menjadi titik perhatian untuk perbaikan lebih lanjut.

3.3.5 Implikasi dan Potensi Pengembangan
Dari hasil ini, dapat disimpulkan bahwa pemilihan representasi fitur teks harus disesuaikan dengan karakteristik algoritma yang digunakan. TF-IDF cocok untuk algoritma berbasis probabilitas sederhana seperti Naive Bayes, sementara Word2Vec menawarkan potensi besar untuk algoritma yang membutuhkan pemahaman semantik, seperti Random Forest. Selain itu, pendekatan berbasis ensemble dapat lebih dieksplorasi untuk mengoptimalkan hasil klasifikasi dengan Text Embedding.

Penelitian ini memberikan informasi pentingnya pemahaman terhadap kekuatan dan kelemahan masing-masing kombinasi algoritma dan fitur. Penggunaan teknik representasi yang tepat tidak hanya meningkatkan performa model tetapi juga memberikan interpretasi yang lebih baik terhadap data teks yang diklasifikasikan.

4 KESIMPULAN

Penelitian ini menunjukkan bahwa pemilihan representasi fitur teks, seperti TF-IDF, TF-IDF + Ngram, dan Text Embedding (Word2Vec), memiliki pengaruh signifikan terhadap kinerja algoritma binary classifier dalam klasifikasi SMS palsu. Berdasarkan hasil eksperimen, Naive Bayes dengan TF-IDF memberikan performa terbaik di semua metrik evaluasi, dengan akurasi 91.26%, precision 90%, recall 91%, dan F1-score 91%. Hal ini menegaskan bahwa representasi TF-IDF sangat efektif dalam menangkap informasi tekstual yang relevan untuk klasifikasi SMS. Namun, penggunaan Text Embedding (Word2Vec) juga memberikan hasil yang kompetitif, terutama ketika digunakan bersama dengan algoritma Random Forest, yang mencatatkan akurasi 89.08% dan F1-score 88%. Representasi Word2Vec mampu menangkap konteks semantik dari teks, sehingga memperkaya informasi yang digunakan untuk klasifikasi, terutama pada algoritma berbasis ensemble seperti Random Forest. Di sisi lain, Logistic Regression menunjukkan performa terendah saat menggunakan TF-IDF + Ngram, menunjukkan bahwa pendekatan ini kurang optimal untuk algoritma tersebut. Kesimpulannya, kombinasi Text Embedding dengan algoritma tertentu, seperti Random Forest, dapat menjadi alternatif yang kuat untuk menangani klasifikasi teks berbasis konteks, sementara TF-IDF tetap menjadi pilihan unggulan untuk algoritma sederhana seperti Naive Bayes. Penelitian ini menegaskan pentingnya pemilihan representasi teks yang sesuai untuk meningkatkan performa model klasifikasi yang lebih tinggi.

DAFTAR PUSTAKA

[1] O. Abayomi-Alli, S. Misra, A. Abayomi-Alli, and M. Odusami, "A review of soft techniques for SMS spam classification: Methods, approaches and applications," *Eng. Appl. Artif. Intell.*, vol. 86, pp. 197–212, 2019, doi: 10.1016/j.engappai.2019.08.024.

[2] A. Theodorus, T. K. Prasetyo, R. Hartono, and D. Suhartono, "Short Message Service (SMS) Spam Filtering using Machine Learning in Bahasa Indonesia," in *2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, 2021, pp. 199–203. doi: 10.1109/EIConCIT50028.2021.9431859.

[3] N. Aulia, "Hate Speech Detection on Indonesian Long Text Documents Using Machine Learning Approach," pp. 164–169, 2019.

[4] Y. Vernanda, S. Hansun, and M. B. Kristanda, "Indonesian language email spam detection using N-gram and Naïve Bayes algorithm," vol. 9, no. 5, pp. 2012–2019, 2020, doi: 10.11591/eei.v9i5.2444.

[5] N. K. Nagwani and A. Sharaff, "SMS spam filtering and thread identification using bi-level text classification and clustering techniques," *J. Inf. Sci.*, vol. 43, no. 1, pp. 75–87, 2017, doi: 10.1177/0165551515616310.

[6] D. Kim, D. Seo, S. Cho, and P. Kang, "Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec," *Inf. Sci. (Ny).*, vol. 477, pp. 15–29, 2019, doi: 10.1016/j.ins.2018.10.006.

[7] Z. Alamin, T. A. Lorosae, and S. Ramadhan, "Improving Performance Sentiment Movie Review Classification Using Hybrid Feature TFIDF , N-Gram , Information Gain and Support Vector Machine," vol. 11, no. 2, pp. 375–384, 2024.

[8] S. Yilmaz and S. Toklu, "A deep learning analysis on question classification task using Word2vec representations," *Neural Comput. Appl.*, vol. 32, no. 7, pp. 2909–2928, 2020, doi: 10.1007/s00521-020-04725-w.

[9] X. Bao, S. Lin, R. Zhang, Z. Yu, and N. Zhang, "Sentiment Analysis of Movie Reviews Based on Improved Word2vec and Ensemble Learning," *J. Phys. Conf. Ser.*, vol. 1693, no. 1, 2020, doi: 10.1088/1742-6596/1693/1/012088.

[10] S. Yilmaz and S. Toklu, "A deep learning analysis on question classification task using Word2vec representations," *Neural Comput. Appl.*, vol. 32, no. 7, pp. 2909–2928, 2020, doi: 10.1007/s00521-020-04725-w.

