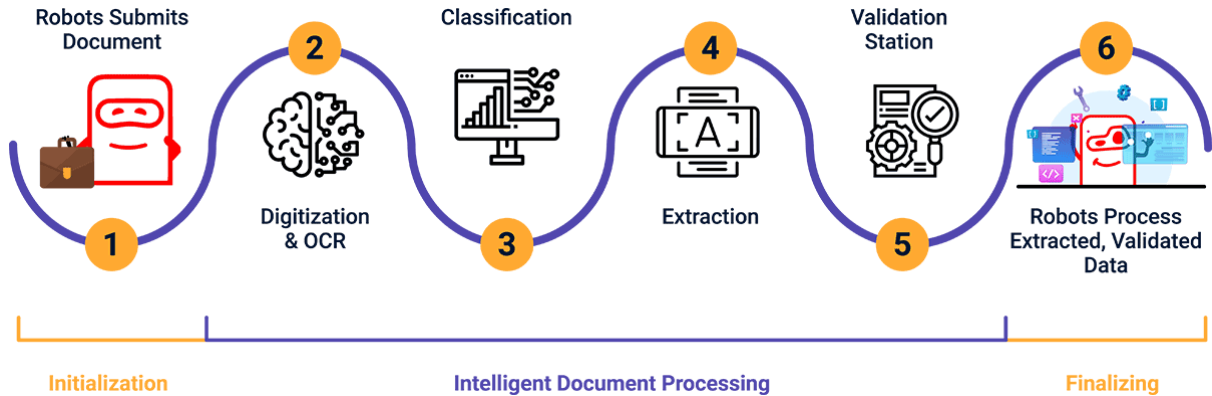


프로젝트 1 : AI 를 이용한 문서 분류

내용 :

AI 모델을 이용하여 입력 받은 문서를 정해진 분류로 구별하여 태깅한다.

클라우드 스토리지나 파일 서버에 저장 된 파일들을 지정된 카테고리로 분류 할 수 있도록 한다.



사용 기술 :

- 공공 행정문서 OCR
 - <https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=88>
- DocBERT / Reg-LSTM
 - <https://github.com/castorini/hedwig>
- fastText
 - <https://fasttext.cc/>
- Document Classification - paperswithcode
 - <https://paperswithcode.com/task/document-classification>
- DataSet - Web of Science / 20 Newsgroups
 - <https://data.mendeley.com/datasets/9rw3vkcfy4/2>
 - <http://qwone.com/~jason/20Newsgroups/>

기능 :

- 문서 파일을 파일 서버로 업로드시 자동으로 분류하여 저장
- 공공 행정문서 OCR 데이터 셋을 기준으로 문서 분류