

Implementasi Metode Multinomial Naive Bayes dalam Klasifikasi Judul Berita Clickbait

by Dicky Satria Mahendra

Submission date: 18-Jul-2024 03:37PM (UTC+0700)

Submission ID: 2418618543

File name: NEPTUNUS_Vol_2_no_3_Agust_2024_hal_303-316.pdf (1.2M)

Word count: 4440

Character count: 27713



Implementasi Metode Multinomial Naive Bayes dalam Klasifikasi Judul Berita Clickbait

Dicky Satria Mahendra

Universitas Pembangunan Nasional “Veteran” Jawa Timur, Indonesia

Basuki Rahmat

Universitas Pembangunan Nasional “Veteran” Jawa Timur, Indonesia

Retno Mumpuni

Universitas Pembangunan Nasional “Veteran” Jawa Timur, Indonesia

Alamat: RungkutMadya No. 1, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur

Abstract. This research aims to classify news headlines into clickbait and non-clickbait using the Multinomial Naive Bayes method. The data used comes from the dataset CLICK-ID: A Novel Dataset for Indonesian Clickbait Headlines. The research process involves stages of data collection, preprocessing, feature extraction, model training, model evaluation, and result analysis. The test results show that the Multinomial Naive Bayes algorithm consistently produces an accuracy rate of around 78%. Optimization using Grid Search did not result in an accuracy improvement. However, there was an improvement in the recall value for the non-clickbait class from 76% to 80%. The best parameter found was an alpha of 0.15. Therefore, the Multinomial Naive Bayes algorithm can be effectively used to address the problem of classifying clickbait news headlines, with the potential to contribute to clickbait prevention efforts in the future.

Keywords: Multinomial Naive Bayes, Grid Search Optimization, Classification, Clickbait News.

Abstrak. Penelitian ini bertujuan untuk mengklasifikasikan judul berita menjadi clickbait dan non-clickbait menggunakan metode Multinomial Naive Bayes. Data yang digunakan berasal dari dataset CLICK-ID: A Novel Dataset for Indonesian Clickbait Headlines. Proses penelitian melibatkan tahapan pengumpulan data, preprocessing, ekstraksi fitur, pelatihan model, evaluasi model, dan analisis hasil. Hasil pengujian menunjukkan bahwa algoritma Multinomial Naive Bayes mampu menghasilkan tingkat akurasi yang konsisten sekitar 78%. Melalui optimasi menggunakan Grid Search, tidak menghasilkan peningkatan akurasi. Meskipun tidak terjadi peningkatan akurasi, terdapat perbaikan dalam nilai recall untuk kelas non-clickbait dari 76% menjadi 80%. Parameter terbaik yang ditemukan adalah alpha sebesar 0.15. Dengan demikian, Algoritma Multinomial Naive Bayes dapat digunakan secara efektif dalam mengatasi masalah klasifikasi judul berita clickbait, dengan potensi untuk memberikan kontribusi dalam upaya pencegahan clickbait di masa mendatang.

Kata kunci: Multinomial Naive Bayes, Optimasi Grid Search, Klasifikasi, Berita Clickbait

1. LATAR BELAKANG

Di era digital saat ini, informasi merupakan sebuah kebutuhan bagi setiap orang. Dengan kemajuan teknologi internet akan menghasilkan peningkatan besar dalam penyebaran informasi, terutama informasi berita elektronik Berbahasa Indonesia. Sebelum memasuki era digital, untuk mendapatkan berita terbaru dan relevan masyarakat biasanya mencari berita melalui media cetak, televisi, koran, ataupun radio. Menurut survei Reuters Institute terbaru bertajuk Digital News Report 2023, Media Online memiliki tingkat survei tertinggi dengan

angka mencapai 84% dibandingkan dengan Media Sosial sebesar 65%, Televisi sebesar 54%, dan Media Cetak sebesar 15% (Annur, 2023). Portal berita online atau perusahaan media digital dapat memanfaatkan kondisi ini untuk bersaing menarik perhatian pembaca agar mengunjungi situs mereka, yang pada akhirnya dapat menyesatkan pembaca. Fenomena tersebut disebut dengan Clickbait yang digunakan media online untuk meningkatkan jumlah kunjungan pengunjung. Clickbait merupakan jenis konten web yang dirancang dengan judul menggunakan formula dan tata bahasa tertentu dengan maksud untuk menarik perhatian pembaca agar mengklik tautan (Palau & Sampio, 2016). Penggunaan clickbait yang berlebihan di media online dianggap berbahaya, mengingat masih banyaknya masyarakat yang tingkat literasinya tidak sebanding dengan jumlah pengguna internet. Maka dari ³itu, perlu adanya ³sistem otomatis yang dapat mengklasifikasi berita clickbait dengan memanfaatkan teknik text mining. Teknik text mining adalah model keterbaruan dari ³metode data mining dan dapat digunakan untuk mengatasi permasalahan dalam mengklasifikasi berita clickbait. Sebelum melakukan analisis ³text menggunakan metode dalam text mining, langkah awal yang perlu dilakukan adalah ³pre-processing teks. ³Text Processing adalah suatu proses pengolahan teks yang melibatkan berbagai teknik dan algoritma dengan tujuan mengekstrak informasi dari teks. Terdapat beberapa tahapan untuk melakukan pengolahan teks atau text processing, seperti pembersihan teks, ekstraksi fitur, dan klasifikasi teks. Setelah itu akan dilakukan metode klasifikasi berita clickbait. Klasifikasi adalah proses di mana ¹⁷suatu dokumen ditetapkan ke dalam ¹⁷kategori atau kelas tertentu sesuai dengan karakteristik yang dimilikinya (Nurhadi, A., 2016).

