전자제조 산학프로젝트 최종보고서

**과제명: 센서 데이터 기반의 냉장고 기능 이상탐지 고도화**

김재호, 박순혁, 서승원, 윤하영

지도교수: 정진우

텍스트이(가) 표시된 사진

자동 생성된 설명

**목 차**

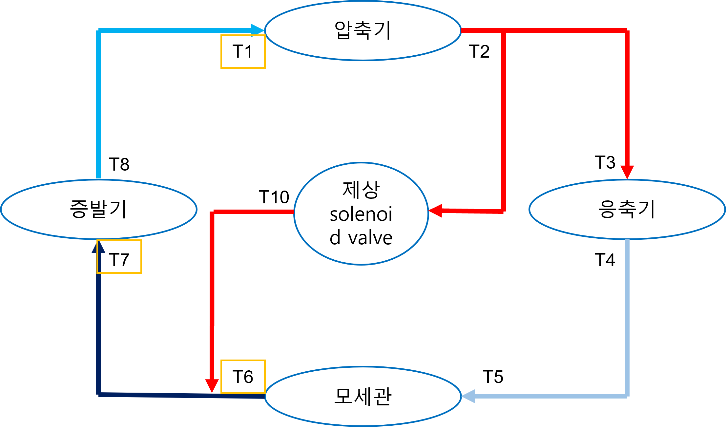
**I. Introduction**

**Backgruond & Motivation**

텍스트이(가) 표시된 사진

자동 생성된 설명

제이오텍은 보다 안전하고 신뢰할 수 있는 실험 환경을 위하여 연구용 실험장비(열풍건조기, 배양기), 환경 신뢰성 검증 장비(항온항습기, 항온수조, 냉동기), 산업용QC장비 제조, 도소매 등 물질 검사, 측정 및 분석기구를 제조하는 연구 실험 장비 기업.

산학협력팀은 이번 전자제조 산학프로젝트로 제이오텍의 많은 연구 실험 장비 중 냉동기에서 발생하는 적상과 냉매 부족 현상을 예측하고자 함.

<그림 1> 냉동기 구조

* 적상 : <그림 1>에서 냉동기의 증발기에 냉매가 증발하여 열교환기에서 주위의 열을 흡수해서 주변 환경보다 온도가 낮아지면 공기 중에 수증기가 얼어 서리가 끼게 되는 현상.
* 제상 : 적상에 위해 냉동기의 효율이 저하. 이를 방지하기 위한 증발기에 낀 서리를 녹이는 작업. 제이오텍에서는 이러한 제상 작업을 <그림 1>의 T2로 압축기에서 나온 고온고압의 냉매를 T6의 증발기 입구로 보내는 고압가스 제상 방식으로 이루어짐.

원래 기업 측에서 요구했던 방향은 냉동기의 적상을 예측하는 것이었으나, 이전 산학프로젝트에서 냉매부족을 예측하는 과제를 수행했기 때문에 프로젝트의 연속성을 감안하여 적상과 냉매부족을 동시에 예측하는 연구를 진행하였음.

**Research Highlights/Contributions**

* 냉장고 센서 데이터센서 데이터 기반으로 적상 및 냉매부족을 자동으로 감지할 수 있는 시스템을 제안함
* 센서들 중 적상과 냉매부족 감지에 유리한 센서 종류를 판별함
* 중복되는 센서인 증발기 출구 센서에 대해서 더 유리한 센서를 제안함
* 제안하는 시스템들은90% 이상의 정확도로 냉장고 적상 및 냉매부족 상태를 판별함

**II. Dataset**

**Description**

1. 1차 년도 dataset

* 실험 환경
* 입력 온도 5도, 외부 온도 25도, without 히터
* 냉매량 60%, 70%, 80%, 90%, 100% 조절
* 변수
* 압력 센서 (3)
  + lp1, hp1, hp2
* 온도 센서 (17)
  + 압축기 : comp in, comp out
  + 응축기 : cond in, cond out, cond air in temp, cond air out temp
  + 증발기 : evap in, evap out1, evap out2, eva air in temp, eva air out temp
  + 팽창기 : Exp in, Exp, out, Sol out
  + 내부, 중앙, 외부 온도 : inside temp, center, outside temp
* 냉매 60%, 70%, 80%, 90%, 100%에 대한 센서 데이터 파일 5개
* 10초 단위로 각 센서의 min, max 값이 저장되어 있음, 총 18971개의 row

