



현대백화점

의사결정나무 & 신경망모형

빅데이터경영통계 20132496 김민지
 20132618 오연경
 20132663 이윤선

C CONTENTS



케이스선정



탐사과정



결과



결론



케이스선정



같은 브랜드의 다른 상품군을 구매하는
고객의 특성간에 차이가 있는지 알아보기 위함

케이스선정

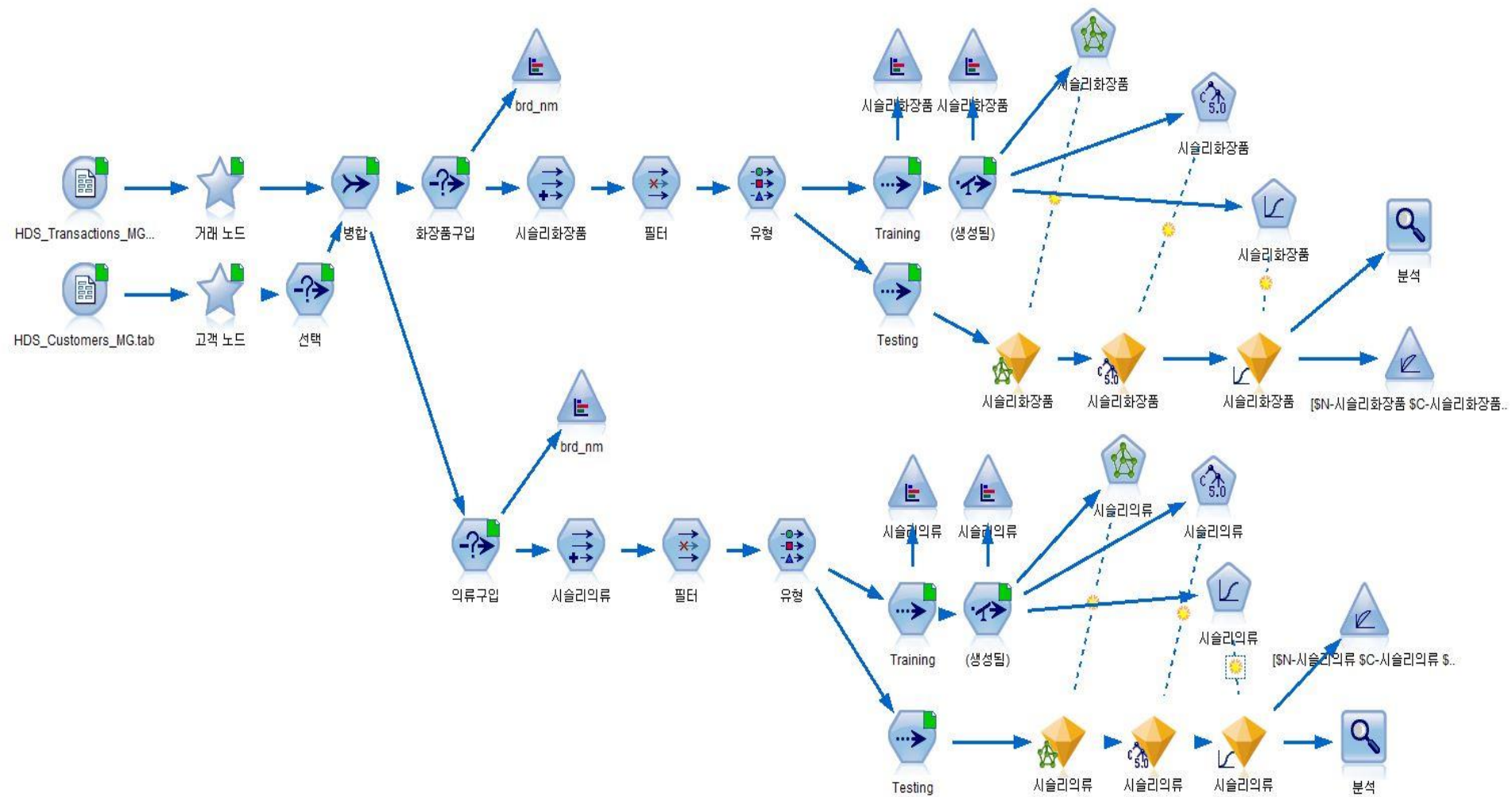


VS



SISLEY **화장품** 제품을 구매한 고객의 특성과
SISLEY **의류** 제품을 구매한 고객의 특성을 비교해본다

100



100

[illegible]

개인사업	개인사업
건설업	무역, 건설업
공사(한전)	공기업
교육기관	교육기관

지정되지 않은 값 사용 시: ☒ 원래 값 ☐ 기본값

undef

확인 취소 적용 (A) 재설정 (R)

탐사과정

화장품구입

미리보기(P)

설정

주석

모드: ☒ 포함 ☐ 삭제

조건: pc_nm='화장품'

확인

취소

적용 (A)

재설정 (R)

파생 필드:
시슬리화장품

파생 유형: 이분형

필드 유형: 이분형

참 값: Y 거짓 값: N

참일 조건:
brd_nm='시슬리'

확인

취소

적용 (A)

재설정 (R)

의류구입

미리보기(P)

설정

주석

모드: ☒ 포함 ☐ 삭제

조건: 상품군='여성의류'

확인

취소

적용 (A)

재설정 (R)

파생 필드:
시슬리의류

파생 유형: 이분형

필드 유형: 이분형

참 값: Y 거짓 값: N

참일 조건:
brd_nm='시슬리'

확인

취소

적용 (A)

재설정 (R)

100



탐사과정



신경망분석

시슬리화장품

목적: 표준 모형

필드 작성 옵션 모형 옵션 주석

사용자 정의 역할

사용자 정의 필드 할당

필드:

정렬: 없음

custid

corner_nm

대상*:

시슬리화장품

예측자(입력)*:

판매지점

계절

요일

고객등급

주거형태

주거현황

성별

연령대

직업군

모두

확인

실행

취소

적용 (A)

재설정 (R)

로지스틱

C5.0

시슬리화장품

대상이 플래그입니다. 이항 프로시저를 사용할 수 있습니다

필드 모델 고급 분석 주석

모델 이름: 자동 (O) 사용자 정의 (M)

