LAPORAN TUGAS MACHINE LEARNING

(Ekstraksi Fitur)

Dibuat Oleh :

NAMA : MUHAMMAD NURUL MUSTOFA

NIM : 2241720022

KELAS : TEKNIK INFORMATIKA-3A

PROGRAM STUDI D-IV TEKNIK INFORMATIKA

JURUSAN TEKNOLOGI INFORMASI

POLITEKNIK NEGERI MALANG

Jl. Soekarno Hatta No.9, Jatimulyo, Kec. Lowokwaru, Kota Malang, Jawa Timur 65141

6. Studi Kasus aEkstraksi Fitur dari Data Teks

praktikum - praktikum6.py

```python
from sklearn.feature_extraction.text import TfidfVectorizer
corpus = [
    'the house had a tiny little mouse',
    'the cat saw the mouse',
    'the mouse ran away from the house',
    'the cat finally ate the mouse',
    'the end of the mouse story'
]

# inisialisasi obyek TFidVectorizer
vect = TfidfVectorizer(stop_words='english')

# Pembobotan TF-IDF
resp = vect.fit_transform(corpus)

# Cetak Hasil
print(resp)

print(vect.get_feature_names_out())
```

Hasil

```
  (0, 7)        0.2808823162882302
  (0, 6)        0.5894630806320427
  (0, 11)       0.5894630806320427
  (0, 5)        0.47557510189256375
  (1, 9)        0.7297183669435993
  (1, 2)        0.5887321837696324
  (1, 7)        0.3477147117091919
  (2, 1)        0.5894630806320427
  (2, 8)        0.5894630806320427
  (2, 7)        0.2808823162882302
  (2, 5)        0.47557510189256375
  (3, 0)        0.5894630806320427
  (3, 4)        0.5894630806320427
  (3, 2)        0.47557510189256375
  (3, 7)        0.2808823162882302
  (4, 10)       0.6700917930430479
  (4, 3)        0.6700917930430479
  (4, 7)        0.3193023297639811
['ate' 'away' 'cat' 'end' 'finally' 'house' 'little' 'mouse' 'ran' 'saw'
 'story' 'tiny']
```

Tugas Praktikum

1. Salin kalimat pada Kode 1-7 dengan tanda baca titik pada setiap kalimatnya dengan menggunakan editor teks.

2. Simpan kalimat tersebut pada file ‘.txt’ dengan nama ‘corpus.txt’.

3. Lakukan proses ektraksi fitur TF-IDF dengan menggunakan file ‘corpus.txt’.

Jawaban :

Corpus.txt

corpus.txt

```
1  the house had a tiny little mouse.
2  the cat saw the mouse.
3  the mouse ran away from the house.
4  the cat finally ate the mouse.
5  the end of the mouse story.
```

Code :

praktikum - tugas.py

```python
from sklearn.feature_extraction.text import TfidfVectorizer

with open('corpus.txt', 'r') as file:
    corpus = file.readlines()


# inisialisasi obyek TFidVectorizer
vect = TfidfVectorizer(stop_words='english')
# Pembobotan TF-IDF
resp = vect.fit_transform(corpus)

# Cetak Hasil
print(resp)

print(vect.get_feature_names_out())
```

Hasil :

```
  (0, 7)        0.2808823162882302
  (0, 6)        0.5894630806320427
  (0, 11)       0.5894630806320427
  (0, 5)        0.47557510189256375
  (1, 9)        0.7297183669435993
  (1, 2)        0.5887321837696324
  (1, 7)        0.3477147117091919
  (2, 1)        0.5894630806320427
  (2, 8)        0.5894630806320427
  (2, 7)        0.2808823162882302
  (2, 5)        0.47557510189256375
  (3, 0)        0.5894630806320427
  (3, 4)        0.5894630806320427
  (3, 2)        0.47557510189256375
  (3, 7)        0.2808823162882302
  (4, 10)       0.6700917930430479
  (4, 3)        0.6700917930430479
  (4, 7)        0.3193023297639811
['ate' 'away' 'cat' 'end' 'finally' 'house' 'little' 'mouse' 'ran' 'saw'
 'story' 'tiny']
```