

PORTFOLIO. |

신동우 포트폴리오



## PROFILE

Name 신동우 (1998.10.03)

## CONTACTS

Phone 010.6309.9824

E-mail doungeoosin@naver.com

Notion

## EDUCATION

2023 동의대학교 컴퓨터공학과 졸업

2017 동의대학교 입학

2017 부산 성지고등학교 졸업

## ACTIVITIES

2022 Intel AI 개발 교육 과정 1기 (수료)

2022 Intel AI Global Impact Festival 2022

2022 부산 디지털 덴티스트리 데이터톤 대회

2024 LG Aimers 교육 및 해커톤 대회

2024 KPI 도출 비즈니스 전략 아이디어 경진대회

## SKILLS

Python

SQL

Tensorflow

Pandas

## AWARDS

2022 Intel AI Global Impact Festival 2022  
Country/Region winner Award

2022 부산 디지털 덴티스트리 데이터톤 대회  
대상

# CONTENTS. |

01

personalized  
food warning system  
image Classification

02

튼튼하니, 튼튼한 이  
image Detection

03

MQL 고객정보를 활용한  
고객영업 전환 예측 프로그램  
정답셋 기반

04

E-commerce Data  
KPI 도출 및 전략 아이디어 제시



01

personalied food warning system



02

튼튼하니, 튼튼한 이



03

MQL 고객정보를 활용한 고객영업 전환 예측 프로그램



04

E-commerce Data  
KPI 도출 및 전략 아이디어 제시

01

personalized food warning system

(Image classification)

# 인텔 AI 국제 대회

# Intel AI Global Impact Festival 2022

# 국제 2등 수상

# 한국 대표

# Country / Reigon Winner

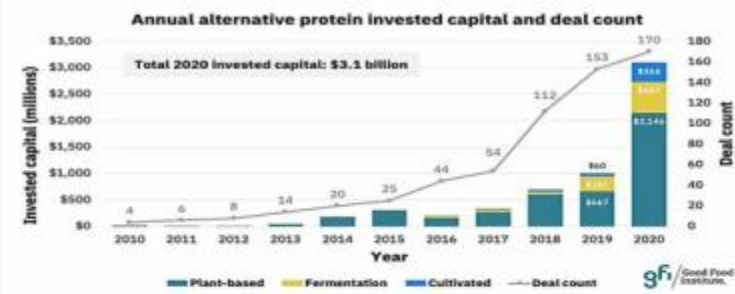
작업기간 2022. 08. 15. ~ 2022. 08. 25.  
참여인원 총 4명  
역할 데이터 전처리 및 모델 학습





## 1. Set problem scope

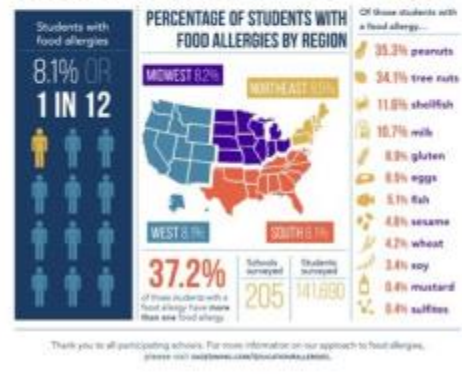
Vegan population of USA



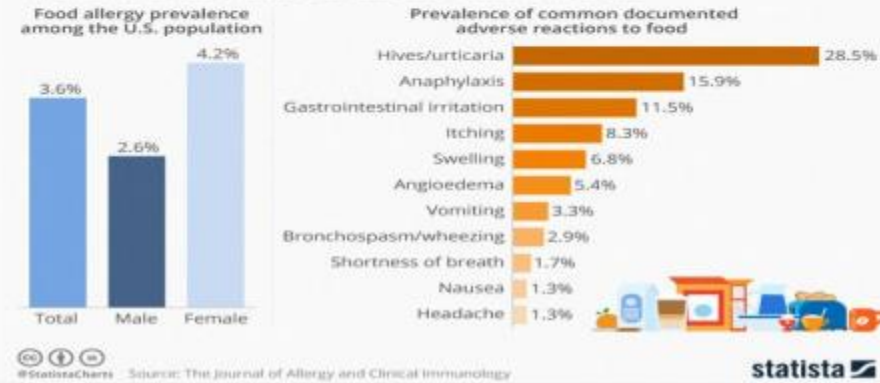
The world religion population



2018-2019 FOOD ALLERGY STUDY



Fewer Than 1 In 25 Americans Have A Food Allergy



## Labeling and purifying collected data

```
batch_size = 32
img_height = 200
img_width = 200
```

```
train_ds = tf.keras.preprocessing.image_dataset_from_directory(
    data_dir,
    validation_split=0.2,
    subset="training",
    seed=256,
    image_size=(img_height, img_width),
    batch_size=batch_size)
```

