Statistician

통계분석학회
P-SAT

27기 신입학회원 모집

목차

1. P-SAT 소개

P-SAT이란?

"Power Statistical Analysis Technics"

성균관대학교 통계학과 소속 유일한 통계분석 학회로,
통계학적 전문성 확보를 통한 실제 데이터 분석능력 함양을 목표로 합니다.

14년째 활동을 이어오고 있으며, 그동안 축적해온 자료들로 분석 이론 공부와 분석을 해볼 수 있는 탄탄한 **커리큘럼**을 지니고 있습니다.

학회 내 프로젝트로 쌓은 실력을 기반으로 매년 많은 **공모전 수상**을 하고 있습니다.

데이터 분석 분야에 진출한 많은 선배님들과 **유기적 네트워크**가 형성되어 있습니다.

팀 구성

"TEAM" 지도교수님, 학회장/부학회장, 그리고 6개의 팀으로 구성되어 있습니다.
전체 학회원은 32명입니다.

지도 교수님

박세영 통계학과 교수님

주 연구 분야 :
고차원데이터 모형 분석,
머신러닝,
분위수 모형분석,
바이오통계

학회장/ 부학회장

- 1팀 딥러닝
- 2팀 범주형 자료분석
- 3팀 선형대수학
- 4팀 데이터 마이닝
- 5팀 시계열 자료분석
- 6팀 회귀분석

P-SAT

팀 구성

"TEAM" 6개 팀으로 구성되어 있으며, 각 팀은 데이터 분석에 필요한 이론들을 공부하고 학회원들에게 발표하여 분석 지식을 전문적으로 학습합니다 .

Deep Learning

1팀 딥러닝

비정형 데이터를 분석할 수 있는 딥러닝 모델들을 다룸. CNN, RNN등 다양한 모델들의 이론을 학습하고 어텐션 기법을 배움. 학습한 이론들을 실습을 통해 구현함

Categorical Data Analysis

2팀 범주형 자료분석

범주형 자료를 분석하는 방법들을 이론과 실습을 통해 다룸. 오즈비, 로지스틱회귀를 비롯한 GLM, 모델 평가지표와 인코딩 방법, 비대칭 데이터를 해결하는 샘플링 방법들을 학습하고 R을 통해 실습.

Linear Algebra

3팀 선형대수학

통계의 근간이 되는 선형대수를 공부함.선형대수 개념을 수식적, 공간적으로 익히고, 딥러닝, Least-Square 등의 원리를 선형대수의 응용 차원에서 이해함. 통계지식에 깊이를 더하는 것이 목표

Data Mining

4팀 데이터 마이닝

데이터 분석의 대표적인 방법론인 지도학습과 비지도학습을 주로 다룸. Tree 모델 전반에 대해 익히고 더불어 추천시스템의 작동 원리를 이해하고자 함. R과 Python을 통해 실제 데이터에 적용하여 실습함.

Time Series

5팀 시계열 자료분석

시간에 따라 관측되는 자료인 시계열 자료에 대한 분석방법을 다룸. 시계열분석의 개념과 다양한 정상, 비정상 시계열 모형을 학습하고 이를 R을 통해 실습.

Regression Analysis

6팀 회귀분석

데이터 분석의 가장 기본이 되는 회귀 모형의 이론과 기법에 대해 다룸. 회귀 모형의 특징과 가정, 변형에 대해 학습하고 R과 python을 통해 이를 구현하는 법을 실습.

P-SAT

학회 활동

"Curriculum" Clean up&Lead off 이후 주제분석을 하는 2-step 커리큘럼으로 진행하고 있습니다. 더불어 Package 과제와 Mentoring을 병행합니다.