1. 2차 년도 dataset

* 실험 환경
* 입력 온도 0도 or 4도, 외부 온도 15도 or 25도, with or without 히터
* 냉매량 100% 고정
* 히터 on/off로 적상 현상 발생 조절
* 변수는 1차 년도와 동일
* 총 14개의 센서 데이터 파일 존재
* (외부 0도, 내부 25도, without 히터)에 대한 센서 데이터 파일 4개
* (외부 4도, 내부 25도, with 히터)에 대한 센서 데이터 파일 3개
* (외부 4도, 내부 15도, with 히터)에 대한 센서 데이터 파일 7개,
* 10초 단위로 각 센서의 min, max 값이 저장되어 있음, 총 110129개의 row

**Preprocessing**

1. (증발기 출구 온도 – 증발기 입구 온도) 변수 추가

* T8\_1-T7\_MIN : (T8\_ evap out1\_MIN – T7\_evap in\_MIN)
* T8\_1-T7\_MAX : (T8\_ evap out1\_MAX – T7\_evap in\_MAX)
* T8\_2-T7\_MIN : (T8\_ evap out2 \_MIN – T7\_evap in \_MIN)
* T8\_2-T7\_MAX : (T8\_ evap out2 \_MAX – T7\_evap in \_MAX)

1. 결측치가 존재하는 column 제거 or 선형 보간하여 사용

* eva air out temp\_MIN, eva air out temp\_MAX, out side temp\_MAX
* 1차 dataset은 결측치 없음
* 2차 dataset
* eva air out temp\_MIN : ‘+OVER’값 0개, ‘-OVER’값 4755개
* eva air out temp\_MAX : ‘+OVER’값 0개, ‘-OVER’값 1936개
* out side temp\_MAX : ‘+OVER’값 1개, ‘-OVER’값 0개

1. Label space

* 1차 dataset : [0(60%), 1(70%), 2(80%), 3(90%), 4(100%)]
* 2차 dataset : [0(정상), 1(적상)]
* 1차 + 2차 dataset : [0(60%), 1(70%), 2(80%), 3(90%), 4(100%), 5(적상)]

**III. Data Analysis & Visualization**

**EDA**

• 시험 조건별 히터 ON/OFF 전후의 변수들의 패턴은 다음 그림과 같은 패턴을 보이고 있음

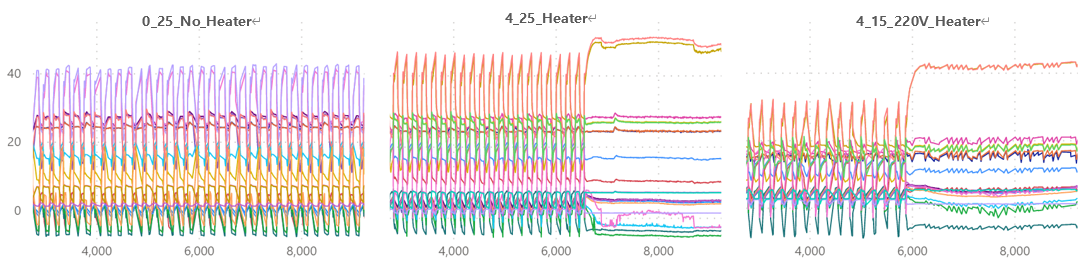
****

그림 1. 시험 조건별 히터 ON/OFF 전후의 센서값 변화 (index 3,000~9,000만 표시 및 범례 생략)

