

## Deteksi Berita Palsu Pada Pemilu 2024 Dengan Menggunakan Algoritma Random Forest

Arizal Sabila Nurhikam<sup>1</sup>, Rivaldi Syaputra<sup>2</sup>, Saepul Rohman<sup>3</sup>, Satrio Rully Priyambodo<sup>4</sup>, Nova Agustina<sup>5</sup>

<sup>1,2,3,4</sup>Universitas Adhirajasa Reswara Sanjaya

<sup>5</sup>Sekolah Tinggi Teknologi Bandung

email: saepulrohman345@gmail.com

**Abstract:** Elections or general elections are an important process in democratic countries used to choose leaders and governments. However, the widespread spread of fake news on social media causes doubts about the information obtained. Therefore, techniques are needed to detect fake news before the news is received and believed by the public. In this research, we evaluate the ability of the random forest algorithm in detecting fake news in the 2024 elections. The dataset used contains 859 records containing id, fact\_status, and news\_title, where fact\_status is the label. The processes carried out on the dataset include cleaning, tokenization, and stemming. Then, training and testing processes are carried out on the model to obtain the accuracy of the model. The results obtained from this test are an accuracy of 84.88%. The implementation of cleaning, tokenization, and stemming on the dataset also improves the quality of the dataset used in the classification process.

**Keywords:** Fake News, General Election 2024, Random Forest Algorithm, News Detection

**Abstrak:** Pemilu atau pemilihan umum merupakan proses penting dalam negara demokrasi yang digunakan untuk memilih pemimpin dan pemerintah. Namun, maraknya berita palsu yang beredar di media sosial menyebabkan keraguan terhadap informasi yang didapat. Oleh karena itu, dibutuhkan teknik untuk mendeteksi berita palsu sebelum berita tersebut diterima dan dipercayai oleh masyarakat. Dalam penelitian ini, kami mengevaluasi kemampuan algoritma random forest dalam mendeteksi berita palsu pada pemilu 2024. Dataset yang digunakan berisikan 859 record yang berisi id, status\_fakta, dan judul\_berita, dimana status\_fakta sebagai label. Proses yang dilakukan pada dataset meliputi cleaning, tokenisasi, dan stemming. Kemudian dilakukan proses training dan testing pada model untuk mendapatkan akurasi dari model tersebut. Hasil yang didapatkan dari pengujian ini adalah akurasi sebesar 84.88%. Penerapan cleaning, tokenisasi, dan stemming pada dataset juga meningkatkan kualitas dataset yang digunakan dalam proses klasifikasi.

**Kata kunci:** Berita Palsu, Pemilihan Umum 2024, Algoritma Random Forest, Deteksi Berita

### Pendahuluan

Pemilu atau pemilihan umum adalah proses penting dalam negara demokrasi yang digunakan untuk memilih pemimpin dan pemerintah. Di Indonesia, pemilu diakui sebagai pilar penting dalam demokrasi dan diatur dalam UUD NRI Tahun 1945. Pemilu harus dilaksanakan secara demokratis (Ansori, 2017). Namun, saat ini maraknya berita palsu yang beredar di media sosial menyebabkan keraguan terhadap informasi yang didapat. Media sosial saat ini diwarnai dengan berita palsu, fitnah, dan sikap intoleran. Teknologi yang maju membuat

informasi sangat cepat beredar di internet. Kondisi ini menyebabkan persebaran informasi yang tidak terverifikasi dengan sangat cepat melalui media sosial. Banyak informasi yang disebar setiap hari dan orang sering belum sempat memahami informasi sebelum bereaksi (Juliswara, 2017). Oleh karena itu, dibutuhkan teknik untuk mendeteksi berita palsu sebelum berita tersebut diterima dan dipercayai oleh masyarakat.

Berita palsu merupakan suatu hal yang sangat merugikan bagi orang yang mudah percaya, karena dapat merusak karakter seseorang dan menimbulkan

dampak negatif dalam jangka panjang. Berita palsu dapat mempengaruhi pemikiran, tindakan, atau keputusan individu atau kelompok yang menyebabkan pemahaman yang salah tanpa disadari, sehingga sangat penting untuk selalu berhati-hati dan mengecek kebenaran berita sebelum mempercayainya (Yunanto et al., 2021). Oleh karena itu, penting untuk menangani masalah ini dengan cara yang efektif. Penelitian sebelumnya telah menunjukkan bahwa algoritma pembelajaran mesin dapat digunakan untuk mendeteksi dan memfilter informasi yang tidak valid (Yonathan et al., 2022). Namun, dalam penelitian ini, kami akan fokus pada kemampuan algoritma random forest dalam mendeteksi berita palsu pada pemilu 2024.

