

---

# SONG JUN HEE

Data Analyst

## CONTACT

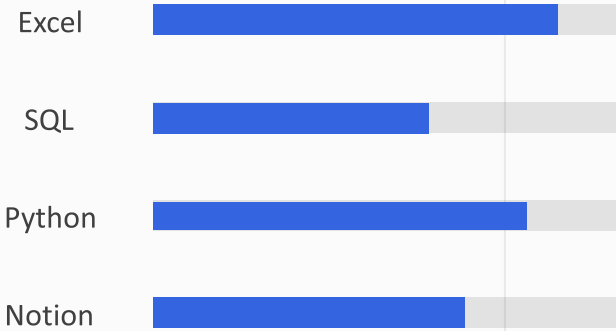
EMAIL : heej0903@Naver.com

TEL : 010 - 4936 - 2465

ADDRESS : 대전광역시 대덕구 신탄진동



### Skill



### Education

2015.02 포항고등학교 졸업  
2019.02 충남대학교 고분자공학과 졸업  
2022.12 포스코 청년 Ai BigData 아카데미 20기 수료  
2023.09 LG Aimers 3기 수료  
2024.07 제로베이스 데이터 분석 취업 스쿨 5기

### Experience

2019.03 ~ 2021.06 ROTC 육군 장교  
2023.01 ~ 2023.03 포항공대 인공지능 연구원  
2023.09 ~ 삼양패키징 대전1공장

### Certification

2023.04 SQLD  
2023.08 Six Sigma Green Belt  
2024.03 ADSP  
2024.03 KT AICE Associate

### Project(포스코 아카데미)

이차전지 화성공정 불량 최적화 팀 프로젝트  
미술관 안내 Chat Bot 제작 팀 프로젝트

### Hackathon(LG Aimers)

온라인 채널 제품 판매량 예측 해커톤

### Project(제로베이스)

은행 고객 Segmentation 개인 프로젝트  
보험 관심 고객 예측 개인 프로젝트  
신용카드 고객 데이터 분석 팀 프로젝트

### Awards

포스코 청년 Ai BigData 아카데미 프로젝트 우수상  
LG Aimers 3기 해커톤 최우수상 수상

프로젝트 제목을 클릭하시면, 해당 프로젝트 페이지로 이동합니다.

## Project(포스코 아카데미)

이차전지 화성공정 불량 최적화 팀 프로젝트 : 제조업 / 공정 이상탐지 / EDA, ML

## Hackathon(LG Aimers)

온라인 채널 제품 판매량 예측 해커톤 : 이커머스 / 시계열 예측 / EDA, DL

## Project(제로베이스)

보험 관심 고객 분석을 통한 응답률 개선 : 보험 / 군집 분석, 분류 / EDA ML

고객군별 소비 패턴 분석을 통한 전환 효과 분석 : 카드사 / 군집 분석, 분류 / EDA ML



# 01 이차전지 화성공정 불량 최적화 프로젝트.

## 주요 목표

이차전지 화성공정 불량률 증가 원인 파악 및 불량 감소를 위한 개선안 도출

## 세부 목표

- 1) 설비 간의 유의차 분석 결과를 기반으로 공정 작업 조건 유의차 도출
- 2) 공정 작업조건(온도, 압력 등)에 의한 영향도 분석
- 3) 공정 개선을 위한 최적 관리 조건 도출 및 관리 방안 수립

## 결과물

최적화 시뮬레이션 분석 결과 불량률 2.38%를 0.62%로 개선.

## 주요 결론

### 1. 공정 정보 기반의 설비 유의차 분석

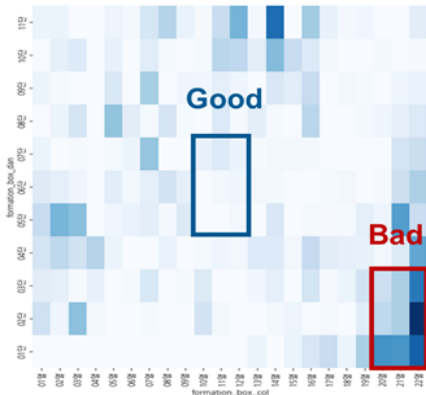
- 1) Aging 단계에서 구조물 가장자리 부근에서 불량률 증가.
- 2) 통계적 검정을 통한 설비 위치에 따른 불량 편차 유의성 확인.
- 3) 온도 변화의 요인 도출 : 구조물 하단의 대류 현상

