

디지털경영론(001)

PBL1 상장법인의 매출액 예측

- CJ제일제당을 중심으로 -



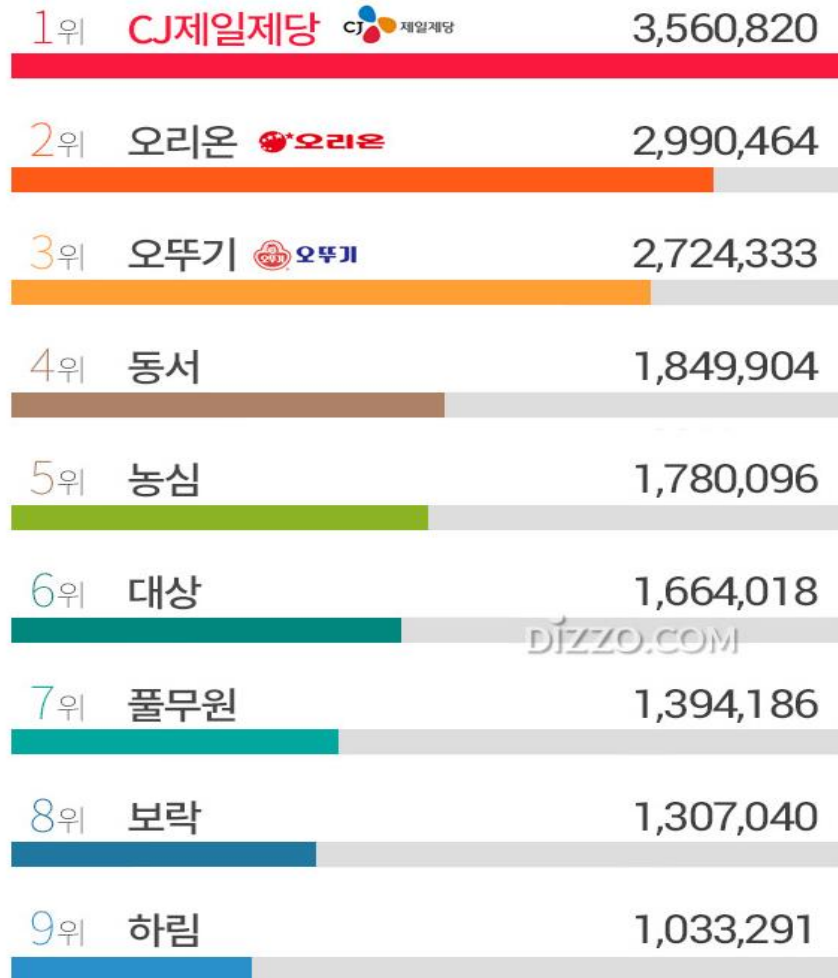
4조

16010057 전현지

16010088 민장태

16012993 이준희

기업 개요 - CJ제일제당



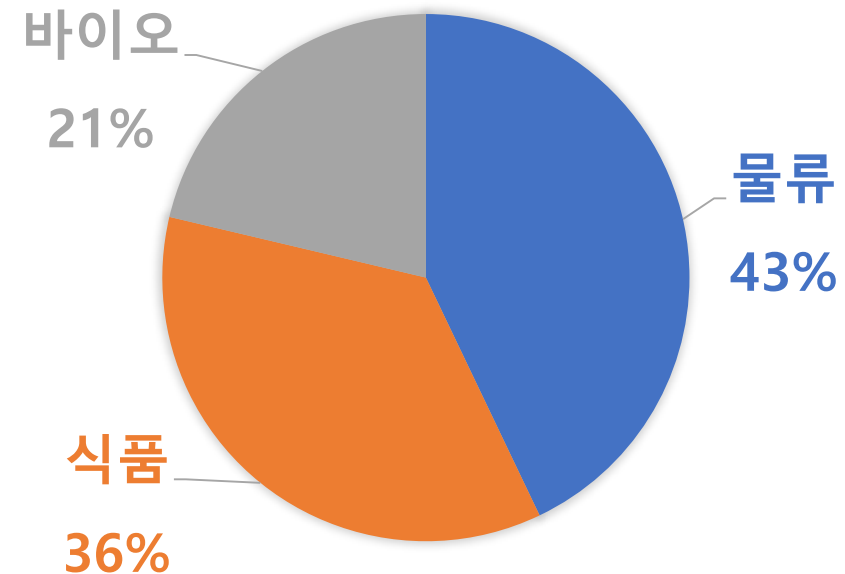
- CJ제일제당은 60여 년 동안 식품산업의 발전을 이끌어 온 국내 1위 식품회사
- 2007년부터 CJ 주식회사에서 기업 분할되어 식품, 생명공학에 집중하는 사업 회사로 새출발

주요 사업 개요

- CJ제일제당은 식품, 바이오, 물류를 주요 사업으로 하고 있음

구분	비중(%)	금액(억원)
물류	42.9	95,857
식품	35.8	80,105
바이오	21.3	47,563
계	100.0	223,525

2019년 사업별 매출 구조(%)



자료출처 : 금융감독원 전자공시시스템

주요 사업 개요 - ①식품

- 설탕, 밀가루, 식용유, 조미료, 장류, 육가공식품, 신선식품, 쌀가공식품, 냉동식품, HMR 등의 생산 · 판매
- CJ제일제당의 대표 식품 브랜드/상품



