Jurnal Teknik Informatika (JUTIF)
Vol. 5, No. 6, December 2024, pp. 1801-1810
DOI: https://doi.org/10.52436/1.jutif.2024.5.6.3434
p-ISSN: 2723-3863
e-ISSN: 2723-3871

SENTIMENT ANALYSIS AND ENTITY DETECTION ON NEWS HEADLINES TO SUPPORT INVESTMENT DECISIONS

Ajar Parama Adhi*[1], Khairil Umuri*[2], Gandung Triyono*[3]
[1,2,3]Fakultas Teknologi Informasi, Magister Ilmu Komputer, Universitas Budi Luhur, Indonesia
E-mail: [1]2211601501@student.budiluhur.ac.id, [2]2211601584@student.budiluhur.ac.id,
[3]gandung.triyono@budiluhur.ac.id

(Article received: September 3, 2024; Revision: October 13, 2024; published: December 29, 2024)

Abstract

Accurate investment decisions are often influenced by information available in the media. News headlines, as part of information media, can provide an initial picture of market sentiment and ongoing trends. This research examines the importance of making appropriate investment decisions with a focus on sentiment analysis and entity detection in news headlines as supporting tools. Through machine learning-based sentiment analysis and Named Entity Recognition (NER) techniques, this study identifies opinions and entities such as company names, stock indices, and industry sectors in news headlines. This research compares three machine learning algorithms, namely SVM, Naive Bayes, and Random Forest using cross-validation. The result shows that the best algorithm is SVM with weighted average F1-score of 76,68%. Furthermore, hyperparameter optimization is performed using Optuna for the SVM algorithm, which is an innovation in the context of sentiment analysis on news headlines in Indonesia. The result shows an increase in weighted average F1-score to 78,14%. For NER, a rule-based method is used by utilizing the Jaro-Winkler string similarity function. The combination of sentiment analysis and NER is then presented in the form of a dashboard using Google Looker Studio tools, with data from sentiment analysis and NER results being processed periodically and automatically using Google Workflows. This research makes a significant contribution by expanding the scope of analysis from just one or a few issuers to all entities published on news portals thanks to NER support, making the results relevant to support investment decisions that are responsive to dynamic market changes.

Keywords: *decision support system, investment decisions, machine learning, named entity recognition (NER), news headlines, sentiment analysis.*

ANALISIS SENTIMEN DAN DETEKSI ENTITAS TERHADAP JUDUL BERITA SEBAGAI PENDUKUNG KEPUTUSAN INVESTASI

Abstrak

Keputusan investasi yang tepat sering kali dipengaruhi oleh informasi yang tersedia di media. Judul berita, sebagai bagian dari media informasi, dapat memberikan gambaran awal mengenai sentimen pasar dan tren yang sedang berlangsung. Penelitian ini mengkaji pentingnya pengambilan keputusan investasi yang tepat dengan fokus pada analisis sentimen dan deteksi entitas dalam judul berita sebagai alat pendukung. Melalui analisis sentimen berbasis *machine learning* dan *Named Entity Recognition* (NER), penelitian ini mengidentifikasi opini dan entitas seperti nama perusahaan, indeks saham, dan sektor industri dalam judul berita. Penelitian ini membandingkan tiga algoritma *machine learning*, yaitu SVM, Naive Bayes, dan Random Forest menggunakan *cross validation*. Hasilnya didapat algoritma terbaik yaitu SVM dengan nilai rata-rata tertimbang *F1-score* sebesar 76,68%. Selanjutnya, dilakukan optimasi *hyperparameter* menggunakan Optuna untuk algoritma SVM, yang merupakan inovasi dalam konteks analisis sentimen pada judul berita di Indonesia. Hasilnya terdapat peningkatan rata-rata tertimbang *F1-score* menjadi 78,14%. Dari sisi NER, digunakan metode *rule-based* dengan memanfaatkan fungsi *string similarity* yaitu Jaro-Winkler. Kombinasi analisis sentimen dan NER ini kemudian disajikan dalam bentuk *dashboard* dengan memanfaatkan *tools* Google Looker Studio, data dari hasil analisis sentimen dan NER diproses periodik secara otomatis memanfaatkan Google Workflows. Penelitian ini memberikan kontribusi signifikan dengan memperluas cakupan analisis dari hanya satu atau beberapa emiten menjadi seluruh entitas yang dimuat pada portal berita berkat dukungan NER, menjadikan hasilnya relevan untuk mendukung keputusan investasi yang responsif terhadap perubahan pasar yang dinamis.

Kata kunci: *analisis sentimen, judul berita, keputusan investasi, named entity recognition (NER), pembelajaran mesin, sistem pendukung keputusan.*

1 PENDAHULUAN

Dalam dunia investasi, pengambilan keputusan yang tepat merupakan kunci utama untuk mencapai keuntungan yang diharapkan dan meminimalkan risiko kerugian [1]. Investor selalu mencari cara untuk memperoleh informasi yang akurat dan relevan untuk mendukung keputusan mereka. Salah satu sumber informasi yang paling berpengaruh adalah media massa, yang mencakup berbagai platform seperti surat kabar, televisi, radio, dan yang semakin dominan, media digital. Di antara berbagai bentuk informasi yang disediakan oleh media massa, judul berita memainkan peran penting karena sering kali menjadi titik fokus pertama yang menarik perhatian pembaca dan terbukti memiliki korelasi yang erat [2].

Judul berita tidak hanya berfungsi sebagai ringkasan singkat dari isi artikel, tetapi juga dapat mencerminkan sentimen pasar dan tren yang sedang berlangsung [3]. Sentimen pasar yang diungkapkan melalui judul berita dapat memiliki dampak signifikan terhadap perilaku investor dan keputusan investasi mereka [4]. Sentimen ini dapat bersifat positif, negatif, atau netral, dan dapat mempengaruhi persepsi investor terhadap kondisi pasar dan prospek masa depan.

Penelitian terdahulu telah menunjukkan bahwa sentimen berita dapat dilakukan untuk memprediksi tren harga saham. Pada penelitian [5] digunakan algoritma *deep learning* untuk memprediksi harga saham berdasarkan data sentimen berita yang dilengkapi dengan indikator teknikal, model terbaik didapatkan dengan memanfaatkan *lexicon* dari Loughran–McDonald Financial Dictionary [6].

Penelitian [7] juga menunjukkan bahwa sentimen berita harian memiliki pengaruh signifikan terhadap pasar saham, penelitian tersebut menggunakan data indeks saham di Brazil dengan periode Juni sampai dengan Desember 2018.

