

KOMPARASI ALGORITMA RANDOM FOREST CLASSIFIER DAN SUPPORT VECTOR MACHINE UNTUK SENTIMEN MASYARAKAT TERHADAP PINJAMAN ONLINE DI MEDIA SOSIAL

Dhea Annisa Fitri ¹⁾, Damayanti*²⁾

1. Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia
2. Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia

Article Info

Kata Kunci: Analisis Sentimen; Fintech; Random Forest; Support Vector Machine

Keywords: Fintech; Online Loans; Random Forest, Sentiment Analysis; Support Vector Machine;

Article history:

Received 29 September 2024

Revised 13 Oktober 2024

Accepted 4 November 2024

Available online 4 December 2024

DOI :

<https://doi.org/10.29100/jupi.v9i4.5608>

* Corresponding author.

Damayanti

E-mail address:

damayanti@teknokrat.ac.id

ABSTRAK

Perkembangan di sektor keuangan, terutama dengan kemunculan teknologi keuangan (*fintech*), telah membawa beragam layanan keuangan yang menggunakan teknologi, termasuk layanan pinjaman online. Fenomena pinjaman online menjadi sorotan utama di Indonesia, baik dari segi hukum maupun dampaknya terhadap masyarakat. Dalam konteks ini, penting untuk mengkaji analisis sentimen masyarakat terhadap pinjaman online yang tersebar di berbagai platform media sosial. Penelitian ini memiliki tujuan untuk membandingkan performa dua algoritma klasifikasi, yakni *Random Forest* dan *Support Vector Machine* (SVM), dalam menganalisis sentimen terhadap opini masyarakat terkait pinjaman online di media sosial, khususnya di platform Twitter. Metode penelitian mencakup proses pengumpulan data melalui pengambilan data pada platform Twitter, pemrosesan data untuk membersihkan dan memformat dataset, serta memberi label pada data untuk mengkategorikan opini sebagai positif atau negatif. Langkah selanjutnya melibatkan klasifikasi data, yang diikuti dengan evaluasi menggunakan *Confusion Matrix* untuk menilai performa kedua algoritma. Temuan penelitian menunjukkan bahwa kedua algoritma mampu melakukan klasifikasi dengan baik setelah data disesuaikan, namun *Random Forest* mencapai akurasi 100% sedangkan SVM mencapai 99%. Walaupun begitu, secara keseluruhan, *Random Forest* menunjukkan hasil yang lebih baik dengan akurasi yang lebih tinggi.

ABSTRACT

Developments in the financial sector, especially with the emergence of financial technology (*fintech*), have brought a variety of financial services that use technology, including online lending services. The phenomenon of online lending is a major concern in Indonesia, both in terms of the law and its impact on society. In this context, it is important to examine the analysis of public sentiment towards online loans spread across various social media platforms. This study aims to compare the performance of two classification algorithms, namely *Random Forest* and *Support Vector Machine* (SVM), in analyzing the sentiment of public opinion related to online loans on social media, especially on the Twitter platform. The research method includes the data collection process through data retrieval on the Twitter platform, data processing to clean and format the dataset, and labeling the data to categorize opinions as positive or negative. The next step involved data classification, followed by evaluation using *Confusion Matrix* to assess the performance of both algorithms. The findings showed that both algorithms were able to perform well in classification after the data was adjusted, but *Random Forest* achieved 100% accuracy while SVM achieved 99%. However, overall, *Random Forest* showed better results with higher accuracy.

I. PENDAHULUAN

KEMAJUAN dalam bidang finansial dalam era globalisasi merupakan salah satu bentuk terciptanya *financial technology*. *Financial technology* adalah kombinasi keuangan dengan sistem teknologi, dimana isu terkait fintech ini telah berkembang dengan pesat di Indonesia [1]. Munculnya fintech di Indonesia

diharapkan mampu mempermudah masyarakat semua lapisan dan golongan serta yang berada di daerah terpencil sekalipun, dalam akses mendapatkan layanan keuangan yang berbasis teknologi. Salah satu bentuk dari fintech itu sendiri adalah pinjaman online. Pinjaman online menjadi fenomena yang saat ini sedang banyak dibicarakan oleh masyarakat Indonesia [2].

Pinjaman online ini terdiri dari berbagai Lembaga baik legal maupun ilegal. Legalitas pinjaman online saat ini sudah diatur sangat baik oleh OJK (Otoritas Jasa Keuangan). Pinjaman online merupakan suatu fasilitas pinjaman uang oleh penyedia jasa keuangan yang terintegrasi dengan teknologi informasi, mulai dari proses pengajuan, persetujuan hingga pencairan dana dilakukan secara online atau melalui konfirmasi SMS atau telepon [3]. Pinjaman online secara ilegal ternyata masih diminati oleh masyarakat Indonesia karena syaratnya yang mudah, mempunyai kecenderungan pencairan yang cepat, dan mempunyai limit yang besar [4]. Pinjaman online ini dapat diatur terkait dengan tenor peminjaman sesuai dengan kebutuhan peminjam. Sebagian besar jenis pinjaman online justru tidak membutuhkan sebuah agunan atau jaminan [5].

Proses peminjaman online yang ditawarkan oleh beberapa Lembaga keuangan begitu mudah dalam pengajuannya. Signifikannya pertumbuhan perusahaan atau Lembaga keuangan sebagai Lembaga pemberi pinjaman diakibatkan dari banyaknya permintaan di pasar. Hal ini juga yang memicu munculnya perusahaan atau Lembaga keuangan yang ilegal. Perusahaan tersebut akan membuat nasabahnya menjadi korban dengan memberikan bunga yang tinggi. Hal ini dapat terjadi dikarenakan pinjaman online ilegal tidak diawasi oleh pihak OJK, sehingga peraturan yang dibuat tidak berdasarkan pada aturan yang telah dibuat oleh OJK. Berdasarkan laporan OJK (2023) nilai penyaluran fintech lending atau pinjaman online (pinjol) di Indonesia mencapai Rp20,53 triliun pada Agustus 2023. Jumlah penyaluran pinjol tersebut naik sedikit 0,78% dari bulan sebelumnya (month-on-month/mom) sebesar Rp20,37 triliun.

