Jurnal TICOM: Technology of Information and Communication
Volume 12, Nomor 3, Mei 2024 Hal: 102-107 E-ISSN: 2962-7982

Penerapan Metode SVM dan Random Forest untuk Mendeteksi Berita Hoaks pada PT. Global Arrow

Rizky Purwanto Fernandes[1], Rizky Tahara Shita[2*]

[1,2]Fakultas Teknologi Informasi, Teknik Informatika, Universitas Budi Luhur, Jakarta, Indonesia
Jl. Raya Ciledug, Petukangan Utara, Pesanggrahan, Jakarta Selatan, 12260
E-mail: [1]pfrizky40@gmail.com, [2]*rizky.taharashita@budiluhur.ac.id
(*: corresponding author)

Abstrak—**Berdasarkan statistik aduan konten yang tercatat di *website* kominfo yaitu https://www.kominfo.go.id pada bulan Maret 2022, total laporan mencapai 16.370 laporan dan sejak tanggal 29 Desember 2023, total laporan ada 1.713.103 laporan isu hoaks. Laporan isu hoaks dapat berupa fitnah, penipuan, kekerasan, perdagangan produk dengan aturan khusus, terorisme atau radikalisme, dan sebagainya. Beberapa karakteristik dan dampak berita hoaks atau palsu di Indonesia meliputi adanya peran media sosial seperti facebook, twitter, dan sebagainya. Pengaruhnya terhadap pemilihan umum terutama terkait pemilu pilpres 2024, serta adanya isu hoaks kesehatan terkait COVID-19 yang masih memiliki dampaknya pada tahun 2023 menuju 2024, walaupun tidak seburuk pada tahun 2020 yang lalu. Banyaknya berita hoaks telah membuat masyarakat menjadi enggan divaksinasi karena disebarkan informasi yang tidak akurat dan sejenisnya. Beberapa masalah yang timbul akibat berita hoaks antara lain ketidakpercayaan masyarakat, kekacauan sosial, ketidakstabilan politik, dampak ekonomi, diskriminasi dan perpecahan sosial. Berdasarkan permasalahan yang timbul, potensi dampak yang dapat terjadi, serta beberapa laporan mengenai isu-isu terkait berita hoaks, telah dikembangkan sebuah sistem yang mampu mendeteksi berita hoaks dengan menggunakan berbagai metode, termasuk di antaranya metode TF-IDF. Sebuah metode atau algoritma yang digunakan untuk menghitung kemunculan kata-kata tertentu pada berita asli, hoaks, dan lain sebaginya. Hasil evaluasi meggunakan algoritma *Support Vector Machine* (SVM) dan algoritma *Random Forest* menunjukkan tingkat akurasi di atas 90% pada pengujian pertama yang menunjukkan tingkat akurasi yang tinggi. Namun setelah dilakukan uji coba kedua, terjadi penurunan skor akurasi menjadi sekitar 55%. Meskipun demikian, hasil evaluasi tersebut menunjukkan bahwa model memeberikan prediksi yang cukup baik dengan performa sedang.**

Kata Kunci—**statistik, algoritma, berita, laporan, skor, akurasi.**

Abstract—***Based on the content complaint statistics recorded on the Kominfo website at https://www.kominfo.go.id, in March 2022, there were a total of 16,370 reports, and since December 29, 2023, the total number of reports regarding hoax issues reached 1,713,103. Hoax reports can involve defamation, fraud, violence, trade in products with specific regulations, terrorism, or radicalism, among others. Some characteristics and impacts of fake news or hoaxes in Indonesia include the role of social media platforms such as Facebook, Twitter, and others. Their influence on general elections, especially regarding the 2024 presidential election, and the existence of health-related hoaxes concerning COVID-19 still have effects from 2023 to 2024, although not as severe as in 2020. The abundance of hoax news has made people reluctant to get vaccinated due to the dissemination of inaccurate information and the like. Some problems arising from hoax news include public distrust, social unrest, political instability, economic impacts, discrimination, and social division. Based on the issues that arise, the potential impacts, and several reports regarding hoax-related issues, a system has been developed capable of detecting hoax news using various methods, including the TF-IDF method. TF-IDF is a method or algorithm used to calculate the frequency of occurrence of specific words in genuine news, hoaxes, and others. Evaluation results using the Support Vector Machine (SVM) algorithm and the Random Forest algorithm showed accuracy rates above 90% in the initial testing phase, indicating a high level of accuracy. However, after the second trial, there was a decrease in accuracy scores to around 55%. Nevertheless, these evaluation results indicate that the model provides predictions that are quite accurate with moderate performance***

