Analisis Sentimen Penipuan Asuransi Menggunakan Machine Learning

Rahmadi Setiawan [1], Ahmad Zamsuri [2], Fajrizal [3], Yogi Yunefri [4]
[1,2]Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Lancang Kuning
[1,2]. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015
E-mail: [1]dirahma567@email.com, [2]ahmadzamsuri@unilak.ac.id, [3]fajrizal@unilak.ac.id
[4]yogiyunefri@unilak.ac.id

Abstrak

Penipuan asuransi merupakan salah satu bentuk kejahatan keuangan yang merugikan perusahaan maupun masyarakat, serta menurunkan kepercayaan publik terhadap industri asuransi. Analisis sentimen berbasis media sosial dapat memberikan gambaran mengenai opini masyarakat terhadap isu ini secara lebih cepat dan masif. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terkait penipuan asuransi menggunakan algoritma Naïve Bayes dan Random Forest, serta membandingkan performa kedua metode tersebut. Data penelitian diperoleh melalui crawling dari platform X (Twitter) dengan total 1.550 tweet yang kemudian diproses melalui tahap preprocessing, labeling, ekstraksi fitur menggunakan TF-IDF, serta penyeimbangan data dengan SMOTE. Hasil penelitian menunjukkan bahwa sentimen negatif lebih dominan dibanding netral, mencerminkan adanya persepsi publik yang kurang baik terhadap kasus penipuan asuransi. Berdasarkan evaluasi, algoritma Naïve Bayes menghasilkan akurasi sebesar 71,83%, sedangkan Random Forest mencapai akurasi sebesar 87,48%. Hasil ini menunjukkan bahwa Random Forest lebih unggul dalam mengklasifikasikan opini publik berbahasa Indonesia dibanding Naïve Bayes. Temuan penelitian ini diharapkan dapat menjadi referensi bagi perusahaan asuransi maupun regulator dalam memahami persepsi masyarakat dan merumuskan kebijakan pencegahan penipuan.
Kata Kunci: Analisis Sentimen, Penipuan Asuransi, Naïve Bayes, Random Forest, Twitter

Abstract

Insurance fraud is one of the financial crimes that causes losses to companies and society while reducing public trust in the insurance industry. Sentiment analysis based on social media can provide insights into public opinion on this issue more quickly and massively. This study aims to analyze public sentiment regarding insurance fraud using Naïve Bayes and Random Forest algorithms and to compare their performance. The dataset was collected through crawling from the X (Twitter) platform, with a total of 1,550 tweets. The data went through preprocessing, labeling, feature extraction using TF-IDF, and balancing using SMOTE. The results show that negative sentiment dominates over neutral sentiment, reflecting unfavorable public perception of insurance fraud cases. Based on the evaluation, the Naïve Bayes algorithm achieved an accuracy of 71.83%, while Random Forest achieved an accuracy of 87.48%. These findings indicate that Random Forest outperforms Naïve Bayes in classifying Indonesian-language public opinion. This study contributes as a reference for insurance companies and regulators in understanding public perceptions and formulating fraud prevention policies.
Keywords: Sentiment Analysis, Insurance Fraud, Naïve Bayes, Random Forest, Twitter