Semua itu merupakan kerja sama penyaluran oleh pemberi pinjaman institusi (super lender) pada Agustus 2023 yang berasal dari 2.196 lembaga jasa keuangan konvensional dengan nilai Rp5,92 triliun. Pemanfaatan secara tidak bijaksana merupakan awal dari permasalahan yang ada pada pinjaman online tersebut. Banyak pengalaman dari masyarakat baik pengalaman positif maupun negatif dari pinjaman online ini yang tersebar diberbagai media sosial baik di Instagram, Facebook, TikTok, Twitter dan lain-lain. Hal ini menimbulkan adanya sentiment dari masyarakat yang selalu menjadi topik pembicaraan yang menarik untuk diulas. Berkembangnya opini dan komentar masyarakat maka diperlukan analisa untuk mengklasifikasikan opini yang ada,

Analisis sentimen adalah cara untuk mendapatkan berbagai opini masyarakat tentang pelayanan publik, kasus, isu publik, serta kejadian permasalahan. Analisis sentimen adalah studi yang berfokus pada cara mengungkapkan sentimen dan opini dalam kalimat atau teks. Klasifikasi sentimen berkaitan dengan aktivitas mengelompokkan sentimen berdasarkan teks opini terkait topik yang dibahas [6]. Analisis sentimen penting dalam konteks pinjaman online untuk memahami opini dan perasaan pelanggan terhadap layanan, yang dapat membantu meningkatkan kepuasan dan memperbaiki layanan. Algoritma seperti *Support Vector Machine* (SVM) dan *Random Forest* digunakan untuk mengklasifikasikan teks menjadi sentimen positif, negatif, atau netral. SVM mencari *hyperplane* optimal untuk memisahkan data, sementara *Random Forest* menggunakan beberapa pohon keputusan untuk memberikan hasil yang akurat dan andal dalam analisis sentimen pelanggan.

Penelitian yang dilakukan sebelumnya menemukan bahwa analisis sentimen mengenai vaksinasi Covid-19 menggunakan algoritma *Support Vector Machine* (SVM) dan *Random Forest*. Perbandingan ini menghasilkan algoritma SVM dengan pelabelan textblob menghasilkan akurasi sebesar 0,8940. Sedangkan hasil sentimen menunjukkan masyarakat cenderung mempunyai opini negatif [7].

Penelitian yang telah dilakukan sebelumnya dengan pemodelan terbaik untuk analisis sentimen adalah dengan menggunakan *Support Vector Machine* dengan pelabelan Textblob. Penelitian selanjutnya menemukan akurasi pada SVM sebesar 97%, *Precision* 90% dan *Recall* 91%, sedangkan untuk *Accuracy in Random Forest* sebesar 92%, *Precision* 71%, dan *Recall* 86%. Nilai yang diperoleh merupakan hasil pengujian menggunakan 526 URL website. Hasil pengujian menunjukkan bahwa *Support Vector Machine* lebih baik dibandingkan *Random Forest* pada penelitian ini. Berdasarkan latar belakang tersebut, pada penelitian ini akan melakukan komparasi atau perbandingan metode mana yang lebih baik dalam mengklasifikasikan sentimen masyarakat mengenai pinjaman online di media sosial khususnya Twitter menggunakan *Random Forest* dan *Support Vector Machine* (SVM) [8].

Penelitian sebelumnya menunjukkan bahwa aplikasi FinTech seperti Dana sangat populer di kalangan generasi milenial Indonesia untuk pembayaran dan pembelian barang atau jasa, dengan pemilihan aplikasi berdasarkan kenyamanan, keamanan, ketepatan transaksi, kemudahan, dan promosi. Namun, ulasan di Google Play Store memengaruhi kepercayaan pengguna, baik melalui ulasan positif maupun negatif. Untuk mengatasi ini, dilakukan analisis menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neighbor, yang menunjukkan akurasi terbaik sebesar 84,76% dalam mengklasifikasi ulasan. Hasil ini menegaskan bahwa ulasan positif dapat meningkatkan kepercayaan pengguna, sementara ulasan negatif dapat menjadi masukan penting bagi perusahaan

untuk perbaikan produk [9].

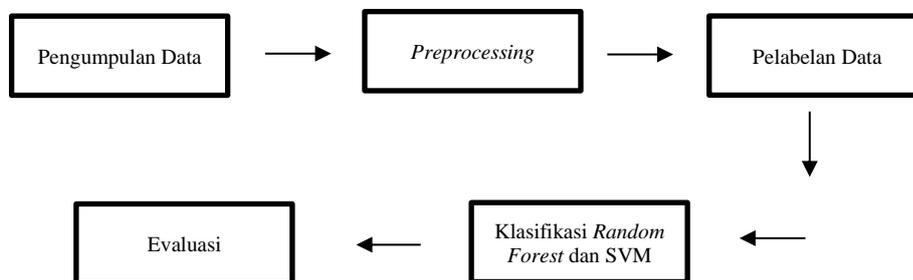
Penelitian sebelumnya menunjukkan bahwa pertumbuhan pesat uang elektronik telah mendorong banyak inovasi layanan baru, termasuk transaksi online menggunakan teknologi FinTech. FinTech yang populer saat ini meliputi e-wallet atau dompet digital seperti Dana, Ovo, dan LinkAja. Penelitian yang menggunakan aplikasi Dana menunjukkan bahwa faktor kenyamanan, keamanan, ketepatan transaksi, kemudahan, dan promosi mempengaruhi pilihan pengguna, dengan ulasan pengguna yang memengaruhi kepercayaan. Pada penelitian ini, aplikasi LinkAja dianalisis menggunakan 100 sampel ulasan yang diklasifikasikan menjadi positif dan negatif dengan metode Naive Bayes Classifier. Hasil analisis sentimen pada aplikasi LinkAja menunjukkan akurasi 75%, precision 83%, recall 75%, dan F1-score 73%, menegaskan pentingnya ulasan dalam membangun kepercayaan pengguna serta memberikan masukan untuk perbaikan layanan [10].

Pemilihan Support Vector Machine (SVM) dan Random Forest untuk analisis sentimen didasarkan pada keunggulan masing-masing algoritma. SVM efektif dalam memisahkan kelas-kelas data yang kompleks dengan margin yang jelas, bahkan dengan data pelatihan yang kecil. Random Forest, sebagai metode ensemble, meningkatkan akurasi dan mengurangi overfitting dengan menggabungkan hasil dari beberapa pohon keputusan, serta menangani data yang berisik dan tidak terstruktur dengan baik. Kombinasi ini memastikan analisis sentimen yang akurat dan andal, memenuhi kebutuhan akan interpretabilitas dan kecepatan pemrosesan.

