DETEKSI BERITA PALSU MENGGUNAKAN ALGORITMA *RANDOM FOREST*

Siti Nurohanisah[1], Rini Astuti[2], Fadhil Muhammad Basysyar[3]
[1]Teknik Informatika, STMIK IKMI Cirebon
[2] Sistem Informasi, STMIK LIKMI Bandung
[3]Sistem Informasi, STMIK IKMI Cirebon
Jalan Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135
sitinurohanisah12@gmail.com

ABSTRAK

Dalam era digital yang dipenuhi oleh penyebaran berita palsu atau yang lebih dikenal dengan istilah "hoaks," tantangan untuk memerangi penyebaran informasi yang salah dan menyesatkan semakin mendesak. Berita palsu memiliki potensi untuk merusak kepercayaan publik, menciptakan kebingungan, dan berdampak negatif pada masyarakat. Oleh karena itu, penelitian ini bertujuan untuk mengklasifikasi berita palsu yang menggunakan algoritma *Random Forest*, dengan tujuan mengeksplorasi cara menggunakan algoritma *Random Forest* dalam mengidentifikasi dan memisahkan berita palsu dari berita yang asli. Penelitian ini dilakukan dengan memanfaatkan dataset yang mencakup berbagai jenis berita, termasuk berita palsu dan berita asli, yang telah dianalisis dan dikategorikan oleh ahli faktual. Teknik pemrosesan bahasa alami (NLP) digunakan untuk mengekstraksi fitur-fitur penting dari teks berita. Penelitian ini menggunakan algoritma *Random Forest* untuk klasifikasi. Random Forest bekerja dengan cara menggabungkan sejumlah besar pohon keputusan yang independen untuk membuat keputusan akhir. Setiap pohon membuat prediksi, dan hasil akhirnya didasarkan pada mayoritas prediksi dari pohon-pohon tersebut. Kelebihan utama Random Forest adalah kemampuannya dalam mengatasi overfitting dan meningkatkan akurasi dengan menggabungkan prediksi dari beberapa model. Evaluasi dilakukan dengan menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa klasifikasi berita palsu menggunakan algoritma *Random Forest* mampu mengklasifikasikan berita dengan tingkat akurasi yang memuaskan. Hasil akhir menunjukkan bahwa sistem ini memiliki kinerja yang baik dalam membedakan berita palsu dari berita asli, dengan akurasi rata-rata di atas 87%.

Kata kunci : *Algoritma Random Forest, Berita Palsu, Dataset, Deteksi*

1. PENDAHULUAN

Dalam era digital yang berkembang pesat, peran Informatika telah mengubah banyak aspek kehidupan manusia. Teknologi informasi telah memengaruhi sektor bisnis, pendidikan, komunikasi, dan hampir semua aspek kehidupan sehari-hari [1]. Penggunaan teknologi informasi, khususnya di era internet, telah memberikan kemudahan akses informasi dan interaksi, tetapi juga membawa tantangan baru terutama dalam hal validitas informasi. Salah satu masalah utama yang muncul dalam konteks ini adalah penyebaran berita palsu atau yang dikenal dengan sebutan "hoaks".

Penyebaran berita palsu dapat merugikan masyarakat, menciptakan kebingungan, dan merusak kepercayaan publik. Dampak negatif dari berita palsu ini dapat berdampak pada keputusan politik, kesehatan masyarakat, dan stabilitas sosial. Oleh karena itu, penelitian ini memfokuskan pada masalah klasifikasi berita palsu dalam konteks Informatika[2].

Berbagai tren terbaru menunjukkan bahwa berita palsu semakin menyebar di platform media sosial dan internet. Peran teknologi informasi dalam menyebarkan berita palsu menjadi semakin penting. Dalam hal ini, algoritma *Random Forest* menjadi salah satu pendekatan yang potensial untuk mengidentifikasi dan memisahkan berita palsu dari berita yang asli [3]. Namun, keberhasilan implementasi teknik ini bergantung pada pemahaman yang mendalam tentang konteks Informatika dan penelitian sebelumnya.

Dalam konteks ini, studi-studi sebelumnya telah berusaha mengatasi masalah berita palsu. Namun, ada potensi untuk meningkatkan dan memperdalam pemahaman tentang masalah ini, serta mengembangkan pendekatan yang lebih efektif dalam klasifikasi berita palsu. Oleh karena itu, penelitian ini akan mengambil langkah lebih lanjut dalam memahami dampak berita palsu di bidang Informatika dan pengembangan solusi yang lebih baik.

