

---

# PORTFOLIO

조수현

+

o

•

# 자기소개서



1997.11.11

조수현

한국외국어대학교 독일어통번역학과

## SKILL SET

Language Python (2), R(2)

OS Linux (2)

Framework+Library Django (2)

Database Oracle (2) SQL Developer (2)

Development Tool Vscode (2) , Jupyter notebook(3), spyder (3)

Data Preprocessing Pandas (2) Numpy (2)

Data Analysis Scikit-learn (2) Tensorflow (2) OpenCV(2), Pytorch(2)

Data Visualization Matplotlib (2) Seaborn (2) Power BI (3)

ETC Git (2) MS-Word (3)

## 핵심 역량

- 롯데 멤버스 주최 빅데이터 경진 대회 참여 (22.06.27~22.08.08)
- 와이즈넷 기업 프로젝트 참여로 교육용 MRC모델 기반 간편 질의응답 시스템 구축 (22.05.30~22.06.24)
- K-Digital Training 빅데이터 기반 AI 응용 솔루션 개발자 전문과정 수료 (22.02.22~22.08.16)
- 2021년 공공데이터 청년인턴 과정 수료 (한국지능정보사회진흥원 주관)
- ADsP 데이터 분석 준전문가 자격 시험 응시 후 결과 대기 중
- TOEIC 895점, 한국사능력검정시험 1급

## CONTACT

PHONE: 010-2236-6875

E-mail: [cookie0605@naver.com](mailto:cookie0605@naver.com)

Git-hub: <https://github.com/cookie0605>

# TEAM PROJECTS

사용자 기반  
협업 필터링

고객 구매 데이터 기반  
추천 시스템 개발  
및 개인화 마케팅 전략 제안

자연어 처리

역사 빅데이터 활용  
한국사 MRC 모델 기반  
간편 질의응답 시스템 구축

Python

Pygame 프로젝트

# 01. 고객구매 데이터 기반 추천 시스템 개발 및 개인화 마케팅 전략 제안

Recommending system based on customer purchase datasets and personalized marketing proposal

## □ OVERVIEW

롯데멤버스 제공 내부 데이터를 기반으로 개인화 마케팅에 활용 가능한 상품 추천 시스템 개발  
사용자 기반 협업 필터링을 활용해 사용자 간 유사도 분석하여 선호도가 비슷해 구매할 것으로 예상되는 상품 추천

□ 기간 22.06.27~22.08.08 (6주)

□ 팀원 <Team project> 주최 - 롯데 멤버스  
김OO, 한OO, 조수현 (3명)

## □ 역할

### EDA 분석

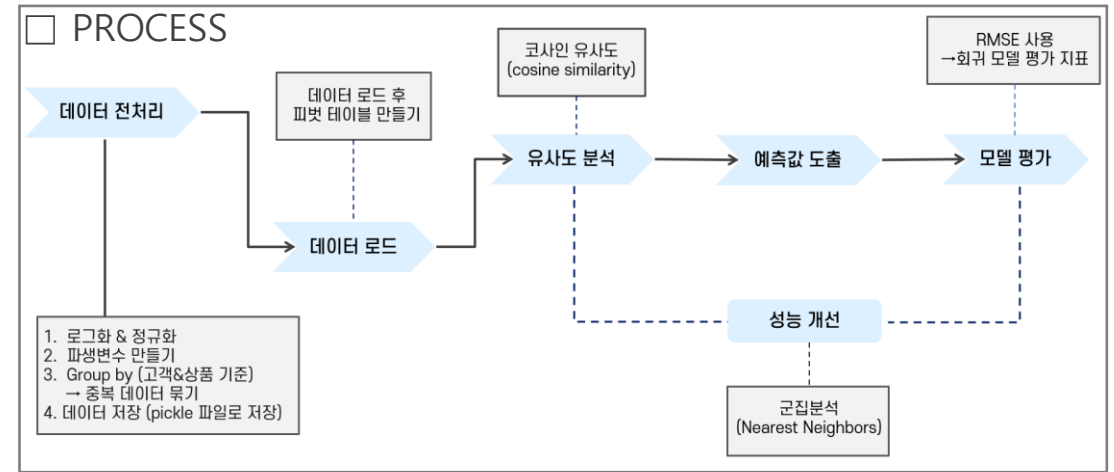
데이터 전처리, 활용 파일 병합 및 가공  
추천 시스템 활용 모델 탐색 및 알고리즘 탐색  
외부 데이터 수집 및 내부 데이터와의 상관관계 분석  
추천 시스템 평가 지표 탐색  
모델 고도화 위한 가중치 탐색

