

# [SCM팀]결과 정리

- 분석 목표
- 활용 데이터
- 분석 Workflow
- 주요 결과
  - 1. 학습 단계
    - 1) 상관관계 분석
    - 2) 모델 학습 결과
    - 3) 변수 영향도 분석 (마감시 결품 SKU수 기준)
  - 2. 검증 단계
  - 3. 타겟 선정 단계
  - 1) 가상 판매량 시뮬레이션 (세부 가설: 군집내 SKU 결품률이 증가하면 가상 판매량이 줄어드는 것이다.)
  - 2) 카테고리별 분석 결과
    - (최종 결론 도출)
  - 3) 결품 대체 가능 카테고리 VS 결품 대체 불가 카테고리 비교 (고객 관점에서 원인 파악)
  - \*참고 페이지 & 파일
- APPENDIX
  - 사용 모델
    - 1. Light GBM 이란?
    - 2. Light GBM을 사용한 이유
  - 세부 분석 결과
    - 1) 변수간 상호 작용 분석 및 시각화
      - (주요 변수 해석) 예측 주문수, S,A등급 상품수, 할인율, 마감할인 행사등은 판매량 상승에 긍정적인 영향을 주나, 가격 상승, 거리제한 손실 주문수 증가등은 판매량 상승에 부정적인 영향을 줌
    - 2) 카테고리 별 세부 분석 - 결품 대체 불가 카테고리(RTE샐러드)와 결품 대체 가능 카테고리(두부부침용) 비교
      - (주요 결론) 카테고리에서 결품률에 따른 FC의 판매량 감소 추세는 대체적으로 동일하나 FC 별로 세부적으로 본다면, 판매량 추이에 대한 경향성은 다를 수 있다

## 분석 목표

- 1차 분석 후 나온 주요 피드백 사항 반영
  - 1) FC 정보가 포함된 데이터의 상세 분석 필요
  - 2) 판매량에 영향을 주는 다른 Feature 도 고려할 것
    - 주문, 노출 여부 등
    - 친환경 제품, 부위별 제품 등 → 데이터 추가 수집 필요
  - 3) 카테고리 별로 세부 분석 진행
  - 4) 결품시 다른 상품으로 대체가 이루어지고 있다는 가설을 검증하기 위해 세부 가설 정립 및 검증
- Feature store 구축(판매량에 영향을 줄 수 있는 변수들 수집 및 전처리 후 하나의 통합된 Dataset 구축)
- 다수의 Feature를 활용하여 판매량 예측 후 관련 SKU 결품률의 상대적 영향도 파악
- 채소 depth5 카테고리 중 군집 결품을 적용 가능 대상 카테고리라 적용 불가 카테고리를 선정(가상의 판매량을 통해 비교 검증)

## 활용 데이터

- 1) 대상 품목: 채소 depth5 카테고리
- 2) 대상 기간: 2022-10-01 ~ 2023-02-13(결품률이 안정되기 시작한 시점인 10월 이후 데이터 활용)

3) 활용 변수: 상품가격, 할인, 재고, 실적, 행사, 상품등급, 배너 유입, 거리제한등

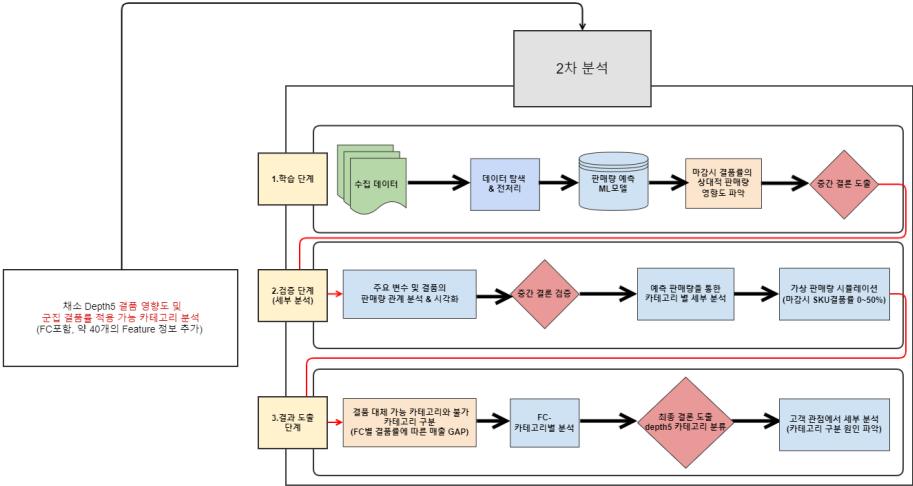
index	데이터 유형	컬럼명	컬럼 의미	데이터타입	비고
1	기본정보	date_cd	일자	datetime	
2		shop_name	FC 점포명	varchar	
3		category2	depth2 카테고리	varchar	
4		category3	depth3 카테고리	varchar	
5		category4	depth4 카테고리	varchar	
6		category5	depth5 카테고리	varchar	경상가 - 할인가
7	가격, 할인	sale_price_avg	판매가격 평균	float	5 depth 카테고리내 가격의 평균
8		sale_price_std	판매가격 표준편차	float	5 depth 카테고리내 가격의 표준 편차
9		discount_rate_avg	할인율 평균	float	5 depth 카테고리내 할인율의 평균
10		discount_rate_std	할인율 표준편차	float	5 depth 카테고리내 할인율의 표준편차
11	실적	sale_qty	매출수량	integer	
12		gpv	매출금액	integer	
13		order_cnt	주문건수	integer	단일 주문건수를 계산
14	재고(결품, 폐기등)	lack_sum_17	군집내 SKU결품수(17시)	integer	
15		lack_sum_20	군집내 SKU결품수(20시)	integer	
16		lack_sum_22	군집내 SKU결품수(22시)	integer	
17		lack_sum_24	군집내 SKU결품수(24시)	integer	
18		no_refersh_discard_cnt	선도저하 폐기수량	integer	scm 기준으로 선도저하, 유통기한만 고려함
19		expired_discard_cnt	유통기한 폐기수량	integer	
20	행사	lightning_sale	번쩍할인 행사 상품수	integer	5 depth 카테고리내 해당 행사를 진행한 상품수
21		exhibition	기획전 행사 상품수	integer	
22		search	검색 상품수	integer	
23		closeout_sale	마감할인 상품수	integer	
24		recommendation	추천 상품수	integer	
25	상품 등급	grade_s_cnt	S등급 상품수	integer	5 depth 카테고리내 S등급의 상품 수
26		grade_a_cnt	A등급 상품수	integer	5 depth 카테고리내 A등급의 상품 수
27		grade_b_cnt	B등급 상품수	integer	5 depth 카테고리내 B등급의 상품 수
28		grade_c_cnt	C등급 상품수	integer	5 depth 카테고리내 C등급의 상품 수
29		grade_d_cnt	D등급 상품수	integer	5 depth 카테고리내 D등급의 상품 수
30		grade_n_cnt	C등급 상품수	integer	5 depth 카테고리내 N등급의 상품 수
31		grade_z_cnt	D등급 상품수	integer	5 depth 카테고리내 Z등급의 상품 수
32	배너 유입	baemin_uv	메인 배너 유입객수	integer	일자- 점포별 - 유입객수(상품 정보 없음)

