**1.2 Tugas 2 Karakteristik Data**

Pada data yang digunakan, berikut karakteristik masing-masing kolom:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Kolom | Penjelasan | Kategori | Range/Atribut Kategori | Unique Values |
| publish\_time | Waktu mengunggah video | Categorical (belum Time Series) | (2021-01-11T12:21:56Z -- 2021-11-19T02:45:03Z) | 10157 |
| channel\_id | ID unik tiap channel pemilik video trending | Categorical | - | 2043 |
| title | Judul/nama video trending yang berada pada database | Categorical | - | 10405 |
| description | Deskripsi dari video trending | Categorical | - | 9588 |
| thumbnail\_url | Link tiap video untuk mengakses video pada youtube | Categorical | - | 1 |
| thumbnail\_widht | Ukuran lebar thumbnail | Kuantitatif | - | 1 |
| thumbnail\_height | Ukuran tinggi thumbnail | Kuantitatif | - | 1 |
| channel\_name | Nama channel yang mengupload video | Categorical | - | 2063 |
| tags | Kumpulan kata kunci (sebagai keyword pencarian) | Categorical | - | 6854 |
| category\_id | Key category (masing-masing video memiliki kategori yang berbeda) | Categorical (Nominal) | 22, 24, 23, 28, 1, 25, 17, 10, 15, 27, 26, 2, 19, 20, 29, 43 | 15 |
| live\_status | Status live video | Categorical | - | 1 |
| local\_title | Judul lokal video | Categorical | - | 10405 |
| local\_description | Deskripsi lokal tentang video | Categorical | - | 9588 |
| duration | Lama durasi video trending | Categorical | - | 2733 |
| dimension | Jenis dimensi dari video | Categorical | - | 1 |
| definition | Kualitas dari video yang dapat dilihat | Categorical (Binary) | HD,/SD | 2 |
| caption | Ketersediaan Subtitle/Caption | Categorical (Binary) | True/False | 2 |
| license\_status | Ketersediaan lisensi dari video | Categorical (Binary) | True/False | 58 |
| allowed\_region | Region yang dapat mengakses video | Categorical | - | 122 |
| blocked\_region | Region yang diblokir untuk mengakses video | Categorical | - | 49229 |
| view | Banyaknya video tersebut ditonton | Kuantitatif | (15282.0 - 335957560.0) | 37510 |
| like | Banyaknya likes setiap video | Kuantitatif | (16.0 - 17005260.0) | 10045 |
| dislike | Banyaknya dislike setiap video | Kuantitatif | (0.0 - 392340.0) | 1 |
| favorite | Banyaknya favorit yang diberikan kepada video | Kuantitatif | - |  |
| comment | Banyaknya komentar pada video | Kuantitatif | (0.0 - 7195733.0) | 16171 |
| trending\_time | Tanggal video tersebut termasuk dalam video trending | Categorical (belum Time Series) | (2021-02-07 05:46:51.832614+00:00 -- 2021-11-19 11:29:39.271647+00:00) | 49722 |
| category\_name | Kategori video berdasarkan key category\_id | Categorical (Nominal) | 'People & Blogs',  'Entertainment',  'Comedy',  'Science &  Technology', 'Film  & Animation',  'News & Politics',  'Sports',  'Music', 'Pets &  Animals',  'Education',  'Howto & Style',  'Autos &  Vehicles', 'Travel  & Events',  'Gaming',  'Nonprofits &  Activism', 'Shows' | 15 |

Keterangan :

\* data json yang telah di-append dan disesuaikan dengan key category\_id (Data lain 40949 sedangkan description hanya 40379)

# Menampilkan ukuran dari data

print('Jumlah Baris :', len(df))

print('Jumlah Kolom :', len(df.columns))

# Menampilkan banyak data unik dari tiap kolom

for i in range(len(df.columnsdf.columns)):

print(f'Data unik pada kolom {df.columns[i]} ada sebanyak {len(df.iloc[:,i].unique())}')

# Mencari range dari tiap kolom dengan tipe data kuantitatif (numerik dan time-series)

kuantitatif = ['trending\_time', 'publish\_time', 'view', 'like', 'dislike']

index = 1

for i in kuantitatif:

print(f'{index}. Kolom {i} memiliki range : ({df[i].min()} - {df[i].max()}) ')

index += 1

print()

# Mencari member category dari kolom dengan tipe data kategorikal

print('Member pada kolom category\_name dan category\_id :')

print('No. category\_name - category\_id')

index = 1

for i in range(len(df['category\_name'].unique())):

print(f"{index}. {df['category\_name'].unique()[i]} - {df['category\_id'].unique()[i]}")

index += 1

# Menampilkan info data

df.info()