Di samping informasi sentimen dari berita, deteksi entitas atau dikenal juga sebagai *Named Entity Recognition* (NER) dapat digunakan untuk memahami lebih baik tentang konteks dan subyek dari sentimen sehingga memperkaya informasi yang dapat digunakan sebagai pendukung keputusan investasi [8]. Penelitian [9] mengombinasikan analisis sentimen dengan sistem deteksi entitas. Berdasarkan penelitian tersebut, sentimen dari judul berita yang muncul pada sore hari setelah pasar saham ditutup akan memengaruhi harga saham di hari esoknya.

Selain menggunakan *machine learning*, deteksi entitas dapat dilakukan menggunakan metode *rule-based* seperti yang pernah dilakukan dalam penelitian [10]. Metode *rule-based* memiliki keunggulan dalam hal kecepatan dan skalabilitas jika digunakan pada data ukuran besar.

Penelitian ini bertujuan untuk menganalisis sentimen dan mendeteksi entitas dalam judul berita guna mendukung pengambilan keputusan investasi. Dalam konteks ini, analisis sentimen digunakan untuk mengklasifikasikan sentimen yang terkandung dalam judul berita. Sebagai penunjang informasi sentimen, deteksi entitas ditambahkan untuk membantu dalam mengidentifikasi faktor-faktor kunci yang mempengaruhi sentimen tersebut, seperti perusahaan yang terlibat, indeks saham yang teridentifikasi, atau sektor industri yang sedang mengalami perubahan signifikan.

Metode yang digunakan dalam penelitian ini meliputi analisis sentimen berbasis *machine learning* dengan membandingkan algoritma SVM, Naive Bayes, dan Random Forest. Teknik NER yang dipakai adalah *rule based* dengan memanfaatkan fungsi *string similarity* yaitu Jaro Winkler. Algoritma tersebut dipilih karena telah terbukti memiliki kinerja yang bagus untuk menangani kasus analisis sentimen, contohnya pada penelitian [11], SVM berhasil meraih akurasi 79% dalam analisis sentimen terhadap inflasi pasca Covid-19 pada Twitter. Naive Bayes memiliki keunggulan pada sisi kecepatan prediksi [12], cocok jika diterapkan untuk klasifikasi *real time*. Random Forest juga memiliki kinerja yang bagus dalam analisis sentimen [13], selain itu Random Forest juga tidak mudah *overfit* terhadap data latih [14].

Jika dibandingkan dengan penelitian-penelitian sebelumnya, salah satu perbedaan utama penelitian ini adalah penekanan pada analisis sentimen pada judul berita daripada platform media sosial Twitter seperti yang digunakan pada [15]. Selain itu, penelitian ini menggunakan metode Optuna untuk mengoptimalkan *hyperparameter* [16], sementara NER menggunakan pendekatan aturan. Kontribusi penting dari penelitian ini adalah penggunaan data dari judul berita, yang lebih inklusif daripada fokus pada satu emiten saja. Secara khusus, penelitian ini juga mencatat bahwa optimasi *hyperparameter* dengan Optuna untuk algoritma SVM pada data judul berita di Indonesia merupakan langkah baru yang belum pernah dilakukan sebelumnya.

Pentingnya penelitian ini terletak pada kemampuannya untuk memberikan alat analisis yang lebih canggih bagi investor untuk menghadapi kompleksitas pasar saham. Integrasi teknik analisis sentimen dan deteksi entitas memungkinkan pengambilan keputusan yang lebih terinformasi dan memperkenalkan strategi investasi yang adaptif dan responsif terhadap perubahan pasar. Dalam era di mana informasi tersebar luas melalui media massa dan tren pasar dapat berubah dengan cepat, kemampuan untuk menganalisis sentimen pasar dan mengidentifikasi entitas relevan dalam judul berita memberikan keunggulan kompetitif bagi investor. Dengan model yang dapat memberikan rekomendasi investasi yang akurat dan informatif, investor dapat meminimalkan risiko dan meningkatkan potensi keuntungan mereka. Sehingga, penelitian ini bukan hanya merupakan kontribusi pada pengembangan alat analisis, tetapi juga langkah maju dalam

memberdayakan investor dalam lingkungan investasi yang dinamis dan kompleks.

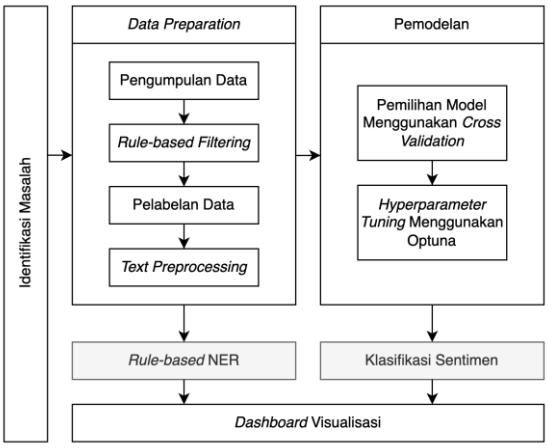
2 METODE PENELITIAN

Dalam melakukan penelitian ini terdapat beberapa tahapan yang dapat dilihat pada Gambar 1. Secara umum, prosesnya meliputi tahap pengumpulan data dari *website* berita menggunakan teknik *crawling*, dilanjutkan dengan *rule based filtering* untuk memisahkan data yang tidak terkait dengan penelitian, kemudian dilakukan pelabelan data secara manual, selanjutnya masuk ke tahap *preprocessing* di mana data dilakukan manipulasi agar dapat digunakan sebagai input saat pemodelan.

Pada tahap pemodelan, dibandingkan beberapa algoritma dengan metode *cross validation.* Algoritma yang dibandingkan yaitu SVM, Naive Bayes, dan Random Forest. Model terbaik dipilih berdasarkan rata-rata akurasi yang diperoleh pada setiap *fold.* Kemudian dilakukan *hyperparameter tuning* terhadap model terbaik menggunakan metode Optuna.

Setelah model untuk klasifikasi sentimen selesai dibuat, proses selanjutnya adalah membuat algoritma deteksi entitas bernama atau *Named Entity Recognition* (NER) menggunakan teknik *rule-based* yaitu dengan *string similarity.*

Proses klasifikasi sentimen dan NER dilakukan untuk setiap data sehingga dapat diketahui sentimen dan emiten yang terkait dengan setiap berita yang muncul. Hasil analisis sentimen dan NER kemudian diintegrasikan dan disajikan menggunakan visualisasi sehingga memungkinkan pemahaman yang lebih komprehensif terhadap judul berita investasi. Kombinasi ini tidak hanya mengidentifikasi sentimen keseluruhan dari judul berita, tetapi juga mengenali entitas spesifik yang terlibat, memberikan konteks yang lebih kaya untuk pengambilan keputusan investasi.



