

Python 활용 데이터 분석

Lecture 18. Project

Dataset

❖ 1. 연기 감지기를 통한 화재 예측

- 연기 감지기는 연기를 감지하는 장치로, 일반적으로 화재의 지표로 사용
- 화재 발생시 연기 감지기를 통해 즉각적인 대응 및 화재 피해 최소화 가능
- 연기 감지기의 종류
 - 광전식 연기 감지기, 이온화 연기 감지기
- IOT 장치와 AI 모델을 활용하여 연기 감지 장치를 개발하기 위해 수집
- 60,000개의 측정값으로 구성
- 모든 센서의 샘플링 속도는 1Hz

Dataset

❖ 1. 연기 감지기를 통한 화재 예측

- index
- UTC: Timestamp UTC seconds
- Temperature[C]: Air Temperature
- Humidity[%]: Air Humidity
- TVOC[ppb]: Total Volatile Organic Compounds; measured in parts per billion
- eCO2[ppm]: CO2 equivalent concentration; calculated from different values like TVCO
- Raw H2: raw molecular hydrogen; not compensated (Bias, temperature, etc.)
- Raw Ethanol: raw ethanol gas
- Pressure[hPa]: Air Pressure

Dataset

❖ 1. 연기 감지기를 통한 화재 예측

- PM1.0: particulate matter size $< 1.0 \mu\text{m}$ (PM1.0). $1.0 \mu\text{m} < 2.5 \mu\text{m}$ (PM2.5)
- PM2.5: particulate matter size $< 1.0 \mu\text{m}$ (PM1.0). $1.0 \mu\text{m} < 2.5 \mu\text{m}$ (PM2.5)
- NC0.5: Number concentration of particulate matter. This differs from PM because NC gives the actual number of particles in
- NC2.5: Number concentration of particulate matter. This differs from PM because NC gives the actual number of particles in
- CNT: Sample counter
- Fire Alarm: ground truth is "1" if a fire is there

Dataset

❖ 2. 엘리베이터 고장 예측

- 엘리베이터는 건물의 층 사이를 이동하는 화물과 사람을 운반하는 교통수단
- 이동 중 엘리베이터가 고장 발생시 많은 불편을 초래하고 심지어는 재난적인 사고
- 엘리베이터의 고장을 예방하고 사고를 방지하기 위해 정기적인 유지보수가 필요
- 2020년 1월 한 달 동안 IoT 센서가 매분 기록한 데이터를 분석
- 총 44,640개의 데이터와 11개의 속성

Dataset

❖ 2. 엘리베이터 고장 예측

- Time
- Temperature
- Humidity
- RPM
- Vibrations
- Pressure
- Sensor1,2,3,4,5,6: 도어 베어링 센서 측정값
- Status : 0, 1, 2로 구성되며 0은 이상무, 1은 도어 개방 딜레이, 2는 운행 중단

Dataset

❖ 3. 공기질에 따른 실내 상황 예측

- 시간에 따른 실내 가스 농도의 변화를 모니터링하고 데이터를 저장하여 방 안에서 수행된 활동의 유형을 평가
- 데이터 수집 도구
 - MQ 센서 (MQ2, MQ9, MQ135, MQ137, MQ138): 높은 감도, 낮은 대기 시간, 저비용을 자랑하며, 각 센서는 다양한 가스에 반응
 - 아날로그 CO2 가스 센서 (MG-811): 이 센서는 이산화탄소에 대한 높은 감도를 가지고 있으며, 공기의 온도와 습도의 영향을 거의 받지 않음
- 1845개의 데이터 및 7개의 속성

Dataset

❖ 3. 공기질에 따른 실내 상황 예측

- 일반 상황 1 : 깨끗한 공기, 사람이 잠을 자거나 공부하거나 휴식을 취하는 상황 - 샘플: 595개;
- 식사 준비 2: 고기나 파스타 요리, 채소 튀김. 방 안에 한두 명이 있으며, 강제 공기 순환이 이루어짐 - 샘플: 515개;
- 연기 발생 3: 창문과 문이 닫힌 방에서 종이와 나무를 짚은 시간 동안 태우는 상황 - 샘플: 195개;
- 청소 4: 암모니아 및/또는 알코올이 포함된 스프레이 및 액체 세제를 사용하는 상황. 강제 공기 순환이 활성화되거나 비활성화될 수 있음 - 샘플: 540개;
- 'MQ2', 'MQ9', 'MQ135', 'MQ137', 'MQ138', 'MG-811', 'label'