## □ WORK RULE

Environment: window 10, Jupyter notebook, Oracle, Power BI  
Language: python  
Framework/Library: pandas, numpy, scikit-learn

□ DATA SET - 롯데멤버스 제공 고객데이터 (약 400만 건)  
- 뉴스 심리지수, 카카오 트렌드 지수

## □ PROCESS



## □ OUTPUT

1. 입력 고객 데이터에 따른 추천 상품 반환
2. Train 데이터의 RMSE 값을 7.3818 → 1.5797 까지 줄임
3. 추천 시스템을 활용한 개인화 마케팅 방안 제시

# 01. 고객구매 데이터 기반 추천 시스템 개발 및 개인화 마케팅 전략 제안

Recommending system based on customer purchase datasets and personalized marketing proposal

## □ OUTPUT

### <시스템 구동 과정>



고객  
M999227380

#### 데이터 입력

<고객 데이터>

30대/여성  
서울  
주말 A01  
여성 의류 및 남성 의류 구매  
|  
L01에서 엘페이 사용  
|  
C01에서 문화 생활 후  
엘포인트 적립  
평일 A06  
- 엘페이 사용 식품군 구매  
(과자, 맥주, 냉동 식품 등)



#### 모델 학습 및 예측

<추천 시스템 알고리즘 가동>

사용자-사용자 (유사도) 비교 후  
고객이 추가적으로 구매할 만한  
상품들 추천

구매 0 상품 VS 구매 X 상품  
비교 후 예측 값이 높은 것을 추천

상품명	예측값
일반 스낵	0.997749
즉석 밥	0.866025
국산 맥주	0.842922
수입 맥주	0.832273
안주용 육포류	0.641555



#### 결과 및 검증

<예측 정확도 검증 및 개선>

평가 지표: RMSE  
→ 절대적인 기준이 없기 때문에  
지속적으로 RMSE 값을 줄여감을 목표

오류율이 1.6 이하인 것으로 결론

기본 모델 군집 분석 전  
→ 최종 모델 군집 분석 후 비교

Train: 7.38188023 → 1.57974330  
Test: 8.24680045 → 2.44789364

학습 데이터의 오류율을 대폭 줄임  
추천 시스템의 성능 검증 완료

### <평가 지표 개선 비교>

train: 7.381880236477308  
Test: 8.246900453482464



train: 1.579743305945178  
Test: 2.4478936412833026

군집 분석을 통한 성능 개선

### <개인화 마케팅 제안>

LOTTE Mart 대한민국 농산물

000 고객님의 위한 맞춤 상품 추천

전체 과일 채소 쌀/잡곡 견과/견과 정육/계란 수산물 견해산물 밀/밀가루/밀가루 냉동/냉동/간편식

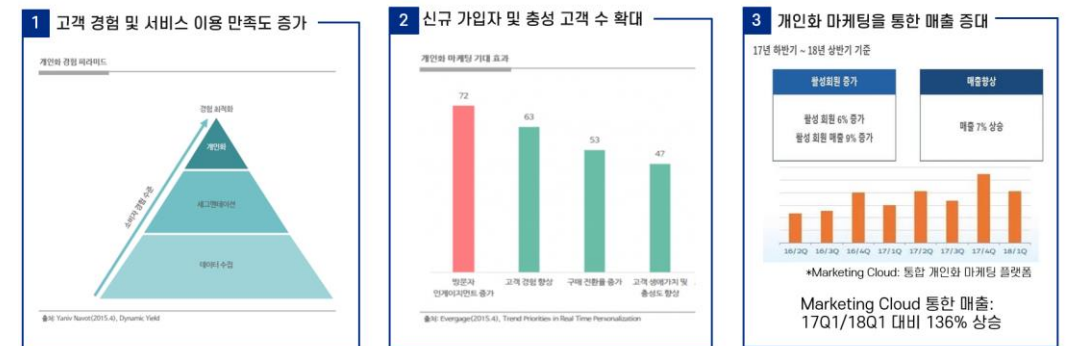
1. [당일배송] [소매포장] 온리프라이스 Only Price 1등급 우유 (930ML\*2입) 3,500원 4.9 (288,596) 후기/리뷰

