**Team Project**

Team Name : MangJa

김병목 심동섭 최준용

**프로젝트 목표**

H 백화점의 기존 방문 고객들의 각종 정보를 입력 받아, 신규 고객의 정보를 입력하였을 때, 그 고객의 성별을 예측하는 프로그램을 만들어본다.

**프로젝트 진행 과정**

- train.csv 의 데이터를 불러와 학습시켜

- test.csv 의 데이터에 적용시킬 예정

1. 데이터 분석

- 시각화

- 각 데이터의 특성 파악

- 각 데이터간의 상관관계 파악

2. 데이터 정제

- 유효한 컬럼의 생성

- 높은 상관관계의 컬럼의 유무 확인

- 정규화, 일반화, 균일화

- 유효하지 않은 컬럼 제거

- 최종 필요한 컬럼의 결정

3. 머신러닝

- 가장 효율적인 머신러닝의 탐색

1. 로지스틱 회귀분석

2. GridSearchCV

3. Desision Tree

4. 최근접이웃 모델

5. 랜덤포레스트

6. 그라디언트 부스팅

7. XGBoosting

8. Extra Tree

9. Gaussian Process

10. Bagging

- 최종 예측모델 결정

4. 학습내용의 적용

- 결정한 예측모델의 적용

- 적용을 위한 train.csv 의 수정

- 모델 적용 및 출력 / 저장

**향후 추가할 내용**

- Voting Classifier

- SMOTE와 Bagging이 불균형 데이터를 정제하는 데 가장 적합하지만 데이터의 성격에 따라 그렇지 않을 가능성도 있기 때문에 이후 지속적인 탐구 예정.