

김세정

연락처 010-8480-2611

이메일 kimsejung@unist.ac.kr



[주요 관심사]

- Position: Data Scientist / Data Engineer / AI 전략 기획
- AI 콘텐츠, 기술, 서비스, 소프트웨어 등을 Research하고 비즈니스 관점에서의 문제인식과 해결을 중점으로 AI 전략을 수립하는 직무
- 데이터의 수집 - 분석 - 모델 - 배포까지 자동화 구현의 MLOps와 GPU 효율 관점에서의 커널 최적화
- 텐스토렌트의 소프트웨어의 AI 프레임워크(PyTorch, TensorFlow 등)와 원활한 통합 여부 실험
- 각 산업에 적합한 생성형 AI 모델을 활용한 서비스 전략 기획(예: 엔터테인먼트 산업의 SNS 게시물 자동 생성)

1) 통계 지식을 기반으로한 머신러닝, 딥러닝 모델 구현

- 선형회귀분석, 로지스틱회귀분석, 의사결정나무, 연관성분석, 군집분석, ARIMA 등 응용통계 모형에 대한 이해 보유
- XGBoost, LightGBM, CatBoost 등 Gradient Boosting Framework 활용 경험
- Rerank, Item2Vec 를 활용한 추천 모델 구현

2) AI 업무 경험

- VLM(Vision-Language Model)모델 데이터 파이프라인 설계 및 구축 프로젝트 리딩 경험 보유
- Vision Detection 모델의 초거대 AI 모델 데이터세트 구축(수집, 전처리, 파이프라인 설계, 분석) 리딩 경험 보유
- 영상의 메타데이터 추출 및 영상 패치화 기술 보유
- RGB 밴드값, 위경도값, 영상의 노데이터값 등 메타데이터에 없는 데이터 추출 기술 보유
- Milvus를 활용한 영상 데이터의 Vector DB 구축 역량 보유
- Transformer 기반의 텍스트 인코더와 이미지 인코더를 사용
- GPU, CPU등 자원 최소화를 위한 멀티 프로세싱 등 코드 최적화 역량 보유

3) 환경 구축 경험

- MLflow: 실험 관리, 재현성, 배포 및 모니터링
- Airflow: 데이터, 모델 파이프라인을 생성, 스케줄링, 모니터링 자동화
- Databub: 데이터 카탈로그 생성 및 Lineage 관리
- MiniIO: 실험 데이터 및 모델 저장소
- Datalore: 모델에 사용된 데이터의 EDA 시각화를 대시보드로 구축

4)Skills

- Language: Python, SQL
- 데이터분석 시각화 대시보드 구축: Datalore
- Tools / Technologies: Docker, Kubernetes, Git
- Library: PyTorch, transformers

5) 창업가 정신

- 2018.05 ~ 2022.04 창업에 도전한 경험으로 창업자의 고충과 국내 창업 생태계를 이해하는 사람으로 성장함
- 다 수의 정부지원과제를 수행하는 과정에서 자금 확보 및 서비스 개발까지 전반적인 영역을 핸들링한 경험 보유
- 스타트업의 중요한 keyword는 '소통' '팀워크' '시장에 대한 이해' '창업자의 경영능력' 이라고 생각함

6) 문서작성 및 발표

- 경진대회 수상 및 정부지원사업 선발 경험은 문서작성 및 발표능력을 입증함
- 문서 및 발표자료는 전체적인 스토리텔링을 중요시하며 구체적이고 현실성 있게 작성함

경력사항

총 12년

(주)엠에스인포테크

팀장 (데이터전략팀)

2024.10 - 2024.12

3개월

[수행 프로젝트 별 보유 역량]

1. 데이터 관리 환경 구축

- DB에 단순 적재된 데이터를 Datahub 도입하여 카탈로그화 수행
- Airflow를 DAG 생성하여 Datahub Lineage 자동 업데이트

2. 머신러닝 모델 실험 및 배포 환경 구축

- MLflow, Airflow를 도입하여 모델 등록, 버전, 추적 관리 기능 구현
- 모델과 서비스를 자동화된 방식으로 배포 및 관리 기능 구현
- MinIO 도입하여 실험 데이터 및 모델을 저장 및 재활용 환경 구현

3. 테이블오더 플랫폼 주문량 예측 모델 개발

(1) 목적: 플랫폼 입점 업체의 주문량 예측을 통한 재고 관리

(2) 수행 내용

- 데이터 feature 엔지니어링으로 DB내 적재되어 있는 데이터 중 적합한 feature 선정
- XGBoost, LightGBM, CatBoost 모델에 적합하도록 인코딩, 스케일링 등 데이터 전처리
- MLflow로 실험결과를 저장하여 파라미터값 별 성능 추적
- 모델의 과적합 여부 검증을 위한 K 폴드 교차검증과 Early Stopping 시행

