



**INSTITUT TEKNOLOGI TELKOM PURWOKERTO**  
**SOAL EVALUASI PROSES PEMBELAJARAN**  
**SEMESTER GANJIL 2023 / 2024**  
**PROGRAM STUDI S1 – SAINS DATA**

|                 |          |                                 |                           |
|-----------------|----------|---------------------------------|---------------------------|
| No. Revisi : 00 | 1 dari 2 | No. Dokumen : IT-TEL/RM/AKA/007 | Tanggal : 22 Oktober 2017 |
|-----------------|----------|---------------------------------|---------------------------|

|                      |  |                |
|----------------------|--|----------------|
| <b>Mata Kuliah</b>   | : Praktikum Pemrosesan Bahasa Alami  | Review Kaprodi |
| <b>SKS / Kelas</b>   | : 1/ S1SD-01-A, S1SD-02-A  |                |
| <b>Semester</b>      | : 5  |                |
| <b>Pengampu</b>      | : Siti Khomsah, S.Kom., M.Cs   | Review KK      |
| <b>Hari/Tanggal</b>  | : 20 Oktober 2023  |                |
| <b>Waktu</b>         | : 150 Menit  |                |
| <b>Sifat Ujian</b>   | : -  |                |
| <b>Metode Ujian</b>  | : <b>Responsi</b>  |                |
| <b>Petunjuk Soal</b> | : 1. Berdoalah sebelum mengerjakan<br>2. Bacalah petunjuk pelaksanaan ujian dengan cermat<br>3. Selama ujian, mahasiswa tidak diperkenankan untuk bekerjasama dengan peserta ujian lainnya ataupun menerima bantuan jawaban dari siapapun, jika kedapatan contek mencotek maka akan diberikan sanksi yaitu nilai 0.<br>4. Selama ujian, HP silence dan dimasukkan tas<br>5. Jawaban program gunakan template yang ada. Simpan file jawaban ipynb menyesuaikan NIM dan Nama Anda.<br>6. Jawaban yang dikumpulkan adalah kode program (ipynb), pdf, dan hasil2 excel/csv. Kirim dalam folder github.<br>7. Anda juga wajib mengirimkan link github ke LMS. |                |

CPMK1: Mahasiswa mampu menjelaskan teori komputasi linguistic dan menggunakan dengan tepat metode-metode pemrosesan bahasa alami (CPL9/P6, CPL6/P3)

Sub-CPMK-1.1 Mahasiswa dapat mendeskripsikan (C2) teori-teori terkait komputasi bahasa, pengetahuan terkait bahasa, dan dasar-dasar ekspresi reguler dan menerapkan (C3) ekspresi regular untuk pengolahan data teks (P6).

Sub-CPMK-1.2 Mahasiswa mampu menjelaskan (C2) dan menerapkan (C3) metode-metode text preprocessing meliputi cleansing, tokenizing, stemming(P6, P3)

**1. CPMK1- Sub CPMK1.1( bobot : 10 )**

Lakukan crawling data teks dari media sosial/ web dan simpan hasilnya dalam bentuk excel/csv. Setiap mahasiswa harus melakukan crawling dengan kata kunci tertentu. Kata kunci tidak boleh sama dengan mahasiswa lainnya.

**2. CPMK1- Sub CPMK1.1. (25)**

Lakukan pembersihan meliputi penghapusan punctuation, angka dan karakter yang tidak penting menggunakan menggunakan regex.

Simpan hasilnya menjadi file csv/excel.

**3. CPMK1- Sub CPMK1.2.**

a. Lakukan parsing dan simpan hasilnya dalam bentuk excel/csv (**bobot:10**)

|                 |          |                                 |                                  |
|-----------------|----------|---------------------------------|----------------------------------|
| No. Revisi : 00 | 2 dari 2 | No. Dokumen : IT-TEL/RM/AKA/007 | Tanggal Terbit : 22 Oktober 2017 |
|-----------------|----------|---------------------------------|----------------------------------|

- b. Carilah kata-kata slangword yang ada dalam dataset Anda, dengan cara mencocokkan dengan kamus KBBI (terlampir). Simpan hasilnya dalam bentuk csv/excel. Tampilkan 100 kata slang yang Anda dapatkan dan tampilkan dalam bentuk Gunakan (**bobot:25**)
- c. Lakukan tokenizing berdasarkan hasil 3b, simpan hasilnya dalam bentuk csv/excel dan tampilkan 100 token pertama. (**bobot:10**)
- d. Lakukan stopwords removing berdasarkan hasil 3c, simpan hasilnya dalam bentuk csv/excel. (**bobot:10**)
- e. Lakukan stemming berdasarkan hasil 3d dan tampilkan 100 stem pertama. (**bobot:10**)

**SELAMAT MENGERJAKAN**