1-step → 2-step

Clean up & Lead off	주제분석
6개 팀별로 해당 과목에 관한 이론을 깊게 공부 한 주 동안 스터디를 한 후 준비해온 것을 세미나 때 발표 ✓ 3주 과정	6개 팀별로 주제를 정하여 데이터 분석 진행 주제 선정부터 모델링까지 데이터 분석의 모든 과정을 주도적으로 기획 및 발표 ✓ 4주 과정

With

Package

R/Python 통계 패키지를 통해 스스로 데이터를 분석하고 유의미한 해석을 해보는 과제
✓ 매주 제출 및 학회장/부학회장 채점

With

Mentoring

매주 데이터 분석 직무에 취업하거나 통계 관련 대학원에 진학한 선배님의 멘토링 진행

2. 활동내역 및 수상내역

P-SAT

26기 주제분석

"Deep Learning"

1팀 딥러닝팀 주제분석 요약입니다.

주제

: 현재 취향에 기반한 다른 시대의 음악 추천

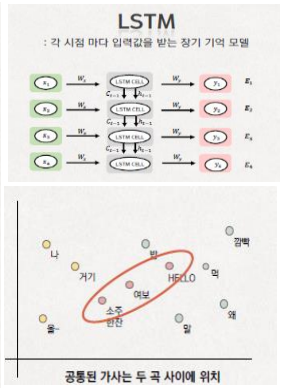
사용 데이터

: 카카오 아레나의 멜로디 데이터,
멜론 가사 + 차트 크롤링 데이터

분석 내용

: 가사 데이터를 CNN + LSTM, Word2Vec, Doc2Vec, Fasttext 모델을 기반으로 임베딩 벡터로 만들고, 멜로디 데이터를 VGG16 모델을 기반으로 임베딩 벡터로 만들어 유사도 행렬을 기준으로 추천 시스템을 구축함

** PPT 예시



27th Recruiting
2021

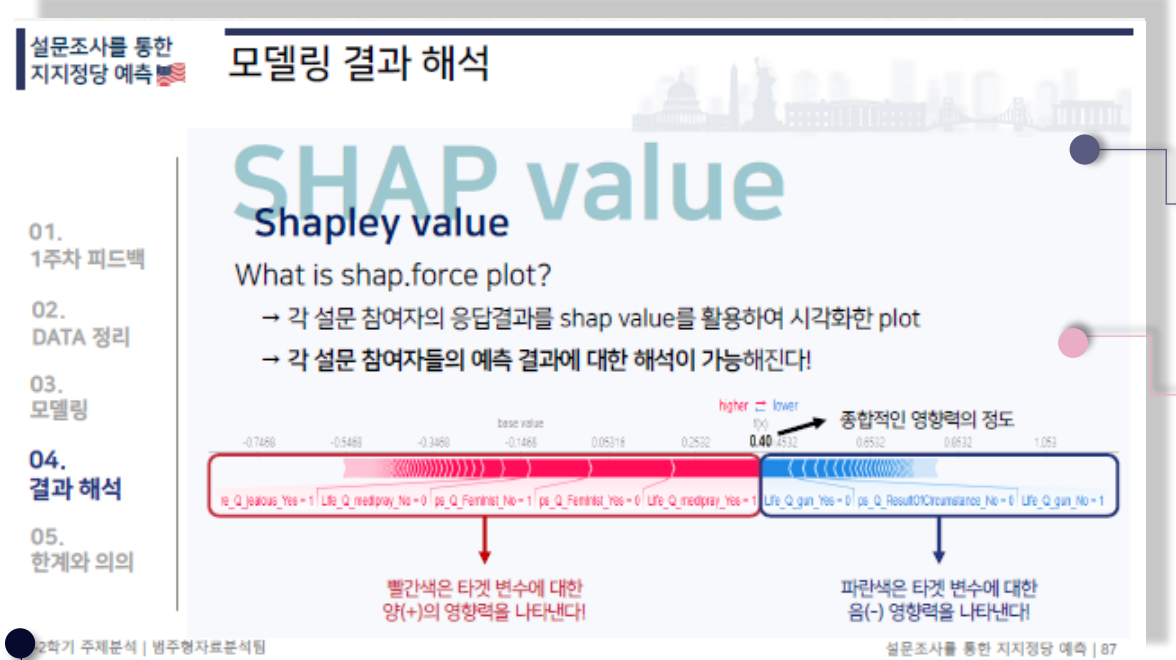
1. P-SAT 소개
2. 활동내역 및 수상내역
   - 활동내역
   - 수상내역
3. 리쿠르팅

P-SAT

26기 주제분석

"Categorical Data Analysis"

2팀 범주형 자료분석팀 주제분석 요약입니다.



주제

: 설문조사를 통한 지지정당 예측

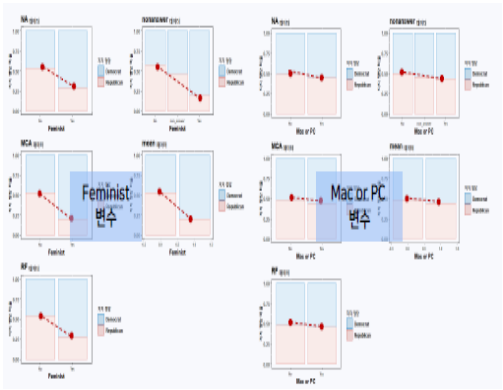
사용 데이터

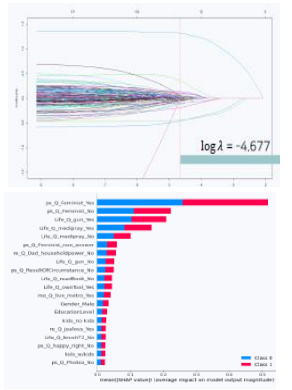
: 101개 질문에 대한 설문조사 데이터

분석 내용

: 설문조사 데이터를 이용하여 지지정당 예측력을 높이는 모델을 만듦.
MICE로 NA 를 처리하고 다양한 파생변수를 생성하여 데이터를 완성 했으며,
GLM, RandomForest, XGBoost, CATBoost, LGBM 모델을 만든 후
SHAP value를 이용하여 모델을 해석함

** PPT 예시





27th Recruiting
2021

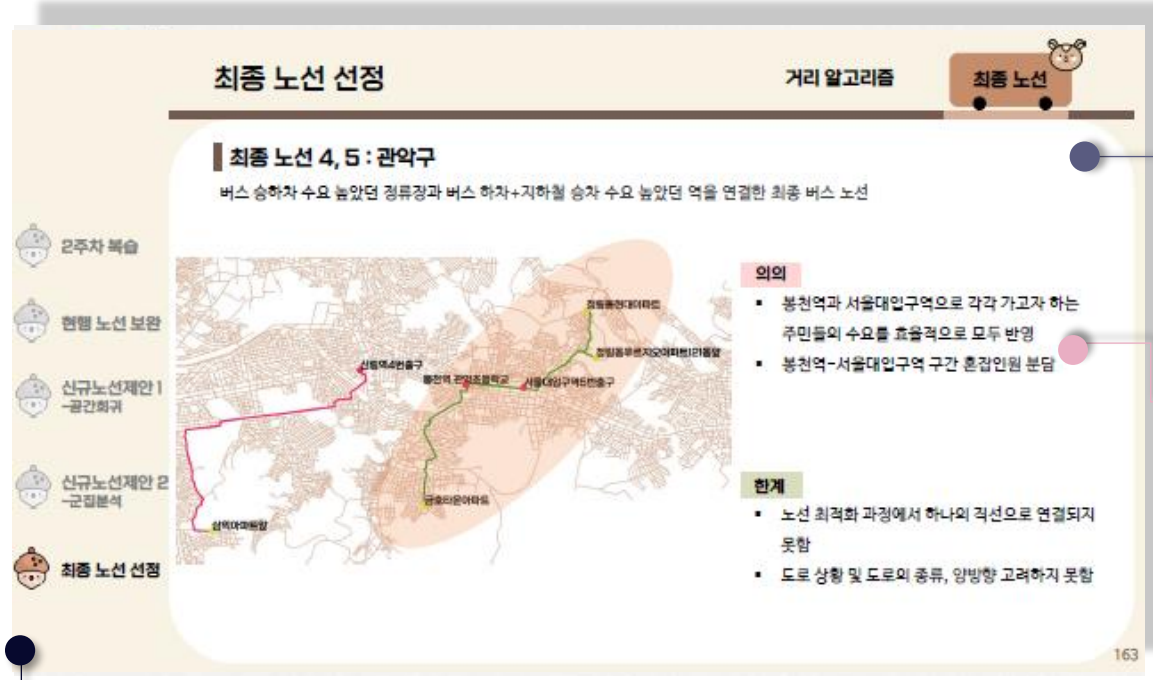
P-SAT

26기 주제분석

"Linear Algebra"

