

Prediksi Depresi Risiko Depresi pada Remaja Berdasarkan Frekuensi Penggunaan Media Sosial yang Tinggi Menggunakan SMOTE dan Algoritma Naïve Bayes

Dian Margi Santoso^{1*}, Ulfah Kusnia², Arif Faizin³

^{1,2,3}Teknik Informatika, Universitas Yudharta Pasuruan

*email: dionms203@gmail.com

DOI: <https://doi.org/10.31603/komtika.v10.i1.17504>

ABSTRACT

The increasing use of social media among adolescents is believed to be associated with the emergence of mental health disorders, including depression. However, identifying depression risk in adolescents remains a challenge, particularly due to imbalanced data that can affect the performance of classification models. This study focuses on developing a naïve Bayes model to classify depression risk in adolescents based on social media usage frequency, and on evaluating the differences in its performance before and after the application of SMOTE using accuracy, precision, recall, and F1-Score metrics. This study uses the Teen Mental Health Dataset obtained from Kaggle. The research process begins with data preprocessing, followed by the separation of features and target. The data is then split into training and testing sets with an 80:20 ratio. In the next stage, the training data is balanced using the SMOTE method before being classified using the naïve Bayes algorithm. A confusion matrix is used as the evaluation tool to measure the resulting model's performance. The results show that the model without SMOTE achieved an accuracy of 99%, precision of 100%, recall of 50%, and F1-Score of 67%. Meanwhile, the model with SMOTE applied achieved an accuracy of 97%, precision of 50%, recall of 33%, and F1-Score of 40%. Based on these results, the naïve Bayes algorithm demonstrates excellent classification performance in predicting depression risk among adolescents. The application of SMOTE successfully balanced the class distribution in the training data but did not provide a significant performance improvement compared to the model without SMOTE. Therefore, the Naïve Bayes model without SMOTE is the most optimal approach for the dataset used in this study.

Keywords: Teen Depression, Social Media, Naive Bayes, SMOTE, Machine Learning.

ABSTRAK

Penggunaan media sosial yang semakin meningkat pada remaja diperkirakan memiliki keterkaitan dengan timbulnya gangguan kesehatan mental, termasuk depresi. Namun, identifikasi risiko depresi pada remaja masih menjadi tantangan, terutama karena adanya ketidakseimbangan data (*imbalanced data*) yang dapat memengaruhi kinerja model klasifikasi. Studi ini berfokus pada pengembangan model *naïve bayes* untuk mengklasifikasikan risiko depresi pada remaja berdasarkan frekuensi penggunaan media sosial serta mengevaluasi perbedaan kinerjanya sebelum dan setelah penerapan SMOTE melalui metrik akurasi, presisi, *recall*, dan *F1-Score*. Penelitian ini menggunakan *Teen Mental Health Dataset* yang diperoleh dari Kaggle. Tahapan penelitian ini diawali dengan melakukan prapemrosesan terhadap data, dilanjutkan dengan pemisahan antara fitur dan target. Selanjutnya, data dibagi menjadi data latih dan data uji dengan proporsi 80:20. Pada tahap berikutnya, data latih diseimbangkan menggunakan metode SMOTE sebelum diklasifikasikan dengan algoritma *naïve bayes*. Untuk mengukur kinerja model yang dihasilkan, digunakan *confusion matrix* sebagai alat evaluasi. Hasil dari penelitian ini menunjukkan bahwa model tanpa SMOTE memperoleh akurasi sebesar 99%, presisi 100%, *recall* 50%, dan *F1-Score* 67%. Sementara itu, model dengan penerapan SMOTE memperoleh akurasi 97%, presisi 50%, *recall* 33%, dan *F1-Score* 40%. Berdasarkan hasil penelitian, algoritma *naïve bayes* menunjukkan kinerja klasifikasi yang sangat baik dalam memprediksi risiko depresi pada remaja. Penerapan SMOTE berhasil menyeimbangkan distribusi kelas pada data *training*, namun tidak memberikan peningkatan performa yang signifikan.

dibandingkan model tanpa penerapan SMOTE. Oleh karena itu, model *Naïve Bayes* tanpa SMOTE menjadi pendekatan paling optimal untuk dataset yang digunakan.

