

# Prediksi Risiko Depresi pada Mahasiswa Menggunakan Algoritma Random Forest Berdasarkan Data Akademik dan Gaya Hidup

Diterima:

10 Juni 2025

Revisi:

19 Juli 2019

Terbit:

25 Desember 2025

<sup>a</sup>Imam Syafi'i Maulana, <sup>a</sup>Sherly Dian Tiara, <sup>a</sup>Ady Yoga Pratama,

<sup>a</sup>Fifin Kumala Sari, <sup>b</sup>Irwan Darmawan

<sup>a</sup>Universitas Nusantara PGRI Kediri

<sup>b</sup>Universitas Madura

**Abstrak— Latar Belakang:** Depresi menjadi isu krusial yang semakin sering dialami oleh mahasiswa akibat tekanan akademik, gaya hidup, dan tuntutan sosial. Deteksi dini sangat mendesak karena depresi yang tidak ditangani dapat berdampak buruk pada performa akademik, hubungan sosial, dan kualitas hidup mahasiswa. **Tujuan:** Penelitian ini bertujuan untuk membangun model prediksi risiko depresi yang akurat dan objektif pada mahasiswa dengan menggunakan algoritma Random Forest berbasis data akademik dan gaya hidup. **Metode:** Penelitian ini merupakan studi kuantitatif yang menerapkan metodologi Knowledge Discovery in Databases (KDD). Data yang digunakan adalah Student Depression Dataset dari Kaggle. Prosesnya meliputi preprocessing data, penyeimbangan data dengan SMOTEENN dan label encoding untuk mengubah data kategorikal menjadi numerik. **Hasil:** Model yang dikembangkan menunjukkan performa yang sangat tinggi pada saat diuji. Hasil evaluasi pada data uji menghasilkan nilai akurasi, precision, recall, dan f1-score yang seragam, yaitu sebesar 97%. **Kesimpulan:** Model Random Forest terbukti efektif untuk mendeteksi risiko depresi secara akurat dan seimbang. Model ini berpotensi menjadi alat bantu yang berharga bagi institusi pendidikan untuk melakukan intervensi preventif yang tepat sasaran. Penelitian selanjutnya dapat berfokus pada validasi model menggunakan data institusional di dunia nyata.

**Kata Kunci—** Depresi; Mahasiswa; Random Forest; Kesehatan Mental

**Abstract— Background:** Depression has become a crucial issue increasingly experienced by students due to academic pressure, lifestyle, and social demands. The urgency of early detection is very high, as failure to address this condition can negatively impact academic performance, social relationships, and students' overall quality of life. **Objective:** The aim of this research is to develop an accurate and objective depression risk prediction model using a Random Forest algorithm based on students' academic and lifestyle data. **Methods:** This research is a quantitative study applying the Knowledge Discovery in Databases (KDD) methodology. The data used is the Student Depression Dataset from Kaggle. The process includes data preprocessing, data balancing with SMOTEENN, and label encoding to convert categorical data to numeric data. **Results:** The developed model showed very high-performance during testing. The model evaluation results on the test data show an accuracy, precision, recall, and f1-score consistently at 97%. **Conclusion:** The Random Forest model proved to be effective in accurately and evenly detecting the risk of depression. It can be a helpful tool for educational institutions to conduct targeted preventive interventions. Future studies should focus on validating this model with real-world institutional data.

**Keywords—** Depression; Students; Random Forest; Mental Health

This is an open access article under the CC BY-SA License.



