



FAKULTAS
**ILMU
KOMPUTER**

CSCE604135 • Temu-Balik Informasi
Semester Gasal 2024/2025
Fakultas Ilmu Komputer, Universitas Indonesia

Tugas Pemrograman Kecil 3: Word Embedding
Tenggat Waktu: Kamis, 7 November 2024, 23.55 WIB

Ketentuan:

1. Anda diberikan sebuah Google Colab *notebook* berisi contoh dan soal yang harus Anda kerjakan.
2. Kumpulkan *file notebook* (IPYNB) yang sudah dilengkapi dengan format penamaan **TPK2_NPM.ipynb** melalui submisi SCell.
Contoh penamaan file: TPK3_2006524290.ipynb
3. Kumpulkan *file notebook* tersebut pada submisi yang telah disediakan di SCell sebelum **Kamis, 7 November 2024, 23.55 WIB**. Keterlambatan pengumpulan akan dikenakan penalti.
4. Tugas ini dirancang sebagai tugas mandiri. **Plagiarisme tidak diperkenankan dalam bentuk apapun.** Adapun kolaborasi berupa diskusi (tanpa menyalin maupun mengambil jawaban orang lain) dan memanfaatkan informasi dari literatur manapun masih diperbolehkan. **Pastikan** untuk mencantumkan nama kolaborator dan referensi literatur.
5. Anda boleh berkonsultasi terkait tugas ini asisten dosen berikut. Asisten dosen diperbolehkan membantu anda dengan memberikan petunjuk.
 - a. Muhammad Ilham Ghozali
Email: muhammad.ilham013@ui.ac.id
Discord: myticalcat

Petunjuk Pengerjaan Tugas

Akses *notebook* berikut:  Tugas Pemrograman Kecil 3

Dalam tugas pemrograman kecil ini, Anda akan mengeksplorasi dan mengimplementasikan beberapa metode untuk mendapatkan word embedding. Tugas ini terbagi menjadi dua bagian utama yang akan menguji pemahaman dan kemampuan implementasi Anda dalam teknik-teknik pemrosesan bahasa alami.

Bagian pertama berfokus pada Embedding with SVD & LSA. Anda akan diminta untuk mengimplementasikan Singular Value Decomposition (SVD) from scratch, hanya menggunakan numpy tanpa memanfaatkan fungsi SVD dari library luar. Selanjutnya, Anda akan menggunakan implementasi SVD tersebut untuk mengembangkan Latent Semantic Analysis (LSA). LSA ini akan digunakan untuk menganalisis hubungan antara dokumen dan term dalam korpus yang diberikan.

Bagian kedua dari tugas ini berfokus pada Neural Embeddings dengan Word2Vec, khususnya model Continuous Bag of Words (CBOW). Anda akan memulai dengan mempersiapkan data training untuk CBOW, yang melibatkan pembuatan pasangan input-output dari dokumen yang diberikan. Selanjutnya, Anda akan mengimplementasikan arsitektur neural network untuk CBOW. Akhirnya, Anda akan mengembangkan loop training lengkap untuk model CBOW, yang mencakup forward pass, perhitungan loss, backward pass, dan update weights.

Catatan Revisi:

-

Rubrik Penilaian

Komponen	Proporsi
Implementasi SVD	30%
Implementasi LSA	20%
Implementasi CBOW input-output pair	25%
Implementasi CBOW training loop	25%

Selamat mengerjakan!