PORTFOLIO

Data Scientist

김서현

seohyun03@hanyang.ac.kr

김서현 Kim SeoHyun

EDUCATION
명지대학교 문헌정보학 학사 (2014.03 – 2019.08)
한양대학교 비즈니스인포매틱스 석사(2021.09 – 2023.08)

WORK EXPERIENCE
미국 University of Michigan의 Asia Library 인턴 근무 (2020.02 - 2020.12)

PUBLICATION
메타버스 연구동향 및 대중인식 비교분석: 토픽모델링을 활용하여

AWARD
제 2회 K-인공지능 제조데이터 분석 경진대회 우수상 수상

CERTIFICATE
ADsP
SQLD
빅데이터분석기사

SKILL
Python, Scikit-Learn, Tensorflow, Pytorch, PostSQL, Git, Docker

METHODOLOGY
KoBERT, DistilBERT, ResNet, EfficientNet, TabNet, SENet, K-means, LDA, BERTopic(SentenceBERT), LSTM, DNN, XGBOOST

핵심역량

커뮤니케이션

데이터 기반
인사이트 창출

문제해결력 및
지속적인 새로운 모델 학습

목차

1.1. 메타버스에 대한 연구동향 및 대중인식 비교분석

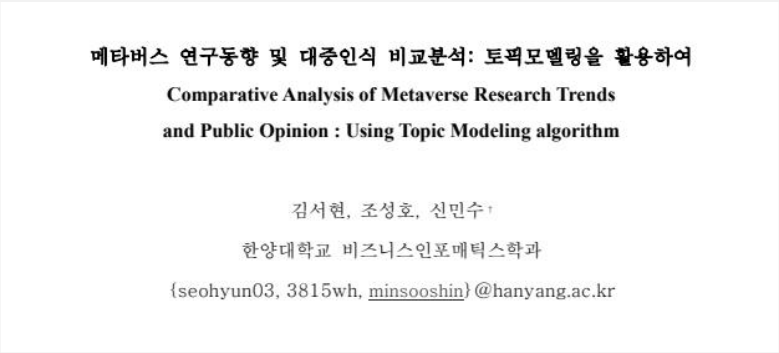
프로젝트 요약	코로나 이후 메타버스에 대한 연구동향과 대중인식을 비교 분석하여 학계 및 산업계에 정보 제공
수행기간	2022.12 - 2023.07
프로젝트 형태	팀 프로젝트
참여인력	2명
참여도	데이터 수집 및 모델 구현 60%, 선행연구 리서치 및 논문 작성 60%

<역할>

- 메타버스 선행연구 논문 리서치 및 논문
- 트위터 데이터 크롤링
- 데이터 전처리 및 BERTopic 모델 구현

<연구 배경>

- COVID-19 사태 안정화 및 거리두기 조치 해제 이후 대중들의 관심은 감소하였으나, 이에 대한 연구들은 지속적으로 늘어나고 있으며 여전히 산업계에서 연구개발 진행되고 있음
- 선행연구에서는 대체로 공급자 관점에서 연구가 진행되어 사용자 측면의 고려가 상대적으로 부족
- 메타버스에 대한 사용자들의 일반적 인식을 확인하고, 이를 반영하여 메타버스 연구가 대중과 지나치게 괴리되는 것을 방지하고 적절한 연구주제 및 방향성 정립 필요



메타버스 연구동향 및 대중인식 비교분석: 토픽모델링을 활용하여

Comparative Analysis of Metaverse Research Trends and Public Opinion : Using Topic Modeling algorithm

김서현, 조성호, 신민수†

한양대학교 비즈니스인포매틱스학과

{seohyun03, 3815wh, minsooshin}@hanyang.ac.kr

<수행과정>

- 메타버스 관련 선행연구 리서치
- **트위터 2021년-2022년 국문 트윗데이터**와 RISS의 학술문헌 데이터를 **크롤링**
- SNS와 학술문헌의 글쓰기 방식의 각기 다른 특징을 고려하여 **트위터 데이터는 semantic gap을 고려하는 SentenceBERT를 기반으로 하는 BERTopic으로 모델을 구현**하였으며, RISS데이터는 LDA로 모델을 구현
- 학술연구동향과 대중인식 간의 공통점과 차이점을 확인하여 **논문을 작성**하였으며, 기존 연구와 다르게 2021년과 2022년의 시간의 흐름에 따라 변화하는 각각의 인식 확인