Keyword— ***statistics, algorithms, news, reports, scores, accuracy***

I. PENDAHULUAN

Latar belakang penelitian ini adalah dipicu oleh meningkatnya jumlah berita yang tersebar setiap harinya, baik melalui situs web, sosial media, dan berbagai platform lainnya [1], [2], [3]. Banyaknya berita tersebut terkadang membuat kita sulit untuk menentukan apakah berita terkait adalah berita terpercaya atau bisa dikatakan hoaks atau menyesatkan [4].

Berdasarkan statistik bulan Maret 2022 di *website* Kominfo [5], terdapat total laporan isu hoaks sejumlah 16.370 laporan sejak diperiksa kembali pada hari Jumat tanggal 29 Desember 2023 dan statistik keseluruhan berjumlah 1.713.103 laporan isu hoaks [6]. Ini merupakan angka yang besar, dimana jika terdapat hanya satu berita hoaks yang tersebar itu bisa menyebabkan perubahan persepsi masyarakat terkait topik yang diangkat berita tersebut[7]. Beberapa laporan isu berita hoaks di *website* Kominfo diklasifikasikan sebagai berikut, yaitu pornografi, perjudian, fitnah, penipuan, sara, kekerasan/kekerasan pada anak, perdagangan produk dengan aturan khusus, terorisme/radikalisme, separatisme/organisasi berbahaya, HKI, pelanggaran keamanan Informasi, Konten negatif yang direkomendasikan Instansi sektor [5], [6]. Konten yang meresahkan masyarakat, konten yang melanggar nilai sosial dan budaya, berita bohong / hoaks, pemerasan, konten yang memfasilitasi diaksesnya konten negatif, dan normalisasi [7]. Berikut terdapat beberapa statistik dari *website* Kominfo terkait berita hoaks [1], [2], [3], [6]. Statistik keseluruhan dari

website Kominfo ditunjukkan Gambar 1, temuan isu hoaks dari *website* Kominfo pada Gambar 2, temuan isu hoaks per kategori dari *website* Kominfo di Gambar 3, rekapitulasi isu hoaks pemilu dari *Website* Kominfo pada Gambar 4, dan penanganan sebaran isu hoaks pemilu dari *Website* Kominfo dapat dilihat pada Gambar 5.

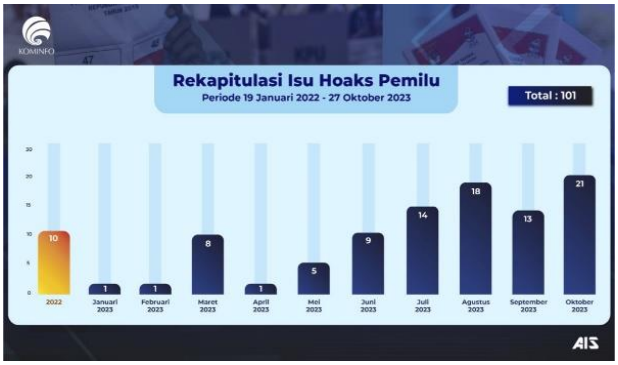
Statistik Keseluruhan

Pornografi	1.142.010
Perjudian	540.410
Fitnah	17
Penipuan	16.461
Sara	189
Kekerasan / Kekerasan Pada Anak	13
Perdagangan Produk Dengan Aturan Khusus	127
Terorisme / Radikalisme	521
Separatisme / Organisasi Berbahaya	5
HKI	9.400
Pelanggaran Keamanan Informasi	325
Konten Negatif yang direkomendasikan Instansi Sektor	4.834
Konten yang meresahkan masyarakat	23
Konten yang melanggar nilai sosial dan budaya	26
Berita Bohong / HOAKS	21
Pemerasan	0
Konten yang memfasilitasi diakses knya konten negatif	0
Normalisasi	1.279
Total	1.713.103

