

multicampus

이민성 Portfolio

| 세상을 이롭게 하는 데이터 분석가, 이민성

CONTENTS

CHAPTER 01 | 인적 사항

CHAPTER 02 | 개인 역량

CHAPTER 03 | 프로젝트 및 수상

CHAPTER 04 | PR

CHAPTER

01

인적사항

인적사항

성명	이민성	성별	남자
휴대폰	010-3641-6141	이메일	lgt302@hanmail.net
생년월일	93.12.03	병역구분	해군 병장만기제대
주소	인천광역시 부평구 부영로 165 우성APT 108-205		

학력사항

기간	학교명	학과	소재지
2013.03~2019.08	총신대학교	영어교육과	서울
2009.03~2012.02	산곡고등학교	인문	인천

경력사항

기간	회사	부서	직위
2021.02~2021.05	티씨컴퍼니	교육기획	사원
2019.11~2020.03	Siller Preferred Company	Administrative team	인턴
2017.12~2018.01	스카우트	전략기획팀	인턴

CHAPTER

02

개인 역량

프로그래밍 및 라이브러리 역량

언어	숙련도	라이브러리	
Python	●●●●○	Pandas	Matplotlib
SQL	●●●○○	Numpy	Seaborn
		TensorFlow	Scipy
		Sklearn	Pytorch
		statsmodels	Keras

SW 및 분석 Tools 역량

종류	숙련도	자격증
MS - Word	●●●●●	ADsP
MS - Powerpoint	●●●●○	컴퓨터 활용능력 1급
MS - Excel	●●●●○	정보처리기능사
Photoshop	●●●○○	
Tableau	●●●●○	
AWS	●●●○○	

외국어역량

언어	숙련도	어학성적
영어	Advanced	TOEIC 950
일본어	Intermediate	OPic - IH

GitHub

<https://github.com/Minsung-commit/>

CHAPTER

03

프로젝트 및 수상

주요 프로젝트 내역

프로젝트	NLP 기반 숙소 추천서비스	기여도	40%
역할	모델링, 파이프라인 구축, 추천 알고리즘 구현, 프론트엔드	사용언어	Python, Mongo DB, HTML/CSS, JavaScript
기간	2020.08.25 ~ 2020.10.08		
내용	<ul style="list-style-type: none">• Spherical K-means 기반 클러스터링 모델 구축• W2V/코사인 유사도를 활용한 콘텐츠 기반 필터링 모델 구축• 엔지니어와의 협업을 통한 파이프라인 내 모델 연결• 콘텐츠 기반 필터링 기법을 적용한 추천 로직 구현		
링크	https://github.com/Minsung-commit/SNS_NLP_PJT		

프로젝트 이미지 -
분석모델정의서, 유즈케이스 다이어그램, 화면설계서, 실제구현화면

분석모델정의서

1. 전체 감성숙소 정보 제공
2. 인기 감성숙소 정보 제공
3. 테마별 감성숙소 정보 제공
4. 유사한 숙소 추천 서비스 제공
5. 코로나 및 감염 안전 정보 제공

유즈케이스 다이어그램

- 로그인 가능
- 회원가입 가능
- 숙소 확인
- 유사 숙소 검색
- 숙소 추천
- 코로나 정보 확인
- 코로나 웹사이트 방문

화면설계서

도구상자, Item Image, Item details, 유사 아이템 추천(내용기반 추천), 하단배너

실제구현화면

구름의아루, 하단 화면, 상단 화면

주요 프로젝트 내역

프로젝트	주택가격 예측 모델링	기여도	30%
역할	데이터 수집, 전처리, 모델링	사용언어	Python
기간	2021.08.09 ~ 2021.08.20		
내용	<ul style="list-style-type: none"> 서울시 아파트 실거래가데이터 수집 연도별 전처리 및 파생변수(평당가격 등) 생성 LGBM 모델을 활용한 분석 결과 해석 및 인사이트 도출 		
링크	https://github.com/Minsung-commit/PJT-list/tree/master/ML_PJT		