주요 사업 개요 - ②바이오



주요 사업 개요 - ③물류

- CL 사업부문
- 택배 사업부문
- 글로벌 사업부문
- 건설 사업부문



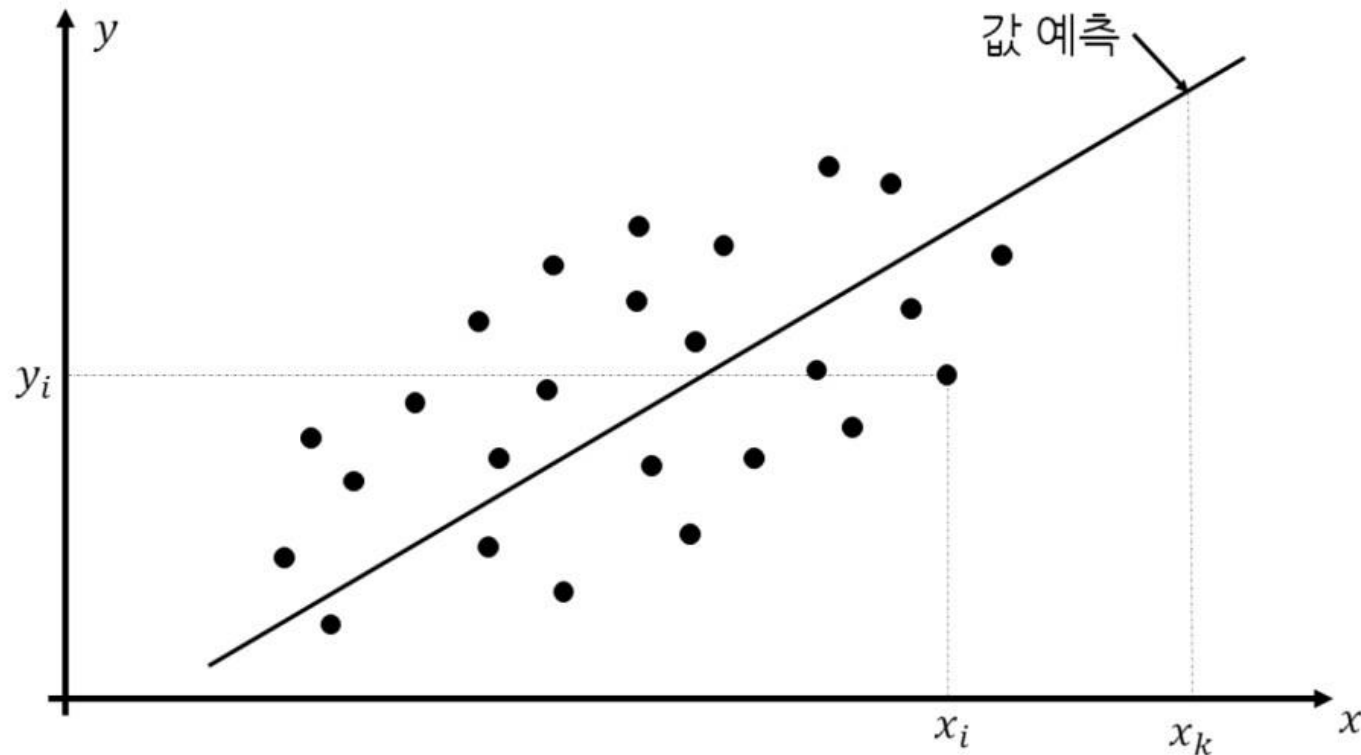
주요 사업 현황

- 사업부문별 현황(2020년 3월 말 기준, 단위 : 백만원)

사업부문	매출유형	품목	매출액(비율)
식품 사업부문	제품/상품	설탕, 밀가루, 식용유 등	2,260,560 (39%)
바이오 사업부문	제품/상품	아미노산 등	1,221,153 (21%)
물류 사업부문	서비스	운송, 하역, 건설 등	2,349,229 (40%)

선형회귀분석

데이터에서 입력변수 X 와 목표변수 Y 사이의 선형관계를 추정한 모형의 정확성 예측



- ☺ 장점
: 복잡하지 않고 다양한 문제에 적용 가능
- ☹ 단점
: 입력변수의 개수에 영향을 받음(오차 발생)

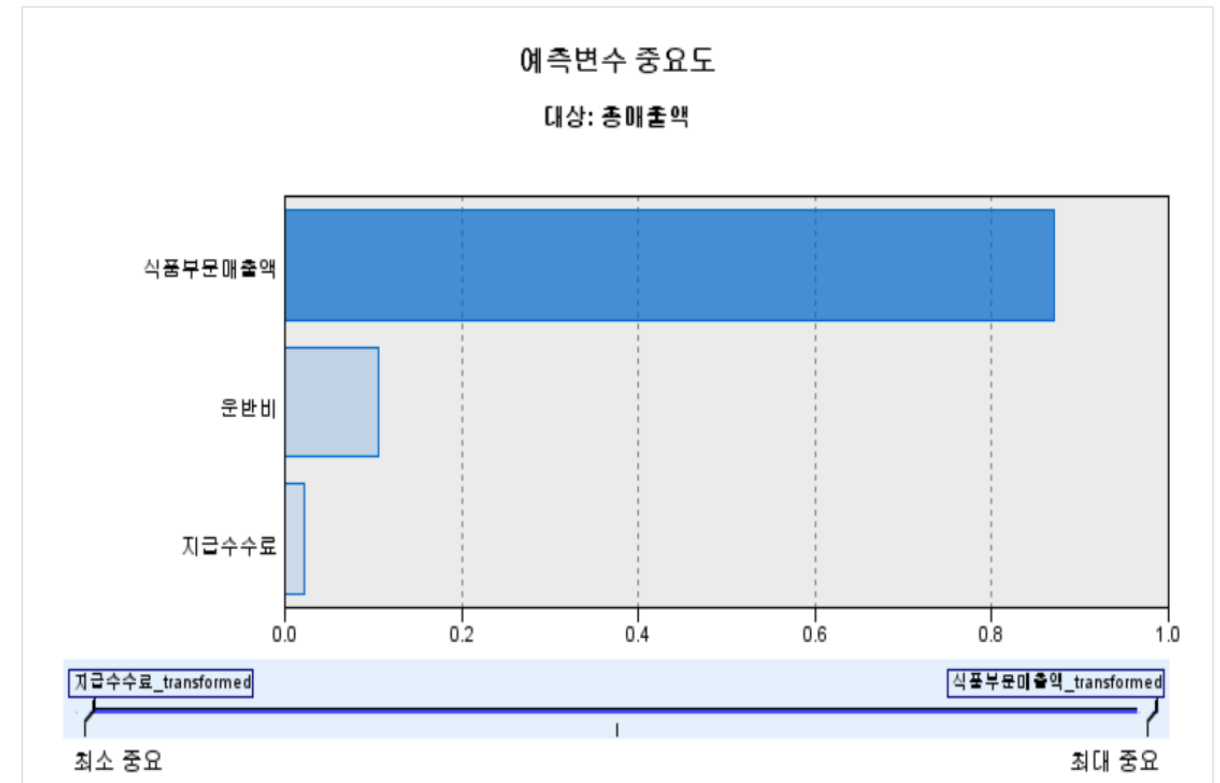
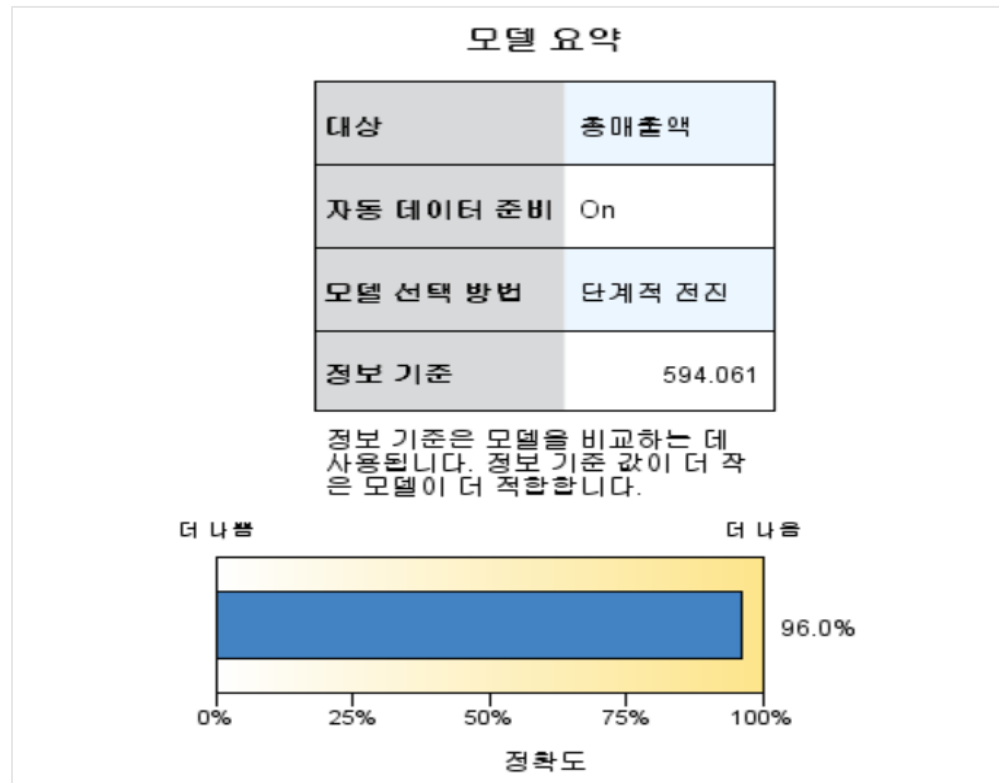
변수 설정

