

## About

- 모바일 광고 회사의 데이터 사이언티스트로서 통계 분석·머신러닝 모델링 및 데이터 시각화를 수행한 실무 경력
- 고에너지 핵물리 실험 연구를 수행하며 얻은 수학, 물리, 공학에 대한 이해와 빅데이터 통계 분석 경험, 분석적 사고방식 보유
- 국내외 연구원들과 미팅·워크샵·학회 등을 통해 논의하고, 대시보드를 개발하여 회사의 운영부서와 실시간으로 정보를 공유하는 등의 다양한 소통 경험

## Experience

### FSN Cauly

데이터 사이언티스트, 매니저

2022. 4 - 2023. 10 (1년 6개월)

- 다양한 광고 캠페인의 목표 성과 달성 및 예산 소진을 위해 데이터를 분석하고 인사이트 제공
- MySQL, Presto, Hadoop 등을 이용하여 데이터 저장 및 처리
- Python 을 활용하여 데이터 수집·전처리, 통계 분석·머신러닝 모델링, 데이터 시각화 등을 수행
- Streamlit 웹 애플리케이션을 활용한 대시보드를 개발하여 운영 부서와 효율적으로 소통하며 협력

### 고에너지 핵물리 연구실

대학원 석사 및 박사 과정

2016. 3 - 2021. 8 (5년 5개월)

- C++ 및 분산·병렬 컴퓨팅을 활용하여 상대론적 중이온 충돌 실험 빅데이터의 전처리, 통계 기반 분석 및 시각화
- C++를 기반으로 반도체 입자 검출기의 신호 검출 시뮬레이션을 제작
- 유럽핵입자물리연구소(CERN) ALICE 그룹의 반도체 입자 검출기를 제작하는 프로젝트를 국내 및 스위스 현지에서 참여

## Publications

- Kim B, Bok J, Cho J, **Kwon J**, Lee H, Kweon M. A Fast Simulation Tool for Silicon-Pad Detectors. J. Korean Phys. Soc. 77, 635-642 (2020).  
<https://doi.org/10.3938/jkps.77.635>
- **Kwon\* J**, Ey S, Kim Y, Kweon M. Simulation of a Silicon-Pixel Detector. NPSM 2018;68:189-195. <https://doi.org/10.3938/NPSM.68.189>  
✓ 해당 이슈의 하이라이트 논문으로 선정

## Skills

- 데이터 분석: AI (ML/DL), Statistical analysis, Time Series Analysis, Computer Vision, Scikit-learn, Tensorflow, Keras, OpenCV
- 프로그래밍 언어/데이터베이스: Python, C++, Matlab, MySQL, Presto, Hadoop
- 데이터 시각화: Streamlit, Plotly, Matplotlib, Apollo
- 운영 체제: macOS, Linux (CentOS, Ubuntu), Windows
- 협업 및 버전 관리: Git, Jira, Slack, Trello, Notion, Mattermost, Confluence
- 기타: VI, VSCode, Jupyter Notebook, Google Colab, RunDeck

## Education

- 인하대학교 물리학 박사 수료  
2018. 3 - 2021. 8; (학점: 4.17 / 4.5)
- 인하대학교 물리학 석사 졸업  
2016. 3 - 2018. 2; (학점: 3.89 / 4.5)
- 인하대학교 물리학 학사 졸업  
2012. 3 - 2016. 2; (학점: 4.00 / 4.5)

## Certifications

- OPic 영어: IH (Intermediate High)  
2024.03; ACTFL
- 데이터 분석 준전문가 (ADsP)  
2021.12; 한국데이터산업진흥원
- SQL 개발자 (SQLD)  
2021.12; 한국데이터산업진흥원
- TOEIC: 925 점  
2021.10; ETS

## Projects

### FSN Cauly

데이터 사이언티스트, 매니저

2022. 4 - 2023. 10

- 체리피커 검출 모델 개발

2023. 8 - 2023. 9

- 기여도: 100%
- 기술: Python, Machine Learning, Scikit-learn, Statsmodels, Pandas, NumPy, Plotly, Matplotlib, MySQL, Presto, Hadoop, Jupyter Notebook
- 목적: 리워드 성향 유저 트래픽을 제외시켜 광고 효율을 향상시키기 위함.
- 방법: 전체 매체 중 리워드 매체를 식별하고, 관련 매개변수를 정의. Gaussian Mixture 클러스터링 알고리즘을 이용하여 체리피커를 검출. 유저가 체리피커일 확률 또한 계산되도록 하여, 원하는 확률 이상의 체리피커만을 선택함으로써 모수 조정에 용이하도록 개발.

- RTB (Real-Time Bidding) 입찰가 최적화

2023. 6 - 2023. 9

- 기여도: 100%
- 기술: Python, Pandas, MySQL, Presto, Hadoop, Jupyter Notebook
- 목적: 광고주가 광고 인벤토리에 입찰할 때 사용되는 입찰가를 조정하여 캠페인의 효율을 최대화.
- 방법: 사용자의 특성 및 성과 지표를 고려하여 입찰가를 조정하는 알고리즘 개발.
- 성과: 초기값 대비 높은 KPI 로 광고주에게 높은 수익을 제공하고 광고 플랫폼의 경쟁력을 향상시킴.