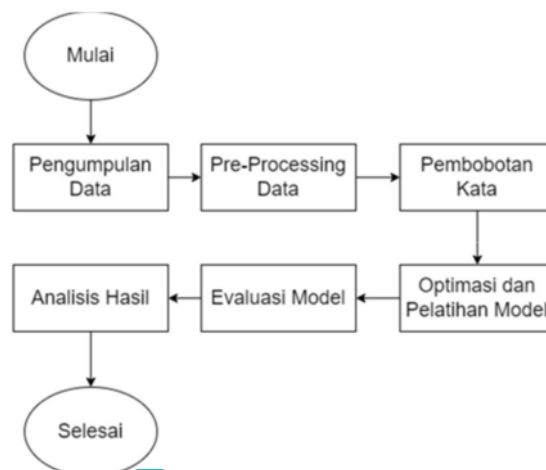
Banyak algoritma yang dapat digunakan dalam klasifikasi text, seperti ³Naïve Bayes Classifier (NBC), K-Nearest Neighbour (KNN), Artificial Neural Network (ANN), dan Support Vector Machines (SVM). Pada ¹⁸penelitian ini, akan menggunakan metode Multinomial Naïve Bayes karena ¹⁸metode ini sangat efisien dalam pembelajaran, terutama pada dataset yang relatif besar. Selain itu, kemudahan implementasi hasil dari metode Naïve Bayes menjadi pilihan yang baik untuk mengklasifikasi judul berita. Harapan dilakukannya penelitian ini semoga bisa memberikan manfaat dalam mengidentifikasi judul berita clickbait, sehingga dapat menjadi tambahan informasi yang berguna dalam upaya pencegahan clickbait di masa mendatang. Selain itu, diharapkan metode ini dapat berperan sebagai sumber pengetahuan yang berharga terkait penggunaan Multinomial Naive Bayes.

2. KAJIAN TEORITIS

Dalam proses menjalankan penelitian ini, penulis perlu merujuk pada literatur penelitian sebelumnya yang terkait dengan fokus penelitian. Tujuannya untuk mendapatkan pemahaman yang mendalam mengenai topik penelitian yang sedang dijalankan. Penelitian tersebut akan menjadi sumber referensi dan panduan dalam pengembangan penelitian ini. Penelitian yang dilakukan oleh Mori Hovipah, Elin Hearani, Fadhilah Syafria berjudul "Klasifikasi Clickbait Menggunakan Transformers" (Mori Hovipah et al., 2023). Melakukan ⁵klasifikasi judul berita ⁵clickbait menggunakan metode Transformers. Penelitian ini melibatkan total 6632 judul berita. Proses klasifikasinya mencakup beberapa tahap, yaitu pengumpulan data, pelabelan data, praproses, Exploratory Data Analysis (EDA), dan ⁵penerapan ⁵klasifikasi menggunakan ⁵transformers. Hasil ⁵terbaik yang dicapai adalah akurasi sebesar 63%, dengan precision 0.63 dan recall 1, yang diperoleh dari ³⁸pembagian data 10% untuk data uji dan 90% untuk data latih.

Penelitian lain yang dilakukan oleh Afandi dan rekan-rekannya pada tahun 2022, berjudul "Klasifikasi Judul Berita Clickbait menggunakan RNN-LSTM" (Asro'i & Febuariyanti, 2022). Penelitian ini bertujuan untuk mengklasifikasikan judul berita, membedakan antara judul berita clickbait dan non-clickbait, dengan menerapkan ¹⁰metode Deep ¹⁰Learning menggunakan arsitektur RNN-LSTM. Hasil klasifikasi menunjukkan akurasi tertinggi sebesar 79% pada data latih dengan epoch 27, dan akurasi pada data validasi mencapai 77%.

3. METODE PENELITIAN



²⁸**Gambar 1. Alur Penelitian**

Pengumpulan Data

Dalam penelitian ini, penggunaan data bersifat sekunder yang didapatkan melalui ⁹dataset CLICK-ID: A Novel Dataset for Indonesian Clickbait Headlines yang dapat diakses melalui ³⁷situs <https://data.mendeley.com/datasets/k42j7x2kpn/1>, dan dataset ini diterbitkan pada 11 Mei 2020. Data yang dianalisis mencakup judul berita dari bulan Maret hingga November tahun 2019, yang berasal ⁷dari 12 situs berita, antara lain kompas.com, liputan6.com, okezone.com, posmetro.com, detiknews.com, fimela.com, kapanlagi.com, republika.co.id, sindonews.com, tempo.com, tribunnews.com, dan wowkeren.com. Dataset yang digunakan untuk 16 penelitian ini sebanyak 300 data. Dataset ini telah dilabeli oleh 3 orang dengan menggunakan sistem voting, dimana label didasarkan pada hasil voting terbanyak. Dataset ini menggunakan 3 kolom yaitu Judul, Label, dan Label Score yang disimpan ke dalam format excel, Setelah itu, data akan diproses dengan bahasa pemrograman Python ¹untuk melakukan klasifikasi berita clickbait menggunakan ¹metode Multinomial Naive Bayes. Berikut adalah contoh dataset yang akan digunakan dalam penelitian

Text Processing

Pada tahap ini, teks akan diproses agar dapat diolah dan dianalisis dengan lebih efektif. Proses ini melibatkan berbagai teknik text preprocessing untuk membersihkan dan mengorganisir data teks seperti ³⁵Case Folding, Cleansing, Tokenizing, Stopword, dan Stemming.