목표 변수 : 총매출액

필드	측정	값	결측값	검사	역할
시간	명목	"1Q17","2Q1...		없음	없음
식품부문매출액	연속형	[1323859.0,...		없음	입력
생명공학 매출액	연속형	[1075590.0,...		없음	입력
물류매출액	연속형	[1467085.0,...		없음	입력
식품매출액 비율	연속형	[0.31,0.34]		없음	입력
생명공학매출액...	연속형	[0.28,0.29]		없음	없음
물류매출액비율	연속형	[0.38,0.4]		없음	없음
단기차입금	연속형	[1.50325798...		없음	입력
자본총계	연속형	[5.90684177...		없음	입력
부채와 자본총계	연속형	[1.53905011...		없음	입력
자산총계	연속형	[1.53905011...		없음	없음
급여	연속형	[3.56738203...		없음	입력
무형자산상각비	연속형	[2.2366798E...		없음	입력
운반비	연속형	[2.73182555...		없음	입력
장비임차료	연속형	[0.0,4.32804...		없음	입력
지급수수료	연속형	[2.18515767...		없음	입력
총매출액	연속형	[1.48298323...		없음	목표

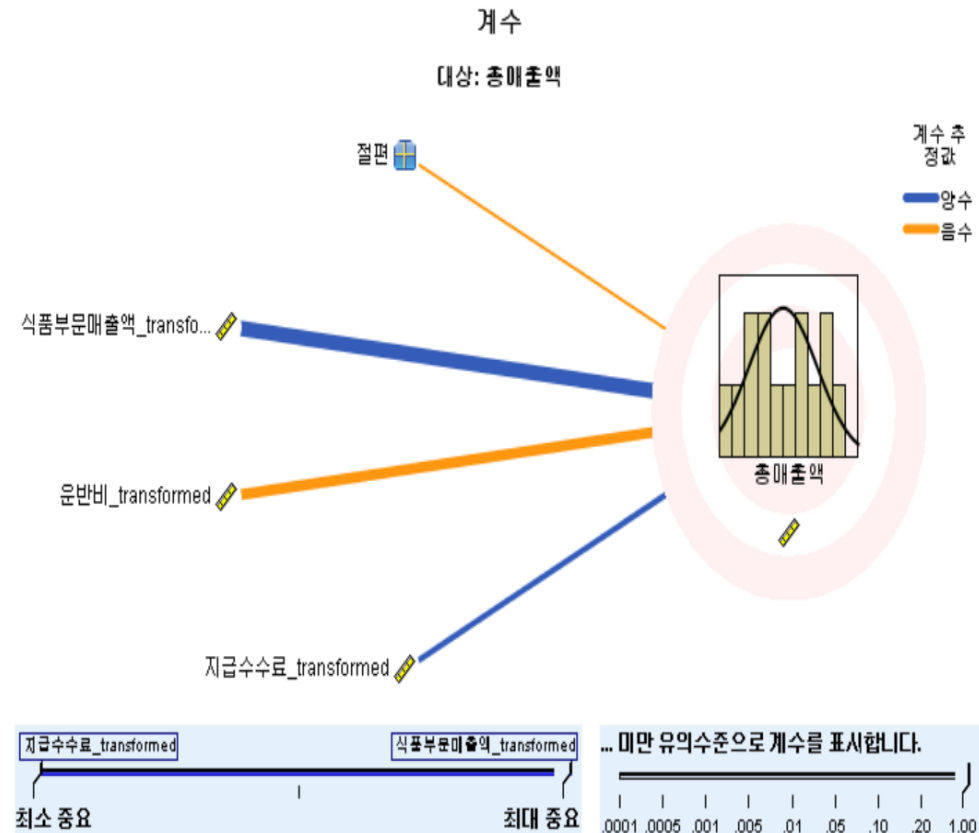
분석 결과

- 모델의 정확도가 96%로 높게 나타남
- 예측변수 중요도 식품부문매출액 > 운반비 > 지급수수료 순으로 높음

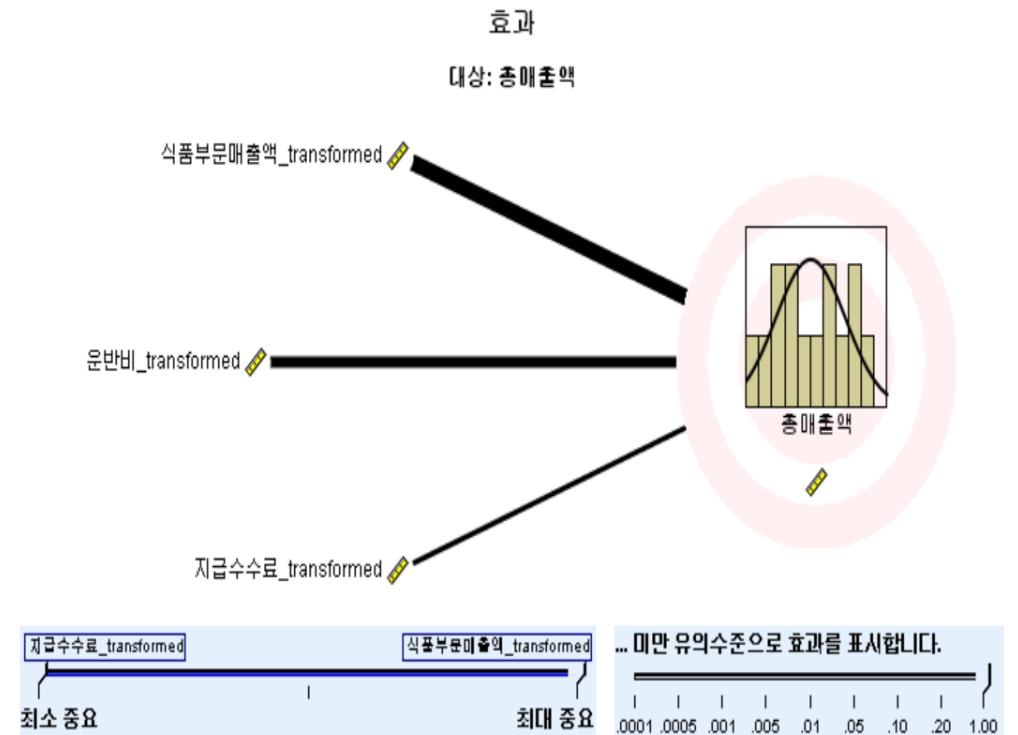


분석 결과

- 계수 추정값 → 식품매출액(양의 선형관계)



