

강 의 계 획 서

학습과정명	구분	교재명	저자명	출판사	출판연도	비고
빅데이터개론	주교재	파이썬을 활용한 빅데이터 분석 개론	안기수	생능출판사	2024	

학습목표	본 교과목은 빅데이터가 모든 분야의 혁신 패러다임으로 등장하게 되었고 이러한 데이터 분석을 통해 엄청난 가치를 창출할 수 있게 되었다. 결론적으로 이러한 빅데이터의 개념, 분석 등을 학습하여 빅데이터 분석 도구를 활용할 수 있는 능력을 배양한다.
------	---

학점	3학점			정원	40명	
수업 기간 (※학사일정 참고)	1. 1학기: 3월 ~ 6월 (15주) 2. 여름 계절학기: 7월 ~ 8월 (8주) 3. 2학기: 9월 ~ 12월 (15주) 4. 겨울 계절학기: 1월 ~ 2월 (8주)			주당 시수 / 총 시수	3시간 / 45시간	
교·강사명	김종국 등 5명			수강료	420,000원	
성적평가 방법(평가요소)						
중간고사	기말고사	과제물	출결	기타	합계	비고
30%	30%	10%	20%	<u>10%</u>	100%	기타-수시시험

학습과정명		빅데이터개론	
■ 주차별 수업(강의 • 실험 • 실습 등) 내용			
주 별	차 시	주차별 수업(강의•실험•실습 등) 내용	과제 및 기타 참고사항
1	1	오리엔테이션: 전반적인 강의계획을 설명하고 학습과정에 대한 소개 및 강의 진행방식(과제물, 조별과제 안내 포함) 설명 강의주제: 빅데이터의 개념과 중요성	
	2	강의목표: - 빅데이터의 정의와 특성을 이해한다.	
	3	- 빅데이터 관련 직업과 자격증을 알아본다. - 빅데이터와 인공지능의 관계를 파악한다. - 개인정보 보호의 중요성을 학습한다. 강의세부내용: - 빅데이터의 5V(Volume, Velocity, Variety, Veracity, Value) 특성과 이를 활용한 사례를 소개한다. - 빅데이터의 처리와 분석에 관련된 기술들에 대하여 학습한다. - 빅데이터 관련 분야와 국내외 자격증에 대하여 학습한다. - 인공지능과의 관계 그리고 개인정보보호와 정보 활용에 대하여 학습한다. - 빅데이터 분석의 전반적인 흐름을 파악한다. 강의방법: 강의, 질의응답, 문제풀이	
2	1	강의주제: 파이썬 언어의 기초	

		강의목표: - 파이썬 언어의 설치 방법과 실행 방법을 익힌다.	
	2	- 기초 문법을 이해하고 간단한 프로그램을 작성할 수 있다. - 파이썬을 활용한 데이터 시각화와 요약 기법을 학습한다.	
	3	강의세부내용: - 파이썬의 특징과 설치에 대하여 학습한다. - print 함수와 연산자를 사용하여 간단한 연산 결과를 알아본다. - 변수 선언, 데이터 입력, 조건문, 반복문 등 기초 문법을 학습한다. - 파이썬에서 빅데이터 분석과 처리를 위해 필요한 라이브러리들과 설치방법에 대하여 학습한다. - 언어 사용법과 시각화 기술에 대하여 학습한다. 강의방법: 강의, 질의응답, 문제풀이	
3	1	강의주제: 파이썬을 활용한 데이터의 시각화와 요약 강의목표: - 다양한 경로로 데이터를 수집하는 방법을 파악한다. 그래프를 통하여 데이터의 특성을 나타내는 방법들에 대하여 학습한다.	
	2	- 공공 데이터 포털을 활용해 실제 데이터를 가져오는 실습을 한다. - 수집된 데이터를 분석용으로 정리하고 준비하는 방법을 익힌다.	
	3	강의세부내용: - 파이썬 설치 방법(Anaconda 사용)을 소개하고 실행 환경을 이해한다. - 기본적인 데이터 처리를 위해서 NumPy와 Pandas를, 통계 계산과 분석 그리고 그래프 작성을 위해 SciPy와 Matplotlib 라이브러리를 사용하여 학습한다. - 시각화 방법인 시간 시각화, 분포 시각화, 관계 시각화, 비교 시각화, 공간 시각화를 활용하여 학습한다. - 히스토그램, 막대그래프, 선 그래프 등을 활용하여 데이터를 시각적인 구조로 나타낸다. 강의방법: 강의, 질의응답, 문제풀이	
4	1	강의주제: 데이터 수집 방법 (Part 1) 강의목표: - 데이터 수집의 중요성과 다양한 수집 방법을 이해한다. - 공공데이터를 활용하여 데이터를 수집하는 방법을 익힌다. - 데이터 제공 사이트를 활용하여 필요한 데이터를 찾을 수 있다.	
	2	강의세부내용: - 데이터 수집의 목적과 방법에 대해 학습한다. - 공공데이터 포털에서 데이터를 검색하고 다운로드하는 방법을 학습한다.	
	3	- 다양한 데이터 제공 사이트를 소개하고, 원하는 데이터를	

