

# **Artificial Intelligence and Data Science**

**Semester: 6**

**Course Name: Mini project (ADD334)**

**Topic :** Converting marks in answer sheet to excel(CSV)

## **Abstract :**

This mini project involves developing a software tool that utilizes Optical Character Recognition (OCR) technology to convert answer paper images, especially the marks that are written in a box into Excel sheets.

The software tool will take input in the form of scanned images(pdf) of answer papers, and split the pdf to separate images. Predefined zones are selected & checked for tables and passed through an OCR system like Tesseract or Paddle to extract the marks. The extracted text will be organized and formatted using frameworks like numpy & pandas and finally to Excel sheet.

The resulting Excel sheets will contain the original content of the answer paper, allowing users to easily analyze and manipulate the data. The project will aim to optimize the accuracy and efficiency of the OCR process, as well as provide customization options for users to tailor the output to their specific needs. The project is expected to have applications in educational institutions, where it can be used to streamline the process of grading exams and generating reports.

**Batch Number : 7**

## **Team Members :**

1. Ajay T Shaju - SJC20AD004
2. Emil Saj Abraham - SJC20AD028
3. Justin Thomas Jo - SJC20AD046
4. Vishnuprasad K G - SJC20AD063

## **Guide :**

Dr. Deepa V  
HOD, AD Dept