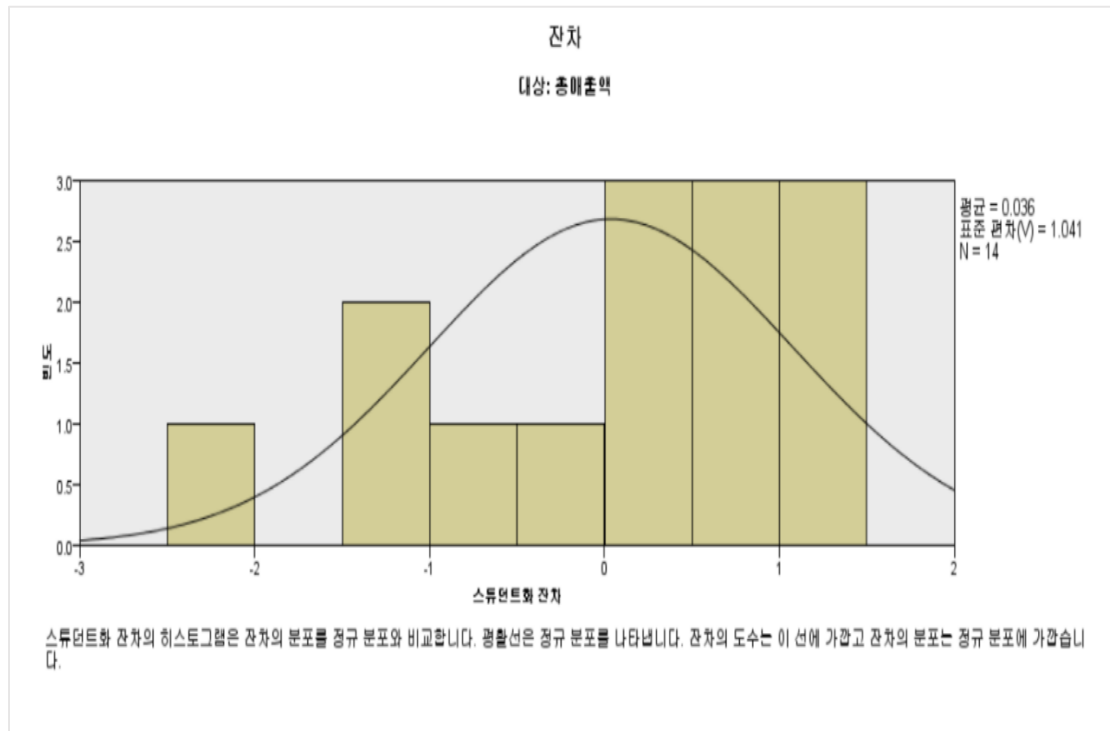
- 유의수준 → 식품매출액(전 구간에 해당)



