

# Analyzing Mental Health Trends and Predictors Among Students

Bootcamp Data Analyst with SQL & Python using Google Platform

Muhammad Fairuz Umar



# Business Understanding

## Latar Belakang:

Tingkat depresi di kalangan pelajar semakin mengkhawatirkan, apalagi setelah pandemi. Universitas dan sekolah ingin mengidentifikasi siswa yang berisiko mengalami depresi lebih dini untuk memberikan intervensi tepat waktu.

## Tujuan Bisnis:

- Memprediksi status depresi siswa (Depresi / Tidak Depresi).
- Memberikan data pendukung untuk program kesehatan mental.
- Meningkatkan efektivitas layanan konseling dan kesehatan kampus.
- Kriteria Sukses: Model prediksi harus akurat minimal 80% dan mampu membedakan siswa yang benar-benar membutuhkan bantuan.

## Data Understanding

### Dataset: Student Depression Dataset

- Dataset berisi data siswa, termasuk atribut seperti gender, usia, akademik, tingkat stress, dll, dengan target label:
- 0 = Tidak Depresi
- 1 = Depresi

### Observasi Awal:

- Data tidak seimbang sempurna tapi masih cukup baik (sedikit lebih banyak yang mengalami depresi).
- Ada atribut-atribut yang relevan dengan psikologi siswa.
- Visualisasi Awal:
- Bar plot distribusi label memperlihatkan banyak siswa mengalami depresi.

## Data Preparation

### Pembersihan Data:

- Cek missing values dan menangani data yang kosong.
- Encoding untuk fitur kategorikal jika ada (label encoding / one-hot encoding).