Gambar 1. Statistik Keseluruhan dari *Website* Kominfo

Gambar 2. Temuan Isu Hoaks dari *Website* Kominfo

Gambar 3. Temuan Isu Hoaks Per Kategori dari *Website* Kominfo



Gambar 4. Rekapitulasi Isu Hoaks Pemilu dari *Website* Kominfo

Gambar 5. Penanganan Sebaran Isu Hoaks Pemilu dari *Website* Kominfo

A. Tujuan

Tujuan dari penelitian ini adalah mengembangkan sebuah sistem pendeteksi berita hoaks yang dapat membantu dalam memilih berita yang hoaks atau tidak hoaks (berita kredibel, terpercaya, dan benar). Sistem ini bertujuan untuk memudahkan proses verifikasi dan klarifikasi terhadap berita yang dipertanyakan kebenarannya. Meskipun tingkat akurasi sistem tidak mencapai 100, sistem ini diharapkan dapat memberikan kemudahan dalam memilah berita yang hoaks dan tidak hoaks di tengah jumlah berita yang sangat besar dan beragam yang tersebar di berbagai platfrom media informasi setiap harinya [6].

B. Identifikasi Masalah

Adapun rumusan masalah terkait topik penelitian ini yaitu bagaimana cara untuk melakukan pemilahan terhadap berita yang muncul dalam jumlah yang sangat besar setiap harinya berdasarkan kebenaran isi berita? Apa saja informasi yang dapat diperoleh dari berita yang dikumpulkan, sehingga dapat memberikan wawasan, pengetahuan, dan pemahaman yang relevan terkait dengan berita yang diperoleh dan muncul?

C. Metode Penelitian

Berbagai algoritma yang diujicoba sebelum dilakukannya pemilihan model algoritma yang sesuai diantaranya:

1) *Algoritma SVM (Support Vector Machine):* Algoritma SVM digunakan untuk klasifikasi dan juga dapat digunakan untuk regresi atau deteksi anomali. Memiliki prinsip kerja dengan mencari *hyperlane* terbaik yang memisahkan dua kelas

dalam ruang fitur. *Hyperlane* ini dipilih agar memiliki margin dari titik-titik terdekat dari kedua kelas [8], [9], [10], [11].

2) *Algoritma Random Forest Classification:* Algoritma *Random Forest* digunakan untuk klasifikasi dan regresi. Dalam konteks klasifikasi, setiap pohon (*tree*) di dalam hutan (*forest*) memberikan suara untuk kelas tertentu, dan kelas dengan suara terbanyak menjadi predikis akhir. Algoritma *Random Forest* adalah *ensemble learning* yang terdiri dari banyak pohon keputusan memiliki prinsip kerja, Setiap pohon (*tree*) dihasilkan dari subset acak dari data dan fitur. Prediksi akhir diambil berdasarkan mayoritas suara dari semua pohon (*tree*) [8], [9], [10], [12].

Metode yamg digunakan pada topik ini adalah metode TF-IDF (*Term Frequency-Inverse Document Frequency*). TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah sebuah metode dalam pengolahan bahasa alami dan pengelompokan dokumen yang digunakan untuk mengevaluasi seberapa penting suatu kata dalam suatu dokumen terhadap sebuah koleksi dokumen. Metode ini umumnya digunakan dalam ekstraksi fitur dan pemodelan teks [13].

D. Penelitian Sebelumnya

Pada penelitian sebelumnya dilakukan *web scrapping* menggunakan bahasa pemrograman python dengan *library* beautifulsoup dengan target *website* sumber berita Indonesia untuk mendapatkan artikel berita terkait jadwal dan hari besar dan hari libur di Indonesia. *Library* beautifulsoup baik digunakan untuk melakukan metode *web scrapping* dimana data yang diambil dalam bentuk HTML, JSON, atau XML. Jika *website* yang akan dilakukan metode *web scrapping* lebih kompleks maka digunakan beberapa metode lainnya.

II. METODE PENELITIAN

A. Analisis Kebutuhan

1) *Langkah-langkah Analisis Kebutuhan:* Identifikasi tantangan dan risiko berdasarkan statistik aduan konten secara keseluruhan, termasuk berita pornografi, perjudian, penipuan, dan hoaks, yang dapat menyebabkan keresahan pada masyarakat. Dengan adanya pemilu pilpres 2024, perlu dilakukan antisipasi terhadap dampak dari berita palsu, termasuk pengaruhnya pada keputusan dan stabilitas politik.

2) *Pemetaan Platform Media Sosial atau Informasi:* meliputi identifikasi platform sumber berita seperti situs web Kompas [14], Antara News [15], dan berita dari instansi terkait seperti PT. Global Arrow [16], yang akan digunakan sebagai sampel untuk analisis berita palsu atau hoaks.

