

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

### UTS Data Mining

1. Topik / Judul

"Analisis Prediktif Faktor-Faktor yang Mempengaruhi Tingkat Depresi Mahasiswa Menggunakan Model Machine Learning"

2. Deskripsi Singkat (Ringkasan BU/DU) Pemahaman Bisnis (BU): Kesehatan mental, khususnya depresi, kecemasan, dan isolasi, semakin menjadi perhatian utama dalam lingkungan akademis. Mahasiswa menghadapi segala macam tekanan akademis, sosial, dan finansial yang dapat memengaruhi kesehatan mental mereka. Analisis prediktif bertujuan untuk membantu lembaga pendidikan mengidentifikasi faktor-faktor yang secara signifikan memengaruhi tingkat depresi di kalangan mahasiswa.

Pemahaman Data (DU): Kumpulan data ini mencakup data survei untuk berbagai faktor yang terkait dengan kesehatan mental mahasiswa, termasuk faktor akademis, sosial, dan finansial, serta tingkat depresi, kecemasan, dan isolasi. Dengan memahami data ini, kita dapat mengetahui variabel mana yang paling berpengaruh dan membangun model prediktif untuk mengungkap faktor-faktor tersebut.

3. Masalah dan Tujuan yang ingin dicapai

Masalah :

- Mengidentifikasi tingkat depresi di kalangan mahasiswa dan faktor-faktor yang berkontribusi terhadap kesehatan mental yang buruk.
- **Data Tidak Seimbang:** Kemungkinan terdapat ketidakseimbangan dalam data (lebih sedikit mahasiswa dengan tingkat depresi tinggi dibandingkan yang rendah), hal ini dapat menyebabkan model bias atau menyimpang.
- **Pemilihan Fitur:** Tidak semua faktor berkontribusi secara signifikan terhadap prediksi depresi. Maka penting untuk memilih fitur yang paling relevan untuk model prediktif.
- **Model Prediksi yang Optimal:** Menemukan model machine learning yang dapat memprediksi tingkat depresi dengan akurasi tinggi dan dapat diandalkan untuk membantu dalam pengambilan keputusan atau intervensi.

Tujuan yang ingin dicapai:

- **Identifikasi Faktor Risiko:** Memahami faktor-faktor yang berkontribusi signifikan terhadap depresi di kalangan mahasiswa
- **Prediksi Tingkat Depresi:** Membangun model machine learning yang dapat secara akurat memprediksi tingkat depresi mahasiswa berdasarkan faktor-faktor yang ada.

- **Mengatasi Ketidakseimbangan Data:** Menggunakan teknik seperti SMOTE untuk menyeimbangkan data sehingga model dapat belajar dari semua kategori secara merata.
- **Penyempurnaan Model:** Menggunakan Grid Search untuk menemukan parameter terbaik untuk Random Forest, memastikan bahwa model seoptimal mungkin.
- **Evaluasi dan Interpretasi Model:** Mengukur kinerja model menggunakan metrik seperti akurasi dan laporan klasifikasi, serta menginterpretasikan fitur mana yang paling penting dalam prediksi depresi.

#### 4. Penjelasan Datasets (Sumber data dan penjelasan atribut)

Sumber Data : Data ini diperoleh dari kaggle “**Student Mental Health Survey: Online Survey on the Mental Health of IT Students**”

Penjelasan Atribut:

- **gender:** Jenis kelamin responden.
- **age:** Usia responden.
- **study\_satisfaction:** Skor kepuasan mahasiswa terhadap proses belajar.
- **academic\_workload:** Beban akademik yang dirasakan oleh mahasiswa.
- **academic\_pressure:** Tekanan akademik yang dialami mahasiswa.
- **financial\_concerns:** Tingkat kekhawatiran terhadap masalah finansial.
- **social\_relationships:** Kualitas hubungan sosial yang dimiliki mahasiswa.
- **depression:** Skor tingkat depresi mahasiswa (variabel target).
- **anxiety:** Skor tingkat kecemasan mahasiswa.
- **isolation:** Skor tingkat isolasi sosial yang dialami mahasiswa.
- **future\_insecurity:** Kekhawatiran mahasiswa terhadap masa depan.
- **stress\_relief\_activities:** Aktivitas yang dilakukan untuk meredakan stres.