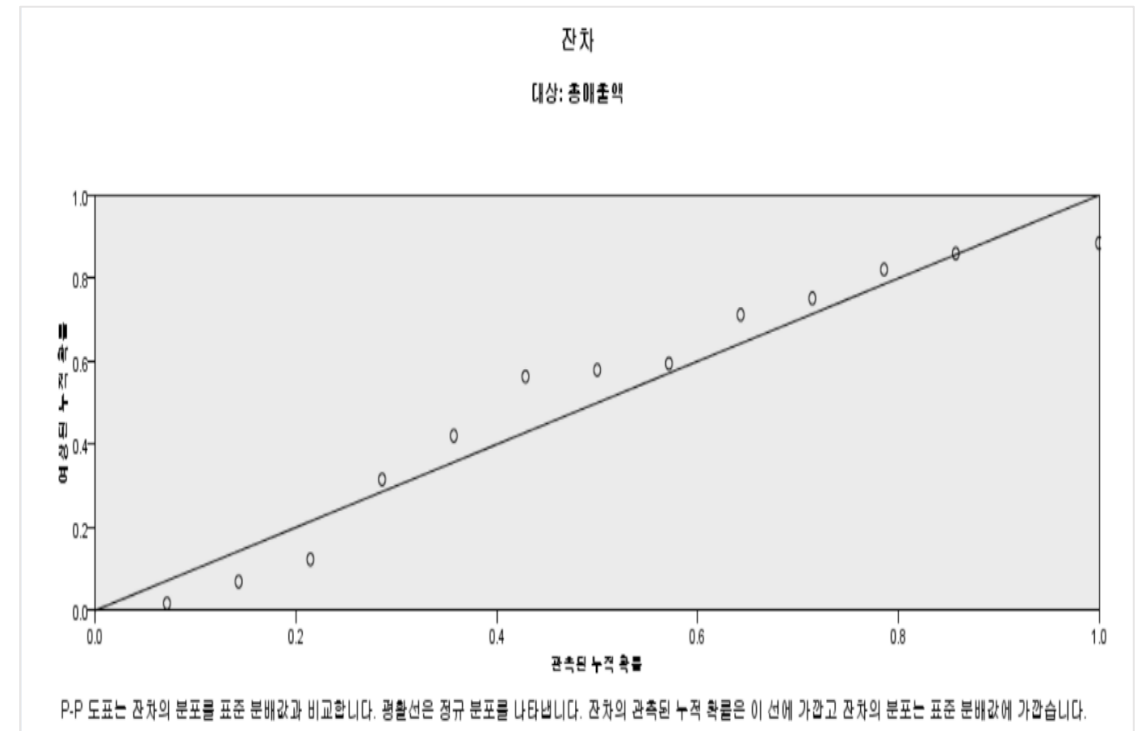
분석 결과

• 잔차의 시각화

- 히스토그램
- → 입력변수의 중요도의 크기에 따른 분포



- P-P도표
- → 직선과 오차(원)의 분포의 차이



분석 결과

- 총매출액과 예상매출액(\$L)의 비교
 - 높은 선형 상관관계를 통한 모형의 예측 정확성 근거 제시

출력 필드 총매출액의 결과

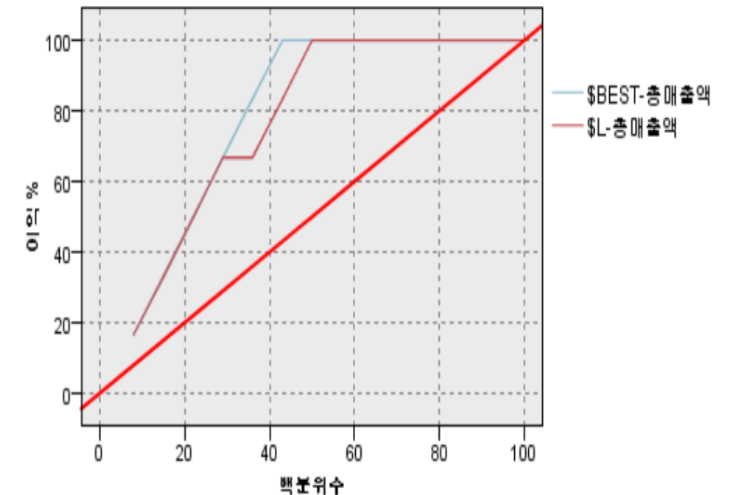
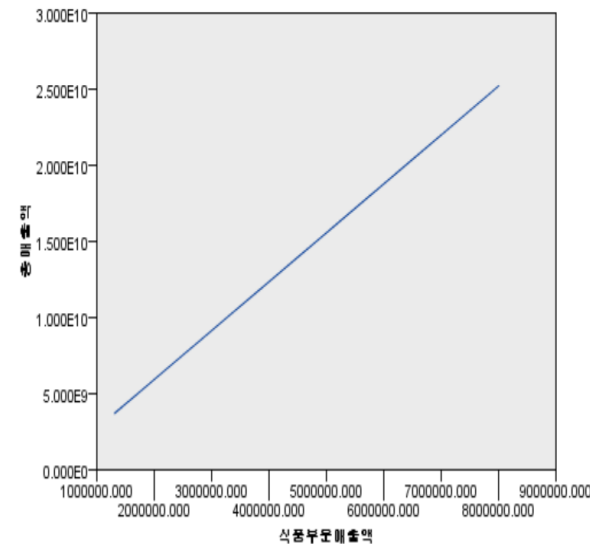
\$L-총매출액과(와) 총매출액 비교

최소 오류	-2351097264.239
최대 오류	1383518670.319
평균 오류	0.0
평균 절대 오차	839452477.177
표준 편차	1089599225.924
선형 상관관계	0.984
발생	14

추정 평균

대상: 총매출액

상위 10개 유의 효과($p < .05$)에 대한 추정 평균 차트가 표시되었습니다.