Tujuan utama dari penelitian ini adalah mengklasifikasi berita palsu yang menggunakan algoritma *Random Forest*. Sistem ini akan memungkinkan mengklasifikasi berita palsu secara otomatis dan efisien, sehingga dapat membantu dalam menjaga integritas informasi di era digital ini. Selain itu, penelitian ini juga bertujuan untuk memberikan manfaat praktis bagi masyarakat luas dalam memerangi penyebaran berita palsu.

Algoritma *Random Forest*, metode penelitian, dan analisis data akan menjadi bagian integral dari pendekatan penelitian ini. Hasil penelitian diharapkan dapat memberikan wawasan baru tentang cara Paper (Adhitya Karel Maulaya & Junadhi, 2022) membahas mengenai sistem klasifikasi opini cuitan Twitter terkait aksi peretasan data oleh hacker bernama Bjorka. Sistem ini menggunakan metode Support Vector Machine (SVM). Dataset yang digunakan dalam

penelitian ini adalah kumpulan cuitan Twitter yang dikumpulkan menggunakan Twitter API. Dataset tersebut berjumlah 1000 cuitan. Hasil penelitian menunjukkan bahwa metode SVM dapat menghasilkan akurasi sebesar 62,33% dalam mengklasifikasi opini cuitan Twitter terkait aksi peretasan data oleh hacker bernama Bjorka.digital.

2. TINJAUAN PUSTAKA

Algoritma Random Forest adalah suatu metode ensemble learning yang digunakan untuk tugas klasifikasi dan regresi. Ensemble learning melibatkan penggabungan hasil beberapa model pembelajaran mesin untuk meningkatkan kinerja dan keakuratan prediksi. Random Forest bekerja dengan menggabungkan sejumlah besar pohon keputusan (*decision trees*) yang dibuat secara acak.[4]

Klasifikasi adalah proses pengembangan model untuk memprediksi kategori atau label kelas dari suatu objek, berdasarkan pada informasi yang terdapat dalam himpunan contoh yang sudah diberi label atau yang telah diklasifikasikan sebelumnya.[5]

Machine Learning adalah adalah bidang ilmu komputer yang berfokus pada pengembangan teknik dan algoritma yang memungkinkan komputer untuk belajar dari data dan pengalaman tanpa di-program secara eksplisit.[3]

Natural Language Processing (NLP) adalah adalah cabang dari kecerdasan buatan yang berfokus pada interaksi antara komputer dengan bahasa manusia secara alami. NLP mencakup pengembangan teknologi dan algoritma untuk memahami, memproses, dan menghasilkan bahasa manusia dalam bentuk yang dapat diperlakukan oleh komputer.[6]

Knowledge Discovery in Databases (KDD) adalah suatu proses integral dalam bidang ilmu data yang bertujuan untuk mengeksplorasi, menggali, dan menemukan pola atau pengetahuan yang bermanfaat dari dataset yang besar dan kompleks. Proses KDD melibatkan langkah-langkah seperti pemahaman masalah yang dihadapi, persiapan dan pembersihan data, serta tahap penambangan data untuk mengidentifikasi pola atau informasi yang berguna. Seiring dengan itu, KDD Cup menambah dimensi kompetitif ke dunia penambangan data dengan mengadakan serangkaian kompetisi tahunan di mana peserta diberikan dataset untuk menyelesaikan tantangan khusus. Di sisi lain, dalam konteks sistem operasi Windows, istilah KDD juga merujuk pada Kernel Data Deduplication, sebuah fitur yang dirancang untuk mengurangi penggunaan ruang penyimpanan dengan menghilangkan salinan data identik pada tingkat blok. Dengan demikian, KDD mencakup spektrum konsep yang luas, membentang dari metode analisis data hingga kompetisi ilmu data dan teknologi penyimpanan data dalam konteks perangkat lunak sistem operasi.

Paper [7] membahas mengenai deteksi hoax pada berita online bahasa inggris menggunakan bernoulli naive bayes dengan ekstraksi fitur TF-IDF yang bertujuan untuk mengetahui performa dari penggunaan algoritma bernoulli naive bayes dengan ekstraksi TF-IDF dalam mendeteksi berita hoax. Hasil impelementasi pada penelitian ini menunjukan model prediksi yang dibangun dengan 8800 data berita, mampu menghasilkan nilai akurasi 98,5% dari jumlah data uji sebanyak 2200 data berita, dimana akurasi dari label fake sebesar 97,8% dan akurasi dari label true sebesar 99,1% yang diikuti dengan nilai precision 99,1%, recall 97,8% dan f1-score 98,4%.