33	거리 제한	minus_ord	거리제한 주문수	float	
34		exp_ord	예상 주문수	float	거리제한 분을 반영한 예상되는 주문수를 뜻함
35		all_duration_min	거리제한 분 비율	float	
32		all_cnt	실제 주문수	integer	

\*거리 제한 & 배너 유입의 경우 B마트사업팀 선행 분석가 분들께서 피드백 주신 사항으로 반영

4) 타겟 변수: 판매량

분석 Workflow



주요 결과

1. 학습 단계

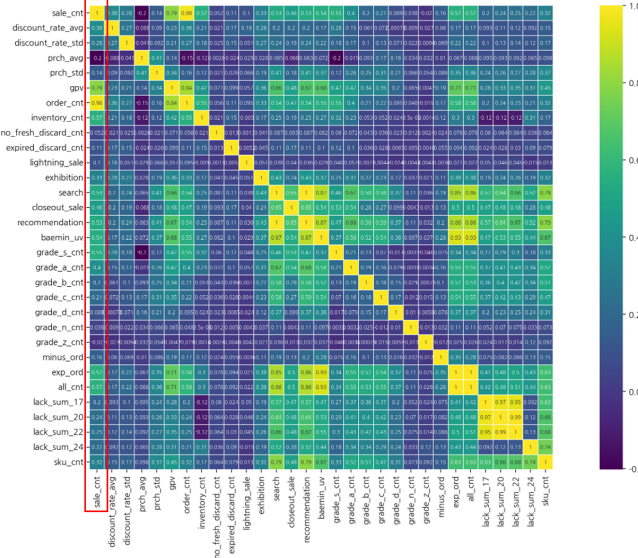
1)상관관계 분석

- (판매량 상관 관계 분석) 카테고리내에서 S, A 등급 SKU가 많을 수록, 마감할인 SKU가 많을 수록 판매량에 긍정적인 영향을 줌
- (결품 상관 관계 분석) 결품과 관련된 변수는 판매량 자체에 큰 영향을 주지 않음\_ 상관계수 0.2 (다만, 아래의 세부 분석을 통해 검증해볼 예정입니다.)

변수명	의미	상관계수	상관관계정도
sale_cnt	판매수량	1.000000	상
discount_rate_cnt	할인율	0.060174	상
gap	재고감액	0.708174	상
all_cnt	총 주문수(거래제한)	0.567153	상
inventory_cnt	재고수량	0.566671	상
exp_cnt	예약정산수(거래제한 순 일반환)	0.560344	상
grade_a_cnt	카테고리내 상품 SKU수	0.553889	상
baemin_uv	배민유입량(SKU수)	0.541144	상
recommendation	카테고리내 추천 SKU수	0.550422	상
search	카테고리내 검색 SKU수	0.527035	상
closeout_sale	카테고리내 마감할인 SKU수	0.464436	상
grade_n_cnt	카테고리내 상품 SKU수	0.393203	상
discount_rate_avg	할인율 평균	0.376862	상
exhibition	카테고리내 기획전 SKU수	0.320126	상
sku_cnt	카테고리내 상품수(전체)	0.321795	상
discount_rate_std	할인율 분산	0.28475	상
lack_sum_20	카테고리내 품절 SKU수(20%)	0.233365	상
lack_sum_20	카테고리내 품절 SKU수(20%)	0.236675	상
lack_sum_24	카테고리내 품절 SKU수(24%)	0.219641	상
grade_c_cnt	카테고리내 C 등급 SKU수(24%)	0.214654	상
grade_b_cnt	카테고리내 B 등급 SKU수(24%)	0.203037	상
lack_sum_17	카테고리내 품절 SKU수(17%)	0.196484	상
minus_cnt	음의 주문수	0.191993	상
prch_std	재입단가 분산	0.141624	상
expired_discard_cnt	상품기회 놓친 횟수	0.111246	상
lightning_sale	카테고리내 번쩍할인 SKU수	0.100106	상
grade_d_cnt	카테고리내 D 등급 SKU수(24%)	0.088893	중
no_fresh_discard_cnt	신도착품 놓친 횟수	0.054991	중
grade_n_cnt	카테고리내 상품 SKU수(24%)	0.037636	중
grade_z_cnt	카테고리내 상품 SKU수(24%)	-0.01983	중
prch_avg	재입단가 평균	-0.20338	중

