

Praktikum Pemrosesan Data

Text Processing

Wahyu Ikbal Maulana

3323600056

D3 SDT B

Politeknik Elektronika Negeri Surabaya

```
In [ ]: kalimat="Wahyu ikbal maulana sedang asyik membaca buku"
        LowerCase = kalimat.lower()
        print(LowerCase)
```

wahyu ikbal maulana sedang asyik membaca buku

```
In [ ]: import re
        import string

        kalimat = "Wahyu ikbal {{maulana}} sedang asyik bermain mobile legends"

        hasil = kalimat.lower()
        hasil = re.sub(r'\d+', '', hasil)
        translator = str.maketrans('', '', string.punctuation)
        hasil = hasil.translate(translator)

        print(kalimat)
        print(hasil)
```

Wahyu ikbal {{maulana}} sedang asyik bermain mobile legends
wahyu ikbal maulana sedang asyik bermain mobile legends

```
In [ ]: kalimat="Wahyu ikbal maulana sedang belajar "
        pisah = kalimat.split()
        print(pisah)
```

['Wahyu', 'ikbal', 'maulana', 'sedang', 'belajar']

```
In [ ]: from nltk.tokenize import word_tokenize
kalimat="Kerakyatan yang dipimpin oleh hihmat dalam pemusyawaratan
tokens=word_tokenize (kalimat)
print(tokens)
```

```
['Kerakyatan', 'yang', 'dipimpin', 'oleh', 'hihmat', 'dalam', 'pe
musyawaratan', 'dan', 'perwakilan']
```

```
In [ ]: from nltk.tokenize import sent_tokenize
kalimat="wahyu ikbal maulana sedang olahraga. Di besoknya beliau
token=sent_tokenize(kalimat)
print(token)
```

```
['Persatuan Indonesia.', 'Kerakayatan yang dipimpin oleh hikmat d
alam permusyawaratan Indonesia']
```

```
In [ ]: import Sastrawi
from Sastrawi.StopWordRemover. StopWordRemoverFactory import Stc
factory=StopWordRemoverFactory()
stopwords=factory.get_stop_words()
print(stopwords)
```

```
['yang', 'untuk', 'pada', 'ke', 'para', 'namun', 'menurut', 'anta
ra', 'dia', 'dua', 'ia', 'seperti', 'jika', 'jika', 'sehingga',
'kembali', 'dan', 'tidak', 'ini', 'karena', 'kepada', 'oleh', 'sa
at', 'harus', 'sementara', 'setelah', 'belum', 'kami', 'sekitar',
'bagi', 'serta', 'di', 'dari', 'telah', 'sebagai', 'masih', 'ha
l', 'ketika', 'adalah', 'itu', 'dalam', 'bisa', 'bahwa', 'atau',
'hanya', 'kita', 'dengan', 'akan', 'juga', 'ada', 'mereka', 'suda
h', 'saya', 'terhadap', 'secara', 'agar', 'lain', 'anda', 'begit
u', 'mengapa', 'kenapa', 'yaitu', 'yakni', 'daripada', 'itulah',
'lagi', 'maka', 'tentang', 'demi', 'dimana', 'kemana', 'pula', 's
ambil', 'sebelum', 'sesudah', 'supaya', 'guna', 'kah', 'pun', 'sa
mpai', 'sedangkan', 'selagi', 'sementara', 'tetapi', 'apakah', 'k
ecuali', 'sebab', 'selain', 'seolah', 'seraya', 'seterusnya', 'ta
npa', 'agak', 'boleh', 'dapat', 'dsb', 'dst', 'dll', 'dahulu', 'd
ulunya', 'anu', 'demikian', 'tapi', 'ingin', 'juga', 'nggak', 'ma
ri', 'nanti', 'melainkan', 'oh', 'ok', 'seharusnya', 'sebetulny
a', 'setiap', 'setidaknya', 'sesuatu', 'pasti', 'saja', 'toh', 'y
a', 'walau', 'tolong', 'tentu', 'amat', 'apalagi', 'bagaimanapu
n']
```

```
In [ ]: from Sastrawi.StopWordRemover. StopWordRemoverFactory import Stc
from nltk.tokenize import word_tokenize
factory=StopWordRemoverFactory()
stopword=factory.create_stop_word_remover()
kalimat="Hampir satu bulan ini semua proses kegiatan perkuliahar
kalimat=kalimat.lower()
stop=stopword.remove(kalimat)
```

```
tokens=word_tokenize (stop)
print(tokens)
```

```
['hampir', 'satu', 'bulan', 'semua', 'proses', 'kegiatan', 'perku  
liahan', 'darung']
```

```
In [ ]: from Sastrawi. Stemmer.StemmerFactory import StemmerFactory
factory=StemmerFactory()
stemmer=factory.create_stemmer()
kalimat="Dampak positif dari virus corona terbukanya peluang pas
katadasar=stemmer.stem(kalimat)
print(katadasar)
```

```
dampak positif dari virus corona buka peluang pasar baru
```