Dataset

❖ 4. 가설 변압기 수명 예측

- 건설 현장에서 사용하는 가설 변압기는 다양한 이유로 고장 발생
- 가장 흔한 원인으로는 번개, 과부하, 마모 및 부식, 전력 서지, 그리고 습기 등이 원인
- 변압기가 고장나면 대형 화재 및 인명피해 발생 가능
- 변압기 고장 예방을 위해 다양한 데이터를 기반으로 변압기의 수명을 예측
- 471개의 데이터 및 16개의 속성

Dataset

❖ 4. 가설 변압기 수명 예측

- Hydrogen
- Oxygen
- Nitrogen
- Methane
- CO
- CO₂
- Ethylene
- Ethane
- Acetylene
- DBDS
- Power factor
- Interfacial V
- Dielectric rigidity
- Water content
- Health index
- Life expectation

Dataset

❖ 5. 질소산화물 배출량 예측

- 화력발전소, 소각로, 건설현장의 덤프트럭 등에서 인위적으로 발생하는 질소산화물은 대기오염의 원인이 되며 사회적인 문제 발생
- 질소산화물은 미세먼지와 결합하여 현장 작업자의 호흡계 질환, 심혈관계 질환을 유발
- 이에 해당 데이터 셋은 가스 터빈에서 수집된 12개의 센서값을 바탕으로 의도치 않게 생성된 질소산화물 발생량을 예측
- 36734개의 데이터 및 12개의 속성

Dataset

❖ 5. 질소산화물 배출량 예측

- Variable (Abbr.) Unit Min Max Mean
- Ambient temperature (AT) C 6.23 37.10 17.71
- Ambient pressure (AP) mbar 985.85 1036.56 1013.07
- Ambient humidity (AH) (%) 24.08 100.20 77.87
- Air filter difference pressure (AFDP) mbar 2.09 7.61 3.93
- Gas turbine exhaust pressure (GTEP) mbar 17.70 40.72 25.56
- Turbine inlet temperature (TIT) C 1000.85 1100.89 1081.43
- Turbine after temperature (TAT) C 511.04 550.61 546.16
- Compressor discharge pressure (CDP) mbar 9.85 15.16 12.06
- Turbine energy yield (TEY) MWH 100.02 179.50 133.51
- Carbon monoxide (CO) mg/m3 0.00 44.10 2.37
- Nitrogen oxides (NOx) mg/m3 25.90 119.91 65.29

Dataset

❖ 5. 질소산화물 배출량 예측

- 변수(약어) 단위 최솟값 최댓값 평균값
- 주변 온도 (AT) °C: -6.23 ~ 37.10 (평균: 17.71)
- 주변 압력 (AP) mbar: 985.85 ~ 1036.56 (평균: 1013.07)
- 주변 습도 (AH) %: 24.08 ~ 100.20 (평균: 77.87)
- 공기 필터 압력 차이 (AFDP) mbar: 2.09 ~ 7.61 (평균: 3.93)
- 가스터빈 배기 압력 (GTEP) mbar: 17.70 ~ 40.72 (평균: 25.56)
- 터빈 입구 온도 (TIT) °C: 1000.85 ~ 1100.89 (평균: 1081.43)
- 터빈 배출 온도 (TAT) °C: 511.04 ~ 550.61 (평균: 546.16)
- 압축기 배출 압력 (CDP) mbar: 9.85 ~ 15.16 (평균: 12.06)
- 터빈 에너지 출력 (TEY) MWH: 100.02 ~ 179.50 (평균: 133.51)
- 일산화탄소 (CO) mg/m3: 0.00 ~ 44.10 (평균: 2.37)
- 질소 산화물 (NOx) mg/m3: 25.90 ~ 119.91 (평균: 65.29)

Dataset

❖ 6. 일사병 환자수 예측

- 기온, 풍속, 강수량, 습도, 기압, 불쾌지수 등의 데이터를 기반으로 일사병 환자수를 예측
- 이러한 날씨 데이터를 기반으로 현장 작업자의 일사병 발생 가능성을 예측하여 사전에 예방할 수 있음
- Train셋 1208개의 데이터 및 41개의 속성
- Test셋 141개의 데이터 및 39개의 속성

