

Perbandingan Naïve Bayes dan CNN yang Dioptimasi PSO pada Identifikasi Berita Hoax Politik Indonesia

 Universiti Teknologi MARA

Document Details

Submission ID

trn:oid::13381:103224220

Submission Date

Jul 2, 2025, 2:58 PM GMT+7

Download Date

Jul 2, 2025, 3:02 PM GMT+7

File Name

Perbandingan Naïve Bayes dan CNN yang Dioptimasi PSO pada Identifikasi Berita Hoax Politik In....pdf

File Size

496.9 KB

13 Pages

7,397 Words

45,743 Characters

11% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- ▶ Bibliography
- ▶ Internet sources
- ▶ Crossref database
- ▶ Crossref posted content database

Top Sources

- 0%  Internet sources
- 1%  Publications
- 10%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Top Sources

- 0% Internet sources
- 1% Publications
- 10% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Submitted works	
Konsorsium Perguruan Tinggi Swasta Indonesia II on 2025-06-25		<1%
<hr/>		
2	Submitted works	
Sriwijaya University on 2020-07-07		<1%
<hr/>		
3	Submitted works	
Universitas Sebelas Maret on 2025-06-15		<1%
<hr/>		
4	Submitted works	
Universitas Muria Kudus on 2024-05-30		<1%
<hr/>		
5	Submitted works	
Perguruan Tinggi Pelita Bangsa on 2020-10-11		<1%
<hr/>		
6	Submitted works	
Universiti Selangor on 2018-10-07		<1%
<hr/>		
7	Submitted works	
STT PLN on 2023-08-14		<1%
<hr/>		
8	Submitted works	
UM Surabaya on 2024-03-23		<1%
<hr/>		
9	Submitted works	
Universitas Budi Luhur on 2023-08-23		<1%
<hr/>		
10	Submitted works	
Padjajaran University on 2023-06-26		<1%
<hr/>		
11	Submitted works	
Politeknik Negeri Bandung on 2019-08-08		<1%
<hr/>		

12	Submitted works	Universitas Dian Nuswantoro on 2015-06-17	<1%
13	Submitted works	Binus University International on 2025-03-21	<1%
14	Submitted works	Bogazici University on 2021-04-29	<1%
15	Submitted works	Universitas Indonesia on 2020-06-26	<1%
16	Submitted works	Universitas Katolik Widya Mandala on 2019-02-27	<1%
17	Submitted works	STT PLN on 2024-08-06	<1%
18	Submitted works	National Institute Of Technology, Tiruchirappalli on 2023-06-13	<1%
19	Submitted works	Universitas Dian Nuswantoro on 2015-08-14	<1%
20	Submitted works	Universitas Muslim Indonesia on 2024-03-25	<1%
21	Submitted works	UIN Sultan Syarif Kasim Riau on 2021-02-16	<1%
22	Submitted works	Universitas Pelita Harapan on 2022-06-09	<1%
23	Submitted works	Forum Perpustakaan Perguruan Tinggi Indonesia Jawa Tengah on 2025-06-18	<1%
24	Submitted works	Sriwijaya University on 2019-07-04	<1%
25	Submitted works	Udayana University on 2025-01-01	<1%

26	Submitted works	Universitas Sumatera Utara on 2024-12-23	<1%
27	Publication	Almasoum, Mohammad Ahmed. "Evaluating Faults Detection and Their Impact o..."	<1%
28	Submitted works	Sriwijaya University on 2023-03-28	<1%
29	Submitted works	Tarumanagara University on 2023-07-13	<1%
30	Submitted works	Universitas Brawijaya on 2017-12-18	<1%
31	Submitted works	Universitas Brawijaya on 2023-12-05	<1%
32	Submitted works	Universitas PGRI Semarang on 2023-07-24	<1%
33	Publication	Durgesh Kumar Mishra, Nilanjan Dey, Bharat Singh Deora, Amit Joshi. "ICT for Co..."	<1%
34	Submitted works	Forum Komunikasi Perpustakaan Perguruan Tinggi Kristen Indonesia (FKPPTKI) o...	<1%
35	Submitted works	Marquette University on 2023-10-24	<1%
36	Submitted works	Sriwijaya University on 2023-07-25	<1%
37	Submitted works	State Islamic University of Alauddin Makassar on 2023-09-19	<1%
38	Submitted works	Universitas Brawijaya on 2021-11-22	<1%
39	Submitted works	Universitas Dian Nuswantoro on 2015-06-17	<1%

40	Submitted works	Universitas Dian Nuswantoro on 2015-07-23	<1%
41	Submitted works	Universitas Dian Nuswantoro on 2015-08-14	<1%
42	Submitted works	Universitas Diponegoro on 2024-05-26	<1%
43	Submitted works	Universitas Maritim Raja Ali Haji on 2023-01-27	<1%
44	Submitted works	Universitas Muhammadiyah Purwokerto on 2024-06-20	<1%
45	Submitted works	Universitas Muslim Indonesia on 2025-04-16	<1%
46	Submitted works	Universitas Muslim Indonesia on 2025-06-10	<1%
47	Submitted works	Universitas Negeri Semarang on 2021-01-11	<1%
48	Submitted works	Universitas Sebelas Maret on 2022-12-22	<1%
49	Submitted works	Dewan Perwakilan Rakyat on 2020-10-04	<1%
50	Submitted works	STT PLN on 2020-09-17	<1%
51	Submitted works	Universitas Brawijaya on 2020-12-22	<1%
52	Submitted works	Universitas Ibn Khaldun on 2019-05-03	<1%
53	Submitted works	Universitas Maritim Raja Ali Haji on 2025-02-04	<1%

54	Submitted works	Landmark University on 2023-11-21	<1%
55	Submitted works	STT PLN on 2020-09-23	<1%
56	Submitted works	STT PLN on 2024-08-28	<1%
57	Publication	Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical ...	<1%
58	Submitted works	Universitas Brawijaya on 2021-04-11	<1%
59	Submitted works	Universitas Islam Indonesia on 2018-03-16	<1%

Perbandingan Naïve Bayes dan CNN yang Dioptimasi PSO pada Identifikasi Berita Hoax Politik Indonesia

Yusuf Kurnia^{1)*}, Ellysha Dwiyanthi Kusuma²⁾, Lianny Wydiastuty Kusuma³⁾, Suwitno⁴⁾,
Welman Apridius⁵⁾

¹⁾²⁾³⁾⁴⁾⁵⁾ Universitas Buddhi Dharma

Jl. Imam Bonjol No.41, Tangerang, Indonesia

¹⁾yusuf.kurnia@ubd.ac.id

²⁾ellysha.dwiyanthi@ubd.ac.id

³⁾lianny.wydiastuty@ubd.ac.id

⁴⁾suwitno.suwitno@ubd.ac.id

⁵⁾we.apridius@gmail.com

Article history:

Received 27 Maret 2024;
Revised 16 April 2024;
Accepted 23 April 2024;
Available online 30 April 2024

Keywords:

Convolutional Neural Network
Hoax Politik
Naïve Bayes
Particle Swarm Optimization
Text Preprocessing

Abstract

Berita palsu dalam politik menjadi ancaman serius terhadap stabilitas demokrasi dan kepercayaan publik terhadap informasi. Fenomena ini tidak hanya meresahkan, tetapi juga memiliki dampak yang dapat mengganggu proses demokrasi serta kepercayaan masyarakat terhadap media dan pemerintah. Penelitian ini bertujuan untuk menyelidiki dan membandingkan kinerja dua algoritma yang berbeda, yaitu *Naïve Bayes* (NB) dan *Convolutional Neural Network* (CNN), yang telah dioptimalkan menggunakan *Particle Swarm Optimization* (PSO), dalam mendeteksi berita palsu di ranah politik Indonesia. Untuk mencapai tujuan ini, penelitian menggunakan dataset berita politik yang telah melalui proses *text preprocessing*, termasuk pembersihan data dan normalisasi teks. Hasil penelitian menunjukkan bahwa model *Naïve Bayes* secara konsisten mampu mengungguli kinerja CNN, baik dengan atau tanpa penerapan PSO. Akurasi model *Naïve Bayes* mencapai 90.71%, sementara CNN mencapai 80.86% tanpa PSO, 79.68% dengan PSO, dan *Naïve Bayes* dengan PSO mencapai 90.25%. Temuan ini menegaskan bahwa pendekatan menggunakan algoritma *Naïve Bayes* memiliki potensi lebih besar dalam mengidentifikasi berita palsu dalam konteks politik Indonesia. Kontribusi signifikan dari penelitian ini terletak pada pemahaman yang lebih mendalam mengenai metode deteksi berita palsu, serta memberikan wawasan yang penting bagi pengembangan strategi yang efektif dalam menanggulangi permasalahan berita palsu di era digital. Oleh karena itu, diharapkan bahwa penelitian ini akan memberikan nilai tambah dalam upaya mempertahankan keaslian informasi politik dan meningkatkan kualitas demokrasi di Indonesia.