1.1. 메타버스에 대한 연구동향 및 대중인식 비교분석

<모델 선정 근거>

- 연구동향 데이터 경우 정제된 글과 단어를 사용하는 학술문헌 글의 특성상 semantic gap 의 영향이 제한적임을 고려하여 LDA를 적용함
- 대중인식 데이터 경우 짧고 단편적인 글을 올리는 소셜미디어의 특성으로 인한 semantic gap 을 고려하여 BERTopic 적용
- 형태소로 구성되어 있는 한국어 텍스트에 맞게 BERTopic에서 기본으로 사용하는 Countvectorizer 대신 TF-IDF vectorizer로 대체

<연구 결과>

학술연구동향 및 대중인식 간의 공통점

- 2021년 : 주로 메타버스 활용이나 일반적 개념에 관한 논의 진행
- 2022년 : 메타버스 세계에서 발생할 수 있는 부작용에 대한 논의 활발

학술연구동향 및 대중인식 간의 차이점

학술연구 동향

- 메타버스 공간 상에서 문화생활, 상호작용과 같이 메타버스를 통해 일상 대체하는 연구 진행
- 저작권, 소비자 이슈 등 경제적 차원의 문제에 보다 집중

대중인식

- 메타버스-가상화폐 기술 간의 융합
- 메타버스시티 등 메타버스를 활용한 신기술 도입 및 혁신
- 성범죄와 같은 윤리적 문제에 집중

<연구 의의>

- 기존 연구에서는 고려되지 않았던 '대중인식'에 주목하여 메타버스가 대중과 지나치게 괴리되는 것을 방지
- 메타버스 연구에 대한 공급자 관점 및 사용자관점에서의 이해도 증진
- 일정기간의 데이터를 한 번에 분석한 선행연구와 달리 연도별 분석 수행하여 시간에 따른 변화 확인
- 학술문헌과 소셜미디어 각 텍스트의 특징에 맞는 방법론 사용

<프로젝트 성과>

- 대한산업공학회, 한국경영과학회 공동 춘계학술대회 참여 (1저자)
- **한국경영과학회 학술지 논문 출판 (1저자)**
- 프로젝트 수주 기여
- 연구실 내에서 석사생은 달성하기 힘든 논문 1저자 목표 달성
- 논문 출판 과정 전반 참여하면서 커뮤니케이션 능력 향상

1.2. 석사학위논문

프로젝트 요약	웹 UI 디자인의 색상조합에 관련한 사용자 평가 연구
수행 기간	2021.12 - 2023.06
프로젝트 형태	개인 프로젝트
참여인력	1명

<역할>

- UI 디자인에 관한 사용자평가 선행연구 리서치와 색상조합 선행연구 리서치
- 보색 조합, 삼원색 조합을 반영한 새로운 변수 고안
- ResNet, SENet과 TabNet 하이브리드 모델 구축

<연구 배경>

- 브랜드의 디지털 채널은 고객의 긍정적인 경험과 브랜드와의 관계에 있어 중요한 역할을 함
- 모바일 어플리케이션 및 웹사이트를 구성하는 다양한 요소 중 시각적 디자인과 사용자 만족도 사이에는 강한 관계 있음
- Deep feature만을 단독으로 사용하는 선행연구의 단점 보완하는 연구 필요
- 디자인 요소로서 색상은 적절한 색상 조합은 매력적이지만, 동시에 시각적 복잡성은 서비스환경에 대해 느끼는 매력을 감소시키도 함
- 선행연구에서는 색상의 조합들에 따라서 달라지는 인식을 반영하지 않음