### 2. 핵심인자 도출을 위한 모델링 분석

- 1) 온도가 불량 여부에 가장 중요한 영향을 미치는 것을 확인
- 2) Decision Tree 모델을 통해 구간별 최적 온도 조건 도출.
- 3) 최적 온도 조건 적용을 위한 열선 설치 개선안 제시.
- 4) 제약 조건 하 최적화 방안 제시

## 화성공정 핵심인자 도출 및 통계적 검증

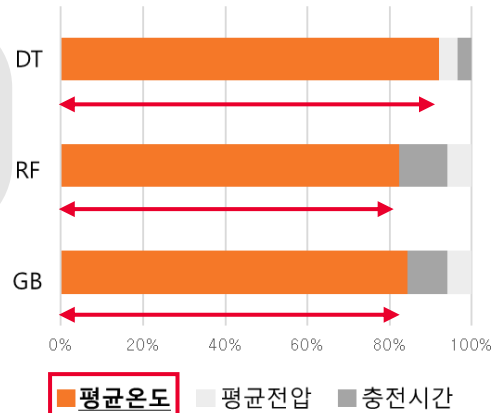
### 설비 위치별 영향인자 분석



Bad	Good
불량 cell 303개 불량률 17.26%	불량 cell 34개 불량률 1.81%

Formation 단계에서  
**가장자리(20열 ~ 22열)**  
불량률이 증가

3 가지 변수  
(평균 온도, 평균 전압, 충전 시간)  
중 **평균 온도**가 영향도 ↑



이를 기반으로 불량 판정 인자 도출

※ 서로 다른 모델에서 영향도 분석을 진행하여 변수 중요도 정확성 향상

### 통계적 가설검정

가설 1 : 설비의 열에 따라 **불량 발생 빈도** 차이가 있음

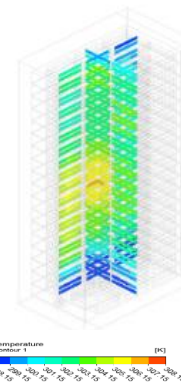
Chi-squared 검정 결과,  $p\text{-value} < 0.001$

**설비의 열에 따른 불량 발생 빈도 차이가 유의미함.**

가설 2 : 설비의 열에 따라 **불량률**에 차이가 있음

ANOVA 검정 결과,  $p\text{-value} < 0.001$

**설비의 열에 따른 불량률 차이가 유의미함.**



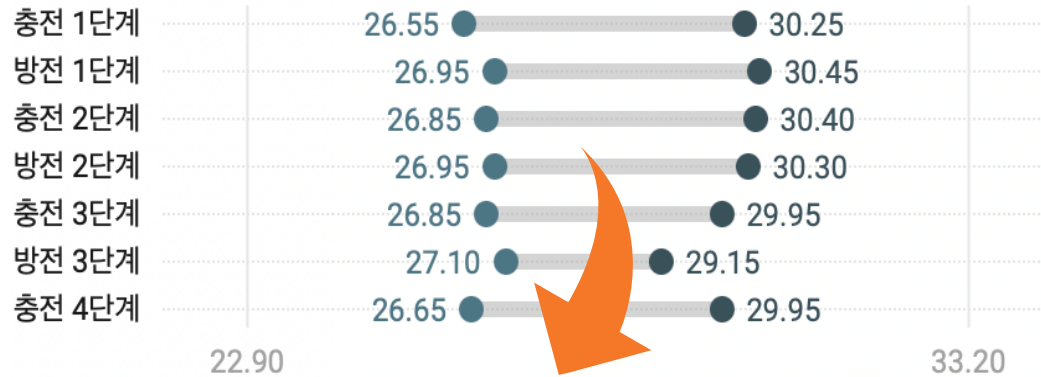
구조물 하단에 배기구가 위치하여 대류에 방해를 받아, 온도 변화에 민감한 불량의 원인으로 추정됨

[참고논문] 강윤희, 이준현 and 이진경. (2020). ESS용 Rack 내부의 공기유동에 따른 배터리 열관리에 관한 해석적 연구. 동력시스템공학회지, 24(2), 37-44.

[참고논문] 장경민 and 김광선. (2017). 급격한 온도 변화에 따른 리튬 이온 배터리의 전해질 내 염 농도 분포 특성. 반도체디스플레이기술학회지, 16(1), 11-15.

