Jurnal Rekayasa Sistem Informasi dan Teknologi
Volume 1, No 3 – Februari 2024
e-ISSN : 3025-888X

IMPLEMENTASI MACHINE LEARNING DALAM MENGANALISIS DAN MENDETEKSI BERITA PALSU PADA PORTAL BERITA BAHASA INGGRIS

Nur Amalia Hasma
Universitas Islam Kebangsaan Indonesia, Indonesia
nuramaliahasma05@gmail.com

Received: 20-01- 2024	Revised: 27-01-2024	Approved: 03-02-2024

ABSTRAK

Kemajuan teknologi informasi di era digital saat ini telah memberikan dampak yang signifikan terhadap akses informasi melalui media online. Hal ini disebabkan semakin populernya penggunaan media online karena biayanya yang murah dan akses yang mudah. Berita palsu dapat disebarkan melalui berbagai platform media, termasuk penggunaan website, media sosial, email, dan platform digital lainnya. Berita palsu disebarkan untuk berbagai tujuan, termasuk menghasilkan pendapatan iklan melalui clickbait, mempengaruhi opini publik mengenai peristiwa terkini, menyebarkan perselisihan, dan mempromosikan agenda tertentu.Dalam penelitian ini diusulkan metodologi untuk mendeteksi berita asli dan berita palsu dengan menggunakan metode seleksi fitur TF-IDF untuk melakukan klasifikasi dan menerapkan algoritma pembelajaran mesin yaitu, Support Vector Machine, Logistic Regression, Random Forest, dan Decision Tree. Hasil dari klasifikasi tersebut ditampilkan dalam bentuk confusion matrix. Berdasarkan hasil penelitian yang telah dilakukan algoritma Support Vector Machine menghasilkan akurasi sebesar 95,65% dengan nilai presisi 94,91%, disusul dengan algoritma Logistic Regression, algoritma Random Forest, dan algoritma Decision Tree yang memperoleh nilai akurasi terkecil sebesar 91,25%.

Kata kunci: Media Online, Berita Palsu, Algoritma Pembelajaran Mesin

PENDAHULUAN

Kemajuan teknologi informasi di era digital saat ini telah memberikan dampak yang signifikan terhadap akses informasi melalui media online. Hal ini disebabkan semakin populernya penggunaan media online karena biayanya yang murah dan akses yang mudah. Media online telah berkembang menjadi platform di mana masyarakat dapat mengakses, mengonsumsi, dan berbagi berita secara online menyebabkan peningkatan penyebaran berita palsu. Penyebaran berita palsu adalah penyebaran informasi palsu atau menyesatkan yang disengaja. Informasi berita palsu dibuat untuk meniru gaya dan format artikel berita sebenarnya, sehingga menyulitkan pembaca media untuk mengetahui secara pasti apakah informasi yang disampaikan benar adanya. Akibatnya, tidak jarang banyak informasi palsu dan dimanipulasi, termasuk rumor dan misinformasi, beredar di media online[1].

Berita palsu dapat disebarkan melalui berbagai platform media, termasuk penggunaan website, media sosial, email, dan platform digital lainnya. Berita palsu disebarkan untuk berbagai tujuan, termasuk menghasilkan pendapatan iklan melalui clickbait, mempengaruhi opini publik mengenai peristiwa terkini, menyebarkan perselisihan, dan mempromosikan agenda tertentu. Penyebaran berita palsu dalam skala besar dapat menimbulkan dampak negatif baik bagi pengguna individu maupun masyarakat secara keseluruhan. Berita palsu dapat

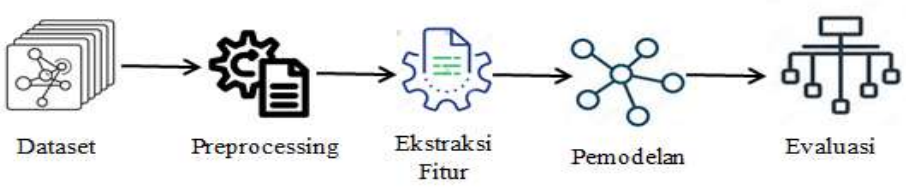
mengubah cara pandang seseorang dalam menafsirkan dan bereaksi terhadap berita nyata. Salah satu contoh penyebaran berita palsu yang paling menonjol adalah penyebaran informasi palsu yang bertujuan mempengaruhi opini publik dalam pemilihan umum presiden yang dilakukan setiap empat tahun sekali[2].

Sebelum meluasnya penggunaan Internet, penyebaran berita palsu sangat terbatas karena jurnalis bertugas memverifikasi dan memeriksa fakta informasi berdasarkan sumber yang relevan. Oleh karena itu, penyebaran berita palsu perlu diwaspadai untuk mencegah dampaknya[3]. Dalam mendeteksi berita palsu pre training model seperti TF-IDF, Count Vectorizer dan Word2Vec sering digunakan sebagai representasi matriks dalam mengubah setiap teks menjadi representasi numerik[4]. Selain itu TF-IDF juga banyak digunakan dalam proses ektraksi fitur yang digabungkan dengan berbagai algoritma pembelajaran mesin [5-9] sehingga mampu menghasilkan performa yang baik dalam mendeteksi berita palsu, selanjutnya hasil tersebut ditampilkan dalam bentuk confusion matrix berdasarkan nilai accuracy, presisi, recall dan F-1 Score [10-12].