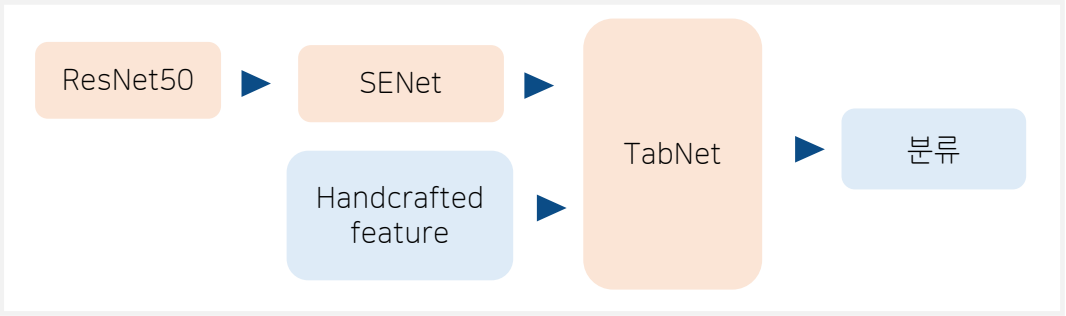
<수행과정>

- UI 디자인에 관한 사용자평가 선행연구 리서치
- 색상조합 관련 선행연구 리서치
- 선행연구 변수 중 색상의 하위요소에 해당하는 채도, 휘도 등을 제외하고, 전체 색상조합을 반영하기 위해 가장 많이 나타나는 색상의 RGB와 전체적 RGB의 평균으로 대체
- 색상조합 중 보색과 삼원색 조합을 반영한 새로운 handcrafted 변수들을 제안

1.2. 석사학위논문

- 사용자가 가장 주목하는 위치의 색상조합 반영을 위해 attention을 활용하여 deep feature 추출하여 활용
- TabNet에서 feature selection 이후에 발생하는 정보손실 문제를 보완하기 위해 각 decision step의 ReLU를 ELU로 튜닝하여 활용

<연구 프레임워크>



<모델 선정 근거>

- RESNet: Attention인 SENet과의 결합시에 시너지효과가 좋음. 실험에서 가장 우수한 결과가 도출된 RESNet 50을 채택함
- SENet: 색상조합에 효과적으로 집중하기 위해 channel 간의 관계를 잘 이해하여 사용
- TabNet: 다양한 modality를 이해하는 모델로, 이미지에서 추출한 데이터를 활용하는 본 연구에 적절하다고 판단

<연구 결과>

Model	Accuracy	F1-score
제안 연구모델	0.643	0.638
ResNet50	0.464	0.317
제안모델 (handcrafted feature 제외)	0.488	0.466
제안모델 (TabNet 변형없이)	0.524	0.517

<연구 의의>

- 모델의 강건성을 위한 handcrafted feature와 deep feature를 활용한 상호보완적 변수 구성
- 기존 Deep learning 사용자평가 연구에서는 feature가 기하급수적으로 많아 모델의 과적합 확률이 높아짐
- 따라서 feature selection을 통해 중요한 feature들만을 활용하도록 하여 선행연구의 단점 보완
- UI디자인에서 적절한 색상조합구성을 찾기 위해 가산혼합의 보색, 삼합색 조합 반영한 파생변수 handcrafted feature로, Attention을 사용하여 사용자가 집중하는 위치의 색상조합 반영한 deep feature 구성