Kata-kata kunci: Depresi Remaja, Media Sosial, *Naïve Bayes*, SMOTE, *Machine Learning*

PENDAHULUAN

Penggunaan media sosial yang terus meningkat di kalangan remaja telah menimbulkan kekhawatiran terkait dampaknya terhadap kesehatan mental. Beberapa faktor yang diduga berperan dalam munculnya depresi akibat penggunaan media sosial secara berlebihan mencakup perilaku penggunaan pasif atau komplusif, *cyberbullying*, gangguan tidur, dan perbandingan sosial [1]. Sebuah studi menunjukkan bahwa remaja yang menggunakan media sosial yang berlebihan cenderung mengalami gejala kecemasan sebesar 78,3% dan gejala depresi sebesar 82,6%, sehingga penggunaan yang tidak terkontrol dapat berdampak negatif terhadap kesehatan mental mereka [2]. Frekuensi penggunaan media sosial yang tinggi menjadi satu indikator penting dalam mendefinisikan risiko depresi pada remaja. Pada penelitian yang dilakukan oleh [3] menjelaskan bahwa, penggunaan media sosial dalam durasi Panjang berkaitan dengan meningkatnya kemungkinan terjadinya stress, kecemasan, kelelahan, dan gangguan tidur.

Permasalahan tersebut kemudian dikaji menggunakan *Teen Mental Health Dataset* yang digunakan sebagai sumber data dalam penelitian ini. Dataset terdiri atas 1.200 data remaja dengan sejumlah karakteristik yang berkaitan dengan kondisi mental dan aktivitas sehari-hari. Beberapa atribut yang digunakan dalam penelitian meliputi *daily social media hours* yang merepresentasikan durasi penggunaan media sosial, *sleep hours* yang menggambarkan durasi tidur, serta *anxiety level* dan *stress level* yang merepresentasikan kondisi psikologis responden. Variabel *depression label* digunakan sebagai target klasifikasi yang membedakan responden ke dalam kelas tidak depresi dan depresi. Dengan demikian, dataset tersebut memiliki keterkaitan langsung dengan permasalahan penelitian karena menyediakan informasi mengenai penggunaan media sosial dan beberapa faktor yang dapat dimanfaatkan dalam pengembangan model klasifikasi risiko depresi pada remaja. Dengan demikian, diperlukan suatu sistem prediksi yang dapat mendukung proses identifikasi risiko depresi secara dini dengan mempertimbangkan pola penggunaan media sosial. Hasil prediksi tersebut diharapkan dapat menjadi dasar bagi orang tua, pendidik, maupun tenaga kesehatan dalam menentukan langkah penanganan yang sesuai..

Untuk memahami permasalahan tersebut, diperlukan analisis terhadap data kesehatan mental terhadap remaja, guna mengidentifikasi faktor apa yang menyebabkan remaja mengalami depresi. Untuk mengatasi permasalahan tersebut, *machine learning* dapat menjadi salah satu solusi yang efektif, mengingat kemampuannya dalam mengenali risiko depresi berdasarkan analisis sejumlah karakteristik data seperti pola perilaku, jejak aktivitas digital, dan kondisi psikologis individu [4].

Penggunaan algoritma *Naïve bayes* diterapkan karena memiliki tingkat efisiensi dan kecepatan komputasi yang tinggi, khususnya ketika menangani dataset yang berukuran besar [5]. Selain itu *naïve bayes* juga menghasilkan proses klasifikasi yang relatif mudah dipahami karena menggunakan pendekatan probabilitas bersyarat, sehingga memungkinkan untuk mengetahui pengaruh masing-masing fitur terhadap hasil klasifikasi yang diperoleh [6].