(3) 성과

- 모델 성능(R^2 Score): 최저 0.86 ~ 최고 0.96

[이직 희망 사유]

1. 유사 직무의 동료 부재

- 사내 데이터엔지니어, ML 엔지니어, 데이터사이언티스트, AI 리서처 등 관련 직무의 동료가 없음
- 신규 채용에 대한 의사가 없어 해소가 어려울 것으로 판단함

2. 기술을 이해하는 C레벨의 부재

- 머신러닝, 딥러닝을 위해 환경 구축과 리소스 지원이 필요함에 대한 소통에 한계가 있음
- 플랫폼을 보유하고 있어 Ops 환경 구축에 대한 필요성을 설득하였으나 기술에 대한 이해가 부족함

[수행 프로젝트 별 보유 역량]

1. 초거대 AI 모델 적용 전 후 객체 탐지 성능 변화 분석

(1) 목적: 초거대 AI 모델의 적용 전 후 성능 검증

(2)수행 내용

- 객체 탐지, 분류, 분할 모델의 상황 별(모델의 버전, 객체의 종류, 영상의 센서 등) 성능을 쉽게 파악하기 위한 시각화 및 대시보드 개발
- 데이터 시각화 툴 Datalore 활용하여 모델의 버전, 객체의 종류 등을 드롭박스로 선택하여 성능을 확인할 수 있도록 대시보드로 개발함

(3)성과

- 초거대 모델을 적용했을 때 성능이 향상되었음을 증명하고 임원진이 초거대모델에 투자 여부를 결정할 수 있는 의사결정을 도움

2. 영상 라벨링 프로세스 효율 향상을 위한 사내 라벨링 툴 log 데이터 분석

(1) 목표: 영상 데이터 라벨링 프로세스 비용 절감

(2) 목적: AI 객체탐지모델(인퍼런스 결과물)을 라벨링 Tool에 적용 시 데이터 생성의 비용 절감 가능성을 검증

(3) 수행 내용

- 영상의 객체를 사람이 100% 수작업으로 라벨링하여 학습데이터셋을 구축하는 비용이 매우 높았음
- 라벨링의 요구사항정의 항목으로 인해 AI 객체 탐지 모델을 활용 시 요구사항을 맞추지 못하였음
- 불필요한 요구사항을 제거하기 위해 인공지능연구소, 라벨링 툴 개발팀, MLOps팀, 각 사업부와 합의가 될 수 있도록 커뮤니케이션을 주도함
- 합의된 요구사항정의서에 따라 객체 탐지 모델을 라벨링 툴에 적용 전과 후의 라벨링 속도, 비용을 라벨링 툴의 log 데이터를 추출하여 검증하여 시각화하였음

(5) 성과

- 라벨링 로그데이터 분석하여 프로세스를 개선하는 결과를 도출할 수 있는 역량이 있으며, 특히 이해관계부서와의 커뮤니케이션 능력을 보여줌
- 라벨링의 속도는 250% 향상되었으며 연간 약 8억원에 달하는 비용 감소로 이어짐

3. VLM(Vision-Language Model)모델을 위한 데이터 파이프라인 설계 및 DB구축

(1) 목적: VLM모델에 영상과 텍스트를 Pair한 학습 데이터로 DB화

(2) 수행 내용

- GPU 과부화로 학습이 불가능하여 영상을 적합한 사이즈로 분할
- 분할된 영상과 텍스트가 Pair로 학습될 수 있도록 영상에서 텍스트 데이터 추출
- 분할된 영상을 Nas 스토리지로 저장하고 추출된 텍스트는 PostgreSQL DB로 저장함
- Nas 스토리지로 저장된 영상은 Milvus를 활용, resnet50모델로 이미지를 벡터로 변환하여 벡터 DB로 별도 구축함

(3) 성과

- 인공지능 모델링을 위한 학습데이터를 RGB밴드값, 픽셀, 촬영날짜, 국가, 도시, 좌표값 등 메타데이터를 추출하여 DB로 구축 할 수 있고 이미지 데이터를 벡터화하여 DB를 별도로 구축하여 모델이 유사값을 계산할 수 있도록 준비되었음.

