Nama : Nifelling Rosmelia Sandewa

NIM : 21110014

Tugas NLP

```
#Load Data Menjadi Dataframe

import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer

# Memuat dataset dari file CSV
df = pd.read_csv('/content/data.csv')
df
```

	Unnamed: 0	userName	content	score
0	0	Gadri Bandu	Saya kasih bintang tiga dulu soalnga gemenya b...	3
1	1	070_Putri Azzahra Nurdin	Bagus bgtt. Dari segi story, grafik sampe puzz...	5
2	2	Novrealita Setiyana	Sejauh ini aku main aku suka banget sama model...	5
3	3	auah gelap	Semuanya sudah sempurna tapi Sangat disayangka...	3
4	4	Adi Nugroho	Grafik bagus,story bagus,terutama archon quest...	5
...	...	...	...	...
995	995	TEGUH TEAM	Masih bingung sih dengan cara game ini ngerend...	5
996	996	Zen Flow	game nya seru dari segi story. saran buat Hoyo...	5

Data diatas terdiri dari 1000 baris dan 4 kolom bernama "Unnamed", "Username", "Content", dan "Score".

```
# Inisialisasi TfidfVectorizer
tfidf_vectorizer = TfidfVectorizer()

# Mengubah kolom "content" menjadi vektor TF-IDF
tfidf_matrix = tfidf_vectorizer.fit_transform(df['content'])

# Menampilkan DataFrame vektor TF-IDF
print(tfidf_matrix)
```

```
  (0, 3251)     0.1055299758616986
  (0, 2602)     0.14500522566335713
  (0, 538)      0.18749785600261215
  (0, 340)      0.22117246882497477
  (0, 953)      0.05838791659996207
  (0, 1680)     0.06532245345690199
  (0, 2368)     0.12440423861415989
  (0, 1785)     0.135243996694897
  (0, 3502)     0.12713065282236005
  (0, 1862)     0.09833987824244249
  (0, 1341)     0.1718400527197851
  (0, 639)      0.12951658229423796
  (0, 3145)     0.12312688906052045
  (0, 3440)     0.1718400527197851
  (0, 2802)     0.15059373755015762
  (0, 1915)     0.3998523073106945
  (0, 370)      0.19992615365534724
  (0, 443)      0.22117246882497477
  (0, 870)      0.058161701894011326
  (0, 2147)     0.16625154083298463
  (0, 428)      0.08481675158152252
  (0, 2560)     0.17867983848571975
  (0, 3355)     0.20874417117223965
  (0, 4099)     0.1756559033334547
  (0, 2933)     0.05691042801487717
  :     :
  (997, 1364)   0.06318278374092187
  (997, 377)    0.14013407953934043
  (998, 329)    0.472267532819263
  (998, 1558)   0.472267532819263
  (998, 1058)   0.472267532819263
  (998, 3674)   0.2908434965194849
  (998, 3488)   0.2469431713090349
```

```
  (998, 4230)   0.21466786789401426
  (998, 1798)   0.2334990565070229
  (998, 1472)   0.23621500609881577
  (998, 870)    0.17005184800078044
  (999, 4228)   0.389031112703819
  (999, 2478)   0.36717036986116447
  (999, 2774)   0.36717036986116447
  (999, 2445)   0.2924279964074674
  (999, 2184)   0.3374695547009886
  (999, 1139)   0.24674574097621849
  (999, 201)    0.16074830460404113
  (999, 2167)   0.14074232244941945
  (999, 2377)   0.22888054618306697
  (999, 1070)   0.2347149060295643
  (999, 840)    0.1881595104507794
  (999, 1364)   0.25831107671776826
  (999, 4037)   0.1803144362269218
  (999, 2385)   0.17599891557202452
```

Interpretasi hasil :

(0, 3251): Ini mengindikasikan bahwa pada kalimat pertama (indeks 0), kata dengan indeks 3251 dalam matriks (kata tertentu) memiliki bobot TF-IDF sebesar 0.1055299758616986.

```
# Daftar nama fitur
feature_names = tfidf_vectorizer.get_feature_names_out()

# Output
print(feature_names)
```

```
['00' '000' '000an' ... 'zhong' 'zhongli' 'zonk']
```

```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer

# Inisialisasi TfidfVectorizer
tfidf_vectorizer = TfidfVectorizer()

# Mengubah kolom "content" menjadi vektor TF-IDF
tfidf_matrix = tfidf_vectorizer.fit_transform(df['content'])

# Menampilkan DataFrame vektor TF-IDF
df_tfidf = pd.DataFrame(tfidf_matrix.toarray(), columns=tfidf_vectorizer.get_feature_names_out())

# Jika Anda ingin melihat hasilnya
df_tfidf
```

	00	000	000an	08102020	0sebelumnya	10	100	1000	100ms	107	...	yoimiya
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...
995	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
996	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
997	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
998	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
999	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0