파티션된 데이터 사용

각 분할에 대한 모형 작성

시슬리화장품

필드 모델 비용 분석 주석

모델 이름: 자동 (O) 사용자 정의 (M)

파티션된 데이터 사용

각 분할에 대한 모형 작성

출력 유형: 의사결정나무 의사결정규칙

그룹 기호

boosting 사용

교차타당성 확인

시도 횟수: 10

접기 횟수: 10

모드: 단순 고급

가치치기 강도: 75

자식마디 최소 레코드 수: 2

전역 가치치기 사용

필드유용성 사전조사

확인

실행

취소

적용 (A)

재설정 (R)

탐사과정

로지스틱

신경망분석

시슬리의류

목적: 표준 모형

필드 작성 옵션 모형 옵션 주석

☒ 사용자 정의 역할
☐ 사용자 정의 필드 할당

필드:

정렬: 없음

custid
 buyer_nm

대상*:

시슬리의류

예측자(입력)*:

판매지점
 계절
 요일
 고객등급
 주거형태
 주거현황
 성별
 연령대
 직업군

모두

확인 실행 취소

적용 (A) 재설정 (R)

시슬리의류

대상이 플러그입니다. 이항 프로시저를 사용할 수 있습니다

필드 모델 고급 분석 주석

모델 이름: ☒ 자동 (O) ☐ 사용자 정의 (M)

☒ 파티션된 데이터 사용
☒ 각 분할에 대한 모형 작성

프로시저: 다항분포

시슬리의류

필드 모델 비용 분석 주석

모델 이름: ☒ 자동 (O) ☐ 사용자 정의 (M)

☒ 파티션된 데이터 사용
☒ 각 분할에 대한 모형 작성

출력 유형:

☒ 의사결정나무 ☐ 의사결정규칙
☐ 그룹 기호
☐ boosting 사용 시도 횟수: 10
☐ 교차타당성 확인 접기 횟수: 10

모드:

☐ 단순 ☒ 고급

가치치기 강도: 75

자식마디 최소 레코드 수: 2

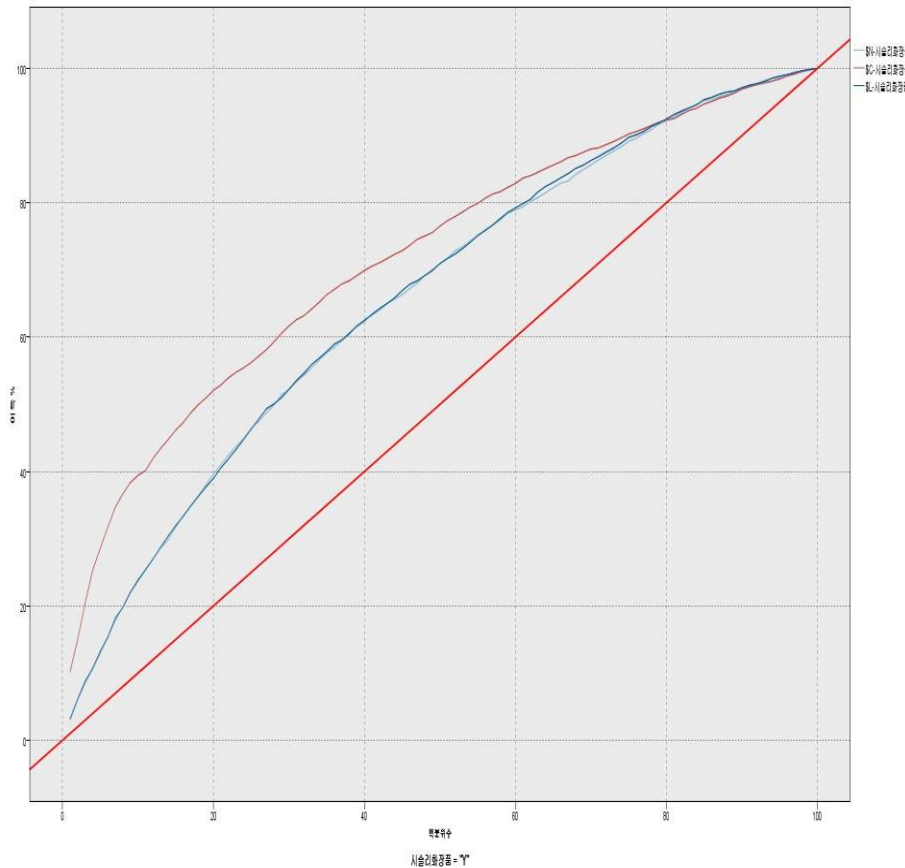
☒ 전역 가치치기 사용 ☐ 필드유용성 사전조사

확인 실행 취소

적용 (A) 재설정 (R)

C5.0

탐사과정



출력 필드 시슬리화장품의 결과

개별 모델

\$N-시슬리화장품와(과) 시슬리화장품 비교

정확	55,288	66.03%
틀림	28,440	33.97%
합계	83,728	

\$N-시슬리화장품의 일치 교차표(행은 실제 항목 표시)

성능 평가

\$C-시슬리화장품와(과) 시슬리화장품 비교

정확	74,453	88.92%
틀림	9,275	11.08%
합계	83,728	

\$C-시슬리화장품의 일치 교차표(행은 실제 항목 표시)

