| **Nama :**  **Nim :** | **Praktikum Penambangan Teks** | **MODUL 3**    **Nama Dosen :**  **Syandra Sari, S.Kom, M.Kom** |
| --- | --- | --- |
| **Hari dan Tanggal :** | **Nama Asisten Laboratorium :**   1. **Ridho Rachmat Giffary (064001800006)** 2. **Tasya Aulia (065001800022)** |

**Praktikum 3 :**

**Tokenizing & Normalize**

**KODE POKOK BAHASAN:**

**DESKRIPSI POKOK BAHASAN:**

Mengenalkan dan mengajarkan cara melakukan salah satu proses pada tahapan Pre-Processing yaitu Tokenizing dan Normalize yang akan digunakan untuk membersihkan data yang telah ditarik dari platform social media twitter sebelumnya.

| No | Elemen Kompetensi | Indikator Kerja | Jml Jam | Hal |
| --- | --- | --- | --- | --- |
| 1 | Melakukan proses Tokenizing | Mampu melakukan proses Tokenizing terhadap data yang telah dihimpun | 30 menit | 2 |
| 2 | Melakukan proses Normalize | Mampu melakukan proses Normalize Terhadap kata dalam kalimat | 60 menit |  |

**TUGAS PENDAHULUAN:**

Hal yang harus dilakukan dan acuan yang harus dibaca sebelum praktikum:

1. Mengetahui pemahaman dasar tentang tahapan text mining pada modul sebelumnya.
2. Mempersiapkan Aplikasi Jupyter Notebook yang sudah terinstall di Laptop masing - masing.

**TEORI SINGKAT:**

| Tuliskan Teori singkat terkait Tokenizing dan Normalize pada Text Mining minimal 2 Paragraf |
| --- |

**LAB SETUP:**

Hal yang harus disiapkan dan dilakukan oleh praktikan untuk menjalankan praktikum modul ini.

**Alat dan Bahan:** Laptop

**Software** : Jupyter Notebook

**ELEMEN KOMPETENSI I:**

**Deskripsi:**

Melakukan proses Tokenizing

**Kompetensi Dasar:**

Mampu melakukan proses Tokenizing terhadap data yang telah dihimpun

**Latihan 1:**

**Penjelasan Singkat:**

Pada latihan ini anda diminta untuk melakukan proses Tokenizing terhadap data yang telah dihimpun pada tugas sebelumnya.

Python

| # Tokenizing nltk.download('punkt')  def word\_tokenize\_wrapper(text):  return word\_tokenize(text)  Nama\_kelompok['tweet'] = Nama\_kelompok['tweet'].apply(word\_tokenize\_wrapper)  Nama\_kelompok['tweet'] |
| --- |

| letakan screenshot codingan kalian disini! |
| --- |

**ELEMEN KOMPETENSI 2:**

**Deskripsi:**

Melakukan proses Normalize

**Kompetensi Dasar:**

Mampu melakukan proses Normalize Terhadap kata dalam kalimat

**Latihan 2:**

**Penjelasan Singkat:**

Pada latihan ini anda diminta untuk melakukan proses Normalize Terhadap kata dalam kalimat

Python

download file normalisasi disini  
<https://drive.google.com/drive/folders/1JlcfIKAs1fl9GWUaZn0kcB88-2W1tN5w?usp=sharing>

| # Normalize  normalizad\_word = pd.read\_excel("normalisasi.xlsx")  normalizad\_word\_dict = {}  for index, row in normalizad\_word.iterrows():  if row[0] not in normalizad\_word\_dict:  normalizad\_word\_dict[row[0]] = row[1]  def normalized\_term(document):  return [normalizad\_word\_dict[term] if term in normalizad\_word\_dict else term for term in document]  Nama\_kelompok['tweet'] = Nama\_kelompok['tweet'].apply(normalized\_term)  Nama\_kelompok['tweet'].head(10) |
| --- |

| letakan screenshot codingan kalian disini! |
| --- |



| **No** | **Elemen Kompetensi** | **Penyelesaian**  **Cek List** | |
| --- | --- | --- | --- |
| **Selesai** | **Tidak** |
| **1** | **Melakukan proses Tokenizing** |  |  |
| **2.** | **Melakukan proses Normalize** |  |  |

1. **Form Umpan Balik**

| **Elemen Kompetensi** | **Waktu Pengerjaan** | **Kriteria** |
| --- | --- | --- |
| **Melakukan proses Tokenizing** |  |  |
| **Melakukan proses Normalize** |  |  |

Kriteria

1.Sangat Menarik

2.Cukup Menarik

3.Kurang Menarik

4.Sangat Kurang Menarik