|  |  |
| --- | --- |
| 교육 제목 | AI 프로젝트 기반 S/W 전문가 양성과정 |
| 교육 일시 | 21.11.10 |
| 교육 장소 | 대면(강의실) |
| **교육 내용** | |
| 오전 | 누락 데이터 처리  #누락 데이터 확인 (NaN)  value\_counts() > 데이터프레임[열 이름].value\_counts(dropna=false)  isnull() > bool 형식으로 누락된 데이터 존대하면 true  notnull() > bool 형식으로 유효한 데이터 존재하면 true  #누락 데이터 제거  dropna() -> thresh 옵션 설정하여 NaN개수 특정수가 넘어가면 제거  #누락 데이터 치환  mean() -> 평균 값으로 누락 데이터 치환  idmax() -> 가장 많이 나타나는 값으로 치환  fillna(method=옵션값) -> ffill입력하면 NaN직전 행의 데이터, bfill은 이후  중복 데이터 처리  #중복 데이터 확인  duplicated() -> 행 인덱스 단위로 중복 데이터 검사  #중복 데이터 제거  drop\_ duplicated() -> 데이터 프레임에 중복 데이터 제거, 특정 열 가능  데이터 표준화  #단위 환산  round() -> 소수점 단위 표기. 잘리는 부분에서 반올림  #자료형 변환  astype() -> 매개 변수에 적힌 형태로 변환 |
| 오후 | #범주형 데이터 처리  # 구간 분할  cut() -> 연속된 데이터를 여러 구간으로 나누고 범주형 데이터로 변환  # 더미 변수  get\_dummies() -> 범주형 변수의 모든 값을 각각 새로운 더미 변수로 변환  #정규화  # 중구난방 데이터 사이즈를 0~1사이로 정렬 하는 기법  ( (각데이터) – min ) / ( max - min )  시계열 데이터  #다른 자료형을 시계열 데이터로 변환  to\_datetime() -> 문자열 데이터를 datetime64 형태로 변환  to\_period() -> timestamp객체를 period로 변환  시계열 데이터 만들기  #Timestamp 배열  date\_range() ->범위시작, 범위끝, period계수, 시작간격, 시간대 입력  #period 배열  period\_range() -> 날짜범위시작, 날짜범위끝, period계수, 기간의 길이  시계열 데이터 활용  날짜 데이터 분리  dt.<날짜형식> -> 연-월-일 정보에서 원하는 날짜형식을 추출 할 수 있다  to\_period() -> 매개변수에 원하는 정보 약어 입력하면 추출 가능  날짜 인덱스 활용  set\_index() 메소드를 사용하여 timestamp형식의 데이터를 인덱스로 지정 |