Found 8574 files belonging to 18 classes.  
Using 6860 files for training.

```
val_ds = tf.keras.preprocessing.image_dataset_from_directory(
    data_dir,
    validation_split=0.2,
    subset="validation",
    seed=256,
    image_size=(img_height, img_width),
    batch_size=batch_size)
```

Found 8574 files belonging to 18 classes.  
Using 1714 files for validation.

## Recall labeled data using a matplotlib

```
import matplotlib.pyplot as plt

plt.figure(figsize=(10, 10))
for images, labels in train_ds.take(1):
    for i in range(9):
        ax = plt.subplot(3, 3, i + 1)
        plt.imshow(images[i].numpy().astype("uint8"))
        plt.title(class_names[labels[i]])
        plt.axis("off")
```



## 4. Test

```
[70]: test_dir = "banana.jpeg"
test_dir = pathlib.Path(test_dir)

img = keras.preprocessing.image.load_img(test_dir, target_size=(img_height, img_width))
img_array = keras.preprocessing.image.img_to_array(img)
img_array = tf.expand_dims(img_array, 0) # Create a batch

predictions = model.predict(img_array)
score = tf.nn.softmax(predictions[0])

print(
    "This image most likely belongs to {} with a {:.2f} percent confidence."
    .format(class_names[np.argmax(score)], 100 * np.max(score))
)

1/1 [=====] - 0s 22ms/step
This image most likely belongs to 저혈압 - 바나나 with a 100.00 percent confidence.
```

- Explanation

- When I entered a picture of banana that was not included in the training data into the model, it was 100% accurate because it was banana.

\* Currently, a prototype has been created to accurately distinguish food \*

```
[73]: test_dir = "pizza(test).jpg"
test_dir = pathlib.Path(test_dir)

img = keras.preprocessing.image.load_img(test_dir, target_size=(img_height, img_width))
img_array = keras.preprocessing.image.img_to_array(img)
img_array = tf.expand_dims(img_array, 0) # Create a batch

predictions = model.predict(img_array)
score = tf.nn.softmax(predictions[0])

print(
    "This image most likely belongs to {} with a {:.2f} percent confidence."
    .format(class_names[np.argmax(score)], 100 * np.max(score))
)

1/1 [=====] - 0s 21ms/step
This image most likely belongs to 고혈압 - 피자 with a 99.27 percent confidence.
```

- Explanation

Accuracy is getting higher than before.  
The loss rate is also much lower than before.

- Explanation

Results of the first training graph  
Problem: somewhat low accuracy, somewhat high loss rate



01

personalied food warning system



02

튼튼하니, 튼튼한 이



03

MQL 고객정보를 활용한 고객영업 전환 예측 프로그램



04

E-commerce Data  
KPI 도출 및 전략 아이디어 제시

02

튼튼하니, 튼튼한 이

(Image Detection)