3팀 선형대수학팀 주제분석 요약입니다.

1. P-SAT 소개
2. 활동내역 및 수상내역
   - 활동내역
   - 수상내역
3. 리쿠르팅



주제

: 서울시 출근시간대 버스 노선 제안

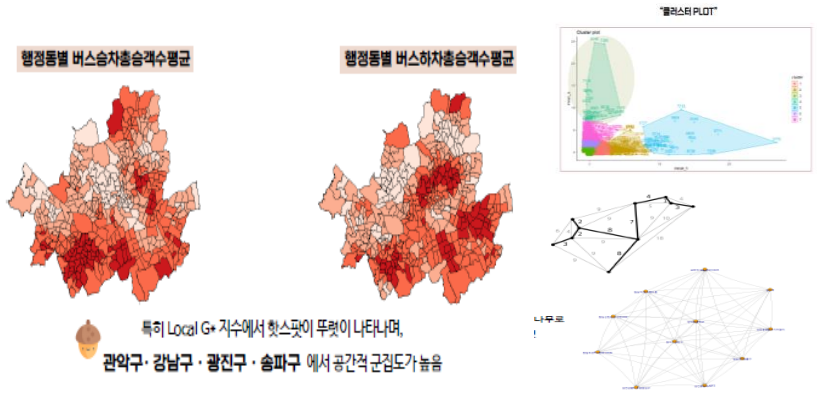
사용 데이터

: 버스 승하차, 유동인구 등 출근시간대 관련 및 서울시 지도데이터

분석 내용

: 서울시 출근시간대 버스인 '다람쥐버스'의 노선을 제안하기 위해 출근시간대에 높은 버스 수요 요인과 지역을 공간회귀와 클러스터링을 통해 알아낸 후, 최소신장나무 알고리즘과 다익스트라 알고리즘을 이용하여 노선을 설계함

** PPT 예시



P-SAT

26기 주제분석

" Data Mining "

4팀 데이터마이닝팀 주제분석 요약입니다.

주제

: 혼자 사는 청년들을 위한 서울시 살 곳 추천

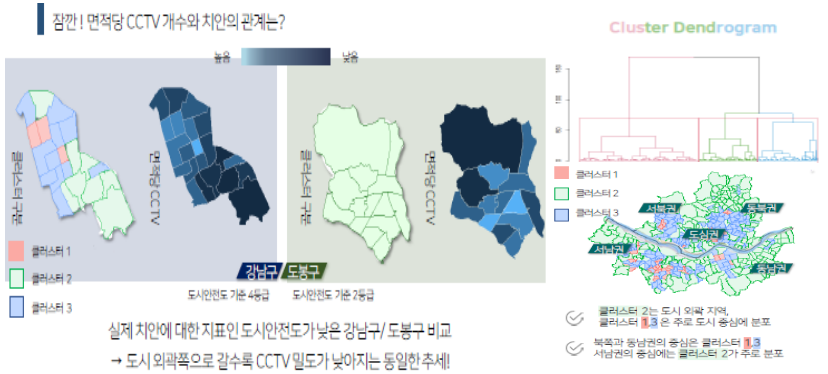
사용 데이터

: 시세, 유동인구, 인프라, CCTV, 범죄 발생 등 서울시 내 주거 관련 데이터

** PPT 예시

분석 내용

: 혼자 사는 청년이 살기 적절한 곳을 추천하기 위하여 주거 관련 데이터를 모아 PCA, Stepwise selection 등으로 다중공선성을 해결한 후, K-means, DBSCAN, Hierarchical 클러스터링 기법으로 지역 특성을 파악한 뒤 추천 알고리즘을 구현함



27th Recruiting
2021

1. P-SAT 소개
2. 활동내역 및 수상내역
   - 활동내역
   - 수상내역
3. 리쿠르팅

P-SAT

26기 주제분석

" Time Series "

5팀 시계열자료분석팀 주제분석 요약입니다.

