

# PORTFOLIO

자기소개 포트폴리오

임정민

# Jeongmin Lim



- 동국대학교 통계학 & 데이터사이언스 졸업 예정 (21년 8월)
- 대학수학능력시험 성적 우수 장학생
- 동국대학교 데이터분석 학회 비어플 소속
- SQL 개발자 자격증 소지 (SQLD)
- 데이터분석 준전문가 자격증 소지 (ADsP)
- 튜링바이오 연구소 인턴 근무 (6개월)
- AI 팀프로젝트 최우수상
- Python, R, SAS, SPSS, SQL, MS Office

# Experience

동국108리더스

19.03

19.05

탐색적자료분석  
프로젝트

19.10

데이터마이닝  
프로젝트

19.11

데이터사이언스  
교육 프로그램 이수

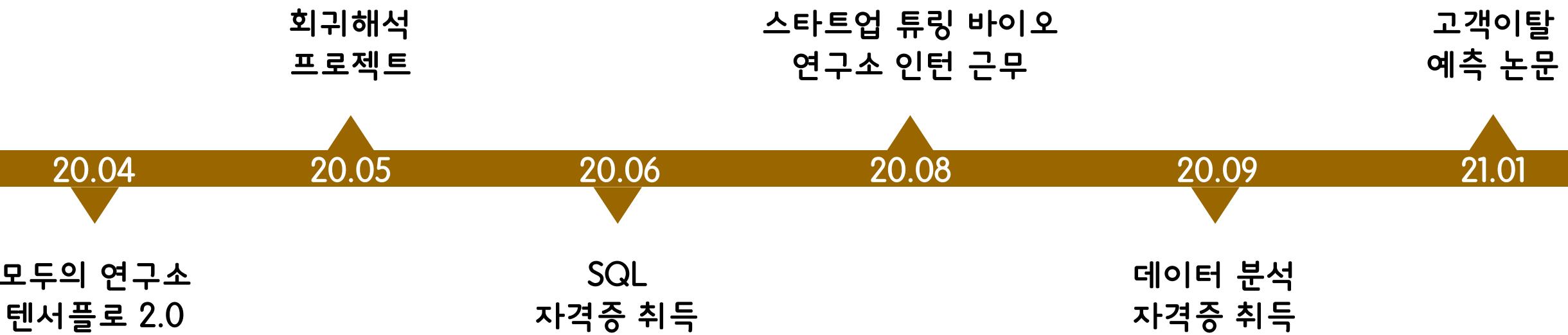
비어플 데이터분석  
프로젝트

19.12

20.01

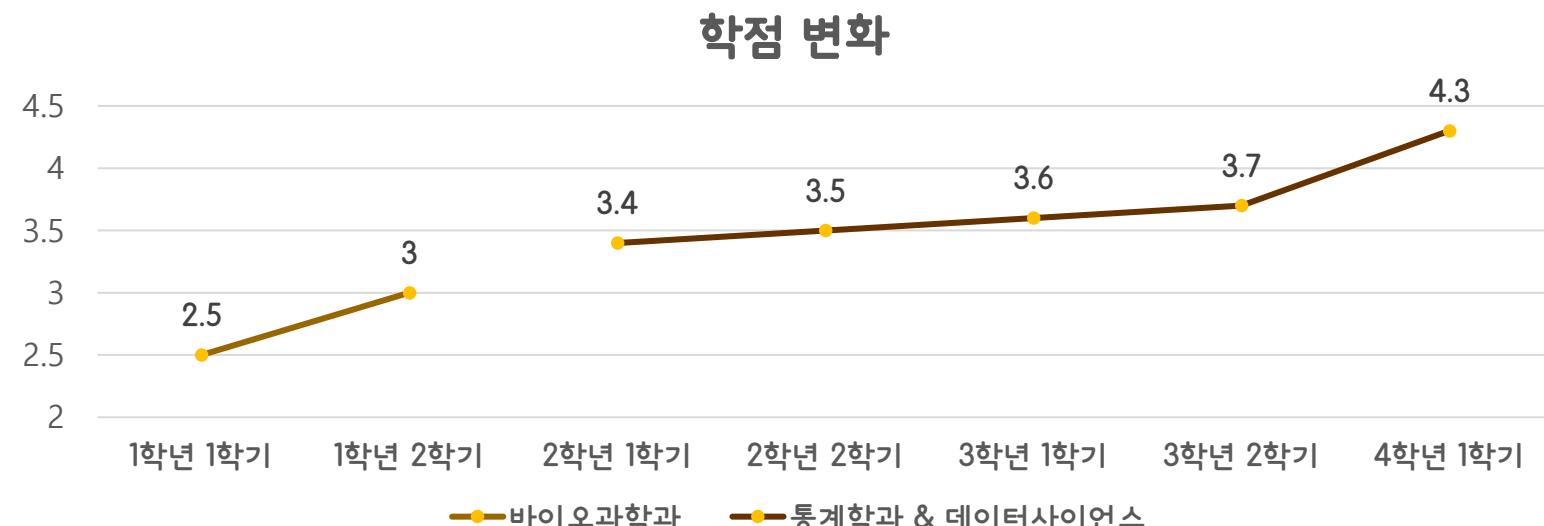
데이터사이언스 과정  
AI 프로젝트 최우수상

# Experience



# 통계학과 전과 & 데이터사이언스 복수전공

- 대학수학능력시험 우수 장학금을 받고 바이오과학과 입학
- 확률론 및 선형대수 독학으로 통계학과 전과 ➔ 장학금 포기
- 통계학, 데이터사이언스 소프트웨어 복수전공 ➔ 학점 상승



## 동국 108리더스

- 동국 108 리더스란?

리더를 양성하기 위해 동국대학교 3학년 학생들 중 108명을 선정하여  
발표, 면접, 토론, 멘토링, 진로 선택 등의 교육을 받는 단체

- 1년 동안 경험한 대표 활동은?

리더십, 자기관리 교육 이수

MS 오피스 교육 이수

스피치 교육 이수

직업 가치관 그룹 코칭

단체봉사활동



## 데이터마이닝 프로젝트

- 첫 데이터 분석 프로젝트

- 보험회사 CRM Data

- 분석목표

- 1) 보험 구매 고객 특성 파악

- 2) Cross Selling

- Python & SAS

- (주사용 패키지: pandas, sklearn, seaborn 등)

- 주요 역할 : 데이터 시각화, 보고서 작성, 발표

## 탐색적자료분석 프로젝트

- 전처리 및 EDA 프로젝트

- 날씨 및 미세먼지 데이터

- 분석목표

- 1) 미세먼지에 영향을 끼치는 기상 특징 발견

- 2) 미세먼지 감소 증거 찾기

- R

