

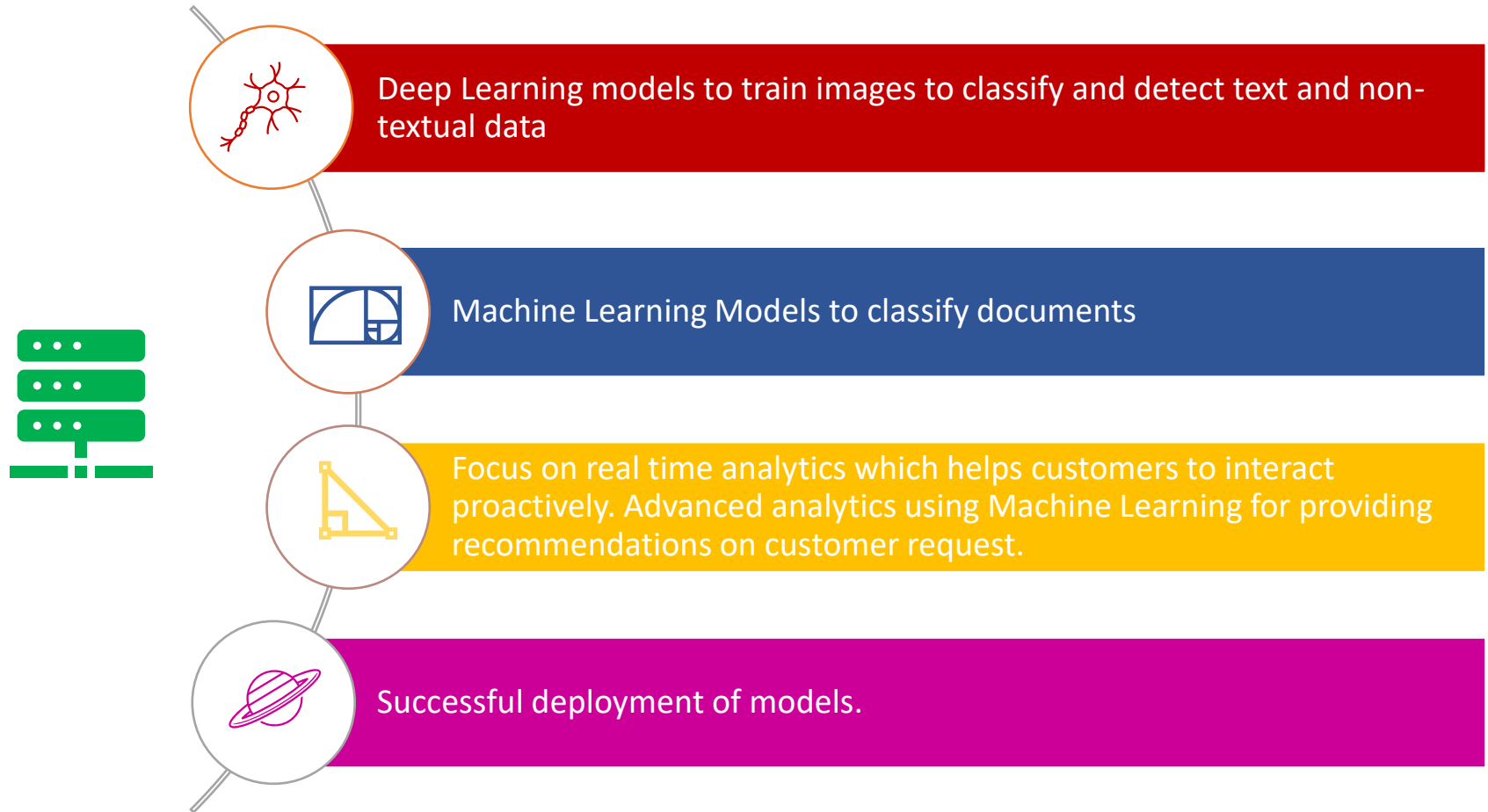
Data Extraction

Check box and signature detection using CNN

Document classification and Field Extraction using Machine Learning



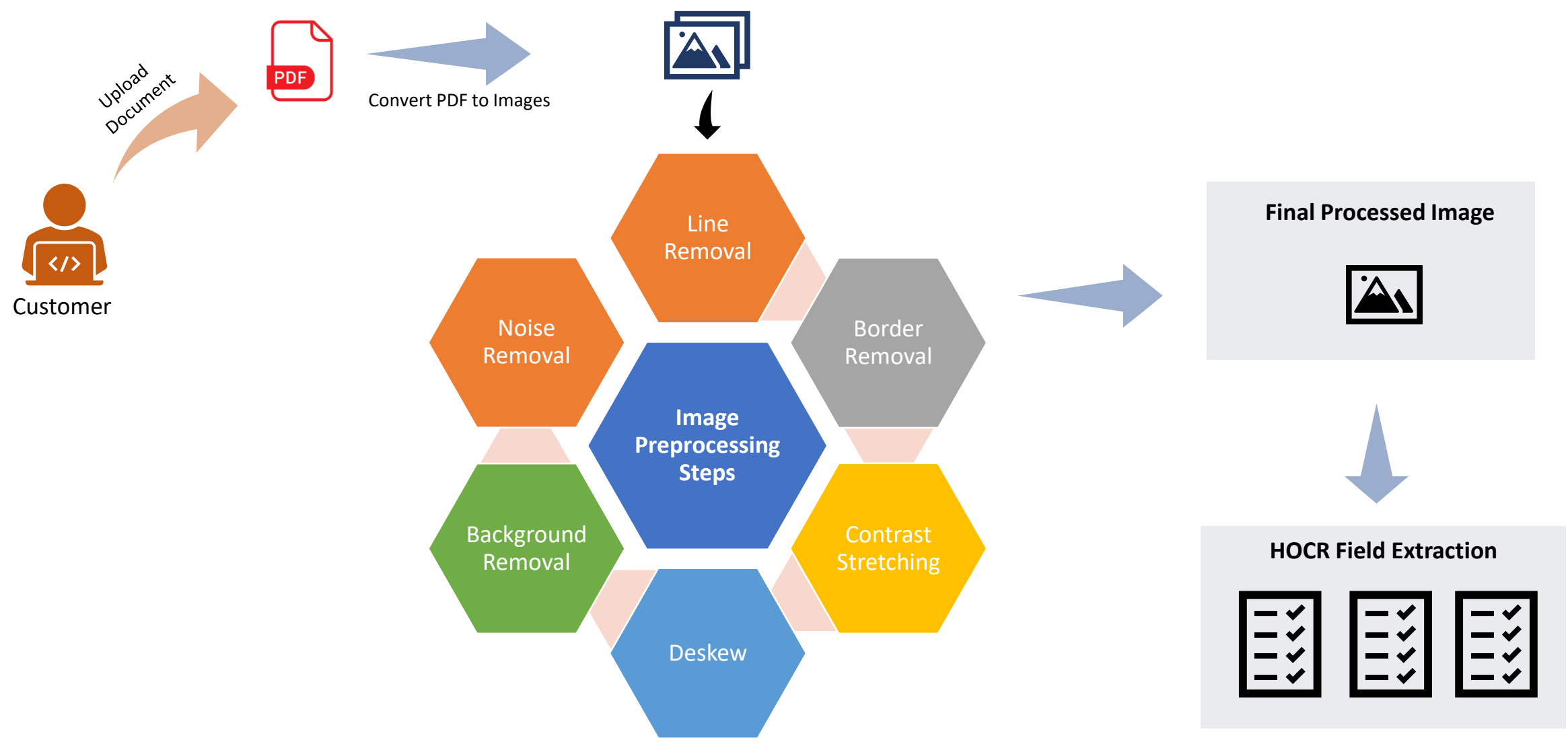
Capabilities