Gambar 1. Diagram Alur Metode Penelitian

2.1 Identifikasi Masalah

Berdasarkan latar belakang penelitian yang sudah dijelaskan, masalah yang hendak dipecahkan pada penelitian ini antara lain:

1. Model apa yang paling tepat digunakan untuk melakukan klasifikasi sentimen terhadap berita?
2. Bagaimana metode NER yang dapat digunakan untuk mengidentifikasi entitas yang terkait dengan berita?
3. Bagaimana *dashboard* visualisasi yang dibangun agar dapat mendukung keputusan investasi?

2.2 Pengumpulan Data

Setelah masalah berhasil diidentifikasi, tahap selanjutnya adalah pengumpulan data. Data diperoleh menggunakan teknik *scraping* yang berasal dari portal berita Kontan dengan alamat *https://kontan.co.id* dan dibatasi hanya pada rubrik investasi.

2.3 Rule Based Filtering

Pemilihan berita yang terkait atau tidak terkait dengan topik penelitian dilakukan menggunakan metode *rule based.* Tujuannya agar berita yang masuk proses analisis sentimen lebih terfokus hanya yang terkait dengan informasi seputar investasi saham saja.

2.4 Pelabelan Data

Sebelum dilakukan *text preprocessing*, teks berita harus diberi label secara manual terlebih dahulu. Alasan utamanya adalah agar label sentimen atau kelas yang diberikan pada data teks tidak terganggu atau berubah setelah dilakukan operasi *preprocessing* seperti penghapusan *stopwords*, *stemming*, tokenisasi, atau normalisasi teks. Kelas label yang digunakan ada tiga yaitu positif, negatif, dan netral.

2.5 Text Preprocessing

Dalam tahap ini dilakukan beberapa metode *natural language processing* (NLP) untuk mengolah data teks agar dapat dijadikan sebagai input pada algoritma *machine learning.* Berikut adalah tahapan *text preprocessing* yang dilakukan:

1. *Case Folding*
   Case folding adalah proses mengonversi semua huruf dalam teks menjadi huruf besar *(uppercase)* atau huruf kecil *(lowercase)* secara konsisten. Tujuan utama *case folding* adalah untuk menyeragamkan representasi teks dan menghilangkan perbedaan antara huruf besar dan huruf kecil.
2. *Stemming*
   Stemming adalah proses menghilangkan imbuhan dari sebuah kata untuk mendapatkan kata dasarnya *(stem)*. Tujuan *stemming* adalah untuk menyederhanakan kata-kata menjadi

bentuk dasarnya sehingga kata-kata yang memiliki akar kata yang sama dapat dikelompokkan bersama.

3. Mengapus *Stopwords*
   Stopwords adalah kata-kata umum yang seringkali muncul dalam teks namun tidak memberikan kontribusi yang signifikan terhadap makna teks. Menghapus stopwords dapat membantu mengurangi dimensi data dan meningkatkan efisiensi pemrosesan.
4. Tokenisasi
   Tokenisasi adalah proses memecah teks menjadi unit-unit yang lebih kecil, seperti kata, frasa, atau kalimat. Tokenisasi adalah langkah awal yang penting dalam pemrosesan bahasa alami karena memungkinkan penanganan dan analisis teks pada level yang lebih granular.
5. Vektorisasi
   Vektorisasi adalah proses mengonversi teks menjadi representasi numerik atau vektor yang dapat diproses oleh algoritma *machine learning*.

2.6 SVM

Support Vector Machine (SVM) adalah algoritma *supervised machine learning* yang bertujuan menemukan *hyperplane* optimal untuk memisahkan kelas-kelas dalam ruang fitur. SVM mencari margin maksimum antara *hyperplane* dan titik-titik data terdekat *(support vectors)*. Untuk data yang tidak dapat dipisahkan secara linear, SVM menggunakan teknik *kernel* untuk menangani data yang tidak linear dengan cara memetakan data ke dimensi yang lebih tinggi [17].

Dalam analisis sentimen, algoritma SVM telah digunakan secara luas. SVM merupakan salah satu algoritma yang efektif dalam melakukan analisis sentimen pada data teks dari media sosial seperti Twitter [18], [19]. Selain itu, SVM juga dapat diterapkan untuk klasifikasi berita, contohnya seperti klasifikasi validitas berita apakah hoax atau tidak terkait virus Corona (Covid-19) [20]. SVM dapat membantu dalam mengklasifikasikan berita berdasarkan fitur-fitur tertentu yang diidentifikasi dalam teks berita.

2.7 Naive Bayes

Naive Bayes merupakan salah satu algoritma probabilistik pada *machine learning* berdasarkan Teorema Bayes yang mengasumsikan hubungan antar fitur bersifat independen [21]. Meskipun asumsi ini sering tidak realistis dan jarang terjadi, Naive Bayes tetap efektif dalam berbagai kasus nyata, terutama untuk klasifikasi teks dan analisis sentimen.

Prinsip dasar dari algoritma Naive Bayes melibatkan perhitungan probabilitas suatu kelas tertentu berdasarkan serangkaian fitur yang diberikan. Metode ini menggunakan Teorema Bayes dengan rumus sebagai:

$$P(A|B) = \frac{(P(B|A) \cdot P(A))}{P(B)}$$

Di mana P(A|B) adalah probabilitas posterior, P(B|A) adalah *likelihood*, P(A) adalah probabilitas *prior*, dan P(B) adalah *evidence*. Naive Bayes bekerja dengan menghitung probabilitas setiap kelas berdasarkan fitur-fitur yang ada kemudian memilih kelas dengan probabilitas tertinggi sebagai hasil klasifikasi [22].

Algoritma ini sangat efisien dalam hal komputasi dan juga umum dipakai untuk mengklasifikasikan teks dari Twitter [23], [24] maupun berita [7], [25].

2.8 Random Forest

Random Forest adalah metode *ensemble* yang menggunakan banyak pohon keputusan untuk melakukan klasifikasi. Setiap pohon dibangun menggunakan *subset* acak dari data *training* dan *subset* acak dari fitur [26]. Metode ini memberikan prediksi dengan akurasi tinggi dan interpretabilitas yang baik [27]. Random Forest melakukan prediksi dengan voting mayoritas dari semua pohon untuk menentukan klasifikasi.

Algoritma ini ikut dibandingkan dalam pemilihan model pada penelitian ini karena umum dipakai pada klasifikasi teks seperti data tekstual laporan keuangan [28], media sosial [24], dan artikel berita [4], [5], [29].

2.9 Cross Validation

Cross validation adalah sebuah metode yang digunakan dalam pengembangan dan evaluasi model *machine learning* untuk mengukur seberapa baik model tersebut dapat digeneralisasi ke data baru yang tidak terlihat sebelumnya dan juga untuk menghindari *overfitting* [30]. Metode ini melibatkan pembagian data menjadi *subset* pelatihan dan pengujian secara berulang, di mana model dilatih pada *subset* pelatihan dan diuji pada *subset* pengujian untuk mengukur kinerjanya. *Cross validation* juga bermanfaat untuk membantu pemilihan model dan *hyperparameter tuning* [31].

