

데이터 분석가 이종원

✉ lee.jongwon.dev@gmail.com

🐱 github.com/jongwoniee

📄 jongwon.net

👜 jongwon.net/portfolio

Summary

미국에서 데이터 사이언스를 전공하며, 머신러닝과 딥러닝 등 이론적인 토대를 쌓았습니다. 또한, 다양한 프로젝트에 참여하여 실전적인 분석 경험을 쌓았습니다. 예를 들어, 부동산 가격 예측 프로젝트에서 대규모 데이터를 효과적으로 처리하고, 데이터의 다양한 특징을 분석하여 예측 성능을 높이는 기술적 역량을 키웠습니다. 또한, 미국 저소득층의 세무 서비스 문제를 해결하기 위해 세액 공제 정책의 효과를 계산하는 Python 모델을 개발했습니다. 최근에는 인기 RPG 게임의 데이터를 분석하는 유튜브 채널을 운영하며 사냥터와 아이템의 효율성을 정량적으로 평가하고 추천하는 콘텐츠를 제작했습니다. 데이터 수집, 실험 설계, 결과 분석 과정을 통해 사용자 맞춤형 정보를 제공하며 채널은 최근 한 달간 조회수 10만 회, 순 시청자 3만 명을 기록했습니다.

Skills

Data Science/Analysis

Python, TensorFlow, PyTorch, Keras, NLP

Diffusion, OpenCV

Tableau, SQL, Google BigQuery/Analytics/Chart, Pandas, D3, Paraview

Software Engineering

RoR, JS, Redis, AWS, RESTful API, React, Swift, Heroku, HTML/CSS

Others

Git, Docker, Unity3D, R, MIS, English (Native)

Work Experience

데이터 분석가 | [PolicyEngine](#) | United States

저소득층과 일반인들을 위한 공공정책 혁신

2023.04 - 2023.11

- **문제 정의 및 데이터 수집:** 미국 저소득층이 세무 업무를 수행하는 데 어려움을 겪는 구조적 문제와 주별 상이한 세제 정책 및 빈번한 개정으로 인한 불확실성을 데이터 기반으로 정의. 이를 해결하기 위해 공공 웹사이트에서 관련 데이터를 수집 및 정제.
- **데이터 모델링 및 분석:** 저소득층 자녀세액공제 정책의 세금 혜택을 소득 구간별로 산정하는 Python 기반 모델을 개발. 이 모델은 복잡한 정책 규정을 수치화하여 개인별 맞춤 혜택을 정확히 계산하도록 설계.
- **경제적 영향 분석:** 정책 개정 및 새로운 제안이 저소득층과 전체 경제에 미치는 영향을 예측하는 모델을 Python으로 구축. 분석 결과를 통해 정책 입안자와 대중이 정책의 혜택과 한계를 명확히 이해하도록 지원.
- **결과 시각화 및 사용자 인터페이스 개발:** 웹 기반 대시보드에서 결과를 시각화하여 데이터 통찰을 효과적으로 전달. 또한, ChatGPT API를 활용하여 국민들이 받을 수 있는 세금 혜택 목록과 예상 수혜액을 쉽게 확인할 수 있는 대화형 기능 구현.

1인 스타트업 | [leetcg](#) | United States

카드게임 컬렉션 리셀링, 유튜브, 플레이 코칭

2020.07 - 2024.07

- **문제 정의 및 데이터 수집:** 일본과 미국의 카드 게임 컬렉션 시장에서 발생하는 시세 차이와 정보 격차 기반의 수익화 가능성을 발견. 이를 위해 각 국가의 주요 거래 플랫폼(eBay, Mercari, TCGplayer 등)에서 카드 게임별 시세 데이터를 **Python, Selenium** 등을 활용해 수집.
- **데이터 분석 및 가격 최적화:** **Pandas**와 시각화 라이브러리를 사용하여 시장 데이터를 분석, 국가별 가격 변동성과 수익률을 계산. 카드 게임의 인기 요소, 트렌드, 배송비 등을 고려하여 최적의 판매 타이밍과 거래 전략을 도출. 연간 X달러 매출 규모와 50% 이상의 마진율을 달성.
- **브랜드 구축 및 디지털 마케팅:** 유튜브 채널과 인스타그램 계정을 운영하며 카드 게임 플레이 코칭 및 리셀링 관련 정보를 제공하여 신뢰 구축, 이를 바탕으로 신규 고객 유입 및 거래량 증대를 실현. 미국 시장의 거래 수수료를 최소화하기 위해 직접 거래 방식을 적극 활용.

