

Analisis Sentimen pada Ulasan Aplikasi myIM3 Menggunakan *Multinomial Naive Bayes* dengan TF-IDF

Ni Komang Ayu Juliana^{a1}, Made Agung Raharja^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
¹juliana.2208561046@student.unud.ac.id
²made.agung@unud.ac.id

Abstract

The digital service adoption in Indonesia has emerged as a primary trend to meet the needs of the millennial generation, seeking greater convenience and speed. Amidst this trend, self-service apps like MyIM3 by Indosat Ooredoo Hutchison have become a trusted solution for users to manage their services more efficiently. Sentiment analysis is crucial for understanding user responses to such apps. This study employs the Multinomial Naïve Bayes algorithm with hyperparameter alpha 0.8 and TF-IDF to analyze sentiment towards user reviews on Google Play Store for MyIM3. The dataset, sourced from Kaggle, consists of 8475 reviews, pre-processed and labeled to 8212 reviews. Model evaluation with an 80:20 split reveals an overall accuracy of 89%, with a precision of 86% for negative (0) and 93% for positive (1) labels. The recall for negative is 95% and positive is 81%. Thus, this research contributes to understanding user perspectives on MyIM3 and provides a basis for enhancing the quality of app-based services.

Keywords: *Multinomial Naive Bayes, myIM3, TF-IDF, Sentiment Analysis*

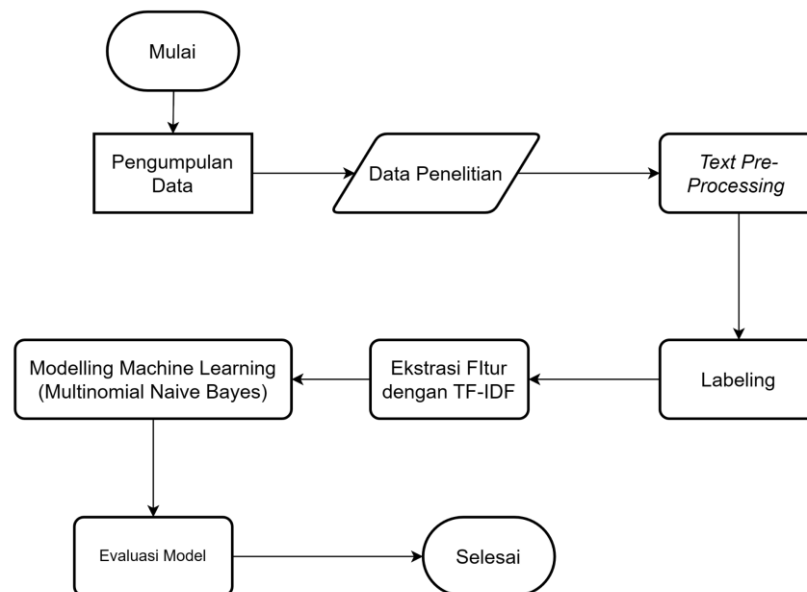
1. Pendahuluan

Perkembangan teknologi saat ini terus menawarkan kemudahan yang signifikan, termasuk dalam bidang telekomunikasi, informasi, dan transaksi pembelian. Fenomena ini tercermin dalam adopsi layanan digital di Indonesia, yang menjadi salah satu tren utama dalam memenuhi kebutuhan generasi milenial yang menginginkan segalanya menjadi lebih praktis dan cepat. Dukungan untuk tren ini terlihat dalam peningkatan jumlah pengguna internet, seperti yang dilaporkan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII). Berdasarkan survei terbaru mereka tahun 2023, total pengguna internet sekitar 215 juta orang yang menunjukkan peningkatan signifikan dari tahun sebelumnya sebesar 1,17% [1]. Di tengah dinamika ini, perusahaan penyedia layanan jaringan telekomunikasi, seperti Indosat Ooredoo Hutchison, terus berinovasi untuk memenuhi kebutuhan pelanggan [2]. Salah satu inovasi mereka adalah aplikasi self-service bernama MyIM3, yang telah menjadi solusi yang sangat diandalkan bagi pengguna untuk mengelola layanan mereka dengan lebih efisien. Dengan lebih dari 50 juta unduhan dan lebih dari 9 juta ulasan di *Google Play Store*, MyIM3 menjadi salah satu aplikasi layanan provider yang terkenal di pasar. Oleh karena itu, ulasan dan penilaian pengguna di platform seperti *Google Play Store* menjadi sangat penting, karena dapat mempengaruhi persepsi pengguna potensial tentang kualitas dan kegunaan suatu aplikasi. Dalam konteks ini, analisis sentimen menjadi alat yang efektif untuk memahami respons pengguna terhadap aplikasi MyIM3. Algoritma *Multinomial Naive Bayes* dipilih karena kecepatan dan kesederhanaannya dalam mengklasifikasikan data teks, serta kemampuannya untuk menangani distribusi frekuensi kata yang besar. Metode TF-IDF (*Term Frequency-Inverse Document Frequency*) digunakan sebagai metode pembobotan kata untuk mengurangi dampak kata-kata yang sering muncul namun kurang informatif, sambil memberikan penekanan pada kata-kata yang lebih jarang tetapi mungkin lebih penting dalam konteks sentimen. Dengan demikian, ini meningkatkan akurasi klasifikasi dengan memberikan bobot lebih pada kata-kata yang lebih signifikan dalam menentukan sentimen ulasan. Terdapat beberapa penelitian terdahulu yang membahas mengenai analisis sentimen menggunakan *Multinomial Naive Bayes* dan TF-IDF seperti misalnya pada penelitian mengenai Analisis

Sentimen terhadap Produk Skin *Game* di Forum *Review Female Daily* Menggunakan Metode *Multinomial Naïve Bayes* dan TF-IDF yang mendapatkan hasil evaluasi sebesar 86,6%, *precision* 89,1%, dan *recall* 94,2% [3]. Selain itu, terdapat penelitian lain mengimplementasikan algoritma *Multinomial Naive Bayes* dalam analisis sentimen opini pengguna aplikasi Brimo berbobot TF-IDF yang mendapatkan hasil evaluasi *confusion matrix* terbaik pada *fold-2* dengan nilai akurasi 98,02%, presisi 97,06%, *recall* 97,06%, dan *f1-score* 97,06% [4]. Terdapat juga penelitian yang berjudul Analisis Sentimen Opini Pengguna Twitter Pada Aplikasi Bibit Menggunakan *Multinomial Naïve Bayes* yang memperoleh akurasi sebesar 91.50% dan tingkat akurasi validasi sebesar 85.35% [5]. Berdasarkan hasil studi sebelumnya, penelitian ini melakukan analisis sentimen pada ulasan pengguna di Google Play Store untuk aplikasi myIM3 menggunakan metode *Multinomial Naive Bayes*. Metode *Multinomial Naive Bayes* dengan pembobotan TF-IDF ini digunakan untuk memilah ulasan ke dalam kategori positif dan negatif dengan lebih presisi. Dengan mendapatkan informasi yang lebih akurat, diharapkan hasilnya dapat digunakan sebagai panduan untuk menjaga kualitas layanan, memperbaiki kekurangan, dan meningkatkan evaluasi ke depan untuk aplikasi myIM3.

2. Metode Penelitian

Dalam penelitian ini, beberapa langkah akan dilakukan untuk menganalisis sentimen dari ulasan di Google Play Store terhadap Aplikasi myIM3 menggunakan metode *Multinomial Naive Bayes* dan TF-IDF (*Term Frequency-Inverse Document Frequency*). Langkah-langkah tersebut dimulai dengan mengumpulkan data, melakukan *text pre-processing* pada data, melabelkan data yang sudah di *pre-processing*, mengekstraksi fitur menggunakan TF-IDF, kemudian memodelkan data menggunakan algoritma *Multinomial Naïve Bayes*, dan terakhir mengevaluasi model yang dihasilkan menggunakan *confusion matrix*. Berikut adalah alur penelitian yang ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Data

Teknik pengumpulan data dalam penelitian dapat beragam, seperti menggunakan data yang tersedia secara umum di berbagai platform online, melakukan web scraping untuk mengambil data dari internet, dan menggabungkan beberapa dataset tertentu melalui data augmentation [6]. Dataset untuk penelitian ini diperoleh dari data publik yang berasal dari situs web Kaggle yang diunggah oleh pengguna dengan nama pengguna Ari Zidane. Kaggle adalah platform yang menyediakan berbagai dataset untuk keperluan penelitian dan pembelajaran. Dataset di Kaggle

