

DATA DISCOVERY 2022

AMI 데이터 활용,
전력사용 이상 고객
예측 모델 개발

CONTENTS

- 01 추진배경
- 02 추진개요
- 03 추진내용
- 04 기대효과/향후계획
- 05 과제를 진행하며



'19~22년(3년간) 제주본부 고압고객 사용량 협정 32건 시행, 총 협정금액 59백만원 ➡

탐지되지 못한 비정상 사용 고객이 잠재되어있을 가능성이 항상 존재



CHAPTER

후진배경 INTRO

01

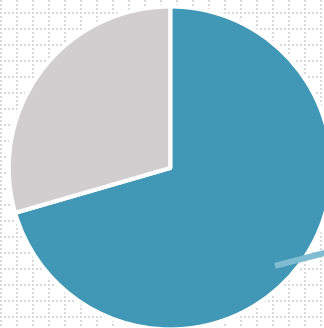
전기위약으로 인한 전기요금 누수도 상당하지만 현장방문 없이 고객의 전기위약사용을 정확하게 검출할 수 있는 수단이 없는 것이 현실이다



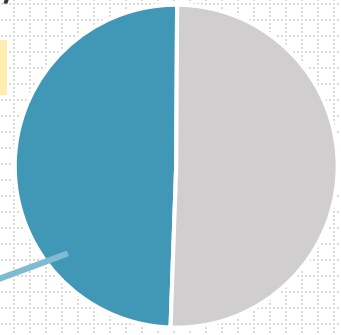
2016~2020년 전기위약 발생 통계

총 2만 1,952건
위약금 1099억

“계약종별 위반”

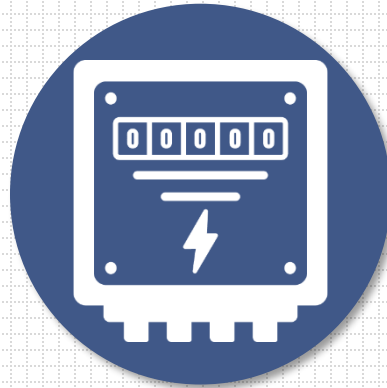


건수



위약금

현행 고압 AMI 시스템에서 제공하는 사용량 관련 기능은 제한적이며, 비정상사용 고객을 판단할 수 있는 서비스가 필요한 상황이다



현행 고압 AMI

24시간 동안 사용량
0/-인 고객 조회

계약전력 초과 사용
고객 조회

전월 사용량 대비
25% 이상 변동 고객

➔ 사용량만으로 비정상 판단 불가

CHAPTER

수전배경 현황

01

특히 수전설비 무단증설, 계약종별 위반 등의 고객사유 비정상상황은 검증할 수 있는 수단이 전혀 없는 것으로 확인되었다

| 비정상상황 | | 상 황 | 계량결과 | 검증수단 |
|----------|-----------|---|----------------------------|------|
| 계기 고장 | 계기 소손 | 고객 사용설비 과다로 계량기 내부 코일이 타서 계량이 되지 않는 것 | 계량 불가 | AMI |
| | 계기 부동 | 전기를 사용하고 있는데 계량기가 회전하지 않는 것 | 과소 계량 | AMI |
| | 계기 잠동 | 전기를 사용하고 있지 않는데 계량기가 회전하는 것 | 과다 계량 | AMI |
| | 계기과회전 | 전기를 사용하고 있는 양과 비교할 때 계량기가 과다하게 회전하는 것 | 과다 계량 | AMI |
| 업무 착오 | 계기결선 착오 | 전력량계 부설시 전원측(1차측)과부하측(2차측) 교차 결선 | 과소 계량 (사용량의 1/3 또는 2/3) | AMI |
| | 계기배수 입력착오 | 전산에 계기의 배수를 잘못 입력함 ex) 360배수 계량기를 240배로 입력 등 | 정상계량, 사용량 왜곡 | AMI |
| MOF 결상 | | MOF 내부에서 3상 중 어느 한 상에 전류가 흐르지 않는 상태 | 과소 계량 | AMI |
| 무정전 계기교환 | | 고압전력량계 교환시 정전시키지 않고 직결로 연결후 계량기 교환 | 계량 불가 | AMI |
| 기 타 | | 비상재해 등 비상시 직결 송전 | 계량 불가 | AMI |

| | | | |
|-----|--------------------|------------------|----|
| 고객측 | 수전설비 무단증설, 계약종별 위반 | 정상 계량, 사용량 왜곡 | 없음 |
|-----|--------------------|------------------|----|



전력사용량 이상고객 예측모델 개발

데이터 수집, 확인

제주본부 고압고객 시간별
검침(사용량)데이터
외(최근3년)
(MDMS, 영배4.0)

데이터전처리

- 분석용 데이터 가공
- 데이터 정제
- 이상값 및 결측치 처리

모델링 및 검증

- 사용량 이상고객 예측모델
모델링
- 알고리즘 검증

탐색적 분석

- 데이터간 분석
- 데이터 시각화

CHAPTER 03 후진내용

데이터 수집, 확인

제주본부 고압고객 시간별
검침(사용량)데이터
외(최근3년)
(MDMS, 영배4.0)

데이터전처리

- 분석용 데이터 가공
- 데이터 정제
- 이상값 및 결측치 처리

모델링 및 검증

- 사용량 이상고객 예측모델
모델링
- 알고리즘 검증

탐색적 분석

- 데이터간 분석
- 데이터 시각화

CHAPTER 03 후진내용 데이터 수집

MDMS, 영배 4.0 등의 시스템을 통해 데이터수집을 시행하였으며 가장 중점적으로 확인하고자 한 고압고객 사용량 데이터는 데이터 규모가 상당히 큰 것을 확인할 수 있었다

| 데이터명 | 수집항목 | 시스템 | 데이터수(규모) |
|----------------|---|---------------|-------------------|
| 고압고객 사용량 (3년간) | 계약종별, 계약방법, 계기번호, 계기배수, 계기유형, 검침일자, 검침시간, 시간별 사용량, 일별사용량, 월별사용량 | MDMS | 118,190,612건 |
| 정복전 이벤트 이력 | 정전발생시간, 정전복구시간, 총정전시간 | MDMS | 52,928건 |
| 고압 계기교환 이벤트 이력 | 계기번호, 부설일, 철거일, 계약번호, 계약일, 종료일, 계약변경사유 | MDMS | 22,695건 |
| 고압고객 정보 | 계약번호, 고객명, 주소, 계약종별, 사용용도, 주생산물, 산업분류, 계기번호 | 영배4.0 | 7,478건 |
| 기타 고압계기 이벤트 이력 | 이벤트 타입, 이벤트 발생시간 | MDMS | 18,005,262건 |
| 고압고객 사용량 협정내역 | 협정시작일, 협정종료일, 협정일수, 중간부하, 최대부하, 경부하, 협정의뢰유형, 협정방법 | 영배4.0 | 943건 |
| 태풍 발생 이력 | 재난명, 발생일시, 종료일시, 비상발생코드 | 재난관리 통합시스템 | 14건 ₁₀ |

CHAPTER 03 후진내용 데이터특성 확인

고객 사용량 데이터는 ①높은 결측치와 ②다양한 변수, ③시계열이라는 특징을 가진다



높은 결측치 비율

- 계약시점에 따라 사용량 측정 불가능한 불완전 데이터
- AMI 통신 이상으로 인한 계량값 없음 데이터 다수 포함



중요 변수 선정 필요

- 시간별, 계약종별, 업종, 계약변경 사유, 정전복구 유무, 교환 유무 등

여러 특성을 반영한 해석 필요



시계열 데이터

- 반복되는 시간의 흐름에 따라 시·일·월·계절별로 법칙성을 가짐
- 법칙성, 즉 패턴을 분석하여 미래의 결과를 Forecasting 가능

CHAPTER 03 후진내용

목표가설 분석방법 설정

데이터 수집, 확인

제주본부 고압고객 시간별
검침(사용량)데이터
외(최근3년)
(MDMS, 영배4.0)

데이터전처리

- 분석용 데이터 가공
- 데이터 정제
- 이상값 및 결측치 처리

모델링 및 검증

- 사용량 이상고객 예측모델
모델링
- 알고리즘 검증

탐색적 분석

- 데이터간 분석
- 데이터 시각화

CHAPTER 03 후진내용 목표가설 설정

과제수행을 위해 두가지 목표가설을 설정해 이번 과제에서 확인해보기로 했다

가설 1. 계약종별 별 패턴 존재

전기사용량 데이터는 '계약종별' 이라는 분류에 따라 각 종별의 특징을 가진 패턴 유사도가 존재할 것

➡ '계약종별' 사용패턴을 확인해 해당 종별의 패턴을 따르지 않는 데이터를 추출하면 위약의심고객 탐지가 가능 할 것

가설 2. 이벤트 전후 동일

동일고객이라면 정전 또는 계기교체 전과 후의 사용패턴의 차이가 크지 않을 것

➡ 정전, 계기교체 전후 유사도가 낮다면 무단증설 또는 계기 오결선 등 이벤트를 탐지해 낼 수 있을 것

WHY

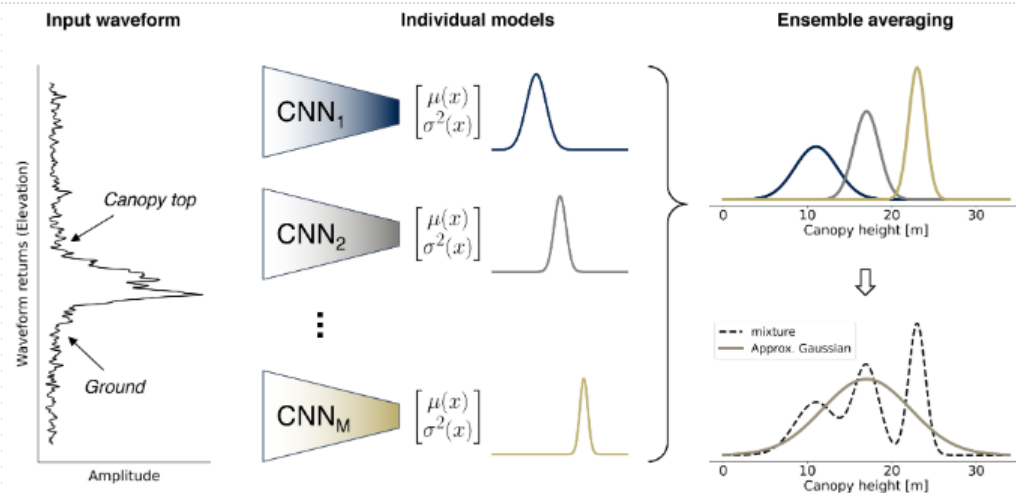
- 고압고객의 설비변동 작업에는 정전이 필수적
- 계기교체 작업시 담당자 착오로 인한 오류 가능성

CHAPTER 03 후진내용 분석방법 결정

목표가설 검증을 위한 분석방법은 각각 1D-CNN과 코사인유사도로 결정했다

1. 1D-CNN

딥러닝 알고리즘의 한 종류로 시계열 데이터나 이미지같은 격자(Grid)가 있는 데이터 분석에 효과적인 신경망



선정
이유

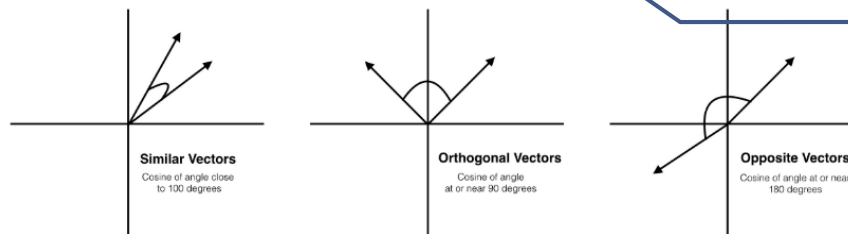
- 1차원 시계열 데이터의 예측분석이 가능
- 고객사용량 데이터는 일정기간의 시계열 데이터이며 계약종별이라는 분류기준이 존재하여 분석에 적합할 것으로 판단

CHAPTER 03 후진내용 분석방법 결정

목표가설 검증을 위한 분석방법은 각각 1D-CNN과 코사인유사도로 결정했다

2. 코사인유사도

DTW(Dynamic Time Warping, 동적시간 워핑) 분석 방법 중 하나로, 벡터간의 코사인 각도를 이용하여 서로간에 얼마나 유사한지를 산정하는 방법



DTW : 두개의 시계열이 존재할 때 상호간의 유사도를 측정하기 위한 방식

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

선택
이유

- 정전, 계기교환 전후 시계열 간의 상호 유사도를 측정할 수 있는 방식
- 연산속도가 빠르고 서로 다른 길이의 시계열 유사도 분석 가능

CHAPTER 03 후진내용

목표가설 분석방법 설정

데이터 수집, 확인

제주본부 고압고객 시간별
검침(사용량)데이터
외(최근3년)
(MDMS, 영배4.0)

데이터전처리

- 분석용 데이터 가공
- 데이터 정제
- 이상값 및 결측치 처리

모델링 및 검증

- 사용량 이상고객 예측모델
모델링
- 알고리즘 검증

탐색적 분석

- 데이터간 분석
- 데이터 시각화

CHAPTER 03 후진내용 데이터 전처리

데이터 전처리는 데이터 정제, 이상값 및 결측값 처리, 데이터 재배열의 순서로 시행하였으며, 이후 분석을 진행하는 동안에도 데이터 로드를 줄이기 위해, 더 정확한 결과값을 얻기 위해 추가적으로 시행하였다

데이터정제

서로다른 시스템에서 추출한
데이터의 일관성 유지를 위한
데이터 정제

- 데이터 타입 변환
 - 정전발생시간 (숫자→list)
- 데이터 형식 변환
 - 계기번호 구성형식 일치화
(00-0000000000)
- 정전발생시간 데이터에서
연도, 일자, 시간 추출 재조립

이상값 결측값 처리

분석결과에 영향을 미칠수 있는
이상값, 결측값의 처리방안결정

- 데이터 이상값
 - 이번 과제에서 데이터
이상값이 유의미한 결과를 가질
수 있음 → 포함
- 데이터 결측값
 - AMI 통신장애로 인한 무작위
결측치, 해지 및 신규 접수로 인한
계량값 미확보 데이터 등 → 제외

데이터재배열

분석로드를 줄이기 위한
데이터 재배열 및 정리

- 중복데이터 삭제
- 정전, 계기교환 이력이 없는
데이터 제외
- 계약번호별 시간별, 일별,
월별 사용량 데이터 구분
- 검침일자 기준 데이터 조인
 - 시간별 사용량, 검침시간
데이터 일자별 1행내 List 형식
구성

CHAPTER 03 후진내용

목표가설 분석방법 설정

데이터 수집, 확인

제주본부 고압고객 시간별
검침(사용량)데이터
외(최근3년)
(MDMS, 영배4.0)

데이터전처리

- 분석용 데이터 가공
- 데이터 정제
- 이상값 및 결측치 처리

모델링 및 검증

- 사용량 이상고객 예측모델
모델링
- 알고리즘 검증

탐색적 분석

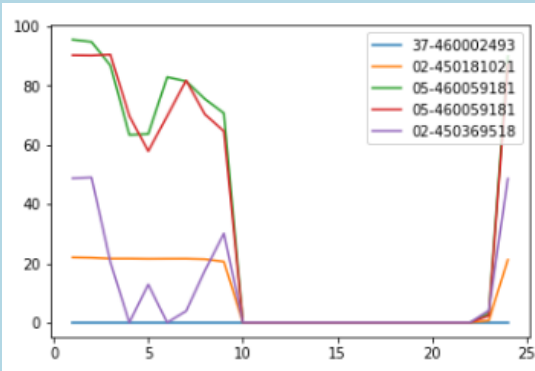
- 데이터간 분석
- 데이터 시각화

CHAPTER 03 후진내용 탐색적 분석

본격적인 데이터 분석 시행 전 랜덤 추출한 데이터를 시각화 하여 데이터 특징을 확인하고 분석방향을 재검토하였다

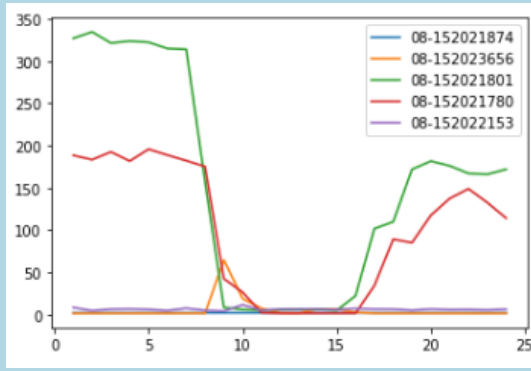
계약종별 별 사용패턴 시각화 - 일별

심야전력



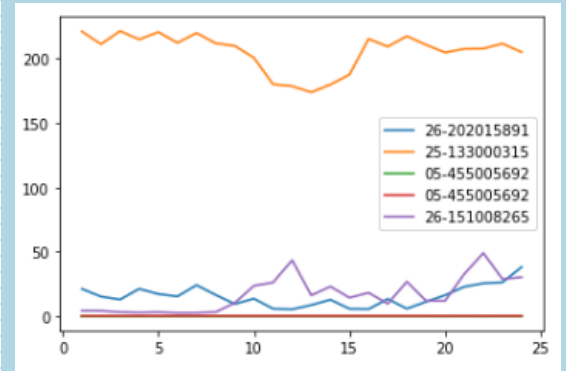
- 패턴 특성이 가장 뚜렷함
- 값이 모두 0인 경우 多

농사용



- 가장 다양한 패턴 보유함
- 심야전력과 유사 패턴 有

산업용



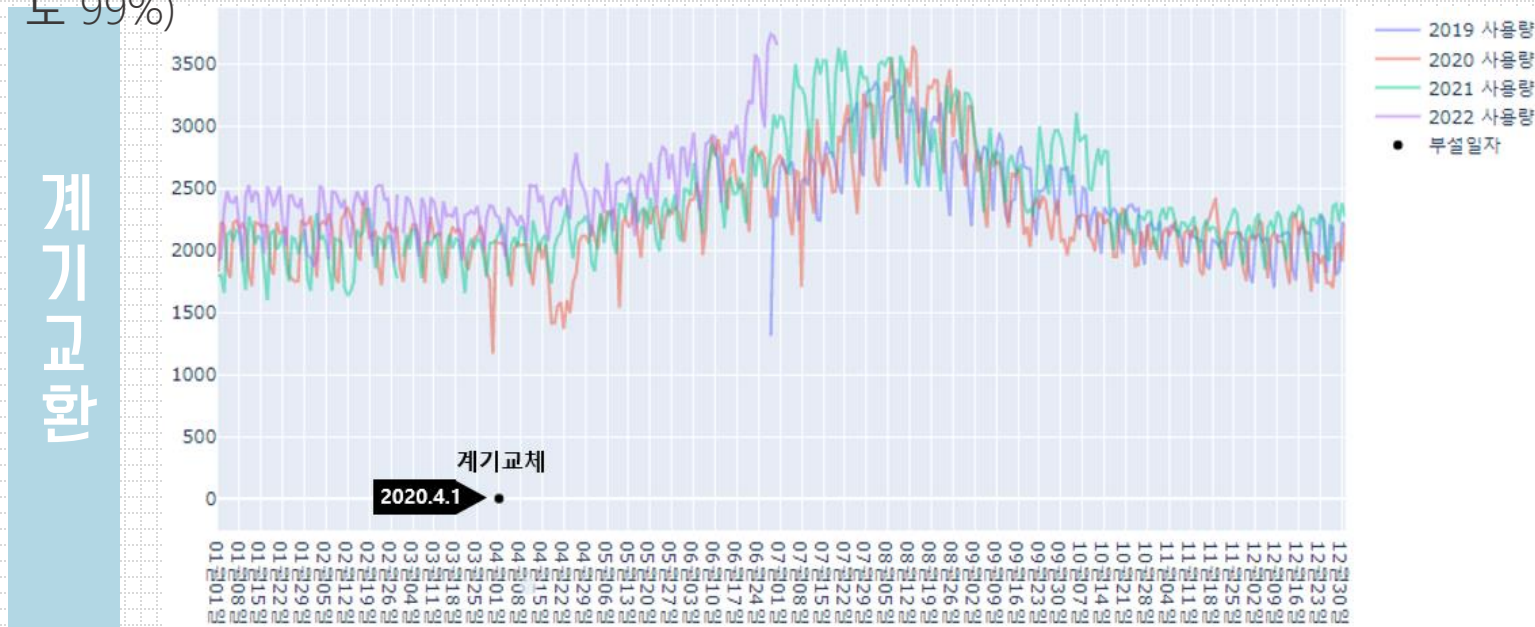
- 워킹아워에 사용량이 높은편
- 시간대별 급격한 차이 無

CHAPTER 03 후진내용 탐색적 분석

본격적인 데이터 분석 시행 전 랜덤 추출한 데이터를 시각화 하여 데이터 특징을 확인하고 분석방향을 재검토하였다

정전/계기교환 전후 사용량 시각화

계기교환에 관계없이 전후 사용량이 일정한 패턴을 보이는 정상데이터 예시 확인(3년간 유사도 99%)