3) *Analisis Data dan Konten Berita Hoaks:* melibatkan identifikasi karakteristik konten berita palsu atau hoaks dalam bentuk teks, dengan menggunakan analisis data dan metode penambangan teks untuk memahami pola dan ciri khas berita hoaks.

4) *Pengumpulan Data Sumber Berita:* melibatkan pengumpulan data yang mencakup nama sumber berita, status, dan tautan yang diperoleh dari berbagai sumber berita yang telah diidentifikasi sebelumnya.

B. Pengumpulan Data

Beberapa metode pengumpulan data yaitu:

1) *Web Crawler dan Web Scrapping:* Pengumpulan data web vrawler dan *web scrapping* menggunakan bahasa pemrograman python dan hasil data yang dikumpulkan dan dismpan dalam bentuk excel yaitu dengan format xlsx atau csv dan dalam format JSON [17], [18]. Data juga disimpan dalam sistem basis data NoSQL (Not Only SQL) yaitu MongoDB.

2) Data Berita Diperoleh Secara Manual dari Situs Web dan Sebagainya*:* Metode ini merupakan cara untuk mengambil berita secara manual tanpa menggunakan *web scrapping* atau *web crawler*. Metode ini digunakan untuk uji coba atau evaluasi model.

Pengambilan data dilakukan melalui *command prompt* atau terminal, serta melalui antarmuka pengguna grafis (GUI) yang telah disiapkan. Setiap *web crawler* menjalani proses pengambilan data yang telah ditentukan. Berikut adalah tautan yang digunakan dalam proses tersebut, serta jumlah artikel dari setiap sumber berita [5], [14], [15], [16], [19], [20]. Tautan sumber berita ditampilkan pada Tabel 1, dan jumlah artikel setiap sumber berita pada Tabel 2.

TABEL 1
TAUTAN SUMBER BERITA

Sumber Berita	Tautan Awal
antara news	https://www.antaranews.com/tag/berita-hoax https://www.antaranews.com/tag/asli
jppn	https://www.jpnn.com/tag/hoax https://www.jpnn.com/tag/asli
kominfo	https://www.kominfo.go.id
kompas	https://www.kompas.com/tag/hoax https://www.kompas.com/tag/asli
tribunnews	https://www.tribunnews.com/tag/hoax https://www.tribunnews.com/tag/asli
globalarrow	https://globalarrow.co.id/id/newst

TABEL 2
JUMLAH ARTIKEL SETIAP SUMBER BERITA

Sumber Berita	Tanggal	Asli	Hoax	Total artikel
antara news	10 Januari 2024	1393	387	1780
jppn	10 Januari 2024	0	41	41
kominfo	-	0	10	10
kompas	10 Januari 2024	901	578	1479
tribunnews	10 Januari 2024	1	198	199
globalarrow	12 Januari 2024	13	0	13

C. Pemrosesan Data

Pada tahap ini pemrosesan data dilakukan saat melakukan *web crawler* atau *web scrapping* dan pada saat melakukan analisis. Pada saat melakukan *web crawler* atau *web scrapping*, pemrosesan data dilakukan di bagian *pipelines* dan pada saat melakukan analisis, dilakukannya pemrosesan data berupa data *cleaning* dan *feature engineering* untuk menyiapkannya saat dilakukan analisis. Tahapan yang digunakan pada tahap ini adalah sebagai berikut:

1) *Data Cleaning:* Cara atau teknik pembersihan data seperti mengatasi *missing value* atau *outliers*. Beberapa hal seperti tokenisasi, penghapusan *stop words* dan *stemming*.

2) *Feature Engineering:* Mengekstrak fitur-fitur yang relevan dengan data. Dalam hal ini fitur yang dipertimbangkan adalah artikel teks dan judul berita.

Metode yang digunakan yaitu ETL (*Extract, Transform, Load*). Mengimplementasikan proses ETL untuk memproses dan mempersiapkan data. Proses setelah data dibersihkan dan disiapkan, data disimpan ke dalam berbagai format dan sistem basis data.

D. Pembagian Dataset

Pembagian dataset dilakukan menjadi dua bagian, yaitu data pelatihan dan data pengujian. Sebanyak 70% dari total 1974 records data dialokasikan untuk data pelatihan, yang setara dengan 1381 records, sementara sisanya, sebanyak 30%, digunakan untuk data pengujian, yang mencakup 593 *records*.