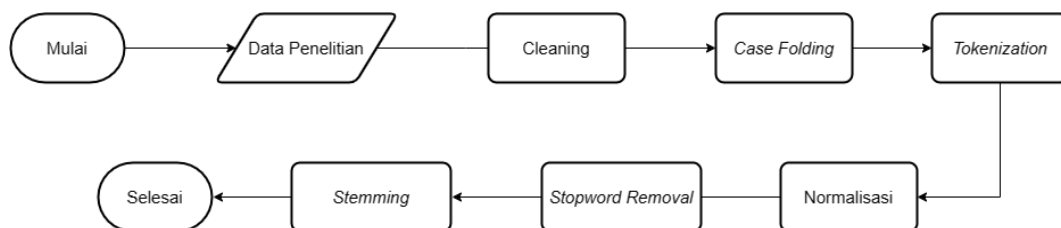
berisi data yang beragam yang telah dikumpulkan dan diproses oleh pengguna lain untuk berbagai tujuan penelitian. Dataset ini terdiri dari 8475 ulasan yang mencakup nama pengguna, skor atau rating, tanggal pembuatan ulasan, dan konten ulasan itu sendiri. Dataset ini terdiri dari 8475 ulasan yang berisi username, score atau rating, tanggal pembuatan ulasan, dan juga content atau ulasannya seperti pada Tabel 1.

Tabel 1. Dataset myIM3

userName	content	score	at
Nurul Aini	klo gk ada data nya gk bisa di buka	2	31/01/2024 23:56:00
Salsa Bila	bagus	5	31/01/2024 23:50:00
JebolGamingYT	Aku beli pulsa. Beli paket data. Di proses terus lama.. Sudah 1 jam. Biasaya cepet. Kok ini lama.. Segeraaa	1	31/01/2024 23:42:00
AREK BERBEK 1H	Apa"an sekarang im3 error bkin ribet.pas beli kartu perdana baru tapi disuruh beli pulsa.hbis beli pulsa harus ada registrasi dulu. Registrasi di WA harus ada kode otp dulu di SMS. Tapi g ada notifnya. Sumpah apaan. Pusing	1	31/01/2024 23:38:00
Denik Rahayu	bagus	5	31/01/2024 23:30:00

2.2 Text Pre-Processing

Tahap selanjutnya adalah pre-processing data adalah langkah awal di mana data disiapkan dan dibersihkan untuk melakukan analisis. Untuk tahapan dari pre-processing dapat dilihat pada Gambar 2.



Gambar 2. Alur Pre-Processing

Terdapat beberapa tahapan yang dilakukan pada *text pre-processing*, berikut adalah penjelasannya:

- Cleaning*, pada tahapan ini pembersihan data dari unsur-unsur yang tidak diperlukan atau berlebihan, seperti data kosong, karakter khusus, dan format yang tidak konsisten. Pada proses ini juga di cek apakah data yang didapatkan terdapat ulasan yang kosong atau tidak apabila ada, maka data tersebut akan dibuang.
- Case Folding* adalah proses dimana semua huruf dalam teks diubah menjadi huruf kecil untuk mencegah adanya duplikasi kata yang sama yang mungkin dianggap berbeda karena perbedaan kapitalisasi.

- c. *Tokenization* merupakan proses dimana teks atau kalimat dipisahkan menjadi unit yang lebih kecil, yaitu kata-kata. Proses tokenisasi membantu dalam mengidentifikasi elemen-elemen yang akan dianalisis dan mempermudah proses perhitungan frekuensi serta analisis konteks.
- d. *Normalisasi*, proses ini melibatkan pengubahan kata-kata yang memiliki variasi bentuk seperti kata imbuhan dan slang menjadi bentuk dasar atau standar.
- e. *Stopword Removal* adalah tahap di mana kata-kata yang sering muncul tetapi tidak memiliki makna signifikan dihapus. Proses penghapusan stopwords bertujuan untuk mengeliminasi kata-kata tersebut agar analisis dapat terfokus pada kata-kata yang lebih informatif.
- f. *Stemming*, proses menghilangkan imbuhan dari kata-kata untuk kembali ke bentuk dasar atau akar kata.

