

Klasifikasi Subjektif Berita Menggunakan Algoritma Machine Learning

Laily Fajria Chasanah¹, Endang Wahyu Pamungkas²✉
¹⁻²Universitas Muhammadiyah Surakarta, Indonesia

✉Corresponding Author: ewp123@ums.ac.id

ABSTRAK

Teknologi informasi telah mengubah penyebaran berita, dengan akses yang lebih mudah melalui media elektronik. Dalam era digital ini, semakin banyak berita yang tersebar di internet, dengan sifat berita yang dapat dibedakan menjadi objektif dan subjektif. Berita subjektif sering kali mengandung pendapat atau keyakinan pribadi penulis, sementara berita objektif menyajikan fakta secara netral. Dengan tingginya volume berita, klasifikasi manual menjadi tidak efisien. Oleh karena itu, penerapan algoritma *machine learning* seperti *Naïve Bayes*, *Decision Tree*, *Support Vector Machine* (SVM), dan *Logistic Regression*. Proses penelitian meliputi pengumpulan data, representasi teks dengan TF-IDF, pembelajaran model dan evaluasi. Dengan dataset 3.000 berita, model diuji menggunakan akurasi, presisi, *recall*, dan F1-Score. Penelitian ini bertujuan untuk mengembangkan sistem yang dapat membantu masyarakat membedakan antara berita subjektif dan objektif melalui penerapan *machine learning*. Berdasarkan hasil pengujian, model yang diterapkan memberikan akurasi yang bervariasi, dengan model *Support Vector Machine* menunjukkan performa terbaik. Hasil penelitian ini diharapkan dapat berkontribusi dalam pengembangan teknologi serta membantu masyarakat dalam memilah berita yang mereka terima untuk mengurangi dampak penyebaran informasi yang salah atau bias.

Kata Kunci: berita, evaluasi model, *machine learning*, objektif, subjektif, TF-IDF.

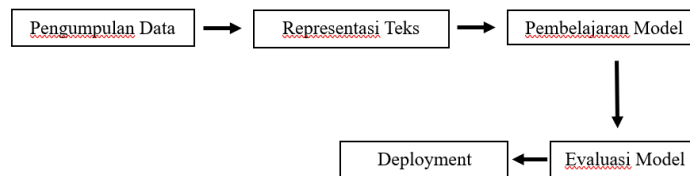
A. Latar belakang

Perkembangan informatika telah mendorong perubahan besar, termasuk dalam konsumsi berita yang kini mudah diakses melalui internet [1]. Jumlah berita yang terus meningkat menjadikan media elektronik sebagai sumber utama informasi masyarakat [2]. Berita dapat bersifat subjektif, yang mengandung opini pribadi, atau objektif, yang menyajikan fakta secara netral [3]. Namun, tingginya volume berita membuat klasifikasi manual tidak efisien. Teknologi *machine learning* menjadi solusi efektif untuk secara otomatis membedakan berita objektif dan subjektif, menghemat waktu, dan meningkatkan akurasi klasifikasi [4]. Kemudahan akses berita menghadirkan tantangan dalam membedakan berita subjektif dan objektif, di mana berita subjektif sering kali mengandung opini yang dapat memengaruhi persepsi publik. Proses klasifikasi manual menjadi tidak praktis, sehingga diperlukan pendekatan otomatis berbasis teknologi. Algoritma seperti *Naïve Bayes* yang sederhana dan efektif [5] serta *Decision Tree* yang logis dan mudah dipahami [6] dapat digunakan untuk mengatasi masalah ini. Penelitian ini bertujuan mengembangkan model *machine learning* untuk membantu masyarakat memilah berita secara cepat dan akurat, sehingga mengurangi penyebaran berita subjektif tanpa fakta yang jelas [7].

