



Program Sarjana Ilmu Komputer

"Perbandingan Model Logistic Regression, Support Vector Machine, dan XGBoost dalam Klasifikasi Depresi Siswa"

Anisa Hayatullah (G6401221009)

Roshan Zakaria (G6401221010)

Sri Arini Ismayasari (G6401221029)

Nurul Fadillah (G6401221078)

A. Latar Belakang

- Masalah kesehatan mental yang banyak menyerang siswa yang saat ini menjadi isu yang serius.
- Siswa banyak mengalami masalah kesehatan mental karena adanya tekanan akademik, sosial media, krisis identitas serta ekspektasi sosial.
- Menurut WHO 1 dari 7 remaja mengalami masalah kesehatan mental namun tidak terdeteksi.
- Pentingnya deteksi dini untuk mencegah dampak panjang pada masa dewasa.



B. Tujuan & Manfaat

- Membandingkan performa dari 3 model machine learning yaitu XGBoost, SVM dan Logistic Regression.
- Membantu dalam proses deteksi dini terhadap gangguan kesehatan mental, khususnya depresi, berdasarkan data karakteristik siswa yang tersedia.
- Mendorong pemanfaatan machine learning di bidang kesehatan mental



C. Ruang Lingkup

- Menggunakan dataset "Student Depression" dari platform Kaggle dalam format CSV dengan 27.901 *rows*, 17 fitur, dan 1 target.
- Setiap data merepresentasikan satu siswa dengan atribut:
- Target: Depression (Yes/No)
- Fitur:
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



D. Batasan dan Fokus Penelitian

Fokus utama:

- Klasifikasi status depresi siswa
- Evaluasi performa model klasifikasi

Tidak termasuk ke dalam penelitian:

- Analisis mendalam mengenai faktor penyebab depresi, faktor geografis atau lokasi siswa, intervensi psikologis, maupun solusi klinis
- Memberikan diagnosis medis, melainkan hanya memanfaatkan pendekatan komputasi melalui teknik machine learning



E. Tinjauan Pustaka

Teknik Machine Learning untuk Prediksi Kesehatan Mental (T. Jain et al (2021))

- Membandingkan 8 model ML: DT, RF, SVM, NB, LR, XGB, GBC, ANN.
- Data yang digunakan: 1429
- Hasil terbaik: SVM dengan akurasi 87,83%

Penerapan XGBoost untuk Deteksi Depresi (Sharma dan Verbeke (2020))

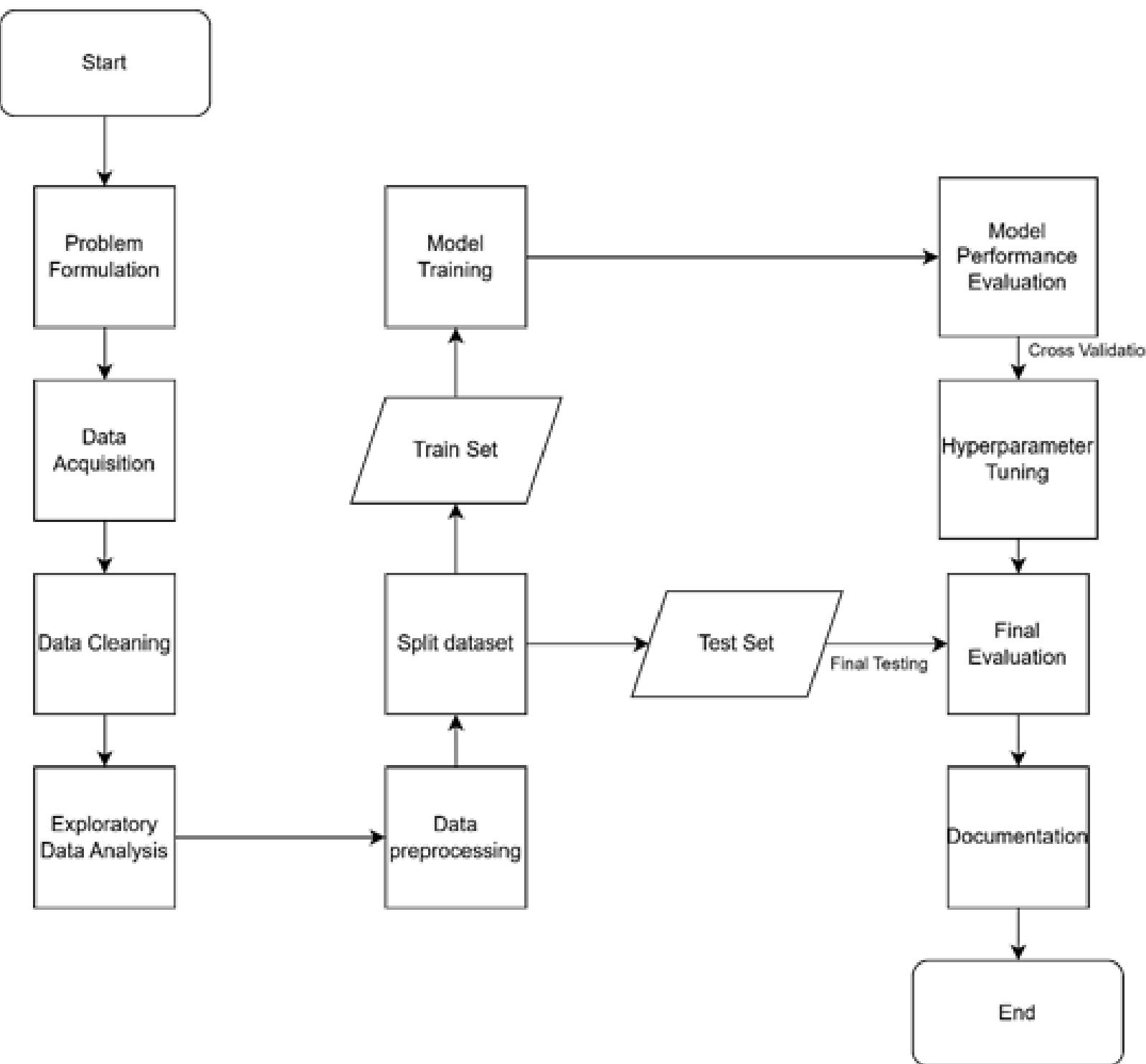
- Menggunakan data imbalanced menerapkan solusi berupa teknik resampling.
- Data yang digunakan: 11,081
- Model XGBoost mendapat hasil F1-Score, precision, recall > 0.90.