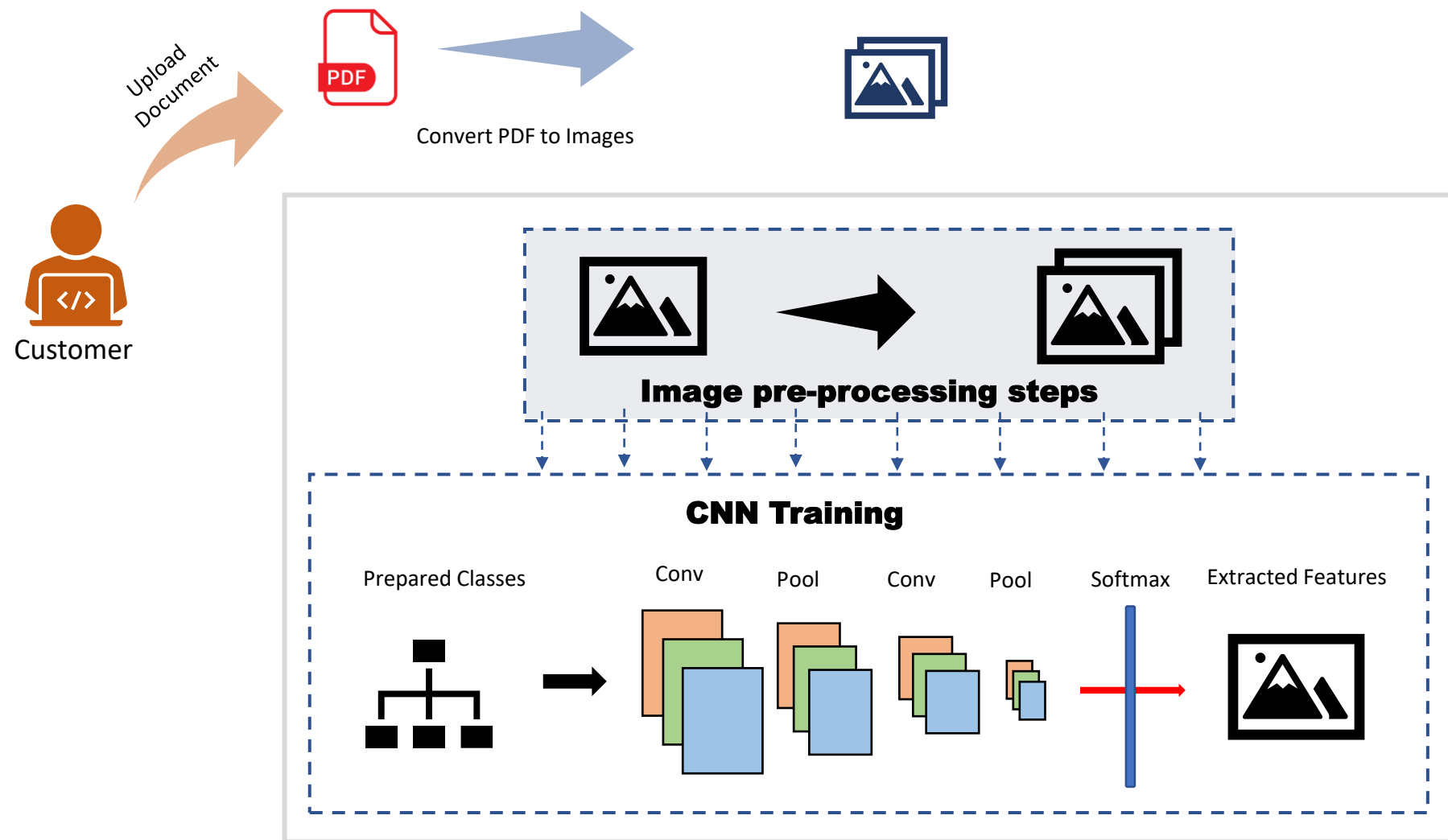
Case Studies

1. Signature and Checkbox Identification in Images which involves training of neural nets to identify the features
2. Identification of Font Attribute in an Image
3. Automatic Information Extraction from document
 - Using OCR (Google-Tesseract) to find key-value pairs
4. Document Classification using two approaches
 - Supervised method using TF-IDF with multiple Machine learning algorithms
 - Unsupervised method using TF-IDF Topic modeling technique

