



Klasifikasi Materi Judul Berita Hasil Siaran TVRI dengan Metode Support Vector Machine

Taufik Masri, Rini Sovia, Agung Ramadhanu*

Fakultas Ilmu Komputer, Magister Teknik Informatika, Universitas Putra Indonesia YPTK Padang, Padang, Indonesia

Email: ¹Topekmasri@gmail.com, ²rinisovia2020@gmail.com, ³*agung_ramadhanu@upiypk.ac.id

Email Penulis Korespondensi: agung_ramadhanu@upiypk.ac.id

Abstrak—Perkembangan teknologi digital telah mendorong peningkatan volume data tekstual, khususnya pada sektor penyiaran berita yang menuntut pengelolaan informasi secara cepat, akurat, dan terstruktur. TVRI Sumatera Barat sebagai lembaga penyiaran publik menghadapi permasalahan dalam pengelompokan judul berita hasil siaran yang masih dilakukan secara manual sehingga berpotensi menimbulkan ketidakefisienan dalam pengelolaan dan pencarian informasi. Penelitian ini bertujuan mengimplementasikan algoritma Support Vector Machine (SVM) untuk mengklasifikasikan judul berita hasil siaran TVRI Sumatera Barat ke dalam kategori politik, ekonomi, dan sosial. Tahapan penelitian meliputi text preprocessing yang terdiri atas *case folding*, *tokenizing*, normalisasi, *stopword removal*, dan *stemming*, dilanjutkan dengan pembobotan kata menggunakan Term Frequency–Inverse Document Frequency (TF-IDF). Dataset yang digunakan berjumlah 900 judul berita yang diperoleh dari dokumen *rundown* program "Sumatera Barat Hari Ini" selama periode Januari–Desember 2025. Dataset dibagi menjadi 80% data latih dan 20% data uji, kemudian diklasifikasikan menggunakan algoritma Support Vector Machine dengan kernel linear. Hasil pengujian menunjukkan bahwa model mampu mencapai akurasi sebesar 88%, dengan precision weighted average sebesar 89%, recall weighted average sebesar 88%, dan F1-score sebesar 88%. Hasil tersebut menunjukkan bahwa metode SVM mampu mengklasifikasikan judul berita secara efektif sehingga dapat meningkatkan efisiensi, konsistensi pengelolaan arsip berita, serta mendukung proses temu kembali informasi pada lembaga penyiaran publik.

Kata Kunci: Klasifikasi; Judul Berita; TF-IDF; Support Vector Machine; TVRI

Abstract—The development of digital technology has driven an increase in the volume of textual data, especially in the news broadcasting sector which demands fast, accurate, and structured information management. TVRI West Sumatra as a public broadcasting institution faces problems in grouping broadcast news titles which are still done manually, potentially causing inefficiencies in information management and retrieval. This study aims to implement the Support Vector Machine (SVM) algorithm to classify TVRI West Sumatra broadcast news titles into political, economic, and social categories. The research stages include text preprocessing consisting of case folding, tokenizing, normalization, stopwords removal, and stemming, followed by word weighting using Term Frequency–Inverse Document Frequency (TF-IDF). The dataset used is 900 news titles obtained from the rundown document of the "West Sumatra Today" program during the period January–December 2025. The dataset is divided into 80% training data and 20% test data, then classified using the Support Vector Machine algorithm with a linear kernel. Test results showed that the model achieved an accuracy of 88%, with a weighted average precision of 89%, a weighted average recall of 88%, and an F1-score of 88%. These results demonstrate that the SVM method is able to classify news titles effectively, thereby improving efficiency and consistency in news archive management, and supporting information retrieval processes in public broadcasting institutions.

Keywords: Classification; News Title; TF-IDF; Support Vector Machine; TVRI

1. PENDAHULUAN

Perkembangan teknologi informasi telah meningkatkan jumlah data tekstual yang dihasilkan oleh berbagai media penyiaran, termasuk lembaga penyiaran publik seperti TVRI Sumatera Barat. Setiap hari, TVRI menghasilkan berbagai judul berita dari program siaran yang harus dikelola sebagai arsip digital untuk mendukung pencarian kembali informasi, penyusunan dokumentasi, serta distribusi konten. Namun, proses pengelompokan judul berita di TVRI Sumatera Barat masih dilakukan secara manual berdasarkan kategori tertentu, seperti politik, ekonomi, dan sosial. Proses tersebut membutuhkan waktu yang relatif lama, bergantung pada subjektivitas petugas, serta berpotensi menimbulkan inkonsistensi dalam pengelompokan berita sehingga menurunkan efisiensi pengelolaan arsip.

