|  |  |
| --- | --- |
| 교육 제목 | **판다스 문법, 머신러닝 데이터 분석 기초** |
| 교육 일시 | 2021-10-05 |
| 교육 장소 | 집 |
| **교육 내용** | |
| 오전 | 시계열 데이터 : 데이터프레임의 행 인덱스로 사용하면, 시간으로 기록된 데이터를 분석하기에 유용함    외부파일 df.info를 활용하여 시간까지 출력가능  문자열을 Timestamp로 변환      데이터 내용을 확인 하고 인덱스 제거    New Data로 새로운 행 추가    새로 추가된 행 그래프로 표현  Pipe 함수 알아보기    해당 출력 값으로 전달된 데이터 값 확인    출력화면 정렬 방법    열 분리 방법 예제 실습 코드      시리즈에서 리스트 각 원소 추출하는 방법    나이가 10세 미만 80세 이상 승객 데이터 정보 출력 예제    많은 값을 사용하기위한 불린 인덱싱 연습 예제  원하는 데이터만 추출 하는 경우 : bollean mask 이용, isin() 메소드 이용  **데이터프레임 합치기**    기존 원본 데이터    위에서 아래로 붙이기    기존 인덱스 무시하고 합치기    열로 붙히기    조인 조건 inner와 outer  시리즈 합치기    좌우 열 방향으로 연결 index값없을 때    **Index 값 존재 할 경우**    **좌우 열 방향으로 연결**    **위 아래로 연결 index 존재할 경우** |
| 오후 | Merge() 함수 : sql 의 join과 유사, 어떤 기준에 의해 두 데이터 프레임을  병합하는 개념    외부 파일에서 불러온 자료 원본    Merge()\_inner -> 교집합으로 출력  Merge()\_outer -> 합집합으로 출력    Merge()\_left -> 왼쪽프레임 기준으로 키 값을 분리  Merge()\_right -> 왼쪽프레임 기준으로 키 값을 분리    불린 인덱싱으로 원하는 조건의 값 출력하기    교집합 결합 -> join() 인덱스로 결합  **그룹연산**  : 복잡한 데이터 어떠한 기준에 따라 여러 그룹으로 나눠 연산  단계   1. 분할(split) : 데이터를 특정 조건에 의해 분할 2. 적용(apply) : 데이터를 집계, 변환, 필터링하는데 필요한 메소드 적용 3. 결합(combine) : 2단계의 처리 결과를 하나로 결합     Class 를 기준으로 분할    분할된 값의 데이터를 데이터프레임으로 반환, 이후 시리즈로 변환    이때 각 그룹 객체에서 개별 그룹으로 선택하는 방법    선택한 그룹에서 원하는 행 선택하는 방법    선택 이후 피벗테이블을 이용한 데이터 인자 전달  **머신 러닝**  데이터 분석이란 : 예측 하기 위해 예측 수식을 만드는 것  1자료를 가져옴  2 누락데이터 처리 및 사전 데이터 전처리  : 데이터 가공(조인, 분리, 결합 등), 이후 문자를  더미 값/ 원핫인 코딩/ NaN값 처리 이 후 x, y값 분리  3 실험데이터와 테스트데이터로 분리  4 모델 학습  5 모델 결과로 타겟 값 계산 결과 나오게 도출  6 예측한 테스트 데이터를 모델 값에 적용해서 결과와,  실제 테스트의 타겟 값과 비교  7 비교한 결과/점수 확인  8 이후 예측된 값과 실제 값을 비교해서 그래프화  9 모델검증 계속 진행 |