성능 평가

\$L-시슬리화장품와(과) 시슬리화장품 비교

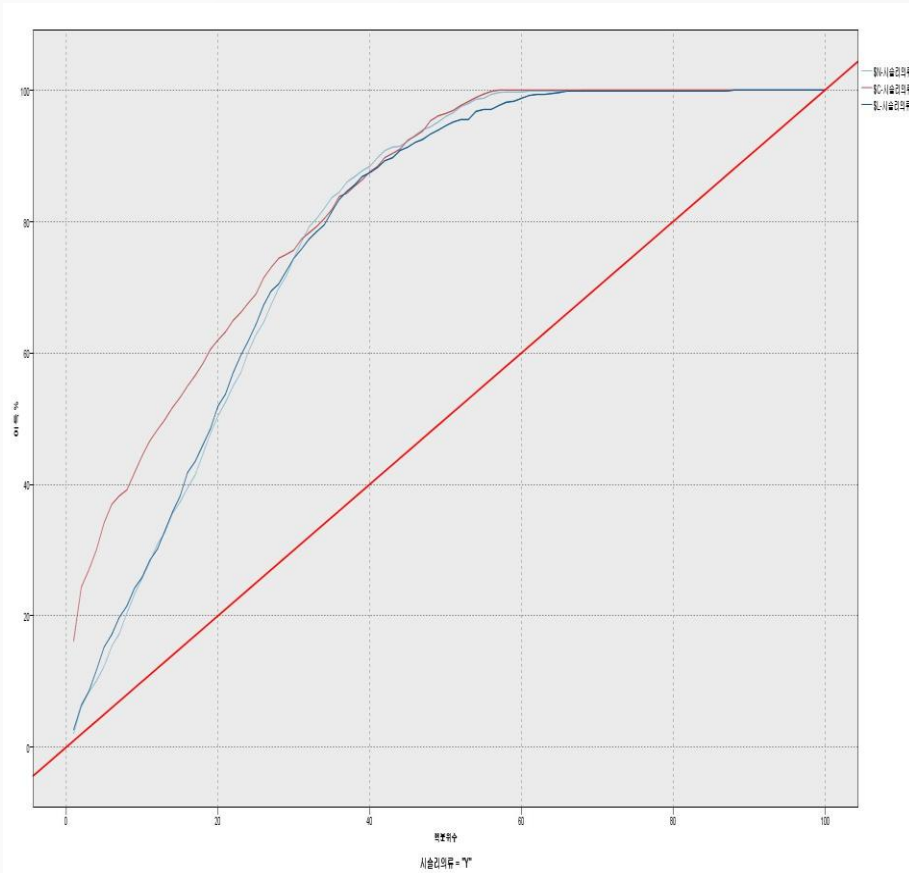
정확	57,117	68.22%
틀림	26,611	31.78%
합계	83,728	

\$L-시슬리화장품의 일치 교차표(행은 실제 항목 표시)

성능 평가

모델 적합도 : 의사결정나무 > 로지스틱 > 신경망분석

탐사과정



출력 필드 시술리의류의 결과

개별 모델

\$N\$-시술리의류와(과) 시술리의류 비교

정확	45,661	59.88%
틀림	30,592	40.12%
합계	76,253	

\$N\$-시술리의류의 일치 교차표(행은 실제 항목 표시)

성능 평가

\$C\$-시술리의류와(과) 시술리의류 비교

정확	73,653	96.59%
틀림	2,600	3.41%
합계	76,253	

\$C\$-시술리의류의 일치 교차표(행은 실제 항목 표시)

성능 평가

\$L\$-시술리의류와(과) 시술리의류 비교

정확	50,784	66.6%
틀림	25,469	33.4%
합계	76,253	

\$L\$-시술리의류의 일치 교차표(행은 실제 항목 표시)

성능 평가

모델 적합도 : 의사결정나무 > 로지스틱 > 신경망분석

탐사과정

TRAIN TEST

출력 필드 시슬리화장품의 결과

\$C-시슬리화장품와(과) 시슬리화장품 비교

정확	149,658	97.24%
틀림	4,255	2.76%
합계	153,913	

\$C-시슬리화장품의 일치 교차표(행은 실제 항목 표시)

	N	Y
N	76,026	4,172
Y	83	73,632

성능 평가

N	0.651
Y	0.681

출력 필드 시슬리화장품의 결과

개별 모델

\$N-시슬리화장품와(과) 시슬리화장품 비교

정확	74,452	88.92%
틀림	9,276	11.08%
합계	83,728	

\$C-시슬리화장품와(과) 시슬리화장품 비교

\$C-시슬리화장품의 일치 교차표(행은 실제 항목 표시)

	N	Y
N	73,075	7,163
Y	2,113	1,377

성능 평가

N	0.014
Y	1.353

\$L-시슬리화장품와(과) 시슬리화장품 비교

\$N-시슬리화장품 \$C-시슬리화장품 \$L-시슬리화장품 간의 일치



출력 필드 시슬리의류의 결과

\$C-시슬리의류와(과) 시슬리의류 비교

정확	148,768	99.25%
틀림	1,124	0.75%
합계	149,892	

\$C-시슬리의류의 일치 교차표(행은 실제 항목 표시)

	N	Y
N	74,394	1,124
Y	0	74,374

성능 평가

N	0.686
Y	0.686

정확	73,636	96.57%
틀림	2,617	3.43%
합계	76,253	

\$C-시슬리의류의 일치 교차표(행은 실제 항목 표시)