Pembobotan Kata

⁴TF-IDF (Term Frequency-Inverse Document Frequency) adalah metode pembobotan yang melibatkan perhitungan jumlah kemunculan suatu term dalam suatu dokumen dan jumlah dokumen yang mengandung kata tersebut (Taantyoko et al., 2019). Dengan menggunakan metode TF-IDF, kita dapat meminimalkan kegunaan kata dalam sebuah dokumen dan memanfaatkan nilai tersebut untuk analisis lebih lanjut pada dokumen tersebut (Herwijayanti, Ratnawati, & Muflikhah, 2018). Berikut merupakan rumus dari TF-IDF.

$$w_{t,d} = TF_{t,d} \times IDF_{t,d} \times \log \frac{N}{dft}$$

Keterangan :

W = Bobot TF-IDF setiap kata

TF = Jumlah frekuensi data yang muncul

IDF = Nilai ²²inverse frekuensi dokumen

DF_t = Jumlah frekuensi dokumen yang terdapat kata t

N = Jumlah total dokumen

Optimasi Grid Search

Tahap optimasi Grid Search dilakukan agar dapat meningkatkan hasil akurasi dari metode pengolahan data yaitu Multinomial Naïve Bayes. Dalam penelitian ini, penulis memilih untuk menggunakan salah satu metode optimasi yaitu Grid Search untuk mengoptimasi model algoritma Multinomial Naïve Bayes. Grid Search adalah salah satu metode optimasi yang mudah untuk memodifikasi beberapa parameter. Grid Search dapat melakukan optimasi dengan cara melakukan seleksi fitur (feature selection) dan tuning parameter untuk meningkatkan kemampuan analisis data. Grid search biasanya dilakukan bersamaan dengan k-fold crossvalidation, yang menghasilkan indeks evaluasi untuk model klasifikasi. (Toha et al., 2022).

Multinomial Naive Bayes

Multinomial Naïve Bayes adalah suatu metode yang memanfaatkan probabilitas dan termasuk dalam supervised learning pada kasus klasifikasi teks. Pendekatan ini mengacu pada prinsip distribusi multinomial dalam probabilitas bersyarat (Manning, et al., 2009). Dalam metode ini, diasumsikan bahwa setiap atribut saling tergantung satu sama lain dengan mempertimbangkan konteks kelas, dan tidak memperhitungkan dependensi antar atribut (Saleh, 2014). Berikut merupakan rumus proses perhitungan dari metode Multinomial Naive Bayes.

$$P(c, d) = \frac{N_c}{N} \times P(t_1, c) \times \dots \times P(t_n, c)$$

1
Keterangan :

$P(c, d)$: Probabilitas suatu dokumen termasuk kelas c

N_c : Jumlah kelas c pada seluruh dokumen

N : Jumlah seluruh dokumen

t_n : Kata dokumen d ke- n

$P(t_n, c)$: Probabilitas kata ke- n dengan diketahui kelas c

Confusion Matrix

Confusion matrix digunakan sebagai alat analisis untuk mengevaluasi seberapa baik klasifikasi dapat mengidentifikasi instance dari kelas yang berbeda. Dalam situasi ini, terdapat dua kelas yang dikenal sebagai true positif dan true negatif. True positif menunjukkan instance dari sampel positif yang secara akurat diberi label sesuai dengan kelasnya, sedangkan true negative

merujuk pada instance dari sampel negatif yang benar-benar diberi label sesuai dengan kelasnya. False positif terjadi ketika instance dari sampel negatif diberi label secara tidak tepat, sedangkan false negatif terjadi ketika instance dari sampel positif diberi label secara tidak tepat. Penjelasan ini diambil dari (Yunita, 2017).

Evaluasi Model

Tahap selanjutnya adalah melakukan evaluasi model untuk menilai sejauh mana kinerja model machine learning yang telah dilatih. Dalam melakukan pengujian terhadap data latih, evaluasi terhadap hasil prediksi menjadi suatu kebutuhan yang penting. Pada penelitian ini menggunakan Confusion Matriks untuk menguji hasil pelatihan data latih dengan memperhatikan akurasi, presisi, dan recall.

Analisis Hasil

Dalam analisis ini, fokus diberikan pada performa model dalam mengklasifikasikan data uji melalui metrik-metrik evaluasi yang telah diaplikasikan, seperti akurasi, presisi, recall, dan F1-score. Analisis lebih lanjut menggambarkan sejauh mana model mampu mengatasi tantangan khusus, seperti ketidakseimbangan kelas atau kompleksitas masalah. Pemahaman yang mendalam terhadap kelemahan dan kelebihan model diperoleh melalui pemeriksaan hasil per kelas dan per metrik evaluasi. Hal ini memberikan pandangan yang komprehensif terhadap kinerja model dalam menghadapi skenario dunia nyata.