2.10 Hyperparameter Tuning

Hyperparameter tuning adalah proses optimasi parameter-parameter yang tidak dipelajari secara langsung oleh model *machine learning* selama proses pelatihan. Parameter-parameter ini mempengaruhi kinerja model dalam melakukan prediksi, pemilihan nilai *hyperparameter* yang tepat dapat secara signifikan meningkatkan kinerja model [32].

Optuna merupakan *framework* optimasi *hyperparameter* yang menggunakan pendekatan *Tree-structured Parzen Estimator* (TPE) dengan metode *bayesian optimization* [16], artinya Optuna memanfaatkan model probabilistik untuk memandu pencarian *hyperparameter* yang optimal [31].

Keunggulan Optuna yaitu ruang pencarian *hyperparameter* dapat berubah selama proses optimasi berlangsung dan prosesnya dapat disajikan dalam bentuk visualisasi untuk mempermudah pemahaman [16].

2.11 Rule-based NER

Named Entity Recognition (NER) adalah pendekatan dalam pemrosesan bahasa alami (*Natural Language Processing*/NLP) yang bertujuan untuk mengidentifikasi entitas bernama (seperti nama orang, organisasi, lokasi, tanggal, dan lainnya) dalam teks berdasarkan kesamaan fitur.

Selain menggunakan *machine learning*, NER juga dapat dilakukan dengan metode *rule-based*, terutama ketika sudah diketahui basis pengetahuan mengenai populasi entitas yang mungkin muncul [33]. Salah satu tekniknya adalah dengan memanfaatkan algoritma *string similarity* seperti Jaro-Winkler [34], [35].

Untuk menghitung kemiripan teks menggunakan Jaro-Winkler, terdapat dua tahap, pertama menghitung Jaro Similarity:

$$JaroSim = \frac{1}{3}\left(\frac{|m|}{|s1|} + \frac{|m|}{|s2|} + \frac{|m-t|}{|m|}\right)$$

di mana $|m|$ adalah jumlah karakter yang cocok, $|s1|$ dan $|s2|$ adalah panjang teks pertama dan kedua, dan $|t|$ adalah jumlah transposisi. Dari hasil hitungan tersebut, dapat diperoleh nilai Jaro-Winkler dengan:

$$JaroWinkler = JaroSim + (l \times p \times (1 - JaroSim))$$

di mana l adalah panjang prefiks yang cocok dan p adalah konstanta.

3 HASIL DAN PEMBAHASAN

3.1 Data yang Dikumpulkan

Scraping dilakukan pada *website* Kontan khusus hanya rubrik Investasi untuk rentang waktu 1 tahun, dimulai dari 1 Januari 2023 sampai dengan 31 Desember 2023. *Tools* yang digunakan adalah Jupyter Notebook dengan bahasa pemrograman Python. Hasilnya, didapatkan sebanyak 19.490 judul berita dengan kutipan datanya seperti yang ditunjukkan pada Tabel 1.

Tabel 1. Kutipan Data Awal yang Berhasil Dikumpulkan

No	Tanggal	Judul Berita
1	2023-01-01	Saham Sektor Teknologi Masih Akan Tertekan di 2023, Begini Rekomendasi Analis
2	2023-01-01	Resesi Ekonomi Mengancam, Prospek Komoditas Logam Mulia Diramal Cerah di 2023
3	2023-01-01	Berikut Prediksi Pergerakan Kurs Rupiah di Hari Pertama Perdagangan 2023
4	2023-01-01	Kinerja Portofolio Investasi 2022 Lebih Rendah Dibandingkan Tahun 2021
5	2023-01-01	Menilik Prospek Saham Emiten Keuangan Non Bank pada 2023

3.2 Rule Based Filtering

Setelah data awal berhasil dikumpulkan, selanjutnya dilakukan *filtering* untuk membuang judul berita yang tidak relevan dengan investasi saham, contohnya judul berita yang mengandung *clickbait* dan tidak merepresentasikan sektor, kondisi ekonomi, atau emiten. Penyaringan judul berita dilakukan menggunakan *regular expression* dengan kriteria sebagai berikut:

1. Untuk membuang judul berita yang mengandung *clickbait*, dibuang judul yang mengandung kata-kata "simak", "intip", "cek", "saham-saham", "pilihan", "begini". Dari setiap kata tersebut, digunakan *word boundary* (\b) dari *regular expression* pada awal dan akhir kata untuk menandai bahwa kata tersebut berbatasan dengan karakter non huruf.
2. Untuk membuang judul berita yang tidak relevan, *regular expression* digunakan untuk membuang judul berita yang tidak mengandung kode emiten, nama emiten, dan yang tidak merepresentasikan sektor atau kondisi ekonomi. Daftar emiten didapat dari *website* Bursa Efek Indonesia (BEI).

Hasil dari proses penyaringan data relevan, tersisa sebanyak 6.995 baris judul berita seperti yang dikutip pada Tabel 2.

Tabel 2. Kutipan Data Setelah Proses *Filtering* Relevansi

No	Judul Berita
1	PPKM Dicabut, IHSG Berpotensi Menguat di Hari Pertama Perdagangan 2023, Senin (2/1)
2	Bisnis Baru Astra Otoparts (AUTO), Pasarkan Alat Pengisian Daya Kendaraan Listrik
3	Analis Prediksi IHSG Berpotensi Menguat ke 8.205 pada Akhir Tahun 2023
4	Menakar Dampak Pencabutan PPKM Tehadap Saham Emiten Properti dan Ritel
5	Data Sinergitama (ELIT) Pasang Harga IPO di Rp 120, Incar Dana Segar Rp 60 Miliar

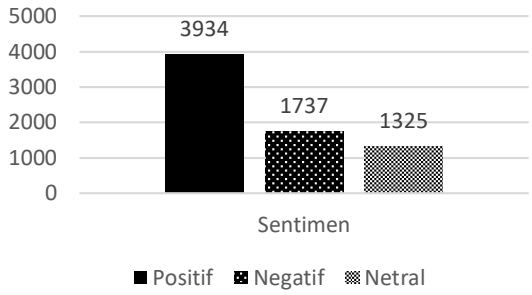
3.3 Pelabelan Data

Judul berita tersaring selanjutnya dilakukan proses pelabelan secara manual dan telah divalidasi oleh 2 orang pakar bergelar S2 Ekonomi, langkah ini diambil untuk menjaga objektivitas penelitian. Label terdiri dari 3 kelas, yaitu positif, negatif, dan netral. Hasil dari proses ini dapat dilihat pada Tabel 3.