Dalam penelitian ini, kami akan mengevaluasi kemampuan algoritma Random Forest dalam mendeteksi berita palsu pada Pemilu 2024. Kami akan mengolah dataset yang berisikan 859 records. Pemilihan dataset yang digunakan dalam penelitian didasarkan pada beberapa kriteria penting, diantaranya adalah relevansi data dengan topik penelitian dan jumlah data yang cukup banyak. Dengan memperhatikan kedua faktor tersebut, diharapkan dapat memperoleh hasil yang akurat dan valid dalam penelitian tersebut. Dalam penelitian yang telah dilakukan, para peneliti mencoba untuk mengatasi masalah menangani penyebaran berita palsu dengan cara melakukan pemeriksaan kebenaran berita untuk memastikan apakah berita tersebut asli atau palsu. Para peneliti menggunakan model supervised learning random forest dan logistic regresi. Model random forest menghasilkan tingkat akurasi sebesar 84%, yang lebih baik dibandingkan dengan model logistic regresi yang hanya menghasilkan tingkat akurasi sebesar 77% (Ramadhan et al., 2022). Penelitian lain membahas tentang masalah berita palsu di media sosial

dan mengimplementasikan algoritma klasifikasi untuk membedakan konten hoax dengan konten fakta. Dalam penelitian tersebut, algoritma random forest digunakan untuk mengklasifikasi berita Bengali sebagai berita palsu atau tidak dan menghasilkan tingkat akurasi sebesar 85%. Sistem ini juga memiliki antarmuka web untuk memeriksa validitas berita dalam bahasa Bengali (Sgurev et al., 2020).

Penelitian ini bertujuan untuk mengembangkan model klasifikasi berita palsu pada pemilu 2024 menggunakan algoritma Random Forest. Penelitian ini akan menggunakan dataset yang berisikan 859 record. Diharapkan bahwa hasil dari penelitian ini dapat memberikan solusi atau jawaban untuk menangani masalah berita palsu pada pemilu yang dapat membantu masyarakat dalam menentukan informasi yang benar dan dapat dipercaya.

### Studi Literatur

Pada penelitian terdahulu, deteksi berita hoax telah menjadi perhatian yang dilakukan oleh (Ratkiewicz et al., 2011) pada jurnal *Proceedings of the 20<sup>th</sup> Internasional Conference Companion on World Wide Web*, berita hoax dapat menyebar dengan cepat dan mempengaruhi opini publik. Hal ini dapat menyebabkan kerugian ekonomi dan dampak sosial yang tidak diinginkan.

Dalam menangani masalah deteksi berita hoax, banyak penelitian yang menggunakan metode klasifikasi teks. Salah satu algoritma yang sering digunakan adalah random forest (RF). Menurut studi yang dilakukan oleh (Ramadhan et al., 2022) telah menunjukkan bahwa penggunaan algoritma supervised learning, seperti Random Forest, dapat memberikan hasil yang baik dalam melakukan deteksi berita palsu. Dalam penelitiannya menunjukkan bahwa random forest memberi hasil yang lebih baik yaitu sebesar 84%

dibandingkan dengan algoritma lainnya seperti Logistic Regression sebesar 77%. Selain itu, pre-processing data seperti tokenisasi, dan stemming juga diperlukan untuk meningkatkan akurasi dalam deteksi berita palsu. Penelitian ini bertujuan untuk menguji efektivitas Random Forest dalam deteksi berita palsu dan mengevaluasi hasil dibandingkan dengan metode lainnya.

