

AI 도입한 기업들, AI 서비스 구입 비용 약 6배 증가

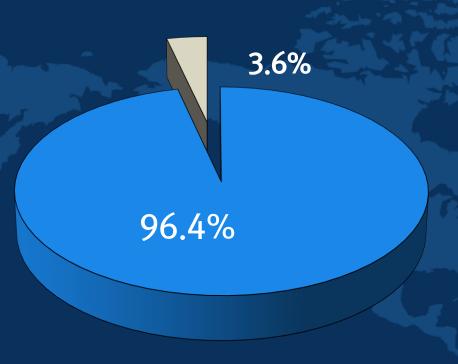
2018 8억

약 6배 증가

2019 50억

산업연구원(2020), 기업의 AI 활용 실태조사

AI를 도입한 기업



96.4% 아직 도입하지 않음

KDI(2021), AI에 대한 기업체 인식 및 실태 조사 결과

Index.

인공지능 모델

2

진행 일정

- 데이터 탐색 및 전처리
- 모델 설명 및 파인 튜닝
- 웹페이지 구현 및 모델 연동

3 발전 및 개선 방향

4

개발 환경 및 팀원 소개





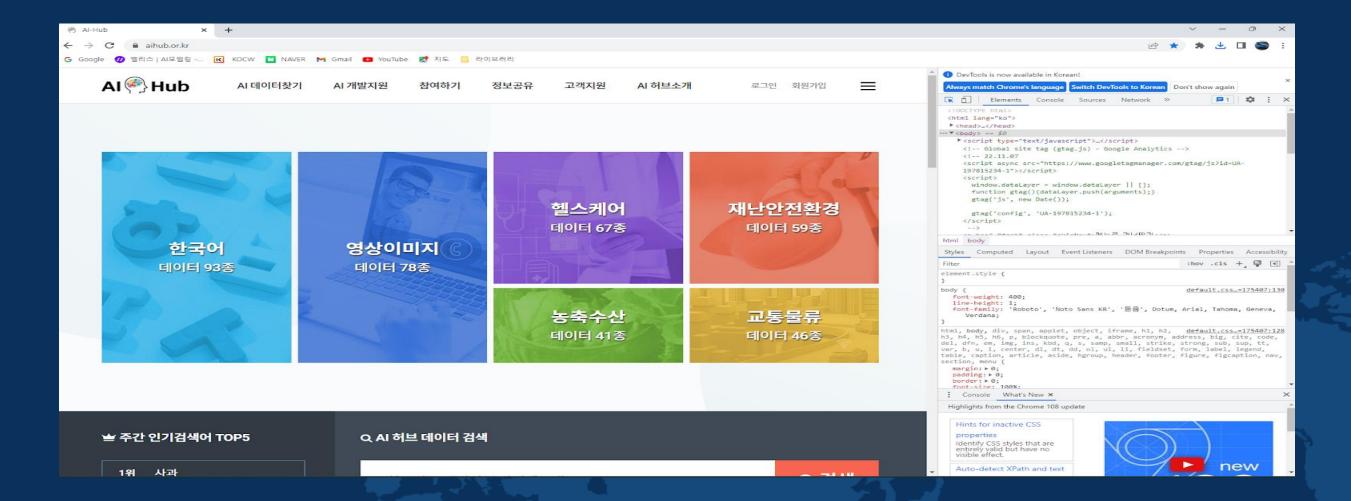


	А	В	С	D
1	소분류(000)	관련도 높은 3개	그 외 n개	
2	곡물 및 기타 식량작물 재배업	작물 디지털 피노타이핑 데이터, 식물 병 유발 통합 데이터_, 노지 작물 질병 진단 이미지_	정밀농업 노지작물 통합 데이터+, 농산물 품질(QC) 이미지	+, 드론 농경작지 🕯
3	채소, 화훼작물 및 종묘 재배업	노지 작물 질병 진단 이미지_, 지능형 스마트팜 통합 데이터(파프리카), 식물 병 유발 통합 데이터_	지능형 수직농장 통합 데이터(엽채류)+, 지능형 스마트팜 통	통합 데이터(토마토
4	과실, 음료용 및 향신용 작물 재배업	식물 병 유발 통합 데이터_, 노지 작물 질병 진단 이미지_, 주요 농작물 생육 이미지_	고품질 과수작물 통합 데이터+, 지능형 수직농장 통합 데이	터(딸기)+, 전북 장
5	기타 작물 재배업	식물 병 유발 통합 데이터_, 동의보감 약초 이미지, 동의보감 독초판별 이미지 데이터	지능형 스마트팜 통합 데이터(버섯)+, 식의약용 자생식물 분	본석 데이터, 정밀농
6	시설작물 재배업	식물 병 유발 통합 데이터_, 작물 디지털 피노타이핑 데이터_, 농산물 품질(QC) 이미지_	시설 작물 개체 이미지+, 시설 작물 질병 진단 이미지+, 지능	형 스마트팜 통합
7	소 사육업	한우 신체중실지수 등급 데이터, 지능형 스마트 축사 통합 데이터(한우), 지능형 스마트 축사(젖소) 데이터	가축 행동 영상9, 축산물 품질(QC) 이미지8, 지능형 곤충 시	l육 데이터7, 지능 ⁴
8	양돈업	축산물 품질(QC) 이미지, 지능형 스마트 축사 통합 데이터(양돈), 가축 행동 영상	지능형 스마트 축사 (젖소) 데이터, 지능형 스마트 축사 데이	[터(육계. 산란계),
9	가금 사육업	지능형 스마트 축사 데이터(육계. 산란계), 축산물 품질(QC) 이미지, 가축 행동 영상	지능형 스마트축사 통합 데이터(양돈), 지능형 스마트축사	통합 데이터(한우),
10	기타 축산업	축산물 품질(QC) 이미지, 지능형 스마트축사 통합 데이터(양돈), 가축 행동 영상_	지능형 스마트축사 통합 데이터(한우)+, 지능형 스마트 축사	ŀ (젖소) 데이터, ㅈ
