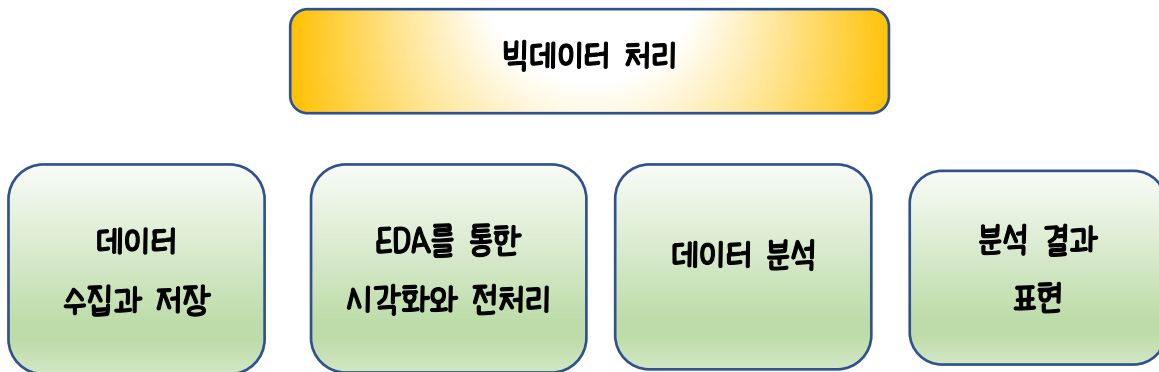


※ 빅데이터반의 전공 수업은 [R 기반의 빅데이터 처리], [Python 기반의 빅데이터 처리] 그리고 [빅데이터 활용 프로젝트] 입니다.



