Pendekatan *Macine Learning* dalam Evaluasi Label Berita Berdasarkan Judul: Studi Kasus Media Online

Rezky Yuranda[1]*, Tata Sutabri[2], Delpiah Wahyuningsih[3]
Program Studi Magister Teknik Informatika Universitas Bina Darma[1], [2],
Program Studi Teknik Informatika ISB Atma Luhur [3]
Palembang[1][2], Indonesia
Pangkalpinang[3], Indonesia
yurandarezky@gmail.com[1], tata.sutabri@gmail.com[2], delphibabel@atmaluhur.ac.id[3]

***Abstract*— In the current digital era, information availability is abundant, and news serves as a primary source of up-to-date and reliable information for the public. However, with the increasing volume of information, a robust evaluation method is necessary to ensure accurate and dependable news labeling. This research employs a machine learning approach, utilizing three common classification algorithms: Naive Bayes, SVM, and Random Forest, to evaluate news labels based on their titles. The dataset utilized in this study is obtained from Jakarta AI Research and consists of 10,000 samples covering various news topics. Evaluation is conducted using accuracy, precision, recall, and F1-Score metrics to gain a comprehensive understanding of the classification algorithm's performance. The results of this research demonstrate that the SVM algorithm exhibits the best performance, achieving an accuracy rate of 92.92%. Random Forest follows with an accuracy rate of 91.21%, and Naive Bayes with an accuracy rate of 89.61%. These findings provide deep insights into the effectiveness of the machine learning approach in evaluating news labels based on their titles. Furthermore, the study highlights the importance of considering other evaluation metrics such as precision, recall, and F1-Score to obtain a more holistic understanding of the algorithm's performance. Further research is encouraged to involve additional classification algorithms and more diverse and extensive datasets to enhance the comprehension of news label evaluation comprehensively. Such endeavors can significantly contribute to the development of automated systems for classifying news with higher accuracy and reliability in the future**

Keywords: News label evaluation, machine learning, Naive Bayes, SVM, Random Forest

***Abstrak*—Dalam era digital saat ini, ketersediaan informasi sangat melimpah. Berita menjadi sumber informasi terkini dan terpercaya bagi masyarakat. Namun, dengan informasi yang semakin banyak, diperlukan sebuah metode evaluasi yang baik untuk memastikan label berita yang akurat dan dapat diandalkan. Penelitian ini menggunakan pendekatan machine learning dengan menggunakan tiga algoritma klasifikasi umum: Naive Bayes, SVM, dan Random Forest untuk mengevaluasi label berita berdasarkan judul. Dataset yang digunakan berasal dari Jakarta AI Research dan mencakup berbagai topik berita sebanyak 8754 sample. Evaluasi dilakukan dengan metrik akurasi, presisi, recall, dan F1-Score guna mendapatkan gambaran komprehensif tentang performa algoritma klasifikasi. Hasil penelitian ini menunjukan bahwa algoritma SVM memiliki kinerja terbaik, dengan tingkat akurasi terbaik mencapat 92.92%. Diikuti Random Forest dengan tingkat akurasi 91.21% dan Naive Bayes dengan tingkat akurasi 89.61%. Temuan ini memberikan wawasan mendalam tentang efektivitas pendekatan machine learning dalam evaluasi label berita berdasarkan judul. Selain itu, hasil penelitian ini juga menekankan pentingkanya mempertimbangkan metrik evaluasi lainnya seperti presisi, recall, dan F1-Score untuk mendapatkan pemahaman yang lebih holistik mengenai kinerja algoritma. Penelitian lebih lanjut dianjurkan untuk melibatkan lebih banyak algortima klasifikasi serta dataset yang lebih luas dan beragam guna meningkatkan pemahaman evaluasi label berita secara lebih komprehensif. Hail ini dapat memberikan sumbangan penting dalam pengembangan sistem otomatis untuk mengklasifikasi berita dengan akurasi dan keandalan yang lebih tinggi di masa depan.**

Kata Kunci: Evaluasi label berita, machine learning, naive bayes, svm, random forest

I. PENDAHULUAN

Dalam era digital saat ini, ketersediaan informasi sangat melimpah. Berita menjadi bagian penting dari informasi tersebut, karena berita menjadi sumber informasi terkini dan terpercaya bagi masyarakat [1].