인턴 | 플랫폼사업부 | 스마일게이트

스토브라는 자체 커뮤니티 유저 활동 증가

2017.06 - 2017.08

- **문제 정의 및 데이터 수집:** 스토브라는 게임 플랫폼/커뮤니티의 DAU 증가를 목표로 사용자 활동 데이터를 분석. 유튜브, 트위터, 구글, 스토브 커뮤니티 데이터를 수집하여 사용자 간 대화의 주요 주제와 트렌드를 모니터링하기 위해 키워드 추출.
- **데이터 분석 및 제안:** NLP 분석을 토대로 커뮤니티 내 인기 있는 주제와 사용자 관심사를 도출하고 플랫폼 활성화를 위한 전략적 방향성을 제안
- **해커톤 및 서비스 개발:** 인턴 프로젝트 경험을 확장하여 약 한 달 뒤 진행된 해커톤에서 **AWS**의 클라우드 기능과 **Ruby on Rails**를 접목한 키워드 분석 웹 서비스를 개발. SNS 및 커뮤니티 데이터를 실시간으로 추출, 분석, 시각화하여 활용도를 높인 서비스로 **AWS Award** 수상.
- **리더십 및 팀워크:** 프로젝트를 리드하며 데이터 수집, 분석, 웹 서비스 개발 과정에서 팀을 조율. 백지상태에서 시작한 프로젝트를 완성하며 도전적이고 창의적인 문제 해결 능력을 발휘.

풀스택 개발, 프로젝트 리드 | **SKOOLBUS** | 성균관대학교

대학교 셔틀버스 위치추적 앱 개발

2017.11 - 2019.03

- **문제 정의 및 설문조사:** 성균관대학교 셔틀버스 위치 추적 앱의 부재로 인해 학생과 교직원이 비효율적으로 등하교하는 문제를 발견. 설문조사를 통해 문제의 심각성을 수치화하고, 교직원과의 논의를 통해 앱 개발의 필요성을 제안.
- **데이터 수집 및 API 설계:** 협력업체와 협업하여 셔틀버스에 GPS 수신기를 설치, 위치와 시간 데이터를 포함하는 API를 설계. 이를 통해 실시간으로 버스의 위치 및 예상 도착 시간을 제공.
- **앱 개발 및 출시:** **Ruby on Rails, AWS, Java, Swift** 등을 활용하여 셔틀버스의 실시간 위치를 시각화하고 예상 소요 시간을 표시하는 앱을 개발. 재학생과 교직원의 60% 이상이 다운로드하며 높은 사용률을 기록.
- **서비스 유지 및 확장:** 앱 유지보수와 팀 멤버 선발, 지원금 조달 및 관리, 앱 퍼블리싱을 주도하여 서비스 안정성을 확보. 2년 이상의 안정적인 운영 후 추가 유저 확보를 위해 미세먼지 및 날씨 정보를 제공하고 간단한 커뮤니티 기능을 추가.
- **리더십 및 프로젝트 관리:** 프로젝트를 리드하며 기술적 도전과 팀 관리를 성공적으로 수행. 사용자 중심의 문제 해결과 서비스 확장을 통해 커뮤니티의 요구를 반영한 성과 창출.

백엔드 엔지니어 인턴 | [Classlion](#) | 멋쟁이사자처럼

강의평가 웹 서비스

2016.10 - 2017.03

- **시장조사 및 기획:** 강의평가 웹 서비스의 필요성을 확인하기 위해 시장조사를 수행하고 서비스의 핵심 기능을 기획. 사용자의 요구를 분석하여 구체적인 서비스 설계 방향을 도출.
- **백엔드 및 서버 개발:** **Ruby on Rails**와 **AWS**를 활용하여 서버 및 백엔드를 개발. 검색 기능 구현과 데이터베이스 관계 설정을 통해 효율적인 데이터 처리 구조를 구축.
- **팀 구성 및 관리:** 동아리 내 네트워킹을 통해 4인의 팀을 구성하고 프로젝트를 리드. 역할 분담 및 프로젝트 일정 관리를 통해 효과적인 협업을 이끌어냄.
- **결과 보고 및 평가:** 완성된 결과물을 동아리 대표에게 보고하며 성과와 향후 개선 방향을 제시. 프로젝트 경험을 통해 서비스 설계, 개발, 팀 관리 역량을 강화.