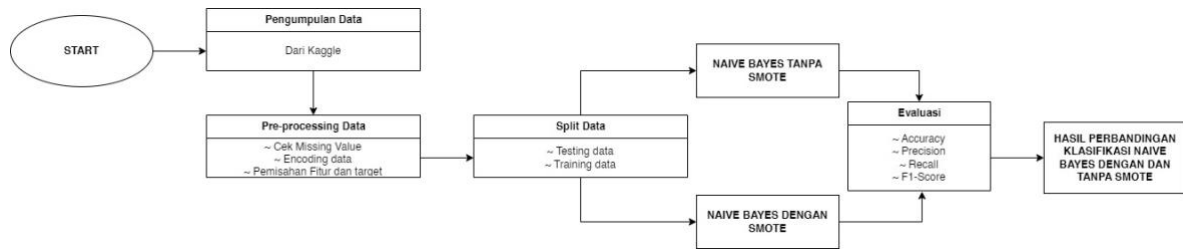
Naïve Bayes memiliki keterbatasan dalam menangani dataset yang tidak seimbang karena lagoritma ini bekerja dengan menghitung probabilitas berdasarkan jumlah frekuensi dan kombinasi data pada setiap kelas [7]. Maka dari itu diterapkannya metode *Synthetic Minority Over-sampling Technique* (SMOTE) untuk mengatasi permasalahan tersebut. Pemilihan metode SMOTE didasarkan pada kemampuannya dalam menyeimbangkan distribusi kelas melalui pembentukan data baru secara sintetis pada kelas minoritas dengan teknik interpolasi [8]. Pada penelitian ini menerapkan metode SMOTE sebelum proses klasifikasi menggunakan *Naïve Bayes* dengan harapan menghasilkan performa klasifikasi yang lebih optimal, sehingga meningkatkan akurasi serta menghasilkan prediksi yang presisi.

Dari penelitian sebelumnya yang memprediksi depresi dari data *Patient Health Questionnaire* (PHQ-9) menggunakan *naïve bayes*, *random forest*, dan *support vector machine* (SVM), hasilnya *naïve bayes* mendapatkan akurasi 77%, *random forest* mendapatkan akurasi 97%, dan SVM mendapatakn akurasi 98% [9]. Pendekatan ekstraksi fitur telah diterapkan dalam beberapa penelitian terdahulu guna mengidentifikasi indikasi depresi dan kecemasan di kalangan pengguna *Twitter*. Perbandingann dilakukan terhadap empat algoritma pembelajaran mesin, yakni *decision tree*, *random forest*, *naïve bayes*, dan *k-nearest neighbor*. Hasil evaluasi yang didapat, algoritma *decision tree* mendapatkan akurasi 95%, algoritma *random forest* mendapatkan 96%, algoritma *naïve bayes* mendapatkan 91%, dan algoritma *k-nearest neighbor* mendapatkan 86% [10]. Penelitian lainnya melakukan perbandingan kinerja algoritma *Naïve Bayes* dan *Random Forest* dalam mengklasifikasikan kondisi depresi pada pengguna platform X dengan menggunakan dataset yang terdiri atas 100 data. Hasilnya algoritma *naïve bayes* mendapatkan akurasi 80%, dan algoritma *random forest* mendapatkan akurasi 90% [11].

Penelitian ini dilakukan untuk merancang model klasifikasi guna mengidentifikasi risiko depresi pada remaja berdasarkan frekuensi penggunaan media sosial menggunakan algoritma *naïve bayes*, serta menganalisis pengaruh ketidakseimbangan data terhadap performa model. Penelitian ini juga dimaksudkan untuk membandingkan performa algoritma *Naïve Bayes* sebelum dan setelah penerapan SMOTE, yang diukur melalui empat indikator evaluasi berupa *accuracy*, *precision*, *recall*, serta *F1-Score*. Evaluasi tersebut dilakukan untuk mengetahui efektivitas SMOTE dalam mengatasi permasalahan ketidakseimbangan kelas pada dataset kesehatan mental remaja.

METODE

Sebagaimana ditunjukkan pada Gambar 1, penelitian ini dilakukan melalui beberapa tahapan yang tersusun secara sistematis, diawali dengan pengumpulan data dan pre-processing, lalu dilanjutkan pada tahap pembagian dataset (*split data*). Setelah itu, diterapkan dua skenario implementasi algoritma *Naïve Bayes*, yaitu dengan menggunakan metode SMOTE dan tanpa penerapan SMOTE. Kedua skenario tersebut kemudian dievaluasi menggunakan *confusion matrix* berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-Score*. Tahapan terakhir dalam penelitian ini adalah membandingkan hasil evaluasi antara kedua skenario tersebut, untuk melihat perbedaan kinerja algoritma dalam menghadapi permasalahan ketidakseimbangan data.