Bersamaan dengan pertumbuhan teknologi saat ini, terjadi pula peningkatan jumlah informasi tekstual yang tersedia di dalamnya. Data dari *Indonesia Digital Association* (IDA) menunjukan bahwa *96%* masyarakat Indonesia mengkonsumsi berita secara online [2].

Pada sebuah berita, judul utama atau headline merupakan elemen penting dari setiap halaman web. Judul tersebut menentukan apakah pembaca akan tetap berinteraksi dengan konten atau pergi ke halaman lain. Begitu juga dengan kategori atau label yang membantu pembaca untuk menemukan konten berdasarkan topik dan memberikan konteks pada artikel yang di baca [3]. Oleh sebab itu, maka penting untuk memiliki metode evaluasi yang baik dalam mengelola label berita untuk memastikan informasi yang disajikan akurat dan dapat diandalkan bagi konsumen [2].

Dalam hal ini, penggunaan dataset berkualitas dan repesentatif sangat penting dalam melakukan penelitian dan pengembangan algoritma *AI* untuk evaluas label berita. Salah satu dataset yang digunakan dalam penelitian ini berasal dari *Jakarta AI Research*. *Jakarta AI Research* atau yang biasa disebuat *Jakarta Research* adalah sebuah komunitas riset dibidang kecerdasan buatan yang meliputi natural language

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v12i3.1808, Copyright ©2023
Submitted : June 19, 2023, Revised : July 20, 2023, Accepted : July 26, 2023, Published : November 7, 2023

processing, computer vision, dan speech processing dan juga intersection antara field tersebut dengan machine learning dan deep learning [13].

Dataset yang disediakan oleh *Jakarta AI Reserach* menawarkan akses ke beberapa media berita online yang mencakut topik yang beragam. Kumpulan data ini mencakup judul-judul berita yang dapat dijadikan dasar untuk melakuka analisis dan evaluasi label dari berita dengan menggunakan pendekatan machine learing. Dalam penelitian ini, akan memanfaatkan dataset tersebut untuk melihat sejauh mana pendekatan machine learning dapat memberikan hasil yang akurat dalam mengevaluasi label berita berdasarkan judul.

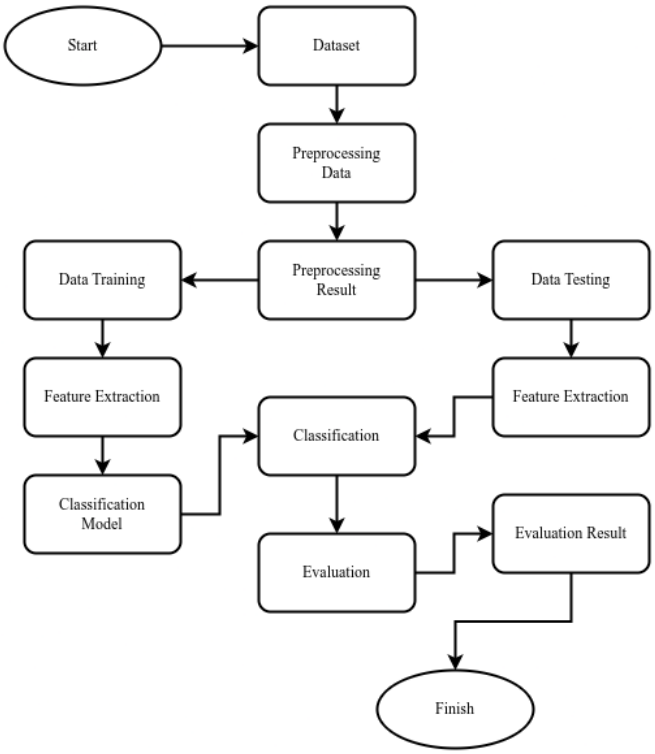
https://journals.usm.ac.id/index.php/transformatika/article/view/2317/1668