Oleh karena itu penelitian ini bertujuan untuk membandingkan kinerja dua algoritma klasifikasi, yaitu Random Forest Classifier dan Support Vector Machine (SVM), dalam melakukan analisis sentimen terhadap opini masyarakat mengenai pinjaman online di media sosial, khususnya pada platform Twitter. Tujuan utama penelitian ini adalah untuk mengevaluasi dan membandingkan akurasi kedua algoritma dalam mengklasifikasikan sentimen dari teks yang dihasilkan oleh pengguna Twitter terkait topik pinjaman online. Selain itu, penelitian ini juga memiliki tujuan tambahan, seperti mengevaluasi waktu pemrosesan, mengidentifikasi fitur-fitur penting yang memengaruhi sentiment.

II. METODE PENELITIAN

Metode penelitian ini dimulai dengan pengumpulan data dari platform Twitter melalui proses *crawling* untuk mengambil data terkait opini masyarakat tentang pinjaman online. Data yang diperoleh kemudian diproses menggunakan teknik preprocessing, yang meliputi langkah-langkah seperti membersihkan data dari karakter yang tidak relevan, mengubah format, dan mempersiapkan dataset untuk analisis lanjutan. Selanjutnya, dilakukan pelabelan data di mana opini masyarakat dikategorikan menjadi positif atau negatif. Setelah proses pelabelan selesai, model klasifikasi, yaitu Random Forest dan Support Vector Machine, diterapkan untuk mengklasifikasikan data menjadi dua kategori sentimen. Evaluasi dilakukan menggunakan Confusion Matrix untuk mengevaluasi kinerja kedua model tersebut, dengan mengukur akurasi, presisi, recall, dan F1-score. Dengan demikian, alur pengumpulan data, preprocessing, pelabelan data, klasifikasi model, dan evaluasi membentuk kerangka metodologi yang komprehensif dalam penelitian ini. Berikut ini merupakan alur atau metode penelitian pada jurnal ini :



Gambar 1. Digram Alir Penelitian

2.1 Pengumpulan Data

Untuk mendapatkan sebuah dataset membutuhkan proses pengumpulan data yang diambil dari hasil *crawling* pada platform Twitter menggunakan API dengan bahasa pemrograman Python. Dari proses *crawling* tersebut dataset yang diperoleh berjumlah 3664 data dengan rentang waktu dari tanggal 31 Oktober 2023 – 6 Mei 2024 menggunakan kata kunci pencarian "pinjaman online" dan kemudian diekstraksi menjadi format.csv

2.2 Preprocessing

Dalam penelitian, proses *preprocessing* data merupakan langkah kritis untuk mempersiapkan dataset mentah sebelum dilakukan analisis lebih lanjut [11]. Tahapan pertama adalah *cleansing*, yang bertujuan untuk membersihkan data dari informasi yang tidak relevan atau mengganggu, seperti tanda baca dan karakter khusus. Hal ini penting karena memastikan bahwa data yang digunakan dalam analisis adalah data yang bersih dan terstruktur dengan baik. Kemudian, dalam proses *case folding*, semua karakter diubah menjadi huruf kecil untuk memastikan konsistensi dalam representasi kata yang sama, sehingga menghindari duplikasi yang tidak perlu dalam analisis. [12]. Tahapan *tokenize* kemudian memecah teks menjadi unit-unit diskrit, yang disebut token, untuk memfasilitasi analisis lanjutan [13]. Setelah itu, dalam proses *filtering*, token-token yang tidak relevan, seperti stop words dan karakter non-alfanumerik, dihapus dari dataset. Terakhir, tahapan *stemming* dilakukan untuk normalisasi morfologi kata dengan menghapus awalan dan akhiran, sehingga kata-kata dengan makna yang sama dapat diperlakukan secara seragam [14]. Hal ini membantu memperbaiki konsistensi data dan mengurangi variasi kata yang mungkin mengganggu dalam analisis sentimen. Keseluruhan, langkah-langkah *preprocessing* ini dipilih untuk mempersiapkan dataset dengan baik sehingga memungkinkan analisis sentimen yang lebih akurat dan efektif. Hasil proses *preprocessing* dapat dilihat pada Tabel 1 berikut ini.

Tabel 1. Hasil Preprocessing

Hasil Preprocessing	
<i>Dataset Awal</i>	Fix kalo kalian cari pinjaman online mending di akulaku aja. Legal udh pasti OJK. Data juga aman ga bakal kesebar kaya di pinjol ilegal. ini aku buat reviewnya. semoga membantu. https://t.co/IBLNsCs8Wr
<i>Cleansing</i>	Fix kalo kalian cari pinjaman online mending di akulaku aja Legal udh pasti OJK Data juga aman ga bakal kesebar kaya di pinjol ilegal ini aku buat reviewnya semoga membantu
<i>Case Folding</i>	fix kalo kalian cari pinjaman online mending di akulaku aja legal udh pasti ojk data juga aman ga bakal kesebar kaya di pinjol ilegal ini aku buat reviewnya semoga membantu
<i>Tokenize</i>	['fix', 'kalo', 'kalian', 'cari', 'pinjaman', 'online', 'mending', 'di', 'akulaku', 'aja', 'legal', 'udh', 'pasti', 'ojk', 'data', 'juga', 'aman', 'ga', 'bakal', 'kesebar', 'kaya', 'di', 'pinjol', 'ilegal', 'ini', 'aku', 'buatkan', 'reviewnya', 'semoga', 'membantu']
<i>Filtering</i>	['fix', 'kalo', 'cari', 'pinjaman', 'online', 'mending', 'akulaku', 'aja', 'legal', 'udh', 'ojk', 'data', 'aman', 'ga', 'kesebar', 'kaya', 'pinjol', 'ilegal', 'buatkan', 'reviewnya', 'semoga', 'membantu']
<i>Stemming</i>	fix kalo cari pinjam online mending akulaku aja legal udh ojk data aman ga sebar kaya pinjol ilegal buat reviewnya moga bantu

2.3 Pelabelan Data

Pada tahap pelabelan data ini, nantinya data akan dikategorikan menjadi dua, yaitu data yang bersifat positif, dan negatif. Data tersebut akan diproses secara otomatis menggunakan Library pada bahasa pemrograman Python.