2. [당일배송] [소매포장] 제주 산다수 그린 (2L\*6입) 5,880원 4.9 (12,408) 후기/리뷰

3. [당일배송] [소매포장] 행복 생생란 (특란, 30입) (1.8KG) 6,790원 4.8 (116,141) 후기/리뷰

4. [당일배송] [소매포장] 농심 농심 올리브 파파게티 (5개입) 4,280원 4.7 (78,188) 후기/리뷰

5. [당일배송] [소매포장] 양반 양반 참기름 식탁김 (4G\*16봉) 8,480원 4.7 (319) 후기/리뷰



개인화 마케팅을 통해 소비자의 경험 수준을 개인에 최적화되어 만족도를 증가시킴

소비자의 경험 수준을 끌어올려 총성고객 확보 및 신규 가입자에 대한 방문 유도

고객 경험 및 만족도 증대와 총성 고객 확보 → 결과적으로 매출 증대 실현

## 02. 역사 빅데이터 활용 한국사 MRC모델 기반 간편 질의응답 시스템 구축

Question and answering system using Korean history 'big data' based on MRC

### □ OVERVIEW

위키 데이터와 국정 교과서 데이터를 데이터 셋으로 사용하여 konlpy의 tag를 통해 분석하고 입력되는 질문과 일치하는 문서 반환  
→ 사용자가 입력한 질문의 답을 매칭하여 반환해주는 한국사 교육용 응용프로그램 개발

□ 기간 22.05.30~22.06.24 (4주)

□ 팀원 <Team project> 참여기업 - 와이즈넷  
김OO, 김OO, 박OO, 한OO, 조수현 (5명)

### □ 역할

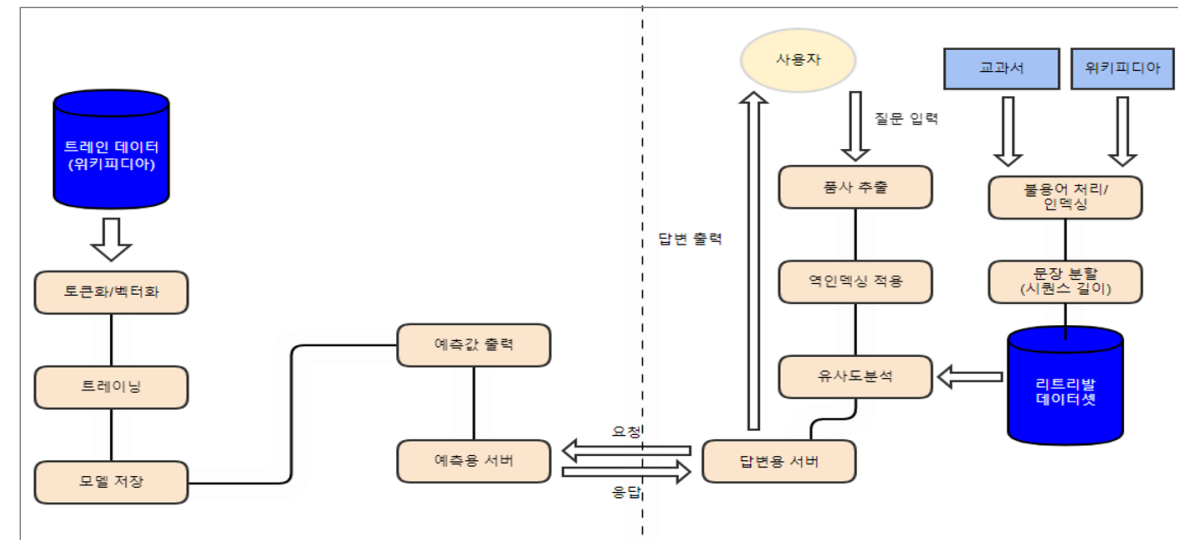
데이터 플로우 작성을 통한 프로젝트 전반 데이터 활용 방안 수립  
데이터 수집(크롤링) 및 가공 - 위키 데이터 수집  
모델 로직 탐색 및 기획  
서비스 배포 단계 flask 활용한 웹 사이트 구성 수정