Pada penelitian ini, digunakan beberapa algoritma klasifikasi yang umum digunakan, seperti *Naive Bayes*, *Support Vector Machines* (*SVM*), dan *Random Forest*, untuk membangun model klasifikasi label berita berdasarkan judul. Algoritma-algoritma ini memiliki keunggukan masing-masing dalam memahami pola dan hubungan dalam data judul berita, yang kemudian akan digunakan untuk mengklasifikasi berita menjadi berbagai label atau kategori.

Dalam eksperimen ini akan dibandingkan performa ketiga algoritma klasifikas tersebut berdasarkan berbagai metrik evaluasi, seperti *akurasi*, *presisi*, *recall*, dan *F1-Score*. Hasil penelitian ini diharapkan dapat memberikan pemahaman lebih baik tentang efektivitas pendekatan machine learning dalam proses evaluasi label berita berdasarkan judul.

II. METODELOGI PENELITIAN

Metodelogi penelitian untuk pendekatan machine learning dalam evaluasi label berita berdasarkan judul, ditampilkan pada gambar 1.



Gambar 1. Alur penelitian

A. *Dataset*

Langkah pertama mempersiapkan dataset yang akan digunakan untuk pelatihan dan pengujian. Dataset yang digunakan pada penelitian ini diambil dari *Jakarta AI Reserach* yang berisikan informasi judul berita beserta label atau kategorinya sebanyak 8754 data. Pada dataset ini, terdapat didefinisikan 5 label, yaitu bola, news, bisnis, tekno, dan otomotif. Untuk distribusi data per label dapat dilihat pada table 1.

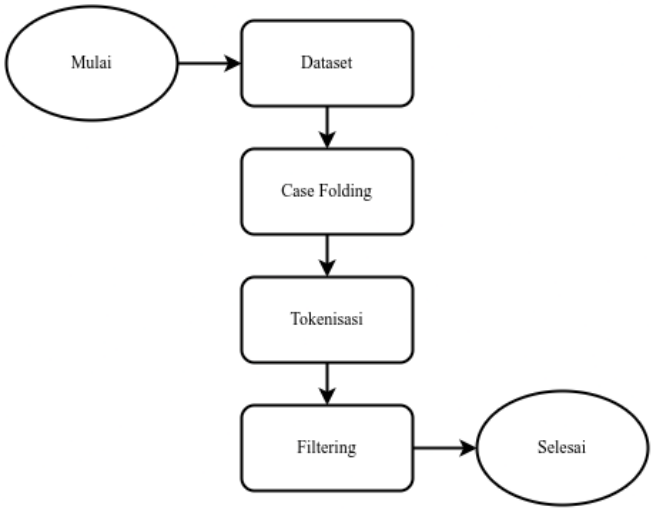
TABEL 1. DISTRIBUSI DATA SETIAP LABEL

No	***Label***	***Jumlah***
1	Bola	2986
2	News	2897
3	Bisnis	1881
4	Tekno	790
5	Otomotif	200
Jumlah Total		**8754**

B. *Preprocessing Data*

Tahapan preprocessing yang sangat penting pada data mining, alasan utamanya adalah karena kualitas dari input data sangatmempengaruhi kualitas output analisis yang dihasilkan [4]. Tahapan Preprocessing data dapat dilihat pada gambar 2.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v12i3.1808, Copyright ©2023
Submitted : June 19, 2023, Revised : July 20, 2023, Accepted : July 26, 2023, Published : November 7, 2023



Gambar 2. Preprocessing data

C. Feature extraction

Pada tahap *feature extraction*, dilakukan extraksi fitur dar teks judul berita menggunakan metode *TF-IDF* (*Term Frequency-Inverse Document Frequency*). Metode ini digunakan untuk mengukur tingkat pentingnya sebuah kata dalam suatu dokumen atau dataset berdasarkan frekuensi kemunculan kata dalam dokumen tersebut [6]. Kata-kata terpilih dianggap lebih informatif dan efektif untuk dieksekusi pada model classifier [7][8].