Paper [8] membahas mengenai deteksi berita palsu menggunakan metode random forest dan logistic regression yang bertujuan untuk deteksi berita palsu menggunakan model supervised learning random forest. Dataset yang digunakan pada penelitian ini berjumlah 6256 baris judul yang memiliki kelas fake dan real. Hasil yang didapatkan menggunakan model random forest sebesar 85%, hasil tersebut lebih tinggi dibandingkan dengan menggunakan model logistic regresi sebesar 75%.

Paper [9] membahas mengenai algoritma Long Short-Term Memory (LSTM) untuk mendeteksi ujaran kebencian berkaitan dengan Pemilihan Presiden (Pilpres) 2019. Algoritma ini dilatih dengan menggunakan dataset yang bersumber dari media sosial Facebook. Hasil uji coba terbaik dengan 190 kalimat memiliki nilai recall sebesar 0,7021. Hal ini berarti bahwa dari keseluruhan data uji ujaran kebencian, 70,21% benar dideteksi sebagai ujaran kebencian, sedangkan sisanya 29,79% salah dideteksi sebagai bukan ujaran kebencian.

Paper [10] membahas mengenai tingkat akurasi metode K-Nearest Neighbor dan naive bayes classifier pada klasifikasi ujaran kebencian terhadap agama islam. Penelitian ini bertujuan untuk membandingkan performa dari dua metode klasifikasi, yaitu metode K-Nearest Neighbor (KNN) dan metode Naïve Bayes Classifier (NBC), dalam mendeteksi ujaran kebencian (hate speech) di Twitter. Penelitian ini menggunakan dataset yang sama dan tahapan preprocessing yang sama dari penelitian sebelumnya. Presentase perbandingan data latih dan data uji adalah 50:50, 60:40, 70:30, 80:20, dan 90:10. Hasil penelitian menunjukkan bahwa metode NBC menghasilkan nilai akurasi yang lebih baik daripada metode KNN. Metode NBC mampu menghasilkan akurasi sebesar 95% dengan pembagian 90% data latih berbanding 10% data uji. Metode KNN mampu menghasilkan akurasi sebesar 94% dengan pembagian 80% data latih berbanding 20% data uji. Sedangkan tanpa fitur threshold, metode KNN hanya menghasilkan akurasi sebesar 82%. Secara keseluruhan, penelitian ini menunjukkan bahwa metode NBC merupakan metode yang lebih efektif untuk mendeteksi ujaran kebencian di Twitter.

Paper [11] membahas mengenai sistem klasifikasi opini cuitan Twitter terkait aksi peretasan data oleh hacker bernama Bjorka. Sistem ini menggunakan metode Support Vector Machine (SVM). Dataset yang digunakan dalam penelitian ini

adalah kumpulan cuitan Twitter yang dikumpulkan menggunakan Twitter API. Dataset tersebut berjumlah 1000 cuitan. Hasil penelitian menunjukkan bahwa metode SVM dapat menghasilkan akurasi sebesar 62,33% dalam mengklasifikasi opini cuitan Twitter terkait aksi peretasan data oleh hacker bernama Bjorka.

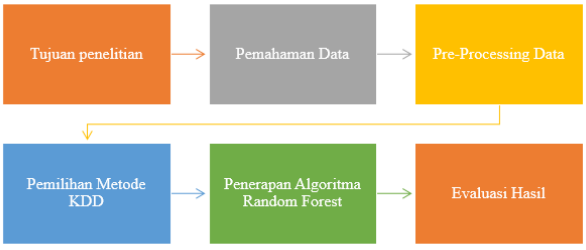
Paper [12] membahas mengenai deteksi berita palsu tentang vaksinasi Covid-19 dengan menggunakan text mining dan algoritma cosine similarity. Penelitian ini bertujuan untuk menguur tingkat kesamaan berita palsu dengan berita benar atau fakta.