<연구 성과>

- 새로운 변수 제안하여 학술적 기여
- 해외저널 출판 준비
- 이미지 연구에 대한 이해도 향상

1.3. Analysis of Sentiment Analysis Research Trends using Text Mining

프로젝트 요약	감성분석연구 동향을 종합적으로 탐색
수행 기간	2022.10 - 2023.01
프로젝트 형태	개인 프로젝트
참여인력	1명

<역할>

- 선행연구 리서치
- 구글 스칼라 및 스코퍼스 데이터 크롤링
- DistilBERT와 K-means를 활용한 클러스터 모델 구축

<연구 배경>

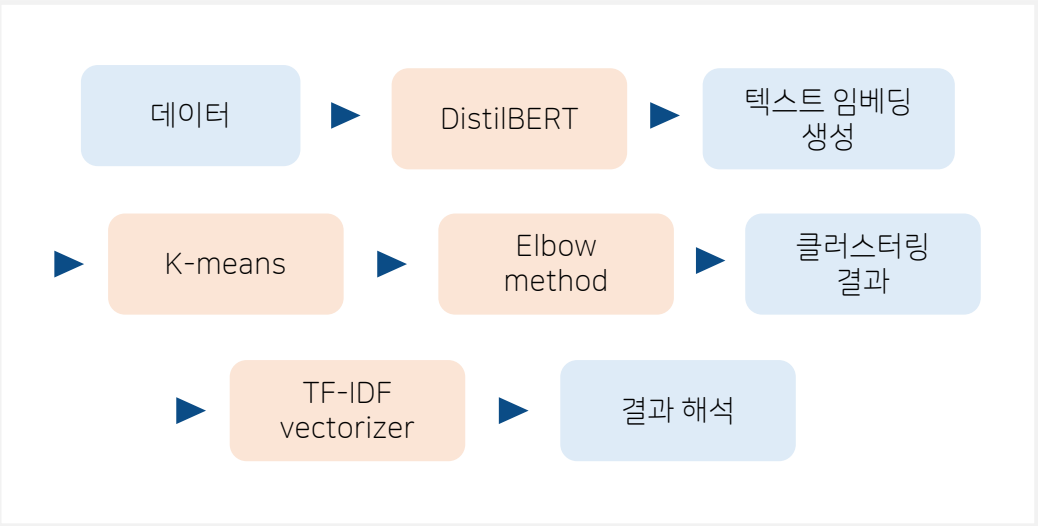
- 텍스트 마이닝은 다양한 분야에서 분석도구로 사용되었지만, 텍스트 마이닝 자체에 대한 연구 부족
- 소셜미디어의 성장으로 감성분석의 중요성은 강조되며, 그 영향력이 지속적으로 증가하기 때문에 감성분석은 지속적으로 연구 가능성 높음
- 감성분석 그 자체에 대한 범위가 넓어 개인 연구자들이 연구흐름을 파악하기 어려움
- 선형연구들은 대체로 연구 방법론과 모델에 대하 연구로 치우쳐져 방법론과 주제 간의 관계 등에 대한 부분이 드러나지 않음

<수행과정>

- 구글 스칼라와 스코퍼스에서 'sentiment analysis'를 키워드로 하여 논문 제목 크롤링
- 데이터 EDA
- DiStilBERT 및 K-means 구축
- 연구결과 확인 및 해석
- 논문 작성 및 Techrxiv투고

1.3. Analysis of Sentiment Analysis Research Trends using Text Mining

<연구 프레임워크>



<모델 선정 근거>

- DistilBERT: 데이터의 수와 시간적 여유 부족으로, 기존 BERT 대비 약 97% 성능을 내는 DistilBERT 선택
- K-means: 텍스트 임베딩 벡터를 고려하여 유사도를 코사인유사도로 변경 가능하여 채택
- TF-IDF: 군집에서 전치사 등을 제외하고 유의미한 결과 도출을 위해 시행

<연구 결과>

- 새로운 연구 대상 텍스트를 찾는 시도 필요함
- 자연어 처리에 국한되어 있던 감성분석 연구가 음성, 비디오 등 다양한 모달리티와 피쳐들을 활용하는 연구로 진화되고, 비중이 증가됨을 확인
- 특히 하나의 모달리티에 집중하여 별개의 연구영역으로 보는 것이 아니라 그에 대한 상호작용에 대한 연구 시작
- COVID-19와 같은 특정 사건 및 기점에 따른 감성변화 분석이 하나의 연구방향성을 형성함을 확인

<연구 의의>

- 연구방법론에 집중한 선행연구들과 달리 주제와 방법론을 종합적으로 살펴봄
- 감성분석 연구 내의 복합적으로 연결된 주제 간의 관계를 고려

<프로젝트 성과>

- 관련 수업에서 A+로 우수한 성적 거둠
- Techrxiv에 투고하여 preprint 논문 발행
- 스스로 연구문제를 정의하여 진행하고, 논문화하는 작업 학습

2.1. KT 골목경제 부활 프로젝트

프로젝트 요약	지역 활성화를 위해 데이터를 기반으로 소상공인 고민 해결 및 발전방안 코칭
수행기간	2022.10 - 2023.01
프로젝트 형태	팀 프로젝트
참여인력	5명
참여도	분석 및 관련 솔루션 제안 40%, 모델 구현 및 신메뉴 제안 70%, 발표안 작성 30%