Penerapan teknik *TF-IDF* dilakukan setelah *preprocessing* data dan dilakukan sebelum proses klasifikasi data. Secara lebih detail, *TF-IDF* menghitung skor atau bobot untuk setiap kata dalam dokumen dengan mempertimbangkan dua faktor yaitu *Term Frequency* (*TF*) dan *Inverse Document Frequency* (*IDF*) [9].

$$TF = \frac{fk}{ft} \qquad (1)$$

Pada persamaan (1), *fk* adalah jumlah kemunculan kata dalam suatu kalimat dan *ft* adalah total kata dalam suatu dokumen. *IDF* menggambarkan seberapa umum atau jarang kata tersebut muncul dalam seluruh dataset. *Inverse Document Frequency* dihitung dengan menggunakan rumus

$$IDF = log \qquad (2)$$

Dimana *n* pada konteks ini merujuk pada total jumlah dokumen pada suatu korpus, sementara *df* adalah singkatan dari jumlah dokumen yang mengandung kata tertentu pada suatu korpus.

D. Classification model

Dalam metode pendekatan machine learing dalam evaluasi label berita berdasarkan judul pada penelitan ini menggunakan tiga algoritma klasifikasi yang berbeda yaitu *Naive Bayes*, *Support Vector Machine* (*SVM*), dan *Random Forest*.

Naive Bayes merupakan salah satu algoritma klasifikasi yang berdasarkan pada *Teorema Bayes* [10]. Algoritma ini memprediksi label berita berdarkan probalitas kelas yang diestimasikan dari fitur-fitur yang ada.

SVM (*Support Vector Machines*) adalah algoritma klasifikasi yang bekerja dengan mencari *hyperplane* terbaik yang memisahkan dua kelas berita berdasarkan fitur-fitur nya. *SVM* memiliki kemampuan untuk menangani data yang kompleks dan memiliki kelebihan dalam menangani dataset yang memiliki dimensi yang tinggi [11].

Random Forest merupakan algoritma klasifikasi ensemble yang membangun sejumlah pohon keputusan secara acak dan mengkombinasikan hasil prediksi dari setiap pohon untuk menghasilkan prediksi akhir. Dengan cara ini, *Random Forest* dapat mengatasi overfitting pada dataset dan menghasilkan hasil prediksi yang lebih stabil [11].

Dengan menggunakan kombinasi dari *Naive Bayes*, *SVM*, dan *Random Forest*, dapat dibandingkan dan dilakukan evaluasi performa dari ketiga algoritma ini dalam memprediksi dan mengklasifikasi label berita berdasarkan fitur-fitur yang diekstraksi.

E. Classification

Pada tahap ini, model klasifikasi yang telah di latih akan digunakan untuk memprediksi label berita berdasarkan fitur-fitur yang telah di ekstraksi sebelumnya. Proses ini melibatkan penggunaan rumus atau algoritma tertentu yang dapat memberikan prediksi yang akurat.

Pada algoritma *Naive Bayes*, prediksi label berita dilakukan dengan menggunakan *Teorema Bayes*. Untuk setiap kategori *k*, nilai probalitas posterior dapat dihitung dengan persamaan 3.

$$P(k \vee x) = \frac{(p(x \vee k) * p(k))}{p(x)} \qquad (3)$$

Dimana P(*k*/*x*) adalah probabilitas kondisional dari *k* terhadap *x*, P(*x*/*k*) adalah probabilitas kondisional dari *x* terhadap *k*, P(*k*) adalah probabilitas prior dari *k*, dan P(*x*) adalah probabilitas margin dari *x*.