Paper [13] membahas mengenai Pentingnya deteksi berita hoaks menjadi semakin mendesak, dan meskipun terdapat beberapa upaya untuk mengatasi masalah ini, sampai saat ini belum ada sistem deteksi hoaks yang efektif dalam Bahasa Indonesia, kecuali menggunakan representasi vektor teks berdasarkan Term Frequency dan Document Frequency serta teknik klasifikasi Support Vector Machine dan Stochastic Gradient Descent dengan tingkat akurasi 60%.Metode yang digunakan dalam penelitian ini melibatkan aplikasi pre-processing data atas informasi yang dikumpulkan dari Twitter dan Facebook selama 3 bulan. Selanjutnya, model dibangun menggunakan Scikit-Learn, dan akurasi model diuji menggunakan berita aktual sebagai data uji. Hasil penelitian menunjukkan bahwa model yang dikembangkan mampu mengidentifikasi apakah suatu berita di media sosial, terutama Twitter dan Facebook, bersifat palsu atau tidak, dengan memanfaatkan TF-IDF, CountVectorizer, PassiveAgressive Classifier, dan Support Vector Classifier.

3. METODE PENELITIAN

3.1. Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode eksperimen. Penelitian ini akan mengeksplorasi cara menggunakan *Random Forest*.



Gambar 1. Langkah-langkah penelitian

Dari langkah-langkah yang terdapat pada gambar 1 dapat dideskripsikan sebagai berikut :

a) Tujuan penelitian bertujuan untuk mendefinisikan dengan jelas apa tujuan dari penelitian.
b) Pemahaman data berguna untuk mendeskripsikan data memiliki beberapa kolom dan data yang memiliki atribut label sebagai penanda.
c) Preprocessing data berguna untuk melakukan penghapusan baris yang memiliki NaN, mengubah teks menjadi *lowercase* dan menghapus karakter special.
d) Pemilihan metode KDD bertujuan untuk memilih metode apa yang digunakan, peneliti memilih metode atau algoritma random forest.
e) Penerapan algoritma random forest berguna untuk menerapkan algoritma tersebut pada dataset yang sudah dibagi atau di-*split*.
f) Evaluasi hasil, menggunakan matrix dan ROC bertujuan untuk melihat keakuratan random forest dalam mengklasifikasi.

3.2. Sumber data

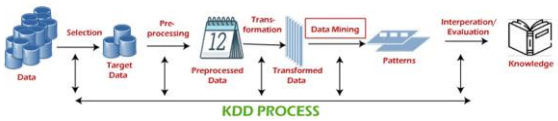
Sumber data yang digunakan pada penelitian ini adalah dari sumber public yakni kaggle.com, pada situs tersebut peneliti mengambil dataset tentang fake news.

3.3. Populasi dan Sampel

Populasi yang digunakan adalah berita palsu dan berita asli. Sampling yang digunakan adalah seluruh data yang terdapat pada dataset fake news.

3.4. Teknik Analasis Data

Algoritma *Random Forest* akan digunakan untuk mengklasifikasikan berita menjadi dua kategori, yaitu berita palsu dan berita asli. Hasil klasifikasi ini akan digunakan untuk mengevaluasi kinerja sistem klasifikasi berita palsu yang telah dikembangkan.



Gambar 2 Tahapan Proses KDD

Dari tahapan-tahapan yang terdapat pada gambar 2 dapat dijelaskan sebagai berikut :

a) Dataset
Dataset merupakan sekumpulan data oprasional dimana dilakukan sebelum tahap penggalian informasi dalam KDD dimulai.
b) Data selection
Data selection merupakan tahap awal yang digunakan untuk mengumpulkan data operasional yang akan digunakan sebelum proses penggalian data operasional yang akan digunakan sebelum proses penggalian informasi KDD dimulai dengan memaparkan data aslinya. Hasil dari data seleksi disimpan dalam suatu berkas yang terpisah dari basis data fungsional.

c) Pre-processing.
Dalam tahap *pre-processing*, proses dilanjutkan dengan membersihkan data dari baris yang mengandung nilai *null* (*NaN*). *Case folding* juga diterapkan untuk mengubah seluruh teks menjadi huruf kecil, sehingga konsistensi dalam pemrosesan teks dapat dipertahankan. Selain itu, dilakukan penghapusan karakter spesial guna mengoptimalkan representasi teks sebelum memasuki tahap analisis dan pemodelan. Proses preprocessing ini bertujuan untuk meningkatkan kualitas data dan memastikan bahwa data yang digunakan dalam proses selanjutnya telah disesuaikan dengan format yang diinginkan.

