

To grow your business contact us now:
info@canspirit.ai

Base OCR -> Text based detection from Images

Client Profile

Client is a legal firm. The Company has a rapidly growing customer base and offers many services to its customers. The client's requirement revolves around gathering data from customer's unique identification card i.e., Aadhaar card, driver's license, pan card etc.

The Business Challenge

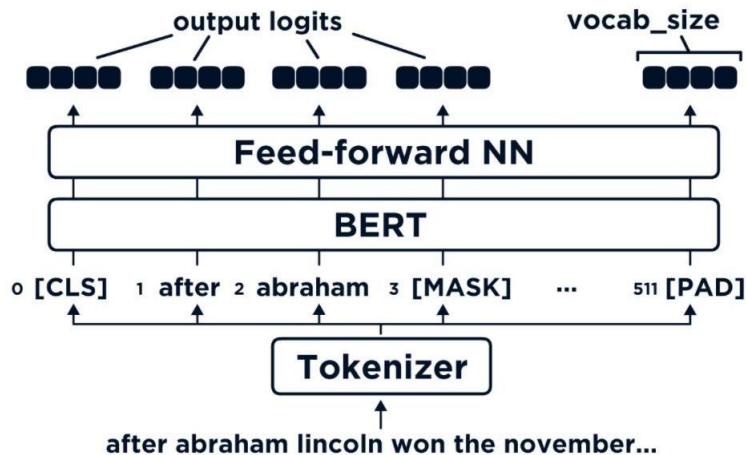
The client must gather all the data manually which consumes a lot of time and effort. To tackle this situation, detecting essential information such as Name of client, address, contact details, unique identification number authorized by government, a Base OCR equipment is required. Base OCR technique will help to automate the system and reduce time and effort of the client and improve efficiency.

Canspirit Predictive Analytics Approach

We proposed a model which augments the data and thereby extracts textual essential content from an image passed using NLP-BERT. This Bert technique requires huge amount of data to be trained upon, thereby augmenting a single training file and testing it. This technique now automates the data-entry process which in turn leverages analysis of the client's data.

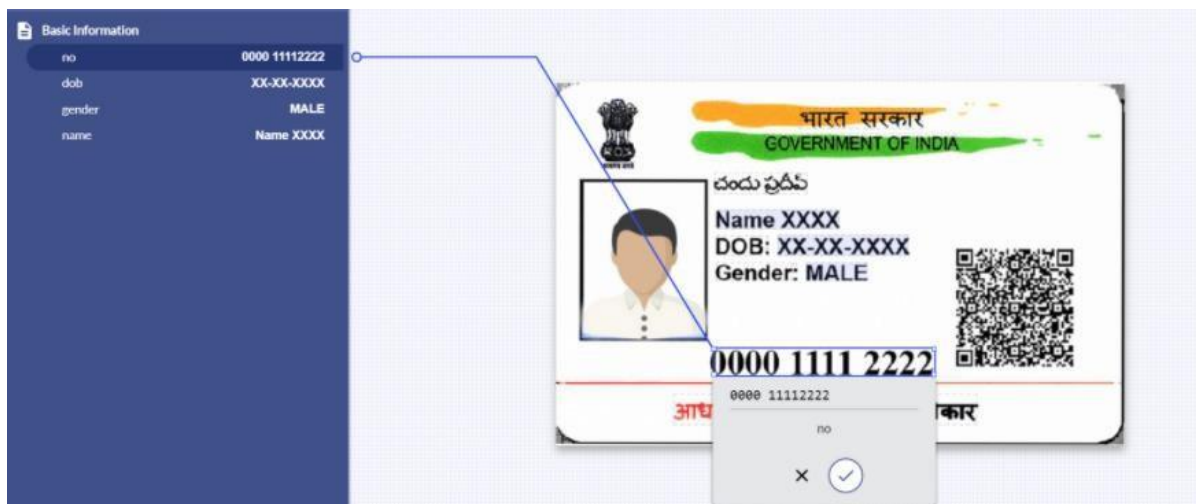
After analyzing the business problem, our focus was on two model techniques to extract textual data from image:

- NLP-Bert entity extraction
- Tensorflow-2.0 implementation



Process flow of NLP-Bert methodology

A sample data of extracting essential text from Aadhaar card images



Necessary data gets extracted from image passed to our model.

Technologies and Tools

- Python {Natural Language Processing – BERT entity extraction method}.
- Ubi.ai tool to augment data
- Tensorflow-2.0, Keras-api to extract textual data
- Python libraries – Pillow to read images, tesseract, EasyOCR.
- Poppler – To convert pdf to image.
- Python Framework – Flask {For deployment}.

Challenges and Solutions

- The customer aimed at implementing Natural Language Processing based technique to extract data and hence to gather humongous data for training and generating the model file is the prominent challenge been faced.
- Later obstacle to be tackled was to eliminate unwanted text if any detected by the model.

Business Benefits

- The company now saved much time of entering client's information with this prominent approach of gathering information.
- The Aadhaar number providing unique identity of an individual stops duplicate entries of the same customer and maintains integrity of the system.
- This automated system now is time-efficient and error-free, client would not report for change in information.
- **{Resources being saved, profiting the market.} This technique being customer-friendly and hassle-free has great __.**

Bottom Line

Canspirit leveraged its Machine learning and Artificial Intelligence skills to build a robust BERT-entity extraction model to draw textual data from images. **This model has extensive functionalities to extract medical records of the client too.**

For a discussion on how we can help you grow your business, email us now at:
info@canspirit.ai