Education

학력

Data Science MS, Indiana University

2020.08 - 2022.12

- 주요 이수 과목: Data Science Capstone, Computational Linguistics, Search
- GPA: 3.94 / 4.0

Scientific Visualization, Human Robot Interaction, and more

Economics BA, Informatics Minor, Indiana University

2018.08 - 2020.05

- 주요 이수 과목: Usable AI, Data Visualization, Information Infrastructure I, II
- GPA: 3.8 / 4.0 (Informatics 과목)

글로벌경제학 학사, 성균관대학교, 삼성전액장학생

2013.03 - 2020.05

게임 데이터 분석 - 레벨에 따른 사냥터 추천

2024.11 - 2024.12

목표 및 문제 정의: 레벨, 스펙, 직업별 적합한 사냥터를 시기와 유행별로 파악하기 어려움. 디스코드 파티 구인글의 비정형 텍스트에서 맵, 직업, 레벨, 스펙을 추출하여 데이터 분석을 통해 [유튜브](#) 시청자에게 맞춤형 사냥터 추천.

주요 수행 내용: 비정형 텍스트를 추출해 가공, 분석, 시각화.

데이터 수집 및 전처리: 디스코드의 구인글 메시지에서 필요한 정보를 정규 표현식을 활용하여 추출.

- 이상치 처리: 잘못 수집된 데이터를 다른 데이터들의 평균값과 비교해 제거. 타당도 (Validity)를 평균의 20%로 정의. t-SNE 군집으로 레벨 직업 맵 조합이 상위 0.1% 이상치 처리.
- 결과 시각화: 분석 결과를 matplotlib을 이용해 시각화하여, 시청자에게 데이터의 패턴과 인사이트를 직관적으로 전달.
- 모델링:
 - Linear Regression으로 직업별 레벨에 해당하는 기대스펙을 공식화.
 - Random Forest, Gradient Boosting (e.g., XGBoost, LightGBM) 등의 분류 알고리즘으로 맵 추천 시스템 개발.
 - t-SNE를 이용한 차원축소 2D 시각화.
 - K-Means를 이용한 사용자 클러스터링.

결과 해석 예시:

- 허밋이라는 직업을 예시로 스펙 = $22.34 * \text{레벨} - 500$ 이라는 관계를 도출.
- 130레벨 허밋은 망용둥이라는 맵이 주력 사냥터임에 이에 대비 권유.

결과 분석 및 성과:

- 레벨이나 스펙, 직업을 인풋으로 최적 사냥터를 추천 가능.
- 시각화로 사냥터의 인기 정도, 직업 및 레벨별 분포를 확인 가능.

추가 개발 예정 사항:

- 셀레니움 등으로의 수집 자동화로 데이터 증량
- 데이터 증량 이후 다양한 머신러닝 기법 적용으로 추천 시스템, 예측 시스템 등 추가 개발
- 지피티 연동으로 사용자들과 상호 작용 가능한 서비스 제공

House Price Prediction Using Machine Learning: Hendricks, Hamilton, Tippecanoe of Indiana, United States

미국 인디애나 주 부동산 가격 예측

2022.08 - 2022.12

목표 및 문제 정의: 6년간의 주택 거래 데이터를 활용하여 주택 가격에 영향을 미치는 주요 요인을 식별하고, 내부적 특성 (면적, 방 개수 등)과 외부적 특성 (학교 접근성, 공항 거리 등)을 분석해 예측 모델 개발.

주요 수행 내용: 다양한 소스의 데이터를 직접 수집, 통합하고 고성능의 모델을 개발.

- 데이터 수집 및 통합:
 - Zillow, Redfin 등의 상업 부동산 웹사이트 데이터를 BeautifulSoup, Selenium 등의 웹 크롤러로 수집하려 했으나 정책, 사이트의 보안상 이슈로 수집의 한계점 발견.
 - Google Places, Melissa, Niche.com, Stats Indiana Data 등에서 API사용, 유료 데이터 구매로 6년간의 부동산 거래 사용.
- 모델링:
 - Regression, SVM, XGBoost, CatBoost, LGBM 등 머신러닝 알고리즘을 활용해 예측 모델링 수행
 - SHAP (Shapley Additive Explanations) 시각화를 통해 모델 결과 해석 및 주요 인사이트 도출.
 - 베스트 모델 선정 후 베이지안 최적화로 하이퍼 파라미터 튜닝.

결과 해석:

- 내부적 특성 중 면적, 건축년도 외부적 특성 중 학교 접근성, 일자리 등이 주택 가격에 중요하게 작용.
- Melissa 출처의 데이터가 가장 성능이 좋음.
- XGBoost가 최적의 모델로, MAPE 11.53% 달성.
- 주택 가격은 면적, 건축년도와 같은 내부적 특성과 학교 접근성, 일자리 분포 등 외부적 특성 간의 복잡한 비선형 관계를 포함. XGBoost는 이러한 비선형 관계를 효과적으로 학습하여 높은 예측 성능을 발휘.

결과 분석 및 성과:

- 여러 출처에서 수집한 데이터 중 고성능을 내는 데이터 출처를 선별하여 데이터 구입 자금 투자방향을 제시.
- 이후 같은 프로젝트를 인디애나 주 전체에 확장.