Tabel 2. Hasil Labeling

Hasil Labeling	
fix kalo cari pinjam online mending akulaku aja legal udh ojk data aman ga sebar kaya pinjol ilegal buat reviewnya moga bantu	Positif
miris banyak lilit utang judi online amp pinjam online pinjol yang picu bunuh indonesia	Negatif

2.4 Klasifikasi

Penelitian ini akan membandingkan algoritma Random Forest dengan algoritma SVM. Algoritma Random Forest merupakan metode klasifikasi dan regresi yang tersusun dari sekumpulan pohon keputusan yang berfungsi sebagai base classifier yang dibangun dan dikombinasikan [14]. Bootstrap sampling merupakan salah satu aspek penting dari metode Random Forest yang berfungsi untuk membuat pohon prediksi, dimana setiap pohon keputusan menggunakan prediksi menggunakan prediktor acak dan rata-rata klasifikasi mayoritas dan regresi. Random Forest sendiri yang membuat prediksi dengan menggabungkan hasil dari setiap pohon Keputusan [15]. Sedangkan Support Vector Machine (SVM) adalah salah satu teknik pembelajaran yang telah di latih untuk membentuk hyperplanes

saat melakukan tugas-tugas seperti regresi, klasifikasi, dan mendeteksi anomali. Salah satu aplikasi utamanya adalah dalam klasifikasi teks dan hypertext [16]. SVM berupaya mencari hiperrata optimal yang memungkinkan pemisahan maksimal antara kelas-kelas titik data, sehingga memungkinkan klasifikasi dan prediksi yang tepat [14].

Pemilihan Random Forest dan Support Vector Machine (SVM) sebagai algoritma klasifikasi dalam penelitian ini didasarkan pada keunggulan masing-masing algoritma yang memenuhi kebutuhan penelitian. Random Forest, sebagai metode ensemble, efektif dalam mengatasi overfitting dan menangani dataset besar dengan banyak fitur, seperti dalam analisis sentimen terhadap opini masyarakat di media sosial. Di sisi lain, SVM terkenal karena kemampuannya dalam memisahkan kelas-kelas data yang kompleks, cocok untuk mengklasifikasikan sentimen dari teks yang kompleks dan beragam, seperti yang dihasilkan oleh pengguna media sosial. Kombinasi kedua algoritma ini memastikan analisis sentimen yang akurat dan handal dalam penelitian ini.

2.5 Evaluasi

Dalam tahap evaluasi kinerja sistem menggunakan *classification report*. Recall, precision, dan F1-score adalah metrik evaluasi yang umum digunakan dalam analisis klasifikasi, termasuk dalam analisis sentimen. Recall mengukur kemampuan model untuk mengidentifikasi semua instance positif yang relevan, sementara precision mengukur kemampuan model untuk mengidentifikasi instance positif secara akurat. F1-score memberikan keseimbangan antara kedua metrik tersebut. Misalnya, recall yang tinggi menunjukkan kemampuan model dalam mengidentifikasi sebagian besar sentimen positif atau negatif secara keseluruhan, sementara precision yang tinggi menunjukkan bahwa model jarang memberikan prediksi yang salah. F1-score menggabungkan kedua aspek ini untuk memberikan gambaran yang lebih komprehensif tentang kinerja model secara keseluruhan. Berikut adalah rumus untuk menghitung accuracy, precision, recall, dan f1-score dapat dilihat pada persamaan 1,2,3 dan 4 dibawah ini.

$$\text{accuracy} = \frac{TP+TN}{TP+FP+TN+TN} \quad (1)$$

$$\text{precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{recall} = \frac{TP}{TP+FN} \quad (3)$$

$$F1 \text{ Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Selain itu juga dilakukan pengujian dengan menggunakan Confusion Matrix, yang merupakan proses penting dalam mengevaluasi kinerja algoritma. Confusion Matrix digunakan untuk menilai seberapa baik algoritma tersebut berkinerja [17]. Tabel Confusion Matrix, seperti yang ditunjukkan dalam Tabel 3, memberikan representasi kelas prediksi pada setiap kolom dan kelas sebenarnya pada setiap baris. Untuk melakukan klasifikasi, empat metrik digunakan, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN), yang mewakili hasil klasifikasi. Dari tahap evaluasi ini, dihasilkan nilai untuk recall, precision, dan precision, yang merupakan evaluasi penting dalam menilai performa algoritma [18].

Tabel 3. Confusion Matrix

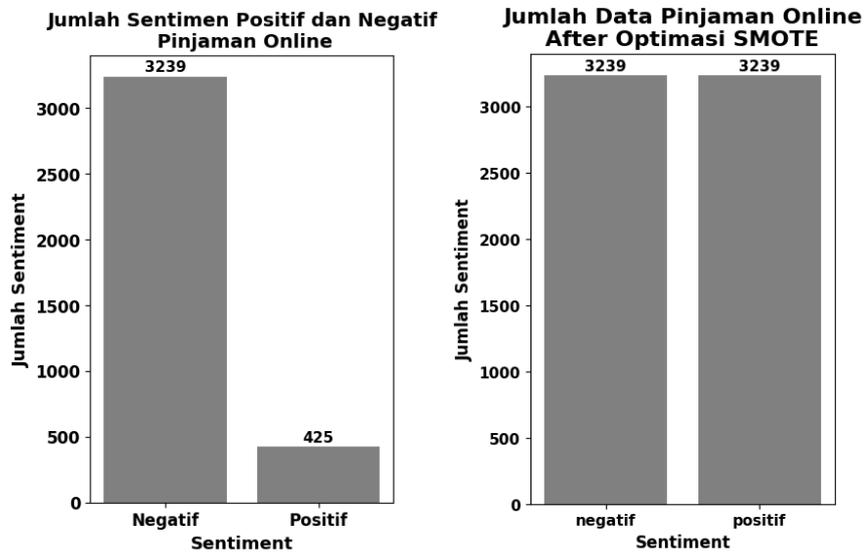
Klasifikasi	Prediksi	
	Positive	Negative
Aktual	Positive	True Positive (TP)
	Negative	True Negative (TN)
	Positive	False Positive (FP)
	Negative	False Negative (FN)

III. HASIL DAN PEMBAHASAN

A. Tahap Pengujian

Setelah melalui tahap pengumpulan data, preprocessing, pelabelan data, klasifikasi model, dan evaluasi, hasil dari penelitian ini terungkap melalui tahap pengujian. Pengujian dilakukan untuk mengevaluasi kinerja kedua model klasifikasi, yaitu Random Forest dan Support Vector Machine, dalam mengklasifikasikan sentimen terkait pinjaman online dari data yang telah diproses. Hasil pengujian mengungkapkan akurasi, presisi, *recall*, dan *F1-score* dari masing-masing model, yang kemudian akan menjadi dasar untuk pembahasan lebih lanjut tentang keefektifan dan keunggulan dari masing-masing model dalam melakukan analisis sentimen. Dengan demikian,

tahap pengujian menjadi pintu masuk yang relevan menuju hasil dan pembahasan dalam penelitian ini, yang akan membahas temuan dan implikasinya terhadap pemahaman kita tentang analisis sentimen terhadap pinjaman online di media sosial.



Gambar 2. Jumlah data sebelum dan sesudah optimasi SMOTE

Gambar 2 merupakan hasil dari tahap pengujian jumlah sentimen positif dan negatif serta proses terjadinya SMOTE. SMOTE adalah teknik yang digunakan untuk mengatasi masalah ketidakseimbangan data, di mana satu kelompok data memiliki jumlah yang jauh lebih besar daripada kelompok lainnya. Ketidakseimbangan ini dapat menyebabkan model yang dikembangkan menjadi terlalu terfokus pada kelompok mayoritas, sehingga performa prediksi terhadap kelompok minoritas menjadi kurang baik. Dengan SMOTE, data tambahan dibuat untuk kelompok minoritas sehingga jumlah data dalam kedua kelompok menjadi seimbang. Pendekatan ini membantu model belajar secara lebih adil dan mengurangi risiko overfitting [19]. Dalam analisis data sentimen terhadap pinjaman online, ditemukan ketidakseimbangan yang signifikan antara jumlah data yang menyatakan sentimen positif dan negatif. Sebelum proses SMOTE, terdapat 425 data dengan sentimen positif dan 3.239 data dengan sentimen negatif. Ketidakseimbangan semacam ini dapat menghasilkan bias dalam model klasifikasi, di mana model mungkin cenderung lebih baik dalam mengklasifikasikan kategori mayoritas (negatif) daripada kategori minoritas (positif).

Untuk mengatasi ketidakseimbangan ini, diterapkan teknik SMOTE (*Synthetic Minority Over-sampling Technique*). SMOTE bekerja dengan cara menciptakan sampel sintesis baru dalam ruang fitur untuk kategori minoritas, sehingga menyamakan jumlah data antara kedua kategori. Dalam konteks ini, SMOTE digunakan untuk menghasilkan sampel sintesis baru dengan sentimen positif hingga jumlahnya menjadi setara dengan jumlah data sentimen negatif, yaitu 3.239 data. Proses SMOTE sendiri terdiri dari beberapa tahapan yang mendalam. Pertama, untuk setiap data dalam kategori minoritas (positif), jarak euclidean dihitung dengan data tetangga terdekatnya. Kemudian, satu atau beberapa tetangga dipilih secara acak, dan titik-titik sintesis baru dihasilkan di sepanjang garis yang menghubungkan data dengan tetangga terpilih. Proses ini dilakukan untuk setiap data dalam kategori minoritas hingga jumlahnya seimbang dengan jumlah data dalam kategori mayoritas.

Hasil dari proses SMOTE adalah peningkatan jumlah data sentimen positif hingga setara dengan jumlah data sentimen negatif, yaitu 3.239 data. Dengan demikian, ketidakseimbangan data telah berhasil diatasi, dan dataset siap untuk digunakan dalam pelatihan model klasifikasi yang adil dan akurat. Dengan menggunakan teknik SMOTE, kita dapat memastikan bahwa model klasifikasi tidak cenderung bias terhadap satu kategori tertentu karena ketidakseimbangan dalam jumlah data.

B. Klasifikasi Model

Setelah penerapan SMOTE untuk menyeimbangkan jumlah data sentimen positif dan negatif dalam dataset pinjaman online, total jumlah data yang tersedia menjadi 6.478, dengan 5.182 data untuk pelatihan model dan 1.296 data untuk pengujian model. Dataset dibagi menjadi data latih (80%) dan data uji (20%). Dua algoritma klasifikasi

yang digunakan, yaitu Random Forest dan Support Vector Machine (SVM), dilatih dengan data latih yang telah disimbangkan melalui SMOTE. Random Forest menggunakan beberapa pohon keputusan dan melakukan voting untuk menentukan kelas prediksi, sementara SVM mencari hyperplane terbaik yang memisahkan kelas dalam ruang fitur. Performa keduanya dievaluasi menggunakan data uji dengan metrik seperti akurasi, presisi, recall, dan F1-score [20]. Analisis ini memberikan pemahaman tentang efektivitas kedua model dalam mengklasifikasikan sentimen pinjaman online setelah penyeimbangan data dengan SMOTE.

```

===== Jumlah Data Keseluruhan =====
6475

===== Jumlah Data Latih =====
5188

===== Jumlah Data Uji =====
1295
    
```

Gambar 3. Jumlah data keseluruhan, data latih dan data uji

Tabel 4. Hasil Accuracy

Model	Accuracy
RandomForest	100 %
Support Vector Machine	99 %

Dalam hasil analisis klasifikasi menggunakan metrik akurasi, disajikan dalam Tabel 4, terlihat bahwa kedua model klasifikasi, yaitu Random Forest dan Support Vector Machine (SVM), menunjukkan performa yang sangat baik. Random Forest mencapai akurasi 100%, menandakan kemampuannya dalam menghasilkan prediksi yang sangat akurat dengan teknik ensemble yang menggunakan beberapa pohon keputusan untuk melakukan voting. Hal ini mengindikasikan bahwa model Random Forest mampu mengatasi overfitting dan meningkatkan generalisasi terhadap data uji dengan baik. Sementara itu, model SVM juga menunjukkan performa yang sangat baik dengan akurasi 99%. Meskipun tidak mencapai akurasi sempurna, nilai akurasi yang mendekati 100% menunjukkan kemampuan SVM dalam mencari hyperplane terbaik yang memisahkan kelas-kelas dalam ruang fitur, yang menghasilkan prediksi yang sangat akurat. Dengan demikian, hasil ini menegaskan bahwa kedua model klasifikasi, Random Forest dan SVM, dapat diandalkan dalam mengklasifikasikan sentimen pinjaman online setelah proses penyeimbangan data dengan SMOTE, dan memberikan kontribusi yang berharga dalam pengembangan model klasifikasi yang andal dan akurat untuk aplikasi di bidang keuangan dan fintech.

Tabel 5. Hasil Klasifikasi Algoritma Random Forest

No	Parameter	Random Forest	
		Nilai	
		Positif	Negatif
1	Recall	99 %	100 %
2	Precision	100 %	99 %
3	F1-Score	100 %	100 %

Hasil klasifikasi menggunakan algoritma Random Forest, seperti yang terdokumentasi dalam Tabel 5, menunjukkan evaluasi performa model yang memadai dengan metrik klasifikasi standar, yaitu Recall, Precision, dan F1-Score, untuk kedua kelas sentimen, baik positif maupun negatif. Untuk kelas sentimen positif, diperoleh nilai Recall sebesar 99%, menunjukkan kemampuan model untuk mengidentifikasi sebagian besar data sentimen positif dengan akurat. Hal ini berarti bahwa model dapat menangkap sebagian besar kasus sentimen positif yang sebenarnya dalam dataset. Selanjutnya, nilai Precision sebesar 100% menunjukkan bahwa dari semua prediksi yang dilakukan oleh model untuk sentimen positif, semuanya benar atau hanya memiliki sedikit false positive. Ini mengindikasikan keakuratan model dalam memprediksi sentimen positif tanpa memberikan banyak prediksi yang salah. F1-Score,

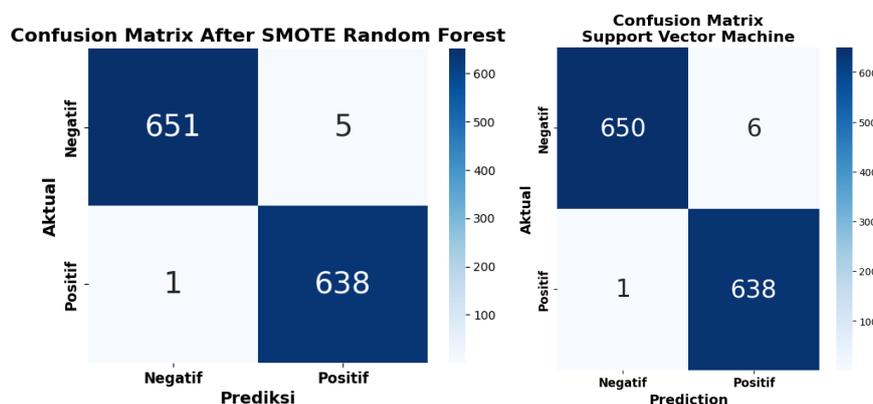
yang merupakan harmonik dari Recall dan Precision, mencapai 100%, menunjukkan keseimbangan yang baik antara Recall dan Precision untuk kelas sentimen positif.

Ini mengonfirmasi bahwa model memiliki kinerja yang sangat baik dalam memprediksi sentimen positif secara keseluruhan. Untuk kelas sentimen negatif, hasilnya juga menjanjikan dengan Recall dan F1-Score mencapai 100%, menunjukkan kemampuan model dalam mengidentifikasi sebagian besar data sentimen negatif secara tepat. Namun, Precision sedikit lebih rendah, yaitu 99%, yang berarti ada sedikit false positive dalam prediksi sentimen negatif. Meskipun demikian, hasil ini tetap menunjukkan bahwa model Random Forest memberikan performa yang sangat baik dalam mengklasifikasikan sentimen positif dan negatif terkait pinjaman online setelah proses penyeimbangan data menggunakan SMOTE.

Tabel 6. Hasil Klasifikasi Algoritma Support Vector Machine

<i>Support Vector Machine</i>			
No	Parameter	Nilai	
		Positif	Negatif
1	<i>Recall</i>	99 %	100 %
2	<i>Precision</i>	100 %	99 %
3	<i>F1-Score</i>	99 %	99 %

Dalam analisis mendalam, ditemukan bahwa SVM mencapai performa yang sangat baik dalam mengklasifikasikan kedua kategori sentimen, baik positif maupun negatif. Dilihat dari nilai Recall, SVM mencapai akurasi sebesar 99% untuk kategori sentimen positif dan 100% untuk kategori sentimen negatif. Ini menunjukkan bahwa SVM mampu dengan sangat baik dalam mengidentifikasi sebagian besar data yang benar-benar termasuk dalam setiap kategori sentimen. Selanjutnya, nilai Precision yang mencapai 100% untuk sentimen positif dan 99% untuk sentimen negatif menunjukkan bahwa dari semua prediksi yang dibuat oleh model SVM, sebagian besar di antaranya adalah benar. Terakhir, F1-Score, yang merupakan perpaduan antara Recall dan Precision, juga menunjukkan performa yang sangat baik untuk kedua kategori sentimen dengan nilai sebesar 99%. Hal ini mengindikasikan bahwa SVM mampu memberikan keseimbangan yang baik antara keakuratan prediksi dan kemampuan untuk mengidentifikasi data yang relevan dalam kedua kategori sentimen. Dengan demikian, hasil analisis secara mendalam menegaskan bahwa SVM merupakan algoritma yang efektif dalam mengklasifikasikan sentimen terhadap pinjaman online dengan akurasi tinggi dan keseimbangan yang baik antara kategori sentimen positif dan negatif.



Gambar 4. Hasil Confusion Matrix

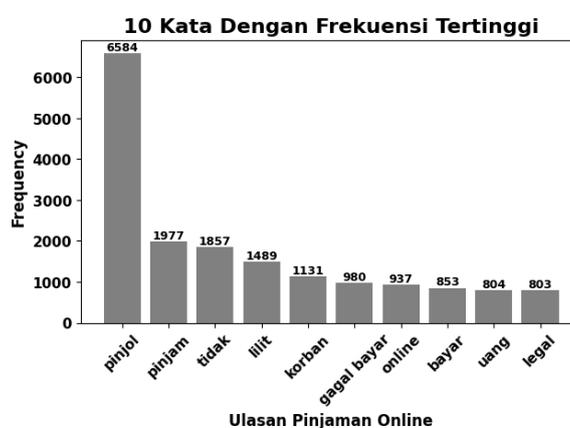
Dari hasil Confusion Matrix setelah penerapan SMOTE Random Forest, didapatkan nilai True Positive (TP) sebesar 638, yang mengindikasikan jumlah data positif yang diprediksi dengan benar. Selain itu, terdapat 651 data yang diklasifikasikan secara tepat sebagai negatif (True Negative/TN), yang menunjukkan jumlah data negatif yang diprediksi dengan benar. Namun, terdapat 5 data positif yang seharusnya diprediksi tetapi salah diklasifikasikan sebagai negatif (False Negative/FN), dan hanya 1 data negatif yang salah diprediksi sebagai positif (False Positive/FP). Dari analisis ini, meskipun model memiliki performa yang baik dalam mengklasifikasikan data positif dan negatif dengan benar (tinggi TP dan TN), masih terdapat beberapa kasus di mana model membuat kesalahan (FN dan FP).

