

Soal Hackathon-02

Buatlah script Python dengan nama **find_categories.py** yang di dalamnya memiliki fungsi **process_ngram_matching**, dan menerima 6 parameter berikut:

- **master_file**: adalah nama file Excel dalam format xlsx yang berisi deskripsi dan kategori akun untuk pencatatan transaksi keuangan.
- **master_sheet**: adalah nama sheet dari file Excel master_akun.xlsx, yaitu secara default adalah **Data**.
- **trx_file**: adalah nama file Excel dalam format xlsx yang mencatat transaksi sebenarnya dengan nama deskripsi dan nilai pendapatan, pengeluaran atau perubahan tarif.
- **trx_sheet**: adalah nama sheet yang akan dibaca dari file trx_file, secara default adalah **Data**.
- **output_file**: adalah file Excel yang akan dihasilkan oleh function. File ini dalam format xlsx dan akan berisi kolom dengan spesifikasi lebih detil diceritakan selanjutnya di dokumen ini.
- **output_sheet**: adalah nama sheet dari file output Excel, yaitu secara default adalah **Data**.

Contoh pemanggilan fungsi secara default adalah sebagai berikut:

```
process_ngram_matching(  
    master_file="master_akun.xlsx",  
    master_sheet="Data",  
    trx_file="transaksi.xlsx",  
    trx_sheet="Data",  
    output_file="transaksi_dengan_kategori.xlsx",  
    output_sheet="Data"  
)
```

Proses Yang Harus Dilakukan

Fungsi `process_ngram_matching` akan membaca file **transaksi.xlsx** pada sheet **Data**, dimana terdapat 4 kolom yaitu No, Tanggal, Deskripsi Transaksi dan Nominal seperti terlihat pada gambar *screenshot* berikut.

	A	B	C	D
1	No	Tanggal	Deskripsi Transaksi	Nominal
2	1	16-09-2015	Parkir motor penghuni	50000
3	2	25-04-2010	Renov kecil area dapur	400000
4	3	10-03-2012	Sewa kos bulanan	1200000
5	4	28-05-2021	Tagihan air kos	150000
6	5	13-01-2024	Sewa ruang usaha warung	850000
7	6	15-12-2020	Denda bayar kamar	25000
8	7	28-12-2010	Denda terlambat bayar	20000
9	8	17-09-2010	Beli obat flu	30000
10	9	19-02-2010	Perubahan Harga Kamar Kos	1500000
11	10	06-01-2011	Sewa area parkir	20000
12	11	14-09-2016	Pendapatan sewa bulanan kos	1200000
13	12	08-12-2024	Internet kos	300000
14	13	19-11-2011	Renovasi cat area ruang tamu	1500000
15	14	02-02-2012	Kunjungan ke dokter	95000
16	15	14-09-2010	Wifi & internet bulanan	350000
17	16	02-01-2023	Pemeliharaan bangunan dan kamar kos	450000
18	17	10-12-2016	Sewa ruang kecil untuk warung	500000
19	18	18-11-2014	Kenaikan Harga Pembuatan / Pembangunan Kamar Kos per unit	20000000
20	19	12-12-2019	Pendapatan Sewa Kamar	1200000
21	20	11-03-2011	Listrik Bulan Februari	600000
22	21	16-09-2010	Sewa parkir mobil	150000
23	22	17-02-2011	Penyesuaian Sewa Kamar Kos	1500000
24	23	13-02-2025	Laundry kos	80000
25	24	02-02-2016	Tagihan listrik	600000
26	25	13-04-2018	Biaya perawatan bangunan	500000

Kemudian proses akan mengambil teks dari kolom Deskripsi Transaksi sebagai basis untuk melihat apakah transaksi ini bisa dikategorikan sebagai **Pendapatan**, **Pengeluaran** atau **Tarif Baru**.

Nama kategori ini dapat dicari pada file master_akun.xlsx seperti terlihat pada screenshot berikut. Dimana terdapat kolom **Deskripsi** yang akan dipadankan dengan teks **Deskripsi Transaksi** pada file **transaksi.xlsx**, dan kemudian deskripsi yang dapat padanannya akan diambil nilai **Kategori**-nya.