# 튼튼하니, 튼튼한 이

# 부산 디지털 덴티스트리 데이터톤 대회

# 인사이트 도출

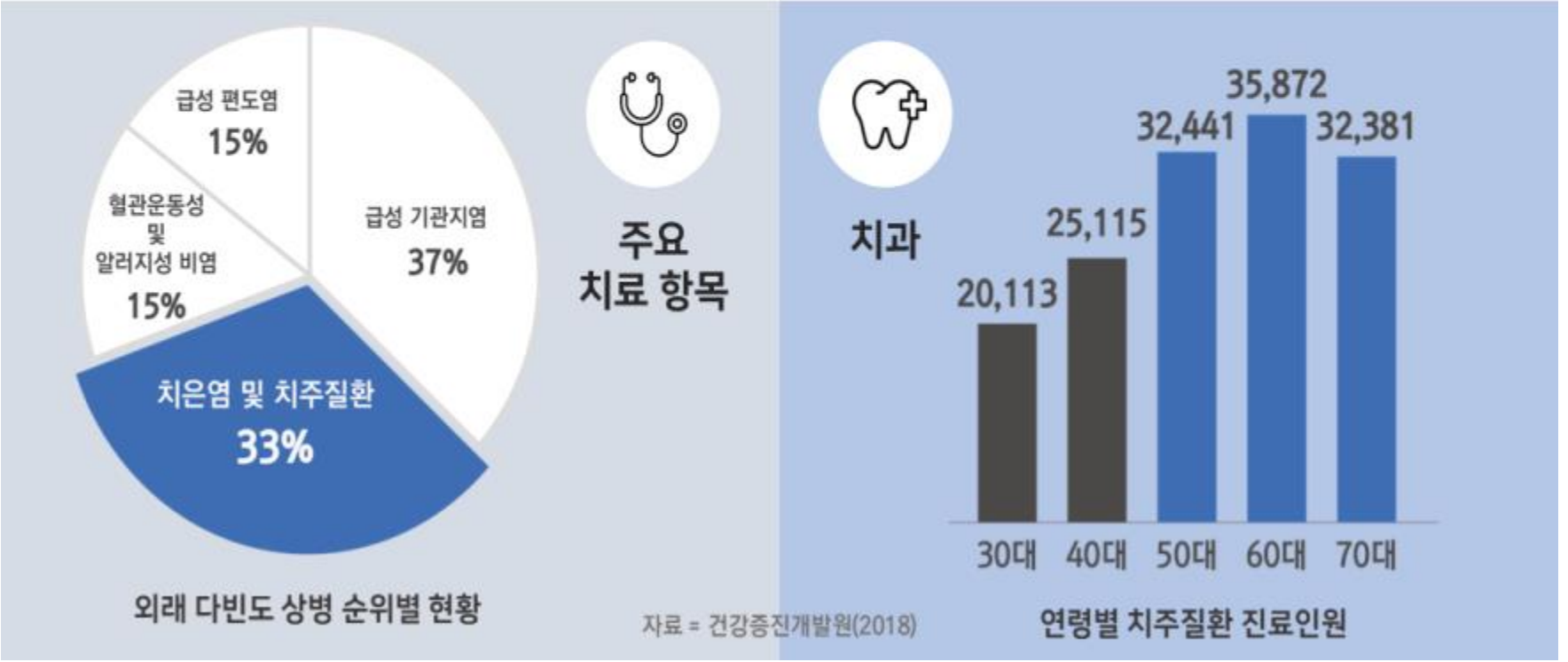
# 대상 수상

작업기간 2022. 09. 28. ~ 2022. 09. 30.  
참여인원 총 4명  
역할 데이터 분석 및 모델 학습

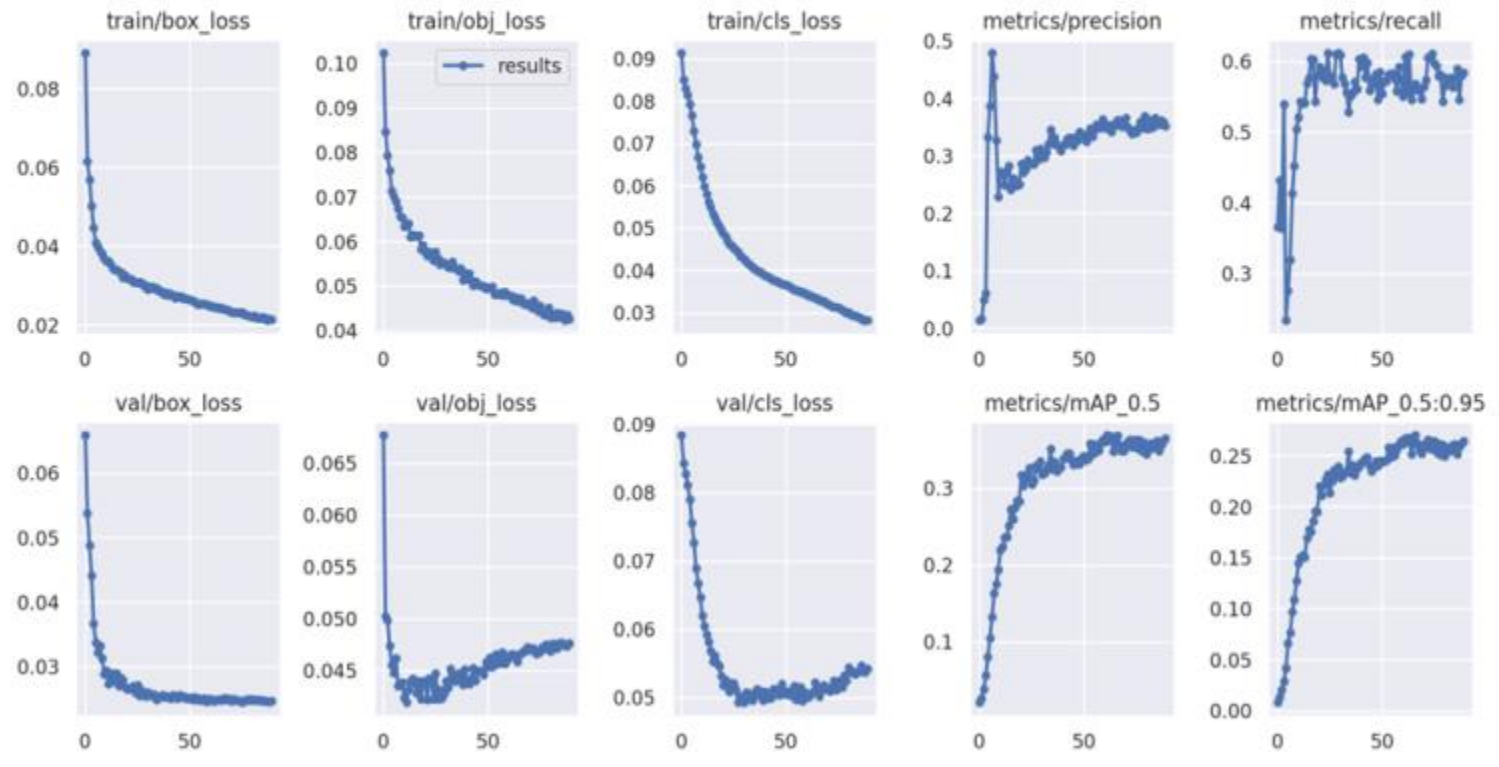




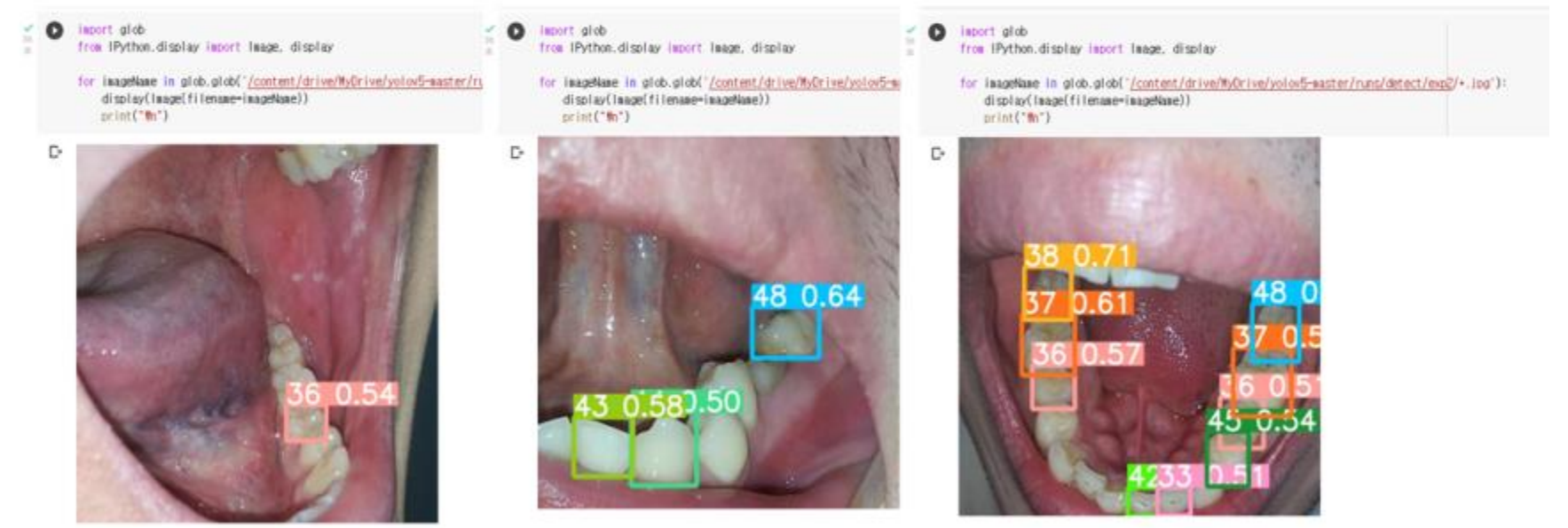
확장성이 높은  
의료분야



Results



그렇다면 어떤 코드를 사용할까?  
Precision & Recall rate



프로젝트를 활용한 서비스  
확장 가능성

<p>딥러닝과 빅데이터를 활용한 치주질환 관리 플랫폼 기술 개발</p>	<p>치의학 의료 데이터 인공지능 의료기기 개발 가능</p>
<p>증상, 질병 이미지로 전문 자료를 검색 및 환자에게 가장 적합한 치료법 제시 가능</p>	<p>다양한 이미지 인식 및 영상 인식 등을 이용한 새로운 형태의 서비스 출시</p>