Gambar 1. Tahapan Penelitian

a. Pengumpulan Data

Data penelitian didapat dari platform *Kaggle* yang berupa file dataset *Teen_mental_health_dataset* yang diakses pada tanggal 6 Mei 2026. Terdapat 13 variabel yang ada pada dataset yang digunakan untuk penelitian ini diantaranya *age*, *gender*, *daily social media hours*, *platform usage*, *sleep hours*, *screen time before sleep*, *academic performance*, *physical activity*, *social interaction level*, *stress level*, *anxiety level*, *addiction level*, dan *depression label*.

b. Pre-processing Data

Pada tahap prapemrosesan, dilakukan pengecekan terhadap *missing value* dan pengodean data kategorikal ke dalam format numerik sehingga data dapat digunakan dalam proses klasifikasi menggunakan algoritma *naïve bayes*. Kemudian dilakukan memisahkan fitur dan target yang bertujuan untuk menentukan variabel yang digunakan sebagai masukan dalam proses klasifikasi serta variabel yang akan diprediksi oleh algoritma.

c. Split Data

Pada tahap *split data* pembagian data menggunakan rasio 80:20. Rasio tersebut diterapkan untuk menyediakan jumlah data yang memadai bagi proses pelatihan sekaligus memastikan tersedianya data yang cukup untuk menguji kinerja model [12].

d. Implementasi *Synthetic Minority Over-sampling Technique* SMOTE

SMOTE digunakan sebagai teknik *oversampling* untuk meningkatkan representasi kelas minoritas pada dataset yang tidak seimbang [13]. SMOTE menciptakan sampel sintesis pada kelas minoritas dengan cara melakukan interpolasi antar sampel yang ada, sehingga proporsi kelas dalam data pelatihan menjadi lebih seimbang [8].

Rumus dari SMOTE adalah sebagai berikut seperti yang digunakan pada penelitian [14].

$$X_{syn} = X_i + (X_{knn} - X_i) + \delta \quad (1)$$

Keterangan :

X_{syn} = Data sintesis yang akan dibuat

X_i = Sampel yang dipilih

X_{knn} = Data tetangga terdekat dari X_i

δ = Angka acak (0,1)

e. Implementasi *Naïve Bayes*

Naïve Bayes merupakan metode klasifikasi statistik yang berlandaskan pada *teorema bayes* serta dikenal karena struktur perhitungannya yang sederhana dan kemampuannya dalam menangani berbagai permasalahan klasifikasi secara efektif [5]. Metode ini memiliki tingkat akurasi, efisiensi dan kemudahan implementasi yang baik dalam proses klasifikasi, sehingga mampu membantu pengambilan keputusan berdasarkan karakteristik data yang dianalisis [15].

Rumus perhitungan *Naïve Bayes* adalah sebagai berikut seperti yang digunakan pada penelitian [16].

$$P(x_i | \gamma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu_\gamma)^2}{2\sigma^2}\right) \quad (2)$$

Keterangan :

$P(x_i | \gamma)$ = Probabilitas kemunculan fitur x_i terhadap kelas γ .

x_i = Nilai fitur.

γ = Kelas target.

μ_γ = Rata-rata (mean) fitur pada kelas γ .

σ^2 = Varians pada kelas γ .

π = Konstanta phi (3,14159).

exp = Fungsi eksponensial.

f. Evaluasi Model

Evaluasi dilakukan pada kedua sekenario menggunakan *Confusion Matrix* dengan aspek pengukuran yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Tahap ini bertujuan untuk mengukur kemampuan model dalam mengklasifikasikan data serta mengetahui pengaruh penggunaan metode oversampling terhadap peningkatan performa klasifikasi [5].

Penilaian kinerja model dilakukan berdasarkan empat elemen dalam *confusion matrix*. *True Negative* (TN) mengindikasikan jumlah data negatif yang terklasifikasi dengan benar, *False Positive* (FP) menunjukkan data negatif yang justru diprediksi sebagai positif, *False Negative* (FN) menunjukkan data positif yang diprediksi sebagai negatif, sedangkan *True Positive* (TP) menunjukkan jumlah data positif yang berhasil diprediksi dengan akurat [5]. Pada tabel 1 merupakan gambaran dari *Confusion Matrix* yang digunakan untuk menilai kinerja algoritma.

Tabel 1. *Confusion Matrix*

<i>Actual</i>	<i>Predicted</i>	
	–	+
–	<i>True Negative</i>	<i>False Positive</i>
+	<i>False Negative</i>	<i>True Positive</i>

Rumus adalah sebagai berikut [14].

