PROJECT REPORT

TITLE: PLAGIARISM DETECTION SYSTEM

Submitted by :Kummara Pallavi

Submitted to :Global Quest Technologies

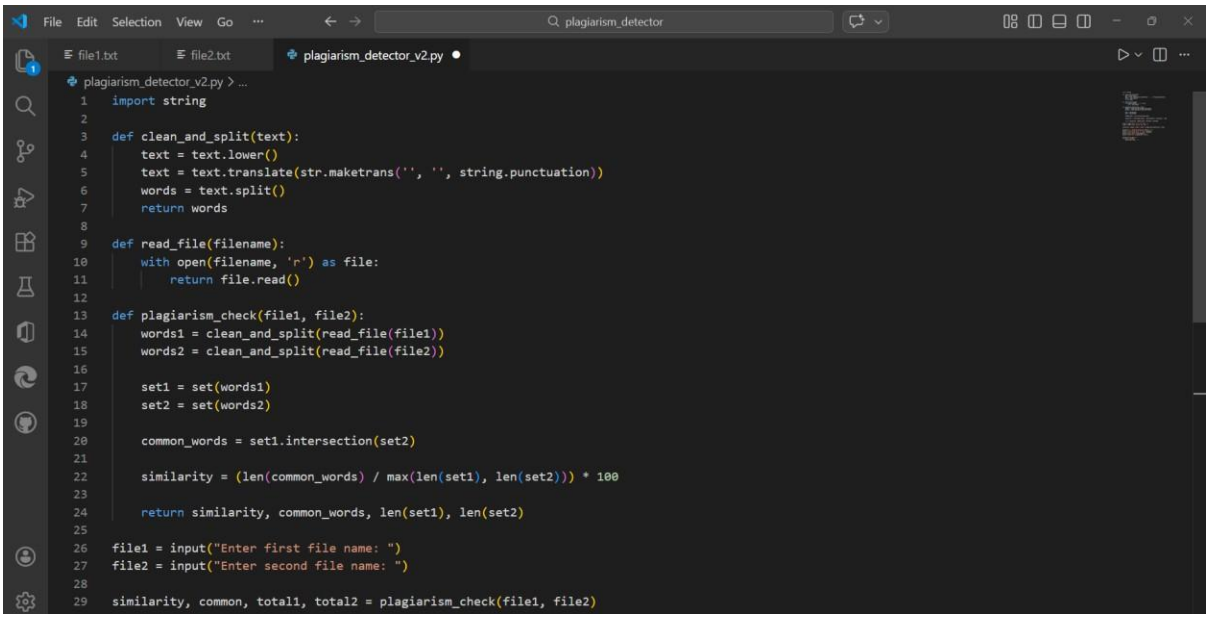
Date :January 2026

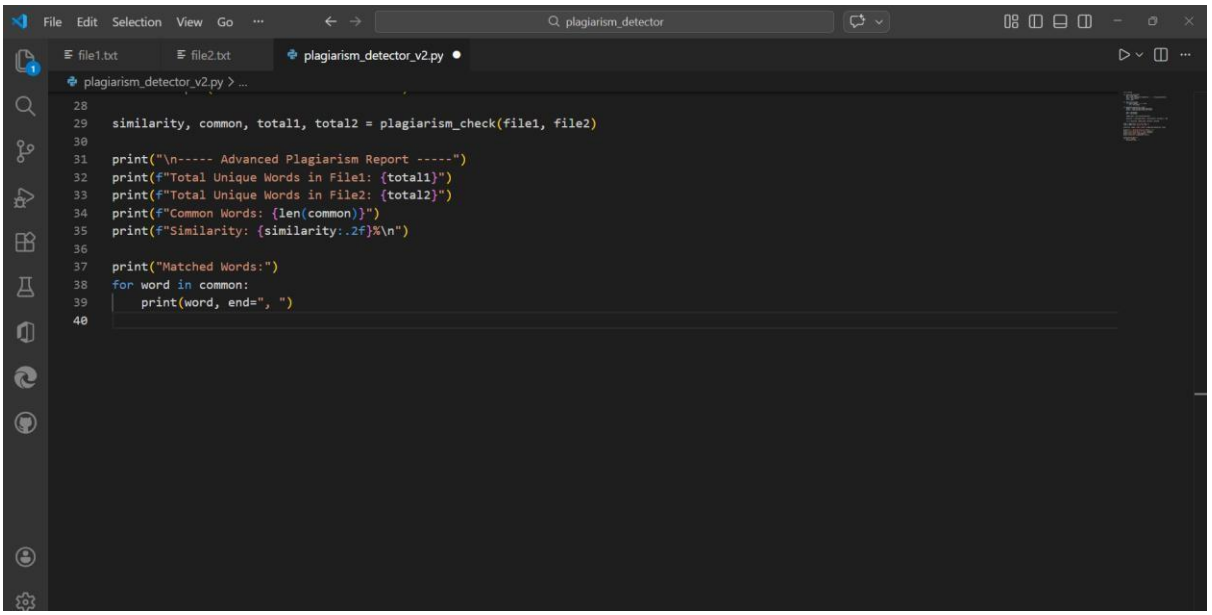
Abstract:

