

Optimasi Klasifikasi Sentimen pada Komentar *Online* menggunakan Multinomial Naïve Bayes dan Ekstraksi Fitur TF-IDF serta N-grams

Optimization of Sentiment Classification on Online Comments using Multinomial Naïve Bayes and TF-IDF Feature Extraction and N-grams

Alfin Gerliandeva¹⁾, Yulison Herry Chrisnanto²⁾, Herdi Ashaury³⁾

^{1,2,3}Fakultas Sains dan Informatika, Program Studi Informatika, Universitas Jenderal Achmad Yani

^{1,2,3}Jl. Terusan Jend. Sudirman, Cibeber, Kec. Cimahi Selatan, Kota Cimahi, Jawa Barat, 40531, Telp. 022-6656190, Indonesia

alvingerliandeva20@if.unjani.ac.id¹⁾, yhc@if.unjani.ac.id²⁾, hay@if.unjani.ac.id³⁾

Diterima: 14 Maret 2024 || Direvisi: 30 Mei 2024 || Disetujui: 30 September 2024

Abstrak – Algoritma Naïve Bayes (NB) merupakan metode pengklasifikasi yang menghitung probabilitas sederhana dan cocok digunakan untuk klasifikasi teks salah satunya dalam konteks analisis sentimen. Varian klasik NB adalah Multinomial Naïve Bayes (MNB). Kelemahan algoritma MNB adalah asumsi independensi terhadap fitur. Penelitian ini menggunakan *dataset* komentar dan ulasan dari berbagai *platform online*. Penelitian ini menggunakan metode yang diusulkan dalam menangani kelemahan dari algoritma MNB yaitu penggunaan ekstraksi fitur TF-IDF dan N-grams (1-gram sampai 5-gram), dan penggunaan seleksi fitur Chi-Square, serta menangani ketidakseimbangan *dataset* menggunakan SMOTE (metode *oversampling* dan *undersampling*). Hasil penelitian ini menunjukkan bahwa penggunaan *pentagram* (5-gram) dengan data yang telah dilakukan *oversampling* SMOTE menghasilkan nilai akurasi tertinggi sebesar 94% dan nilai Area Under Curve (AUC) sebesar 100%.

Kata Kunci: Chi-Square, Komentar Online, Multinomial Naïve Bayes (MNB), N-grams, SMOTE, TF-IDF

Abstract – The Naïve Bayes (NB) algorithm is a classifier method that calculates simple probabilities and it is suitable for text classification in the context of sentiment analysis. The classic variant of NB is Multinomial Naïve Bayes (MNB). The weakness of the MNB algorithm is the assumption of feature independence. This research uses a dataset of comments and reviews from various online platforms. This study uses the proposed method to handle the weakness of the MNB algorithm, namely the use of TF-IDF feature extraction and N-grams (1-gram to 5-gram), and the use of Chi-Square feature selection, as well as handling dataset imbalance using SMOTE (oversampling and undersampling method). The results of this study show that the use of pentagram (5-gram) with data that has been oversampled by SMOTE produces the highest accuracy value of 94% and an Area Under Curve (AUC) value of 100%.

Keywords: Chi-Square, Online Comments, Multinomial Naïve Bayes (MNB), N-grams, SMOTE, TF-IDF

PENDAHULUAN

Analisis sentimen adalah domain pengetahuan yang dikenal dengan klasifikasi sentimen, atau penambangan opini, meneliti pandangan, dan sentimen terhadap berbagai hal yang direpresentasikan oleh bahasa tertulis (Taufiqi & Nugroho, 2023). Tujuan analisis sentimen merupakan untuk mengkategorikan sebuah dokumen ke dalam kelas sentimen positif, negatif, ataupun netral agar mengetahui sebuah polaritas kontekstualnya. Analisis sentimen dapat digunakan untuk menilai sebuah kebijakan, keputusan, dan produk yang bermanfaat bagi organisasi bisnis, dan masyarakat secara menyeluruh untuk mengklasifikasikan data dengan jumlah yang besar. Komentar dan ulasan dari berbagai *platform online*

merupakan contoh teks atau dokumen yang dapat dilakukan klasifikasi sentimen. *Platform* seperti Twitter (X), Zomato, TripAdvisor, Facebook, Instagram, dan Qraved merupakan sumber dari komentar - komentar serta ulasan pengguna yang dapat diperoleh dengan proses “*crawling data*” (Purwarianti & Crisdayanti, 2019).

Pada *platform* Twitter (X) data yang diambil berupa “*tweets*” atau komentar yang ada pada Twitter (X). Kemudian *platform* Zomato adalah aplikasi untuk pengiriman makanan dan pencarian restoran, sehingga data yang diambil berupa ulasan pengguna tentang pengalaman dari pemesanan dan pencarian restoran pada aplikasi Zomato tersebut. Selanjutnya *platform* TripAdvisor merupakan aplikasi untuk perjalanan atau

travel yang menyediakan pemesanan hotel dan pencarian restoran di seluruh dunia, sehingga data yang diambil berupa hasil ulasan dari pengalaman pengguna dalam menggunakan aplikasi TripAdvisor. Berikutnya *platform* Qraved merupakan aplikasi untuk pencarian restoran termasuk ulasan restoran, lokasi dan detail kontak, data yang diambil merupakan pengalaman pengguna dalam menggunakan aplikasi dari Qraved. Terakhir ada *platform* Facebook, Instagram, dan Youtube. Pada masing-masing *platform* tersebut data yang diambil berupa komentar *online* daripada pengguna saat memberikan sebuah komentar terhadap video, gambar, dan teks. *Multiple online platform* adalah *dataset* gabungan dari berbagai *platform* di atas dengan mengumpulkan keseluruhan *dataset* dari komentar dan ulasan *online* seluruh *platform* tersebut.

Satu dari algoritma populer dalam mengimplementasikan analisis sentimen yaitu Naïve Bayes Classifier (NBC). Algoritma ini termasuk dari teknik *classification* yang didasarkan dari Bayes Theorem. Karakteristik penting NBC merupakan dugaan yang kokoh (naif) terhadap independensi dari setiap situasinya (Sholehurrohman & Sabda Iman, 2022). Naïve Bayes Classifier bekerja menggunakan probabilitas bersyarat, dengan mempertimbangkan independensi bersyarat dari fitur-fiturnya. Multinomial Naïve Bayes Classifier (MNB) adalah sebuah versi modifikasi atau tingkat lanjut dari Naïve Bayes Classifier dan memiliki strategi probabilistik yang mirip dengan NBC. MNB digunakan secara khusus untuk menentukan frekuensi setiap kata dalam dokumen sebuah teks. Pemilihan algoritma Multinomial Naïve Bayes Classifier dalam penelitian ini didasarkan pada MNB merupakan versi tingkat lanjut dari Naïve Bayes Classifier yang berfungsi dengan sangat baik untuk memanipulasi jumlah kata dengan mencari tahu seberapa sering setiap kata muncul. Oleh karena itu, MNB dianggap cocok sebagai algoritma yang optimal untuk mengklasifikasi teks khususnya pada analisis sentimen (Surya et al., 2019).

Penelitian terdahulu yang menggunakan metode klasifikasi teks Multinomial Naïve Bayes (MNB) dengan ekstraksi fitur TF-IDF dan Bag of Words (BoW) menghasilkan akurasi sebesar 90%. Kelemahan algoritma MNB terletak pada asumsi independensi antar fitur (kata), yang menjadi celah penelitian untuk klasifikasi sentimen (Abbas et al., 2019). Penelitian sebelumnya yang menggunakan algoritma MNB pada klasifikasi sentimen ulasan buku bahasa Bengali. Ekstraksi fitur N-grams, meliputi *unigram*, *bigram*, dan

trigram, digunakan dalam penelitian tersebut. Hasil menunjukkan bahwa algoritma MNB dengan fitur *unigram* menghasilkan akurasi tertinggi yaitu sebesar 84%. Celah penelitian yang dapat dieksplorasi adalah penggunaan fitur N-grams yang lebih kompleks seperti *quadgram* dan *pentagram*, yang diharapkan dapat meningkatkan kinerja algoritma MNB (Hossain et al., n.d.). Penelitian lain berikutnya telah mengeksplorasi algoritma MNB untuk klasifikasi sentimen pada ulasan hotel. Ekstraksi fitur BoW dan dua metode seleksi fitur, yaitu seleksi fitur berbasis frekuensi (kemunculan fitur) dan seleksi fitur dengan menghapus fitur yang memiliki nilai perbedaan probabilitas positif dan negatif minimum, diimplementasikan dalam penelitian tersebut. Hasil penelitian menunjukkan bahwa MNB dengan seleksi fitur berbasis frekuensi (kemunculan fitur) menghasilkan *average* F1-Score yaitu 91,4% (Farisi et al., 2019). Penelitian terdahulu selanjutnya menggunakan metode seleksi fitur Chi-Square untuk menguji independensi fitur (kata) dengan kategorinya. Hasil penelitian menunjukkan bahwa penggunaan seleksi fitur Chi-Square memberikan peningkatan akurasi yang signifikan dalam klasifikasi sentimen teks dengan algoritma MNB (Ernayanti et al., 2023).