---

## Penulis Korespondensi:

Imam Syafi'i Maulana,  
Sistem Informasi,  
Universitas Nusantara PGRI Kediri,  
Email: [imamsmaulana07@gmail.com](mailto:imamsmaulana07@gmail.com)

---

## I. PENDAHULUAN

Depresi merupakan salah satu gangguan psikologis yang paling sering terjadi pada remaja dan mahasiswa di berbagai level pendidikan. Faktor-faktor pemicu utamanya meliputi tekanan akademik yang tinggi, perubahan gaya hidup yang drastis, serta tuntutan sosial yang terus meningkat di kalangan mahasiswa (Dianovinina, 2018). Kondisi ini, apabila tidak dideteksi dan ditangani secara dini, dapat berdampak serius tidak hanya pada performa akademik dan hubungan sosial, tetapi juga pada kualitas hidup mahasiswa secara keseluruhan. Institusi pendidikan sebenarnya memiliki akses terhadap berbagai data yang berpotensi menjadi indikator awal gejala depresi, seperti nilai akademik, kehadiran, dan kebiasaan hidup lainnya (Putri et al., 2022).

Namun, dalam konteks pendidikan saat ini, identifikasi risiko depresi masih banyak bergantung pada metode konvensional seperti observasi manual atau pengisian kuesioner yang subjektif dan kurang efektif. Padahal, institusi pendidikan sebenarnya memiliki akses terhadap berbagai data yang berpotensi menjadi indikator awal munculnya gejala depresi, seperti nilai akademik, kehadiran, pola tidur, dan kebiasaan hidup mahasiswa lainnya. Kesenjangan inilah yang coba diatasi oleh kemajuan dalam bidang ilmu data, khususnya (Wijaya & Lunanta, 2024).

Kemajuan dalam bidang ilmu data, khususnya *data mining*, membuka peluang baru dalam analisis masalah-masalah psikologis berbasis data. Salah satu metode yang terbukti efektif dalam klasifikasi dan prediksi adalah algoritma *Random Forest*, yaitu algoritma ansambel yang mampu menangani data dengan banyak variabel, menangkap pola-pola kompleks, serta memberikan interpretasi terhadap pentingnya tiap fitur (Health et al., 2025). *Machine learning* membandingkan beberapa algoritma termasuk Random Forest dan menemukan bahwa Random Forest memiliki kinerja yang kompetitif dalam memprediksi depresi mahasiswa (Fadhilla et al., 2025).

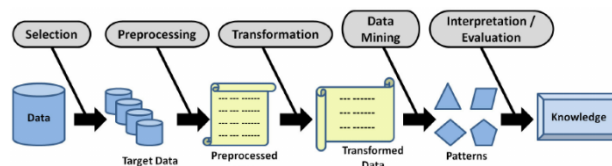
Perbedaan penelitian ini dengan penelitian sebelumnya adalah fokus pada penerapan proses *Knowledge Discovery in Databases* (KDD) yang komprehensif, mulai dari *preprocessing* data yang teliti dengan teknik *hybrid* SMOTE untuk mengatasi data tidak seimbang, hingga evaluasi model yang mendalam. Sementara penelitian lain sering kali hanya membandingkan performa algoritma, penelitian ini membangun sebuah alur kerja utuh yang dapat direplikasi untuk menghasilkan model prediksi yang tidak hanya akurat tetapi juga seimbang dalam mengenali kelas minoritas. Keunikan lainnya adalah pemanfaatan *Student Depression Dataset* dari Kaggle yang mencakup variabel gaya hidup dan akademik secara bersamaan, sehingga memberikan pandangan yang lebih holistik.

Tujuan penelitian ini adalah untuk membangun model prediksi risiko depresi pada mahasiswa menggunakan algoritma Random Forest dengan memanfaatkan data dari platform publik Kaggle tersebut. Dengan demikian, diharapkan institusi pendidikan dapat memiliki alat bantu dalam

mendeteksi risiko depresi secara lebih cepat, akurat, dan objektif, serta mendukung upaya intervensi dini bagi mahasiswa yang berisiko.

## II. METODE

Penelitian ini menerapkan pendekatan *Knowledge Discovery in Databases* (KDD) untuk menemukan pola dan pengetahuan yang tersembunyi dari data. KDD merupakan sebuah proses multi-tahap yang digunakan dalam penambangan data untuk menghasilkan pengetahuan yang dapat mendukung pengambilan keputusan. Seluruh proses pengolahan data dan pemodelan dalam penelitian ini diimplementasikan menggunakan bahasa pemrograman Python dengan dukungan dari platform Google Colab. Diagram Penelitian yang mengilustrasikan alur kerja KDD disajikan pada Gambar 1.



