

Digital Transformation of a Realty Company Looking to Optimize Data Collection

Introduction

The client is a realty/housing domain company. When an agent organizes an open house event for a property, interested buyers drop their visiting cards into a glass jar. The data from these cards is digitized manually into the lead database.

Need

Client was looking for better and faster approach for digitizing the data as the current approach meant, submitting the card jar to a data entry operator who would take sometimes a full day to capture all data from cards. Accuracy was also a problem as error would creep in while typing out the information from the card.

Solution

We used AI based smart optical character recognition algorithm which can recognize characters from an image or a hand written note. Given a contact card this can detect character on the card and save the number in the contact list against the name on the card. Along with extraction, the algorithm was able to validate email and phone numbers against an external API.

A pre-trained Azure OCR model from Azure cognitive services module was used . Model can handle both handwritten and different fonts characters as well as detect and identify characters even from a bad quality image.

All card jars submitted to data-entry operator are clicked via mobile camera and uploaded to blob storage on Azure cloud.

This is followed by pre-processing of the image. This step involves image quality enhancing techniques. We use thresholding and later apply a small Gaussian blur to the threshold image followed by use of erosion and dilation methods to remove and an obstructing lines and any background noise. Rotation and Translation errors are fixed and the characters are placed 0 degrees to the paper plane and at the centre of the image.

This is sent into the Azure OCR API as a batch to get characters from the image like person name, phone number and email and submitted to CRM API after validation.

Advantage

99.9% accuracy

Compared to data entry which had typos which were difficult to spot and correct. Data capture with the automated solution resulted in very high accuracy. And the captured data was validated with 3rd party APIs.

No data entry

Solution component is fully dynamic and efficient it requires no monitoring and eliminated the need for data entry operator.

Multi-language

Support is present. Detecting text from many different languages and capturing data correctly was an additional plus point over data entry operators.

Industry

Realty

Digitization

Our Services

Data Warehouse

Data Orchestration

File Transfer Services

OCR

Technology Stack

Azure Data Factory

Azure Cognitive Services

Azure Machine Learning

Azure Blob Storage