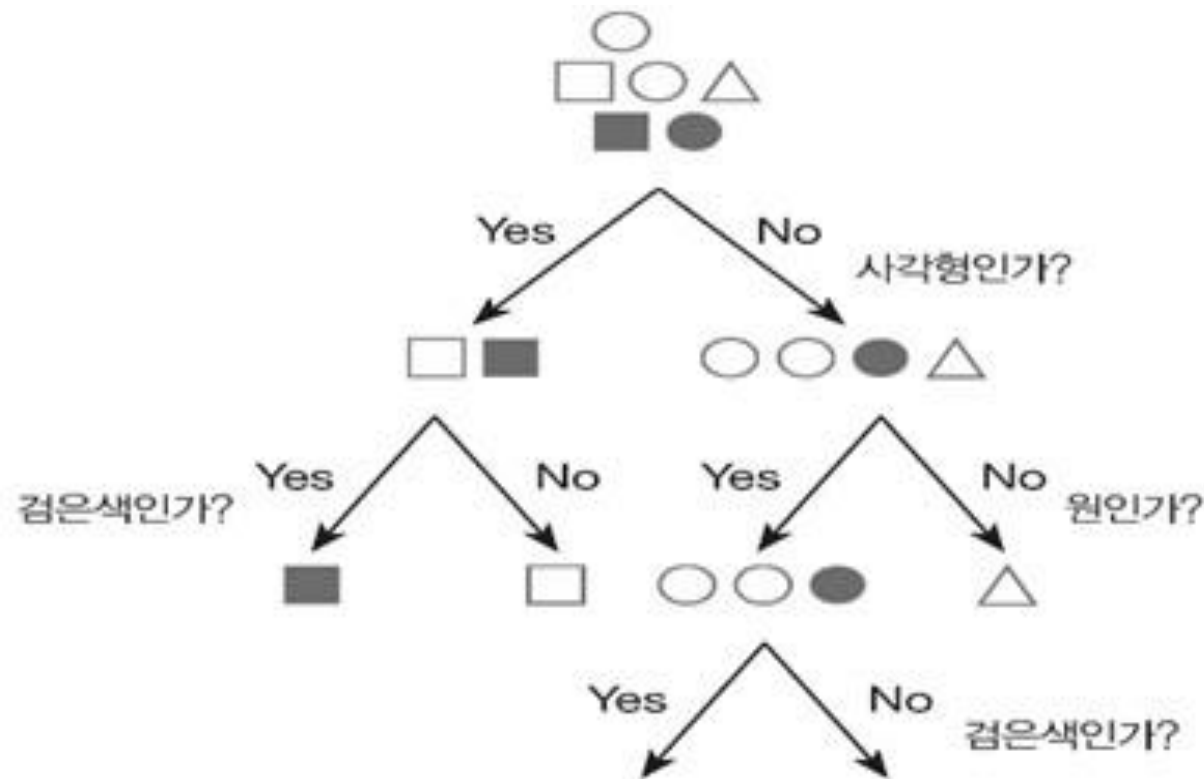
총매출액 > 1.1917722765E10

- 해석의 한계

2Q20	오차	총매출액	\$L-총매출액
선형회귀	11.65%	11,751,821,403	13,120,525,494.91

의사결정나무

데이터 속에서 분류대상이 되는 관측치들 간의 규칙을 통해서 분류대상들을 n 개의 소집단으로 분류하는 규칙을 나무구조로 나타내서 목표변수를 예측

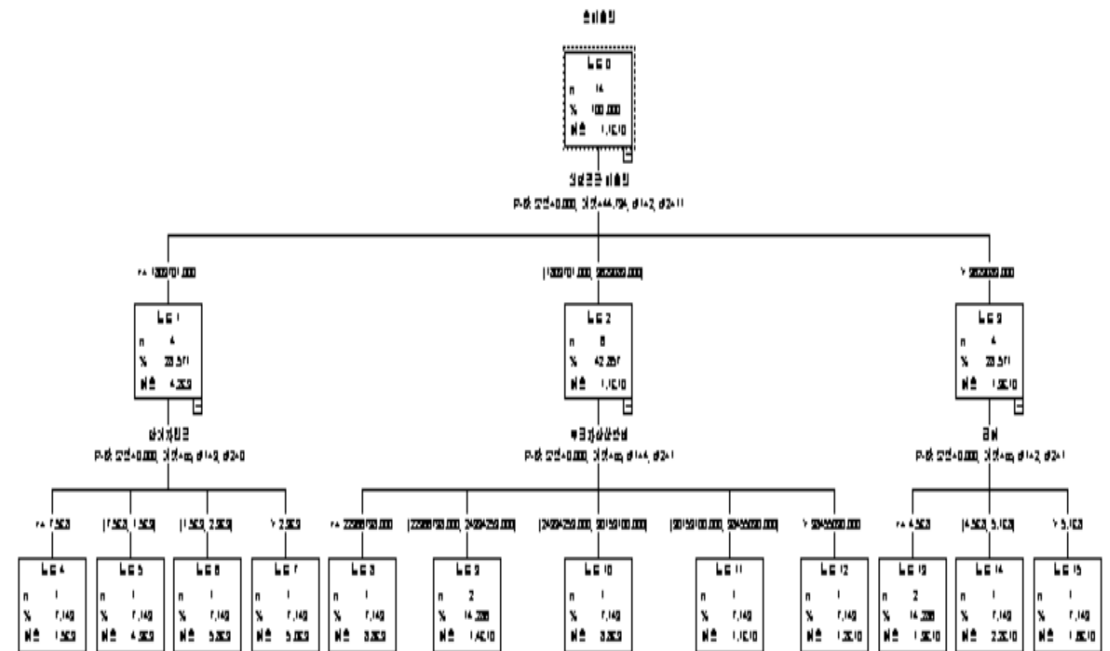
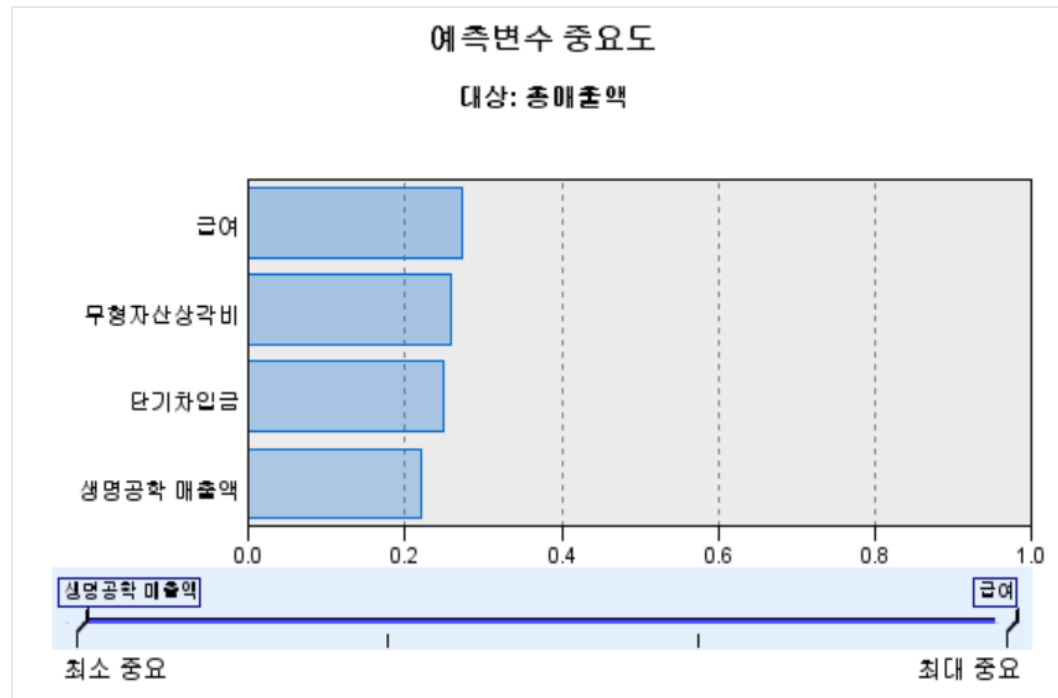


☺ **장점**
: 시각적으로 쉬운 이해
높은 예측 정확도

☹ **단점**
: 과대적합 발생 위험
표본의 크기에 민감

CHAID

- 변수들의 속성 → 연속형
- 예측변수 중요도 급여 > 무형자산상각비 > 단기차입금 > 생명공학매출액 순으로 높음



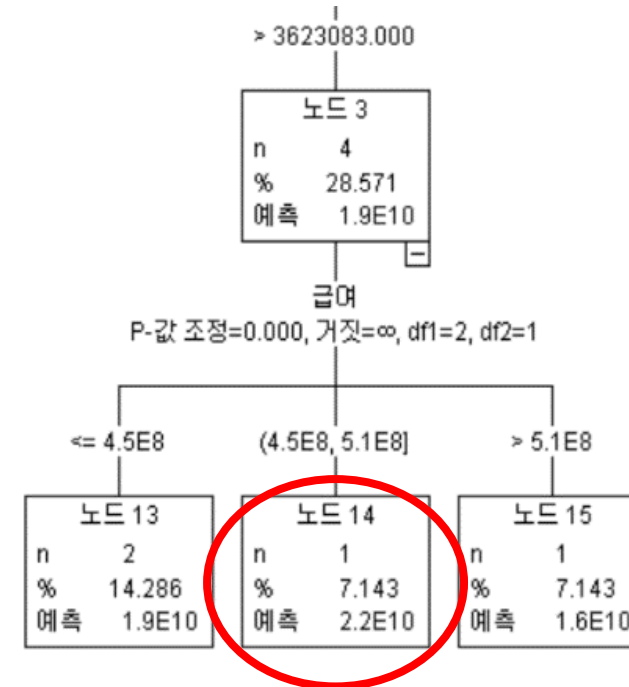
CHAID

- 다지분리를 통한 예상매출액의 최대값에 해당하는 노드 선택
:생명공학매출액(1수준) → 급여(2수준)
- 실제 총매출액과의 비교

생명공학매출액 <= 1,809,701 [평균: 4,170,066,697.25, 효과: -7,269,176,609.036]
단기차입금 <= 754,006,634 [평균: 1,482,983,234, 효과: -2,687,083,463.25] ⇒ 1482983234.0
단기차입금 > 754,006,634 and 단기차입금 <= 1,503,257,981 [평균: 4,348,588,146, 효과: 178,521,448.75] ⇒ 4348588146.0
단기차입금 > 1,503,257,981 and 단기차입금 <= 1,951,127,690 [평균: 5,830,942,047, 효과: 1,660,875,349.75] ⇒ 5830942047.0
단기차입금 > 1,951,127,690 [평균: 5,017,753,362, 효과: 847,686,664.75] ⇒ 5017753362.0
생명공학매출액 > 1,809,701 and 생명공학매출액 <= 3,623,083 [평균: 11,230,897,068.5, 효과: -208,346,237.786]
무형자산상각비 <= 22,366,798 [평균: 8,802,336,934, 효과: -2,428,560,134.5] ⇒ 8802336934.0
무형자산상각비 > 22,366,798 and 무형자산상각비 <= 24,334,259 [평균: 13,747,911,781, 효과: 2,517,014,712.5] ⇒ 13747911781.0
무형자산상각비 > 24,334,259 and 무형자산상각비 <= 30,159,100 [평균: 8,802,336,934, 효과: -2,428,560,134.5] ⇒ 8802336934.0
무형자산상각비 > 30,159,100 and 무형자산상각비 <= 38,455,030 [평균: 10,533,063,578, 효과: -697,833,490.5] ⇒ 10533063578.0
무형자산상각비 > 38,455,030 [평균: 11,751,821,403, 효과: 520,924,334.5] ⇒ 11751821403.0
생명공학매출액 > 3,623,083 [평균: 19,020,939,272, 효과: 7,581,695,965.714]
급여 <= 445,425,004 [평균: 18,670,059,989, 효과: -350,879,283] ⇒ 18670059989.0
급여 > 445,425,004 and 급여 <= 511,666,970 [평균: 22,352,462,296, 효과: 3,331,523,024] ⇒ 22352462296.0
급여 > 511,666,970 [평균: 16,391,174,814, 효과: -2,629,764,458] ⇒ 16391174814.0