프로젝트 이미지

모델 비교&선택

적용 모델 목록

- KNNRegression
- Decision Tree
- Random Forest
- GBM
- LGBM
- XGboost

각 모델 별 MSE & RMSE 시각화
XGB > DT > RF > KNN > GBM > LGBM

평균 오차(MSE)가 가장 낮게 나오는 LGBM 모델 선정

순으로 모델 성능 확인

모델 설명

전체 데이터로 LGBM을 실행

```

lgb = LGBMRegressor(linear_tree = True,
                    boosting_type = 'gbdt',
                    objective = 'regression',
                    n_estimators = 9000,
                    learning_rate = 0.001,
                    max_depth = 4,
                    n_jobs = -1)

lgb.fit(X_train, y_train)

MSE : 1829758.059959414
RMSE : 1352.6854992789026
                    
```

RMSE가 1352(만원) 정도로 측정된다. 다른 모델보다는 오차가 낮게 측정되지만 더 낮게 할 순 있을까?
어떻게 하면 오차를 더 낮게 만들 수 있을까?

- 변수 선택이 잘못 되었을 가능성 확인
- 하이퍼 파라미터 조정

위의 변수중요도를 참고하여, 변수 선택 실시해보자

하이퍼 파라미터

하이퍼 파라미터 튜닝 과정

```

# GridSearchCV 활용
from sklearn.model_selection import GridSearchCV

params = [
    {'learning_rate': [0.1, 0.01, 0.001, 0.0001],
     'max_depth': [1, 2, 3, 4]}
]

grid_cv = GridSearchCV(lgb, param_grid=params, cv=5, scoring='neg_mean_squared_error', verbose=1)
grid_cv.fit(X_train, y_train)

CPU times: user 0.46s, sys: 0 ns, total: 0.46s
Wall time: 0.72s
Fitting 4 folds for each of 16 candidates, totalling 64 fits
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.
[Parallel(n_jobs=1)]: Done 64 out of 64 | elapsed: 25.5min finished

grid_cv.best_estimator_
estimator=LGBMRegressor(learning_rate=0.001, linear_tree=True,
                        max_depth=4, n_estimators=9000,
                        objective='regression',
                        param_grid={'learning_rate': [0.1, 0.01, 0.001, 0.0001],
                                   'max_depth': [1, 2, 3, 4]},
                        scoring='neg_mean_squared_error', verbose=1)

최적의 하이퍼 파라미터 : {'learning_rate': 0.1, 'max_depth': 4}
예측 오차 : -140839.36042269424
                    
```

- GridSearchCV 활용
- learning_rate, max_depth 전란 대입
- 교차검증 진행
- 최적의 하이퍼 파라미터를 도출

실제 적용

최적의 하이퍼 파라미터를 적용하여 모델링 후 실제값과 예측값 비교

```

lgb_r = LGBMRegressor(linear_tree = True,
                    boosting_type = 'gbdt',
                    objective = 'regression',
                    n_estimators = 9000,
                    learning_rate = 0.1,
                    max_depth = 4,
                    n_jobs = -1)

lgb_r.fit(X_train, y_train)

MSE : 537552.7584483256
RMSE : 733.1798950109895
                    
```

초기 Dataset	1차 하이퍼변수 제거	2차 하이퍼변수 제거	파라미터 튜닝 후
1352	1352	1355	733

튜닝한 하이퍼 파라미터를 통해서 test 데이터 확인

```

pred[0]
62496.679686278745

y_test.iloc[0]
62500
                    
```

RMSE가 733(만원) 정도로 오차가 파라미터 튜닝 전보다 절반정도로 줄었다.

예측값은 62496.68정도이고, 실제값이 62500 정도로 유의미한 예측을 하는 것으로 판단할 수 있다.

주요 프로젝트 내역

프로젝트	코로나 이후 경제적 손실 분석	기여도	30%
역할	데이터 수집, 전처리, 연구모델 설정, 데이터 시각화, PPT 제작	사용언어	Python, Tableau
기간	2021.07.09 ~ 2021.07.20		
내용	<ul style="list-style-type: none"> 서울시 코로나 증감 및 유동인구 변화 시각화 업종별, 상권별 폐업률, 주요상권 공실률 시각화 이태원 지역 상권의 현황을 분석하여 시각화하고, 인사이트 도출 및 해결방안 제언 		
링크	https://github.com/Minsung-commit/PJT-list/tree/master/Visualization_PJT		