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

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (6)$$

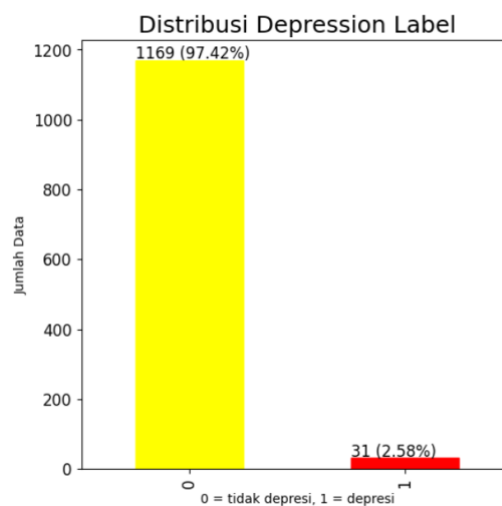
g. Hasil Perbandingan Klasifikasi *Naïve Bayes* dengan dan tanpa SMOTE

Tahap terakhir adalah melakukan analisis perbandingan antara dua model yang telah dibangun yaitu *Naïve Bayes* tanpa SMOTE dan *Naïve Bayes* dengan SMOTE. Perbandingan kinerja dilakukan menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score* guna

mengetahui apakah penerapan SMOTE mampu meningkatkan kemampuan algoritma *Naïve Bayes* dalam mengklasifikasikan risiko depresi pada remaja.

HASIL DAN PEMBAHASAN

Pada gambar 2 merupakan visualisasi dari data target. Identifikasi ketidakseimbangan kelas dilakukan dengan menganalisis distribusi kelas pada data target menggunakan perhitungan jumlah dan persentase setiap kelas. Hasil analisis menunjukkan bahwa kelas tidak depresi yang disimbolkan dengan angka (0) berjumlah 1.169 data (97,42%), sedangkan kelas depresi yang disimbolkan dengan angka (1) hanya berjumlah 31 data (2,58%). Ketimpangan jumlah data yang signifikan antara kedua kelas mengindikasikan bahwa dataset berada dalam kondisi tidak seimbang, sehingga diperlukan metode penanganan ketidaksimbangan data yaitu *Synthetic Minority Over-sampling Technique* (SMOTE).



Gambar 2. Visualisasi Data Target

Tahap awal *pre-processing* data dilakukan dengan memeriksa keberadaan *missing value* pada setiap variabel dalam dataset. Pemeriksaan data menunjukkan bahwa tidak terdapat *missing value* pada seluruh variabel dalam dataset. Tahap selanjutnya adalah melakukan proses *encoding* terhadap variabel yang bertipe kategorikal menjadi data numerik. Hasilnya terdapat tiga variabel yang memerlukan proses *encoding* antara lain variabel *gender*, *platform usage*, dan *social interaction level*. Kemudian dilakukan tahap memisahkan fitur dan target, pada penelitian ini variabel *depression_label* digunakan sebagai variabel target, sementara itu variabel lainnya digunakan sebagai variabel fitur.

Langkah selanjutnya adalah dataset dibagi menjadi data latih dan data uji menggunakan rasio 80:20. Pada tabel 2 merupakan visualisasi dari pembagian data yang sudah dilakukan.

Tabel 2. Visualisasi Hasil Pembagian Data

Split Data		Jumlah Data	
Data Training	Data Testing	Data Training	Data Testing
80%	20%	960	240
Jumlah Data		1200	

Berdasarkan hasil pengujian yang telah dilakukan, SMOTE berhasil meningkatkan data minoritas pada data *training* dari 25 data, menjadi 935 data seperti yang disajikan pada tabel 3 yang merupakan tabel hasil perbandingan sebelum dan setelah penerapan SMOTE. Sehingga jumlah data pada kedua kelas menjadi seimbang.

Tabel 3. Hasil Penerapan SMOTE Pada Data *Training*

Kelas	Sebelum SMOTE	Setelah SMOTE
Tidak Depresi (0)	935	935
Depresi (1)	25	935
Total Data	960	1.870

Berdasarkan hasil evaluasi pada tabel 4, penerapan *naïve bayes* dengan SMOTE menunjukkan kemampuan model dalam mengklasifikasikan sejumlah data dengan benar. Berdasarkan 240 data pengujian, sebanyak 232 data dari kelas tidak depresi (0) berhasil diklasifikasikan secara tepat sebagai (TN). Sementara itu, terdapat 2 data dari kelas tidak depresi yang keliru diprediksi sebagai depresi atau (FP). Pada kelas depresi (1), model berhasil mengidentifikasi 2 data secara benar sebagai (TP), sedangkan 4 data lainnya salah diklasifikasikan sebagai tidak depresi atau (FN).