d) Transformation
Pada tahap transformasi data, dilakukan proses pembagian data menjadi dua bagian, yaitu data latih (train) sebanyak 70% dan data uji (test) sebanyak 30%. Pembagian ini bertujuan untuk mempersiapkan dataset yang akan digunakan dalam proses pelatihan model dan pengujian model. Data latih digunakan untuk melatih model, sementara data uji digunakan untuk menguji sejauh mana model dapat memberikan prediksi yang baik pada data yang belum pernah dilihat sebelumnya.

e) Data mining
Dalam tahap eksplorasi data mining, dilakukan penerapan algoritma Random Forest untuk memodelkan dataset yang telah melalui tahap seleksi, preprocessing, dan transformasi sebelumnya.

f) Evaluasi
Dalam tahap evaluasi, kinerja algoritma Random Forest dievaluasi menggunakan metode *Confusion Matrix* dan *Receiver Operating Characteristic* (ROC) untuk mendapatkan gambaran seberapa akurat prediksi model. *Confusion Matrix* memberikan rincian jumlah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN), yang memungkinkan penilaian yang lebih rinci terhadap performa model klasifikasi. Sementara itu, analisis ROC digunakan untuk mengukur kemampuan model dalam membedakan antara kelas positif dan negatif, dengan memberikan nilai Area *Under the Curve* (AUC) sebagai indikator tingkat akurasi secara keseluruhan. Dengan kombinasi kedua metode ini, evaluasi menyeluruh dilakukan untuk menyimpulkan sejauh mana algoritma Random Forest dapat memberikan prediksi yang akurat dan reliabel.

4. HASIL DAN PEMBAHASAN

Pada bagian ini menjelaskan tentang hasil dan analisa hasil dari penelitian yang telah dilakukan. Percobaan dilakukan dengan menggunakan library bahasa pemrograman python dengan editor jupyter notebook. Spesifikasi perangkat keras yang digunakan untuk percobaan yaitu laptop dengan Intel(R) Core(TM) i5-2450M CPU @ 2.50GHz (4 CPUs).

4.1. Hasil Penelitian

Dalam penelitian ini, dilakukan pemilihan data sampel dengan metode random sampling sebanyak 6000 entitas. Pendekatan ini digunakan untuk memastikan representativitas dan keberagaman data yang diambil secara acak, guna melanjutkan analisis lebih lanjut.

Gambar 3. Random Sampling

Tahap selanjutnya adalah proses pra-pemrosesan data, dilakukan eliminasi data yang hilang (missing) guna meningkatkan kualitas dataset. Selain itu, dataset menjalani tahap case folding dan filtering untuk mengubah teks menjadi huruf kecil serta menghilangkan karakter khusus.

Setelah tahap preprocessing dilakukan selanjutnya adalah transformasi data, dataset dipecah menjadi data latih (train) sebanyak 70% dan data uji (test) sebanyak 30%. Selain itu, dilakukan proses count vectorization guna mengkonversi teks menjadi representasi numerik.

Gambar 4. Transformasi Data

Berdasarkan evaluasi model klasifikasi, ditemukan bahwa model memiliki tingkat akurasi sebesar 87%. Analisis performa klasifikasi lebih lanjut menyoroti parameter-parameter penting, yaitu precision, recall, dan f1-score, untuk kedua kategori yaitu 'FAKE' dan 'REAL'. Untuk kategori 'FAKE', nilai precision, recall, dan f1-score berturut-turut

adalah 0.86, 0.89, dan 0.87. Sementara itu, untuk kategori 'REAL', parameter-parameter tersebut memiliki nilai masing-masing 0.88, 0.86, dan 0.87.

Hasil & Evaluasi

```
[43] # Buat prediksi pada set pengujian
     y_pred = clf.predict(X_test_vectorized)

     # Evaluasi kinerja model
     accuracy = accuracy_score(y_test, y_pred)
     print("Accuracy:", accuracy)

     Accuracy: 0.8705555555555555
```

```
[44] # Tampilkan classification report
     print("Classification Report:\n", classification_report(y_test, y_pred))

     Classification Report:
                   precision    recall  f1-score   support

             FAKE       0.87      0.88      0.87       914
             REAL       0.87      0.86      0.87       886

         accuracy                           0.87      1800
        macro avg       0.87      0.87      0.87      1800
     weighted avg       0.87      0.87      0.87      1800
```

