Tugas Kelompok “Mesin Pembelajar untuk Bisnis”

Anggota:

1. Wahyu Prasetya Adi – 23.01.85.0037
2. RR Fitri Damaryanti – 23.01.85.0038
3. Alfin Hilmy N – 23.01.85.0039
4. Randi Afif – 23.01.85.0035

Semester 2

**“Scrapping Data Aplikasi Jobstreet di Google Play Store”**

1. Berikut Kode dan Penjelasannya

A screen shot of a computer program

Description automatically generated

Berikut adalah penjelasan mengenai script tersebut:

**1. Instalasi Library**

!pip install google\_play\_scraper

* **Fungsi**: Menginstal library google-play-scraper untuk melakukan scraping ulasan aplikasi di Google Play Store.
* Library ini mempermudah pengambilan data seperti ulasan, rating, dan informasi aplikasi secara langsung dari Play Store.

**2. Impor Library**

from google\_play\_scraper import app, reviews\_all, Sort

import pandas as pd

import numpy as np

* **Library yang digunakan**:
  + google\_play\_scraper: Untuk scraping data dari Google Play Store.
    - app: Untuk mengambil informasi aplikasi.
    - reviews\_all: Untuk mengambil semua ulasan aplikasi.
    - Sort: Untuk menentukan urutan ulasan (berdasarkan yang terbaru atau yang paling relevan).
  + pandas: Untuk manipulasi data dalam format DataFrame.
  + numpy: Untuk mempermudah manipulasi array.

**3. Scraping Semua Ulasan**

review = reviews\_all(

'com.jobstreet.jobstreet',

sleep\_milliseconds=0,

lang='id',

country='id',

sort=Sort.NEWEST

)

* **Fungsi**: Mengambil semua ulasan aplikasi JobStreet (dengan ID com.jobstreet.jobstreet) dari Play Store.
* **Parameter**:
  + sleep\_milliseconds: Waktu jeda antar permintaan (default 0).
  + lang='id': Bahasa ulasan yang diambil (Bahasa Indonesia).
  + country='id': Hanya mengambil ulasan dari negara Indonesia.
  + sort=Sort.NEWEST: Mengambil ulasan terbaru terlebih dahulu.

**4. Membuat DataFrame dari Ulasan**

df\_reku = pd.DataFrame(np.array(review), columns=['content'])

df\_reku.to\_csv('jobstreet\_all\_column\_review\_mentah\_.csv', index=False)

* **Fungsi**:
  + Mengonversi ulasan (review) menjadi DataFrame menggunakan pandas.
  + Menyimpan data mentah ke file CSV dengan nama jobstreet\_all\_column\_review\_mentah\_.csv.

**5. Memilih Kolom Penting**

df\_reku = pd.DataFrame(np.array(review), columns=['content'])

df\_reku = df\_reku.join(pd.DataFrame(df\_reku.pop('content').tolist()))

df\_selected = df\_reku[['content', 'score']]

df\_selected.to\_csv('jobstreet\_selected\_column\_review\_mentah.csv', index=False)

* **Fungsi**:
  1. **Mengurai Konten**:
     + df\_reku.pop('content').tolist(): Mengekstrak detail dari kolom content (karena struktur content adalah nested dictionary).
     + pd.DataFrame(...).join(...): Menggabungkan detail ke DataFrame utama.
  2. **Memilih Kolom Penting**:
     + df\_reku[['content', 'score']]: Memilih kolom content (isi ulasan) dan score (rating ulasan).
  3. **Menyimpan Data**:
     + Data dengan kolom terpilih disimpan ke file CSV bernama jobstreet\_selected\_column\_review\_mentah.csv.

**Output File**

1. **jobstreet\_all\_column\_review\_mentah\_.csv**:
   * Berisi data mentah dari scraping (semua kolom dan detail).
2. **jobstreet\_selected\_column\_review\_mentah.csv**:
   * Berisi data yang lebih ringkas dengan hanya kolom content dan score.

**Catatan Tambahan**

* **Kemungkinan Error**:
  + Jika aplikasi memiliki terlalu banyak ulasan, scraping mungkin memakan waktu lebih lama.
  + Jika terjadi perubahan API Play Store, library bisa tidak bekerja optimal.
* **Pentingnya Data Mentah**:
  + Menyimpan file mentah membantu jika ingin memanipulasi ulang data di masa mendatang.

2. Berikut Hasilnya:

A screenshot of a computer

Description automatically generated

Berikut adalah penjelasan dari kode yang diberikan:

**1. Membaca Data dari File CSV**

df\_js = pd.read\_csv('jobstreet\_selected\_column\_review\_mentah.csv')

* **Fungsi**: Membaca file CSV (jobstreet\_selected\_column\_review\_mentah.csv) dan mengonversinya menjadi DataFrame df\_js menggunakan pandas.
  + **pd.read\_csv()**: Fungsi ini digunakan untuk membaca file CSV dan mengembalikan data dalam bentuk DataFrame yang bisa digunakan untuk analisis lebih lanjut.
  + **File yang dibaca**: Dalam hal ini, file yang dibaca berisi data ulasan aplikasi JobStreet dengan dua kolom utama: content dan score.

**2. Menampilkan 5 Baris Teratas DataFrame**

df\_js.head()

* **Fungsi**: Menampilkan 5 baris pertama dari DataFrame df\_js menggunakan metode .head().
  + **head()**: Secara default, fungsi ini menampilkan 5 baris pertama dari DataFrame. Jika Anda ingin menampilkan lebih banyak baris, Anda bisa menambahkan parameter seperti df\_js.head(10) untuk menampilkan 10 baris pertama.

**Output yang Dihasilkan**

Output yang ditampilkan adalah 5 baris pertama dari DataFrame, yang berisi dua kolom: content dan score.

Contoh data yang ditampilkan:

content score

0 Tolong perbaiki masalah tidak bisa login ulang... 4

1 Sudah kirim lamaran sebanyak apapun tetap aja ... 5

2 Susah login, tampilan berubah seperti web, apa... 1

3 Aplikasi sering ngebug saat dipakai login 1

4 Sebelumnya applikasi ini oke, tapi kok ini jad... 2

* **content**: Berisi teks ulasan pengguna mengenai aplikasi JobStreet.
* **score**: Berisi rating yang diberikan oleh pengguna (dalam angka dari 1 sampai 5).

Kode ini digunakan untuk memuat dan melihat sampel data dari file CSV untuk memastikan bahwa data telah berhasil dibaca dengan benar dan siap untuk dianalisis lebih lanjut.