E. Pemilihan dan Pengembangan Model

Pada tahap pemilihan dan pengembangan model atau algoritma, disesuaikan dengan algoritma yang memiliki dan menunjukkan tingkat akursai yang lebih tinggi. Berbagai algoritma yang diujicoba sebelum dilakukannya pemilihan model algoritma yang sesuai diantaranya:

1) *Algoritma SVM (Support Vector Machine):* Algortima SVM digunakan untuk klasifikasi dan juga dapat digunakan untuk regresi atau deteksi anomali [8], [9], [10], [11].

2) *Algoritma Random Forest Classification:* Algoritma *Random Forest* digunakan untuk klasifikasi dan regresi. Dalam konteks klasifikasi, Setiap pohon (tree) di dalam hutan (forest) memberikan suara untuk kelas tertentu, dan kelas dengan suara terbanyak menjadi predikis akhir [8], [9], [10], [12].

F. Evaluasi dan Optimisasi Model

Evaluasi dan peningkatan pada model yang digunakan untuk memprediksi berita palsu yang dilakukan sebagai berikut:

1) *Evaluasi Model Berdasarkan Akurasi Skor Algoritma. Menggunakan SVM dan Random Forest untuk Klasifikasi:* Hasil evaluasi, model dievaluasi menggunakan metrik akurasi menggunakan algoritma SVM dan *Random Forest* untuk klasifikasi. Saran untuk perbaikan, Pertimbangkan untuk melihat metrik evaluasi lainnya seperti presisi, recall, f1-score, dan area di bawah kurva ROC (AUC-ROC) untuk mendapatkan pemahaman yang lebih komprehensif tentang kinerja model.

2) *Evaluasi Input Artikel Berita Secara Manual:* Hasil evaluasi, artikel berita dari berbagai sumber dikumpulkan dan diinputkan secara manual melalui *form input* GUI. Model dievaluasi terhadap artikel yang dimasukkan untuk memeriksa apakah model memberikan prediksi yang sesuai dengan label atau fakta sebenarnya. Saran untuk perbaikan, perlu memvalidasi dan menginterpretasikan hasil model secara lebih mendalam untuk memastikan prediksi yang lebih akurat dan konsisten.

3) *Visualisasi data:* Hasil evaluasi, data dan hasil prediksi divisualisasikan menggunakan matplotlib untuk mempermudah pemahaman. Saran untuk perbaikan, memastikan visualisasi data dapat memberikan wawasan yang jelas dan mudah dimengerti.

4) *Pertimbangan untuk Menambahkan Fitur Judul:* Hasil evaluasi, dipertimbangkan untuk menambah fitur judul untuk meningkatkan akurasi model prediksi. Saran untuk perbaikan, perlu dilakukan penelitian lebih lanjut untuk memastikan bahwa penambahan fitur judul dapat memberikan peningkatan yang signifikan dalam kinerja model.

5) *Optimisasi Model:* Melakukan penyesuaian hyperparameter untuk memperbaiki kinerja model. Hasil evaluasi, Melakukan penyesuaian hyperparameter dan analisis fitur untuk meningkatkan kinerja model. Saran untuk perbaikan, selalu mempertimbangkan peningkatan data dengan meningkatkan jumlah data serta memperhatikan tuning parameter pada algoritma yang digunakan untuk meningkatkan kinerja model secara keseluruhan.

Dengan memperhatikan evaluasi yang komprehensif dan saran untuk perbaikan tersebut, diharapkan dapat membangun model yang lebih baik, tangguh, akurat dan dapat diandalkan.

G. Pengembangan Antarmuka Pengguna (GUI)

GUI dibuat menggunakan tkinter yang merupakan bawaan dari python. Tkinter digunakan untuk membuat antarmuka pengguna grafis (GUI). Dengan menggunakan tkinter, bisa membuat jendela (*windows*), tombol, label, input teks dan elemen GUI lainnya. GUI ini digunakan untuk menampilkan notifikasi jika proses *crawler* selesai.

Web GUI atau GUI berbasis web menggunakan streamlit yang merupakan sebuah perpustakaan (*library*) dalam Bahasa pemrograman python yang memungkinkan pengembang dengan cepat membuat antarmuka (UI) web sederhana untuk aplikasi data dan prototipe interaktif. Web GUI ini digunakan untuk proses uji coba dan validasi artikel berita [21].