⁴¹4. HASIL DAN PEMBAHASAN

Pengumpulan Data

Pada tahap pertama untuk melakukan klasifikasi dalam penelitian ini adalah mengumpulkan dataset ⁵yang diperoleh dari CLICK-ID: A Novel Dataset for Indonesian Clickbait Headlines. Seperti yang dijelaskan dalam bab-bab sebelumnya. ¹⁶Data yang digunakan pada penelitian ini bersifat sekunder dan berasal dari 12 situs berita yang digabung menjadi satu dalam file excel .csv. ³³pada penelitian ini menggunakan 300 data yang terdiri dari 2 kelas, yaitu clickbait dan non clickbait. Berikut merupakan potongan kode untuk menampilkan data yang akan digunakan untuk proses klasifikasi.

Evaluasi Model

Setelah melalui proses pra-pemrosesan data, pembagian data, optimalisasi, dan pengujian model, tahapan selanjutnya adalah melakukan evaluasi model untuk mengevaluasi seberapa baik klasifikasi yang telah dilakukan. Evaluasi model dilakukan menggunakan confusion matrix untuk memahami seberapa baik model dapat memprediksi ²³True Positive, True Negative,

False Positive, dan False Negative. Dari nilai-nilai tersebut, dapat dihitung akurasi (Accuracy), presisi (Precision), recall (Recall), dan F1-Score.

1. Evaluasi Model Skenario 1

Pada skenario 1 akan dilakukan pengujian klasifikasi judul berita menggunakan algoritma multinomial naive bayes. Pengujian ini dilakukan dengan membagi data menggunakan perbandingan 90% untuk data latih dan 10% untuk data uji. Berikut merupakan hasil tabel confusion matrix pada skenario 1.

Tabel 1. Hasil Confussion Matrix Skenario 1

Total Data	Clickbait	Non-clickbait	Data Salah	Akurasi
30	14	12	4	86%

Dari hasil klasifikasi, menunjukkan bahwa model memiliki tingkat akurasi sebesar 86.67%. Berdasarkan nilai presisi masing-masing kelas yang diperoleh, untuk kelas non-clickbait (0) mendapatkan 92%, sementara untuk kelas clickbait (0) mendapatkan 82%. Hasil ini menunjukkan bahwa berdasarkan 15 total data yang ada, 13 data yang benar-benar non-clickbait berhasil diprediksi, sedangkan 2 data lainnya salah diprediksi. Lalu, untuk kelas clickbait (1) dari 15 total data yang ada, 12 data yang benar-benar clickbait berhasil diprediksi, sedangkan 3 data lainnya salah diprediksi. Kemudian untuk recall setiap kelas mendapatkan hasil sebesar 80% untuk non-clickbait, dan 93% untuk clickbait. Hasil ini menunjukkan bahwa berdasarkan 15 total data yang ada seharusnya non-clickbait, terdapat 12 data yang diketahui oleh model dan 3 data yang lain tidak diketahui oleh model. Lalu, berdasarkan 15 data yang ada seharusnya clickbait, terdapat 14 data yang berhasil diketahui oleh model dan 1 data lainnya tidak diketahui oleh model. Setelah didapatkan nilai dari precision dan recall, hasil lain didapatkan pada F1-Score yang merupakan nilai rata-rata harmonik dari precision dan recall. Untuk hasil F1-Score pada skenario 1 ini sebesar 86% untuk judul berita non-clickbait dan 87% untuk judul berita clickbait.

Tabel 2. Hasil Classification Report Skenario 1

Kelas	Precision	Recall	F1-Score	Support
Clickbait	92%	80%	86%	15
Non-clickbait	82%	93%	87%	15

1. Evaluasi Model Skenario 2

Selanjutnya pada skenario 2 akan dilakukan pengujian klasifikasi judul berita ²¹menggunakan algoritma multinomial naive bayes. Pengujian ini dilakukan dengan ¹¹membagi data menggunakan perbandingan 80% untuk data latih dan 20% untuk data uji. Berikut merupakan hasil tabel confusion matrix pada skenario 2.

Tabel 3. Hasil Confussion Matrix Skenario 2

Total Data	Clickbait	Non-clickbait	Data Salah	Akurasi
60	24	20	16	73%

Dari hasil klasifikasi, menunjukkan bahwa model memiliki tingkat akurasi sebesar 73%. Berdasarkan nilai presisi masing-masing kelas yang diperoleh, untuk kelas non-clickbait (0) mendapatkan 74%, sementara untuk kelas clickbait (0) 52 mendapatkan 73%. Hasil ini menunjukkan bahwa berdasarkan 29 total data yang ada, 21 data yang benar-benar non-clickbait berhasil diprediksi, sedangkan 8 data lainnya salah diprediksi. Lalu, untuk kelas clickbait (1) dari 31 total data yang ada, 23 data yang benar-benar clickbait berhasil diprediksi, sedangkan 8 data lainnya salah diprediksi. Kemudian untuk recall setiap kelas mendapatkan hasil sebesar 69% untuk non-clickbait, dan 77% untuk clickbait. Hasil ini menunjukkan bahwa berdasarkan 29 total data yang ada seharusnya non-clickbait, terdapat 20 data yang diketahui oleh model dan 88 data yang lain tidak diketahui oleh model. Lalu, berdasarkan 469 data yang ada seharusnya clickbait, terdapat 446 data yang berhasil diketahui oleh model dan 23 data lainnya tidak diketahui oleh model. Setelah didapatkan nilai dari precision dan recall, hasil lain didapatkan pada ²F1-Score yang merupakan nilai rata-rata harmonik dari precision dan recall. Untuk hasil F1-Score pada skenario 2 ini sebesar 44% untuk judul berita non-clickbait dan 89% untuk judul berita clickbait.

