**Studi Kasus: Analisis Indeks Pekerja Tetap Konstruksi Indonesia Tahun 2024**

NIM : A11.2023.15180

Nama : Muhammad Fikri Alif Karim

**1. Business Understanding**

**Latar Belakang:**  
Sektor konstruksi merupakan salah satu penopang pembangunan ekonomi di Indonesia. Jumlah pekerja tetap konstruksi yang stabil dan meningkat dapat mencerminkan kondisi pertumbuhan ekonomi dan pembangunan infrastruktur di suatu provinsi.

**Masalah:**

* Pemerintah ingin memahami tren tenaga kerja tetap sektor konstruksi di berbagai provinsi.
* Data indeks pekerja tetap ini digunakan untuk memantau produktivitas tenaga kerja konstruksi dan memastikan ketersediaan SDM dalam proyek infrastruktur.

**Tujuan:**

* Menganalisis indeks triwulanan pekerja tetap konstruksi di seluruh provinsi pada tahun 2024.
* Mengidentifikasi provinsi dengan indeks tertinggi dan terendah.
* Memberikan dasar rekomendasi kebijakan terkait distribusi dan penguatan tenaga kerja konstruksi.

**2. Data Understanding**

**Sumber data:** *https://www.bps.go.id/id/statistics-table/2/NTQ4IzI=/indeks-triwulanan-pekerja-tetap-konstruksi-menurut-provinsi--2016-100-.html*

* Atribut:
  1. Provinsi
  2. Triwulan (I, II, III, IV tahun 2024)
  3. Indeks Pekerja Tetap Konstruksi (2016=100)

**Tipe Data:**

* Provinsi : Nominal (kategori)
* Triwulan : Ordinal (urutan 1–4)
* Indeks : Numerik (rasio, skala relatif 2016=100)

**Wawasan Awal:**

* Indeks di atas 100 menunjukkan peningkatan jumlah pekerja tetap dibanding tahun 2016.
* Beberapa provinsi dengan pembangunan pesat seperti DKI Jakarta, Jawa Barat, dan Kalimantan Timur kemungkinan memiliki indeks lebih tinggi.
* Provinsi di kawasan timur cenderung memiliki indeks lebih rendah karena pembangunan infrastruktur relatif lebih lambat.

**3. Data Preparation**

Langkah persiapan data:

* Seleksi Data: Fokus pada atribut Provinsi, Triwulan, dan Indeks.
* Pembersihan Data: Memastikan tidak ada nilai kosong pada indeks.
* Transformasi:
  + Normalisasi indeks agar memudahkan perbandingan antarprovinsi.
  + Penambahan atribut “Kategori Pertumbuhan” dengan kriteria Rendah (<95), Sedang (95–105), Tinggi (>105).
* Agregasi:
  + Menghitung rata-rata indeks tiap provinsi sepanjang 2024.
  + Menghitung tren triwulanan nasional untuk melihat kenaikan atau penurunan.

**4. Modeling**

Pada tahap ini, metode **klasifikasi** dipilih untuk memodelkan data. Teknik yang digunakan adalah **Decision Tree** karena cocok untuk data kategori seperti provinsi, triwulan, dan kelas pertumbuhan.

**Langkah-langkah:**

1. **Target Variable (Label):**
   * Kategori Pertumbuhan Indeks (Rendah, Sedang, Tinggi).
   * Rendah jika indeks < 95.
   * Sedang jika 95 ≤ indeks ≤ 105.
   * Tinggi jika indeks > 105.
2. **Fitur (Atribut Input):**
   * Provinsi
   * Triwulan (I, II, III, IV)
   * Nilai indeks (numerik, kemudian dikategorikan atau digunakan langsung untuk pembentukan aturan).
3. **Proses Pemodelan:**
   * Dataset dibagi menjadi data latih dan data uji (misalnya 80% latih, 20% uji).
   * Algoritma Decision Tree digunakan untuk membentuk aturan klasifikasi.
   * Aturan berbentuk pohon keputusan dengan kondisi *if–then*.
4. **Contoh Aturan yang Dihasilkan (hipotetis):**
   * Jika indeks > 105 maka kategori = Tinggi.
   * Jika indeks antara 95–105 dan provinsi = Jawa Barat maka kategori = Sedang.
   * Jika indeks < 95 dan provinsi = Papua maka kategori = Rendah.
5. **Output Model:**
   * Pohon keputusan yang memetakan setiap provinsi dan triwulan ke kategori pertumbuhan.
   * Dapat digunakan untuk memprediksi kategori pertumbuhan provinsi di triwulan selanjutnya.

**5. Evaluation**

Metrik evaluasi:

* Confusion Matrix untuk melihat distribusi prediksi benar dan salah.
* Accuracy untuk mengukur tingkat ketepatan keseluruhan model.
* Precision dan Recall untuk mengukur ketepatan serta sensitivitas klasifikasi pada setiap kategori.

