INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

Natural Language Processing Based Doc Q&A System

Rishabh Dwivedi, Varsha, Prankit Sahu, Prachi Sahu, Sanchita Chourawar

Department of Computer Science and Engineering, Government Engineering College, Koni, Bilaspur,

Chhattisgarh, India

ABSTRACT: Doc Q&A is a web application developed in Python that utilises natural language processing and computer vision to extract information from images uploaded by the user in any format. It can identify and decode text from distorted images or those with low quality, making it easier for users to find answers to their questions. With Doc Q&A, users can effortlessly obtain accurate information from images without the need for manual transcription or data entry.

In addition to our image analysis capabilities, our project also includes a feature to extract text from hotel bills and other similar documents, and present the data in a JSON (Javascript Object Notation) format. This allows for the easy integration of data into other models and systems, enabling real-time use and analysis of the extracted data. Our team has worked diligently to ensure that the extraction process is both accurate and efficient, providing a useful tool for a range of applications.

KEYWORDS: Natural Language Processing, Computer Vision, JSON and Python.

I. INTRODUCTION

Doc Q&A uses state of the art natural language processing and computer vision for image encoding and text decoding process. It is a python based webApp built with Gradio library for providing a smooth user interface for processing requests in real time. Users have the privilege to upload any kind of image and a question, based upon that our model returns respective output along with the accuracy score for better clarity. We have trained our model on hugging faces dataset, hence accuracy is really good. We can also analyse blurry images or distorted images in our models.

Our model returns answers in JSON (Javascript Object Notation) format, which gives users a privilege to use that data in real time in any of his services.

II. LITERATURE REVIEW

What if you have a distorted and blurry image which contains very important information, but you are not able to identify the information because the image is not clear to understand. In such situations, Doc Q&A helps you to identify the information which is available in the image, all you need to do is to upload an image and a set of questions for which you want answers.

The models which are used are trained on hugging faces dataset and the literature survey is done on various research papers which were available in similar domains.

Going through various research papers helps us identify this use case and encourages us to build respective models.

III. ARCHITECTURE

Doc Q&A is a natural language processing based webApp which is based on the core concepts of natural language processing and computer vision. The main purpose of our project is to make the life of peoples easier by saving their valuable time and money.

The frontend of the webApp is built with Gradio Python which is mostly used to provide user interface for machine learning and NLP models. The backend of the webApp is written in pure python with hugging faces models like

LayoutLM and **Donut Cord.**

The core libraries which are used behind our model are LayoutLM and Donut Cord models.

The reason behind choosing such library are:-
- LayoutLM models are highly accurate in recognizing text and layout information from scanned documents, handwritten documents, and other unstructured data. This is because they are specifically designed to leverage both visual and textual information in the document, which leads to better accuracy
- LayoutLM models are highly adaptable to different languages and document types. They can be trained on a wide range of document layouts and fonts, making them suitable for a wide range of document analysis tasks
- LayoutLM models are highly efficient in processing large volumes of documents quickly and accurately.
- Donut (Document understanding transformer) models are highly efficient in understanding document images fastly and freely.
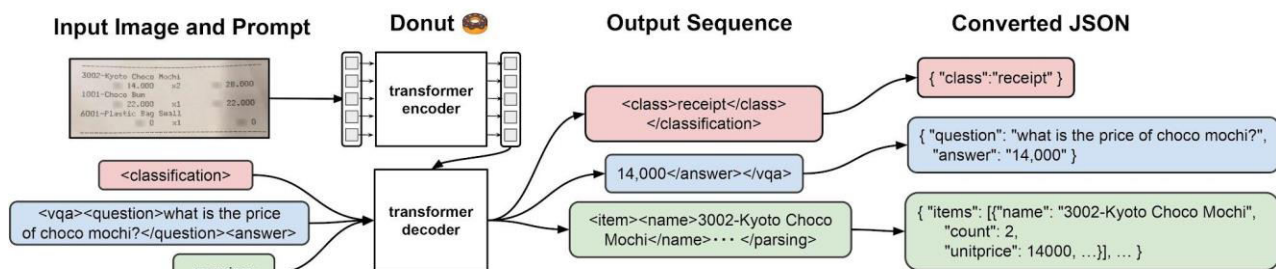