• ‘eva air out temp\_MIN’ 변수의 ‘0-25-히터없음’ 조건에서 ‘-OVER’ 입력오류 및 이상치 다수 발생

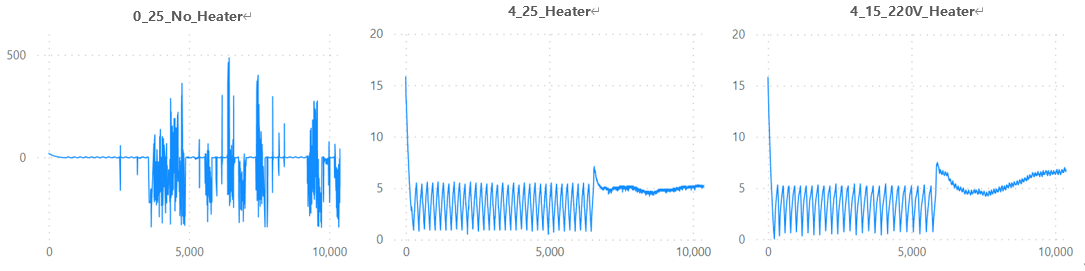


그림 2. 시험 조건별 ‘eva air out temp\_MIN’ 온도 변화 (index 3,000~10,000만 표시)

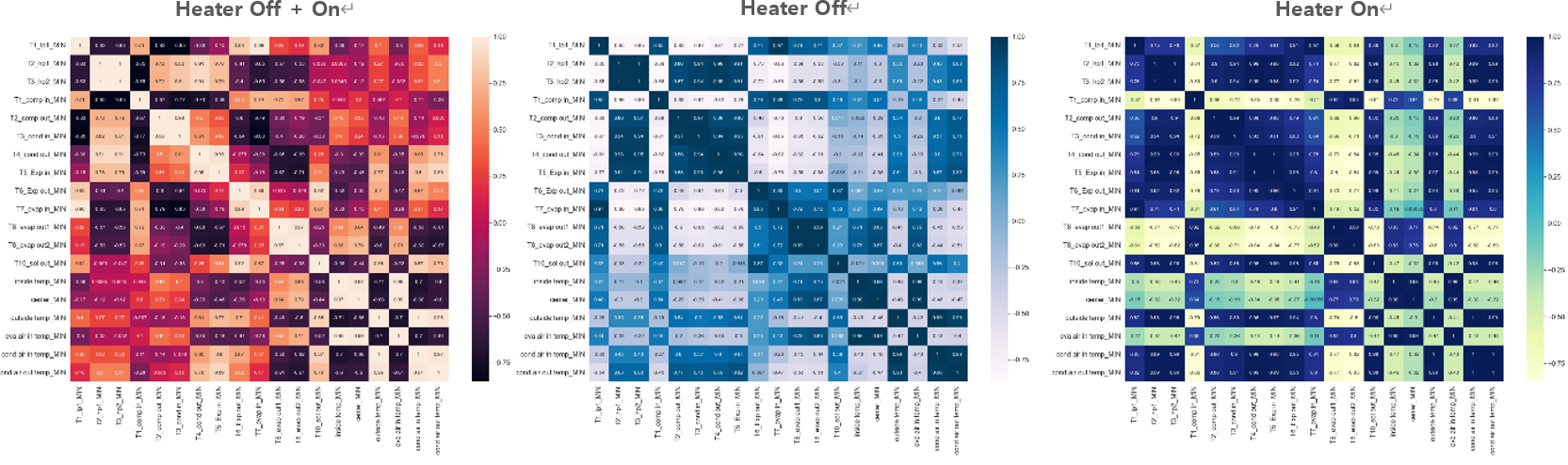


그림 3. 변수별 상관관계 확인을 위한 히트맵 (전체, 히터OFF, 히터ON 비교)

