



Analisis sentimen terhadap berita hoaks lowongan kerja dengan metode *naive bayes*

Dwi Sartika Simatupang¹, Siti Nursinta²

¹dwi.sartika@esaunggul.ac.id, ²siti.nursinta_ti20@nusaputra.ac.id

¹Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Esa Unggul

²Program Studi Teknik Informatika, Fakultas Computer Science, Universitas Bina Nusantara

Diterima: 21 Agustus 2024 | Direvisi: 22 Agustus 2024 | Disetujui: 25 Agustus 2024

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Berita hoaks atau berita palsu telah menjadi masalah yang signifikan di era digital saat ini, berita yang dimaksud adalah informasi lowongan pekerjaan di media sosial salah satunya Twitter yang banyak diperbincangkan oleh para pencari kerja. Penyebaran berita hoaks terkait dengan lowongan kerja dapat menyebabkan kerugian bagi pencari kerja yang mengandalkan informasi tersebut untuk mencari pekerjaan. Oleh karena itu, analisis sentimen terhadap berita hoaks lowongan kerja menjadi penting untuk membantu pengguna membedakan antara informasi yang valid dan hoaks. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap berita hoaks yang terkait dengan lowongan kerja pada twitter. Metode yang digunakan adalah metode Naive Bayes, yang merupakan metode klasifikasi yang umum digunakan dalam analisis sentimen. Dataset berita hoaks yang dikumpulkan dari informasi lowongan kerja akan diproses dan fitur-fitur yang relevan dengan lowongan kerja akan diidentifikasi. Selanjutnya, model analisis sentimen Naive Bayes akan dikembangkan untuk mengklasifikasikan sentimen berita hoaks sebagai positif atau negatif. Hasil analisis sentimen akan membantu pengguna dalam mengidentifikasi dan menghindari berita hoaks yang berkaitan dengan lowongan kerja. Selain itu, hasil analisis sentimen juga meningkatkan kualitas konten yang ditampilkan dan memperkuat kepercayaan pengguna terhadap informasi yang disediakan. Diharapkan bahwa penelitian ini akan memberikan kontribusi dalam mengatasi masalah berita hoaks pada twitter, meningkatkan kualitas informasi yang tersedia, dan membantu pencari kerja dalam membuat keputusan yang lebih baik dalam mencari pekerjaan.

Kata kunci: *analisis sentimen, berita hoaks, twitter, lowongan kerja, naive bayes.*

Sentiment Analysis of Job Vacancy Hoax News Using The Naive Bayes Method

Abstract

Hoax news or fake news has become a significant problem in today's digital era, the news in question is information on job vacancies on social media, one of which is Twitter, which is widely discussed by job seekers. The spread of hoax news related to job vacancies can cause harm to job seekers who rely on this information to find work. Therefore, sentiment analysis of hoax job vacancies is important to help users distinguish between valid and hoax information. This study aims to analyze sentiment towards hoax news related to job vacancies on Twitter. The method used is the Naive Bayes method, which is a classification method commonly used in sentiment analysis. Hoax news dataset collected from job vacancy information will be processed and features that are relevant to job vacancies will be identified. Furthermore, the Naive Bayes sentiment analysis model will be developed to classify hoax news sentiment as positive or negative. The results of sentiment analysis will help users identify and avoid hoax news related to job vacancies. In addition, the results of sentiment analysis also improve the quality of the content displayed and strengthen user confidence in the information provided. It is hoped that this research will contribute to overcoming the problem of hoax news on Twitter, improve the quality of available information, and assist job seekers in making better decisions in finding work.