Pada *SVM*, prediksi label berita dilakukan dengan mencari hyperplane terbaik yang memisahkan dua kelas berdasarkan fitur-fitur yang ada. Rumus untuk memprediksi label pada SVM terdapat pada persamaan 4.

$$f(x) = sign(\sum_{i=1}^{n} \alpha_i \, y_i K(x_i, x) + b) \qquad (4)$$

Dimana *x* adalah input sample yang ingin di prediksi, *xi* adalah training sample ke-*i*, *yi* adalah class label untuk training sample ke-*i* (mengambil nilai -1 atau 1), α*i* adalah koefisien untuk traiing sample ke-*i* yang dihitung oleh algoritma SVM selama proses training, K(*xi*,*x*) adalah kernel function yang mengukur kesamaan anatar input sample *x* dan training sample ke-*i*, *b* adalah bias term yang juga dihitung oleh algoritma *SVM* selama proses training, dan *n* adalah jumlah total training sample.

Dalam *Random Forest*, prediksi label berita dilakukan dengan menggabungkan hasil prediksi dari banyak pohon

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v12i3.1808, Copyright ©2023
Submitted : June 19, 2023, Revised : July 20, 2023, Accepted : July 26, 2023, Published : November 7, 2023

keputusan yang terbentuk. Setiap pohon memberikan suara dalam memprediksi label, dan label yang paling banyak terpilih menjadi prediksi akhir. Rumus yang digunakan dalam prediksi Random Forest tidak memiliki matematis khusus, tetapi melibatkan agregasi hasil prediksi dari setiap cabang. .

F. Evaluasi

Pada tahap ini, dilakukan pengukuran performa model klasifikasi yang telah dibangun menggunakan metrik-metrik seperti *akurasi*, *presisi*, *recall*, dan *F1-Score*.

Akurasi adalah metrik yang mengukur sejauh mana model mampu memberikan prediksi yang benar secara keseluruhan. *Presisi* mengukur sejauh mana prediksi positif yang diberikan oleh model adalah benar, sementara *recall* mengukur sejauh mana model mampu mendeteksi semua kasus positif yang sebenarnya. *F1-Score* adalah metrik yang menggabungkan presisi dan recal untuk memberikan gambaran yang lebih holistik tentang performa model.

Dalam tahapan evaluasi, kita menganalisis hasil prediksi model klasifikasi dengan membandingkan prediksi dengan label sebenarnya. Dengan menghitung *akurasi*, *presisi*, *recal* dan *F1-score*, kita dapat memperoleh pemahaman yang lebih mendalam tentang kualitas model klasifikasi dalam tugas evaluasi label berita berdasarkan judul. Metrik-metrik ini akan membantu kita dalam hal mengevaluasi tingkat keberhasilan model dalam memprediksi label berita dengan akurasi tinggi, dan juga membantu kita memahami seberapa baik model dapat menghindari kesalahan prediksi yang mungkin terjadi [12].

III. HASIL DAN PEMBAHASAN

A. Preprocessing Data

Case-folding, yaitu tahapan untuk mengkonversi semua karakter dengan huruf kapital menjadi huruf kecil. Proses l*owercasing* dapat dilihat pada table 2.

TABEL 2. PROSEES CASE-FOLDING

Data Asli	***Case-folding***
London - Lee Dixon khawatir Arsenal tak bisa merekrut Denis Suarez secara permanen musim panas nanti!!	london - lee dixon khawatir arsenal tak bisa merekrut denis suarez secara permanen musim panas nanti!!

Punctuation Removal, karakter khusus atau tanda baca seperti titik, koma, tanda tanya, tanda seru, dan karakter khusus lainnya dihapus dari teks. Proses punctuation removal dapat dilihat pada table 3.

TABEL 3. PROSES PUNCTUATION REMOVAL

Data Hasil Case-folding	***Punctuation Removal***
london - lee dixon khawatir arsenal tak bisa merekrut denis suarez secara permanen musim panas nanti!!	london lee dixon khawatir arsenal tak bisa merekrut denis suarez secara permanen musim panas nanti

Tokenisasi, adalah proses pemotongan kata menjadi satu kata pada setiap data. Proses tokenisasi dapat dilihat pada table 4.