Dalam penelitian ini deteksi berita palsu menggunakan proses ekstraksi fitur TF-IDF dengan menggabungkan empat algoritma pembelajaran mesin yaitu, Support Vector Machine, Logistic Regression, Random Forest dan Decision Tree dalam mendeteksi berita palsu berdasarkan kata, frasa, sumber berita, dan judulnya yang diperoleh melalui dataset deteksi berita palsu yang dapat diunduh pada situs Zenodo.org[13]. Kemudian hasil seleksi fitur yang diperoleh dapat digunakan untuk mendeteksi dan mengklasifikasikan berita asli dan berita palsu yang ditampilkan dalam bentuk confusion matrix.

METODE PENELITIAN

Dalam mendeteksi berita palsu dengan menggunakan teknik pembelajaran mesin pada platform berita online bahasa Inggris ditunjukkan pada gambar berikut ini :



Gambar 1. Metode Pengembangan Sistem

Dapat dilihat pada gambar 1, metode pengembangan sistem untuk mendeteksi berita valid dan berita palsu dilakukan secara berurutan dimulai dengan akuisisi data, diikuti dengan tahap data pra-pemrosesan (preprocessing), kemudian dilanjutkan dengan proses ekstraksi fitur, tahap pemodelan data dan yang terakhir adalah tahap klasifikasi data. Masing-masing proses tersebut dijelaskan lebih lanjut pada sub-bab berikut ini.

a. Dataset

Langkah pertama dalam penelitian ini adalah pengumpulan data dan pelabelan. Data yang digunakan dalam penelitian ini adalah WelFake dataset yang dapat diunduh website Zenodo.org[13]. Dataset ini dikumpulkan menggunakan teknologi *crawling* dari berbagai platform berita online berbahasa Inggris dan terdiri dari dataset *real news* dan dataset *fake news*. Setiap kumpulan data berisi lebih dari 72.134 item berita.

Tepatnya, dataset *real news* berisi 35.028 item data dan dataset *fake news* berisi 37.106 item data yang terdiri dari empat atribut: indexing (yang dimulai dari 0), title (yang berisi judul berita), text (isi berita) dan label. Data pada label yang semula bertanda "FAKE" atau "REAL" diubah dalam bentuk biner menjadi 0 untuk berita palsu (hoaks) dan 1 untuk berita asli.

b. Preprocessing
Tahap kedua dalam pengembangan sistem adalah preprocessing. Tujuan dari proses ini adalah untuk memahami properti data, memeriksa data, dan membersihkan data untuk digunakan pada tahap berikutnya. Tahap prapemrosesan ini melakukan sensitivitas huruf besar-kecil, mengubah seluruh karakter dalam data menjadi huruf kecil untuk memastikan konsistensi dan menggeneralisasi struktur data teks yang digunakan. Langkah selanjutnya setelah case folding adalah proses tokenisasi. Proses ini digunakan untuk menghapus karakter tertentu yang tidak digunakan dalam data. Tujuannya adalah untuk mengetahui kata mana yang mewakili kelas data tertentu. Proses terakhir yang dilakukan pada tahap preprocessing data adalah penghapusan stopword untuk mempercepat kinerja proses dan meningkatkan kinerja model. Fitur-fitur kata kemudian disusun kembali dan kata-kata yang diproses digabungkan menjadi unit lengkap yang dapat diproses menggunakan model pembelajaran mesin.

c. Ekstraksi Fitur
Pada tahap ini, proses ekstraksi fitur dilakukan setelah tahap preprocessing. Pada penelitian ini, proses ekstraksi fitur dilakukan dengan menggunakan metode TF-IDF (Term Frekuensi – Inverse Document Frekuensi). Metode TF-IDF digunakan untuk menentukan nilai frekuensi kata yang muncul dalam suatu dokumen atau artikel. Perhitungan yang dilakukan oleh fitur TF-IDF menggabungkan frekuensi kata untuk menentukan seberapa sering suatu kata muncul dalam suatu dokumen, kemudian melakukan perhitungan invers frekuensi dokumen untuk menentukan seberapa sering suatu kata muncul dalam suatu dokumen, dan kemudian mengetahui seberapa penting suatu kata dalam dokumen tersebut. Rumus penghitungan TF-IDF ditunjukkan di bawah ini.

