



Classification of Online News Using The Multinomial Naïve Bayes

Klasifikasi Judul Berita Online Menggunakan Multinomial Naïve Bayes

¹Lilis Laome, ²Dian Christien Arisona, ³Zelianti, ⁴Gusti Ngurah Adhi Wibawa, ⁵Irma Yahya, ⁶Mukhsar

^{1,2,3,4,5,6}Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Halu Oleo

Jl. H.E.A Mokodompit, Kota Kendari, Sulawesi Tenggara, Indonesia

Email: darisona@uho.ac.id

Article History:

Submitted: 17-12-2024; Received in Revised: 18-02-2025; Accepted: 23-03-2025

Abstract

News is defined as stories or information about current events. News can be classified into two categories, namely hard news and soft news. Generally, hard news refers to news topics that are timely, important and consequential. Meanwhile, soft news refers to information that is interesting, unique and entertaining. This research investigates online news headline classification using an enhanced Multinomial Naïve Bayes approach combined with SMOTE (Synthetic Minority Oversampling Technique) to address class imbalance issues. The results showed that out of 4,122 data collected, 3821 or 92.70% of news headlines were classified as hard news categories and 301 or 7.30% of other news headlines were classified as soft news categories. With an accuracy value of 91.5%, precision of 96.8%, recall of 93.8%, F1-Score of 95.2% and AUC value of 0.78 which shows Multinomial Naïve Bayes is good enough in distinguishing hard news and soft news categories, although it has not reached the optimal level.

Keywords: AUC; Multinomial Naïve Bayes; Online News; SMOTE.

Abstrak

Berita didefinisikan sebagai cerita atau informasi tentang peristiwa terkini. Berita dapat diklasifikasikan ke dalam dua kategori, yaitu berita keras (hard news) dan berita lunak (soft news). Umumnya, berita keras mengacu pada topik berita yang tepat waktu, penting, dan konsekuen. Sementara itu, berita lunak mengacu pada informasi yang menarik, unik, dan menghibur. Penelitian ini menyelidiki klasifikasi judul berita online menggunakan pendekatan Multinomial Naïve Bayes yang disempurnakan yang dikombinasikan dengan SMOTE (Synthetic Minority Oversampling Technique) untuk mengatasi masalah ketidakseimbangan kelas. Hasil penelitian menunjukkan bahwa dari 4.122 data yang dikumpulkan, 3821 atau 92,70% judul berita diklasifikasikan sebagai kategori hard news dan 301 atau 7,30% judul berita lainnya diklasifikasikan sebagai kategori soft news. Nilai akurasi sebesar 91.5%, precision sebesar 96.8%, recall sebesar 93.8%, F1-Score sebesar 95.2% dan nilai AUC sebesar 0.78 yang menunjukkan Multinomial Naïve Bayes sudah cukup baik dalam membedakan kategori hard news dan soft news, meskipun belum mencapai tingkat optimal.

Kata Kunci: AUC; Berita Online; Multinomial Naïve Bayes; SMOTE.

Pendahuluan

Platform digital saat ini menawarkan informasi yang tersedia dalam jumlah yang besar termasuk berita, artikel ilmiah, dan blog. Menurut KBBI, berita didefinisikan sebagai cerita atau informasi mengenai kejadian yang sedang terjadi. Berita secara umum adalah informasi yang dialirkan dari berbagai sumber, definisinya sangat banyak namun dapat ditarik beberapa kesimpulan tentang arti berita ¹. Berita selalu menjadi kebutuhan yang diperlukan bagi masyarakat awam untuk memiliki pengetahuan mengenai apa yang sedang terjadi di sekitarnya. Setiap harinya, puluhan hingga ratusan berita *online* dipublikasikan, sehingga pencarian informasi yang relevan menjadi tantangan tersendiri. Banyaknya sumber berita, mulai dari portal berita ternama hingga blog pribadi, menyulitkan pembaca dalam menentukan mana yang kredibel dan sesuai dengan kebutuhan mereka.

Klasifikasi merupakan proses mengidentifikasi kategori yang sesuai untuk observasi baru berdasarkan data latih yang telah ada. Klasifikasi adalah bentuk analisis yang mengekstrak model untuk menggambarkan kelas data yang penting. Klasifikasi data terdiri dari proses dua tahap. Tahap pertama adalah tahap pelatihan, di mana algoritma klasifikasi dibuat mempelajari data pelatihan dan kemudian direferensikan dalam bentuk aturan klasifikasi. Sedangkan tahap kedua adalah data uji diaplikasikan untuk mengukur akurasi aturan klasifikasi pada data pelatihan².