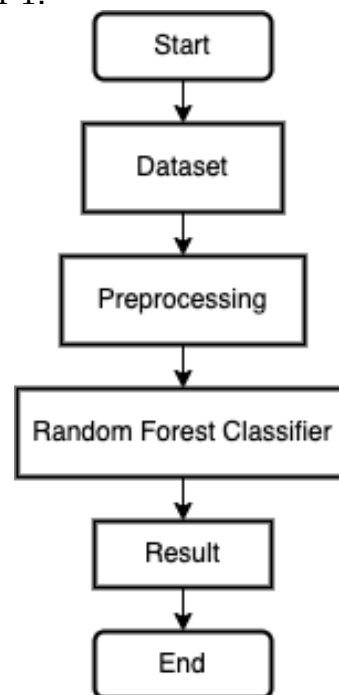
Selain itu, penelitian lain yang dilakukan oleh (Sgurev et al., 2020) menunjukkan bahwa deteksi berita palsu saat ini menjadi perhatian penting karena banyaknya penyebaran berita palsu di internet. Penelitian sebelumnya telah melakukan klasifikasi berita palsu pada bahasa Inggris, tetapi sedikit yang dilakukan pada bahasa Bengali. Dalam penelitian ini, menggunakan algoritma data mining untuk mengklasifikasikan berita palsu dan asli pada bahasa Bengali dan menunjukkan hasil yang akurat sebesar 85% menggunakan algoritma random forest. Tujuan dari penelitian ini adalah untuk membuat sistem deteksi berita palsu pada bahasa Bengali dan menyediakan antarmuka web untuk memeriksa keaslian berita dalam bahasa tersebut.

### Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah metode supervised learning menggunakan algoritma random forest. Dataset yang digunakan berisikan 859 record yang berisi id, status\_fakta, dan judul\_berita, dimana status\_fakta sebagai label. Proses yang dilakukan pada dataset meliputi proses cleaning, tokenisasi, dan stemming untuk memisahkan kalimat menjadi kata-kata yang independen (Ardiani et al., 2020). Kemudian melakukan proses pelatihan dan pengujian pada model untuk menentukan tingkat akurasi yang diperoleh dari model tersebut (Nugroho, 2014). Analisis yang dilakukan pada hasil pengujian

mencakup evaluasi dari metrik precision, recall, f1-score, dan akurasi. Hasil yang diharapkan dari penelitian ini adalah mendapatkan model yang mampu mendeteksi berita palsu dengan tingkat akurasi yang cukup tinggi.

Pada penelitian ini akan dijelaskan alur sistem yang digunakan, seperti yang ditunjukkan dalam gambar 1.



Gambar 1. Alur penelitian

### 1. Dataset

Penelitian ini menggunakan dataset yang berisi 859 record berita, yang terdiri dari 670 berita fakta dan 189 berita palsu yang diambil dari berbagai sumber berita online yang membedakan antara berita fakta dan berita palsu seperti detik.com, liputan6, okezone, kominfo, turnbackhoax dan lain-lain. Setiap record berisi informasi tentang id, status\_fakta, dan judul\_berita. Status\_fakta digunakan sebagai label dalam penelitian ini, yang digunakan untuk menentukan apakah berita tersebut merupakan berita asli atau berita palsu. Dataset ini digunakan untuk melakukan analisis dan

klasifikasi berita palsu pada pemilu 2024.

Tabel 1. Contoh data berita yang terverifikasi sebagai berita fakta atau berita hoax

Id	Judul_berita	Status_fakta
1	Optimis Menang Pemilu 2024 Puan: Kita Bersama Wong Licik	Hoax
2	Pemilu 2024 Ditunda	Hoax
3	Bawaslu RI Rilis 5 Provinsi Paling Rawan dalam Pemilu 2024	Fakta
4	Bawaslu Luncurkan Indeks Kerawanan Pemilu 2024	Fakta

## 2. Preprocessing

Dalam penelitian ini, preprocessing data merupakan tahap penting yang dilakukan sebelum melakukan proses training menggunakan algoritma random forest. Preprocessing data terdiri dari tiga tahap yaitu cleaning, tokenisasi, dan stemming (Bayhaqy et al., 2018). Pada tahap cleaning, dilakukan penghapusan simbol-simbol yang tidak diperlukan seperti tanda baca, dan karakter unik. Tujuan dari tahap ini adalah untuk membersihkan teks dari noise yang tidak diperlukan (Harahap & Muslim, 2018). Pada tahap tokenisasi, dilakukan pemecahan teks menjadi token-token yang lebih kecil seperti kata-kata atau frase. Tujuan dari tahap ini adalah untuk membuat teks lebih mudah diolah dan dianalisis (Mustikasari, 2017). Kemudian pada tahap stemming, dilakukan pemotongan kata dasar dari sebuah kata. Tujuan dari tahap ini adalah untuk menghilangkan imbuhan dan mengubah kata menjadi kata dasar sehingga memudahkan

dalam melakukan analisis (Sardjono et al., 2018).