## Formation 공정 최적 조건 도출 및 시뮬레이션 분석

### 각 단계별 최적 온도 조건

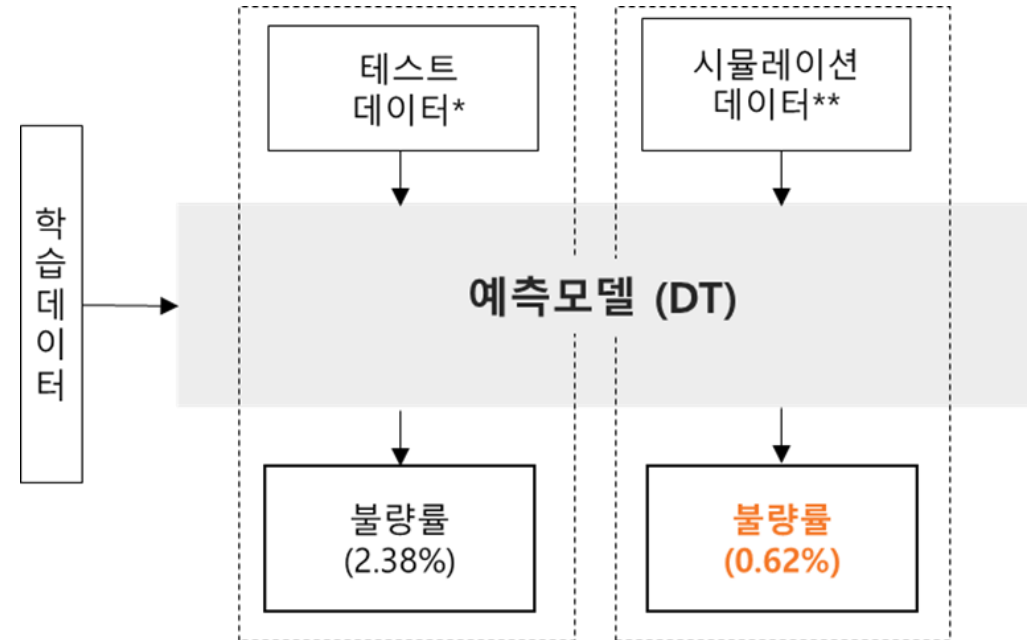


**최적 온도 조건 도출**

### 온도 구간 별 불량률 비교

Formation 충전 1단계			
온도 조건	26.55°C 이하	26.55 ~ 30.25°C	30.25°C 이상
불량률	8.82% (683/7,741)	1.59% (532/33,290)	9.15% (86/940)

### 시뮬레이션 분석 결과



**1.76%p 감소**

\* 테스트 데이터 : 기존 온도 적용

\*\* 시뮬레이션 데이터 : 도출한 최적 온도를 적용

## 개선 효과 분석

### 설비 위치별 온도 최적화 시, 품질 기회비용 측정



온도 조건을 최적화 시 불량률이 2.0%p 감소 추정  
품질 기회비용 15.8 억 원 절감 효과 예상

- 1) 기존 불량수 36,753개 - 시뮬레이션 효과 적용 시 불량수 5,751개  
2) 배터리 1kWh당 가격 = 150달러, cell 1개 용량 = 0.5kWh, 1달러 = 1,300원

### 제약 조건 하 최적화를 통한 불량 최소화

#### 공정 순서

Formation		Power Charging	
온도(°C)	불량률(%)	온도(°C)	불량률(%)
6단 28.60	0.81	9열 27.65	2.17
5단 28.08	0.01	8열 27.78	1.55
11열		7열 27.30	2.36
:		:	
3단 25.00	19.00	3열 26.17	4.12
2단 25.36	26.40	2열 25.75	3.40
1단 25.24	21.40	1열 26.00	6.21
22열		1단	

Power Charging 공정에서  
Rack 재배치를 통한  
불량 감소

#### 설비 위치에 따른 불량률 감소

	최적 조건이 아닌 Formation	Power Charging에서 온도 조절
공정 조건	27.10 ~ 29.00 °C 이외의 범위	27.10 ~ 30.25°C
불량률	6.51% (969/14,879)	3.11% (462/14,879)



# 02 온라인 판매 채널 제품 판매량 예측.

## 주요 목표

온라인 쇼핑몰 일별 제품별 판매량을 예측하는 AI 모델

## 세부 목표

- 1) 온라인 쇼핑몰 제품 판매 특성을 분석
- 2) 시계열 데이터 예측 모델 구현
- 3) 모델 입력 구조 및 학습 방법 개선을 통한 성능 향상