Multinomial Naive Bayes (MNB) merupakan metode yang memanfaatkan peluang serta termasuk dalam pembelajaran yang diawasi dalam kasus klasifikasi teks ³. MNB adalah metode klasifikasi yang mengacu pada teorema Bayes yang digunakan untuk mengelompokkan data dengan fitur diskrit dan memodelkan berapa kali sebuah kata muncul dalam sebuah dokumen ⁴. Klasifikasi kategori sebuah dokumen dalam MNB bergantung pada jumlah kata yang ada dalam dokumen dan dipengaruhi oleh seberapa sering

¹ I Made Riartha Prawira, Adiwijaya, and Mohamad Syahrul Mubarak, 'Klasifikasi Multi-Label Pada Topik Berita Berbahasa Indonesia Menggunakan Multinomial Naive Bayes', in *Proceeding of Engineering*, vol. 5, 2018, 7774-81, <https://doi.org/oai:openlibrary.telkomuniversity.ac.id:18.04.2507>.

² Umbar Riyanto, 'Analisis Perbandingan Algoritma Naive Bayes Dan Support Vector Machine Dalam Mengklasifikasikan Jumlah Pembaca Artikel Online', *JIKA (Jurnal Informatika)* 2, no. 2 (9 October 2019): 62-72, <https://doi.org/10.31000/v2i2.1521>.

³ Dicky Satria Mahendra, Basuki Rahmat, and Retno Mumpuni, 'Implementasi Metode Multinomial Naive Bayes Dalam Klasifikasi Judul Berita Clickbait', *Neptunus: Jurnal Ilmu Komputer Dan Teknologi Informasi* 2, no. 3 (18 July 2024): 303-16, <https://doi.org/10.61132/neptunus.v2i3.249>.

⁴ Taslim Taslim, Susi Handayani, and Fajrizal Fajrizal, 'Kinerja Komparatif Optimasi Algoritma Naive Bayes dalam Klasifikasi Teks untuk Uji Klinis Kanker', *Jurnal Eksplorasi Informatika* 13, no. 1 (30 September 2023): 113-23, <https://doi.org/10.30864/eksplora.v13i1.994>.

kata-kata tersebut muncul ⁵. Efektivitas MNB dalam klasifikasi teks terutama berasal dari kemampuannya memanfaatkan informasi frekuensi kemunculan kata sebagai dasar perhitungan probabilitas, sehingga menghasilkan akurasi yang lebih baik dibanding pendekatan yang hanya mempertimbangkan keberadaan kata.

Penelitian yang dilakukan Prawira tentang klasifikasi multi-label pada berita berbahasa Indonesia dengan MNB menunjukkan bahwa nilai *hamming loss* yang diperoleh adalah 0.18 dan hasil ini mengindikasikan bahwa metode MNB efektif dalam menangani masalah klasifikasi teks dalam konteks multi-label ⁶. Dalam penelitian yang dilakukan oleh Sabrani, MNB diterapkan untuk mengklasifikasikan artikel daring mengenai gempa di Indonesia. Hasil yang diperoleh menunjukkan *F-measure* tertinggi sebesar 95.20%, yang dicapai melalui skenario pengujian yang menyatukan fitur unigram dan bigram ⁷. Kemudian, penelitian Sanrilla yang melakukan analisis sentimen pengguna terhadap aplikasi Shopee dengan menerapkan algoritma MNB diperoleh nilai akurasi 91% yang menunjukkan bahwa metode MNB terbukti berfungsi dalam melakukan klasifikasi teks ⁸. Sementara itu, Rahman melakukan klasifikasi berita *online* dengan menerapkan MNB diperoleh hasil untuk pembobotan TF-IDF menghasilkan akurasi sebesar 94,29% sedangkan untuk pembobotan DF-*Thresholding* mencapai 92,38% ⁹. Lalu, penelitian terbaru oleh Mahendra yang mengklasifikasikan judul berita *clickbait* menggunakan MNB memperoleh nilai akurasi yang konsisten sekitar 78% sehingga algoritma MNB dapat digunakan untuk mengklasifikasikan judul berita *clickbait* ¹⁰.

