



HS코드 추천 시스템 개발 공정도

♡ 류	🔧 개발 이름	≡ 상세 설명	🕒 진행 상황	📋 개발 순서	📅 시작일자	📅 종료일자	👤 담당자
	<u>부 이미지 자료조사</u>	이미지가 잘 조사되는 품목들은 어떤 것들이 있는지 살펴봄. 이후 이미지의 패턴이 찾기 어려워 이미지로 분류하는 것은 폐기하기로 결정	완료	자료조사	@12/11/2023	@12/12/2023	강민서 심세현 오지현 원현식 이현서 전하늘
84류	16부 84류 (영문)해설서(86개)	크롤링+파일저장	완료	데이터 수집 테스트	@12/12/2023	@12/13/2023	강민서 심세현 오지현 원현식 이현서 전하늘
84류	16부 84류 국내 품목사례 (국문)	크롤링+파일저장(4000개) 기간 -2013/1/1 ~ 2023/12/31-	완료	데이터 수집 테스트	@12/13/2023	@12/14/2023	원현식 이현서 전하늘
84류	16부 84류 미국 품목사례	크롤링+파일저장 기간 -2019/1/1 ~ 2023/12/31-	멈춤	데이터 수집 테스트	@12/13/2023	@12/14/2023	심세현 오지현
84류	16부 84류 (영문)해설서 전처리	불용어제거, 토큰화, 어간추출, 피쳐 벡터화	완료	데이터 전처리 테스트	@12/13/2023	@12/14/2023	심세현
84류	16부 84류 미국 품목사례 전처리	불필요한 데이터 삭제(아이템이 두개 이상인 것. 하지만 두 번째 아이템이 99인 것은 포함)	멈춤	데이터 분석 테스트	@12/14/2023	@12/15/2023	심세현
84류	16부 84류 국내 품목사례 (국문) 데이터 전처리(초)	불필요한 컬럼 삭제/중복행 삭제/ hs code 4자리 통일	완료	데이터 전처리 테스트	@12/15/2023	@12/15/2023	심세현 원현식 이현서 전하늘
84류	16부 84류 국내 품목사례 (국문) 데이터 전처리(중)	구글 번역기로 모든 데이터 번역	완료	데이터 전처리 테스트	@12/15/2023	@12/16/2023	심세현 오지현 원현식 이현서 전하늘
84류	16부 84류 국내 품목사례 (영문) 데이터 전처리(후)	영어 제거/소문자화/구두점 제거/불용어(TF-IDF)/어근 추출	완료	데이터 전처리 테스트	@12/17/2023	@12/18/2023	심세현 원현식 전하늘

♡ 류	🔧 개발 이름	≡ 상세 설명	🕒 진행 상황	📋 개발 순서	📅 시작일자	📅 종료일자	👤 담당자
84류	16부 84류 국내 품목사례 (영문) ML	모델 학습/예측/확정	완료	ML 테스트	@12/18/2023	@12/18/2023	심세현 원현식 전하늘
85류	16부 85류 (영문)해설서(48개)	크롤링+파일저장	완료	데이터 수집 테스트	@12/18/2023	@12/18/2023	심세현
85류	16부 85류 국내 품목사례 (국문)	크롤링+파일저장(5116개) 기간 -2013/1/1 ~ 2023/12/31-	완료	데이터 수집 테스트	@12/18/2023	@12/18/2023	심세현 오지현 이현서
85류	16부 85류 국내 품목사례 (국문) 데이터 전처리(초)	HS 코드 4자리로 자르기	완료	데이터 전처리	@12/18/2023	@12/18/2023	전하늘
84류	16부 84류 국내 품목사례 (영문) ML	코사인 유사도 측정	완료	ML 테스트	@12/19/2023	@12/20/2023	원현식 전하늘
84류	16부 84류 물품설명에 없는 hs코드 임의생성(가데이터)	사례 없는거 7개/사례 1개뿐 인거 10개 가데이터 생성	완료	데이터 수집 테스트	@12/19/2023	@12/19/2023	이현서
85류	16부 85류 영문해설서 전처리	불용어제거, 토큰화, 어간추출, 피쳐 벡터화	완료	데이터 전처리 테스트	@12/19/2023	@12/19/2023	강민서 오지현
85류	16부 85류 국내 품목사례 (국문) 전처리 (초/중)	HS코드 4자리/중복행 삭제/85류가 아닌 데이터 삭제/구글 번역기로 모든 데이터 번역/가데이터 합침	완료	데이터 전처리 테스트	@12/19/2023	@12/20/2023	심세현 이현서
84류	16부 84류 가 데이터 넣어서 LSVM 모델 돌려보기	정확도 0.836	완료	ML 테스트	@12/20/2023	@12/20/2023	심세현
39류	07부 39류 (영문)해설서 25개	크롤링+파일저장	완료	데이터 수집	@12/20/2023	@12/20/2023	강민서
39류	07부 39류 국내 품목사례(국문) 크롤링&전처리(초)	크롤링+파일저장(3209개) 기간 -2013/1/1 ~ 2023/12/31- HS코드 4자리/중복행 삭제/39류가 아닌 데이터 삭제	완료	데이터 수집 데이터 전처리	@12/20/2023	@12/20/2023	원현식
85류	16부 85류 물품설명에 없는 hs코드 임의생성(가데이터)	사례 없는거 3개 (8522,8546,8549)	완료	데이터 수집	@12/20/2023	@12/20/2023	이현서
85류	16부 85류 국내 품목사례 (영문) 전처리 (후)&가데이터 넣고 모델 돌려보기	영어 제거/소문자화/구두점 제거/불용어(TF-IDF)/어근 추출	완료	ML 데이터 전처리	@12/20/2023	@12/21/2023	오지현