Permasalahan klasifikasi dokumen teks telah banyak diteliti menggunakan berbagai pendekatan machine learning. Penelitian sebelumnya menunjukkan bahwa algoritma klasifikasi mampu meningkatkan efisiensi pengelolaan dokumen dibandingkan metode manual. Guruh Wijaya *et al.* membandingkan beberapa algoritma klasifikasi berita dan menunjukkan bahwa pendekatan machine learning menghasilkan akurasi yang lebih stabil dibandingkan metode konvensional [3]. Dewi *et al.* juga berhasil menerapkan teknik klasifikasi pada judul berita pariwisata Indonesia menggunakan metode ekstraksi fitur sehingga mampu meningkatkan ketepatan pengelompokan berita [4]. Selain itu, Nanda *et al.* menunjukkan bahwa sistem klasifikasi berita dapat membantu meningkatkan efektivitas pengelolaan konten digital [5].

Meskipun demikian, sebagian besar penelitian terdahulu menggunakan dokumen berita dengan isi yang relatif panjang. Berbeda dengan penelitian tersebut, objek penelitian ini berupa judul berita, yang termasuk kategori short text karena hanya terdiri atas beberapa kata. Karakteristik tersebut menyebabkan representasi fitur menjadi sangat jarang (*sparse*) sehingga informasi yang tersedia lebih sedikit dibandingkan dokumen lengkap. Kondisi ini menjadi tantangan tersendiri dalam proses klasifikasi karena algoritma harus mampu membedakan kategori hanya berdasarkan sejumlah kecil kata yang muncul pada judul berita.

Salah satu metode yang banyak digunakan untuk menangani klasifikasi teks adalah kombinasi Term Frequency–Inverse Document Frequency (TF-IDF) dan Support Vector Machine (SVM). TF-IDF berfungsi mengubah data teks menjadi representasi numerik berdasarkan tingkat kepentingan setiap kata [11]–[14], sedangkan SVM digunakan untuk

mencari *hyperplane* optimal yang mampu memisahkan data ke dalam kelas yang berbeda [15], [16]. Dibandingkan algoritma berbasis probabilitas seperti Naïve Bayes, SVM memiliki kemampuan yang lebih baik dalam menangani data berdimensi tinggi dan representasi fitur yang bersifat *sparse*. Oleh karena itu, SVM dengan kernel linear dinilai lebih sesuai untuk klasifikasi judul berita yang memiliki jumlah kata terbatas, karena mampu menghasilkan batas pemisah kelas yang lebih stabil tanpa bergantung pada distribusi probabilitas data [17].