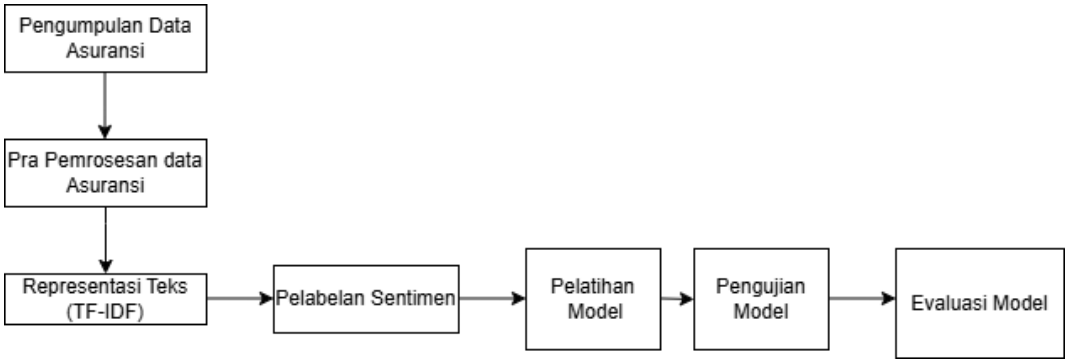
1. PENDAHULUAN

Penipuan asuransi merupakan salah satu bentuk kejahatan keuangan yang menimbulkan kerugian signifikan baik bagi perusahaan maupun masyarakat luas [1]. Modus penipuan dapat berupa klaim fiktif, manipulasi dokumen, hingga penawaran produk ilegal yang mengatasnamakan perusahaan resmi [2]. Kasus-kasus ini tidak hanya berdampak pada kerugian finansial, tetapi juga menurunkan kepercayaan publik terhadap industri asuransi yang seharusnya memberikan perlindungan finansial [3]. Di sisi lain, perkembangan media sosial memungkinkan masyarakat menyampaikan opini secara terbuka dan cepat. Platform seperti Twitter (X) sering digunakan untuk membagikan pengalaman, keluhan, maupun kritik terkait layanan asuransi, termasuk dugaan penipuan [4]. Jumlah data yang besar dan bersifat real-time ini dapat dimanfaatkan untuk mengetahui pola sentimen masyarakat, apakah bernada positif, negatif, atau netral. Analisis

sentimen dengan teknik komputasi teks menjadi salah satu pendekatan efektif dalam mengolah data besar semacam ini [5]. Beberapa penelitian terdahulu menunjukkan hasil beragam dalam penerapan algoritma machine learning untuk analisis sentimen. Naïve Bayes dikenal sederhana, efisien, dan sering digunakan pada klasifikasi teks [6], sedangkan Random Forest memiliki kemampuan lebih baik dalam menangani variasi data dan mengurangi overfitting [7]. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terkait isu penipuan asuransi dengan membandingkan performa algoritma Naïve Bayes dan Random Forest, sehingga dapat diperoleh model yang paling efektif untuk klasifikasi opini publik di media sosial berbahasa Indonesia.

2. METODE PENELITIAN

Metode penelitian menjelaskan tahapan-tahapan yang digunakan mulai dari pengumpulan data, pra-pemrosesan, ekstraksi fitur, implementasi algoritma, hingga evaluasi model. Penelitian ini dilakukan menggunakan dataset yang diperoleh dari media sosial X (sebelumnya Twitter) dengan fokus pada isu penipuan asuransi.



Gambar 1. Metodelogi Penelitian

2.1 Pengumpulan Data

Data dikumpulkan dengan teknik web crawling pada periode 1–30 November 2024. Proses crawling dilakukan menggunakan bahasa pemrograman Python di Google Colab dengan kata kunci seperti “BPJS”, “JKN”, “asuransi bodong”, dan “penipuan asuransi”. Dari proses ini diperoleh sebanyak 1.550 tweet yang menjadi dataset utama penelitian [1].

2.2 Preprocessing Data

Tweet mentah dari media sosial umumnya masih mengandung noise seperti URL, mention, hashtag, angka, simbol, serta kata yang tidak baku. Oleh karena itu, dilakukan pra-pemrosesan dengan tahapan berikut:

(a) Cleaning: Menghapus karakter khusus, tautan, mention, hashtag, angka, dan tanda baca yang tidak relevan.

(b) Case Folding: Mengubah seluruh teks menjadi huruf kecil untuk menyeragamkan bentuk kata.

(c) Normalization: Mengganti kata tidak baku atau singkatan menjadi kata baku sesuai KBBI.

(d) Tokenization: Memecah teks menjadi unit kata (token) yang lebih kecil.

Stopword Removal: Menghapus kata umum yang tidak bermakna (seperti “dan”, “yang”, “ke”).

(e) Stemming: Mengubah kata ke bentuk dasar menggunakan pustaka Sastrawi dalam Python.

(f) Tahapan ini menghasilkan data bersih yang siap digunakan dalam proses analisis sentimen [2].