프로젝트 이미지

2. 분석 개요

연구 모형

위치 특성 -상권 ↔ 경제적 손실 -매출 증감 ↔ 업체 특성 -업종

01 이태원 현황 시각화

02 매출 증감 분석 -상권으로

03 매출 증감 분석 -업종으로

04 업종, 상권 군집분석

2.1. 이태원 현황 시각화

외식업 폐업률

소매업 폐업률

서울시 주요상권 공실률

△ 코로나로 인한 두드러진 폐업을 증가하는 보이지 않음.
△ 20년도 1분기에 폐업을 상승시킨 것으로 다시 정상수치로 돌아옴.
△ 서울시 모든 주요 상권에서 20년도 1분기를 기준으로 공실률이 상승함.
△ 경제적 손실이 발생했음에도 불구하고 폐업률로 나타나지 않음.

2.4. 분석 결과

K-means Clustering

전년도 대비 매출 증감률과 유동인구와의 관계를 통해 48개동 업종에 대해 5개의 군집을 지향 (n_clusters = 5)

매출 손실액 기준 업종 상위 5종

- 편의점, 식당
- 일반음식점
- 커피, 음료
- 편의점, 식당
- 동프, 피자점

전년도 대비 매출 증감률과 유동인구와의 관계를 통해 이태원 내 11개의 지역에 대해 군집을 지향 (n_clusters = 2)

매출 손실액 기준 지역 상위 2종

- 서울 동산구 이태원역 (발달상권)
- 이태원 관동동 (관동동)

3. 결론 - 활용 예시 및 한계점

활용 예시 - 나의 지원금은?

지원금 알아보기

+1 이태원역 인근
+2 슈퍼마켓

합계 : 3pts

지원금 : 250만원

이태원역 인근에서 슈퍼마켓을 하고 있는 자영업자 A

연구 한계점은?

- 판단 Feature가 적음
- 각 기준 간 가중치 분석
- 영업자의 개별특성을 제외

위 세 가지를 함께 연구한다면 향상된 판단기준 제시 가능

2.4. 분석 결과

K-means Clustering

전년도 대비 매출 증감률과 유동인구와의 관계를 통해 48개동 업종에 대해 5개의 군집을 지향 (n_clusters = 5)

매출 손실액 기준 업종 상위 5종

- 편의점, 식당
- 일반음식점
- 커피, 음료
- 편의점, 식당
- 동프, 피자점

전년도 대비 매출 증감률과 유동인구와의 관계를 통해 이태원 내 11개의 지역에 대해 군집을 지향 (n_clusters = 2)

매출 손실액 기준 지역 상위 2종

- 서울 동산구 이태원역 (발달상권)
- 이태원 관동동 (관동동)

3. 결론 - 활용 예시 및 한계점

활용 예시 - 나의 지원금은?

지원금 알아보기

+1 이태원역 인근
+2 슈퍼마켓

합계 : 3pts

지원금 : 250만원

이태원역 인근에서 슈퍼마켓을 하고 있는 자영업자 A

연구 한계점은?