I. PENDAHULUAN

Tingginya pertumbuhan pengguna internet di Indonesia telah menjadi perhatian utama di tengah era digital yang sedang berkembang pesat. Menurut informasi yang dirilis oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), jumlah pengguna internet di Indonesia terus meningkat secara signifikan dari tahun ke tahun. Antara 2021 dan 2022, jumlah pengguna internet mencapai 210,03 juta, mengalami kenaikan sebesar 6,78% dari periode sebelumnya yang mencatat 196,7 juta pengguna. Fenomena ini menandakan adopsi yang semakin meluas terhadap teknologi internet di kalangan masyarakat Indonesia. Faktor-faktor seperti peningkatan aksesibilitas internet, berkembangnya penetrasi perangkat pintar, serta adopsi teknologi digital di berbagai sektor kehidupan, termasuk pendidikan, bisnis, dan pemerintahan, telah turut mendorong pertumbuhan yang pesat dalam jumlah pengguna internet di negara ini. Sementara itu, hal ini juga menandakan perlunya perhatian lebih lanjut terhadap infrastruktur teknologi informasi dan komunikasi (TIK) guna memastikan kelangsungan dan peningkatan aksesibilitas internet yang merata di seluruh wilayah Indonesia, termasuk di daerah-daerah yang masih terpencil. Ini juga mengakibatkan tingkat penetrasi internet di Indonesia mencapai 77,02%. [1]. Dalam periode tiga tahun

*Corresponding

48 terakhir, mulai dari Agustus 2018 hingga awal 2022, Kominfo telah menemukan 9.546 berita palsu tersebar di internet. [2].

Ketika masyarakat menerima informasi palsu tanpa pertimbangan kritis yang memadai, hal ini dapat mengakibatkan penurunan kualitas debat publik, munculnya konflik sosial, serta perpecahan yang lebih besar di masyarakat. Di samping itu, beberapa pihak juga memanfaatkan tindakan tersebut untuk menghasut masyarakat dan mengancam kestabilan persatuan negara [3]. Teknik text mining merupakan metode yang efektif untuk memeriksa apakah suatu berita termasuk dalam kategori hoax dengan cara mengklasifikasikan teks. Selain itu, teknik ini juga bermanfaat untuk menganalisis sentimen atau mengelompokkan teks berdasarkan pola dan ciri-ciri tertentu. [4].

Teknologi *Text Mining* telah membuktikan kemampuannya dalam mengidentifikasi, mengekstraksi, dan menganalisis informasi semantik dari berbagai sumber teks dengan tingkat kualitas yang lebih tinggi. Melalui penggunaan algoritma dan metode analisis bahasa alami yang canggih, *Text Mining* mampu secara otomatis mengevaluasi pola dan tren yang muncul dalam teks-teks yang dianalisis, termasuk literatur, dokumen, dan data teks lainnya. Dengan demikian, teknologi ini memungkinkan pengguna untuk mendapatkan pemahaman konten yang lebih mendalam dan komprehensif daripada sekadar menganalisis jumlah kata dasar [5].

Melalui penggunaan algoritma *Machine Learning* yang terintegrasi dengan teknik analisis data yang cerdas, sistem tersebut dapat diarahkan untuk mengenali pola-pola khas dan ciri-ciri khusus yang terkait dengan konten hoax, termasuk pemahaman terhadap konteks sosial dan historis di balik berita tersebut. Dengan demikian, metode ini diharapkan dapat memberikan pemahaman mendalam terhadap aspek-aspek konten yang mungkin menandakan adanya kecenderungan berita *hoax* [6].

32 *Naïve Bayes Classifier* adalah suatu model *Machine Learning* yang bersifat probabilistik dan digunakan untuk melakukan klasifikasi berdasarkan Teorema *Bayes*. [7]. Penggunaan algoritma *Naïve Bayes* sebagai metode utama dalam klasifikasi data telah meluas dan mendapat penerimaan luas dari berbagai disiplin ilmu, mulai dari bidang medis hingga pengelolaan data dalam jaringan komputer, serta analisis teks. Dengan kemampuannya yang sederhana dan kemudahan implementasinya, *Naïve Bayes* terus menjadi salah satu pilihan utama dalam klasifikasi data di berbagai domain, meskipun juga diakui bahwa kemampuan model ini masih memerlukan pemahaman yang mendalam terhadap asumsi-asumsi yang diberlakukan. Ini karena kemudahannya, efektivitasnya, dan kapabilitasnya dalam mengenali pola data melalui representasi grafis [8].

58 K-Nearest Neighbor (KNN) merupakan algoritma yang digunakan untuk mengklasifikasikan objek berdasarkan data pembelajaran yang memiliki jarak terdekat dengan objek yang ingin diklasifikasikan [9]. Dengan memanfaatkan konsep jarak antara objek-objek dalam ruang fitur, KNN mengidentifikasi dan mengklasifikasikan suatu objek berdasarkan mayoritas tetangga terdekatnya. Dengan demikian, KNN memungkinkan pengguna untuk mengklasifikasikan data baru berdasarkan kesamaan fitur dengan data pembelajaran yang telah ada, tanpa memerlukan proses pembelajaran yang rumit.

46 Ada banyak teknik yang bisa diterapkan untuk mengklasifikasikan dokumen, seperti *Support Vector Machine*, *Artificial Neural Network*, *Naïve Bayes*, dan metode lainnya yang banyak tersedia. Penelitian yang dilakukan oleh Prakoso dkk pada tahun 2019 dalam jurnalnya yang berjudul “Klasifikasi Berita Menggunakan Algoritma *Naïve Bayes Classifier* Dengan Seleksi Fitur Dan *Boosting*” [10], menyatakan bahwa *Naïve Bayes* tanpa seleksi fitur dan *Boosting* mampu mengklasifikasikan berita dengan nilai akurasi sebesar 72%. Sedangkan untuk *Naïve Bayes* menggunakan *Information Gain* (IG) memperoleh akurasi sebesar 69,5%, *Bayesian Boosting* + NBC memperoleh akurasi sebesar 73,2%, dan IG + *Bayesian Boosting* + NBC memperoleh akurasi sebesar 73,2%.

12 Dari penjelasan tersebut, dapat disimpulkan bahwa tiap metode klasifikasi memiliki keunggulan dan kelemahan masing-masing. Penelitian ini bertujuan untuk membandingkan kinerja antara algoritma *Naïve Bayes* dan CNN yang dioptimasi dengan PSO dalam mengidentifikasi berita hoax politik berbahasa Indonesia (*Text Mining*). Harapannya, hasil penelitian ini dapat memberikan manfaat konkret dalam menentukan algoritma terbaik antara *Naïve Bayes* dan CNN yang dioptimasi dengan PSO untuk mengidentifikasi berita *hoax* politik berbahasa Indonesia (*Text Mining*).

2 Penelitian ini memiliki keunikan tersendiri dengan fokus utama pada komparasi performa antara algoritma *Naïve Bayes* dan CNN Dioptimasi dengan PSO dalam mengidentifikasi berita hoax, memiliki urgensi atau keutamaan yang signifikan, terutama dalam konteks berita politik Indonesia. Penelitian ini menawarkan kontribusi signifikan melalui analisis komparatif antara algoritma *Naïve Bayes* dan CNN yang dioptimasi dengan PSO dalam mengidentifikasi berita *hoax* politik, menyediakan wawasan penting mengenai pendekatan terbaik untuk mendeteksi dan menangani konten berita palsu. Selain itu, analisis perbandingan ini berfungsi sebagai panduan berharga untuk pengembangan metode yang efektif dalam mengatasi berita *hoax* politik secara cepat dan efisien, menekankan pentingnya pemahaman mendalam tentang kelebihan dan kelemahan masing-masing algoritma. Dengan demikian, pemahaman yang mendalam tentang kemampuan relatif dari kedua algoritma ini akan memberikan landasan yang kuat untuk pengambilan keputusan yang cerdas dan strategi yang lebih efektif dalam menangani fenomena berita palsu, terutama dalam konteks politik yang sensitif.