Oleh karena itu, dengan mengklasifikasikan berita *online* berdasarkan judulnya, penelitian ini diharapkan dapat menjadi solusi dalam membantu pembaca memahami konteks, menilai tujuan, serta mengantisipasi dampak dari informasi yang mereka terima.

⁵ Lydia Mayasari and Dina Indarti, 'Klasifikasi Topik Tweet Mengenai Covid Menggunakan Metode Multinomial Naive Bayes Dengan Pembobotan TF-IDF', *Jurnal Ilmiah Informatika Komputer* 27, no. 1 (2022): 43–53, <https://doi.org/10.35760/ik.2022.v27i1.6184>.

⁶ Prawira, Adiwijaya, and Mubarok, 'Klasifikasi Multi-Label Pada Topik Berita Berbahasa Indonesia Menggunakan Multinomial Naive Bayes'.

⁷ Alif Sabrani, I Gede Wirarama Wedashwara W, and Fitri Bimantoro, 'Multinomial Naïve Bayes Untuk Klasifikasi Artikel Online Tentang Gempa Di Indonesia', *Jurnal Teknologi Informasi, Komputer, Dan Aplikasinya (JTIKA)* 2, no. 1 (31 March 2020): 89–100, <https://doi.org/10.29303/jtika.v2i1.87>.

⁸ Sanrilla et al., 'Analisis Sentimen Masyarakat Terhadap Toko Online Aplikasi Shopee Menggunakan Metode Multinomial Naïve Bayes', *Jurnal Matematika Komputasi Dan Statistika* 2, no. 2 (24 August 2022): 68–75, <https://doi.org/10.33772/jmks.v2i2.9>.

⁹ Amelia Rahman, Wiranto Wiranto, and Afrizal Doewes, 'Online News Classification Using Multinomial Naive Bayes', *ITSMArt: Jurnal Teknologi dan Informasi* 6, no. 1 (10 August 2017): 32–38, <https://doi.org/10.20961/itsmart.v6i1.11310>.

¹⁰ Mahendra, Rahmat, and Mumpuni, 'Implementasi Metode Multinomial Naive Bayes Dalam Klasifikasi Judul Berita Clickbait'.

Metode

Data yang digunakan merupakan data sekunder yang diperoleh melalui metode *scraping*. Data yang dikumpulkan berupa judul berita *online* sebanyak 4122 judul mulai dari bulan Januari 2024 sampai September 2024. Variabel dalam penelitian ini adalah teks yang berisi judul berita *online* yang telah dipublikasikan. Variabel ini merupakan input utama yang dianalisis menggunakan algoritma Multinomial Naive Bayes.

Tabel 1. Variabel Penelitian

Variabel	Skala Data	Definisi Operasional
X	Rasio	Bobot kata dasar yang terdapat dalam dataset dokumen, hasil pembobotan TF-IDF.
Y	Nominal	Kategori berita <i>hard news</i> atau <i>soft news</i> .

Penelitian ini dilakukan dengan beberapa tahapan menggunakan *Python* sebagai dalam proses klasifikasi. Uraian tahapan tersebut antara lain:

- 1) Melakukan pengumpulan data dengan teknik *scraping* yang menghasilkan 4122 data judul berita *online* yang kemudian disimpan dalam penyimpanan lokal.
- 2) Melakukan *preprocessing* data agar data yang akan dianalisis menjadi berkualitas. Tahapannya seperti data *cleansing*, tokenisasi, *stopword removal* dan *stemming*.
- 3) Melakukan pelabelan data secara manual dengan mempertimbangkan nuansa, konteks, dan makna yang terdapat dalam data teks.
- 4) Melakukan eksplorasi pada data untuk mendapatkan informasi yang terdapat dalam dataset. Visualisasi yang digunakan berupa diagram batang dan *wordcloud*.
- 5) Melakukan pembagian dataset, yang membagi jumlah keseluruhan data menjadi dua bagian menjadi data *training* dan data *testing* dengan rasio sebesar 80% dan 20%.
- 6) Melakukan pembobotan kata dengan TF-IDF, proses ini melakukan transformasi hasil data judul berita *online* yang telah melewati rangkaian *preprocessing* menjadi data numerik.
- 7) Melakukan penambahan data sintesis untuk mengatasi *imbalanced* data dengan menggunakan teknik SMOTE.
- 8) Melakukan klasifikasi menggunakan algoritma MNB untuk membedakan teks menjadi kategori *hard news* atau *soft news*.
- 9) Melakukan evaluasi pemodelan menggunakan *confusion matrix*, yaitu nilai *accuracy*, nilai *precision*, nilai *recall*, nilai *F1-score* dan nilai AUC.