Gambar 5Hasil dan evaluasi

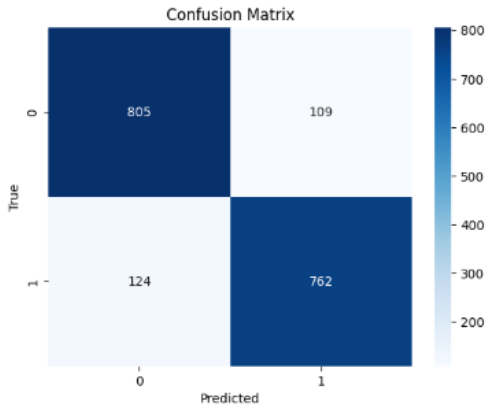
Confusion matrix adalah alat evaluasi kritis dalam analisis kinerja model klasifikasi. Dalam matriks ini, empat komponen utama dapat diidentifikasi. *True Positive* (TP) adalah jumlah kasus yang benar-benar termasuk dalam kelas positif dan diidentifikasi dengan benar. *True Negative* (TN) mencerminkan jumlah kasus yang benar-benar termasuk dalam kelas negatif dan diprediksi dengan benar. Di sisi lain, *False Positive* (FP) adalah jumlah kasus yang sebenarnya negatif, tetapi diprediksi sebagai positif, sementara *False Negative* (FN) mencakup kasus yang sebenarnya positif, tetapi diprediksi sebagai negatif. Sebagai contoh, dalam confusion matrix [[805 109] [124 762]], model memiliki 805 prediksi benar-benar negatif (TN), 762 prediksi benar-benar positif (TP), 109 prediksi palsu positif (FP), dan 124 prediksi palsu negatif (FN). Analisis lebih lanjut tergantung pada konteks aplikasi klasifikasi, tetapi matriks ini memberikan gambaran komprehensif tentang kinerja model.

Kurva ROC (*Receiver Operating Characteristic*) menggambarkan kinerja model klasifikasi, menunjukkan hubungan antara Tingkat *True Positive* dan Tingkat *False Positive* pada berbagai ambang batas keputusan. Dalam hal ini, Area di Bawah Kurva ROC (ROC AUC) sebesar 0.87, mencerminkan kemampuan model dalam membedakan antara kelas positif dan negatif. Semakin tinggi nilai ROC AUC, semakin baik kemampuan diskriminatif model, dengan 1.0 menandakan klasifikasi sempurna.

```
confusion_mat = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:\n", confusion_mat)

sns.heatmap(confusion_mat, annot=True, fmt='d', cmap='Blues')
plt.xlabel('Predicted')
plt.ylabel('True')
plt.title('Confusion Matrix')
plt.show()

Confusion Matrix:
 [[805 109]
 [124 762]]
```

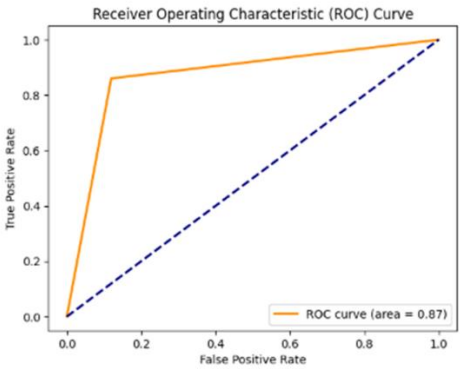


Gambar 6.Confusion Matrix

```
[47] y_test_binary = lb.fit_transform(y_test)
     y_pred_binary = lb.transform(y_pred)

     # Compute ROC curve and ROC area
     fpr, tpr, _ = roc_curve(y_test_binary, y_pred_binary)
     roc_auc = auc(fpr, tpr)

     # Plot ROC curve
     plt.figure()
     plt.plot(fpr, tpr, color='darkorange', lw=2, label='ROC curve (area = {:.2f})'.format
     plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
     plt.xlabel('False Positive Rate')
     plt.ylabel('True Positive Rate')
     plt.title('Receiver Operating Characteristic (ROC) Curve')
     plt.legend(loc="lower right")
     plt.show()
```



Gambar 7.Curve ROC

Word Cloud untuk Berita Palsu adalah Representasi visual dari kata-kata kunci yang sering muncul dalam konteks berita palsu, menyoroti perhatian pada isu penyebaran informasi palsu dan pentingnya verifikasi sebelum dipercayai.