Berikut adalah contoh dari hasil pre-processing data yang ditunjukkan pada Tabel 2.

Tabel 2. Hasil *Pre-Processing*

Data	Hasil <i>Pre-Processing</i>
tolong diperbaiki transferan pulsa nomor baru bisa langsung buat TF	['tolong', 'perbaiki', 'transfer', 'pulsa', 'nomor', 'baru', 'bisa', 'langsung', 'buat', 'transfer']
klo gk ada data nya gk bisa di buka	['kalau', 'tidak', 'ada', 'data', 'nya', 'tidak', 'bisa', 'di', 'buka']
Aplikasi yang bagus	['aplikasi', 'bagus']
mantap ??	['mantap']
Im3 sinyal busuk . waktunya ganti nih. Payah	['im3', 'sinyal', 'busuk', 'waktu', 'ganti', 'ini', 'payah']

2.3 Labeling

Dalam tahap labeling dataset, setiap data diberi label berdasarkan skor yang terkait dengan masing-masing data. Skor ini berasal dari rating yang diberikan oleh pengguna pada aplikasi myIM3. Prosedur ini bertujuan untuk mengkategorikan data ke dalam dua kelas yang berbeda berdasarkan tingkat kepuasan atau pendapat pengguna terhadap aplikasi. Dalam tahap labeling ini, data dengan skor 3 dianggap tidak dapat diinterpretasikan secara tegas, sehingga dianggap tidak cukup untuk dimasukkan ke dalam kelas sentimen netral. Selain itu, data dengan skor 3 juga tidak dapat dianggap sebagai sentimen negatif atau positif karena adanya ketidakpastian dalam interpretasi kata-kata pada ulasan dengan skor tersebut. Sehingga data dengan skor 3 dihapus untuk memudahkan proses analisis dan klasifikasi, menghindari ketidakpastian dalam interpretasi ulasan, serta dataset menjadi lebih terfokus pada sentimen positif dan negatif yang lebih jelas. Lalu untuk skor antara 1 hingga 2, data akan diberi label negatif atau 0, yang menandakan bahwa pengguna tidak puas dengan pengalaman mereka menggunakan aplikasi. Sementara untuk skor antara 4 hingga 5, data akan diberi label positif atau 1, yang menandakan bahwa pengguna puas atau sangat puas dengan pengalaman mereka.

2.4 Ekstraksi Fitur

Ekstraksi fitur adalah proses untuk mengubah data teks menjadi representasi numerik yang dapat digunakan oleh algoritma pembelajaran mesin. Dalam penelitian ini digunakan ekstraksi fitur dengan TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk mengukur seberapa penting suatu kata dalam suatu dokumen atau korpus teks. TF-IDF menghitung nilai bobot untuk setiap kata berdasarkan frekuensi kata dalam dokumen tersebut dan kebalikannya dengan

frekuensi kata dalam seluruh korpus [7]. Nilai TF-IDF secara matematis direpresentasikan dalam Persamaan 1 berikut.

$$TF - IDF = w_{d,t} = tf_{d,t} \times \log\left(\frac{N}{df_t}\right) \quad (1)$$

Adapun nilai $tf_{d,t}$ merupakan TF (*Term Frequency*) digunakan untuk mengevaluasi seberapa sering sebuah kata muncul dalam teks atau jumlah kemunculan istilah d dalam dokumen t . Sedangkan nilai $df_{i,j}$ merupakan *Document Frequency* Document Frequency (DF) untuk menilai relevansi sebuah istilah dalam setiap kumpulan dokumen atau jumlah dokumen yang mencakup kemunculan istilah tersebut. Ini membantu dalam mengevaluasi istilah yang sering muncul dan yang jarang muncul dalam teks. Lalu nilai $w_{d,t}$ merupakan nilai dari TF-IDF nya dan nilai N adalah jumlah dokumen. Dengan memanfaatkan metode TF-IDF, kita bisa memberikan nilai numerik kepada kata-kata dalam dokumen yang mencerminkan signifikansi kata-kata tersebut untuk dokumen tersebut dalam konteks keseluruhan kumpulan dokumen.