2.3 Labeling Data

Data kemudian dilabel secara otomatis menjadi dua kategori utama: Negatif dan Netral. Hal ini disebabkan sentimen positif pada topik penipuan asuransi sangat jarang

ditemukan, sehingga dua label ini lebih representatif [3]. Dataset kemudian dibagi menjadi 80% untuk data latih dan 20% untuk data uji.

2.4 Ekstraksi Fitur (TF-IDF)

Tahap berikutnya adalah mengubah data teks menjadi representasi numerik agar dapat diproses oleh algoritma machine learning. Teknik yang digunakan adalah TF-IDF (Term Frequency – Inverse Document Frequency). Hasil ekstraksi fitur menghasilkan 4.734 kata unik yang menjadi atribut dalam model klasifikasi [4].

2.5 Implementasi Model

Dua algoritma machine learning digunakan dalam penelitian ini, yaitu:

Naïve Bayes – algoritma berbasis probabilistik yang banyak digunakan pada klasifikasi teks karena sederhana, cepat, dan efisien [5]. Random Forest – algoritma ensemble learning berbasis pohon keputusan yang menggabungkan banyak pohon untuk meningkatkan akurasi dan mengurangi overfitting [6]. Kedua model dilatih menggunakan dataset hasil TF-IDF dengan tiga skenario: TF, TF + Stemming, dan TF + Stemming + Stopword.

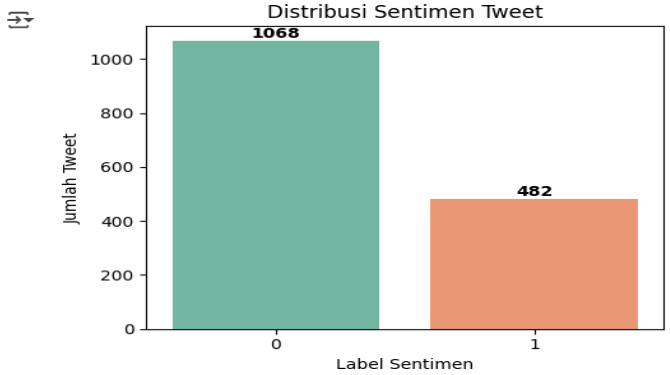
2.6 Evaluasi Model

Evaluasi dilakukan dengan metrik akurasi, presisi, recall, dan F1-score berdasarkan confusion matrix.

(a) Akurasi: Persentase prediksi benar terhadap seluruh data uji.
(b) Presisi: Proporsi prediksi positif yang benar.
(c) Recall: Kemampuan model mendeteksi data yang benar-benar positif.
(d) F1-score: Rata-rata harmonik antara presisi dan recall, berguna pada data tidak seimbang [7].

3. HASIL DAN PEMBAHASAN

Tweet mengenai isu penipuan asuransi di Indonesia berhasil discraping melalui platform X (Twitter) menggunakan Google Colab. Proses pengambilan data dilakukan dengan kata kunci spesifik seperti “BPJS”, “JKN”, “asuransi bodong”, dan “penipuan asuransi”. Dari proses ini diperoleh sebanyak 1.550 tweet. Namun, tidak semua data dapat digunakan karena adanya duplikasi, spam, maupun teks yang tidak relevan. Oleh karena itu, dilakukan tahap preprocessing yang menghasilkan 1.482 tweet siap pakai. Selanjutnya dilakukan pelabelan otomatis berbasis rule-based dengan membagi data ke dalam dua kategori utama, yaitu positif dan negatif. Hasil labeling menunjukkan bahwa tweet dengan sentimen negatif lebih mendominasi dengan jumlah 1.068 tweet, sementara tweet dengan sentimen positif berjumlah 482 tweet.