Tabel 4. Hasil Classification Report Skenario 2

Kelas	Precision	Recall	F1-Score	Support
Clickbait	74%	69%	71%	29
Non-clickbait	73%	77%	75%	31

2. Evaluasi Model Skenario 3

Selanjutnya pada skenario 3 akan dilakukan pengujian klasifikasi judul berita menggunakan algoritma multinomial naive bayes. Pengujian ini dilakukan dengan membagi data menggunakan perbandingan 80% untuk data latih dan 20% untuk data uji. Berikut merupakan hasil tabel confusion matrix pada skenario 3.

Tabel 5. Hasil Confussion Matrix Skenario 3

Total Data	Clickbait	Non-clickbait	Data Salah	Akurasi
90	34	34	22	75%

Dari hasil klasifikasi, menunjukkan bahwa model memiliki tingkat akurasi sebesar 75%. Berdasarkan nilai presisi masing-masing kelas yang diperoleh, untuk 54 kelas non-clickbait (0) mendapatkan 77%, sementara untuk kelas clickbait (1) mendapatkan 74%. Hasil ini menunjukkan bahwa berdasarkan 46 total data yang ada, 35 data yang benar-benar non-clickbait berhasil diprediksi, sedangkan 11 data lainnya salah diprediksi. Lalu, untuk kelas clickbait (1) dari 44 total data yang ada, 32 data yang benar-benar clickbait berhasil diprediksi, sedangkan 12 data lainnya salah diprediksi. Kemudian untuk recall setiap kelas mendapatkan hasil sebesar 74% untuk non-clickbait, dan 77% untuk clickbait. Hasil ini menunjukkan bahwa berdasarkan 46 total data yang ada seharusnya non-clickbait, terdapat 34 data yang diketahui oleh model dan 12 data yang lain tidak diketahui oleh model. Lalu, berdasarkan 44 data yang ada seharusnya clickbait, terdapat 34 data yang berhasil diketahui oleh model dan 10 data lainnya tidak diketahui oleh model. Setelah didapatkan nilai dari precision dan recall, hasil lain didapatkan pada **F1-Score** yang merupakan nilai rata-rata harmonik dari precision dan recall. Untuk hasil **F1-Score** pada skenario 2 ini sebesar 76% untuk judul berita non-clickbait dan 76% untuk judul berita clickbait.

Tabel 6. Hasil Classification Report Skenario 3

Kelas	Precision	Recall	F1-Score	Support
Clickbait	74%	69%	71%	29
Non-clickbait	73%	77%	75%	31

3. Evaluasi Model Skenario 4

Selanjutnya pada skenario 4 akan dilakukan pengujian klasifikasi judul berita menggunakan algoritma multinomial naive bayes dan optimasi grid search. Pengujian ini dilakukan dengan membagi data menggunakan perbandingan 90% untuk data latih dan 10% untuk data uji. Berikut merupakan hasil tabel confusion matrix pada skenario 4.

Tabel 7. Hasil Confussion Matrix Skenario 4

Total Data	Clickbait	Non-clickbait	Data Salah	Akurasi
30	13	11	6	80%

Dari hasil klasifikasi, menunjukkan bahwa model memiliki tingkat akurasi sebesar 80%. Berdasarkan nilai presisi masing-masing kelas yang diperoleh, untuk kelas non-clickbait (0) mendapatkan 85%, sementara untuk kelas clickbait (1) 56 mendapatkan

76%. Hasil ini menunjukkan bahwa berdasarkan 15 total data yang ada, 13 data yang benar-benar non-clickbait berhasil diprediksi, sedangkan 2 data lainnya salah diprediksi. Lalu, untuk kelas clickbait (1) dari 15 total data yang ada, 11 data yang benar-benar clickbait berhasil diprediksi, sedangkan 4 data lainnya salah diprediksi. Kemudian untuk recall setiap kelas mendapatkan hasil sebesar 73% untuk non-clickbait, dan 87% untuk clickbait. Hasil ini menunjukkan bahwa berdasarkan 15 total data yang ada seharusnya non-clickbait, terdapat 11 data yang diketahui oleh model dan 4 data yang lain tidak diketahui oleh model. Lalu, berdasarkan 15 data yang ada seharusnya clickbait, terdapat 13 data yang berhasil diketahui oleh model dan 2 data lainnya tidak diketahui oleh model. Setelah didapatkan nilai dari precision dan recall, hasil lain didapatkan pada ²F1-Score yang merupakan nilai rata-rata harmonik dari precision dan recall. Untuk hasil F1-Score pada skenario 4 ini sebesar 79% untuk judul berita non-clickbait dan 81% untuk judul berita clickbait.