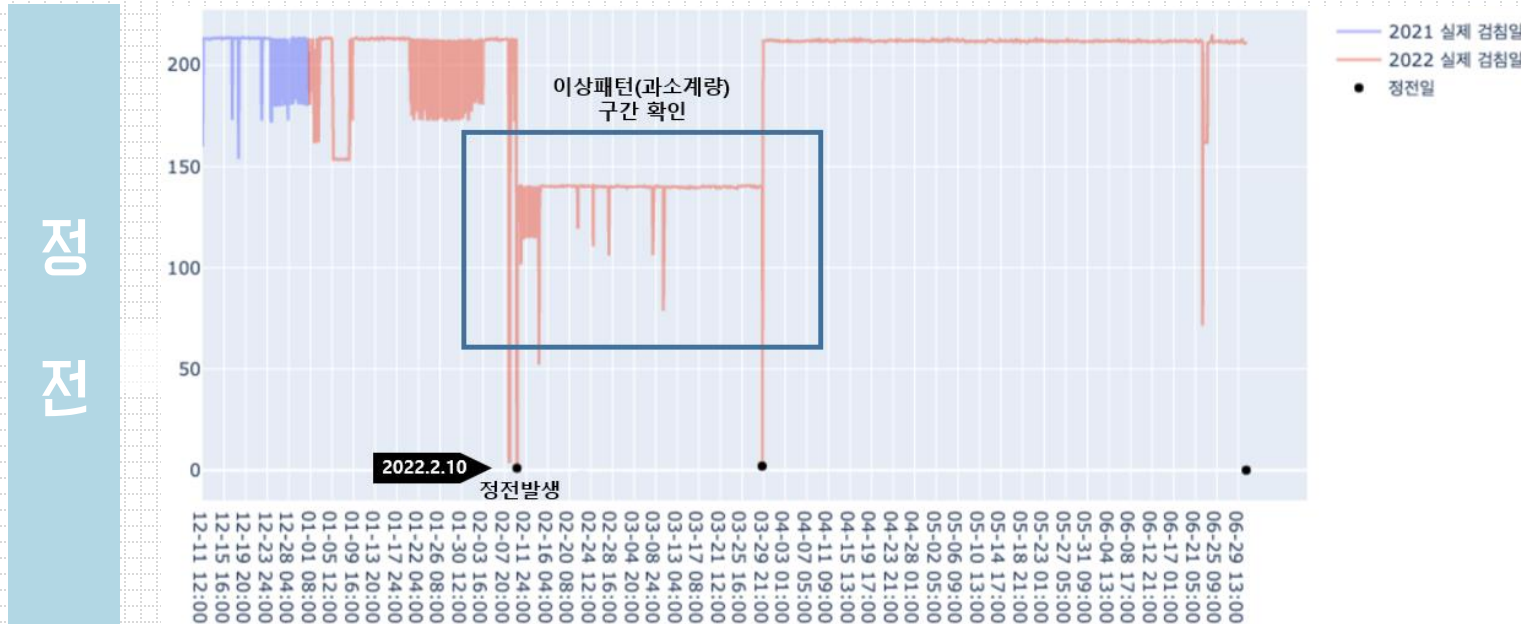


CHAPTER 03 후진내용 탐색적 분석

본격적인 데이터 분석 시행 전 랜덤 추출한 데이터를 시각화 하여 데이터 특징을 확인하고 분석방향을 재검토하였다

정전/계기교환 전후 사용량 시각화

정전 발생 후 뚜렷한 이상패턴이 확인되는 비정상데이터 예시 확인 → 이후 세부분석 시행



목표가설 분석방법 설정

데이터 수집, 확인

제주본부 고압고객 시간별
검침(사용량)데이터
외(최근3년)
(MDMS, 영배4.0)

데이터전처리

- 분석용 데이터 가공
- 데이터 정제
- 이상값 및 결측치 처리

모델링 및 검증

- 사용량 이상고객 예측모델
모델링
- 알고리즘 검증

탐색적 분석

- 데이터간 분석
- 데이터 시각화

CHAPTER 03 후진내용 CNN 모델링

03

일별사용량데이터를 활용한 계약종별 예측 모델링을 1D-CNN알고리즘을 활용하여 진행하였으며 최종 정확도 77.51%의 결과를 얻었다

1. 일별사용량데이터 활용 계약종별 예측 모델링

학습데이터 설정

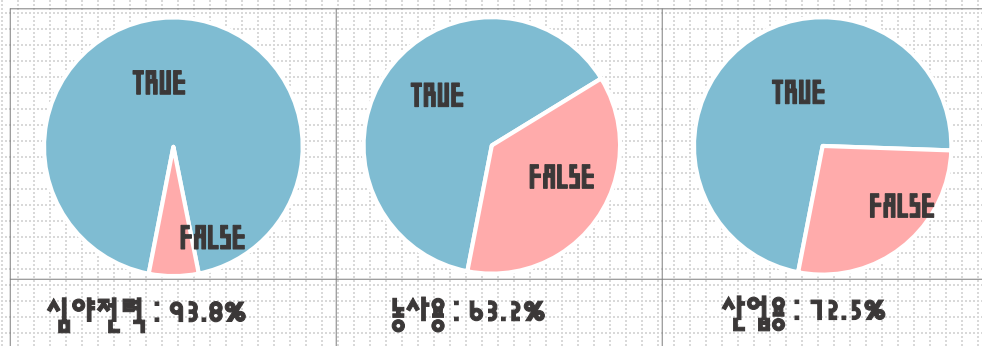
- '21년도 심야전력, 산업용, 농사용 데이터 사용 (일반용 : 데이터 過多로 부적합)

모델링

- 1D-CNN 사용
- 각 계약종별 일별 사용량 패턴을 학습하여 검침데이터의 일별 사용량 예측 및 계약종별 적정여부 확인

모델링 결과

- 최종 정확도 : 77.51%



➡ 계약종별이 달라도 패턴이 같은 케이스 존재, 추가 데이터 가공 필요

CHAPTER 03 후진내용 CNN 모델링

03

전력 사용량의 계절적 영향등을 반영하기위해 일별데이터를 월별데이터로 재가공하여 2차 모델링을 시행하였고 66.15%의 정확도를 확인했다

2. 월별사용량데이터 활용 계약종별 예측 모델링 -

학습데이터 설정, 모델링

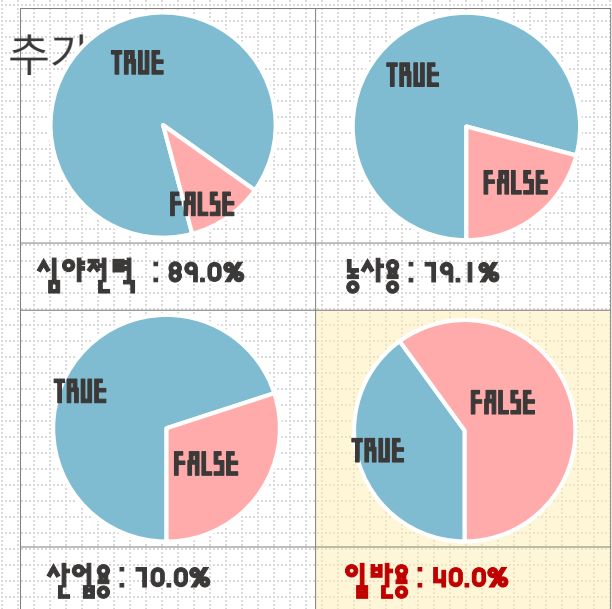
- 기존 데이터를 1개월 단위로 묶어 학습용 데이터 구성
- 각 계약종별 월별 사용량 패턴을 학습하여 검침데이터의 월별 사용량을 예측, 계약종별의 적정여부를 확인하는 모델
- 전체 계약종별 중 가장 큰 비중을 차지하는 일반용 데이터 추가
- 1D-CNN 사용