### □ WORK RULE

Environment: window, jupyter notebook, Flask, Pycharm  
Language: python  
Framework/Library: pandas, numpy, konlpy(Hannanum),  
Wikipedia,

□ DATA SET 2007년 국정 교과서  
WIKIPEDIA 모듈을 활용한 위키데이터

### □ PROCESS



### □ OUTPUT

MRC 모델 이해 - 사용자가 입력한 질문에 대한 정답을 학습된 데이터(Retrieval 서버)에서 찾아 Reading 서버에 연결하여 제작한 flask 웹 화면에 띄워 정답 확인

## 02. 역사 빅데이터 활용 한국사 MRC모델 기반 간편 질의응답 시스템 구축

Question and answering system using Korean history 'big data' based on MRC

### □ OUTPUT

데이터-활용 모델-결과(시각화) 과정

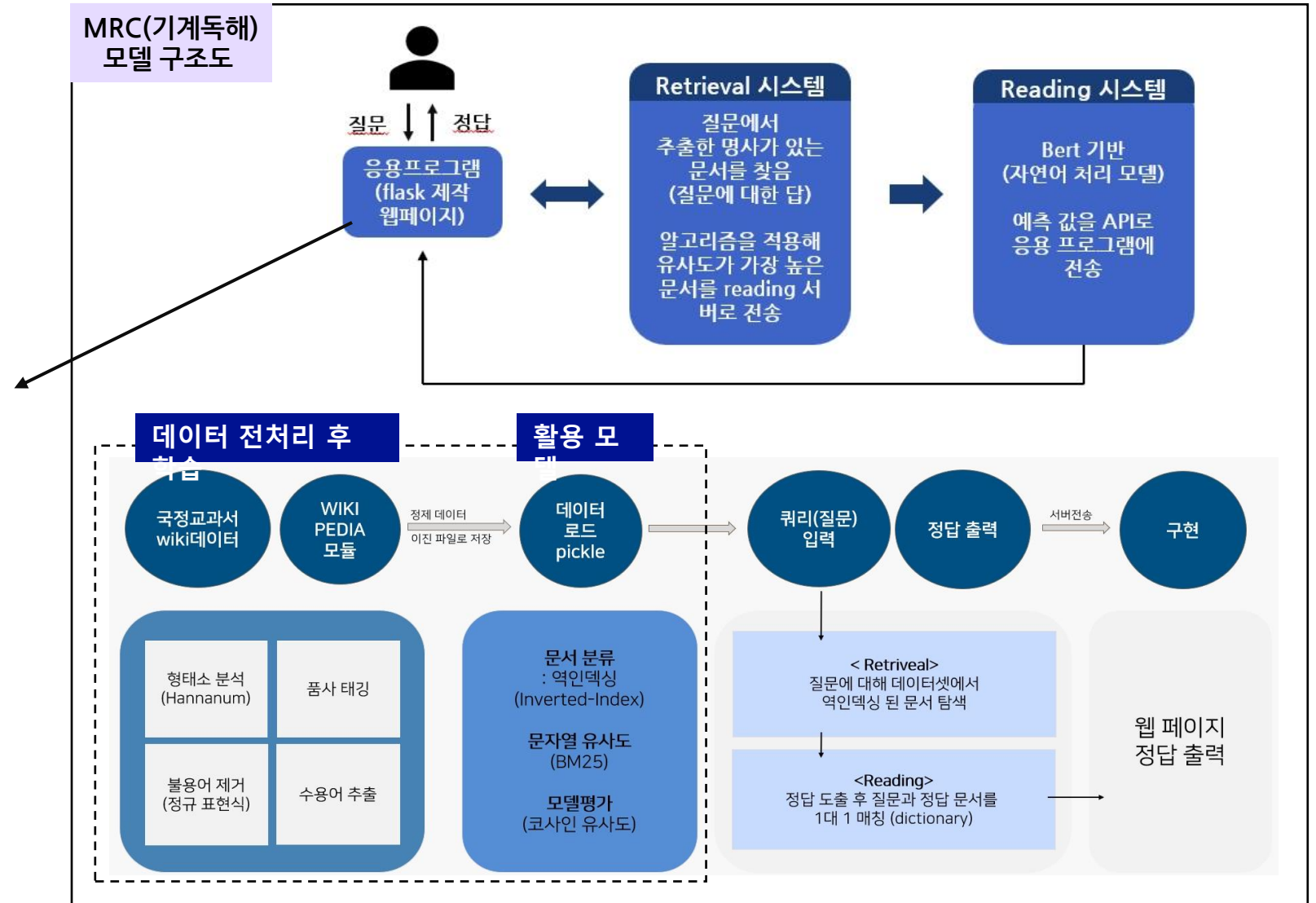
### 시현 화면

시일야방성대곡을 게재했던 신문은?
 

분석수행

정답	스코어
황성신문	3.4391341
를 발견, 황성신문	2.0458148
이 신채호를 발견, 황성신문	2.0434432
신채호를 발견, 황성신문	2.0348263
, 황성신문	2.027994

### MRC(기계독해) 모델 구조도





## 03. Pygame 프로젝트

shooting game project