\*판매수량과 변수들간의 상관관계

판매량과 주요 변수들의 상관관계



## 2) 모델 학습 결과

카테고리명	평가기준	지표값
depth5 채소	MAE	2.06
	MAPE	40.95

MAE: 2.06, 실제 판매량과 예측 판매량의 오차가 약 2개 정도 난다는 뜻으로 학습이 잘되었다고 볼 수 있음

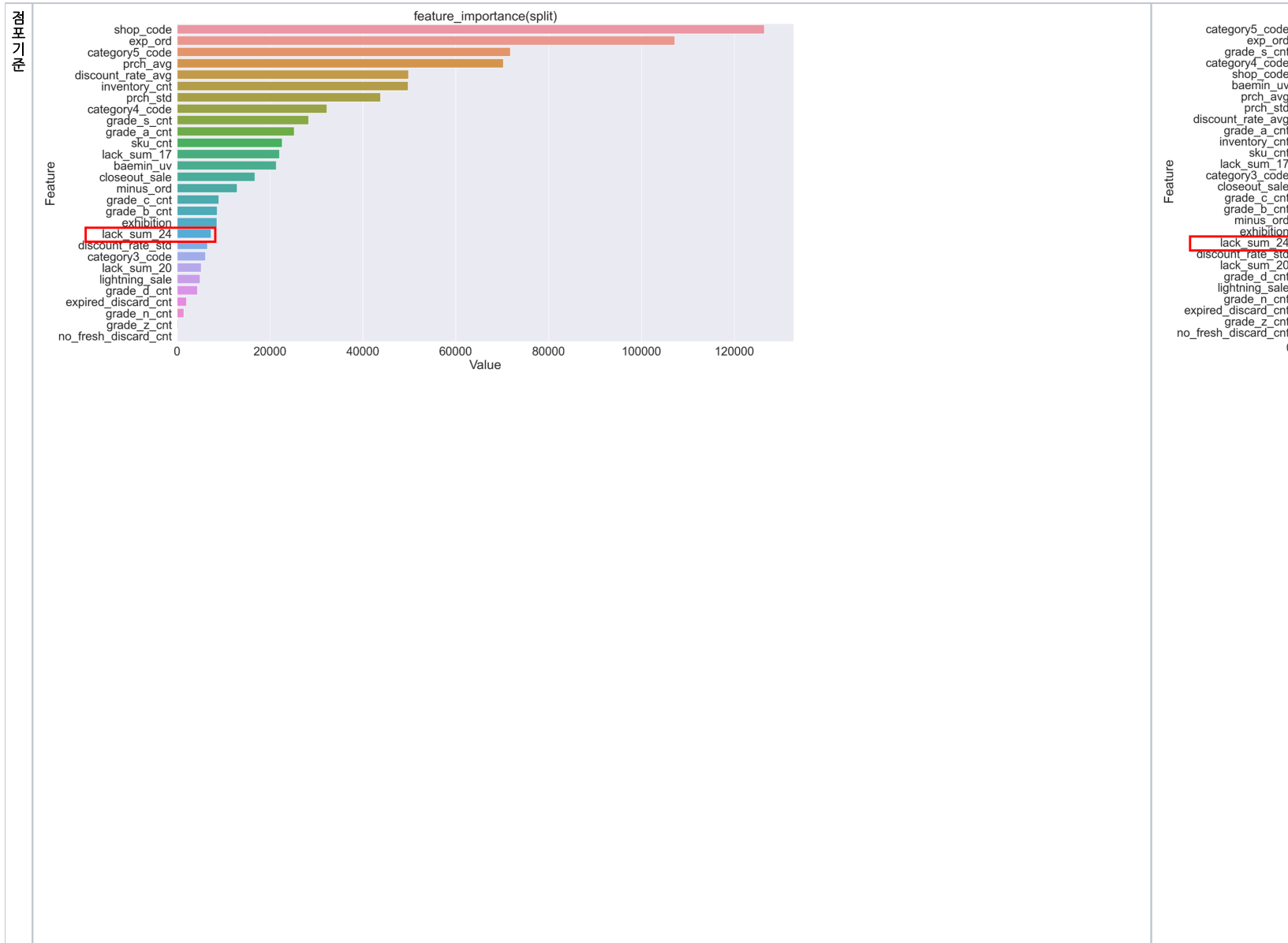
MAPE: 40.95, 실제 판매량과 예측 판매량의 오차가 실제 판매량 대비 40% 정도 차이 하고 있다는 뜻으로, 큰 차이가 나지 않는다고 해석할 수 있음(이마트에서는 70-80% 정도 MAPE 성능을 보여주었음)

(학습결과 요약) 두 지표를 종합하면 향후 결과들이 유의미하다고 판단 수 있을 정도로 학습이 잘 이루어졌다고 볼 수 있음 (\*\*값 자체가 낮을 수록 학습이 잘되었다는 설명력을 지니지만, 절대적인 기준이 없음을 참고 부탁드립니다.)

## 3)변수 영향도 분석 (마감시 결품 SKU수 기준)

구분	Feature importance method			중요도 평균
	변수 중요도 방법1	변수 중요도 방법2	변수 중요도 방법3	
중요도 순위	18/28(상위 65%)	19/28(상위 68%)	20/28(상위 71%)	상위 68%

구분	Feature Importance(split) → 데이터를 구분짓는데 사용하게된 변수의 영향도	



\*feature importance 값이 높을 수록 모델의 판매량 예측에 영향력을 준 변수로 볼 수 있음

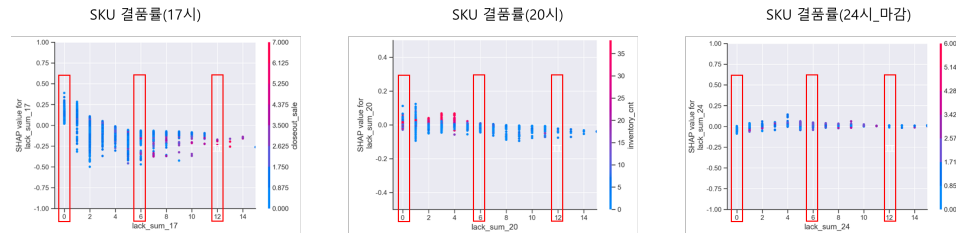
\*shap XAI 방법은 해석 방법 (색상: Feature 값의 크기, 길이: 예측 판매량의 크기)

- exp\_ord(예측 주문수): 예측 주문수 값이 높을 수록 (빨간색에 근접) 판매량 증가에 긍정적인 영향을 주며, 예측 주문수가 낮을 수록 (파란색에 근접) 판매량에 부정적인 영향을 줄 수 있다.
- prch\_avg(매입 단가 평균): 매입 단가, 즉 가격이 낮을 수록 (파란색에 근접) 판매량 증가에 긍정적인 영향을 주며, 가격이 높을 수록 (빨간색에 근접) 판매량에 부정적인 영향을 줄 수 있다.
- (마감시 SKU 결품률) 판매량 예측에서 해당 변수가 큰영향을 주지 않음 (상위 68%)
- (17시 기준 SKU 결품) 마감시 결품 SKU수 대비 변수가 영향도가 높으나 그렇게 높은 영향을 주지 않음 (상위 46%)
- (중요도 기준 Top5 변수) 1.점포 특성 2.카테고리 특성 3.거리제한이 포함된 예측 주문수 4.카테고리내 S,A등급 상품수 5.상품 가격

(중간결론 도출) 결품여부, 결품을 자체에 대한 변수의 영향도는 타 변수 대비 상대적으로 미비하다.

## 2. 검증 단계

- SKU 결품률에 따른 판매량 증감 분석



변수구분	결품률 0% 시	결품률 50% 시	결품률 100% 시	50%시 판매량 감소 비율(추정)	100%시 판매 량 감소 비율(추정)
SKU결품률(17시)	0.45	-0.05	-0.1	6.73%	7.40%
SKU결품률(20시)	0.15	0.01	0.01	1.88%	1.88%
SKU결품률(24시)	0	0.01	0	-0.13%	0.00%

- \*판매량 감소 비율은 결품을 증가로 인해 감소되는 판매량의 비율을 뜻함
- (17시 SKU결품률) 군집내 SKU 결품률이 증가함에 따라 판매량이 감소하는 추세를 보임 (약 7% 전체 판매량 감소)
- (20시 SKU결품률) 군집내 SKU 결품률이 증가함에 따라 판매량이 소폭 감소하는 추세를 보임 (약 2% 전체 판매량 감소)
- (24시SKU결품률) 군집내 결품 SKU수 증감에 상관없이 판매량의 큰 영향도는 없음 (판매량 감소 없음)

(중간결론 검증) 마감 기준 SKU 결품률은 판매수량에 큰 영향을 주지 않으나, 일부 영향을 주는 카테고리가 있을 것이므로 카테고리별 세부 분석을 진행하도록 한다.