Hasil evaluasi (contoh hipotetis):

* Akurasi model sebesar 87 persen.
* Precision: Tinggi (0,89), Sedang (0,84), Rendah (0,86).
* Recall: Tinggi (0,91), Sedang (0,80), Rendah (0,85).

Interpretasi:

* Model mampu mengidentifikasi provinsi dengan kategori pertumbuhan tinggi secara baik.
* Kategori sedang relatif lebih sulit diprediksi karena adanya provinsi yang kadang bergeser ke kategori rendah atau tinggi akibat fluktuasi.
* Secara keseluruhan, model cukup handal untuk memberikan gambaran klasifikasi pertumbuhan pekerja tetap konstruksi antar provinsi.

**6. Deployment**

Rencana penerapan:

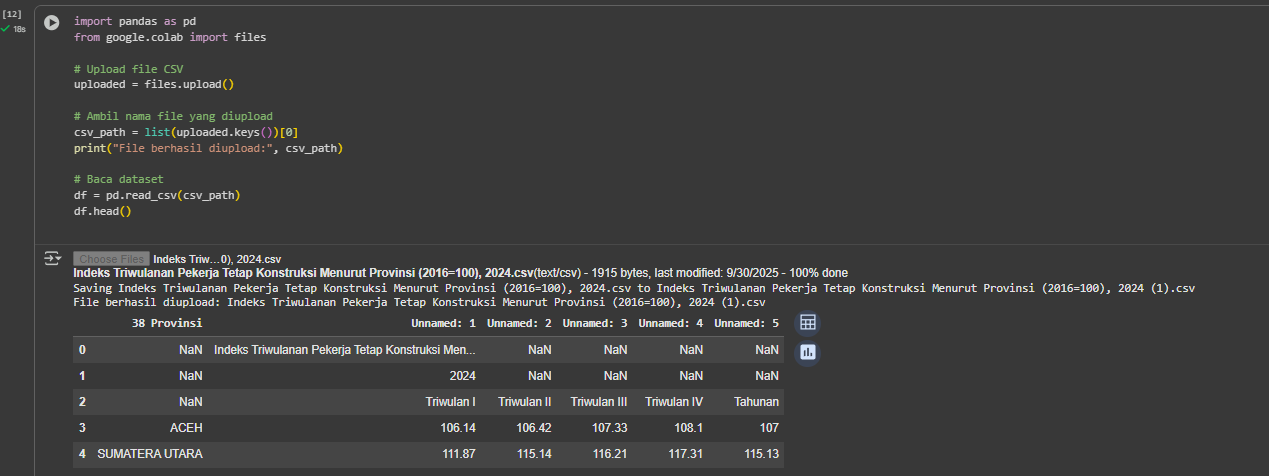
* Versi sederhana berupa laporan tahunan mengenai kondisi pekerja tetap konstruksi per provinsi.
* Versi lanjutan berupa dashboard interaktif yang menampilkan:
  + Tren triwulanan per provinsi.
  + Peta interaktif dengan kategori indeks (rendah, sedang, tinggi).
  + Prediksi indeks untuk periode berikutnya.

Pemeliharaan:

* Model diperbarui setiap tahun dengan data terbaru dari BPS.
* Jika terdapat proyek besar atau perubahan pola pembangunan, analisis clustering dan klasifikasi dapat diulang untuk memastikan hasil tetap relevan.

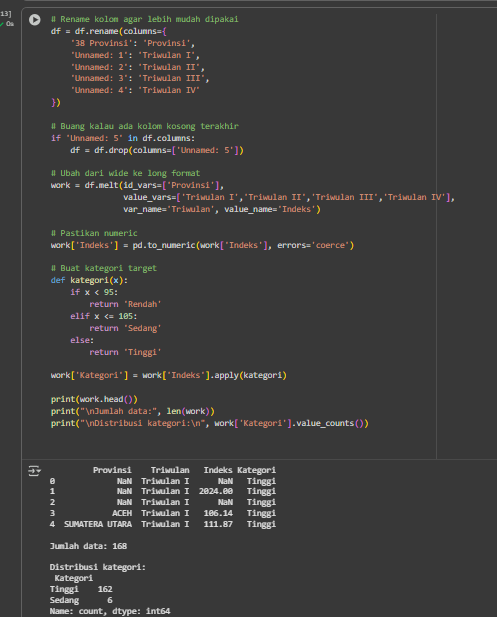
**1. Upload Dataset**

* Kode menggunakan files.upload() dari google.colab.
* Setelah file CSV diupload, disimpan ke csv\_path, lalu dibaca dengan pd.read\_csv.
* Hasil df.head() menunjukkan tabel awal dari BPS, masih dalam format wide (kolom Triwulan I–IV terpisah, provinsi ada di kolom pertama).