### □ OVERVIEW

Pygame 기본 포맷에서 다양한 기능 추가 및 게임 시나리오 설정을 통한 게임 화면 구현 및 슈팅게임 구동을 위한 변수와 클래스 활용 실습

### □ 기간 22.05.11~22.05.17 (1주)

### □ 팀원 <Team project>

김OO, 김OO, 김OO, 장OO, 조수현 (5명)

### □ 역할

게임 UI 화면 구현에 활용할 데이터 수집  
슈팅게임 구동(게임 시작과 종료, 진행) 위한 이미지 및 화면 구현  
클래스 함수 구현 - 아이템 하강 속도 및 개수 조절  
최종 코드 검토 및 오류 수정

### □ WORK RULE

Environment: window 10, spyder

Language: python

Framework/Library: pandas, numpy

### □ PROCESS

Pygame 슈팅게임 기본 과정 구현

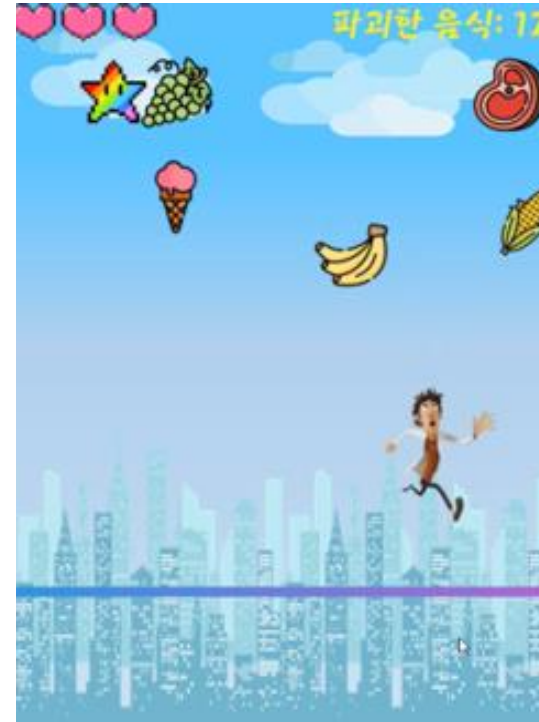
보스/필살기 추가

캐릭터, 음식, 배경, 사운드 등 시나리오에 대한 게임 전반 컨셉 수정

### □ 소감

처음 진행한 프로젝트로 팀원 간 역할 분담과 팀워크의 중요성을 느낌  
처음이라 모르는 부분이나 헤매는 부분이 있었으나 팀원들이 도와준 덕에 무사히 프로젝트를 마무리 할 수 있었음

### □ OUTPUT



1. 점수(파괴한 음식)에 따라 아이템 하강속도 ↑
2. 파괴한 음식 갯수에 따라 레벨 상승, 배경 화면 바뀜
3. 점수(파괴한 음식)가 75점이 되면 보스 등장
4. 지나친 음식이 3개 or 캐릭터에 음식이 닿으면 게임 오버
5. 필살기 사용 시 아이템 모두 제거 가능



