**WORD EMBEDDING DALAM NLP**

**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.
4. Memahami teori *word embedding*, seperti Word2Vec, GloVe, atau FastText, serta perbedaan dengan representasi tradisional seperti *one-hot encoding*.
5. Menggunakan pustaka NLP seperti Gensim atau *TensorFlow* untuk melatih model *word embedding* menggunakan dataset yang tersedia.
6. Mengeksplorasi representasi vektor dari kata-kata dalam dataset dan melihat hubungan semantik antara kata-kata (misalnya, menemukan kata-kata yang paling mirip dengan "bahagia").
7. Menggunakan teknik seperti PCA atau t-SNE untuk memvisualisasikan vektor kata dalam ruang dimensi rendah.
8. Menganalisis pola-pola semantik yang muncul dari visualisasi tersebut.

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. Melakukan import library

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Memuat Model Word2Vec

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Mengambil Representasi Vektor

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Menghitung tingkat kesamaan antara kata ‘science’ dengan kata lainnya

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Menghitung kesamaan antar dua frasa

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Menggunakan metode most\_similar dalam menemukan kata yang paling mirip dengan kata ‘Jelly’

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Memvisualisasikan hubungan sejenis antar kata menggunakan PCA

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Menggunakan metode most\_similar dalam melengkapi hubungan mirip dengan pola hubungan dari 3 kata yang diberikan

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Membaca data dan menampilkan data

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Menampilkan data histogram dari kolom review

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan cleaning pada data

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Menampilkan grafik countplot distribusi jumlah review berdasarkan rating

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan pemisahan data fitur dan labels

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan split data dan menampilkan hasilnya

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan proses teks dan membangun kosa kata menggunakan Keras Tokenizer

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Melakukan proses vectorize

|  |
| --- |

Penjelasan :

|  |
| --- |

1. Membuat Matriks Embedding dari Model Word2Vec dan menampilkan hasilnya

|  |
| --- |

Penjelasan :

|  |
| --- |

**Kesimpulan : Minimal 5 baris**

|  |
| --- |