모델링 결과

- 최종 정확도 : 66.15%

➔ 패턴이 더욱 명확해질 것이라는 예상과 달리 정확도가 10% 이상 감소

일반용 : 종별 내 다양한 사용패턴(업종) → 분류 정확도 LOW



CHAPTER 03 후진내용 CNN 모델링

03

최초 일별사용량데이터 모델과 동일한 조건으로 3차 모델링을 시행하였고 최종 정확도 74.97%를 확인하였다. 이를 바탕으로 위약의심고객 리스트를 추출한다

3. 월별사용량데이터 활용 계약종별 예측 모델링 -

학습데이터 재설정

- 일반용 데이터 제외하여 모델 재설정

모델링 결과

- 최종 정확도 : 74.97%

결과 활용

- 결과(위약의심고객 리스트) EXCEL 추출하여 위약 현장 확인 기초자료로 활용 가능

의심고객 추출기준

- 실제 계약종별 \neq 예측 계약종별
- 실제 계약종별과의 유사도 20% 이하
- 예측 계약종별과의 유사도 90% 이상

```
Epoch 1/20
31/31 [=====] - ETA: 0s - loss: 1.2215 - acc: 0.4765
Epoch 1: val_acc improved from -inf to 0.54642, saving model to best_month_model.h5
31/31 [=====] - 42s 1s/step - loss: 1.2215 - acc: 0.4765 - val_loss: 1.1007 - val_acc: 0.5464
Epoch 2/20
31/31 [=====] - ETA: 0s - loss: 1.0410 - acc: 0.5842
Epoch 2: val_acc improved from 0.54642 to 0.59585, saving model to best_month_model.h5
31/31 [=====] - 46s 2s/step - loss: 1.0410 - acc: 0.5842 - val_loss: 1.0210 - val_acc: 0.5958
Epoch 3/20
31/31 [=====] - ETA: 0s - loss: 0.9601 - acc: 0.6138
Epoch 3: val_acc improved from 0.59585 to 0.59350, saving model to best_month_model.h5
31/31 [=====] - 47s 2s/step - loss: 0.9601 - acc: 0.6138 - val_loss: 0.9685 - val_acc: 0.5935
Epoch 4/20
31/31 [=====] - ETA: 0s - loss: 0.9049 - acc: 0.6415
Epoch 4: val_acc did not improve from 0.59350
31/31 [=====] - 47s 2s/step - loss: 0.9049 - acc: 0.6415 - val_loss: 0.9727 - val_acc: 0.5922
Epoch 5: val_acc improved from 0.59350 to 0.62798, saving model to best_month_model.h5
31/31 [=====] - 46s 2s/step - loss: 0.8582 - acc: 0.6590 - val_loss: 0.8984 - val_acc: 0.6280
Epoch 6/20
31/31 [=====] - ETA: 0s - loss: 0.8011 - acc: 0.6859
Epoch 6: val_acc improved from 0.62798 to 0.63263, saving model to best_month_model.h5
31/31 [=====] - 47s 2s/step - loss: 0.8011 - acc: 0.6859 - val_loss: 0.9061 - val_acc: 0.6326
Epoch 7/20
31/31 [=====] - ETA: 0s - loss: 0.7842 - acc: 0.6876
Epoch 7: val_acc did not improve from 0.63263
31/31 [=====] - 48s 2s/step - loss: 0.7842 - acc: 0.6876 - val_loss: 0.8945 - val_acc: 0.6320
Epoch 8/20
31/31 [=====] - ETA: 0s - loss: 0.7200 - acc: 0.7216
Epoch 8: val_acc improved from 0.63263 to 0.65053, saving model to best_month_model.h5
31/31 [=====] - 51s 2s/step - loss: 0.7200 - acc: 0.7216 - val_loss: 0.8816 - val_acc: 0.6505
Epoch 9/20
31/31 [=====] - ETA: 0s - loss: 0.6719 - acc: 0.7436
Epoch 9: val_acc did not improve from 0.65053
31/31 [=====] - 49s 2s/step - loss: 0.6719 - acc: 0.7436 - val_loss: 0.8741 - val_acc: 0.6459
Epoch 10/20
31/31 [=====] - ETA: 0s - loss: 0.6407 - acc: 0.7600
Epoch 10: val_acc did not improve from 0.65053
31/31 [=====] - 50s 2s/step - loss: 0.6407 - acc: 0.7600 - val_loss: 0.9271 - val_acc: 0.6386
Epoch 11/20
31/31 [=====] - ETA: 0s - loss: 0.6200 - acc: 0.7653
Epoch 11: val_acc did not improve from 0.65053
31/31 [=====] - 47s 2s/step - loss: 0.6200 - acc: 0.7653 - val_loss: 0.9030 - val_acc: 0.6499
Epoch 12/20
31/31 [=====] - ETA: 0s - loss: 0.5737 - acc: 0.7889
Epoch 12: val_acc did not improve from 0.65053
31/31 [=====] - 47s 2s/step - loss: 0.5737 - acc: 0.7889 - val_loss: 0.9430 - val_acc: 0.6227
Epoch 13/20
31/31 [=====] - ETA: 0s - loss: 0.5343 - acc: 0.8055
Epoch 13: val_acc did not improve from 0.65053
31/31 [=====] - 46s 1s/step - loss: 0.5343 - acc: 0.8055 - val_loss: 0.9812 - val_acc: 0.6233
Epoch 14/20
31/31 [=====] - ETA: 0s - loss: 0.5184 - acc: 0.8055
Epoch 14: val_acc improved from 0.65053 to 0.66910, saving model to best_month_model.h5
31/31 [=====] - 46s 1s/step - loss: 0.5184 - acc: 0.8055 - val_loss: 0.8951 - val_acc: 0.6691
Epoch 14: early stopping
```

월 단위 검침 데이터 입력시 계약종별 및 위약 의심고객 자동 추출

CHAPTER 03 후진내용 CNN 모델링

03

2022년 제주본부 고압고객 데이터를 대상으로 모델링 검증을 시행하였고
그결과 위약의심 대상 104건을 추출할 수 있었다

3. 월별사용량데이터 활용 계약종별 예측 모델링 -

검증 시행

- 2022년 제주본부 고압고객 사용량 데이터를 입력하여 계약종별 예측모델 검증 시행
- 전체 7,668 데이터 중 최종 현장확인필요(위약의심) 대상 104건 추출