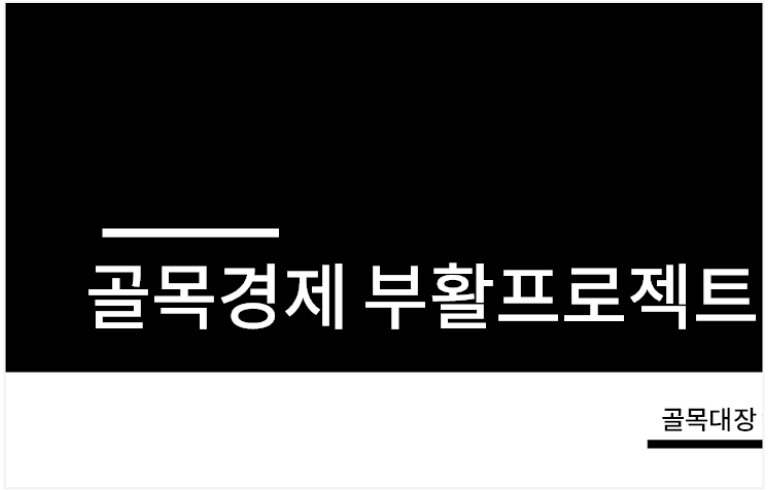
<역할>

- KT 잘나가게 프로그램 및 직접 방문, 관찰하여 유동인구, 지역 매출 및 상권분석
- 브런치, 인스타그램 리뷰 수집 데이터로 LDA 구현
- 분석 결과를 토대로 휴일, 배달서비스 및 신메뉴 카테고리 제안

<수행과정>

1. 잠실 L레스토랑 고민 확인
- 인테리어 변화
- 홍보 성과 향상, 관련 콘텐츠 작성법 가이드
- 신메뉴 기획 고민

2. 문제점 정의 및 분석
- 마케팅 채널 : 활발한 마케팅 활동에 비해서 저조한 성과
- 네이버 및 인스타그램 콘텐츠 문제점 분석
- 인테리어 : 모던함 속의 산만함



2.1. KT 골목경제 부활 프로젝트

상권분석

- 상권분석과 주변 지역 확인하였을 때, 타겟고객층이지만 유동인구비율이 낮은 20대 방문 비율 향상 필요
- 지역 매출을 분석했을 때 남성 매출이 높으나, 그 차이가 크지 않아 데이트라는 특성으로 발생했을 수 있음을 고려해야 함
- 주변 서양식 상권 대비 1회 평균 결제 금액이 비싼 가격

LDA

- 인스타그램, 브런치에서 키워드 '레스토랑'으로, 키워드 '와인, 하이볼, 에이드'로 게시물 크롤링하여 LDA를 진행

3. 분석기반 솔루션 제안

상권기반

- 마케팅 또는 배달을 통한 고객 유치 중요
- 상권 주변 특성 고려하여 아파트 단지내의 평일 고객유치를 위해 유아용 의자 및 아이용 식기의 필요성
- 휴일변경

메뉴

- 샹그리아 및 논알코올 메뉴 제안
- 글라스 와인 판매 제안

<프로젝트 성과>

- 인스타그램 팔로워 약 16% 증가
- 네이버 방문자 리뷰 수 약 50% 증가
- 가장 매출 높던 개점달 대비 26% 매출 증가
- 전월 대비 90% 매출 증가
- 실제 비즈니스에서 데이터 및 기타요소를 복합적으로 고려하여 인사이트 도출하는 능력 향상

2.2. 결함사전대응을 위한 품질오류 예지분석 모델 GLA-3개발

프로젝트 요약	용핵 공정에 발생하는 대량 결함 구간 사전 대응을 위한 예지분석 모델
수행기간	2022.07 - 2022.12
프로젝트 형태	팀 프로젝트
참여인력	3명
참여도	선행연구 리서치 80%, 모델 구현 30%, 보고서 및 발표 자료 작성 70%

<역할>

- 선행연구 리서치
- 데이터 전처리 및 변수 선정
- 변수 유의미성 파악을 위한 DNN모델 구현
- 보고서 및 발표자료 작성

<연구배경>

- 용해공정 불량은 낮은 온도, 속도 등으로 발생
- 1초에도 여러 번 측정되는 제조데이터의 특성 상 불량은 바로 발생하지 않고, 양품으로 측정되어 정상으로 보이나 실상은 이상징조를 내포하는 구간일 확률이 높음
- 이상징조가 누적되어 불량 다량 발생구간 초래
- 불량 누적 구간 방지를 위해 불량 발생 이전 불량 발생확률 예측 필요