Interpretasi terperinci terhadap hasil evaluasi model menyoroti beberapa faktor yang memengaruhi kinerja model dan implikasinya dalam konteks aplikasi praktis. Pertama, meskipun kedua model, Random Forest dan Support

Vector Machine (SVM), menunjukkan akurasi yang tinggi, ada perbedaan dalam jumlah kasus yang salah diprediksi, terutama dalam hal false positive dan false negative. Ini menandakan bahwa meskipun model memiliki performa yang solid, masih ada ruang untuk perbaikan untuk mengurangi kesalahan prediksi yang mungkin terjadi. Faktor-faktor yang memengaruhi kinerja model meliputi kualitas data awal, pengaturan parameter model, serta kompleksitas dan kecocokan algoritma dengan struktur data.

Implikasinya adalah bahwa dalam aplikasi praktis, penyedia layanan atau peneliti perlu melakukan evaluasi yang cermat terhadap faktor-faktor ini untuk memastikan kinerja model yang optimal. Selain itu, hasil evaluasi juga menunjukkan bahwa teknik seperti SMOTE dapat membantu mengatasi ketidakseimbangan data dan meningkatkan kinerja model dalam mengklasifikasikan kategori minoritas. Dengan demikian, dalam konteks aplikasi praktis, pemahaman tentang teknik preprocessing data seperti SMOTE dapat membantu meningkatkan kualitas prediksi model. Kesimpulannya, interpretasi terperinci terhadap hasil evaluasi model ini memberikan wawasan yang berharga tentang faktor-faktor yang memengaruhi kinerja model dan implikasinya dalam pengembangan aplikasi praktis di berbagai bidang, termasuk keuangan dan fintech.

C. Visualisasi Data



Gambar 5. Frekuensi kata yang sering muncul

Analisis terhadap sepuluh kata dengan frekuensi tertinggi dalam ulasan pinjaman online menyoroti tren dan topik yang dominan dalam percakapan pengguna. Dari hasil visualisasi data dalam bentuk bar chart, terlihat bahwa kata "pinjol" mendominasi dengan frekuensi tertinggi, menandakan dominasi industri pinjaman online dalam ulasan pengguna. Diikuti oleh "pinjam", yang mencerminkan keinginan atau tindakan pengguna untuk memanfaatkan layanan pinjaman tersebut. Namun, di tengah dominasi kata-kata yang mengindikasikan interaksi positif, terdapat juga kata-kata dengan konotasi negatif seperti "tidak" dan "gagal bayar", yang mengisyaratkan adanya pengalaman negatif atau kegagalan dalam pembayaran pinjaman.

Kata "lilit" dan "korban" juga mencerminkan pengalaman yang mungkin sulit atau tidak menyenangkan bagi sebagian pengguna. Selain itu, fokus pengguna pada aspek finansial tercermin melalui kemunculan kata-kata seperti "online", "bayar", dan "uang". Terakhir, kebutuhan akan kepastian dan keabsahan layanan pinjaman online ditunjukkan melalui frekuensi kata "legal". Analisis ini memberikan pemahaman yang lebih dalam tentang perspektif dan kebutuhan pengguna dalam pengalaman mereka dengan layanan pinjaman online, yang dapat menjadi landasan bagi penyedia layanan untuk meningkatkan kualitas layanan mereka.

Word cloud merupakan sebuah gambaran grafis yang menampilkan kata-kata tertentu yang sering muncul dalam suatu teks, dan sering digunakan sebagai cara yang menarik dan efektif untuk memvisualisasikan informasi. Dalam konteks penelitian ini, penggunaan word cloud menjadi penting sebagai alat visual yang sangat bermanfaat dalam memvisualisasikan hasil analisis data yang diperoleh melalui proses web scraping [6].

WordCloud Analisis Sentiment Pinjaman Online



Gambar 6. Hasil Wordcloud semua sentimen

Analisis word cloud mengenai sentimen pinjaman online dengan kata-kata yang paling dominan menunjukkan pola-pola penting dalam pengalaman pengguna. Kata-kata yang muncul dengan ukuran yang lebih besar dalam word cloud, seperti "pinjol", "lilit", "gagal", dan "bayar", menyoroti topik-topik yang menjadi fokus utama percakapan pengguna. "Pinjol" sebagai kata yang paling besar menunjukkan dominasi industri pinjaman online dalam diskusi tersebut. Kemudian, kata-kata seperti "pinjam online" dan "judi online" muncul sebagai bagian dari konteks penggunaan layanan secara daring. "Gagal" dan "bayar" memberikan indikasi adanya pengalaman negatif atau kesulitan dalam proses pembayaran pinjaman. Di sisi lain, kata "berkah" dan "minat" memberikan indikasi adanya aspek positif atau keinginan pengguna dalam menggunakan layanan tersebut. "Legal" muncul sebagai pertanda bahwa pengguna mungkin mencari kepastian dan keabsahan dalam menggunakan layanan pinjaman online. Analisis ini memberikan pemahaman yang lebih dalam tentang fokus, sentimen, dan kebutuhan pengguna dalam konteks pengalaman mereka dengan layanan pinjaman online, yang dapat menjadi dasar untuk perbaikan layanan dan pengembangan strategi pemasaran yang lebih efektif.

WordCloud Positif Analisis Sentiment Pinjaman Online



Gambar 7. Hasil Wordcloud sentiment positif dan negatif

Analisis terhadap hasil wordcloud analisis sentimen positif terhadap pinjaman online menampilkan kata-kata yang paling sering muncul dalam konteks positif. Kata-kata yang memiliki ukuran paling besar dalam wordcloud adalah "pinjam" dan "minat", menunjukkan bahwa pengguna menyatakan minat yang tinggi dalam melakukan pinjaman online. Kata-kata ini mencerminkan keinginan dan tindakan positif dari pengguna terhadap layanan

model membuat kesalahan, terutama terlihat dari jumlah *False Negative* dan *False Positive*. Meskipun demikian, *Random Forest* menunjukkan hasil yang lebih baik secara keseluruhan dengan mencapai akurasi sempurna. Oleh karena itu, dalam memilih algoritma yang terbaik untuk kasus ini, *Random Forest* merupakan pilihan yang lebih unggul.