## 3. 타겟 선정 단계

### 1) 가상 판매량 시뮬레이션(세부 가설: 군집내 SKU 결품률이 증가하면 가상 판매량이 줄어들 것이다.)

- 모든 상황 통제(다른 변수는 모두 동일)하에 마감시 SKU결품 임의로 변화를 주어 결품률에 따른 판매량의 추이 확인 (SKU 마감 결품률기준: 0%, 10%, 20%, 30%, 40%, 50%)
- 해당 시뮬레이션의 장점
  - 실제 데이터에서 결품 SKU수를 제외하고 다른 변수들이 모두 동일한 값을 가질 수 있는 경우는 현실적으로 존재하지 않음

- 해당 시뮬레이션에서는 순수하게 마감시 결과에 따른 영향도만 파악 가능

카테고리명	점포명	재고 수량	선도저하 배기수량	...	\$등급 상품수	군집내 SKU 결과물 (마감시)
무	강남논현	40	1	...	1	0%
감자	강남논현	6	0	...	2	0%
고구마스낵	강남논현	100	0	...	0	0%
순두부	강남논현	9	0	...	0	0%
양파	마포	1341	18	...	3	0%
오이	잠실	186	5	...	4	0%
...	...	...	...	...	...	...

...

카테고리명	점포명	재고 수량	선도저하 배기수량	...	\$등급 상품수	군집내 SKU 결과물 (마감시)
무	강남논현	40	1	...	1	50%
감자	강남논현	6	0	...	2	50%
고구마스낵	강남논현	100	0	...	0	50%
순두부	강남논현	9	0	...	0	50%
양파	마포	1341	18	...	3	50%
오이	잠실	186	5	...	4	50%
...	...	...	...	...	...	...

동일한 가상의 데이터 셋에 결과물 정보만 바뀌서가상의 판매량 예측  
-> 결과물의 영향도에 민감한 카테고리 추출

## 2) 카테고리별 분석 결과

- 채소 5depth 카테고리 내 sku수가 5개 이상인 29개 카테고리 선정

카테고리명	매출수량 비중(결과물 0% 대비)					매출 기회로스 결과물 구간 (10%이상 매출 손실)	결과 대체 가능여부 (최종)	기회로스 (판매수량)	기회로스 (매출금액)
	매출비중 변화 주이 (SKU 결과물 10%)	매출비중 변화 주이 (SKU 결과물 20%)	매출비중 변화 주이 (SKU 결과물 30%)	매출비중 변화 주이 (SKU 결과물 40%)	매출비중 변화 주이 (SKU 결과물 50%)				
무	100%	100%	86%	86%	70%	30%	Y	536	₩1,199,553
RTF샐러드	100%	100%	90%	91%	71%	20%	N	299	₩1,411,848
가공샐러드	100%	100%	96%	96%	71%	없음	Y	41	₩101,341
감자	100%	100%	89%	89%	72%	30%	Y	629	₩4,055,695
고구마스낵	100%	100%	96%	96%	96%	없음	Y	119	₩354,972
고구마말반	100%	89%	89%	73%	73%	20%	N	465	₩3,188,865
깻잎	100%	100%	92%	92%	92%	없음	Y	425	₩1,393,487
냉동채소	85%	68%	68%	68%	68%	10%	N	1,081	₩3,929,908
다진마늘	100%	100%	94%	94%	94%	없음	Y	266	₩1,494,417
당근	100%	100%	89%	89%	73%	30%	Y	488	₩2,534,553
대파	100%	100%	94%	94%	76%	50%	Y	112	₩880,594
두부부침용	100%	100%	89%	89%	89%	30%	Y	575	₩1,442,977
두부부침찌개용	100%	88%	88%	88%	70%	20%	N	684	₩1,712,262
두부찌개용	100%	100%	98%	98%	85%	50%	Y	23	₩77,695
모듬콩	100%	100%	91%	91%	91%	없음	Y	206	₩1,193,626
배추	100%	100%	91%	91%	91%	없음	Y	421	₩1,801,063
부추	100%	100%	92%	92%	92%	없음	Y	193	₩738,684
브로콜리	100%	100%	93%	93%	93%	없음	Y	128	₩494,113
샐러드채소	88%	70%	70%	70%	70%	10%	N	690	₩5,106,714
생나물	100%	84%	84%	63%	63%	20%	N	468	₩1,570,896
손질채소	100%	86%	86%	69%	69%	20%	N	630	₩2,489,539
순두부	100%	100%	93%	93%	93%	없음	Y	396	₩712,145
쌈채소	100%	100%	96%	96%	96%	없음	Y	25	₩110,757
양배추	100%	88%	88%	69%	69%	20%	N	703	₩2,105,596
양파	100%	89%	89%	70%	70%	20%	N	612	₩5,001,436
오이	100%	100%	90%	90%	90%	없음	Y	537	₩2,739,652
청양고추	100%	100%	88%	88%	70%	30%	Y	635	₩2,976,804
콩나물	100%	100%	98%	98%	98%	없음	Y	26	₩52,113
파프리카	100%	100%	95%	95%	95%	없음	Y	115	₩500,747
허브류	100%	100%	95%	95%	95%	없음	Y	53	₩172,442

카테고리명	카테고리명	변태	가상매출량					판매량 대비 매출 비중 (20% 이상)					판매량 대비 매출 비중 (20% 이상)					기회로스 (판매수량)	기회로스 (매출금액)
			군집내 10%결과물 10%	군집내 10%결과물 20%	군집내 10%결과물 30%	군집내 10%결과물 40%	군집내 10%결과물 50%	군집내 10%결과물 10%	군집내 10%결과물 20%	군집내 10%결과물 30%	군집내 10%결과물 40%	군집내 10%결과물 50%	군집내 10%결과물 10%	군집내 10%결과물 20%	군집내 10%결과물 30%	군집내 10%결과물 40%	군집내 10%결과물 50%		
N	무	25	25	25	25	25	25	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	536	₩1,199,553
	RTF샐러드	14	14	14	14	14	14	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	299	₩1,411,848
	가공샐러드	12	12	12	12	12	12	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	41	₩101,341
	감자	26	26	26	26	26	26	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	629	₩4,055,695
	고구마스낵	46	46	46	46	46	46	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	119	₩354,972
Y	고구마말반	25	25	25	25	25	25	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	465	₩3,188,865
	깻잎	41	41	41	41	41	41	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	425	₩1,393,487
	냉동채소	28	28	28	28	28	28	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	1,081	₩3,929,908
	다진마늘	27	27	27	27	27	27	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	266	₩1,494,417
	당근	11	11	11	11	11	11	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	488	₩2,534,553