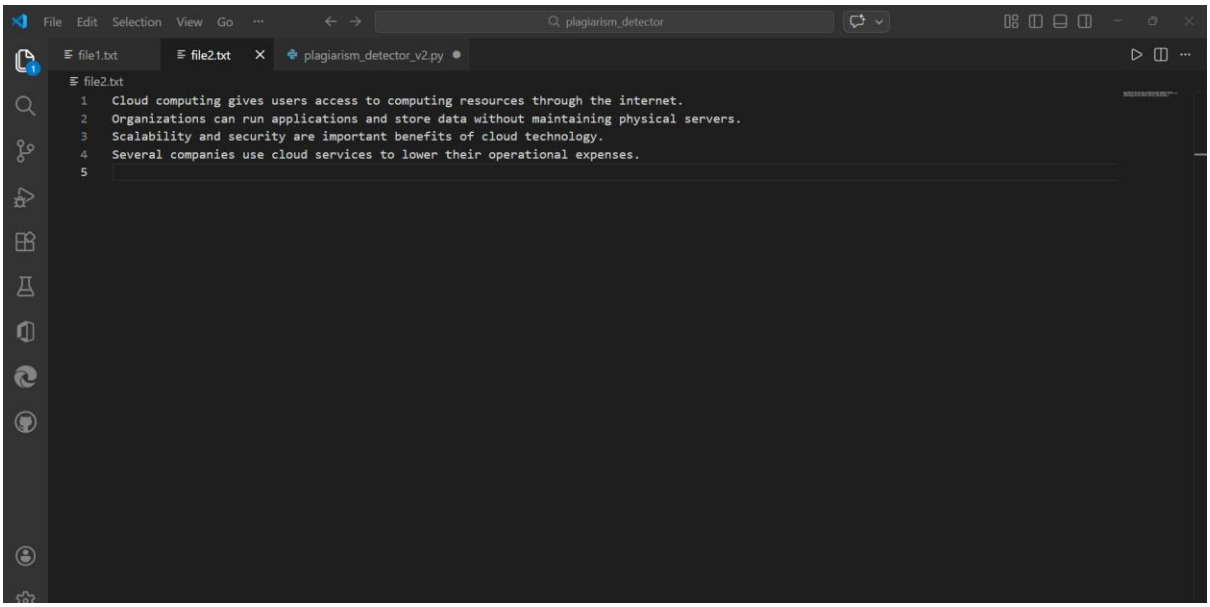
The Advanced Plagiarism Detection System is a console-based Python application designed to compare the content of two text files and determine the level of similarity between them. The system analyzes the text by converting it to lowercase, removing punctuation, and splitting it into words to ensure uniformity in comparison. It then identifies common words between the two documents and calculates a similarity percentage using the Jaccard similarity index.