Setelah melalui tahap-tahap preprocessing, data yang telah diolah akan digunakan dalam proses training dan testing menggunakan algoritma random forest untuk mendeteksi fakta atau hoax dari berita-berita yang dianalisis. Pada table 2, hasil dari penerapan cleaning data kalimat dalam dataset ditampilkan. Pada table 3, hasil dari penerapan tokenisasi kalimat ditampilkan, yang memecah kalimat menjadi kata-kata individu. Dan pada table 4 hasil dari penerapan stemming ditampilkan, yang mengubah kata-kata menjadi kata dasar untuk tujuan analisis.

Tabel 2. Hasil aplikasi dari proses cleaning

Id	Judul_berita	Status_fakta
1	optimis menang pemilu 2024 puan kita bersama wong licik	Hoax
2	pemilu 2024 ditunda	Hoax
3	bawaslu ri rilis 5 provinsi paling rawan dalam pemilu 2024	Fakta
4	bawaslu luncurkan indeks kerawanan pemilu 2024	Fakta

Tabel 3. Hasil aplikasi dari proses tokenisasi

Id	Judul_berita	Status_fakta
1	“optimis”, “menang”, “pemilu”, “2024”, “puan”, “licik”	Hoax
2	“pemilu”, “2024”, “ditunda”	Hoax
3	“bawaslu”, “ri” “rilis”, “5”, “provinsi”, “rawan”, “pemilu”, “2024”	Fakta
4	“bawaslu”,	Fakta

“luncurkan”,  
“indeks”,  
“kerawanan”,  
“pemilu”, “2024”

Tabel 4. Hasil aplikasi dari proses stemming

Id	Judul_berita	Status_fakta
1	“optimis”, “menang”, “milu”, “2024”, “puan”, “licik”	Hoax
2	“milu”, “2024”, “tunda”	Hoax
3	“bawaslu”, “ri”, “rilis”, “5”, “rawan”, “milu”, “2024”	Fakta
4	“bawaslu”, “luncur”, “indeks”, “rawan”, “milu”, “2024”	Fakta

### 3. Random Forest

Random forest adalah algoritma yang diterapkan dalam pembelajaran mesin untuk melakukan klasifikasi dan analisis regresi (Religia et al., 2021). Algoritma ini menggabungkan beberapa model pembelajaran yang independen dalam bentuk pohon keputusan untuk membuat satu model yang lebih kuat dan menghasilkan keputusan akhir. Ini didasarkan pada konsep ensemble learning (Siregar, 2020).

Algoritma Random Forest menggunakan teknik bootstrap aggregating (bagging) untuk menghasilkan sejumlah pohon yang dibangun dari sub-sample dari dataset yang digunakan (Religia et al., 2021). Selain itu, algoritma ini juga menggunakan teknik random feature selection, yaitu dengan memilih fitur secara acak dari dataset untuk digunakan dalam pembentukan pohon (Irawan & Samopa, 2018).

Random Forest digunakan untuk meningkatkan akurasi dari model pembelajaran mesin dan menangani

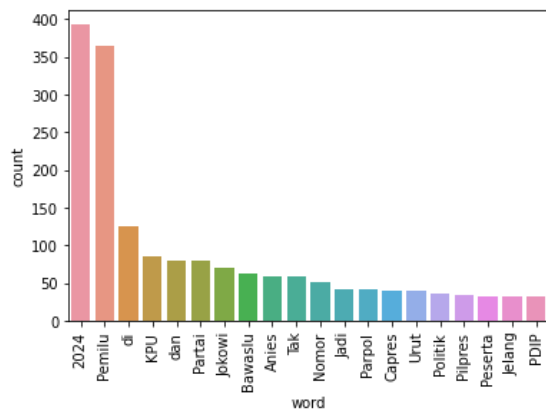
overfitting yang sering terjadi pada algoritma pohon keputusan biasa (Agustiani et al., 2022). Algoritma ini juga mampu menangani masalah missing value dan outlier dalam dataset. Meskipun demikian, algoritma ini memerlukan waktu yang cukup lama untuk melakukan proses penelitian dan memerlukan memori yang cukup besar (Gislason et al., 2006).