검증 결과

| | 원본계약종 | 계약번호 | 예측분 | 심야전력확률 | 농사용확률 | 산업용 확률 | 휴지확률 | 확인필요 |
|------|-----------|-----------|------|--------------|--------------|--------------|--------------|------|
| 7644 | 산업용(을)고압A | 813282069 | 산업용 | 1.70000009% | 10.99999994% | 87.31499910% | 0.00000000% | X |
| 7645 | 농사용(을)고압A | 813282087 | 산업용 | 2.60000005% | 28.70000005% | 68.74200106% | 0.00000000% | X |
| 7646 | 산업용(갑)고압A | 813288697 | 산업용 | 17.29999930% | 27.39999890% | 55.15099764% | 0.20000001% | X |
| 7647 | 산업용(갑)고압A | 813394288 | 심야전력 | 89.60000277% | 2.60000005% | 7.77800009% | 0.00000000% | X |
| 7648 | 농사용(갑) | 813443118 | 농사용 | 31.00000024% | 43.79999936% | 24.93699938% | 0.20000001% | X |
| 7649 | 농사용(을)고압A | 813455444 | 산업용 | 13.79999965% | 5.70000000% | 80.46699762% | 0.00000000% | X |
| 7650 | 농사용(을)고압A | 813503972 | 심야전력 | 63.80000114% | 30.70000112% | 5.15800007% | 0.30000000% | X |
| 7651 | 산업용(을)고압A | 813643456 | 심야전력 | 97.39999771% | 0.70000002% | 1.14799999% | 0.70000002% | ○ |
| 7652 | 산업용(을)고압A | 813643465 | 심야전력 | 98.10000062% | 1.30000003% | 0.52800002% | 0.10000000% | ○ |
| 7653 | 심야전력(을)II | 813644936 | 산업용 | 2.40000002% | 4.69999984% | 92.93900132% | 0.00000000% | ○ |
| 7654 | 산업용(을)고압A | 813657672 | 산업용 | 2.19999999% | 25.90000033% | 71.84600234% | 0.00000000% | X |
| 7655 | 산업용(갑)고압A | 813670167 | 산업용 | 11.99999973% | 11.50000021% | 76.34999752% | 0.20000001% | X |
| 7656 | 농사용(을)고압A | 813695122 | 농사용 | 0.80000004% | 97.10000157% | 2.05700006% | 0.00000000% | X |
| 7657 | 산업용(갑)고압A | 813719080 | 산업용 | 5.09999990% | 10.59999987% | 84.22200084% | 0.00000000% | X |
| 7658 | 산업용(갑)고압A | 813722227 | 심야전력 | 43.79999936% | 20.00000030% | 36.13100052% | 0.00000000% | X |
| 7659 | 심야전력(갑) | 813750204 | 휴지 | 10.80000028% | 0.89999996% | 3.66600007% | 84.70000029% | X |
| 7660 | 산업용(을)고압A | 813765341 | 산업용 | 2.40000002% | 10.70000008% | 86.87899709% | 0.00000000% | X |
| 7661 | 산업용(을)고압A | 813773653 | 산업용 | 15.99999964% | 14.69999999% | 69.18600202% | 0.00000000% | X |
| 7662 | 산업용(을)고압A | 813773671 | 산업용 | 26.49999857% | 25.29999912% | 48.12999964% | 0.00000000% | X |
| 7663 | 산업용(을)고압A | 813906449 | 산업용 | 1.49999997% | 5.09999990% | 93.41800213% | 0.00000000% | X |
| 7664 | 농사용(을)고압A | 813917491 | 농사용 | 0.40000002% | 85.60000062% | 13.95200044% | 0.00000000% | X |
| 7665 | 산업용(갑)고압A | 813967677 | 산업용 | 22.90000021% | 31.90000057% | 44.65099871% | 0.60000001% | X |
| 7666 | 산업용(을)고압A | 813981189 | 농사용 | 30.30000031% | 56.99999928% | 12.67299950% | 0.00000000% | X |
| 7667 | 농사용(을)고압A | 813990384 | 농사용 | 20.80000043% | 63.89999986% | 15.27999938% | 0.00000000% | X |
| | | | | | | | | |
| | | | | | | | 전체 | 7668 |
| | | | | | | | 확인필요(unique) | 104 |



추출 의심고객 리스트
현장부서 제공하여
현장 검증 활동 요청

CHAPTER 03 후진내용 분포 유사도

03

정전 또는 계기교환 전후의 사용량 분포유사도를 확인하기 위해 코사인 유사도 알고리즘과 상자그림 분포도를 활용해보았다

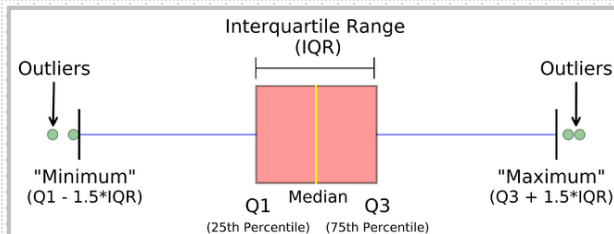
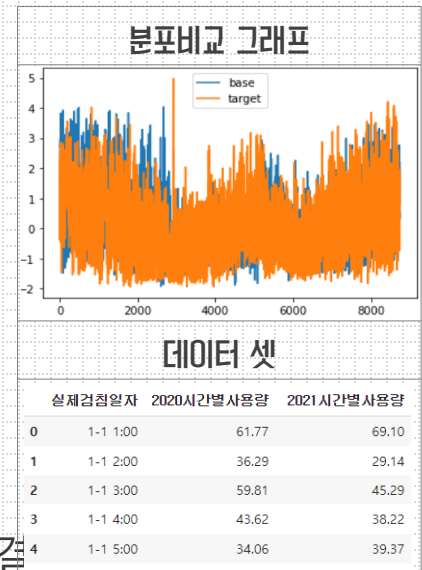
정전/계기교환 전후 유사도

데이터 정제

- 고객 사용량 연도별 분리
- 위약 가능성 낮은 계약종별 제외 : 임시전력, 주택용, 가로등
- 30분 이내 정전 데이터 제외
- 태풍내습 후 24시간 이내 정전데이터 제외
- 정전/계기교환 이력이 없는 고객 사용량 데이터 삭제

분석 방법

- 각각 19-20년, 20-21년, 21-22년 사용량을 코사인 유사도로 비교
- 상자그림(Box Plot)을 활용한 시각화, 위약 의심고객 유사도 한계점 결정



상자그림(Box Plot): 측정값들의 중심위치와 산포를 확인 할 수 있는 그림

- 사분위범위수(IQR) : $Q1 \sim Q3$

- 이상점(Outlier) : 울타리(Fence 바깥의 값)