H. Visualisasi Data

Visualisasi data menggunakan matplotlib yang merupakan perpustakaan (*library*) dalam bahasa pemrograman python yang digunakan untuk membuat visualisasi grafik dua dimensi dan tiga dimensi. Dengan matplotlib, dapat membuat berbagai jenis plot, grafik garis, *scatter plot*, histogram, *bar chart*, dan jenis visualisasi lainnya [22].

I. Pengujian Sistem

Pengujian sistem dilakukan dengan 30% dari total *dataset* yang sudah dilakukan pra-pemrosesan untuk dilakukan uji coba model. Uji coba yang lainnya adalah secara manual menguji apakah hasil yang ditampilkan sesuai yang diharapkan dalam hal ini apakah hasil yang muncul merupakan berita fakta atau berita palsu.

J. Validasi dan Interpretasi Model

Melakukan valuasi model dengan menggunakan *dataset* yang berbeda dalam hal ini adalah 30% dari total *dataset* yang sudah dilakukan pra-pemrosesan dan menggunakan metrik evaluasi lainnya seperti presisi, *recall*, f1-*score*, dan area di bawah kurva ROC (AUC-ROC) disesuaikan dengan kebutuhan spesifik pada topik ini.

K. Dokumentasi

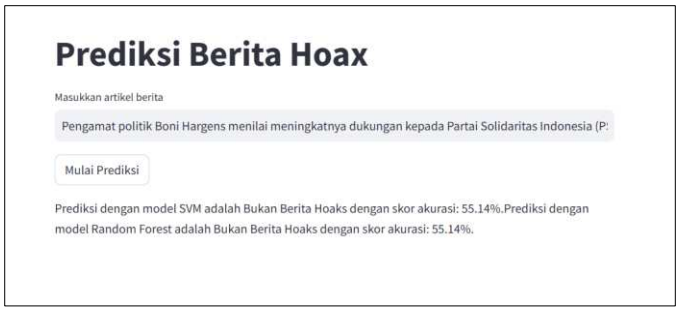
Untuk dokumentasi penggunaan sistem terdapat pada *file readme* salah satunya informasi dan penggunaan *web crawler*.

III. HASIL DAN PEMBAHASAN

A. *Pengujian Prediksi Berita*

Jika Pengujian dilakukan dengan model yang sudah dilakukan tahap evaluasi, setiap artikel berita yang dimasukkan pengguna akan menghasilkan pesan yang berbeda. Jika pengguna tidak memasukkan artikel berita dan menekan tombol untuk melakukan prediksi maka akan muncul pesan "Masukkan Artikel Berita". Jika hasil prediksi dari model SVM (*Support Vector Machine*) adalah berita hoaks, maka akan muncul pesan "Prediksi dengan model SVM adalah Berita Hoaks dengan skor akurasi: 55.14%. Jika hasil prediksi dari model SVM (*Support Vector Machine*) adalah bukan berita hoaks makan akan muncul pesan "Prediksi dengan model SVM adalah Bukan Berita Hoaks dengan skor akurasi: 55.14%. Jika hasil prediksi dari model Random Forest adalah berita hoaks, maka akan muncul pesan "Prediksi dengan model Random Forest adalah Berita Hoaks dengan skor akurasi: 55.14%. Jika hasil prediksi dari model *Random Forest* adalah berita bukan hoaks maka akan muncul pesan "Prediksi dengan model *Random Forest* adalah Bukan Berita Hoaks dengan skor akurasi: 55.14%."

Gambar 6 adalah tampilan hasil prediksi pada aplikasi web deteksi berita hoaks berbasis *machine learning*.



Gambar 6. Tampilan Hasil Prediksi Pada Aplikasi Web Prediksi Berita Hoaks

B. *Analisa Dalam Penjabatan Pengujian*

1) *Analisa Dalam Penjabaran Pengujian Pada Web Crawler:* Dari data yang diperoleh, sumber berita dari AntaraNews dan Kominfo telah digunakan karena keduanya menawarkan jumlah artikel yang signifikan, termasuk artikel dengan status asli dan hoaks. AntaraNews memiliki total 1780 artikel, di mana 1393 di antaranya teridentifikasi sebagai artikel asli dan 387 di antaranya sebagai artikel hoaks. Sementara itu, Kompas menyediakan total 1479 artikel, dengan 901 artikel asli dan 578 artikel hoaks.

