



# 계속해서 도전하는 데이터 엔지니어 이슬기입니다

Lee Seul Gi

Data Engineer

서경대학교 컴퓨터공학과 졸업

Phone. +82 010-7184-3589

E-mail. tmfrl3589@gmail.com

GitHub. <https://github.com/stacy-sg>

## Education & Experience

- 2022.04.04. 서경대 Data Research Lab  
~ 2024.01.15. 학부연구원(2022)/선임연구원(2023)
- 2020.02.01. Business Foundation -Co-op  
~ 2021.02.01. SELC CAREER College (Vancouver/Canada)
- 2021.02.01. 컴퓨터공학과 로봇스터디  
~ 2022.11.30. 스터디원

## Awards

- 2023.06.02. 2023년 한국정보기술학회 대학생  
논문경진대회 우수논문상(은상)
- 2022.11.25. 2022년 융복합지식학회 추계학술대회  
대학생 논문경진대회 우수논문상
- 2021.11.17. 제 19회 임베디드 SW 경진대회[지능  
형 휴머노이드] 입선

## Certification

- 2024.09.10. 정보처리기사
- 2024.04.05. SQL개발자(SQLD)
- 2024.03.22. 데이터 분석 준전문가(ADsP)
- 2023.10.01. TOEIC SPEAKING TEST(IM3)
- 2022.09.23. 리눅스마스터 2급

## Skills

- Python ● ● ● ● ●
- MySQL,  
JavaScript ● ● ● ● ●
- R,  
PostgreSQL ● ● ● ● ●
- C, XML ● ● ● ● ●
- Azure, Java ● ● ● ● ●
- Git ● ● ● ● ●
- Linux, Notion ● ● ● ● ●

- 1 - 기본적인 코드를 읽을 수 있음, 레퍼런스를 참고하여 수정  
작업 가능
- 2 - 시스템 동작 방식을 알고 기본적인 기능 구현 가능
- 3 - 중간 규모의 프로그램을 설계 및 개발 가능, 주요 이슈 트  
러블 슈팅 가능
4. 대규모 프로그램 및 시스템을 레퍼런스 없이 개발 가능

# Project & Competition

## 1. 스포츠 경기 결과 예측 및 선수 데이터 통계 시각화 서비스(Github)

2023.03.02. ~ 2023.11.23. 개인프로젝트

### 담당 개발 내용

1. 데이터
  - 데이터 크롤링을 활용한 데이터 수집
  - 데이터 전처리
  - 데이터 시각화 진행
2. AI
  - XGBoost 알고리즘을 활용한 예측 모델링
3. Back/Front end
  - Flask를 활용한 기능 구현
  - 메인, 결과, 시각화 3가지 화면 구성
  - JavaScript를 활용한 그래프 화면 구성

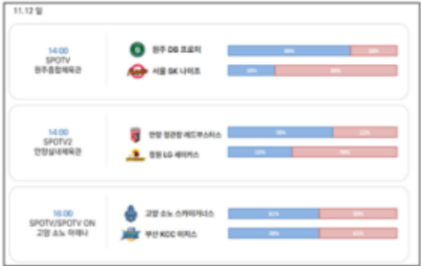
### 이슈

문제1. 데이터 수집 시, 데이터가 없는 날짜의 경우 오류 발생  
해결방안1. 데이터 없는 날짜는 예외처리

문제2. 결과 예측 정확도가 저하되는 현상  
해결방안2. 시계열 데이터라는 점을 이용하여 트렌드 반영을 위해 데이터 입력값 변경

### 배운 점

1. 혼자서 처음으로 프로젝트를 수행하면서 모든 과정을 직접 경험하고 발생한 문제들을 해결해 나가면서 데이터의 전체 사이클, 예외 처리, 예측 모델링 등 다양한 능력을 키울 수 있었습니다.
2. 백엔드와 프론트엔드 분야에서 기능 구현 경험이 부족했지만, 이번 프로젝트를 통해 자바스크립트, HTML 등의 언어를 활용해볼 수 있었습니다. 이 과정에서 데이터와 백엔드 기능, 프론트엔드 간의 연결 과정을 더 잘 이해하게 되었고, 자바스크립트에 대한 이해도도 향상되었습니다.
3. 정해진 예측 모델 알고리즘이 아닌 해당 프로젝트에 맞는 알고리즘을 찾기 위하여 다양한 알고리즘에 대하여 공부하고 적용해보는 기회였습니다.



### 개발 환경

언어	Python3.8, HTML/CSS, JavaScript
프레임워크 및 라이브러리	Flask, BeautifulSoup, Selenium, Matplotlib, Plotly, Scikit-learn etc
AI	XGBoost
협업	Linux, Window

# Project & Competition

## 2. 배달팁, 과연 합당한가(Github)

2022.11.01. ~ 2022.12.23. 팀프로젝트(3인)

- 배달 업체에서 책정하는 배달팁이 합당한지 분석 -

### 담당 개발 내용

1. 데이터
  - CSV 데이터 수집
  - 결측치 제거, 고유 번호 생성 등 데이터 전처리
  - 데이터 시각화 진행
2. AI
  - Kmeans를 활용한 배달팁 클러스터 분석

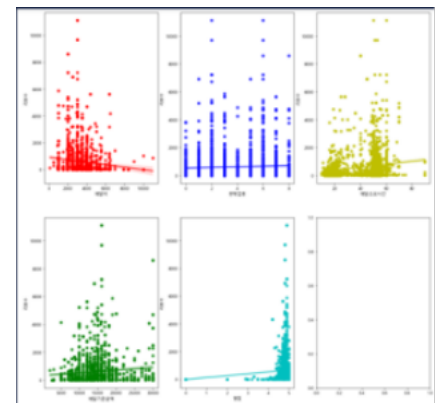
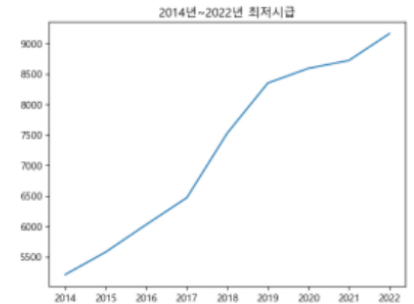
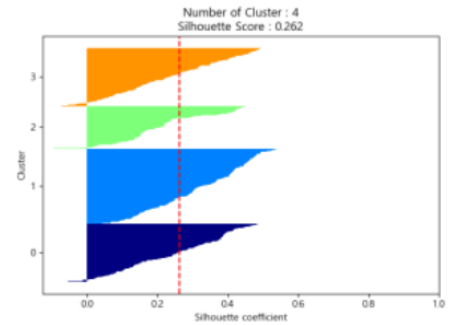
### 이슈

문제1. 데이터 분석 결과와 예상 값이 다름  
해결방안1. 근본적으로 배달팁을 측정할 때 핵심 요소를 다시 정함

문제2. 결과 예측 정확도가 저하되는 현상  
해결방안2. 배달팁 측정 핵심 요소를 다시 정함

### 배운 점

1. 군집분석에 대해서 경험이 부족하였지만, 이번 프로젝트를 통하여 Scikit-learn이라는 라이브러리를 처음으로 공부하고 구현해보았습니다. 군집 분석을 통하여 결과를 어떻게 해석해야하는가도 배웠습니다.
2. 데이터 분석 결과와 예상 값이 다르게 나오면서 당황하였지만 이 부분에서 근본적인 문제를 찾아 나가고 어떻게 해결해야하는지를 배웠습니다. 데이터 분석이란 저희가 보지 못한 인사이트를 도출할 수 있는 것을 깨닫게 되었습니다.
3. 회귀 분석을 하면서 상관도 분석을 같이 하게 되면서 결과값을 설명할 수 있는 부분을 직접 찾아볼 수 있어 회귀 분석과 상관도 분석의 중요도를 깨닫게 되었습니다.



### 개발 환경

언어	Python3.8
프레임워크 및 라이브러리	BeautifulSoup, Selenium, Matplotlib, Seaborn, Scikit-learn et c
AI	Kmeans, Linear Regression(Scikit-learn)
협업	Github, Google Doc, Colab

# Project & Competition

## 3. 지선 버스의 배차 쏠림 현상 원인 도출

2022.09.08. ~ 2022.10.27. 팀프로젝트(3인)

- 서경대학교 주변을 배회하는 지선 버스의 배차 쏠림 현상 분석 -

### 담당 개발 내용

1. 데이터
  - CSV 데이터 수집
  - 데이터 전처리
  - 데이터 시각화 진행

### 이슈

문제1. 필요하지만 검색되지 않는 데이터(강의 과목 데이터, 도로 별 좌표)

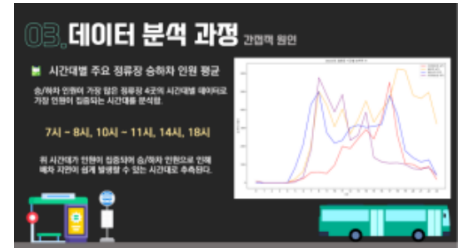
해결방안1. 프로젝트를 위해 관련 홈페이지에서 직접 수집

문제2. 정릉로 데이터를 구할 수 없었음

해결방안2. 정릉로와 연결되어 있는 4개의 도로 교통량 평균을 구하여 유추

### 배운 점

1. 데이터를 단순 csv 다운로드 후, pandas를 통한 활용이 아닌 공공데이터 api를 사용하는 방법을 학습하였습니다.
2. 단순한 데이터 분석에서 끝나는 것이 아닌 직접적인 문제와 간접적인 문제를 해결할 수 있는 방법까지 제안해보았습니다.
3. 간접적인 원인과 이를 해결하는 방법을 찾는 데에 있어서 회의의 중요성을 깨닫게 되었습니다. 회의에서 서로의 의견을 통하여 문제를 해결하는 능력을 기르게 되었습니다.



### 개발 환경

언어	Python3.8
프레임워크 및 라이브러리	pandas, numpy, Matplotlib, Seaborn, Scikit-learn etc
협업	Google Doc, Colab

# Project & Competition

## 4. 롯데멤버스 빅데이터 공모전 2022

2022.06.20. ~ 2022.08.12. 팀프로젝트(3인)

- '롯데' 유통사의 200만건 데이터 분석, 개인화 상품 추천 시스템 구현 -

### 담당 개발 내용

1. 데이터  
외부 데이터 수집  
결측치 보완, 중복값 확인 등 데이터 전처리  
데이터 시각화 진행  
EDA를 통한 패턴 분석

2. 마케팅  
• SWOT 분석  
• 마케팅 방법 제시

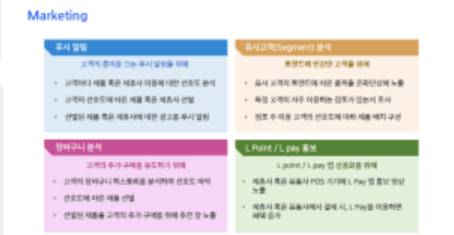
### 이슈

문제1. 팬데믹(코로나19)로 인한 데이터 격차로 생긴 정확도 이슈  
해결방안1. 코로나19 관련 데이터 수집 후, 테이블을 만들어서 병합

문제2. 예측 정확도 향상을 위한 의견 충돌  
해결방안2. 상호 존중 하에, 서로의 의견을 편하게 어필할 수 있는 시간을 가졌음

### 배운 점

1. EDA를 처음 활용하면서 패턴 분석을 위하여 여러 가지 데이터 테이블을 병합하여 다차원 그래프를 그려보았습니다. 이 과정에서 EDA 분석 방법과 다차원 그래프를 그리는 방법을 배웠습니다.
2. 단순한 데이터 분석이 아닌 이를 이용한 SWOT 분석과 마케팅 방법 제시까지 연결할 수 있는 것을 배웠습니다.
3. 팀 프로젝트에서 의견 충돌이 발생하였을 때, 의사소통을 하고 이 과정에서 상호 존중을 하는 것이 가장 중요하다는 것을 깨닫게 되었습니다. 협력을 함으로써 공동의 목표를 이루는 데 크게 기여할 수 있는 가치라는 것 또한 배웠습니다.



## 개발 환경

언어	Python3.8
프레임워크 및 라이브러리	pandas, numpy, Matplotlib, Seaborn, Scikit-learn etc
AI	Apriori, Collaboration Filtering
협업	Google Doc, Colab

## Project & Competition

### 5. 미니드론 자율 비행 경진 대회2022(Github / 본선 진출)

2022.04.20. ~ 2022.07.13. 팀프로젝트(2인)

- 영상 처리를 이용한 드론 미션 수행 -

#### 담당 개발 내용(개발 총괄 및 팀장)

- 영상처리
  - hsv를 이용하여 원 식별
  - CannyEdge를 통한 원 둘레 식별
  - rgb 인식
- 드론 알고리즘 개발
  - 원링 통과 알고리즘
  - 마크 인식 알고리즘

#### 이슈

문제1. 정확한 원의 좌표가 나오지 않음  
해결방안1. hsv를 이용하여 원만 판별하고 평균값을 내려 좌표를 출력

문제2. Github와 Gitbash 이해도 부족으로 인한 활용 불가  
해결방안2. 프로젝트 기간동안 주 1회 Github와 Gitbash 학습과 실습 시간을 가짐

#### 배운점

- Github와 Gitbash를 사용하는 기초적인 방법을 학습하였습니다.
- 제가 만든 알고리즘에 대한 흐름도 작성과 설명을 할 수 있는 능력을 향상시켰습니다.
- MATLAB을 처음 활용하면서 새로운 언어를 배우게 되었습니다.
- 기본적인 영상처리 방법은 알고있었지만, 실시간으로 카메라를 이용하여 실시간 영상처리에 대한 능력을 키웠습니다.