4. AI 모델 학습 및 검증 데이터세트 분석

(1) 목적: 사내 보유한 인공지능 모델의 학습 및 검증 데이터 셋 분석을 통한 모델의 성능 향상 방안 모색

(2) 수행 내용

- 탐지, 분류, 분할 모델의 특정 국가 및 도시, 센서, RGB값 등 모델의 성능에 영향이 미치는 요인을 분석함
- 상관관계분석, 교차분석, 로지스틱 회귀분석, 다중선형회귀분석 등의 분석법을 사용함
- 유의미한 변수를 채택하여 모델 성능을 높이기 위한 데이터 수급 방향을 제시하였음
- 위성영상의 메타데이터 추출(위도, 경도, 좌표계, RGB값 등)하여 속성 정의
- GT 생성 후 AI 모델 인퍼런스 결과 추출 -> 모델 성능 확인
- 각 객체별(항공기, 선박 등) 모델 성능 결과값(f1 score 등) 매칭
- 상관관계 분석 / 교차분석 / 로지스틱 회귀분석 / 다중 선형 회귀분석 진행

(3) 성과

- 모델 성능에 유의미한 변수 파악 후 데이터 세트 재구축 후 모델 재학습을 통해 모델 성능 향상

5. 위성영상 Detection 라벨 데이터 품질 분석

(1) 목적: 사내 보유한 위성영상 객체 탐자 라벨 데이터의 품질을 정량적으로 분석하여 라벨 데이터 생성 및 검수 프로세스의 개선 방향 마련

(2) 수행내용

- [영상 및 라벨 데이터 수집] -> [라벨링 SW에 업로드] -> [검수] -> [검수 전후 SW 로그 데이터 수집]
- > [통계분석] 프로세스 구축
- 위성영상 객체별 라벨 품질을 AI 모델 성능지표인 Recall, Precision, F1-score 를 도입하여 정량적 분석
- 인적, 프로세스적 라벨 품질 영향 요인 분석
- 분석 결과 DB 구축 및 시각화하여 대시보드로 개발

(3) 성과

- 영상의 메타데이터 속성 별 라벨 품질 영향 요인 분석 후 프로세스 개선
- 영상단위의 라벨 데이터 품질을 정량적 분석 결과에 따라 모델에 투입할 데이터 선별하여 AI 모델 성능 개선

울산과학기술원

2021.03 - 2022.05

연구원 (기술경영전문대학원)

1년 3개월

1. 수행 과제

- 과제명: 스마트 해산물류 스타트업 ignite
- 사업명: 실무형 해산물류 일자리 지원사업
- 사업 주체: 해양수산부
- 전담기관 및 참여기관: 울산항만공사 / 울산과학기술원 / 한국정보산업연합회

2. 과제 상세 내용

- 해양수산부 - 울산항만공사 - 울산과학기술원의 스마트 항만물류 지원센터 협력 센터 구축
- 해양수산부 과제 20억원 운용
- 스마트 해상물류 창업오디션 프로그램 기획 및 운영(창업멘토링)
- 빅데이터 분석 경진대회 프로그램 기획 및 운영
- 해양수산빅데이터플랫폼 기획 및 개발 과제 참여

3. 수행 업무

- (주)데이터하우스 창업경험을 바탕으로 창업멘토링 프로그램 운영
- 창업멘토링 참여 기업(학생) 창업 전반의 경영, 시장분석, 아이템 개발, 전반의 멘토링 수행
- 빅데이터 분석 경진대회 프로그램 기획 및 운영
- 경진대회 출품작 심사 및 멘토링 수행
- 과제 예산 운용 및 회계법인 감사 절차 진행

4. 특이사항

- 해상물류 발전에 기여한 공로를 인정받아 울산항만공사 표창장 수여받음

(주)데이터하우스

2018.05 - 2022.04

사내이사 (제품개발)

4년

1. 주요 사업

(1) 사업명

아파트 매물 추천 웹서비스 개발

(2) 주요 업무

- 아파트 매물 추천시스템 Myzib 기획 및 개발
- UI/UX 웹서비스 디자인 기획
- API 연동 주요 구조물(학교, 마트, 지하철역 등 위치데이터) 수집 및 DB구축
- Conjoint, Maxdiff 기법을 활용한 설문조사 및 매물 추천 알고리즘 개발

(3) 보유 역량

- 프론트, 백엔드 개발자와 협력하여 통계기업으로 개발한 추천 알고리즘을 웹사이트에 적용 경험 보유
- 웹서비스의 기능 도출, 서비스 흐름, 유저시나리오, 정보마인드맵 등 웹서비스 기획 역량 보유
- Conjoint, Maxdiff 응용통계기법 알고리즘 설계를 위한 설문조사 설계 역량 보유
- 데이터 수집 및 DB 구축 역량 보유
- 공인중개사 자격증 보유하여 비즈니스 확장가능성 확보

2. 주요 역량

(1) 운영 자금 확보

- 청년창업사관학교 8기 선발 운영 자금 1억원 확보
- 기술보증기금 청년 운영 자금 1억원 확보
- 부산경제진흥원 공유경제기업 선정
- 정부 창업지원금 과제 선정을 위한 사업계획서 및 발표 역량을 보유함

