**LAPORAN ANALISIS DATA BERITA DETIK**

**Nama : Faliqul Ishbah  
Nim : 202210370311148**

**1. Pendahuluan**

Dalam era digital, informasi tersebar luas di berbagai platform berita online. Salah satu cara untuk menganalisis informasi ini adalah dengan melakukan ekstraksi data dan pengelompokan berdasarkan kategori tertentu. Dalam proyek ini, kami mengumpulkan berita dari Detik.com untuk dianalisis lebih lanjut guna mengetahui distribusi berita berdasarkan kategori tertentu.

**2. Tujuan**

Tujuan dari proyek ini adalah:

* Mengumpulkan data berita dari Detik.com dalam periode tertentu.
* Memproses data yang telah dikumpulkan dengan metode Extract, Load, Transform (ELT).
* Menganalisis distribusi berita berdasarkan kategori yang telah ditentukan.
* Menampilkan hasil analisis dalam bentuk visualisasi data (grafik lingkaran).

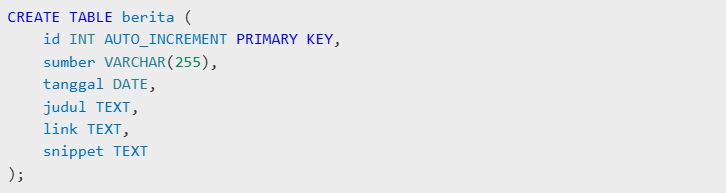
**3. Metode ELT (Extract, Load, Transform)**

**a. Extract (Ekstraksi Data)**

* Data berita dikumpulkan dalam format CSV dengan atribut: sumber, tanggal, judul, link, dan snippet.
* Data diambil dari periode Januari hingga Februari 2025.

**b. Load (Memasukkan Data ke Database MySQL)**

* Data berita dimasukkan ke dalam database MySQL dengan menggunakan skrip Python yang terhubung ke MySQL.
* Struktur tabel dalam database berita\_detik:

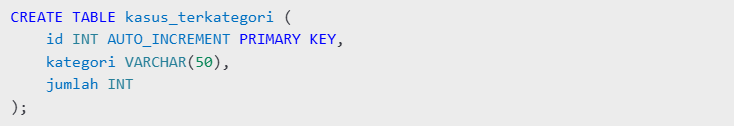


* Data dari CSV dimasukkan menggunakan skrip berikut:



**c. Transform (Pengolahan Data)**

* Berita dikelompokkan berdasarkan kategori seperti Kecelakaan, Korupsi, Pencurian, Pembunuhan, Narkoba, dan Lainnya.
* Data yang telah dikategorikan disimpan dalam tabel kasus\_terkategori:



**4. Analisis Data**

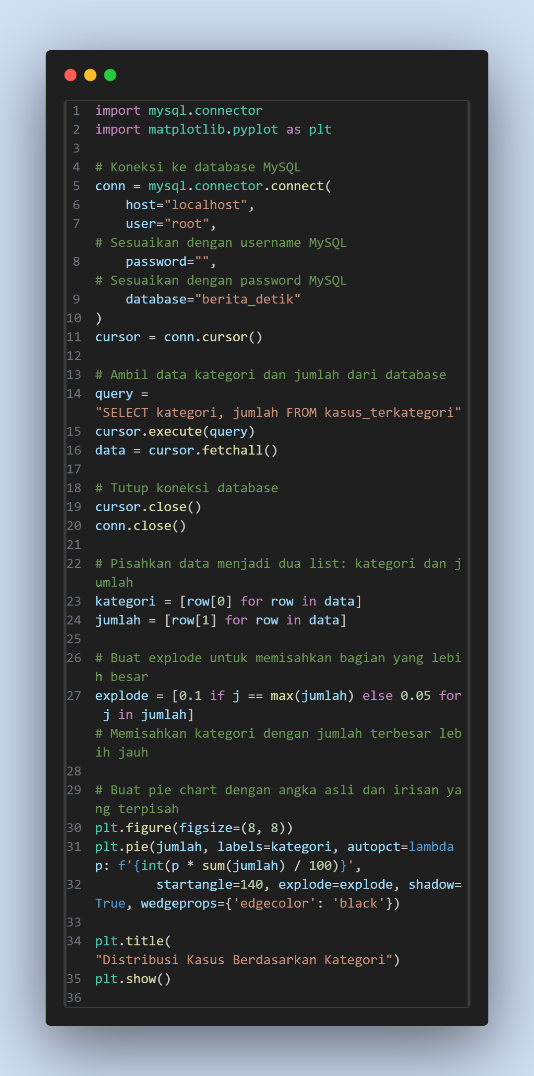
* Berdasarkan data yang telah diproses, distribusi berita dalam kategori-kategori tertentu adalah sebagai berikut:

|  |  |  |  |
| --- | --- | --- | --- |
| **Kategori** | |  |  | | --- | --- | |  | **Jumlah** | |
| Kecelakaan | 43 |
| Korupsi | 21 |
| Pencurian | 24 |
| Pembunuhan | 24 |
| Narkoba | |  | | --- | |  |   18 |
| Lainnya | 1050 |

* Hasil ini menunjukkan bahwa kategori **Lainnya** memiliki jumlah berita terbanyak, sedangkan kategori **Narkoba** memiliki jumlah berita paling sedikit.

**5. Visualisasi Data**

* Data hasil transformasi divisualisasikan dalam bentuk grafik lingkaran menggunakan Python dengan Matplotlib:



**6. Kesimpulan**

* Data berita dari Detik.com berhasil dikumpulkan dan dikategorikan.
* Hasil analisis menunjukkan bahwa berita dalam kategori **Lainnya** mendominasi jumlah total berita.
* Visualisasi dengan grafik lingkaran membantu memahami distribusi kategori berita dengan lebih jelas.
* Metode ELT (Extract, Load, Transform) digunakan untuk mengelola data dari tahap ekstraksi hingga analisis.

**7. Rekomendasi**

* Untuk penelitian lebih lanjut, dapat dilakukan analisis tren berdasarkan waktu untuk melihat bagaimana jumlah berita dalam kategori tertentu berubah dari bulan ke bulan.
* Dapat diterapkan analisis sentimen untuk mengetahui apakah berita cenderung positif, negatif, atau netral.
* Sistem dapat ditingkatkan dengan menerapkan **Machine Learning (NLP)** untuk mengelompokkan berita secara otomatis berdasarkan isi teks.

**8. Lampiran**

* [Github](https://github.com/fall1sh/Data-Informasi-dan-Pengetahuan)
* Hasil grafik analisis data. 