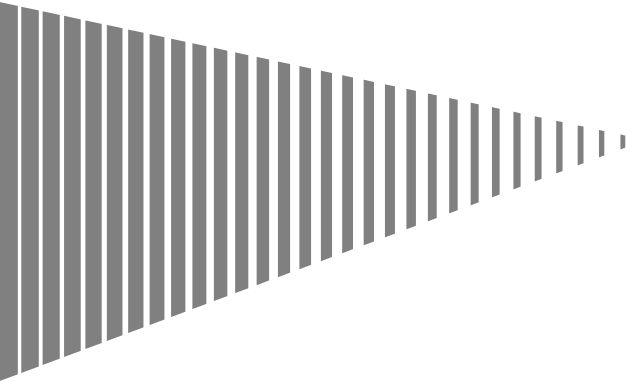


# Document Digitization of Indian PAN card

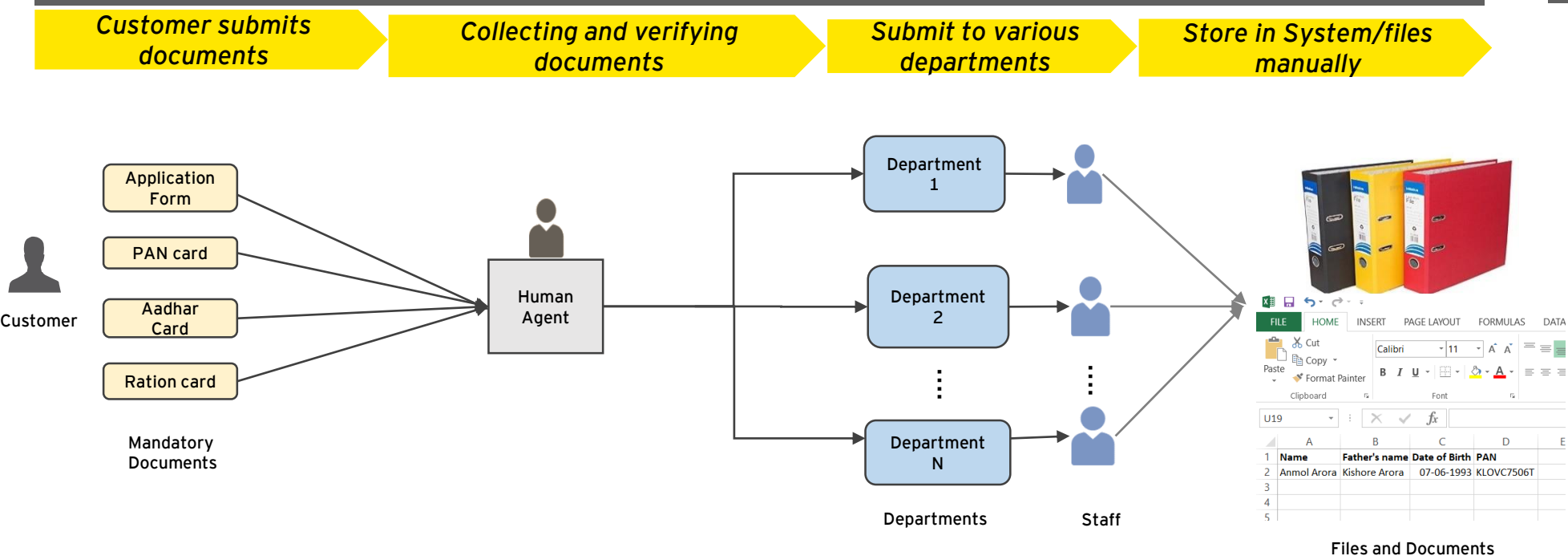
## Business Case



# Problem Statement

Process of extracting data from PAN card and storing it manually in files is very time consuming task which requires lot of human effort.