Fig1: Donut Cord

The image and prompt are simultaneously passed in the webApp, the image gets encoded and segment into smaller pieces, simultaneously the prompt are decoded. The process is gone through an output sequence. Once the data matches, the respective output is converted into JSON data, which helps users to use it in realtime.
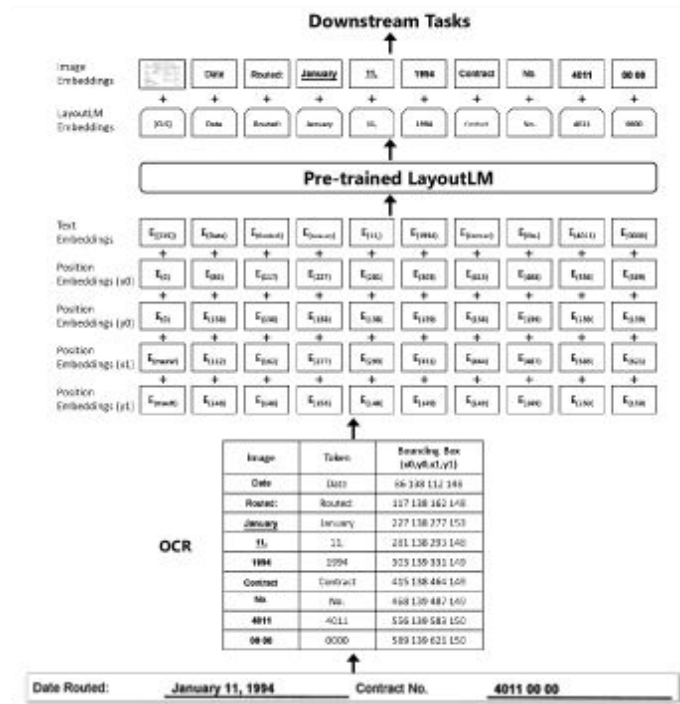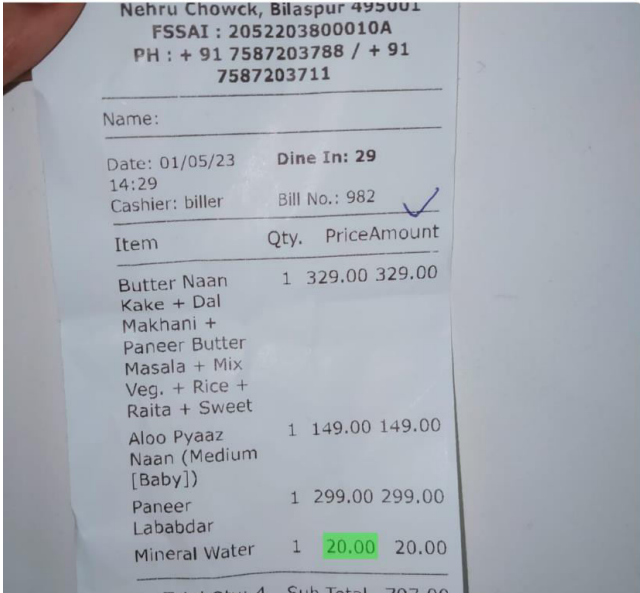
Fig2: LayoutLM

The webApp is deployed on Hugging Faces space and can be used by anyone for free.