Tabel 8. Hasil Classification Report Skenario 4

Kelas	Precision	Recall	F1-Score	Support
Clickbait	85%	73%	79%	15
Non-clickbait	76%	87%	81%	15

4. Evaluasi Model Skenario 5

Selanjutnya pada skenario 5 akan dilakukan pengujian ¹⁵klasifikasi judul berita menggunakan algoritma multinomial naive bayes dan optimasi grid search. Pengujian ini dilakukan dengan membagi ¹¹data menggunakan perbandingan 80% untuk data latih dan 20% untuk data uji. Berikut merupakan hasil tabel confusion matrix pada skenario 5.

Tabel 9. Hasil Confussion Matrix Skenario 5

Total Data	Clickbait	Non-clickbait	Data Salah	Akurasi
60	23	21	16	73%

Dari hasil klasifikasi, menunjukkan bahwa model memiliki tingkat akurasi sebesar 73%. Berdasarkan nilai presisi masing-masing kelas yang diperoleh, untuk kelas non-clickbait (0) mendapatkan 72%, sementara untuk kelas clickbait (1) mendapatkan 74%. Hasil ini menunjukkan bahwa berdasarkan 29 total data yang ada, 21 data yang benar-benar non-clickbait berhasil diprediksi, sedangkan 8 data lainnya salah diprediksi. Lalu, untuk kelas clickbait (1) dari 31 total data yang ada, 23 data yang benar-benar clickbait berhasil diprediksi, sedangkan 8 data lainnya salah diprediksi. Kemudian untuk recall setiap kelas mendapatkan hasil sebesar 72% untuk non-clickbait, dan 74% untuk clickbait. Hasil ini menunjukkan bahwa berdasarkan 29 total data yang ada seharusnya

non-clickbait, terdapat 21 data yang diketahui oleh model dan 8 data yang lain tidak diketahui oleh model. Lalu, berdasarkan 31 data yang ada seharusnya clickbait, terdapat 23 data yang berhasil diketahui oleh model dan 8 data lainnya tidak diketahui oleh model. Setelah didapatkan nilai dari precision dan recall, hasil lain didapatkan pada **F1-Score** yang merupakan nilai rata-rata harmonik dari precision dan recall. Untuk hasil **F1-Score** pada skenario 5 ini sebesar 72% untuk judul berita non-clickbait dan 74% untuk judul berita clickbait.

Tabel 10. Hasil Classification Report Skenario 5

Kelas	Precision	Recall	F1-Score	Support
Clickbait	72%	72%	72%	29
Non-clickbait	74%	74%	74%	31

5. Evaluasi Model Skenario 6

Selanjutnya pada skenario 6 akan dilakukan pengujian klasifikasi judul berita menggunakan algoritma multinomial naive bayes dan optimasi grid search. Pengujian ini dilakukan dengan membagi data menggunakan perbandingan 70% untuk data latih dan 30% untuk data uji. Berikut merupakan hasil tabel confusion matrix pada skenario 6.

Tabel 11. Hasil Confussion Matrix Skenario 6

Total Data	Clickbait	Non-clickbait	Data Salah	Akurasi
90	32	36	22	75%

Dari hasil klasifikasi, menunjukkan bahwa model memiliki tingkat akurasi sebesar 75%. Berdasarkan nilai presisi masing-masing kelas yang diperoleh, untuk 60 kelas non-clickbait (0) mendapatkan 75%, sementara untuk kelas clickbait (1) mendapatkan 76%. Hasil ini menunjukkan bahwa berdasarkan 46 total data yang ada, 34 data yang benar-benar non-clickbait berhasil diprediksi, sedangkan 12 data lainnya salah diprediksi. Lalu, untuk kelas clickbait (1) dari 44 total data yang ada, 33 data yang benar-benar clickbait berhasil diprediksi, sedangkan 11 data lainnya salah diprediksi. Kemudian untuk recall setiap kelas mendapatkan hasil sebesar 78% untuk non-clickbait, dan 73% untuk clickbait. Hasil ini menunjukkan bahwa berdasarkan 46 total data yang ada seharusnya non-clickbait, terdapat 105 data yang diketahui oleh model dan 104 data yang lain tidak diketahui oleh model. Lalu, berdasarkan 44 data yang ada seharusnya clickbait, terdapat 36 data yang berhasil diketahui oleh model dan 8 data lainnya tidak diketahui oleh model. Setelah didapatkan nilai dari precision dan recall, hasil lain didapatkan pada **F1-Score** yang merupakan nilai rata-rata harmonik dari

precision dan recall. Untuk hasil F1-Score pada skenario 6 ini sebesar 77% untuk judul berita non-clickbait dan 74% untuk judul berita clickbait

Tabel 12. Hasil Classification Report Skenario 6

Kelas	Precision	Recall	F1-Score	Support
Clickbait	76%	73%	74%	46
Non-clickbait	75%	78%	77%	44

Analisis Hasil

Setelah melakukan klasifikasi dan evaluasi model, selanjutnya dilakukan analisis hasil dari pengujian model berdasarkan skenario yang telah dibuat. Proses ini dilakukan untuk mengidentifikasi perbedaan dan pengaruh metode optimasi Grid Search pada klasifikasi judul berita clickbait menggunakan Multinomial Naïve Bayes.

