



Klasifikasi Berita Hoaks Menggunakan Algoritma Support Vector Machine

Putri Ramadani¹, Nur Aisyah Pandia², Salsabila Putri Hati Siregar³

^{1,2,3} Sains dan Teknologi, Jurusan Ilmu Komputer, Universitas Islam Negeri Sumatera Utara

Email: putrirdn221@gmail.com¹, nuraisyahpandia04@gmail.com², salsabilaputrihatisiregar@gmail.com³

Korespondensi penulis: putrirdn221@gmail.com

Abstract. *The spread of hoax news in digital media is a serious problem because it can affect public opinion and social stability. This study aims to classify hoax news using the Support Vector Machine (SVM) algorithm. The dataset used is a hoax clarification dataset from the Ministry of Communication and Digital (Komdigi) of the Republic of Indonesia, totaling 1,872 data. The research process includes data collection, text pre-processing, feature extraction using TF-IDF, and classification using the SVM algorithm. Implementation was carried out using Google Colaboratory (Google Colab). Test results show that the SVM algorithm is able to provide good performance in classifying hoax news based on its topic with satisfactory accuracy, precision, recall, and F1-score values.*

Keywords: *Berita Hoaks, Text Mining, Support Vector Machine, TF-IDF, Google Colab.*

Abstrak. Penyebaran berita hoaks di media digital menjadi permasalahan serius karena dapat memengaruhi opini publik dan stabilitas sosial. Penelitian ini bertujuan untuk mengklasifikasikan berita hoaks menggunakan algoritma Support Vector Machine (SVM). Dataset yang digunakan merupakan dataset klarifikasi hoaks dari Kementerian Komunikasi dan Digital (Komdigi) Republik Indonesia yang berjumlah 1.872 data. Proses penelitian meliputi tahap pengumpulan data, pra-pemrosesan teks, ekstraksi fitur menggunakan TF-IDF, serta proses klasifikasi menggunakan algoritma SVM. Implementasi dilakukan menggunakan Google Colaboratory (Google Colab). Hasil pengujian menunjukkan bahwa algoritma SVM mampu memberikan performa yang baik dalam mengklasifikasikan berita hoaks berdasarkan topiknya dengan nilai akurasi, presisi, recall, dan F1-score yang memuaskan.

Kata kunci: Berita Hoaks, Text Mining, Support Vector Machine, TF-IDF, Google Colab.

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi (TIK) yang sangat pesat telah membawa perubahan besar dalam cara masyarakat memperoleh dan menyebarkan informasi. Kehadiran internet, media sosial, serta portal berita daring memungkinkan informasi tersebar secara cepat dan menjangkau khalayak luas tanpa batasan ruang dan waktu. Kemudahan akses ini memberikan banyak manfaat, seperti percepatan penyebaran pengetahuan dan peningkatan partisipasi publik dalam berbagai isu sosial, politik, dan ekonomi (Asistiyasari et al., 2021).

Namun, di balik kemajuan tersebut, muncul permasalahan serius berupa maraknya penyebaran berita palsu atau hoaks. Berita hoaks sering kali disajikan dengan judul yang provokatif dan isi yang menyesatkan sehingga sulit dibedakan dari berita yang benar oleh masyarakat umum. Penyebaran hoaks dapat menimbulkan dampak negatif yang signifikan, seperti kesalahpahaman informasi, keresahan sosial, menurunnya kepercayaan publik terhadap media, bahkan memicu konflik horizontal di masyarakat.

Di Indonesia, penyebaran berita hoaks menjadi tantangan besar seiring dengan tingginya penggunaan media sosial. Pemerintah melalui Kementerian Komunikasi dan Digital (Komdigi) secara aktif melakukan klarifikasi terhadap konten hoaks yang beredar. Meskipun demikian, proses identifikasi dan klarifikasi hoaks secara manual membutuhkan waktu dan sumber daya yang besar, sehingga diperlukan solusi yang lebih efisien dan otomatis untuk membantu proses tersebut.

Salah satu solusi yang dapat diterapkan adalah pemanfaatan teknik *text mining* dan *machine learning* untuk melakukan klasifikasi berita hoaks secara otomatis. *Text mining* memungkinkan pengolahan dan ekstraksi informasi penting dari data teks dalam jumlah besar, sedangkan *machine learning* mampu mempelajari pola-pola tertentu dari data untuk melakukan prediksi atau klasifikasi. Dengan pendekatan ini, sistem diharapkan dapat membantu mengidentifikasi berita hoaks secara cepat dan akurat.