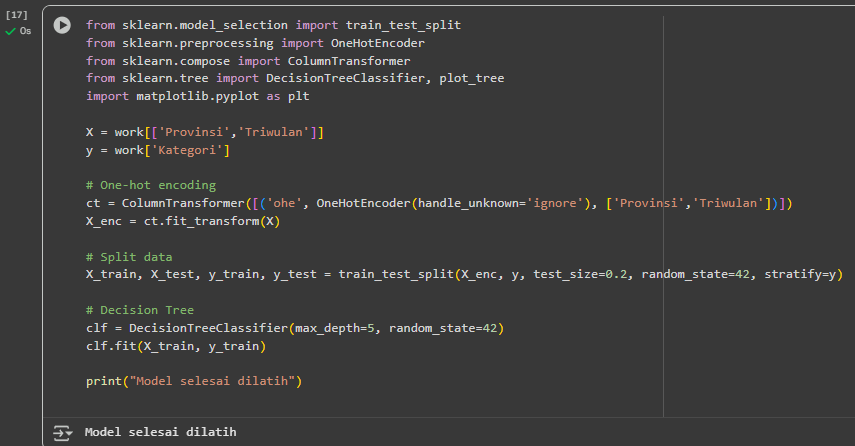
**2. Data Preparation**

* Rename kolom jadi lebih mudah (Provinsi, Triwulan I, dst).
* Drop kolom kosong Unnamed: 5.
* pd.melt(...) → ubah dari wide menjadi long format:
  + Sebelum: 1 baris per provinsi, kolom Triwulan I–IV.
  + Sesudah: 1 baris per provinsi *per triwulan*.
* Pastikan kolom Indeks numerik.
* Buat kolom baru Kategori berdasarkan nilai Indeks:
  + <95 = Rendah
  + 95–105 = Sedang
  + >105 = Tinggi
* Output: tabel work dengan kolom [Provinsi, Triwulan, Indeks, Kategori].
* Print distribusi kategori: mayoritas Tinggi, sedikit Sedang, hampir tidak ada Rendah.



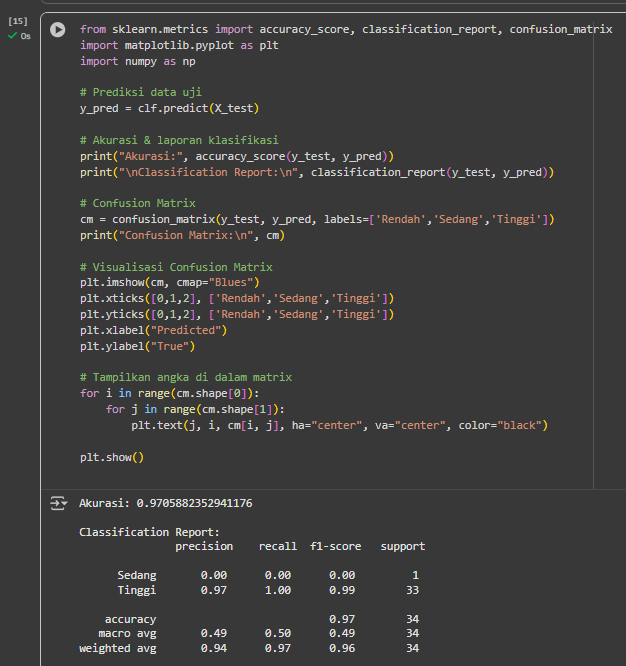
**3. Training Decision Tree**

* Proses training:
  + OneHotEncoder → ubah kategori (Provinsi, Triwulan) ke bentuk numerik.
  + train\_test\_split → bagi data latih & uji (80:20).
  + DecisionTreeClassifier(max\_depth=5) → melatih pohon keputusan dengan kedalaman maksimal 5.
* Output: “Model selesai dilatih”.



**4. Evaluasi Model (Decision Tree)**

* Import metric: accuracy\_score, classification\_report, confusion\_matrix.
* y\_pred = clf.predict(X\_test) → model memprediksi data uji.
* Print akurasi: 97%.
* Classification Report menunjukkan:
  + Sedang: precision, recall = 0 (model gagal memprediksi kelas ini).
  + Tinggi: precision 0.97, recall 1.0 (model sangat bagus di kelas ini).
* Confusion Matrix divisualisasikan:
  + Hampir semua prediksi masuk kelas Tinggi.
  + Ada 1 data “Sedang” salah diprediksi.
* Kesimpulan: model bias ke kelas mayoritas.



**5. Visualisasi Decision Tree**

* Pohon keputusan divisualisasikan dengan plot\_tree.
* Node utama memisahkan berdasarkan **provinsi tertentu** (misalnya Sulawesi Tenggara, Sulawesi Tengah, Sulawesi Utara).
* Hampir semua cabang jatuh ke kelas Tinggi (warna biru).
* Ada sedikit cabang ke kelas Sedang (warna oranye), tapi sangat jarang.
* Ini menunjukkan bahwa model hanya mengenali pola kelas Tinggi, tidak seimbang.