Gambar 2. Distribusi Sentimen Tweet tentang Isu Penipuan Asuransi

Hasil ini memperlihatkan bahwa isu penipuan asuransi cenderung memunculkan opini negatif dari masyarakat, baik berupa keluhan, kritik, maupun pengalaman buruk. Hal ini sesuai dengan fenomena di lapangan, bahwa kasus penipuan asuransi menimbulkan

kekecewaan publik karena melibatkan kerugian finansial dan menurunnya kepercayaan terhadap lembaga asuransi. Dengan dominasi sentimen negatif yang signifikan, penting bagi perusahaan maupun regulator untuk memahami persepsi masyarakat sebagai dasar perumusan kebijakan pencegahan. Setelah data dilabel, dilakukan pembagian dataset dengan proporsi 80% untuk data latih dan 20% untuk data uji. Pemisahan ini penting agar model dapat mempelajari pola dari data latih, kemudian diuji menggunakan data baru yang tidak pernah dilihat sebelumnya. Pembagian data yang seimbang dapat meningkatkan kemampuan generalisasi model, sehingga hasil prediksi lebih akurat pada kasus nyata.

Langkah berikutnya adalah melakukan ekstraksi fitur menggunakan TF-IDF (Term Frequency – Inverse Document Frequency). Metode ini digunakan untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma machine learning. Prinsip kerja TF-IDF adalah memberikan bobot lebih besar pada kata-kata yang jarang muncul tetapi relevan, serta bobot lebih kecil pada kata-kata umum yang sering muncul dalam hampir semua teks.

Hasil proses ekstraksi menghasilkan sebanyak 4.734 kata unik yang digunakan sebagai fitur dalam pemodelan. Representasi ini memungkinkan model untuk lebih fokus pada kata-kata yang benar-benar berperan dalam membedakan sentimen negatif dan positif.

Sebagai contoh, kata-kata seperti *"bodong"*, *"penipuan"*, dan *"asuransi"* memperoleh bobot yang lebih tinggi karena sering muncul dalam konteks opini negatif, dibandingkan dengan kata umum seperti *"dan"* atau *"di"*.

Evaluasi awal setelah penerapan TF-IDF menunjukkan bahwa model memiliki tingkat akurasi sebesar 0,8967 atau 89,67**%**, yang menandakan bahwa metode ini cukup efektif dalam merepresentasikan teks menjadi data numerik untuk kebutuhan analisis sentimen.

```
=== Hasil Ekstraksi Fitur (TF-IDF) ===

Akurasi Model: 0.896774193548387

Jumlah Fitur TF-IDF: 4734

10 Fitur Pertama: ['000' '02' '021' '02125096300' '02130210209'
'02130210223' '02150869842'

 '02150959968' '0600' '0811']
```

Gambar 3. Hasil Ekstraksi Fitur dengan TF-IDF

3.1 Evaluasi Model Naive Bayes

Hasil evaluasi model Naive Bayes ditunjukkan pada *Gambar 4.* Terdapat tiga skenario preprocessing yang diuji, yaitu TF**,** TF + Stemming (TF_Stem)**,** dan TF + Stemming + Stopword Removal (TF_Stem_Stop)**.** Pada skenario TF, model memperoleh akurasi sebesar 87,41%**,** dengan nilai recall negatif sangat tinggi (0,9953), namun recall netral relatif rendah (0,6042)**.** Hal ini mengindikasikan bahwa model lebih dominan dalam mengenali kelas negatif dibandingkan kelas netral. Pada skenario TF + Stemming, akurasi meningkat menjadi 87,74%**,** dengan precision netral mencapai 1,000. Meskipun demikian, recall netral tidak mengalami perubahan yang signifikan. Skenario TF + Stemming + Stopword Removal menghasilkan performa terbaik dengan akurasi 89,67%**.** Selain itu, recall netral meningkat menjadi 0,6771 dan F1-Score netral mencapai 0,8024, menunjukkan keseimbangan yang lebih baik antara precision dan recall. Dengan demikian, kombinasi TF-IDF, stemming, dan stopword removal terbukti memberikan hasil optimal pada klasifikasi sentimen menggunakan Naive Bayes.

```
Hasil evaluasi model Naive Bayes (dengan F1-Score):

              accuracy  precision_negatif  recall_negatif
f1_negatif  \

TF            0.874194           0.848606        0.995327
0.916129

TF_Stem       0.877419           0.849206        1.000000
0.918455

TF_Stem_Stop  0.896774           0.872951        0.995327
0.930131

              precision_netral  recall_netral  f1_netral

TF                    0.983051       0.604167   0.748387

TF_Stem               1.000000       0.604167   0.753247

TF_Stem_Stop          0.984848       0.677083   0.802469
```