# TOY PROJECTS

Power BI

삼성전자 영업관리  
대쉬보드 시각화

TensorFlow

피마 인디언  
당뇨병 예측

WordCloud

웹 크롤링을 통한  
맛집 키워드  
워드 클라우드 만들기

# 01. S전자 영업관리 대시보드 시각화

## OVERVIEW

S전자 영업관리 데이터를 활용하여 판매 실적, 판매 계획, 달성율 등을 슬라이서, 맵, 원형 차트, 꺾은선 그래프 등 다양하게 시각화하여 1년 간의 실적 파악

## WORK RULE

Environment: window 10, Power BI

## DATA SET

2018-19년 S전자 1년간의 영업관리 데이터

## PROCESS

csv 파일 로드 후 테이블 간 릴레이션hip ship 매칭  
새 측정값 작성하여 카드 및 꺾은 선 그래프 작성  
직원 실적 별 원형 차트, 맵

<새 측정값: DAX>

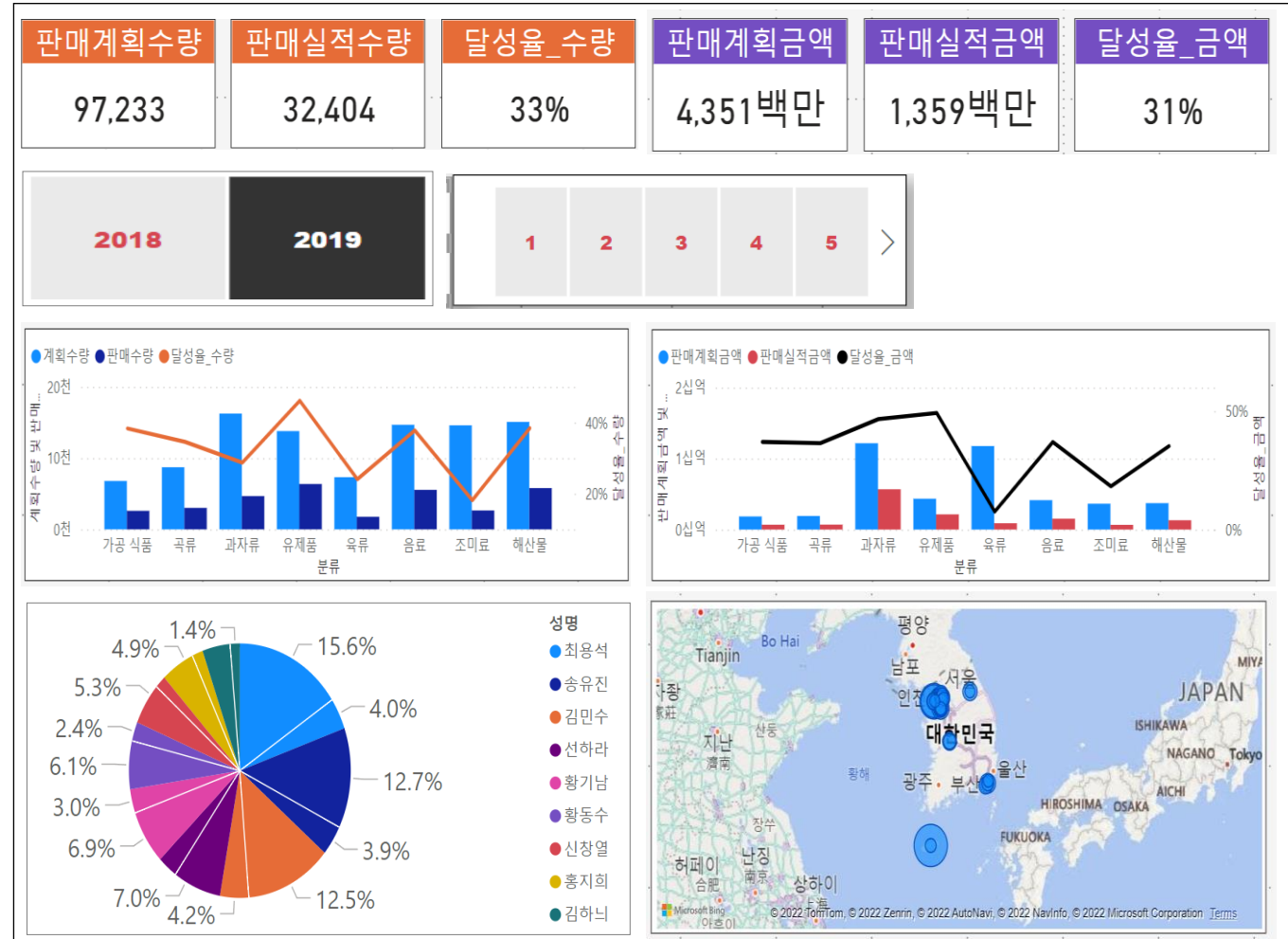
달성율\_수량

=  $\text{sum}(\text{'판매' [판매수량]}) / \text{sum}(\text{'판매계획' [계획수량]})$

달성율\_금액

=  $\text{sum}(\text{'판매' [판매실적금액]}) / \text{sum}(\text{'판매계획' [판매계획금액]})$

## OUTPUT



## 02. 피마 인디언 당뇨병 예측 분류 실습

### □ OVERVIEW

피마 인디언 당뇨병 유발 요소에 대한 상관 관계  
분석을 통한 회귀 실습, 당뇨병 진단 여부에 대한 예측 실시

### □ WORK RULE

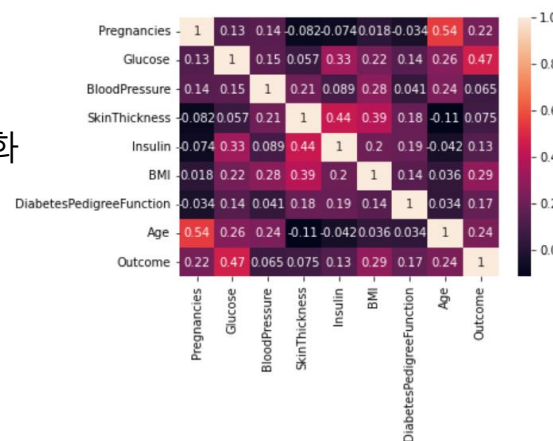
Environment: window 10, jupyter notebook  
Language: python  
Framework/Library: tensorflow(keras), scikit-learn, pandas,  
numpy, matplotlib, seaborn

### □ DATA SET

피마 인디언 당뇨병 유발 요소와 결과 칼럼 (769건)

### □ PROCESS

데이터 칼럼 간 상관관계  
heatmap으로 분석  
관련도 낮은 칼럼 삭제 후 정규화



### □ OUTPUT

```