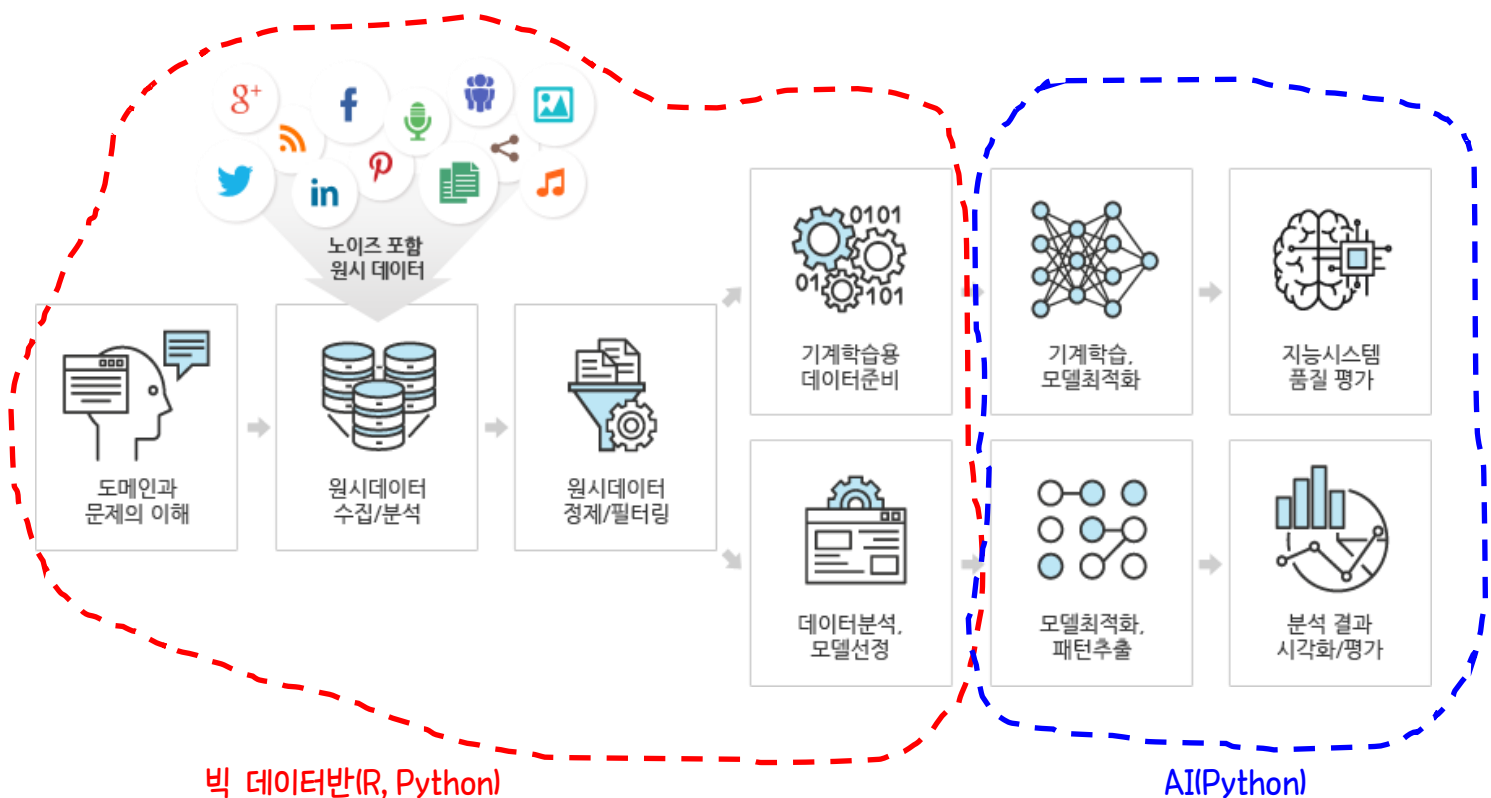
※학습 기간

9월 7일 ~ 10월 6일 - R 기반의 빅데이터 처리(15일)

9월 24일 ~ 10월 8일 - R 기반의 빅데이터 활용 프로젝트(4일)

10월 14일 ~ 11월 6일 - Python 기반의 빅데이터 처리, SQL, Spark(16일)

10월 26일 ~ 11월 17일 - Python 기반의 빅데이터 활용 프로젝트(10일)



빅 데이터반 전공과목 학습 계획(9월9일수정)

9/7 R 구문	9/8 R 구문	9/9 R 구문	9/10 R 구문 날짜 데이터 처리	9/11 정적 크롤링 문자열 데이터 처리
9/14 포트폴리오 특강 (원격진행)	9/15 포트폴리오 특강 (원격진행)	9/16 정적 크롤링 실습 정규표현식 활용 정적 크롤링 수행평가	9/17 Open API를 활용한 공공DB와 SNS 데이터 수집	9/18 팀 프로젝트 발표 동적 크롤링 동적 크롤링 실습
9/21 동적 크롤링 실습 동적 크롤링 수행평가 MariaDB 설치	9/22 MariaDB를 활용한 정형데이터 저장 데이터 전처리시 사용되는 주요 API	9/23 기본시각화 미니 프로젝트 조 나눔	9/24 R을 활용한 데이터 분석 실습 dplyr 패키지를 활 용한 데이터 전처리	9/25 dplyr 패키지를 활 용한 데이터 전처리 R 기반의 데이터 분석 미니 프로젝트 (2인 1조) - 2시간
9/28 EDA와 데이터 분석 실습 텍스트 분석 텍스트분석수행평가	9/29 고급 시각화 - ggplot 활용 - 지도 시각화	9/30 	10/1 	10/2 
R 기반의 데이터 분석 미니 프로젝트 (2인 1조) - 2시간	R 기반의 데이터 분석 미니 프로젝트 (2인 1조) - 4시간			
10/5 R을 활용한 데이터 분석 응용	10/6 R을 활용한 데이터 분석 응용	10/7 R 기반의 데이터 분석 미니 프로젝트 (2인 1조)	10/8 R 기반의 데이터 분석 미니 프로젝트 (2인 1조)	10/9 
R 기반의 데이터 분석 미니 프로젝트 (2인 1조) - 4시간	R 기반의 데이터 분석 미니 프로젝트 (2인 1조) - 4시간			

