

Airbee Data Insight

* What to do? DT 추진단

1. Airbee의 광산란법의 한계점 -> 머신러닝 알고리즘으로 극복 (관련 논문 다수 존재)
2. 기타 insight

[[OUTDOOR]]

- About data table

- 각 컬럼에 대한 구체적인 설명은 이전에 받으셨던 '공기질 데이터 DB 정보 _everyair.xlsx', 'AIRBEE 전송 데이터 테이블_From AirBee.xlsx' 파일을 보면 아실 수 있습니다.
- Outdoor 데이터 중에 날짜가 2018년이 아닌 년도로 표기된 것들이 조금 포함되어 있습니다. DB상 오류라고 합니다.
- ILLUMINATION 컬럼의 값들은 기본적으로 두자리 숫자로 표현되어야 합니다. (00, 01, 10, 11) 데이터에 0과 1로만 표기된 것들은 본래 00과 01이었을 것이라고 추측합니다.
- 실외데이터의 경우 현재 일별로 3일전 데이터를 삭제하는 로직으로 되어 있다고 합니다.

- About insight

- 광산란법으로 측정한 PM2 수치가 습도에 영향을 받는다는 논문이 있었습니다. 실제 Airbee 데이터에서도 PM2와 습도가 상관성을 보이는 것 같다고 예상하고 있습니다.
- 일반적으로 출, 퇴근 시간대에 PM2가 높은 수치로 나타나는 경향이 있습니다. 각 센서의 각 날짜별 그래프도, 각 센서별 통합 데이터 그래프도, 전체 통합 데이터 그래프도 이러한 비슷한 경향성을 보였습니다.
- Airbee 데이터는 환경부 데이터와 비교했을 때 평균과 분산이 매우 큼니다. 그래서 미세먼지 수치가 낮을 때는 비교적 환경부 센서와 비슷하게 예측할 수 있지만, 미세먼지 수치가 높을 때에는 오차가 훨씬 커질 수 있습니다. 하지만 환경부 데이터와의 시계열 경향성을 비교해 보았을 때, 그 경향성은 비슷했습니다.
- 근조도와 관련하여 '근접 어두움'(10)으로 ILLUMINATION 컬럼이 표시된 항목들은 미세먼지 수치가 낮게 나왔을 가능성이 크고, 이러한 경우는 가방 안이나 주머니 안에 들어있는 특수한 상황일 가능성이 있습니다.

[[INDOOR]]

- About data table

- 실내 데이터는 평균치를 가지고 평점을 매겨 추천 장소로 띄워주는 방식을 채택하고 있기 때문에 실외와 조금 다르게 데이터가 저장되어 있습니다.
- TOTAL_CNT 컬럼은 해당 지역에 데이터가 몇 번 올라왔는지를 나타내고, AVG는 그 지역의 평균 실내 미세먼지 농도를 나타냅니다.

- About insight

- 하지만 가끔 실내 데이터임에도 불구하고 주소지가 실내가 아닌 경우 (ex. 한강공원)를 발견할 수 있는데, 이는 자동차 안에서 그 지역을 지나가면서 찍힌 것이 아닐까 추측합니다. TOTAL_CNT가 1 정도인 지역을 위와 같은 상황의 가비지 데이터로 생각하고 지우는 방안도 좋을 것 같습니다.

[[ETC]]

- Outdoor 데이터와 Indoor 데이터는 전송되는 위치 데이터 값을 기준으로 분류된다고 합니다. 실외에 있을 때는 GPS가 작동하여 블루투스 페어링만 되어있어도 위치 정보를 가져올 수 있지만, 실내에 있을 때는 GPS가 작동하지 않아 블루투스 페어링만 되어있을 경우 위치 정보를 가져올 수 없습니다. 실내의 위치 정보를 가져오기 위해서는 어플과 Airbee가 와이파이 페어링까지 되어있어야 합니다. 또는 블루투스 페어링만 되어있을 경우 직접 어플에 주소를 작성하여 DB로 보낼 수 있습니다.
- 속도와 가속도의 경우, 매칭이 정확히 이루어지지 않은 케이스가 있었습니다. (가속도는 1로 정지상태이지만 속도가 20이상인 경우)