# 모델의 설정
model = Sequential()
model.add(Dense(12, input_dim=8, activation='relu'))
model.add(Dense(8, activation='relu'))
model.add(Dense(1, activation='sigmoid'))

# 모델 컴파일
model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
```

Model: "sequential\_9"

Layer (type)	Output Shape	Param #
dense_27 (Dense)	(None, 12)	108
dense_28 (Dense)	(None, 8)	104
dense_29 (Dense)	(None, 1)	9

Total params: 221  
Trainable params: 221  
Non-trainable params: 0

Sequential 을 이용하여 모델 만들기  
Dense층 설정: activation='relu', activation='sigmoid'

```
model.evaluate(x_data, y_data)
```

24/24 [=====] - 0s 679us/step - loss: 0.4592 - accuracy: 0.7721

[0.459158331155777, 0.7721354365348816]

Loss값: 0.4592, 정확도: 0.7721  
개선 방안: dense 층 추가 후 정확도 상승

### 03. 웹 크롤링을 통한 맛집 키워드 기반 워드 클라우드

#### □ OVERVIEW

포털 사이트 맛집 검색을 통한 크롤링 데이터를  
활용하여 키워드 기반 워드 클라우드 제작

#### □ WORK RULE

Environment: window 10, jupyter notebook

Language: python

Framework/Library: numpy, wordcloud, konlpy

#### □ DATA SET

“상암맛집” 검색 시 검색 결과 크롤링 된 데이터

#### □ PROCESS

크롤링 데이터 전처리

- 형태소 분석기를 활용해 형용사, 명사 추출
- 최다 출연 단어 100개만 저장
- 시각화: 하트 모양 이미지를 활용해서 출력  
한글 폰트 저장 후 설정, 출력 이미지 크기 설정