Keywords: *sentiment analysis, hoax news, twitter, job vacancies, naive bayes.*

1. PENDAHULUAN

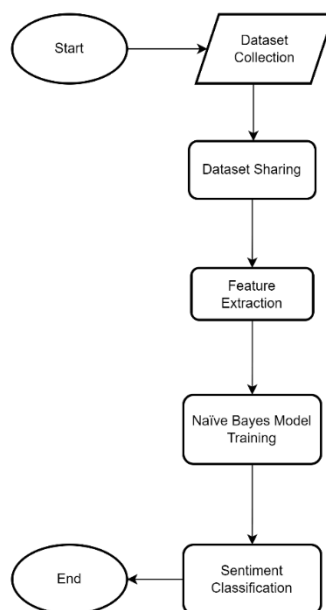
Penggunaan media sosial sudah menjadi kebutuhan pribadi bagi masyarakat, salah satunya yaitu Twitter yang menjadi sumber informasi bagi para pencari kerja. Banyaknya informasi yang diperoleh dengan cepat menjadi dasar tersebarnya berita bohong atau hoaks yang dilakukan oleh oknum yang tidak bertanggung jawab. Salah satunya penyebaran berita hoaks terkait dengan lowongan kerja [1]. Dampak dari berita hoax itu sendiri bisa merubah persepsi orang menjadi sesat. Penyebaran berita hoaks sangat merugikan masyarakat karena banyak pihak yang merasa dirugikan oleh tersebarnya berita hoaks tersebut. Hal ini dilihat dari banyaknya informasi lowongan kerja yang tidak sesuai tersebar pada media sosial twitter [2]. Analisis sentimen memungkinkan pengguna dalam mengevaluasi dan memahami sentimen yang terkandung dalam berita hoaks. Dengan pemahaman tersebut, pengguna dapat membedakan antara berita hoaks dan informasi yang valid, sehingga dapat membuat keputusan yang lebih informatif dan akurat dalam mencari pekerjaan [3].

Selain itu, analisis sentimen dapat membantu dalam meningkatkan kualitas konten yang ditampilkan. Dengan memahami sentimen yang muncul terkait dengan berita hoaks, pengelola dapat mengidentifikasi dan menghapus konten yang tidak valid atau merugikan pelamar. Hal ini akan meningkatkan kepercayaan pengguna terhadap informasi lowongan kerja yang berada pada twitter dan meningkatkan kredibilitas sebagai sumber informasi pekerjaan yang handal [4]. Namun, analisis sentimen terhadap berita hoaks pada lowongan kerja memiliki tantangan tersendiri. Berita hoaks sering kali menggunakan strategi manipulatif dan emosional. Oleh karena itu, diperlukan pengembangan metode analisis sentimen yang efektif dan akurat [5]. Melalui penelitian ini diharapkan dapat dikembangkan teknik dan metode yang lebih baik dalam mengidentifikasi dan mengklasifikasikan sentimen pada konten hoaks. Hal ini akan membantu pengguna dalam menghindari berita hoaks, meningkatkan kualitas informasi terkait lowongan kerja, dan meningkatkan pengalaman pencari kerja dalam mencari pekerjaan yang sesuai dengan kebutuhan dan harapan mereka [6].

METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode naive bayes. Metode **Naive Bayes** adalah sebuah metode klasifikasi yang berdasarkan pada teorema Bayes dengan asumsi bahwa semua fitur yang digunakan adalah independen satu sama lain [7].

Adapun proses dalam metode ini yaitu sebagai berikut.



Gambar 1. Flowchart Metodologi

1. Pengumpulan Dataset

Pengumpulan dataset berita hoaks yang terkait dengan lowongan kerja dari sosial media atau sumber lain yang relevan. Dataset ini mencakup berbagai contoh berita hoaks dengan sentimen yang berbeda, seperti positif atau negatif [8].

2. Pembagian Dataset

Membagi dataset menjadi dua subset: subset pelatihan (training set) yang digunakan untuk melatih model *Naïve Bayes* dan subset pengujian (testing set) untuk menguji kinerja model yang telah dilatih.

3. Ekstraksi Fitur

Ekstraksi fitur dilakukan untuk mengubah teks menjadi representasi numerik yang dapat digunakan oleh model.

4. Pelatihan Model Naïve Bayes

Pelatihan model Naive Bayes dengan menggunakan subset pelatihan. Model Naive Bayes mempelajari pola sentimen dari fitur-fitur dan label sentimen pada data pelatihan. Model akan menghitung probabilitas sentimen (positif atau negatif) berdasarkan fitur -fitur yang ada [9].

5. Kalsifikasi Sentimen

Menggunakan model yang telah dilatih untuk melakukan klasifikasi sentimen pada subset pengujian. Model akan menghitung probabilitas sentimen untuk setiap teks berita hoaks dan memberikan label sentimen berdasarkan probabilitas tertinggi [10].

3. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan dalam beberapa tahap. Pada tahap pengumpulan data, data berupa kalimat-kalimat yang diperoleh dari tweets berbahasa indonesia mengenai hasil persepsi masyarakat. Setelah data terkumpul, kemudian dilakukan preprocessing agar data dapat diolah menggunakan analisis sentimen. Selanjutnya, mengklasifikasi kalimat dengan analisis sentimen sehingga output dapat berupa nilai sentimen tersebut.

A. Tahap Preprocessing

1. *Cleansing*

Cleansing yaitu proses membersihkan kata-kata yang tidak diperlukan. Contohnya seperti username(@), email, hastag(#), dan URL [1].