Penelitian ini menggunakan sekitar 10264 data yang didapatkan dari GitHub IndoNLU. IndoNLU adalah kumpulan sumber daya (Natural Language Understanding/NLU) untuk Bahasa Indonesia (GitHub). Data tersebut sudah memiliki label dengan ketentuan 2 *class* terdiri dari kategori: "*positive*" dan "*negative*". *Dataset* tersebut dilabeli oleh beberapa ahli bahasa dari Indonesia (Purwarianti & Crisdayanti, 2019). Data dengan kategori "*positive*" berjumlah 6624 data atau sekitar 64,5%, dan kategori "*negative*" berjumlah 3640 data atau sekitar 35,5%. Data yang diperoleh mengalami ketidakseimbangan *datasets* atau *imbalance datasets*. *Class* data positif lebih banyak jumlahnya daripada *class* data negatif. Maka dari itu, diperlukan untuk mengatasi masalah *imbalance datasets* ini memakai SMOTE. Pada penelitian yang tidak diterapkan SMOTE nilai akurasinya adalah 72%. Setelah itu teknik SMOTE diterapkan, nilai akurasinya meningkat signifikan menjadi 90% atau sebesar 18% (Anam et al., 2023). Oleh karena itu penelitian ini menggunakan metode *oversampling* dan *undersampling* dalam konteks eksperimental untuk membandingkan hasil performa klasifikasi setelah diterapkan pada algoritma MNB. Di samping itu, untuk pendekatan metode validasi silang yang sudah umum dilakukan di penelitian sebelumnya adalah metode K-

Fold Cross Validation. Metode ini terbukti mampu untuk mengoptimalkan hasil akurasi, pada penelitian sebelumnya nilai akurasi 78,19%. Setelah diterapkan K-Fold Cross Validation nilai akurasi meningkat sedikit menjadi 78,45% (Handayani et al., 2020). Selain untuk menguji agar nilai akurasi tetap objektif, K-Fold Cross Validation terbukti mampu meningkatkan nilai akurasinya.

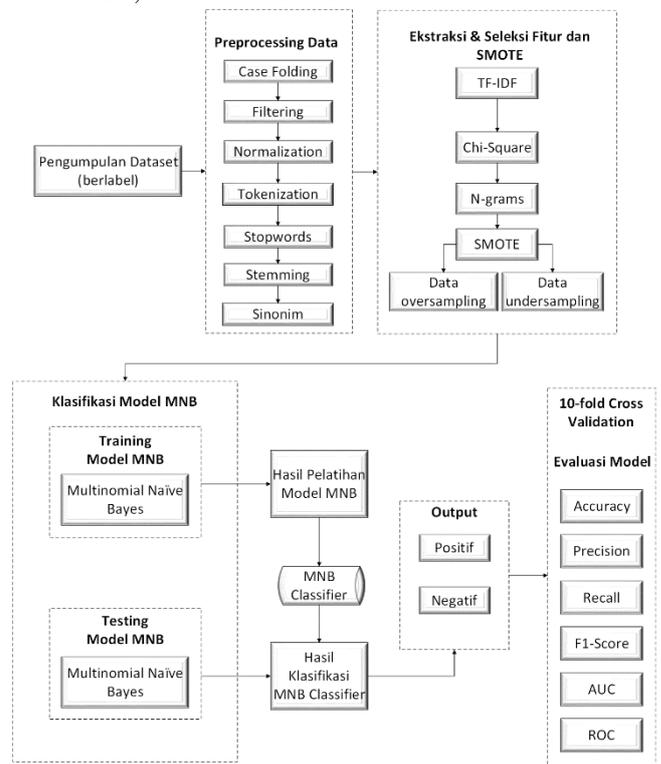
Penelitian ini bertujuan untuk peningkatan kinerja model klasifikasi MNB dengan memperbaiki kelemahan utamanya, yaitu asumsi independensi terhadap fitur. Secara khusus, penelitian ini mengusulkan pendekatan yang komprehensif. Pertama, ekstraksi fitur akan dilakukan menggunakan teknik Term Frequency-Inverse Document Frequency (TF-IDF) bersama dengan N-grams yang kompleks, seperti *unigram*, *bigram*, *trigram*, *quadgram*, dan *pentagram*. Penggunaan kombinasi ini diharapkan dapat menangkap hubungan dan konteks yang lebih dalam antara fitur-fitur (kata-kata) dalam teks, meningkatkan kemampuan model untuk memahami makna dokumen secara lebih baik. Selain itu, penelitian ini akan menerapkan seleksi fitur dengan menggunakan metode Chi-Square. Metode ini memungkinkan untuk mengidentifikasi fitur-fitur yang paling berhubungan dengan kelas target, sehingga hanya fitur-fitur yang relevan yang akan digunakan dalam pembentukan model klasifikasi. Selanjutnya, untuk menangani ketidakseimbangan *dataset* yang umumnya terjadi dalam praktik klasifikasi teks, penelitian ini akan menggunakan Synthetic Minority Over-sampling Technique (SMOTE) dengan metode *oversampling* dan *undersampling*. Hal ini diharapkan dapat mengurangi bias yang mungkin timbul akibat ketidakseimbangan tersebut dan meningkatkan keakuratan serta generalisasi model MNB. Dengan demikian hasil penelitian ini diharapkan dapat menghasilkan model klasifikasi MNB yang lebih akurat dan *robust*.

Celah penelitian yang diperoleh untuk mengatasi kelemahan algoritma MNB yaitu “asumsi independensi fitur” dengan menggunakan fitur N-grams yang lebih kompleks. Kontribusi utama dari penelitian adalah penggunaan N-grams yang lebih kompleks, seperti *quadgram* dan *pentagram* yang jarang digunakan dalam konteks penelitian klasifikasi sentimen. Dengan memperluas cakupan N-grams tersebut, diharapkan mampu menangkap lebih banyak informasi tentang hubungan kata-kata, sehingga dapat mereduksi asumsi independensi fiturnya. Kemudian penelitian ini

menerapkan pendekatan “sinonim” pada pra-proses data yaitu menghapus kata-kata yang memiliki arti sama dalam dataset, sehingga mengurangi redundansi dan memperbaiki representasi teks yang lebih relevan. Dengan demikian, dari kedua hal diatas diharapkan mampu memberikan kontribusi dalam penanganan asumsi independensi fitur dari algoritma Multinomial Naïve Bayes, sehingga mampu meningkatkan hasil dari kinerja model MNB dalam klasifikasi sentimen.

METODE PENELITIAN

Penelitian ini mengembangkan alur penelitian yang terdiri dari 5 tahap : pengumpulan *dataset*, *preprocessing data*, ekstraksi dan seleksi fitur, serta implementasi *oversampling* (SMOTE) dan *undersampling*, klasifikasi model MNB dengan *cross validation*, dan evaluasi model.

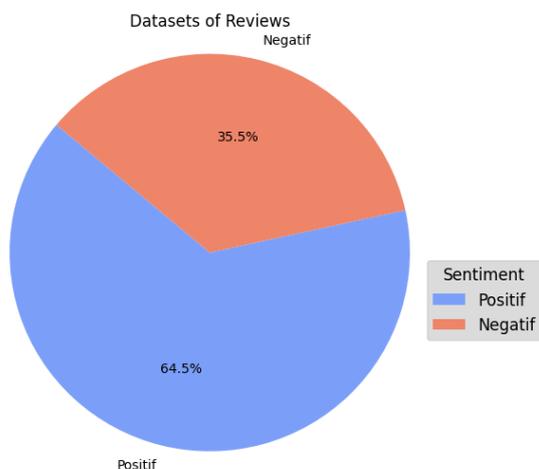


Gambar 1 Metode Penelitian

Pada Gambar 1, merupakan alur dari penelitian yang dilakukan dimulai dari tahapan pengumpulan *dataset*, *preprocessing data* (*case folding* sampai dengan *sinonim*), proses ekstraksi fitur (TF-IDF dan N-grams, proses seleksi fitur (*Chi-Square*), implementasi SMOTE dan *undersampling*. Setelah itu tahapan berikutnya adalah klasifikasi model MNB dan proses evaluasi model dengan Confusion Matrix.

Pengumpulan Dataset

Total terdapat 10264 data dari IndoNLU GitHub yang digunakan dalam penelitian ini, terdiri dari 2 kelas yaitu kategori "positif" dan "negatif". Proses pelabelan *dataset* dilakukan oleh beberapa ahli bahasa Indonesia. Terdapat 6624 data, atau sekitar 64,5%, dalam kategori "positif" dan 3640 data, atau sekitar 35,5%, dalam kategori "negatif".



Gambar 2 Rasio Dataset of Reviews

Pada Gambar 2, terdapat jumlah rasio dari *dataset reviews* yaitu sebanyak 64,5% kelas positif dan sebanyak 35,5% kelas negatif. Hal tersebut mengindikasikan bahwa *dataset* tidak seimbang, karena jumlah data positif lebih superior dibandingkan jumlah data negatif.

Preprocessing Data

Tahapan *preprocessing* data pada penelitian ini terdiri dari 7 tahap: *Case folding*, *Filtering*, *Normalization*, *Tokenization*, *Stopwords*, *Stemming*, dan *Sinonim*.