2.5 Multinomial Naive Bayes

Multinomial Naive Bayes adalah variasi dari algoritma Naive Bayes yang sangat cocok untuk klasifikasi teks, di mana fitur-fitur yang dihitung adalah frekuensi kata. Algoritma ini menerapkan prinsip distribusi multinomial digunakan untuk menggambarkan probabilitas dari kombinasi frekuensi kata-kata dalam dokumen. Ini berarti bahwa algoritma ini memperhitungkan frekuensi kemunculan kata dalam dokumen, bukan hanya keberadaan atau ketiadaan kata tersebut [8]. Algoritma Multinomial Naive Bayes dengan pembobotan TF-IDF ini dapat dihitung dengan Persamaan 2.

$$P(X_n | C) = \frac{\sum_{d \in C} tf(X_n, d) + \alpha}{\sum_{d \in C} N_{d \in C} + V} \quad (2)$$

Dimana $\sum_{d \in C} tf(X_n, d)$ adalah jumlah pembobotan kata X_n dari seluruh data training pada kelas C dan $\sum_{d \in C} N_{d \in C}$ adalah jumlah bobot seluruh term pada data training pada kelas C . Dalam penelitian ini, dataset yang telah melalui text preprocessing akan dibagi menjadi 80% data training dan sebanyak 20% data testing.

2.6 Evaluasi Model

Untuk tahapan evaluasi model, digunakan *confusion matrix*. *Confusion matrix* merupakan alat evaluasi kinerja yang penting dalam pengembangan model klasifikasi, termasuk dalam evaluasi model *Multinomial Naive Bayes*. *Confusion matrix* adalah tabel yang digunakan untuk membandingkan hasil prediksi dari model dengan label yang sebenarnya dari data uji. Pada tahapan ini, data tes sebanyak 20% akan akan diujikan dengan 80% data trainingnya. *Confusion matrix* memiliki empat sel atau komponen utama, yaitu:

- a. *True Positives* (TP): Semua total data yang diprediksi benar sebagai positif atau berlabel 1 dan benar berlabel positif atau 1.
- b. *False Positives* (FP): Semua total data yang diperkirakan sebagai positif atau berlabel 1, tetapi sebenarnya adalah negatif atau berlabel 0.
- c. *True Negatives* (TN): Semua total data yang diperkirakan benar sebagai negatif atau berlabel 0 dan benar berlabel negatif atau 0.
- d. *False Negatives* (FN): Jumlah data yang diprediksi sebagai negatif atau berlabel 0 tetapi sebenarnya adalah berlabel positif atau 1.

Dari matriks ini, kita dapat menghitung berbagai metrik evaluasi lainnya seperti precision, recall, dan F1-score. Berikut adalah rumus perhitungan precision yang ditunjukkan oleh Persamaan 3, perhitungan *recall* pada Persamaan 4, dan F1-score ditunjukkan oleh Persamaan 5:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

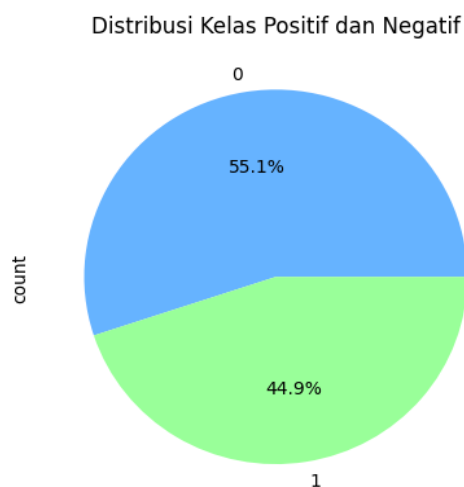
$$F1\ Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (5)$$

Precision (presisi) mengukur kualitas positif yang diprediksi oleh model menggunakan nilai TP (*True Positives*) dan FP (*False Positives*), sedangkan *recall* untuk mengukur seberapa baik model mengidentifikasi semua positif yang sebenarnya menggunakan nilai TP (*True Positives*) dan FN (*False Negative*), dan *F1-Score* digunakan menggabungkan *precision* dan *recall* menjadi suatu nilai yang berguna untuk mengevaluasi model klasifikasi.