Tabel 1. Proses Cleansing

Data Latih	
Input	Output
@sarifudhinz Hati-hati gaes.. ini nomernya barusan sy di wa +91 6291 369 236	Hati-hati gaes. Ini nomernya barusan sy di wa +91 6291 369 236

2. *Case Folding*

Case Folding yaitu proses mengubah kata menjadi bentuk yang sama, baik menjadi huruf kapital atau huruf kecil.

3. *Tokenizing*

Tokenizing yaitu proses sebuah teks yang terpotong oleh

Data Latih	
Input	Output
Harap berhati-hati terhadap pihak yang mengatasnamakan Pertamina. Tetap semangat dan sukses selalu ya sob	[Harap, berhati-hati, terhadap, pihak, yang,mengatasnamakan, Pertamina, tetap semangat, dan sukses, selalu, ya]

mengidentifikasi kata-kata dalam spasi atau karakter special.

Tabel 2. Proses Tokenizing

4. *Stopword Removal*

Stopword removal yaitu proses penghapusan kata-kata yang sering muncul dan bersifat umum. Seperti penggunaan kata penghubung : dan, yang, ya, dan lainnya.

5. *Stemming*



Stemming yaitu proses untuk membuat kata yang berimbuhan (imbuhan me-, meng-, pe-, peng- , dan lain-lain) menjadi kata dasar.

Tabel 3. Proses Stemming

Data Latih	
Input	Output
membuat	buat
mengatasnamakan	atasnama
pelatihan	latih
penglihatan	lihat

Data Sumber

Sumber data yang dipakai pada penelitian ini yaitu berbentuk komentar yang diambil langsung dari *Twitter*. Berikut contoh yang digunakan :

Tabel 4. Contoh Data yang Digunakan

Komentar
@Linda66663301 Saya sudah kena tipu...
@r1en3goo Jangan sampai tertipu yaa
@ginanjarnur_p Bulan maret mulai beraksi lagi ni penipu nya
@pvprivasi Wihh mesti hati2 nih
@nindi_kirani Pintar2 dlm memilih informasi ya guys

Implementasi Preprocessing

Sebelum mengklasifikasikan naive bayes, data akan diolah terlebih dahulu pada tahap ini agar pada tahap klasifikasi hasilnya akan lebih optimal.

Tabel 5. Data Hasil Preprocessing

Komentar
[Saya, sudah, kena, tipu]
[Jangan, sampai, tertipu, yaa]
[Bulan_maret, mulai_beraksi, penipu_nya]
[Wihh, mesti, hati-hati]
[Pintar-pintar, dalam_memilih, informasi]

Penerapan Naive Bayes

Pada tahap ini merupakan proses pengklasifikasin berdasarkan sentimen yang telah ada. Tahap ini meliputi dua proses sebagai berikut.

1. Proses Learning Naive Bayes Classifier (NBC)

Terdapat tiga langkah dalam proses learning NBC ini, yaitu :

- Membentuk Fitur

Fitur yang dimaksud adalah kata kunci yang menjadi parameter satuan data latih, yaitu komentar yang akan diklasifikasikan ke dalam kelas sentimen yang telah ditentukan (positif dan negatif) [11].

Tabel 6. Pembentukan Fitur Data Latih

Data	Fitur	Sentimen
D1	kena(1), tipu(1)	negatif
D2	tertipu(1)	negatif
D3	beraksi(1), penipu(1)	negatif
D4	hati-hati(1)	positif
D5	pintar-pintar(1), memilih_informasi(1)	positif

- Menghitung Probabilitas

Setelah membentuk fitur dari kemunculan data latih, selanjutnya menghitung probabilitas dari setiap kelas dengan cara berikut ini :

$$p(C_i) = \frac{fd(C_i)}{|D|} \dots (1)$$

Tabel 7. Pelatihan Probabilitas Naive Bayes

Kelas Sentimen	Data					fd	p
	D1	D2	D3	D4	D5		
Negatif	1	1	1	0	0	3	3/5
Positif	0	0	0	1	1	2	2/5

