

# Portfolio

빅데이터 분석  
Machine Learning

---

손지우

---

# 손지우, Jiwoo Son



## > Introduction

심리학을 포함한 다각적인 시선으로 데이터를 바라보는  
통계데이터분석가가 되고 싶은 손지우입니다.

## > Profile & Education

[생년월일] 1996.05.17

[고등학교] 하나고등학교

[대학교] 연세대학교 심리학과 & 응용통계학과

[대학원] 연세대학교 통계데이터사이언스학과

## > Characteristic

[차분함] “이것 또한 지나가리라”

[창의성] 남들과 다른 풀이를 하는 데에 흥미를 느낍니다.

[계획성] 꼼꼼하게 단계별로 일을 정리합니다.

[낙관성] “무의미에 의미 주기”

## > Interests

[재즈피아노] 엄밀하면서도 자유로운.

[농구] 거칠면서도 아름다운.

[사진] 정적이지만 동적인.

# 손지우, Jiwoo Son

## › Activity

2015.03 ~ 2018.02 | KSCY 한국청소년학술대회 컨퍼런스 총괄  
2015.07 ~ 2017.07 | 고등학교 심리학 멘토링  
2015.07 ~ 2015.08 | 쿨재팬 리포터  
2015.09 ~ 2015.12 | Psy-world 교내 심리학 학술모임 창립  
2016.01 ~ 2017.01 | 위메이저 전공소개 강연봉사  
2016.01 ~ 2016.04 | Seekers 청년창업  
2016.03 ~ 2016.09 | 멋쟁이사자처럼 4기  
2016.09 ~ 2016.12 | 서은국 교수님 성격사회심리학 연구실 RA  
2017.03 ~ 2019.02 | 대한민국 공군 병장 만기 전역

2019.03 ~ 2019.12 | 통계청 대학생 기자단

2019.03 ~ 2019.05 | 연세대학교 YDMC 전공소개봉사단체

2019.03 ~ 2019.12 | 양천구 대학생 멘토단

2019.09 ~ 2020.08 | 연세대학교 통계학회 ESC 학술부 및 총무

2021.09 ~ NOW | 김현중 교수님 지도제자

2022.03 ~ NOW | 연세대학교 데이터사이언스연구소 통계상담 조교

## › Awards

2015.12 | 연세대학교 Freshman Seminar 창업 101 최우수상  
2021.01 | NH투자증권 Y&Z 투자자 프로파일링 시각화 경진대회 입선상 (공동 5위)  
2021.12 | 연세대학교 빅데이터분석 경진대회 우수상

## › Certificate

데이터분석 준전문가 (ADsP)  
SQL 개발자 (SQLD)  
한자능력검정시험 3급

## › Skill



그외 MS Office(Powerpoint, Excel, Word)도 잘 다룹니다.