1. *Case Folding* yaitu menjadikan karakter *uppercase* pada *tweet* diubah dengan *lowercase* dengan menggunakan metode *lower()* dalam *Python*.
2. *Filtering* yaitu elemen-elemen yang dihapus dari komentar dan ulasan, yaitu: alamat situs web (<https://>, <http://>, atau www), tanda tangan pengguna (@namapengguna), gambar (<pic.twitter.com>), karakter, dan tanda baca.
3. *Normalization* adalah sebuah proses normalisasi sebuah kata "informal" dan "formal" menjadi bentuk yang dapat diterima atau normal.
4. *Tokenization* adalah proses penguraian *tweet* menjadi kata-kata disebut tokenisasi. Proses ini dilakukan dengan menggunakan *library* Python NLTK.
5. *Stopwords* adalah proses menghilangkan kata-kata yang tidak relevan dari *tweet* dengan menggunakan

daftar *stopwords* yang telah didefinisikan sebelumnya, dan mempertahankan kata-kata yang bermakna.

6. *Stemming* adalah proses mengubah kata menjadi bentuk dasarnya dengan cara mengurangi akhiran dan awalan. *Stemming* berfungsi untuk mengubah kata turunan menjadi kata dasar.
7. *Sinonim* adalah proses mengeliminasi fitur (kata) yang memiliki arti serupa, berfungsi agar kamus data lebih padat (fitur-fitur yang penting) sehingga persebaran dalam vektor kata tidak terlalu banyak.

Ekstraksi dan Seleksi Fitur serta Implementasi SMOTE dan Undersampling

TF-IDF yang digunakan adalah menghitung Term Frequency (TF) dan Inverse Document Frequency (IDF), kemudian menghasilkan skor pembobotan dari setiap fiturnya. Kemudian proses mengubah teks menjadi vektor kata dilakukan berdasarkan skor pembobotannya menggunakan TF-IDF Vectorizer. Berikut ini adalah proses perhitungan bobot menggunakan TF-IDF (Prastyo et al., 2020):

1. Persamaan untuk menghitung TF

$$TF_{(d,t)} = \frac{f_{t,d}}{\sum t, d} \quad (1)$$

Keterangan:

$\sum t, d$ Jumlah total kata dalam dokumen (d) dan $f_{t,d}$ Frekuensi kemunculan setiap kata (t) dalam dokumen (d) ditampilkan.

2. Persamaan untuk menghitung IDF

$$IDF_{(t)} = \log \frac{|D|}{f_{t,d}} + 1 \quad (2)$$

Keterangan:

$f_{t,d}$ Jumlah dokumen di mana dokumen (d) berisi teks (t), dan |D| Jumlah rekaman dalam kumpulan.

3. Persamaan untuk menghitung TF-IDF

$$W_{(d,t)} = TF_{(d,t)} \cdot IDF_{(t)} \quad (3)$$

Keterangan:

$TF_{(d,t)}$ merupakan hasil perhitungan TF, dan $IDF_{(t)}$ merupakan hasil perhitungan IDF.

Model bahasa *n-gram*, jenis model bahasa yang paling dasar 1-gram, juga dikenal sebagai *unigram* merupakan urutan satu kata. 2-gram, juga dikenal sebagai *bigram*, adalah urutan dua kata dari kata-kata seperti "putar balik," atau "pekerjaan rumah." 3-gram, juga dikenal sebagai *trigram*, adalah urutan tiga kata dari frasa seperti "tolong putar musiknya," atau

“serahkan pekerjaan rumahmu.” Di samping itu *n-gram* adalah rangkaian *n* kata. Kemudian dalam praktiknya, penggunaan model 4-gram (*quadgram*) atau bahkan 5-gram (*pentagram*) dapat diimplementasikan jika data pelatihannya mencukupi [Buku].

Fitur adalah kolom apa pun dalam kumpulan data. Kata-kata yang ditemukan dalam sebuah ulasan adalah fitur yang digunakan dalam penelitian ini. Proses memilih fitur yang relevan menggunakan kriteria untuk menghasilkan fitur yang optimal dikenal sebagai seleksi fitur. Salah satu teknik untuk seleksi fitur yang menguji independensi istilah dari kategorinya menggunakan teori statistik adalah Chi-Square. Berikut ini adalah proses perhitungan menggunakan *Chi-Square* (Ernayanti et al., 2023):

H_0 : Term/fitur independen terhadap label (*positive* dan *negative*)

H_1 : Term/fitur dependen terhadap label (*positive* dan *negative*)

Taraf signifikansi : α

$$X^2 = \sum_{i=1}^n \sum_{j=1}^m \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (4)$$

Pada baris ke-*i* dan kolom ke-*j*, O_{ij} menunjukkan nilai sebenarnya dan E_{ij} menunjukkan nilai prediksi.

Dua variabel yang diamati dalam skenario ini adalah fitur sebagai baris dan kelas (positif dan negatif) sebagai kolom. O_{ij} adalah jumlah atau nilai aktual dari dua variabel. Mencari *expected count* dari O_{ij} atau mengisi sel E_{ij} dilakukan dengan persamaan ke-5 sebagai berikut:

$$E_{ij} = \frac{b_i k_j}{N} \quad (5)$$

Keterangan Notasi:

b_i : Total perhitungan baris ke-*i*

k_j : Total perhitungan baris ke-*j*

N : Total seluruh nilai observasi

Kaidah pengambilan keputusannya yaitu : Tolak H_0 jika nilai $X^2 \geq X^2_{\alpha, (n-1)(m-1)}$ atau nilai signifikan $\leq \alpha$.

Synthetic Minority Oversampling Technique (SMOTE) adalah teknik yang digunakan untuk menyeimbangkan *dataset* dengan menghasilkan data sintetis untuk kelas minoritas, agar jumlah kelas minoritas tersebut seimbang dengan kelas mayoritas (Wibowo et al., 2021). Selanjutnya skenario implementasi SMOTE dibagi menjadi dua yaitu: *oversampling* atau proses peningkatan sample kelas

minoritas agar seimbang dengan kelas mayoritas, dan *undersampling* adalah proses penurunan sampel kelas mayoritas agar seimbang dengan kelas minoritas.

Klasifikasi Model MNB dengan *Cross Validation*

Algoritma MNB bekerja berdasarkan gagasan frekuensi istilah, yang mengacu pada frekuensi sebuah kata yang terkandung pada dokumen. Teknik tersebut menyajikan dua fakta: frekuensi sebuah kata dalam dokumen dan apakah kata tersebut muncul di dalamnya atau tidak (Singh et al., 2019). *MNB* mempertimbangkan frekuensi kata dan informasi dengan mengasumsikan bahwa dokumen adalah sekumpulan kata (Abbas et al., 2019)(Abbas et al., 2019). Algoritma MNB terbukti dan cocok digunakan pada deteksi polaritas sentimen dari ulasan buku, di mana menghasilkan nilai akurasi yang paling tinggi dibandingkan algoritma lainnya (Hossain et al., n.d.).

K-Fold Cross Validation memecah data dengan *k subset* untuk mengevaluasi kinerja model. *Subset* diulang dengan *k* kali. Kemudian, untuk iterasi awal, 1 *subset* digunakan menjadi *data testing* dari *subset* yang berbeda dipakai untuk *data training*. Selama iterasi, data pengujian dan data pelatihan ditukar. Setiap *subset* berfungsi sebagai data pengujian satu kali dan data pelatihan hingga *k-1* kali (Zul et al., 2018).

Evaluasi Model dengan *Confusion Matrix*

Confusion Matrix metode untuk mengukur dan mengevaluasi kinerja model. *Confusion Matrix* menunjukkan *value* TP, FP, FN, dan TN. Dalam konteks klasifikasi, *True Positive* merupakan klasifikasi yang tepat, yaitu sampel positif diklasifikasikan sebagai positif. *False Negative* merupakan klasifikasi yang keliru, yaitu sampel positif diperkirakan nilai *negative*. *False Positive* yaitu klasifikasi yang keliru, yaitu sampel negatif diperkirakan nilai *positive*. *True Negative* yaitu klasifikasi yang tepat, yaitu sampel negatif diklasifikasikan sebagai *negative* (Amal et al., 2022).

Accuracy algoritma dapat ditentukan dengan membandingkan proporsi data yang dikategorikan dengan benar terhadap seluruh data. Persamaan ke-6 menjelaskan metrik untuk menghitung nilai *accuracy*:

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (6)$$

Precision algoritma dapat didefinisikan sebagai proporsi data terhadap seluruh jumlah dan diperkirakan secara akurat. Persamaan ke-7 menjelaskan metrik untuk menghitung nilai *precision*:

$$Precision = \frac{(TP)}{(TP + FP)} \tag{7}$$

Recall algoritma dapat ditentukan dengan membandingkan proporsi data terhadap jumlah dan yang telah diperkirakan secara akurat. Persamaan ke-8 menjelaskan metrik untuk menghitung nilai *recall*:

$$Recall = \frac{(TP)}{(TP + FN)} \tag{8}$$

F1 Score merupakan ukuran keselarasan antara *precision* dan *recall* yang dapat digunakan sebagai ukuran tambahan. Persamaan ke-9 menjelaskan metrik untuk menghitung nilai *F1 Score*:

$$F1\ Score = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{9}$$

ROC merupakan kurva yang menunjukkan kinerja model klasifikasi dari berbagai nilai ambang batas. Persamaan ke-10 dan ke-11 menjelaskan metrik untuk menghitung nilai *ROC*:

$$TPR = \frac{TP}{(TP + FN)} \tag{10}$$

$$FPR = \frac{FP}{(FP + TN)} \tag{11}$$

AUC mengukur area di bawah *ROC*. Jika nilai *AUC* semakin tinggi, maka kinerja klasifikasi model semakin baik. Persamaan ke-12 menjelaskan metrik untuk menghitung nilai *AUC*:

$$AUC = \int_0^1 TPR\ d(FPR) \tag{12}$$

Analisis dan Perancangan

Tahapan analisis dan perancangan ini akan membahas sesuai dengan alur penelitian yang terdiri dari pengumpulan *dataset*, *preprocessing data*, ekstraksi dan seleksi fitur, serta implementasi *oversampling* (SMOTE) dan *undersampling*, klasifikasi model MNB dengan *cross validation*, dan evaluasi model.