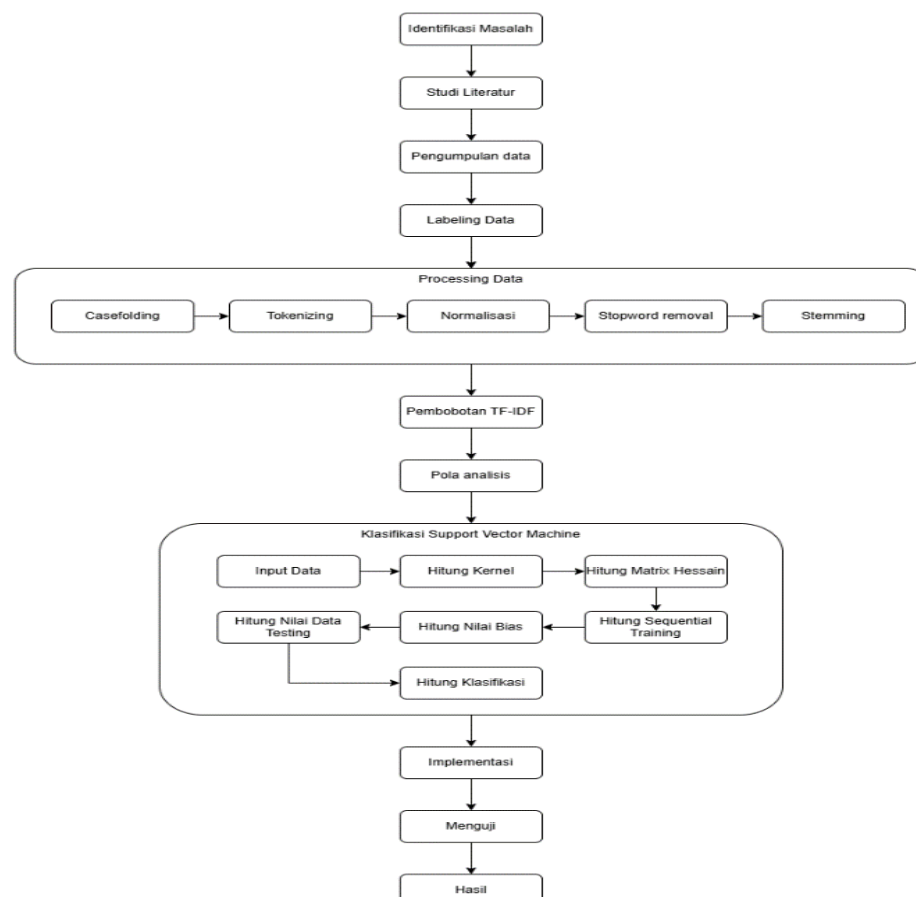
Berdasarkan kondisi tersebut, masih terdapat research gap, yaitu belum banyak penelitian yang secara khusus menerapkan kombinasi TF-IDF dan Support Vector Machine untuk mengklasifikasikan judul berita hasil siaran TVRI Sumatera Barat yang memiliki karakteristik short text. Selain itu, penelitian sebelumnya lebih banyak berfokus pada berita umum atau dokumen lengkap, sehingga belum secara spesifik membahas pengelolaan arsip berita pada lembaga penyiaran publik.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengimplementasikan dan menganalisis kinerja kombinasi metode TF-IDF dan Support Vector Machine dalam mengklasifikasikan judul berita hasil siaran TVRI Sumatera Barat ke dalam kategori politik, ekonomi, dan sosial. Kontribusi penelitian ini adalah menghasilkan model klasifikasi otomatis yang dapat meningkatkan efisiensi pengelolaan arsip berita, mempercepat proses temu kembali informasi, serta memberikan alternatif penerapan machine learning pada pengelolaan data teks di lingkungan lembaga penyiaran publik.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan klasifikasi teks berbasis Support Vector Machine (SVM) untuk mengelompokkan judul berita hasil siaran TVRI Sumatera Barat ke dalam kategori politik, ekonomi, dan sosial. Tahapan penelitian diawali dengan pengumpulan dataset, dilanjutkan dengan proses *text preprocessing*, pembobotan kata menggunakan Term Frequency–Inverse Document Frequency (TF-IDF), pelatihan model menggunakan Support Vector Machine (SVM), serta evaluasi performa model menggunakan Confusion Matrix. Dataset kemudian dibagi menjadi 80% data latih (training data) dan 20% data uji (testing data). Data latih digunakan untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengevaluasi kemampuan model dalam mengklasifikasikan data baru. Alur penelitian secara lengkap ditunjukkan pada Gambar 1.



Gambar 1. Kerangka Kerja Penelitian



2.2 Dataset Penelitian

Dataset penelitian berasal dari dokumen rundown berita Program "Sumatera Barat Hari Ini" yang telah disiarkan oleh TVRI Sumatera Barat selama periode 1 Januari sampai 31 Desember 2025. Total dataset yang digunakan sebanyak 900 judul berita yang telah dikategorikan ke dalam tiga kelas, yaitu Politik, Ekonomi, dan Sosial.

Distribusi dataset ditunjukkan pada Tabel 1

Tabel 1. Distribusi dataset

Kategori	Jumlah
Politik	300
Ekonomi	300
Sosial	300
Total	900

Sebagai ilustrasi data penelitian, beberapa contoh judul berita ditunjukkan pada Tabel 2.