3. Hasil dan Pembahasan

3.1 Hasil Analisis Data dan Pre-Processing

Dari hasil pengumpulan data melalui kaggle didapatkan data sebanyak 8470, lalu setelah dilakukan pre-processing dan labeling data yang digunakan adalah sebanyak 8420 karena data dengan skor 3 dihapus. Lalu untuk dilakukan penghitungan jumlah data yang telah diberi label 0 (negatif) dan label 1 (positif). Distribusi kelas positif dan negatif ini kemudian divisualisasikan dalam bentuk diagram pie untuk memberikan gambaran yang lebih jelas tentang proporsi atau perbandingan antara kedua kelas tersebut dalam dataset. Dengan visualisasi ini, dapat dilihat proporsi ulasan yang positif dan negatif dalam dataset. Adapun jumlah data setelah di hapuskan skor 3 adalah sebanyak 8212 dengan distribusi kelas positif atau berlabel 1 sebanyak 3690 atau sebesar 44.9% dan distribusi kelas negatif dengan label 0 sebanyak 4552 atau sebesar 55.1%, seperti pada Gambar 3.



Gambar 3. Distribusi Kelas Positif dan Negatif

3.2 Hasil Ekstraksi Fitur TF - IDF

Pada tahap ekstraksi fitur ini, metode TF-IDF yang digunakan. Proses ekstraksi fitur melibatkan penggunaan CountVectorizer dan TfidfTransformer dari *library* scikit-learn yang dijalankan dengan bahasa pemrograman *Python* di *Google Colab*. Total vocabulary yang dihasilkan adalah sebanyak 8420 kata. Untuk hasil akhir dari matriks TF-IDF diperlihatkan melalui Gambar 4.

```
dokumen = term_fit.transform(Ulasan)
tfidf_transformer = TfidfTransformer().fit(dokumen)
print(tfidf_transformer.idf_)

tfidf = tfidf_transformer.transform(term_frequency)
print(tfidf)

[9.32032636 9.32032636 9.32032636 ... 8.62717918 9.32032636 9.32032636]
(0, 7872) 0.17226885116600582
(0, 7481) 0.14068588532301415
(0, 6994) 0.4001940272639384
(0, 6518) 0.1569646540283102
(0, 6472) 0.26463947011826816
(0, 5937) 0.15138953877958244
(0, 4365) 0.36700462978223203
(0, 4105) 0.20288982922099022
(0, 3372) 0.24288886088196873
(0, 3252) 0.1358725903342128
(0, 1985) 0.12295473854258003
(0, 1915) 0.1881645464642987
(0, 1737) 0.24753531086175948
(0, 1391) 0.4001940272639384
(0, 1142) 0.31573411436871335
(0, 563) 0.21870321318939745
```

Gambar 4. Hasil TF-IDF

3.3 Hasil Modeling *Multinomial Naive Bayes*

Model *Multinomial Naive Bayes* dibangun dengan menggunakan bantuan *library* scikit-learn dalam bahasa pemrograman Python. Salah satu hyperparameter yang digunakan adalah alpha sebesar 0,8. Setelah itu model akan dilatih dengan data training sebanyak 80% atau 6569 data dan akan diuji menggunakan data test sebanyak 20% atau 1643 data yang ditunjukkan pada Gambar 5.

```
# Split data training dan data testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Membuat CountVectorizer untuk mengonversi data teks menjadi fitur numerik
vectorizer = CountVectorizer()
X_train_vec = vectorizer.fit_transform(X_train)
X_test_vec = vectorizer.transform(X_test)

# Melatih klasifikasi Naive Bayes Multinomial
mnb = MultinomialNB(alpha=0.8, fit_prior=True, force_alpha=True)
mnb.fit(X_train_vec, y_train)
```

MultinomialNB
MultinomialNB(alpha=0.8, force_alpha=True)