2) *Analisa Dalam Penjabaran Pengujian Pada Prediksi Berita:* Pada pengujian pertama didapatkan skor akurasi untuk kedua lebih dari 90%, namun setelah diuji coba manual apa yang dihasilkan tidak sesuai dengan skor akurasi yang baik yaitu 90%. Sehingga perlu dilakukan evaluasi dan pada uji coba kedua dilakukan tahap pra-pemrosesan lagi pada artikel berita walaupun hasil skor akurasinya menurun menjadi kurang lebih 55.14%. Dari hasil pelatihan dan pengujian model yang kedua didapat hasil prediksi adalah 55.14% menunjukkan prediksi cukup baik dalam memprediksi. Memiliki model dengan performa cukup baik. Walaupun model ini memiliki skor prediksi 55.14% dan memiliki prediksi yang cukup baik, perlu untuk dilakukan penyesuaian agar mendapatkan model prediksi dengan skor akurasi yang labih baik.

C. *Evaluasi*

Berikut merupakan evaluasi dari solusi dari program yang ada:

1) *Kelebihan:* Memiliki tingkat akurasi tinggi. Kelebihan dari program ini adalah memiliki tingkat akurasi model yang cukup tinggi yaitu diatas 90%. Dengan tingkat akurasi yang tinggi diharapkan, hasil prediksi juga bisa memberikan hasil prediksi yang akurat. Program atau sistem prediksi berita bisa digunakan di web secara publik. Program ini bisa diakses secara publik dengan menuju ke tautan terkait. Dengan menuju ke tautan terkait, siapa saja bisa menggunakan program prediksi tersebut.

2) *Kekurangan:* Dataset kurang banyak dan bervariasi. Salah satu kekurangan dari program ini mungkin kurangnya variasi dataset dari berbagai sumber berita. Dalam penelitian ini tidak digunakan seluruh data dari berbagai sumber berita karena adanya keterbatasan seperti tidak relevannya berita. Tidak adanya respon yang sesuai saat melakukan web scrapping atau *web crawling*, sehingga respon data yang diterima tidak bisa diproses dan disimpan di dalam *dataset* di dalam *database* maupun file lainnya dengan format csv atau json. *Dataset* tidak diperbarui secara otomatis dan *realtime*. *Dataset* diperbarui secara manual agar tetap relevan dengan berita terkini. Hasil prediksi terkadang tidak benar.

IV. PENUTUP

Secara keseluruhan, program prediksi berita palsu atau hoaks bertujuan untuk memilah dan mengidentifikasi berita yang dapat dipastikan sebagai hoaks atau tidak. Pada tahap pelatihan, pengujian, dan prediksi pertama, evaluasi dan klarifikasi terhadap hasil prediksi menjadi langkah penting setelah dilakukan prediksi terhadap kebenaran berita. Meskipun algoritma seperti SVM (*Support Vector Machine*) dan klasifikasi *Random Forest* mampu mencapai skor prediksi di atas 90%, masih terdapat sekitar 10% hasil prediksi yang mungkin salah. Pada uji coba model kedua, meskipun skor akurasi mencapai sekitar 55%, model prediksi menunjukkan performa yang cukup baik. Tahap pra-pemrosesan yang diperdalam pada fitur teks artikel berita merupakan salah satu upaya untuk meningkatkan akurasi model. Proses dimulai dengan melakukan *crawling* data berita dari berbagai sumber, seleksi sumber berita yang relevan, penyimpanan hasil *crawling*, dan penerapan metode seperti TF-IDF. Langkah-langkah ini kemudian diikuti dengan pemodelan menggunakan algoritma SVM dan *Random Forest*, evaluasi model, serta implementasi dalam sistem berbasis GUI berbasis web. Meskipun masih terdapat kekurangan seperti dataset yang tidak ter-update secara otomatis, diharapkan adanya masukan dan kritik membangun untuk pengembangan program atau sistem yang lebih baik di masa mendatang.

