# Cosine Similarity For plagiarism:

This code is a Flask web application that allows the user to upload two text files and calculates the similarity between them using cosine similarity. It then displays the similarity percentage and the most similar text.

**Libraries used:**

* Flask: a micro web framework for building web applications using Python.
* numpy: a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.
* pandas: a library for data manipulation and analysis.
* scikit-learn: a machine learning library for Python that provides tools for data analysis and modeling.
* pip install Flask
* pip install numpy
* pip install scikit-learn
* pip install pandas

**The code defines several functions**:

* create\_dataframe: creates a Pandas DataFrame from a matrix and a list of tokens.
* calculate\_similarity: takes in two texts, calculates their similarity using cosine similarity, and returns the most similar text and the similarity percentage.

The Flask application has one route:

* '/' (root): if the request method is GET, it renders the home page template. If the request method is POST, it reads the uploaded files, calculates their similarity using calculate\_similarity function, and renders the results page template.

The home page template (index.html) contains a form for uploading two files and a threshold value.

The results page template (results.html) displays the most similar text, the similarity percentage, and the threshold value.

To run this application, you need to have Python and the required libraries installed. You also need to have two text files to upload. When you run the script, the Flask development server starts and listens for HTTP requests on port 5000 by default. You can access the application by opening a web browser and navigating to <http://localhost:5000/>.