Hasil uji menunjukkan bahwa model meraih tingkat akurasi sebesar 97%, yang menunjukkan bahwa mayoritas data dapat diklasifikasikan dengan benar. Pada kelas tidak depresi (0), model mendapatkan nilai *precision* sebesar 98%, *recall* sebesar 99%, dan *F1-Score* sebesar 99%. Hasil tersebut mengindikasikan bahwa model memiliki kinerja yang sangat baik dalam mengenali data yang tergolong dalam kategori tidak depresi. Berbeda halnya dengan kelas depresi (1), yang menghasilkan *precision* sebesar 50%, *recall* 33%, dan *F1-Score* 40%, sehingga dapat dikatakan bahwa performa model dalam mengidentifikasi kelas tersebut belum optimal.

Tabel 4. Hasil Evaluasi *Naïve Bayes* Dengan SMOTE

Data Aktual	Data Prediksi		<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
	0	1			
0	232(TN)	2(FP)	98%	99%	99%
1	4(FN)	2(TP)	50%	33%	40%
<i>Accuracy</i>			97%		

Berdasarkan hasil pengujian pada Tabel 5, algoritma *naïve bayes* tanpa penerapan SMOTE menunjukkan kinerja yang baik dalam mengklasifikasikan data. Dari total 240 data pengujian, sebanyak 234 data dari kelas tidak depresi (0) berhasil diklasifikasikan secara tepat sebagai (TN), tanpa adanya data tidak depresi yang keliru diprediksi sebagai depresi atau FP). Pada kelas depresi (1), sebanyak 3 data berhasil diklasifikasikan secara tepat sebagai (TP), sementara 3 data lainnya tidak berhasil dikenali sebagai kelas depresi dan justru dikategorikan sebagai tidak depresi, sehingga termasuk dalam (FN).

Hasil evaluasi menunjukkan bahwa model mencapai akurasi 99%, yang menunjukkan bahwa mayoritas data pengujian berhasil diprediksi dengan tepat. Untuk kelas tidak depresi (0),

model memperoleh nilai *precision* sebesar 99%, *recall* sebesar 100%, dan *F1-Score* sebesar 99%. Hasil ini mengindikasikan bahwa model mampu mengenali kelas mayoritas dengan sangat baik. Adapun pada kelas depresi (1), model menghasilkan nilai *precision* sebesar 100%, *recall* sebesar 50%, dan *F1-Score* sebesar 67%, yang menunjukkan bahwa meskipun prediksi terhadap kelas depresi yang dilakukan model tergolong tepat, masih terdapat sebagian data depresi yang belum berhasil teridentifikasi.

Tabel 5. Hasil Evaluasi *Naïve Bayes* Tanpa SMOTE

Data Aktual	Data Prediksi		<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
	0	1			
0	234(TN)	0(FP)	99%	100%	99%
1	3(FN)	3(TP)	100%	50%	67%
<i>Accuracy</i>			99%		

Penelitian ini menerapkan dua skenario pengujian, yaitu implementasi algoritma *Naïve Bayes* dengan metode SMOTE dan implementasi *Naïve Bayes* tanpa menggunakan SMOTE. Kedua skenario tersebut diterapkan untuk menganalisis dan membandingkan perubahan kinerja algoritma *Naïve Bayes* sebelum dan setelah proses *oversampling* menggunakan SMOTE..

Pada tabel 6 merupakan perbandingan hasil evaluasi algoritma *naive bayes* dengan dan tanpa penerapan metode SMOTE pada kelas depresi. Model *naive bayes* dengan penerapan SMOTE memperoleh *accuracy* sebesar 97%, *precision* 50%, *recall* 33%, dan *F1-Score* 40%. Sementara itu, model *naive bayes* tanpa penerapan SMOTE memperoleh nilai *accuracy* sebesar 99%, *precision* 100%, *recall* 50%, dan *F1-Score* 67%. Hasil penelitian mengindikasikan bahwa penerapan SMOTE belum meningkatkan kinerja model, karena model yang dilatih dengan data asli menunjukkan kemampuan identifikasi yang lebih baik. Hal ini dikarenakan kondisi ketidakseimbangan pada dataset yang digunakan berpotensi memengaruhi kinerja algoritma klasifikasi *naive bayes*, karena model cenderung lebih banyak mempelajari pola pada kelas mayoritas. Akibatnya, model dapat menghasilkan tingkat akurasi yang tinggi, tetapi memiliki kemampuan yang rendah dalam mendeteksi kelas minoritas.