Pengumpulan Dataset

Dataset diperoleh dari ulasan dan komentar dari *platform* Twitter (X), Zomato, TripAdvisor, Facebook, Instagram, dan Qraved. Total terdapat 10264 data dengan 6624 data, atau sekitar 64,5%, dalam kategori "positif" dan 3640 data, atau sekitar 35,5%, dalam kategori "negatif". Berikut ini adalah contoh dari *record dataset* yang digunakan dalam penelitian ini pada Gambar 3:

text	label
warung ini dimiliki oleh pengusaha pabrik tahu yang sudah puluhan tahun terkenal membuat tahu putih di Bandung tahu berkualitas	positive
lokasi strategis di jalan Sumatera Bandung tempat nya nyaman terutama sofa di lantai paella nya enak sangat pas dimakan dengan	positive
betapa bahagia nya diri ini saat unboxing paket dan barang nya bagus menetapkan beli lagi	positive
duh jadi mahasiswa jangan sombong dong kasih kartu kuning segala belajar dulu yang baik tidak usahlah ikut politik nanti sudu	negative
makanan beragam harga makanan di food stall akan ditambahkan lagi di kasir suasana ramai dan perlu perhatian untuk mendapat	positive
pakai kartu kredit bca tidak untung malah rugi besar	negative
tempat unik bagus buat foto makanan enak pegawai ramah bersih dan luas wifi kencang harga standar dan sesuai dengan tempat	positive
saya bersama keluarga baru saja menikmati pengalaman kuliner yang menyenangkan di rm sari sunda di jalan setiabudi Bandung	positive
bersyukur	positive
simcard indosat inaktif garagara lupa isi pulsa dan kabar nya aktif jika pinda ke pasca bayar ribet banget	negative
sifat iri sering muncul pada orang yang tidak punya tujuan hidup	negative
sekadar menceritakan pengalaman saya pesan steak grilled beef tidak disajikan steak tidak tersebut dalam potongan tidak tidak	positive
pengalaman bersama indosat hari ini semoga tidak terjadi pada pelanggan lain sempat sampai marahmarah dengan pelayanan pe	negative

Gambar 3 Record Datasets of Reviews

Pada Gambar 3, terdapat contoh dari *dataset of reviews* yang digunakan pada penelitian ini. Data tersebut memiliki 2 atribut yaitu "text" dan "label". Jumlah total *record* data sebanyak 10264 data.

Preprocessing Data

Pada tahapan ini merupakan proses awal yaitu melakukan *case folding*. Proses *case folding* dapat dilihat pada Tabel 1:

Tabel 1 Proses Case Folding

Reviews	Case Folding
Makan shabu-shabu sambil ngobrol sama teman-teman. Tempat nya asik buat makan. Harga nya tidak terlalu mahal. Capek banget sumpah menunggu paket dari JNE. Wifi di sini selalu eror , makanan dan minuman agak mahal , lumpiah goreng asin sekali , teh susu hanya sedikit enak , lama sekali untuk pesanan Saya biasa coba makan bento porsi cukup banyak , pilihan bento cukup beragam mulai dari salmon , ayam , daging sapi . hidangan penutup biasa saya pesan es strawberry Jenis makanan variatif banget dan tidak membosankan , yang paling menarik <i>dessert</i> nya , banyak jenis dan tampilan nya juga menarik sekali	makan shabu-shabu sambil ngobrol sama teman-teman. tempat nya asik buat makan. harga nya tidak terlalu mahal. capek banget sumpah menunggu paket dari jne. wifi di sini selalu eror , makanan dan minuman agak mahal , lumpiah goreng asin sekali , teh susu hanya sedikit enak , lama sekali untuk pesanan saya biasa coba makan bento porsi cukup banyak , pilihan bento cukup beragam mulai dari salmon , ayam , daging sapi . hidangan penutup biasa saya pesan es strawberry jenis makanan variatif banget dan tidak membosankan , yang paling menarik <i>dessert</i> nya , banyak jenis dan tampilan nya juga menarik sekali

Pada Tabel 1, merupakan tahapan *case folding* atau pengubahan seluruh teks yang memiliki huruf kapital diubah menjadi "lowercase" atau huruf kecil semuanya.

Tahap yang kedua yaitu *filtering*. Proses *filtering* dapat dilihat pada Tabel 2 :

Tabel 2 Proses *Filtering*

<i>Case Folding</i>	<i>Filtering</i>
makan shabu-shabu sambil ngobrol sama teman-teman. tempat nya asik buat makan. harga nya tidak terlalu mahal. capek banget sumpah menunggu paket dari jne. wifi di sini selalu eror , makanan dan minuman agak mahal , lumpiah goreng asin sekali , teh susu hanya sedikit enak , lama sekali untuk pesanan saya biasa coba makan bento porsi cukup banyak , pilihan bento cukup beragam mulai dari salmon , ayam , daging sapi . hidangan penutup biasa saya pesan es strawberry jenis makanan variatif banget dan tidak membosankan , yang paling menarik <i>dessert</i> nya , banyak jenis dan tampilan nya juga menarik sekali	makan shabu shabu sambil ngobrol sama teman teman tempat nya asik buat makan harga nya tidak terlalu mahal capek banget sumpah menunggu paket dari jne wifi di sini selalu eror makanan dan minuman agak mahal lumpiah goreng asin sekali teh susu hanya sedikit enak lama sekali untuk pesanan saya biasa coba makan bento porsi cukup banyak pilihan bento cukup beragam mulai dari salmon ayam daging sapi hidangan penutup biasa saya pesan es strawberry jenis makanan variatif banget dan tidak membosankan yang paling menarik <i>dessert</i> nya banyak jenis dan tampilan nya juga menarik sekali

Pada Tabel 2, adalah proses penyaringan dari keseluruhan teks seperti tanda baca titik, koma dan lain-lain. Setelah dilakukan proses *filtering*, maka datanya telah dihilangkan tanda baca seperti titik atau koma.

Tahap yang ketiga yaitu *normalization*. Proses *normalization* dapat dilihat pada Tabel 3 :

Tabel 3 Proses *Normalization*

<i>Filtering</i>	<i>Normalization</i>
sabar itu memang cape tapi kalau dibarengin ikhlas semua pasti bakal indah pada waktu nya bongkar saja pak sby sekalian toh ini negeri juga sudah hancur kok wifi di sini selalu eror makanan dan minuman agak mahal lumpiah goreng asin sekali teh susu hanya sedikit enak lama sekali untuk pesanan saya biasa coba makan bento porsi cukup banyak	sabar itu memang capek tapi kalau dibarengin ikhlas semua pasti bakal indah pada waktu nya bongkar saja pak sby sekalian toh ini negeri juga sudah hancur kok wifi di sini selalu eror makanan dan minuman agak mahal lumpiah goreng asin sekali teh susu hanya sedikit enak lama sekali untuk pesanan saya biasa coba makan bento porsi cukup banyak

<i>Filtering</i>	<i>Normalization</i>
pilihan bento cukup beragam mulai dari salmon ayam daging sapi hidangan penutup biasa saya pesan es strawberry jenis makanan variatif banget dan tidak membosankan yang paling menarik <i>dessert</i> nya banyak jenis dan tampilan nya juga menarik sekali	pilihan bento cukup beragam mulai dari salmon ayam daging sapi hidangan penutup biasa saya pesan es strawberry jenis makanan variatif banget dan tidak membosankan yang paling menarik <i>dessert</i> nya banyak jenis dan tampilan nya juga menarik sekali

Pada Tabel 3, adalah proses menormalisasikan kata yang tidak baku menjadi baku. Contohnya adalah kata “cape” setelah melalui proses *normalization*, maka kata tersebut menjadi kata “capek”.