Gambar 1. Alur Knowledge Discovery In Databases (Aqillah Fadia Haya et al., 2022).

Penelitian ini akan mengolah data menggunakan bahasa pemrograman Python yang didukung oleh Google Colab. Untuk memudahkan pemahaman, akan dijelaskan langkah-langkah dalam proses KDD., berikut ini:

### 2.1 Data Section

Pemilihan data harus dilakukan sebelum melanjutkan ke tahap berikutnya. Pemilihan data adalah langkah di mana proses untuk menentukan data atau atribut yang akan dipakai dalam penelitian dilakukan.

### 2.2 Data Preprocessing

Data Preprocessing perlu dilakukan sebelum melanjutkan ke tahap KDD selanjutnya. Proses Data Preprocessing adalah kegiatan membersihkan data atau yang kerap disebut sebagai cleansing, di mana pada fase ini, data yang tidak berguna akan dihapus dan data yang akan digunakan dalam proses penambangan data akan diperbaiki.

### 2.3 Data Transformation

Tahap berikutnya setelah preprocessing adalah transformasi, yaitu mengubah format data agar bisa dimanfaatkan (Pradikdo & Ristyan, 2018). Proses transformasi mencakup pemrosesan skala data, normalisasi, dan pengurangan dimensi (Khoirunnisa Hamidah & Voutama, 2023). Encoding juga merupakan bagian dari transformasi data, yaitu mengkonversi data kategorikal menjadi data numerik. Hal ini penting karena algoritma Random Forest tidak dapat memproses data dengan tipe kategorikal.

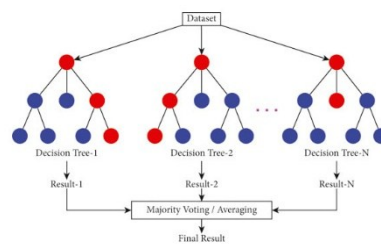
### 2.4 Data Mining

Penggalan data adalah suatu proses untuk menemukan informasi yang istimewa dalam sekumpulan data dengan memanfaatkan metode dan algoritma tertentu (Karimah et al., 2024).

Pada tahap ini, akan dilakukan pembentukan model untuk meramalkan faktor-faktor penyebab depresi dengan menggunakan algoritma dalam penambangan data, yaitu Random Forest.

## 2.5 Algoritma Random Forest

Random forest ialah sekumpulan pohon klasifikasi atau regresi tanpa mengalami pemangkasan serta dibentuk melalui metode mengambil contoh yang dipilih secara acak dari dataset. Proses perkiraan hasil dilakukan menggunakan cara menyatukan seluruh pohon dari regresi dan klasifikasi yang ada. Metode hutan acak mempunyai beberapa kelebihan berupa kemampuan mengenali kekeliruan yang signifikan, hasil proses pengelompokan yang memuaskan, dapat mengatasi dataset yang memiliki sedikit sampel, serta menjadi metode yang efisien untuk memperkirakan data yang hilang.



Gambar 2. Random Forest (Prakoso Indaryono, 2024).

Alur dari metode Random Forest pada gambar 2 (1) memanfaatkan pendekatan proses pemilihan data contoh secara acak guna membangun masing-masing pohon keputusan, (2) setiap pohon memanfaatkan sekelompok atribut yang diambil secara acak dalam hasil perkiraan, serta (3) mengumpulkan output estimasi dari seluruh model pohon yang dimiliki menggunakan cara memilih suara yang paling sering muncul untuk klasifikasi atau menghitung nilai tengah regresi (Adrian et al., 2021).

Singkatnya, metode Random Forest beroperasi dengan cara membuat berbagai decision tree yang beragam serta mengumpulkan prediksi akhir dari tiap-tiap pohon dengan cara suara terbanyak baik dalam klasifikasi maupun dengan menghitung nilai rata-rata pada regresi (Ristyan et al., 2025).