	A	B	C	
1	No	Kategori	Deskripsi	
2	1	Pendapatan	Pendapatan sewa kos bulanan	
3	2	Pendapatan	Layanan catering bulanan	
4	3	Pendapatan	Pendapatan laundry penghuni	
5	4	Pendapatan	Pendapatan sewa parkir motor	
6	5	Pendapatan	Pendapatan sewa parkir mobil	
7	6	Pendapatan	Pendapatan denda keterlambatan bayar	
8	7	Pendapatan	Pendapatan sewa ruang usaha warung	
9	8	Pengeluaran	Biaya pemeliharaan bangunan dan kamar kos	
10	9	Pengeluaran	Biaya kunjungan dokter karena sakit	
11	10	Pengeluaran	Pengeluaran / Pembelian obat	
12	11	Pengeluaran	Biaya listrik bulanan	
13	12	Pengeluaran	Tagihan / biaya air PAM Bulanan	
14	13	Pengeluaran	Biaya bulanan layanan internet & Wi-Fi	
15	14	Pengeluaran	Biaya renovasi kecil (cat, perbaikan pintu)	
16	15	Pengeluaran	Pembelian perabotan kamar kos (kasur, kipas, meja)	
17	16	Pengeluaran	Pajak properti / PBB	
18	17	Tarif Baru	Penyesuaian / Perubahan Biaya Pembangunan Kamar Kos	
19	18	Tarif Baru	Penyesuaian / Perubahan Tarif atau Harga Sewa Kos	
20				

Output terakhir dari proses ini adalah suatu file Excel yang memiliki sheet Data seperti terlihat pada screenshot berikut. Nama file ini dapat dilihat pada contoh file yang diberikan dengan nama file **transaksi_dengan_kategori.xlsx**.

A	B	C	D	E	F
Nomor	Deskripsi Transaksi	Akun	Kategori	Nominal	Tarif Baru
1	Parkir motor penghuni	Pendapatan sewa parkir motor	Pendapatan	50000	0
2	Renov kecil area dapur	Biaya renovasi kecil (cat, perbaikan pintu)	Pengeluaran	-400000	0
3	Sewa kos bulanan	Pendapatan sewa kos bulanan	Pendapatan	1200000	0
4	Tagihan air kos	Tagihan / biaya air PAM Bulanan	Pengeluaran	-150000	0
5	Sewa ruang usaha warung	Pendapatan sewa ruang usaha warung	Pendapatan	850000	0
6	Denda bayar kamar	Pendapatan denda keterlambatan bayar	Pendapatan	25000	0
7	Denda terlambat bayar	Pendapatan denda keterlambatan bayar	Pendapatan	20000	0
8	Beli obat flu	Pengeluaran / Pembelian obat	Pengeluaran	-30000	0
9	Perubahan Harga Kamar Kos	Penyesuaian / Perubahan Biaya Pembangunan Kamar Kos	Tarif Baru	1500000	1500000
10	Sewa area parkir	Pendapatan sewa parkir motor	Pendapatan	20000	0
11	Pendapatan sewa bulanan kos	Pendapatan sewa kos bulanan	Pendapatan	1200000	0
12	Internet kos	Biaya bulanan layanan internet & Wi-Fi	Pengeluaran	-300000	0
13	Renovasi cat area ruang tamu	Biaya renovasi kecil (cat, perbaikan pintu)	Pengeluaran	-1500000	0
14	Kunjungan ke dokter	Biaya kunjungan dokter karena sakit	Pengeluaran	-95000	0
15	Wifi & internet bulanan	Biaya bulanan layanan internet & Wi-Fi	Pengeluaran	-350000	0
16	Pemeliharaan bangunan dan kamar kos	Biaya pemeliharaan bangunan dan kamar kos	Pengeluaran	-450000	0
17	Sewa ruang kecil untuk warung	Pendapatan sewa ruang usaha warung	Pendapatan	500000	0
18	Kenaikan Harga Pembuatan / Pembangunan Kamar Kos per unit	Penyesuaian / Perubahan Biaya Pembangunan Kamar Kos	Tarif Baru	20000000	20000000
19	Pendapatan Sewa Kamar	Pendapatan sewa kos bulanan	Pendapatan	1200000	0
20	Listrik Bulan Februari	Biaya listrik bulanan	Pengeluaran	-600000	0
21	Sewa parkir mobil	Pendapatan sewa parkir mobil	Pendapatan	150000	0
22	Penyesuaian Sewa Kamar Kos	Penyesuaian / Perubahan Tarif atau Harga Sewa Kos	Tarif Baru	1500000	1500000
23	Laundry kos	Pendapatan laundry penghuni	Pendapatan	80000	0
24	Tagihan listrik	Biaya listrik bulanan	Pengeluaran	-600000	0
25	Biaya perawatan bangunan	Biaya pemeliharaan bangunan dan kamar kos	Pengeluaran	-500000	0