01

personalied food warning system



02

튼튼하니, 튼튼한 이



03

MQL 고객정보를 활용한 고객영업 전환 예측 프로그램



04

E-commerce Data  
KPI 도출 및 전략 아이디어 제시

03

MQL 고객정보를 활용한 고객영업 전환 예측  
(정답셋 기반)

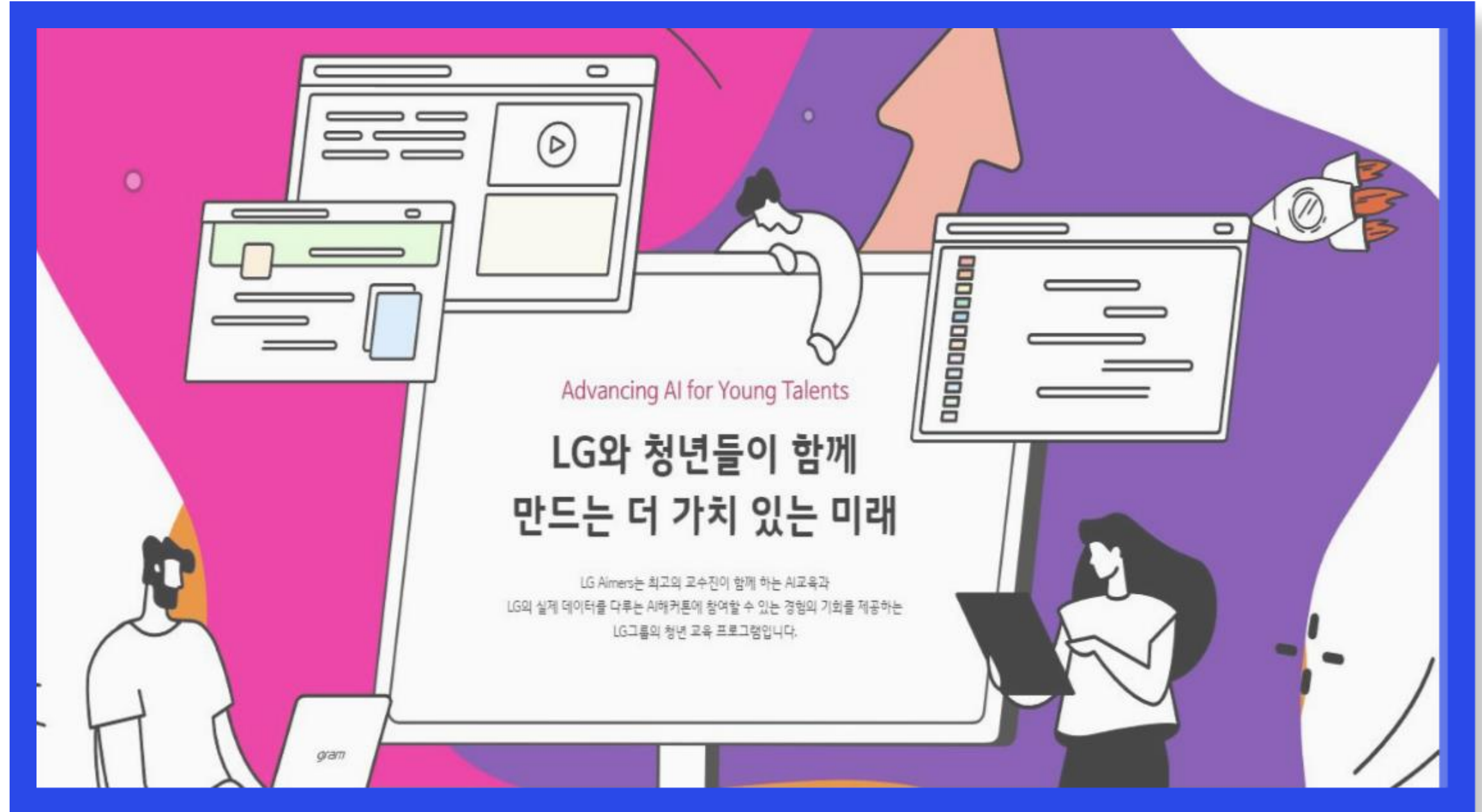
## MQL 고객정보를 활용한 고객영업전환 예측

# 고객지수 산출

# 영업전환 수 UP

#고객 맞춤형 서비스 제공

작업기간 2022. 02.13. ~ 2022. 02. 26  
참여인원 총 4명  
작업 기여도 데이터 요구사항 정의 및 분석





## 데이터 소개

### 정답셋 기반 예측 알고리즘 모델 개발



MQL 고객 정보에는  
개인 정보(회사/직급 등), 구매요청 정보(제품/예산/니즈/기한) 및 영업사원, 마케팅 활동  
정보등이 포함

#### Table data

고객 ID	직책	제품	유입채널	...	Y
A	MQL 정보 변수 약 30개				영업전환이력 성공 : 1 실패 : 0
B					
C					
...					
Z					

#### Text data

고객이 직접 작성한  
요청 메시지 (영어)

예시)

we need Air Ventilation Solution  
ASAP for our new building...