## III. HASIL DAN PEMBAHASAN

Di dalam penelitian ini, digunakan pendekatan yang mirip dengan penelitian sebelumnya, namun terdapat beberapa perbedaan, terutama pada jenis data dan tahapan pengolahannya. Data penelitian diambil dari situs Kaggle, yaitu dataset berjudul "*Student Depression Dataset*" (<https://www.kaggle.com/datasets/hopesb/student-depression-dataset/data>), yang berisi informasi mengenai tekanan akademik, tekanan kerja, IPK, durasi tidur, pola makan, kepuasan belajar, stres finansial, serta status depresi siswa sebagai label target (Setiawan et al., 2025). Total data yang digunakan sebanyak 3.929 entri dengan label biner, yaitu 0 untuk tidak depresi dan 1 untuk depresi.

### 3.1 Data Section

Tabel 1. Nama Fitur

| Nama Fitur                            | Deskripsi  |
|---------------------------------------|--|
| Id                                    | ID unik untuk masing-masing responden                                |
| Gender                                | Jenis kelamin responden (Male/Female/Other)                          |
| Age                                   | Usia responden (numerik)   |
| City                                  | Kota tempat tinggal responden  |
| Profession                            | Status profesi (Mahasiswa, Pekerja, dsb.)                            |
| Academic Pressure                     | Tingkat tekanan akademik yang dirasakan (Low/Medium/High)            |
| Work Pressure                         | Tingkat tekanan kerja (jika bekerja) (Low/Medium/High)               |
| CGPA                                  | Nilai IPK sebagai indikator performa akademik (0–4)                  |
| Study Satisfaction                    | Tingkat kepuasan terhadap kegiatan belajar (Low/Medium/High)         |
| Job Satisfaction                      | Tingkat kepuasan terhadap pekerjaan (jika bekerja) (Low/Medium/High) |
| Sleep Duration                        | Lama tidur setiap hari (dalam jam)                                   |
| Dietary Habits                        | Pola makan (Healthy/Unhealthy/Moderate)                              |
| Degree                                | Jenjang pendidikan (Bachelor, Master, dll.)                          |
| Have you ever had suicidal thoughts ? | Riwayat pemikiran bunuh diri (Yes/No)                                |
| Work/Study Hours                      | Lama jam belajar atau bekerja per hari                               |
| Financial Stress                      | Tingkat tekanan finansial (Low/Medium/High)                          |
| Family History of Mental Illness      | Riwayat gangguan mental dalam keluarga (Yes/No)                      |

Sesudah melaksanakan analisis berdasarkan seluruh fitur tabel 1, diperoleh hasil bahwa semua fitur yang tersedia akan diaplikasikan pada tahap penambangan data. Hal ini dikarenakan setiap atribut memiliki potensi sebagai faktor yang memengaruhi risiko depresi pada siswa. Empat urutan awal dari kumpulan data ditampilkan pada Gambar 3.

| id | Gender | Age    | City | Profession    | Academic Pressure | Work Pressure | CGPA | Study Satisfaction | Job Satisfaction | Sleep Duration | Dietary Habits    | Degree   | Have you ever had suicidal thoughts ? | Work/Study Hours | Financial Stress | Family History of Mental Illness | Depression |   |
|----|--------|--------|------|---------------|-------------------|---------------|------|--------------------|------------------|----------------|-------------------|----------|---------------------------------------|------------------|------------------|----------------------------------|------------|---|
| 0  | 2      | Male   | 33.0 | Visakhapatnam | Student           | 5.0           | 0.0  | 8.97               | 2.0              | 0.0            | 5-6 hours         | Healthy  | B.Pharm                               | Yes              | 3.0              | 1.0                              | No         | 1 |
| 1  | 8      | Female | 24.0 | Bangalore     | Student           | 2.0           | 0.0  | 5.90               | 5.0              | 0.0            | 5-6 hours         | Moderate | BSc                                   | No               | 3.0              | 2.0                              | Yes        | 0 |
| 2  | 26     | Male   | 31.0 | Srinagar      | Student           | 3.0           | 0.0  | 7.03               | 5.0              | 0.0            | Less than 5 hours | Healthy  | BA                                    | No               | 9.0              | 1.0                              | Yes        | 0 |
| 3  | 30     | Female | 28.0 | Varanasi      | Student           | 3.0           | 0.0  | 5.59               | 2.0              | 0.0            | 7-8 hours         | Moderate | BCA                                   | Yes              | 4.0              | 5.0                              | Yes        | 1 |
| 4  | 32     | Female | 25.0 | Jaipur        | Student           | 4.0           | 0.0  | 8.13               | 3.0              | 0.0            | 5-6 hours         | Moderate | M.Tech                                | Yes              | 1.0              | 1.0                              | No         | 0 |