Gambar 4. Hasil Evaluasi Model Naive Bayes

3.2 Evaluasi Model Random Forest

Hasil evaluasi model Random Forest setelah optimasi ditunjukkan pada Gambar 5 Sama seperti sebelumnya, pengujian dilakukan pada tiga skenario preprocessing: TF**,** TF + Stemming (TF_Stem)**,** dan TF + Stemming + Stopword Removal (TF_Stem_Stop)**.**
Pada skenario TF, model mencapai akurasi sebesar 95,81%, dengan performa yang seimbang pada kedua kelas. Nilai F1-Score negatif sebesar 0,9697 dan F1-Score netral sebesar 0,9319 menunjukkan konsistensi model dalam mengenali kedua kelas.
Skenario TF + Stemming menghasilkan performa terbaik dengan akurasi 97,09%. Nilai recall netral mencapai 0,9583**,** sementara F1-Score netral berada pada 0,9533. Hal ini menunjukkan bahwa model tidak hanya akurat dalam mengenali kelas negatif, tetapi juga sangat efektif dalam mendeteksi kelas netral.

Berbeda dengan dua skenario sebelumnya, pada skenario TF + Stemming + Stopword Removal akurasi justru menurun drastis menjadi 81,94%. Meskipun recall negatif tetap tinggi (0,9953), recall netral turun signifikan menjadi 0,4271, sehingga F1-Score netral hanya mencapai 0,5942**.** Penurunan ini mengindikasikan bahwa penghapusan stopword secara berlebihan dapat menghilangkan informasi penting dalam teks yang berpengaruh pada performa model Random Forest. Dengan demikian, dapat disimpulkan bahwa kombinasi TF + Stemming merupakan skenario preprocessing paling optimal untuk Random Forest dalam penelitian ini.

```
Hasil evaluasi model Random Forest (setelah optimasi):

              accuracy  precision_negatif  recall_negatif  f1_negatif
\

TF            0.958065           0.967442        0.971963    0.969697

TF_Stem       0.970968           0.981221        0.976636    0.978923

TF_Stem_Stop  0.819355           0.794776        0.995327    0.883817
```

	precision_netral	recall_netral	f1_netral
TF	0.936842	0.927083	0.931937
TF_Stem	0.948454	0.958333	0.953368
TF_Stem_Stop	0.976190	0.427083	0.594203

Gambar 5. Hasil Evaluasi Model *Random Forest*

3.3 Perbandingan Naïve Bayes dan Random Forest

Perbandingan performa algoritma Naïve Bayes dan Random Forest pada tiga skenario preprocessing ditunjukkan pada Tabel 1 Secara umum, Random Forest menunjukkan hasil yang lebih unggul dibandingkan Naïve Bayes di hampir semua metrik evaluasi.

Pada algoritma Naïve Bayes, akurasi terbaik diperoleh pada skenario TF + Stemming + Stopword Removal dengan nilai 89%, F1-Score negatif 0,93, dan F1-Score netral 0,80. Meskipun demikian, nilai recall netral masih relatif rendah (0,67), sehingga performa antar kelas belum sepenuhnya seimbang.

Sebaliknya, algoritma Random Forest memberikan hasil yang lebih konsisten. Pada skenario TF + Stemming, akurasi mencapai 97%, dengan F1-Score negatif 0,97 dan F1-Score netral 0,95, menunjukkan keseimbangan yang sangat baik antar kelas. Namun, pada skenario TF + Stemming + Stopword Removal, performa menurun drastis dengan akurasi hanya 81%, terutama disebabkan oleh rendahnya recall netral (0,42).