## 결과물

매출 상위 80% 제품 기준 예측 정확도 88% 모델 구현

## 주요 결론

### 1. 온라인 쇼핑몰 제품 판매 특성 확인

- 1) 파레토 분석 : 판매량 상위 20%의 제품이 전체 매출의 80% 차지
- 2) 시계열 패턴 분석 : 제품군별 개별적인 구매 주기 파악

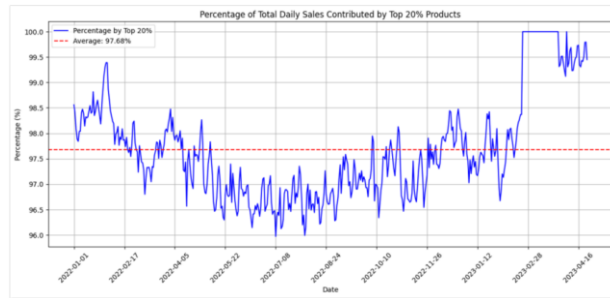
### 2. 모델 알고리즘 개선 및 학습 구조 변경

- 1) 시점별로 판매량과 제품 특성(파레토, 시계열, 카테고리 등)이 학습되도록 다변량 LSTM 구조 설계
- 2) 판매량이 높은 제품군에 집중할 수 있도록 RMSE 방식의 Loss Function에 일별 판매량 가중치를 부여



## 시장의 수요·공급 원리 기반의 제품 판매 분석

### 일별 / 월별 제품 판매 현황 분석



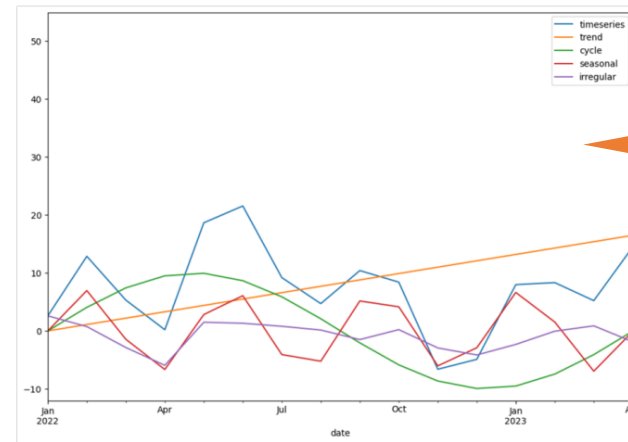
판매량 상위 20% 제품  
매출 비율 평균 79.10%

\* 파레토 법칙 : 전체 매출의 80%는 20%의 제품에서 발생  
\* 2022년 01월 01일부터 2023년 04월 24일까지 상위 20% 제품의 매출 비율(일별) 그래프

- 0041 : 유아용 기저귀
- 0003 : 유산균
- 0025 : 물티슈
- 0004 : 단백질보충제
- 0001 : 건강기능식품
- 0022 : 치약