TABEL 4. PROSES TOKENISASI

Data Hasil Punctuation Removal	***Tokenisasi***
london - lee dixon khawatir arsenal tak bisa merekrut denis suarez secara permanen musim panas nanti!!	['london', 'lee', 'dixon', 'khawatir', 'arsenal', 'tak', 'bisa', 'merekrut', 'denis', 'suarez', 'secara', 'permanen', 'musim', 'panas', 'nanti']

Filtering, adalah proses terakhir dalam *preprocessing* data, pada tahap ini akan menyaring kata penghubung yang bersifat umum dan tidak berpengaruh terhadap proses pembagian kelas sehingga kata penghubung tersebut tidak diperlukan. Pada tahap ini juga hasil tokenisasi akan digabungkan kembali menjadi kalimat utuh kembali. Proses filtering dapat dilihat pada table 5.

TABEL 5. PROSES TOKENISASI

Data Hasil Tokenisasi	***Filtering***	***Finishing***
['london', 'lee', 'dixon', 'khawatir', 'arsenal', '**tak**', '**bisa**', 'merekrut', 'denis', 'suarez', '**secara**', 'permanen', 'musim', 'panas', '**nanti**']	['london', 'lee', 'dixon', 'khawatir', 'arsenal', 'merekrut', 'denis', 'suarez', 'permanen', 'musim', 'panas']	Lodon lee dixon khawatir arsenal merekrut denis suarez permanan musim panas

B. Feature extraction

Berikut ini adalah sebagian hasil output dari pembobotan kata menggunakan metode *Term Frequency - Inverse Document Frequency* (TF-IDF) yang dapat dilihat pada Gambar 3.

```
(0, 43669)    0.057985814282321524
(0, 9269)     0.08402188022875495
(0, 37360)    0.08161037736900699
(0, 6297)     0.038826642368675224
(0, 7200)     0.04984697959846997
(0, 21190)    0.051249282037551386
(0, 35941)    0.08402188022875495
(0, 34644)    0.09151269230153071
(0, 2301)     0.061719635833768875
(0, 34640)    0.07359229206906777
(0, 2165)     0.07964003606623181
:     :
(7002, 41888) 0.07923955782363404
(7002, 526)   0.05869590141705259
(7002, 20763) 0.05009213389423332
(7002, 2744)  0.06877615482504694
(7002, 2300)  0.05522019326981233
(7002, 33195) 0.04301106169596901
(7002, 22879) 0.05789511454843924
(7002, 28947) 0.04496987612275906
```

Gambar 3. *Feature extraction*

C. Evaluasi

Naive bayes.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v12i3.1808, Copyright ©2023
Submitted : June 19, 2023, Revised : July 20, 2023, Accepted : July 26, 2023, Published : November 7, 2023

Pada tahap klasifikasi menggunakan Multinomial Naïve Bayes, pengujian akan dilakukan sebanyak 5 (lima) kali dengan membagi data menjadi data pelatihan (train) dan data uji (test) yang berbeda secara acak. Hal ini bertujuan untuk melihat hasil terbaik dari pengujian dengan variasi kombinasi data train dan data test. Hasil pengujian dapat dilihat pada table 6.

TABEL 6. HASIL PENGUJIAN NAIVE BAYES

Bobot data		***Score***			
Latih	**Uji**	**Akurasi**	**Presisi**	**Recall**	**F1-Score**
90%	10%	0.89612	0.89612	0.89612	0.88800
80%	20%	0.88007	0.88963	0.88007	0.87062
70%	30%	0.87514	0.88572	0.87514	0.86347
60%	40%	0.86465	0.87673	0.86465	0.85094
50%	50%	0.85995	0.85088	0.85995	0.84505

Berdasarkan hasil klasifikasi *Naive Bayes* diatas, nilai akurasi terbaik adalah 89.61% dengan pembagian data latih 90% dan data uji 10% dari jumlah data

Support Vector Machines (SVM)

Pada tahap klasifikasi dengan *Support Vector Machine* dilakukan juga pengujian sebanyak 5 (lima) kali seperti pengeujian pada naive bayes sebelumnya. Hasil pengujian dapat dilihat pada table 7.