- Menentukan Probabilitas

Setelah mendapatkan probabilitas, selanjutnya menghitung fitur-fitur dari setiap kelas dengan cara sebagai berikut :

$$p(W_k|C_i) = \frac{f(W_k, C_i) + 1}{f(C_i) + |W|} \dots (2)$$

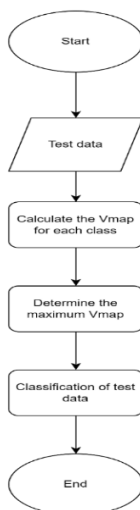
Tabel 8. Data Latih Probabilitas

Data f(W _k , C _i)	Kelas Sentimen

	Negative	Positif
kena	$\frac{1+1}{3+8} = \frac{2}{11}$	$\frac{0+1}{2+7} = \frac{1}{9}$
penipu	$\frac{1+1}{3+8} = \frac{2}{11}$	$\frac{0+1}{2+7} = \frac{1}{9}$
beraksi	$\frac{1+1}{3+8} = \frac{2}{11}$	$\frac{0+1}{2+7} = \frac{1}{9}$
hati-hati	$\frac{0+1}{3+8} = \frac{1}{11}$	$\frac{1+1}{2+7} = \frac{2}{9}$
informasi	$\frac{0+1}{3+8} = \frac{1}{11}$	$\frac{1+1}{2+7} = \frac{2}{9}$

Proses Klasifikasi Naïve Bayes Classifier (NBC)

Adapun alur proses klasifikasi naive bayes classifier nya yaitu sebagai berikut.



Gambar 2. Flowchart Klasifikasi Naive Bayes

Contoh salah satu komentar yang dijadikan data uji menggunakan model probabilitas.

Tabel 9. Data Uji Klasifikasi

Komentar	
Sebelum Preprocessing	Sesudah Preprocessing
Harap berhati-hati terhadap pihak yang mengatasnamakan Pertamina. Tetap semangat dan sukses selalu ya	berhati-hati pihak mengatasnamakan

	semangat_sukses
--	-----------------

Setelah menentukan data uji, langkah selanjutnya yaitu :

Menghitung Vmap

Vmap yaitu perhitungan yang digunakan naive bayes classifier untuk menentukan probabilitas dari masing-masing kelas berdasarkan proses learning. Nilai tertinggi akan dipilih sebagai nilai probabilitas.

Berdasarkan hasil training, berikut perhitungannya :

$$V_{map} = \underset{\{negatif, positif\}}{\operatorname{argmax}} p(wk|c) \times p(c) \dots (3)$$

$$V_{map} = \underset{\{negatif, positif\}}{\operatorname{argmax}} p(wk|c) \times p(c)$$

$$P(\text{"berhati-hati"}|ci) \quad p(\text{"pihak"}|ci)$$

$$p(\text{"mengatasnamakan"}|ci) \quad p(\text{"semangat"}|ci)$$

$$p(\text{"sukses"}|ci).$$

Vmap sentimen negatif

$$V_{map} (\text{"negatif"}) =$$

$$P(\text{"berhati-hati"}|negatif) \quad p(\text{"pihak"}|negatif) \quad p(\text{"mengatasnamakan"}|negatif)$$

$$p(\text{"semangat"}|negatif) \quad p(\text{"sukses"}|negatif).$$

$$= 2/5 \times 2/11 \times 2/11 \times 2/11 \times 1/11 \times 1/11 = 0,018$$

Vmap sentimen positif

$$V_{map} (\text{"positif"}) =$$

$$P(\text{"berhati-hati"}|positif) \quad p(\text{"pihak"}|positif)$$

$$p(\text{"mengatasnamakan"}|positif)$$

$$p(\text{"semangat"}|positif) \quad p(\text{"sukses"}|positif).$$

$$= 25 \times 1/9 \times 1/9 \times 1/9 \times 2/9 \times 2/9 = 0,012$$

- Menentukan Vmap Maksimum

Dari hasil perhitungan Vmap diatas, maka dapat diketahui bahwa nilai Vmap negatif lebih besar dibandingkan nilai Vmap positif. Jadi, dapat disimpulkan bahwa komentar tersebut diklasifikasikan ke dalam kelas sentimen negatif.

Pengujian Naïve Bayes

Pengujian naïve bayes ini dilakukan untuk mengevaluasi dengan menggunakan confusion matrix. Tujuan dilakukannya evaluasi untuk mengukur kinerja klasifikasi yang dilakukan, Berikut adalah tabel confusion matrix hasil klasifikasi menggunakan model naïve bayes.

Tabel 10. Confusion Matrix

		True Value	
		Positive	Negative
Prediction	Positive	120	25
	Negative	3	7

Hasil Perhitungan nilai akurasi, presicion, recall, specificity dideskripsikan sebagai berikut:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots(3)$$

$$\text{Akurasi} = \frac{120+7}{120+7+25+3} = 0,81$$

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{120}{120 + 25} = 0,827$$

$$\text{recall} = \frac{TP}{TP + FN} = \frac{120}{120 + 3} = 0,97$$

$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{7}{7 + 25} = 0,21$$

Akurasi yang dihasilkan 0,81, seperti yang dikerjakan sesuai rumus diatas. Dari 155 data tes, 120 diklasifikasikan dengan benar sebagai positif, dan 7 diklasifikasikan dengan benar sebagai negatif. Sementara itu, 3 titik data positif salah diklasifikasikan sebagai positif, dan 25 titik data negatif salah diklasifikasikan sebagai negatif.