11	작물 재배 및 축산 복합농업	지능형 스마트 축사 통합 데이터(양돈), 드론 농경작지 촬영 영상_, 시설 작물 질병 진단 이미지_	제주 주요작물 자동탐지 데이터, 지능형 수직농장 통합 데이	l터(엽채류), 지능 ³
12	조경수 식재 및 관리 서비스업	식물 병 유발 통합 데이터, 드론 농경작지 촬영 영상, 토지 피복지도 항공위성 이미지 데이터(전라)_	토지 피복지도 항공위성 이미지 데이터(제주도), 토지 피복	지도 항공위성 이미
13	작물재배 관련 서비스업	식물 병 유발 통합 데이터_, 드론 농경작지 촬영 영상_, 시설 작물 질병 진단 이미지_	정밀농업 노지작물 통합 데이터+, 정밀농업 농기계 잡초 인	식 데이터+, 정밀능
14	축산관련 서비스업	축산물 품질(QC) 이미지, 가축 행동 영상, 한우 신체충실지수 등급 데이터	야생동물 활동 영상 데이터, 지능형 스마트축사 통합 데이터	대(양돈), 지능형 스
15	수렵 및 관련 서비스업	야생동물 활동 영상 데이터, 가축 행동 영상, 멀티센서 동선 추적 데이터_	드론 이동체 인지 영상(전방 고정)+, Small object detection	을 위한 이미지 데(
16	영림업	산림 수종 이미지 데이터(전라), 동의보감 약초 이미지, 식물 병 유발 통합 데이터_	동의보감 독초판별 이미지 데이터+, 산림 수종 이미지 데이	터(제주), 산림 수충
17	벌목업	산림 수종 이미지 데이터(전라), 산림 수종 이미지 데이터(제주), 산림 수종 이미지 데이터(강원 및 충청)	산림 수종 이미지 데이터(경상), 산림 수종 이미지 데이터(수	누도권), 태풍 및 홍
18	임업관련 서비스업	산림 수종 이미지 데이터(전라), 토지 피복지도 항공위성 이미지(강원 및 중청), 산림 수종 이미지 데이터(경상)_	이상행동 CCTV 영상+, 산림 수종 이미지 데이터(수도권), 터	내풍 및 홍수로 인형
19	해면 어업	수산양식업 먹이생물 AI 학습 데이터, 기상 정보 데이터_, 해안 오염물질 데이터_	어류 행동 영상+, 지능형 스마트양식장 통합 데이터(뱀장어)+, 지능형 스마트
20	내수면 어업	수산양식업 먹이생물 AI 학습 데이터, 기상 정보 데이터_, 해안 오염물질 데이터_	어류 행동 영상+, 지능형 스마트양식장 통합 데이터(뱀장어)+, 지능형 스마트
21	양식어업	어류 행동 영상, 지능형 스마트양식장 통합 데이터(뱀장어), 어류 개체 촬영 영상_	수산양식업 먹이생물 AI 학습 데이터+, 지능형 스마트양식경	장 통합 데이터(가리
22	어업관련 서비스업	수산양식업 먹이생물 AI 학습 데이터, 어류 행동 영상, 기상 정보 데이터_	지능형 스마트양식장 통합 데이터(뱀장어)+, 지능형 스마트	양식장 통합 데이
23	석탄 광업	산업시설 열화상 CCTV 데이터_, 고소작업 현장 실시간 영상 데이터_, 기계시설물 고장 예지 센서_	자연재해로 인한 생활시설 안전 데이터, 과적차량 도로 위험	함 데이터, 물류창고