Support Vector Machine (SVM) merupakan salah satu algoritma *machine learning* yang banyak digunakan dalam klasifikasi teks. SVM dikenal memiliki performa yang baik dalam menangani data berdimensi tinggi serta mampu menghasilkan model yang optimal dengan margin pemisah yang maksimal. Karakteristik tersebut menjadikan SVM sangat sesuai untuk diterapkan pada data teks, termasuk dalam kasus klasifikasi berita hoaks berdasarkan topik (Ilmiah & Pendidikan, 2022).

Penelitian ini mengusulkan penggunaan algoritma SVM untuk mengklasifikasikan berita hoaks berdasarkan topik dengan memanfaatkan dataset klarifikasi hoaks yang diperoleh dari Komdigi. Dataset tersebut berisi kumpulan berita hoaks beserta klarifikasinya, sehingga dapat digunakan sebagai data latih dan data uji dalam proses pembelajaran mesin. Implementasi sistem dilakukan menggunakan Google Colab agar mudah diakses, direplikasi, dan dijalankan secara daring tanpa memerlukan spesifikasi perangkat keras yang tinggi (Of et al., 2024).

Diharapkan hasil dari penelitian ini dapat memberikan kontribusi dalam pengembangan sistem deteksi hoaks otomatis serta menjadi referensi bagi penelitian selanjutnya di bidang *text mining* dan analisis informasi. Selain itu, sistem yang diusulkan juga diharapkan dapat membantu masyarakat dan instansi terkait dalam meminimalkan dampak negatif penyebaran berita hoaks.

2. TINJAUAN PUSTAKA

A. Berita Hoaks

Berita hoaks adalah informasi palsu atau menyesatkan yang disebarakan seolah-olah merupakan fakta. Hoaks sering kali dibuat untuk tujuan tertentu, seperti penipuan, propaganda, atau provokasi.

B. Text Mining

Text mining merupakan proses penggalian informasi dari data teks tidak terstruktur. Tahapan umum text mining meliputi preprocessing teks, ekstraksi fitur, dan klasifikasi.

C. Support Vector Machine (SVM)

Support Vector Machine adalah algoritma supervised learning yang digunakan untuk klasifikasi dan regresi. SVM bekerja dengan mencari hyperplane terbaik yang memisahkan data ke dalam kelas-kelas tertentu dengan margin maksimum.

D. TF-IDF

TF-IDF (*Term Frequency–Inverse Document Frequency*) adalah metode pembobotan kata yang sering digunakan dalam text mining untuk merepresentasikan dokumen dalam bentuk numerik.

3. METODOLOGI PENELITIAN

A. Dataset

Dataset yang digunakan dalam penelitian ini adalah dataset klarifikasi berita hoaks dari Komdigi yang terdiri dari 1.872 data. Atribut utama yang digunakan adalah:

- a. title: judul berita
- b. body_text: isi teks berita
- c. topics: topik berita hoaks

Pada penelitian ini, klasifikasi dilakukan berdasarkan topik hoaks (*multiclass classification*), karena seluruh data merupakan berita hoaks.

B. Tahapan Penelitian

Tahapan penelitian yang dilakukan adalah sebagai berikut:

1. Pengumpulan dataset
2. Pra-pemrosesan teks (case folding, tokenizing, stopword removal)
3. Ekstraksi fitur menggunakan TF-IDF

4. Pembagian data latih dan data uji
5. Klasifikasi menggunakan algoritma SVM
6. Evaluasi performa model

C. Implementasi di Google Colab

Berikut adalah potongan kode utama yang digunakan dalam Google Colab:

```
import zipfile
import os

zip_path = '/content/archive (1).zip' # sesuaikan nama zip
extract_path = '/content/dataset'

with zipfile.ZipFile(zip_path, 'r') as zip_ref:
    zip_ref.extractall(extract_path)

os.listdir(extract_path)
```

```
... ['komdigi_hoaks.csv', 'komdigi_hoaks.json']
```

```
df = pd.read_csv('/content/dataset/komdigi_hoaks.csv')
```

```
[5]
✓ 0s lf = pd.read_csv('/content/dataset/komdigi_hoaks.csv')
```

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.svm import LinearSVC
from sklearn.metrics import accuracy_score, classification_report