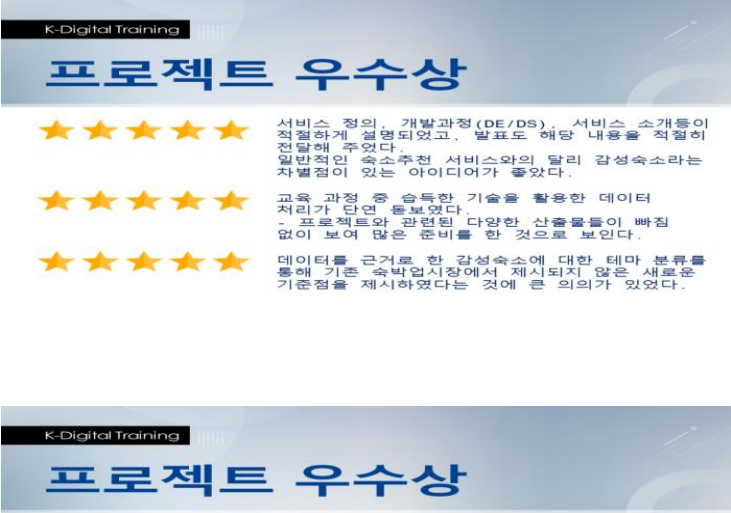
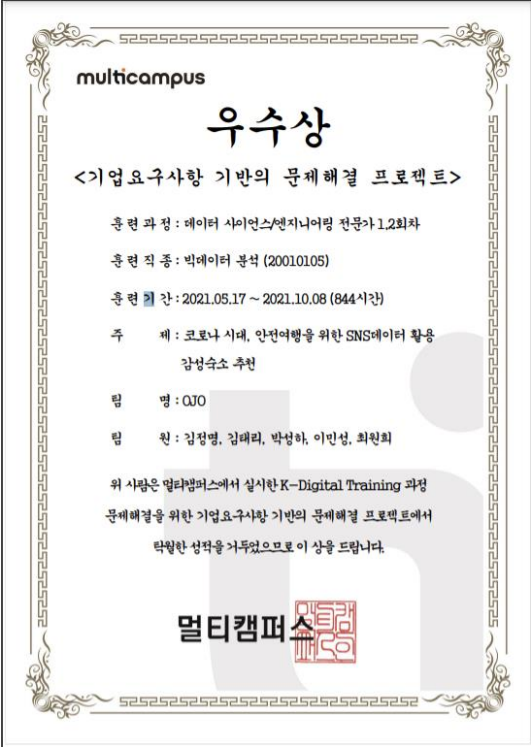
- 판단 Feature가 적음
- 각 기준 간 가중치 분석
- 영업자의 개별특성을 제외

위 세 가지를 함께 연구한다면 향상된 판단기준 제시 가능

주요수상및자격사항

일자	주요내용	발급기관
2021.10.08	기업요구사항 기반 PJT 경진대회 2등상 수상	멀티캠퍼스
2021.09.24	ADsP 자격 취득	데이터산업진흥원
2020.10.09	컴퓨터활용능력 1급 취득	대한상공회의소
2015.12.14	정보처리기능사 취득	산업인력공단
2020.07.12	TOEIC 950	ETS
2021.01.27	OPIc IH	OPIc

수상증빙및 기타



CHAPTER

04

PR

I 지원 분야와 관련하여 전문성을 키우기 위해 노력한 경험

"Soft & Hard"

Digital Transformation에 맞추어 데이터 분석에 필요한 Soft & Hard skills 역량을 키우고자 노력하였습니다. 교육 스타트업 기획팀 수습기간 당시, 사수의 추천으로 데이터 분석을 처음 접하게 되었습니다. 다가올 시대에서 데이터를 다루고 해석하는 역량을 필수라고 판단하였기때문에 수습기간이 종료됨과 동시에 데이터 사이언스라는 새로운 영역에 도전하기로 하였습니다. 멀티 캠퍼스 K-digital 과정을 수강하며, 5개월 하루 8시간씩 학업에 집중하며 풀스택 프로그래밍, 파이썬을 활용한 통계, ML/DL의 개념과 기술을 익혔고, 이에 더하여 유튜브 채널 'PLANIT DV'의 교육영상을 활용하여 Tableau 사용법을 배워 PJT 데이터 시각화 부분을 담당하였습니다. 기획과 관련하여 전화상담실 전화상담원으로 일하면서 일 평균 100개의 콜을 처리하며 문제대응을 위한 소통능력을 길렀고, 스카우트 전략기획팀 인턴을 통해 공공사업의 기획과 운영관리 지식을 함양하였습니다. 또 한 미국 HR 기업의 인턴을 통해 국제 감각을 기르며 문서 작성 및 DB 관리 능력을 인정받으며, 개인정보 수집 및 파기에 관련한 정보윤리를 습득할 수 있었습니다. 현재는 Git Hub를 관리하며 1일 1커밋을 하고자 노력중에 있고, 매일 프로그래머스를 통해 파이썬 기초 알고리즘 연습을 하여 코딩 기본기를 완성하고 있습니다.