54 8

II. TINJAUAN PUSTAKA

Penelitian ini dilatarbelakangi oleh sejumlah penelitian sebelumnya yang relevan. Penelitian [11] membahas isu pengelolaan layanan pemerintah elektronik, khususnya dalam pengelolaan pengaduan *online* yang saat ini masih bergantung pada penanganan manual oleh sumber daya manusia. Pendekatan manual ini terbukti tidak efektif karena dapat menyulitkan respons cepat terhadap keluhan yang mendesak. Oleh karena itu, untuk meningkatkan efisiensi dan responsivitas dalam mengelola pengaduan, diperlukan sistem otomatis yang dapat mengklasifikasikan pengaduan sesuai dengan organisasi perangkat daerah (OPD) yang bersangkutan. Tujuan dari penelitian ini adalah untuk meningkatkan kemampuan klasifikasi teks pengaduan dengan mengoptimalkan NBC dan k-NN agar dapat mengklasifikasikan lebih dari satu label secara bersamaan. Teknik ini mencapai hasil yang optimal. Hasil dari uji coba dengan validasi silang 10 kali menunjukkan bahwa optimasi NBC dengan PSO berhasil mencapai tingkat akurasi sebesar 87,44%, yang jauh melampaui akurasi k-NN sebesar 75% dan NBC tanpa optimasi sebesar 64,38% [12]. Dengan demikian, penelitian ini menyoroti pentingnya penggunaan teknologi untuk meningkatkan efektivitas layanan pemerintah elektronik dalam menangani pengaduan secara efisien. Data adalah hasil dari informasi yang telah diolah dengan efisien dalam suatu bahasa, baik dalam bentuk tunggal maupun jamak, dan digunakan dalam berbagai proses [13]. Data dapat berbentuk informasi digital yang tersusun dari bit biner atau data mentah. Informasi tersebut disimpan dalam file dan dipergunakan dalam berbagai tahapan seperti penyimpanan, konversi, dan pemrosesan data. Data pada dasarnya adalah sejumlah informasi yang belum memiliki signifikansi atau hubungan yang jelas sebelum melalui proses evaluasi.

Penelitian [14] mengangkat isu serius tentang penyebaran berita palsu yang semakin marak di berbagai negara. Dengan meluasnya penggunaan perangkat seluler dan akses mudah ke internet, masyarakat menghadapi tantangan besar dalam mengonsumsi informasi. Di era media sosial, di mana konten yang diproduksi oleh pengguna tersebar luas, risiko menerima informasi yang salah menjadi semakin tinggi. Oleh karena itu, kredibilitas informasi dan sumbernya menjadi sangat penting untuk menghindari penyebaran berita palsu. Namun, seringkali informasi yang valid tidak diterima sebagai benar oleh individu, terutama jika tidak sejalan dengan pandangan atau opini mereka. Dalam penelitian ini, peneliti bertujuan untuk mengurangi jumlah misinformasi dengan menyangring berita palsu. Mereka menggunakan klasifikasi teks terawasi dengan berbagai model Deep Neural Network (DNN), serta membandingkannya dengan pendekatan konvensional seperti SVM dan Naïve Bayes. Penelitian ini mengevaluasi dan menemukan hasil yang beragam antara berbagai pendekatan terhadap Bahasa Indonesia. Menariknya, penelitian ini adalah yang pertama kali membandingkan hasil dari berbagai model DNN, seperti Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), dan CNN. Meskipun 1D-CNN, yang biasanya lebih unggul dalam pengenalan gambar, mencapai hasil tertinggi sebesar 97,9%, ini menunjukkan kemampuannya yang luar biasa dalam mengklasifikasikan data yang besar. Selain itu, hasil antara LSTM dan GRU cukup sebanding, dengan BI-LSTM mencapai akurasi yang lebih tinggi dibandingkan dengan BI-GRU, yang dapat dijelaskan oleh kemampuan LSTM dalam mengingat nilai pada interval waktu yang berubah-ubah. Namun, untuk pengklasifikasi konvensional seperti SVM dan Naïve Bayes, skor akurasi relatif rendah, dengan Naïve Bayes mencapai akurasi sebesar 86,2%. Hal ini menunjukkan bahwa penggunaan model *Deep Learning* dapat memberikan hasil yang lebih baik dalam menangani masalah klasifikasi berita palsu dibandingkan dengan metode konvensional [14].

Dalam penelitian, Perdebatan mengenai kompleksitas platform berita online yang menyediakan akses cepat, mudah, dan murah bagi pengguna telah mendorong mereka untuk mencari dan meningkatkan pemahaman mereka melalui konsumsi berita yang tersedia. Di platform online seperti media sosial dan blog berita, pengidentifikasian berita palsu seringkali sulit bagi pembaca. Untuk mengatasi masalah ini, penelitian ini mengusulkan penggunaan model CNN dengan margin loss dan berbagai metode embedding untuk mengenali berita palsu. Penelitian juga membandingkan penggunaan penyematan kata statis dan non-statis, yang memungkinkan penyesuaian selama proses pelatihan. Hasilnya menunjukkan peningkatan signifikan dalam kinerja model terbaik, meningkatkan performa sebesar 7,9% pada dataset ISOT dan 2,1% pada dataset uji LIAR. [15]. Hoax adalah kata yang digunakan untuk menyebut berita palsu, fitnah, atau informasi yang tidak benar. [16]. Berita palsu telah menimbulkan kekhawatiran di kalangan masyarakat dan mengganggu persatuan. Solusi untuk mengatasi masalah tersebut melibatkan penguatan pemikiran masyarakat, peningkatan literasi informasi, dan verifikasi serta akreditasi yang baik terhadap media utama dan penyedia berita. Pemerintah juga harus responsif terhadap penyebaran berita palsu, dengan mengambil tindakan untuk menutup situs web yang menjadi sumber penyebaran berita palsu, serta memberlakukan hukum yang berlaku dengan tegas seperti UU ITE.

Berdasarkan hasil tiga penelitian yang disebutkan sebelumnya, penelitian ini mengambil inspirasi dari temuan-temuan tersebut untuk mengarahkan fokus pada perbandingan kinerja antara algoritma Naïve Bayes dan CNN yang dioptimasi dengan PSO. Penelitian ini bertujuan untuk mengeksplorasi dan menentukan kinerja terbaik antara algoritma Naïve Bayes dan CNN yang dioptimasi dengan PSO dalam konteks identifikasi berita *hoax* di ranah politik Indonesia.

III. METODE

Penelitian ini menggunakan pendekatan utama dalam analisis data yang disebut **CRISP-DM (Cross-Industry Standard Process for Data Mining)**. Metodologi ini terdiri dari enam tahapan yang mencakup *Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment*. CRISP-DM digunakan untuk menangani permasalahan yang kompleks dalam analisis data. [17]. Proses metodologi ini terdiri dari 6 tahapan yang dapat dijelaskan sebagai berikut:

1. *Business Understanding* (Pemahaman Bisnis)

Pada tahap *Business Understanding*, melibatkan pemahaman yang mendalam terhadap tujuan analisis data terkait identifikasi berita hoax politik, termasuk pemahaman terhadap lingkungan politik dan sosial yang memengaruhi persepsi publik. Tahap ini juga melibatkan identifikasi kebutuhan analisis data spesifik, serta penetapan ruang lingkup dan batasan analisis untuk memastikan fokus yang tepat dalam pengembangan model identifikasi berita hoax politik. Dengan demikian, tahap ini merupakan landasan kritis untuk merumuskan rencana analisis data yang efektif dan sesuai dengan kebutuhan penelitian. Batasan masalah pada penelitian ini meliputi penggunaan data berupa berita politik hoax yang dapat diproses dan berita berbahasa Indonesia. Data yang diunduh berasal dari Kaggle dengan sumber berita dari CNN, Kompas, Tempo, dan Turnbackhoax. Berita dari CNN tersedia dari 01 Juni 2021 hingga 21 Februari 2023, Kompas dari 20 April 2017 hingga 21 Februari 2023, Tempo dari 01 Januari 2021 hingga 04 Februari 2023, dan TurnBackHoax dari 08 September 2015 hingga 28 Februari 2023. Algoritma yang digunakan untuk perbandingan performa adalah naïve bayes dan CNN yang dioptimalkan dengan PSO. Pengolahan data dilakukan menggunakan aplikasi Rapid Miner. *Data Understanding* (Pemahaman Data)

Pada Tabel 1 dibawah ini menjelaskan tentang dataset yang digunakan dalam penelitian ini, berasal dari platform Kaggle (www.kaggle.com) dengan topik "Indonesian Fact and Hoax Political News". Dataset ini dikembangkan oleh M. R. Rizqullah, dkk. melalui metode pengumpulan data web scraping yang mengambil informasi dari beberapa sumber, termasuk CNN, Kompas, Tempo, dan Turnbackhoax. Informasi yang terdapat dalam dataset berasal dari sumber-sumber tersebut:

TABEL 1
INFORMASI DATA PENELITIAN

Sumber	Tanggal		Total Berita
	Awal	Akhir	
CNN	01 Juni 2021	21 Februari 2023	9.630
Kompas	20 April 2017	21 Februari 2023	4.750
Tempo	01 Januari 2021	04 Februari 2023	6.592
Turnbackhoax	08 September 2015	28 Februari 2023	10.381