IV. RESULTS- SCREENSHOTS

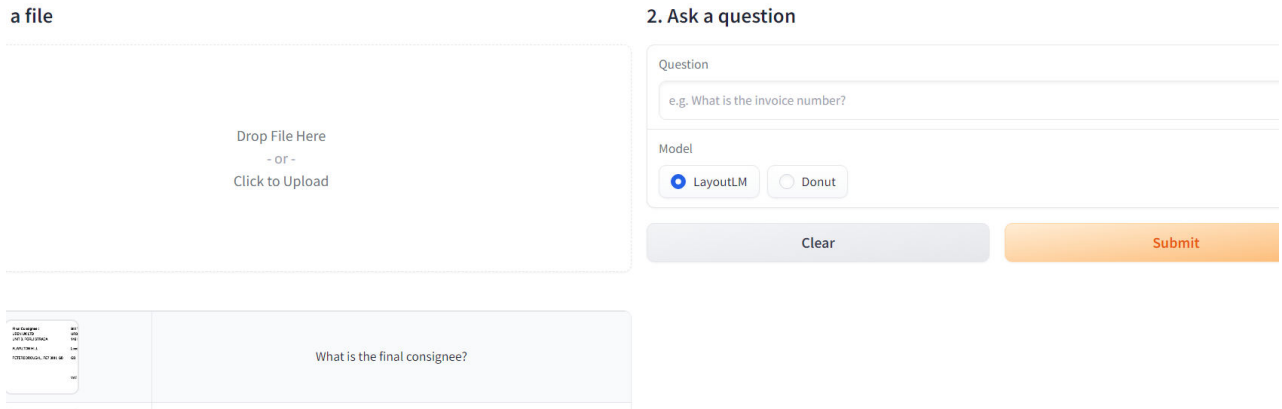

Fig3: Doc Q&A Example-1

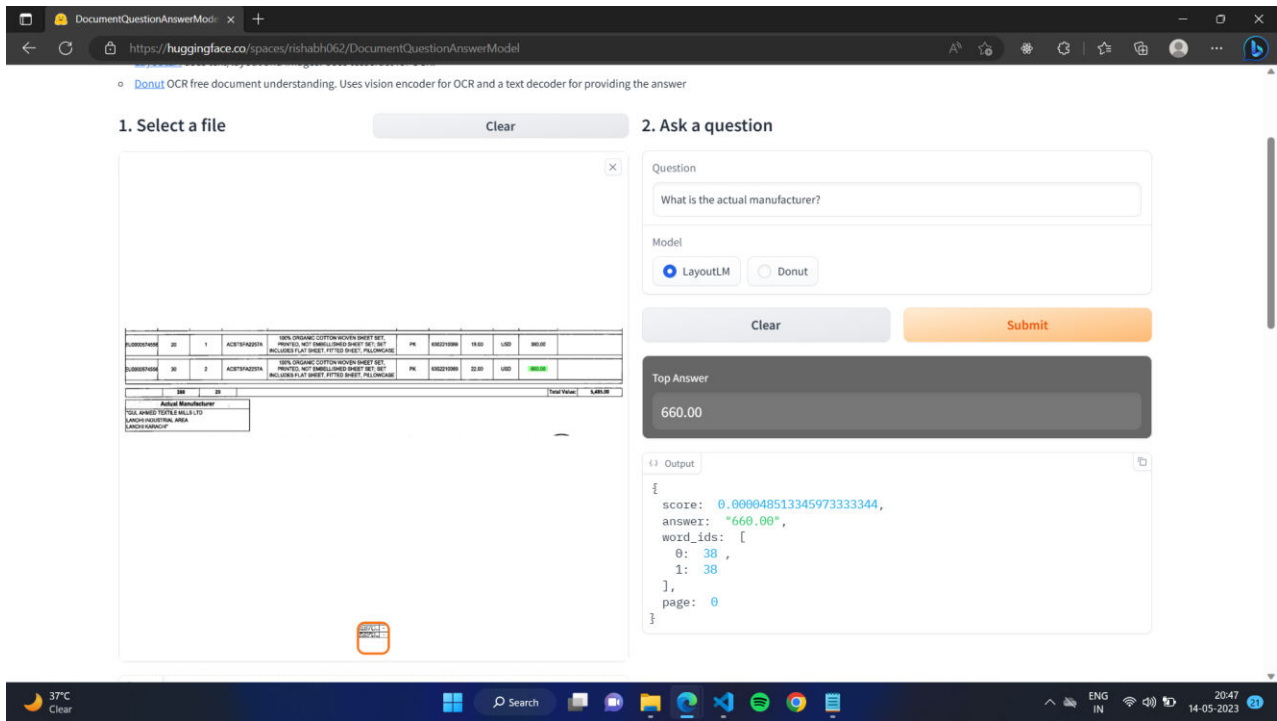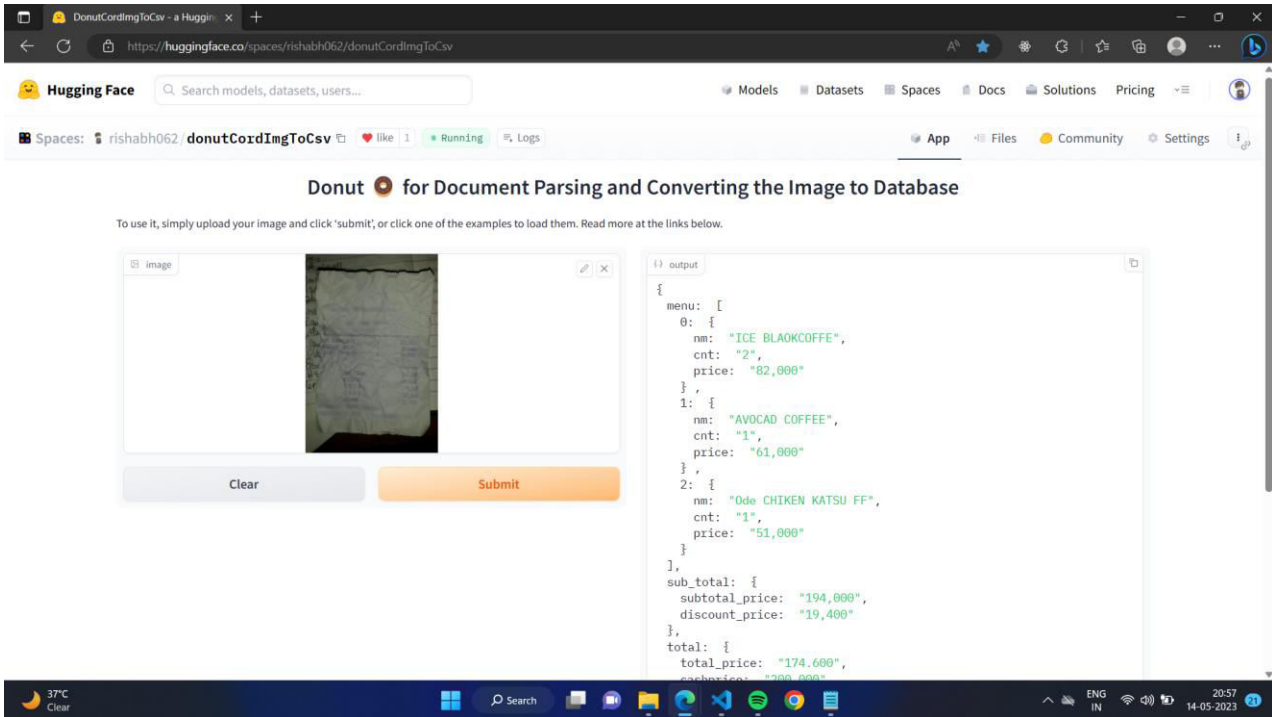Fig4: Doc Q&A Upload File Example