- \*매출 비중 변화 추이란: 결과물 SKU수가 증가할때마다 판매량이 비중이 줄어드는 정도 (Ex. 무 결과물 0%시 판매량이 100개이면, 해당 카테고리내 결과물 30%시 86개로 줄어든다고 볼 수 있음)
- \*예상 기회로스: 주단위 실제 판매량을 통해 통해 잃게 되는 예상 매출 수량과 매출 금액 산출(결과물 30%기준)
- \*결과 대체 가능 여부 선정 기준: FC- 카테고리 단위 세부 정보와, SKU정보를 종합적으로 판단

- 결품 대체 가능(Y): 결품을 20% 내에 매출 GAP이 10% 이상 **줄어들지 않는** FC가 50% 이상을 차지할 경우
- 결품 대체 불가(N): 결품을 20% 내에 매출 GAP이 10% 이상 **줄어드는** FC가 50% 이상을 차지할 경우

- 결품 대체 가능 카테고리(군집 결품을 적용대상): 무, 가공샐러드,감자, 고구마스넬, 잿잎, 다진마늘, 당근, 대파, 두부부침용, 두부찌개용, 모듬쌈,배추, 부추, 브로콜리,순두부, 쌈채소, 오이, 청양고추, 콩나물, 파프리카, 허브류 이상 20개
- 결품 대체 불가 카테고리(단일 결품을 적용대상): RTE샐러드, 고구마일반, 냉동채소, 두부 부침찌개겸용, 샐러드채소, 생나물, 손질채소, 양배추, 양파 이상 9개

### (최종 결론 도출)

- 전체 84개 카테고리 중 한 카테고리안에서 SKU가 5개 이상 존재 하는 카테고리는 29개이다
- 29개 카테고리 중 결품시 대체가 가능한 카테고리, 즉 군집 결품을 적용이 가능한 **카테고리는 20개** 존재한다(70%)
- 29개 카테고리 중 결품시 대체가 불가능한 카테고리, 즉 단일 결품으로 관리해야하는 **카테고리는 9개** 존재한다(30%)

### 3) 결품 대체 가능 카테고리 VS 결품 대체 불가 카테고리 비교 (고객 관점에서 원인 파악)

- 2번에서 나온 결과 중 결품 대체 가능 카테고리와 결품 대체 불가 카테고리의 세부 분석 진행
- PLC 등급, 상품 가격, 무게, 품목의 다양성등을 비교 분석하여 구분 지어진 주요 원인 도출

#### 1. 카테고리 내 S,A 등급 상품수에 따른 차이

결품 대체 가능 여부	카테고리명	상품명	PLC등급	상품가격(매입 원가)
대체가능 (군집 결품 적용대상)	고구마스넬	고구마루 바스 80g	A	1,728
		고구마루바스 240g	D	4,887
		상하농원 고창 황토고구마 스틱 60g	S	1,621
		한울푸드 순수고구마 130g	A	1,833
		한울푸드 축축고구마스틱 20g	A	700

결품 대체 가능 여부	카테고리명	상품명	PLC등급	상품가격(매입 원가)
대체불가 (SKU 결품 적용대상)	고구마 일반	GAP 익산 방고구마 1.2kg	D	3,825
		GAP 익산 방고구마 800g	D	2,859
		물밭 한일 물방고구마 1.5kg	C	6,036
		일발한 물방 고구마 1kg (봉)	S	3,519
		일발한 물방고구마 3kg	N	10,450
		일발한 호박고구마 1kg(봉)	B	4,180
		우농악 물방 고구마 1kg 봉	B	3,752
		새벽채소 방고구마 1kg	C	3,700
		새벽채소 호박고구마 1kg	B	4,600
		화산농림 해남 물방고구마 2kg	C	6,900
		화산농림 해남 호박고구마 2kg	C	7,690
		화산농림 해남 호박고구마 800g	B	3,295

특징비교	고구마 스넬	고구마 일반
S, A 등급 상품 비율	80%	8%
상품 가격 편차	작음(1,700원)	큼(3,000 ~ 7,500원)
상품 무게 편차	작음(60 ~ 200g)	큼(800 ~ 3kg)
품목의 다양성	비슷한 스펙류 상품만 존재	호박 고구마, 방고구마로 크게 구분



2. 무게 편차에 따른 차이

결품 대체 가능 여부	카테고리명	상품명	PLC등급	상품가격(매입 원가)
대체가능 (군집 결품을 적용대상)	다진마늘	베민이지 일자 그대로 살린 다진마늘 150g	S	2,722
		베민이지 일자 그대로 살린 다진마늘 250g	B	4,067
		다진마늘 큐브 180g	D	3,926
		열절삭을 써서쓰는 다진마늘 180g	B	2,750
		종가집 국산 다진마늘 150g	C	3,774
		물우원 국산 다진마늘 150g	A	4,043
		물우원 국산 다진마늘 260g	C	5,454
		물우원 갈끔하게 써서 쓰는 국산 다진 마늘 80g	A	2,781

결품 대체 가능 여부	카테고리명	상품명	PLC등급	상품가격(매입 원가)
대체불가 (SKU 결품을 적용대상)	양배추	강원대화농업 반통 적채 400g	D	1,650
		강원대화농업 비트 1입 300g	D	1,649
		강원대화농업 조각양배추 400g	B	1,650
		강원대화농업 반통 양배추 900g	C	2,688
		국내산 양양배추 1/2갓 500g 이상	A	1,600
		국내산 양양배추 1통 900g 이상	B	2,100
		디얼지 친환경 양배추 400g내외	S	1,554
		디얼지 친환경 양배추 800g내외 (1통)	B	2,506
		만인애 양배추 1통	B	2,277
		만인애 친환경 적채 1/2통	D	2,694
		정보 친환경 양배추 300g내외	A	1,701

특징비교	다진마늘	양배추
S, A 등급 상품 비율	38%	27%
상품 가격 편차	작음(3,000 ~ 4,000원)	작음(1,600 ~ 2,500원)
상품 무게 편차	적음(150 ~ 260g)	중(2배치이, 400g ~ 900g)
품목의 다양성	다진마늘 상품일 존재	양배추, 친환경 상품 존재

3. 품목의 다양성 차이

결품 대체 가능 여부	카테고리명	상품명	PLC등급	상품가격(매입 원가)
대체가능 (군집 결품을 적용대상)	가공샐러드	PIG 그린믹스 풀라진 샐러드 KIT 5입	통급없음	5,250
		누리탈 빵에 발라먹는 에그샐러드200g	통급없음	1,500
		부서먹는 에그샐러드 단호박 150g	통급없음	1,410
		부서먹는 에그샐러드 볼레인 150g	통급없음	1,410
		세프랩 고구마와운드샐러드110g	통급없음	990
		세프랩 크림베리단호박샐러드110g	통급없음	990
		아일탈 베지 송송 리얼 프테이르 샐러드 110g	통급없음	1,271
		아일탈 어니언 특촉 콘샐러드 100g	통급없음	1,271
		아일탈 통 단호박 크림베리 특촉 샐러드 110g	통급없음	1,271
		물우원 5mix 브란치샐러드 리코타고구마 200g	통급없음	2,307
		물우원 5mix 브란치샐러드 리코타단호박 200g	D	2,072
		물우원 5mix 브란치샐러드 베이컨에그프테이르 200g	C	2,032
		물우원 5mix 브란치샐러드 홍계상콘 200g	통급없음	2,284
		물우원 상가득 감자 샐러드 100g	B	1,135
		물우원 상가득 고구마 샐러드 100g	C	1,147
		물우원 상가득 단호박 샐러드 100g	B	1,128
		물우원 상가득 옥수수 샐러드 100g	B	1,150
		피그인더가든 에그마요 스프샐러드 120g	통급없음	1,350