2. *Data Preparation* (Persiapan Data)

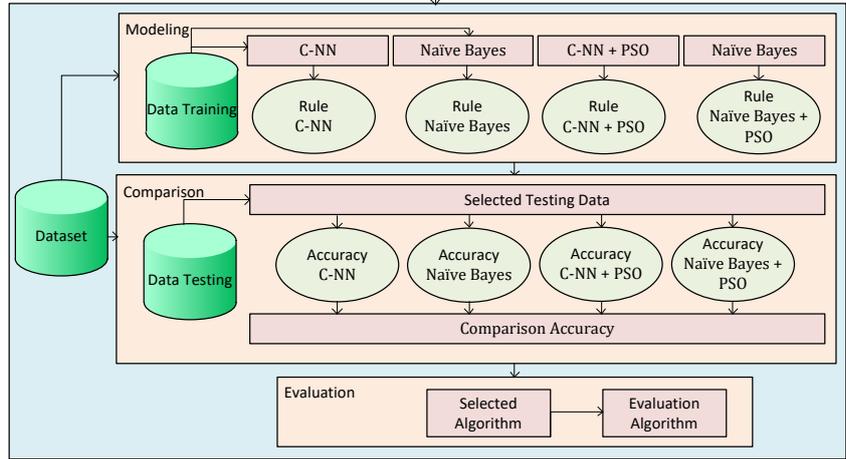
Dalam tahap *Data Preparation*, penelitian ini menggunakan *Text Preprocessing* untuk mengubah data tidak terstruktur menjadi data terstruktur yang sesuai dengan kebutuhan. Langkah-langkahnya mencakup *Tokenization* dan *Case Folding* untuk memisahkan kata-kata dan mengubah huruf menjadi huruf kecil. *Stopword Removal* untuk memilih kata-kata kunci. Dan *Stemming* untuk mengubah kata menjadi bentuk dasarnya. Selain itu, tahap *Modelling* melibatkan pemilihan algoritma dan penyesuaian parameter untuk merancang model analisis yang efektif dalam mengatasi permasalahan identifikasi berita hoax politik. Tahap ini bertujuan untuk mengembangkan model analisis yang akurat dan signifikan dalam konteks penelitian yang dilakukan.

Seperti pada Gambar 1, dataset digunakan untuk membentuk data training dan data testing dalam tahap modeling. Proses pembentukan aturan dilakukan menggunakan metode CNN dan Naïve Bayes Dioptimasi dengan PSO pada data training. Setelahnya, pada tahap komparasi, data testing digunakan untuk menguji performa metode CNN dan Naïve Bayes yang Dioptimasi dengan PSO, membandingkan nilai akurasi guna evaluasi yang lebih mendalam. Pada tahap evaluasi, dilakukan evaluasi berdasarkan metode terbaik yang dipilih dari hasil perbandingan sebelumnya, memberikan gambaran yang komprehensif tentang kinerja dan efektivitas model dalam mengidentifikasi berita hoax politik.

40

Masalah:

- Bagaimana perbandingan kinerja antara algoritma Naïve Bayes dan Convolutional Neural Network (CNN) Dioptimasi dengan Particle Swarm Optimization (PSO) dalam mengidentifikasi berita hoax politik berbahasa Indonesia?
- Apa saja kelebihan dan kelemahan masing-masing algoritma dalam mengenali ciri-ciri berita palsu dalam konteks politik yang sensitif?
- Bagaimana pengaruh kompleksitas bahasa dan struktur konten politik terhadap kemampuan kedua algoritma dalam mengidentifikasi dan memproses berita hoax politik?
- Bagaimana pengembangan metode yang efektif dan efisien dalam mendeteksi dan menangani berita palsu politik berdasarkan pemahaman mendalam mengenai kinerja relatif dari kedua algoritma tersebut?



Tujuan:

- Memberikan pemahaman yang mendalam mengenai perbandingan kinerja antara algoritma dalam mengidentifikasi berita hoax politik berbahasa Indonesia.
- Memberikan pemahaman secara holistik kelebihan dan kelemahan masing-masing algoritma dalam mengenali ciri-ciri berita palsu dalam konteks politik.
- Memberikan pemahaman bagaimana kedua algoritma menangani konten berita politik.
- Memberikan pemahaman mendalam mengenai karakteristik masing-masing algoritma untuk mengembangkan metodologi yang lebih optimal dalam mendeteksi dan menangani berita hoax politik.

GAMBAR 1 PEMODELAN PENELITIAN

3. Evaluation (Evaluasi)

Proses evaluasi menjadi kritis untuk mengukur kinerja dan efektivitas model identifikasi berita hoax politik. Beberapa metrik evaluasi yang relevan untuk digunakan dalam penelitian ini melibatkan perbandingan antara hasil prediksi model dengan keadaan sebenarnya dari data testing. Berikut adalah penjelasan mengenai beberapa metrik evaluasi yang mungkin digunakan:

- 1) Akurasi (Accuracy): Merupakan rasio prediksi yang benar (true positive dan true negative) terhadap total sampel. Akurasi memberikan gambaran keseluruhan seberapa baik model dapat mengklasifikasikan berita sebagai hoax atau bukan.
- 2) Presisi (Precision): Menunjukkan seberapa banyak dari berita yang diklasifikasikan sebagai hoax oleh model yang benar-benar merupakan berita hoax. Presisi berguna untuk mengukur keakuratan model dalam mengidentifikasi berita yang sebenarnya palsu.
- 3) Recall (Sensitivitas): Menunjukkan seberapa banyak berita hoax yang sebenarnya berhasil diidentifikasi oleh model. Recall berguna untuk mengevaluasi sejauh mana model dapat menangkap seluruh berita hoax yang ada.

Proses evaluasi masing-masing algoritma dalam penelitian ini dijalankan melalui serangkaian tahapan sebagai berikut:

- 1) Pengolahan Data dengan Naïve Bayes

TABEL 2
 HASIL PENGOLAHAN DATA NAÏVE BAYES

	Algoritma Naïve Bayes
Accuracy	90.71%
Precision	91.97%
Recall	77.06%

29

36

28

31

4

4

39

Hasil evaluasi pada Tabel 2 dalam pengolahan data Naïve Bayes menunjukkan bahwa model ini memiliki akurasi sebesar 90,71%. Nilai akurasi yang tinggi ini menandakan bahwa model memiliki kemampuan yang baik dalam mengklasifikasikan data ke dalam kategori "valid" dan "hoax". Berdasarkan hasil confusion matrix, hasil tersebut menunjukkan bahwa model klasifikasi memiliki tingkat presisi sebesar 91,97%. Dengan menggunakan confusion matrix yang diberikan, nilai recall untuk kelas positif "hoax" dihitung sebesar 77,06%. Hasil ini mengindikasikan kemampuan model dalam mengidentifikasi sejumlah 77,06% dari total kasus yang sebenarnya termasuk dalam kategori "hoax".

2) Pengolahan Data dengan Naïve Bayes + PSO

TABEL 3
HASIL PENGOLAHAN DATA NAÏVE BAYES + PSO

	Algoritma Naïve Bayes + PSO
Accuracy	90.25%
Precision	90.99%
Recall	76.38%

Hasil evaluasi pada Tabel 3 dalam pengolahan data Naïve Bayes + PSO menampilkan bahwa model tersebut mencapai akurasi sekitar 90,25%, sebuah indikasi kuat akan efisiensi model dalam membedakan antara data yang "valid" dan yang tergolong sebagai "hoax". Dari analisis confusion matrix, didapati bahwa tingkat kepresisian model berada pada angka 90,99%. Berdasarkan confusion matrix tersebut, ditemukan bahwa proporsi recall untuk kategori positif "hoax" adalah sekitar 76,38%. Hal ini menunjukkan bahwa model berhasil mengenali sekitar 76,38% dari keseluruhan contoh yang sejatinya merupakan "hoax".

3) Pengolahan Data dengan Convolutional Neural Network

TABEL 4
HASIL PENGOLAHAN DATA CONVOLUTIONAL NEURAL NETWORK

	Algoritma C-NN
Accuracy	80.86%
Precision	80.72%
Recall	79.53%

Hasil evaluasi pada Tabel 4 dalam pengolahan data CNN mengungkap bahwa model ini berhasil mencatat akurasi sebesar 80,86%, menunjukkan kemampuannya yang signifikan dalam memisahkan data yang "valid" dari yang dikategorikan sebagai "hoax". Melalui analisis pada confusion matrix, kita menemukan bahwa model ini memiliki presisi sebesar 80,72%. Lebih lanjut, dari confusion matrix dapat dilihat bahwa tingkat recall untuk kelas positif "hoax" mencapai 79,53%. Ini menandakan bahwa model ini efektif dalam mengidentifikasi sekitar 79,53% dari total kasus yang benar-benar adalah "hoax".

4) Pengolahan Data dengan Convolutional Neural Network + PSO

TABEL 5
HASIL PENGOLAHAN DATA CONVOLUTIONAL NEURAL NETWORK + PSO

	Algoritma C-NN + PSO
Accuracy	79.68%
Precision	78.94%
Recall	75.25%

Temuan evaluasi pada Tabel 5 dalam pengolahan data CNN+PSO menunjukkan bahwa model berhasil meraih akurasi sekitar 79,68%, yang mencerminkan kapasitasnya yang luar biasa dalam membedakan antara informasi "valid" dan "hoax". Dari pemeriksaan confusion matrix, tercatat presisi model sebesar 78,94%. Selanjutnya, berdasarkan data dari confusion matrix, proporsi recall untuk kategori "hoax" terhitung sebesar 75,25%. Hal ini menegaskan efektivitas model dalam mengenali kira-kira 75,25% dari keseluruhan kasus yang nyatanya termasuk dalam kategori "hoax".

4. Deployment (Penerapan)

Dalam fase pelaksanaan penelitian, tahap penerapan memegang peranan penting dalam menguji metode yang telah dipilih dan menguji keabsahan hasilnya untuk memastikan bahwa kinerja metode tersebut sesuai dengan harapan. Melalui proses pengujian dan validasi ini, penelitian berupaya untuk memastikan bahwa metode yang digunakan memiliki kemampuan yang optimal dalam menghasilkan identifikasi berita hoax dengan tingkat akurasi yang tinggi. Pendekatan ini melibatkan serangkaian uji coba dan evaluasi yang cermat terhadap metode yang diterapkan, termasuk penggunaan dataset uji yang sesuai, pengaturan parameter yang tepat, serta analisis mendalam terhadap hasil identifikasi yang dihasilkan. Dengan demikian, tahap penerapan ini bukan hanya merupakan langkah penting dalam proses penelitian, tetapi juga menjadi titik fokus untuk memastikan bahwa penelitian ini menghasilkan hasil yang berkualitas dan dapat diandalkan dalam mendeteksi dan menangani berita hoax secara efektif.

IV. HASIL

Dalam penelitian ini, akan diterapkan algoritma PSO sebagai metode optimasi untuk memilih atribut yang paling relevan dalam proses pembobotan atribut. Setelah atribut-atribut tersebut terbobot, perbandingan akan dilakukan antara algoritma Naïve Bayes dan CNN untuk menentukan algoritma yang paling efektif dalam mengidentifikasi berita hoax politik berbahasa Indonesia. Dengan demikian, penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja kedua algoritma tersebut dalam konteks pengidentifikasian berita hoax politik, dengan harapan menemukan pendekatan yang paling optimal untuk tujuan tersebut.

TABEL 6
PERBANDINGAN ALGORITMA NB, NB+PSO, CNN, CNN+PSO

	Algoritma NB	Algoritma NB + PSO	Algoritma CNN	Algoritma CNN + PSO
Accuracy	90.71%	90.25%	80.86%	79.68%
Precision	91.97%	90.99%	80.72%	78.94%
Recall	77.06%	76.38%	79.53%	75.25%
AUC	0.827	0.860	0.837	0.842

Tabel 6 yang disajikan memberikan perbandingan yang sangat informatif tentang performa empat algoritma yang digunakan dalam pengklasifikasian berita hoax dan tidak hoax, yaitu Algoritma NB, Algoritma NB+PSO, Algoritma CNN, dan Algoritma CNN+PSO. Mari kita lihat dengan lebih detail hasil dari masing-masing algoritma berdasarkan tiga metrik evaluasi utama, yaitu *accuracy*, *precision*, dan *recall*.

Pertama-tama, dari segi akurasi, Algoritma NB menunjukkan hasil yang sangat baik dengan tingkat akurasi sebesar 90.71%, yang merupakan yang tertinggi di antara keempat algoritma. Disusul oleh Algoritma NB+PSO dengan akurasi 90.25%, menunjukkan bahwa penambahan Particle Swarm Optimization memberikan kontribusi yang signifikan namun tidak cukup untuk mengungguli Algoritma NB secara signifikan dalam hal akurasi. Di sisi lain, Algoritma CNN menunjukkan akurasi sebesar 80.86%, sedangkan Algoritma CNN+PSO memiliki akurasi yang sedikit lebih rendah, yaitu 79.68%. Hal ini menandakan bahwa pendekatan berbasis probabilitas seperti yang digunakan dalam Algoritma NB lebih efektif dalam hal memprediksi kebenaran berita daripada pendekatan berbasis neural network seperti CNN.

Selanjutnya, jika kita melihat dari sisi *precision*, yang mengukur seberapa banyak dari prediksi positif yang sebenarnya benar-benar positif, Algoritma NB juga menunjukkan hasil tertinggi dengan nilai *precision* sebesar 91.97%. Diikuti oleh Algoritma NB+PSO dengan *precision* sebesar 90.99%, menunjukkan bahwa penambahan Particle Swarm Optimization sedikit meningkatkan kemampuan algoritma dalam menghasilkan prediksi yang benar. Sementara itu, Algoritma CNN memiliki *precision* sebesar 80.72%, sedangkan Algoritma CNN+PSO memiliki *precision* yang sedikit lebih rendah, yaitu 78.94%. Meskipun demikian, hasil *precision* dari Algoritma NB dan NB+PSO masih jauh lebih tinggi dibandingkan dengan Algoritma CNN dan CNN+PSO, menegaskan bahwa pendekatan probabilitas lebih efektif dalam hal menghasilkan prediksi yang akurat.

Terakhir, dari segi *recall*, yang mengukur seberapa banyak dari data positif yang berhasil diprediksi dengan benar oleh model, Algoritma NB tetap memimpin dengan nilai *recall* sebesar 77.06%. Algoritma CNN menunjukkan hasil yang cukup baik dalam hal *recall* dengan nilai 79.53%, namun masih kalah dari Algoritma NB. Sementara Algoritma NB+PSO dan CNN+PSO memiliki nilai *recall* yang lebih rendah, yaitu 76.38% dan 75.25% secara berturut-turut.

Penelitian ini mengungkapkan bahwa Naïve Bayes menunjukkan kinerja yang lebih baik daripada CNN standar, CNN yang dioptimasi dengan PSO, dan Naïve Bayes yang dioptimasi dengan PSO dalam mengenali berita hoax politik berbahasa Indonesia.

V. PEMBAHASAN

Proses evaluasi dan validasi dalam penelitian ini akan melibatkan perhitungan manual yang kemudian akan dibandingkan dengan hasil yang diolah menggunakan aplikasi RapidMiner 10.2. Metode ini memungkinkan kami untuk mengukur kinerja algoritma secara langsung dengan melakukan perhitungan manual, serta memvalidasi hasil tersebut dengan menggunakan alat analisis data yang canggih seperti RapidMiner. Dengan pendekatan ini, kami dapat memastikan keakuratan dan kehandalan algoritma dalam mengidentifikasi berita hoax politik berbahasa Indonesia secara efektif.

Dataset yang digunakan pada penelitian ini adalah data berita hoax dan tidak hoax berbahasa Indonesia, setelah itu dilakukan olah data secara manual dengan bantuan Microsoft Office Excel.

Proses untuk melakukan perhitungan manual Naïve Bayes pada berita hoax menggunakan 15 data sampel dari dataset yang digunakan. Pada Tabel 7 berikut merupakan 15 data sampel yang diambil dari dataset untuk melakukan perhitungan manual.