♡ 류	🔧 개발 이름	≡ 상세 설명	🕒 진행 상황	📋 개발 순서	📅 시작일자	📅 종료일자	👤 담당자
84류 85류	16부 84류, 85류, 학습되지 않은 데이터 ML	84, 85와 학습되지 않은 데이터를 넣어 3개를 잘 분류할 수 있는지 확인	완료	ML	@12/21/2023	@12/21/2023	심세현 원현식 이현서 전하늘
39류	07부 39류 국 내품목사례(국문) 전처리(중)	구글 번역기로 모든 데이터 번역	완료	데이터 전처리	@12/21/2023	@12/22/2023	심세현 원현식 전하늘
39류	07부 39류 국 내품목사례(영문) 전처리(후)&ML	영어 제거/소문자화/구두점 제거/불용어(TF-IDF)/어근 추출/LSVM	시작 전	ML 데이터 전처리	@12/21/2023		
84류 85류	16부 영국, EU, 필리핀 크롤링		완료	데이터 수집	@12/21/2023	@12/24/2023	강민서 심세현 오지현 원현식 이현서 전하늘
84류 85류	84, 85 DATA 취합&전처리&모델링	EU/영국/한국 84, 85류 전체 취합&전처리(초/후)&사례 300개 이하 삭제/모델링/미국사례 predict	완료	ML 데이터 전처리	@12/26/2023	@12/27/2023	강민서 심세현 오지현 원현식 이현서 전하늘
84류 85류	84, 85류 총 데이터 모델링	필리핀 사례를 predict로 사용해본다.(미국사례는 예측을 못하니까)	완료	ML	@12/28/2023	@12/28/2023	심세현
84류 85류	키워드 사전 만들기	84, 85 해설서에서 불필요한 단어들을 사람이 보고 삭제 한 뒤 굵어옴/TF-IDF를 활용하여 각 호의 키워드를 추출/HS코드 사례가 별로 없는 HS CODE들을 키워드로 구분하기 위해 만들	완료	데이터 수집 데이터 전처리	@12/28/2023	@12/28/2023	강민서 오지현 원현식 이현서 전하늘
84류 85류	불용어 사전 만들기	ML은 불용어 사전이 따로 있지만 LSTM에는 불용어 사전이 따로 없어 우리가 만들어야 함/키워드 사전 만들 때도 불용어를 제거한 후 TF-IDF를 해야 더 잘 뽑히기 때문	완료	데이터 수집 데이터 전처리	@12/28/2023	@12/28/2023	이현서
84류 85류	84, 85류 사례 데이터셋 최종	HS코드 4자리/중복행 삭제/84류가 아닌 데이터 삭제/구글 번역기로 모든 데이터 번역/가데이터 합침/구두점+불용어 제거	완료	데이터 전처리	@12/28/2023	@12/29/2023	이현서

♡ 류	🔧 개발 이름	≡ 상세 설명	🕒 진행 상황	📅 개발 순서	📅 시작일자	📅 종료일자	👤 담당자
84류 85류	84, 85류 총 데이터 LSTM 모델링	불용어 사전 전처리/LSTM 모델링&키워드 검증	완료	ML 데이터 전처리	@12/28/2023	@01/02/2024	전하늘
84류 85류	코사인 유사도 개발	코사인 유사도 코드 개발	완료	코사인유사도	@12/29/2023	@12/31/2023	이현서
84류 85류	84, 85류 총 데이터 LSTM 모델 정제&모델 시각화 &LSVC 모델링	acc/F1 score/RECALL/Precision 등 여러 평가지표를 넣어서 LSTM 모델링/	완료	ML	@01/02/2024	@01/03/2024	심세현 원현식 이현서 전하늘
84류 85류	84, 85류 총 데이터 LSVC/LSTM/ 유사도 3개 총 합	3개를 사용하여 HS 코드에 대한 예측률이 높은 순대로 가중치 부여 후 최종 값이 높은 순대로 HS CODE를 추천하는 코드 제작	완료	ML	@01/03/2024		심세현 이현서 전하늘
PPT	PPT 순서 제작&서론 제작		완료		@01/03/2024	@01/03/2024	강민서 오지현
PPT	PPT 전처리 제작		완료		@01/04/2024	@01/05/2024	강민서 오지현
PPT	PPT 모델링 제작		완료		@01/05/2024	@01/06/2024	강민서 오지현
PPT	PPT 결론 제작		완료		@01/08/2024	@01/08/2024	강민서 오지현