

유통 빅데이터 기반

# 가을시즌 매출 성장을 위한 AI 기반 구매감소 예방 솔루션

L1 – 박도은, 예상일, 장재석, 전병수, 전정재, 한겨레, 김상현

# CONTENTS

01 과제 정의

02 데이터 수집 및 전처리

03 모델링

04 분석 결과 활용

CONTENTS

# 1. 과제정의

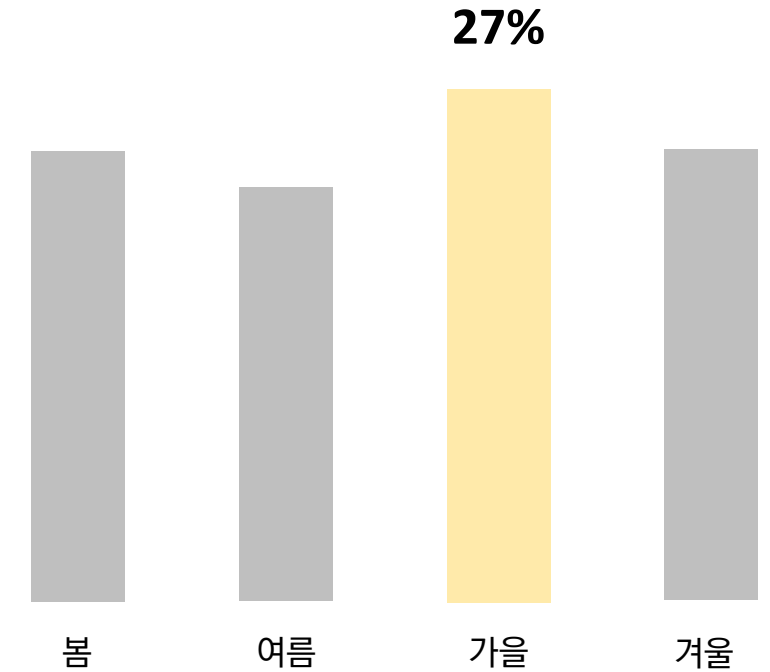
## 비즈니스 이슈

매년 가장 큰 매출비중을 차지하는 가을 매출이 정체되어있음

2014 - 2015 매출 추이



시즌별 매출비중



# 1. 과제정의

## 비즈니스 이슈

의류와 취미상품군에서 매출감소가 가을 정체의 주 요인

KRWm	2014 가을	2015 가을	증가량
상품 카테고리			
의류	28,269	27,671	(599)
기타	982	1,096	113
음식	21,067	23,616	2,548
취미	7,423	7,211	(212)
사치	7,987	9,489	1,502
생활용품	23,970	24,499	528
	89,701	93,585	3,883

- 의류와 취미상품군에서의 구매감소가 가을시즌 매출정체의 원인
- 14년도 대비 15년도 오락/문화 평균소비액이 증가 (통계청)

# 1. 과제정의

## 핵심 아이디어

### 비즈니스 이슈 데이터 과제화

#### 비즈니스 이슈

KRWm	2014 가을	2015 가을	증가량
상품 카테고리			
의류	28,269	27,671	(599)
기타	982	1,096	113
음식	21,067	23,616	2,548
취미	7,423	7,211	(212)
사치	7,987	9,489	1,502
생활용품	23,970	24,499	528
	89,701	93,585	3,883

- 가을시즌 의류와 취미상품 구매감소

비즈니스 이슈에 대응하기 위해  
2년간 수집한 고객 구매 데이터 활용



#### 데이터 과제화

- 고객데이터 활용
  - 고객 16295명
  - 의류 취미 관련 구매데이터 339017건
- 구매감소고객 예측모델을 개발
- 고객 행동 및 특성 별로 군집화
- 각 고객군별 대응방안 마련

Raw data

고객정보

경쟁사

멤버십

영수증

상품분류

채널

# 1. 과제정의

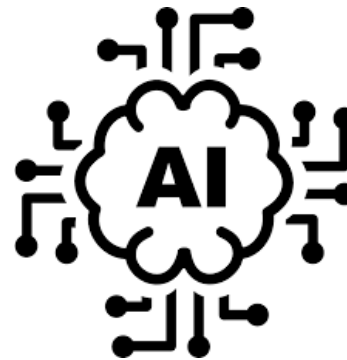
핵심 아이디어  
매출 성장을 위한 AI 기반  
**구매감소 예방 솔루션**