Hasil dan Diskusi

1. Data Preprocessing

Preprocessing adalah tahap untuk mengubah data sehingga berada dalam format seharusnya dan dapat diproses¹¹. *Preprocessing* dimulai dengan data *cleansing* untuk menghilangkan atau membersihkan data dari *noise*, kemudian tokenisasi untuk memecah kalimat menjadi token-token, dilanjutkan penghapusan *stopword* untuk menghilangkan kata-kata yang tidak bermakna dan *stemming* untuk mengubah kata menjadi bentuk dasar.

Tabel 2. *Data Preprocessing*

<i>Before data preprocessing</i>	<i>After data preprocessing</i>
Kala Ganjar Tanya Influencer: Mau Makan Siang Gratis atau Internet Gratis?	[influencer, makan, siang, gratis, internet, gratis]
Penumpang KRL Capai 726 Ribu Saat Perayaan Tahun Baru 2024	[penumpang, krl, capai, ribu, perayaan]
Siber Bareskrim Tangkap Pemilik Akun TikTok @presiden_ono_niha	[siber, bareskrim, tangkap, pemilik, akun, tiktok, presidenononiha]

2. Pelabelan Data

Pelabelan kategori judul berita *online* dilakukan secara manual pada 4122 data. Kategori diberikan secara langsung pada data judul berita dengan memperhatikan konteks dan makna yang terkandung dalam judul berita.

Tabel 3. Kategori Judul Berita *Online*

Kategori	Jumlah
<i>Hard News</i>	3821
<i>Soft News</i>	301
Total	4122

3. Eksplorasi Data

Eksplorasi data yang dilakukan berupa *wordcloud* untuk mengidentifikasi kata-kata yang paling banyak muncul serta menjadi poin utama dalam judul berita *online*. Lihat Gambar 1.



Gambar 1. (a) *Wordcloud* kategori *hard news*; (b) *Wordcloud* kategori *soft news*

¹¹ Sheila Shevira, I. Made Agus Dwi Suarjaya, and Putu Wira Buana, 'Pengaruh Kombinasi Dan Urutan Pre-Processing Pada Tweets Bahasa Indonesia', *JITTER : Jurnal Ilmiah Teknologi Dan Komputer* 3, no. 2 (12 July 2022): 1074–81, <https://doi.org/10.24843/JTRTI.2022.v03.i02.p06>.

Gambar 1.a menyajikan *word cloud* dari kategori *hard news*. Kata-kata yang menonjol, antara lain "polisi," "duga," "tewas," "KPU," "tangkap," "korban," dan "bakar," mengindikasikan karakteristik utama berita *hard news*, yaitu laporan tentang peristiwa-peristiwa aktual yang berdampak besar, menarik perhatian luas, dan membutuhkan publikasi segera. Sedangkan Gambar 1.b menyajikan *word cloud* untuk kategori *soft news* yang didominasi oleh kata-kata seperti "bareng," "momen," "cerita," "lebaran," "gelar," "gaspoll," "bro," dan "libur". Pilihan kata ini mengindikasikan bahwa berita dalam kategori ini cenderung berfokus pada peristiwa yang menarik, unik, dan menghibur.

4. Pembagian Data

Split data merupakan metode yang digunakan untuk membagi dataset dimana sebagai salah satu dari beberapa aspek yang mempengaruhi seberapa baik kinerja model klasifikasi pada algoritma pembelajaran mesin ¹². Dalam penelitian ini, data dibagi dengan rasio 80% untuk data *training* dan 20% untuk data *testing*. Dari total 4122 data, sebanyak 3297 dokumen digunakan sebagai data *training* dan 824 dokumen digunakan sebagai data *testing*.