10/12 R 기반의 데이터 분석 미니 프로젝트 발표 파이썬&장고 복습 필답평가	10/13 취업특강	10/14 파이썬 기반의 정적 크롤링	10/15 파이썬 기반의 동적 크롤링 동적크롤링수행평가	10/16 Open API를 활용한 공공DB와 SNS 데이터 수집
10/19 교재1-파이썬 머신 러닝 판다스 데이터 분석	10/20 교재1-파이썬 머신 러닝 판다스 데이터 분석	10/21 교재1-파이썬 머신 러닝 판다스 데이터 분석	10/22 교재1-파이썬 머신 러닝 판다스 데이터 분석 데이터분석수행평가	10/23 교재1-파이썬 머신 러닝 판다스 데이터 분석 팀플젝조나눔발표
10/26 교재2-수치,범주형, 텍스트,날짜,이미지 타입 데이터 전처리 빅데이터 활용 프로젝트(4)	10/27 교재2-수치,범주형, 텍스트,날짜,이미지 타입 데이터 전처리	10/28 교재2-수치,범주형, 텍스트,날짜,이미지 타입 데이터 전처리 빅데이터 활용 프로젝트(4)	10/29 파이썬을 활용한 데이터 분석 응용 (센서, 이미지, 시계열 분석 관련 실습) 필답평가,수행평가제출	10/30 빅데이터 활용 프로젝트(8) 앞에서 과정 진행이 좀 늦어 지면 이날 수업을 일부 진행 할 수도 있으며 강사는 AWS 교육환경 세팅을 진행함
11/2 AWS에서 진행 Linux 명령어 MongoDB 활용 SQL을 활용한 데이터 전처리	10/30 AWS에서 진행 SQL을 활용한 데이터 전처리	11/4 AWS에서 진행 Spark 개요 빅데이터 활용 프로젝트(4)	11/5 AWS에서 진행 Spark을 활용한 데이터 처리	11/6 Spark을 활 용한 데이터 처리 빅데이터 활용 프로젝트(4)(4)
11/9 AWS에서 진행 Spark ML을 활용한 머신러닝 소개 빅데이터 활용 프로젝트(4)	11/10 빅데이터 활용 프로젝트	11/11 빅데이터 활용 프로젝트	11/12 빅데이터 활용 프로젝트	11/13 빅데이터 활용 프로젝트
11/16 빅데이터 활용 프로젝트	11/17 빅데이터 활용 프로젝트 프로젝트 발표	융합프로젝트를 위한 순환교육 머신러닝과 딥러닝 개요	융합프로젝트를 위한 순환교육 IoT 프로그래밍 개요	융합프로젝트를 위한 순환교육 클라우드 개요

※제시된 학습 스케줄은 코로나 상황에 따라서 약간의 변경이 있을 수 있으며 변경이 있더라도 우리 전공 수업이 11월 17일까지인 것은 변함없습니다.

※10월 10일(토)은 특강으로 알고리즘 수업이 예정되어 있습니다. 취업 지원때 온라인 코딩테스트를 실시하는 업체들이 많습니다. 원하는 분들은 틈틈이 코딩테스트 준비하시는데 한빛미디어의 “[이것이 취업을 위한 코딩 테스트다 with 파이썬](#)”을 추천합니다.

추천 유튜브 URL :

https://www.youtube.com/watch?v=qQ5iLNjpxSk&list=PLRxOvPvLEmdDHxCvAQS1_6XV4deOwfVrz

코딩 테스트 준비와 관련 정보를 공유하는 사이트의 URL :

<http://blog.samstdio.com/coding-test/>

<https://ebbnflow.tistory.com/190>

[미니 프로젝트 개요]

이번 프로젝트는 데이터 수집, 전처리(EDA), 분석, 시각화를 이번에 학습한 R 이라는 구현하고 결과를 도출해 보는 겁니다.

여러분들이 그동안 궁금하게 생각했더거 예를 들면,

- 올해엔 집값이 많이 올랐다고 하는데 정말 그럴까? 서울 집값이 얼마나 올랐을까?
- 지역별로 병원수의 격차가 크다 하는데 정말 그럴까? 지역별로 규모가 있는 병원들이 얼마나 분포하고 있을까?
- 1년중 꽃게 가격이 제일 비싼 시기는 언제일까? 꽃게 가격을 알아보자.
- 영유아들이 병원 진료를 받게되는 병질환은 어떤 것들이 있을까?
- 각 구별 CCTV 설치 현황을 알아보자.
- 내가 좋아하는 도서들에 대한 소비자들 반응을 텍스트 분석하여 워드클라우드로 시각화해 보자.

4일을 기준으로 해서 다음과 같이 대략적으로 시간을 정해서 진행하면 좋을 거 같아요.(이건 예시입니다.)

기획 : 4시간

수집 : 10시간

EDA : 4시간

분석 : 10시간

분석 결과 시각화와 문서 작성 : 4시간

빅데이터반 친구들의 역할

- 프로젝트 주제와 관련된 데이터 수집(공공DB Open API, SNS Open API, 웹 크롤링&스크래핑)
(공공데이터 포털, 서울시 빅데이터, 기상청 기상자료개발포털, 국가대기측정망, 국가공간정보포털, 환경공간정보서비스 등)
- 수집된 데이터 저장
- 수집된 데이터 EDA와 전처리
- 수집된 센서데이터 전처리
- 데이터 분석과 시각화
- AWS 기반의 Spark, MongoDB, MariaDB 활용