고객 구매 데이터  
수집



데이터 전처리  
및 변수선정



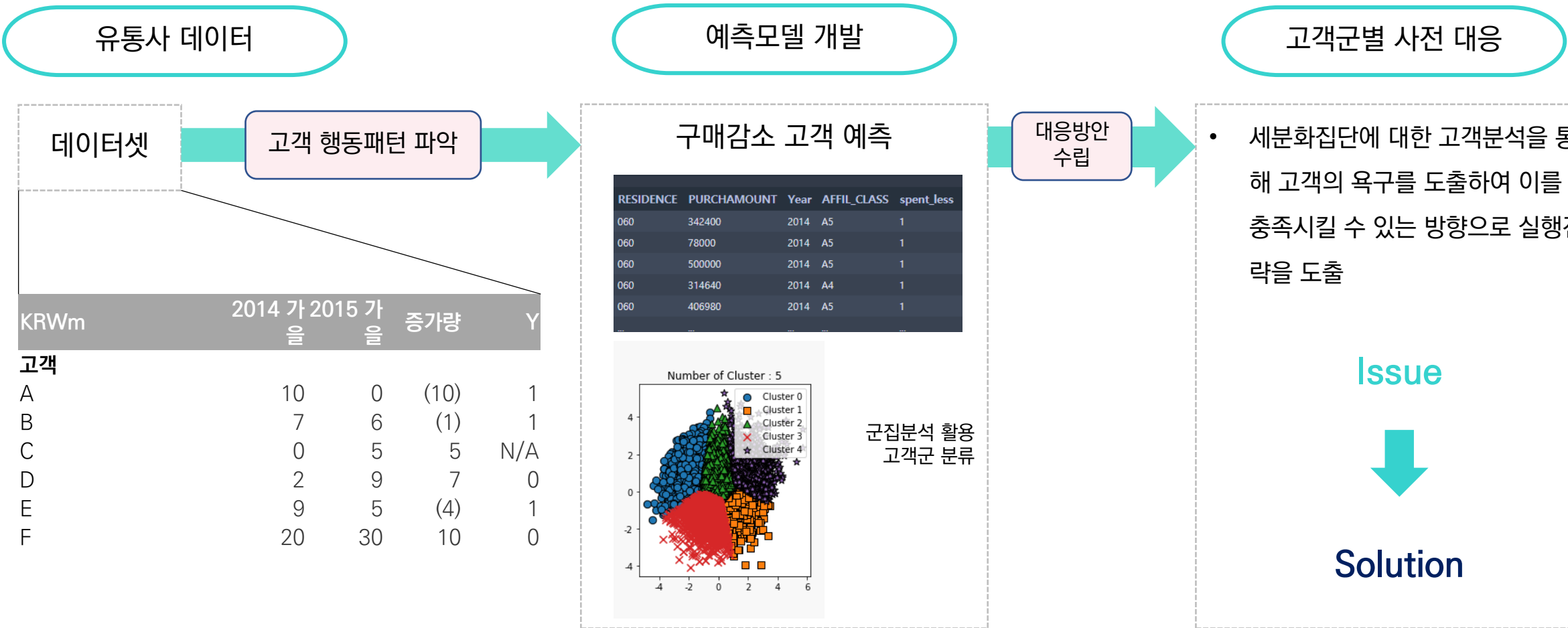
구매감소 예측 모  
델 개발 및 분석



감소 고객군별 마케팅  
전략 수립

# 1. 과제정의

## 핵심 아이디어 구매감소 예방 솔루션 Flowchart



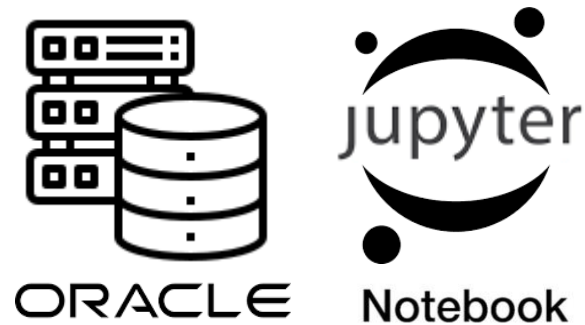
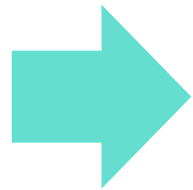
## 2. 데이터 수집 및 전처리

### 데이터 전처리

고객 행동패턴을 변수화하여 데이터셋 생성



Raw data:  
고객정보, 경쟁사, 멤버십,  
영수증, 상품분류, 채널



Oracle을 과 python을 활용한  
데이터 전처리, 변환, 마이닝



분석용 데이터셋

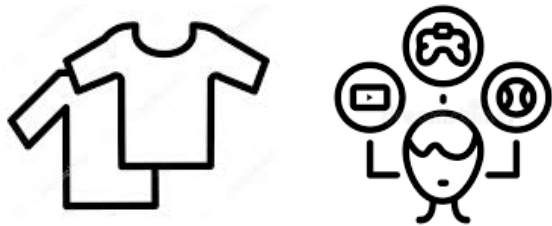


## 2. 데이터 수집 및 전처리

### 변수 생성

고객의 행동패턴을 파악하기 위한 변수 개발

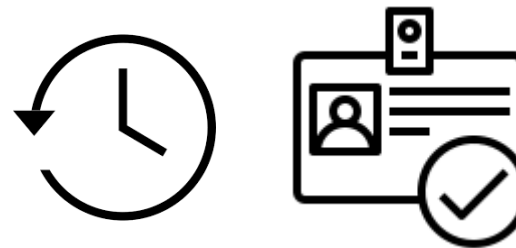
#### 상품



##### 주요변수

- 상품 카테고리별 구매증감
- 상품 카테고리별 하루평균 구매빈도
- 상품 가격대별 구매변동

#### 고객



##### 주요변수

- 최근 구매 경과일 수
- 멤버십 가입여부

#### 제휴사



##### 주요변수

- A 제휴사(백화점) 구매 증감
- B 제휴사(마트) 구매 증감

## 2. 데이터 수집 및 전처리

### 변수 선정

상관관계 분석 후 다중공선성 해결을 위해 변수제거

	spent_less	cloth_men_purch_diff	cloth_kids_purch_diff	cloth_casual_purch_diff	cloth_lingerie_purch_diff
spent_less	1.000000	0.144972	0.009342	0.090646	0.117385
cloth_men_purch_diff	0.144972	1.000000	0.002149	0.175638	0.063822
cloth_kids_purch_diff	0.009342	0.002149	1.000000	-0.001466	-0.000797
cloth_casual_purch_diff	0.090646	0.175638	-0.001466	1.000000	0.062554
cloth_lingerie_purch_diff	0.117385	0.063822	-0.000797	0.062554	1.000000
cloth_sports_purch_diff	0.020047	0.001278	-0.000463	0.002121	0.009059
cloth_etc_purch_diff	0.121593	0.037846	0.002179	0.215004	0.149921
cloth_business_purch_diff	0.071909	0.022476	-0.001523	0.168422	0.159291
hobby_media_purch_diff	0.045958	-0.001770	-0.002815	0.002064	0.003301
hobby_child_purch_diff	0.026015	0.000769	0.002317	-0.000151	0.000851
hobby_pet_purch_diff	0.049685	0.002823	0.006931	0.002920	0.003708
hobby_outdoor_purch_diff	0.106303	0.003225	0.002845	0.000081	0.015796
hobby_sports_purch_diff	0.014321	-0.000481	-0.022806	0.000386	0.005278
hobby_health_purch_diff	0.024328	0.000064	-0.002178	0.000901	0.009915
hobby_instex_purch_diff	0.029550	-0.003279	-0.003621	0.001117	-0.004684
cloth_women_purch_diff	0.008842	-0.003364	-0.003528	-0.002052	0.014894
hobby_golf_purch_diff	0.102919	-0.009734	0.016044	-0.010770	0.020124
A_purch_diff	0.137276	0.248660	0.001015	0.940269	0.428784
B_purch_diff	0.143663	0.006524	0.100901	0.010263	0.036520
Recency	0.062516	0.014763	0.015332	0.020224	0.025388
cluster_label	-0.055620	0.001741	-0.017946	-0.001224	-0.008784

### 변수 삭제

- 상관도 높은 독립변수 중 종속 변수와 상관관계가 낮은 것 제거
- 독립변수끼리 상관관계가 높은 것 추출(0.7기준)

## 2. 데이터 수집 및 전처리

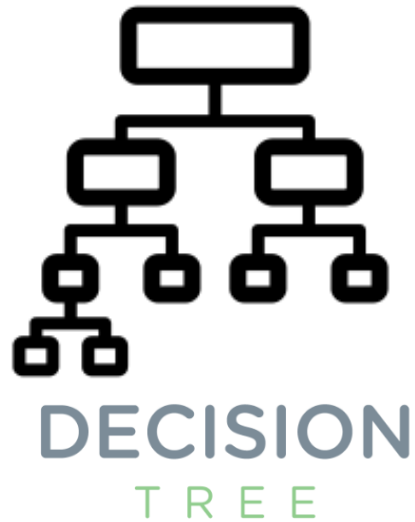
### 변수 선정

#### 최종 독립변수 목록

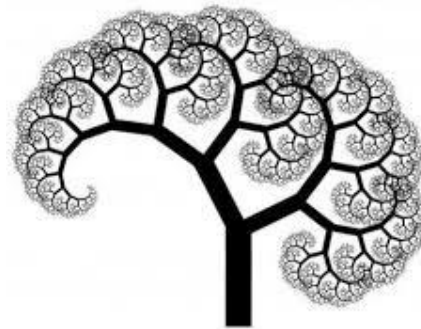
고객속성/구매변동 변수		구매금액증감 변수		구매빈도증감 변수	
변수명	설명	변수명	설명	변수명	설명
CUST_NO	고객번호	A_purch_diff	A사 증감	CLOTH_MEN_PERDAY_VISITING	남성의류_구매횟수
Recency	최근구매_경과일수	B_purch_diff	B사 증감	CLOTH_CASUAL_PERDAY_VISITING	캐주얼의류_구매횟수
cloth_highvalue	고가의류구매빈도_변동	cloth_kids_purch_diff	아동복 증감	CLOTH_ETC_PERDAY_VISITING	기타의류_구매횟수
cloth_mediumvalue	중가의류구매빈도_변동	cloth_men_purch_diff	남성복 증감	CLOTH_BUSINESS_PERDAY_VISITING	비즈니스의류_구매횟수
cloth_lowvalue	저가의류구매빈도_변동	cloth_women_purch_diff	여성복 증감	CLOTH_LINGERIE_PERDAY_VISITING	내의_구매횟수
hobby_highvalue	고가취미구매빈도_변동	cloth_business_purch_diff	비즈니스복 증감	CLOTH_SPORTS_PERDAY_VISITING	스포츠의류_구매횟수
hobby_mediumvalue	중가취미구매빈도_변동	cloth_etc_purch_diff	기타의류 증감	CLOTH_WOMEN_PERDAY_VISITING	여성의류_구매횟수
hobby_lowvalue	저가취미구매빈도_변동	cloth_lingerie_purch_diff	내의 증감	HOBBY_PET_PERDAY_VISITING	펫용품_구매횟수
MEMBERSHIP_CAT_2	14년 가을 멤버십 가입자	cloth_sports_purch_diff	스포츠의류 증감	HOBBY_GOLF_PERDAY_VISITING	골프용품_구매횟수
MEMBERSHIP_CAT_3	15년 가을 멤버십 가입자	hobby_media_purch_diff	음반_서적류 증감	HOBBY_MEDIA_PERDAY_VISITING	음반_서적류_구매횟수
		hobby_child_purch_diff	완구류 증감	HOBBY_CHILD_PERDAY_VISITING	완구류_구매횟수
		hobby_pet_purch_diff	펫용품 증감	HOBBY_HEALTH_PERDAY_VISITING	건강보조_구매횟수
		hobby_outdoor_purch_diff	아웃도어 증감	HOBBY_OUTDOOR_PERDAY_VISITING	아웃도어_구매횟수
		hobby_health_purch_diff	건강보조 증감	HOBBY_INSTEX_PERDAY_VISITING	운동기구_구매횟수
		hobby_golf_purch_diff	골프용품 증감		
		hobby_sports_purch_diff	골프용품 증감		
		hobby_instex_purch_diff	운동기구 증감		

#### 알고리즘 선정

##### 분류모델

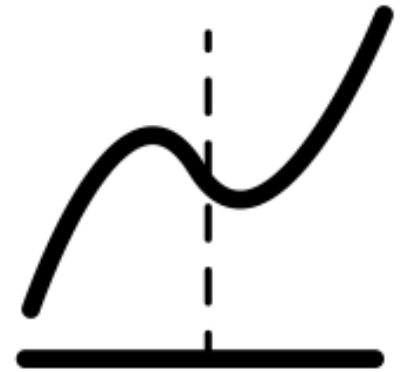


- 가장 기본적인 Classifier
- 'White Box'
- 과적합 리스크가 존재  
→ 다른 Classifier 참조



 LightGBM

- 현재 가장 성능이 뛰어나고 빠른 모델
- 정확한 구매 감소 고객 예측을 위하여 활용
- 변수 중요도 활용

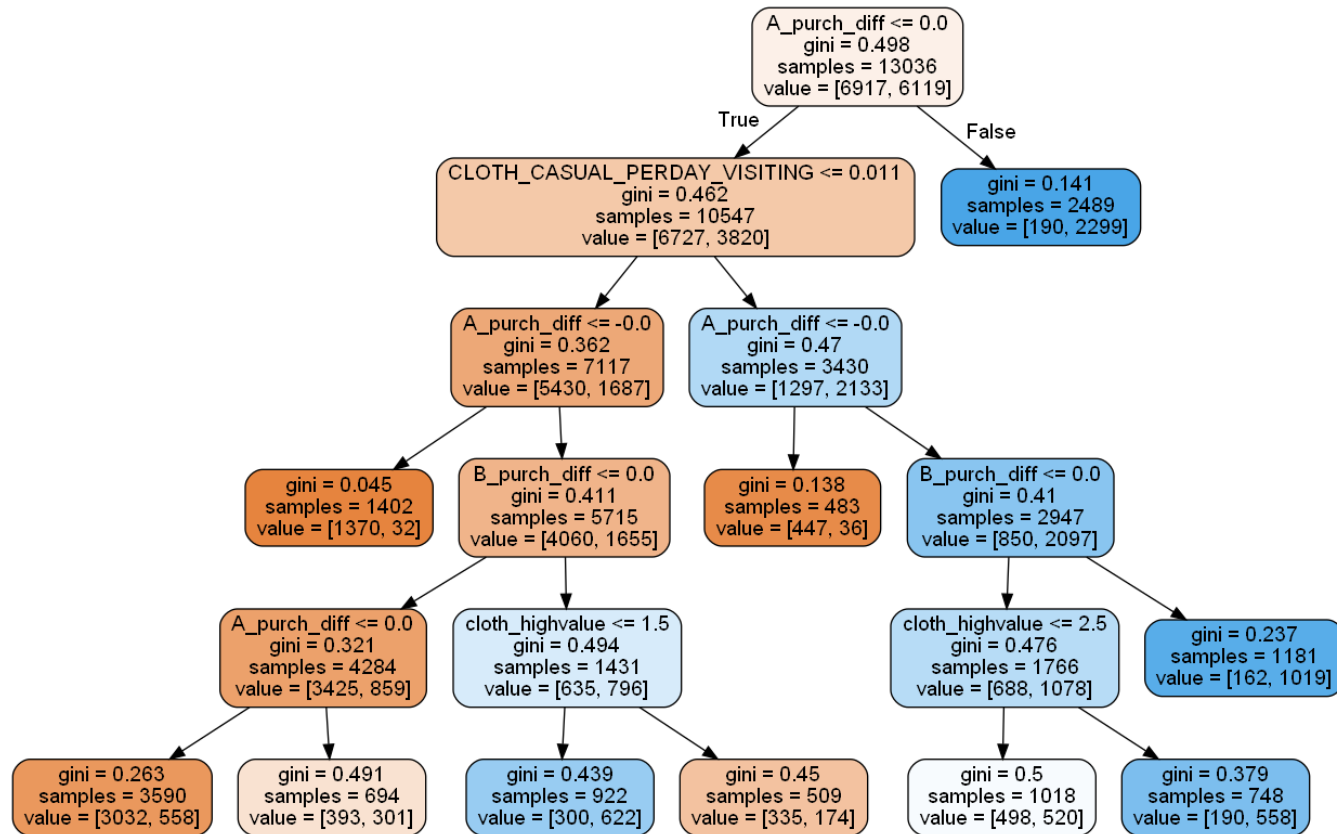


**Logistic  
Regression**

- 이진분류에 적합한 회귀모델
- 예측 정확도 활용
- 회귀계수와 p-값 활용

## Decision Tree

### 분석 결과



#### 평가 결과

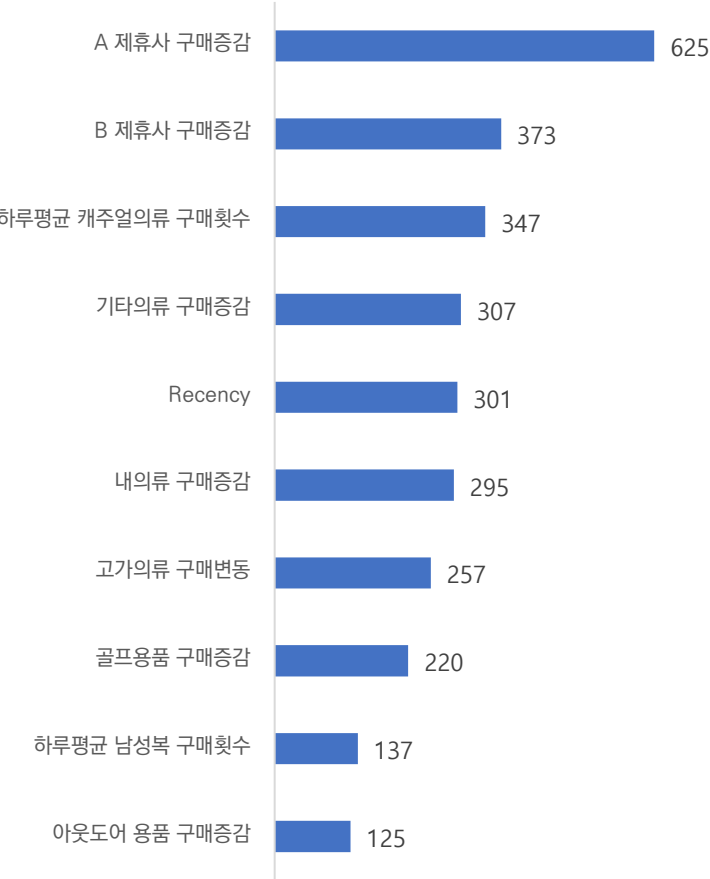
	Score
Accuracy	0.81
Precision	0.83
Sensitivity	0.82
F1-score	0.82
Area under the ROC	0.87

#### 주요변수

- 제휴사별 구매 증감
- 하루평균 캐주얼의류 구매횟수
- 하루평균 남성의류 구매횟수
- 고가의류 구매변동

LGBM Classifier  
분석 결과

변수 중요도 (Top 10)



제품군 별 변수 중요도 (Top 10)

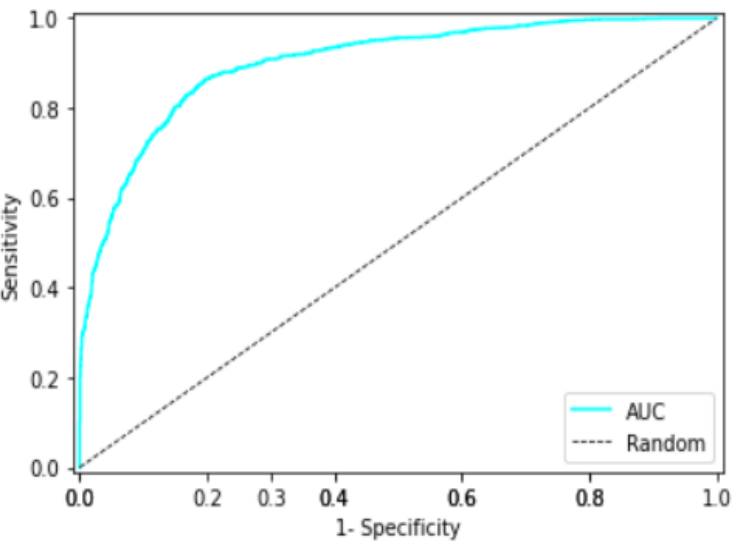
순위	제품군		
1	기타의류_증감	307	25.0%
2	내의_증감	295	24.0%
3	골프용품_증감	220	17.9%
4	아웃도어_증감	125	10.2%
5	비즈니스복_증감	108	8.8%
6	음반_서적류_증감	79	6.4%
7	펫용품_증감	53	4.3%
8	건강보조_증감	19	1.5%
9	스포츠용품_증감	11	0.9%
10	운동기구_증감	10	0.8%

평가 결과	Score
Accuracy	0.84
Precision	0.85
Sensitivity	0.86
F1-score	0.85
Area under the ROC	0.92

주요변수

- 제휴사별 구매 증감
- 하루평균 캐주얼의류 구매횟수
- 기타의류 구매증감
- 내의류 구매증감
- 골프용품 구매증감

Logistic Regression  
분석 결과



	coef	P >  t
cloth_lingerie_purch_diff	-0.2135	0.6139
cloth_sports_purch_diff	-0.1138	0.7845
cloth_etc_purch_diff	-1.5594	0.0000
cloth_business_purch_diff	-1.0399	0.0036
hobby_media_purch_diff	-0.3679	0.5621
hobby_pet_purch_diff	-0.0360	0.0046
hobby_outdoor_purch_diff	-0.2062	0.0000
hobby_sports_purch_diff	-0.0197	0.5397
hobby_health_purch_diff	-0.0641	0.0204
hobby_instex_purch_diff	-0.0934	0.3247
cloth_women_purch_diff	-0.0546	0.6537
hobby_golf_purch_diff	-0.7024	0.0000
A_purch_diff	-16.6877	0.0061
B_purch_diff	-0.8490	0.0030
Recency	-0.0476	0.0075
cloth_highvalue	0.2324	0.0000
cloth_lowvalue	-0.0030	0.6178
cloth_mediumvalue	-0.1132	0.0121
hobby_highvalue	-0.0648	0.5170
hobby_lowvalue	-0.0424	0.9355
hobby_mediumvalue	-0.0280	0.0008

평가 결과	Score
Accuracy	0.81
Precision	0.79
Sensitivity	0.88
F1-score	0.84
Area under the ROC	0.90

주요변수

- 제휴사별 구매증감
- 기타의류 구매증감
- 비즈니스 의류 구매증감
- 골프용품 구매증감
- 고가의류 구매변동
- 아웃도어용품 구매증감
- PERDAY\_VISITING

인사이트 도출

모델별 평가결과와 인사이트 요약

구분	LGBM Classifier	Logistic Regression	Decision Tree
예측모형 평가	Accuracy: 0.84 Sensitivity: 0.85 Precision : 0.86 AUC : 0.92	Accuracy: 0.81 Sensitivity: 0.79 Precision: 0.88 AUC : 0.90	Accuracy: 0.81 Sensitivity: 0.83 Precision: 0.82 AUC : 0.87
활용변수	37개	37개	37개

Insights

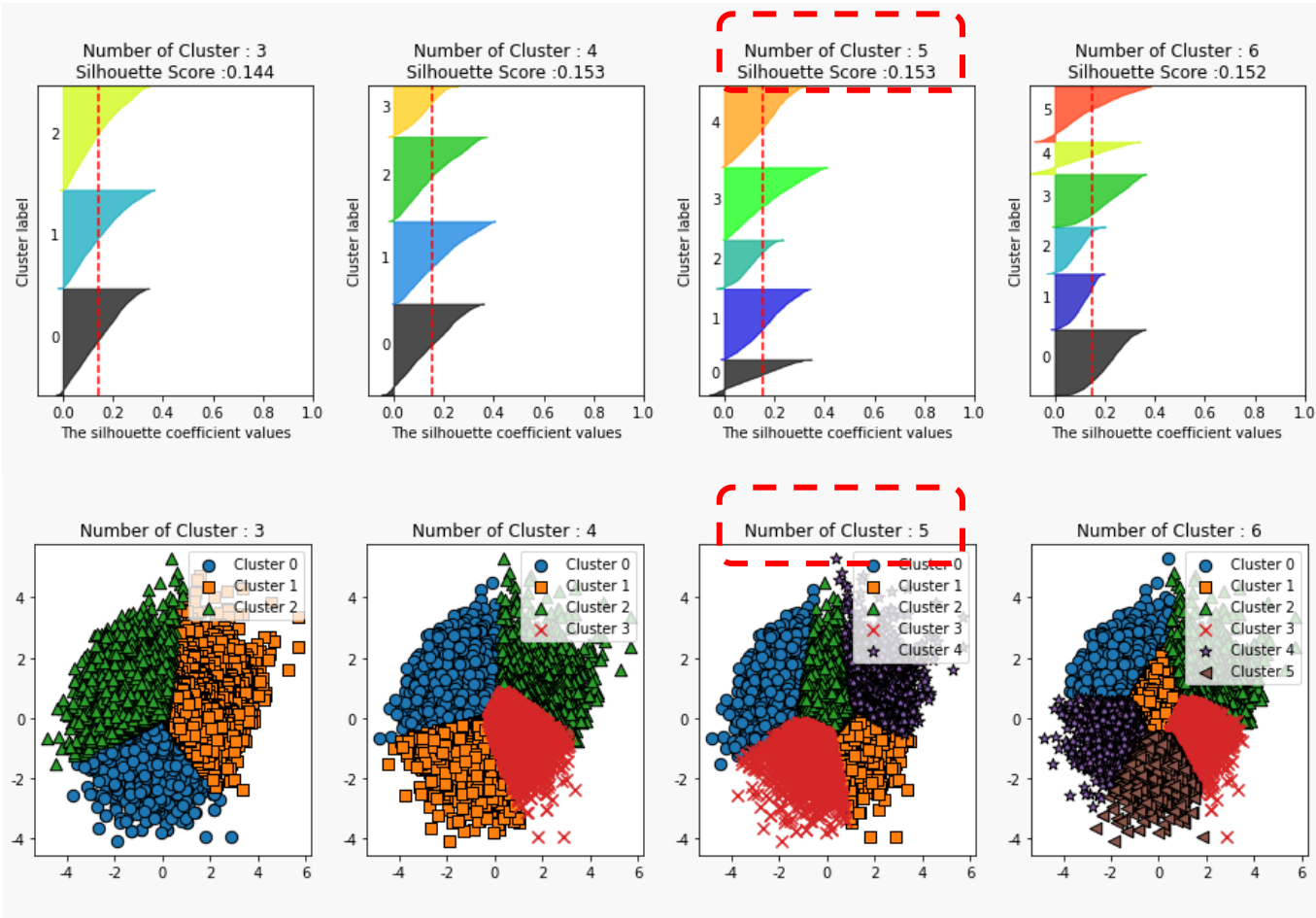
- A제휴사의 증감변수가 중요도 1순위
- 취미 - 골프>아웃도어 순으로 유의미
- 의류 - 기타>정장>고가의류 순으로 유의미
- A제휴사의 증감변수가 중요도 1순위
- 하루평균 캐주얼의류 구매횟수
- 하루평균 남성의류 구매횟수
- 고가의류 구매변동
- 기여도 기타의류(양말)>정장>골프>고가의류>아웃도어 순
- PERDAY\_VISITING 기여도 캐주얼 > 남성의류 순으로 유의미
- A제휴사의 증감변수가 중요도 1순위
- 하루평균 캐주얼의류 구매 횟수는 다른 제품들보다 우위
- 기타의류 ,내의,골프용품 등의 증감 변수 중요도가 큼



## 4. 분석 결과 활용

### 개선 방안

고객 특성 및 구매 패턴 파악을 위한 분석 기법 활용 - 군집분석



### 주요변수

- 나이
- 성별
- 구매액
- 구매빈도
- 거주지
- 온라인 채널 사용 여부
- 경쟁사 이용 여부
- 멤버십 가입 여부

개선 방안

고객 특성 및 구매 패턴 파악을 위한 분석 기법 활용 – 연관성 분석

antecedents	consequents	support	support	support	confidence	lift	leverage	conviction
(hobby_cat_hobby_sports)	(cloth_cat_cloth_sports)	0.001484	0.007455	0.001484	1	134.134328	0.001473	inf
(hobby_cat_hobby_media, hobby_cat_hobby_outdoor)	(hobby_cat_hobby_child)	0.003616	0.051797	0.001122	0.310256	5.989848	0.000935	1.374718
(cloth_cat_cloth_sports)	(hobby_cat_hobby_sports)	0.008567	0.001894	0.001894	0.221064	116.724713	0.001878	1.28137
(hobby_cat_hobby_sports)	(cloth_cat_cloth_sports)	0.001894	0.008567	0.001894	1	116.724713	0.001878	inf
(hobby_cat_hobby_health)	(hobby_cat_hobby_pet)	0.006843	0.05051	0.00201	0.293734	5.815397	0.001664	1.34438
(cloth_cat_cloth_women)	(cloth_cat_cloth_casual)	0.007023	0.267113	0.001947	0.277193	1.037735	0.000071	1.013945
(hobby_cat_hobby_sports)	(cloth_cat_cloth_sports)	0.003943	0.019812	0.003943	1	50.475124	0.003865	inf
(hobby_cat_hobby_health)	(hobby_cat_hobby_pet)	0.018358	0.133655	0.004509	0.245638	1.837844	0.002056	1.148446
(hobby_cat_hobby_media, hobby_cat_hobby_outdoor)	(hobby_cat_hobby_child)	0.00961	0.137647	0.002982	0.310256	2.253997	0.001659	1.250251
(hobby_cat_hobby_outdoor, hobby_cat_hobby_child)	(hobby_cat_hobby_media)	0.012198	0.11234	0.002982	0.244444	2.175926	0.001611	1.174844

연관성 분석 선정기준

- support(지지도) : 0.001
- Confidence(신뢰도): 0.2
- Lift(향상도) : 1

## 4. 분석 결과 활용

### 개선 방안

#### VIP고객 타겟 마케팅

#### Attributes

- 나이 40대 중반
- 구매액 최다
- 90% 여성
- 멤버십 가입율 저조 (1명당 0.2개 가입)
- 경쟁사 이용률 1위 (1명당 1.3회)

#### Issue

- 여성의류, 남성의류, 캐주얼 의류 부문에서 큰 폭의 구매 감소 발생
- 특히 캐주얼 의류 부문에서 상당한 구매 감소 발생  
(고객 전체의 70% 수준)
- 반려동물 용품 부문 구매 감소 발생

#### Solution

- VIP에게만 제공하는 제품 선공개 등의 마케팅으로 해당 고객군을 인플루언서로 활용
- 기타의류와 남성의류를 활용한 캐주얼 의류 구매 촉진

## 4. 분석 결과 활용

### 개선 방안

중장년층 여성고객 타겟 마케팅

#### Attributes

- 평균 나이 50대 후반
- 여성 90%
- 멤버십 1인당 평균 0.21개 가입
- 구매 빈도 / 구매액 두번째로 높음
- 저조한 온라인 채널 이용 (1인당 평균 3회)

#### Issue

- 캐주얼 의류, 비즈니스 의류 부문에서 주로 구매 감소 발생
- 주로 가족 단위 소비
- 취미 부문 소비가 저조
- 연령층이 높아 온라인 채널 활용에 한계 존재

#### Solution

- ‘패밀리존’에서 아동복과 영캐주얼, 성인복 등 가족 구성원 각자에게 필요한 모든 종류의 옷을 한자리에서 판매
- 취미 부문 소비로 이어질 수 있는 문화/예술/스포츠 강의 강화
- [건강용품-반려동물]의 상품조합을 활용한 마케팅 실시 및 매장 개선

## 4. 분석 결과 활용

### 개선 방안

남성고객 타겟 마케팅

#### Attributes

- 전원 남성
- 40 초반
- 멤버십 40% 가입
- 구매 금액 / 빈도 3번째

#### Issue

- 이용비중이 여성고객에 비해 현저히 적음
- 캐주얼 의류에서 큰 폭으로 구매감소 발생
- B 제휴사 구매금액 감소

#### Solution

- 남성고객 전반에 대해 온라인 채널 홍보 및 프로모션 진행
- 상대적으로 목적지향적이므로 온라인 주문 오프라인 수령 시스템 도입

## 4. 분석 결과 활용

### 개선 방안

충성고객 타겟 마케팅

#### Attributes

- 30대 후반 여성
- 멤버십 전원 가입
- 구매금액 4위
- 온라인 채널 최다 이용  
(1인당 약 10회)
- 고객군 중 경쟁사 이용 최저  
(1인당 0.7회)

#### Issue

- 캐주얼 의류 부문 구매 감소 발생
- 문화 예술 상품, 아웃도어 상품 부문  
구매횟수 감소 발생
- 제휴사 A사에서 보다 제휴사 B사에  
서 큰 폭의 구매 감소 발생

#### Solution

- 건강용품과 애완동물 용품을 연계해 소비  
할 수 있도록 하는 이벤트 시행 및 매장 환  
경 구축
- C사 온라인 이용 고객에 한해 할인쿠폰 제  
공, 접근성 제고
- B사의 스포츠 용품 구매 촉진 시 A사의 의  
류 구매 또한 촉진 가능, 이를 활용해 해당  
고객군의 A사 이용 비중 확대
- 온라인으로 구입하는 멤버에 대해 포인트  
적립등의 시스템 도입으로 충성도 유지

## 4. 분석 결과 활용

### 개선 방안

비충성고객 타겟 마케팅

#### Attributes

- 멤버십 전원 미가입
- 여성
- 40대 초반
- 구매금액 최하
- 구매빈도 최하

#### Issue

- C사에 대한 충성도가 떨어짐
- 여성의류 구매횟수 감소
- A, B 제휴사 구매금액 감소
- 아웃도어용품 구매횟수 감소

#### Solution

- A 제휴사의 고가의류, 취미용품에 대한 프로모션
- 온라인으로 스포츠용품을 구매하면 A 제휴사의 고가의류를 할인하는 프로모션

Q & A

Q&A



THANK YOU –

감사합니다.