Tabel 2. Contoh Sampel Dataset Judul Berita

No	Judul	Kategori
1.	KPU PASBAR PERSIAPKAN SIDANG GUGATAN PHPU DI MK	Politik
2.	HARGA CABAI MERAH KERITING TURUN DRASTIS	Ekonomi
3.	WARGA KELUHKAN KONDISI JALAN RUSAK	Sosial

2.3 Preprocessing

Tahapan text Preprocessing bertujuan membersihkan dan menyeragamkan data sehingga siap diproses oleh algoritma klasifikasi. Preprocessing merupakan tahapan dimana data dibersihkan dan dirapikan agar nantinya dapat dikonversi ke bentuk angka atau numerik.

- Casefolding* yaitu, mengubah semua huruf menjadi huruf kecil atau *lowercase* sehingga tidak terdapat perbedaan antara huruf kapital dan huruf kecil.
- Tokenizing* yaitu pemenggalan kata per kata dari sebuah kalimat. Kalimat dipecah menjadi token-token kata tunggal.
- Normalisasi berfungsi untuk menyeragamkan bentuk teks agar data lebih bersih, konsisten, dan mudah diolah oleh algoritma.
- Stopwords* adalah proses menghilangkan kata-kata yang tidak berguna yang dimana dapat menghambat proses analisis, seperti "dan", "oleh", "pada" dan sebagainya.
- Stemming* data akan diubah ke dalam bentuk kata dasarnya, dilakukan dengan cara menghilangkan awalan, akhiran, sisipan, dan confixes (kombinasi awalan dan akhiran).

2.4 TF-IDF

TF-IDF merupakan metode pembobotan kata yang digunakan untuk menilai tingkat kepentingan suatu kata dalam sebuah dokumen terhadap keseluruhan kumpulan dokumen. Metode TF-IDF digunakan untuk pembobotan dan pemilihan fitur kata[19]. Metode ini terdiri dari dua komponen utama, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). TF mengukur seberapa sering suatu kata muncul dalam sebuah dokumen, sedangkan IDF mengukur seberapa jarang kata tersebut muncul di seluruh dokumen.

Term Frequency

$$TF(t, d) = \frac{f_{t,d}}{\sum_k f_{k,d}} \quad (1)$$

Inverse Document Frequency

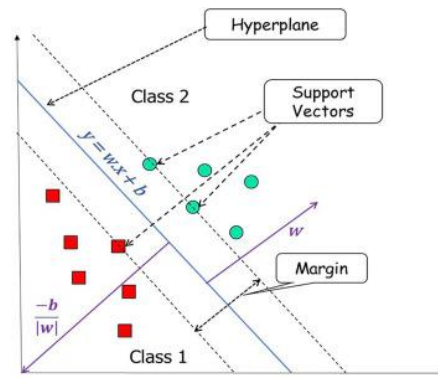
$$IDF(t) = \log\left(\frac{N}{df_t}\right) \quad (2)$$

Kombinasi Tf-IDF

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \quad (3)$$

2.5 Support Vector Machine

Support Vector Machine (SVM) merupakan algoritma *supervised learning* yang digunakan untuk menyelesaikan permasalahan klasifikasi dengan cara mencari hyperplane optimal yang memisahkan data ke dalam kelas-kelas tertentu. Support Vector Machine (SVM) merupakan metode machine learning yang mengelompokkan data dengan kelasnya dengan menggunakan hyperplane[20]. SVM bekerja dengan memaksimalkan margin, yaitu jarak antara hyperplane dan titik data terdekat dari masing-masing kelas (*support vector*). Proses kerja SVM dapat dilihat di Gambar 2.



Gambar 2. Konsep Support Vector machine

Pada penelitian ini digunakan Kernel Linear, karena data teks hasil pembobotan TF-IDF memiliki dimensi tinggi dan cenderung bersifat linearly separable, sehingga kernel linear mampu menghasilkan proses klasifikasi yang lebih sederhana, stabil, dan efisien dibandingkan kernel non-linear. Proses pelatihan dilakukan menggunakan data latih sebanyak 80%, sedangkan evaluasi model dilakukan menggunakan data uji sebanyak 20%. Persamaan fungsi keputusan SVM ditunjukkan persamaan (4) Dimana x adalah titik data masukan Support Vector Machine, α_i adalah nilai bobot setiap titik data, $K(x, x_i)$ adalah fungsi kernel dan b adalah parameter hyperplane yang dicari (nilai bias).