Perbandingan Kernel Support Vector Machine (SVM) (Aulia et al. (2021))

- Membandingkan empat kernel SVM
- SVM Linear dan Sigmoid meraih akurasi 87% dibandingkan RBF dan Polynomial.



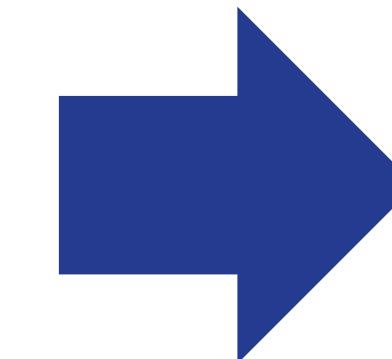
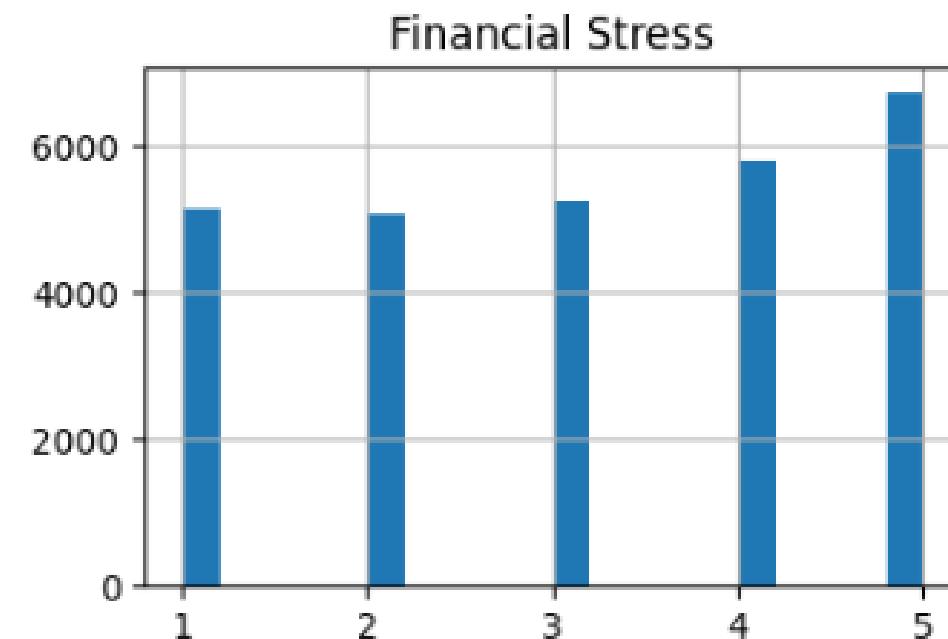
F. Metode (Langkah)