**실험 조건별 변수 비교 - Box-Plot**

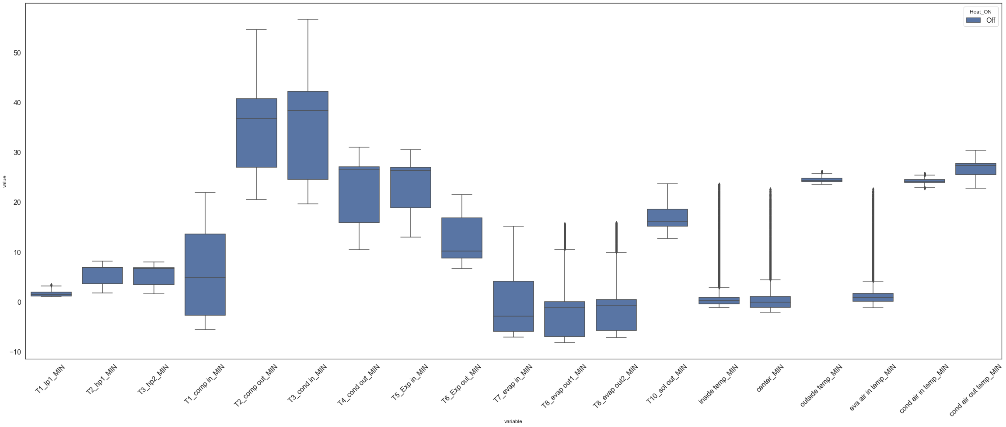


그림 3. 0\_25\_NoHeater 실험 변수 비교 Box-plot

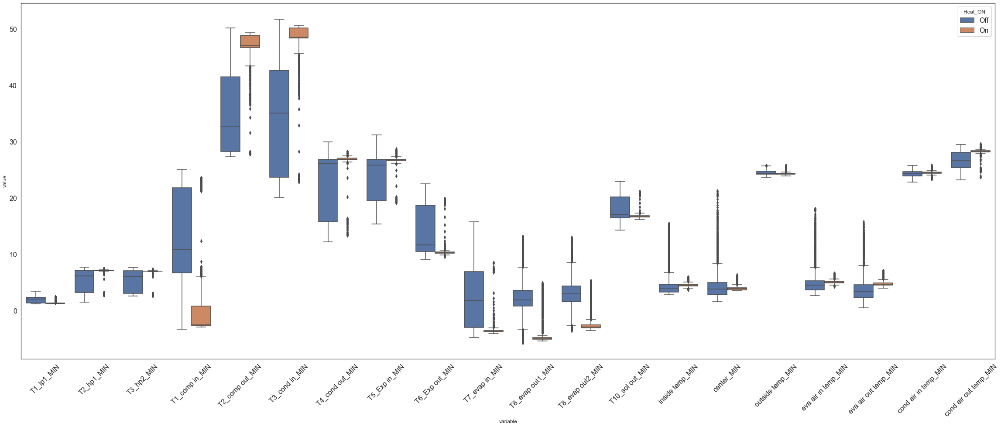


그림 4. 4\_25\_Heater 실험 변수들의 히터 ON/OFF 비교 Box-plot