### Hasil Dan Pembahasan

Dalam penelitian ini, kami menggunakan algoritma random forest untuk mengklasifikasikan berita palsu pemilihan umum 2024. Dataset yang digunakan berjumlah 859 record yang terdiri dari id, status\_fakta, dan judul\_berita. Algoritma random forest ini diuji dengan data training dan data testing dengan desity kelas yang seimbang. Hasil yang didapatkan dari pengujian ini adalah akurasi sebesar 84.88%.

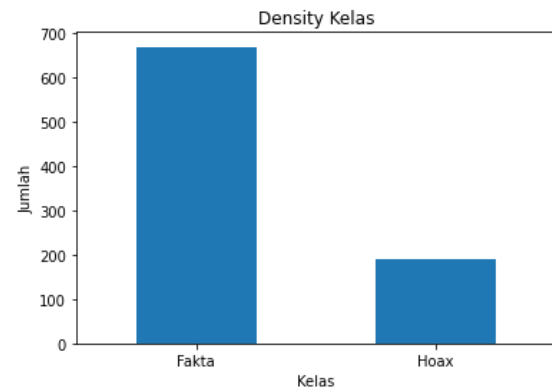
Selain itu, kami juga menampilkan metrik precision, recall, dan f1-score untuk mengevaluasi kinerja model. Metrik ini menunjukkan kinerja model dalam menangkap berita palsu dan berita yang benar.

Penerapan cleaning, tokenisasi, dan stemming pada dataset juga meningkatkan kualitas dataset yang digunakan dalam proses klasifikasi. Hal ini dapat dilihat dari grafik common words yang menunjukkan kata-kata yang paling sering muncul dalam dataset. Kami juga menampilkan word cloud berdasarkan status fakta yaitu hoax dan fakta serta menampilkan confusion matrix untuk mengetahui kinerja model dalam mengklasifikasikan berita palsu dan berita yang benar.



Gambar 2. Grafik common word

Gambar 2 merupakan grafik common word, yaitu kata-kata yang sering muncul dalam dataset yang digunakan. Common word digunakan dalam penelitian ini untuk menganalisis kata-kata yang paling sering muncul dalam dataset yang digunakan dengan tujuan untuk mengetahui kata-kata yang paling banyak digunakan. Hasil dari analisis common word dapat dilihat pada grafik diatas yang menampilkan 20 kata yang paling sering muncul dalam dataset. Dari hasil analisi, dapat diketahui bahwa kata-kata seperti '2024', 'pemilu', 'di', 'KPU', dan 'dan' merupakan kata-kata yang paling sering muncul dalam dataset. Hal ini menunjukkan bahwa topik utama dalam dataset adalah mengenai pemilu tahun 2024. Selain itu, kata-kata seperti 'Jokowi', 'Anies', 'Parpol', dan 'Capres' juga sering muncul dalam dataset, menunjukkan bahwa dataset juga membahas mengenai figure-figur politik yang akan terlibat dalam pemilu tahun 2024. Analisis common word ini dapat digunakan sebagai acuan dalam proses klasifikasi untuk mengetahui kata-kata yang paling relevan dalam dataset yang digunakan.



Gambar 3. Grafik density kelas

Gambar 3 menunjukkan distribusi jumlah data dalam setiap kelas yang digunakan dalam penelitian ini. Dalam penelitian ini, data yang digunakan adalah sesuai dengan persebaran density yang ditunjukkan dalam gambar 3 yaitu sebanyak 859 record dengan label status fakta. Dari 859 record tersebut, terdapat 670 record yang memiliki label fakta dan 189 record yang memiliki label hoax. Hal ini menyatakan bahwa dalam dataset yang digunakan, jumlah record berita fakta lebih banyak dibandingkan dengan jumlah record berita hoax. Hal ini penting untuk diperhatikan dalam proses klasifikasi karena dapat mempengaruhi hasil akhir dari klasifikasi yang dilakukan. Namun, jumlah ini dapat dianggap cukup representative untuk digunakan dalam proses klasifikasi.