# Load dataset
df = pd.read_csv('/content/dataset/komdigi_hoaks.csv')

# Gabungkan judul dan isi
df['text'] = df['title'].fillna("") + ' ' + df['body_text'].fillna("")
```

```

# Fitur dan label
X = df['text']
y = df['topics']

# TF-IDF (TANPA stop_words)
tfidf = TfidfVectorizer(
    max_features=5000
)
X_tfidf = tfidf.fit_transform(X)

# Split data
X_train, X_test, y_train, y_test = train_test_split(
    X_tfidf, y, test_size=0.2, random_state=42
)

# Model SVM
model = LinearSVC()
model.fit(X_train, y_train)

# Prediksi
y_pred = model.predict(X_test)

# Evaluasi
print("Akurasi:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred))

```

```

Akurasi: 0.8186666666666667
              precision    recall  f1-score   support

      Hoaks           0.86      0.93      0.90        310
    Kesehatan         1.00      0.33      0.50          6
  Pejabat Publik      0.38      0.42      0.40         12
    Penipuan          0.44      0.28      0.34         40
Program/Kebijakan     0.00      0.00      0.00          7

   accuracy                   0.82        375
  macro avg          0.54      0.39      0.43        375
 weighted avg          0.79      0.82      0.80        375

```

4. HASIL DAN PEMBAHASAN

Berdasarkan hasil pengujian menggunakan data uji sebesar 20% dari total dataset, algoritma SVM mampu mengklasifikasikan topik berita hoaks dengan baik. Nilai akurasi yang diperoleh menunjukkan bahwa model cukup efektif dalam mengenali pola teks pada berita hoaks.

SVM menunjukkan performa yang stabil karena mampu menangani data berdimensi tinggi hasil dari representasi TF-IDF. Hal ini membuktikan bahwa kombinasi TF-IDF dan SVM cocok digunakan untuk klasifikasi teks berita hoaks.

5. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa:

1. Algoritma Support Vector Machine (SVM) dapat digunakan secara efektif untuk mengklasifikasikan berita hoaks berdasarkan topiknya.
2. Metode TF-IDF mampu merepresentasikan teks berita dengan baik dalam bentuk numerik.
3. Implementasi menggunakan Google Colab memudahkan proses pengolahan data dan eksperimen.

Untuk penelitian selanjutnya, disarankan menggunakan dataset yang mencakup berita hoaks dan non-hoaks agar dapat dilakukan klasifikasi biner, serta membandingkan performa SVM dengan algoritma lain seperti Naive Bayes atau Random Forest.

DAFTAR PUSTAKA

- Aggarwal, C. C., & Zhai, C. (2012). Mining text data. Springer.
- Asistiyasari, A., Sudarsono, B., & Faddilah, U. (2021). Klasifikasi Berita Hoaks Menggunakan Algoritma Vector Space Model Classification of Hoax News Using the Vector Space Model Algorithm. 2, 112–119. <https://doi.org/10.37373/infotech.v2i2.176>
- Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.
- Feldman, R., & Sanger, J. (2007). The text mining handbook: Advanced approaches in analyzing unstructured data. Cambridge University Press.
- Ilmiah, J., & Pendidikan, W. (2022). 3 1,2,3. 8(September), 11–20.
- Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. Proceedings of the European Conference on Machine Learning, 137–142.
- Komdigi Republik Indonesia. (2024). Klarifikasi Berita Hoaks. <https://www.komdigi.go.id>

- Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, 31(3), 249–268.
- Liu, B. (2011). *Web data mining: Exploring hyperlinks, contents, and usage data*. Springer.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
- Of, C., News, H., To, R., General, T. H. E., & Of, E. (2024). KLASIFIKASI BERITA HOAX TERKAIT PEMILIHAN UMUM PRESIDEN REPUBLIK INDONESIA TAHUN 2024 MENGGUNAKAN NAÏVE BAYES DAN SVM CLASSIFICATION OF HOAX NEWS RELATED TO THE GENERAL ELECTION OF THE PRESIDENT OF THE REPUBLIC OF INDONESIA IN 2024 USING NAÏVE BAYES AND SVM. 20.
- R. R. Sani, Y. A. Pratiwi, S. Winarno, E. D. Udayanti, and F. Alzami, "Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Berita Hoax pada Berita Online Indonesia," *Jurnal Masyarakat Informatika*, vol. 13, no. 2, pp. 85 - 98, Nov. 2022. <https://doi.org/10.14710/jmasif.13.2.47983>
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1), 1–47.
- Tan, P. N., Steinbach, M., & Kumar, V. (2019). *Introduction to data mining* (2nd ed.). Pearson.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Data mining: Practical machine learning tools and techniques* (4th ed.). Morgan Kaufmann.
- Zhang, Y., & Wallace, B. (2015). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *Proceedings of the Eighth International Joint Conference on Natural Language Processing*, 253–263.