**Tugas 6**

**MK WI2002 Literasi Data dan Inteligensi Artifisial**

**Preparasi Data**

**Waktu Pengumpulan: 9 April 2025**

**Petunjuk Pengerjaan**

1. Tugas dikumpulkan maksimal pada tanggal 9 April 2025, Pukul 12.00 WIB
2. Tugas dikumpulkan dalam format PDF
3. Tugas dikumpulkan dengan cara mengunggahnya ke Edunex
4. Tugas dikerjakan perkelompok (maksimal 3 orang perkelompok)

**Studi Kasus “Preparasi Data pada Sejarah Produksi Sumur Minyak”**

Dalam industri perminyakan, data sejarah produksi minyak dan gas digunakan untuk analisis performa sumur, estimasi cadangan, dan optimasi produksi. Namun, data ini sering kali memiliki **missing values**, **outlier**, **inkonsistensi**, dan **duplikasi** sehingga memerlukan proses pembersihan (*data cleansing*) sebelum digunakan lebih lanjut.

Anda diberikan dataset yang berisi **data produksi harian** dari beberapa sumur minyak dan gas di lapangan X. Tugas Anda adalah membersihkan data tersebut agar siap digunakan untuk analisis lebih lanjut.

## **Instruksi Pengerjaan**

### **1. Menangani Missing Data**

* Identifikasi jumlah dan distribusi data yang hilang.
* Gunakan metode berikut untuk menangani missing values:
  + **Interpolasi (linear, spline, dll.)** untuk data waktu seperti tekanan dan temperatur.
  + **Mean/Median Imputation** untuk nilai produksi minyak/gas jika data yang hilang kecil (<5%).
  + **Filling Forward/Backward** untuk mengisi data produksi berdasarkan hari sebelumnya jika masuk akal.
  + **Pembuangan Baris (Row Deletion)** jika data yang hilang lebih dari 30% dari total kolom dalam satu baris.

### **2. Menghapus atau Menangani Outlier**

* Gunakan **boxplot** atau **Z-score** untuk mendeteksi outlier pada produksi minyak, gas, dan tekanan sumur.
* Jika outlier terjadi karena kesalahan input, ganti dengan nilai interpolasi atau rata-rata sumur dalam zona yang sama.
* Jika outlier merupakan kejadian nyata (misalnya lonjakan produksi karena stimulasi sumur), tandai sebagai event khusus tanpa menghapusnya.

### **3. Mengatasi Data Duplikasi**

* Periksa apakah ada data duplikat pada kolom **(Date, Well\_ID)**. Jika ya, simpan hanya satu data yang valid.

### **4. Standardisasi Format Data**

* Pastikan format tanggal seragam (YYYY-MM-DD atau DD/MM/YYYY).
* Konversi unit yang tidak seragam (misalnya tekanan dalam bar → psi).

### **5. Validasi dan Simpan Data Bersih**

* Cek apakah ada anomali setelah pembersihan.