$$f(x) = \text{sign} \sum_{i=0}^n (\alpha_i y_i K(x, y_i) + b) \quad (4)$$

3. HASIL DAN PEMBAHASAN

Pada tahapan ini dijelaskan bagaimana tahapan pengumpulan data sampai data tersebut berhasil diklasifikasikan. Diawali dengan pengumpulan data berupa teks di TVRI Sumatera Barat. Data didapat dari rundown mata acara berita dengan jumlah total 900 data. Berikutnya tahapan yang dilakukan adalah preprocessing data, yang mencakup case folding, tokenisasi, stopwords dan stemming. Proses ini bertujuan untuk menghasilkan teks yang bersih sehingga dapat diolah secara komputasional. Kemudian dilakukan pembobotan Tf-IDF untuk mengubah teks menjadi representasi numerik yang menggambarkan tingkat kepentingan setiap kata dalam dokumen. Bobot Tf-IDF selanjutnya diproses menggunakan metode Support Vector Machine (SVM) untuk mengklasifikasikan teks judul berita menjadi tiga kategori, yaitu politik, ekonomi dan sosial.

3.1 Preprocessing

Tahapan preprocessing bertujuan menghasilkan data teks yang bersih sehingga dapat diproses oleh algoritma klasifikasi. Tahapan yang dilakukan meliputi case folding, tokenizing, normalisasi, stopword removal, dan stemming. Contoh hasil preprocessing ditunjukkan pada Tabel 3.

Tabel 3. Tahapan Preprocessing

Tahapan	Sebelum Proses	Setelah Proses
Case Folding	REALISASI JANJI POLITIK, PENUHI KEBUTUHAN MASYARAKAT	realisasi janji politik, penuhi kebutuhan masyarakat
Tokenizing	realisasi janji politik, penuhi kebutuhan masyarakat	'realisasi', 'janji', 'politik', 'penuhi', 'kebutuhan', 'masyarakat'
Stopword	'realisasi', 'janji', 'politik', 'penuhi', 'kebutuhan', 'masyarakat'	'realisasi', 'janji', 'politik', 'penuhi', 'kebutuhan', 'masyarakat'
Stemming	'realisasi', 'janji', 'politik', 'penuhi', 'kebutuhan', 'masyarakat'	'realisasi', 'janji', 'politik', 'penuhi', 'butuh, masyarakat'

Berdasarkan Tabel 3 terlihat bahwa setiap tahapan preprocessing memberikan perubahan terhadap representasi teks. Proses case folding menyeragamkan seluruh huruf menjadi huruf kecil, tokenizing memisahkan kalimat menjadi token-token kata, sedangkan stopword removal menghilangkan kata-kata umum yang tidak memiliki kontribusi terhadap proses klasifikasi. Selanjutnya proses stemming menggunakan library Sastrawi mengubah kata berimbuhan menjadi bentuk dasar sehingga jumlah variasi kata menjadi lebih sedikit. Tahapan tersebut menghasilkan representasi teks yang lebih konsisten sehingga mampu meningkatkan kualitas pembobotan TF-IDF.

3.2 TF-IDF

Setelah data bersih, dilakukan pembobotan menggunakan metode TF-IDF (*Term Frequency – Inverse Document Frequency*). Metode ini memberikan bobot numerik pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen (TF) dan tingkat keunikannya di seluruh korpus data (IDF). Kata yang sering muncul di satu kategori tetapi



jarang di kategori lain akan memiliki bobot tinggi, yang menjadi fitur pembeda utama antar kelas (Politik, Ekonomi, Sosial)

Tabel 4. Tahapan Preprocessing

Term (Kata)	Bobot TF-IDF	Bobot TF-IDF
realisasi	0,9542	0
janji	0,9542	0
politik	0,6532	0
tani	0	0,6532
bawang	0	0,9542
merah	0	0,9642
harga	0	0,9642

Pada Tabel 4 terlihat bahwa kata "realisasi", "janji", dan "politik" hanya muncul pada Dokumen A, sehingga memiliki bobot TF-IDF yang tinggi pada dokumen tersebut dan bernilai 0 pada Dokumen B. Sebaliknya kata "tani", "bawang", "merah", dan "harga" hanya muncul pada Dokumen B, sehingga memiliki bobot TF-IDF pada Dokumen B dan bernilai 0 pada Dokumen A. Nilai 0 menunjukkan bahwa istilah tersebut tidak muncul pada dokumen yang bersangkutan, sehingga tidak memberikan kontribusi terhadap representasi fitur dokumen tersebut.

