



Penerapan Machine Learning untuk Pengklasifikasian Hoaks pada Platform Media Sosial: Studi Literatur

Ahmad Wafi Fathurrahman¹, Muhammad Thoriqulhaq², Farhan Dequika Pradana Putra Arianto³

^{1,2,3} Program Studi Sains Data, UPN “Veteran” Jawa Timur

²21083010094@student.upnjatim.ac.id

³21083010062@student.upnjatim.ac.id

*Corresponding author email: 21083010011@student.upnjatim.ac.id

Abstract: Fake news or hoaxes are often discovered as technology develops. The purpose of making a hoax is to create a negative perspective on its readers. Therefore, a study was conducted to classify hoax news or not. One of them is by applying Machine Learning (ML) algorithm. The literature study that we use is to collect as many as articles that have been selected. The result is the application of Machine Learning algorithms can be applied to classify hoax news.

Keywords: Hoax, Machine Learning, Classification

Abstrak: Berita palsu atau hoax sering kali ditemukan seiring berkembangnya teknologi. Tujuan dibuatnya hoax yakni untuk menimbulkan perspektif yang negatif terhadap para pembacanya. Maka dari itu, dilakukan sebuah penelitian untuk mengklasifikasi berita hoax atau bukan. Salah satunya dengan menerapkan algoritma Machine Learning (ML). Studi pustaka yang kami gunakan adalah dengan mengumpulkan sebanyak artikel-artikel yang telah dipilih. Hasilnya adalah penerapan algoritma Machine Learning dapat diterapkan untuk mengklasifikasikan berita hoax.

Kata kunci: Hoaks, Machine Learning, Klasifikasi

I. PENDAHULUAN

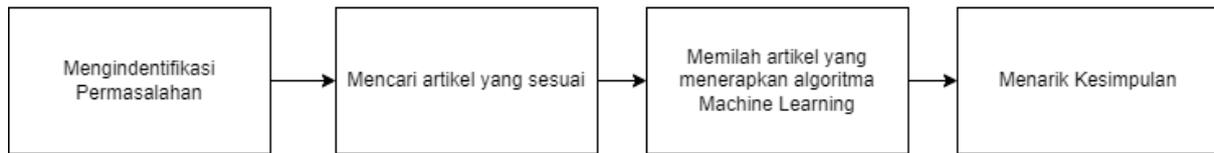
Salah satu perwujudan perkembangan teknologi adalah terdapat pada bidang teknologi dan informasi. Sekarang, teknologi dan informasi telah menjadi sarana atau media komunikasi dan sumber informasi yang begitu cepat. Teknologi tersebut telah dimanfaatkan hampir di setiap aspek kehidupan manusia, mulai dari pekerjaan, pendidikan, kesehatan, hingga ke pemerintahan. Dengan ditemukannya perangkat media yang berbasis internet, seperti media sosial (Instagram, Twitter, Facebook, dll.) berbagai jenis informasi menjadi dengan mudah ditemukan hanya dengan mengaksesnya melalui jaringan internet.

Menurut survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) [1], pengguna internet di Indonesia pada tahun 2021-2022 mencapai 210,03 juta. Dalam artian, 77,02% dari penduduk di Indonesia adalah pengguna internet. Dengan begitu banyaknya pengguna internet di Indonesia, tentu saja berdampak kepada aktivitas di media sosial. Banyaknya aktivitas terkait berita di media sosial dapat menyebabkan masalah terkait akan validnya sebuah informasi.

Berita palsu atau hoaks masih menjadi permasalahan besar di Indonesia. Hoaks dapat merugikan siapapun yang menjadi target dari pembuatnya. Hoaks merupakan informasi atau berita yang berisi hal-hal yang belum pasti atau benar-benar bukan merupakan fakta yang terjadi [2]. Untuk mengurangi penyebaran hoaks, diperlukan sebuah sistem untuk mengklasifikasikan berita yang tersebar di media sosial. Dalam penelitian ini, kami mengumpulkan beberapa artikel yang di dalamnya menerapkan algoritma Machine Learning (ML). Sehingga, bertujuan untuk mengetahui tingkat akurasi hoaks suatu berita yang sebelumnya telah diolah.