*Fence = IQR \pm 1.5로 지정

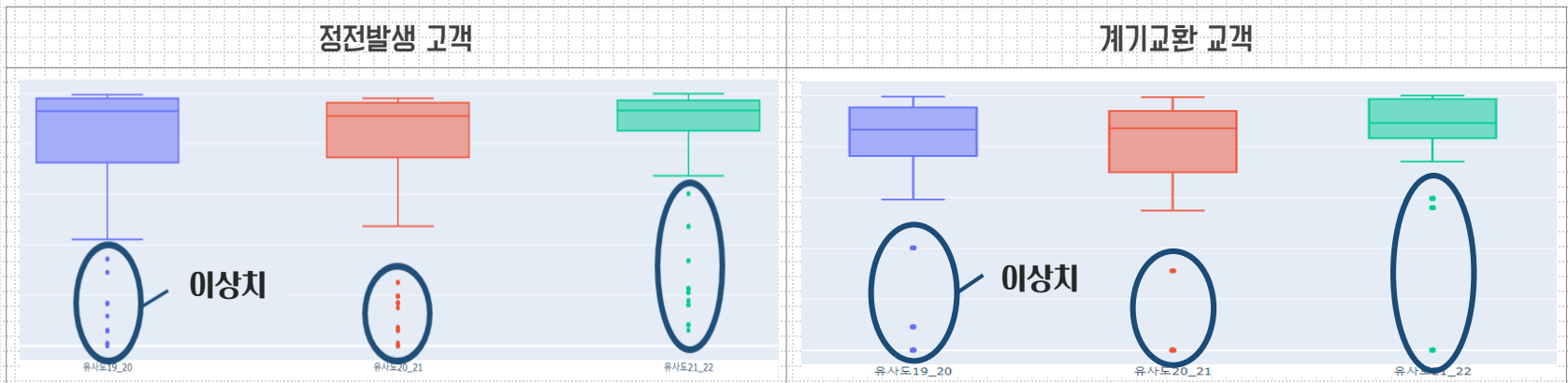
CHAPTER 03 후진내용 분포 유사도

코사인유사도 분석결과 상자분포를 통해 전체데이터 및 이상치 분포를 확인하고 이를 바탕으로 최종 위약의심고객으로 선정하기 위한 기준을 설정하였다

정전/계기교환 전후 유사도

분석 결과

- 상자그림 확인시 LP(IQRx1.5) 바깥 이상치 다수 확인됨
- 현장확인 등 추가 검증이 필요한 대상 기준 설정 : Q1(하위25%)



➔ 이상치 : 경고, Q1 이하 : 주의 → 위약 의심고객 사전 탐지로 활용

CHAPTER 03 **주진내용** 분포 유사도

3년간 정전 및 계기교환 데이터를 대상으로 검증을 시행해 위약의심고객을 추출하였으며 현장방문 여부를 최종 검토할 수 있는 시각화 분석도 함께 시행했다.

정전/계기교환 전후 유사도

검증 시행

- 2019~2022년(3년간) 제주본부 정전 및 계기교환 데이터를 입력하여 전후 데이터 유사도 비교분석 검증 시행
- (정전) 전체 4,996건 데이터 중 현장확인필요(위약의심-경고) 대상 162건 추출
- (계기교환) 전체 2,670건 데이터 중 현장확인필요(위약의심-경고) 대상 115건 추출

검증 결과

| | 개기년 | 정발발행년 | 출판정발사(간) | 유도수 19.20 | 유도수 20.21 | 유도수 21.22 | 19.20 | 20.21 | 21.22 |
|-------|-----------|---------------------|---------------|-------------|-------------|-------------|-------|-------|-------|
| 47459 | 27-250111 | 2020-09-02 19:47:18 | 590.174895389 | 0.683725918 | 0.681745968 | | | | |
| 47747 | 27-250111 | 2020-09-02 19:47:18 | 590.13333 | 0.578591029 | 0.647426663 | 0.697782826 | 주의 | | |
| 48024 | 29-250111 | 2020-09-02 19:47:18 | 590.13333 | 0.839464737 | 0.844052983 | 0.819722355 | | | |
| 48080 | 29-250111 | 2020-09-02 19:47:18 | 300.06667 | 0.839464737 | 0.844052983 | 0.819722355 | | | |
| 48279 | 27-250111 | 2020-09-02 19:47:19 | 590.1 | 0.610014405 | 0.620924216 | 0.685103152 | | | |
| 48355 | 29-250111 | 2020-09-02 19:47:19 | 590.11667 | 0.917654205 | 0.867436664 | 0.93204852 | | | 주의 |
| 49176 | 29-250111 | 2020-09-02 19:47:19 | 590.11667 | 0.803148297 | 0.786726088 | 0.885128204 | | | |
| 49180 | 29-250111 | 2021-04-17 16:22:06 | 300.33333 | 0.803148297 | 0.786726088 | 0.885128204 | | | |
| 49314 | 29-250111 | 2020-09-02 19:47:19 | 590.1 | 0.854987970 | 0.817750497 | 0.865125119 | | | |
| 49611 | 29-250111 | 2020-09-02 19:47:18 | 590.13333 | 0.729050855 | 0.578850509 | 0.765277019 | | | |
| 50524 | 29-250111 | 2021-06-21 09:55:57 | 34.43333 | 0.114591633 | 0.110375745 | 0.132861698 | 주의 | | 검교 |
| 50526 | 29-250111 | 2021-06-21 10:30:50 | 116.66667 | 0.114591633 | 0.110375745 | 0.132861698 | 주의 | | 검교 |
| 50579 | 29-250111 | 2020-12-05 09:16:21 | 78.11667 | 0.124303573 | 0.139369512 | 0.246055140 | 주의 | | 주의 |
| 50580 | 29-250111 | 2021-04-04 17:11:50 | 579.00000 | 0.124303573 | 0.139369512 | 0.246055140 | 주의 | | 주의 |
| 50588 | 29-250111 | 2019-06-27 17:11:50 | 57.36667 | 0.088762576 | 0.495162527 | 0.963152519 | 주의 | | |
| 50589 | 29-250111 | 2019-06-27 18:06:43 | 9608.83333 | 0.088762576 | 0.495162527 | 0.963152519 | 주의 | | |
| 50622 | 29-250111 | 2021-02-05 03:01:45 | 73.8 | 0.088762576 | 0.495162527 | 0.963152519 | | | |
| 50623 | 29-250111 | 2021-02-05 03:01:45 | 168.00000 | 0.088762576 | 0.495162527 | 0.963152519 | 검교 | | 검교 |
| 50699 | 29-250111 | 2019-06-10 11:12:00 | 114.33333 | 0.995924995 | 0.995924995 | 0.999290024 | | | |
| 50900 | 29-250111 | 2019-06-20 01:02:28 | 78.81667 | 0.995924995 | 0.995924995 | 0.999290024 | | | |
| 50901 | 29-250111 | 2020-10-08 01:01:28 | 65.78333 | 0.995924995 | 0.995924995 | 0.999290024 | | | |
| 50903 | 29-250111 | 2019-06-11 08:22:49 | 2969.20333 | 0.650142683 | 0.611359434 | 0.476780529 | | | |
| 50904 | 29-250111 | 2020-06-01 04:43:33 | 346.86667 | 0.650142683 | 0.611359434 | 0.476780529 | | | |
| 51022 | 29-250111 | 2019-12-19 10:02:29 | 37.78333 | 0.929162687 | 0.917704453 | 0.994362629 | | | |

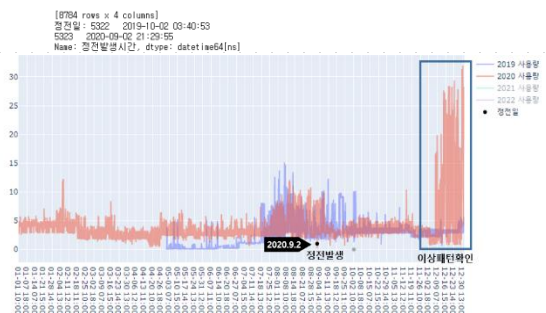
1.추출 의심고객 리스트 현장부서 제공

```
In [67]: num = input("제기번호를 입력하세요(제기번호는 '-' 있는 연대파, '-'로 바꿔서 입력해 주세요.)")
numDF = globa1('c:\exam\자료2\제기번호(-).forat(n=1)')
pr int(numDF)
그때 후기파가(numDF,num)
```