Fig5: Doc Q&A Example-2

Fig6: Donut Example-1

V. CONCLUSION

The conclusion of the project is to make the life of peoples easier by saving their time and money in OCR related things. The webApp is highly responsive and quick for handling images and prompts. The code is clean, highly readable and easy to adapt. The models are platform independent and can be run in any machine with respective configurations. We can run webApp on any device, hence it is useful in every environment.

VI. ACKNOWLEDGEMENT

REFERENCES

1. Analytics Vidhya Blogs : https://www.analyticsvidhya.com
2. Kaggle : https://www.kaggle.com/
3. Github: https://github.com/
4. Hugging Faces: https://huggingface.co/
5. Medium Blogs: https://medium.com
6. arXiv Research Paper: https://arxiv.org/abs/1912.13318

BIOGRAPHY

[1] **Rishabh Dwivedi** pursuing Bachelor of Technology with a major in Computer Science and Engineering from Government Engineering College, Bilaspur (C.G) Affiliated to Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G).

[2] **Varsha** pursuing Bachelor of Technology with a major in Computer Science and Engineering from Government Engineering College, Bilaspur (C.G) Affiliated to Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G).

[3] **Prankit Sahu** pursuing Bachelor of Technology with a major in Computer Science and Engineering from Government Engineering College, Bilaspur (C.G) Affiliated to Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G).

[4] **Prachi Sahu** pursuing Bachelor of Technology with a major in Computer Science and Engineering from Government Engineering College, Bilaspur (C.G) Affiliated to Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G).

[4] **Sanchita Chourawar** Assistant Professor Computer Science and Engineering Department Government Engineering College, Bilaspur (C.G) Affiliated to Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G).

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details