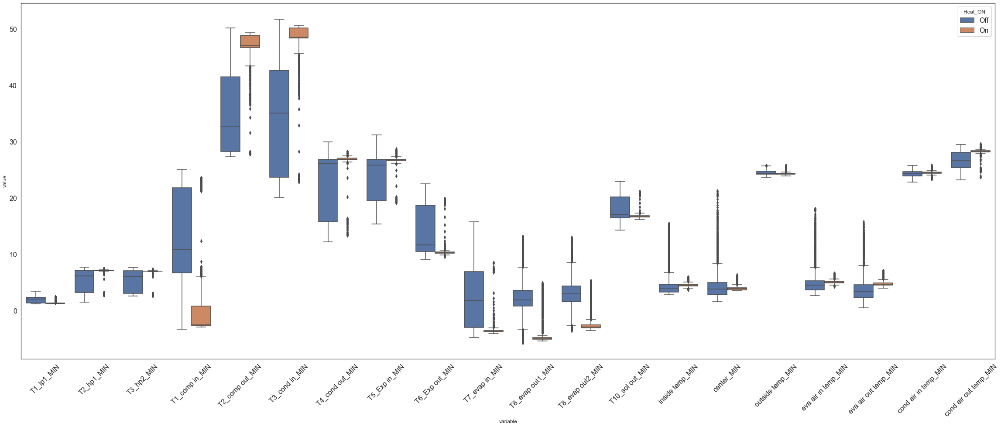


그림 5. 4\_15\_220V\_Heater 실험 변수들의 히터 ON/OFF 비교 Box-plot

**실험 조건별 변수 비교 - Rader-chart**

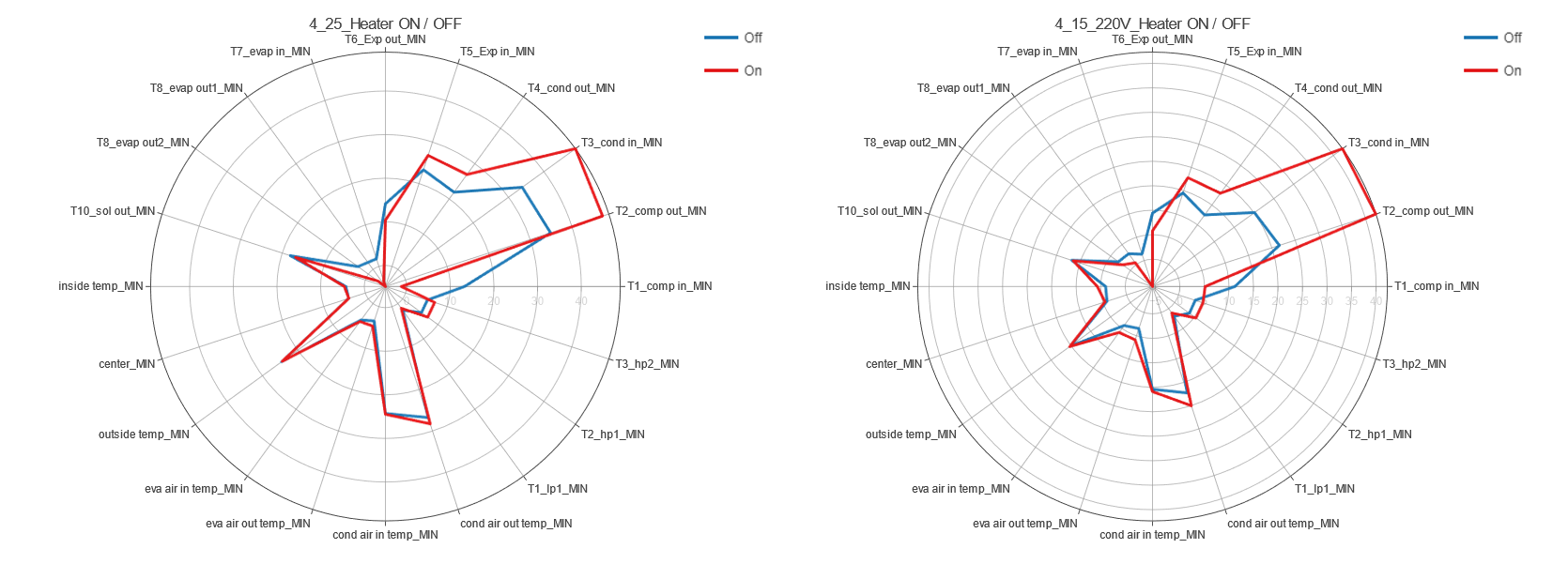


그림4. 히터 ON/OFF에 따른 센서 평균값 비교 (좌측:4\_25\_히터, 우측: 4\_15\_220V히터)