Dataset

❖ 6. 일사병 환자수 예측

- 年月日 (년월일): 날짜
- 搬送人員 (計) (반송 인원 (계)): 이송 인원 (총계)
- 最高気温(℃) (최고 기온(℃)): 최고 기온(℃)
- 平均気温(℃) (평균 기온(℃)): 평균 기온(℃)
- 最低気温(℃) (최저 기온(℃)): 최저 기온(℃)
- 日照時間(時間) (일조 시간(시간)): 일조 시간(시간)
- 平均風速(m/s) (평균 풍속(m/s)): 평균 풍속(m/s)
- 平均雲量(10分比) (평균 운량(10분비)): 평균 운량(10분비)
- 平均湿度(%) (평균 습도(%)): 평균 습도(%)
- 降水量の合計(mm) (강수량의 합계(mm)): 강수량의 합계(mm)
- 最小相對湿度(%) (최소 상대 습도(%)): 최소 상대 습도(%)
- 合計全天日射量(MJ/m²) (합계 전천 일사량(MJ/m²)): 합계 전천 일사량(MJ/m²)

Dataset

❖ 6. 일사병 환자수 예측

- 平均蒸気圧(hPa) (평균 증기압(hPa)): 평균 증기압(hPa)
- 平均現地気圧(hPa) (평균 현지 기압(hPa)): 평균 현지 기압(hPa)
- 平均海面気圧(hPa) (평균 해면 기압(hPa)): 평균 해면 기압(hPa)
- 最大風速(m/s) (최대 풍속(m/s)): 최대 풍속(m/s)
- 最大瞬間風速(m/s) (최대 순간 풍속(m/s)): 최대 순간 풍속(m/s)
- c最高最低気温差 (최고-최저 기온차): 최고-최저 기온차
- c体感温度(°C) (체감 온도(°C)): 체감 온도(°C)
- c不快指数 (불쾌지수): 불쾌지수
- 月 (월): 월
- 曜日 (요일): 요일
- 土日祝日 (토일 및 공휴일): 토일 및 공휴일
- 昼_晴割合 (낮_맑음 비율): 낮_맑음 비율

Dataset

❖ 6. 일사병 환자수 예측

- 昼_曇割合 (낮_흐림 비율): 낮_흐림 비율
- 昼_雨割合 (낮_비 비율): 낮_비 비율
- 昼_雷あり (낮_번개 있음): 낮_번개 있음
- 夜_晴割合 (밤_맑음 비율): 밤_맑음 비율
- 夜_曇割合 (밤_흐림 비율): 밤_흐림 비율
- 夜_雨割合 (밤_비 비율): 밤_비 비율
- 夜_雷あり (밤_번개 있음): 밤_번개 있음
- m前日最高気温との差 (전일 최고 기온 차이): 전일 최고 기온 차이
- m前日平均気温との差 (전일 평균 기온 차이): 전일 평균 기온 차이
- m前日最低気温との差 (전일 최저 기온 차이): 전일 최저 기온 차이
- m最高気温移動平均(5日間) (최고 기온 이동 평균(5일간)): 최고 기온 이동 평균(5일간)
- m平均気温移動平均(5日間) (평균 기온 이동 평균(5일간)): 평균 기온 이동 평균(5일간)

Dataset

❖ 6. 일사병 환자수 예측

- m体感温度移動平均(5日間) (체감 온도 이동 평균(5일간)): 체감 온도 이동 평균(5일간)
- m不快指数移動平均(5日間) (불쾌지수 이동 평균(5일간)): 불쾌지수 이동 평균(5일간)
- m前日の搬送人数 (전일의 이송 인원수): 전일의 이송 인원수
- m搬送人数移動平均(5日間) (이송 인원수 이동 평균(5일간)): 이송 인원수 이동 평균(5일간)
- 年 (년): 년

Dataset

❖ 7. LPG 누출 탐지

- 다양한 화학물질을 다루는 산업현장에서 LPG 가스의 누출 여부를 판단하여 사전에 화재 예방
- IoT 센서를 통해 탐지되는 다양한 화학물질 속에서 LPG 가스 누출 여부를 판단
 - 현장에서 평소에 존재하는 LPG의 양인지 혹은 가스관 파열로 인한 가스 과다 누출인지 여부를 판단
- 1001개의 데이터 및 9개의 속성

Dataset

❖ 7. LPG 누출 탐지

- Alcohol
- CH₄
- CO
- H₂
- LPG
- Propane
- Smoke
- Temp
- LPG_Leakage