F. Metode (Langkah)

1. Penanganan *Missing Value*

| | |
|---------------------------------------|---|
| Degree | 0 |
| Have you ever had suicidal thoughts ? | 0 |
| Work/Study Hours | 0 |
| Financial Stress | 3 |
| Family History of Mental Illness | 0 |
| Depression | 0 |



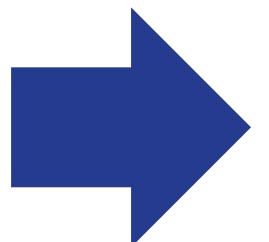
- Kolom *Financial Stress* bertipe kategorik dan Distribusi cukup rata
- Nilai Null tidak signifikan (3/27870)

Nilai null diganti dengan nilai modus *financial stress* (5)

F. Metode (Langkah)

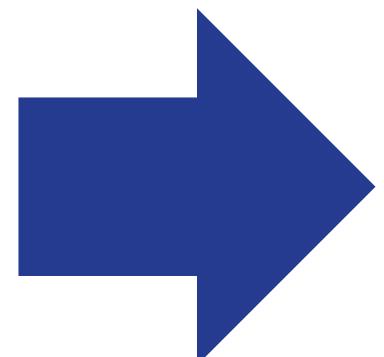
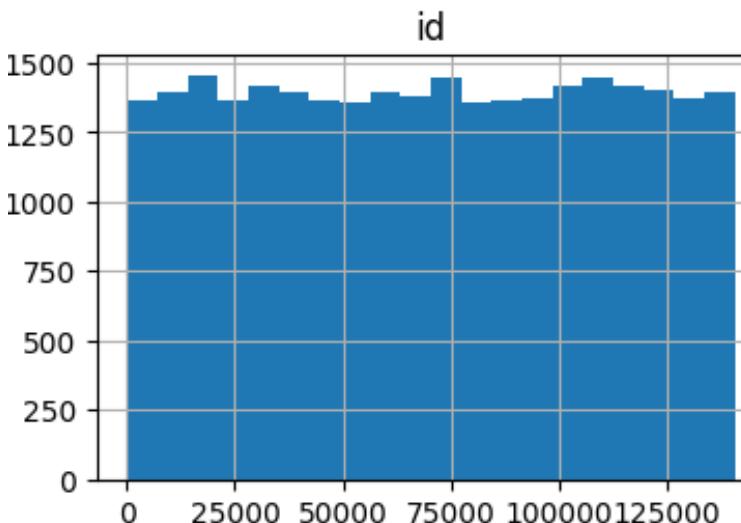
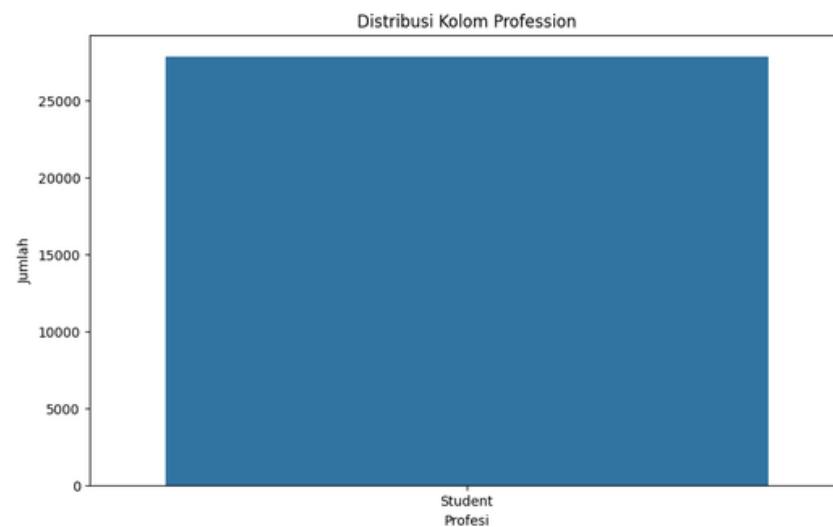
2. Deteksi Duplikasi

Jumlah data duplikat: 0



Tidak ada nilai null. seluruh entri bersifat unik

3. Seleksi Fitur



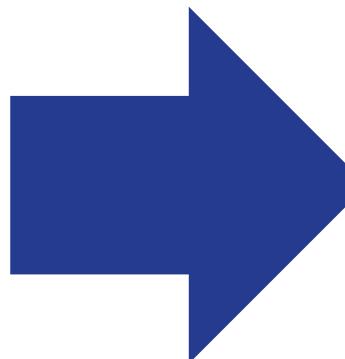
- Kolom “Profession” Homogen.
- Kolom “Id” Tidak memiliki informasi untuk model