Gambar 3. 4 baris pertama dataset.

### 3.2 Data Preprocessing

Tahapan ini mencakup pemeriksaan pada duplikat, nilai yang hilang, *drop* kolom *null values* pada dataset *Student Depression Dataset*. Setelah dilakukan analisis ada tahapan pra-pemrosesan yang diterapkan meliputi pemeriksaan serta *missing value*, dan informasi *imbalance* (Schutijser et al., 2020).

#### 3.2.1 Data Duplikat

Prosedur ini bertujuan menjamin tidak terdapat data ganda dalam dataset, agar proses analisis menjadi lebih efisien (Nugraha et al., 2022). Pemeriksaan duplikasi data ditampilkan pada gambar 4.

```
## duplicates.
data.duplicated().sum()
np.int64(0)
```

Gambar 4. Pengecekan data duplikasi

#### 3.2.2 Missing Value

Kemudian, dilakukan verifikasi bahwa *dataset* bebas dari nilai hilang; hasilnya terlihat pada gambar 5.

```
## null values.
null_col = data.isna().sum() / data.shape[0] * 100
null_col[null_col > 0].to_frame().rename(columns = {'0': 'null_values'})
```

| null_values      |          |
|------------------|----------|
| Financial Stress | 0.010752 |

Gambar 5. Pengecekan missing value

Ditemukan hasil bahwa dataset memiliki *Missing value*. Tahap selanjutnya berupa penanggulangan *missing value* yaitu menyelesaikan permasalahan yang divisualisasikan pada gambar 6 dan gambar 7.

```
## drop null values.
data.dropna(inplace = True)
data.shape
```

(27898, 15)

Gambar 6. Penanganan *missing value*

#### 3.2.3 Imbalance Dataset

| proportion |       |
|------------|-------|
| Depression |       |
| 1          | 0.585 |
| 0          | 0.415 |

Gambar 7. Hasil pengecekan *dataset imbalance*

Berdasarkan analisis awal terhadap distribusi kelas pada atribut target Depression, diketahui bahwa dataset bersifat tidak seimbang. Kelas 1 (siswa yang mengalami depresi) memiliki proporsi sebesar **58,5%**, sedangkan kelas 0 (siswa yang tidak mengalami depresi) hanya sebesar **41,5%**. Distribusi yang tidak merata ini berpotensi menjadikan model pembelajaran mesin cenderung berat sebelah kekelas dominan, yang berdampak pada penurunan kemampuan model dalam mengidentifikasi kelompok data yang lebih sedikit secara akurat (Arifiyanti & Wahyuni, 2020).

Dalam rangka menyelesaikan permasalahan ini, dihadapi dengan teknik SMOTEENN (*Synthetic Minority Over-sampling Technique Edited Nearest Neighbours*), yaitu metode gabungan antara oversampling pada kelas minoritas dan undersampling pada data yang dianggap noise atau kurang informatif. SMOTE berfungsi untuk menambah sampel sintetis dari kelas minoritas, sedangkan ENN menghapus data yang ambigu pada kedua kelas.

```
Resampled dataset shape: Depression
0    0.523446
1    0.476554
Name: proportion, dtype: float64
```

Gambar 8. *Dataset balance*

Setelah diterapkan SMOTEENN pada gambar 8, distribusi kelas menjadi lebih seimbang, dengan proporsi kelas 0 sebesar **52,34%** dan kelas 1 sebesar **47,65%**. Hal ini menunjukkan bahwa proses resampling berhasil mengurangi dominasi kelas mayoritas dan membuat model lebih mampu melakukan klasifikasi secara adil dan akurat terhadap kedua kelas.