Proses klasifikasi data latih

TABEL 7
DATA YANG DIGUNAKAN UNTUK CONTOH PERHITUNGAN MANUAL

No	Berita	Label
1	Hanya memberitahukan. Ada video yang disebar di WhatsApp yang menunjukkan grafik Covid19 yang sudah mendarat di Argentina. Video itu berjudul Argentina is doing it, jangan dibuka maupun ditonton, video itu akan meretas ponsel dalam 10 detik dan tidak akan bisa dihentikan. Sebarkan informasi ini ke keluarga dan teman-teman anda. Sekarang kabar ini juga telah diberitakan oleh CNN	Hoax
2	Hati-hati semua. Sekarang banyak kasus penculikan anak. Dapat video dari grup whatsapp. Mereka sayat anak itu hidup2 untuk diambil organ dalamnya. Banyak kasus yang dibawa ke thailand. Jika kalian ingin berlibur ke thailand harap hati hati.	Hoax
3	Christian Eriksen, seorang pesepakbola yang sehat dan bermain untuk salah satu klub sepakbola terbaik di dunia dengan tim medis terbaik, terjatuh di lapangan hari ini karena serangan jantung. Syukurlah mereka dapat menyelamatkan nyawanya. Ia baru saja disuntik vaksin Pfizer beberapa hari yang lalu Kepala Tim Medis dan Dokter Spesialis Jantung Inter Milan membenarkan bahwa ia baru saja disuntik vaksin Pfizer 12 hari yang lalu. Dokter itu bicara satu jam yang lalu di Radio Sportiva Italia.	Hoax
4	Jika kamu memotong semangka dan melihat retakan pada permukaannya, lebih baik buang saja. Retakan ini mungkin disebabkan oleh bahan kimia yang digunakan petani untuk membuat semangka tumbuh lebih besar. Ketika tertelan, bahan kimia ini dapat menyebabkan masalah kesehatan yang serius termasuk keracunan.	Hoax
5	Untuk sesama ibu menyusui keluarga kami terpapar covid sekarang. Kantung di sebelah kiri berisi ASI yang dipompa minggu lalu ketika kami masih sehat, dan kantung di sebelah kanan berisi ASI yang dipompa hari ini (saat sedang sangat sakit lol). Penelitian menunjukkan bahwa antibodi covid disalurkan kepada bayi melalui ASI dan aku yakin itulah yang menyebabkan perbedaan warna pada dua kantung ini (tidak ada perubahan konsumsi makanan dan obat, dll.) Aku tidak pernah melihat ASI berwarna hijau selama 10,5 bulan memompa ASI Hanya pengingat rutin bahwa tubuh kita menakjubkan	Hoax
6	Inilah yang terjadi ketika seseorang menerima terlalu banyak vaksin Covid-19 dan vaksin booster. Mereka bertindak seperti zombie, kehilangan kemampuan fisik dan mental. Untuk menemukan kebenaran, pengetahuan, bimbingan & kebijaksanaan bergabunglah dengan grup telegram kami.	Hoax
7	*HATI2 MAKAN NASI PADANG* Nasi Padang sumber penularan virus Covid-19 . Coba bayangi tiap meja tamu2 mkn dak habis sisa2 dikembalikan lagi dan di sajikan lag ke tamu berikutnya ! pun demikian yg bungkus juga bekas2 air liur tamu2 yg mungkin ada yg virus corona ? biasa selesai mkn tamu2 ngobrol2 dulu dan hidangan di meja blm diangkat ? hujan rintik2 lah di hidangan tsb . paling rentang penularan virus corona.	Hoax
8	Ada email yang beredar dan menyertakan sebuah foto yang menunjukkan bahwa Covid-19 telah berhasil ditangani di Meksiko dan foto tersebut berjudul Mexico did it , jangan buka foto itu karena foto tersebut akan meretas ponsel dalam 5 detik dan tidak akan dapat dihentikan. Itu adalah sebuah virus.	Hoax
9	Airlangga Nilai Tahun Politik Jadi Vitamin untuk Pemulihan Ekonomi 2023Menko Bidang Perekonomian Airlangga Hartarto menilai tahun politik menjadi vitamin, untuk menjaga momentum pemulihan ekonomi pada 2023. Dengan tahun politik, justru ekonomi Indonesia akan terdorong berkat belanja politik, dan Airlangga yakin belanja politik itu akan menggerakkan perekonomian di kuartal 4 2023. Airlangga menilai kestabilan politik di Indonesia saat ini tidak terlepas dari kepemimpinan Presiden Joko Widodo. Hal itu dilihat dari keberhasilan menangani pandemi Covid-19, dan penanganan itu dibutuhkan keterlibatan dari pihak eksekutif dan legislatif. Ia menyebut negara-negara yang kondisi politiknya tidak stabil, justru menyebabkan pemerintahannya kesulitan dalam menangani pandemi itu.	Valid
10	Jokowi Bicara Soal Vaksinasi Hingga Pembiayaan Infrastruktur Kesehatan di AS 19 yang terjadi di sejumlah negara di dunia saat ini. Dia menyatakan momentum ini harus dimanfaatkan untuk meluncurkan pukulan terakhir demi menuntaskan pandemi yang telah berlangsung lebih dari dua tahun terakhir.	Valid
11	GMNI Turut Suarakan Tolak Penundaan Pemilu dan Kenaikan Harga BBM refocusing untuk menghadapi pandemi Covid. Sehingga GMNI khawatir postur APBN hanya akan menjadi bancakan bagi segelintir elit untuk menambah modal kontestasi pemilu. Adapun terkait dengan kenaikan bahan bakar, GMNI melihat ada 'skema licik' karena pemerintah menaikkan harga Pertamina di satu sisi, namun membatasi ketersediaan Pertalite di sisi lain. "Masyarakat dipaksa secara bertahap untuk pindah menggunakan Pertamina," katanya. Baca Juga: Demo 11 April: BEM SI Tetap Tuntut Tolak Jokowi 3 Periode	Valid
12	Golkar Klaim Kinerja Airlangga Jadi Alasan Ridwan Kamil Bergabung Ketua DPD Partai Golkar Jawa Barat, Ace Hasan Syadzily mengungkapkan proses Gubernur Jawa Barat Ridwan Kamil akhirnya menjadi kader Golkar. Menurutnya, hal itu tak lepas dari kinerja apik Ketua Umum Partai Golkar Airlangga Hartarto dalam pemulihan ekonomi saat pandemi Covid-19. Selain menjabat sebagai Menteri Koordinator Bidang Perekonomian, Airlangga juga diamanatkan Presiden Joko Widodo (Jokowi) sebagai ketua tim penanganan Covid-19, dan pemulihan ekonomi nasional (PEN).	Valid
13	IMI Menolak Ikutan Pusaran Formula E 19 tahun 2022, Formula E dipastikan akan digelar sesuai target. Optimisme itu juga didorong dengan program vaksinasi Covid-19 di Jakarta yang terus berjalan dengan baik. "Apalagi vaksin kan dalam satu bulan ke depan sudah selesai di DKI Jakarta. Mudah-mudahan ini meyakinkan kita semua bahwa kegiatan-kegiatan ke depan dapat dilaksanakan dengan lebih baik lagi," kata Riza.	Valid

14	Dukungan Petisi Tolak IKN Terus Bertambah, Inisiator: Tandanya Cacat Aspirasi 19 yang membutuhkan dana besar dari APBN dan Program Pemulihan Ekonomi Nasiobal (PEN)."Publik merasakan penderitaan yang luar biasa dari pandemi dan kesulitan ekonomi namun pilihan pemerintah malah menghamburkan uang dan bukan menangani kesehatan publik malah justru memprioritaskan proyek yang syarat kepentingan elit oligarki," tutur Achmad.	Valid
15	Hasto: PDIP Kokoh dan Tak Ada Ruang Tunda Pemilu 2024 71 persen warga yang menolak perpanjangan masa jabatan presiden baik karena alasan pandemi, pemulihan ekonomi akibat pandemi, atau pembangunan Ibu Kota Negara (IKN) baru."Mayoritas warga juga lebih setuju bahwa pergantian kepemimpinan nasional melalui Pemilu2024 harus tetap diselenggarakan meski masih dalam kondisi pandemi (64 persen), ketimbang harus ditunda karena alasan pemulihan ekonomi nasional akibat pandemi (26,9 persen). Jadi wacana ini sebaiknya diakhiri dan kembali ke jadwal pemilu yang sudah ditetapkan," ujar dia.Baca juga: Peneliti LSI Sarankan Elite Politik Sudah Keinginan Tunda Pemilu 2024	Valid

Perhitungan Naïve Bayes adalah nilai probabilitas. Setelah melalui Text Preprocessing, dilakukan pembobotan kata dengan menggunakan Raw Term Frequency Weighting. Selanjutnya Proses klasifikasi data latih sebagai berikut:

1. Menghitung Probabilitas prior dari setiap kelas, yang terbagi menjadi 2 kelas yaitu Hoax dan Valid. Pada penelitian ini menggunakan 15 dokumen training dan 2 kelas.

$$P(c) = \frac{Nc}{N} \tag{1}$$

$$P(Hoax) = \frac{8}{15} = 0,5333$$

$$P(Valid) = \frac{7}{15} = 0,4667$$

2. Menghitung conditional probability dengan Mutinomial Model masing-masing fitur pada masing-masing kategori. Misalnya jumlah seluruh kata 174, Hoax = 131, Valid = 180, dari semua kelas yang ada. Berikut merupakan contoh perhitungan likelihood dari term nomor 1 (video):

$$P(w|c) = \frac{\text{count}(w,c) + 1}{\text{count } c + |V|} \tag{2}$$

$$P(\text{video}|Hoax) = \frac{4 + 1}{131 + 174} = 0,016393$$

$$P(\text{video}|Valid) = \frac{0 + 1}{180 + 174} = 0,002825$$

Dihitung hingga semua term mendapatkan nilainya masing-masing.