II. METODE PENELITIAN

Studi pustaka yang kami lakukan merupakan studi pustaka sistematis. Dimana dalam studi pustaka tersebut terdapat langkah-langkah berurutan yang dilakukan. Metode pada penelitian kami yakni mengidentifikasi permasalahan, mencari artikel yang sesuai, memilah artikel yang menerapkan algoritma Machine Learning dan menarik kesimpulan. Berikut diagram metode penelitian kami.



Mengidentifikasi permasalahan yang dilakukan yakni memerlukan penerapan algoritma Machine Learning dalam pengklasifikasian berita palsu atau hoaks. Diharapkan nantinya tingkat akurasi dari pengaplikasian machine learning dapat membantu untuk mengetahui berita yang diolah tersebut termasuk ke dalam berita hoaks atau bukan.

III. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan pengolahan *text mining* yang sudah dilakukan dengan cara menerapkan algoritma berupa *Support Vector Machine* (SVM) dan dipadukan dengan dua fungsi *kernel* seperti *kernel linear* dan *kernel RBF* yang menyebabkan berita *hoax* dan bukan *hoax* dengan berita Covid-19 yang diambil sampelnya dari berbagai situs *online*.

Tabel 1. Daftar Literatur

Author	Metode	Hasil
[3]	SGD, Naïve Bayes, dan Random Forest	Diperoleh hasil akurasi yang paling tinggi terdapat pada algoritma SGD dengan nilai akurasi 84,92%.
[4]	SVM	Nilai akurasi terbaik adalah 92,90% pada scenario tiga yang memiliki perbandingan data training dan data testing (80:20)
[5]	KNN	Akurasi sebesar 48% dengan perbandingan data train dan data test (80:20) serta nilai $k = 5$.
[6]	Naïve Bayes	Pada pengujian statis terhadap 600 berita memperoleh hasil akurasi senilai 82,6%
[7]		Pada pengujian dinamis terhadap

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan beberapa hal, antara lain penerapan algoritma Support Vector Machine dengan kernel linear dan RBF untuk klasifikasi berita hoax Covid-19 dilakukan dengan menggunakan beberapa proses seperti case folding, tokenizing, filtering, stemming, pembobotan kata (TF-IDF) dan melakukan pemodelan dengan 4 skenario pengujian secara acak untuk mendapatkan model terbaik dari kedua kernel algoritma Support Vector Machine. Algoritma Support Vector Machine (SVM) dengan kernel linear memiliki hasil prediksi yang bagus pada skenario 3 (80:20) karena model sanggup dalam mengklasifikasikan berita hoax dan bukan hoax covid-19. Hasil yang didapat yaitu ada 111 data hoax yang diprediksi hoax, ada 61 data hoax yang diprediksi bukan hoax. Sedangkan data bukan hoax yang diprediksi hoax ada 55 dan data bukan hoax diprediksi bukan hoax ada 1408. Selain itu akurasi yang didapat pada skenario 80:20 juga memiliki nilai akurasi tertinggi yaitu sebesar 92,90%.



REFERENSI

1. “Laporan Hasil Survei Profil Internet 2022”. <https://apjii.or.id/content/read/39/559/Hasil-Survei-Profil-Internet-Indonesia-2022> (Diakses pada 5 September 2022).
2. C. Juditha. “Interaksi Komunikasi Hoax di Media Sosial serta Antisipasinya *Hoax Communication Interactivity in Social Media and Anticipation*”. 2018, *Vol. 3 No. 1*, pp 31-44.
3. S. M. Tambunan; Y. Nataliani; E. S. Lestari. “Perbandingan Klasifikasi dengan Pendekatan Pembelajaran Mesin untuk Mengidentifikasi Tweet Hoaks di Media Sosial Twitter”. 2021, *Vol. 7 No. 2*, pp. 112-120.
4. Author 1, A.B. (University, City, State, Country); Author 2, C. (Institute, City, State, Country). Personal communication, 2012.
5. Author 1, A.B.; Author 2, C.D.; Author 3, E.F. Title of Presentation. In Title of the Collected Work (if available), Proceedings of the Name of the Conference, Location of Conference, Country, Date of Conference; Editor 1, Editor 2, Eds. (if available); Publisher: City, Country, Year (if available); Abstract Number (optional), Pagination (optional).
6. Author 1, A.B. Title of Thesis. Level of Thesis, Degree-Granting University, Location of University, Date of Completion.
7. Title of Site. Available online: URL (accessed on Day Month Year).