월별 판매량 상위 Top 6 제품 종류 확인

### 판매 주기성 분석



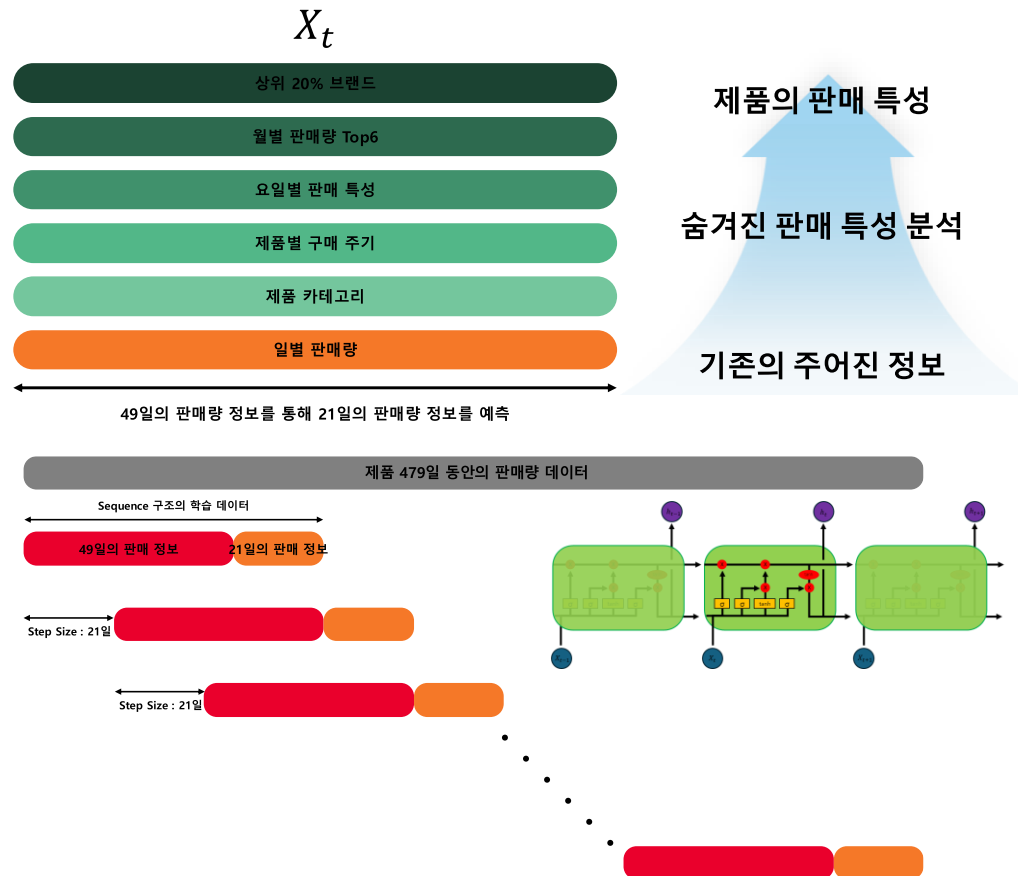
계절성과 불규칙성이  
전체 시계열에 영향을  
미침

- 3개월 : 제지/위생용품
- 4개월 : 욕실용품
- 5개월 : 유아 생활용품, 유아식품, 주방/청소/세탁세제 등
- 1년 : 탈취/방충/살충/제습/방향

고객들은 제품별로 개별적인 구매 주기를 가짐

## 시계열 예측 모델 구조 설계

### 다변량 LSTM 모델 구조 설계



### Loss Function 개선

$$1 - \frac{1}{n} \sum_{day=1}^n \sum_{i=1}^N \left( \frac{|y_i^{day} - p_i^{day}|}{\max(y_i^{day}, p_i^{day})} \right) \times \frac{y_i^{day}}{\sum_{i=1}^N y_i^{day}}$$

[오차]                      [판매비중]

$i$  : 제품 index

$y_i^{day}$  : i번째 제품의 day일의 판매량

$p_i^{day}$  : i번째 제품의 day일의 예측량

일별 제품 판매 비중을 추가한 MSE 기반의 Loss Function 구현

판매량 상위 20% 제품에 대하여 예측 정확도 88% 달성



# 03 보험 관심 고객 분석을 통한 응답률 개선

## 주요 목표

보험 상품에 관심을 보일 가능성이 높은 고객 식별을 위한 분류 모델을 구축

## 세부 목표

- 1) 고객 데이터 분석을 통한 관심도 지표 생성
- 2) 오버 샘플링 처리를 통한 데이터 Imbalance 해결
- 3) 고객 응답률 상승을 위한 개선 전략 도출

## 결과물

고객에게 접근하는 채널 변경을 통한 응답 비율 개선(11.14%에서 19.14% 개선)

## 주요 결론

### 1. 관심도 지표 생성

- 1) 응답률과 높은 상관관계를 나타내는 변수 선정
- 2) 특성 별 응답 비율을 기반으로 관심도 지표 계산

### 2. 판매 채널별 평균 응답률 분석

- 1) 채널별 응답현황 분석
- 2) 152번 채널의 경우, 가장 많은 고객에게 접근했지만, 응답률이 매우 낮음(평균 응답률 0.17 대비 0.02)