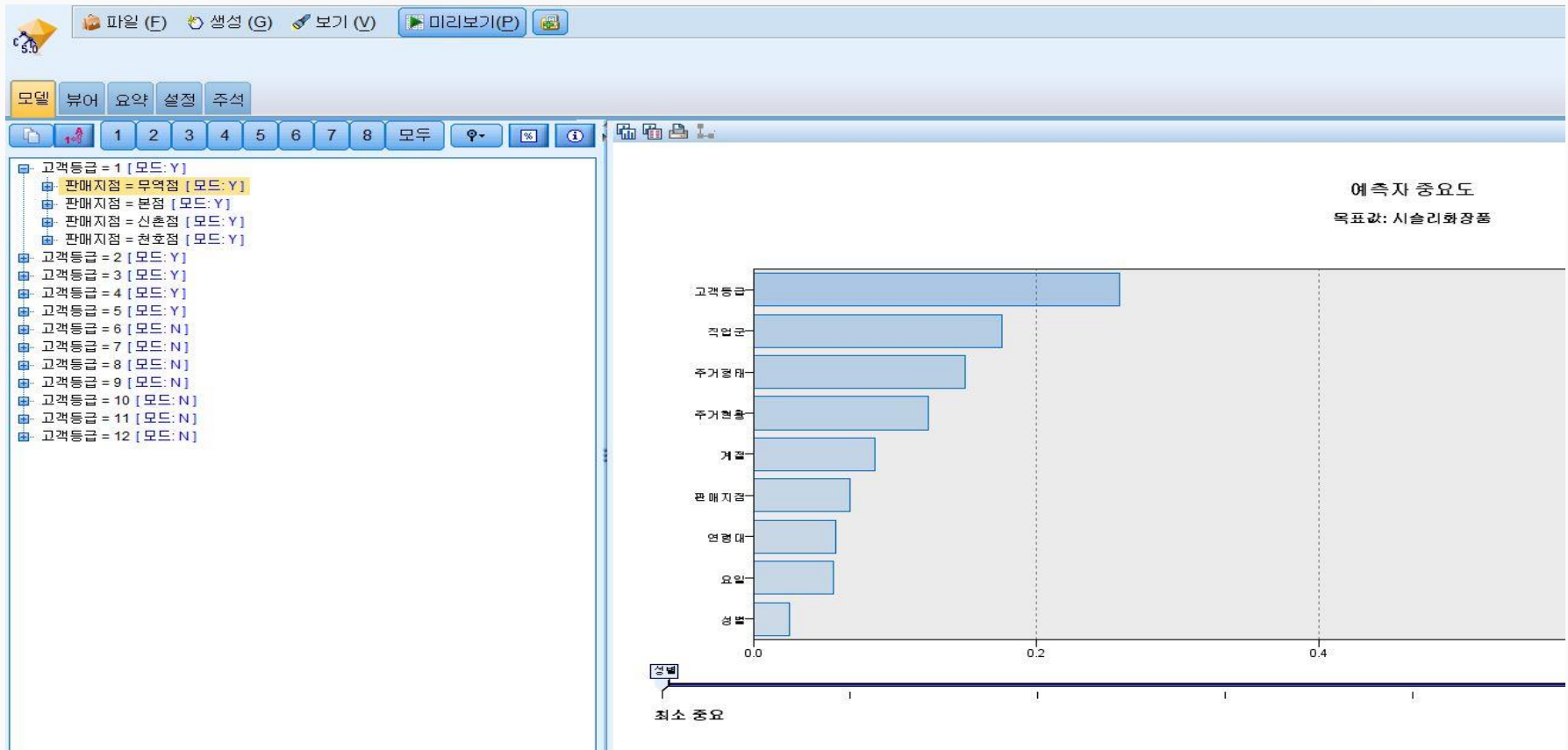
	N	Y
N	73,444	2,094
Y	523	192

성능 평가

N	0.002
Y	2.192



결과



SISLEY 화장품 구매 변수 중요도

:고객등급 > 직업군 > 주거형태 > 주거현황 > 구매계절

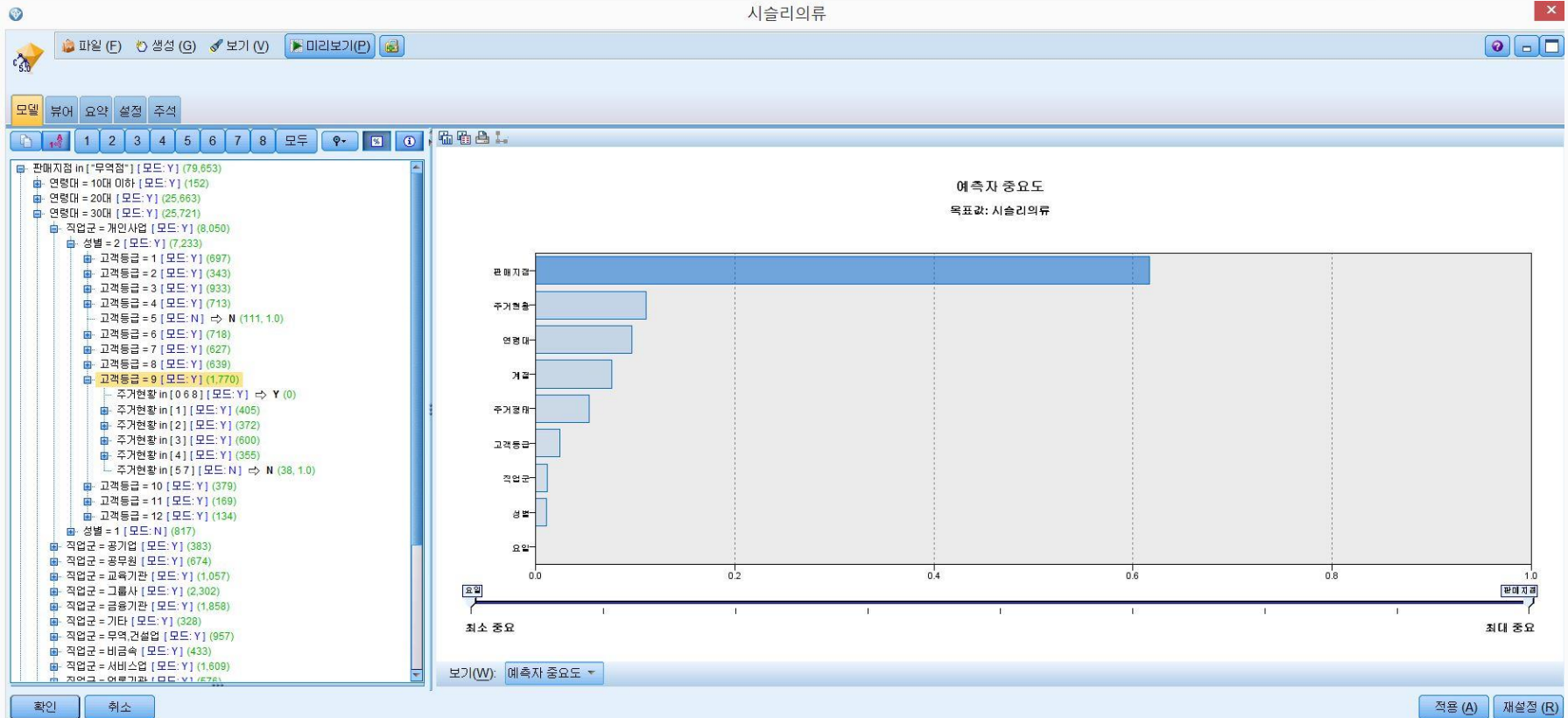
결과



```
고객등급 = 1 [모드: Y] (14,741)
고객등급 = 2 [모드: Y] (9,157)
고객등급 = 3 [모드: Y] (17,088)
고객등급 = 4 [모드: Y] (13,882.957)
고객등급 = 5 [모드: Y] (8,415)
고객등급 = 6 [모드: N] (9,302)
고객등급 = 7 [모드: N] (11,294)
고객등급 = 8 [모드: N] (13,341)
고객등급 = 9 [모드: N] (35,603)
연령대 = 10대 이하 [모드: N] => N (15, 1.0)
연령대 = 20대 [모드: N] (18,274)
직업군 in ["개인사업"] [모드: N] (8,439)
주거현황 in [0] [모드: N] => N (0)
주거현황 in [1] [모드: Y] (794)
주거현황 in [2] [모드: Y] (323)
주거현황 in [3] [모드: N] (4,257)
주거형태 in ["$null$" "B" "H"] [모드: N] => N (0)
주거형태 in ["A"] [모드: N] (1,646)
주거형태 in ["N"] [모드: N] (1,735)
요일 = 금 [모드: N] (294)
요일 = 목 [모드: N] (209)
요일 = 수 [모드: N] (183)
요일 = 월 [모드: N] (102)
요일 = 일 [모드: N] (280)
요일 = 토 [모드: N] (404)
성별 = 2 [모드: N] (381)
계절 in ["가을" "겨울"] [모드: N] => N (197, 0.68)
계절 in ["봄"] [모드: Y] (119)
판매지점 in ["무역점" "천호점"] [모드: Y] => Y (91, 0.692)
판매지점 in ["본점" "신촌점"] [모드: N] => N (28, 1.0)
계절 in ["여름"] [모드: N] (65)
성별 = 1 [모드: N] => N (23, 1.0)
요일 = 화 [모드: N] (263)
주거형태 in ["V"] [모드: N] (454)
주거형태 in ["X"] [모드: N] => N (1, 1.0)
주거형태 in ["Z"] [모드: N] (421)
```