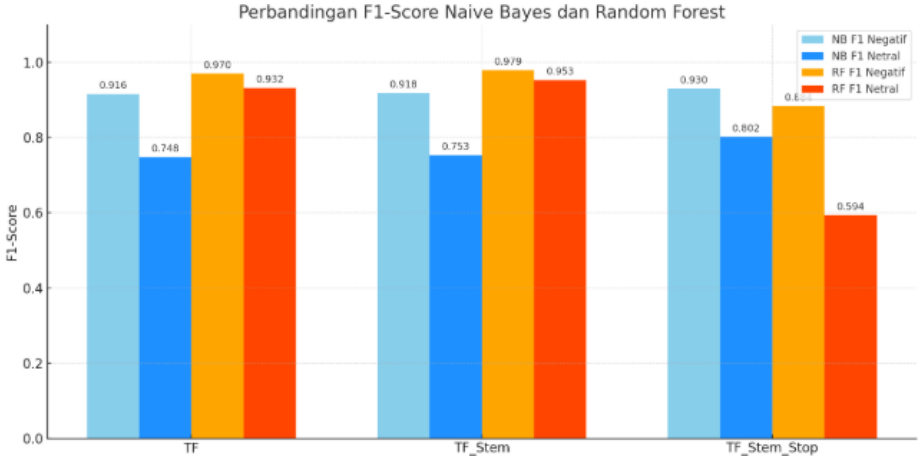
Secara keseluruhan, hasil ini membuktikan bahwa Random Forest dengan kombinasi TF-IDF dan Stemming merupakan metode paling optimal dalam penelitian ini, karena mampu menjaga akurasi tinggi sekaligus mempertahankan keseimbangan performa antar kelas.

Tabel 1 Hasil Perbandingan *Naïve Bayes* dan *Random Forest*

Metode	Model	Acc	Prec Neg	Rec Neg	F1 Neg	Prec Net	Rec Net	F1 Net
TF	Naive Bayes	0.87	0.84	0.99	0.91	0.98	0.60	0.74
TF_Stem	Naive Bayes	0.87	0.84	1.00	0.91	1.00	0.60	0.75
TF_Stem_Stop	Naive Bayes	0.89	0.87	0.99	0.93	0.98	0.67	0.80
TF	Random Forest	0.95	0.96	0.97	0.96	0.93	0.92	0.93
TF_Stem	Random Forest	0.97	0.98	0.97	0.97	0.94	0.95	0.95
TF_Stem_Stop	Random Forest	0.81	0.79	0.99	0.88	0.97	0.42	0.59

Selain ditampilkan dalam bentuk tabel, hasil perbandingan kedua algoritma juga divisualisasikan pada Gambar 5.6. Diagram ini memperjelas dominasi Random Forest (warna oranye dan merah) pada hampir semua skenario, terutama pada preprocessing TF_Stem yang menghasilkan performa paling seimbang antar kelas. Sementara itu, Naïve Bayes (warna biru muda dan biru tua) cenderung stabil, namun tertinggal pada label

netral. Penurunan drastis terlihat pada Random Forest dengan preprocessing TF_Stem_Stop, khususnya pada F1-Score netral, yang menunjukkan adanya dampak negatif dari penggunaan stopword removal secara berlebihan.