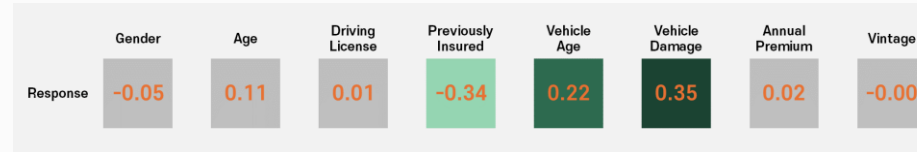
### 3. 분류 모델 구축 및 A/B Test

- 1) SMOTE 처리 이후, LightGBM 기반의 분류 모델 구축
- 2) 응답률이 높은 채널로 이동할 경우, 전환 효과 분석

## 사용한 데이터



## Response와 변수 간의 상관 관계 분석

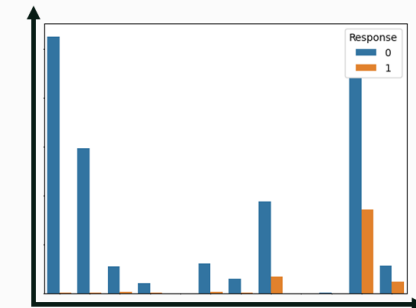


## 관심도 지표 계산

$$\text{Interest Index} = \text{Previously Insured} + \text{Vehicle Age} + \text{Vehicle Damage}$$

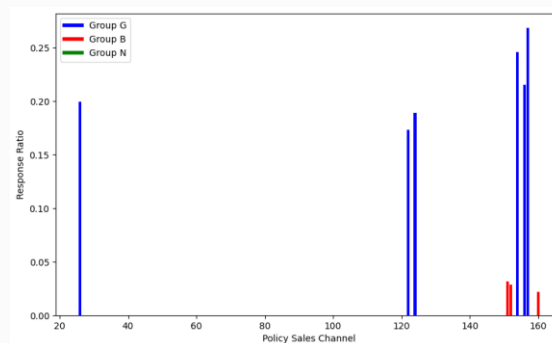
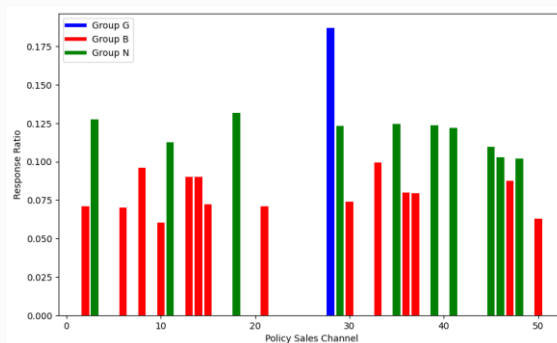
※ 특성별 경우에 따라(Yes, No, 1-2 Year 등) Response가 1인 비율을 계산

※ 각각의 특성별 비율의 합을 구하여 관심도 지표 계산



Interest Index 지표가 증가할 수록,  
Response가 1인 비율이 증가

## 지역별, 채널별 Response 분석



※ 응답률이 좋은 고객 Segment 발굴 : 전체 고객 대상 평균 긍정 응답비율(12%) 기준

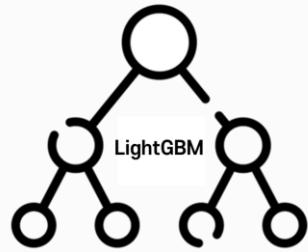
- 14% 이상인 경우 응답률 양호한 그룹(G)
- 10% 이하인 경우 응답률이 저조한 그룹(B)
- 나머지 그룹(N)

※ 전체 응답 횟수 대비 지역별(채널별) 응답 횟수가 1% 미만인 경우

표본 크기가 부족하여 대표성이 부족하므로 분석에서 제외

1. 지역별 응답현황 분석 결과, 평균 응답률(0.17)보다 높은 지역이 하나의 지역이 있음.
2. 채널별 응답현황 분석 결과, 평균 응답률(0.17)보다 높은 채널이 존재함.(0.26 ~ 0.24)
3. 152번 채널의 경우, 가장 많은 고객들에게 접근했지만 응답률은 매우 낮음(0.02)

## 응답 여부 분류 모델

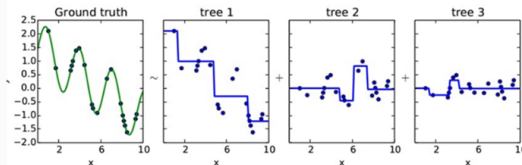


고객 응답 분류 모델

Accuracy : 0.84    Recall : 0.92

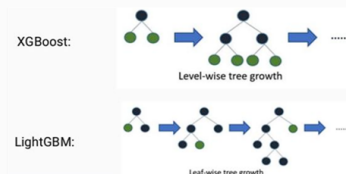
### 1. GBM(Gradient Boosting Machine)

