**1. Pemahaman Data**

Dataset ini berisi informasi tentang pelamar program MBA untuk kelas tahun 2025. Data ini mencakup berbagai atribut yang berkaitan dengan latar belakang akademis, pengalaman profesional, dan karakteristik demografis pelamar.

**1.1 Informasi Dataset**

* **Sumber**: Kaggle
* **Jumlah Baris**: 6.194
* **Jumlah Kolom**: 10
* **Variabel Target**: admission (Status penerimaan: Admit, Waitlist)

**1.2 Deskripsi Variabel**

* application\_id: Identifikasi unik untuk setiap pelamar.
* gender: Jenis kelamin pelamar (Male, Female).
* international: Status internasional pelamar (True jika pelamar internasional, False jika domestik).
* gpa: Indeks Prestasi Kumulatif (IPK) dari pendidikan sebelumnya.
* major: Jurusan sarjana pelamar (misalnya, STEM, Business, Humanities).
* race: Etnisitas pelamar.
* gmat: Skor GMAT pelamar.
* work\_exp: Jumlah tahun pengalaman kerja.
* work\_industry: Industri tempat pelamar bekerja sebelumnya.
* admission: Status penerimaan (Admit, Deny, Waitlist).

**1.3 Tujuan Dataset**

Dataset ini dirancang untuk:

* Menganalisis profil pelamar MBA berdasarkan latar belakang akademis, pengalaman profesional, dan karakteristik demografis.
* Mengidentifikasi tren dan pola yang dapat membantu dalam proses seleksi penerimaan.
* Mengembangkan model prediktif untuk memprediksi hasil penerimaan berdasarkan atribut pelamar.

**2. Penentuan Tujuan**

**2.1 Tujuan Bisnis**

Memahami faktor-faktor utama yang mempengaruhi penerimaan pelamar ke program MBA. Hasil analisis dapat digunakan oleh institusi pendidikan untuk meningkatkan proses seleksi dan strategi penerimaan.

**2.2 Tujuan Analisis**

* **Prediksi Penerimaan MBA**: Mengembangkan model prediktif untuk menentukan apakah seorang pelamar akan diterima, ditolak, atau dimasukkan ke daftar tunggu berdasarkan atribut yang tersedia.
* **Analisis Faktor Pengaruh**: Mengidentifikasi variabel utama yang mempengaruhi keputusan penerimaan, seperti IPK, skor GMAT, pengalaman kerja, dan lainnya.
* **Segmentasi Pelamar**: Melakukan segmentasi pelamar berdasarkan atribut tertentu untuk memahami kelompok pelamar yang memiliki peluang penerimaan lebih tinggi.
* **Rekomendasi Berbasis Data**: Memberikan wawasan berbasis data bagi calon pelamar untuk membantu mereka meningkatkan peluang diterima.

**2.3 Langkah-Langkah Proyek**

* Melakukan analisis data eksploratif (EDA) untuk memahami distribusi data dan hubungan antar variabel.
* Membersihkan data untuk menangani nilai yang hilang, outlier, dan inkonsistensi.
* Membuat visualisasi data untuk menyajikan wawasan utama secara intuitif.
* Melatih model prediktif (misalnya, Logistic Regression, Decision Tree, Random Forest) untuk memprediksi penerimaan MBA.
* Mengevaluasi performa model menggunakan metrik seperti akurasi, F1-scor.
* Menyusun laporan akhir yang mencakup wawasan analisis, model prediktif, dan rekomendasi.

Dengan pendekatan ini, diharapkan dapat diperoleh pemahaman yang komprehensif mengenai faktor-faktor yang mempengaruhi penerimaan pelamar ke program MBA dan bagaimana meningkatkan proses seleksi penerimaan.