TABEL 8
 20 DATA CONTOH SKOR PELUANG (CONDITIONAL PROBABILITY/LIKELIHOOD)

Term	P(HOAX)	P(VALID)
video	0,016393	0,002825
sebar	0,009836	0,002825
whatsapp	0,009836	0,002825
tunjuk	0,009836	0,002825
grafik	0,006557	0,002825
covid	0,022951	0,016949
datar	0,006557	0,002825
argentina	0,006557	0,002825
judul	0,009836	0,002825
buka	0,009836	0,002825
tonton	0,006557	0,002825
retas	0,009836	0,002825
ponsel	0,009836	0,002825
henti	0,009836	0,002825
informasi	0,006557	0,002825
keluarga	0,009836	0,002825
teman	0,006557	0,002825
berita	0,006557	0,002825
kasus	0,009836	0,002825
culik	0,006557	0,002825

51

Proses klasifikasi data uji

Proses klasifikasi data uji dengan mencari nilai posterior dengan cara mengalikan nilai conditional probability dari setiap kategori (kelas). Hasil dari conditional probability dengan skor yang terbesar akan menjadi kelas baru dari data tersebut.

$$P(c|d) = P(c) * \prod_{w \in d} P(w|c) \tag{3}$$

$$P(d|c) = P(w1, w2, w3, w4|c) = P(w1|c) * P(w2|c) * P(w3|c) * P(w4|c) \tag{4}$$

Berikut contoh perhitungan untuk mencari nilai posterior pada dokumen 1 dengan menggunakan rumus 4:
 $P(uji|Hoax) = P(Hoax) * P(video|Hoax) * P(sebar|Hoax) * P(whatsapp|Hoax) * P(tunjuk|Hoax) * P(grafik|Hoax)$

$$\begin{aligned}
 & * P(covid|Hoax) * P(datar|Hoax) * P(argentina|Hoax) * P(judul|Hoax) * P(buka|Hoax) \\
 & * P(tonton|Hoax) * P(retas|Hoax) * P(ponsel|Hoax) * P(henti|Hoax) * P(informasi|Hoax) \\
 & * P(keluarga|Hoax) * P(teman|Hoax) * P(berita|Hoax) \\
 & = 0,5333 * 0,016393 * 0,009836 * 0,009836 * 0,009836 * 0,006557 * 0,022951 * 0,006557 * 0,006557 \\
 & * 0,009836 * 0,009836 * 0,006557 * 0,009836 * 0,009836 * 0,009836 * 0,006557 * 0,009836 * 0,006557 \\
 & * 0,006557 \\
 & = 9,01511E-38
 \end{aligned}$$

$$\begin{aligned}
 P(uji|Valid) &= P(Valid) * P(video|Valid) * P(sebar|Valid) * P(whatsapp|Valid) * P(tunjuk|Valid) * P(grafik|Valid) \\
 & * P(covid|Valid) * P(datar|Valid) * P(argentina|Valid) * P(judul|Valid) * P(buka|Valid) \\
 & * P(tonton|Valid) * P(retas|Valid) * P(ponsel|Valid) * P(henti|Valid) * P(informasi|Valid) \\
 & * P(keluarga|Valid) * P(teman|Valid) * P(berita|Valid) \\
 & = 0,4667 * 0,002825 * 0,002825 * 0,002825 * 0,002825 * 0,002825 * 0,016949 * 0,002825 * 0,002825 \\
 & * 0,002825 * 0,002825 * 0,002825 * 0,002825 * 0,002825 * 0,002825 * 0,002825 * 0,002825 \\
 & * 0,002825 \\
 & = 3,67366E-46
 \end{aligned}$$

TABEL 9
 PERHITUNGAN HASIL PREDIKSI DATA UJI

Data Uji			
Dokumen	Peluang	Skor Probabilitas	Kelas Baru
Dokumen 1	P(Hoax)	9,01511E-38	Hoax
	P(Valid)	3,67366E-46	
Dokumen 2	P(Hoax)	1,51313E-19	Hoax
	P(Valid)	5,34537E-24	
Dokumen 3	P(Hoax)	4,26644E-26	Hoax
	P(Valid)	1,44594E-30	
Dokumen 4	P(Hoax)	4,035E-20	Hoax
	P(Valid)	1,60361E-23	
Dokumen 5	P(Hoax)	5,30935E-26	Hoax
	P(Valid)	7,22968E-31	
Dokumen 6	P(Hoax)	8,18051E-35	Hoax
	P(Valid)	3,68295E-40	
Dokumen 7	P(Hoax)	1,5789E-31	Hoax
	P(Valid)	1,62971E-38	
Dokumen 8	P(Hoax)	5,5079E-27	Hoax
	P(Valid)	4,08456E-33	
Dokumen 9	P(Hoax)	1,4313E-44	Valid
	P(Valid)	4,56331E-36	
Dokumen 10	P(Hoax)	4,67727E-42	Valid
	P(Valid)	6,3921E-37	
Dokumen 11	P(Hoax)	1,44245E-49	Valid
	P(Valid)	2,0749E-42	

Dokumen 12	Peluang	Skor Probabilitas	Kelas Baru
	P(Hoax)	5,42289E-57	4,9251E-48
P(Valid)			
Dokumen 13	Peluang	Skor Probabilitas	Kelas Baru
	P(Hoax)	1,01524E-36	Valid
	P(Valid)	1,08139E-32	
Dokumen 14	Peluang	Skor Probabilitas	Kelas Baru
	P(Hoax)	1,88549E-47	Valid
	P(Valid)	7,67304E-40	
Dokumen 15	Peluang	Skor Probabilitas	Kelas Baru
	P(Hoax)	3,11818E-43	Valid
	P(Valid)	7,8527E-34	

Keterangan: Pada dokumen uji, skor atau nilai peluang terbesar ada pada kelas P(uji|Hoax) yang memiliki nilai sebesar 9,01511E-38. Sehingga berita tersebut masuk pada prediksi kelas "Hoax".

Pada Tabel 9 diatas dapat dilihat bahwa nilai terbesar yang akan menjadi kelas baru atau hasil prediksinya. Hasil prediksi menyatakan benar dengan data yang belum diolah (Tabel 7).

Hasil Pengujian, dilakukan dengan menguji model dari total 10.000 data dengan 80% sebagai data training dan 20% sebagai data testing. Dilakukan dengan confusion matriks untuk mengukur hasil yang dari percobaan. Untuk mengetahui nilai *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). Tabel 10 dibawah ini merupakan tabel *confusion matrix*.

TABEL 10
 PERHITUNGAN CONFUSION MATRIX DATA UJI

	Pred. Valid	Pred. Hoax
True Valid	5329	169
True Hoax	574	1928

Hasil dari *Confusion Matrix* merupakan gambaran dari hasil evaluasi model klasifikasi Naïve Bayes. Pada kasus ini terdapat dua kelas, yaitu Valid dan Hoax. Total jumlah sampel yang dievaluasi adalah sebanyak 8.000. Dalam kasus ini, True Positive = 5329, True Negative = 1928, False Positive = 574, False Negative = 169. Dari sini dapat diketahui nilai Accuracy, Precision, dan Recall.

5) Accuracy

$$\begin{aligned}
 Accuracy &= \frac{(TP + TN)}{(TP + FP + TN + FN)} & (5) \\
 &= \frac{(5329 + 1928)}{(5329 + 1928 + 574 + 169)} \\
 &= \frac{7257}{8000} \\
 &= 0,9071 \\
 &= 90,71\%
 \end{aligned}$$

6) Precision

$$\begin{aligned}
 Precision &= \frac{TP}{(TP + FP)} & (6) \\
 &= \frac{5329}{(5329 + 574)} \\
 &= 0,9022 \\
 &= 90,22\%
 \end{aligned}$$

7) Recall

$$\begin{aligned}
 Recall &= \frac{TP}{(TP + FN)} & (7) \\
 &= \frac{5329}{(5329 + 169)} \\
 &= 0,9697 \\
 &= 96,97\%
 \end{aligned}$$

TABEL 11
PERBANDINGAN EVALUASI DATA LATIH DAN DATA UJI

Metode Evaluasi	Accuracy	Precision	Recall
Manual (Data Uji)	90.71%	90.22%	96.97%
10-fold Cross Validation (Data Latih)	90.71%	91.97%	97.06%

Perbedaan hasil antara perhitungan manual dan metode 10-fold Cross Validation pada Tabel 11, meskipun ada, tidak begitu signifikan sehingga masih dapat diterima. Perbedaan ini muncul karena kedua pendekatan memiliki tujuan dan proses evaluasi yang berbeda. Dalam perhitungan manual, evaluasi dilakukan pada satu set data yang tidak berubah, sementara dalam metode 10-fold Cross Validation, pengujian dilakukan secara iteratif dengan membagi dataset ke dalam K (jumlah lipatan) yang ditentukan sebelumnya, diikuti dengan penghitungan nilai rata-rata dari semua iterasi tersebut. Dengan demikian, perbedaan ini mencerminkan penggunaan pendekatan yang berbeda dalam mengevaluasi kinerja model, namun keduanya tetap memberikan gambaran yang berguna tentang keandalan model yang dikembangkan.