### Splitting Data:

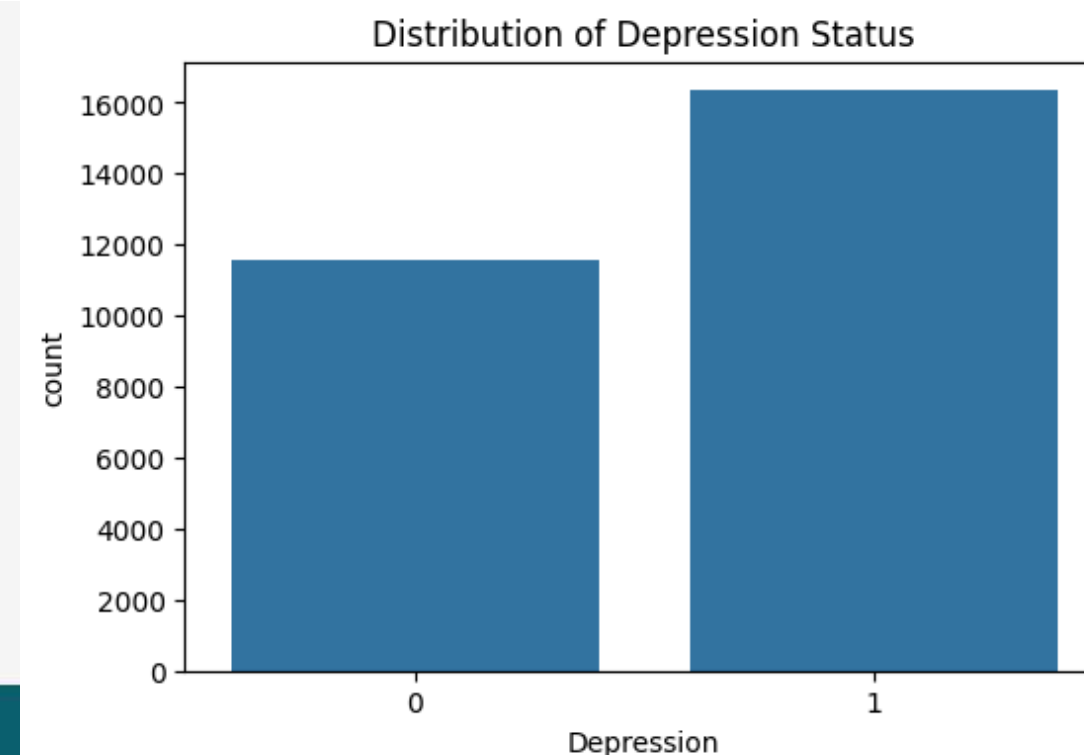
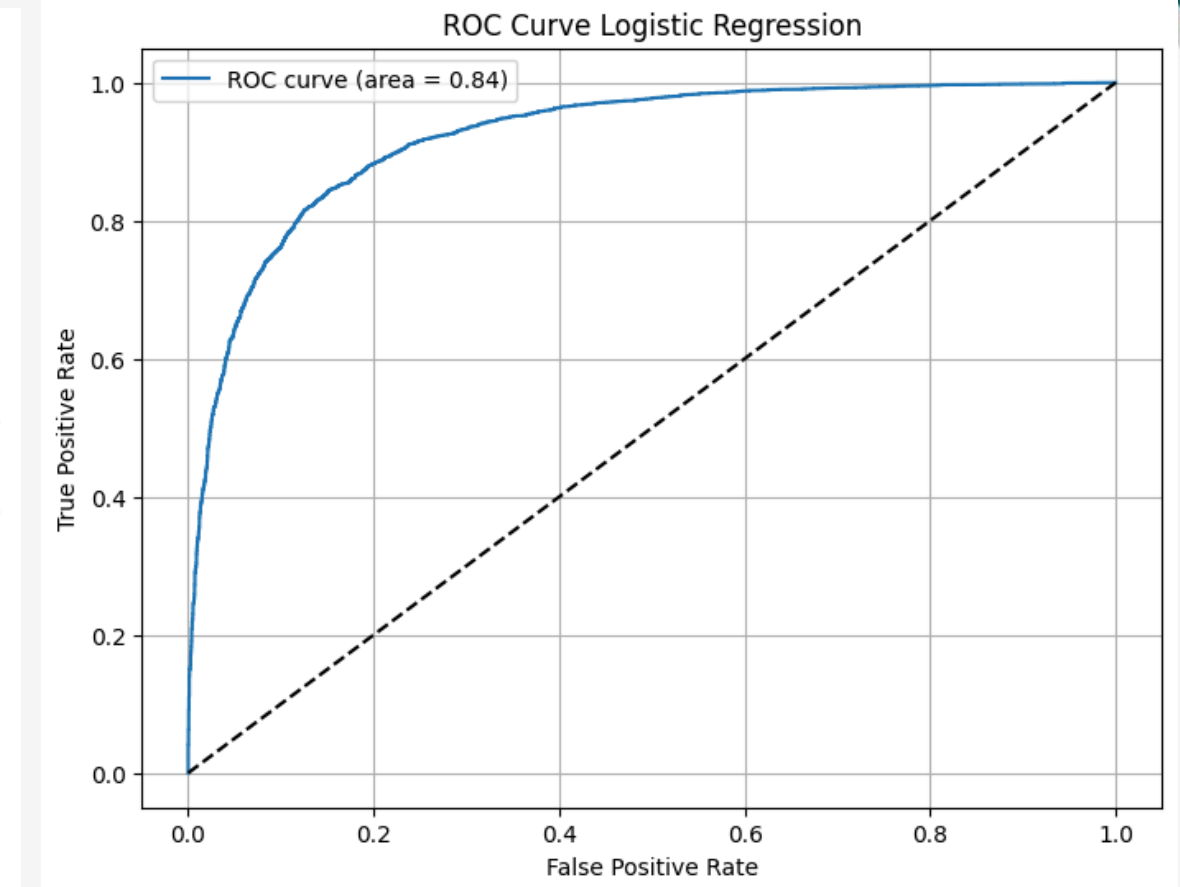
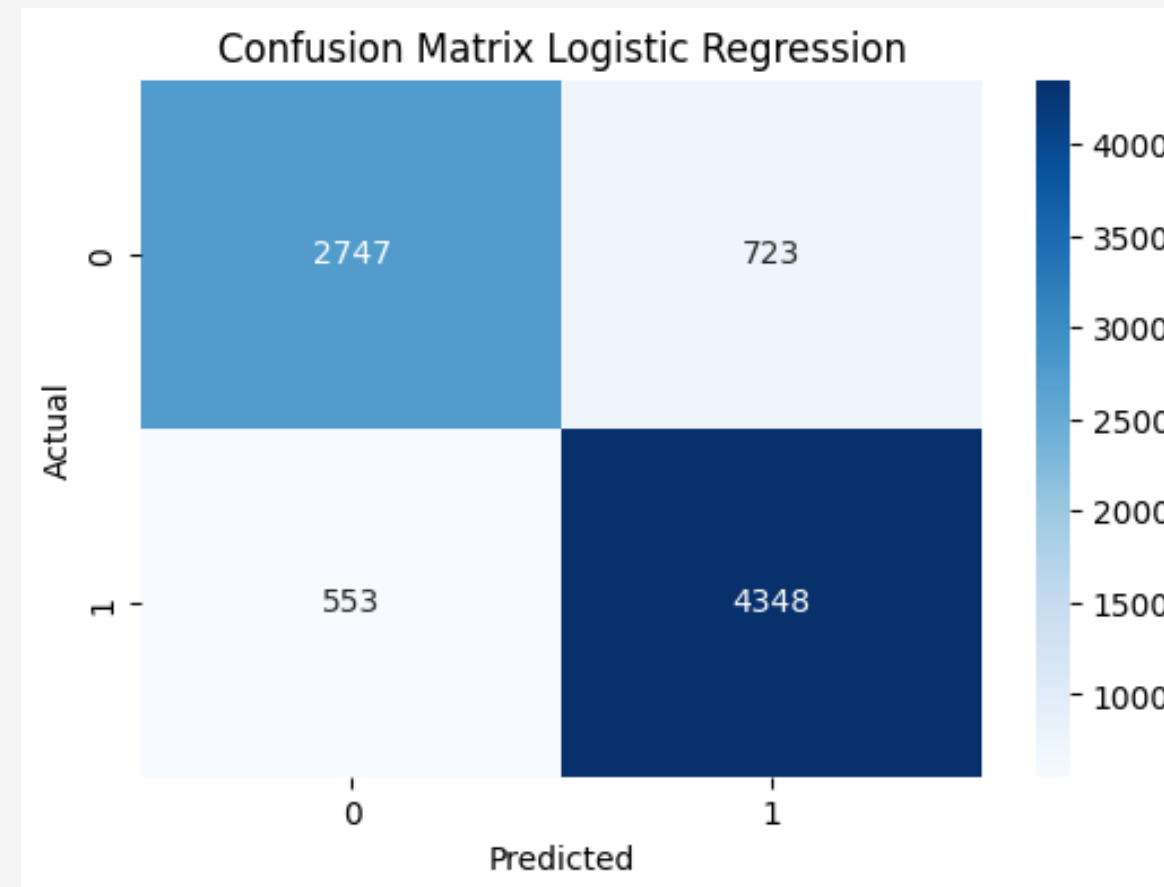
- Split data menjadi train dan test .

