CASI_ASSIGNMENT

Primary Objective Statement:

The objective is to extract key layout elements from the provided PDF file like titles, section headings, textual content, equations, figures, tables, notes, references, and formatting details. The extracted information will be stored in a structured format for easy reference. You can decide what should be the final structure for the layout data.

Key Layout Elements Example:

- **Title:** "Algorithms and Inference"
- **Subtitle:** "Statistics is the science of learning from experience..."
- **Sections:**
  - "1. Algorithms and Inference"
  - "1.1 A Regression Example"
- **Textual Content:**
  - Paragraphs discussing statistical theories, algorithms, and inference.
- **Equations:**
  - (1.1) Mean calculation
  - (1.2) Standard error calculation:
- **Figures:**
  - Figure 1.1: Kidney fitness vs age for 157 volunteers.
  - Figure 1.2: Local polynomial lowess fit to kidney-fitness data.
  - Figure 1.3: Bootstrap replications of lowess fit.
  - Figure 1.4: Scores for gene 136, leukemia data.
  - Figure 1.5: Two-sample t-statistics for 7128 genes, leukemia data.
- **Tables:**
  - Table 1.1: Regression analysis of kidney data.
- **Notes:**
  - Note [p. 6]: Information about the origin of lowess and kidney data.
- **References:**
  - Legendre (1805), Gauss (1795), Tukey (exploratory/confirmatory system), Cleveland (1981), Efron (2004).
- **Formatting:**
  - Consistent font and spacing.
  - Numbered lists for equations, figures, and tables.

Secondary Objective (PDF Recreation):

Utilizing the extracted information, a PDF document will be created with the original text, incorporating titles, sections, equations, figures, tables, notes, and references. The formatting will be maintained for a clear and organized presentation of statistical concepts, algorithms, and inference methods.

Submission Instructions

- **Code/ Script to replicate the process with sample PDF given.**
- **Output of the Script in JSON or any other similar formats.**
- **You may also want to extract and store images in different folders.**
- **It's not necessary to extract everything but one should try to get as many things as possible.**
- **Any library can be used and there is no limitation.**
- **Please try adding comments as necessary.**