VI. KESIMPULAN

Kesimpulan yang dihasilkan dari penelitian ini memberikan pemahaman yang mendalam tentang efektivitas algoritma Naïve Bayes dalam mengidentifikasi berita hoax politik berbahasa Indonesia dibandingkan dengan Convolutional Neural Network (CNN) dalam berbagai skenario, termasuk penggunaan CNN standar, CNN yang dioptimasi dengan Particle Swarm Optimization (PSO), serta Naïve Bayes yang dioptimasi dengan PSO. Hasil penelitian menunjukkan bahwa meskipun CNN telah menjadi salah satu pendekatan yang populer dalam pengolahan bahasa alami dan analisis teks, algoritma Naïve Bayes ternyata unggul dalam menangani permasalahan spesifik ini.

Pentingnya penelitian ini tidak hanya terletak pada penemuan bahwa Naïve Bayes memiliki kinerja yang lebih baik dalam konteks identifikasi berita hoax politik berbahasa Indonesia, tetapi juga dalam menyediakan landasan untuk pengembangan dan implementasi sistem yang lebih efektif dalam memerangi penyebaran berita hoax. Dengan memahami bahwa Naïve Bayes mungkin merupakan pilihan yang lebih tepat dalam skenario ini, para pengembang sistem dan pengambil keputusan memiliki panduan yang jelas untuk merancang solusi yang lebih akurat dan efisien.

Dengan demikian, kesimpulan ini menegaskan bahwa penggunaan algoritma Naïve Bayes dapat memberikan kontribusi yang signifikan dalam upaya melawan penyebaran berita hoax politik berbahasa Indonesia. Namun, penting juga untuk mempertimbangkan bahwa penelitian ini hanya merupakan langkah awal dalam upaya ini, dan penelitian lebih lanjut diperlukan untuk memperdalam pemahaman kita tentang fenomena ini dan mengembangkan solusi yang lebih komprehensif dan efektif.

Walaupun algoritma Naïve Bayes telah terbukti memiliki tingkat akurasi yang memadai, ada beberapa aspek yang bisa diperhatikan dalam penelitian selanjutnya untuk meningkatkan akurasi dan kinerja secara keseluruhan. Salah satu saran adalah melakukan eksplorasi terhadap algoritma lain seperti Decision Trees, Random Forest, atau Support Vector Machines selain Naïve Bayes dan Convolutional Neural Network. Membandingkan kinerja berbagai algoritma akan memberikan wawasan yang lebih komprehensif tentang pendekatan terbaik untuk masalah identifikasi berita hoax. Selain itu, dalam pengembangan lebih lanjut, penelitian dapat mempertimbangkan validasi model yang lebih komprehensif seperti k-fold stratified cross validation untuk memastikan keandalan dan generalisasi model. Terakhir, menggunakan teknik ensemble learning seperti penggabungan Naïve Bayes dengan model-model lain atau menggunakan metode voting dapat menjadi pendekatan yang menarik untuk meningkatkan akurasi dan kestabilan prediksi pada penelitian lanjutan.

REFERENCES

- [1] D. Bayu, "Apjii: pengguna internet indonesia tembus 210 juta pada 2022," *Data Indonesia*. 2022. [Online]. Available: <https://dataindonesia.id/digital/detail/apjii-pengguna-internet-indonesia-tembus-210-juta-pada-2022>
- [2] F. C. Rosana and A. Y. Widyastuti, "Hingga Awal 2022, Kominfo Temukan 9.546 Hoaks di Internet." *tempo.co*, 2022. [Online]. Available: <https://bisnis.tempo.co/read/1558213/hingga-awal-2022-kominfo-temukan-9-546-hoaks-di-internet>
- [3] K. H. Siswoko, "Kebijakan Pemerintah Menangkal Penyebaran Berita Palsu atau 'Hoax,'" *J. Muara Ilmu Sos. Humaniora, dan Seni*, vol. 6, no. 1, pp. 13–19, 2017.
- [4] A. Wijaya, C. Rozikin, and B. N. Sari, "Penerapan Text Mining Untuk Klasifikasi Judul Berita Hoax Vaksinasi COVID-19 Menggunakan Algoritma Support Vector Machine," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 16, pp. 11–20, 2022, [Online]. Available: <https://doi.org/10.5281/zenodo.7058890>
- [5] T. Silwattananusarn and P. Kulkanjanapiban, "A text mining and topic modeling based bibliometric exploration of information science research," *IAES Int. J. Artif. Intell.*, vol. 11, no. 3, pp. 1057–1065, 2022, doi: 10.11591/ijai.v11.i3.pp1057-1065.

- [6] P. Utama, "Identifikasi Hoax pada Media Sosial dengan Pendekatan Machine Learning," *Widya Duta J. Ilm. Ilmu Agama dan Ilmu Sos. Budaya*, vol. 13, p. 69, 2018, doi: 10.25078/wd.v13i1.436.
- [7] Archit Verma, "Evaluation of Classification Algorithms with Solutions to Class Imbalance Problem on Bank Marketing Dataset using WEKA," *Int. Res. J. Eng. Technol.*, vol. 06, no. 03, 2019.
- [8] Y. D. Pramudita, S. S. Putro, and N. Makhmud, "Klasifikasi Berita Olahraga Menggunakan Metode Naïve Bayes dengan Enhanced Confix Stripping Stemmer," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 3, p. 269, 2018, doi: 10.25126/jtiik.201853810.
- [9] A. J. T. D. Yanosma, and K. Anggriani, "IMPLEMENTASI METODE K-NEAREST NEIGHBOR (KNN) DAN SIMPLE ADDITIVE WEIGHTING (SAW) DALAM PENGAMBILAN KEPUTUSAN SELEKSI PENERIMAAN ANGGOTA PASKIBRAKA," *Pseudocode*, vol. 3, no. 2, pp. 98–112, Jan. 2016, doi: 10.33369/PSEUDOCODE.3.2.98-112.
- [10] B. S. Prakoso, D. Rosiyadi, H. S. Utama, and D. Aridarma, "Klasifikasi Berita Menggunakan Algoritma Naive Bayes Classifier Dengan Seleksi Fitur Dan Boosting," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 2, pp. 227–232, 2019, doi: 10.29207/resti.v3i2.1042.
- [11] K. S. Nugroho, I. Istiadi, and F. Marisa, "Naive Bayes classifier optimization for text classification on e-government using particle swarm optimization," *J. Teknol. dan Sist. Komput.*, vol. 8, no. 1, pp. 21–26, 2020, doi: 10.14710/jtsiskom.8.1.2020.21-26.
- [12] K. S. Nugroho, I. Istiadi, and F. Marisa, "Optimasi naive Bayes classifier untuk klasifikasi teks pada e-government menggunakan particle swarm optimization," *J. Teknol. dan Sist. Komput.*, vol. 8, no. 1, pp. 21–26, Jan. 2020, doi: 10.14710/JTSISKOM.8.1.2020.21-26.
- [13] J. dalleh, Akrim, and Baharuddin, *Pengantar Teknologi Farmasi*. 2020. [Online]. Available: http://eprints.ulm.ac.id/9247/1/Buku_Pengantar_Teknologi_Informasi.pdf
- [14] B. P. Nayoga, R. Adipradana, R. Suryadi, and D. Suhartono, "Hoax Analyzer for Indonesian News Using Deep Learning Models," *Procedia Comput. Sci.*, vol. 179, no. 2020, pp. 704–712, 2021, doi: 10.1016/j.procs.2021.01.059.
- [15] M. H. Goldani, R. Safabakhsh, and S. Momtazi, "Convolutional neural network with margin loss for fake news detection," *Inf. Process. Manag.*, vol. 58, no. 1, p. 102418, Jan. 2021, doi: 10.1016/j.ipm.2020.102418.
- [16] R. Pakpahan, "Analisis Fenomena Hoax Diberbagai Media Sosial Dan Cara Menanggulangi Hoax," *Konf. Nas. Ilmu Sos. Teknol.*, vol. 1, no. 1, pp. 479–484, 2017, [Online]. Available: <http://seminar.bsi.ac.id/knist/index.php/UnivBSI/article/view/184>
- [17] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103–108, Oct. 2021, doi: 10.30871/JAIC.V5I2.3200.