3.3 Proses Metode SVM

Tahapan selanjutnya adalah proses klasifikasi menggunakan algoritma Support Vector Machine dengan Kernel Linear. Metode utama yang digunakan adalah *Support Vector Machine* (SVM) dengan *Kernel Linier*. Klasifikasi dilakukan menggunakan algoritma Support Vector Machine (SVM) dengan kernel linier. Proses ini bertujuan mencari *hyperplane* optimal yang dapat memisahkan data antar kelas. Analisis dilakukan dengan menghitung nilai Kernel, membentuk Matriks Hessian, melakukan *Sequential Training* untuk mencari nilai Alpha (α), dan menghitung nilai Bias (β) guna membentuk fungsi keputusan. Hasil perhitungan parameter SVM

Nilai kernel

$$K(D2, D2) \ 4,9792$$

$$K(D4, D4) \ 4,0687$$

$$K(D2, D4) = 0 \text{ (Karena tidak ada kata yang sama antar kedua dokumen)}$$

Pembentukan Matriks Hessian (H) Matriks Hessian dibentuk dari perkalian label kelas (y) dan nilai kernel:

$$H_{22} = (+1)(+1)(4,9792) = 4,9792$$

$$H_{44} = (-1)(-1)(4,0687) = 4,0687$$

Nilai Alpha

$$\alpha_2 = \alpha_4 = 0.2210$$

Nilai Bias (b)

$$b = 1 - (0.2210 \times 4.9792) - 0.1006$$

Dengan parameter tersebut, fungsi keputusan $f(x)$ terbentuk.

Hasil uji $f(x)$ untuk D2 (Politik) menghasilkan nilai +1.0 (positif)

Hasil uji $f(x)$ untuk D4 (Ekonomi) menghasilkan nilai -1.0 (Negatif)

Kernel linear dipilih karena data hasil pembobotan TF-IDF memiliki dimensi tinggi dan bersifat **sparse**, sehingga lebih sesuai dipisahkan menggunakan *hyperplane* linear. Nilai kernel yang diperoleh digunakan untuk membentuk Matriks Hessian, kemudian dilakukan proses *Sequential Training* untuk memperoleh nilai alpha dan bias sebagai pembentuk fungsi keputusan.

3.4 Proses Evaluasi

Tahap evaluasi dilakukan untuk mengukur performa sistem dalam mengklasifikasikan seluruh 900 data judul berita. Evaluasi menggunakan metode *Confusion Matrix* untuk menghitung nilai *Accuracy*, *Precision*, dan *Recall*. Pengujian dilakukan dengan skenario pembagian data latih dan data uji 80:20 untuk melihat konsistensi model terhadap data baru.

Tabel 5. Evaluasi

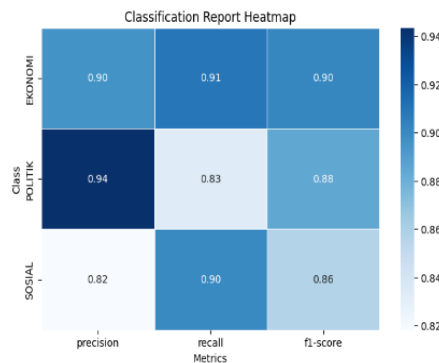
	Precision	Recall	F1-score	support
EKONOMI	0.90	0.91	0.90	66
POLITIK	0.94	0.83	0.88	60
SOSIAL	0.82	0.90	0.86	60

	Precision	Recall	F1-score	support
Accuray			0.88	186
Macro avg	0.89	0.88	0.88	186
Weghted avg	0.89	0.88	0.88	186

Berdasarkan Tabel 5 diperoleh nilai akurasi sebesar 88%, menunjukkan bahwa model mampu mengklasifikasikan sebagian besar judul berita dengan benar. Nilai precision weighted average sebesar 89% menunjukkan bahwa sebagian besar prediksi yang dihasilkan model sesuai dengan kelas sebenarnya. Nilai recall weighted average sebesar 88% menunjukkan kemampuan model dalam mengenali data pada setiap kategori, sedangkan F1-score sebesar 88% menunjukkan keseimbangan antara precision dan recall.