- (주사용 패키지: dplyr, stringr, fpc, cluster 등)

- 주요 역할 : 프로그래밍, 보고서 작성, 발표

# 데이터사이언스 교육 프로그램 이수

- 데이콘 주관 데이터사이언스 오프라인 교육
- 파이썬을 이용한 데이터 전처리, 시각화, 모델 구축
- 매주 새로운 데이터로 분석 실습
- feature 생성, 파라미터 튜닝으로 모델 성능 향상 훈련
- 타이타닉 데이터, 웹 데이터, 서울시 따릉이 데이터 분석



Dacon Education Certificate

HAS BEEN AWARDED TO  
임정민

임정민 HAS SUCCESSFULLY COMPLETED IN  
데이콘 오프라인 교육

DATE  
2019.11.18~2019.12.09



PORTFOLIO 임정민

# 비어풀 데이터분석 프로젝트

\* 코드 및 발표자료: <https://github.com/datajudy20/jejubus>

- 동국대학교 유일 데이터분석 학회 비어풀
- 분석 이론, 시각화, 머신러닝, 프로그래밍 세미나 진행
- 데이터 분석 팀 프로젝트 진행 ➔ 비어풀 내의 2등 팀으로 선정
  - 제주시 버스 데이터 분석
  - feature engineering으로 10개 파생 변수 생성
  - 최근 시간 버스 승차인원 예측 모델 구축
- Python & R
  - (주사용 패키지 : pandas, sklearn, seaborn 등 / dplyr, ggplot2 등)
- 주요 역할 : feature engineering, Modeling



# 데이터사이언스 지능소프트웨어 과정 수료

\* 코드 및 발표자료: <https://github.com/datajudy20/AINADU>

- 동국대학교 융합교육원 주관 빅데이터 교육 및 프로젝트 진행
- 교내에서 선발한 AI 인재 30명 중 한 명으로 선발됨
- 인공지능 프로젝트 A!NADU ➔ **프로젝트 대회 최우수상 (1등)**
  - 객체 인식을 이용하여 영상 속 인물 모자이크 처리 자동화 프로그램 개발
  - Ultra Light Fast Generic Face Detector 활용
- Python (사용라이브러리 : pandas, sklearn, seaborn, opencv, pytorch 등)
- 주요 역할 : 자료수집, 데이터 분석 프로그래밍, 핵심 기능 보완, 발표



# 데이터사이언스 기반 지능소프트웨어 과정 수료



# 모두의 연구소 풀잎스쿨 텐서플로 2.0 과정

- 모두의 연구소에서 진행하는 풀잎스쿨 텐서플로 2.0 과정 수료
- TensorFlow 2.0으로 구현한 신경망 코드 공부



PORTFOLIO 임정민

# 회귀해석 프로젝트

\* 코드 및 발표자료: <https://github.com/datajudy20/lifespan>

- 회귀분석 및 요인분석 프로젝트
- WHO에 의해 수집된 193개국 기대수명 데이터 & UN에 의해 수집된 각 나라의 경제 및 사회적 요인 데이터
- 분석목표
  - 1) 선진국 여부에 따른 국가 기대수명 예측
  - 2) 아시아 국가의 기대수명에 영향을 주는 요인 분석
- Python & R
  - (사용라이브러리 : pandas, sklearn, seaborn, scipy 등)
- 주요 역할 : 프로그래밍, 보고서 작성

회귀분석을 통한  
기대수명 예측 및 요인 분석  
- 아시아 국가 중심으로 -

회귀해석

2020년 6월 26일

동국대학교 이과대학 통계학과

201711 유  
2017111748 임정민

PORTFOLIO 임정민

# 인턴 근무

2020.08

## ● AI 기반 바이오-헬스케어-의료 스타트업 인턴 근무 (6개월)

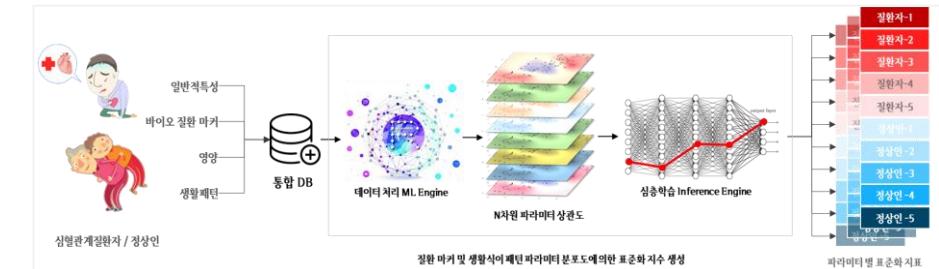
## ● 판교 소재 (주) 투링바이오 - 연구소

## ● 직무 : 의료-헬스케어 인공지능 시스템 개발

## ● 성과 : 을지대학교 관련 프로젝트 머신러닝 모델 설계

목표 - 심혈관 질환 중증도 분류    ➡    연구 성과 홈페이지 게재

\* 투링바이오 URL 참조 : <http://www.turingbio.com/service/clinical-decision-support/>



### 표준화 지수 정량화 개발 절차

#### KNN\* Modeling

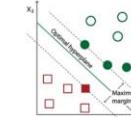
정상인 / 심혈관계질환자 분류



심혈관계질환자와의 유사도를  
기반으로 하는 분류 모델

#### SVM\* Modeling

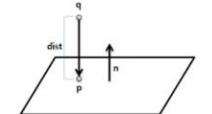
질환/건강 판별 기준 구축



정상인과 심혈관계질환자를  
구분하는 평면 구축

#### Clustering

질환/건강 척도 예측



평면(직선)으로부터 떨어진 거리를 기반으로  
질환/건강 척도 정량화

# 고객 이탈 예측 논문 2021.01

- Ensemble 모델을 활용한 월정액 서비스 가입 고객 이탈 예측 연구 - MMORPG 유저를 중심으로
- 논문에서 공개된 WOW Avatar History Dataset 사용 (유저의 로그데이터)
- 앙상블 모델 제안 – KNN, SVM, Random Forest, XGBoost 모델 soft-voting
- Feature engineering으로 모델 성능을 개선하는 9개 파생 변수 생성
- 이탈에 영향을 미치는 핵심 요인 3가지 도출
- Python  
(사용라이브러리 : pandas, sklearn, seaborn, scipy, xgboost 등)

감사합니다