Atribut yang memiliki kepentingan untuk memprediksi tingkat depresi yaitu ;

- **Future Insecurity** (Ketidakpastian Masa Depan) adalah fitur yang paling penting dalam memprediksi tingkat depresi mahasiswa. Ini menunjukkan bahwa ketidakpastian tentang masa depan sangat memengaruhi kesehatan mental mereka.
- **Financial Concerns** (Kekhawatiran Finansial) dan **Social Relationships** (Hubungan Sosial) juga menempati posisi penting. Kekhawatiran finansial dapat menimbulkan tekanan tambahan, sedangkan kualitas hubungan sosial berperan besar dalam mendukung atau membebani kesehatan mental.
- **Academic Pressure** (Tekanan Akademik) juga merupakan faktor signifikan. Tekanan dari tugas-tugas dan tanggung jawab akademik dapat berkontribusi pada peningkatan tingkat stres dan depresi.
- **Study Satisfaction** (Kepuasan Studi), meskipun memiliki tingkat kepentingan yang lebih rendah dibandingkan faktor lainnya, tetap relevan. Hal ini menunjukkan bahwa kepuasan terhadap pengalaman akademik mereka memiliki pengaruh, meski tidak sebesar faktor-faktor lainnya.

#### 5. Alur / Tahapan / Kerangka Eksperimen

- Langkah 1: Pengumpulan dan Pemuatan Data  
Memuat dataset yang sebelumnya telah tersimpan di google drive

- Langkah 2: Praproses dan Pembersihan Data  
Bersihkan data dengan menghapus spasi pada nama kolom dan menangani ketidakseimbangan kelas menggunakan teknik SMOTE.
- Langkah 3: Analisis Data Eksploratori (EDA)  
Lakukan analisis awal untuk memahami distribusi data dan korelasi antar variabel. Ini termasuk memvisualisasikan distribusi skor depresi, kecemasan, dan isolasi, serta menghitung matriks korelasi untuk memahami hubungan antar factor
- Langkah 4: Pemilihan dan Rekayasa Fitur  
Gunakan teknik Recursive Feature Elimination (RFE) untuk memilih fitur yang paling relevan yang memengaruhi tingkat depresi.
- Langkah 5: Pembuatan Model dan Penyetelan Hiperparameter  
Buat model prediktif menggunakan Random Forest Classifier. Model-model ini kemudian disempurnakan dengan penyetelan hiperparameter dengan Pencarian Grid untuk mengetahui campuran parameter apa yang membentuk kombinasi terbaik.
- Langkah 6: Evaluasi Model  
Evaluasi performa model dengan metrik evaluasi dan validasi silang (akurasi dan laporan klasifikasi) untuk memastikan keandalan model.
- Langkah 7: Interpretasi dan Analisis Pentingnya Fitur  
Analisis fitur-fitur yang paling berpengaruh dari prakiraan depresi untuk menawarkan rekomendasi yang didasarkan pada data.

#### 6. Timeline Eksperimen (hingga akhir semester)

Minggu	Kegiatan	Detail Kegiatan	Output
Minggu 8	Eksplorasi Data Tambahan dan Analisis Kausalitas	- Lakukan analisis statistik mendalam terhadap data - Gunakan SEM untuk analisis kausalitas antara variabel-variabel kunci	Wawasan kausalitas antara faktor-faktor kesehatan mental mahasiswa
Minggu 9	Eksperimen Augmentasi Data dengan Synthetic Data Generation	- Gunakan VAEs atau SMOTE-ENN untuk menghasilkan data sintetik tambahan yang bervariasi	Dataset dengan variasi data yang lebih luas untuk pelatihan model
Minggu 10	Clustering Mahasiswa Berdasarkan Profil Kesehatan Mental	- Terapkan K-means atau Hierarchical Clustering untuk mengelompokkan mahasiswa	Klaster mahasiswa dengan risiko kesehatan mental yang berbeda

		berdasarkan profil risiko	
Minggu 11	Eksperimen Feature Engineering dengan Fitur Kombinasi	<ul style="list-style-type: none"> <li>- Buat fitur baru dari kombinasi atau interaksi antar variabel</li> <li>- Evaluasi apakah fitur ini meningkatkan performa model</li> </ul>	Fitur baru yang dapat meningkatkan akurasi model
Minggu 12	Implementasi Model Neural Network dan Ensemble Learning	<ul style="list-style-type: none"> <li>- Implementasikan Multilayer Perceptron (MLP)</li> <li>- Gunakan Voting Classifier atau Stacking untuk ensemble learning</li> </ul>	Model yang lebih kuat dengan peningkatan akurasi melalui ensemble
Minggu 13	Validasi Lanjutan dengan Cross-Validation dan ROC-AUC	- Lakukan cross-validation dengan metrik tambahan (ROC-AUC dan Precision-Recall Curve)	Validasi menyeluruh dengan metrik tambahan untuk keandalan model
Minggu 14	Dokumentasi dan Penyelesaian Laporan Akhir	- Susun laporan akhir dengan hasil eksperimen yang paling signifikan	Laporan akhir proyek UTS