Secara umum kategori Politik memiliki precision tertinggi yaitu 94%, sedangkan kategori Ekonomi memperoleh recall tertinggi sebesar 91%. Perbedaan nilai pada masing-masing kategori dipengaruhi oleh karakteristik kata yang digunakan pada setiap judul berita. Walaupun demikian, seluruh kategori menunjukkan nilai F1-score di atas 0,85, sehingga model dapat dikatakan memiliki performa klasifikasi yang baik.

Hasil penelitian ini sejalan dengan penelitian Nanda et al. [1] dan Salsabillah et al. [17] yang menunjukkan bahwa Support Vector Machine memiliki performa yang baik pada permasalahan klasifikasi teks. Pada penelitian ini diperoleh tingkat akurasi 88%, yang menunjukkan bahwa kombinasi TF-IDF dan SVM efektif digunakan untuk mengklasifikasikan judul berita hasil siaran TVRI Sumatera Barat yang memiliki karakteristik short text.



Gambar 3. Confusion Matrix

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa kombinasi metode Term Frequency–Inverse Document Frequency (TF-IDF) dan Support Vector Machine (SVM) dengan kernel linear berhasil diimplementasikan untuk mengklasifikasikan judul berita hasil siaran TVRI Sumatera Barat ke dalam kategori Politik, Ekonomi, dan Sosial. Pengujian menggunakan skenario pembagian data latih dan data uji sebesar 80:20 menunjukkan bahwa model mampu mencapai tingkat akurasi sebesar 88,17%, dengan nilai *weighted average precision* sebesar 89%, *weighted average recall* sebesar 88%, dan *weighted average F1-score* sebesar 88%. Hasil tersebut menunjukkan bahwa metode yang diusulkan mampu mengklasifikasikan judul berita secara efektif meskipun data yang digunakan berupa *short text*, sehingga dapat meningkatkan efisiensi, konsistensi, dan ketepatan dalam pengelolaan arsip berita di TVRI Sumatera Barat. Dengan demikian, penelitian ini memberikan kontribusi dalam penerapan metode klasifikasi teks berbasis machine learning untuk mendukung pengelolaan informasi pada lembaga penyiaran publik.

REFERENCES

- [1] R. Nanda, E. Haerani, S. K. Gusti, and S. Ramadhani, "Klasifikasi Berita Menggunakan Metode Support Vector Machine," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 2, pp. 269–278, 2022, doi: 10.32672/jnkti.v5i2.4193.
- [2] N. Silalahi and Guidio Leonarde Ginting, "Rekomendasi Berita Berkaitan dengan Menerapkan Algoritma Text Mining dan TF-IDF," *Bull. Comput. Sci. Res.*, vol. 3, no. 4, pp. 276–282, 2023, doi: 10.47065/bulletincsr.v3i4.266.
- [3] Guruh Wijaya, Dudi Irawan, Zainul Arifin, Hardian Oktavianto, Miftahur Rahman, and Ginanjar Abdurrahman, "Studi Klasifikasi Topik Berita Dengan Algoritma Machine Learning," *J-Ensirec*, vol. 11, no. 01, pp. 10202–10206, 2024, doi: 10.31949/jensirec.v11i01.12037.
- [4] N. L. P. R. Dewi, I. N. S. W. Wijaya, I. K. Purnamawan, and N. W. Marti, "Model Classifier Judul Berita Pariwisata Indonesia Berdasarkan Sentimen," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 1, pp. 117–124, 2024, doi: 10.25126/jtiik.20241117617.
- [5] A. Tejawati, A. Septiarini, R. Rismawati, and N. Puspitasari, "Comparison of K-Nearest Neighbor and Naive Bayes Methods for Classification of News Content," *J. Tek. Inform.*, vol. 4, no. 2, pp. 401–412, 2023, doi: 10.52436/1.jutif.2023.4.2.676.
- [6] R. S. rahmaddeni, Debok Wulandari, Masni Renova, ilang Ramadhan, *Machine learning*. Serasi Media Teknologi, 2024.
- [7] M. Ardiansyah, *Kecerdasan Buatan: Teori Algoritma dan Aplikasi*. Agam: Yayasan Tri Edukasi Ilmiah, 2024.
- [8] O. M. Alyasiri and Y. N. Cheah, "Multi-Class Text Classification using Machine Learning Techniques," *Eng. Technol. Appl. Sci. Res.*, vol. 15, no. 3, pp. 22598–22604, 2025, doi: 10.48084/etasr.9994.