Tabel 3. Kutipan Data Setelah Proses Pelabelan Manual

No	Judul Berita	Label
1	PPKM Dicabut, IHSG Berpotensi Menguat di Hari Pertama Perdagangan 2023, Senin (2/1)	Positif
2	Bisnis Baru Astra Otoparts (AUTO), Pasarkan Alat Pengisian Daya Kendaraan Listrik	Positif
3	Analis Prediksi IHSG Berpotensi Menguat ke 8.205 pada Akhir Tahun 2023	Positif
4	Menakar Dampak Pencabutan PPKM Tehadap Saham Emiten Properti dan Ritel	Netral
5	IHSG Merosot 1,10% ke 6.813 Rabu (4/1), MIKA, UNTR dan ADRO Top Losers di LQ45	Negatif

Berdasarkan hasil pelabelan, didapatkan sebanyak 3.934 judul berita dengan sentimen positif, 1.325 judul berita dengan sentimen negatif, dan 1.737 judul berita dengan sentimen netral. Proporsi setiap kelas sentimen dapat dilihat pada Gambar 2.



Gambar 2. Proporsi Setiap Kelas Sentimen Hasil Pelabelan Manual

Karena jumlah kelas sentimen tidak seimbang, maka diputuskan untuk menggunakan *weighted average* atau rata-rata tertimbang *F1-score* sebagai pengukuran kinerja model, metrik ini menghitung *F1-score* untuk setiap kelas secara terpisah dan kemudian mengambil rata-ratanya, sehingga memberikan bobot yang proporsional untuk setiap kelas.

3.4 Text Preprocessing

Sebelum dijadikan sebagai input algoritma *machine learning*, judul berita perlu melewati tahap pra-pemrosesan untuk dibersihkan dan disiapkan sesuai kebutuhan pemodelan. Tahap ini mencakup proses *case folding*, *stemming*, penghapusan *stopwords*, tokenisasi, dan vektorisasi. Tabel 4 menunjukkan kutipan data judul berita sebelum dan sesudah melewati proses *text preprocessing* sampai tahap penghapusan *stopwords*.

Tabel 4. Kutipan Data Setelah Proses Pelabelan Manual

No	Judul Berita	Judul Berita *Prepocessed*
1	PPKM Dicabut, IHSG Berpotensi Menguat di Hari Pertama Perdagangan 2023, Senin (2/1)	ppkm cabut ihsg potensi kuat hari pertama dagang 2023 senin
2	Bisnis Baru Astra Otoparts (AUTO), Pasarkan Alat Pengisian Daya Kendaraan Listrik	bisnis baru astra otoparts auto pasar alat isi daya kendara listrik
3	Analis Prediksi IHSG Berpotensi Menguat ke 8.205 pada Akhir Tahun 2023	analis prediksi ihsg potensi kuat akhir tahun 2023
4	Menakar Dampak Pencabutan PPKM Tehadap Saham Emiten Properti dan Ritel	takar dampak cabut ppkm tehadap saham emiten properti ritel
5	IHSG Merosot 1,10% ke 6.813 Rabu (4/1), MIKA, UNTR dan ADRO Top Losers di LQ45	ihsg merosot rabu mika untr adro top losers lq45

3.5 Pemilihan Model

Dalam penelitian ini, dibandingkan tiga algoritma *machine learning* yaitu: SVM, Naive Bayes, dan Random Forest. Perbandingan dilakukan menggunakan metode *cross validation*, lebih tepatnya yaitu *Stratified K-Fold Cross Validation*. Metode ini dipilih karena tetap mempertimbangkan proporsi masing-masing kelas untuk setiap *fold* yang dijalankan. Jumlah *fold* yang dipakai sebanyak 5 kali. Hasil iterasi yang diperoleh dapat dilihat pada Tabel 5.

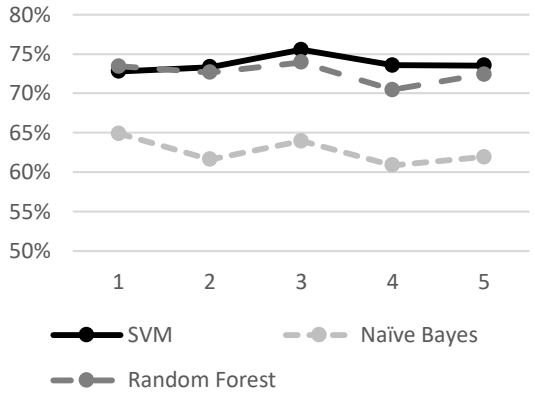
Tabel 5. Hasil *Cross Validation*

Algoritma	*Fold*	Waktu Prediksi (detik)	*Accuracy*	*Precision*	*Recall*	*F1-score*
SVM	1	0.8065	0.7693	0.7681	0.7693	0.7576
	2	0.8028	0.7748	0.7692	0.7748	0.7632
	3	0.8098	0.7963	0.7927	0.7963	0.7835
	4	0.7944	0.7806	0.7812	0.7806	0.7642
	5	0.8054	0.7798	0.7766	0.7798	0.7652
Rata-rata SVM		**0.8038**	**0.7802**	**0.7776**	**0.7802**	**0.7668**
Naive Bayes	1	0.0069	0.7264	0.7298	0.7264	0.6993
	2	0.0068	0.7026	0.7002	0.7026	0.6699
	3	0.0068	0.7234	0.7287	0.7234	0.6923
	4	0.0069	0.7019	0.7045	0.7019	0.6648
	5	0.0069	0.7119	0.7230	0.7119	0.6766
Rata-rata Naive Bayes		**0.0069**	**0.7133**	**0.7173**	**0.7133**	**0.6806**
Random Forest	1	0.0868	0.7729	0.7705	0.7729	0.7632
	2	0.0858	0.7691	0.7618	0.7691	0.7564
	3	0.0856	0.7834	0.7775	0.7834	0.7706
	4	0.0861	0.7584	0.7539	0.7584	0.7410
	5	0.0869	0.7663	0.7606	0.7663	0.7535
Rata-rata Random Forest		**0.0862**	**0.7700**	**0.7649**	**0.7700**	**0.7569**

Selanjutnya, model terbaik dipilih berdasarkan nilai rata-rata skor yang diperoleh untuk setiap *fold*. Perbandingan kinerja setiap model dari hasil *cross validation* dapat dilihat pada Gambar 3. Berdasarkan eksperimen yang telah dilakukan, model yang dipilih yaitu SVM dengan nilai rata-rata tertimbang *F1-score* sebesar 76,68%. Kekurangan dari model SVM adalah waktu prediksinya paling lama dibanding model lainnya, namun hal ini tidak menjadi masalah karena sistem yang dibangun menggunakan metode *batch* untuk memproses klasifikasi sentimen.