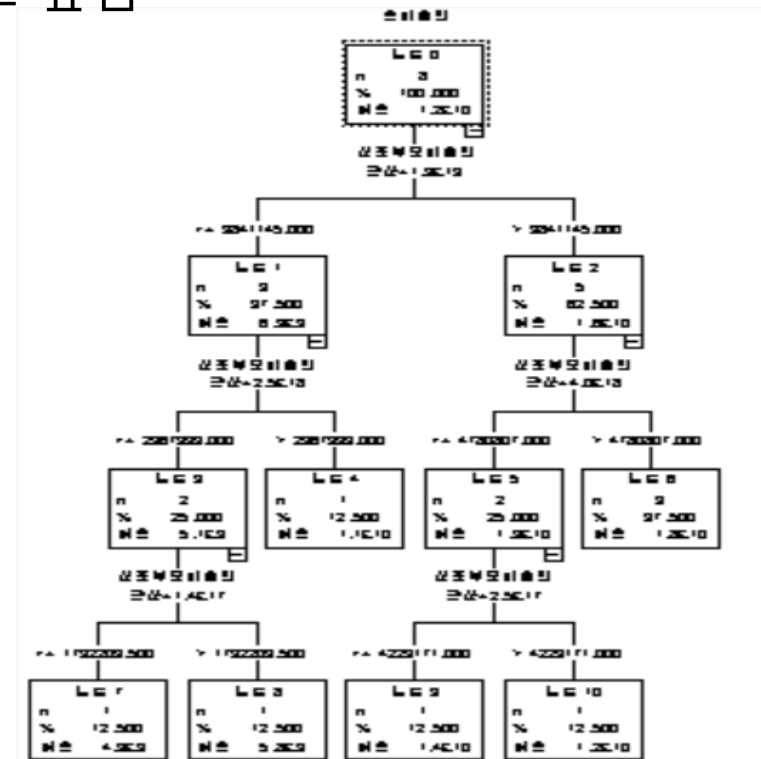
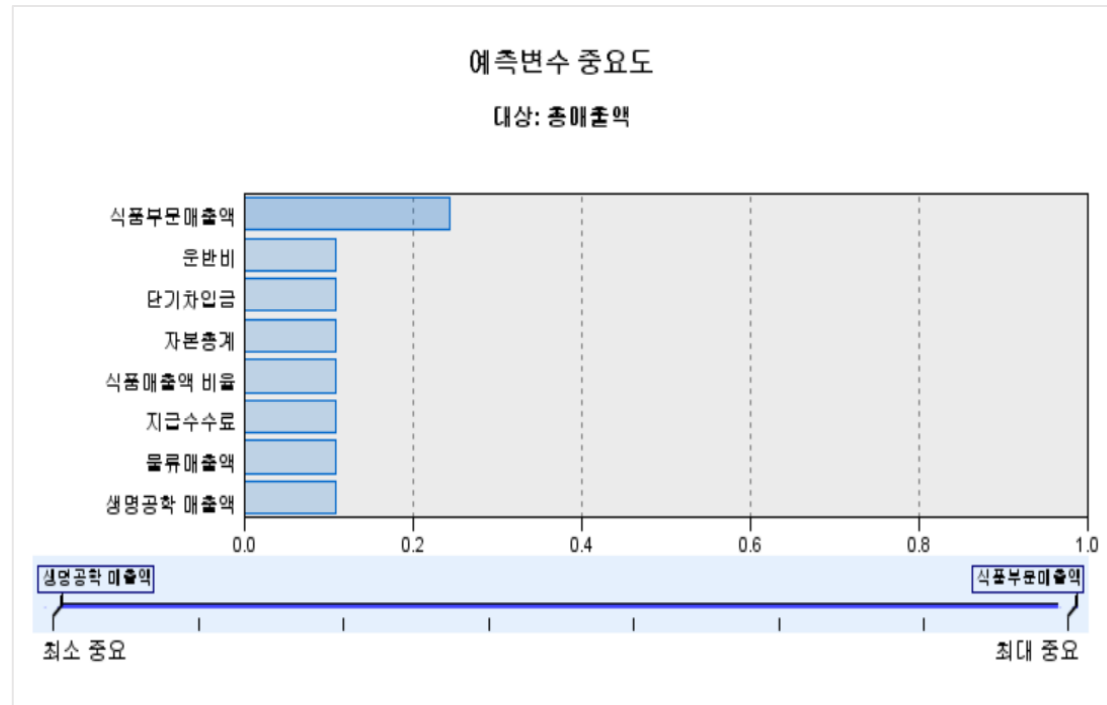
수준	조건
1	생명공학매출액 > 3623083
2	급여 > 445425004 and 급여 <= 511666970

대응



CART

- 변수들의 속성 → 연속형
- 예측변수 중요도 식품부문매출액 > 운반비 > 단기차입금 > 자본총계 > 식품매출액 비율 > 지급수수료 > 물류매출액 > 생명공학 매출액 순으로 높음