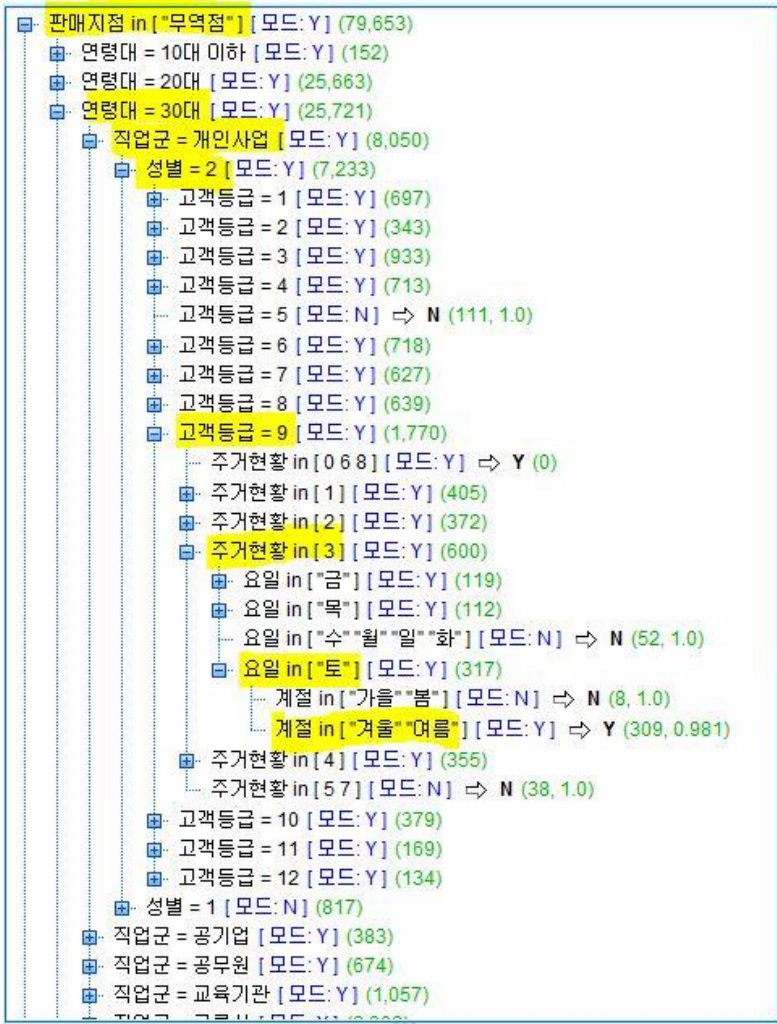
- + 고객 등급 9등급 (고정고객)
- + 연령대 20대
- + 개인사업
- + 주거현황 3 (부모소유)
- + 주거형태 N (단독주택)
- + 토요일 구매
- + 여성
- + 봄 구매
- + 무역점 or 천호점 구매
-> 시슬리 화장품 구매
(신뢰도 0.692)

결과



SISLEY 의류 구매 변수 중요도
 :판매지점 > 주거현황 > 연령대 > 구매계절


결과



- + 무역점 구매
- + 연령대 30대
- + 개인사업
- + 여성
- + 고객 9등급 (고정 고객)
- + 주거현황 3 (부모 소유)
- + 토요일 구매
- + 겨울 or 여름 구매
- > 시슬리 의류 구매 (신뢰도 0.981)

결론



SISELY 화장품

- ! 화장품 구매 고객 주요 특성 :
고객등급 > 직업군 > 주거형태 > 주거현황 > 구매계절
- ! 고객 등급 : 9 > 3 > 1 > 4
- ! 직업군 : 개인사업 > 제조업 > 그룹사 > 전문직 종사자
- ! 주거형태 : A(아파트) > N(단독주택) > V(빌라)
- ! 계절 : 계절간 큰 차이 없음 

SISELY 의류

- △ 의류 구매 고객 주요 특성 :
판매지점 > 주거현황 > 연령대 > 계절 > 주거형태
- △ 최대 판매 지점 : 무역점
- △ 주거현황: 3(부모소유) > 1(본인소유) > 4(전세)
- △ 연령대: 20대 > 30대 > 40대
- △ 계절 : 가을 > 봄 > 여름 > 겨울