$$W\ dt = TF\ dt \times IDF\ dt \quad (1)$$

Simbol W pada persamaan (1) menunjukkan bobot yang dimiliki dokumen ke-d ketika kata ke-t muncul di seluruh dokumen data. Di sisi lain, TF menunjukkan frekuensi kemunculan sebuah kata dalam dokumen, dan IDF menghitung kemunculan kata-kata ke-t pada seluruh dokumen. Untuk mendapatkan nilai IDF ditunjukkan pada persamaan berikut :

$$IDF = \log\log\frac{N}{dft} \quad (2)$$

d. Pemodelan
Langkah selanjutnya yang dilakukan dalam mendeteksi berita hoaks adalah pemodelan setelah tahap ekstraksi fitur. Alagoritma pembelajaran mesin digunakan untuk memodelkan data sehingga sistem dapat memprediksi apakah berita termasuk dalam kategori "berita asli" atau "berita palsu". Empat algoritma pembelajaran mesin yang digunakan dalam penelitian ini :

Support Vector Machine, Logistic Regression, Random Forest, dan Decision Tree

e. Evaluasi
Langkah terakhir dalam analisis dan deteksi berita asli dan palsu adalah confusion matrix untuk mengetahui kinerja model yang diusulkan dalam penelitian ini. Confusion Matrix merupakan salah satu metrik standar yang biasa digunakan untuk mengevaluasi kinerja hasil klasifikasi. Dalam penelitian ini digunakan pustaka Python Scikit Learn untuk memeriksa hasil prediksi dalam mendeteksi berita asli dan berita palsu.

Tabel 1. Confusion Matriks

Kelas yang dirediksi	**Kelas yang sebenarnya**	
	Positive	Negative
Positive	True Positive [TP]	False Positive [FP]
Negative	False Negative [FN]	True Negative [TN]

Berdasarkan tabel 1, confusion matrix memiliki bentuk persegi yang berfungsi untuk memetakan setiap kelas lainnya, baik dari sisi aktual maupun sisi prediksi dari sistem. Secara umum terdapat empat bagian proses pemetaan kelas dalam confusion matriks, yaitu :

a. True Positive
True Positive (TP) menunjukkan berapa banyak data yang diprediksi oleh sistem adalah positif, dengan nilai aktualnya juga positif.

b. True Negative
True Negative (TN) menunjukkan berapa banyak data yang diprediksi oleh sistem sebagai nilai negatif, dan nilai aktualnya juga negatif.

c. False Positive
False Postive (FP) menunjukkan berapa banyak data yang diprediksi oleh sistem sebagai nilai positif, sementara nilai aktualnya adalah negatif

d. False Negative
False Neagative (FN) menunjukkan berapa banyak data yang diprediksi oleh sistem sebagai nilai negatif, sedangkan nilai sebenarnya adalah positif.

Confusion matrix digunakan untuk menentukan indikator kinerja model klasifikasi, seperti akurasi, presisi, sensitivitas, dan F-1 Score untuk menentukan keberhasilan model pembelajaran mesin dan kinerja klasifikasi melalui persamaan (3), selanjutnya kinerja model klasifikasi ditampilkan berdasarkan matriks akurasi (Manning et al, 2008)

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (3)$$

Perhitungan akurasi berdasarkan persamaan diatas juga dapat dituliskan sebagai berikut :

$$Accuracy = \frac{data\ yang\ diklasifikasikan\ benar}{total\ data\ uji} \quad (4)$$

Ukuran akurasi menunjukkan seberapa efektif kinerja klasifikasi yang digunakan, yang diukur dari nilai persentase yang diberikan oleh algoritma, memprediksi dengan benar berdasarkan semua temuan prediksi. Selain itu, metrik akurasi dapat digunakan untuk mengevaluasi kinerja model klasifikasi, seperti halnya nilai presisi yang dihitung berdasarkan persamaan 5.

$$precision = \frac{TP}{TP + FP} \quad (5)$$

Nilai presisi merupakan ukuran untuk mengevaluasi kinerja klasifikasi dalam hal ketergantungan model dalam membuat prediksi positif, sehingga dapat menentukan seberapa efektif model mengidentifikasi berita palsu dari total hasil yang diperoleh sebagai berita palsu. Metrik selanjutnya yang dapat memberikan informasi mengenai tingkat keberhasilan model klasifikasi adalah metrik recall yang digunakan untuk menghitung proporsi data yang diprediksi positif oleh sistem dari seluruh data yang berlabel positif yang ditunjukkan pada persamaan 6.

$$recall = \frac{TP}{TP + FN} \quad (6)$$

Selanjutnya digunakan metrik F1-Score untuk menilai proporsi nilai presisi dan nilai recall yang dihitung

$$F1\ Score = \frac{2 \times precision \times recall}{precision + recall} \quad (7)$$

