COMPUTER SCIENCE: IMAGE TEXT DETECTOR

DESIGN PROBLEM:

Being able to differentiate handwritten and machine generated documents can be a useful pre-processing tool across many industries. This technology can be useful for OCR, doctors, pharmacists and analysts.

PROBLEM STATEMENT:

Without technology hat can differentiate between handwritten and machine generated text on documents, sorting, scoring, and translating documents can be a less efficient and accurate process.

DESIGN OBJECTIVE:

Create a deep learning model that can differentiate between handwritten and machine generated text on documents.

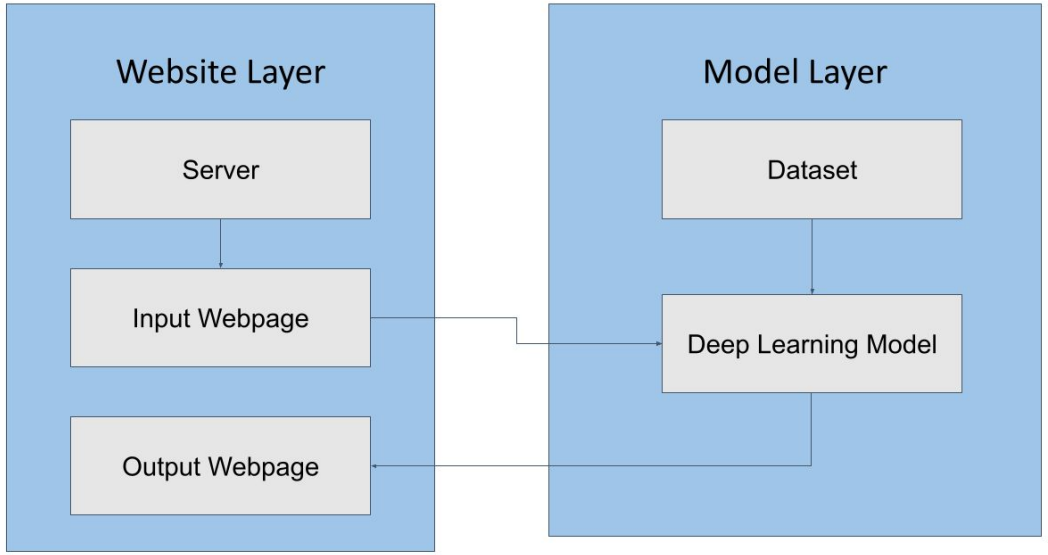
DESIGN CRITERIA/ENGINEERING REQUIREMENTS:

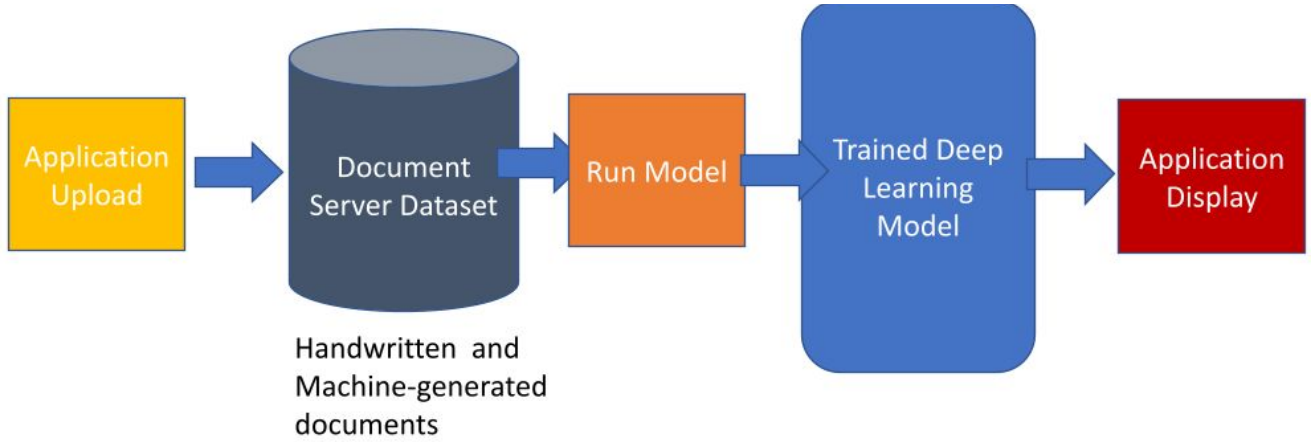
- ***Pull images from a dataset***
- ***Train a deep learning model***
- ***Measure the model's accuracy metrics***
- ***Display the result categorizations on a webpage***

TEAM PICTURE:

DESIGN ARCHITECTURE

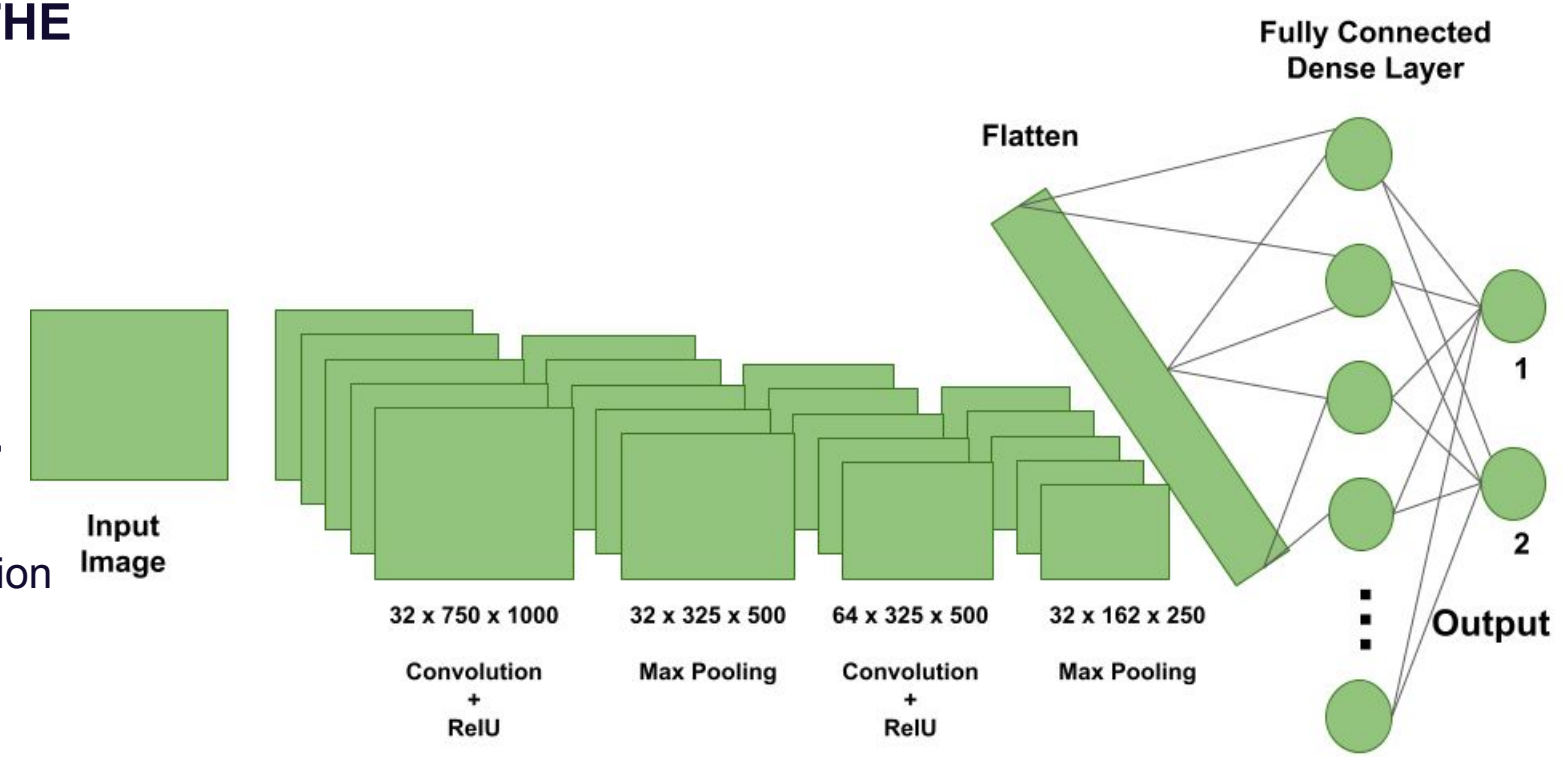
- *The input webpage is hosted on a Heroku server*
- *Files uploaded by the user are stored on the server*
- *The users files are used as an input for the trained deep learning model*
- *The deep learning model produces a list of predictions*
- *Those productions are displayed on the output webpage*





DESIGN IMPLEMENTATION

- **MODELING, BUILDING, TESTING THE SOLUTION**
  - Convolution Neural Network (CNN)
  - Flask Website for Display
  - Data-Engineered Data Set
  - Model Validation Testing
  - Raw input testing
- **DESCRIBE HOW YOU ARRIVED AT THE SOLUTION**
  - CNN models are adept for image classification
  - Website is a user friendly tool to display the results
  - Build Dataset handwritten and printed documents



RESULTS:

- **DISCUSSION/CONCLUSION:**
- **THESE RESULTS SUGGEST…..**
  - **We were able to achieve at least 80% accuracy in distinguishing between handwritten and machine generated**
- **THIS DEVELOPMENT IS IMPORTANT (WHAT WAS LEARNED?)…**
  - **Having a robust data set is crucial for deep learning models**
  - **Importance: The ability to distinguish between handwritten and machine documents for (OCR, doctors and students)**
- **FUTURE DEVELOPMENT SHOULD FOCUS (IMPROVEMENTS?)….**
  - **Improving overall accuracy through: Larger data set, more robust model/architecture, More GPU power**