결론

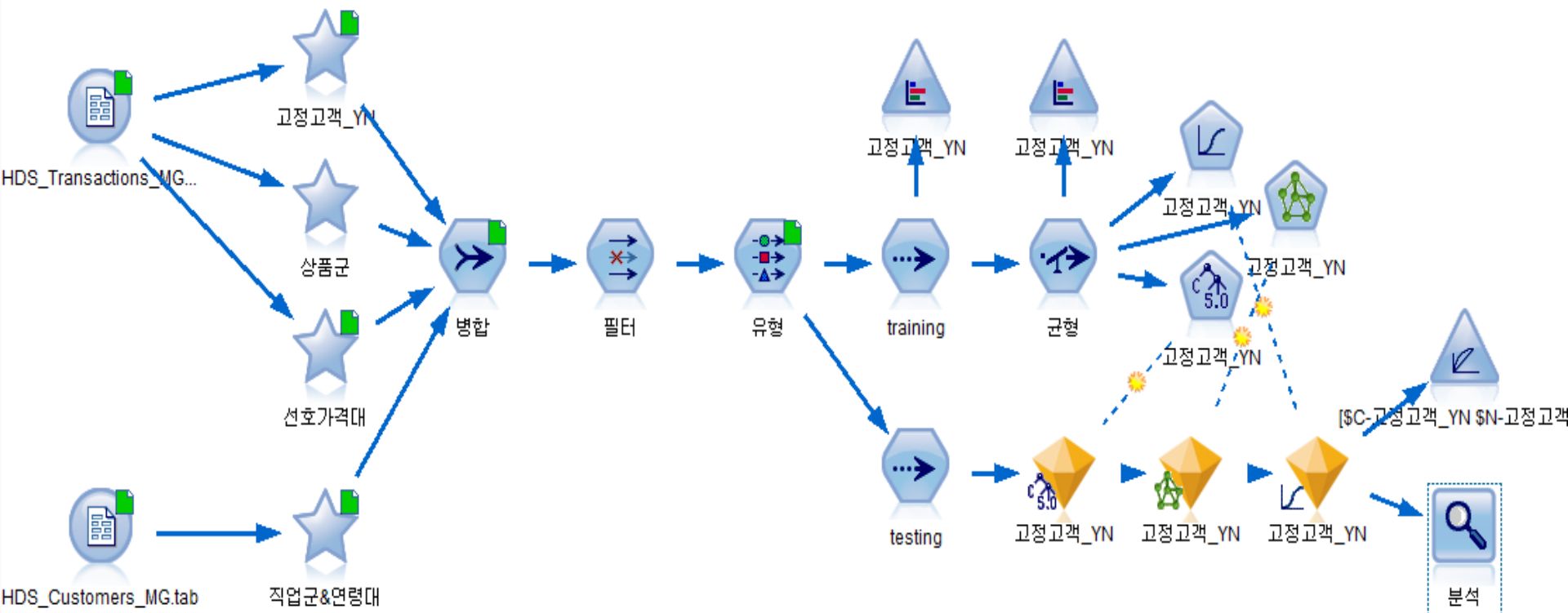
의류는 화장품과 달리
 계절, 판매지점 
에서 차별적인 구매 특성을 가진다

케이스선정

향후 고정고객이 될 가능성이 높은 고객들은 누구이며
그들의 특징은 무엇인가?

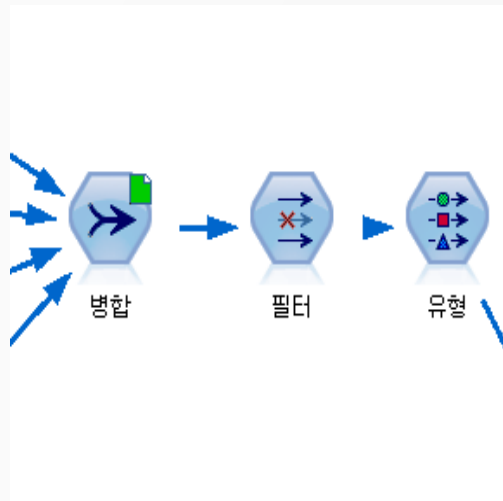
탐사과정

전체노드



탐사과정

유형노드 - 각 변수의 역할을 지정해줌



유형

미리보기(P)

유형 형식 주석

값 읽기 값 지우기 모든 값 지우기

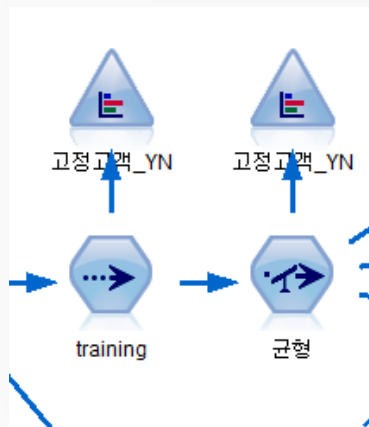
필드	초점	값	누락	화이	연함
sex	명목	0,1,2		EO	EO
mrg_flg	명목	0,1,2,7		EO	EO
hobby	명목	1,2,3,4,5,...		EO	EO
직업군	명목	기타,무역,...		EO	EO
연령대	명목	*10대 미...		EO	EO
str_nm	명목	무역점,본...		EO	EO
상품군	명목	스포츠캐...		EO	EO
고정고객_YN	이분형	1/0		EO	EO
custid	연속형	[1,50000]		EO	EO
birth	유형 없음			EO	EO
birth_flg	연속형	[-1,2]		EO	EO
home_zip1	연속형	[100,730]		EO	EO

☒ 현재 필드 보기 ☐ 사용되지 않은 필드 설정 보기

확인 취소 적용 (A) 재설정 (R)

탐사과정


균형노드 - 고정고객 기준에 따라 나뉜 고정고객 여부 분포를 조정



고정고객_YN 분포

값	비율	%	빈도
0		50.52	538185
1		49.48	527040

탐사과정



로지스틱

대상이 플러그입니다. 이항 프로시저를 사용할 수 있습니다.