주제

: 취향에 맞는 향수 추천

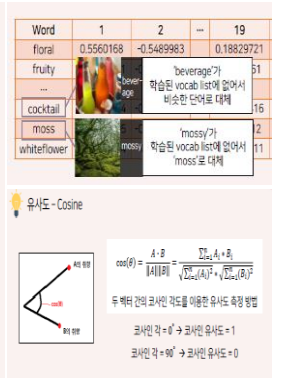
사용 데이터

: 향수 정보를 제공하는 웹사이트 크롤링 데이터

분석 내용

: 향수 정보 웹사이트에서 색, 특징, 향과 관련한 데이터를 크롤링한 후 K-means로 향수 이미지에서 대표 색을 추출하고, Word2Vec을 통해 향과 관련한 리뷰 데이터를 가공하여 데이터를 완성 한 다음, 유사도 행렬을 구축하여 추천시스템을 구현함

** PPT 예시

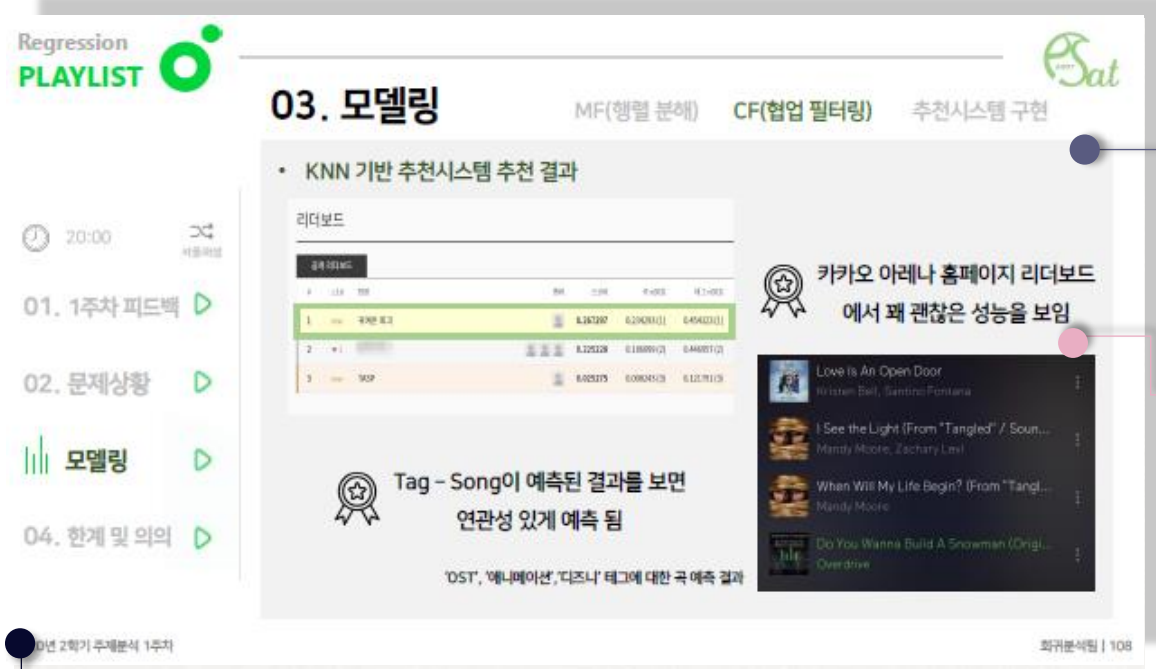


P-SAT

26기 주제분석

" Regression Analysis"

6팀 회귀분석팀 주제분석 요약입니다.