Tabel 13. Analisis Hasil

No.	Aspek	Data Latih	Data Uji	Precision	Recall	F1-Score	Akurasi
1	MNB	90%	10%	87%	87%	87%	87%
2		80%	20%	73%	73%	73%	73%
3		70%	30%	76%	76%	76%	76%
4	MNB + GridSearch	90%	10%	81%	80%	80%	80%
5		80%	20%	73%	73%	73%	73%
6		70%	30%	76%	75%	76%	76%

Berdasarkan tabel 4.1, pengujian pada skenario 1, 2, dan 3 diproses dengan model Multinomial Naïve Bayes tanpa metode optimasi. Pengujian MNB dengan rasio data latih dan data uji sebesar 90%:10% menghasilkan akurasi 87% dan F1-Score 87%. Selanjutnya, pengujian MNB dengan rasio data latih dan data uji sebesar 80%:20% menghasilkan akurasi 73% dan F1-Score 73%. Sementara itu, pengujian MNB dengan rasio data latih dan data uji sebesar 70%:30% menghasilkan akurasi 76% dan F1-Score 76%. Selanjutnya pada skenario 4, 5, dan 6, dilakukan pengujian model menggunakan Multinomial Naïve Bayes dengan optimasi Grid Search. Untuk rasio data latih dan data uji sebesar 90%:10%, nilai akurasi yang diperoleh adalah 80% dengan F1-Score sebesar 80%. Pengujian dengan rasio data latih dan data uji sebesar 80%:20% menghasilkan akurasi 73% dan F1-Score 73%. Sedangkan, untuk rasio data latih dan data uji sebesar 70%:30%, model MNB dengan optimasi Grid Search menghasilkan akurasi 76% dan F1-Score 76%.

5. KESIMPULAN DAN SARAN

Berdasarkan pada rumusan masalah dan penelitian yang telah dilakukan dapat disimpulkan bahwa Algoritma Multinomial Naive Bayes dapat diimplementasikan dengan baik untuk mengklasifikasikan judul berita clickbait dan non-clickbait. Hasil pengujian pada berbagai skenario menunjukkan bahwa model Multinomial Naive Bayes memiliki tingkat akurasi yang konsisten di sekitar 78%. Pada skenario 1 (90% data latih, 10% data uji), skenario 2 (80% data latih, 20% data uji), dan skenario 3 (80% data latih, 20% data uji), model berhasil memprediksi dengan benar sebagian besar data, dengan tingkat kesalahan prediksi yang cukup rendah. Optimasi menggunakan Grid Search tidak menghasilkan peningkatan akurasi. Meskipun tidak mengalami peningkatan akurasi, terdapat perbaikan dalam nilai recall untuk kelas non-clickbait dari 76% menjadi 80%, hal ini memberikan jawaban bahwa model menjadi lebih baik dalam mengenali berita non-clickbait setelah optimasi. Disarankan untuk memperluas penelitian dengan menggunakan dataset yang lebih besar dan lebih beragam. Hal ini dapat membantu meningkatkan generalisasi model dan memastikan bahwa model dapat bekerja efektif pada berbagai jenis berita dan sumber

DAFTAR REFERENSI

- Afandi, W., Saputro, S. N., Kusumaningrum, A. M., Ardiansyah, H., Kafabi, M. H., & Sudianto, S. (2022). Klasifikasi Judul Berita Clickbait menggunakan RNN-LSTM. *Jurnal Pengembangan IT*, 85-89.
- Alhaq, Z., Mustopa, A., Mulyatun, S., & Santoso, J. D. (2021). Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter. *Jurnal Sistem Informasi dan Manajemen*, 44-49.
- Annur, C. M. (2023, Juni Jumat). Databooks. Diambil kembali dari databooks: <https://databoks.katadata.co.id/datapublish/2023/06/16/meski-trennya-turun-media-online-tetap-jadi-sumber-berita-utama-masyarakat-indonesia>
- Betseda. (2020). Peningkatan Optimasi Sentimen dalam Pelaksanaan Proses Pemilihan Presiden Berdasarkan Opini Publik dengan Menggunakan Algoritma Naïve Bayes dan Particle Swarm Optimization. *pp. 101-114*.
- Cios, K. J., Pedrycz, W., S., R., W., & Kurgan, L. A. (2007). *Data Mining A Knowledge Discovery Approach*. Springer.
- Dewi, F. K., & Aji, T. P. (2021). Klasifikasi Berita Menggunakan Metode Multinomial Naïve Bayes. *SCAN*, 1-8.
- Gupta, V., & Lehal, G. S. (2009). A Survey of Text Mining Techniques and Application. *Journal of Emerging*.
- Hadiyat, Y. D. (2019). Clickbait di Media Online Indonesia. *Pekommas*, 1-10.