Menghapus kolom “Profession” dan “id”



F. Metode (Langkah)

3. Seleksi Fitur

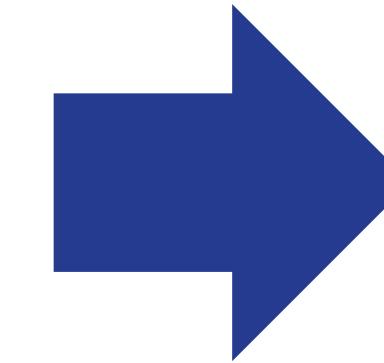
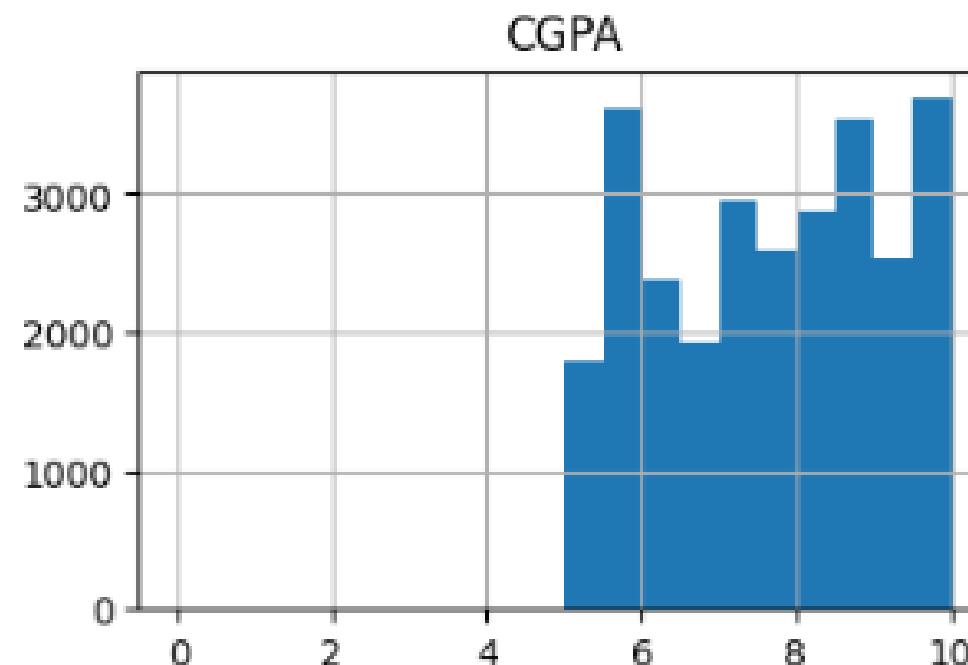
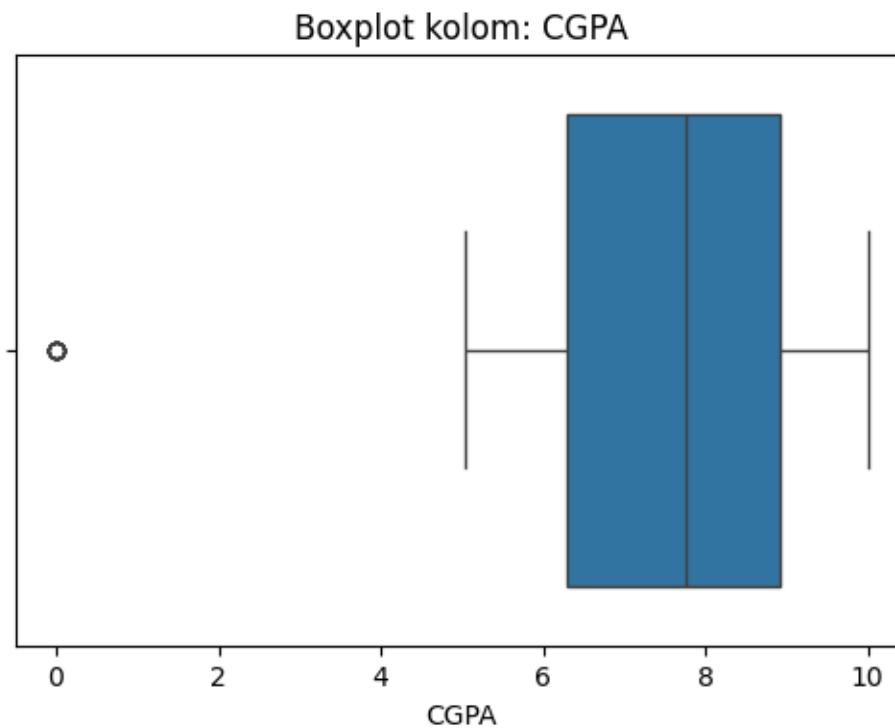