# Project LIST

## | Team Project |

2020.04 ~ 2020.06 | 수소차 충전소 입지 추천

- Role: RandomForest 모델링

2020.05 ~ 2020.06 | 서울주민 가구원 행복도 예측

- Role: 행복에 대한 학술자료 정리, FA 모델링

2020.08 ~ 2020.09 | NS Shop+ 홈쇼핑 실적 예측 및 편성표 제안

- Role: 상품군별 데이터분석, 노출시간 전처리, 상품명 전처리

2020.11 ~ 2020.12 | NH투자증권 YZ 투자자 프로파일링 시각화 경진대회

- Role: 주식데이터 크롤링

2021.05 ~ 2021.06 | 아파트 경매가격 예측

- Role: 베이지안 회귀분석 코딩, 명목형 변수 전처리

2021.09 ~ 2021.12 | 효돌 사용자로그 기반 고객 세분화

- Role: EDA, 데이터 전처리 및 일부 통합, 시각화

## | Personal Project |

2021.05 | 빅맥과 노동가치

- Point: 회귀분석 with Stepwise selection

2021.05 | 수질오염총량 관리제에 대한 지역주민 인식 분석

- Point: 로지스틱 회귀분석 with Factor Analysis

2021.06 | 택배회사직원 지각시간

- Point: 회귀분석을 로지스틱 회귀분석 문제로 바꿈

2021.12 | 큰돌고래 상호작용 네트워크분석

- Point: 커뮤니티 탐지, ERGM

2021.10 ~ 2021.12 | 아파트 경매가격 예측2

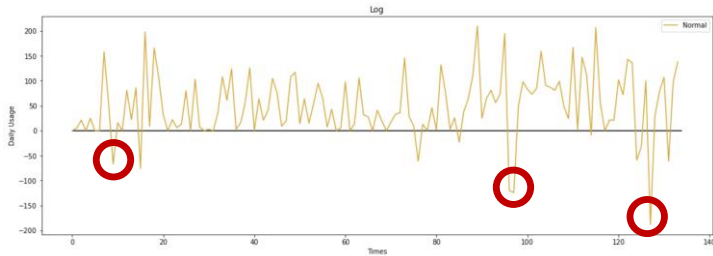
- Point: 독립변수 Oversampling

# Project #1. 효돌 (어르신 생활관리 AI 로봇인형)

## Problem

### 데이터 오류

센서의 고장 등으로 인해  
음수 또는 지나치게 큰 값(이상치)



## Solution

### iForest

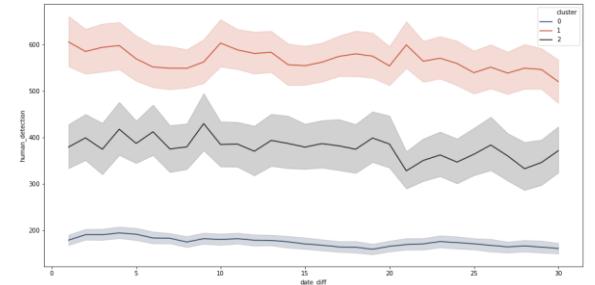
Random feature,  
Random Split을 이용한  
이상치 탐지 알고리즘

#### [ Process ]

1. 개인별 시계열 iForest 분석
2. MA(moving average) 계산
  - (1) 음수 → 음수: 제거
  - (2) 음수 → 양수: MA로 대체
  - (3) 이상치: MA로 대체

## Result

### 액티브 시니어 세부분류



1. 친밀감 위주 사용자
2. 귀기능 위주 사용자
3. 소극적 사용자

## My Role

EDA, 데이터 전처리 및 일부 통합, 시각화

## What I Have Learned

이상치 또는 모델에 맞지 않는 데이터는 분석력을 흐리기 때문에 데이터 전처리는 필수적이다.  
올바른 분석으로 나아가기 어렵다. GIGO(Garbage In, Garbage Out)

# Project #2. 아파트 경매가격 예측

## GOAL

2년간의 서울과 부산 지역 아파트 경매물의 임차, 감정가, 유찰 횟수 등의 정보를 통해 낙찰가를 예측하는 것이 목적이다. 부동산 가격이 타겟변수이기에 잘못된 예측을 했을 때 RMSE\*의 변동폭이 상당히 크다는 어려움이 있었다. 그래서 성능 향상을 위해 다양한 시도를 해보았다.

“1등 대비 1.95배 RMSE”

### 1. Bayesian Linear Regression

Bayesian Model Averaging (BMA)을 통해 최종 변수 선택

회귀계수가 분포를 갖는 모수로 판단.  
Conditional probability, g-prior, BMA  
과정 모두를 직접 일일이 코딩하였다.

### 2. Platt's Probability

가변수를 그대로 활용하지 않고,  
각 가변수를 종속변수로 하여 로지스틱  
회귀분석을 하여 cutoff를 통과하기 전의  
확률값으로 대체해주었다.

예를 들어, Seoul이라는 변수가  
1일 경우에는 서울이라는 의미이고,  
0일 경우 부산이라는 의미라면,  
해당 변수의 값을 '서울일 확률',  
즉 0과 1 사이의 값으로 대체해주었다.