TABEL 7. HASIL PENGUJIAN SUPPORT VECTOR MACHINES (SVM)

Bobot data		***Score***			
Latih	**Uji**	**Akurasi**	**Presisi**	**Recall**	**F1-Score**
90%	10%	0.92922	0.93098	0.92922	0.92932
80%	20%	0.92404	0.92606	0.92404	0.92416
70%	30%	0.91740	0.92005	0.91740	0.91719
60%	40%	0.91034	0.91375	0.91034	0.90962
50%	50%	0.90199	0.90709	0.90199	0.90108

Berdasarkan hasil pada table 7, nilai akurasi terbaik pada *Support Vector Machines* adalah 92.92% dengan pembagian data latih 90% dan data uji 10% dari jumlah data.

Random Forest

Pada tahap klasifikasi dengan *Random Forest*, dilakukan juga pengujian sebanyak 5 (lima) kali seperti pengeujian pada naive bayes sebelumnya. Hasil pengujian dapat dilihat pada table 8.

TABEL 8. HASIL PENGUJIAN RANDOM FOREST

Bobot data		***Score***			
Latih	**Uji**	**Akurasi**	**Presisi**	**Recall**	**F1-Score**
90%	10%	0.91210	0.91582	0.91210	0.91027
80%	20%	0.90520	0.90876	0.90520	0.90386
70%	30%	0.89912	0.90397	0.89912	0.91719
60%	40%	0.89063	0.89630	0.89063	0.88762
50%	50%	0.88805	0.89397	0.88805	0.88468

Berdasrakan hasil pengujian pada table 8, nilai akurasi terbaik pada *Random Forest* adalah 91.21% dengan pembagian data latih 90% dan data uji 10% dari jumlah data.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan terhadap tiga model kelasifikasi, yaitu *Naive Bayes*, *SVM*, dan *Random Forest*, Hasil penelitian menunjukkan bahwa algoritma *SVM* (*Support Vector Machine*) memiliki kinerja terbaik dalam evaluasi label berita, dengan tingkat akurasi mencapai 92.92%. Disusul oleh algoritma Random Forest dengan tingkat akurasi 91.21%, dan algoritma Naive Bayes dengan tingkat akurasi 89.61%.

Dengan demikian, penggunaan algoritma *SVM* dalam mengelola label berita berdasarkan judul utama menunjukkan hasil yang paling mengesankan, memberikan tingkat akurasi tertinggi dalam klasifikasi label berita. Hasil ini menunjukkan bahwa *SVM* adalah pilihan yang baik untuk memastikan label berita yang akurat dan dapat diandalkan bagi konsumen. Penelitian ini memberikan sumbangan penting bagi pengelolaan informasi berita dalam era digital yang penuh dengan informasi tekstual. Dengan metode evaluasi yang baik, diharapkan informasi berita yang disajikan dapat menjadi lebih relevan dan bermanfaat bagi pembaca, membantu mereka menemukan konten yang sesuai dengan minat dan kebutuhan mereka. Selain itu, penelitian ini juga dapat menjadi dasar bagi penelitian selanjutnya dalam mengembangkan dan meningkatkan kinerja algoritma Machine Learning lainnya dalam pengelolaan label berita

Penelitian ini memiliki kelebihan dalam melibatkan tiga model klasifikasi yang berbeda, memberikan perbandingan yang komprehensif dalam hal performa. Metode pemrosesan teks yang baik, termasuk tokenisasi, pengahpusan *stop words*, dan vektorisasi *TF-IDF*, juga digunakan dalam penelitian ini.