### 3.3 Data Transformation

Pengkodean label digunakan guna mengubah informasi kategorikal diubah menjadi format data angka. Proses ini diperlukan karena algoritma Random Forest dan metode SMOTEENN hanya dapat memproses data numerik. *Encoding* dilakukan hanya pada variabel independen, sementara label target tetap tidak diubah. Hasil encoding ditampilkan pada Gambar 9.

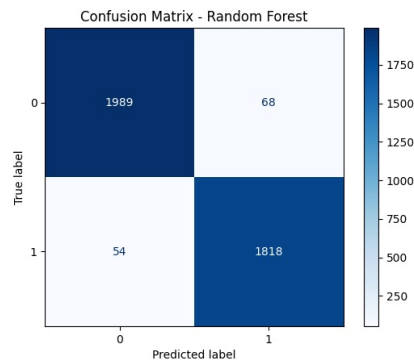
|   | Gender | Age | Academic_Pressure | Gpa  | Study_Satisfaction | Sleep_Duration | Dietary_Habits | Suicidal_Thoughts | WrkStdy_Hours | Financial_Stress | Family_Mental_History | Depression |
|---|--------|-----|-------------------|------|--------------------|----------------|----------------|-------------------|---------------|------------------|-----------------------|------------|
| 0 | 0      | 24  | 2                 | 5.90 | 5                  | 0              | 1              | 0                 | 3             | 2                | 1                     | 0          |
| 1 | 1      | 31  | 3                 | 7.03 | 5                  | 2              | 0              | 0                 | 9             | 1                | 1                     | 0          |
| 2 | 0      | 25  | 4                 | 8.13 | 3                  | 0              | 1              | 1                 | 1             | 1                | 0                     | 0          |
| 3 | 1      | 29  | 2                 | 5.70 | 3                  | 2              | 0              | 0                 | 4             | 1                | 0                     | 0          |
| 4 | 1      | 30  | 3                 | 9.54 | 4                  | 1              | 0              | 0                 | 1             | 2                | 0                     | 0          |

Gambar 9. Hasil encoding Dataset

### 3.4 Data Mining

Kumpulan data *Student Depression Dataset* termasuk jenis klasifikasi karena memiliki atribut numerik dan kategorikal dengan label biner (0 = tidak depresi, 1 = depresi). Penelitian ini bertujuan mengklasifikasikan siswa ke dalam dua kategori tersebut. Sebanyak 80% data digunakan sebagai data latih setelah melalui proses preprocessing dan penyeimbangan. Algoritma yang diterapkan adalah Random Forest karena keandalannya dalam menangani informasi yang rumit dengan performa yang baik (Herdika et al., 2024).

Setelah model dibangun dan dilatih, dilakukan evaluasi terhadap data uji menggunakan confusion matrix, yang hasilnya ditampilkan pada Gambar 10.



**Gambar 10.** Confusion matrix

Berdasarkan confusion matrix, diperoleh hasil sebagai berikut:

- **True Positive (TP):** 1818 (siswa depresi yang berhasil diprediksi dengan benar),
- **True Negative (TN):** 1989 (siswa tidak depresi yang diperkirakan dengan akurat),
- **False Positive (FP):** 68 (siswa tidak depresi yang keliru diprediksi sebagai depresi),
- **False Negative (FN):** 54 (siswa depresi yang salah diprediksi sebagai tidak depresi).

Hasil ini menunjukkan bahwa model memiliki tingkat akurasi dan ketepatan klasifikasi yang tinggi dalam memprediksi status depresi siswa.