“1등 대비 1.22배 RMSE”

### 3. 독립변수 Oversampling

명목형 독립변수를 타겟변수로 보고  
SMOTE를 통해 Oversampling하는 방법

기존의 불균형데이터는 타겟변수의  
불균형도에만 초점을 맞추었다.  
그러나 타겟변수의 불균형도가  
독립변수의 불균형도에도  
영향을 미친다는 연구결과가 있다.  
그렇다면 이를 보정해주었을 시  
성능향상을 기대해볼 수 있다.

\*RMSE(Root Mean Squared Error)는 작을수록 좋다.



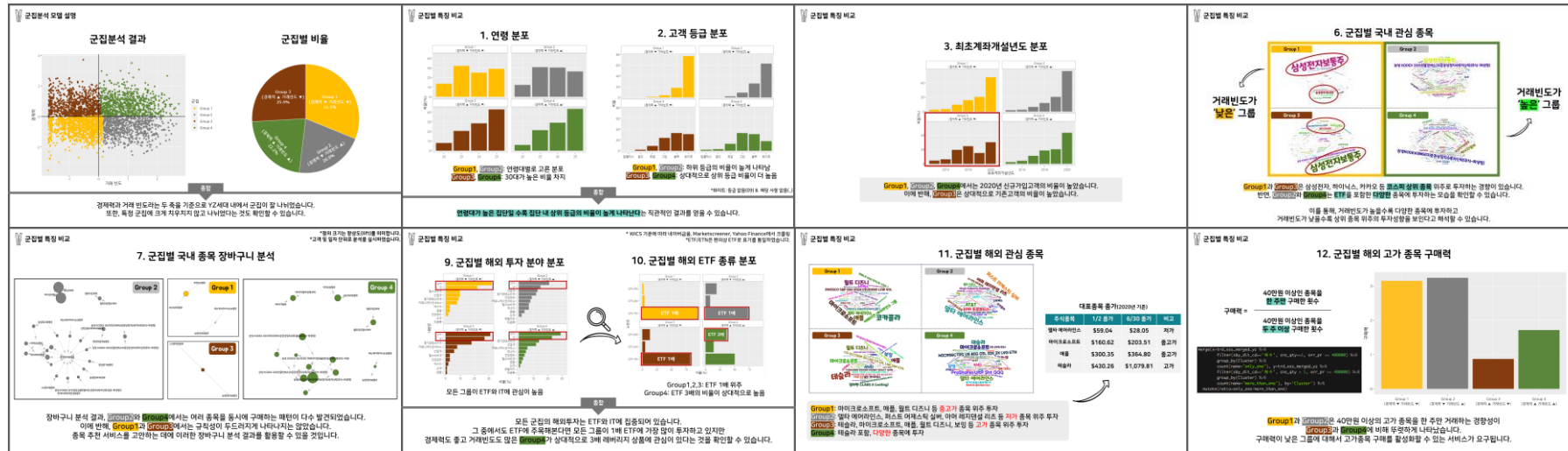
# Project #3. NH투자증권 YZ 투자자 시각화 대회

## Data Description

2020년 급격하게 늘어난 2030 투자자들에 대한 분석을 하고 이를 시각화하기 위한 데이터이다.

## Clustering & Visualization

네 집단으로 구분하여 특징을 살펴보았다. Random Forest와 Factor Analysis를 활용하였다.



## Service Suggestion

집단별 특징을 요약하였으며, 집단별로 적합한 서비스를 최소 하나씩 제안하였다.



# Thank you for reading my portfolio.

## Contact.

| Phone.  
010. 9120. 6675

| E-Mail.  
[bungaedm@naver.com](mailto:bungaedm@naver.com)

| Blog  
[통계학/ML] [jiwooblog.netlify.app](http://jiwooblog.netlify.app)  
[심리학] [blog.naver.com/bungaedm](http://blog.naver.com/bungaedm)

| KakaoTalk  
@bungaedm

| Instagram  
@5on\_jiwoo