Jika dibutuhkan klasifikasi sentimen secara *realtime*, Naive Bayes dapat dipertimbangkan karena memiliki waktu prediksi yang paling cepat.

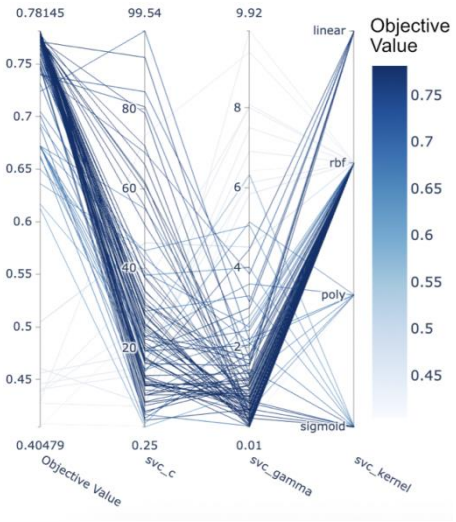
Di samping itu, Random Forest juga dapat dipertimbangkan jika membutuhkan model dengan kinerja prediksi yang cukup baik dan waktu yang cepat.



Gambar 3. Perbandingan Akurasi Setiap Model

3.6 Hyperparameter Tuning

Setelah model terbaik dipilih, dilanjutkan dengan optimasi melalui proses *hyperparameter tuning*. Metode yang digunakan adalah Optuna dengan *hyperparameter* yang dioptimasi yaitu nilai C, gamma, dan kernel pada algoritma SVM. Metode *cross validation* yang digunakan masih sama dengan proses sebelumnya agar dapat diperbandingkan hasilnya. Jumlah *trial* yang dilakukan pada Optuna adalah 100 percobaan untuk mencari model dengan nilai rata-rata tertimbang *F1-score* terbaik. Hasilnya dapat dilihat pada Gambar 4. Berdasarkan gambar tersebut, dapat dilihat bahwa skor rata-rata tertimbang *F1-score* yang tinggi didapat dari nilai C yang agak rendah, gamma yang rendah, dan pada kernel RBF. Hasil akhir optimasi didapat rangkaian *hyperparameter* berikut: C = 12,97, gamma = 0,42, kernel = RBF. Berdasarkan optimasi tersebut, nilai rata-rata tertimbang *F1-score* berhasil naik menjadi 78,14%.



Gambar 4. Proses *Hyperparameter Tuning* Menggunakan Optuna

3.7 *Rule-based* NER

Untuk melakukan deteksi entitas, tahap pertama yang dilakukan adalah mendapatkan daftar emiten yang terdaftar pada Bursa Efek Indonesia (BEI) untuk dijadikan sebagai *dictionary* entitas. Data tersebut didapatkan dari *website* BEI, kemudian dipilih dengan rincian terdapat 934 perusahaan tercatat, 7 indeks saham, dan 11 sektor. Data tersebut kemudian diolah menjadi 1.897 baris daftar entitas dengan kutipan yang dapat dilihat pada Tabel 8.

Tabel 8. Kutipan Data *Dictionary* Entitas

No	Nama	Jenis Entitas
1	AALI	emiten
2	ACES	emiten
3	BBRI	emiten
4	IHSG	indeks
5	LQ45	indeks
6	IDX30	indeks
7	Sektor Kesehatan	sektor
8	Sektor Keuangan	sektor
9	Sektor Teknologi	sektor

Dalam melakukan pencocokan entitas, digunakan fungsi Jaro-Winkler seperti yang dilakukan pada penelitian [36]. Hasil dari proses ini dapat dilihat pada Tabel 9.

Tabel 9. Kutipan Data Setelah Proses Pelabelan Manual

No	Judul Berita	Entitas Terdeteksi
1	PPKM Dicabut, IHSG Berpotensi Menguat di Hari Pertama Perdagangan 2023, Senin (2/1)	IHSG
2	Bisnis Baru Astra Otoparts (AUTO), Pasarkan Alat Pengisian Daya Kendaraan Listrik	Astra Otoparts, AUTO
3	Analis Prediksi IHSG Berpotensi Menguat ke 8.205 pada Akhir Tahun 2023	IHSG
4	Menakar Dampak Pencabutan PPKM Tehadap Saham Emiten Properti dan Ritel	Sektor Properti
5	IHSG Merosot 1,10% ke 6.813 Rabu (4/1), MIKA, UNTR dan ADRO Top Losers di LQ45	IHSG, MIKA, UNTR, ADRO, LQ45

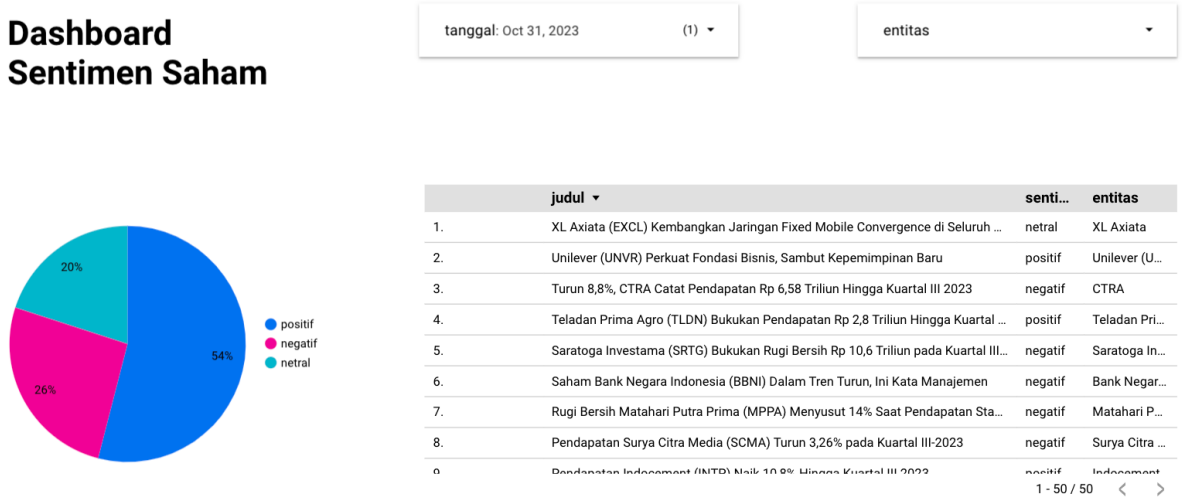
3.8 *Dashboard* Pendukung Keputusan Investasi

Hasil dari analisis sentimen dan NER kemudian diintegrasikan dan disajikan dalam bentuk *dashboard* visualisasi yang informatif dan mudah dipahami oleh pengguna. Alur proses integrasi disusun menggunakan Google Workflows dengan tahapan sebagai berikut:

1. *Scraping* dilakukan secara periodik, yaitu setiap 1 jam sekali dengan *link* berita dijadikan sebagai *primary key* untuk menghindari duplikasi.
2. Hasil *scraping* kemudian melalui proses *text preprocessing*, langkahnya sama seperti pada bagian 2.5.
3. Setelah melalui *text preprocessing*, selanjutnya dilakukan klasifikasi sentimen dan NER secara paralel.