Gambar 8.Word Cloud

Word Cloud untuk Berita Asli adalah gambaran visual kata-kata kunci yang sering muncul dalam konteks berita yang dapat dipercayai. Menonjolkan keakuratan dan keandalan informasi dalam liputan berita.

Gambar 9. word Cloud Berita Asli

4.2. Pembahasan

Berdasarkan hasil penelitian yang tercatat dalam jurnal berjudul "Deteksi Berita Palsu Menggunakan Metode Random Forest dan Logistic Regression," ditemukan bahwa model deteksi berita palsu yang menerapkan Metode Random Forest mencapai tingkat akurasi sebesar 85%. Sebagai perbandingan, dalam penelitian ini yang dilakukan, implementasi Metode Random Forest untuk mengklasifikasi berita palsu menunjukkan peningkatan kinerja dengan mencapai tingkat akurasi sebesar 87%. Hasil ini menggambarkan kemajuan dalam efektivitas model deteksi berita palsu menggunakan pendekatan yang serupa, di mana model yang dikembangkan dalam penelitian ini berhasil mencapai tingkat akurasi yang lebih tinggi dibandingkan dengan penelitian sebelumnya. Peningkatan ini dapat diatribusikan pada perbedaan dalam dataset, metode penelitian, atau faktor-faktor lainnya yang memengaruhi kinerja model. Evaluasi lebih lanjut terhadap faktor-faktor tersebut diharapkan dapat memberikan wawasan lebih mendalam terkait dengan efektivitas model deteksi berita palsu.

Model klasifikasi memberikan hasil yang mengesankan dengan 929 berita yang akurat diidentifikasi sebagai berita asli, dan 871 berita yang dikenali sebagai palsu. Dari total 1800 berita yang dievaluasi, sebanyak 929 diidentifikasi dengan benar sebagai berita faktual (True Positives), dan hanya 871 berita yang salah dikategorikan sebagai palsu (False Positives). Sebaran ini mencerminkan kemampuan model untuk membedakan dengan baik antara berita yang sesuai dengan fakta dan berita palsu, menunjukkan tingkat akurasi yang tinggi dalam klasifikasi berita. Meskipun demikian, evaluasi dan perbaikan model secara terus-menerus tetap penting untuk memastikan ketepatan hasil klasifikasi pada waktu yang akan datang.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang dilakukan untuk mengklasifikasi berita antara yang asli (real) dan yang palsu (fake), model yang dikembangkan dalam penelitian ini berhasil mencapai tingkat akurasi sebesar 87% dengan menerapkan algoritma Random Forest. Hasil tersebut menunjukkan keberhasilan model dalam melakukan prediksi dengan tingkat keakuratan yang tinggi. Selain itu, nilai False Positive (FP) dan False Negative (FN) dalam confusion matrix juga tergolong rendah, menunjukkan bahwa model cenderung memiliki tingkat kesalahan yang minim dalam mengklasifikasikan berita asli dan palsu, Model klasifikasi memberikan hasil yang mengesankan dengan 929 berita yang akurat diidentifikasi sebagai berita real asli, dan 871 berita yang dikenali sebagai fake palsu, Berdasarkan hasil klasifikasi random forest ditemukan karakteristik fake atau palsu (negatif) terdapat mengandung banyak kata yaitu will, one, people, sedangkan untuk karakteristik real atau asli (positif) terdapat mengandung banyak kata trump, said, dan clinton.

Dalam rangka memberikan arahan bagi peneliti atau praktisi yang berminat untuk melanjutkan penelitian klasifikasi berita palsu menggunakan algoritma Random Forest, penulis memberikan beberapa saran yang diharapkan dapat memperkaya dan memperluas pemahaman terkait. Pertama, penelitian lebih lanjut dapat dieksplorasi dengan menerapkan algoritma klasifikasi berita palsu selain Random Forest, guna memperoleh pemahaman yang lebih mendalam tentang opsi yang paling efektif. Selanjutnya, integrasi analisis sentimen dapat menjadi langkah yang potensial untuk meningkatkan ketepatan model dalam mengklasifikasikan berita, mengingat aspek emosional seringkali terdapat dalam teks berita. Validasi eksternal dengan menggunakan dataset yang lebih baru juga diusulkan untuk menilai kemampuan model dalam mengikuti perkembangan tren dan perubahan perilaku penyebaran berita palsu yang terkini. Selain itu, pengembangan antarmuka pengguna yang ramah pengguna dapat mempermudah pihak non-teknis dalam memanfaatkan model deteksi berita palsu. Terakhir, kolaborasi dengan pihak terkait,

seperti media massa, lembaga riset, atau pihak berwenang, disarankan untuk menguji dan mengimplementasikan model dalam konteks dunia nyata. Dengan mengikuti saran-saran ini, diharapkan penelitian dapat lebih relevan dan kontributif dalam mengatasi tantangan deteksi berita palsu dengan pendekatan yang holistik dan up-to-date.