HASIL DAN PEMBAHASAN

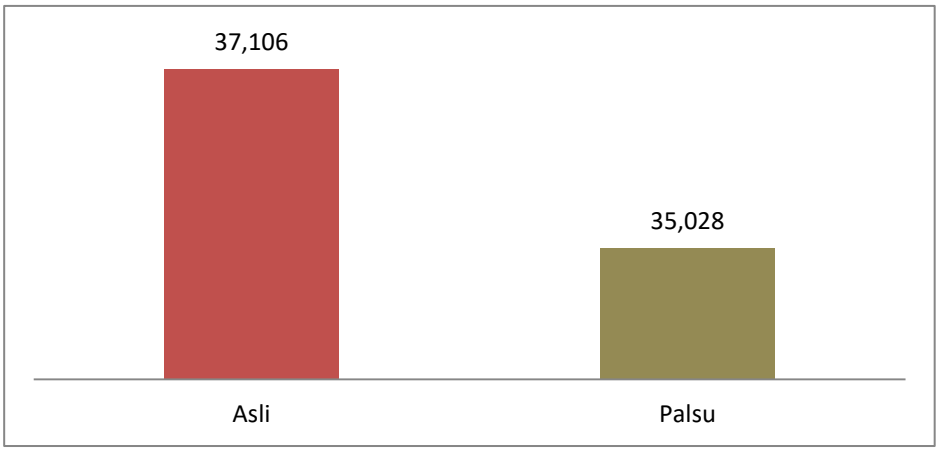
Penelitian terkait analisis dan deteksi berita palsu pada situs daring berbahasa Inggris telah dilakukan sesuai dengan penjelasan pada bagian metodologi. Dalam penelitian ini dilakukan pendekatan analitis untuk mengklasifikasikan berita valid dan berita hoaks berdasarkan hasil yang diperoleh dari seleksi fitur dan confusion matrix. Dataset yang digunakan dalam penelitian ini adalah dataset Hoax News Detection dari situs Zenodo.org yang ditunjukkan pada tabel 2 berikut ini.

Tabel 2. Contoh dataset berita hoaks bahasa Inggris

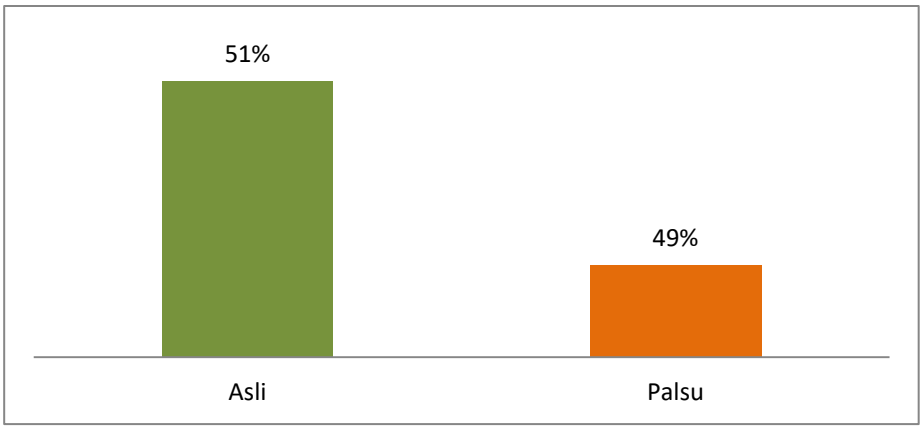
	Unnamed: 0	title	text	label
0	0	LAW ENFORCEMENT ON HIGH ALERT Following Threat...	No comment is expected from Barack Obama Membe...	1
1	1	NaN	Did they post their votes for Hillary already?	1
2	2	UNBELIEVABLE! OBAMA'S ATTORNEY GENERAL SAYS MO...	Now, most of the demonstrators gathered last ...	1
3	3	Bobby Jindal, raised Hindu, uses story of Chri...	A dozen politically active pastors came here f...	0
4	4	SATAN 2: Russia unvelis an image of its terrif...	The RS-28 Sarmat missile, dubbed Satan 2, will...	1

Tabel 2 menunjukkan beberapa contoh data berita pada media online berbahasa Inggris untuk deteksi berita hoaks. Pada dataset tersebut terdapat dua kolom utama yaitu *text* (berita) dan *label* (penandaan). Kolom text adalah isi artikel berita yang diambil dari berbagai sumber berita pada media online berbahasa Inggris, sedangkan kolom label merupakan pemberian label berita asli dan berita palsu

dengan melakuan proses tag voting yang dilakukan oleh annotator pada saat melakukan pelabelan data. Label data yang awalnya berlabel 'Real' dan 'Fake' diubah menjadi bilangan biner 0 untuk berita palsu dan 1 untuk berita asli. Dari dataset yang diperoleh, langkah selanjutnya yang perlu dilakukan adalah memperoleh infomasi penting yang dapat membantu proses analisis berjalan lebih optimal.



