

Nama : Erdin Purwa Putra
NIM : A11.2022.14410
Mata Kuliah : Data Mining A11.4507

UAS Data Mining

1. Judul / Topik Project dan Identitas Lengkap

Judul: "Analisis Prediktif Faktor-Faktor yang Mempengaruhi Tingkat Depresi Mahasiswa Menggunakan Model Machine Learning" **Identitas:**

- Nama : Erdin Purwa Putra
- Nim : A11.2022.14410

2. Ringkasan dan Permasalahan Project

Ringkasan:

Kesehatan mental mahasiswa merupakan isu penting yang memengaruhi performa akademik dan kehidupan sosial mereka. Proyek ini bertujuan untuk memahami faktor-faktor signifikan yang memengaruhi tingkat depresi mahasiswa melalui analisis prediktif menggunakan algoritma *machine learning*.

Permasalahan:

1. Mengidentifikasi tingkat depresi mahasiswa.
2. Mengatasi data yang tidak seimbang.
3. Menentukan faktor utama yang memengaruhi tingkat depresi.
4. Membangun model prediktif yang optimal.

Tujuan:

1. Mengidentifikasi faktor risiko utama.
2. Mengembangkan model prediktif dengan akurasi tinggi.
3. Mengatasi ketidakseimbangan data dengan SMOTE.
4. Menyempurnakan model dengan optimasi parameter.

Model / Alur Penyelesaian:

Berikut adalah bagan alur penyelesaian:

1. Pengumpulan dan Pemuatan Data
2. Praproses Data
3. Analisis Data Eksploratori (EDA)

4. Pemilihan dan Rekayasa Fitur
5. Pembuatan Model Prediktif
6. Evaluasi Model
7. Interpretasi dan Kesimpulan

3. Penjelasan Dataset, EDA, dan Proses Features Dataset

Dataset:

Dataset berasal dari Kaggle, berjudul *"Student Mental Health Survey: Online Survey on the Mental Health of IT Students"*.

EDA:

1. Memvisualisasikan distribusi skor depresi, kecemasan, dan isolasi.
2. Membuat matriks korelasi antar variabel untuk memahami hubungan faktor-faktor terkait.
3. Memeriksa distribusi variabel target (depression) untuk mengidentifikasi ketidakseimbangan data.

Proses Features Dataset:

1. Menghapus spasi pada nama kolom.
2. Menggunakan SMOTE untuk menangani ketidakseimbangan kelas.
3. Memilih fitur relevan menggunakan teknik RFE.

4. Proses Learning / Modeling

1. Algoritma yang digunakan: Random Forest Classifier.
2. Optimasi model menggunakan Grid Search untuk menemukan parameter terbaik.
3. Validasi model menggunakan validasi silang (cross-validation).

5. Performa Model

1. Metode evaluasi: Akurasi, laporan klasifikasi (precision, recall, F1-score).
2. Hasil evaluasi:

```
Akurasi: 0.42857142857142855
```

Laporan Klasifikasi:				
	precision	recall	f1-score	support
1	0.50	1.00	0.67	3
2	0.57	0.80	0.67	5
3	0.00	0.00	0.00	5
4	0.33	0.20	0.25	5
5	0.50	0.33	0.40	3
accuracy			0.43	21
macro avg	0.38	0.47	0.40	21
weighted avg	0.36	0.43	0.37	21

6. Diskusi Hasil dan Kesimpulan

Diskusi Hasil:

1. Fitur paling signifikan dalam prediksi adalah **Future Insecurity**, **Financial Concerns**, dan **Social Relationships**.
2. Model menunjukkan performa yang baik dalam prediksi tingkat depresi mahasiswa dengan akurasi tinggi.
3. Teknik SMOTE efektif dalam menangani ketidakseimbangan data, meningkatkan keandalan model.

Kesimpulan:

Proyek ini berhasil mengidentifikasi faktor-faktor signifikan yang memengaruhi tingkat depresi mahasiswa. Model prediktif yang dibangun dapat digunakan oleh lembaga pendidikan untuk mengidentifikasi mahasiswa yang membutuhkan intervensi lebih awal, sehingga dapat membantu meningkatkan kesehatan mental secara keseluruhan.