DAFTAR PUSTAKA

[1] A. Kadir, *Pengenalan Sistem Informasi*, Revisi. Yogyakarta: Penerbit Andi Yogyakarta, 2014.

[2] A. R. Purnajaya and Y. Pernando, "Analisa Sentimen Informasi Hoaks Pasca Pandemi Covid-19 dengan Text Mining," *J. Comput. Syst. Informatics*, vol. 4, no. 3, pp. 460–469, 2023, doi: 10.47065/josyc.v4i3.3358.

[3] I. I. Sholikhah, A. T. J. Harjanta, and K. Latifah, "Machine Learning Untuk Deteksi Berita Hoax Menggunakan BERT," *Pros. Semin. Nas. Inform.*, vol. 1, no. 1, pp. 524–531, 2023.

[4] N. G. Ramadhan, F. D. Adhinata, A. J. T. Segara, and D. P. Rakhmadani, "Deteksi Berita Palsu Menggunakan Metode Random Forest dan Logistic Regression," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, p. 251, 2022, doi: 10.30865/jurikom.v9i2.3979.

[5] N. Widjiyati, "Implementasi Algoritme Random Forest Pada Klasifikasi Dataset Credit Approval Implementation of Random Forest Algorithm in The Classification of Credit Approval Dataset," vol. 1, no. 1, pp. 1–7, 2021, doi: 10.25008/janitra.v1i1.118.

[6] N. Agustina and M. Hermawati, "Implementasi Algoritma Naïve Bayes Classifier untuk Mendeteksi Berita Palsu pada Sosial Media," *Fakt. Exacta*, vol. 14, no. 4, pp. 1979–276, 2021, doi: 10.30998/faktorexacta.v14i4.11259.

[7] A. Yodi Prayoga, A. Id Hadiana, and F. Rakhmat Umbara, "Deteksi Hoax pada Berita Online Bahasa Inggris Menggunakan Bernoulli Naïve Bayes dengan Ekstraksi Fitur Tf-Idf," *J. Syntax Admiration*, vol. 2, no. 10, pp. 1808–1823, 2021, doi: 10.46799/jsa.v2i10.327.

[8] N. Kamal, Y. Ramdhani, P. Studi, T. Informatika, U. Adhirajasa, and R. Sanjaya, "Optimasi Algoritma Neural Network berbasis Fitur Seleksi Menggunakan Algoritma Genetika Untuk Prediksi Curah Hujan," vol. 4, no. 2, pp. 269–282, 2023.

[9] A. S. Talita and A. Wiguna, "Implementasi Algoritma Long Short-Term Memory (LSTM) Untuk Mendeteksi Ujaran Kebencian (Hate Speech) Pada Kasus Pilpres 2019," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 19, no. 1, pp. 37–44, 2019, doi: 10.30812/matrik.v19i1.495.

[10] A. Fitri, "Tingkat Akurasi Metode K-Nearest Neighbor," pp. 1–213, 2019.

[11] Adhitya Karel Maulaya and Junadhi, "Analisis Sentimen Menggunakan Support Vector Machine Masyarakat Indonesia Di Twitter Terkait Bjorka," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 3, no. 3, pp. 495–500, 2022, doi: 10.37859/coscitech.v3i3.4358.

[12] D. Marta, G. L. Ginting, and A. M. H. Sihite, "Deteksi Berita Palsu Tentang Vaksinasi Covid-19 Dengan Menggunakan Text Mining Dan Algoritma Cosine Similarity," *KOMIK (Konferensi …*, vol. 6, no. November, pp. 129–139, 2023, doi: 10.30865/komik.v6i1.5738.

[13] Y. Riadi Silitonga, "Sistem Pendeteksi Berita Hoax di Media Sosial dengan Teknik Data Mining Scikit Learn," *J. Ilmu Komput.*, vol. 4, p. 173, 2019, [Online]. Available: www.beritasatu.com,