Gambar 2 Distribusi berita asli dan berita palsu

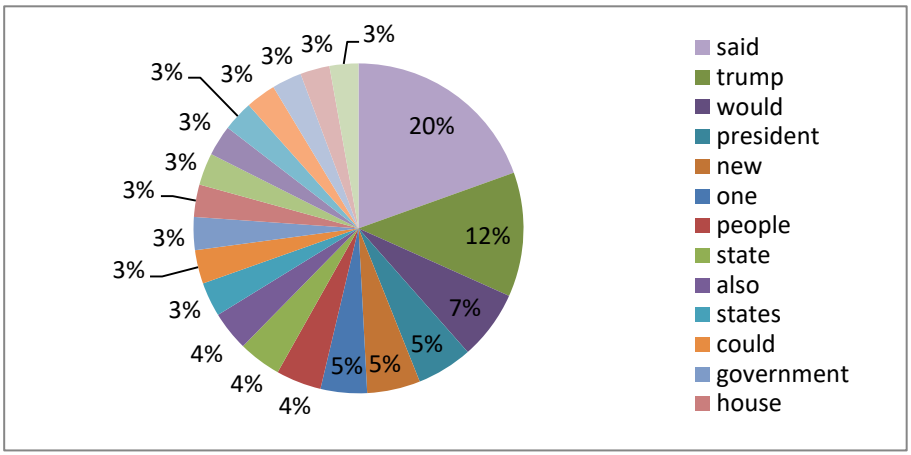


Gambar 3 Rasio penyebaran berita asli dan berita palsu

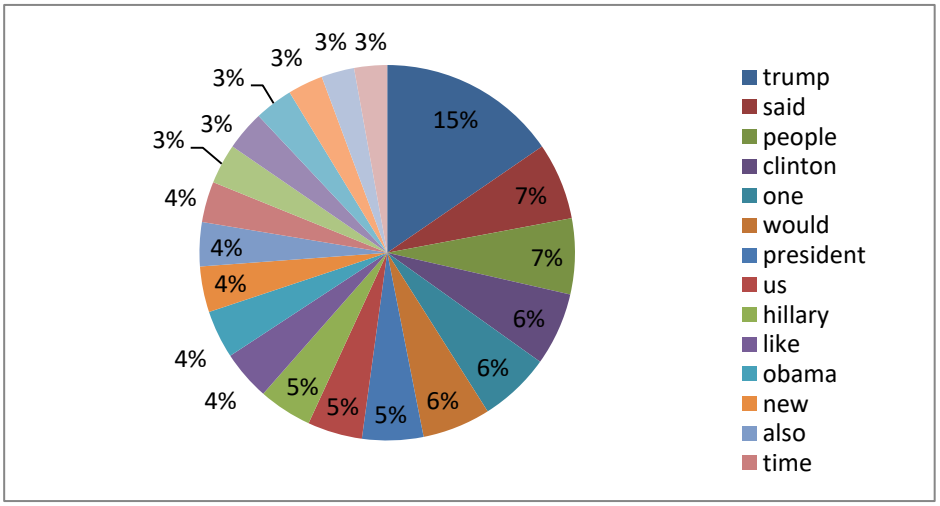
Berdasarkan informasi yang ditunjukkan pada gambar 2 dan gambar 3, perolehan distribusi berita asli sebanyak 37.106 dan berita palsu sebanyak 35.028 dengan rasio penyebaran berita yang tidak jauh berbeda, dimana rasio penyebaran berita asli sebesar 51% sedangkan rasio penyebaran berita palsu sebesar 49%. Hal yang perlu dilakukan selanjutnya adalah melakukan pre-processing data. Tahapan pre-processing dilakukan dengan cara melakukan case folding untuk mengubah huruf yang terdapat pada artikel berita menjadi huruf kecil untuk memastikan keseragaman isi konten, selain itu karakter khusus, angka dan tanda baca juga dihapus. Proses selanjutnya yang dilakukan adalah melakukan proses tokenisasi, proses ini berfungsi untuk membagi teks menjadi blok yang ditampilkan per kata. Dilanjutkan dengan proses stopword removal untuk membantu mengurangi *noise* atau informasi yang tidak relevan yang terdapat pada sebuah teks dan stemming yang berfungsi untuk mengubah kata yang digunakan kembali ke bentuk dasarnya.

Tahapan pre processing dilakukan dengan menggunakan library NLP

NLTK dan modul Spacy. Setelah pre processing selesai dapat dilihat informasi yang ditampilkan pada dataset terkait frekuensi kemunculan kata yang terdapat pada berita asli dan berita palsu. Untuk menentukan frekuensi kemunculan kata tersebut dapat menggunakan library counter dengan cara menghitung jumlah kata yang sering muncul pada berita asli dan berita palsu misalnya, sebanyak 20 kata untuk melihat kata-kata dan informasi apa saja yang dapat diberikan.