## Current Process / Solution



## Challenges

Misplacement

Disarrangement

Searching

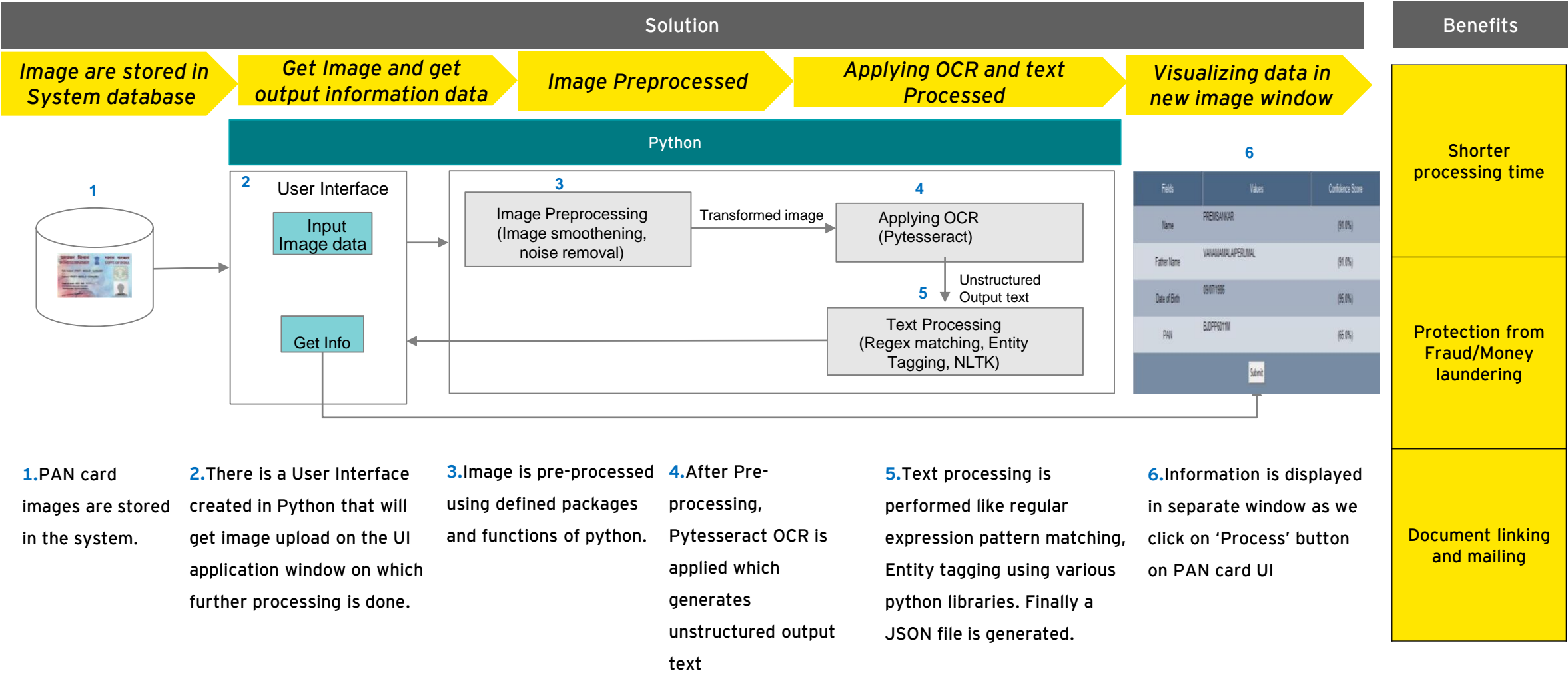
Segregation

Collaboration

Document security problems

# KYC of PAN card Approach

Objective: To develop and run an effective KYC application on Indian PAN card images, in order to extract data for in-person verification with least human intervention.



1.PAN card images are stored in the system.

2.There is a User Interface created in Python that will get image upload on the UI application window on which further processing is done.

3.Image is pre-processed using defined packages and functions of python.