필드 모델 고급 분석 주석

모델 이름: ☐ 자동 (O) ☐ 사용자 정의 (M)

☒ 파티션된 데이터 사용

☒ 각 분할에 대한 모형 작성

프로시저: ☐ 다항분포 ☐ 이항검정

다항 프로시저

방법: 모든 변수 사용

대상의 기본 범주: 0 지정...

모델 유형: ☒ 주 효과 ☐ 완전 요인 ☐ 사용자 정의

모델 용어:

☒ 방정식에 상수항 포함

확인 실행 취소



신경망

목적: 표준 모형

필드 작성 옵션 모형 옵션 주석

☒ 사용자 정의 역할

☐ 사용자 정의 필드 할당

필드:

정렬: 없음

custid

선택가격대

대상*:

고정고객_YN

예측자(입력)*:

sex

mrg_flg

hobby

직업군

연령대

str_nm

상품군

모두

확인 실행 취소

적용 (A) 재설정 (R)

탐사과정

C5.0

필드 모델 비용 분석 주석

모델 이름: ☐ 자동 (O) ☐ 사용자 정의 (M)

☒ 파티션된 데이터 사용

☒ 각 분할에 대한 모델 작성

출력 유형: ☒ 의사결정나무 ☐ 의사결정규칙

☐ 그룹 기호

☐ boosting 사용 시도 횟수:

☐ 교차타당성 확인 접기 횟수:

모드: ☐ 단순 ☒ 고급

가지치기 강도:

자식마디 최소 레코드 수:

☒ 전역 가지치기 사용 ☐ 필드유용성 사전조사

확인 실행 취소 적용 (A) 재설정 (R)

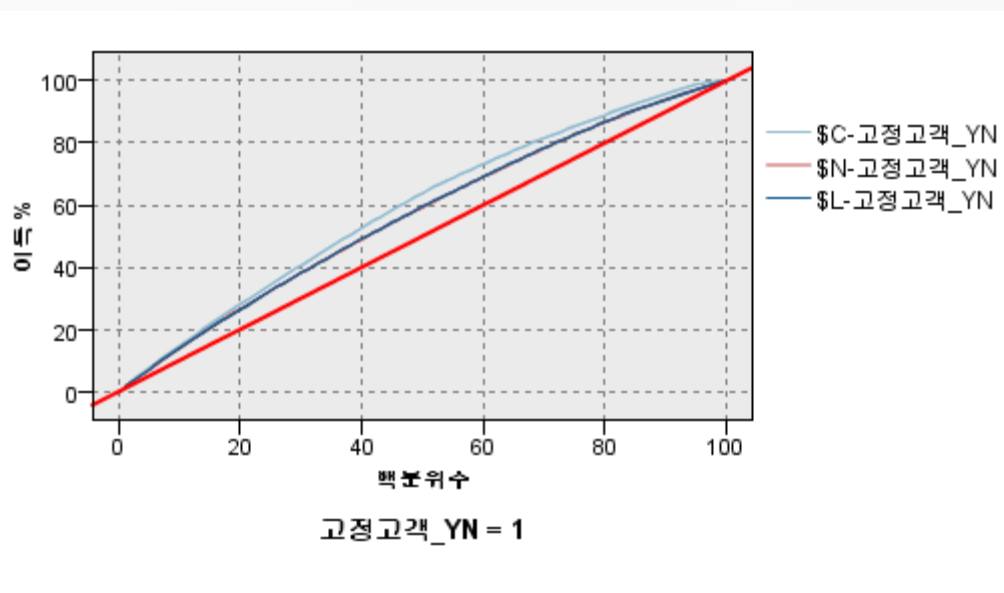
가지치기 강도 조정

가지치기 강도를 85로 늘려서 Over-fitting의 가능성을 적게 한다.

또한 의사결정 나무를 적절히 단순하게 하여 예측오차가 커지지 않도록 한다.