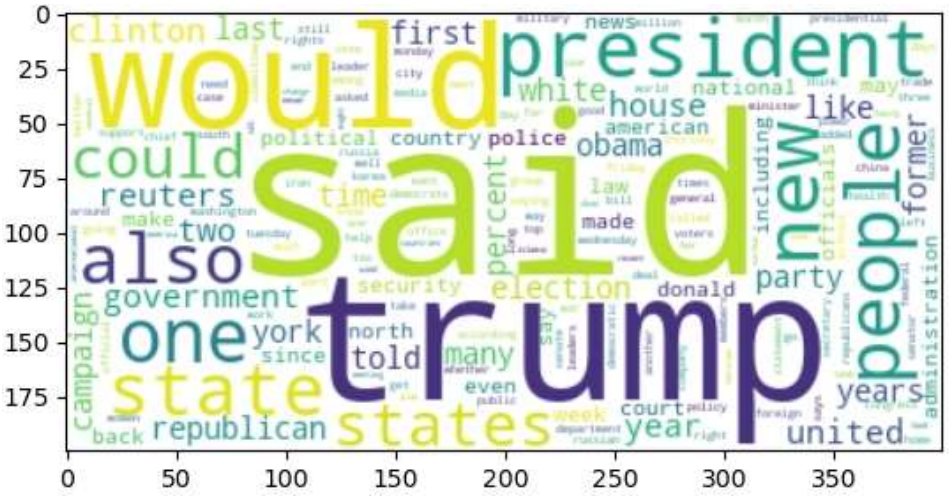
Gambar 7 Kata teratas yang muncul dalam berita asli



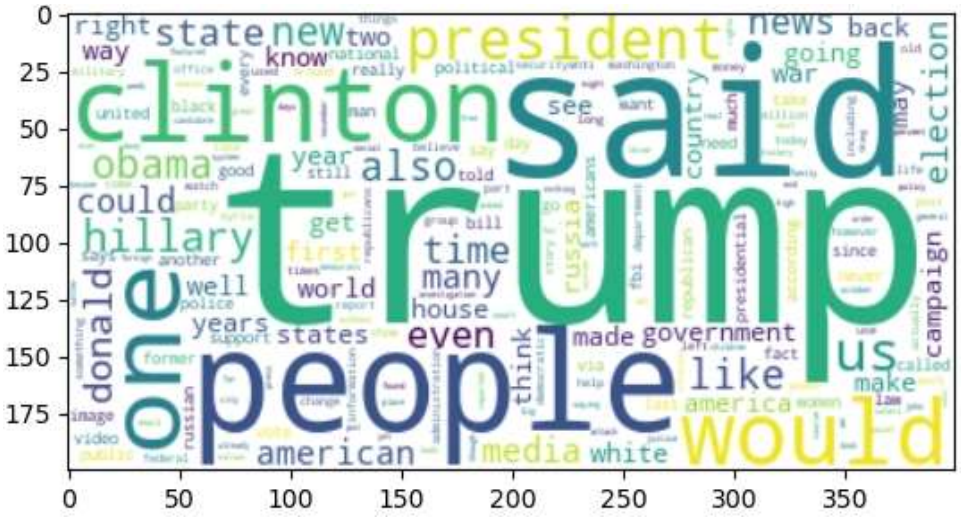
Gambar 8 Kata teratas yang muncul dalam berita palsu

Berdasarkan hasil diagram yang ditampilkan pada gambar 7 dan gambar 8 informasi yang diberikan terkait frekuensi kemunculan kata dalam berita asli dan berita palsu adalah mengenai pemilihan presiden AS. Terdapat penggunaan kata yang sama digunakan dalam berita asli dan berita palsu. Pada berita asli frekuensi kemunculan kata *said* adalah yang tertinggi sebesar 20% diikuti dengan kata *trump* sebesar 12%. Sementara itu pada berita palsu frekuensi kemunculan kata *trump* adalah yang tertinggi sebesar 15% diikuti dengan kemunculan kata *said* sebesar 7%. Selain itu juga terdapat beberapa kata yang muncul dalam berita asli dan berita palsu, seperti kata *people, would, could, one, new, also, president,* dan *state*. Selanjutnya dengan menggunakan library WordCloud dapat menampilkan visualisasi frekuensi kemunculan kata-kata yang ditunjukan pada gambar 7 dan gambar 8 dalam bentuk sekumpulan kata yang dapat dilihat pada gambar 9 dan

gambar 10.