### 6. 제 19회 임베디드 소프트웨어 경진대회[지능형 휴머노이드](입선 수상)

2021.06.21. ~ 2021.11.17. 팀프로젝트(4인)

- 임베디드 보드와 제어보드, 센서를 통한 휴머노이드 모션 알고리즘 개발 및 모터 제어 -

#### 담당 개발 내용

- 모션제어
  - 사물 집어 올리기 모션
  - 안정적으로 걷기 모션
  - 문 열기 모션
- 모션 알고리즘 개발
  - 시퀀스미션 알고리즘
  - 문열기 알고리즘
  - 두리번거리기 알고리즘

#### 이슈

문제1. 개발하던 알고리즘 날라감  
해결방안1. 두뇌 보드에 저장되어 있던 데이터 복구 작업

문제2. 불안정한 모션으로 로봇 넘어짐 발생  
해결방안2. 로봇 관절 값 수평 맞추기, 다양한 모션 수행을 촬영하고 가장 안정적인 모션으로 진행

#### 배운점

- roboBASIC2.8을 처음 활용하게 되면서 새로운 언어를 학습하였습니다.
- 개발 시, 백업의 중요성을 깨닫게 되었습니다. 개발 코드 백업을 진행하지 않고 두뇌 보드에서 바로 개발 진행하다가 데이터가 손실된 경험을 하였습니다. 경험 이후에는 항상 백업을 진행하고 있습니다.
- UART 시리얼 통신에 대하여 학습하게 되었습니다. 두뇌보드에서는 카메라를 이용해 영상 처리를 하고 이를 데이터화하여 제어보드로 전송하고 제어보드에선 수신한 데이터를 통하여 적절한 모션을 취합니다.

# Research Publications & Contributions

## 1. Prophet 모델을 이용한 일일 지하철 이용객 인원 예측(은상)

- **저자** : 고희석, 이슬기, 이동규, 이준석, 하정원, 김재현
- **발표** : 한국정보기술학회 하계종합학술대회, 2023
- **요약** : COVID-19 팬데믹 이후 일상생활로 복귀함에 따라 지하철 이용객 수의 변화가 나타나고 있으며 특히 주중과 주말 이용객 수의 차이가 상업지역에서 더 두드러진다. 본 논문은 3개월간의 지하철 이용 데이터를 기반으로 LSTM과 Prophet 기법을 적용해 예측 정확도가 높은 모델을 제안한다.

## 2. TSlearn 알고리즘을 적용한 미세먼지 데이터 클러스터링 분석

- **저자** : 하정원, 이슬기, 이동규, 고희석, 이준석, 김재현
- **발표** : 한국전기전자학회 하계학술대회, 2023
- **요약** : 서울시 25개 자치구의 8년간 미세먼지 데이터를 활용하여 K-means 클러스터링(클러스터 개수 5)을 수행, 미세먼지 농도 패턴을 지역별로 분류하고 시각화를 통해 각 클러스터의 특성을 분석하였다. 이를 통해 지역별 미세먼지 특성을 도출하고 환경적 대책 마련에 기여할 수 있는 정보를 제공하였다.

## 3. XGBoost 기법을 활용한 서울 권역별 미세먼지 데이터 예측

- **저자** : 이동규, 이슬기, 하정원, 고희석, 이준석, 김재현
- **발표** : 한국전기전자학회 하계학술대회, 2023
- **요약** : 서울시 25개 자치구를 4개 권역으로 구분하고, 8년간의 미세먼지(PM10) 데이터를 정규화(minimax 변환)하여 시계열 패턴을 분석하였다. XGBoost 기법을 활용해 권역별 미세먼지(PM10) 수치를 예측함으로써 미래 미세먼지 수준과 등급을 예측하고, 효과적인 정책 수립에 기여할 수 있는 정보를 제공하였다.

## 4. 학교 근접 측정소 미세먼지 데이터 시각화 웹 서비스(우수논문상)

- **저자** : 이슬기, 이지선, 박관희, 김민우, 고희석, 김재현
- **발표** : 융복합지식학회 추계학술대회, 2022
- **요약** : 어린 학생들이 재학 중인 학교 인근의 미세먼지 정보를 효율적으로 관리하기 위해, 학교 기본 정보와 에어코리아에서 제공하는 실시간 측정 데이터를 결합하여 시각화하는 웹 서비스를 제안하였다. 이는 모든 학교가 실시간 미세먼지 측정기를 보유하지 않은 현실을 보완하고, 학교 미세먼지 관리의 효율성을 높이는 데 기여하고자 한다.

## 5. ALS 협업 필터링을 적용한 상품 구매 데이터 분석

- **저자** : 문성욱, 이슬기, 민현기, 조영완, 김재현
- **발표** : 융복합지식학회 추계학술대회, 2022
- **요약** : 추천 시스템 중 ALS(Alternative Least Square)를 적용한 사용자 기반의 협업 필터링(User based Collaboration Filtering)을 활용하여 고객 정보와 장바구니 구매 데이터를 분석하고, 이를 토대로 특정 고객에게 적합한 상품 추천 목록을 제시한다.