<연구 목적>

- 불량 발생 이전 불량 발생확률 예측

<연구 목적>

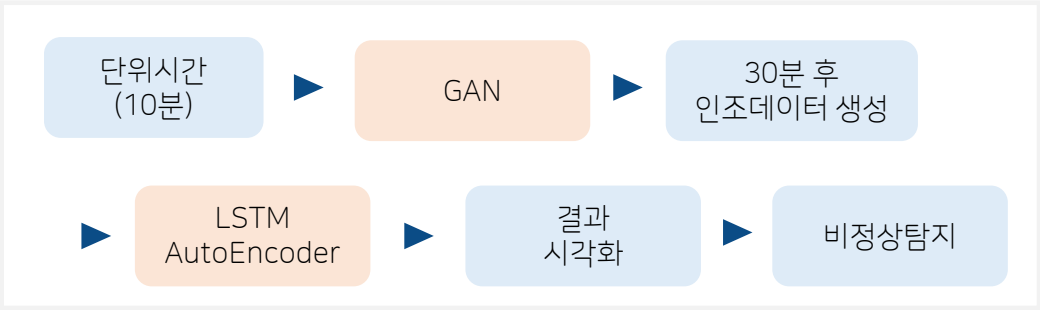
- 불량 발생 이전 불량 발생확률 예측
- 선제적 결함을 예측하여 작업자의 조정가능
- 결함 발생 시 발견하는 기존 분류 모델과 달리 불량 발생을 방지하여 손실비용 절약

2.2. 결함사전대응을 위한 품질오류 예지분석 모델 GLA-3개발

<수행과정>

- **용해탱크 및 용해과정 입자에 관한 선행연구 리서치**
- 데이터 EDA 및 데이터 분석
- **선행연구 및 데이터 특성 기반으로 기획**
- **데이터 전처리 및 변수 선정 및 변수 유의미성을 위한 DNN 모델 구현**
- 단위시간 10분으로 30분 후 인조데이터 생성하는 GAN모델 구현
- 결함발생 위험 구간 특성 학습하여 발생 전조 예측하는 LSTM AE 구현
- 연구결과 확인 및 결과 시각화, **보고서 작성 및 발표자료 작성**
- 발표 프레젠테이션

<연구 프레임워크>



<모델 선정 근거>

- DNN: 1초에 여러 개의 데이터가 존재하는 데이터의 복잡한 특성과 비선형적 관계를 고려하여 선정
- GAN: 이미지 분포에서 많이 활용되나, 제조데이터에서 드러나는 데이터의 분포를 학습하여 유사데이터를 생성하는 데 적합하다고 판단
- LSTM AutoEncoder: 시계열적 형태를 보이는 데이터의 이상을 감지하기 위해 활용

<연구결과>

DNN

- Precision score 0.974, Recall score 0.977 (불량 기준)

LSTM AE 학습지표 : MSE

- Epoch 28에서 MSE는 train_loss는 0.00621, val_loss는 0.00628로 나타남

GAN 학습지표 : D-loss, G-loss

Epoch 1000으로 설정시 MELT_TEMP는 D-loss는 0.4088, G-loss가 1.15137로 나타남

MOTOR_SPEED는 D-loss가 0.3518, G-loss는 1.6272로 나타남

<기대효과>

- 타산업에도 적용 가능한 확장성
- 모델 사용성 향상 – 중소제조기업의 데이터 전문성 부족
- 선제적 결함 방지를 통한 비용절감 및 효율성 극대화

<프로젝트 성과>

- **제 2회 K-인공지능 제조데이터 분석 경진대회 우수상 수상**
- 프로젝트 내용 및 배경 전달 커뮤니케이션 능력 향상

2.3. 거래량 예측을 통한 유망 기업 선정 모델 개발

프로젝트 요약	소규모 투자자들의 투자 의사결정을 돕기 위한 개별 기업 주식 거래량 예측 모델 개발
수행기간	2022.06 - 2022.07
프로젝트 형태	팀 프로젝트
참여인력	3명
참여도	기획 및 제안서 작성 50%, 데이터 전처리 및 모델 구현 100%

