**Thu thập dữ liệu:**

Đọc file pdf, doc, docx sang text

Chia nhỏ theo từng bộ luật riêng, điều khoản trong các folder, file text riêng lẻ

**Xử lý dữ liệu:**

Convert dữ liệu từng điều khoản, chia cấu trúc dữ liệu bằng file.json ( nhằm mục đích truy suất dữ liệu dễ dnag hơn)

Cấu trú file JSON:

{

"instruction": "Nội dung của Điều 43 trong Chương III của luat\_lao\_dong (bo\_luat\_lao\_dong) là gì?",

"input": "",

"output": "Nội dung đầy đủ của Điều 43 ...",

"meta": { "law\_category": "bo\_luat\_lao\_dong", "law\_file": "luat\_lao\_dong", "chapter": "Chương III", "clause": "Điều 43. Nghĩa vụ của người sử dụng..." }

}

**Xây dựng Mô Hình:**

Lựa chọn Mô hình mt5: