

Tugas Kelompok “Mesin Pembelajaran 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

```
!pip install google_play_scraper

from google_play_scraper import app
import pandas as pd
import numpy as np

# Scrap all reviews aplikasi reku di playstore
from google_play_scraper import Sort, reviews_all

review = reviews_all(
    'com.jobstreet.jobstreet',
    sleep_milliseconds=0, # defaults to 0
    lang='id', # defaults to 'en'
    country='id', # defaults to 'us'
    sort=Sort.NEWEST # defaults to Sort.MOST_RELEVANT
)

df_reku = pd.DataFrame(np.array(review), columns=['content'])
# df_reku.to_csv('reku.csv', index=False)
df_reku.to_csv('jobstreet_all_column_review_mentah.csv', index=False)

# Memilih kolom 'score' dan 'content' dalam content
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']]

# Menyimpan hasilnya ke dalam file CSV
df_selected.to_csv('jobstreet_selected_column_review_mentah.csv', index=False)
```

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:

```
# panggil dan tampilkan all column
df_js = pd.read_csv('jobstreet_selected_column_review_mentah.csv')
df_js.head()
```

✓ 0.1s

	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

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:

contentscore

```
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.