## 모델링

### 정답셋 기반 예측 알고리즘 모델 개발



#### 결측치 처리

: 결측치가 있는 컬럼에 대해서 삭제하거나 특정 값으로 채우기

#### 인코딩

: 범주형 변수에 대해서 인코딩을 통해 수치로 변환

#### 가공변수 생성

: 주요 feature를 파악하고 비즈니스 도메인 지식을  
활용하여 다양한 파생 변수 생성

#### 이진 분류 모델 생성

: 주어진 데이터에 적합한 ML/DL 알고리즘을 활용하고,  
모델의 예측 성능을 향상시킬 수 있는 다양한 방법론을  
적용

#### 텍스트 마이닝

: 자연어처리 과정을 통해 고객이 작성한 메시지 정보에서  
유의미한 정보를 추출하여 예측 과정에 활용

## 평가 지표

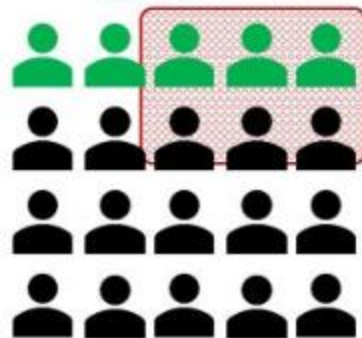
### 정답셋 기반 예측 알고리즘 모델 개발



#### 영업 전환 고객으로 예측

영업 전환  
고객 (y=1)

영업 비전환  
고객 (y=0)

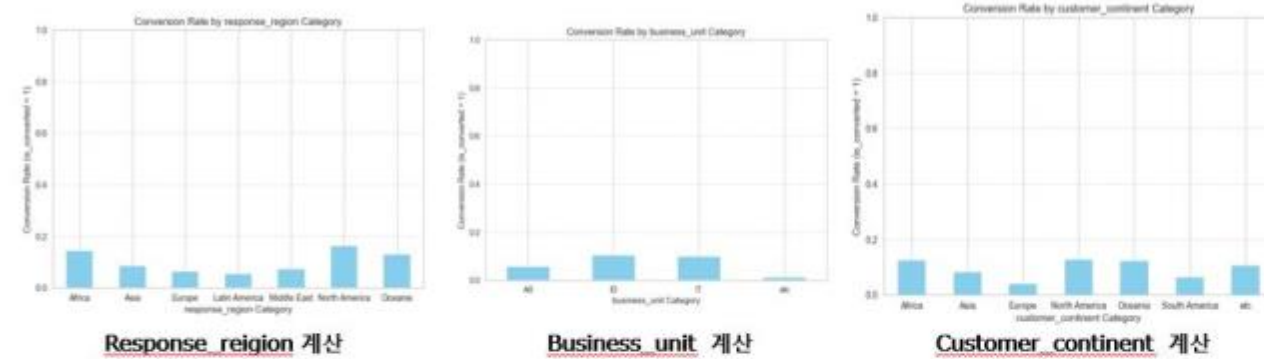


#### Precision

: 모델이 영업 전환 고객이라고 예측한 고객 중 실제 영업 전환 고객 비율

$$\frac{\text{정확히 예측한 영업 전환 고객 수}}{\text{영업 전환 고객으로 예측한 고객 수}} = \frac{3}{6} = 0.5$$