Tahap yang keempat yaitu *tokenization*. Proses *tokenization* dapat dilihat pada Tabel 4 :

Tabel 4 Proses *Tokenization*

<i>Normalization</i>	<i>Tokenization</i>
sabar itu memang cape tapi kalau dibarengin ikhlas semua pasti bakal indah pada waktu nya	['sabar', 'itu', 'memang', 'cape k', 'tapi', 'kalau', 'dibarengin', 'ikhlas', 'semua', 'pasti', 'bakal', 'indah', 'pada', 'waktu', 'nya']
bongkar saja pak sby sekalian toh ini negeri juga sudah hancur kok	['bongkar', 'saja', 'pak', 'sby', 'sekalian', 'toh', 'ini', 'negeri', 'juga', 'sudah', 'hancur', 'kok']
wifi di sini selalu eror makanan dan minuman agak mahal lumpiah goreng asin sekali teh susu hanya sedikit enak lama sekali untuk pesanan	['wifi', 'di', 'sini', 'selalu', 'error', 'makanan', 'dan', 'minuman', 'agak', 'mahal', 'lumpiah', 'goreng', 'asin', 'sekali', 'teh', 'susu', 'hanya', 'sedikit', 'enak', 'lama', 'sekali', 'untuk', 'pesanan']
saya biasa coba makan bento porsi cukup banyak pilihan bento cukup beragam mulai dari salmon ayam daging sapi hidangan penutup biasa saya pesan es strawberry jenis makanan variatif banget dan tidak membosankan yang paling menarik <i>dessert</i> nya banyak jenis dan tampilan nya juga menarik sekali	['saya', 'biasa', 'coba', 'makan', 'bento', 'porsi', 'cukup', 'banyak', 'pilihan', 'bento', 'cukup', 'beragam', 'mulai', 'dari', 'salmon', 'ayam', 'daging', 'sapi', 'hidangan', 'penutup', 'biasa', 'saya', 'pesan', 'es', 'strawberry'] ['jenis', 'makanan', 'variatif', 'banget', 'dan', 'tidak', 'membosankan', 'yang', 'paling', 'menarik', 'dessert', 'nya', 'banyak', 'jenis', 'dan', 'tampilan', 'nya', 'juga', 'menarik', 'sekali']

Pada Tabel 4, adalah proses untuk membuat setiap masing-masing kata menjadi *token-token*. Proses ini membuat setiap kata akan diwakili oleh satu *token* atau masing-masing *token* per-kata.

Tahap yang kelima yaitu *stopwords*. Proses *stopwords* dapat dilihat pada Tabel 5 :

Tabel 5 Proses *Stopwords*

<i>Tokenization</i>	<i>Stopwords</i>
['sabar', 'itu', 'memang', 'capek', 'tapi', 'kalau', 'dibarengin', 'ikhlas', 'semua', 'pasti', 'bakal', 'indah', 'p ada', 'waktu', 'nya']	['sabar', 'capek', 'dibarengin', 'ikhlas', 'indah']
['bongkar', 'saja', 'pak', 'sby', 'sekalian', 'toh', 'ini', 'negeri', 'juga', 'sudah', 'hancur', 'kok']	['bongkar', 'pak', 'sby', 'sekalian', 'toh', 'negeri', 'sudah', 'hancur']
['wifi', 'di', 'sini', 'selalu', 'eror', 'makanan', 'dan', 'minuman', 'agak', 'mahal', 'lumpiah', 'goreng', 'asin', 'sekali', 'teh', 'susu', 'hanya', 'sedikit', 'enak', 'lama', 'sekali', 'untuk', 'pesanan']	['wifi', 'error', 'makanan', 'minuman', 'mahal', 'lumpiah', 'goreng', 'asin', 'teh', 'susu', 'enak', 'pesanan']
['saya', 'biasa', 'coba', 'makan', 'bento', 'porsi', 'cukup', 'banyak', 'pilihan', 'bento', 'cukup', 'beragam', 'mulai', 'dari', 'salmon', 'ayam', 'daging', 'sapi', 'hidangan', 'penutup', 'biasa', 'saya', 'pesan', 'es', 'strawberry']	['bento', 'coba', 'makan', 'porosi', 'pilihan', 'beragam', 'salmon', 'ayam', 'daging', 'sapi', 'hidangan', 'penutup', 'pesan', 'es', 'strawberry']
['jenis', 'makanan', 'variatif', 'banget', 'dan', 'tidak', 'membosankan', 'yang', 'paling', 'menarik', 'dessert', 'nya', 'banyak', 'jenis', 'dan', 'tampilan', 'nya', 'juga', 'menarik', 'sekali']	['jenis', 'menarik', 'makanan', 'variatif', 'banget', 'membosankan', 'dessert', 'tampilan']

Pada Tabel 5, adalah proses penghapusan dari kata-kata yang termasuk dalam kamus “stopwords” bahasa Indonesia dari *library* python. Contohnya adalah kata “itu” setelah melalui proses *stopwords*, maka kata “itu” dihapus dari teks.

Tahap yang keenam yaitu *Stemming*. Proses *stemming* dapat dilihat pada Tabel 6 :

Tabel 6 Proses *Stemming*

<i>Stopwords</i>	<i>Stemming</i>
['sabar', 'capek', 'dibarengin', 'ikhlas', 'indah']	['sabar', 'capek', 'bareng', 'ikhlas', 'indah']
['capek', 'banget', 'sumpah', 'menunggu', 'paket', 'dari', 'jne']	['capek', 'banget', 'sumpah', 'tunggu', 'paket', 'dari', 'jne']

<i>Stopwords</i>	<i>Stemming</i>
['wifi', 'error', 'makanan', 'minuman', 'mahal', 'lumpiah', 'goreng', 'asin', 'teh', 'susu', 'enak', 'pesanan']	['wifi', 'error', 'makan', 'minum', 'mahal', 'lumpiah', 'goreng', 'asin', 'teh', 'susu', 'enak', 'pesan']
['bento', 'coba', 'makan', 'porosi', 'pilihan', 'beragam', 'salmon', 'ayam', 'daging', 'sapi', 'hidangan', 'penutup', 'pesan', 'es', 'strawberry']	['bento', 'coba', 'makan', 'porosi', 'pilih', 'agam', 'salmon', 'ayam', 'daging', 'sapi', 'hidang', 'tutup', 'pesan', 'es', 'strawberry']
['jenis', 'menarik', 'makanan', 'variatif', 'banget', 'membosankan', 'dessert', 'tampilan']	['jenis', 'tarik', 'makan', 'variatif', 'banget', 'bosan', 'dessert', 'tampil']

Pada Tabel 6, adalah proses untuk menghapus kata *prefix* seperti, “nya”, “di”, “in” dan lain-lain. Contohnya adalah kata “dibarengin”, maka setelah proses *stemming* menjadi “bareng” atau diubah menjadi kata dasar.

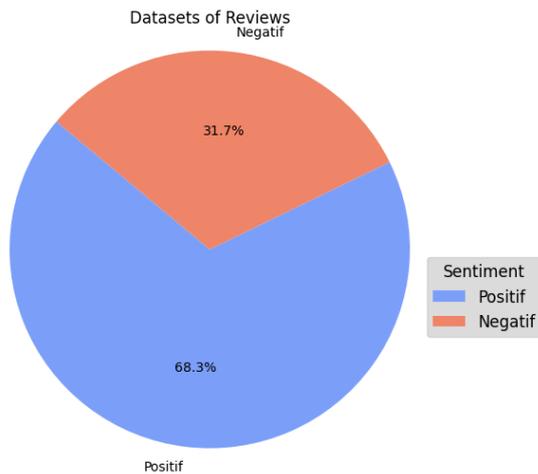
Tahap yang ke-tujuh yaitu Sinonim. Proses sinonim dapat dilihat pada Tabel 7 :

Tabel 7 Proses Sinonim

<i>Stemming</i>	<i>Sinonim</i>
enak banget roti <i>ice cream</i>	enak banget roti <i>ice cream</i>
lokasi jalan braga nyaman	lokasi jalan braga nyaman
asik <i>hang out</i> teman	asik <i>hang out</i> teman
kolega rekomendasi	rekomendasi
enak makan banget wajib	enak makan banget wajib
coba cinta ayam bakar	coba cinta ayam bakar
harga padan bareng	harga padan bareng
keluarga cinta	keluarga
makan restoran enak	makan restoran enak
variasi lokal interlokal jus	variasi lokal interlokal jus
segar pilih pandang hadap	pilih pandang hadap hutan
hutan cantik sejuk makan	cantik sejuk makan
keluarga pasang	keluarga pasang
mantap bakso empuk	mantap bakso empuk enak
lembut enak variasi harga	variasi harga murah pokok
murah pokok wajib coba	wajib coba deh
deh	

Pada Tabel 7, adalah proses sinonim atau menghapus setiap kata yang memiliki arti yang sama pada dataset. Contohnya adalah “asik hang out teman kolega”, kata “teman” dan “kolega” memiliki arti yang sama, sehingga kata “kolega” dihapus.

Hasil akhir dari *preprocessing* data adalah total terdapat 9177 data dengan 6272 data, atau sekitar 68,3%, dalam kategori "positif" dan 2905 data, atau sekitar 31,7%, dalam kategori "negatif".



Gambar 4 Rasio Dataset of Reviews Setelah Pra-proses

Pada Gambar 4, merupakan rasio jumlah dari dataset yang telah melalui proses pra-proses data. Jumlah data positif sebanyak 68,3% dan data negatif sebanyak 31,7%. Data tersebut masih tidak seimbang karena jumlah data positif lebih banyak daripada jumlah data negatif.