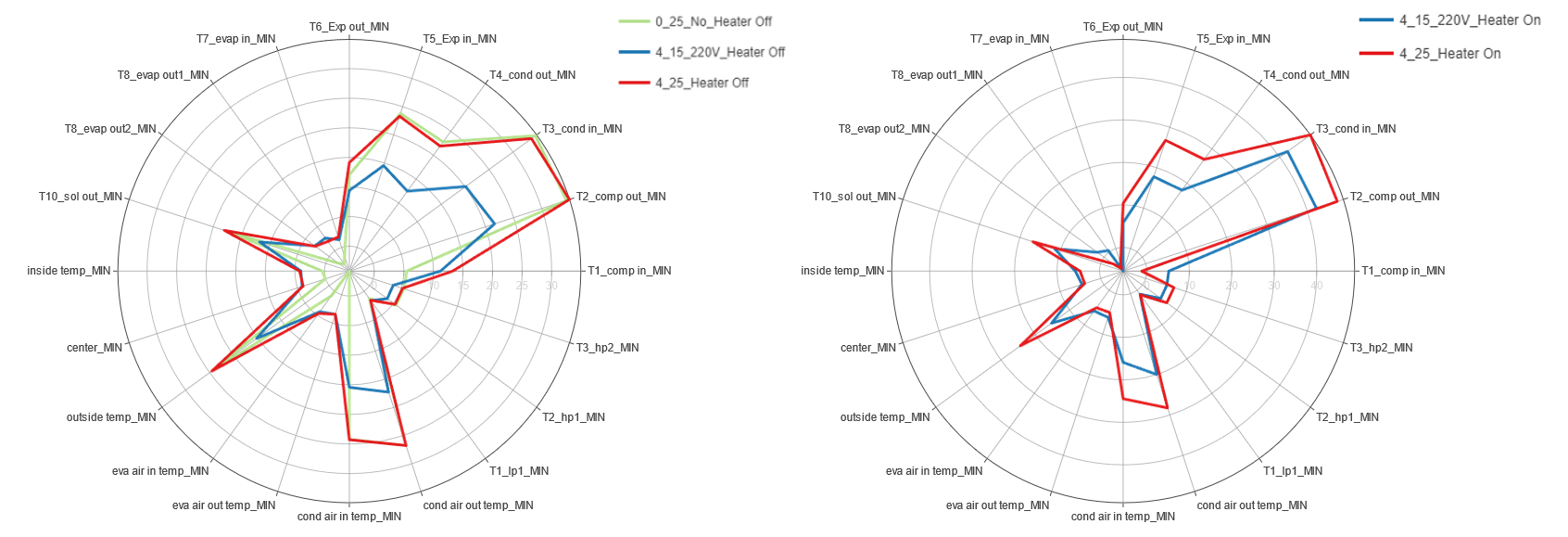


그림5. 히터 ON/OFF별 실험 조건에 따른 센서 평균값 비교 (좌측 : 히터 OFF, 우측 : 히터 ON)

**IV. Model Development & Result**

**Model**

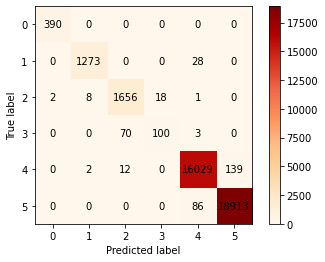
* 목표 : Multi-Class Classification
* Method : Logistic Regression, LightGBM, SVC, Random forest,
* 독립 변수
* 결측치가 존재하는 변수를 제외한 나머지
* 종속 변수
* Heat\_ON: [0(60%), 1(70%), 2(80%), 3(90%), 4(100%), 5(적상)]

**Performance**

1. **Logistic Regression**

* 1st Model
* 사용 변수 : All features
* Accuracy : 0.9905, Micro F1 : 0.9905, Macro F1: 0.9372
* 2nd Model
* Feature selection : RFECF (cv=3)
* 사용 변수 : 'T1\_comp in\_MAX', 'T2\_comp out\_MAX', 'T3\_cond in\_MIN', 'T3\_cond in\_MAX', 'T5\_Exp in\_MIN', 'T5\_Exp in\_MAX', 'T6\_Exp out\_MIN', 'T8\_evap out1\_MAX', 'inside temp\_MAX', 'eva air in temp\_MIN', 'eva air in temp\_MAX', 'cond air out temp\_MIN'
* Accuracy : 0.9851, Micro F1 : 0.9851, Macro F1: 0.9051

**Confusion matrix**



* Summary
* 냉매부족 70% 와 80% 를 많이 헷갈려함
* 많은 feature를 사용하는 것이 feature selection으로 얻은 feature들만 사용하는 것보다 성능적인 측면에서 유리함

1. **LightGBM**

* 1st Model

|  |  |  |
| --- | --- | --- |
| accuracy | Micro F1 | Time(sec) |
| 0.9996 | 0.9995 | 79.64 |

* + 사용 변수 : All features
* Feature importance

**테이블이(가) 표시된 사진

자동 생성된 설명**

* 2nd Model(Based on Feature importance)

|  |  |  |
| --- | --- | --- |
| accuracy | Micro F1 | Time(sec) |
| 0.9996 | 0.9996 | 71.29 |

