Meta Data Extraction from Documents

1. **Problem Statement**

    - Build an AI/ML system to extract some metadata from a document irrespective of its template format.
    - Users can either upload a scanned image or a docx file format.
    - The AI/ML system should be able to extract the following fields from the documents
        - Agreement Value
        - Agreement Start Date
        - Agreement End Date
        - Renewal Notice (Days)
        - Party One
        - Party Two

NOTE: Please do not use a rule-based/oriented approach (RegEx, static conditions, etc).

2. **Dataset Details**

    Please find the data/ folder containing a few .docx and .png files with the following folder structure
    - **train.csv** = Contains the details/metadata of files under train/ folder
    - **test.csv** = Contains the details/metadata of files under test/ folder
    - **train/** = Contains .docx and .png files that can be used for training if required
    - **test/** = Contains .docx and .png files used for testing/evaluation

3. **Evaluation Criteria**
    - Per field Recall
    - Recall here refers to (Per Field)
      True = Number of exact values matches for a document's
      metadata given in the training/validation set to the extracted
      value by the system
      False = Number of Did not match or Not Extracted
      Recall = (True)/(True + False)

4. **Submission Requirements**

    - Properly structured Codebase containing python scripts or Jupiter notebook either in a zipped folder or just share a Github repo link
    - Make sure your codebase contains a **README** file that properly explains your **Solution approach** and instructions to run the code to replicate the predictions
    - Predictions for the files in the test set (test/ folder) (you can mention it in the above README)

- Per field Recall metric score (you can mention it in the above README)
- Wrap your AI/ML system as a RESTful web service to consume it as API. Please mention the instruction to use API in the above README **(Optional)**