Ekstraksi dan Seleksi Fitur serta Implementasi SMOTE dan Undersampling

TF-IDF yang digunakan adalah menghitung TF dan IDF, kemudian menghasilkan skor pembobotan dari setiap fiturnya. Kemudian proses mengubah teks menjadi vektor kata dilakukan berdasarkan skor pembobotannya menggunakan TF-IDF Vectorizer. Penggunaan N-grams pada penelitian ini yaitu dimulai dari 1-gram (*unigram*) sampai 5-gram (*pentagram*). Seleksi fitur yang digunakan adalah dengan uji Chi-Square. Selanjutnya skenario implementasi SMOTE dibagi menjadi dua yaitu: *oversampling* atau proses peningkatan sampel kelas minoritas agar seimbang dengan kelas mayoritas, dan *undersampling* adalah proses penurunan sampel kelas mayoritas agar seimbang dengan kelas minoritas.

Klasifikasi Model MNB dengan Cross Validation

Klasifikasi MNB merupakan proses pelatihan model yang menghasilkan *output* training yaitu : prediksi kelas dan prediksi frekuensi kemunculan fitur. Hasil dari pelatihan model adalah menguji model yang sudah dilatih yang menghasilkan *output testing* yaitu: positif, netral, dan negatif. *Cross Validation* diterapkan pada tahapan ini dengan k=10, sehingga terjadi 10 kali perulangan proses pelatihan dan pengujian model MNB.

Evaluasi Model dengan Confusion Matrix

Confusion matrix akan digunakan dalam pengujian untuk mengevaluasi kinerja algoritma Multinomial Naïve Bayes. Pada penelitian ini menggunakan enam indikator kinerja dalam evaluasi ini, yaitu *accuracy*, *precision*, *recall*, *f1-score*, *AUC* dan *ROC*.

HASIL DAN PEMBAHASAN

Penelitian ini melakukan klasifikasi sentimen dengan metode yang diusulkan yaitu: Klasifikasi menggunakan algoritma MNB dengan ekstraksi fitur TF-IDF dan N-grams (1-gram sampai 5-gram), serta proses seleksi fitur menggunakan Chi-Square dan penanganan ketidakseimbangan dataset dengan *oversampling* SMOTE dan *undersampling*.

Klasifikasi dan Evaluasi Model

Eksperimen dilakukan pada tahap klasifikasi model dengan terbagi menjadi 3 tahap : Pertama, proses klasifikasi “No sampling” menggunakan dataset tanpa dilakukan *oversampling* atau *undersampling*. Kedua, proses klasifikasi “Oversampling” yaitu dataset yang telah dilakukan *oversampling*. Ketiga, “Undersampling” yaitu dataset yang telah dilakukan *undersampling*.

Hasil terbaik dari klasifikasi MNB dengan “No sampling” dapat dilihat pada Gambar 5 :

```

Training Accuracy (No Sampling): 0.9043260324724856

Accuracy (No Sampling): 0.8806796340249416
Precision (No Sampling): 0.9113768347414695
Recall (No Sampling): 0.8651432259650849
F1 Score (No Sampling): 0.8833306304433068
Classification Report (No Sampling):

```

	precision	recall	f1-score	support
negative	0.93	0.76	0.83	2905
positive	0.90	0.97	0.93	6272
accuracy			0.90	9177
macro avg	0.91	0.87	0.88	9177
weighted avg	0.91	0.90	0.90	9177

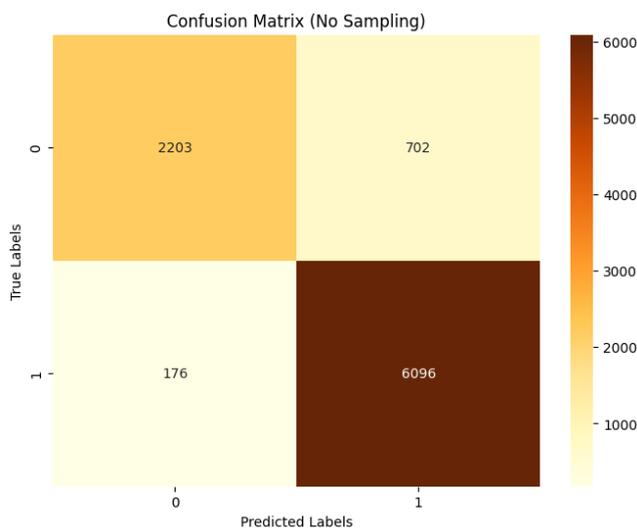
```

AUC Score (No Sampling): 0.9553927078576698

```

Gambar 5 Hasil Akurasi Klasifikasi MNB dengan “No sampling”

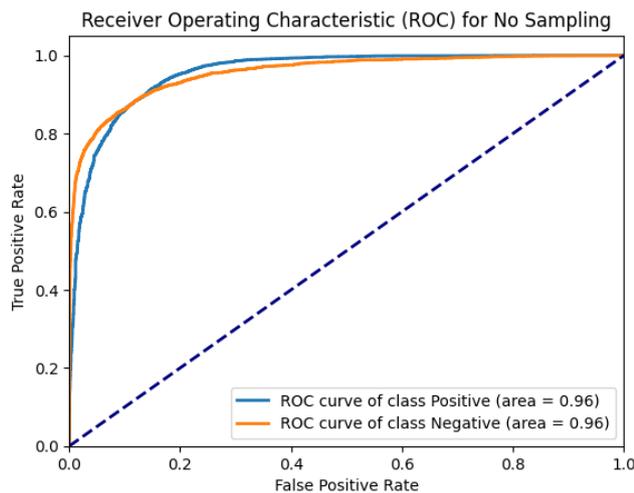
Pada Gambar 5, Nilai akurasi sebesar 88% didapatkan dari penggunaan *unigram* dan untuk penggunaan N-grams lainnya pada data “no sampling” menunjukkan akurasi yang lebih rendah. Kemudian terdapat nilai Precision sebesar 91%, nilai Recall sebesar 87%, nilai F1-Score sebesar 88%, dan nilai AUC Score sebesar 96%. Secara keseluruhan model sudah cukup baik dalam mengklasifikasikan kelas positif dan negatif. *Confusion Matrix* dari data dengan “No sampling” dilihat pada Gambar 6 :



Gambar 6 Confusion Matrix dari MNB “No sampling”

Pada Gambar 6, nilai dari true positive (TP) sebanyak 6096, false negative (FN) sebanyak 176, artinya model sudah sangat baik dalam mengklasifikasi kelas data positif. Kemudian nilai dari true negative (TN) sebanyak 2203, false positive (FP) sebanyak 702, artinya model cukup baik dalam mengklasifikasi kelas data negatif.

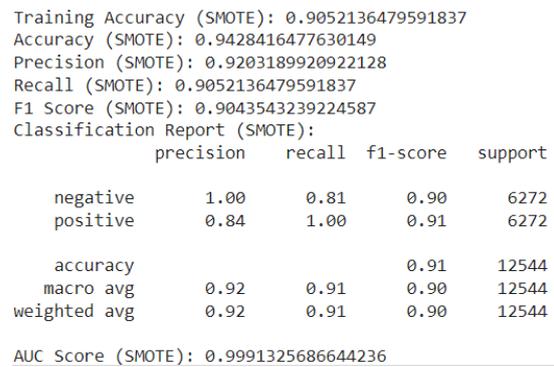
Nilai AUC atau *Area Under Curve* seperti pada Gambar 5 yaitu sebesar 96%, artinya model sudah sangat baik dalam memisahkan antar dua kelas yang berbeda. Nilai ROC atau Receiver Operating Characteristic dapat dilihat pada Gambar 7 :



Gambar 7 Kurva ROC dari MNB “No sampling”

Pada Gambar 7, nilai ROC pada kelas positif sebesar 96%, dan nilai ROC pada kelas negatif sebesar 96%, sehingga model klasifikasi mampu membedakan antara kelas positif dan negatif dengan sangat baik.

Hasil terbaik dari klasifikasi MNB dengan “Oversampling” dapat dilihat pada Gambar 8 :



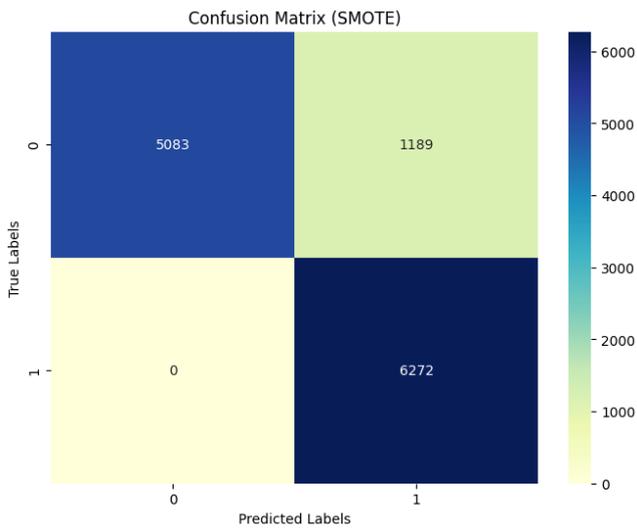
Gambar 8 Hasil akurasi klasifikasi MNB dengan “Oversampling”

Pada Gambar 8, hasil yang didapatkan pada data dengan *oversampling* dan penggunaan *pentagram* atau *5-gram* mendapatkan akurasi sebesar 94%. Kemudian terdapat nilai Precision sebesar 92%, nilai Recall sebesar 91%, nilai F1-Score sebesar 90%, dan nilai AUC Score sebesar 100%. Secara keseluruhan model sudah sangat baik dalam mengklasifikasikan kelas positif dan negatif. Hasil ini merupakan yang tertinggi dari eksperimen klasifikasi yang dilakukan.