이전 트리에서 예측하고 남은 Loss를 다음 트리들이 차례로 학습하면서  
Loss를 줄여가는 Boosting 알고리즘



### 2. Leaf - Wise 트리 분할

Max Delta Loss를 가지는 리프 노드로 분할시키는 방법  
주어진 데이터의 크기가 충분(10,000개 이상)하다면 높은 정확도를 보장



### 3. 내장된 범주형 변수 지원 기능으로 별도의 인코딩 없이도 범주형 변수 처리

## Rule Base 기반 응답률 개선 시뮬레이션 분석

Rule Base 기반의 응답률 개선

초기 평균 응답률 12.26%

이전에 보험을 가입 하지 않은 고객

20~30대 남성 고객    차량 손상이 있는 고객

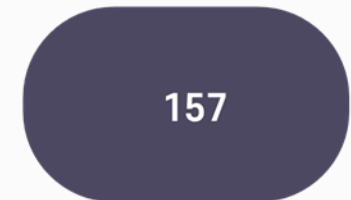
개선 평균 응답률 17.21%

## 채널 변경을 통한 응답률 개선 시뮬레이션 분석

응답률이 낮은 채널    12.26%



응답률이 높은 채널    19.15%



※ 기존 채널(151, 152, 160)과 변경한 채널(157) 간의 긍정  
응답 비율의 차이를 분석한 결과(t-test) p-Value 0.5이하의  
유의미한 결과를 확인



# 04 고객군별 소비 패턴 분석을 통한 전환 분석

## 주요 목표

고객 군집별 소비 패턴 파악을 통한 이탈 고객 전환

## 세부 목표

- 1) 신용카드 가입 기간, 사용 금액 등을 활용한 RFM 지표 선정
- 2) 계층적 군집 분석을 통한 고객 Segmentation 및 소비 패턴 파악
- 3) 소비 패턴 전환에 따른 이탈 감소 효과 분석

## 결과물

기존 1627명 대상으로, 군집별 소비패턴 적용을 통해 이탈 고객을 54명으로 감소 (16.06% → 0.53%)

## 주요 결론

### 1. RFM 지표 선정

- 1) 비활성화 월 수, 가입 유지기간, 총 거래 금액(횟수)를 활용하여 RFM 지표 계산

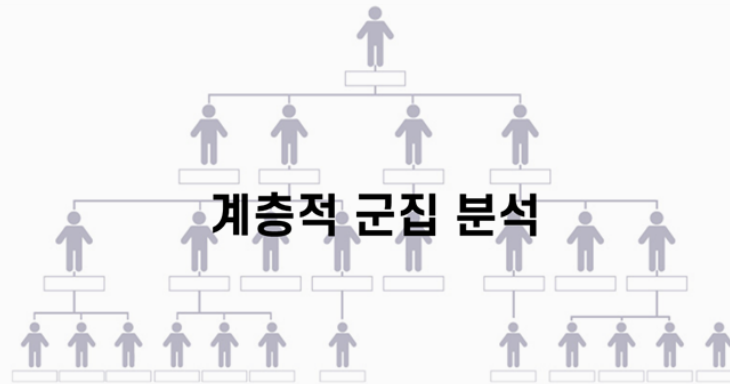
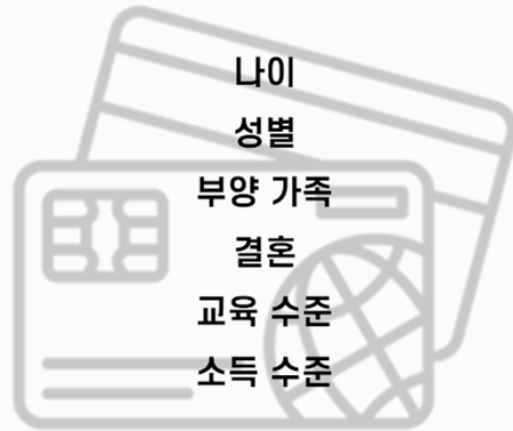
### 2. 계층적 군집 분석 및 t-Test 검정을 통한 이탈 고객과의 차이점 분석

- 1) 인구통계학적 데이터 기반의 계층적 군집 분석.
- 2) 각 군집별로 소비 특성 데이터 기반의 세부 군집 분석
- 3) 이탈한 고객과 소비 군집 간의 유의차 분석(t-test 검정)