주제

: 음원 사이트 멜론 플레이리스트 예측 및 추천

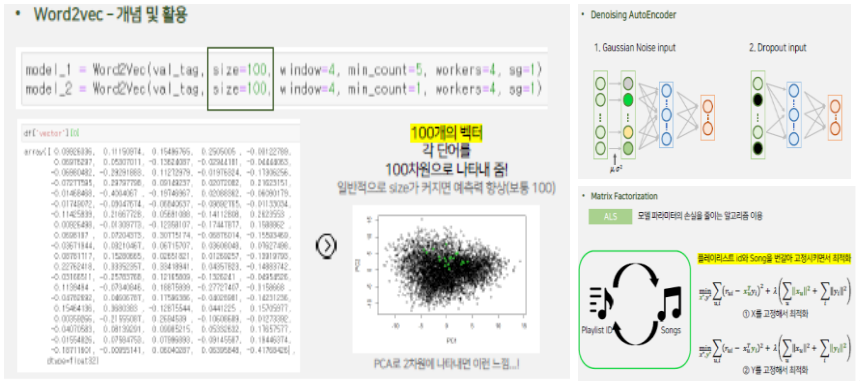
사용 데이터

: 멜론 곡 정보 및 플레이리스트 데이터

분석 내용

: 주어진 플레이리스트 수록곡과 태그를 예측하기 위해, 협업필터링 모델들을 사용해 추천시스템 구현함. 추천 과정에서 Matrix Factorization, Factorization Machine, Denoising AutoEncoder, Word2Vec, knn 등 다양한 모델들을 활용함

** PPT 예시



P-SAT

Package

"Package" 실제 데이터분석 시 도움되는 시각화, 전처리, 모델링, 해석과 관련된 난이도 있는 문제들을 해결하면서 코딩 실력을 기를 수 있습니다.

** 과제 제출 코드 예시(R/PYTHON)

Ch.2 문제풀이

문제1

```
set.seed(1)
n_split <-5
cv <- createFolds(train$click,k=n_split)
tune_parameter <- expand.grid(k=c(3,4,5))
tune_parameter$logloss <- NA

for(k in 1:NROW(tune_parameter)){
  logloss_result <- c()
  for( i in 1:n_split){
    idx <- cv[[i]]

    train_x<- train[-idx,]
    val_x <- train[idx,]

    set.seed(1)
    model_rf_2 <-randomForest(click ~., data=train_x,mtry=tune_parameter[k,'k'])
    prediction <- predict(model_rf_2, newdata=val_x,type="prob")

    logloss_temp <- LogLoss(prediction[,2],as.numeric(val_x$click)-1)
    logloss_result <- c(logloss_temp, logloss_result)
    result <- mean(logloss_result)

  }
  tune_parameter[k,'logloss'] <- result
}
tune_parameter
```

```
##   k   logloss
## 1 3 0.2594486
## 2 4 0.2628726
## 3 5 0.2611207
```

```
Audience_rating = []
Audience_rat_num = []
Net_rating = []
Net_num = []

#관람객 평점, 참여 수, 네티즌 평점, 참여 수

for i in range(0, 142):
    driver.get(url + str(link_df.loc[i, "movie_url"]))
    driver.find_element_by_xpath('//*[@id="movieEndTabMenu"]/li[5]/a')
    driver.find_element_by_xpath('//*[@id="movieEndTabMenu"]/li[5]/a').click()
    time.sleep(1)

    source = driver.page_source
    soup = BeautifulSoup(source, 'html.parser')

    #관람객
    audience_content = soup.select(".st_on")
    audience_num_content = soup.select("#actual_point_tab_inner > span")

    try:
        Audience_rating.append(audience_content[0].text)
        Audience_rat_num.append(audience_num_content[0].text)
    except:
        Audience_rating.append(" ")
        Audience_rat_num.append(" ")

    #네티즌
    content = soup.select("#netizen_point_tab_inner > em")
    netizen_num_content = soup.select("#graph_area > div.grade_netizen > div.title_area.grade_tit > div.sc_area > span > em")

    try:
        Net_rating.append(str(content[0].text) + str(content[1].text) + str(content[2].text) + str(content[3].text))
        Net_num.append(netizen_num_content[0].text)
    except:
        Net_rating.append(" ")
        Net_num.append(" ")
```

P-SAT

2020년 수상내역

"Prizes"

2020년에 총 7번 데이터분석 공모전 수상을 하였습니다.