Terlihat pada file ini isi **Deskripsi Transaksi** yang berasal dari transaksi.xlsx telah dipadankan dengan master_akun.xlsx untuk mendapatkan **kategori**, disini kemudian ditempatkan di bawah kolom **Akun**.

Kemudian perhatikan kolom nilai berikut:

- Kolom **Nominal**: jika kategorinya Pendapatan maka nilainya akan positif. Tapi jika kategorinya Pengeluaran maka nilainya akan menjadi negatif. Jika kategori Tarif Baru maka selalu bernilai 0.
- Kolom **Tarif Baru**: Jika kategori merupakan Tarif Baru, maka nilai akan diisi dengan nilai positif dari nilai asal. Untuk kategori lain maka diisi dengan nilai 0.

Tantangan

Tantangan dari proses ini adalah pada saat pemadanan atau matching data yang harus dilakukan, dimana kedua deskripsi yang dipadankan tidak persis sama. Untuk contoh, maka kita perhatikan screenshot berikut.

B	C
Deskripsi Transaksi	Akun
Parkir motor penghuni	Pendapatan sewa parkir motor
Renov kecil area dapur	Biaya renovasi kecil (cat, perbaikan pintu)
Sewa kos bulanan	Pendapatan sewa kos bulanan
Tagihan air kos	Tagihan / biaya air PAM Bulanan
Sewa ruang usaha warung	Pendapatan sewa ruang usaha warung
Denda bayar kamar	Pendapatan denda keterlambatan bayar
Denda terlambat bayar	Pendapatan denda keterlambatan bayar
Beli obat flu	Pengeluaran / Pembelian obat
Perubahan Harga Kamar Kos	Perubahan / Perubahan Biaya Pembangunan Kamar Kos

- Terlihat “**Parkir motor penghuni**” tidak ada padanan teks yang persis sama di file master_akun.xlsx. Tapi yang ada adalah “**Pendapatan sewa parkir motor**”, dan inilah yang kita ambil.
- Terlihat juga “**Renov kecil area dapur**” tidak ada padanan teks yang persis sama di file master_akun.xlsx. Tapi yang ada adalah “**Pendapatan sewa parkir motor**”, dan inilah yang kita ambil.
- Dan seterusnya

Padanan ini bisa dilakukan dengan menggunakan algoritma similarity yang merupakan tugas para peserta untuk mencari algoritma yang tepat (dengan batasan yang ada) di bawah ini.

Batasan dan Tenggat Waktu

Batasan yang harus diperhatikan pada saat pembuatan script python ini:

- Library yang boleh digunakan hanya yang built-in dari Python 3.10, panda dan openpyxl, tidak boleh ada library lain.
- Nama file script yang dikirimkan harus bernama find_categories.py
- Format file output yang dihasilkan tidak boleh diubah, perhatikan dengan seksama nama dan format data yang ada pada transaksi_dengan_kategori.xlsx yang dikirimkan sebagai contoh data dari soal ini.
- File script yang dihasilkan harus dikirimkan sebagai attachmen ke email tugastest@ujikompetensi.com dengan subjek

Solusi Hackathon untuk soal 'HACKATHON-02'

- Submission atau pengiriman hasil tidak boleh melewati tanggal 18 November 2025, pukul 10.00 WIB. Di atas waktu tersebut, maka pengiriman hasil tidak akan diproses dan tidak berlaku.