

경력기술서

jaewoo.so22@gmail.com / (82)010-6355-6339 / 소재우, 데이터사이언티스트

경력

LLM 기반 보고서 생성 프롬프트 연구

대신경제연구소 / 데이터사이언티스트

2023.05 - 2023.07 (3개월) 프로젝트 기획, 생성AI 기술 및 프롬프트 연구 수행.

[성과]

- 문장 및 이미지 작성 시간 약 80% 단축.
- 평균 보고서 작성 시간 약 70시간에서 40시간 미만으로 단축.

[역할]

- 인건비 절감을 위한 업무 자동화 프로젝트 기획.
- 내무 업무 내용 및 비용 조사 및 데이터화.
- LLM기반 개인화된 문장, 문구, 이미지 생성 연구개발.

[획득역량]

- 프롬프트 엔지니어링 연구 경험.
- OpenAI API 기반 기능 개발 및 Output 튜닝 경험.
- 사내 전용 마이크로 서비스 구축 경험.

ESG 미디어 실시간 분석 시스템

대신경제연구소 / 데이터사이언티스트

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 파이프라인의 세부 기능들을 각각 마이크로서비스로 배포.

[획득 역량]

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

ESG 미디어 특화 언어 모델 개발

대신경제연구소 / 데이터사이언티스트

2021.04 - 2022.09 (18개월)

[성과]

- 머신러닝기반 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.
- 실제 모델 실험 및 구축. 평가 지표 설계 부터 최종 훈련 및 컨셉 시프트 테스트까지 전 과정 진행.

[획득역량]

- 정확도만이 필요한 학계 상황과는 다른 여러가지 기준을 모델 설계에 적용시킨 경험. (인적자원의 한계로 시간당 샘플 수 대비 precision & recall 의 최적점이 나오도록 설계)
- 전문가의 관점과 비전문가의 관점의 차이를 경험했고, 비전문가와와의 원활한 소통방식과 설득 역량 강화.

Tax Loss Harvesting 서비스의 최적화 모델 설계

대신경제연구소 / 데이터사이언티스트

2020.09 - 2021.04 (8개월)

[성과]

- 2023년 예정된 법 개정이 2025년으로 연기되어 프로젝트 일시중단.

[역할]

- Reinforcement Learning 모델 및 RNN 기반 자산 관리시 세금 최적화 모델 연구.
- 시장 상황을 클러스터링 후, 각 클러스터별 최적 전략을 찾는 방법으로 기존 전통적이 방법론 대비 누적수익률 0.8% 초과 달성.
- Tax Loss harvesting 실현 전략에 대해서 전통적 방법, 강화학습 기반, 확률과정 기반의 기법을 각각 비교 실험.
- 트레이딩 방식을 학습하는 것이 아닌, 시장 국면을 클러스터링 후 클러스터 별 최적 전략을 찾는 방식으로 관점 전환시킴.

[획득역량]

- 시계열 데이터인 주가 데이터, 경제 지표, 기업 정보 등의 여러 데이터로 강화학습 환경 설계 경험.

Drug Activity & Chemosensitivity 예측 모델 연구

Theragen Bio / 데이터사이언티스트

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 위한 representation learning 및 dimensional reduction 연구
- Drug response data의 missing value imputation 방법 연구
- 예측 모델 실험 및 연구.
- Blending, Stacking 등의 모델 앙상블 전략 연구.

비소세포폐암 환자에 대한 PD-1 면역항암제 반응성 예측 모델 연구

Theragen Bio / 데이터사이언티스트

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 consistent 보장하기 위한 missing value imputation 연구.
- 모델 해석 알고리즘 개선 연구. (Local Interpretable Model-Agnostic Explanation(LIME) 방법론에서 샘플 주변부 추정 개선.)
- 예측 모델간 성능 검증 방법 디자인 및 실험 진행.

[획득 역량]

- 의료나 유전자 처럼 코호트 기반의 연구가 많은 분야에서, 통계적으로 도출되는 연구결과를 데이터 기반의 머신러닝 분야에 접목시키는 경험.
- 모델 해석을 데이터 기반이 아닌 도메인 분야 관점에서도 만족시킬 수 있는 프로세스 연구 경험.

Photoluminescence 현상을 이용한 웨이퍼 두께 및 조성비 알고리즘 연구개발

Etamax / 머신러닝 연구원

2017.10 - 2018.03 (6개월)

[성과]

- AI 기술 접목으로 경쟁사 대비 원가 32%로 비용절감 , 검사 속도 3배, 성능 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.

사이드 프로젝트

[ChatGPT + DALL-E 기반 중소기업 SR 리포트 자동 생성 + DALL-E 중소기업 SR 리포트 대화형 생성 시스템]

- report.jaewoo-so.online

[Data Scientist Competition Library]

- https://github.com/jaewoo-so/ML_LIB

[C# Functional Programming Nuget Package]

- <https://github.com/jaewoo-so/SpeedyCoding>

링크

[포트폴리오 사이트] post.jaewoo-so.site

[GitHub] <https://github.com/jaewoo-so>

[SR보고서 생성 데모] report.jaewoo-so.online