

# Word Embeddings & Visualization

## Objective

The objective of this assignment is to **explore, compute, and visualize word embeddings** using classical techniques such as TF-IDF and GloVe. This is paired with an **interactive REST API and web application** that enables real-time exploration of word relationships.

## Tasks Completed

### 1. Word Embedding on Custom Corpus

- Developed a small custom text corpus.
- Applied **TF-IDF** vectorization and optionally integrated **pretrained GloVe embeddings**.
- Preprocessed text using standard NLP techniques: tokenization, stopword removal, lowercasing.

### 2. Visualization of Embeddings

- Used **Principal Component Analysis (PCA)** and **t-Distributed Stochastic Neighbor Embedding (t-SNE)** to reduce high-dimensional embeddings into 2D space.
- Generated scatter plots to visualize word proximity and clustering.

### Libraries Used:

scikit-learn, matplotlib, seaborn, numpy, pandas

### 3. REST API for Embedding Computation

- Created a RESTful API using **FastAPI**.
- API endpoint `/embedding?word=<input>` returns:
  - Embedding vector
  - List of top-N nearest neighbors

### 4. Interactive Web Application

- Built a user interface with **Streamlit**.
- Users can:
  - Input a word
  - View 2D embedding position on a scatter plot
  - See nearest neighbors dynamically

#### Features:

- Responsive layout
- Embedding plot updates on user input
- REST API integration for real-time feedback