#### □ OUTPUT



단어 크기가 클수록 핵심 키워드 (언급 정도↑)

# 학습 내용 요약



- 파이썬 기본 연산 - 슬라이싱, 인덱싱 등
- 제어문 - if문/for문/while문, 함수/class, 정규표현식
- Pandas : 리스트, 튜플, 딕셔너리, series, dataframe
- append(), sort(), count(), len(), iloc()
- Numpy : 선형 대수 기반 수학적 연산 활용. Nddarray



- 데이터 전처리 - dplyr 활용: 데이터 불러오기, 조건부 출력, 오름차순/내림차순 정렬
- 데이터 가공 - 통계 함수 활용, mutate/summarise
- 데이터 프레임 만들기, join 활용, 이상치/결측치 확인
- 데이터 시각화 - ggplot2 (산점도, 박스플롯, 히스토그램 등)
- 워드 클라우드, 인구 분포 시각화



- 기본키(Primary Key), 외래키(Foreign Key)
- 데이터 정의어(DDL): CREATE, ALTER, DROP
- 데이터 조작어(DML): SELECT, INSERT, DELETE, UPDATE
- 데이터 제어어(DCL): GRANT, REVOKE
- 활용: SELECT/FROM절, WHERE절 (LIKE 와일드 문자), ORDER BY (ASC/DESC), SUM/COUNT, GROUP BY절, HAVING절
- 조인: INNER JOIN, OUTER JOIN, JOIN~ON
- 서브 쿼리: WHERE 절 이하 단일행 연산자 (부호)/다중행 연산자 (IN/EXISTS 등등), FROM 절/WITH절, SELECT 절



- Csv, xlsx 파일 대시보드 시각화
- Python과 R Studio와 연동하여 코드 시각화
- 파일 간 관계 분석(릴레이션 십) 및 새 측정값 작성
- 슬라이서, 카드, 맵, 원형 차트, 꺾은 선 그래프, 테이블 등
- Power BI를 활용하여 달력 만들기

# 학습 내용 요약



- 데이터 전처리 과정 - 정규화
- 지도 학습
- 분류: K-NN, 결정 트리&랜덤 포레스트, 나이브 베이즈, SVM
- 회귀: 선형 회귀, 로지스틱 회귀, 규제 선형 모델 - 릿지/라쏘/엘라스틱 넷
- 비지도 학습: K-Means, GMM, DBSCAN, PCA
- <학습 결과 평가>
- 분류: Precision/Accuracy/Recall/F1-score
- 회귀: MSE/RMSE/MAE/MAPE



- 딥러닝: ANN, DNN, CNN, RNN, LSTM, GRU
- 기본 신경망 모델 → 은닉층을 넣으면서 모델 고도화 → 합성곱 신경망(학습/분류): 이미지 처리에 활용 → 기울기 소멸 문제 해결
- 자연어 처리
- [전처리] 토큰화 - 불용어 제거 - 어간 추출 - 정규화 → [임베딩]
- Open CV: 딥러닝을 활용한 이미지 분석. 비전(vision)
- cv2.imshow(), cv2.imread(), cv2.resize(), cv2.CascadeClassifier() 등
- PyTorch를 통한 딥러닝 학습: 모델 생성 후 가설 수립 → 데이터 판별하기(모델 초기화, optimizer, sigmoid)

## django

- 파이썬 웹 프레임워크 (자바 프레임워크: spring)
- 장고 프로젝트와 앱 생성 - 모델 생성 - URL 작성 - 템플릿 작성 - 뷰 작성
- 테이블 릴레이션 : 공통된 필드로 2개의 테이블 릴레이션
- 장고 CRUD 구현 (CREATE, READ, UPDATE, DELETE)
- ORM - DB 브라우저와 연동하여 게시판 만들기
- Bootstrap 통하여 게시판 꾸미기



- 오픈 소스 운영 체제
- 기본 명령어: pwd, cd, ls (ls 명령어: 파일 타입, 모드, 소유자 등을 확인할 수 있음)
- 파일과 디렉토리 생성 및 편집 명령어: mkdir, touch, rm/rmdir, cat, less, cp, mv, ln 명령어
- find, locate, man, which 명령어
- 텍스트 에디터(Vim 기본 에디터에서 명령어 입력)