

소재우

jaewoo.so22@gmail.com
(82)010-6355-6339
수원, 경기도

요약

8년간 다양한 산업에서 AI 기술 및 서비스를 연구했습니다.

AI를 이론이 아닌 실제 비즈니스적 니즈도 충족되도록 연구 및 서비스 개발을 해왔습니다.

여러 분야의 전문가들과 협업을 하면서 의사소통 능력을 키우고, 니즈를 파악해 실제 사용 가능하도록 한 연구 개발 경험이 많습니다.

경력

ESG 미디어 실시간 분석 시스템

대신경제연구소 / machine learning researcher
2022.06 - 2022.11 (6개월)

[성과]

- 외부 서비스를 위한 머신러닝 서비스 파이프라인 구축.
- Kubernetes 기반으로 시스템 개선. 서비스 관리 인원 70% 감축.

[역할]

- 기존 윈도우 기반 서비스를 쿠버네티스 기반으로 개선.
- 머신러닝 파이프라인 구축하여 Restful 마이크로서비스로 배포.

[획득 역량]

- On-premise 환경에서 쿠버네티스 기반의 머신러닝 서비스 파이프라인을 구축하는 전 과정 경험.
- Kubernetes, Github-Action, 마이크로서비스, Restful API.

ESG 미디어 이슈 분류 모델 연구

대신경제연구소 / machine learning researcher
2021.04 - 2022.05 (14개월)

[성과]

- 머신러닝 기반 분류 모델 개발로 인건비 연간 2억3000만원 절감.
- 실무 및 머신러닝 모델 양쪽에서 유효한 ESG 이슈 클래스 분류체계 개발.

[역할]

- supervised 및 unsupervised clustering 분석으로 multi-label 및 single-label 데이터가 혼합되어 있음을 발견. multi-label 이 되는 데이터가 최소화 되면서 실사용이 가능한 분류체계로 개선. f -beta (beta = 3) score 0.76 → 0.87로 향상.

Tax Loss harvesting 실현 알고리즘 연구

대신경제연구소 / machine learning researcher
2020.09 - 2021.04 (8개월)

[성과]

- BM 대비 평균 1.7% 초과 수익률 달성.

[역할]

- 시장 상황을 클러스터링 후, 각 클러스터별 최적 전략을 찾는 방법으로 기존 전통적이 방법론 대비 수익률 0.8% 초과 달성.
- Tax Loss harvesting 실현 전략에 대해서 전통적 방법, 강화학습 기반, 확률과정 기반의 기법을 각각 비교하여 No free lunch theorem이 적용됨을 보임. 시장 국면별 최적 전략을 예측하는 문제로 관점을 바꾸었다.

Drug Activity & Chemosensitivity 예측 모델 연구

Theragen Bio / machine learning researcher

2019.12 - 2020.07 (8개월)

[성과]

- 국제 바이오분야 대회 드림챌린지 Pancancer Drug Activity, Chemosensitivity 각각 2위
- Cell Reports Medicine에 방법론 등재 (A Community Challenge for Pancancer Drug Mechanism of Action Inference from Perturbational Profile Data, Cell Reports Medicine, 3(1), 100492.)

[역할]

- 다른 성격의 데이터를 통합 사용하기 위해 representation learning 및 dimensional reduction 연구
- Drug response data의 missing value imputation 방법 연구
- Blending, Stacking 등의 모델 앙상블 전략 연구.

비소세포폐암 환자에 대한 PD-1 면역항암제 반응성 예측 모델 연구

Theragen Bio / machine learning researcher

2018.12 - 2019.06 (7개월)

[성과]

- 기존 면역항암제 반응성 예측 모델 (AUC = 0.57) 대비 AUC=0.86으로 정확도 향상.
- 연간 면역항암제 관련 의료비 37억원 (100명 기준) 절감.

[역할]

- 예측에 주요한 피쳐 선정의 consistent를 보장하기 위한 missing value imputation 연구.
- 면역항암제 투여를 위해서 PD-L1검사를 선행한다는 점에 착안, PD-L1값을 샘플 주변 값 추정에 사용해 feature attribution model의 local fidelity를 34% 개선했다. 본 문제에서 Local Interpretable Model-Agnostic Explanation(LIME)을 사용했다. 일반적으로 훈련용 데이터셋 전체에 대해서 around_x 를 추정한다. 그러나 본 문제에서는 면역항암제 사용을 위해 바이오마커 PD-L1을 검사한다. 따라서 이러한 사전 지식을 around_x 를 추정시 사용했다. 즉 PD-L1의 발현여부에 따른 조건부로 확률 분포를 설정하여 around_x 를 추정.
- 예측 모델간 성능 검증 방법 디자인 및 실험 진행.

Photoluminescence 현상을 이용한 웨이퍼 두께 및 조성비 알고리즘 연구개발

Etamax / machine learning researcher

2017.10 - 2018.03 (6개월)

[성과]

- AI 기술 접목으로 경쟁사 검사장치 대비 원가는 20% , 검사 속도는 3배, 성능은 87% 달성.

[역할]

- 고정밀 부품을 저정밀 부품으로 대체. 부족한 성능은 머신러닝 모델로 보완.
- 레이저 파장, 각도, 동선에 따른 머신러닝 모델의 성능 분석.

[적용 포인트]

- 하드웨어의 특성 및 AI기술의 적용 가능 포인트를 종합적으로 고려한 문제해결 능력.

학력

Kyushu University

Master of Engineering

2012.04 - 2015.03

정보과학 전기공학

연구분야 : Multi-agent Reinforcement Learning, Optimization Theory

Kyushu University

Bachelor of Engineering

2008.04 - 2012.03

전기공학 컴퓨터 사이언스

대회

[2nd place] CTD-squared Pancancer Chemosensitivity DREAM Challenge

Issued by Columbia University, Sage Bionetworks, NIH/NCI, CTD2, Mount Sinai School of Medicine, Heidelberg University, Jun 2020

[2nd place] CTD-squared Pancancer Drug Activity DREAM Challenge

Issued by Columbia University, Sage Bionetworks, NIH/NCI, CTD2, Mount Sinai School of Medicine, Heidelberg University, Apr 2020

[2nd place] Gimhae Fire Prediction Competition

Issued by Korea Land and Housing Corporation (LH) · Dec 2019

자격증

Architecting with Google Kubernetes Engine

Google Cloud, Jun 2020

Machine Learning on Google Cloud

Google Cloud, May 2020

논문

Clinical decision support algorithm based on machine learning to assess the clinical response to anti-programmed death-1 therapy in patients with non-small-cell lung cancer Clinical decision support algorithm based on machine learning to assess the clinical response to anti-programmed death-1 therapy in patients with non-small-cell lung cancer

European Journal of Cancer · Jun 25, 2022

특허

A tissue origin prediction device, method of predicting the tissue origin using a genome data, and computer program

1020200076756 · Filed Jun 23, 2020

Methods for providing information about responses to cancer immunotherapy and devices using the same

means for providing information about responses to cancer immunotherapy and devices using the same

10-2021-0081547 · Filed Dec 24, 2019

Languages

한국어, 일본어, 영어