Please install all the libraries used in the code before running it

|  |  |
| --- | --- |
| **PyMuPDF** | Text and Meta data extraction and used with Pillow for Images. |
| **Tabula** | Allows users to the extraction of the table and converts the PDF file directly into Data frames or JSON. |
| **Pdfminer.six** | It is a library for extracting text more accurately. It is considered to be the best in text extraction |
| **Pillow** | The Pillow library contains all the basic image processing functionality. |
| **Textblob** | The textblob library allows us to easily determine the sentiment of a piece of text |
| **FPDF** | Writing the Images and text to a PDF. |
| **os** | Provides functions for creating and removing a directory (folder), fetching its contents, changing and identifying the current directory, etc |
| **re** | Built in python library for creating regex patterns |

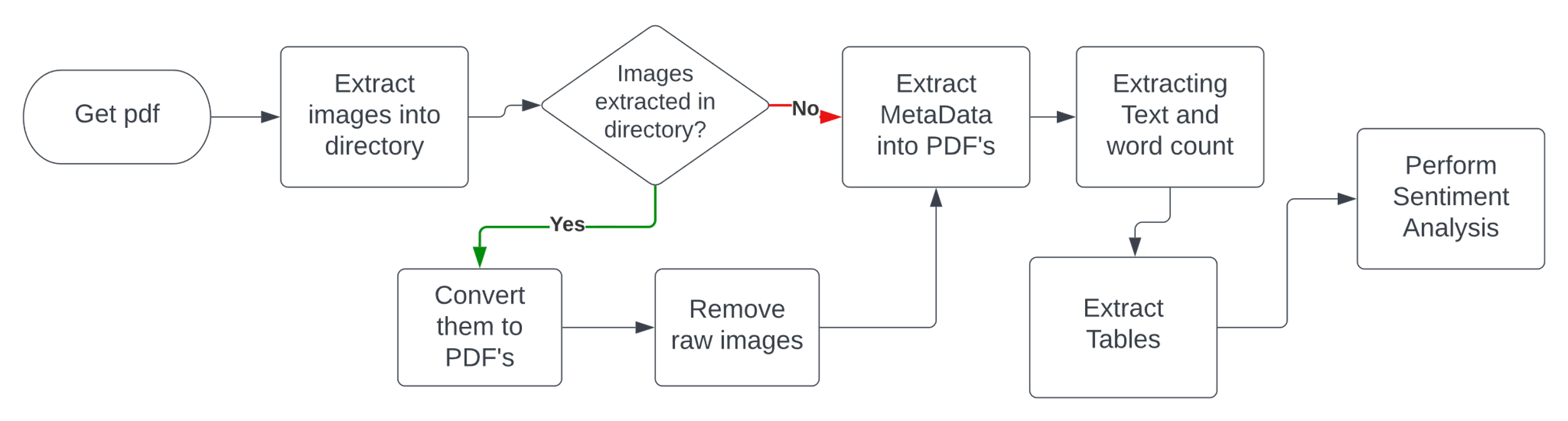


Figure Flow chart of its working