- 광고 캠페인/그룹의 성과 및 이슈 모니터링을 위한 대시보드 개발

2023. 6 - 2023. 8

- 기여도: 100%
- 기술: Python, Pandas, NumPy, MySQL, Presto, Hadoop, Apollo, Streamlit, Plotly, Jupyter Notebook, RunDeck
- 목적: 광고 캠페인 및 그룹 효율의 추이를 모니터링하고 이슈를 감지하는 대시보드를 개발.
- 방법: Streamlit 웹 애플리케이션을 구현하여 실시간 데이터 시각화를 제공하며, 누적일수 데이터를 선택한 기간에 따라 검토하여 성과 변화를 탐지.
- 성과: 광고 운영 부서에 시각적인 정보를 제공하여 쉽고 빠른 의사결정을 지원하며, 광고 예산을 효율적으로 관리할 수 있도록 함.

- 매체 성과 지표 기준 grade 부여 기능 개발

2023. 1 - 2023. 5

- 기여도: 100%
- 기술: Python, Machine Learning, Scikit-learn, Statsmodels, Pandas, NumPy, SciPy, MySQL, Presto, Hadoop, Apollo, Streamlit, Plotly, Matplotlib, Jupyter Notebook, RunDeck
- 목적: 각 매체의 CTR, CPC, CPM 등 성과 지표별 등급을 부여.
- 방법: 매체별 성과 지표를 평가하여 등급을 부여하는 시스템을 구현. 이 때 표본의 크기에 따른 성과 지표의 신뢰도를 고려하며 등급을 부여. Rundeck 에 해당 작업을 등록하여 매일 갱신된 테이블을 hadoop 에 저장.
- 성과: 다양한 업무에 이 등급을 활용.

## 고에너지 핵물리 연구실

### 학부 및 대학원 과정

2013. 3 - 2021. 8

- 매체 CTR 예측 모델 개발

2022. 9 - 2023. 1

- 기여도: 100%
- 기술: Python, Machine Learning, Regression, Scikit-learn, Statsmodels, One-hot encoding, Pandas, NumPy, Matplotlib, MySQL, Presto, Hadoop, Web scrapping, Jupyter Notebook
- 목적: 신규 매체 및 저노출 매체의 CTR 을 예측.
- 방법: 회사 내부 데이터 및 웹 스크래핑을 통해 수집한 다양한 특성을 분석하고, 여러 머신러닝 모델을 비교하여 가장 성능이 우수한 모델을 선택.
- 성과: 예측한 CTR 을 이용하여 더 효율적으로 광고 예산을 할당하고 캠페인의 성과를 높임.

- 상대론적 중이온 충돌 실험에서의 뷰티 쿼크 생성량 분석

2016 - 2021

- 기여도: 100%
- 기술: C++, Docker, ROOT, Grid, HTCondor, PROOF, VI
- 목적: 유럽핵입자물리연구소(CERN)의 거대 강입자 가속기(LHC)에서 입자의 원자핵을 충돌시켜 빅뱅과 유사한 상태를 재현했을 때 생성되는 뷰티 쿼크 연구.
- 방법: 약 300 TB 의 실험 데이터를 분산 및 병렬 컴퓨팅을 이용하여 분석. 15 억 개 실험 데이터에서 적절한 데이터 전처리를 통해 얻은 50 억 개 이상의 입자들로부터 전자를 분류하고, 그 중 뷰티 쿼크로부터 온 전자의 양을 계산.
- 성과: 실험 데이터 분석 결과를 이론적 결과와 비교하여 타당성을 확인.  
✓ 해외학회발표: Quark Matter 2018, Italy  
(<https://indico.cern.ch/event/656452/contributions/2859719>)

- 반도체 입자 검출기 제작

2016 - 2021

- 기여도: 1%
- 기술: Mattermost, Trello, Google Sheets, Google Docs
- 목적: CERN LHC 의 반도체 입자 검출기를 제작.
- 방법: 국내 및 스위스 현지에서 반도체 칩 성능테스트, 매스테스트, HIC 어셈블리, HIC 성능테스트, QA 등을 수행. 상세한 설명과 troubleshooting 등을 다루는 매뉴얼을 제작하고 매 교대 근무마다 로그북을 작성하여 팀원들과 효율적으로 정보 공유.
- 성과: LHC 에 성공적으로 설치되어 안정적으로 운영 중.

- 실리콘 픽셀 검출기 시뮬레이션

2014 - 2015, 2019 - 2020

- 기여도: 80%
- 기술: C++, ROOT, Silvaco TCAD, VI
- 목적: 사용자가 입력하는 실리콘 픽셀 검출기의 구조에 따른 내부 전기 포텐셜과 전기장을 계산하고, 입자가 검출기에 입사할 때 만드는 신호를 분석하는 시뮬레이션을 제작.
- 방법: C++를 기반으로 제작했으며, GUI 를 이용하여 사용자가 쉽게 이용할 수 있도록 구성.
- 성과: 특정한 구조의 검출기를 대상으로 최적화하여 시중의 시뮬레이션과 같은 결과를 약 10 배 더 빠르게 도출.  
✓ 수상: 한국물리학회 2015 봄 학술논문발표회 학부생 작품발표회 부문 장려상  
([http://www.kps.or.kr/content/community/post\\_view.php?bt=2&post\\_id=37](http://www.kps.or.kr/content/community/post_view.php?bt=2&post_id=37))

- 다수의 조교 경험

2013 - 2021

- 기여도: 100%
- 기술: C++, Matlab, PPT
- 대학물리학, 일반수학, 수리물리학, 고체물리학, Matlab 실습, 고에너지 핵물리 연구 실습 등
- 성과: 청자의 지식 수준에 따른 소통 능력을 기름.