Menghapus work pressure dan Job Satisfaction

- Kolom "City" dihapus karena model yang dikembangkan tidak mempertimbangkan lokasi individu sebagai variabel yang relevan

F. Metode (Langkah)

4. Penanganan Outlier



Kolom CGPA memiliki 9 outlier dengan nilai 0, yang dianggap tidak wajar bagi siswa/mahasiswa aktif.

Nilai outlier dihapus agar analisis lebih akurat dan representatif

F. Metode (Langkah)

5. Encode Variabel Kategorik Ordinal

| Sleep Duration | Ordinal Encoding |
|-------------------|------------------|
| Others | 0 |
| Less than 5 hours | 1 |
| 5-6 hours | 2 |
| 7-8 hours | 3 |
| More than 8 hours | 4 |

Sleep Duration

| Dietary Habits | Ordinal Encoding |
|----------------|------------------|
| Others | 0 |
| Unhealthy | 1 |
| Moderate | 2 |
| Healthy | 3 |

Dietary Habits

| Degree | Ordinal Encoding |
|----------|------------------|
| Others | 0 |
| Class 12 | 1 |
| Bachelor | 2 |
| Magister | 3 |
| PhD | 4 |

Degree



F. Metode (Langkah)

6. Encode Variabel Kategorik Nominal

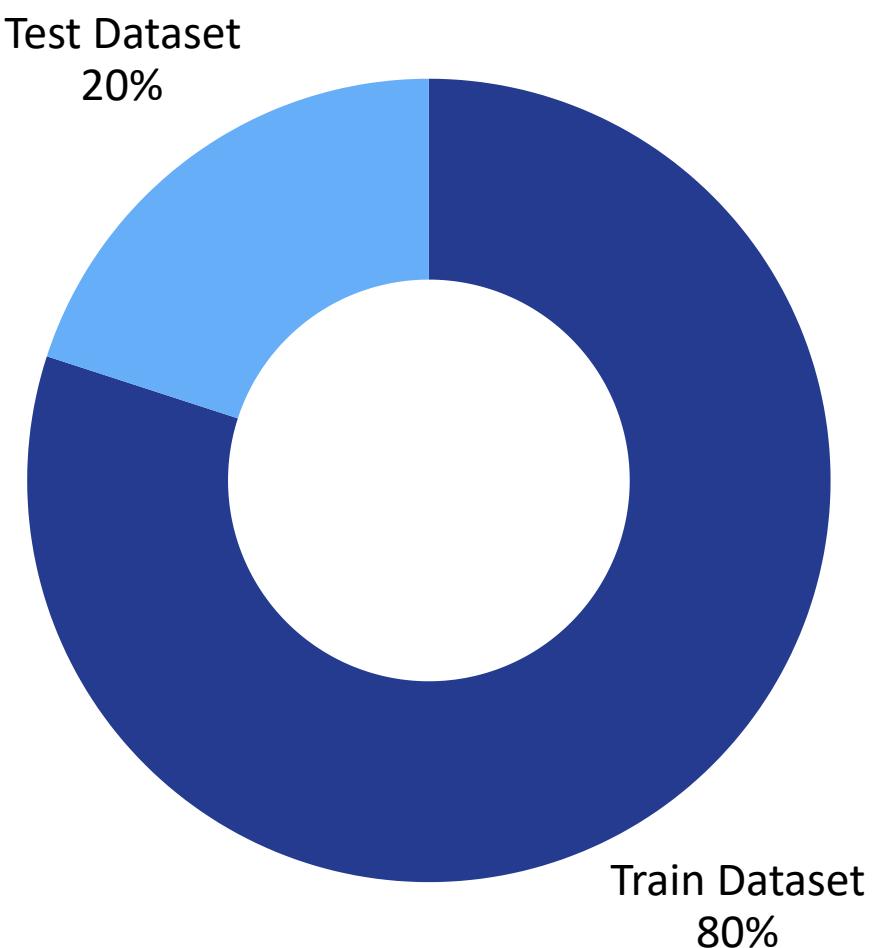
Variabel dengan tipe nominal di-*encode* menggunakan **One-Hot Encoding** (`get_dummies`) pada library pandas. Adapun hasilnya

- Gender → Gender_Male
- Have you ever had suicidal thoughts ? → Have you ever had suicidal thoughts ?_Yes
- Family History of Mental Illness → Family History of Mental Illness_Yes



F. Metode (Langkah)

7. Split dan Scaling Data



Standarisasi data dilakukan pada prediktor dengan menggunakan fungsi *standardscaler* dari *library sklearn python*.