Pada penelitian yang dilakukan oleh (Hossain et al., n.d.), penggunaan N-grams pada *unigram* menghasilkan nilai akurasi yang lebih baik dibandingkan *bigram* dan *trigram*. Dalam penelitian ini penggunaan N-grams yang lebih kompleks yaitu *quadgram* hingga *pentagram*. Hasilnya terbukti bahwa dengan perluasan cakupan N-grams mampu meningkatkan nilai akurasi dari hasil klasifikasi model MNB. Nilai akurasi penggunaan *pentagram* pada skenario dataset Oversampling SMOTE menunjukkan akurasi yang tinggi sebesar 94%.

Berikutnya penelitian yang dilakukan oleh (Anam et al., 2023) bahwa pengimplementasian teknik SMOTE berdampak signifikan terhadap peningkatan nilai akurasi klasifikasi sentimen. Kemudian hasil dari skenario dataset “No Sampling” dengan *pentagram* memiliki nilai akurasi sebesar 78% dan untuk perbandingan yaitu skenario dataset Oversampling SMOTE dengan *pentagram* menghasilkan nilai akurasi sebesar 94%, sehingga teknik Oversampling SMOTE terbukti meningkatkan hasil akurasi sebesar 16%.

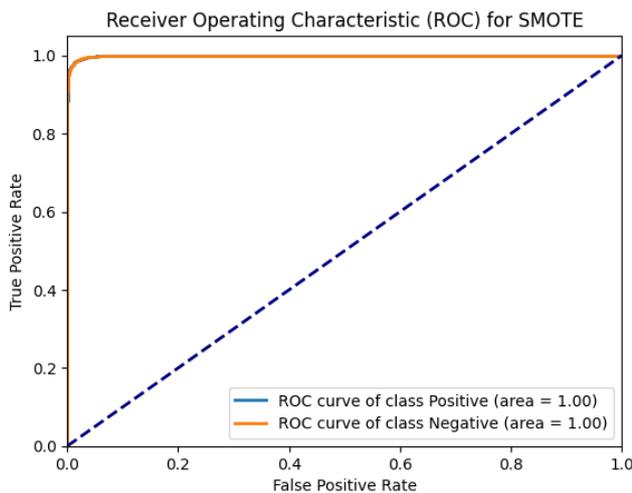
Confusion Matrix dari data dengan “Oversampling” dilihat pada Gambar 9 :



Gambar 9 Confusion Matrix dari MNB “Oversampling”

Pada Gambar 9, nilai dari true positive (TP) sebanyak 6272, false negative (FN) sebanyak 0, artinya model sudah sangat baik dalam mengklasifikasi kelas data positif. Kemudian nilai dari true negative (TN) sebanyak 5083, false positive (FP) sebanyak 1189, artinya model cukup baik dalam mengklasifikasi kelas data negatif.

Nilai AUC atau Area Under Curve seperti pada Gambar 8 yaitu sebesar 100%, artinya model sudah sangat baik dalam memisahkan antar dua kelas yang berbeda. Nilai ROC atau Receiver Operating Characteristic dapat dilihat pada Gambar 10 :



Gambar 10 Kurva ROC dari MNB “Oversampling”

Pada Gambar 10, nilai ROC pada kelas positif sebesar 100%, dan nilai ROC pada kelas negatif sebesar 100%, sehingga model klasifikasi mampu membedakan antara kelas positif dan negatif dengan sangat baik.

Hasil terbaik dari klasifikasi MNB dengan “Undersampling” dapat dilihat pada Gambar 11 :

Training Accuracy (Undersampling): 0.8986230636833047

Accuracy (Undersampling): 0.8650602409638555
 Precision (Undersampling): 0.9157126502575844
 Recall (Undersampling): 0.8986230636833046
 F1 Score (Undersampling): 0.8975703653178666

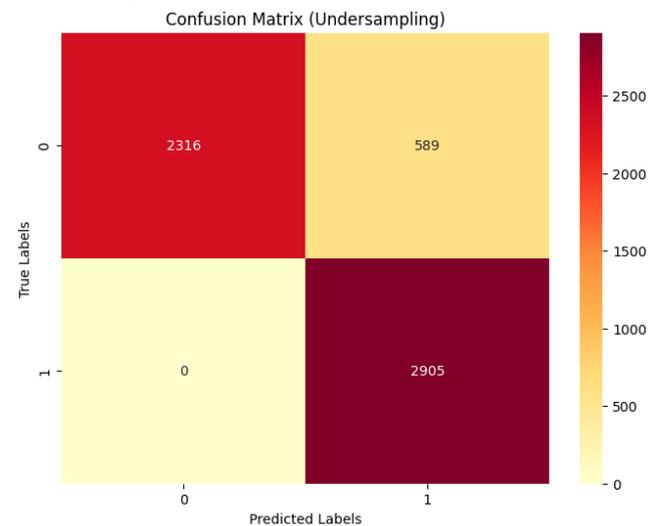
Classification Report (Undersampling):

	precision	recall	f1-score	support
negative	1.00	0.80	0.89	2905
positive	0.83	1.00	0.91	2905
accuracy			0.90	5810
macro avg	0.92	0.90	0.90	5810
weighted avg	0.92	0.90	0.90	5810

AUC Score (Undersampling): 0.999095156134743

Gambar 11 Hasil akurasi klasifikasi MNB dengan “Undersampling”

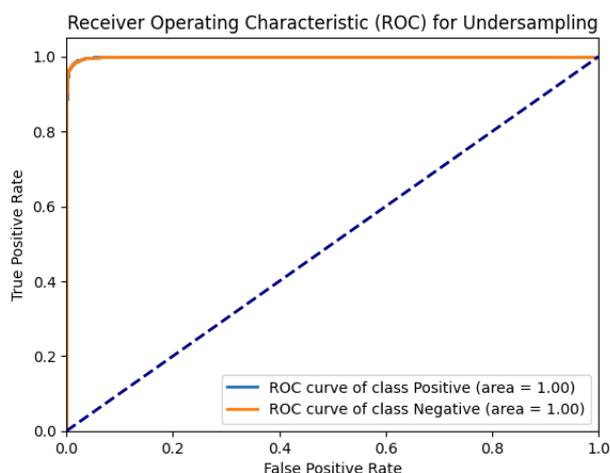
Pada Gambar 11, metode dengan data “undersampling” memiliki hasil akurasi yang sama dari penggunaan *unigram* hingga *pentagram*, yaitu menghasilkan akurasi sebesar 86%. Selanjutnya terdapat nilai Precision sebesar 92%, nilai Recall sebesar 90%, nilai F1-Score sebesar 90%, dan nilai AUC Score sebesar 100%. Secara keseluruhan model sudah sangat baik dalam mengklasifikasikan kelas positif dan negatif. Confusion Matrix dari data dengan “Oversampling” dilihat pada Gambar 12 :



Gambar 12 Confusion Matrix dari MNB “Undersampling”

Pada Gambar 12, nilai dari true positive (TP) sebanyak 2905, false negative (FN) sebanyak 0, artinya model sudah sangat baik dalam mengklasifikasi kelas data positif. Kemudian nilai dari true negative (TN) sebanyak 2316, false positive (FP) sebanyak 589, artinya model cukup baik dalam mengklasifikasi kelas data negatif.

Nilai AUC atau Area Under Curve seperti pada Gambar 11 yaitu sebesar 100%, artinya model sudah sangat baik dalam memisahkan antar dua kelas yang berbeda. Nilai ROC atau Receiver Operating Characteristic dapat dilihat pada Gambar 13 :



Gambar 13 Kurva ROC dari MNB “Undersampling”

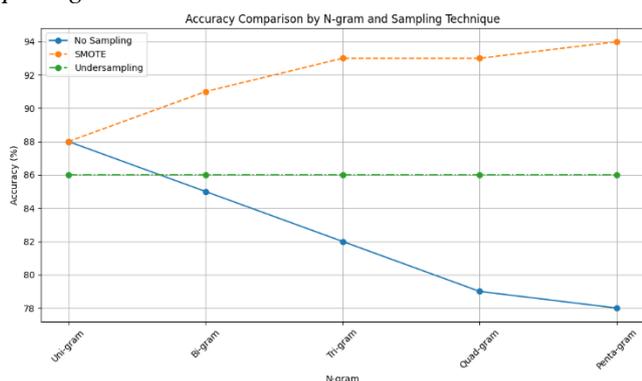
Pada Gambar 13, nilai ROC pada kelas positif sebesar 100%, dan nilai ROC pada kelas negatif sebesar 100%, sehingga model klasifikasi mampu membedakan antara kelas positif dan negatif dengan sangat baik.