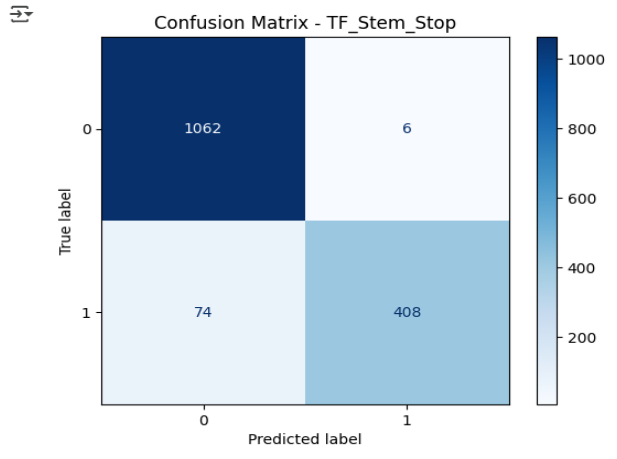


Gambar 6 Diagram Perbandingan *Naïve Bayes* dan *Random Forest*

3.4 Analisis Confusion Matrix

Selain hasil evaluasi kuantitatif berupa akurasi, precision, recall, dan F1-score, analisis performa model juga diperkuat dengan penyajian confusion matrix. Confusion matrix memberikan gambaran detail mengenai distribusi prediksi benar maupun salah pada tiap kelas. Gambar 6 menampilkan confusion matrix untuk skenario TF + Stemming + Stopword Removal. Terlihat bahwa kelas negatif (label 0) berhasil diklasifikasikan dengan sangat baik, di mana 1.062 data terprediksi benar dan hanya 6 data yang salah prediksi. Sebaliknya, pada kelas netral (label 1) terdapat 408 data yang benar diklasifikasikan, namun masih terdapat 74 data yang salah diprediksi sebagai kelas negatif.

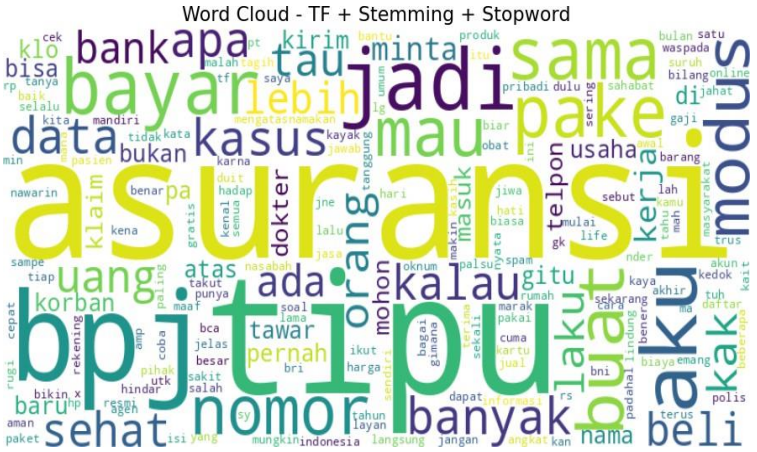
Temuan ini menguatkan hasil evaluasi pada Tabel 1, bahwa meskipun akurasi keseluruhan tergolong baik, performa antar kelas belum seimbang. Hal ini terutama tercermin pada nilai recall kelas netral yang lebih rendah, sehingga model cenderung lebih "kuat" mengenali kelas negatif dibandingkan kelas netral.



Gambar 6 Confusion Matrix pada Skenario *TF_Stem_Stop*

3.6 Visualisasi Word Cloud

Untuk memberikan gambaran umum mengenai distribusi kata yang sering muncul dalam dataset, dilakukan visualisasi menggunakan Word Cloud pada skenario TF + Stemming + Stopword Removal. Gambar 7 menunjukkan bahwa kata-kata seperti *"asuransi"*, *"bpjs"*, *"bayar"*, *"penipuan"*, dan *"nomor"* muncul dengan frekuensi yang tinggi. Hal ini menegaskan bahwa isu utama dalam dataset berkaitan erat dengan topik layanan asuransi, pembayaran, serta indikasi penipuan. Kehadiran kata-kata tersebut memperkuat hasil klasifikasi sebelumnya, karena secara semantik memang berhubungan erat dengan sentimen negatif dalam konteks penipuan asuransi. Dengan demikian, Word Cloud tidak hanya berfungsi sebagai visualisasi tambahan, tetapi juga memberikan validasi kualitatif terhadap pola yang ditangkap oleh model.



