IET HACK AND CODE 5.0

TEAM NAME : THE STARKS

TEAM LEADER : AJAY R

PROBLEM STATEMENT : A MECHANISM IS REQUIRED THAT CAN ANALYSE

HANDWRITTEN TEXT AND THEN CONVERT IT INTO

THE EDITABLE TEXT.

TEAM MEMBERS : DEVA PRASATH R

ROHIN S

DHANUSH KARTHIK K R

COLLEGE CODE : 1-3516209872



- ➤ Handwritten text is unique for everyone. The handwritten text for doctors and judges are very complex and not easily understandable.
- > The handwritten text document is scanned and it is should be converted into an editable document, converted into other regional language of users choice.

DEPENDENCIES

- > Optical character recognition (OCR) lets you turn scanned images into text so you can turn paper-based documents into editable, searchable, digital documents.
- > OCR analyses the patterns of light and dark that make up the letters and numbers to turn the scanned image into text.
- > The OCR we are using is Pytesseract which gives accurate results of handwritten text from images.



TECHNOLOGY STACK

Process

Grid Flow

Chart

- > First the text is extracted from the image using pytesseract OCR.
- ➤ Now the extracted text undergoes feature extraction and segmentation.
- > The recognized text will be displayed in the editable document format.
- > The content which is in the document. can be changed into other regional languages.

01

Pre-proprocessing

Extract the document text from the image without performing some kind of

05 **Publish the** Outcome

The extracted text is displayed in an editable document format and it can be changed into other regional languages.

04

Classification

02 Segmentation

being used is the key factor in deciding the accuracy of OCR system.

In feature extraction stage each character is represented as a feature vector, which becomes its identity. The major goal of feature extraction is to extract a set of features, which maximizes the recognition rate with the least amount of elements.

03

Feature Extraction