### 3.5. Evaluation

Evaluation Setelah itu diukur Sensitifitas, Spesifisitas, Presisi, serta area yang luas Under the Curve (AUC), Accuracy terhadap algoritma Random Forest yang telah diterapkan, Tabel 2 merupakan hasil pengukuran dari pengujian terhadap algoritma Random Forest.

**Tabel 2. Accuracy**

| Metrik Keseluruhan | Nilai |
|--------------------|-------|
| Accuracy           | 0.97  |
| Macro Avg          | 0.97  |
| Weighted Avg       | 0.97  |

Hasil penelitian ini sejalan atau didukung oleh beberapa studi sebelumnya yang juga menyoroti keunggulan Random Forest dalam konteks kesehatan mental. Sebagai contoh, penelitian oleh Fadhilla dkk (Fadhilla et al., 2025). juga menemukan bahwa Random Forest menunjukkan kinerja yang sangat kompetitif dan bahkan unggul dibandingkan algoritma lain seperti Decision Tree dan Naïve Bayes untuk prediksi depresi mahasiswa. Hal ini memperkuat justifikasi pemilihan algoritma dalam penelitian ini dan mengonfirmasi bahwa Random Forest adalah pilihan yang solid untuk tugas klasifikasi pada data psikologis yang kompleks. Keberhasilan dalam mencapai akurasi tinggi juga menegaskan pentingnya tahapan *preprocessing* yang teliti, terutama penyeimbangan data dengan SMOTEENN yang terbukti efektif mengatasi bias pada model.



#### IV KESIMPULAN

Penelitian ini berhasil membangun sebuah model prediksi risiko depresi pada mahasiswa yang sangat efektif menggunakan algoritma Random Forest. Berdasarkan hasil penelitian, model yang dikembangkan mampu mencapai performa luar biasa dengan akurasi, presisi, dan *f1-score* seragam sebesar 97% setelah melalui proses KDD yang sistematis, termasuk penanganan *missing value* dan penyeimbangan data menggunakan SMOTEENN. Meskipun demikian, penelitian ini memiliki keterbatasan, yaitu penggunaan dataset publik dari Kaggle yang mungkin tidak sepenuhnya merepresentasikan keragaman populasi mahasiswa secara global atau di institusi spesifik, serta data yang bersifat *self-reported* yang dapat mengandung bias subjektif. Untuk penelitian selanjutnya, disarankan untuk melakukan validasi model menggunakan data riil dari institusi pendidikan untuk menguji generalisasinya. Selain itu, penelitian di masa depan dapat mengeksplorasi penambahan fitur lain seperti data aktivitas media sosial atau analisis sentimen, serta membandingkan performa Random Forest dengan algoritma *deep learning* untuk melihat potensi peningkatan akurasi.

#### DAFTAR PUSTAKA

- Adrian, M. R., Putra, M. P., Rafialdy, M. H., & Rakhmawati, N. A. (2021). Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB. *Jurnal Informatika Upgris*, 7(1), 36–40. <https://doi.org/10.26877/jiu.v7i1.7099>
- Aqillah Fadia Haya, A., Reynaldi Azhar, Muhamad Khandava Mulyadien, & Betha Nurina Sari. (2022). Klasifikasi Minat Beli Pelanggan Terhadap Udang Vaname Menggunakan Algoritma Naïve Bayes. *Jurnal Ilmiah Betrik*, 13(1), 59–65. <https://doi.org/10.36050/betrik.v13i1.452>
- Arifiyanti, A. A., & Wahyuni, E. D. (2020). Smote: Metode Penyeimbang Kelas Pada Klasifikasi Data Mining. *SCAN - Jurnal Teknologi Informasi Dan Komunikasi*, 15(1), 34–39. <https://doi.org/10.33005/scan.v15i1.1850>
- Dianovinina, K. (2018). Depresi pada Remaja: Gejala dan Permasalahannya. *Journal Psikogenesis*, 6(1), 69–78. <https://doi.org/10.24854/jps.v6i1.634>
- Fadhilla, M., Wandri, R., Hanafiah, A., Setiawan, P. R., Arta, Y., & Daulay, S. (2025). Analisis Performa Algoritma Machine Learning Untuk Identifikasi Depresi Pada Mahasiswa. 5(1), 40–47. <https://doi.org/10.47065/jimat.v5i1.473>
- Health, M., Mental, K., & Technique, S. M. O. (2025). PREDIKSI KESEHATAN MENTAL MENGGUNAKAN RANDOM FOREST BERDASARKAN FAKTOR DEMOGRAFI DAN LINGKUNGAN KERJA. 9(3), 4515–4522.
- Herdika, E., Aulia, S., Azzahra, D., Ristyawan, A., Daniati, E., Teknik, F., & Informasi, S. (2024). Klasifikasi Risiko Kambuhnya Kanker Tiroid Menggunakan Algoritma Random Forest. 8, JSITIK: Jurnal Sistem Informasi dan Teknologi Informasi Komputer | Hal:1-10