Gambar 5. Modeling dan *Split* Data

Setelah itu, model yang dibangun dengan model *Multinomial Naive Bayes* dan ekstraksi fitur TF-IDF akan diuji modelnya menggunakan *confusion matrix* untuk mengetahui nilai presisi, akurasi, dan F1 scorenya. Proses evaluasi atau pengujian model ini mendapatkan hasil akurasi keseluruhan sebesar 89%. Dengan precision untuk kelas negatif atau berlabel 0 adalah 86%, recall sebesar 95%, dan f1-score sebesar 91%. Sedangkan untuk kelas positif atau berlabel 1

didapatkan hasil precision sebesar 93%, recall sebesar 81%, dan f1-score sebesar 86%. Hasil dari confusion matrix dapat dilihat pada Gambar 6.

```
----- confusion matrix -----  
[[319 160]  
 [ 26 343]]  
----- classification report -----  
              precision    recall  f1-score   support  
  
     0           0.92       0.67       0.77         479  
     1           0.68       0.93       0.79         369  
  
 accuracy              0.78         848  
 macro avg           0.80         0.80         0.78         848  
 weighted avg        0.82         0.78         0.78         848
```

Gambar 6. Hasil *Confusion Matrix*

4. Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa analisis sentimen pada aplikasi myIM3 menggunakan metode Multinomial Naive Bayes yang memanfaatkan *hyperparameter alpha* sebesar 0.8 dengan ekstraksi fitur TF-IDF menunjukkan nilai precision, recall, akurasi, dan F1-score untuk label negatif berturut-turut adalah 86%, 95%, dan 91%. Sementara untuk label positif, nilai precision, recall, akurasi, dan F1-score berturut-turut adalah 93%, 81%, dan 86%. Akurasi pengujian model menggunakan confusion matrix juga signifikan, mencapai 89%. Selain menggunakan ekstraksi fitur, peningkatan performa model juga dapat dilakukan dengan melakukan penyesuaian terhadap *hyperparameter* pada model.

Daftar Pustaka

- [1] Asosiasi Penyelenggara Jasa Internet Indonesia, "Survey APJII 2023," 2023. [Daring]. Tersedia pada: <https://survei.apjii.or.id/>. Diakses pada 4 Mei 2024.
- [2] A. P. Pratama, A. R. Destriawan, E. K. Panggiarti, "Analisis Merger Pada PT Indosat Tbk (ISAT) Terhadap PT Hutchison 3 Indonesia (H3I)," *J. Akuntan Publik*, vol. 1, no. 4, pp 292-298. 2023.
- [3] Q. A. N. Prakoso, A. Muliawati, I. N. Isnainiyah, "Analisis Sentimen terhadap Produk Skin Game di Forum Review Female Daily Menggunakan Metode Multinomial Naïve Bayes," *J. Informatik*, vol. 18, no. 3, pp. 198-207, 2022.
- [4] N. F. Arminda, N. Sulistyowati, T. N. Padilah, "Implementasi Algoritma Multinomial Naive Bayes Pada Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Brimo," *J. Mahasiswa Teknik Informatika*, vol. 7, no. 3, pp. 1817-1822, 2023.
- [5] Z. G. Ngilo, Nuryulani, "Analisis Sentimen Opini Pengguna Twitter Pada Aplikasi Bibit Menggunakan Multinomial Naïve Bayes," *J. Teknik dan Science*, vol. 2, no. 1, pp. 8-15, 2023.
- [6] P. A. N. Aryanti, I. B. M. Mahendra, "Analisis Sentimen Opini Berbahasa Indonesia Pada Sosial Media Menggunakan TF-IDF dan Support Vector Machine," *J. Elektronik Ilmu Komputer Udayana*, vol.12, no. 1, pp. 45-52, 2023.
- [7] I. Widyaningrum, D. Mustikasari, R. Arifin, S. L. Tsaqila, D. Fatmawati, "Algoritma Term Frequency – Inverse Document Frequency (TF-IDF) dan K-Means Clustering Untuk Menentukan Kategori Dokumen," *Prosiding Seminar Nasional Sistem Informasi dan Teknologi*, vol. 2022, no. 6, pp. 145-149, 2022.
- [8] M. N. Randhika, J. C. Young, A. Suryadibrata, H. Mandala, "Implementasi Algoritma Complement dan Multinomial Naïve Bayes Classifier Pada Klasifikasi Kategori Berita Media Online," *J Teknik Informatika*, vol. 13, no. 1, pp. 19-25, 2021.