B. Metode

Penelitian ini bertujuan untuk mengklasifikasikan berita menjadi subjektif atau objektif menggunakan algoritma *machine learning*. Data penelitian berasal dari dataset Indonesia News dan Daily Insights dengan total 3.000 berita yang telah dilabeli secara manual. Teks berita

direpresentasikan dalam format numerik menggunakan metode TF-IDF dan Word Embedding untuk memahami pola linguistik dan hubungan semantik antar kata [8]. Algoritma yang digunakan mencakup Decision Tree, Naïve Bayes, Support Vector Machine (SVM), dan Logistic Regression karena keunggulan masing-masing dalam klasifikasi teks [9]. Evaluasi dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score [4]. Model yang telah dilatih kemudian diimplementasikan dalam aplikasi interaktif untuk memudahkan pengguna mengidentifikasi berita subjektif atau objektif secara real-time. Gambar 1 menunjukkan alur tahapan yang perlu dilakukan saat penelitian.



Gambar 1. Diagram alur tahap penelitian.

C. Hasil dan Pembahasan

Penelitian ini menggunakan 3.000 data dari Indonesia News dan Daily Insights yang diklasifikasikan sebagai subjektif dan objektif. Proses representasi teks dilakukan dengan metode TF-IDF dan data dibagi menggunakan train_test_split. Pengujian dilakukan dengan algoritma Decision Tree, Naïve Bayes, SVM, dan Logistic Regression. Hasil terbaik diperoleh dari SVM dengan TF-IDF dan Random Over Sampling (ROS) yang mencapai akurasi 82%, meskipun performanya lebih rendah pada kelas 'Subjective'. Model Naïve Bayes dan Logistic Regression juga menunjukkan kinerja baik pada label 'Objective'.

Hasil menunjukkan bahwa TF-IDF dengan ROS memberikan akurasi tertinggi, sementara Word Embedding dengan Undersampling (US) lebih baik dalam recall dan F1-Score. Teknik Undersampling meningkatkan keseimbangan model, terutama dalam mendeteksi kelas minoritas. Penerapan Streamlit memungkinkan pengguna untuk menganalisis dan mengklasifikasikan berita secara otomatis sebagai subjektif atau objektif menggunakan algoritma pembelajaran mesin.

Model	Representasi	Metode	Accuracy	Precision	Recall	F1-Score
Naïve	TF-IDF	ROS	0.69	0.62	0.66	0.62
Bayes	Word Embedding	ROS	0.66	0.63	0.69	0.61
	TF-IDF	US	0.59	0.62	0.67	0.56
	Word Embedding	US	0.69	0.65	0.71	0.64
	TF-IDF	Original	0.79	0.39	0.50	0.44
	Word Embedding	Original	0.79	0.39	0.50	0.44
Decision	TF-IDF	ROS	0.72	0.60	0.62	0.61
Tree	Word Embedding	ROS	0.79	0.39	0.50	0.44
	TF-IDF	US	0.70	0.55	0.55	0.55
	Word Embedding	US	0.66	0.58	0.61	0.58
	TF-IDF	Original	0.75	0.54	0.52	0.51
	Word Embedding	Original	0.75	0.60	0.57	0.57
Support	TF-IDF	ROS	0.82	0.73	0.65	0.67
Vector	Word Embedding	ROS	0.79	0.39	0.50	0.44
Machine (SVM)	TF-IDF	US	0.68	0.62	0.67	0.62
	Word Embedding	US	0.71	0.72	0.71	0.71
	TF-IDF	Original	0.79	0.67	0.53	0.50
	Word Embedding	Original	0.79	0.39	0.50	0.44
Logistic	TF-IDF	ROS	0.80	0.70	0.69	0.69
Regression	Word Embedding	ROS	0.79	0.65	0.52	0.48
	TF-IDF	US	0.70	0.63	0.69	0.64
	Word Embedding	US	0.72	0.72	0.72	0.72
	TF-IDF	Original	0.79	0.65	0.52	0.48
	Word Embedding	Original	0.79	0.65	0.54	0.53

Gambar 2. Hasil *Random Over Sampling*, *Under Sampling* dan original.

