| **2021 혁신성장 빅데이터 분석 프로젝트 요약서** | | | |
| --- | --- | --- | --- |
| **팀 명** | Da.Gul | **팀 장** | 김승건 |
| **팀 원** | 김민송, 김정연, 남승규, 오지현 | | |
| **주 제 명** | 해양 플랫폼 데이터의 상관관계 분석 | | |

| 구분 | 프로젝트 상세내용 | |
| --- | --- | --- |
| 프로젝트명 | 해양 플랫폼 데이터의 상관관계 분석 | |
| 구현 목표 | 1. 해양 데이터 수집을 위해 설치한 부표에서 얻어온 데이터셋으로 설치된 체인의 교체시기를 예측 | |
| 구현 내용 | \*해양 플랫폼이란 무엇인가  -계류시스템, 피로 파괴  \*부표로부터 얻은 해양 데이터 전처리  \*전처리 데이터셋을 이용하여 의미 있는 상관관계 찾기  (날짜별로 데이터셋을 확인하기)  \*해당 부표에 설치된 체인의 교체시기 예측 시각화한 결과를 해석하기  \*Machine Learning(Linear Support Vector, XGboost)  \*모델링 과정을 거친 후 예측  \*모델 예측을 통한 그래프와 실제 데이터와의 유사성 확인 | |
|
| 상세 설명 | **1. 연구 과제**  1)연구 과제 설명  -해양 데이터를 수집하기 위해 설치해 둔 부표로부터 데이터 셋을 가져옴  -가져온 데이터셋으로부터 의미 있는  상관관계를 발견  -해당 부표에 설치된 체인의 교체 시기를 예측하는 것이 목표  -이를 위해 날짜별로 데이터에 대한 검증, 필요한 부분에 대한 데이터로 제한 및 모델 예측을 통해 체인 장력 피로파괴도/임계점을 추측  1-2) 솔루션  모델링 작업을 통해 모델 검증과정을 거치고 앞으로 예상되는 결과에 대한 예측모델을 만듬  **2. 데이터 설명 및 모델링**  2-1) 사용 데이터  **-Dataset Columns-**  **>AIRGAP 공기영향 데이터**  >CURRENT\_DEPTH, CURRENT\_SPEED 조류 데이터  >MOORING\_TENSION, MOORING\_LENGTH 부표 데이터  >MOTION\_TILTX, MOTION\_TILTY부표의 움직임  >MOTION\_YAW 부표의 기울기  >POSITION\_X, POSITION\_Y 부표의 위치  >WIND1\_DIRECTION, WIND1\_SPEED바람1데이터  >Wind2\_DIRECTION, WIND2\_SPEED바람2 데이터  **-Tested Datasets-**  >2015-08-07  >2015-08-08  >2015-09-03  >2015-09-09  >2015-09-11  2-2) 이상치 발견(Outlier)   1. 8월 7일, 8일 데이터 경향이 다르며, 센서의 오작동이 보이고 불규칙하게 센서의 동작된 것이 많이 보였음 2. 8월 8일, 9월 3일 데이터 경향이 다름을 확인 3. 9월 3일, 9월 7일 데이터 경향이 다름을 확인 4. **9월 7일, 9월 11일 데이터 경향이 비슷함을 확인** 5. 9월 7일 데이터를 바탕으로 ML을 실행하여 11일을 예측 설정   -이에 9월3일 혹은 9월 9일의 데이터 중 선택해야함  **-**두 날짜의 데이터를 살펴보면 9월3일의 YAW값, POSITION\_X, POSITION\_Y, TILT\_X, TILT\_Y의 폭이 9월9일 보다 더 넓게 오르내리며 찍혔기 때문에 9월 3일의 데이터 선택 | 2-3) 상관관계 분석(Correlation)  -데이터의 추세 확인  -각각의 Mooring Tension 그룹은 상이한 부분에서 상관관계를 나타냄을 확인  -모든 영역에서 상관관계가 0인 Feature가 없음, 이에 Machine Learning을 나눠서 진행하지 않고 전체 데이터를 통해 진행  2-4) Machine Learning  -데이터셋 전체 경향 파악  \*전처리와 주의점  -주의점: 과적합  2-5) Machine Learning-분석 도구   1. Linear Support Vector 2. XGboost   -MLScore  -HyperParameter  **3. 결론 및 시사점**   1. Df4 2015-09-07 를 바탕으로 만든 ML-모델로 df5 2015-09-11을 예측 2. ML을 통해 예측한 그래프와 실제 데이터 그래프의 성향이 유사함을 확인 3. 보유한 데이터를 통해 ML을 통한 예측 가능성 4. 이를 위한 적합한 모델링의 중요성 5. 기본 데이터와 예측을 위한 데이터의 경향의 중요성 6. 시간의 흐름에 따라 바뀌는, 반복되는 자연현상을 관찰, 데이터화 하여 유사한 경향을 가진 데이터셋간의 ML을 통해 피로도를 예측가능   **6. 활용방안 및 기대효과**  -모든 센서에 대한데이터의 예측과정을 통해 결측치를 보간하여 완성도를 높일 수 있음 |