### 3. 이탈 고객 분류 모델을 활용한 전환 효과 분석(A/B Test)

- 1) 인구통계학적 군집별로 이탈 고객 분리
- 2) 소비 특성 군집의 평균값으로 이탈 고객의 소비 특성 전환
- 3) Randomforest 모델 기반의 분류 모델로 A/B Test 진행

## 분석 Flow



신용한도  
신용한도 대비 사용한 한도의 비율  
1Q 대비 4Q 총 금액 변동 비율  
1Q 대비 4Q 총 거래 횟수 변동 비율  
RFM Score  
(비활성화 월 수, 계정유지기간, 총 거래 금액, 총 거래 횟수)



인구통계학 기반의 군집 내에

소비 특성 기반의 세부 군집 분석



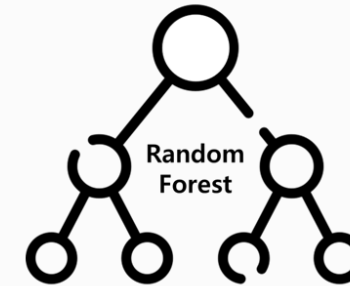
이탈 고객 대상으로  
소비 패턴 개선을 통한  
유지 고객으로 전환

## 모델링 분석

신용한도  
신용한도 대비 사용한 한도의 비율  
1Q 대비 4Q 총 금액 변동 비율  
1Q 대비 4Q 총 거래 횟수 변동 비율  
RFM Score  
(비활성화 월 수, 계정유지기간, 총 거래 금액, 총 거래 횟수)

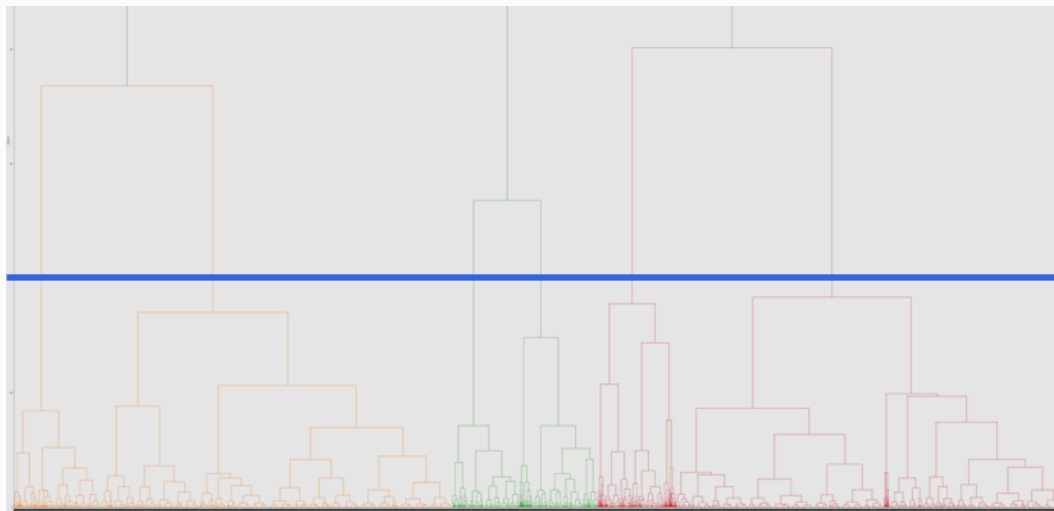


나이  
성별  
부양 가족  
결혼  
교육 수준  
소득 수준



이탈 고객 분류 모델

Accuracy : 0.90 Recall(이탈 고객) : 0.72



※ 고객의 정보를 알 수 없는 경우, 군집 분석에서 제외함.

※ Distance 30 기준, 군집 분석으로 6개의 고객 Group 도출

기존 이탈고객

Group	All	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
%	16.06	16.53	17.07	14.91	19.18	17.22	14.40
명	1627	140	576	96	145	129	541

인구통계학적 Group 별 이탈 고객

Group	All	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
%	3.37	0.24	1.70	4.30	5.20	3.70	5.00
명	342	2	57	28	39	28	188

※ 이탈 군집의 소비 특성을 세부 군집 소비 특성으로 전환 시, 전환 효과 분석

※ 세부 군집과 이탈 군집의 독립성 검정 결과, 소비 특성 6개에서 모두 p-Value 0.5 이하



---

# THANK YOU

## CONTACT

EMAIL : [heesj0903@Naver.com](mailto:heesj0903@Naver.com)

TEL : 010 - 4936 - 2465

ADDRESS : 대전광역시 대덕구 신탄진동