최우수상	빅콘테스트 혁신아이디어 분야
서울특별시장상 (대상)	서울특별시 빅데이터 캠퍼스 공모전
3위	KB 국민은행 제 2회 Future Finance Ai Challenge
최우수상	한국정보화진흥원 데이터 크리에이터 캠프
우수상	포스트 코로나 데이터 시각화 경진대회 공모전
최우수상 (1위)	제3회 상권분석 빅데이터 경진대회
최우수상	빅데이터 활용 정책 아이디어 공모전

P-SAT

15-19년 수상내역

2019 IGAWorks BIG DATA Competition 입상
2019 제6회L.POINT Big Data Competition 우수상
2019 미래에셋대학생디지털금융페스티벌 대상/ 은상/ 동상
2019 디지털금융혁신과금융보안공모전 장려상
2019 KCB x 데이콘: 금융스타일시각화대회 5위
2019 기상청날씨빅데이터콘테스트 우수상
2019 제5회L.POINT Big Data Competition 대상/ 우수상(2)
2019 프로농구데이터활용경진대회 최우수상/ 우수상
2018 빅콘테스트 최우수상(신한은행)
2018 네이버Data Science Competition 3위
2018 미래에셋대우빅데이터페스티벌 대상

"Prizes"

2015-19년 5년간 총 24번
데이터분석 공모전 수상을 하였습니다.

2018 기상청날씨빅데이터콘테스트 우수상, 장려상
2018 한국수자원공사빅데이터공모전 입선
2018 성균C-School 프로젝트성과발표회 대상
2017 빅콘테스트SK텔레콤데이터사업 본부장상
2017 기상청날씨빅데이터콘테스트 우수상
2016 기상청날씨빅데이터콘테스트 최우수상
2015 SAS 분석챔피언십공모전 입선

3. 리쿠르팅

P-SAT

지원 자격

"Recruiting" 아래 세 조건을 만족하시는 분은 언제든 지원이 가능합니다.
조건을 반드시 숙지해 주시길 바랍니다.

통계학과 원전공/복수전공 &

✓ 이번학기는 학과차원의 요구로 통계학과 **원전공 및 복수전공생**만 지원 가능합니다.

1년의 의무기간 &

✓ 1년간 활동이 **반드시** 가능한 분
✓ **금요일 저녁 6시-9시** 세미나 참가 가능한 분
**온라인으로 진행할 예정이며,
팀 별로 **3시**부터 미리 세미나를 준비합니다.
✓ **주 2회 팀별 스터디** 참가 가능한 분
**확진자 추이 및 사회적거리두기 정책 상 소규모 스터디가
가능할 경우 혜화에서 오프라인 스터디를 진행합니다.

열정과 관심

✓ 데이터분석에 관심있는 분

P-SAT

지원 일정

"Recruiting" 아래 일정을 확인하시어 지원해주시길 바랍니다.

지원서 접수 기간

: 2021년 2월 22일(월) –
2021년 2월 25일(목) 15시

온라인면접 기간

: 2021년 2월 27일(토) –
2021년 2월 28일(일)

** 서류합격자에 한에 개별적으로 면접 일정을 알려드릴 예정입니다.

최종 발표

: 2021년 2월 28일(일) 밤 /
2021년 3월 1일(월) 예정

27th Recruiting
2021

카페 주소

: https://cafe.naver.com/powersat

인스타 아이디

: skku_psat2007

깃헙 주소

: https://github.com/P-Sat

카카오 채널 이름

: SKKU P-SAT (혹은 "피셋"으로 검색)

문의

: 학회장 권남택 010-7518-8810

27기 여러분을
기다립니다.