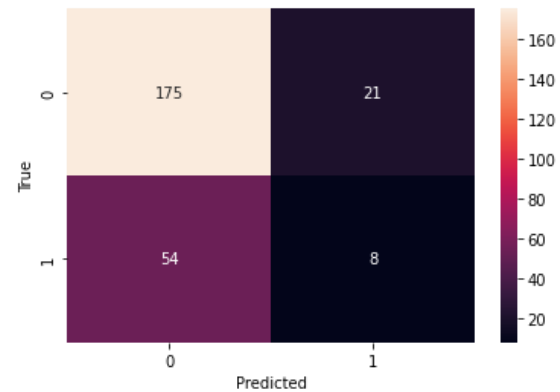


Gambar 4. Word cloud berita hoax pemilu 2024

Gambar 4 merupakan word cloud untuk menampilkan kata berdasarkan status fakta yaitu hoax dan fakta. Word cloud berita hoax digunakan untuk menampilkan kata-kata yang

paling sering muncul dalam berita hoax yang ada dalam dataset. Ini dilakukan untuk mengetahui pola kata-kata yang paling umum digunakan dalam berita hoax dan memberikan pandangan mengenai tren dari kata-kata yang sering digunakan dalam berita hoax. Dalam hal ini angka '2024' tidak muncul karena tidak dianggap sebagai kata yang signifikan dalam konteks klasifikasi berita fakta atau hoax. Angka 2024 mungkin hanya digunakan sebagai referensi waktu dalam judul berita, bukan sebagai kata kunci dalam pembahasa berita itu sendiri. Selain itu, algoritma yang digunakan dalam penelitian mungkin tidak memberikan bobot yang cukup besar pada angka dibandingkan dengan kata-kata lain dalam dataset yang digunakan. Namun, angka 2024 muncul dalam grafik common word, karena grafik ini menunjukkan jumlah kemunculan setiap kata dalam dataset.

angka '2024' tidak ditampilkan dalam word cloud berita fakta karena seperti yang telah dijelaskan sebelumnya.



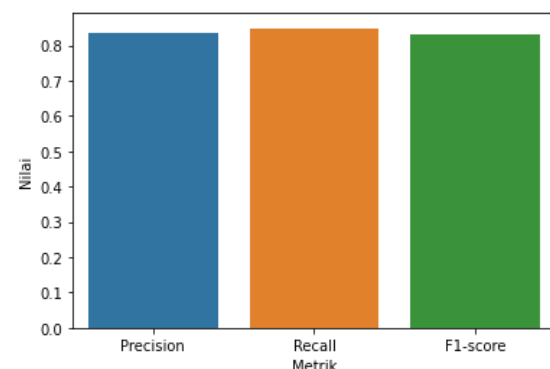
Gambar 6. Confusion matrix

Gambar 6 adalah confusion matrix, digunakan untuk mengevaluasi kinerja model klasifikasi dalam penelitian ini. Matrix ini menunjukkan jumlah prediksi yang benar dan salah dari model tersebut. Pada matrix di atas, di baris pertama, 175 adalah jumlah prediksi yang benar untuk kelas fakta dan 21 adalah jumlah prediksi yang salah. Pada baris kedua, 54 adalah jumlah prediksi yang salah untuk kelas hoax dan 8 adalah jumlah prediksi yang benar. Dapat dilihat bahwa model klasifikasi dalam penelitian ini memiliki performa yang cukup baik dalam memprediksi kelas fakta, namun memiliki performa yang kurang baik dalam memprediksi kelas hoax. Hal ini dapat diperbaiki dengan menggunakan metode lain atau dengan menambah jumlah data pada dataset.