(2) 빅데이터 분석

- 빅데이터 분석 기술을 기반으로 한 아파트 매물 추천 시스템 기획 및 개발
- 추천 알고리즘 개발 후 웹 서비스로 구현 및 배포

(3) 서비스 기획 및 개발 능력

- 시장 분석, 트렌드 분석을 기반으로 한 서비스 및 콘텐츠 기획 능력 보유

(주)나무커뮤니케이션

2012.10 - 2013.10

총괄 팀장 (부산지사총괄)

1년 1개월

1. 주요 업적

(1) 카카오톡, 페이스북 광고 매체 확장

- 네이버, 다음 등 키워드광고 및 배너광고에 집중되어 있던 매체를 SNS 카카오톡, 페이스북으로 확장
- 부산지사의 포트폴리오를 다양화하는 전략으로 추가 매출 및 안정성 확보

(2) 조직운영 및 리더십

- 온라인 마케팅 업계의 고질적 문제인 높은 퇴사율은 조직의 성장을 방해하는 요소로 판단함
- 부산지사 퇴사율 0% 달성 (정규직 기준)
- 텔레마케팅 영업방식을 탈피하여 인바운드 영업 방식 도입
- 사내 문화 재정립 (만성적인 야근, 회식문화)
- 영업부터 광고관리까지 1인이 모두 감당했던 시스템을 광고성과분석에 구글 애널리틱스를 도입하여 영업과 관리를 분리

(3) 목표매출 달성

- 서울 본사로 부터 부여받은 목표매출 매월 초과 달성
- 2012. 09 부산지사 오픈 후 1년내 매출 2억원 달성

(주)이엠넷

2009.03 - 2012.07

주임 (부산지사)

3년 5개월

1. 주요 업무

(1) 온라인 광고 운영

- 네이버 키워드광고 및 배너 SNS 온라인 광고 운영
- 각 분기별 광고성과 분석 후 개선안 제시
- Google Analytics 활용
- 매체, 키워드별 클릭률 및 구매전환율 분석
- SNS 마케팅 콘텐츠 기획 및 개발
- 광고주 경쟁사 분석 후 온라인마케팅 전략 수립

(2) 신규 광고주 발굴

- 신규 광고주 발굴
- 시그니개 여이오 의하 오라이과기 제아서 기획

- 산업공학 전공을 위한 데이터분석에 대한 이해
- 타 경쟁사와 경쟁 PT를 통해 신규 광고주 영입

학력	울산과학기술원(UNIST)	2016 - 2018
	비즈니스분석 / 석사	
	<p>1. 연구 참여 프로젝트</p> <p>(1)프로젝트 명: 수면 습관의 패턴 및 수면 퀄리티 영향 요인 분석</p> <p>(2)수행 기간: 2016.09 ~ 2016.12</p> <p>(3) 과제 형태: 산업체위탁과제</p> <p>(4) 산업체: (주)테일러넬슨소프레스코리아</p> <p>2. 연구 수행 내용</p> <p>(1) 설문조사 설계: 실험 대상으로 부터 수면 습관 및 퀄리티에 대한 설문조사 항목 설계</p> <p>(2) 데이터 수집: 설문조사 응답 데이터 및 가전기기(온습도기, 정수기, 매트리스) 센서 데이터에서 데이터 수집</p> <p>(3) 데이터 전처리: 설문조사 데이터 및 센서데이터의 null, 오류 데이터(설문응답지와 센서 데이터의 상이함)등 보정</p> <p>(4)데이터 분석</p> <p>- 실험대상자의 수면 패턴을 의사결정나무와 군집분석 방식으로 수면 패턴 타입 설정</p> <p>- 각 수면 패턴 타입 별 영향요인을 로지스틱회귀분석, 선형회귀분석, t-검증, f-검증 등 통계기법을 적용하여 가설 검증</p> <p>- 페르소나 생성 후 시뮬레이션 데이터 제공</p>	
	부경대학교	2004 - 2009
	경제학 / 경제학사	
	경제학 전공, 경영학 부전공	
전문분야 및 스킬	인공지능(AI) · 머신러닝 · 데이터 사이언티스트 · Python · 빅데이터 · 데이터분석	
자격증	<p>데이터 분석 준전문가(ADsP) 취득</p> <p>공인중개사 자격 취득</p>	

기타이력	제2회 스타트업 발굴 공모전 우수상_(주)선보엔젤파트너스 울산창조경제혁신센터장상 U2A 창업경진대회 최우수상_울산지방중소벤처기업청장상 K-Global 300 기업 선정_과학기술정보통신부장관 인증 대학원생 창업경진대회 장려상_동국대학교 창업지원단장상 청년창업사관학교 8기 선발_중소기업진흥공단 2020 공유경제 아이디어 공모전 수상 _부산경제진흥원장상 표창상 _ 울산항만공사
------	--

외국어	영어 비즈니스 회화 가능
-----	---------------

웹사이트	https://kimsejung84.github.io/
------	---