G. Hasil & Pembahasan

| | Logistic Regression | SVM | XGBoost |
|--------------------------------------|----------------------------|--|--|
| Hyperparameter (Grid Search) | C = 0.1 | C = 0.01; gamma = 0.1; kernel = sigmoid | max_depth = 3; learning_rate = 0.1, subsample = 0.5 |
| Training accuracy | 84,68783% | 84,66989% | 85,2081% |
| Average cross-validation accuracy | 84,67434% | 84,66987% | 84,78647% |
| Prediction accuracy | 84,66092% | 84,49946% | 84,76856% |



G. Hasil & Pembahasan

| Confusion Matrix: | | | | |
|-----------------------------|-----------|--------|----------|---------|
| [[1844 503] [350 2877]] | | | | |
| Classification Report: | | | | |
| | precision | recall | f1-score | support |
| 0 | 0.84 | 0.79 | 0.81 | 2347 |
| 1 | 0.85 | 0.89 | 0.87 | 3227 |
| accuracy | | | 0.85 | 5574 |
| macro avg | 0.85 | 0.84 | 0.84 | 5574 |
| weighted avg | 0.85 | 0.85 | 0.85 | 5574 |

Logistic Regression

Precision dan recall seimbang di kedua kelas, dengan recall positif 89% dan F1-score positif 87%, menunjukkan performa baik dalam mendeteksi kasus depresi tanpa mengorbankan terlalu banyak akurasi pada kelas negatif.

| Confusion Matrix: | | | | |
|-----------------------------|-----------|--------|----------|---------|
| [[1807 540] [324 2903]] | | | | |
| Classification Report: | | | | |
| | precision | recall | f1-score | support |
| 0 | 0.85 | 0.77 | 0.81 | 2347 |
| 1 | 0.84 | 0.90 | 0.87 | 3227 |
| accuracy | | | 0.84 | 5574 |
| macro avg | 0.85 | 0.83 | 0.84 | 5574 |
| weighted avg | 0.85 | 0.84 | 0.84 | 5574 |

SVM Sigmoid

Menunjukkan recall tertinggi pada kelas positif (90%), sehingga sangat efektif dalam mendeteksi kasus depresi, namun sedikit mengorbankan recall pada kelas negatif (77%).

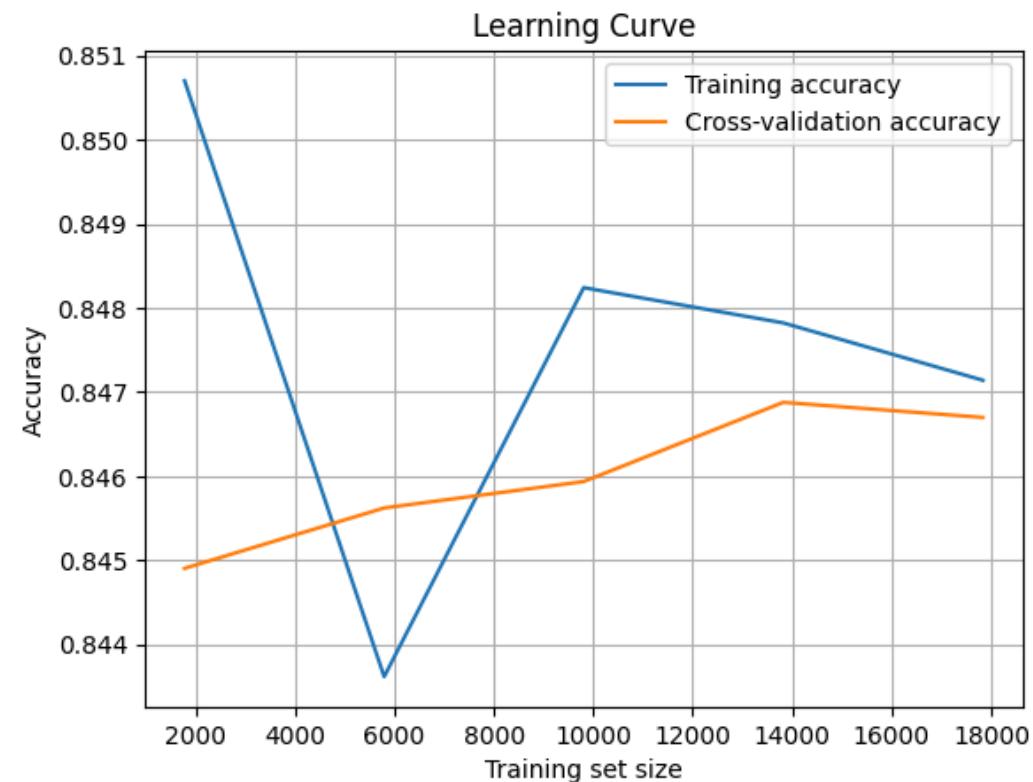
| Confusion Matrix: | | | | |
|-----------------------------|-----------|--------|----------|---------|
| [[1844 503] [346 2881]] | | | | |
| Classification Report: | | | | |
| | precision | recall | f1-score | support |
| 0 | 0.84 | 0.79 | 0.81 | 2347 |
| 1 | 0.85 | 0.89 | 0.87 | 3227 |
| accuracy | | | 0.85 | 5574 |
| macro avg | 0.85 | 0.84 | 0.84 | 5574 |
| weighted avg | 0.85 | 0.85 | 0.85 | 5574 |