5. Pembobotan dengan TF-IDF

Pembobotan TF-IDF digunakan untuk memberi pembobotan pada setiap kata dalam dokumen berdasarkan sebesar sering dalam kata tersebut muncul dalam sebuah dokumen dan keunikannya dalam korpus teks secara keseluruhan ¹³. Dari 3.297 dokumen yang digunakan sebagai data pelatihan, sebanyak 5.433 kata telah diproses menggunakan metode TF-IDF dan ditetapkan sebagai variabel independen dalam penelitian ini

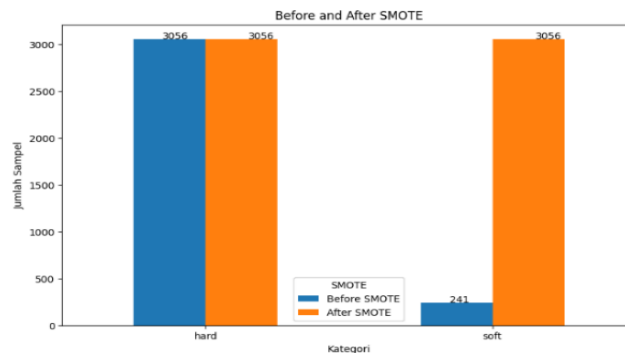
6. SMOTE (Synthetic Minority Oversampling Technique)

Dalam penelitian ini, teknik SMOTE diterapkan untuk mengatasi ketidakseimbangan kelas antara *hard news* dan *soft news* dalam data pelatihan, dengan cara meningkatkan jumlah data untuk kelas minoritas

¹² Rian Oktafiani, Arief Hermawan, and Donny Avianto, 'Pengaruh Komposisi Split Data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning', *Jurnal Sains Dan Informatika* 9, no. 1 (30 June 2023): 19–28, <https://doi.org/10.34128/jsi.v9i1.622>.

¹³ Nurul Hidayah and Dodiman Dodiman, 'Implementasi Algoritma Multinomial Naïve Bayes, TF-IDF dan Confusion Matrix dalam Pengklasifikasian Saran Monitoring dan Evaluasi Mahasiswa Terhadap Dosen Teknik Informatika Universitas Dayanu Ikhsanuddin', *Jurnal Akademik Pendidikan Matematika* 10, no. 2 (5 May 2024): 8–15, <https://doi.org/10.55340/japm.v10i1.1491>.

melalui pembangkitan data baru yang didasarkan pada k tetangga terdekat¹⁴, sebagaimana ditampilkan pada Gambar 2.



Gambar 2. Perbandingan sebelum dan sesudah SMOTE

Gambar 2 menunjukkan perbandingan jumlah sampel sebelum dan sesudah teknik SMOTE. Sebelum SMOTE, jumlah sampel pada kategori *hard news* mencapai 3056 sampel, sedangkan pada kategori *soft news* hanya 241 sampel. Namun, setelah dilakukan SMOTE, jumlah sampel pada kedua kategori tersebut menjadi seimbang, yaitu masing-masing berjumlah 3056 sampel.

7. Multinomial Naïve Bayes

Algoritma *Multinomial Naïve Bayes* menganggap kemunculan tiap kata bersifat independen sehingga tidak memperhitungkan struktur kalimat, posisi kata atau informasi kontekstual lainnya¹⁵. Klasifikasi menggunakan algoritma MNB diawali dengan pelatihan terhadap 3056 data latih, yang melibatkan proses TF-IDF dan SMOTE. Berikut ini adalah langkah-langkah dalam menerapkan algoritma MNB.

- 1) Menghitung nilai *prior probability*. *Prior probability* digunakan untuk menemukan probabilitas dokumen tertentu dari jumlah total dokumen. Dengan rumus:

$$P(c) = \frac{N_c}{N}$$

Dimana N_c adalah total dokumen di kelas c dan N adalah jumlah keseluruhan dokumen yang tersedia⁸.

- 2) Kemudian, dilanjutkan dengan menghitung nilai *laplace smoothing*. *Laplace smoothing* membantu mengatasi masalah *sparsity* pada data, di mana

¹⁴ Ni Putu Yulika Trisna Wijayanti, Eka N. Kencana, and I Wayan Sumarjaya, 'Smote: Potensi Dan Kekurangannya Pada Survei', *E-jurnal Matematika* 10, no. 4 (30 November 2021): 235–40, <https://doi.org/10.24843/MTK.2021.v10.i04.p348>.