[11] S. Abdulateef, N. A. Khan, B. Chen, and X. Shang, "Multidocument Arabic text summarization based on clustering and word2vec to reduce redundancy," *Inf.*, vol. 11, no. 2, 2020, doi: 10.3390/info11020059.

[12] S. Hosseinpour and M. R. Keyvanpour, "A Comprehensive Approach to SMS Spam Filtering Integrating Embedded and Statistical Features," in *2023 13th International Conference on Computer and Knowledge Engineering (ICCKE)*, 2023, pp. 7–12. doi: 10.1109/ICCKE60553.2023.10326281.

[13] P. Joseph and S. Y. Yerima, "A comparative study of word embedding techniques for SMS spam detection," in *2022 14th International Conference on Computational Intelligence and Communication Networks (CICN)*, 2022, pp. 149–155. doi: 10.1109/CICN56167.2022.10008245.

[14] A. Rusli, J. C. Young, and N. M. S. Iswari, "Identifying Fake News in Indonesian via Supervised Binary Text Classification," in *2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, 2020, pp. 86–90. doi: 10.1109/IAICT50021.2020.9172020.

[15] M. . Abbashi, A. . Beltyukov, H. Lal, and A. . Abbasi, "Spam Detection in Short Text Messages (Sms) Using Word Embedding and Term Frequency - Inverse Document Frequency (Tf-Idf)," *XXI Century Resumes Past Challenges Present plus*, vol. 9, no. 50, pp. 143–148, 2020, doi: 10.46548/21vek-2020-0950-0026.

[16] T. Singh, T. A. Kumar, and P. G. Shambharkar, "Enhancing Spam Detection on SMS performance using several Machine Learning Classification Models," in *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, 2022, pp. 1472–1478. doi: 10.1109/ICOEI53556.2022.9777157.

[17] A. Ghourabi and M. Alohaly, "Enhancing Spam Message Classification and Detection Using Transformer-Based Embedding and Ensemble Learning," *Sensors*, vol. 23, no. 8, pp. 1–17, 2023, doi: 10.3390/s23083861.

[18] A. E. Qasem and M. Sajid, "Exploring the Effect of N-grams with BOW and TF-IDF Representations on Detecting Fake News," in *2022 International Conference on Data Analytics for Business and Industry (ICDABI)*, 2022, pp. 741–746. doi: 10.1109/ICDABI56818.2022.10041537.

[19] N. Sharma, "A Methodological Study of SMS Spam Classification Using Machine Learning Algorithms," in *2022 2nd International Conference on Intelligent Technologies (CONIT)*, 2022, pp. 1–5. doi: 10.1109/CONIT55038.2022.9848171.

[20] A. A. Ramaditia Dwiyansaputra, Gibran Satya Nugraha, Fitri Bimantoro, "Indonesian SMS Spam Detection using TF-IDF and Stochastic Gradient Descent," vol. 3, no. 2, pp. 200–207, 2021.

[21] B. Das and S. Chakraborty, "An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation," 2018, [Online]. Available: http://arxiv.org/abs/1806.06407

[22] S. Alam and N. Yao, "The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis," *Comput. Math. Organ. Theory*, vol. 25, no. 3, pp. 319–335, 2019, doi: 10.1007/s10588-018-9266-8.

[23] A. I. Kadhim, "Survey on supervised machine learning techniques for automatic text classification," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 273–292, 2019, doi: 10.1007/s10462-018-09677-1.

[24] I. M. Mubaroq and E. B. Setiawan, "The Effect of Information Gain Feature Selection for Hoax Identification in Twitter Using Classification Method Support Vector Machine," *Indones. J. ...*, vol. 5, no. September, pp. 107–118, 2020, doi: 10.21108/indojc.2020.5.2.499.

[25] J. Asian, H. E. Williams, and S. M. M. Tahaghoghi, "Stemming Indonesian," *Conf. Res. Pract. Inf. Technol. Ser.*, vol. 38, pp. 307–314, 2005.

[26] T. Shaik *et al.*, "A Review of the Trends and Challenges in Adopting Natural Language Processing Methods for Education Feedback Analysis," *IEEE Access*, vol. 10, pp. 56720–56739, 2022, doi: 10.1109/ACCESS.2022.3177752.

[27] H. H. Saeed, K. Shahzad, and F. Kamiran, "Overlapping toxic sentiment classification using deep neural architectures," *IEEE Int. Conf. Data Min. Work. ICDMW*, vol. 2018-Novem, pp. 1361–1366, 2019, doi: 10.1109/ICDMW.2018.00193.

[28] I. Express and L. Part, "Short message service (sms) spam filtering using deep learning in bahasa indonesia," vol. 13, no. 10, pp. 1093–1100, 2022, doi: 10.24507/icicelb.13.10.1093.

[29] P. N. Andono and R. A. Pramunendar, "Performance Evaluation of Classification Algorithm for Movie Review Sentiment Analysis," *Int. J. Comput.*, vol. 22, no. 1, pp. 7–14, 2023, doi: 10.47839/ijc.22.1.2873.

[30] M. A. Fauzi, "Random forest approach fo sentiment analysis in Indonesian language," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 12, no. 1, pp. 46–50, 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.

[31] A. P. Widyassari, E. Noersasongko, A. Syukur, and Affandy, "An Extractive Text Summarization based on Candidate Summary Sentences using Fuzzy-Decision Tree," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 7, pp. 572–579, 2022, doi: 10.14569/IJACSA.2022.0130768.