**SENTIMEN ANALISIS**

**PRAKTIKUM PEMROSESAN DATA**

**Oleh**

**Nama**

**NIM**

**(Program Studi Informatika)**

**A black background with a black square

Description automatically generated with medium confidence**

**UNIVERSITAS TRISAKTI  
November 2024**

**I. Pendahuluan**

Pada praktikum ini, mahasiswa akan mempelajari dasar-dasar pemrosesan data dan analisis eksploratif data (EDA) dalam konteks analisis sentimen. Analisis sentimen adalah proses untuk mengidentifikasi dan mengklasifikasikan emosi atau opini dari teks, seperti ulasan, komentar, atau post di media sosial. Analisis ini berguna untuk memahami sentimen publik terhadap suatu produk, layanan, atau topik tertentu.

Praktikum ini bertujuan untuk memberikan pemahaman mendalam tentang langkah-langkah yang harus dilakukan saat dihadapkan dengan dataset dalam konteks analisis sentimen, termasuk pembersihan data dan transformasi yang diperlukan. Adapun beberapa langkah yang akan dilakukan adalah sebagai berikut:

1. Mengimpor data teks yang diberikan ke dalam komputer.
2. Mengidentifikasi kolom mana yang merupakan fitur (seperti teks, waktu posting, dll.) dan mana yang menjadi target (sentimen positif, negatif, atau netral).
3. Melakukan pembersihan dan penanganan data jika diperlukan, seperti menghapus teks yang tidak relevan, menangani nilai yang hilang, mengelola data outlier, dan menyelesaikan ketidakkonsistenan data lainnya.

Melalui praktikum ini, diharapkan mahasiswa dapat memahami dasar-dasar pemrosesan data dalam analisis sentimen serta memiliki keterampilan dasar untuk melakukan EDA pada dataset sentimen.

**II. Hasil Praktikum**

1. Instalasi paket google-play-scraper untuk melakukan scrapping pada aplikasi Google Play Store.

|  |
| --- |

**Penjelasan:**

|  |
| --- |

2. Melakukan import library

|  |
| --- |

**Penjelasan:**

|  |
| --- |

3. Mengakses Google Drive di colab

|  |
| --- |

**Penjelasan:**

|  |
| --- |

4. Menyimpan data ulasan (reviews) ke dalam file CSV (Jika ada, jika tidak menggunakan maka lewati saja)

|  |
| --- |

**Penjelasan:**

|  |
| --- |

5. Membaca data dari file CSV

|  |
| --- |

**Penjelasan:**

|  |
| --- |

6. Menampilkan informasi dari dataframe.

|  |
| --- |

**Penjelasan:**

|  |
| --- |

7. Menghapus missing value

|  |
| --- |

**Penjelasan:**

|  |
| --- |

8. Menghapus duplikat dari dataframe

|  |
| --- |

**Penjelasan:**

|  |
| --- |

9. Membersihkan teks agar lebih rapi dan siap untuk diproses lebih lanjut

|  |
| --- |

**Penjelasan:**

|  |
| --- |

10. Mengubah semua karakter dalam teks menjadi huruf kecil

|  |
| --- |

**Penjelasan:**

|  |
| --- |

11. Memecah atau membagi string, teks menjadi daftar token

|  |
| --- |

**Penjelasan:**

|  |
| --- |

12. Menghapus stopwords dalam teks

|  |
| --- |

**Penjelasan:**

|  |
| --- |

13. Melakukan proses Stemming.

|  |
| --- |

**Penjelasan:**

|  |
| --- |

14. Mengubah daftar kata menjadi kalimat

|  |
| --- |

**Penjelasan:**

|  |
| --- |

15. Mengganti kata-kata slang atau tidak baku dalam teks dengan padanan kata yang lebih formal atau baku.

|  |
| --- |

**Penjelasan:**

|  |
| --- |

16. Membersihkan teks dan menyimpannya di kolom ‘text\_clean’

|  |
| --- |

**Penjelasan:**

|  |
| --- |

17. Mengubah huruf dalam teks menjadi huruf kecil dan menyimpannya di 'text\_casefoldingText’

|  |
| --- |

**Penjelasan:**

|  |
| --- |

18. Mengganti kata-kata slang dengan kata-kata standar dan menyimpannya di 'text\_slangwords'

|  |
| --- |

**Penjelasan:**

|  |
| --- |

19. Memecah teks menjadi token (kata-kata) dan menyimpannya di 'text\_tokenizingText'

|  |
| --- |

**Penjelasan:**

|  |
| --- |

20. Menghapus kata-kata stop (kata-kata umum) dan menyimpannya di 'text\_stopword'

|  |
| --- |

**Penjelasan:**

|  |
| --- |

21. Menggabungkan token-token menjadi kalimat dan menyimpannya di 'text\_akhir'

|  |
| --- |

**Penjelasan:**

|  |
| --- |

22. Menyimpan file ke dalam bentuk CSV setelah dilakukan preprocessing

|  |
| --- |

**Penjelasan:**

|  |
| --- |

23. Membaca data kembali setelah data dilakukan preprocessing.

|  |
| --- |

**Penjelasan:**

|  |
| --- |

24. Membaca data kamus kata-kata positif dari GitHub

|  |
| --- |

**Penjelasan:**

|  |
| --- |

25. Membaca data kamus kata-kata Negatif dari GitHub

|  |
| --- |

**Penjelasan:**

|  |
| --- |

26. Menentukan polaritas sentimen dari tweet

|  |
| --- |

**Penjelasan:**

|  |
| --- |

27. Melakukan analisis sentimen pada teks yang sudah dibersihkan dari stopword.

|  |
| --- |

**Penjelasan:**

|  |
| --- |

28. Membangun model word2Vec untuk representasi kata dalam analisis sentimen

|  |
| --- |

**Penjelasan:**

|  |
| --- |

29. Menampilkan shape dan data token

|  |
| --- |

**Penjelasan:**

|  |
| --- |

30. Ekstraksi fitur word2Vec dan pembagian data untuk klasifikasi sentimen

|  |
| --- |

**Penjelasan:**

|  |
| --- |

31. Menampilkan shape data fitur dan split train-test dalam klasifikasi sentimen

|  |
| --- |

**Penjelasan:**

|  |
| --- |

**Kesimpulan: (minimal 4 baris)**

|  |
| --- |