XGBoost

Memiliki keseimbangan paling stabil antara precision dan recall di kedua kelas, dengan nilai F1-score yang konsisten, menunjukkan performa andal dan seimbang dalam klasifikasi kedua kelas target.

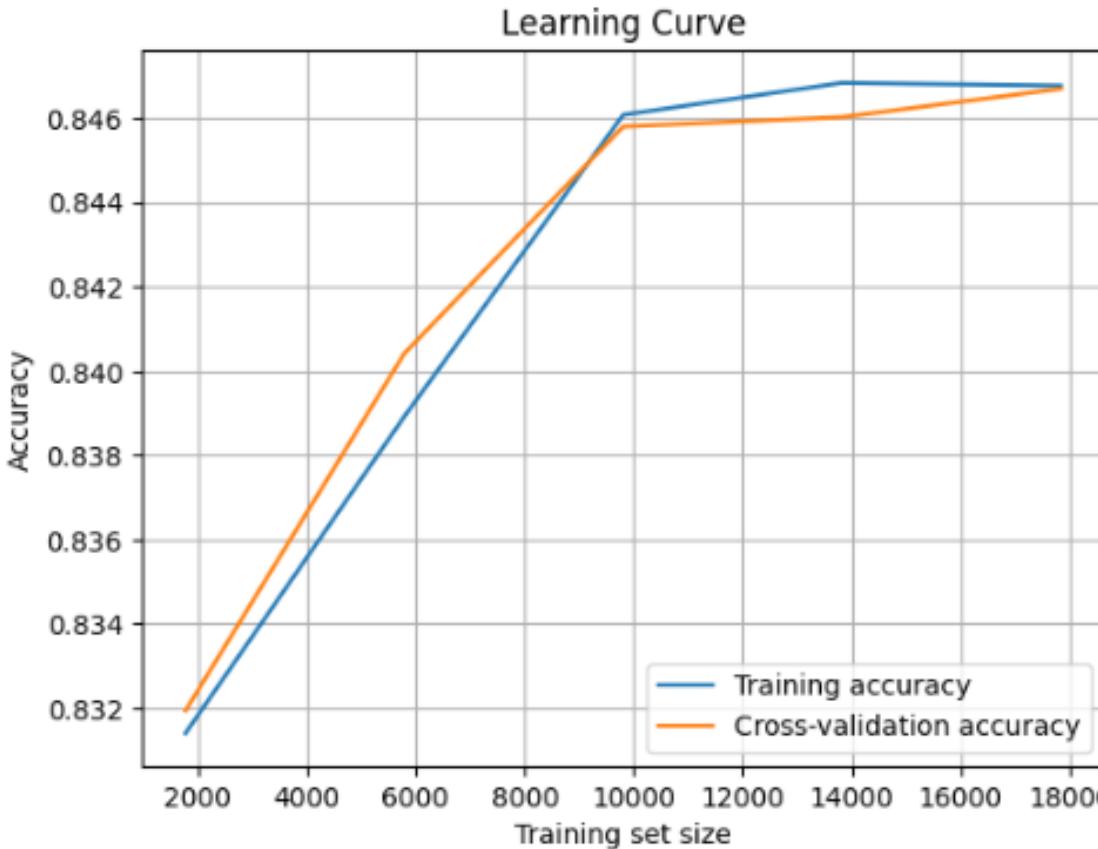


G. Hasil & Pembahasan



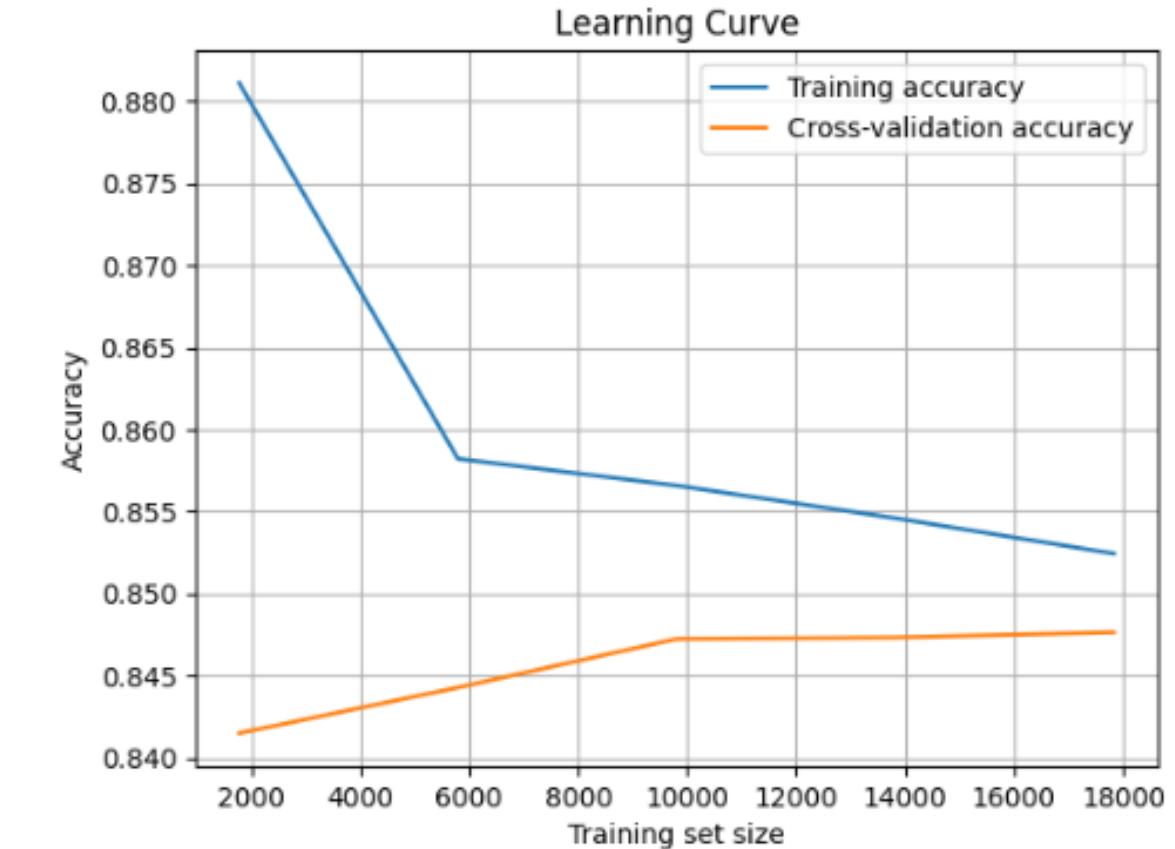
Logistic Regression

- Overfitting: hanya di jumlah data kecil
- Underfitting: tidak
- Generalization: cukup baik



SVM Sigmoid

- Overfitting: Tidak
- Underfitting: Tidak
- Generalization Sangat baik



XGBoost

- Overfitting: Ya
- Underfitting: Tidak
- Generalization Buruk



H. Kesimpulan & Saran

Kesimpulan

Alasan model SVM Sigmoid dipilih:

- **Stabil dan Konsisten:** Akurasi train & validasi hampir sama (84,67%)
- **Kurva pembelajaran konvergen:** tidak *overfitting* dan *underfitting*
- **F1-Score seimbang:** Negatif: 0.81 | Positif: 0.87 | Rata-rata: 0.85
- **Recall tertinggi:** mencegah kasus positif luput dari hasil deteksi
- **Generalizable**

Saran

Pengembangan yang dapat dilakukan

- **Feature Engineering**
- **Hyperparameter tuning**
- **Upgrade arsitektur model**



Thank you



IPB University
— Bogor Indonesia —

**Department of Computer Science
FMIPA-IPB Kampus Darmaga
Jl. Meranti Wing 20 Level V, Bogor, Indonesia
Phone/Fax: +62 251 8625584
<http://cs.ipb.ac.id/>**