## 프로젝트 결과



#### 데이터 정제를 바탕으로 이진분류 모델을 사용한 결과

```
DecisionTreeClassifier 최적 하이퍼파라미터: {'max_depth': 15}
RandomForestClassifier 최적 하이퍼파라미터: {'max_depth': 30, 'n_estimators': 100}
Ensemble Accuracy: 0.967622596964587
Ensemble Precision: 0.7989802984162633
Ensemble Recall: 0.8294416243654822
Ensemble F1 Score: 0.8097125867195242
Model 1 교차 검증 정확도 평균: 0.9739866301845826
Model 2 교차 검증 정확도 평균: 0.9672742481126682
Model 3 교차 검증 정확도 평균: 0.9834888563046264
Model 4 교차 검증 정확도 평균: 0.9788884681701982
Model 5 교차 검증 정확도 평균: 0.9838437483942389
True로 예측된 개수: 1833
```

기존 True로 예측된 개수 : 50  
현재 True로 예측된 개수 : 1000



선제적인  
고객 이슈 대응



고객 맞춤형  
서비스 제공



비즈니스  
전략 수립



01

personalied food warning system



02

튼튼하니, 튼튼한 이



03

MQL 고객정보를 활용한 고객영업 전환 예측 프로그램



04

E-commerce Data  
KPI 도출 및 전략 아이디어 제시

04

E-commerce Data KPI 도출 전략 아이디어 제시



PROJECT\_04

# E-commerce 데이터 분석

# 마케팅 전략 아이디어

# KPI 도출

# 만족도 시계열 예측

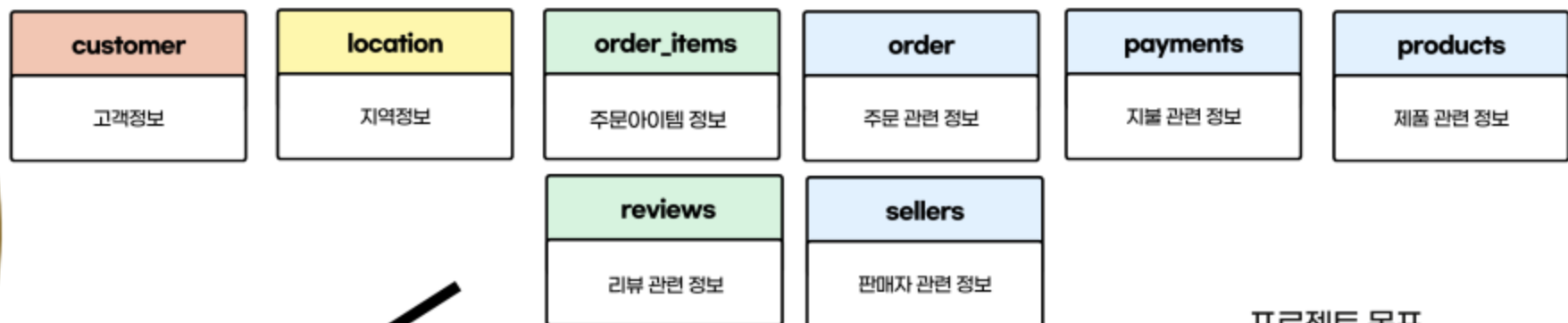
작업기간 2022. 04. ~ 2022. 04.  
참여인원 총 1명  
작업 기여도 데이터 분석 및 KPI 및 전략 아이디어 제시





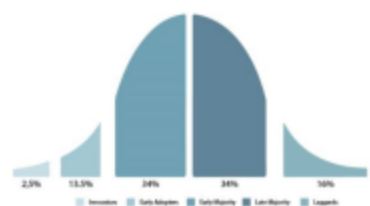
## 프로젝트 개요

### E-commerce Data



### 프로젝트 목표

#### 기업 성장 및 고객 만족



### 상품 리뷰 만족도 분석



### KPI 성과지표 도출

#### 비즈니스 전략 아이디어 제시

## 고객 만족도 조사 데이터 전처리 과정

### 1. 결측치를 확인해줍니다.

```

1: isna_value = master_product_08596.isnull().sum()
print(isna_value)

Product_id      0
Customer_id     0
Order_id        0
Product_category_name  0
Price           0
Freight_value   0
Order_item_id   0
Payment_type    0
Review_id       9
Review_score     9
Review_creation_date  9
dtype: int64
  