특징비교	가공 샐러드	샐러드 채소
S, A 등급 상품 비율	0%	30%
상품 가격 편차	작음(3,000 ~ 4,000원)	작음(1,600 ~ 2,500원)
상품 무게 편차	적음(100 ~ 200g)	중(65g ~ 500g)
품목의 다양성	감자, 호박, 고구마 상품 위주	양상추, 양배추, 스킨채소, 당근, 적채, 파프리카등 다양한 채소 존재

결품 대체 가능 여부	카테고리명	상품명	PLC등급	상품가격(매입 원가)
대체불가 (SKU(단일) 결품을 적용대상)	샐러드채소	네멕 잔용 양배추 250g	S	1,729
		네멕 잔용 양상추 200g	A	1,790
		네멕 부드러운 샐러드 65g	B	1,083
		네멕 스틱채소 볼레터 250g	통급없음	1,990
		네멕 아삭아삭 샐러드 65g	A	1,127
		네멕 캐비지 샐러드 65g	A	1,127
		네멕 파프리카 샐러드 75g	S	1,152
		네멕 브릴라이즈 1입 (180g 내외)	통급없음	1,495
		네멕 하루샐러드 버트와당근 105g	A	1,252
		네멕 하루샐러드 양배추와샐러리&당근 95g	B	1,245
		네멕 하루샐러드 양배추&적채 105g	D	1,251
		에그 샐러리 통합달 180g	C	1,936
		사물 샐러리 200g	B	1,834
		사물 샐러리 400g	통급없음	3,000
		스마트달 버터릭드 120g내외	D	2,188
		스마트달 스틱믹스 80g내외	통급없음	2,200
		스마트달 아가페 120g내외	통급없음	2,200
		스마트달 크리스피아노 100g내외	통급없음	2,200
		지스트그린 베이비믹스 100g	C	2,035
		지스트그린 유로피안 풀라진 믹스 200g	C	2,830
		정보 친환경 샐러드채소 100g	C	2,236
		정보 친환경 샐러드채소 200g	통급없음	4,500
		정보 친환경 샐러드채소 50g	통급없음	1,200
		플레이토 뉴그린믹스 75g	통급없음	840
		플레이토 레드믹스 80g	통급없음	840
		플레이토 레드채소믹스 300g	D	3,180
		플레이토 비건안동 채소믹스 1kg	C	5,050
		플레이토 송송채소 통합합물 300g	통급없음	3,320
		플레이토 양배추와 적채 100g	B	863
		플레이토 양배추와 적채 300g	D	2,870
		플레이토 유리잔인 샐러드 130g	C	2,752
		플레이토 친환경 송송 유리잔인믹스 35g	D	1,346
		플레이토 친환경 채소믹스 500g	A	3,998
		플레이토 친환경 후레리가는 샐러드 250g	A	3,121
		플레이토 파프리카믹스 80g	C	1,113

(상품 등급에 따른 비교)

- (가설) 카테고리내에서 매출이 편중되어있는 상품이 있다면, 해당 상품 결품시 대체가 힘들것이다.
- (결과) 고구마스낵은 S,A등급 상품이 고르게 분포되어 있는 반면, 고구마일반은 S등급 상품이 편중되어 있으므로 해당 상품이 결품일 경우 결품 대체가 불가능 할 수 있음

(무게 편차에 따른 비교)

- (가설) 무게 편차가 큰 카테고리는 해당 상품 결품시 대체가 힘들 것이다. (Ex. 양배추 반통을 사려했으나 한통 밖에 없는 경우)
- (결과) 다진마늘 카테고리의 경우 결품시 대체할 수 있는 비슷한 용량의 상품이 있는 반면, 양배추의 경우 SKU별로 용량이 다양하여 결품 대체가 이루어지기 힘든것으로 판단됨

(품목의 다양성 차이 비교)

- (가설) 카테고리내에서 품목이 다양할수록 해당 상품 결품시 대체가 힘들 것이다. (샐러드 채소 내 파프리카 채소 SKU가 하나라면, 사실상 결품시 대체 불가)
- (결과) 가공샐러드의 경우 4-5개의 제조사에서 판매중인 샐러드의 종류는 대략 3가지 정도로 결품시 다른 품목을 구매할 수 있으나, 샐러드 채소의 경우 양상추, 양배추, 스틱채소, 파프리카 등 다양한 종류의 채소가 존재하여 결품시 대체가 불가능한것으로 보임

## \*참고 페이지 & 파일

Raw Excel sheet: [결품대체가능카테고리선정\\_0220.xlsx](#)

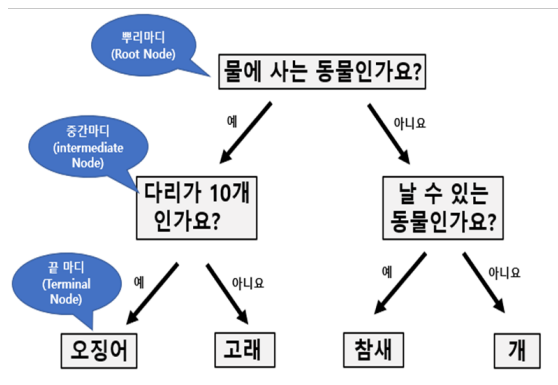
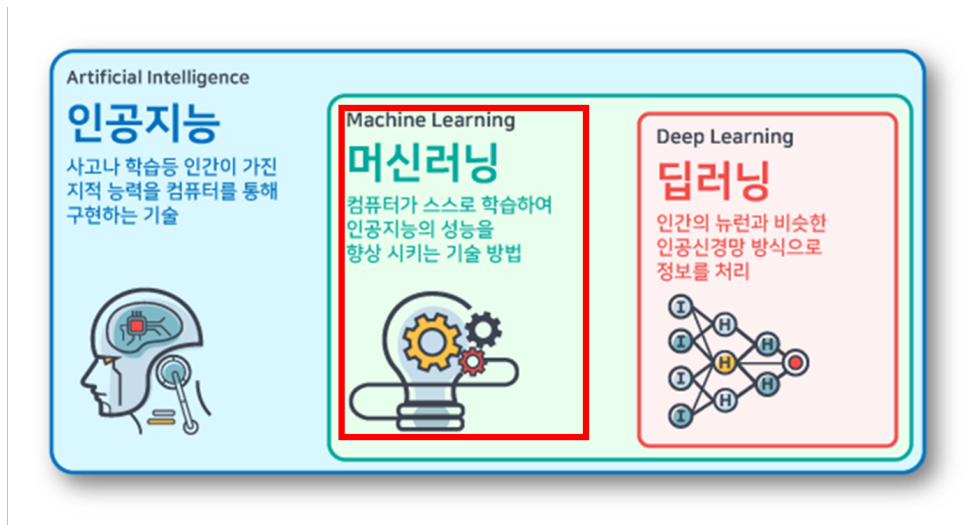
분석결과 도출 과정 세부페이지: [\[SCM팀\]결품 대체 가능성 2차 분석](#)