탐사과정

Gain-Chart



분석노드

출력 필드 고정고객_YN의 결과

개별 모델

\$C-고정고객_YN와(과) 고정고객_YN 비교

정확	582,034	67.43%
틀림	281,181	32.57%
합계	863,215	

\$N-고정고객_YN와(과) 고정고객_YN 비교

정확	525,720	60.9%
틀림	337,495	39.1%
합계	863,215	

\$L-고정고객_YN와(과) 고정고객_YN 비교

정확	533,166	61.77%
틀림	330,049	38.23%
합계	863,215	

\$C-고정고객_YN \$N-고정고객_YN \$L-고정고객_YN 간의 일치

일치	655,505	75.94%
불일치	207,710	24.06%
합계	863,215	

일치와(과) 고정고객_YN 비교

정확	447,731	68.3%
틀림	207,774	31.7%
합계	655,505	

의사결정나무 모형의 정확도

67.43%



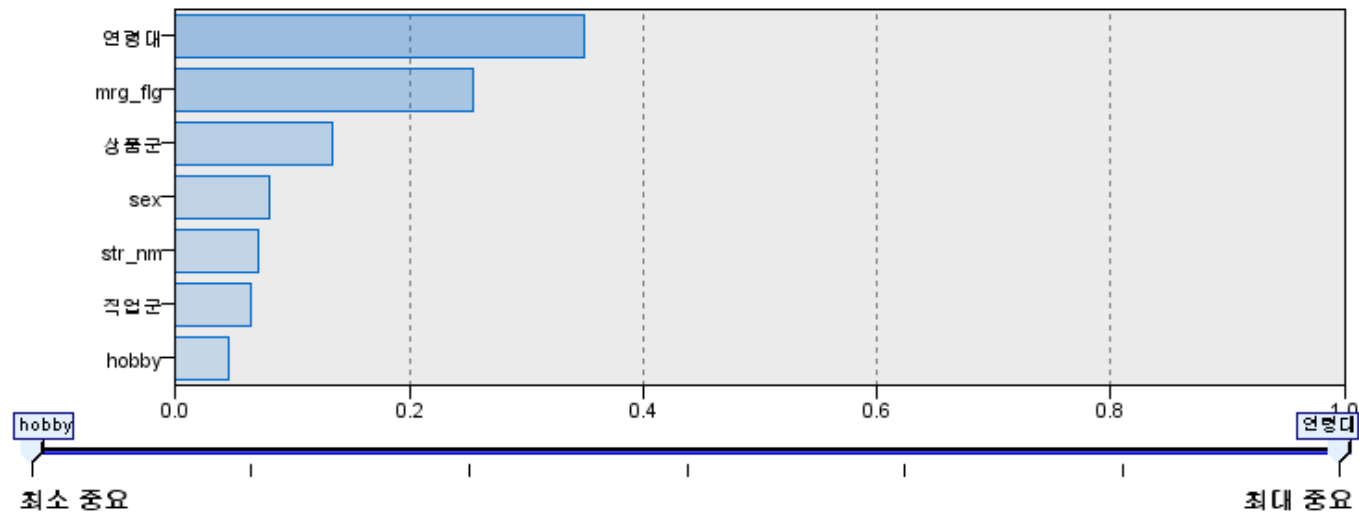
예측모델을

C5.0의 결과로 생긴 모델로 선택한다

결과

C5.0

예측자 중요도
목표값: 고정고객_YN

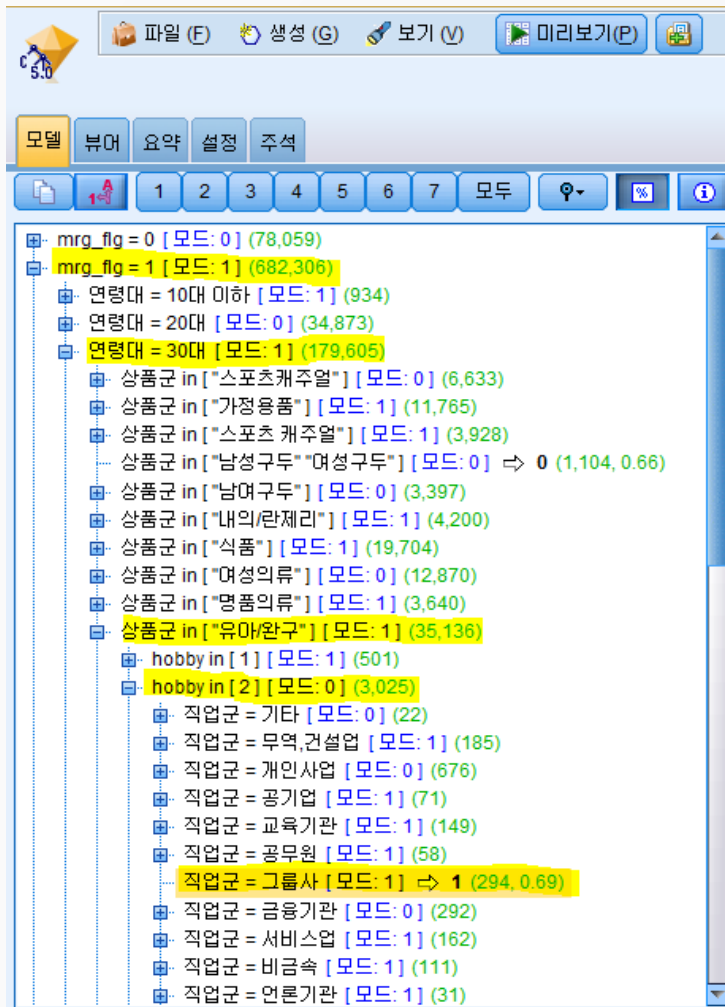


고정고객 여부의 변수 중요도

: 연령대 >> 결혼여부 > 상품군 >

성별 > 지점 > 직업군 > 취미

결과



- + 결혼여부: 기혼
- + 연령대: 30대
- + 유아/완구 상품군을 구매
- + 취미: 스포츠(취미코드=2)
- + 직업군: 그룹사 종사자

결과

```
mrg_flg = 0 [모드: 0] (78,112)
mrg_flg = 1 [모드: 1] (682,321)
  연령대 = 10대 이하 [모드: 1] (944)
  연령대 = 20대 [모드: 0] (34,744)
  연령대 = 30대 [모드: 1] (179,393)
    상품군 in ["스포츠캐주얼"] [모드: 0] (6,586)
    상품군 in ["가정용품"] [모드: 1] (11,724)
    상품군 in ["스포츠 캐주얼"] [모드: 1] (3,910)
    상품군 in ["남성구두" "여성구두"] [모드: 0] => 0 (1,105, 0.661)
    상품군 in ["남여구두"] [모드: 0] (3,390)
    상품군 in ["내의/란제리"] [모드: 0] (4,240)
    상품군 in ["식품"] [모드: 1] (19,694)
    상품군 in ["여성의류"] [모드: 0] (12,829)
    상품군 in ["명품의류"] [모드: 1] (3,624)
    상품군 in ["유아/완구"] [모드: 1] (35,100)
      hobby in [1] [모드: 1] (499)
      hobby in [2] [모드: 1] (3,007)
        직업군 = 기타 [모드: 0] (26)
        직업군 = 무역,건설업 [모드: 1] (188)
        직업군 = 개인사업 [모드: 0] (674)
          str_nm in ["무역점"] [모드: 1] (155)
            sex = 0 [모드: 1] => 1 (0)
            sex = 1 [모드: 1] => 1 (74, 0.784)
            sex = 2 [모드: 0] => 0 (81, 0.593)
          str_nm in ["본점"] [모드: 0] (46)
```

- + 결혼여부: 기혼
- + 연령대: 30대
- + 유아/완구 상품군을 구매
- + 취미: 스포츠(취미코드=2)
- + 직업군: 개인사업
- + 성별: 남성 (성별코드=1)

결론

- 고정고객이 될 가능성이 높은 고객들을 대상으로 세분화 된 마케팅을 실시할 수 있다.
- 향후 고정고객이 될 가능성이 있는 잠재적인 고객을 고정고객으로 끌어들이 수 있는 방안을 모색할 수 있다.

-
1. 기혼이고 30대인 고객들을 대상으로 스포츠 브랜드와 유아/완구 브랜드 DM을 적극적으로 발송한다.
 2. 두 상품군의 할인행사를 다른 상품군 보다 자주 진행한다.
 3. 백화점에서 스포츠 관련한 문화 행사를 열어 고객의 관심을 유도한다.

Thank

you 