- Heliyanti, S. (2022). Penerapan Model Klasifikasi Metode Naive Bayes Terhadap 38 Penggunaan Akses Internet. *Jurnal Riset Sistem Informasi Dan Teknologi Informasi*.
- Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. (2018). Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 306-312.
- Hovipah, M., Hearani, E., Jasril, & Syafria, F. (2023). Klasifikasi Clickbait Menggunakan Transformers. *Jurnal Computer Science and Information Technology*, 172-181.
- Indraloka, D. S., & Santosa, B. (2017). Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia. *Jurnal Sains dan Seni ITS*, 2337-3520.
- Kusumanongrat, H., & Kusumaningrat, P. (2005). *Jurnalistik: Teori dan Praktik*. Bandung: Remaja Rosdakarya.
- Maulidi, R., Ayilillahi, M. F., Isyiriyah, L., & Palandi, J. F. (2018). PENERAPAN NEURAL NETWORK BACKPROGPGATION UNTUK KLASIFIKASI ARTIKEL CLICKBAIT. *Seminar Nasional FST*, 751-757.
- Mihalcea, R. (2007). Advanced Approaches to Analyzing Unstructured Data. Dalam R. Feldman, & J. Sanger, *The Text Mining Handbook* (hal. 125-127). Cambridge: Computational Linguistics.
- Mustofa, H., & Mahfudh, A. A. (2019). Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes. *Walisongo Journal of Information Technology*.

Implementasi Metode Multinomial Naive Bayes dalam Klasifikasi Judul Berita Clickbait

ORIGINALITY REPORT

20%
SIMILARITY INDEX

17%
INTERNET SOURCES

10%
PUBLICATIONS

6%
STUDENT PAPERS

PRIMARY SOURCES

1 jurnal.untan.ac.id 2%
Internet Source

2 Submitted to Universitas Muria Kudus 2%
Student Paper

3 repository.its.ac.id 1%
Internet Source

4 123dok.com 1%
Internet Source

5 ejurnal.umri.ac.id 1%
Internet Source

6 Submitted to Universitas Sebelas Maret 1%
Student Paper

7 journal.isas.or.id 1%
Internet Source

8 repository.ub.ac.id 1%
Internet Source

9 www.researchgate.net 1%
Internet Source

10	ejournal.poltektegal.ac.id Internet Source	1 %
11	publishing-widyagama.ac.id Internet Source	1 %
12	ejournal.uniramalang.ac.id Internet Source	<1 %
13	journal.widyakarya.ac.id Internet Source	<1 %
14	repository.ubharajaya.ac.id Internet Source	<1 %
15	repository.unsri.ac.id Internet Source	<1 %
16	text-id.123dok.com Internet Source	<1 %
17	dspace.uii.ac.id Internet Source	<1 %
18	ejurnal.itats.ac.id Internet Source	<1 %
19	journal.lppspsemarang.org Internet Source	<1 %
20	Taslim Taslim, Susi Handayani, Fajrizal Fajrizal. "Kinerja Komparatif Optimasi Algoritma Naive Bayes dalam Klasifikasi Teks untuk Uji Klinis Kanker", Jurnal Eksplora Informatika, 2023 Publication	<1 %

-
- 21 Ernianti Hasibuan, Elmo Allistair Heriyanto. "ANALISIS SENTIMEN PADA ULASAN APLIKASI AMAZON SHOPPING DI GOOGLE PLAY STORE MENGGUNAKAN NAIVE BAYES CLASSIFIER", Jurnal Teknik dan Science, 2022
Publication <1 %
-
- 22 Tania Puspa Rahayu Sanjaya, Ahmad Fauzi, Anis Fitri Nur Masruriyah. "Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma naive bayes dan support vector machine", INFOTECH : Jurnal Informatika & Teknologi, 2023
Publication <1 %
-
- 23 Submitted to UIN Sultan Syarif Kasim Riau
Student Paper <1 %
-
- 24 ejournal.bsi.ac.id
Internet Source <1 %
-
- 25 bright-journal.org
Internet Source <1 %
-
- 26 databoks.katadata.co.id
Internet Source <1 %
-
- 27 jurnal.umb.ac.id
Internet Source <1 %
-
- 28 media.neliti.com
Internet Source <1 %
-

29

Internet Source

<1 %

30

Submitted to Sriwijaya University

Student Paper

<1 %

31

jip.polinema.ac.id

Internet Source

<1 %

32

Rizky Aziz, Tresna Maulana Fahrudin, Wahyu Syaifullah Jauharis Saputra. "Analisis Sentimen Kepuasan Pengguna OYO DiPlaystore Dengan Multinoial Naive Bayes dan Chi-square", JURNAL FASILKOM, 2024

Publication

<1 %

33

repository.uin-suska.ac.id

Internet Source

<1 %

34

Mohammad Liyananta Septipalan, Muhammad Shata' Hibrizi, Nurun Latifah, Rosa Lina, Fitri Bimantoro. "Klasifikasi Tumor Otak Menggunakan CNN Dengan Arsitektur Resnet50", Seminar Nasional Teknologi & Sains, 2024

Publication

<1 %

35

conference.upnvj.ac.id

Internet Source

<1 %

36

openlibrarypublications.telkomuniversity.ac.id

Internet Source

<1 %

37	Andika William, Yunita Sari. "CLICK-ID: A novel dataset for Indonesian clickbait headlines", Data in Brief, 2020	<1 %
Publication		

38	Rahadi Ramlan, Neva Satyahadewi, Wirda Andani. "Analisis Sentimen Pengguna Twitter Menggunakan Support Vector Machine Pada Kasus Kenaikan Harga BBM", Jambura Journal of Mathematics, 2023	<1 %
Publication		

39	doku.pub	<1 %
Internet Source		

40	ejournal.upnvj.ac.id	<1 %
Internet Source		

41	journal.ipb.ac.id	<1 %
Internet Source		

Exclude quotes On

Exclude matches Off

Exclude bibliography On