Namun penelitian ini memiliki beberapa kekurangan. Keterbatasan pada dataset yang digunakan, seperti jumlah sample, distribusi label yang tidak seimbang dan representasi kelas, sehingga tidak dapat memberikan pemahaman yang lebih baik tentang performa model. Penelitian ini hanya menggunakan algoritma klasifikasi yang umum digunakan dan tidak mempertimbangkan model kasifikasi lainnya yang mungkin memberikan performa yang lebih baik.

Dalam penelitian selanjutnya, disarankan untuk memperluas analisis dengan melibatkan lebih banyak model klasifikasi dan mempertimbangkan variasi parameter untuk meningkatkan performa. Selain itu, dataset yang lebih beragam dan representatif juga dapat membantu untuk mengevaluasi dan membandingkan model secara lebih akurat.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v12i3.1808, Copyright ©2023
Submitted : June 19, 2023, Revised : July 20, 2023, Accepted : July 26, 2023, Published : November 7, 2023

REFERENCES

[1] D. Rani dan S. D. Setiawati, "Penyajian Jurnalistik Online Infobdg untuk Menjadi Sumber Informasi Kredibel," J. Jurnalisa, vol. 6, no. 2, pp. 233–247, 2020.

[2] F. S. Nurfikri dan M. S. Mubarok, "Klasifikasi Topik Berita Menggunakan," vol. 5, no. 1, pp. 1579–1588, 2018.

[3] Briggs, Mark. Journalism next: A practical guide to digital reporting dan publishing. CQ Press, 2013.

[4] H. Junaedi, H. Budianto, I. Maryati, dan Y. Melani, "Data Transformation pada Data Mining," Pros. Konf. Nas. Inov. dalam Desain dan Teknol., vol. 7, pp. 93–99, 2011.

[5] T. Jamaluddin dan dkk, "Perbdaningan Algoritma Sentencepiece BPE dan Unigram Pada Tokenisasi Artikel Bahasa Indonesia Pendahuluan Studi Terkait," e-Proceeding Eng., vol. 7, no. 2, pp. 8323–8331, 2020.

[6] M. J. Lavin, Z. Leblanc, dan Q. Dombrowski, "The Programming Historian Analyzing Documents with TF-IDF," Program. Hist., pp. 1–21, 2020.

[7] D. Wahyuningsih dan E. Patima, "Penerapan Naive Bayes Untuk Penerimaan Beasiswa," Telematika, vol. 11, no. 1, p. 135, 2018, doi: 10.35671/telematika.v11i1.665.

[8] Y. Zhai, W. Song, X. Liu, L. Liu, dan X. Zhao, "A Chi-square Statistics Based Feature Selection," 2018 IEEE 9th Int. Conf. Softw. Eng. Serv. Sci., pp. 160–163, 2018.

[9] Schutze, Hinrich, Christopher D. Manning, dan Prabhakar Raghavan. Introduction to information retrieval. Cambridge University Press, 2008.

[10] R. Wati, "Penerapan Algoritma Naive Bayes Dan Particle Swarm Optimization Untuk Klasifikasi Berita Hoax Pada Media Sosial," JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer), vol. 5, no. 2, pp. 159–164, 2020, doi: 10.33480/jitk.v5i2.1034.

[11] Han, Jiawei, Jian Pei, dan Hanghang Tong. Data mining: concepts and techniques. Morgan kaufmann, 2022.

[12] S. Sudianto, A. D. Sripamuji, I. Ramadhanti, R. R. Amalia, J. Saputra, dan B. Prihatnowo, "Penerapan Algoritma Support Vector Machine dan Multi-Layer Perceptron pada Klasisifikasi Topik Berita," J. Nas. Pendidik. Tek. Inform. JANAPATI, vol. 11, no. 2, pp. 84–91, 2022.

[13] A. Candra, "Jakarta Artificial Intelligence Research is Now Open!" [Online]. Tersedia: https://medium.com/data-folks-indonesia/jakarta-artificial-intelligence-research-is-now-open-f404763867b1. Diakses pada: July 15, 2023.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v12i3.1808, Copyright ©2023
Submitted : June 19, 2023, Revised : July 20, 2023, Accepted : July 26, 2023, Published : November 7, 2023