Team-210
Project Proposal

Problem
Plagiarism in programming assignments is hard to detect especially if the instructor has to manually grade the assignments for many students.

Objective
To develop a program that determines the percentage similarity between a set of uploaded **Python** source code files, with the purpose of finding the likelihood that plagiarism has occurred. The Program should be able to detect plagiarism even in the following scenarios:
- Variable names are changed.
- Function names are changed.
- Extracting code into functions.
- Moving code around in the file.

Program Features
- Choosing a role
  - Teaching Assistant
  - Professor
- Uploading files
- Running Plagiarism tests
- Displaying the files with similar code highlighted
- Providing Similarity percentage and graphs
- Downloading the report

Technology Stack

Backend	Java
Frontend	AngularJS, HTML, CSS

Algorithms
We have decided to use a number of algorithms and average the percentages found by each.
- Longest Common Subsequence
- Levenshtein Distance Formula
- Symantec sequence matching

Project Timeline
- **Week 1 -** Develop Use cases, Study about algorithms used in plagiarism detection.
- **Week 2 -** Create UML diagram, Implement Interfaces supporting the design.
- **Week 3 -** Build basic UI features, Create Abstract syntax tree, Implement LCS algorithm.
- **Week 4 -** Improve UI by adding text highlighting feature, Implement remaining algorithms.
- **Week 5 -** Finish Development, Start Functional testing.
- **Week 6 -** Start Structural testing, Resolve issues found if any.
- **Week 7 -** Code Reviews, Code Refactoring.
- **Week 8 -** Presentations.