**1. PROJECT TITLE – SMART OCR FOR DOCUMENT DIGITIZATION**

**2. INTRODUCTION**

With the advent of OCR techniques, much time has been saved by automatically extracting the text out of a digital image of any invoice or a document. Currently, this is where most organisations that use OCR for any form of automation are.

Digital copies of invoices or documents are obtained by scanning or taking pictures. The text is extracted from these documents using OCR. Relevant data is extracted and the irrelevant data is discarded from the OCR results. The data is entered into a template based data entry software. The data entered is put through manual review to correct errors.

**2.1 Overview**

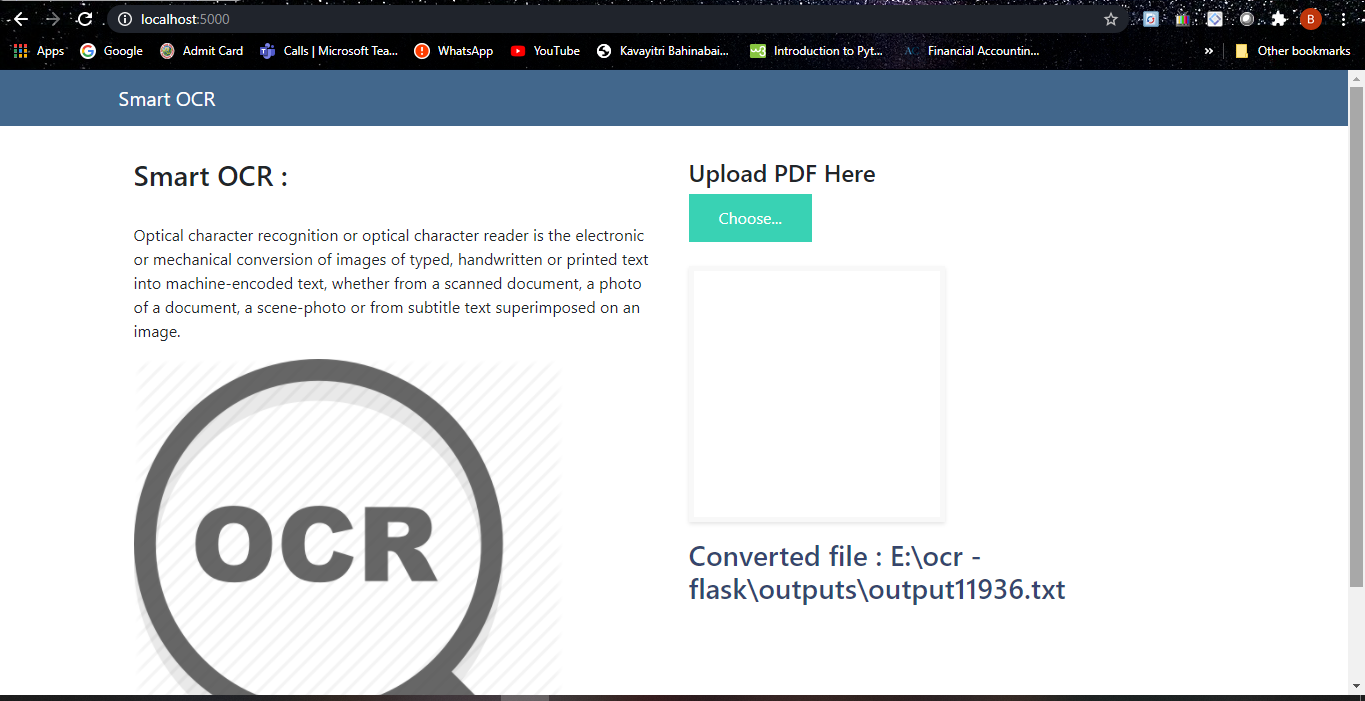
This templates used are unique to each use case, organisation and mostly for each different kind of document. While the OCR process helps in digitization, **it doesn't solve many tedious parts due to the unstructured results of OCR.**

**2.2 Purpose**

Deep learning enabled OCR technology allows you to digitize

1. Forms - legal forms, government procedures, tax fillings, etc.
2. ID cards - driver’s license, passport, aadhar card, etc.
3. Legal documents - affidavits, tickets, bonds, etc.
4. Bank statements - passbooks, account statements, cheques, etc.
5. KYC information - ID cards, address proof.

**3. RESULT**

**3.1 Screenshots of output **

* **APPLICATIONS**
  + - * 1. Digitization of information can help your organisation move towards a paperless workflow. It can help your organisation enable quicker and more convenient processes, enhance customer experience, increase employee satisfaction and reduce costs. It can help you drive better compliance practices in your company while also providing better customer service and increasing the transparency in your organisation.
* **CONCLUSION**
  + - 1. With the advent of OCR techniques, much time has been saved by automatically extracting the text out of a digital image of any invoice or a document. Currently, this is where most organisations that use OCR for any form of automation are.
* **FUTURE SCOPE**

Digital copies of invoices or documents are obtained by scanning or taking pictures. The text is extracted from these documents using OCR. Relevant data is extracted and the irrelevant data is discarded from the OCR results. The data is entered into a template based data entry software. The data entered is put through manual review to correct errors. Forms - legal forms, government procedures, tax fillings, etc.