제기번호를 입력하세요(제기번호는 '-' 있는 연대파, '-'로 바꿔서 입력해 주세요.: 07_455055771
2019SI건발사용량 2020SI건발사용량 2021SI건발사용량 2022SI건발사용량

| | 실용검출시험 | 0.08 | 2.37 | 2.36 | 1.79 |
|---------|--------|------|------|------|------|
| 전사도 (%) | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 |
| 분리도 (%) | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 |
| 정확도 (%) | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 |
| 신뢰도 (%) | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 |

2. 방문 전 사용량 시각화자료 검토 (현장방문여부 최종 결정)



3. 현장확인

CHAPTER 03 후진내용 분석결과

월별전력량 예측 모델 : 정확도 74.97%

월별 전력량을 이용한 계약종별 예측 모델의 정확도는 74.97%로 위약 의심 고객의 1차 검토자료로 활용이 가능



한계점 및 개선방향

1. 최초학습데이터의 오류가능성

ID-CPM 알고리즘은 지도방식의 딥러닝모델이지만 최초 학습데이터가 정답이 아닐 가능성이 존재함

→ 비지도방식 딥러닝 알고리즘을 활용해 모델링 결과를 비교해 볼 수 있을 것

2. 대상데이터의 한계

예측모델 향상을 위해 계약종별 내 세부분류별 패턴을 나누어 확인할 수 있지만 제주본부내 고객데이터 규모가 충분하지 않아 구현이 어려움

→ 추후 전 지역본부 데이터를 대상으로 과제범위의 확장해 확인해 볼 수 있을 것

정전/계기교환 전후 사용량 유사도 분석

각 분석결과 중앙값 : (정전)91~93%
(계기교환)87~89%인 것으로 미루어
'정전 및 계기교환 전후의 사용량 패턴이 유사할 것'이라는 본과제의 가설설정이 적절했음 확인



한계점 및 개선방향

1. 계기교환 사유 미반영

계기교환 전후 사용패턴 차이가 다양한 계기교환사유가 존재(증설, 종변, 일부해지)하지만 위 모델은 구분해내지 못함

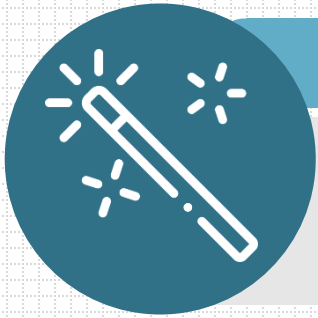
→ 계기교환 사유를 반영하여 명확한 사용패턴 변화가 예상되는 사유는 분석대상에서 제외하여 모델 개선

2. 이벤트 다양화

정전, 계기교환 이외 전후데이터 유사도 분석이 가능한 이벤트를 확인해 분석대상을 확대할 수 있음

CHAPTER 04 기대효과 / 향후계획

본과제 결과를 바탕으로 사용량이상고객을 조기 발견하고 조치하여 담당자 업무를 경감하고 업무착오를 방지하는 Fool-Proof를 확보할 수 있으며, 위약 전기요금 누수를 방지하는 효과를 기대할 수 있을 것으로 보인다



정성적 효과

- 사용량 이상고객 조기 발견·조치로 담당자 업무경감
- 당사 업무처리 착오 발생시 보완기능(Fool Proof) 확보



정량적 효과

- 사용량 이상고객 조기적출로 협정기간 단축 및 협정요금 최소화
- 위약 전기요금 누수 방지



향후계획 (고도화)

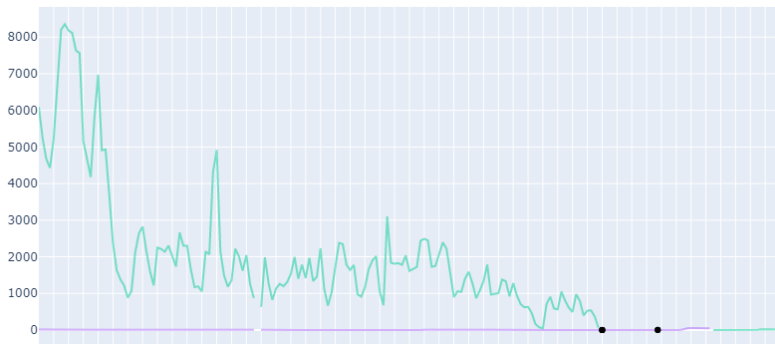
- 고압AMI시스템 내 " 사용확인 대상 고객 탐지 화면 " 구현
→ 고객별 일치율 산출, 일치율 % 기준 이하 디스플레이

CHAPTER 05 과계를 진행하며

과제수행을 통해 실제로 적발된 위약고객 2건에 대해서 현장검증과 사용량 협정을 통해 총 27백만원의 위약요금을 회수하였다

현장검증 결과

1. 배수입력 착오 요금 과소계산 고객 검출



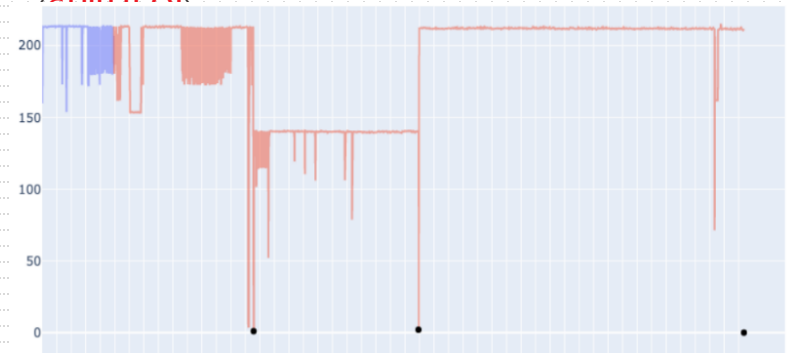
계기교체 고객 현장확인

필요대상 현장검증 시행 과정에서 발견 (휴지후 재사용 처리시 계기변수 미변동)

→ 정정조치 및 과소요금 청구

계기교체 고객 현장확인
필요대상 현장검증 시행 과정에서 발견
(휴지후 재사용 처리시 계기변수 미변동)
→ 정정조치 및 과소요금 청구

2. MOF 불량 요금 과소계산 고객 검출



정전발생 고객 데이터분석 과정에서 발견 (정전 발생시 MOF 파급고장 미인지)

→ 정정조치 및 과소요금 청구

정전발생 고객 데이터분석 과정에서 발견
(정전 발생시 MOF 파급고장 미인지)
→ 정정조치 및 과소요금 청구

시스템 개선 건의

→ 휴지업무 입력화면 기능 개선

- 휴먼 에러 예방을 위한 영배 4.0시스템 개선 건의
(업무담당) 배수 변동시 확인 필요 팝업창 신설기능 요청
(확인자) 공급방식, 계약전력기준 및 배수 상이시 팝업 체크화면 신설기능 요청



Q&A

감사합니다