4. Hasil klasifikasi sentimen dan NER disimpan dalam basis data dan ditampilkan pada *dashboard* visualisasi.

Dashboard tersebut dapat digunakan untuk mendapatkan informasi terkait keputusan investasi yang akan diambil. Gambar 5 menunjukkan tampilan visualisasi yang dapat dimanfaatkan. Dengan menggunakan *dashboard* tersebut, investor diberikan fasilitas untuk melakukan filter data berdasarkan rentang waktu dan entitas yang terdeteksi, visualisasi *pie chart* sebagai agregasi sentimen pada rentang waktu dan entitas terkait, serta tabel untuk informasi yang lebih detail.



Gambar 5. Tampilan *Dashboard* Visualisasi

4 DISKUSI

Banyak penelitian terdahulu yang telah menggunakan analisis sentimen untuk mendukung keputusan investasi. Penelitian ini berusaha menggabungkan beberapa metode yang telah dilakukan dan juga menambahkan optimasi sehingga hasilnya diharapkan dapat dijadikan sebagai acuan baik bagi para pengambil keputusan atau bagi para peneliti selanjutnya.

Dibandingkan dengan penelitian [5], [7], [15] persamaannya adalah penggunaan analisis sentimen untuk membantu keputusan investasi. Pada penelitian [8], [9], [10] juga dilakukan penambahan metode NER untuk mendeteksi entitas yang terkait. Perbedaannya adalah penelitian tidak menggunakan analisis sentimen pada Twitter, melainkan dari judul berita. Perbedaan selanjutnya adalah penggunaan metode Optuna untuk melakukan optimasi *hyperparameter*. Selain itu, NER yang dilakukan menggunakan metode *rule-based*. Kontribusi dari penelitian ini adalah penggunaan data yang bersumber dari judul berita, sehingga tidak spesifik dari emiten tertentu saja, memungkinkan analisis yang lebih luas menggunakan seluruh populasi data judul berita. Kemudian dari sisi analisis sentimen dan NER, *hyperparameter tuning* menggunakan Optuna untuk algoritma SVM dan pemanfaatan *string similarity* untuk NER pada data judul berita di Indonesia masih belum pernah dilakukan sebelumnya.

5 KESIMPULAN

Penelitian ini telah mengembangkan sebuah sistem pendukung keputusan untuk membantu investor dalam mengambil keputusan investasi saham secara lebih tepat dan terinformasi. Sistem ini mengintegrasikan teknik analisis sentimen dan NER untuk mengekstraksi informasi yang relevan dari berita dan opini terkait perusahaan atau sektor industri tertentu. Berdasarkan eksperimen yang telah dilakukan, model terbaik untuk analisis sentimen dari judul berita merupakan model berbasis SVM dengan nilai rata-rata tertimbang *F1-score* sebesar 76,68%. Optimasi *hyperparameter* menggunakan Optuna berhasil meningkatkan skor menjadi 78,14%. Deteksi entitas berhasil dilakukan menggunakan metode *rule-based* dengan memanfaatkan fungsi Jaro-Winkler. Terdapat peluang untuk mengembangkan penelitian ini lebih lanjut, seperti penerapan algoritma lain contohnya *deep learning* dengan arsitektur Transformer untuk meningkatkan kinerja klasifikasi sentimen, menambahkan data lain seperti data finansial dan data teknikal saham, serta integrasi dengan sumber data atau teknik lain yang relevan. Hasil penelitian ini diharapkan dapat memberikan manfaat praktis dalam dunia investasi saham dan berkontribusi terhadap pengembangan bidang ilmu terkait.

DAFTAR PUSTAKA

[1] K. Singh dan S. S. Narta, "Investor's Considerations Towards Investment Decisions in Stock Market," *International Journal of Advanced Research*, 2020, doi: 10.21474/IJAR01/11906.

[2] M. P. Cristescu, D. A. Mara, R. A. Nerişanu, L. C. Culda, dan I. Maniu, "Analyzing the Impact of Financial News Sentiments on Stock Prices—A Wavelet Correlation," *Mathematics*, vol. 11, no. 23, 2023, doi: 10.3390/math11234830.

[3] K. R. Dahal *dkk.*, "A comparative study on effect of news sentiment on stock price prediction with deep learning architecture," *PLOS ONE*, vol. 18, no. 4, hlm. e0284695, Apr 2023, doi: 10.1371/journal.pone.0284695.

[4] M. N. Ashtiani dan B. Raahemi, "News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review," *Expert Systems with Applications*, vol. 217, hlm. 119509, Mei 2023, doi: 10.1016/j.eswa.2023.119509.

[5] X. Li, P. Wu, dan W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong," *Information Processing & Management*, vol. 57, no. 5, hlm. 102212, Sep 2020, doi: 10.1016/j.ipm.2020.102212.

[6] T. Loughran dan B. Mcdonald, "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *The Journal of Finance*, vol. 66, no. 1, hlm. 35–65, 2011, doi: 10.1111/j.1540-6261.2010.01625.x.

[7] A. E. de Oliveira Carosia, G. P. Coelho, dan A. E. A. da Silva, "Investment strategies applied to the Brazilian stock market: A methodology based on Sentiment Analysis with deep learning," *Expert Systems with Applications*, vol. 184, hlm. 115470, Des 2021, doi: 10.1016/j.eswa.2021.115470.

[8] A. Jabbari, O. Sauvage, H. Zeine, dan H. Chergui, "A French Corpus and Annotation Schema for Named Entity Recognition and Relation Extraction of Financial News," dalam *Proceedings of the Twelfth Language Resources and Evaluation Conference*, N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, dan S. Piperidis, Ed., Marseille, France: European Language Resources Association, Mei 2020, hlm. 2293–2299. Diakses: 11 Mei 2024. [Daring]. Tersedia pada: https://aclanthology.org/2020.lrec-1.279

[9] A. Sinha, S. Kedas, R. Kumar, dan P. Malo, "SEntFiN 1.0: Entity-aware sentiment analysis for financial news," *Journal of the Association for Information Science and Technology*, vol. 73, no. 9, hlm. 1314–1335, 2022, doi: 10.1002/asi.24634.

[10] E. T. Khaing, M. M. Thein, dan M. M. Lwin, "Stock Trend Extraction using Rule-based and Syntactic Feature-based Relationships between Named Entities," dalam *2019 International Conference on Advanced Information Technologies (ICAIT)*, Nov 2019, hlm. 78–83. doi: 10.1109/AITC.2019.8920986.

