|  |  |
| --- | --- |
| 교육 제목 | **파이썬 머신러닝 판다스 데이터 분석** |
| 교육 일시 | 2021년 11월 08일 |
| 교육 장소 | 영우글로벌 강의실 |
| **교육 내용** | |
| 오전 | 1. 판다스를 배우는 이유   -빅데이터의 시대. 데이터 과학이라는 새로운 영역의 출현  -데이터 과학은 데이터를 연구하는 분야이고, 데이터 자체가 가장 중요한 자원  -판다스는 데이터를 수집하고 정리하는데 최적화된 도구   1. 판다스 자료 구조   -목적 : 분석을 위해 다양한 소스로부터 수집하는 데이터는 형태나 속성이 매우 다양한데 특히, 서로 다른 형식을 갖는 여러 종류의 데이터를 컴퓨터가 이해할 수 있더록 형식을 갖는 구조로 통합할 필요가 있다. 판다스의 일차적인 목적은 형식적으로 서로다른 여러가지 유형의 데이터를 공통의 포맷으로 정리하는 것이다.   1. 시리즈는 데이터가 순차적으로 나열된 1차원 배열의 형태를 갖는다.   딕셔너리와 시르즈의 구조가 비슷하기 때문에 딕셔너리를 시리즈로 변환하는 방법을 많이 사용한다. 판다스 내장 함수인 Series()를 이용한다.  공통의 포맷으로 정리하는 것이다.   1. 데이터 프레임은 행과 열로 이루어진 2차원 배열이다. |
| 오후 | 1. 인덱스 활용 : 열 기준 정렬로는 DataFrame 객체.sort\_values()로 특정 열의 데이터를 기준으로 새롭게 정렬된 데이터프레임 객체를 반환한다. 2. 산술 연산으로 시리즈vs숫자, 시리즈vs시리즈, 데이터프레임vs숫자, 데이터프레임vs데이터프레임이 있다. 3. 외부 파일 읽어오기 : 데이터 분석을 위해 다양한 곳에서 자료를 수집하다 보면, 여러가지 파일 형식을 마주치게 되는데 파일 확장자로는 .csv, .json, .xlsx등이 있다. |