**Project: Plagiarism Checker**

**Abstract:**

Plagiarism is defined as the representation of another person’s work as our own. In academic institutions, students are often eager to copy each other’s assignments or homework and submit it as their own. This leads to several negative consequences, such as decrease in the ability to create or generate one’s own work. Manual plagiarism detection has not been easy. To go through each file individually and evaluate whether it contains copied work from another file is very time consuming. Furthermore, text can be fabricated through the use of synonyms and still be copied. This use of synonyms can generate new text that has been initially copied. In order to overcome this problem, there can be a classifier that classifies files as containing plagiarism or not. The classifier computes a similarity score between a file and the rest of the corpus and then decides whether that contains text copied from another file. Any file containing a similarity score more than a given percentage for example 20% will be labelled as ‘contains plagiarism’. It also checks for synonym and identifies if the text in two files is same but the structure has been changed.