

데이터사이언스 교수법 개발 (R 과 파이썬 중심)

한국데이터정보과학회지(2023년 1월 20일 게재)

## 1. 서론

4차 산업 혁명 중 중요한 분야인 데이터사이언스는 교육에도 영향을 많은 영향을 미치고 있어 학부 학생들의 대상으로 데이터를 이해할 수 있는 ‘데이터사이언스의 이해’ 교수법을 파이썬과 R을 중심으로 연구하였다.

## 2. 데이터사이언스의 이해

학생들이 데이터를 이해하고 다루는데 필수적인 분야를 6개로 나누어 연구하였다.

### 2.1 Data Wrangling

데이터를 예측분석하고 머신러닝을 위한 전처리하는 과정이다. 다음과 같은 작업으로 이루어진다.

- (1) 결측값 처리 (2) 특정 조건을 만족하는 관측값 선택
- (3) 변수 선택 (4) 새로운 변수 만들기과 표준화
- (5) 그룹별 요약 통계량 구하기
- (6) 복수 데이터 셋의 병합

### 2.2 Data Visualization

자료의 분석단계나 의사결정 진의 레포트를 통한 커뮤니케이션 단계에서 필수적인 단계이다. 최근 파이썬에서 각광받고 있는 시각화 구현방법으로 Dash의 사용이 급증하고 있다.

**Dash란?** 사용자가 필요한 데이터/정보를 한 눈에 확인하여 의사결정을 돕는 도구이며 특히 파이썬에서 Dash를 구현하기 위해 Plotly(파이썬 라이브러리 중 하나로 온라인 데이터 분석 및 시각화 툴을 제공함.)를 사용한다. 이를 통해 데이터 분석, 레포트 작성, 모델링 등을 수행하는 애플리케이션을 파이썬 사용자들이 구축할 수 있다.

### 2.3 Predictive analytics

예측분석은 현재 및 과거 사실을 분석하여 미래 또는 알려지지 않은 사건을 예측하는 데이터 마이닝, 머신러닝의 다양한 통계 기술을 포함한다.

### 2.4 Supervised learning(지도 학습)

지도 학습은 훈련데이터에서 모델을 학습하여, 본적 없는 데이터 또는 미래 데이터에 대한 예측을 만드는 것이며, 분류(종류를 예측)와 회귀(연속된 값을 예측)로 나뉜다.

### 2.5 Unsupervised learning(비지도 학습)

비지도 학습은 정답 값(label)이 없는 데이터를 다루며 알려진 출력 값이나 보상 함수의 도움을 받지 않고 의미있는 정보를 추출하기 위해 데이터 구조를 탐색한다.

### 2.6 Text mining

텍스트 마이닝은 통상의 데이터 마이닝과는 다르게 텍스트 데이터를 분석하여 ‘word cloud’와 같은 방법으로 시각화 한 것이다.

## 3. 데이터사이언스 이해 교수법

### 3.1 수업목표

‘데이터사이언스’의 필요성을 인식하게 하고 파이썬 언어를 습득 한다.

### 3.3 주차별 강의 방식

#### (1) 1 - 4주차

파이썬 언어를 습득을 목적으로 수업하며 파이썬의 자료구조, 프로그래밍, 라이브러리(numpy, pandas), 메서드(concat, merge) 피벗테이블 또는 groupby 와 같은 데이터 분석을 위한 코드를 학습한 후 공공 데이터 저장소, 공공 데이터 포털 사이트에서 자료를 얻어 ① Dash ② 3차원 그래프 ③ 동적그래픽스 중 하나를 선택하여 데이터를 시각화 한다.

#### (2) 5 - 9주차

파이썬 라이브러리(folium, plotly express 등)을 사용하여 통계맵을 체험하고 7주차부터 지도학습을 중심으로 scikit-learn을 사용한 의사결정나무, k-NN, 랜덤포레스트를 실습하고 딥러닝의 기본개념과 딥러닝의 대표적인 프레임워크인 텐서플로, 케라스, 파이토치로 데이터 분석을 실습한다.

#### (3) 10 - 15주차

심화된 데이터 분석을 하는 것을 목표로 scikit-learn을 중점적으로 다루며 scikit-learn의 계층적 집락분석, 차원축소 방법 등 개념을 배우고 실습을 진행한다. 텍스트 분류를 하기 위한 텍스트 마이닝의 개념을 배운다.

특히 마지막 14, 15주차에는 사회 연결망 분석을 위해 networkX(파이썬 라이브러리)를 이용한 ‘왕자의 게임’ 등장인물 네트워크를 분석하고, 추이행렬, PageRank, 네트워크 분할하기, Four color theorem의 개념을 배우고 실습하는 것으로 마무리 한다.