```

이를 위해서 copy() 메서드를 사용하여 복사본을 만들고, 해당 복사본을 수정하고, 다시 원본 데이터로 불러옵니다.

```

# 데이터프레임 복사
master_product_08596_copy = master_product_08596.copy()
master_product_08596_copy['Review_id'].fillna('unknown', inplace=True) # Review_id 값의 결측치를 'unknown'으로 대체
master_product_08596_copy['Review_score'].fillna(0, inplace=True) # Review_score 값의 결측치를 0으로 대체
master_product_08596_copy['Review_creation_date'].fillna('0000-00-00 00:00:00', inplace=True)

isna_value = master_product_08596_copy.isnull().sum()
print(isna_value)

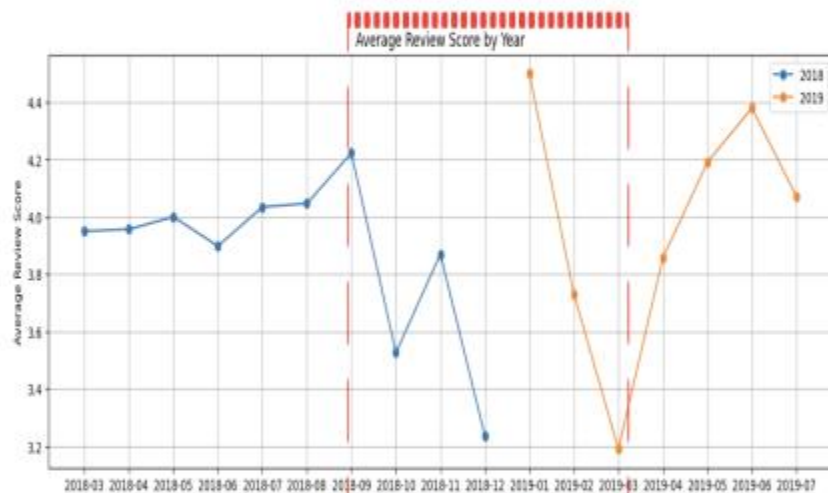
Product_id      0
Customer_id     0
Order_id        0
Product_category_name  0
Price           0
Freight_value   0
Order_item_id   0
Payment_type    0
Review_id       0
Review_score     0
Review_creation_date  0
dtype: int64
  
```

### 결측치: 특정값으로 대체

이유: 리뷰를 쓰는 사람도 있는 반면, 안 쓰는 사람도 존재.

데이터 전처리 완료									
Customer_id	Order_id	Product_category_name	Price	Freight_value	Order_item_id	Payment_type	Review_id	Review_score	Review_creation_date
JSTOMER_76205	ORDER_19931	bed_bath_table	74.0	23.32	1	credit_card	REVIEW_50349	5.0	2019-07-31
JSTOMER_36954	ORDER_87547	bed_bath_table	74.0	23.32	1	credit_card	REVIEW_77181	4.0	2019-06-14
USTOMER_71470	ORDER_27319	bed_bath_table	74.0	23.32	1	credit_card	REVIEW_78269	5.0	2018-05-24
USTOMER_71470	ORDER_27319	bed_bath_table	74.0	23.32	2	credit_card	REVIEW_78269	5.0	2018-05-24
USTOMER_06519	ORDER_68790	bed_bath_table	74.0	23.32	1	credit_card	unknown	0.0	NaT

### 날짜별 고객 만족도(평균) 시각화



### 1. 고객 만족도 조사

#### 1.1 날짜별 고객만족도 점수를 통한 고객만족도 측정 결과

#### 1.2 kpi 도출 아이디어 전략

#### 1.1 날짜별 상품 고객만족도 점수를 통한 고객만족도(평균값) 측정 결과

```

master_product_08596.groupby('Review_creation_date_yn')['Review_score'].mean()

2018-03    3.950000
2018-04    3.956522
2018-05    4.000000
2018-06    3.897959
2018-07    4.034483
2018-08    4.046512
2018-09    4.222222
2018-10    3.526316
2018-11    3.869565
2018-12    3.235294
2019-01    4.500000
2019-02    3.730769
2019-03    3.190476
2019-04    3.857143
2019-05    4.190476
2019-06    4.379310
2019-07    4.071429
Name: Review score, dtype: float64
  
```

### 제품 개선

- 리뷰 만족도가 낮은 기간 제품 개선점 파악
- 품질 향상 집중
- 고객 요구에 맞는 제품 개발 전략 수립

### 접근성

- SNS 유튜브를 활용 광고
- 사람들과 상품의 접근성을 완화

### 가격

- 경쟁기업 대비 경쟁력 있는 상품 가격
- VIP 전용 멤버십 활성화
- 제휴 할인카드 확대를 통한 가격 할인

### 시기별 프로모션

- 리뷰 만족도 높을 때 프로모션 진행
- 인기 제품 홍보 강화
- 마케팅 캠페인 진행

Product  
P1

Price  
P2

P3  
Place

P4  
Promotion

4P  
DIAGRAM