DAFTAR PUSTAKA

- [1] F. Syifa *Et Al.*, "Analysis Of Bsi Readiness In Facing The Development Of Fintech-Based," Vol. 10, No. 1, Pp. 1–10, 2023.
- [2] I. Zuchroh, "Fintech Syariah: Kolaborasi Teknologi Dan Moral Sebagai Instrumen Pembiayaan Di Masa Depan," *Ecoplan*, Vol. 4, No. 2, Pp. 122–130, 2021, Doi: 10.20527/Ecoplan.V4i2.383.
- [3] A. Abdullah, "Analisis Pengetahuan Pinjaman Online Pada Masyarakat Surakarta," *Jesi (Jurnal Ekon. Syariah Indones.)*, Vol. 11, No. 2, P. 108, 2021, Doi: 10.21927/Jesi.2021.11(2).108-114.
- [4] A. Savitri, A. Syahputra, H. Hayati, And H. Rofizar, "Pinjaman Online Di Masa Pandemi Covid-19 Bagi Masyarakat Aceh," *E-Mabis J. Ekon. Manaj. Dan Bisnis*, Vol. 22, No. 2, Pp. 116–124, 2021, Doi: 10.29103/E-Mabis.V22i2.693.
- [5] M. F. Azis And N. D. Rahmawati, "Tinjauan Hukum Terhadap Perjanjian Pinjaman Online Dan Penggunaan Data Konsumen Aplikasi 'Kredit Pintar,'" *Fortioro Law J.*, Vol. 1, No. 1, P. 5, 2021.
- [6] D. Surya Sayogo, B. Irawan, And A. Bahtiar, "Analisis Sentimen Ulasan Instagram Di Google Play Store Menggunakan Algoritma Naïve Bayes," *Jati (Jurnal Mhs. Tek. Inform.)*, Vol. 7, No. 6, Pp. 3314–3319, 2024, Doi: 10.36040/Jati.V7i6.8178.
- [7] B. P. Meliani, O. N. Pratiwi, And R. Andreswari, "Comparison Of Support Vector Machine And Random Forest Algorithms In Sentiment Analysis On Covid-19 Vaccination On Twitter Using Vader And Textblob Labelling," *Proc. Int. Conf. Appl. Sci. Technol. Soc. Sci. 2022 (Icast-Ss 2022)*, Pp. 620–626, 2022, Doi: 10.2991/978-2-494069-83-1_108.
- [8] H. Syahputra And A. Wibowo, "Comparison Of Support Vector Machine (Svm) And Random Forest Algorithm For Detection Of Negative Content On Websites," *J. Ilm. Tek. Elektro Komput. Dan Inform.*, Vol. 9, No. 1, Pp. 165–173, 2023, Doi: 10.26555/Jiteki.V9i1.25861.
- [9] C. Dan, K.-N. Neighbor, S. Aji, And F. F. Wati, "Analisa Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes," Vol. 8, No. 1, Pp. 93–105, 2020.
- [10] M. Sitepu, M. Yohanna, And S. V. B. H. Manurung, "Metode Naïve Bayes Classifier," Vol. 8, No. 1, Pp. 44–50, 2024.
- [11] N. Wahyuningsih And H. Hendry, "Perbandingan Metode Klasifikasi Dalam Analisis Sentimen Masyarakat Terhadap Identitas Kependudukan Digital (Ikd)," *Jipi (Jurnal Ilm. Penelit. Dan Pembelajaran Inform.)*, Vol. 8, No. 4, Pp. 1218–1227, 2023, Doi: 10.29100/Jipi.V8i4.4155.
- [12] D. C. Ramadhan And F. Irwiensyah, "Analisis Sentimen Pengguna Terhadap Aplikasi Bing Chat Di Google Play Store Dengan Metode Naïve Bayes," Vol. 4, No. 5, Pp. 2410–2418, 2024, Doi: 10.30865/Klik.V4i5.1769.
- [13] S. Wulandari And F. N. Hasan, "Analisis Sentimen Masyarakat Indonesia Terhadap Pengalaman Belanja Thrifting Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes," Vol. 8, No. April, Pp. 768–776, 2024, Doi: 10.30865/Mib.V8i2.7520.
- [14] P. M. S. Ardinata, A. A. J. Permana, I. N. S. W. Wijaya, F. Teknik, And D. Kejuruan, "Identifikasi Dan Normalisasi Teks Slang Dengan Fasttext Pada Twitter Dalam Bahasa Indonesia," *J. Pendidik. Teknol. Dan Kejuru.*, Vol. 21, No. 1, 2024.
- [15] Z. H. Nst *Et Al.*, "Perbandingan Algoritma Random Forest Dan K-Nearest," No. 01, Pp. 31–37.
- [16] D. Ananda And R. R. Suryono, "Analisis Sentimen Publik Terhadap Pengungsi Rohingya Di Indonesia Dengan Metode Support Vector Machine Dan Naïve Bayes," Vol. 8, No. April, Pp. 748–757, 2024, Doi: 10.30865/Mib.V8i2.7517.
- [17] Jan Melvin Ayu Soraya Dachi And Pardomuan Sitompul, "Analisis Perbandingan Algoritma Xgboost Dan Algoritma Random Forest Ensemble Learning Pada Klasifikasi Keputusan Kredit," *J. Ris. Rumpun Mat. Dan Ilmu Pengetah. Alam*, Vol. 2, No. 2, Pp. 87–103, 2023, Doi: 10.55606/Jurrimipa.V2i2.1470.
- [18] M. R. Sholahuddin, F. Atqiya, S. R. Wulan, M. Harika, S. Fitriani, And Y. Sofyan, "Implementasi Sistem Identifikasi Senjata Real Time Menggunakan Yolov7 Dan Notifikasi Chat Telegram," *J. Inf. Syst. Res.*, Vol. 4, No. 2, Pp. 598–606, 2023, Doi: 10.47065/Josh.V4i2.2774.
- [19] H. Hidayatullah *Et Al.*, "Penerapan Naïve Bayes Dengan Optimasi Information Gain Dan," Vol. 7, No. 3, 2023.
- [20] A. Ramadhan And M. M. Rose, "Komparasi Algoritma Neural Network Dan K-Nearest Neighbor Dalam Mendeteksi Malware Android," Vol. 5, No. 1, Pp. 191–199, 2023, Doi: 10.47065/Bits.V5i1.3538.