I 다양한 자원 활용, 협력을 이끌어 내어 팀웍을 발휘하여 목표를 달성한 경험

"고민없이 도전정신 발휘 "

공공데이터를 활용한 데이터 시각화 PJT를 진행하며, 다양한 시각화 도구를 시도하며 PJT를 성공적으로 마무리한 경험이 있습니다. 데이터를 활용한 첫 PJT로서, 4인 1조를 구성하여 진행하였습니다. 하지만 팀원들의 데이터 관련 PJT에 대한 경험이 적었고, 특히 ML/DL에 대한 지식이나 기술이 부족한 상황이었습니다. 그에 반해 타 팀들의 경우, 다양한 기계학습을 적용한 PJT로 방향을 정해 진행하고 있었습니다. 저는 팀의 상황과 능력에 대한 자기 객관화가 필요하다고 판단했고, 팀원들에게 주어진 기간 안에서 바로 작업을 진행할 수 있는 부분을 진행하고 진행된 부분을 어필하는 방향으로 PJT 목표를 제시하였습니다. 그에 따라 당시 배운 주된 내용이었던 상관관계분석, EDA 위주로 분석을 진행하였습니다. 시각화 부분에선 Matplotlib과 Seaborn을 주로 배웠지만, 시각화 툴로 주목받고 있는 Tableau를 활용할 수만 있다면 분석 결과를 더욱 효과적으로 표현할 수 있으리라 생각하여 Tableau를 유튜브와 블로그 자료를 통해 사용법을 익혀 활용하였습니다. 마감일이 다가왔을 때, 비지도학습인 클러스터링 분석을 제안하였고 파이썬 활용이 가장 능숙했던 팀원과 함께 비지도학습인 클러스터링 분석을 진행하여, 기간 내에 분석 부분의 완성도를 높여 PJT를 마무리할 수 있었습니다. 수상은 하지 못하였지만, 지도 강사님께서 유의미한 결과를 냈다는 평가를 얻을 수 있었습니다.

I 목표를 세우고 끈질기게 성취한 경험

" 마지막으로 웃는 자가 진정한 승자 "

K-digital training을 수강하면서 세운 2가지 목표가 있습니다. 첫 번째는 좋은 인연을 유지하는 것이었고, 다른 하나는 3번의 PJT 중 반드시 한 PJT에선 수상을 하는 것이었습니다. 이것을 목표로 삼은 이유는 다양하게 있지만, 가장 큰 이유는 구체적인 성과를 만들고 싶었기 때문입니다. 하지만 1차, 2차 프로젝트 모두 아쉬게 수상을 하진 못했습니다. 최종 프로젝트 경진대회에서 수상을 하기 위해, 이전 PJT를 피드백해봤습니다. 1차 PJT의 경우, 아이디어는 좋았지만, 분석과정의 정교함이 부족하였고, 2차 PJT의 경우, 반대로 아이디어의 독창성이 부족하였다는 점을 알았습니다. 마지막 PJT에선 같은 실수를 반복하지 않기 위해, '감성숙소' 라는 독창성 있는 서비스를 구현하고자 하였고, 분석과정의 다양성과 정교함을 위해 다양한 방법을 활용하여 분석을 시도해보기로 하였습니다. 하지만 '감성숙소'를 연구한 선행연구의 부족으로 기본적인 연구모델을 설정하는 것부터 난항을 겪게 되었습니다. 이 문제를 해결하기 위하여, 키워드를 확장하여 연구를 조사하였고, 공간디자인에 관한 연구를 찾고, 공간 및 디자인의 분류에 대해 알아보며, 도메인 지식을 쌓았습니다. 이후 영화, 소설 등 다양한 미디어 콘텐츠에 관한 연구를 조사하여, 군집분석의 방법적 기초를 구성하여 PJT를 진행하였습니다. TF-IDF, W2V, D2V, LDA, K-means, KoBERT 등 다양한 기법을 적용하며 주제 분류 및 키워드 추출을 시도해봤습니다. 최종적으로 W2V - K-means - TextRank를 통한 군집/키워드 분석모델을 수립하여 모델링을 진행하였고 이를 파이프라인과 연결하여 자동화하는 작업을 진행하였습니다. 그 후 추천 로직과 추천 로직을 구현하여 새로운 콘텐츠에 대한 분류 및 추천시스템이 연동된 서비스를 구현하였습니다. 이 산출물을 통해 경진대회에서 2등상을 수상할 수 있었습니다. 또 한 같은 PJT를 진행한 팀원들과도 지속해서 교류하게 되어 제가 목표로 세웠던 2가지를 모두 이룰 수 있었습니다.