Tabel 6. Perbandingan Performa *Naïve Bayes* Dengan dan Tanpa SMOTE

Parameter	<i>Naïve Bayes</i> dengan SMOTE	<i>Naïve Bayes</i> tanpa SMOTE
Akurasi	97%	99%
Presisi	50%	100%
<i>Recall</i>	33%	50%
<i>F1-Score</i>	40%	67%

KESIMPULAN

Dalam penelitian ini melakukan perbandingan performa algoritma *naive bayes* dengan dan tanpa penerapan SMOTE untuk memprediksi risiko depresi pada remaja. Penelitian ini dilaksanakan melalui tahap prapemrosesan data, pemisahan antara fitur dan target, pembagian data menjadi data latih dan data uji dengan perbandingan 80:20, dilanjutkan dengan penerapan

SMOTE pada data latih, kemudian diklasifikasikan menggunakan algoritma naïve bayes. Dengan evaluasi menggunakan *confusion matrix*. Hasil pengujian menunjukkan bahwa model tanpa smote memperoleh akurasi sebesar 99%, presisi 100%, *recall* 50%, dan *F1-Score* 67%. Sementara itu, model dengan penerapan SMOTE memperoleh akurasi 97%, presisi 50%, *recall* 33%, dan *F1-Score* 40%. Berdasarkan hasil evaluasi yang diperoleh, model yang dilatih menggunakan dataset asli menunjukkan kinerja yang lebih baik dibandingkan dengan model yang menggunakan dataset hasil proses *oversampling* SMOTE. Meskipun SMOTE dirancang untuk mengatasi ketidakseimbangan kelas dengan menambahkan data sintetis pada kelas minoritas, metode ini tidak selalu memberikan peningkatan performa klasifikasi. Temuan tersebut mengindikasikan bahwa efektivitas penerapan SMOTE tidak bersifat *universal*, melainkan dipengaruhi oleh karakteristik dataset dan tingkat kesesuaiannya dengan algoritma klasifikasi yang diterapkan. Penelitian berikutnya sebaiknya menggunakan dataset dengan ukuran sampel yang lebih besar dan variasi karakteristik data yang lebih luas, sehingga model dapat lebih baik dalam mengenali dan merepresentasikan pola depresi pada remaja.. Selain itu, penelitian selanjutnya dapat menggunakan metode *oversampling* lainnya seperti Borderline-SMOTE maupun SMOTEENN. Dengan pengembangan tersebut, diharapkan penelitian selanjutnya dapat menghasilkan model klasifikasi yang lebih akurat dan mampu mendeteksi risiko depresi pada remaja secara lebih optimal.