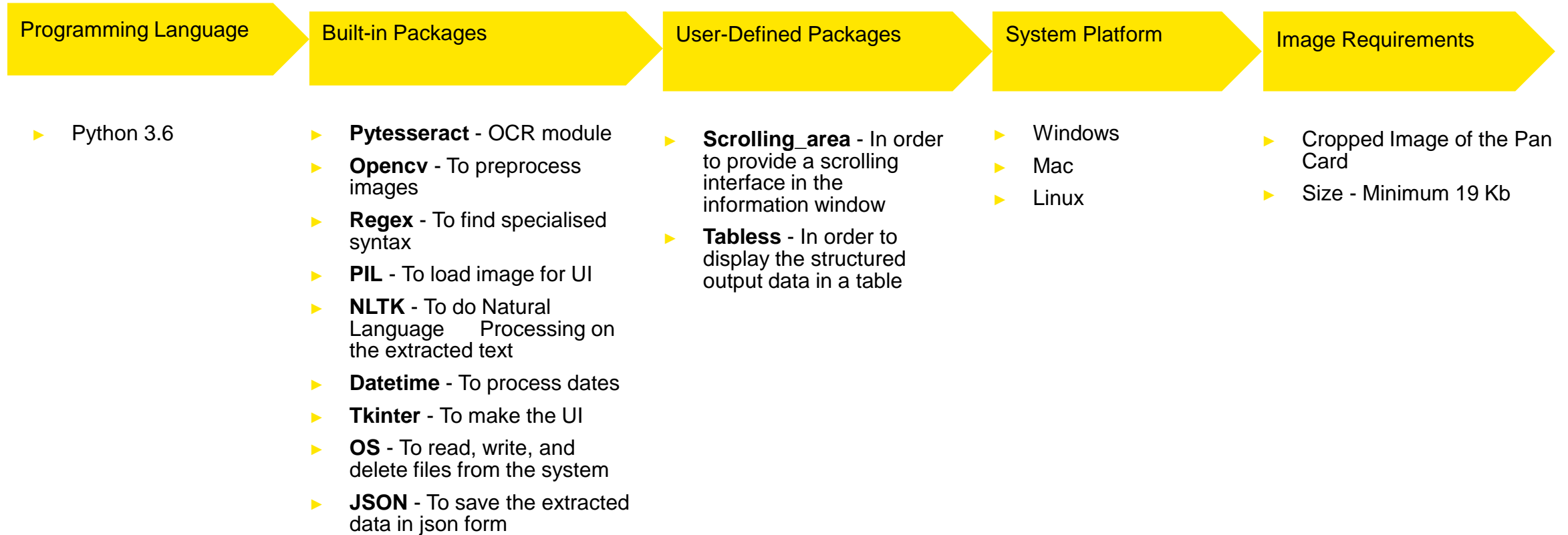
4.After Pre-processing, Pytesseract OCR is applied which generates unstructured output text

5.Text processing is performed like regular expression pattern matching, Entity tagging using various python libraries. Finally a JSON file is generated.

6.Information is displayed in separate window as we click on 'Process' button on PAN card UI

# Prerequisites :

---



# Application Execution:

Step 1:

PAN Card Application Interface Window

Grid Manager

No Image Selected

Open New Image

Reset

Process

Fields	Values	Confidence Score
	NaN	
	NaN	
	NaN	

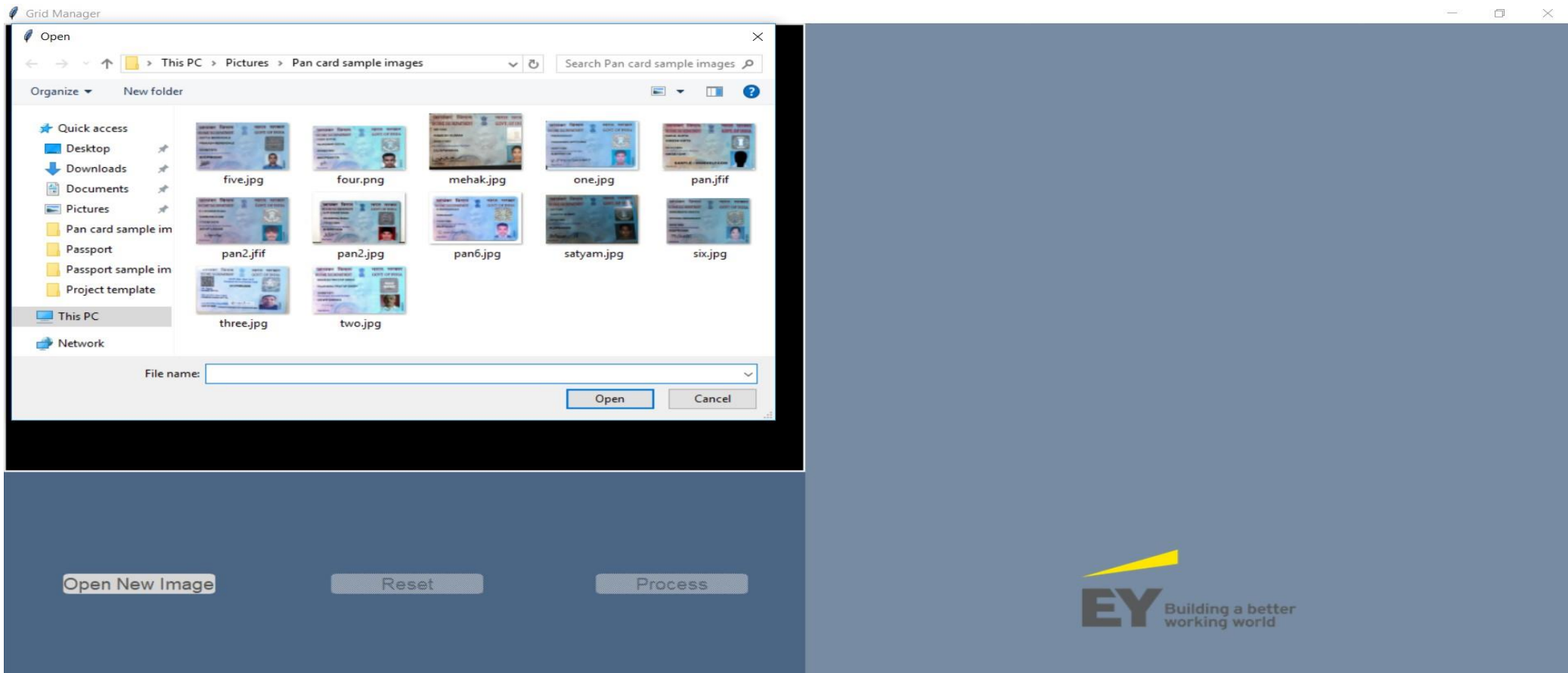
EY

Building a better working world

# Application Execution:

**Step 2:**

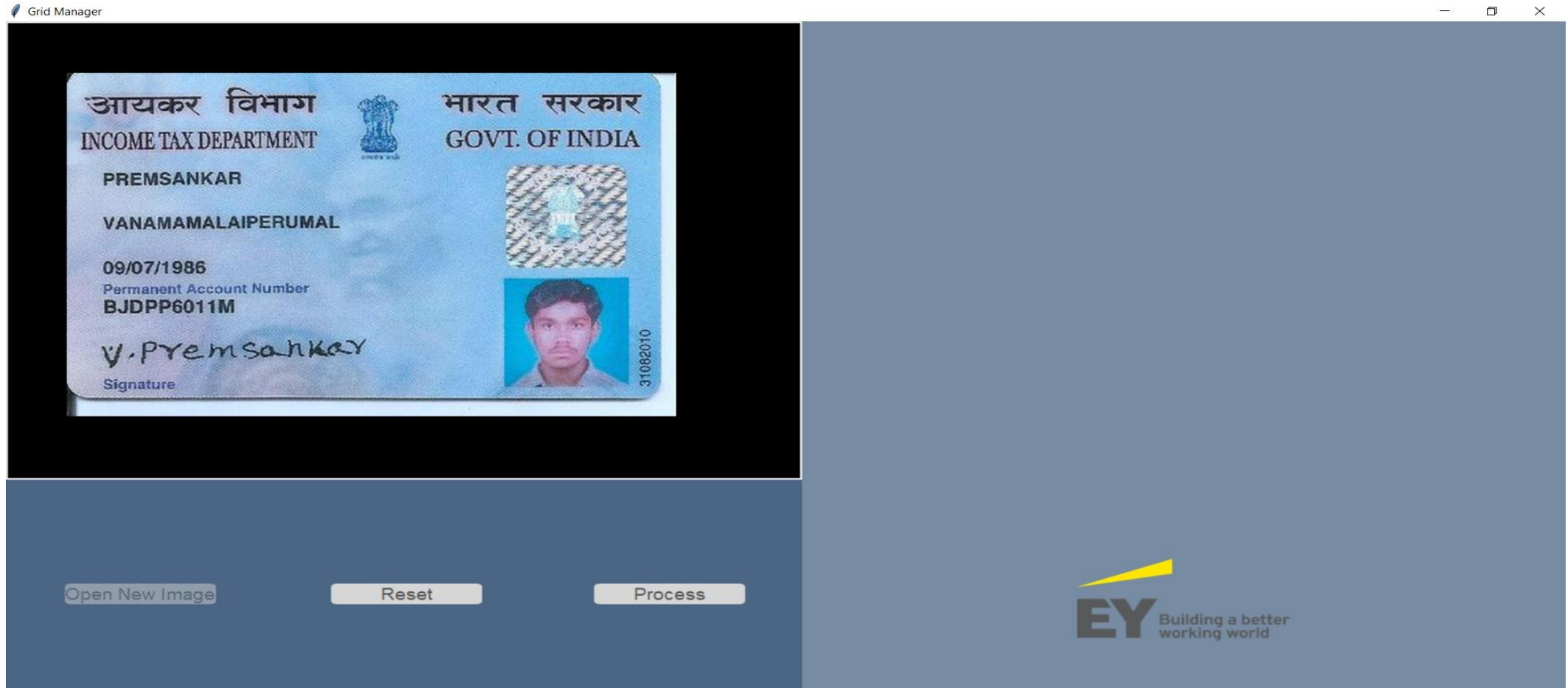
Browse image and open it on Application window