REFERENSI

[1] Kominfo. [Online]. Available:https://www.kominfo.go.id/content/detail/52570/siaran-pers-no-422hmkominfo102023-tentang-menkominfo-isu-hoaks-pemilu-meningkat-hampir-10-kali-lipat/0/siaran_pers.
[2] Infopublik. [Online]. Available: https://www.infopublik.id/kategori/nasional-sosial-budaya/729879/kominfo-identifikasi-425-isu-hoaks-di-triwulan-pertama-2023
[3] Kominfo. [Online]. Available:https://www.kominfo.go.id/content/detail/48363/siaran-pers-no-50hmkominfo042023-tentang-triwulan-pertama-2023-kominfo-identifikasi-425-isu-hoaks/0/siaran_pers
[4] A. H. Subarjo and W. Setianingsih, "Literasi Berita Hoaxs Di Internet Dan Implikasinya Terhadap Ketahanan Pribadi Mahasiswa (Studi Tentang Penggunaan Media Sosial Pada Mahasiswa STT Adisutjipto Yogyakarta)," Jurnal Ketahanan Nasional, vol. 26, p. 1, Apr. 2020.
[5] kominfo. [Online]. Available: https://www.kominfo.go.id/
[6] kominfo. [Online]. Available: https://www.kominfo.go.id/statistik
[7] T. N. Faturohmah and T. A. S. Salim, "Perilaku Masyarakat Terhadap Penyebaran Hoax Selama Pandemi Covid-19 melalui Media di Indonesia: Tinjauan Literatur Sistematis", *Tik Ilmeu: Jurnal Ilmu Perpustakaan dan Informasi*, vol. 6, no. 1, pp. 121-138, 2022.
[8] M. Fachriza and M. Munawar, "Analisis Sentimen Kalimat Depresi pada Pengguna Twitter dengan Naive Bayes, Support Vector Machine, Random Forest", *KOMPUTEK*, vol. 7 no. 2, pp. 49-58, 2023.
[9] H. Syahputra and A. Wibowo, "Comparison of Support Vector Machine (SVM) and Random Forest Algorithm for Detection of Negative Content on *Website* s", *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no.1, pp. 165–173, 2023.
[10] M. R. Adrian, et al., "Perbandingan Metode Klasifikasi Random Forest dan SVM pada Analisis Sentimen PSBB", *Jurnal Informatika Upgris*, vol. 7, no.1, pp. 36-40,2021.
[11] M. K. Tamami and I. Kharisudin, "Komparasi Metode Support Vector Machine dan Naive Bayes Classifier untuk Pemodelan Kualitas Pengajuan Kredit", *Indonesian Journal of Mathematics and Natural Sciences*, vol. 46, no. 1, pp. 38-44 2023.
[12] A. R. Masdian, N. Bashit, and F. Hadi, "Analisis Produktivitas Padi Menggunakan Algoritma Machine Learning Random Forest di Kabupaten Batang Tahun 2018 - 2022", *Jurnal Geodesi dan Geomatika*, vol. 06, no. 01, pp. 43-51, 2023.
[13] S. A. Helmayanti, F. Hamami, and R. Y. Fa'rifah, "Penerapan Algoritma Tf-Idf dan Naïve Bayes untuk Analisis Sentimen Berbasis Aspek Ulasan Aplikasi Flip pada Google Play Store", *Jurnal Indonesia: Manajemen Informatika dan Komunikasi (JIMIK)*, vol. 4, no.3, pp. 1822–1834, Sep. 2023.
[14] kompas. [Online]. Available: https://www.kompas.com/
[15] antaranews. [Online]. Available: https://www.antaranews.com/
[16] globalarrow. [Online]. Available: https://globalarrow.co.id/id/news/
[17] V. A. Flores, P. A. Permatasari, and L. Jasa, "Penerapan Web Scraping Sebagai Media Pencarian dan Menyimpan Artikel Ilmiah Secara Otomatis Berdasarkan Keyword," *Majalah Ilmiah Teknologi Elektro*, vol. 19 no. 2, pp. 157-162, 2020.
[18] I. Y. Prabhaswara, I. M. A. D. Suarjaya, and N. K. D. Rusjayanthi, "Pengembangan Engine Web Crawler Sebagai Pencari Jejak Serangan Cyber Stored Cross-Site Scripting," *JITTER- Jurnal Ilmiah Teknologi dan Komputer*, vol. 4, no.2, 2023.
[19] jpnn. [Online]. Available: https://www.jpnn.com/
[20] tribunnews. [Online]. Available: https://www.tribunnews.com/
[21] G. Chudra, A. Yohannis, and R. Setiawan, "Development of Streamlit-Based Higher Education Ranking Instrument Boards," *JUSIKOM PRIMA (Jurnal Sistem Informasi dan Ilmu Komputer Prima)*, vol. 7, 2023.
[22] "A. H. Sial, S. Y. S. Rashdi, A. H. Khan, "Comparative Analysis of Data Visualization Libraries Matplotlib and Seaborn in Python," International *Journal of Advanced Trends in Computer Science and Engineering*, vol. 10, no. 1, pp. 277–281, 2021.