[11] R. Puspitasari, Y. Findawati, dan M. A. Rosid, "Sentiment Analysis of Post-Covid-19 Inflation Based On Twitter Using the K-Nearest Neighbor and Support Vector Machine Classification Methods," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 4, Art. no. 4, Agu 2023, doi: 10.52436/1.jutif.2023.4.4.801.

[12] S. A. H. Bahtiar, C. K. Dewa, dan A. Luthfi, "Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling," *Journal of Information Systems and Informatics*, vol. 5, no. 3, Art. no. 3, Agu 2023, doi: 10.51519/journalisi.v5i3.539.

[13] P. Nandwani dan R. Verma, "A review on sentiment analysis and emotion detection from text," *Soc. Netw. Anal. Min.*, vol. 11, no. 1, hlm. 81, Agu 2021, doi: 10.1007/s13278-021-00776-6.

[14] L. Barreñada, P. Dhiman, D. Timmerman, A.-L. Boulesteix, dan B. V. Calster, "Understanding overfitting in random forest for probability estimation: a visualization and simulation study," 30 September 2024, *arXiv*: arXiv:2402.18612. doi: 10.48550/arXiv.2402.18612.

[15] N. Afrianto, D. H. Fudholi, dan S. Rani, "Prediksi Harga Saham Menggunakan BiLSTM dengan Faktor Sentimen Publik," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, Feb 2022, Diakses: 11 Mei 2024. [Daring]. Tersedia pada: https://jurnal.iaii.or.id/index.php/RESTI/article/view/3676

[16] T. Akiba, S. Sano, T. Yanase, T. Ohta, dan M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework," dalam *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, dalam KDD '19. New York, NY, USA: Association for Computing Machinery, Jul 2019, hlm. 2623–2631. doi: 10.1145/3292500.3330701.

[17] T. T. Ngoc, C. M. T. Le Van Dai, dan C. M. Thuyen, "Support vector regression based on grid search method of hyperparameters for load forecasting," *Acta Polytechnica Hungarica*, vol. 18, no. 2, hlm. 143–158, 2021.

[18] A. C. Najib, A. Irsyad, G. A. Qandi, dan N. A. Rakhmawati, "Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter," *Fountain of Informatics Journal*, vol. 4, no. 2, Art. no. 2, Nov 2019, doi: 10.21111/fij.v4i2.3573.

[19] H. C. Husada dan A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, 2021, Diakses: 3 Juni 2024. [Daring]. Tersedia pada: https://ejournal.ikado.ac.id/index.php/teknika/article/view/311

[20] R. K. Putri dan M. Athoillah, "Support Vector Machine untuk Identifikasi Berita Hoax Terkait Virus Corona (Covid-19)," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 6, no. 3, Art. no. 3, Okt 2021, doi: 10.30591/jpit.v6i3.2489.

[21] Y. Qi dan Z. Shabrina, "Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach," *Soc. Netw. Anal. Min.*, vol. 13, no. 1, hlm. 31, Feb 2023, doi: 10.1007/s13278-023-01030-x.

[22] C. C. Aggarwal dan C. C. Aggarwal, *Mining text data*. Springer, 2015.

[23] E. I. Setiawan, S. Johanes, A. T. Hermawan, dan Y. Yamasari, "Deteksi Validitas Berita pada Media Sosial Twitter dengan Algoritma Naive Bayes," *INSYST: Journal of Intelligent System and Computation*, vol. 3, no. 2, Art. no. 2, Okt 2021, doi: 10.52985/insyst.v3i2.164.

[24] S. S. dan P. K.v., "Sentiment analysis of malayalam tweets using machine learning techniques," *ICT Express*, vol. 6, no. 4, hlm. 300–305, Des 2020, doi: 10.1016/j.icte.2020.04.003.

[25] T. Winarti, H. Indriyawati, V. Vydia, dan F. W. Christanto, "Performance comparison between naive bayes and k-nearest neighbor algorithm for the classification of Indonesian language articles," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 2, hlm. 452, 2021.

[26] L. Breiman, "Random forests," *Machine learning*, vol. 45, hlm. 5–32, 2001.

[27] M. Rapp, E. Mencía, dan J. Fürnkranz, *Simplifying Random Forests: On the Trade-off between Interpretability and Accuracy*. 2019. doi: 10.48550/arXiv.1911.04393.

[28] X. Chen, T. (Yang H. Cho, Y. Dou, dan B. Lev, "Predicting Future Earnings Changes Using Machine Learning and Detailed Financial Data," 12 Februari 2022, *Social Science Research Network, Rochester, NY*: 3741015. doi: 10.2139/ssrn.3741015.

[29] N. M. Nhat, "Applied Random Forest Algorithm for News and Article Features on The Stock Price Movement: An Empirical Study of The Banking Sector in Vietnam," *Journal of Applied Data Sciences*, vol. 5, no. 3, Art. no. 3, Sep 2024, doi: 10.47738/jads.v5i3.338.

[30] B. Ghojogh dan M. Crowley, *The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial*. 2019. doi: 10.48550/arXiv.1905.12787.

[31] B. Bischl *dkk.*, "Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges," *WIREs Data Mining and Knowledge Discovery*, vol. 13, no. 2, hlm. e1484, 2023, doi: 10.1002/widm.1484.

[32] J. Bergstra, R. Bardenet, Y. Bengio, dan B. Kégl, "Algorithms for Hyper-Parameter Optimization," *Advances in Neural Information Processing Systems*, vol. 24, 2011.

[33] J. Wang, C. Lin, M. Li, dan C. Zaniolo, "Boosting approximate dictionary-based entity extraction with synonyms," *Information Sciences*, vol. 530, hlm. 1–21, Agu 2020, doi: 10.1016/j.ins.2020.04.025.

[34] W. Cohen, P. Ravikumar, dan S. Fienberg, "A comparison of string metrics for matching names and records," dalam *Kdd workshop on data cleaning and object consolidation*, 2003, hlm. 73–78.

[35] O. Rozinek dan J. Mares, "Fast and Precise Convolutional Jaro and Jaro-Winkler Similarity," Mei 2024. doi: 10.23919/FRUCT61870.2024.10516360.

[36] Y. Wang, J. Qin, dan W. Wang, "Efficient Approximate Entity Matching Using Jaro-Winkler Distance," dalam *Web Information Systems Engineering – WISE 2017*, A. Bouguettaya, Y. Gao, A. Klimenko, L. Chen, X. Zhang, F. Dzerzhinskiy, W. Jia, S. V. Klimenko, dan Q. Li, Ed., Cham: Springer International Publishing, 2017, hlm. 231–239. doi: 10.1007/978-3-319-68783-4_16.