1000 rows × 4353 columns

Output tersebut adalah matriks vektor TF-IDF yang merepresentasikan bobot kata-kata dalam kalimat-kalimat dari data awal. Sebagian besar nilai dalam matriks adalah 0 (nol), yang menunjukkan kata-kata yang tidak signifikan dalam dokumen tersebut. Nilai non-nol mewakili bobot kata-kata yang signifikan dalam kalimat atau dokumen tertentu, tetapi informasi lebih lanjut diperlukan untuk menentukan kata-kata mana yang memiliki bobot tertinggi dalam setiap kalimat.

```
# Membuat DataFrame dari vektor TF-IDF dengan nama fitur dan indeks sesuai dengan konten
tfidf_df = pd.DataFrame(tfidf_matrix.toarray(), columns=feature_names, index=df['content'])

#DataFrame vektor TF-IDF
print("Matrik TFIDF:")
```

```
print(tfidf_df)
```

```
                                                  1000  100ms  107  ...  \
content                                                               ...
Saya kasih bintang tiga dulu soalnga gemenya ba...   0.0    0.0  0.0  ...
Bagus bgtt. Dari segi story, grafik sampe puzzl...   0.0    0.0  0.0  ...
Sejauh ini aku main aku suka banget sama model ...   0.0    0.0  0.0  ...
Semuanya sudah sempurna tapi Sangat disayangkan...   0.0    0.0  0.0  ...
Grafik bagus,story bagus,terutama archon quest,...   0.0    0.0  0.0  ...
...                                                  ...    ...  ...  ...
Masih bingung sih dengan cara game ini ngerende...   0.0    0.0  0.0  ...
game nya seru dari segi story. saran buat Hoyov...   0.0    0.0  0.0  ...
Game ini sangat bagus dengan map yg sangat luas...   0.0    0.0  0.0  ...
Sangat seru dan asik untuk dimainkan grafik hd ...   0.0    0.0  0.0  ...
Memuat game terlalu lama . Mau masuk game aja l...   0.0    0.0  0.0  ...

                                                  yoimiya  you  youtube  \
content
Saya kasih bintang tiga dulu soalnga gemenya ba...      0.0  0.0      0.0
Bagus bgtt. Dari segi story, grafik sampe puzzl...      0.0  0.0      0.0
Sejauh ini aku main aku suka banget sama model ...      0.0  0.0      0.0
Semuanya sudah sempurna tapi Sangat disayangkan...      0.0  0.0      0.0
Grafik bagus,story bagus,terutama archon quest,...      0.0  0.0      0.0
...                                                     ...  ...      ...
Masih bingung sih dengan cara game ini ngerende...      0.0  0.0      0.0
game nya seru dari segi story. saran buat Hoyov...      0.0  0.0      0.0
Game ini sangat bagus dengan map yg sangat luas...      0.0  0.0      0.0
Sangat seru dan asik untuk dimainkan grafik hd ...      0.0  0.0      0.0
Memuat game terlalu lama . Mau masuk game aja l...      0.0  0.0      0.0

                                                   yt   yu  yutub  zaman  \
content
Saya kasih bintang tiga dulu soalnga gemenya ba...  0.0  0.0    0.0    0.0
Bagus bgtt. Dari segi story, grafik sampe puzzl...  0.0  0.0    0.0    0.0
Sejauh ini aku main aku suka banget sama model ...  0.0  0.0    0.0    0.0
Semuanya sudah sempurna tapi Sangat disayangkan...  0.0  0.0    0.0    0.0
Grafik bagus,story bagus,terutama archon quest,...  0.0  0.0    0.0    0.0
...                                                 ...  ...    ...    ...
Masih bingung sih dengan cara game ini ngerende...  0.0  0.0    0.0    0.0
game nya seru dari segi story. saran buat Hoyov...  0.0  0.0    0.0    0.0
Game ini sangat bagus dengan map yg sangat luas...  0.0  0.0    0.0    0.0
Sangat seru dan asik untuk dimainkan grafik hd ...  0.0  0.0    0.0    0.0
Memuat game terlalu lama . Mau masuk game aja l...  0.0  0.0    0.0    0.0

                                                  zhong  zhongli  zonk
content
Saya kasih bintang tiga dulu soalnga gemenya ba...   0.0      0.0   0.0
Bagus bgtt. Dari segi story, grafik sampe puzzl...   0.0      0.0   0.0
Sejauh ini aku main aku suka banget sama model ...   0.0      0.0   0.0
Semuanya sudah sempurna tapi Sangat disayangkan...   0.0      0.0   0.0
Grafik bagus,story bagus,terutama archon quest,...   0.0      0.0   0.0
...                                                  ...      ...   ...
Masih bingung sih dengan cara game ini ngerende...   0.0      0.0   0.0
game nya seru dari segi story. saran buat Hoyov...   0.0      0.0   0.0
Game ini sangat bagus dengan map yg sangat luas...   0.0      0.0   0.0
Sangat seru dan asik untuk dimainkan grafik hd ...   0.0      0.0   0.0
Memuat game terlalu lama . Mau masuk game aja l...   0.0      0.0   0.0

[1000 rows x 4353 columns]
```

Output tersebut adalah DataFrame yang menggambarkan matriks vektor TF-IDF. Setiap baris mewakili satu kalimat atau dokumen dari data asli, dan setiap kolom mewakili kata-kata atau fitur. Nilai dalam sel-sel DataFrame adalah bobot TF-IDF untuk kata-kata tersebut dalam kalimat yang sesuai. DataFrame ini memungkinkan analisis dan pemahaman lebih lanjut tentang pentingnya kata-kata dalam setiap dokumen.

```
tfidf_df.to_excel("Tugas_Prak_NLP.xlsx", index=False)
```