Adapun tabel perbandingan akurasi dari hasil keseluruhan eksperimen penggunaan *N-grams* yang dilakukan dapat dilihat pada Tabel 8 :

Tabel 8 Komparasi hasil akurasi dengan N-grams

	No Sampling	SMOTE	Undersampling
Unigram	88%	88%	86%
Bigram	85%	91%	86%
Trigram	82%	93%	86%
Quadgram	79%	93%	86%
Pentagram	78%	94%	86%

Pada Tabel 8, adalah tabel perbandingan dari hasil akurasi dengan *N-grams* dari masing-masing skenario dataset. Untuk hasil akurasi yang paling rendah adalah pada data “No Sampling” dengan penggunaan *pentagram* dan nilai akurasi sebesar 78%. Kemudian untuk hasil akurasi yang paling tinggi adalah pada data “Oversampling SMOTE” dengan penggunaan *pentagram* dan nilai akurasi sebesar 94%.



Gambar 14 Komparasi hasil akurasi klasifikasi MNB

Pada Gambar 14, merupakan hasil perbandingan akurasi dari masing-masing *dataset* dan penggunaan *N-grams*. Pada *dataset No-Sampling* hasilnya lebih baik pada penggunaan *n-grams* yang lebih sederhana atau *unigram* semakin kompleks *n-grams* maka nilai akurasinya pun turun. Kemudian pada *dataset Undersampling*, penggunaan *n-grams* tidak mempengaruhi sama sekali, karena hasil akurasinya tetap stagnan. Selanjutnya pada *dataset Oversampling SMOTE*, terlihat bahwa semakin kompleks penggunaan *n-grams* atau *pentagram* maka nilai akurasinya pun semakin tinggi. Maka hasil akurasi tertinggi dan yang paling baik dari dataset yang telah diterapkan SMOTE dan penggunaan *pentagram*.

Adapun tabel perbandingan akurasi dari hasil keseluruhan eksperimen tanpa penggunaan *N-grams* yang dilakukan dapat dilihat pada Tabel 9 :

Tabel 9 Komparasi hasil akurasi tanpa N-grams

	No Sampling	SMOTE	Undersampling
Akurasi	89%	88%	86%

Pada Tabel 9, merupakan tabel perbandingan dari hasil akurasi tanpa *N-grams* dari masing-masing skenario *dataset*. Hasilnya adalah bahwa *dataset “No Sampling”* menunjukkan hasil yang paling tinggi yaitu sebesar 89% di antara skenario *dataset* lainnya. Untuk perbandingan dengan yang mengimplementasikan *N-grams*. Skenario *dataset* SMOTE dengan *pentagram* lebih unggul dari nilai akurasinya sebesar 94%.

KESIMPULAN

Berdasarkan uraian dari hasil dan pembahasan di atas, dapat disimpulkan bahwa:

1. Kombinasi dari penggunaan ekstraksi fitur TF-IDF dan *N-grams (unigram sampai pentagram)* berhasil dalam meningkatkan kinerja akurasi algoritma MNB.
2. Penggunaan *N-grams* yang lebih kompleks (*pentagram*) dan teknik *Oversampling SMOTE* terbukti dapat meningkatkan nilai akurasi, serta mengatasi ketidakseimbangan kelas data.
3. Penggunaan “sinonim” atau proses menghapus kata-kata yang memiliki arti sama dalam dataset terbukti efektif dalam menyeleksi dataset sehingga representasi teks dalam dataset relevan.
4. Hasil eksperimen dengan skenario *dataset “No Sampling”* pada skenario *unigram* dengan nilai akurasi sebesar 88% dan nilai AUC sebesar 96%. Kemudian pada skenario data

“Undersampling” menunjukkan hasil stagnan yaitu nilai akurasi sebesar 86%. Sementara itu, skenario data “Oversampling” SMOTE dengan *pentagram* menunjukkan hasil akurasi tertinggi sebesar 94%.

Dengan demikian, secara keseluruhan penelitian ini berhasil dalam mengatasi kelemahan algoritma Multinomial Naïve Bayes yaitu “asumsi independensi fitur” dengan mampu meningkatkan kinerja hasil akurasi. Penelitian selanjutnya diharapkan mampu mengeksplorasi cara-cara yang lebih efektif dalam mengatasi kelemahan algoritma MNB, termasuk penggunaan opsi *hyperparameter*, pemilihan teknik ekstraksi dan seleksi fitur lainnya, serta pra-proses data yang lebih optimal.

UCAPAN TERIMA KASIH

Penulis berterima kasih terhadap dukungan yang diberikan oleh seluruh pihak dari Universitas Jenderal Achmad Yani.

DAFTAR PUSTAKA

- Abbas, M., Ali, K., Jamali, A., Ali Memon, K., & Aleem Jamali, A. (2019). Multinomial Naïve Bayes Classification Model for Sentiment Analysis. *IJCSNS International Journal of Computer Science and Network Security*, 19(3), 62. <https://doi.org/10.13140/RG.2.2.30021.40169>
- Amal, M. I., Rahmasita, E. S., Suryaputra, E., & Rakhmawati, N. A. (2022). Analisis Klasifikasi Sentimen Terhadap Isu Kebocoran Data Kartu Identitas Ponsel di Twitter. *Jurnal Teknik Informatika Dan Sistem Informasi*, 8(3), 645–660. <https://doi.org/10.28932/jutisi.v8i3.5483>
- Anam, M. K., Triyani, ;, Fitri, A., Agustin, ;, Lusiana, ;, Muhammad, ;, Firdaus, B., Agus, ;, & Nurhuda, T. (2023). Sentiment Analysis for Online Learning using The Lexicon-Based Method and The Support Vector Machine Algorithm. *ILKOM Jurnal Ilmiah*, 15(2), 290–302. <http://dx.doi.org/10.33096/ilkom.v15i2.1590.290-302>
- Ernayanti, T., Mustafid, M., Rusgiyono, A., & Hakim, A. R. (2023). Penggunaan Seleksi Fitur Chi-Square Dan Algoritma Multinomial Naïve Bayes Untuk Analisis Sentimen Pelanggan Tokopedia. *Jurnal Gaussian*, 11(4), 562–571. <https://doi.org/10.14710/j.gauss.11.4.562-571>
- Farisi, A. A., Sibaroni, Y., & Faraby, S. Al. (2019). Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier. *Journal of Physics: Conference Series*, 1192(1). <https://doi.org/10.1088/1742-6596/1192/1/012024>
- Handayani, Y., Hakim, A. R., & Muljono. (2020). Sentiment analysis of Bank BNI user comments using the support vector machine method. *Proceedings - 2020 International Seminar on Application for Technology of Information and Communication: IT Challenges for Sustainability, Scalability, and Security in the Age of Digital Disruption, ISemantic 2020*, 202–207. <https://doi.org/10.1109/iSemantic50169.2020.9234230>
- Hossain, E., Sharif, O., & Hoque, M. M. (n.d.). *Book Reviews Using Multinomial Naïve Bayes*.
- Prastyo, P. H., Ardiyanto, I., & Hidayat, R. (2020). Indonesian Sentiment Analysis: An Experimental Study of Four Kernel Functions on SVM Algorithm with TF-IDF. *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy, ICDABI 2020*. <https://doi.org/10.1109/ICDABI51230.2020.9325685>
- Purwarianti, A., & Crisdayanti, I. A. P. A. (2019). Improving Bi-LSTM Performance for Indonesian Sentiment Analysis Using Paragraph Vector. *Proceedings - 2019 International Conference on Advanced Informatics: Concepts, Theory, and Applications, ICAICTA 2019*. <https://doi.org/10.1109/ICAICTA.2019.8904199>
- Sholehurrohman, R., & Sabda Ilman, I. (2022). Analisis Sentimen Tweet Kasus Kebocoran Data Penggunaan Facebook Oleh Cambrigde Analytica. *Jurnal Pepadun*, 3(1), 140–147. <https://doi.org/10.23960/pepadun.v3i1.108>
- Singh, G., Kumar, B., Gaur, L., & Tyagi, A. (2019). Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification. *2019 International Conference on Automation, Computational and Technology Management, ICACTM 2019*, 593–596. <https://doi.org/10.1109/ICACTM.2019.8776800>
- Surya, P. P. M., Seetha, L. V., & Subbulakshmi, B. (2019). Analysis of user emotions and opinion using Multinomial Naïve Bayes Classifier. *Proceedings of the 3rd International Conference on Electronics and Communication and Aerospace Technology, ICECA 2019*, 410–415. <https://doi.org/10.1109/ICECA.2019.8822096>
- Taufiqi, A. M., & Nugroho, A. (2023). Sentimen Pengguna Twitter Mengenai Isu Kebocoran Data Dengan Algoritma Naïve Bayes. *Jurnal Nasional Ilmu Komputer*, 4(1), 1–11. <https://doi.org/10.47747/jurnalnik.v4i1.1091>
- Wibowo, N. I., Maulana, T. A., Muhammad, H., & Rakhmawati, N. A. (2021). Perbandingan Algoritma Klasifikasi Sentimen Twitter Terhadap Insiden Kebocoran Data Tokopedia. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 6(2), 120–129. <https://doi.org/10.14421/jiska.2021.6.2.120-129>
- Zul, M. I., Yulia, F., & Nurmalasari, D. (2018). Social media sentiment analysis using K-means and naïve bayes algorithm. *Proceedings - 2018 2nd International Conference on Electrical Engineering and Informatics: Toward the Most Efficient Way of Making and Dealing with Future Electrical Power System and*

Big Data Analysis, ICon EEI 2018, October, 24–29.
<https://doi.org/10.1109/ICon-EEI.2018.8784326>