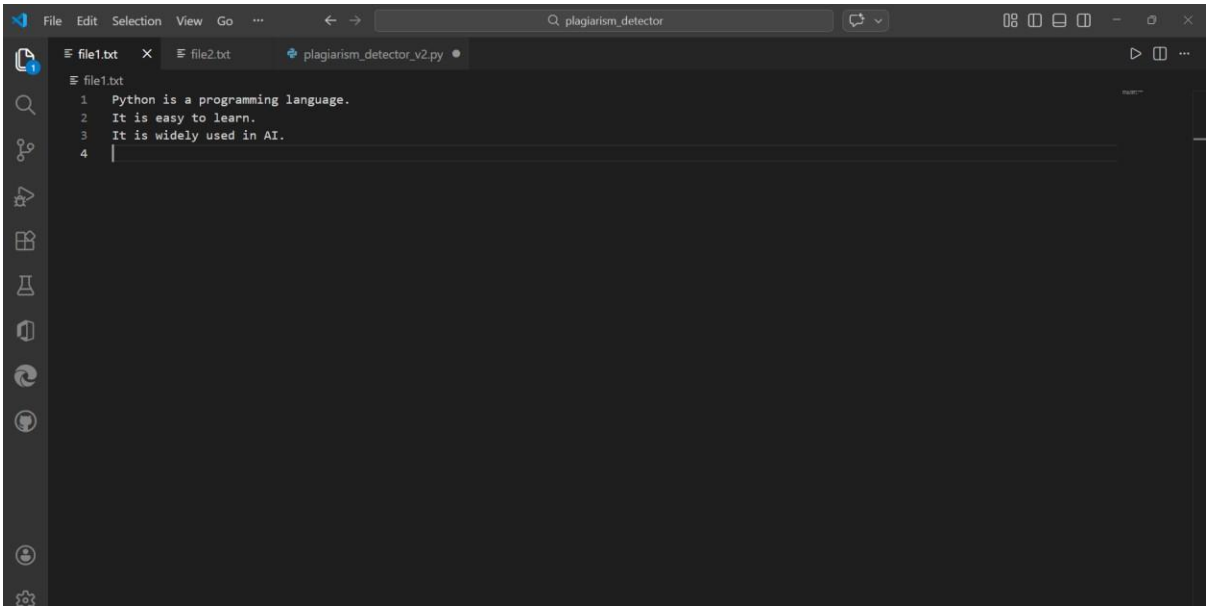
This project helps users detect potential plagiarism by providing details such as the total unique words in each file, the count of common words, and a list of matched words. It is particularly useful for students, educators, and researchers to check assignments, reports, or research papers for content overlap. The system is easy to use, requires no internet connection, and emphasizes the practical application of Python concepts such as file handling, sets, string manipulation, and user input.

Overall, the project demonstrates how programming can be applied to real-world problems like plagiarism detection in an efficient and reliable way.