<역할>

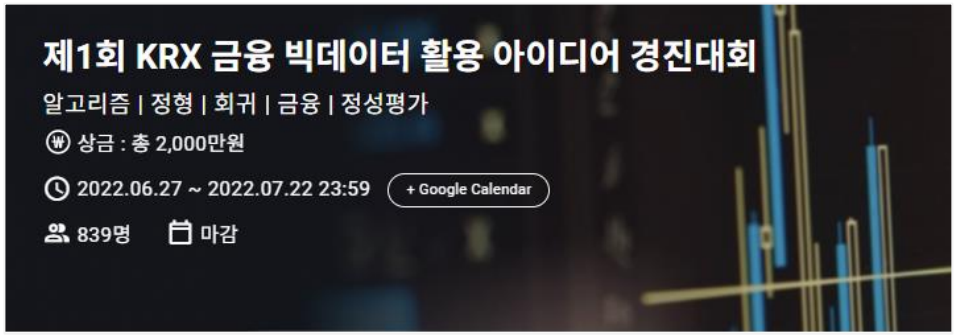
- 기획 및 제안서 작성
- 데이터 전처리 및 모델 구현

<프로젝트 배경>

- 주식시장의 불확실성으로 인한 소규모 투자자들의 피해 증가
- 소규모 투자자들의 투자 의사결정을 도울 수 있는 도구 필요
- 주식의 높은 거래량은 변동성 위허 헷지를 가능하게 함
- 주식의 유동성은 기업의 혁신과 양의 관련성 가짐
- 거래량 예측을 통해 투자자들이 안정적이고, 혁신가능성 높은 유망기업을 선정하는데 기여

<수행과정>

- 선행연구 리서치 및 **기획**
- **데이터 전처리 및 prophet 모델 구현**
- **결과 시각화 및 제안서 작성**



<모델 선정 근거>

- 주식데이터에서 나타나는 주기적 계절성을 반영할 수 있는 모델
- 특정 기업 이슈로 일시적으로 향상한 가격을 일반적 분포와 시각적으로 비교 가능

<프로젝트 성과>

- 팀장의 역할로 프로젝트 전반 과정을 전개하게 됨
- 분업화가 잘 이루어지지 않아 시간 관리에서 아쉬움 느낌
- 시계열 모델에 대한 이해 향상
- 리더보드 71팀 중 23위 달성

3.1. 여행 상품 신청 여부 예측 경진대회

수행 기간	2022.08 - 2022.09
프로젝트 형태	개인 프로젝트
참여인력	1명

<베이스라인>

- CatBoostClassifier + 5 K-Fold
- 대다수가 카테고리컬 칼럼이어서 그런 카테고리컬 칼럼들의 상호작용을 잘 파악하는 CatBoost 선택

<모델 향상>

- 여권보유여부(Passport)와 선호 호텔 숙박업소 등급 (PreferredPropertyStar) 칼럼 합하여 새로운 파생변수 생성

➔ 레이블인 여행패키지 신청여부 칼럼과 가장 양의 상관관계가 높은 칼럼들이어서 조합

- 영업 사원의 프레젠테이션 후 이루어진 후속 조치 수 (NumberOfFollowups)와 영업 사원의 프레젠테이션 만족도 (PitchSatisfactionScore) 합하여 새로운 파생변수 생성

➔ 레이블인 여행패키지 신청여부 칼럼과 가장 양의 상관관계 상위 2위, 3위인 칼럼들이어서 조합

- 직업의 직급 기준(Designation)으로 최소 월 급여(MonthlyIncome) 산출하여 새로운 파생변수 생성

➔ 직업의 직급에 따른 급여의 최소금액이 금전적 여유를 보여주어 레이블인 여행패키지 신청을 결정하는데 영향을 줄 수 있다고 판단

- 직업(Occupation)과 직급(Designation) 문자열로 합쳐 새로운 파생변수 생성

➔ 직업에서 직급이 경제적, 시간적 여유를 보여주어 레이블인 여행패키지 신청을 결정하는데 영향을 줄 수 있다고 판단

- 고객의 제품 인지 방법(TypeofContact)과 영업사원이 제시한 상품 (ProductPitched) 문자열로 합쳐서 새로운 파생변수 추가