- 비즈니스 대학에 상업적 목적으로 활용 권유.

Evaluating Textual Features and Oversampling for Automatic Stance Detection

자동 스탠스 감지를 위한 텍스트 피쳐 분석과 오버샘플링

2021.08 - 2021.12

목표 및 문제 정의: Twitter 데이터를 활용하여 특정 주제나 대상에 대한 스탠스 (찬성, 반대, 중립)를 자동으로 분류하는 모델 개발. Bag-of-Words (BoW), 감정 사전, 구문 분석 등 다양한 텍스트 특성의 효과를 평가하고, 클래스 불균형 문제를 해결하기 위한 방법론 탐구.

주요 수행 내용:

- 데이터 전처리 및 특징 엔지니어링:
 - 2016 SemEval 데이터셋 (5개의 타겟, 5,000+ 트윗) 사용, URL 제거, 해시태그 처리, 구문 분석 등 전처리 수행.
 - BoW, 감정 사전, 구문 의존 관계 (Bag-of-Dependencies), Syntactic n-grams 등 다양한 텍스트 특징 정의 및 실험.
- 모델링:
 - Support Vector Machine (SVM) 기반 분류기 개발.
 - 5-폴드 교차검증으로 최적의 하이퍼파라미터 튜닝 (예시: 정규화 매개변수, 커널 종류).
 - SMOTE 기법을 활용해 클래스 불균형 문제 해결.

결과 해석:

- BoW와 n-grams가 가장 효과적인 텍스트 특징으로 확인.
- 감정 사전과 구문 분석 특징 추가 시 일부 타겟에서 성능 향상 확인.
- SMOTE 적용 시, 데이터 불균형 타겟 (예시: Climate Change)에 대한 예측 성능 개선.
- BoW와 감정 사전 특징 조합으로 Macro-averaged F1-score 0.57 이상 달성 (일부 타겟 기준).

결과 분석 및 성과:

- 클래스 불균형 해결 및 데이터 품질 개선을 통해 모델의 실제 적용 가능성을 높임.
- 추가적으로 데이터 증강 및 특성 확장을 통한 개선 가능성 확인.

3D Customer Segmentation Visualization using Paraview

Paraview를 활용한 3D 고객 세분화 시각화

2021.08 - 2021.12

목표 및 문제 정의: 3D 시각화 기법을 활용하여 고객 세분화 인사이트를 향상시키고, 고객 분류에 영향을 미치는 주요 요인을 식별. 머신러닝 기법과 2D시각화 분석의 한계를 3D로 확장하여 추가적 인사이트 도출.

주요 수행 내용:

- 데이터 분석 및 시각화:
 - Python과 머신러닝 기법을 사용하여 8,000개 이상의 고객 행동 데이터를 전처리하고 분석.
 - 3D 시각화해 인사이트를 분석하고, Sketchfab을 통해 인터랙티브하고 접근성이 우수한 3D 모델을 임베드.
 - 결혼여부, 나이 등이 유의미한 변인이었으나 성별 경력 등은 상대적으로 유의미 하지 않음을 3D 시각화 모델로 설명.

결과 분석 및 성과:

- 마케팅 팀과 이해관계자들이 2D 시각화나 머신러닝에 의존한 분석보다 직관적인 데이터 기반 인사이트를 제공.

Awards & Recognition

아마존 웹 서비스 | 파워유저 어워드 | 대상 (Best 3)

AWS를 활용한 게임 커뮤니티 자연어 처리 분석 및 시각화 2017.08

스마일 게이트 | 게임잼 | 대상 (Best 3)

유니티 엔진을 활용한 3D 퍼즐게임을 3일간 제작, 수상 후 추가 개발 2017.03

프로그래밍 동아리 멋쟁이 사자처럼 성균관대학교 지부 회장

회원 모집, 중앙 동아리와 소통, 교육 콘텐츠 개발, 실전 프로젝트 리드 2017.03

씨스쿨 프로젝트 | 성균관대학교 | 대상

인디영화 관객수 예측을 위해 데이터 수집, 회귀분석 2017.03

멋쟁이 사자처럼 | 해커톤 | 5위 (200팀)

영어 자격증 시험 어학원 데이터를 수집 후 모아주는 웹 서비스 개발 2016.09

씨스쿨 프로젝트 | 성균관대학교 | 최우수상

데이터 분석을 통한 단말기 유통법 (단통법) 시행 효과 분석 2016.08

융합기초프로젝트 | 성균관대학교 | 대상, 특허출원

화재 조기 진압을 위한 해결책을 EV3로 제작한 로봇으로 시연 2016.08