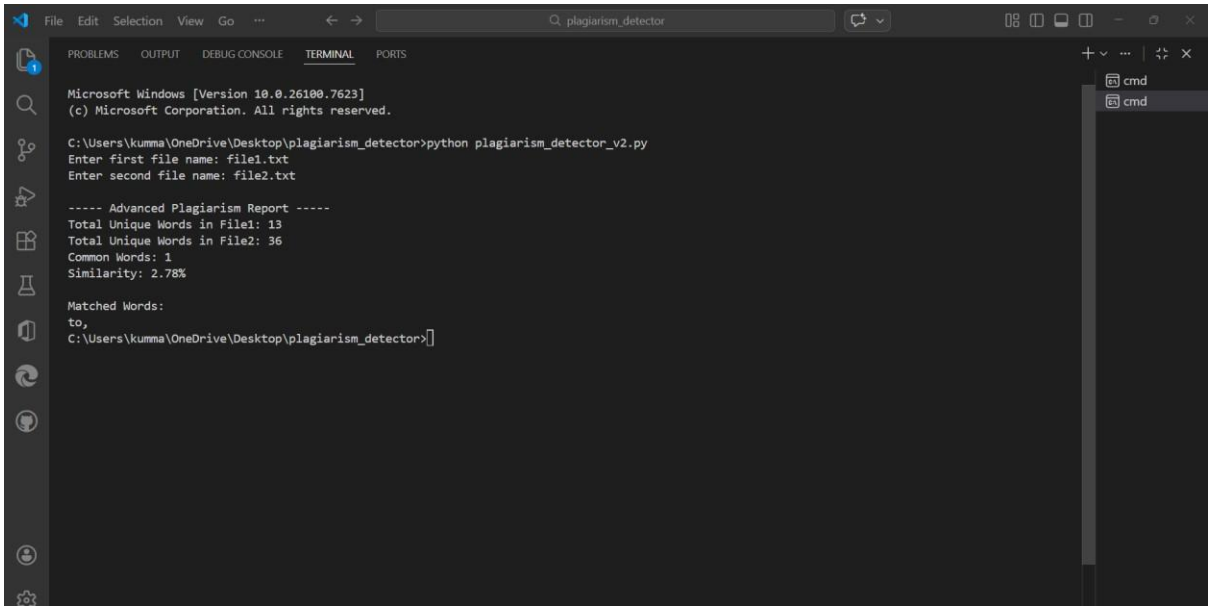

File Edit Selection View Go
plagiarism_detector
file1.txt
file2.txt
plagiarism_detector_v2.py
plagiarism_detector_v2.py > ...
import string
def clean_and_split(text):
text = text.lower()
text = text.translate(str.maketrans('', '', string.punctuation))
words = text.split()
return words
def read_file(filename):
with open(filename, 'r') as file:
return file.read()
def plagiarism_check(file1, file2):
words1 = clean_and_split(read_file(file1))
words2 = clean_and_split(read_file(file2))
set1 = set(words1)
set2 = set(words2)
common_words = set1.intersection(set2)
similarity = (len(common_words) / max(len(set1), len(set2))) * 100
return similarity, common_words, len(set1), len(set2)
file1 = input("Enter first file name: ")
file2 = input("Enter second file name: ")
similarity, common, total1, total2 = plagiarism_check(file1, file2)


File Edit Selection View Go
plagiarism_detector
file1.txt
file2.txt
plagiarism_detector_v2.py
plagiarism_detector_v2.py > ...
similarity, common, total1, total2 = plagiarism_check(file1, file2)
print("\n----- Advanced Plagiarism Report -----")
print(f"Total Unique Words in File1: {total1}")
print(f"Total Unique Words in File2: {total2}")
print(f"Common Words: {len(common)}")
print(f"Similarity: {similarity:.2f}%\n")
print("Matched Words:")
for word in common:
print(word, end=", ")


File Edit Selection View Go
plagiarism_detector
file1.txt
file2.txt
plagiarism_detector_v2.py
file2.txt
1 Cloud computing gives users access to computing resources through the internet.
2 Organizations can run applications and store data without maintaining physical servers.
3 Scalability and security are important benefits of cloud technology.
4 Several companies use cloud services to lower their operational expenses.
5


File Edit Selection View Go
plagiarism_detector
file1.txt
file2.txt
plagiarism_detector_v2.py
file1.txt
1 Python is a programming language.
2 It is easy to learn.
3 It is widely used in AI.
4

OUTPUT:


File Edit Selection View Go
plagiarism_detector
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
cmd
cmd
Microsoft Windows [Version 10.0.26100.7623]
(c) Microsoft Corporation. All rights reserved.
C:\Users\kumma\OneDrive\Desktop\plagiarism_detector>python plagiarism_detector_v2.py
Enter first file name: file1.txt
Enter second file name: file2.txt
----- Advanced Plagiarism Report -----
Total Unique Words in File1: 13
Total Unique Words in File2: 36
Common Words: 1
Similarity: 2.78%
Matched Words:
to,
C:\Users\kumma\OneDrive\Desktop\plagiarism_detector>

GLOBAL QUEST
TECHNOLOGIES

Global Quest Technologies

#324,2nd Floor, 3 A Cross, Near
Seshadripuram First Grade College,
Above City Union Bank,
Yelahanka New Town,
Bengaluru-560064

+91 9448 403 469 | 080-49720009
info@gqtech.in | www.gqtech.in