➔ 고객이 제품을 스스로 인지방식에 따라서 영업사원이 제시한 상품을 선택하는지가 달라질 수 있다고 판단

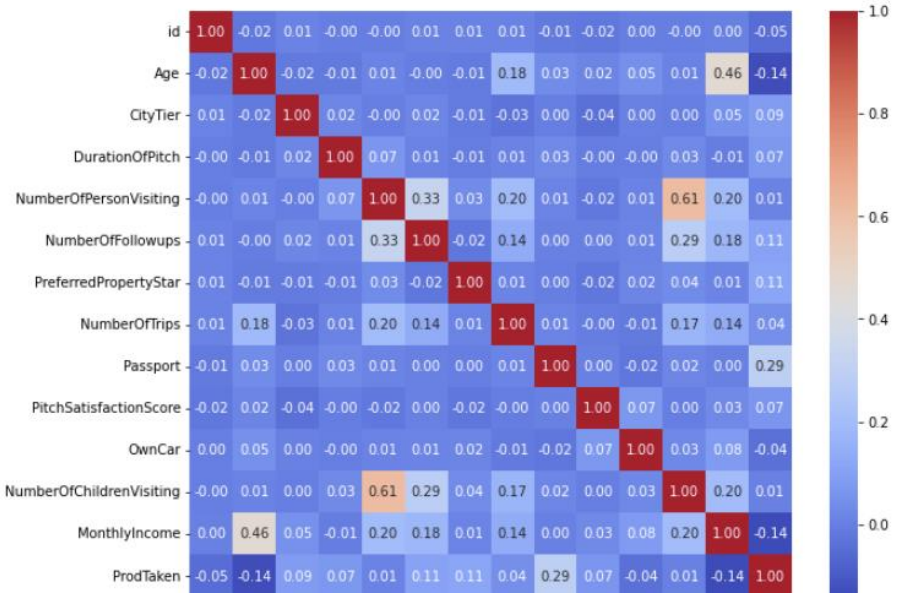
3.1. 여행 상품 신청 여부 예측 경진대회

- 영업 사원이 고객에게 제공하는 프레젠테이션 기간 기준으로 주거 중인 도시의 등급 최소 산출하여 새로운 파생변수 추가

➔ 레이블과의 상관관계를 확인하였을 때 비슷한 양의 상관관계를 보이는 칼럼들이라서 연결지어 파생변수 생성

- 영업 사원이 제시한 상품(ProductPitched)기준으로 평균 월 급여 (MontlyIncome) 산출하여 새로운 파생변수 추가

➔ 월 급여와 상관관계가 가장 높은 독립변수라서 영업 사원이 제시한 상품과 조합하여 파생변수 반영



- 도시의 등급과 고객과 함께 여행을 계획 중인 총 인원 조합하여 새로운 파생변수로 추가하였으나, 오히려 점수 개선에 도움이 되지 않음

➔ 도시등급에 따라 많은 인원의 여행, 적은 인원의 여행이 변동될 수 있어 여행패키지 선택에 영향을 줄 수 있다고 판단

- 결측치 -1로 대체

➔ 0 값이 의미있는 경우가 있을 수 있어 그러한 부분을 보존하기 위해 -1로 대체

- XGBoostClassifier, CatBoostClassifier, LGBMClassifier 스테이킹

➔ 카테고리컬 칼럼과 숫자형 칼럼이 함께 있는 데이터라서 각 칼럼에 맞는 모델의 장점을 가져와 결합시키고자 스테이킹 선택

<결과>

- 187위(0.8721227621)

➔ public 74위, private 38위(0.8985507246) / 603

➔ 상위 7%

3.2. 자율주행 센서의 안테나 성능 예측 AI 경진대회

수행 기간	2022.08 - 2022.08
프로젝트 형태	팀 프로젝트
참여인력	2명

<베이스라인>

- MultiOuputRegressor에 XGBoostRegressor

➔ 종속변수도, 독립변수도 많은 데이터이기 때문에 중요한 변수를 잘 식별하고, 빠른 학습속도를 가진 XGBoostRegressor 기반 MultiOutputRegressor 선택

➔ 하지만 학습이 균일하게 되지 않아 오버피팅, 언더피팅 위험이 있다고 판단

<모델 향상>

- 방열 재료들 무게 통합 칼럼 파생변수로 추가

➔ 방열재료들에 대한 종합적 정보를 제공하면 함께 고려할 수 있을 것이라고 판단

- RF부분들 SMT 납량 통합 칼럼 파생변수 추가

➔ 전자회로보드에 있어 납량이 많을 수록 부품이 많이 부착되었다고 생각하여 납량 통합 칼럼이 영향을 미친다고 판단

- 5번 안테나 패드 위치와 4번 스크류 삽입 깊이 파생변수 추가

➔ 패드와 스크류 위치가 인접할 경우 스크류 깊이가 안정성에 영향을 미친다고 판단

- 값의 분포를 확인하였을때 단일 값만 존재하는 칼럼(검사통과여부)인 X_04, X_23, X_47, X_48칼럼 제거

- XGBoostRegressor, CatBoostRegressor, LGBMRegressor 스테이킹

➔ 카테고리컬 칼럼과 숫자형 칼럼이 함께 있는 데이터라서 각 칼럼에 맞는 모델의 장점을 가져와 결합시키고자 스테이킹 선택

〈결과〉

753위 (2.178218952)

➔ public 115위, private 130위(1.9398802592)/2030

➔ 상위 7%