Gambar 5. Word cloud berita fakta pemilu 2024

Gambar 5 adalah representasi visual dari word cloud yang digunakan untuk menunjukkan kata-kata yang paling sering muncul dalam berita yang diklasifikasikan sebagai fakta yang digunakan dalam dataset. Ini digunakan untuk mengetahui tren kata-kata yang paling sering digunakan dalam berita fakta dan memberikan pandangan mengenai pola kata-kata yang sering digunakan dalam berita fakta yang bersifat valid dan dapat dipertanggung jawabkan keabsahannya, dan dapat dipercayai serta dapat diandalkan. Dalam hal ini,



Gambar 7. Hasil precision, recall, dan f1-score

Gambar 7 menampilkan hasil dari precision, recall, dan f1-score. Precision dalam penelitian ini menunjukkan bahwa dari seluruh prediksi yang diberikan oleh algoritma random forest sebagai berita fakta, sebanyak 83% diantaranya benar-benar berita fakta. Recall menunjukkan bahwa dari seluruh berita fakta yang sebenarnya dalam dataset, algoritma random forest mampu mengidentifikasi sebanyak 85% diantaranya sebagai berita fakta. Dan f1-score menunjukkan keseimbangan antara precision dan recall, dengan nilai sebesar 83%. Hal ini menunjukkan bahwa algoritma random forest dalam penelitian ini memiliki kinerja yang baik dalam mengidentifikasi berita fakta dan berita hoax. Namun, masih ada sekitar 17% prediksi yang salah dalam klasifikasi berita fakta dan sekitar 15% berita fakta yang tidak terdeteksi oleh algoritma.

Tabel 5. Hasil akurasi dari algoritma random forest

Model	Akurasi
Random Forest	84.88%

Tabel 5 merupakan hasil akurasi dari algoritma random forest dalam melakukan klasifikasi berita palsu pemilu 2024. Hasil akurasi dari algoritma random forest yang digunakan dalam penelitian ini sebesar 84.88%. Akurasi ini menunjukkan seberapa baik model dalam mengklasifikasikan dataset yang digunakan. Dalam hal ini, model mampu mengklasifikasikan sebagian besar berita fakta dan berita hoax dengan tepat. Namun, akurasi yang didapatkan masih memiliki ruang untuk perbaikan, seperti dengan menambah jumlah dataset yang digunakan atau mencoba metode lain dalam proses klasifikasi.

Dalam keseluruhan, penelitian ini berhasil menunjukkan bahwa algoritma random forest dapat digunakan untuk mengklasifikasikan

berita fakta dan berita hoax pada pemilu 2024. Melalui preprocessing data, cleaning, tokenisasi, dan stemming, kualitas dataset yang digunakan dapat ditingkatkan. Selain itu, analisis common word, word cloud, dan confusion matrix juga dapat digunakan sebagai acuan dalam proses klasifikasi. Namun, masih ada ruang untuk perbaikan dalam kinerja model dan jumlah dataset yang digunakan.

### Simpulan Dan Saran

Penelitian ini dilakukan untuk mengklasifikasikan berita palsu pemilihan umum 2024 menggunakan algoritma random forest. Dataset yang digunakan dalam penelitian ini berisi 859 record yang terdiri dari 670 berita fakta dan 189 berita palsu. Algoritma random forest diuji dengan data training dan data testing dengan desity kelas yang seimbang. Hasil yang didapatkan dari pengujian ini adalah akurasi sebesar 84.88%.

Selain itu, dalam penelitian ini juga dilakukan preprocessing data yaitu cleaning, tokenisasi, dan stemming untuk meningkatkan kualitas dataset yang digunakan dalam proses klasifikasi. Analisis common word juga dilakukan untuk mengetahui kata-kata yang paling relevan dalam dataset yang digunakan.

Dari hasil yang didapatkan, dapat disimpulkan bahwa algoritma random forest dapat digunakan untuk mengklasifikasikan berita palsu pemilihan umum 2024 dengan akurasi yang cukup tinggi. Namun, masih ada sekitar 17% prediksi yang salah dalam klasifikasi berita fakta dan sekitar 15% berita fakta yang tidak terdeteksi oleh algoritma.

Saran penelitian masa depan dapat meliputi pengembangan dataset yang lebih besar dan memperluas variasi topik berita. Selain itu, dapat juga mencoba menggunakan algoritma klasifikasi seperti SVM atau neural network untuk meningkatkan akurasi



dalam mendeteksi berita palsu. Serta dapat juga mencoba mengintegrasikan metode lain seperti analisis sentiment dalam proses klasifikasi. Penelitian juga dapat melakukan analisis terhadap tingkat kepercayaan dari berita tersebut untuk memberikan rekomendasi kepada pembaca berita.

