**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.

1. Instalasi paket google-play-scraper

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan import library

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Lalu, membuat sebuah dataframe dari hasil scraping tersebut.

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Menampilkan kolom dari variabel reviews

|  |
| --- |

Penjelasan :

|  |
| --- |

Melakukan sorting kembali terhadap dataframe dan menyimpannya dalam bentuk CSV

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan import kembali pada program kita dan juga menggunakan dataset dari ‘csv’ yang telah kita buat sebelumnya.

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan drop pada 3 kolom yang ada pada dataset.

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan import dan penginstalan modul/library “Sastrawi”

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan pengecekan pada dataset, apakah ada yang “NULL”, “NaN”, atau pun duplikasi pada dataset.

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan preprocessing seperti “Cleaning”, “Penghapusan Emoji”, “ReplaceTOM” dan “CaseFolding” pada dataset. Lalu, dibuat kolom baru untuk melihat hasil dari setiap preprocessing yang dilakukan.

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Menggunakan module nltk untuk mengunduh tokenisasi lalu menerapkan tokenisasi tersebut ke dalam dataset dan disimpan dalam kolom baru ‘Tokenizing’.

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan normalisasi dengan fungsi convertToSlangWord(), untuk mengubah setiap kata dari dataset yang memiliki singkatan. (Kamus :

[https://github.com/louisowen6/NLP\_bahasa\_resources/blob/master/combined\_slang \_words.txt](https://github.com/louisowen6/NLP_bahasa_resources/blob/master/combined_slang%20_words.txt)).

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan proses stopword, proses ini berguna untuk mengurangi jumlah kata dalam dataset yang kita miliki.

|  |
| --- |

Penjelasan :

|  |
| --- |

Menyimpan data ulasan dari hasil preprocessing menjadi Data Ulasan.csv

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan proses Stemming dengan mengambil hanya 1000 data

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan penginstallan module/library googletrans

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan import module/library, lalu membuat variabel baru yaitu “translator” dan “translations”. Lalu kita akan mengambil setiap nama kolom yang ada dan menambahkannya ke dalam variabel ‘translations’.

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Pelabelan data menggunakan VADER, lalu membuat sebuah “Compound\_Score” untuk memberikan nilai dari setiap kata yang ada. Nilai tersebut di dapat dengan menganalisa kata pada kolom “Stopword”.

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Membuat sebuah pelabelan berdasarkan dari hasil “Compound\_Score”. “Negatif” untuk nilai dibawah 0, “Positif” untuk nilai diatas 0, dan “Netral” untuk nilai sama dengan 0.

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Menampilkan diagram chart-pie untuk melihat jumlah setiap pelabelan

|  |
| --- |

Penjelasan :

|  |
| --- |

Menyimpan data hasil dari sentimen diatas dengan nama Hasil Sentiment.csv

|  |
| --- |

1. Melakukan import module “WordCloud” untuk melihat kata yang paling banyak digunakan pada pelabelan “Positif” dan “Negatif”.

|  |
| --- |

Penjelasan :

|  |
| --- |

**Latihan!**

Mengubah kolom stopword menjadi angka seperti tugas yang sudah diberikan dikelas!

|  |
| --- |

1. Wordcloud untuk “Positif”

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Wordcloud untuk hasil “Negatif”

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan ekstraksi fitur menggunakan Chi-Square dan LabelEncoder untuk mengubah setiap sentiment yang dimiliki menjadi integer.

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Menggunakan CountVectorizer() untuk melihat jumlah representasi teks yang ada pada dataset.

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Memisahkan setiap sentiment yang ada pada dataset menjadi variabel yang berbeda.

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Menghitung jumlah kata dari semua total\_positif, total\_netral, dan total\_positif yang ada pada dataset. Buat menjadi satu kolom baru “Grand\_total”, lalu lakukan sorting.

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Buat dataframe baru dengan ukuran yang sesuai dengan ‘valid\_indices’. Dataframe ini berisikan fitur yang telah diseleksi menggunakan chi-square.

|  |
| --- |

Penjelasan :

|  |
| --- |

**Kesimpulan : Minimal 5 baris**

|  |
| --- |