		효과적으로 찾는 방법을 학습한다. 강의방법: 강의, 질의응답, 문제풀이	
5	1	강의주제: 데이터 수집 방법 (Part 2) 강의목표: - 스크레이핑, 크롤링, RSS, Open API 등의 데이터 수집 방법을 이해한다.	
	2	- 국내외 유용한 데이터 제공 사이트를 살펴본다.	
	3	- 파이썬 라이브러리에서 제공하는 데이터를 활용하는 방법을 알아본다. - 원하는 데이터를 효과적으로 찾는 방법을 학습한다. 강의세부내용: - 특정한 웹사이트에서 정보를 자동으로 추출하는 기술을 학습한다. - 국가통계포털, 통계청, 공공데이터포털 등을 활용하여 국내 데이터 수집 방법을 학습한다. - 파이썬 라이브러리에서 matplotlib을 기반으로 데이터를 활용하는 방법을 학습한다. 강의방법: 강의, 질의응답, 문제풀이	
6	1	강의주제: 데이터 구조와 전처리 강의목표: - 리스트와 배열의 개념과 사용 방법을 이해한다.	
	2	- 데이터의 구조를 변환하는 방법들을 살펴본다. 강의세부내용: - 데이터 구조인 리스트의 생성과 요소에 대한 접근을 살펴보고 리스트의 메소드 사용법을 학습한다.	
	3	- 1차원 배열의 연산과 2차원 배열의 합과 곱을 이해하고 특정 조건을 사용하여 배열을 생성하는 것을 학습한다. - 데이터를 효율적으로 저장, 처리, 관리하는 방법을 학습한다. - 수집한 자료를 파이썬으로 불러들일 때 적절한 데이터 구조를 사용하는 방법과 필요에 따라 데이터 구조를 변환하는 방법을 학습한다. 강의방법: 강의, 질의응답, 문제풀이	※수시시험 (5점) - 쪽지시험 실시
7	1	강의주제: 데이터 파일 다루기 강의목표: - 데이터 파일을 불러오고 생성하는 방법을 이해한다. - 결측값 처리 방법을 익힌다.	
	2	강의세부내용: - 외부 파일을 파이썬으로 불러오는 방법을 (sample1.csv, sample2.txt, sample3.txt, sample4.xlsx 등) 4가지 파일 형식으로 학습한다.	
	3	- 프로그램 안에서 데이터를 엑셀 파일로 저장하는 방법을 함수 to_excel()을 이용하여 학습한다. - 결측값을 처리하거나 다루는 메소드들을 Pandas, DataFrame을 이용하여 학습한다.	※1차 과제 (5점) : 1~6장까지의 보고서 과제 (9주차 제출, 기간 내 미제출 시 감점)

		- 중간고사 전 정리 강의방법: 강의, 질의응답, 문제풀이	
8	1 2 3	중간고사	
9	1 2 3	강의주제: 탐색적 데이터 분석과 시각화 강의목표: - 탐색적 데이터 분석(EDA)의 정의와 필요성을 이해한다. - 탐색적 데이터 분석(EDA)을 통해 데이터를 요약하고 패턴을 발견하는 방법을 학습한다. 강의세부내용: - 간단한 데이터셋을 이용해 EDA 과정을 학습한다. 강의세부내용: - 탐색적 데이터 분석(EDA)의 개념, 목적(데이터 요약, 이상치 탐지, 패턴 발견 등)을 살펴본다. - 타이타닉 데이터셋을 불러오고, 기본적인 통계 요약(describe 등)을 학습한다. - 변수 간의 관계를 파악하는 다양한 기초 시각화 기법을 학습한다. 강의방법: 강의, 질의응답, 문제풀이	
10	1 2 3	강의주제: 텍스트 마이닝 탐색과 시각화 강의목표: - 텍스트 데이터를 수집하고 기본적인 탐색을 할 수 있다. 강의세부내용: - 텍스트 마이닝 기법을 통해 텍스트를 분석하고 요약할 수 있다. - 단어구름과 히트맵, 네트워크 다이어그램을 활용한 시각화 방법을 학습한다. 강의세부내용: - 텍스트 데이터의 특성과 구조(비정형 데이터)를 이해하고, 전처리 과정(토큰화, 정제, 불용어 제거 등)을 학습한다. - 공통적으로 등장하는 단어들의 상관 관계나 문서 간 유사도를 시각화하는 히트맵을 학습한다. - 텍스트 마이닝 결과를 기반으로 어떤 인사이트를 도출할 수 있는지 다양한 사례를 통해 파악한다. - 강의방법: 강의, 질의응답, 문제풀이	
11	1 2 3	강의주제: 통계 기반 데이터 분석 모형의 이해 강의목표: - 통계 기반 데이터 분석 모형의 개념과 구성요소를 이해한다. 강의세부내용: - 기술통계와 추론통계를 구분하고 각 기법의 사용 목적을 설명할 수 있다. - 추정과 가설검정의 기초 원리를 학습하고, 다변량 분석 기초를 습득한다. 강의세부내용: - 평균, 분산, 분포 등 데이터를 요약·정리하는 기법을 학습한다. - 표본에서 모집단 특성을 추론하는 추정과 가설검정 개념 및 적용 사례를 살펴본다.	

		- 정규성 검정, t-검정, 카이제곱 검정 등 주요 추론통계 기법의 구조와 의미를 알아본다. - 다변량 분석 개요(변수가 2개 이상일 때의 분석 방법) 및 적용 분야를 파악한다. 강의방법: 강의, 질의응답, 문제풀이	
12	1	강의주제: 데이터 마이닝 및 머신러닝 기반 분석 모형 강의목표: - 데이터 마이닝과 머신러닝의 차이점을 이해하고, 분석 흐름을 파악한다.	
	2	- 주요 분석 기법(분류, 군집화, 회귀 등)의 특징을 파악하고 적용 사례를 학습한다. - 머신러닝 기반 분석 절차를 간단하게 학습한다.	
	3	강의세부내용: - 대량의 데이터로부터 유의미한 패턴을 발견하는 과정을 파악한다. - 대표 기법인 분류(의사결정나무, 나이브 베이즈), 군집화(K-means), 연관분석(Apriori) 등의 개념을 학습한다. - 지도학습 vs 비지도학습 구조 비교 및 머신러닝 기반 분석 모형과 해당되는 알고리즘을 통해 학습한다. 강의방법: 강의, 질의응답, 문제풀이	
13	1	강의주제: 상관분석과 회귀분석 강의목표: - 상관분석과 회귀분석의 개념과 차이를 이해한다. - 실제 데이터를 이용해 상관계수와 회귀모형을 살펴본다.	
	2	- 분석 결과를 해석하고 의미를 나타낸다. 강의세부내용: - 변수 간 관계를 분석하는 상관분석(피어슨 상관계수 등)을 학습한다.	※수시시험 (5점) - 꼭지시험 실시
	3	- 단순 선형 회귀분석을 통해 두 변수 간의 관계를 학습한다. - 회귀분석 결과(계수, 결정계수 등)를 해석하는 방법을 학습한다. 강의방법: 강의, 질의응답, 문제풀이	
14	1	강의주제: 다변량 분석의 이해와 활용 강의목표: - 다변량 분석의 필요성과 기본 개념을 이해한다. - 주성분 분석(PCA) 등 주요 다변량 분석 기법을 익힌다.	
	2	- 다변량 분석 결과를 해석하고 시각화할 수 있다. 강의세부내용: - 다변량 데이터의 개념과 다변량 분석이 필요한 이유를 알아본다.	※2차 과제 (5점) : 6~10장까지의 보고서 과제
	3	- 주성분 분석(PCA)의 이론을 배우고, sklearn을 이용해 PCA를 학습한다. - 차원 축소된 데이터를 시각화하고, 의미를 도출하는 방법을 학습한다. - 기말고사 전 정리	기한 내 미 제출시 감점

		강의방법: 강의, 질의응답, 문제풀이	
15	1	기말고사	
	2		
	3		

※ 강의계획서 주차별 내용은 교·강사에 따라 변동될 수 있습니다.