* + 사용 변수 : 'eva air out temp\_MIN', 'inside temp\_MIN', 'T3\_cond in\_MIN', 'eva air out temp\_MAX', 'T2\_comp out\_MAX', 'T1\_comp in\_MAX', 'T3\_cond in\_MAX', 'T5\_Exp in\_MAX', 'T8\_2-T7\_MIN', 'T5\_Exp in\_MIN'
* 3rd Model
  + Feature selection : RFECV(Recursive Feature Elimination with Cross Validation)
  + 사용 변수 : 'T1\_lp1\_MIN', 'T2\_hp1\_MIN', 'T1\_comp in\_MIN', 'T1\_comp in\_MAX', 'T2\_comp out\_MIN', 'T2\_comp out\_MAX', 'T3\_cond in\_MIN', 'T3\_cond in\_MAX', 'T4\_cond out\_MAX', 'T5\_Exp in\_MIN', 'T5\_Exp in\_MAX', 'T6\_Exp out\_MIN', 'T7\_evap in\_MAX', 'T8\_evap out1\_MIN', 'T10\_sol out\_MIN', 'inside temp\_MIN', 'center\_MIN', 'eva air out temp\_MIN', 'eva air out temp\_MAX', 'T8\_1-T7\_MIN'

|  |  |  |
| --- | --- | --- |
| accuracy | Micro F1 | Time(sec) |
| 0.9997 | 0.9997 | 75.76 |

* 테이블이(가) 표시된 사진

  자동 생성된 설명Feature importance

|  |  |  |  |
| --- | --- | --- | --- |
| Model | accuracy | Micro F1 | Time(sec) |
| LightGBM  (All Feature) | 0.9996 | 0.9995 | 79.64 |
| LightGBM  (Top-10 feature) | 0.9996 | 0.9996 | **71.29** |
| LightGBM(RFECV) | **0.9997** | **0.9997** | 75.26 |

* Summary
* 피처를 적게 사용할수록 time cost측면에서 가장 빠름.
* 모델 성능은 acc와 f1 측면에서 REFEV방식으로 피처 추출해서 구축한 모델이 가장 좋음(0.0001~2 차이)

1. **SVC**

* 1st Model
* 사용 변수 : all feature
* 사용 kernel : linear kernel
* 정확도 : 0.993057
* 2nd Model
* 사용 변수 : (REFEV feature selection) T1\_lp1\_MAX, T2\_hp1\_MAX, T7\_evap in\_MIN, T7\_evap in\_MAX, T10\_sol out\_MAX, outside temp\_MIN, outside temp\_MAX, cond air in temp\_MAX 를 제외한 나머지 feature
* 사용 kernel : linear kernel
* 정확도 : 0.992923
* 3rd Model
* 사용 변수 : (REFEV feature selection) T1\_lp1\_MAX, T2\_hp1\_MAX, T7\_evap in\_MIN, T7\_evap in\_MAX, T10\_sol out\_MAX, outside temp\_MIN, outside temp\_MAX, cond air in temp\_MAX 를 제외한 나머지 feature
* 사용 kernel : rdf kernel
* 정확도 : 0.961050
* 4th Model
* 사용 변수 : (A) T7\_evap in\_MIN, T7\_evap in\_MAX, T8\_evap out1\_MIN, T8\_evap out1\_MAX, T8\_evap out2\_MIN, T8\_evap out2\_MAX, T8\_1-T7\_MIN, T8\_1-T7\_MAX, T8\_2-T7\_MIN, T8\_2-T7\_MAX, inside temp\_MIN, inside temp\_MAX
* 사용 kernel : linear kernel
* 정확도 : 0.907570
* 5th Model
* 사용 변수 : T7\_evap in\_MIN, T7\_evap in\_MAX, T8\_evap out1\_MIN, T8\_evap out1\_MAX, T8\_1-T7\_MIN, T8\_1-T7\_MAX, inside temp\_MIN, inside temp\_MAX
* 사용 kernel : linear kernel
* 정확도 : 0.892987
* 6th Model
* 사용 변수 : T7\_evap in\_MIN, T7\_evap in\_MAX, T8\_evap out2\_MIN, T8\_evap out2\_MAX, T8\_2-T7\_MIN, T8\_2-T7\_MAX, inside temp\_MIN, inside temp\_MAX
* 사용 kernel : linear kernel
* 정확도 : 0.886849
* Summary
  + - 전체 feature를 사용하는 경우 정확도가 증가하는 경향을 가짐
    - RDF 커널보다 Linear 커널이 성능 향상에 더 유리함
    - 출구센서는 1번 센서가 성능향상에 유리함
    - Feature를 사용할 경우, feature selection 보다 0.085353성능이 하락함을 볼 수 있음