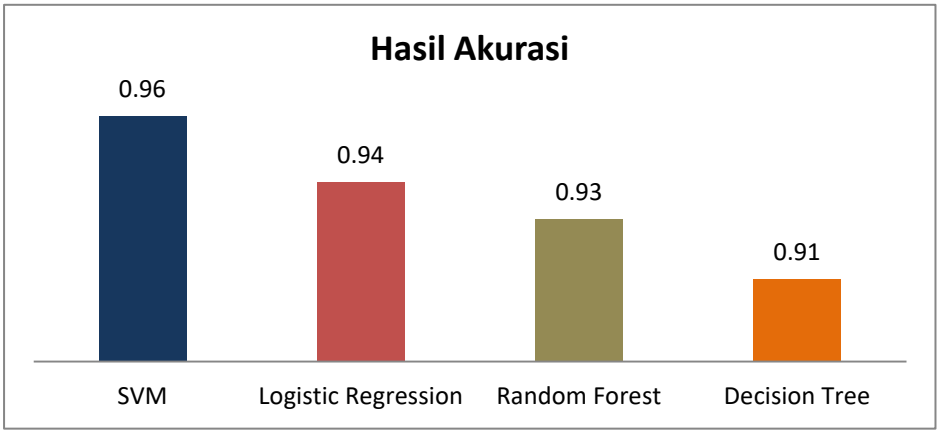


Gambar 9 Visualisasi sekumpulan kata berita asli



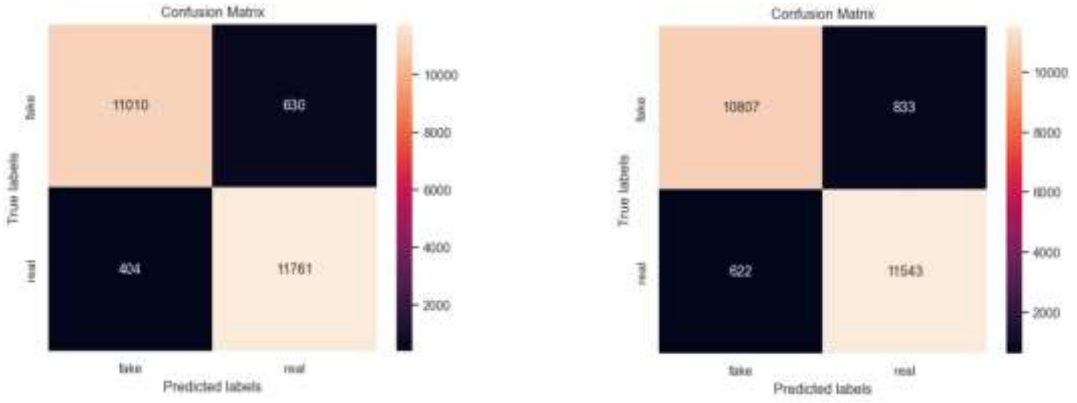
Gambar 10 Visualisasi sekumpulan kata berita palsu

Proses selanjutnya yang dilakukan adalah ekstraksi fitur menggunakan TF-IDF vectorization yang berfungsi untuk mengkonversi teks menjadi representasi numerik berdasarkan nilai frekuensi kata yang muncul dalam suatu dokumen atau artikel. Selanjutnya dari proses ekstraksi fitur tersebut dilanjutkan dengan proses klasifikasi dengan menerapkan empat algoritma pembelajaran mesin, yaitu : Support Vector Machine, Logistic Regression, Random Forest, dan Decision Tree. Proses dilakukan dengan membagi dataset menjadi 80 untuk proses training dan 20 untuk proses testing.

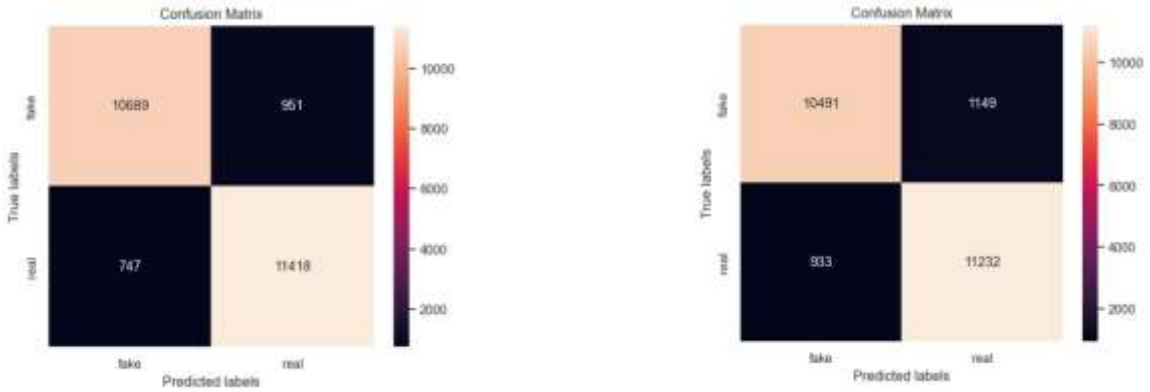


Gambar 11 Perolehan Akurasi

Berdasarkan perolehan akurasi pada gambar 11, algoritma Support Vector Machine memperoleh hasil akurasi tertinggi sebesar 95,65%, diikuti dengan algoritma Logistic Regression sebesar 93,88%, kemudian algoritma Random Forest menghasilkan nilai akurasi sebesar 92,68% dan algoritma Decision Tree menghasilkan nilai akurasi paling kecil yaitu sebesar 91,25%. Berdasarkan perolehan hasil akurasi tersebut diperoleh hasil berupa confusion matrix ketika proses klasifikasi dilakukan. Hasil dari confusion matrix untuk masing-masing algoritma pembelajaran mesin yang digunakan pada penelitian ini ditunjukkan pada gambar 12 dan gambar 13.