DAFTAR PUSTAKA

- [1] S. M. Saleem and S. S. Jan, “A Cross-sectional Study of Mental Health Effects of Excessive Screen Time and Social Media Use among Indian Adolescents and Young Adults,” *J. Nat. Sci. Med.*, vol. 3, p. 299, Jun. 2024, doi: 10.4103/jnsn.jnsn_37_24.
- [2] S. Damodar *et al.*, “Trending: A Systematic Review of Social Media Use’s Influence on Adolescent Anxiety and Depression,” *Adolesc. Psychiatry (Hilversum)*., vol. 12, no. 1, pp. 11–22, Mar. 2022, doi: 10.2174/2210676612666220225122720.
- [3] N. L. P. A. S. Putri, N. M. S. Iswari, and I. G. P. K. Juliharta, “Prediksi Kesehatan Mental Pengguna Berdasarkan Konsumsi Konten Pada Media Sosial Menggunakan Metode Random Forest,” *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 4, no. 4, pp. 12007–12022, Jan. 2026, doi: 10.31004/riggs.v4i4.5440.
- [4] D. Nickson, C. Meyer, L. Walasek, and C. Toro, *Prediction and diagnosis of depression using machine learning with electronic health records data: a systematic review*, vol. 23, no. 1. BioMed Central, 2023. doi: 10.1186/s12911-023-02341-x.
- [5] J. O. Leandro and M. I. Fianty, “Evaluation of Sentiment Analysis Methods for Social Media Applications: A Comparison of Support Vector Machines and Naïve Bayes,” *JOIV Int. J. Informatics Vis.*, vol. 9, no. 2, p. 796, Mar. 2025, doi: 10.62527/joiv.9.2.2905.
- [6] A. Wibowo, A. F. N. Masruriyah, and S. Rahmawati, “Refining Diabetes Diagnosis Models: The Impact of SMOTE on SVM, Logistic Regression, and Naïve Bayes,” *J. Electron. Electromed. Eng. Med. Informatics*, vol. 7, no. 1, pp. 197–207, Jan. 2025, doi: 10.35882/jeeemi.v7i1.596.
- [7] J. J. Wibowo, D. A. Kristiyanti, and J. Wiratama, “Enhancing Heart Disease Classification: A Comparative Analysis of SMOTE and Naïve Bayes on Imbalanced Data,” *JOIV Int. J. Informatics Vis.*, vol. 9, no. 5, p. 2072, Sep. 2025, doi: 10.62527/joiv.9.5.3248.

- [8] Dedi Trisnawarman and Reza Mahendra, “Improving Vehicle Payment Method Classification Using XGBoost with SMOTE and SHAP Interpretation,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 10, no. 1, pp. 11–19, Feb. 2026, doi: 10.29207/resti.v10i1.6935.
- [9] U. Budiyanto, T. Fatimah, B. A. Pratama, and P. Korespondensi, “Prediksi Tingkat Depresi (Kesehatan Mental) Menggunakan Phq-9, Naïve Bayes, Random Forest Dan Support Vector Machine,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 13, no. 1, pp. 2355–7699, 2026, doi: 10.25126/jtiik.2026131.
- [10] K. Rahayu, V. Fitria, D. Septhya, R. Rahmaddeni, and L. Efrizoni, “Klasifikasi Teks untuk Mendeteksi Depresi dan Kecemasan pada Pengguna Twitter Berbasis Machine Learning,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 108–114, Sep. 2023, doi: 10.57152/malcom.v3i2.780.
- [11] I. S. Romadon and C. Kirana, “Analisis Komparatif Naive Bayes, Random Forest: Indikator Depresi Teks Pada Twitter,” *J. Inf. Syst. Manag.*, vol. 8, no. 1, pp. 70–74, Jul. 2026, doi: 10.24076/joism.2026v8i1.2547.
- [12] M. Sivakumar, S. Parthasarathy, and T. Padmapriya, “Trade-off between training and testing ratio in machine learning for medical image processing,” *PeerJ Comput. Sci.*, vol. 10, p. e2245, Sep. 2024, doi: 10.7717/peerj-cs.2245.
- [13] A. N. Cahyani, J. Zeniarja, S. Winarno, R. T. E. Putri, and A. A. Maulani, “Heart Disease Classification Using Deep Neural Network with SMOTE Technique for Balancing Data,” *Adv. Sustain. Sci. Eng. Technol.*, vol. 6, no. 1, p. 0240108, Dec. 2023, doi: 10.26877/asset.v6i1.17521.
- [14] S. O. Ngesthi and L. A. Wulandhari, “Cassava Diseases Classification using EfficientNet Model with Imbalance Data Handling,” *J. Online Inform.*, vol. 9, no. 2, pp. 148–158, Aug. 2024, doi: 10.15575/join.v9i2.1300.
- [15] L. P. Wanti, N. W. Adi Prasetya, I. Awaludin, M. B. Aditya Saputra, S. N. Furi, and D. M. Dwi Kumara, “Application of data mining for diagnosis of ENT diseases using the Naïve Bayes method with genetic algorithm feature selection,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 37, no. 1, p. 398, Jan. 2025, doi: 10.11591/ijeecs.v37.i1.pp398-405.
- [16] M. K. Zuhanda, Lisya Permata, Hartono, Erianto Ongko, and Desniarti, “Impact of Adaptive Synthetic on Naïve Bayes Accuracy in Imbalanced Anemia Detection Datasets,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 9, no. 1, pp. 85–93, Jan. 2025, doi: 10.29207/resti.v9i1.6031.



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)