데이터 탐색 및 전처리

• 업종 별 관련성 분류

- 통계청의 한국표준산업분류표 소분류를 첫번째 컬럼으로 사용
- 각 소분류 업종과 상대적으로 Al-Hub의 가장 유사한 3가지 데이터 셋을 두 번째 컬럼으로 사용
- 그 외 유사한 데이터셋을 세 번째 컬럼으로 사용
- 이 자료는 모델(Word2Vec, SBERT)의 사전학습을 위해 사용

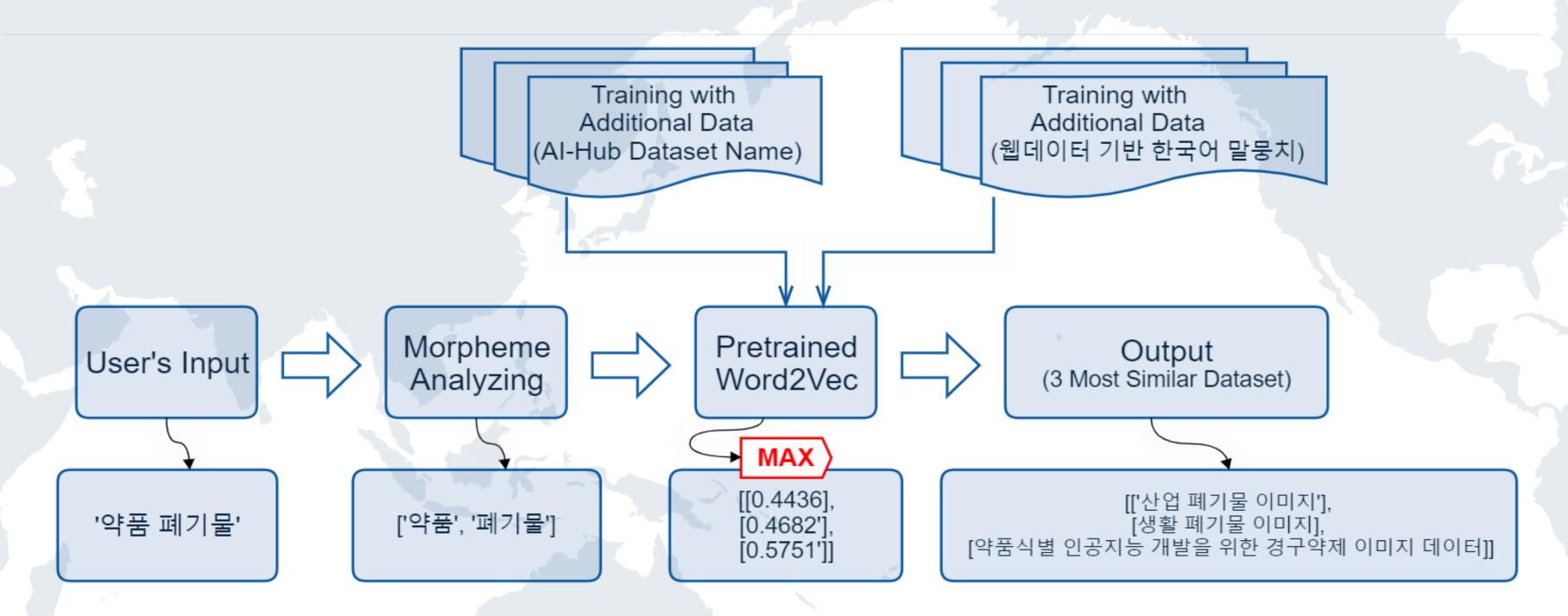


데이터 탐색 및 전처리

• 데이터 크롤링

- Al-Hub 데이터 찾기에 접속
- 모든 데이터셋 제목과 각각의 링크 가져오기
- 커리어 사이트 기업정보란 접속
- 기업명 및 주요 사업 가져오기

모델 1 – Word2Vec 유사도 활용



모델 2 - SBERT 모델 활용



모델 성능 비교

	Word2Vec	Kor_SRoBERTa
정확도	70%	76% 예측이 정확할 확률이 높아요.
Spearman 상관 계수	0.42	0.79 더 유사한 결과를 얻을 수 있어요.
Sentence per Sec	17817 속도가 더 빨라요.	133

고려대학교 (2021) ,'Word2Vec, GloVe 및 RoBERTa 등의 모델을 활용한 한국어 문장 임베딩 성능 비교 연구'





진행 일정

주차	기간	세부 진행 내용
1주차	11/07 ~ 11/11	프로젝트 주제선정 및 기획
2주차	11/14 ~ 11/18	데이터 확보 및 전처리
3주차	11/21 ~ 11/25	모델 1 (Word2Vec)
4주차	11/28 ~ 12/02	모델 2(KoSroBERTa)
5주차	12/05 ~ 12/09	웹페이지 제작
6주차	12/12 ~ 12/14	최종점검 및 발표준비





메인터넌스 앱서비스 웹 개선 패키지선택

시간이 지남에 따라 Al Hub에 데이터셋의 수가 늘어나므로, 모델의 정확성이 올라감 사용자의 반응에 따라 앱 서비스 고려 바로가기 창이 아닌 해당 페이지에서 데이터를 정제하여 사용자가 이해하기 쉽게 제시 사용자에게 Al 모델 추천, 데이터셋 구축, 올인원 3가지 패키지 선택권 부여