¹⁵ Windha Mega P. Duhita and Fritz Zone, 'Perbandingan Kinerja Algoritma Multinomial dan Bernoulli Naïve Bayes dalam Mengklasifikasikan Komentar Cyberbullying', *Komputika: Jurnal Sistem Komputer* 12, no. 2 (4 October 2023): 109–17, <https://doi.org/10.34010/komputika.v12i2.9767>.

sebuah kata muncul pada satu dokumen namun tidak muncul pada dokumen lain. Rumus yang digunakan:

$$\hat{P}(t|c) = \frac{T_{ct} + 1}{\sum_{t' \in V} (T_{ct'} + 1)} = \frac{T_{ct} + 1}{(\sum_{t' \in V} T_{ct'}) + B'}$$

Dengan T_{ct} adalah frekuensi kemunculan kata t dalam dokumen latih dikelas c yang diamati, $\sum_{t' \in V} (T_{ct'})$ adalah jumlah seluruh kata dalam kelas c dan B' adalah total kosakata unik dalam seluruh data latih. Dimana $B' = |V|$ adalah jumlah istilah dalam kosakata ⁸.

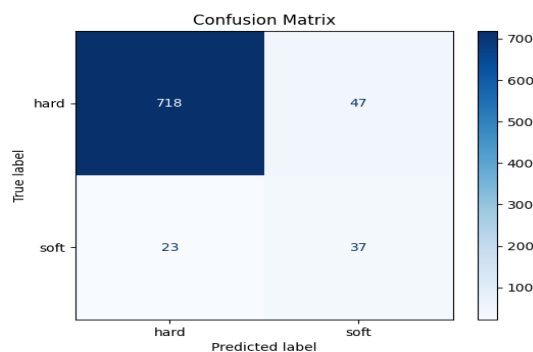
- 3) Terakhir, menghitung nilai probabilitas *posterior*. Nilai *posterior probability* ditentukan dengan menggunakan pendekatan *Maximum A Posteriori* yang telah dimaksimalkan untuk menghindari *floating point underflow*. Dengan rumus:

$$C_{MAP} = \arg \max_{c \in C} [\log \hat{P}(c) + \sum_{1 \leq k \leq n_d} \log \hat{P}(t_k | c)]$$

- 4) Menentukan kategori judul berita *online* dengan melihat nilai *posterior* tertinggi.

8. Confusion Matrix

Dalam penelitian ini, *confusion matrix* digunakan untuk menilai kinerja model dalam melakukan klasifikasi data uji, baik yang benar maupun yang salah ¹⁶, antara kategori *hard news* dan *soft news*, sebagaimana ditunjukkan pada Gambar 3.



Gambar 3. *Confusion matrix*

Berdasarkan nilai *confusion matrix* di Gambar 3, matriks evaluasi akurasi, presisi, *recall*, dan *F1-Score* dihitung.

¹⁶ Dwi Normawati and Surya Allit Prayogi, 'Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter', *J-SAKTI (Jurnal Sains Komputer Dan Informatika)* 5, no. 2 (30 September 2021): 697–711, <https://doi.org/10.30645/j-sakti.v5i2.369>.

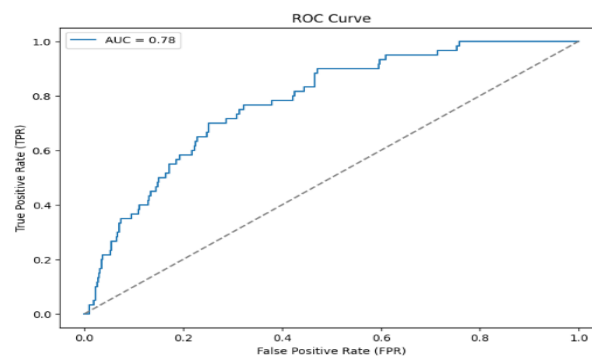
Tabel 4. Parameter Pengukuran

No.	Measurement Parameters	Value
1.	<i>Accuracy</i>	91,5%
2.	<i>Precision</i>	96,8%
3.	<i>Recall</i>	93,8%
4.	<i>F1-Score</i>	95,2%

Berdasarkan Tabel 4, metode MNB yang diterapkan mencapai nilai akurasi sebesar 91.5%. Hal ini mengindikasikan bahwa metode ini cukup mampu mengklasifikasikan data dengan baik, dengan tingkat keberhasilan yang tinggi dalam memprediksi kategori yang benar. Selain itu, nilai *precision* sebesar 96,8% menunjukkan bahwa model sangat efektif dalam mengidentifikasi kategori yang diprediksi, serta memiliki tingkat kesalahan yang rendah dalam memprediksi hard news. Nilai *recall* sebesar 93.8% menunjukkan kemampuan model untuk menemukan sebagian besar kategori *hard news*, meskipun ada beberapa yang terlewatkan. Terakhir, nilai *F1-Score* sebesar 95.2% memperlihatkan kesetaraan yang baik antara presisi dan *recall*.