Gambar 7 *Word Cloud* pada *Skenario TF_Stem_Stop*

4. KESIMPULAN

Penelitian ini berhasil menganalisis sentimen masyarakat terkait isu penipuan asuransi dengan menggunakan algoritma Naïve Bayes dan Random Forest. Dari total 1.550 tweet yang dikumpulkan, setelah tahap preprocessing tersisa 1.482 data bersih yang siap diproses. Proses pelabelan otomatis berbasis rule-based menunjukkan bahwa sentimen negatif jauh lebih dominan dibandingkan sentimen netral. Hal ini memperlihatkan bahwa isu penipuan asuransi memunculkan persepsi publik yang cenderung negatif terhadap industri asuransi. Ekstraksi fitur menggunakan TF-IDF menghasilkan 4.734 kata unik yang digunakan sebagai representasi numerik dalam pemodelan. Visualisasi word cloud memperkuat hasil analisis, di mana kata-kata yang sering muncul seperti *asuransi*, *bpjs*, *bayar*, dan *penipuan* berkaitan erat dengan konteks penelitian. Hasil evaluasi menunjukkan bahwa algoritma Naïve Bayes memberikan performa terbaik pada skenario TF + Stemming + Stopword Removal dengan akurasi 89,67%, F1-score negatif 0,93, dan F1-score netral 0,80. Namun, nilai recall netral yang relatif rendah menunjukkan bahwa model ini masih cenderung lebih kuat dalam mengenali kelas negatif dibanding kelas netral. Berbeda dengan itu, algoritma Random Forest menghasilkan performa lebih optimal pada skenario TF + Stemming dengan akurasi mencapai 97,09%, F1-score negatif 0,97, dan F1-score netral 0,95. Hasil ini membuktikan bahwa Random Forest tidak hanya unggul dalam akurasi, tetapi juga lebih seimbang dalam mengenali kedua kelas. Meski demikian, performa Random Forest menurun signifikan ketika stopword removal digunakan secara berlebihan, sehingga proses preprocessing yang tepat menjadi faktor penting dalam klasifikasi.

Secara keseluruhan, penelitian ini menunjukkan bahwa Random Forest dengan kombinasi TF-IDF dan stemming merupakan metode paling efektif dalam menganalisis sentimen terkait penipuan asuransi. Hasil ini diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai opini publik, serta menjadi bahan pertimbangan bagi perusahaan asuransi maupun regulator dalam merumuskan strategi pencegahan dan penanganan kasus penipuan.

DAFTAR PUSTAKA

[1] F. Fahlapi, D. A. Prasetyo, and S. Febriyanti, "Analisis Sentimen Twitter Menggunakan Metode Naive Bayes dan K-Nearest Neighbor," Jurnal Ilmiah Media Sisfo, vol. 18, no. 1, pp. 1–9, 2024.

[2] M. Merdiansah and A. Ridha, "Analisis Sentimen Terhadap Asuransi Menggunakan Algoritma Naive Bayes," Jurnal Informanika, vol. 10, no. 2, pp. 56–65, 2024.

[3] R. D. Firmansyah and A. Lestariningsih, "Analisis Sentimen dengan Metode Naïve Bayes dan KNN pada Ulasan Layanan BPJS Kesehatan," Jurnal Teknik Komputer AMIK BSI, vol. 10, no. 1, pp. 45–54, 2024.

[4] A. Amrullah, M. B. Nugroho, and R. F. R. Rachman, "Analisis Sentimen Review Google Playstore Menggunakan Metode Naïve Bayes Classifier," Jurnal Informatika Polinema, vol. 6, no. 2, pp. 89–96, 2020.

[5] D. Apriani and D. Gustian, "Penerapan Algoritma Naïve Bayes untuk Analisis Sentimen Komentar Instagram," Jurnal Teknologi dan Sistem Informasi, vol. 5, no. 1, pp. 77–86, 2019.

[6] D. Artanto, "Penerapan Algoritma Random Forest dalam Analisis Sentimen Media Sosial," Jurnal Ilmu Komputer dan Teknologi Informasi, vol. 12, no. 2, pp. 210–219, 2024.

[7] N. Sidauruk, N. Riza, and R. N. Siti Fatonah, "Penggunaan Metode SVM dan Random Forest untuk Analisis Sentimen Ulasan Pengguna Terhadap KAI Access di Google Playstore," Jurnal JATI, vol. 7, no. 3, pp. 1901–1906, 2023.

***Prosiding- SEMASTER**: Seminar Nasional Teknologi Informasi & Ilmu Komputer* is licensed under a Creative Commons Attribution International (CC BY-SA 4.0)