|  |  |  |
| --- | --- | --- |
| 모델 | Feature | 분류 정확도(Accuracy) |
| SVC (linear kernel) | All | 0.993057 |
| SVC (linear kernel) | REFEV Selected feature | 0.992923 |
| SVC (rdf kernel) | REFEV Selected feature | 0.961050 |
| SVC (linear kernel) | (A) | 0.907570 |
| SVC (linear kernel) | (A), 1번 출구센서만 | 0.892987 |
| SVC (linear kernel) | (A), 2번 출구센서만 | 0.886849 |

1. **Random forest**

* 1st Model
* 사용 변수 : All feature
* 정확도 : 0.995202
* 2nd Model
* 사용 변수 : (REFEV feature selection) T1\_lp1\_MIN, T1\_lp1\_MAX, T1\_comp in\_MIN, T1\_comp in\_MAX, T2\_comp out\_MIN, T2\_comp out\_MAX, T3\_cond in\_MIN, T3\_cond in\_MAX, T5\_Exp in\_MIN, T5\_Exp in\_MAX, T6\_Exp out\_MAX, T8\_evap out1\_MIN, T8\_evap out1\_MAX, T10\_sol out\_MIN, inside temp\_MIN, inside temp\_MAX, center\_MIN, eva air in temp\_MIN, eva air in temp\_MAX, eva air out temp\_MAX, cond air out temp\_MIN
* 정확도 : 0.995228
* 3rd Model
* 사용 변수 : (A) T7\_evap in\_MIN, T7\_evap in\_MAX, T8\_evap out1\_MIN, T8\_evap out1\_MAX, T8\_evap out2\_MIN, T8\_evap out2\_MAX, T8\_1-T7\_MIN, T8\_1-T7\_MAX, T8\_2-T7\_MIN, T8\_2-T7\_MAX, inside temp\_MIN, inside temp\_MAX
* 정확도 : 0.946440
* 4th Model
* 사용 변수 : T7\_evap in\_MIN, T7\_evap in\_MAX, T8\_evap out1\_MIN, T8\_evap out1\_MAX, T8\_1-T7\_MIN, T8\_1-T7\_MAX, inside temp\_MIN, inside temp\_MAX
* 정확도 : 0.943518’’
* 5th Model
* 사용 변수 : T7\_evap in\_MIN, T7\_evap in\_MAXT8\_evap out2\_MIN, T8\_evap out2\_MAX, T8\_2-T7\_MIN, T8\_2-T7\_MAX, inside temp\_MIN, inside temp\_MAX 정확도 : 0.943518
* 정확도 : 0.928319
* Summary

|  |  |  |
| --- | --- | --- |
| 모델 | Feature | 분류 정확도(Accuracy) |
| Random forest | All | 0.995202 |
| Random forest | Selected feature | 0.995228 |
| Random forest | (A) | 0.946440 |
| Random forest | (A), 1번 출구센서만 | 0.943518 |
| Random forest | (A), 2번 출구센서만 | 0.928319 |

* + 전체 feature를 사용하는 것 보다 feature selection으로 선택된 feautre가 성능향상에 더 유리함
  + 출구센서는 1번 센서가 성능향상에 유리함
* Feature를 사용할 경우, feature selection 보다 0.048788 성능이 하락함을 볼 수 있음

**V. Future Research Directions**

* 적은 데이터양을 극복하기 위해 대용량 데이터세트 구축
* 냉매 부족 상태에서 적상이 발생한 데이터 수집
* 실제환경에서의 냉매부족 수치 및 적상 상황의 정밀한 감지를 위해 냉매상태와 적상상황 각각 감지하는 Multi-label classification 모델을 개발

**Acknowledgement**

• 이 문서는 서울과학기술대학교 데이터사이언스 학과의 전자제조전문인력 양성사업의 참여학생들이 전자제조업체인 (주)제이오텍과 함께 2022년에 수행한 산학협력과제의 최종보고서입니다.

• 소중한 실험 데이터를 제공해주신 (주)제이오텍에게 감사드립니다.