TOOL







소통







팀장 이재혁

프로젝트 총괄 데이터 관리 프론트 엔드 피피티 제작



팀원 임철성

데이터 전처리 웹 크롤링 모델 2 제작 Flask



팀원 이민흠

데이터 전처리 웹 크롤링 모델 2 제작



팀원 정우현

데이터 전처리 프론트 엔드



팀원 손재형

데이터 전처리 서비스 기획 모델 1 제작 피피티 제작



팀원 주성민

데이터 전처리 서비스 기획 모델 1 제작

출처

REIMERS, Nils; GUREVYCH, Iryna. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.

Junghyun Gan, 2021년 08월 16일, https://huggingface.co/jhgan/ko-sroberta-multitask

BM-K, 2022년 12월 09일, https://github.com/BM-K/Sentence-Embedding-Is-All-You-Need#readme

choonsik_mom 2022년 11월 03일, https://velog.io/@choonsik_mom/SBERTSentenceBERT-%EB%85%BC%EB%AC%B8-%EB%A6%AC%EB%B7%B0

kookeej, 2022년 02월 11일, https://velog.io/@kookeej/%EB%85%BC%EB%AC%B8-%EB%A6%AC%EB%B7%B0Sentence-BERT-Sentence-Embeddings-using-Siamese-BERT-Networks

이기창, 2019년 12월 12일, https://ratsgo.github.io/

monetd, 2021년 3월 27일, https://monetd.github.io/python/nlp/Word-Embedding-Word2Vec-%EC%8B%A4%EC%8A%B5/

고려대학교 컴퓨터학과, 'Word2Vec, GloVe 및 RoBERTa 등의 모델을 활용한 한국어 문장 임베딩 성능 비교 연구' (2021), p1-4.

감소한테니다