# Application Execution:

Step 3 :

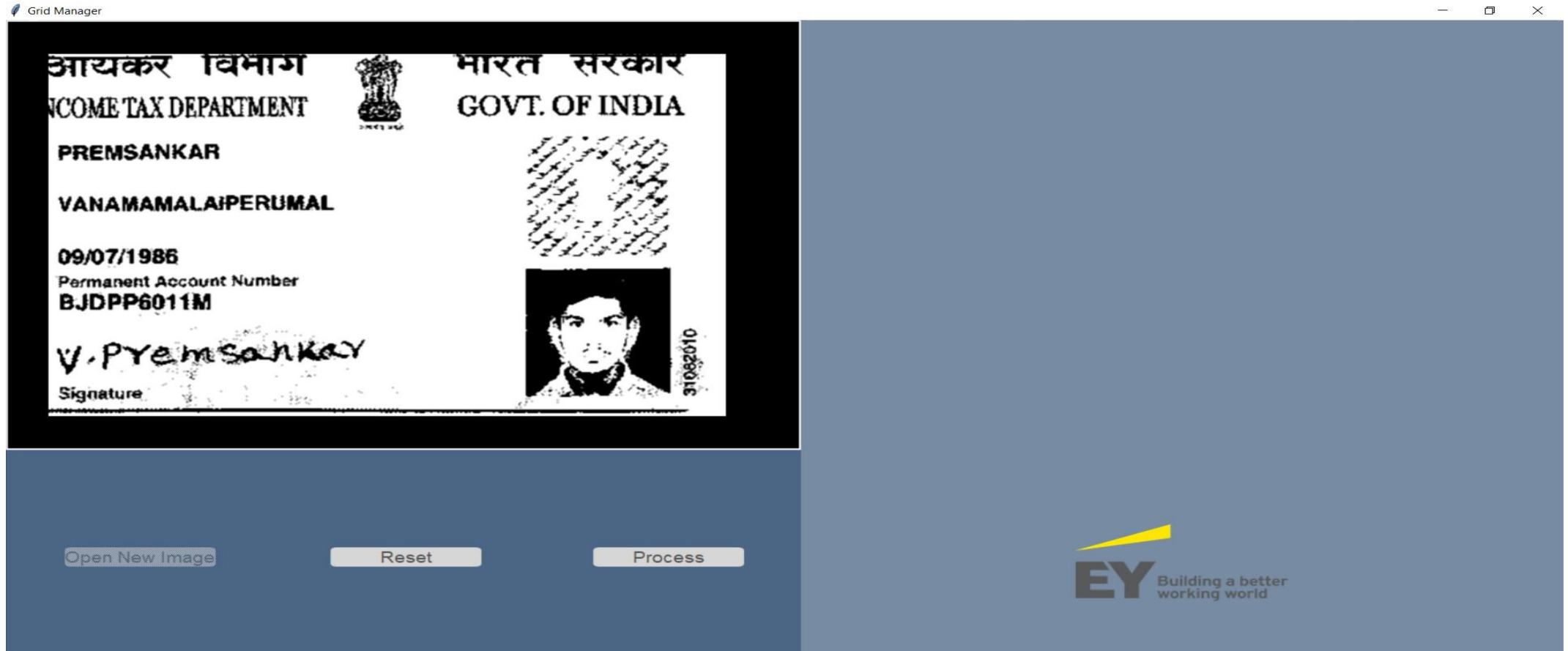
Image opens in Image viewer



# Application Execution:

Step 4 :

Grayscale Image opens in Image viewer when process button is clicked





# Application Execution:

Step 5 : PAN card Information is displayed on new window with confidence score

Grid Manager

आयकर विभाग  
INCOME TAX DEPARTMENT

भारत सरकार  
GOVT. OF INDIA

PREMSANKAR

VANAMAMALAIPERUMAL

09/07/1986  
Permanent Account Number  
BJDPP6011M

V. Prem Sankar  
Signature

31082010

Open New Image

Reset

Process

Fields	Values	Confidence Score
Name	PREMSANKAR	(91.0%)
Father Name	VANAMAMALAIPERUMAL	(91.0%)
Date of Birth	09/07/1986	(95.0%)
PAN	BJDPP6011M	(65.0%)

Submit

EY Building a better working world

# Appendix A : Python Script for PAN card information extraction

---



UI2.py

## Appendix B : PAN card KYC media file

---



Pandemo.mp4