9. Kurva ROC (Receiver Operating Characteristic)

Untuk melihat seberapa baik *MNB* dalam membedakan antara dua kelas kategori, disajikan kurva ROC (*Receiver Operating Characteristic*) beserta nilai AUC (*Area Under Curve*) seperti yang terlihat pada Gambar 4.



Gambar 4. Kurva ROC dan nilai AUC

Kurva ROC pada Gambar 4 menunjukkan bahwa titik-titik berada di atas garis diagonal, yang mengindikasikan hasil klasifikasi yang baik ¹⁷. Selain itu, nilai AUC (*Area Under Curve*) sebesar 0.78 menunjukkan bahwa model ini masuk ke dalam kategori klasifikasi yang cukup baik. Hal ini berarti multinomial naïve bayes memiliki kemampuan yang cukup baik dalam

¹⁷ Wickly Gusthvi, Afrionaldi A. Roza, and Caecilia Bintang Girik Allo, 'Perbandingan Metode Klasifikasi Decision Tree, Naïve Bayes, K-Nearest-Neighbor, Dan Logistic Regression Pada Dataset Phishing', *CENDRAWASIH: Journal of Statistics and Data Science* 1, no. 2 (2023): 72–76.

membedakan kategori *hard news* dan *soft news*, meskipun belum mencapai tingkat optimal.

Kesimpulan

Dari analisis terhadap 4122 judul berita *online* yang dikumpulkan antara Januari dan September 2024, ditemukan bahwa 3821 judul atau 92,70% termasuk dalam kategori *hard news*, yang menyoroti peristiwa-peristiwa signifikan dan mencuri perhatian. Sementara itu, 301 judul atau 7,30% lainnya tergolong *soft news*, yang lebih mengedepankan unsur hiburan. Metode MNB yang diterapkan dalam penelitian ini menunjukkan efektivitas dalam mengklasifikasikan judul, dengan tingkat akurasi mencapai 91,5%. Dalam hal presisi, model ini memperoleh nilai 96,8%, nilai *recall* sebesar 93,8%, dan nilai *F1-Score* mencapai 95,2%. Selain itu, AUC (*Area Under Curve*) sebesar 0,78 menunjukkan bahwa metode ini cukup baik dalam membedakan antara *hard news* dan *soft news*, meskipun masih memiliki potensi untuk perbaikan lebih lanjut. Penelitian ini masih dapat dilanjutkan dengan penambahan jumlah data yang dianalisis agar hasil analisis lebih akurat.

Daftar Pustaka

- Dhuhita, Windha Mega P., and Fritz Zone. 'Perbandingan Kinerja Algoritma Multinomial dan Bernoulli Naïve Bayes dalam Mengklasifikasikan Komentar Cyberbullying'. *Komputika : Jurnal Sistem Komputer* 12, no. 2 (4 October 2023): 109–17. <https://doi.org/10.34010/komputika.v12i2.9767>.
- Gusthvi, Wickly, Afrionaldi A. Roza, and Caecilia Bintang Girik Allo. 'Perbandingan Metode Klasifikasi Decision Tree, Naïve Bayes, K-Nearest-Neighbor, Dan Logistic Regression Pada Dataset Phishing'. *CENDRAWASIH: Journal of Statistics and Data Science* 1, no. 2 (2023): 72–76.
- Hidayah, Nurul, and Dodiman Dodiman. 'Implementasi Algoritma Multinomial Naïve Bayes, TF-IDF dan Confusion Matrix dalam Pengklasifikasian Saran Monitoring dan Evaluasi Mahasiswa Terhadap Dosen Teknik Informatika Universitas Dayanu Ikhsanuddin'. *Jurnal Akademik Pendidikan Matematika* 10, no. 2 (5 May 2024): 8–15. <https://doi.org/10.55340/japm.v10i1.1491>.
- Mahendra, Dicky Satria, Basuki Rahmat, and Retno Mumpuni. 'Implementasi Metode Multinomial Naive Bayes Dalam Klasifikasi Judul Berita Clickbait'. *Neptunus: Jurnal Ilmu Komputer Dan Teknologi Informasi* 2, no. 3 (18 July 2024): 303–16. <https://doi.org/10.61132/neptunus.v2i3.249>.
- Mayasari, Lydia, and Dina Indarti. 'Klasifikasi Topik Tweet Mengenai Covid Menggunakan Metode Multinomial Naive Bayes Dengan Pembobotan TF-IDF'. *Jurnal Ilmiah Informatika Komputer* 27, no. 1 (2022): 43–53. <https://doi.org/10.35760/ik.2022.v27i1.6184>.

- Normawati, Dwi, and Surya Allit Prayogi. 'Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter'. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)* 5, no. 2 (30 September 2021): 697–711. <https://doi.org/10.30645/j-sakti.v5i2.369>.
- Oktafiani, Rian, Arief Hermawan, and Donny Avianto. 'Pengaruh Komposisi Split Data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning'. *Jurnal Sains Dan Informatika* 9, no. 1 (30 June 2023): 19–28. <https://doi.org/10.34128/jsi.v9i1.622>.
- Prawira, I Made Riarta, Adiwijaya, and Mohamad Syahrul Mubarak. 'Klasifikasi Multi-Label Pada Topik Berita Berbahasa Indonesia Menggunakan Multinomial Naive Bayes'. In *E-Proceeding of Engineering*, 5:7774–81, 2018. <https://doi.org/oai:openlibrary.telkomuniversity.ac.id:18.04.2507>.
- Rahman, Amelia, Wiranto Wiranto, and Afrizal Doewes. 'Online News Classification Using Multinomial Naive Bayes'. *ITSMArt: Jurnal Teknologi dan Informasi* 6, no. 1 (10 August 2017): 32–38. <https://doi.org/10.20961/itsmart.v6i1.11310>.
- Riyanto, Umbar. 'Analisis Perbandingan Algoritma Naive Bayes Dan Support Vector Machine Dalam Mengklasifikasikan Jumlah Pembaca Artikel Online'. *JIKA (Jurnal Informatika)* 2, no. 2 (9 October 2019): 62–72. <https://doi.org/10.31000/v2i2.1521>.
- Sabrani, Alif, I Gede Wirarama Wedashwara W, and Fitri Bimantoro. 'Multinomial Naïve Bayes Untuk Klasifikasi Artikel Online Tentang Gempa Di Indonesia'. *Jurnal Teknologi Informasi, Komputer, Dan Aplikasinya (JTika)* 2, no. 1 (31 March 2020): 89–100. <https://doi.org/10.29303/jtika.v2i1.87>.
- Sanrilla, Natalis Ransi, La Surimi, Andi Tenriawaru Andi Tenriawaru, and La Ode Saidi. 'Analisis Sentimen Masyarakat Terhadap Toko Online Aplikasi Shopee Menggunakan Metode Multinomial Naïve Bayes'. *Jurnal Matematika Komputasi Dan Statistika* 2, no. 2 (24 August 2022): 68–75. <https://doi.org/10.33772/jmks.v2i2.9>.
- Shevira, Sheila, I. Made Agus Dwi Suarjaya, and Putu Wira Buana. 'Pengaruh Kombinasi Dan Urutan Pre-Processing Pada Tweets Bahasa Indonesia'. *JITTER: Jurnal Ilmiah Teknologi Dan Komputer* 3, no. 2 (12 July 2022): 1074–81. <https://doi.org/10.24843/JTRTI.2022.v03.i02.p06>.
- Taslim, Taslim, Susi Handayani, and Fajrizal Fajrizal. 'Kinerja Komparatif Optimasi Algoritma Naive Bayes dalam Klasifikasi Teks untuk Uji Klinis Kanker'. *Jurnal Eksplora Informatika* 13, no. 1 (30 September 2023): 113–23. <https://doi.org/10.30864/eksplora.v13i1.994>.
- Wijayanti, Ni Putu Yulika Trisna, Eka N. Kencana, and I Wayan Sumarjaya. 'Smote: Potensi Dan Kekurangannya Pada Survei'. *E-Jurnal Matematika* 10, no. 4 (30 November 2021): 235–40. <https://doi.org/10.24843/MTK.2021.v10.i04.p348>.