## Referensi

- Agustiani, S., Arifin, Y. T., Junaidi, A., Wildah, S. K., & Mustopa, A. (2022). Klasifikasi Penyakit Daun Padi menggunakan Random Forest dan Color Histogram. *Jurnal Komputasi*, 10(1), 65-74.
- Ansori, L. (2017). Telaah Terhadap Presidential Threshold Dalam Pemilu Serentak 2019. *Jurnal Yuridis*, 4(1), 15-27.
- Ardiani, L., Sujaini, H., & Tursina, T. (2020). Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak. *JUSTIN (Jurnal Sistem Dan Teknologi Informasi)*, 8(2), 183-190.
- Bayhaqy, A., Sfenrianto, S., Nainggolan, K., & Kaburuan, E. R. (2018). Sentiment analysis about E-commerce from tweets using decision tree, K-nearest neighbor, and naïve bayes. *2018 International Conference on Orange Technologies (ICOT)*, 1-6.
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294-300.
- Harahap, R. N., & Muslim, K. (2018). Peningkatan Akurasi Pada Prediksi Kepribadian MBTI Pengguna Twitter Menggunakan Augmentasi Data. *Teknologi Informasi Dan Ilmu Komputer*, 815-822.
- Irawan, F., & Samopa, F. (2018). A Comparative Assessment of the Random Forest and SVM Algorithms Using Combination of Principal Component Analysis and SMOTE For Accounts Receivable Seamless Prediction case study company X in Surabaya. *Proceedings of the 2nd International Seminar Of Contemporary ONBUSINESS Research & Management*.
- Juliswara, V. (2017). Mengembangkan model literasi media yang berkebhinnekaan dalam menganalisis informasi berita palsu (hoax) di media sosial. *Jurnal Pemikiran Sosiologi*, 4(2).
- Mustikasari, D. (2017). Analisis Tema Skripsi Mahasiswa Menggunakan Document Clustering Dengan Algoritma LINGO. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 131-140.
- Nugroho, Y. S. (2014). Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro. *Dian Nuswantoro Fakultas Ilmu Komputer Skripsi*.
- Ramadhan, N. G., Adhinata, F. D., Segara, A. J. T., & Rakhmadani, D. P. (2022). Deteksi Berita Palsu Menggunakan Metode Random Forest dan Logistic Regression. *JURIKOM (Jurnal Riset Komputer)*, 9(2), 251. <https://doi.org/10.30865/jurikom.v9i2.3979>
- Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., & Menczer, F. (2011). Truthy: mapping the spread of astroturf in microblog streams. *Proceedings of the 20th International Conference Companion on World Wide Web*, 249-252.
- Religia, Y., Nugroho, A., & Hadikristanto, W. (2021). Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 187-192.

Sardjono, M. W., Cahyanti, M., Mujahidin, M., & Arianty, R. (2018). Pendeteksi Kesamaan Kata untuk Judul Penulisan Berbahasa Indonesia Menggunakan Algoritma Stemming Nazief-Adriani. *Sebatik*, 22(2), 138-146.

Sgurev, V., IEEE Systems, M., Institute of Electrical and Electronics Engineers. Bulgaria Section. IM/CS/SMC Joint Chapter, Institute of Electrical and Electronics Engineers. Bulgaria Section. CIS Chapter, & Institute of Electrical and Electronics Engineers. (n.d.). *2020 IEEE 10th International Conference on Intelligent Systems (IS): proceedings*.

Siregar, A. M. (2020). *Klasifikasi Untuk Prediksi Cuaca Menggunakan Esemble Learning*.

Yonathan, A., Sujaini, H., Esyudha Pratama, E., Nawawi, J. H., & Barat, K. (2022). *Perbandingan Algoritma Klasifikasi dalam Pendeteksian Hoax pada Media Sosial*.

<https://doi.org/10.26418/juara.v1i1.53126>

Yunanto, R., Purfini, A. P., & Prabuwisesa, A. (2021). *Jurnal Manajemen Informatika (JAMIKA) Survei Literatur: Deteksi Berita Palsu Menggunakan Pendekatan Deep Learning*.

<https://doi.org/10.34010/jamika.v1i1i2.493>