D. Simpulan

Penelitian ini berhasil mengklasifikasikan berita subjektif dan objektif menggunakan algoritma machine learning dengan representasi teks *TF-IDF* dan *Word Embedding*, serta teknik sampling *Random Over Sampling (ROS)* dan *Undersampling (US)*. Hasil menunjukkan bahwa *SVM* dengan *Undersampling (US)* memberikan akurasi tertinggi (82%) dan keseimbangan terbaik antara *precision* dan *recall*. Meskipun *TF-IDF* dengan *ROS* memberikan akurasi tinggi, *Word Embedding* lebih efektif dalam menangani data dengan pola linguistik kompleks, terutama untuk *Recall* dan *F1-Score*. Implementasi berbasis *Streamlit* memungkinkan pengguna untuk menganalisis berita secara langsung, mendukung literasi digital, dan membantu memilah berita objektif dan subjektif. Pengembangan lebih lanjut dengan penambahan data dan penggunaan teknik seperti *SMOTE* serta eksplorasi algoritma lain seperti *Random Forest* atau *XGBoost* diharapkan dapat meningkatkan kinerja model. Penelitian ini berpotensi mengurangi penyebaran informasi yang tidak akurat dan mendukung masyarakat dalam menyaring berita yang diterima.

Daftar Pustaka

- [1] Maulana, M. I., Martanto, M., & Hayati, U. (2024). PERBANDINGAN ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBORS UNTUK KLASIFIKASI TOPIK BERITA PADA SITUS DETIK. COM. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3733-3742.
- [2] Bhakti, I. N., Sholikhin, A. Z., Abi Lukman, M., Daniati, E., & Ristyawan, A. (2024, July). Klasifikasi Kategori Berita Menggunakan Naive Bayes. In *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)* (Vol. 8, No. 2, pp. 1155-1164).
- [3] Dey, K., Tarannum, P., Hasan, M. A., & Noori, S. R. H. (2023). NN at CheckThat!-2023: Subjectivity in News Articles Classification with Transformer Based Models. In *CLEF (Working Notes)* (pp. 318-328).
- [4] Salas-Jimenez, K., Díaz, I., Gómez-Adorno, H., Bel-Enguix, G., & Sierra, G. (2024). JK_PCIC_UNAM at CheckThat! 2024: analysis of subjectivity in news sentences using transformers based models. *Faggioli et al.[22]*.
- [5] Kusuma, Z. A., Widayat, W., & Kom, S. (2024). Analisis Sentimen Pengguna Twitter Terhadap Program Makan Siang Gratis Menggunakan Metode Naïve Bayes (Doctoral dissertation, Universitas Muhammadiyah Surakarta).
- [6] Maryam, M., & Ariono, H. W. (2022). Sistem Pakar Pengklasifikasi Stadium Kanker Serviks Berbasis Mobile Menggunakan Metode Decision Tree. *Jurnal Kajian Ilmiah*, 22(3), 267-278.
- [7] Dewi, N. L. P. R., Wijaya, I. N. S. W., Purnamawan, I. K., & Marti, N. W. (2024). Model Classifer Judul Berita Pariwisata Indonesia Berdasarkan Sentimen. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(1), 117-124.
- [8] Dirgantara Yudhanata, I., & Endah Sudarmilah, S. T. (2024). Analisis Sentimen Terhadap Isu Islamofobia Pada Platform Twitter Menggunakan Metode Klasifikasi Naïve Bayes (Doctoral dissertation, Universitas Muhammadiyah Surakarta).
- [9] Bensi, M. E., & Esquivel, R. A. (2023). Unraveling the Significance of the Classification Tree Algorithm in Machine Learning: A Literature Review. *European Journal of Theoretical and Applied Sciences*, 1(5), 604-611.

- [10] Kusuma, F. A., Pamungkas, E. W., Kom, S., & Kom, M. (2023). *Pendeteksian Hate Speech Pada Sosial Media Indonesia Dengan Algoritma Support Vector Machine (Svm) Dan Decision Tree* (Doctoral dissertation, Universitas Muhammadiyah Surakarta).