4. KESIMPULAN

Analisis sentimen terhadap berita hoaks lowongan kerja dengan metode Naive Bayes dapat memberikan kontribusi yang signifikan dalam mengidentifikasi dan membedakan berita hoaks dari informasi valid. Dengan memahami sentimen yang terkandung dalam berita hoaks tersebut, pelamar dapat membuat keputusan yang lebih informatif dan akurat dalam mencari pekerjaan. Namun, penting untuk melihat hasil analisis dengan konteks yang tepat dan mempertimbangkan faktor lain yang dapat mempengaruhi kualitas dan keakuratan analisis sentimen.

Adapun beberapa hal yang perlu ditambahkan dalam pengembangan analisis sentimen terhadap berita hoaks lowongan kerja dengan metode Naive Bayes ini yaitu sebagai berikut.

1. Melakukan pengembangan dataset yang lebih besar dan representif.
2. Menggunakan metode lain untuk meningkatkan akurasi dan keandalan analisis setimen.
3. Memperhatikan konteks dan kebaruan informasi.

DAFTAR PUSTAKA

[1] K. Widodo, Bobby, H. Matondang, Nur, S. Prasvita, and Desta, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Penggunaan Aplikasi Jobstreet Implementation of Naive Bayes Algorithm For Sentiment Analysis Of The Use Of Jobstreet Application," *Agustus*, vol. 21, no. 3, pp. 523–533, 2022.

[2] Wahyu Sejati, Ankur Singh Bist, and Amirsyah Tambunan, "Pengembangan Analisis Sentimen dalam Rekayasa Software Engineering menggunakan tinjauan literatur sistematis," *J. MENTARI Manajemen, Pendidik. dan Teknol. Inf.*, vol. 2, no. 1, pp. 95–103, 2023, doi: 10.33050/mentari.v2i1.377.

[3] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.

[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://jurnal.peneliti.net/index.php/JIWP/article/view/2235>
- [5] A. F. Rahmah, Dwi Sartika Simatupang, and Alun Sujjada, “Sistem Monitoring Dan Kontrol Tanaman Pada Greenhouse Berbasis Android Menggunakan Fuzzy Sugeno,” *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 4, no. 2, pp. 332–340, 2023, doi: 10.37859/coscitech.v4i2.5088.
- [6] M. H. G. Dede Herman and Dwi Sartika Simatupang, “Sistem monitoring siswa bermasalah berbasis web di smp insan cendekia arrasyid,” *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 3, no. 3, pp. 510–517, 2022, doi: 10.37859/coscitech.v3i3.4307.
- [7] R. Nurlaely, S. D. Sartika, Kamdan, and I. L. Kharisma, “Analisis Sentimen Twitter Terhadap Cyberbullying Menggunakan Metode Support Vector Machine (SVM),” *J. Comput. Sci. Inf. Technol.*, vol. 4, no. 2, pp. 376–384, 2023, [Online]. Available: <http://ejournal.umri.ac.id/index.php/coscitech/indexhttps://doi.org/10.37859/coscitech.v4i2.5161>
- [8] D. S. Simatupang, “The Pengembangan Data Warehouse untuk Mengetahui Tren Aktivitas Trouble Ticket,” *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 3, no. 3, pp. 341–349, 2022, doi: 10.37859/coscitech.v3i3.4287.
- [9] H. S. Utama, D. Rosiyadi, D. Aridarma, and B. S. Prakoso, “Sentimen Analisis Kebijakan Ganjil Genap Di Tol Bekasi Menggunakan Algoritma Naive Bayes Dengan Optimalisasi Information Gain,” *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 247–254, 2019, doi: 10.33480/pilar.v15i2.705.
- [10] E. B. Santoso and A. Nugroho, “Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook,” *Eksplora Inform.*, vol. 9, no. 1, pp. 60–69, 2019, doi: 10.30864/eksplora.v9i1.254.
- [11] R. Aziz *et al.*, “Rizky Aziz 1), Tresna Maulana Fahrudin 2),” *J. Fasilkom*, vol. 14 No. 1 |, no. April 2024:, pp. 1–10, 2024.