### Standarisasi:

- Karena menggunakan Logistic Regression, fitur numerik perlu distandarisasi agar model lebih optimal.

# Modeling

- Model yang Digunakan:
  - Logistic Regression (konsep dasar machine learning untuk klasifikasi biner).
  - Logistic Regression memodelkan probabilitas suatu instance masuk ke kelas tertentu berdasarkan fungsi logit (sigmoid).
- Hasil Model:
  - Confusion Matrix:
  - Model cukup seimbang:
    - True Positives (4348), True Negatives (2747)
    - False Positives (723), False Negatives (553)
  - ROC-AUC Score:
    - AUC sebesar 0.84, artinya model cukup bagus dalam membedakan siswa yang depresi dan tidak.



Student Logistic Regression

## Evaluation

### Confusion Matrix Analysis:

- Model lebih banyak berhasil mengklasifikasikan siswa yang depresi.
- Beberapa kesalahan terjadi, namun false negative (missed case of depression) masih relatif kecil.

### ROC Curve Analysis:

- Kurva ROC model jauh di atas garis random, dengan area luas (AUC 0.84), membuktikan kemampuan klasifikasinya kuat.

### Metric Summary:

- Akurasi, Precision, Recall semua cukup tinggi.
- Recall lebih penting di kasus ini: lebih baik sedikit over-predict (false positive) daripada under-predict (false negative untuk kasus kesehatan mental).



## Deployment / Action Plan

- Penggunaan Nyata:
  - Model bisa diintegrasikan dalam sistem monitoring kampus.
  - Screening awal berbasis data sebelum konseling manual dilakukan.
- Rekomendasi Tambahan:
  - Monitoring Berkala: update model dengan data baru secara rutin.
  - Feature Engineering: kembangkan variabel baru (misal: durasi tidur, kebiasaan olahraga).
  - Pengujian Model Lain: coba Random Forest, Gradient Boosting untuk bandingan.

-

let's connect!



Muhammad Fairuz Umar



Muhammad Fairuz Umar



GitHub

Muhammad Fairuz Umar



dqlab



www.dqlab.id

Thanks!