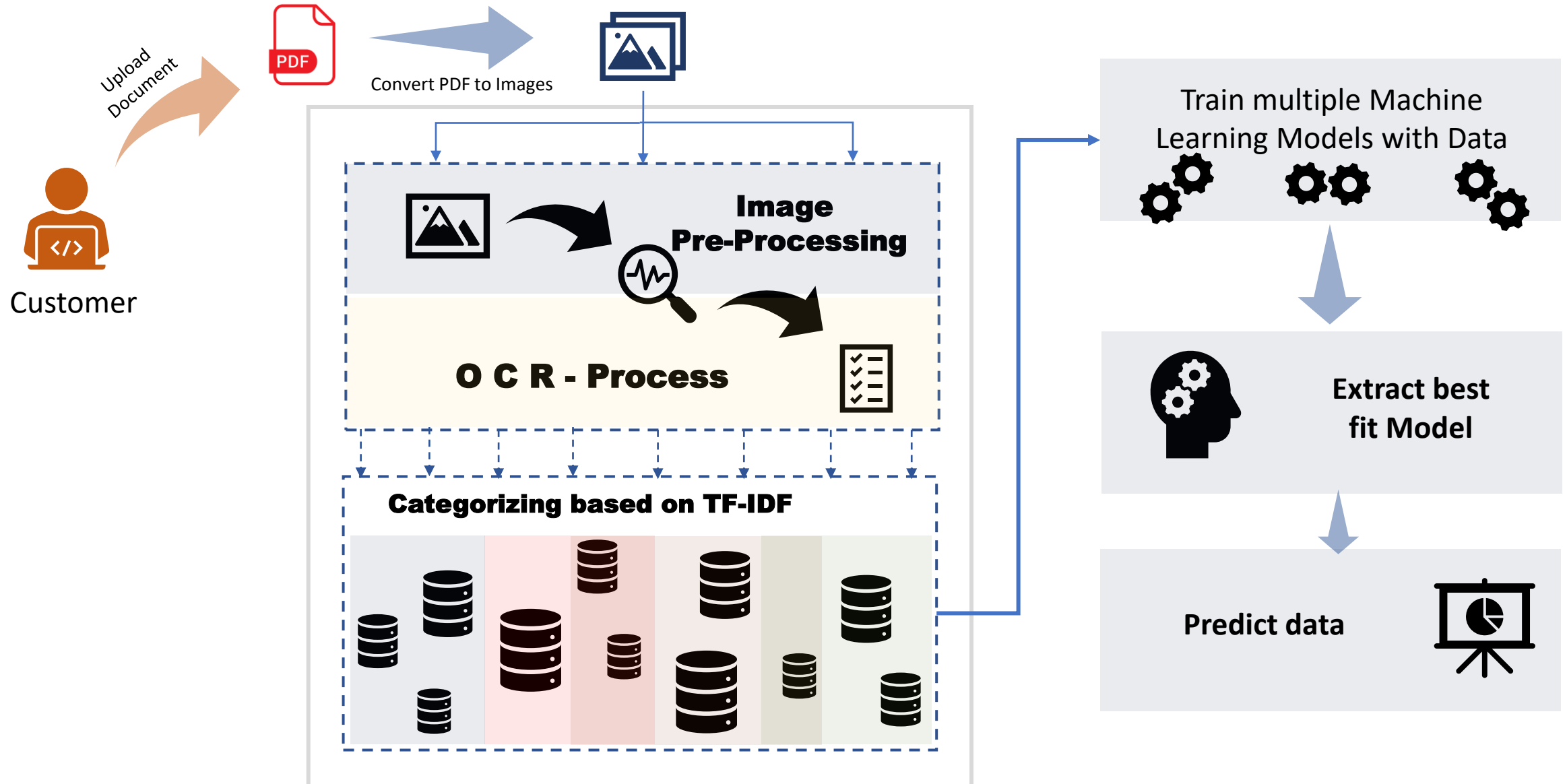
Image Preprocessing Steps For Field Extraction



Checkbox and Signature Identification



Machine Learning TF-IDF



Document Clustering