Gambar 12 Hasil confusion matrix SVM dan Logistic Regression



Gambar 13 Hasil confusion matrix Random Forest dan Decision Tree

Dari hasil confusion matrix yang ditunjukkan pada gambar 12 dan gambar 13 memperlihatkan bahwa label true positive memiliki nilai lebih besar dibandingkan nilai yang didapatkan pada label true negative. Hal tersebut dapat diartikan bahwa, model yang digunakan dalam mendeteksi berita asli dan berita palsu yang dilakukan dalam penelitian ini sudah berjalan dengan cukup baik dengan perolehan akurasi tertinggi dari algoritma Support Vector Machine sebesar 95,65%. Hasil evaluasi keseluruhan confusion matrix yang berisi nilai accuracy, precision, recall, dan F-1 Score ditampilkan pada tabel 3.

Tabel 3 Hasil Evaluasi Confusion Matrix

Algoritma	Accuracy	Precision	Recall	F1-Score
Support Vector Machine	95,65%	94,91%	99,67%	95,00%
Logistic Regression	93,88%	93,26%	94,88%	94,04%
Random Forest	92,68%	92,31%	93,85%	93,00%
Decision Tree	91,25%	90,71%	92,33%	91,00%

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat diambil kesimpulan bahwa pemodelan yang dilakukan dalam mendeteksi berita palsu pada media online bahasa Inggris dengan menerapkan algoritma pembelajaran mesin sudah berjalan dengan cukup baik. Hal tersebut ditunjukkan dengan hasil yang diperoleh pada label true positive memiliki nilai lebih besar dibandingkan nilai yang didapatkan pada label true negative dengan perolehan akurasi tertinggi menggunakan algoritma Support Vector Machine sebesar 95,65% dengan nilai presisi 94,91%, disusul dengan algoritma Logistic Regression, algoritma Random Forest, dan algoritma Decision Tree yang memperoleh nilai akurasi terkecil sebesar 91,25%.

DAFTAR PUSTAKA

[1]Mridha, M. F., Member, S., Keya, A. J., Hamid, A., Monowar, M. M., & Rahman, S. (2021). A Comprehensive Review on Fake News Detection with Deep Learning. *IEEE Access*, *PP*, 1. https://doi.org/10.1109/ACCESS.2021.3129329

[2]Mahyoob, M., Algaraady, J., & Alrahaili, M. (2020). Linguistic-Based Detection of Fake News in Social Media Linguistic-Based Detection of Fake News in Social Media, (November). https://doi.org/10.5539/ijel.v11n1p99

[3]Seddari, N., Derhab, A., Belaoued, M., Halboob, W., Al-muhtadi, J., & Bouras, A. (2022). A Hybrid Linguistic and Knowledge-Based Analysis Approach for Fake News Detection on Social Media. *IEEE Access*, *10*, 62097–62109. https://doi.org/10.1109/ACCESS.2022.3181184

[4]Vijayaraghavan, S., Wang, Y., Voong, J., Nasseri, A., Li, L., & Xu, W. (n.d.). Fake News Detection with Different Models.

[5]Aljwari, F. (2022). Multi-scale Machine Learning Prediction of the Spread of Arabic Online Fake News, *13*, 1–14.

[6]Lakshmi, V. D. (2022). Detection of Fake News using Machine Learning Models, *183*(47), 22–27.

[7]Muslim, I., Karo, K., Dewi, S., & Fadilah, P. M. (2023). Hoax Detection on Indonesian Tweets using Naïve Bayes Classifier with TF-IDF, *4*(3), 914–919. https://doi.org/10.47065/josh.v4i3.3317

[8]Prachi, N. N., Rafi, E. H., Alam, E., & Khan, R. (2022). Detection of Fake News Using Machine Learning and Natural Language Processing Algorithms, *13*(6), 652–661. https://doi.org/10.12720/jait.13.6.652-661

[9]Jatain, A., & Vashisht, P. (2023). Fake News Detection Model Using Machine Learning, *1*(1), 60–66.

[10]Yanuar, I., Pratiwi, R., & Nugraha, A. F. (2022). Hoax news identification using machine learning model from online media in Bahasa Indonesia, *12*(2), 58–67.

[11]Pandey, S., Prabhakaran, S., Huang, J., & Zhao, Z. (2021). Fake News Detection Using Machine Learning Approaches. https://doi.org/10.1088/1757-899X/1099/1/012040

[12]Verma, P. K., Agrawal, P., Amorim, I., & Prodan, R. (n.d.). WELFake : Word Embedding Over Linguistic Features for Fake News Detection, 1–13. https://doi.org/10.1109/TCSS.2021.3068519

[13]Verma, P. K., Agrawal, P., Amorim, I., & Prodan, R. (n.d.). WELFake dataset for fake news detection in text data. 2021[Online]. Available https://doi.org/10.5281/zenodo.4561253. [Accessed : 18-10-2023].