---

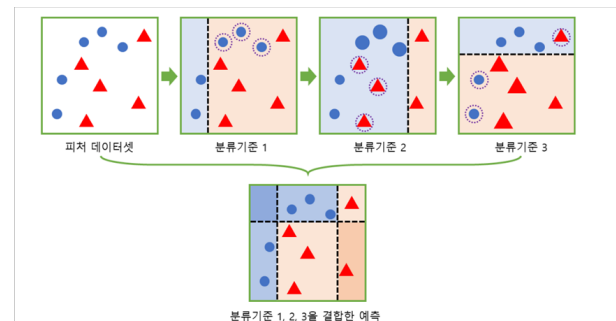
# APPENDIX

## 사용 모델

### 1. Light GBM 이란?



의사결정 나무(Tree)

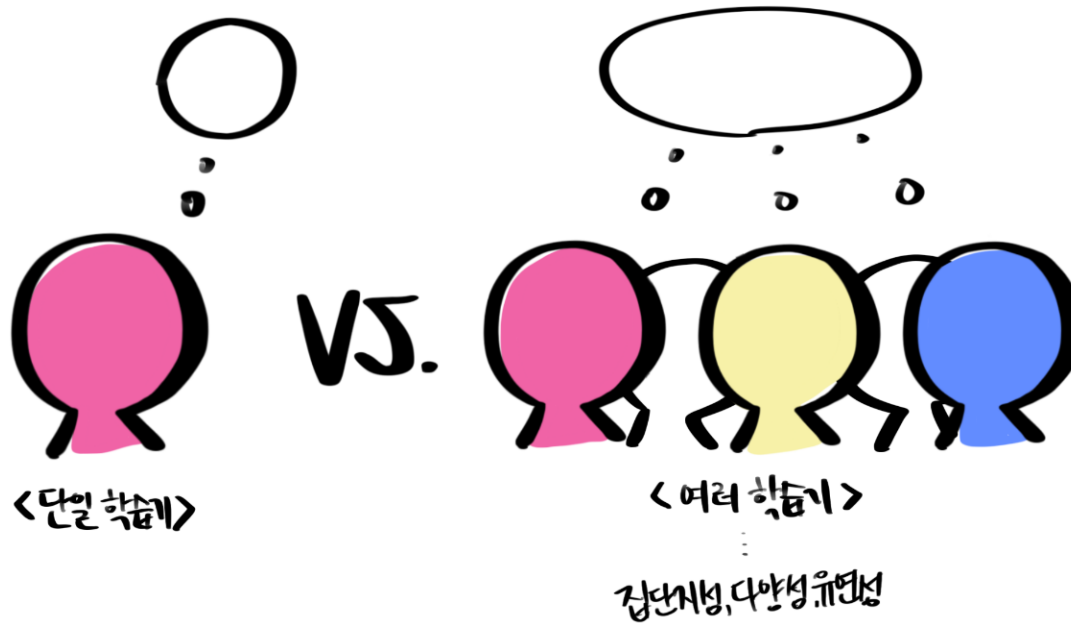


학습을 통해 낮은 예측력 보완(≒Forest)

-여러 개 학습기를 통해 판매량을 예측. 매 학습 과정마다 학습 능력이 부족한 학습기를 보완하여 강건한(Robust)한 모델로 학습

## 2. Light GBM을 사용한 이유

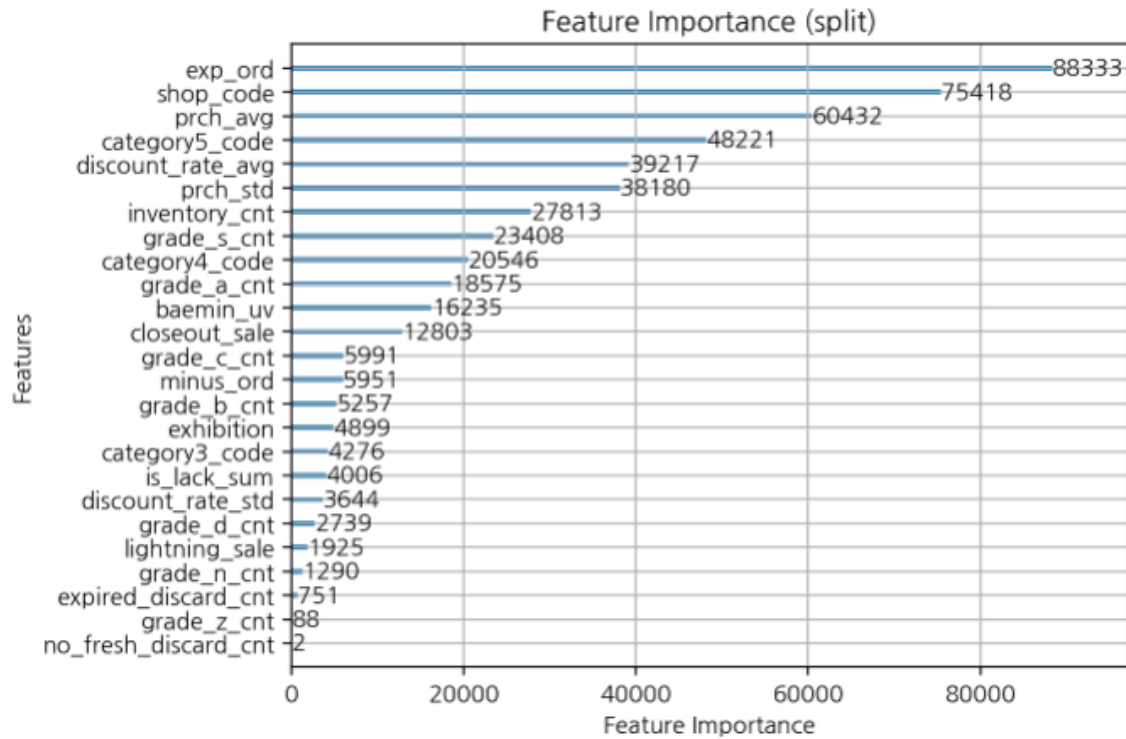
1) 성능이 뛰어난 모델



- 다수의 학습기를 활용 (예측력이 낮은 학습기의 예측 정확도를 높이는 방향으로 학습)
- 빠른 학습 속도, 높은 예측력, 과적합(over fitting) 문제에 강건함