- [9] A. R. Hanum *et al.*, “Analisis Kinerja Algoritma Klasifikasi Teks Bert dalam Mendeteksi Berita Hoaks,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 3, pp. 537–546, 2024, doi: 10.25126/jtiik.938093.
- [10] A. D. P. Randi Rian Putra, Nadya Anthika Putri., “Teknik Cosine Similarity Dan TF - IDF Dalam Analisis Data,” *Medan: Serasi Media Teknologi*, 2024.
- [11] L. Efrizoni, S. Defit, M. Tajuddin, and A. Anggrawan, “Komparasi Ekstraksi Fitur dalam Klasifikasi Teks Multilabel Menggunakan Algoritma Machine Learning,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 653–666, 2022, doi: 10.30812/matrik.v21i3.1851.
- [12] N. Fathirachman Mahing, A. Lazuardi Gunawan, A. Foresta Azhar Zen, F. Abdurrahman Bachtiar, and S. Agung Wicaksono, “Klasifikasi Tingkat Stress dari Data Berbentuk Teks dengan Menggunakan Algoritma Support Vector Machine (SVM) dan Random Forest,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 5, pp. 1067–1076, 2024, doi: 10.25126/jtiik.1078010.
- [13] D. F. Salsabillah, D. E. Ratnawati, and N. Y. Setiawan, “Analisis Sentimen Ulasan Rumah Makan Menggunakan Perbandingan Algoritma *Support Vector Machine* dengan *Naive bayes* (Studi Kasus: Ayam Goreng Nelongso Cabang Singosari, Malang),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 1, pp. 107–116, 2024, doi: 10.25126/jtiik.20241117584.
- [14] T. Dzulkarnain, D. E. Ratnawati, and B. Rahayudi, “Penggunaan Metode Naïve Bayes Classifier pada Analisis Sentimen Penilaian Masyarakat Terhadap Pelayanan Rumah Sakit di Malang,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 5, pp. 993–1000, 2024, doi: 10.25126/jtiik.2024117979.
- [15] R. R. Dede Ardian Tarigan, Zakarias Situmorang, “Analisis Sentimen Aplikasi Playstore Sirekap 2024 Pasca Pilpres Dengan Perbandingan Metode Support Vector Machine (SVM),” *Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 3, pp. 661–670, 2025, doi: 10.25126/jtiik.2025129608
- [16] C. Sri, K. Aditya, F. Dwi, and S. Sumadi, “Combination of term weighting with class distribution and centroid- based approach for document classification,” vol. 4, no. 4, pp. 781-788, 2023, doi: 10.22219/kinetik.v8i4'.1793
- [17] H. Mustakim and S. Priyanta, “Aspect-Based Sentiment Analysis of KAI Access Reviews Using NBC and SVM,” vol. 16, no. 2, pp. 113–124, 2022, doi: 10.22146/ijccs.68903.
- [18] C. Wang, F. You, and Y. Wang, “Text intelligent classification model based on improved KNN algorithm in library service transformation,” *Syst. Soft Comput.*, vol. 7, no. June, p. 200313, 2025, doi: 10.1016/j.sasc.2025.200313.
- [19] C. V. Angkoso *et al.*, “Optimasi Algoritma Support Vector Machine Berbasis Kernel Radial Basis Function (RbBF) Menggunakan Metode Particle Optimizing Support Vector Machine Algorithm Using Particle Swarm Optimization Method For Sentiment Analysis Of Financial,” *Jurnal Teknologi Informasi dan Komputer*, vol. 12, no. 3, pp. 705–718, 2025, doi: 10.25126/jtiik.2025129317.
- [20] A. T. Akbar *et al.*, “Identifikasi Gagal Ginjal Kronis Dengan Mengimplementasikan Metode Support Vector Machine Beserta K-Nearest Neighbour Identification Of Chronic Kidney Disease By Implementing The Support Vector Machine Method And K-Nearest Neighbor (SVM-KNN),” vol. 10, no. 2, pp. 301–308, 2023, doi: 10.25126/jtiik.2023106059.