63–74.

- Karimah, A., Dwilestari, G., & Mulyawan, M. (2024). Analisis Sentimen Komentar Video Mobil Listrik Di Platform Youtube Dengan Metode Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 767–737. <https://doi.org/10.36040/jati.v8i1.8373>
- Khoirunnisa Hamidah, & Voutama, A. (2023). Analisis Faktor Tingkat Kebahagiaan Negara Menggunakan Data World Happiness Report dengan Metode Regresi Linier. *Explore IT: Jurnal Keilmuan Dan Aplikasi Teknik Informatika*, 15(1), 1–7. <https://doi.org/10.35891/explorit.v15i1.3874>
- Nugraha, A., Nurdiawan, O., & Dwilestari, G. (2022). Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 849–855. <https://doi.org/10.36040/jati.v6i2.5755>
- Pradikdo, A. C., & Ristyawan, A. (2018). Model Klasifikasi Abstrak Skripsi Menggunakan Text Mining Untuk Pengkategorian Skripsi Sesuai Bidang Kajian. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 9(2), 1091–1098. <https://doi.org/10.24176/simet.v9i2.2513>
- Prakoso Indaryono, N. A. (2024). Analisa Perbandingan Algoritma Random Forest Dan Naïve Bayes Untuk Klasifikasi Curah Hujan Berdasarkan Iklim Di Indonesia. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 9(1), 158–167. <https://doi.org/10.29100/jupi.v9i1.4421>
- Putri, F. S., Nazihah, Z., Ariningrum, D. P., Celesta, S., & Kharin Herbawani, C. (2022). Depresi Remaja di Indonesia: Penyebab dan Dampaknya Adolescent Depression in Indonesia: Causes and Effects. *Jurnalkesehatanpoltekkeskemenkesripangkalpinang*, 10(2)(2), 99–108.
- Ristyawan, A., Nugroho, A., & Amarya, T. K. (2025). *Optimasi Preprocessing Model Random Forest Untuk Prediksi Stroke*. 12(1).
- Schutijser, B. C. F. M., Klopotoska, J. E., Jongerden, I. P., Wagner, C., & de Bruijne, M. C. (2020). Feasibility of reusing routinely recorded data to monitor the safe preparation and administration of injectable medication: A multicenter cross-sectional study. *International Journal of Medical Informatics*, 141(May), 104201. <https://doi.org/10.1016/j.ijmedinf.2020.104201>
- Setiawan, I., Yasin, I. F., Desianti, Y. T., & Surakarta, A. (2025). *Komparasi Kinerja Algoritma Random Forest , Decision Tree , Naïve Bayes , dan KNN dalam Prediksi Tingkat Depresi Mahasiswa Menggunakan Student Depression Dataset*. 6(1), 47–58.
- Wijaya, Y., & Lunanta, L. (2024). Stres, kecemasan, dan depresi pada mahasiswa di Jakarta. *Forum Ilmiah Indonusa*, 21(1), 26–31.