2) XAI(Explainable AI) 가능한 모델: 모델 학습에 영향을 주는 변수를 확인할 수 있으며, 변수별로 상대적 중요도를 파악할 수 있음

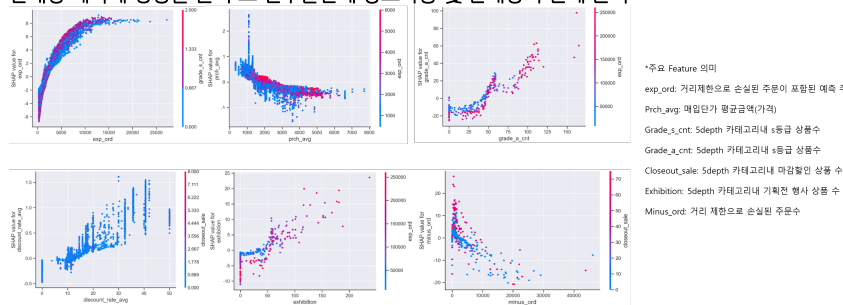
판매량 예측에 영향력을 미치는 변수 파악이 가능하며 결함에 대한 변수가 상대적으로 얼마나 중요한지 체크 할 수 있음



## 세부 분석 결과

### 1) 변수간 상호 작용 분석 및 시각화

- 판매량 예측에 영향을 준 주요 변수들간에 상호작용 및 판매량의 관계 분석



\*축 설명(첫번째 그림 예시) → X축: 주문수(예측), Y축(왼쪽): 판매량(예측), Y축(오른쪽) 5depth 카테고리 내 S등급 상품 수(예측)

- (예측 주문수 \* S 등급 상품 수) 상품 등급과 예측 주문수 간에 양의 상호 작용이 존재하며, 예측 주문수가 높고 카테고리 내 S 등급 상품 수가 많을 수록 판매량 상승에 긍정적인 영향을 줌
- (가격 평균(매입단가) \* 예측 주문 수) 평균 가격과 예측 주문수 간에 음의 상호 작용이 존재하며, 카테고리내 평균 가격이 높을 수록 판매량 상승에 부정적인 영향을 줌
- (A 등급 상품 수 \* 예측 주문수) A 상품 등급 상품수와 예측 주문수 간에 양의 상호 작용이 존재하며, 예측 주문수가 높고 카테고리 내 A등급 상품 수가 많을 수록 판매량 상승에 긍정적인 영향을 줌
- (할인을 평균 \* 마감할인 상품 수) 할인율과 마감할인 상품 수간에 양의 상호 작용이 존재하며, 할인율이 높고 카테고리내 마감할인 상품 수가 많을 수록 판매량 상승에 긍정적인 영향을 줌
- (거리제한 주문 수 \* 마감할인 상품수) 거리제한 주문수와 마감할인 상품 수간에 음의 상호 작용이 존재하며, 거리제한으로 손실된 주문수가 많아질 수록 판매량 예측에 부정적인 영향을 줌

(주요 변수 해석) 예측 주문수, S,A등급 상품수, 할인율, 마감할인 행사등은 판매량 상승에 긍정적인 영향을 주나, 가격 상승, 거리제한 손실 주문수 증가등은 판매량 상승에 부정적인 영향을 줌

## 2) 카테고리 별 세부 분석 -결품 대체 불가 카테고리(RTE샐러드)와 결품 대체 가능 카테고리(두부부침용) 비교

카테고리단위 결품 대체 가능여부	카테고리명	선택명	가상판매량						판매량 비동기(결품율 0% 대비)						판매수량	결품 대체 가능여부	매상 가치로소 (판매수량)	매상 가치로소
			군집내 SKU/결품율 0%	군집내 SKU/결품율 10%	군집내 SKU/결품율 20%	군집내 SKU/결품율 30%	군집내 SKU/결품율 40%	군집내 SKU/결품율 50%	군집내 SKU/결품율 30%	군집내 SKU/결품율 40%	군집내 SKU/결품율 50%	군집내 SKU/결품율 60%	군집내 SKU/결품율 70%	군집내 SKU/결품율 80%				
N	RTE샐러드	강남양정	20	20	20	20	17	17	100%	100%	100%	88%	40%	Y	-	-	₩0	
	RTE샐러드	강남양정	14	14	14	14	13	13	100%	100%	100%	84%	40%	Y	-	-	₩0	
	RTE샐러드	강남양정	17	17	17	17	16	16	100%	100%	100%	89%	40%	Y	-	-	₩0	
	RTE샐러드	강남양정	26	26	26	26	18	18	100%	87%	97%	79%	40%	Y	-	-	₩0	
	RTE샐러드	강동천호	68	68	54	54	44	44	100%	85%	80%	64%	20%	N	13	₩62,136		
	RTE샐러드	강서	25	25	25	25	18	18	100%	100%	100%	71%	40%	Y	5	₩28,954		
Y	RTE샐러드	강서외곽	61	61	67	67	53	53	100%	83%	89%	66%	20%	N	17	₩50,280		
	두부부침용	강남대치	26	26	26	26	26	26	100%	100%	100%	100%	100%	Y	-	-	₩0	
	두부부침용	강남입구정	21	21	21	21	21	21	100%	100%	100%	100%	100%	Y	-	-	₩0	
	두부부침용	강남대치	11	11	11	11	11	11	100%	100%	100%	100%	100%	Y	-	-	₩0	
	두부부침용	강동천호	24	24	24	20	20	20	100%	100%	82%	82%	30%	Y	11	₩57,161		
	두부부침용	강서가양	19	19	17	17	17	17	100%	100%	88%	88%	30%	Y	5	₩44,612		
	두부부침용	광진차랑	36	36	36	30	30	30	100%	100%	89%	89%	30%	Y	14	₩50,032		

- (RTE 샐러드) 강남입구정점에서는 결품 SKU수가 50%까지 증가하여도 판매량의 감소가 이루어지지 않는 반면, 강동천호점에서는 결품 SKU가 20%만 되어도 판매량의 20%가 감소함
- (두부 부침용) 모든 FC에서 결품을 20% 전에 판매 수량이 10%이상 감소하는 경우는 없으나, 강남 입구정 FC와 강서가양 FC에서는 결품률에 따른 판매수량의 추이에서 차이를 보이고 있음

(주요 결론) 카테고리에서 결품률에 따른 FC의 판매량 감소 추세는 대체적으로 동일하나 FC 별로 세부적으로 본다면, 판매량 추이에 대한 경향성은 다를 수 있다

- 좀 더 세부적으로 FC- 카테고리별로 군집 결품을 적용 대상을 분석할 수 있을 것으로 보임