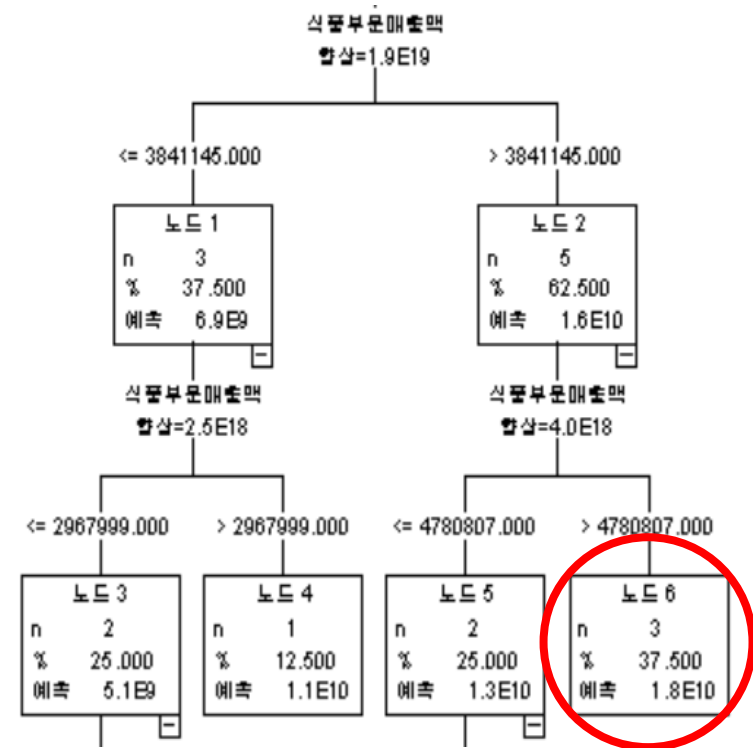


CART

- 이진분리를 통한 예상매출액의 최대값에 해당하는 노드 선택
:식품공학매출액의 금액 범위로 모든 수준에서 분리
- 실제 총매출액과의 비교

식품부문매출액 <= 3,841,145 [평균: 6,904,197,923.667, 효과: -5,588,754,794.708]		
<ul style="list-style-type: none"> 식품부문매출액 <= 2,967,999 [평균: 5,089,765,096.5, 효과: -1,814,432,827.167] <ul style="list-style-type: none"> 식품부문매출액 <= 1792209.500 [평균: 4,348,588,146, 효과: -741,176,950.5] ⇒ 4348588146.0 식품부문매출액 > 1792209.500 [평균: 5,830,942,047, 효과: 741,176,950.5] ⇒ 5830942047.0 식품부문매출액 > 2,967,999 [평균: 10,533,063,578, 효과: 3,628,865,654.333] ⇒ 10533063578.0 		
<ul style="list-style-type: none"> 식품부문매출액 > 3,841,145 [평균: 15,846,205,595.2, 효과: 3,353,252,876.825] <ul style="list-style-type: none"> 식품부문매출액 <= 4,780,807 [평균: 12,749,866,592, 효과: -3,096,339,003.2] <ul style="list-style-type: none"> 식품부문매출액 <= 4,229,171 [평균: 13,747,911,781, 효과: 998,045,189] ⇒ 13747911781.0 식품부문매출액 > 4,229,171 [평균: 11,751,821,403, 효과: -998,045,189] ⇒ 11751821403.0 식품부문매출액 > 4,780,807 [평균: 17,910,431,597.333, 효과: 2,064,226,002.133] ⇒ 17910431597.333 		

대응	히스토리	수준	조건
		1	식품부문매출액 > 3841145
		2	식품부문매출액 > 4780807
빈도			



CHAID & CART 분석 결과 비교

- 총매출액과 예상매출액(\$R,\$R1)의 비교
 - 높은 선형 상관관계를 통한 모형의 예측 정확성 근거 제시
- [CHAID > CART] → CHAID의 정확성 & 적합성
 - CART의 불안정한 예측 오차

출력 필드 총매출액의 결과

개별 모델

\$R-총매출액과(와) 총매출액 비교

최소 오류	0.0
최대 오류	0.0
평균 오류	0.0
평균 절대 오차	0.0
표준 편차	0.0
선형 상관관계	1.0
발생	14

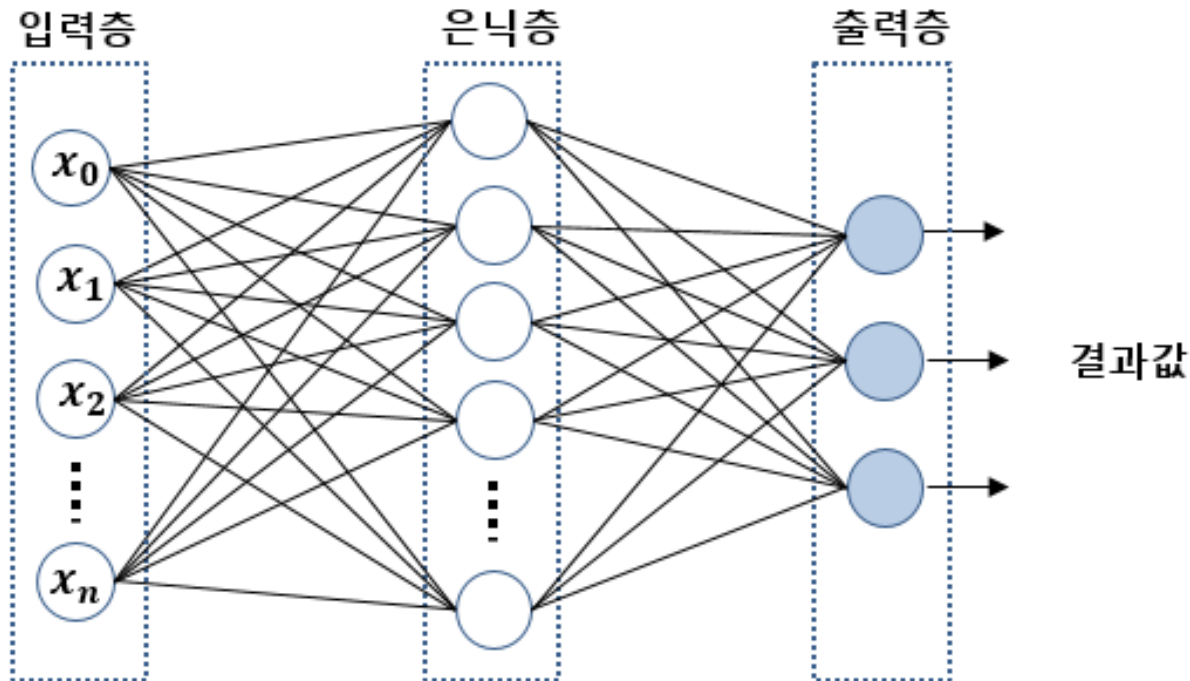
\$R1-총매출액과(와) 총매출액 비교

최소 오류	-2865604912.0
최대 오류	4442030698.667
평균 오류	584884341.19
평균 절대 오차	1211293154.81
표준 편차	1854774328.787
선형 상관관계	0.954
발생	14

총매출액	\$R-총매출액	\$R1-총매출액
4348588146.000	4348588146.000	4348588146.000
8802336934.000	8802336934.000	5830942047.000
13747911781.000	13747911781.000	13747911781.000
18670059989.000	18670059989.000	17910431597.333
1482983234.000	1482983234.000	4348588146.000
8802336934.000	8802336934.000	5830942047.000
13747911781.000	13747911781.000	13747911781.000
18670059989.000	18670059989.000	17910431597.333
5017753362.000	5017753362.000	4348588146.000
10533063578.000	10533063578.000	10533063578.000
16391174814.000	16391174814.000	17910431597.333
22352462296.000	22352462296.000	17910431597.333
5830942047.000	5830942047.000	5830942047.000
11751821403.000	11751821403.000	11751821403.000

인공신경망 분석

데이터에서 알 수 없고 복잡할 수 있는 패턴을 찾아 하나 이상의 예측변수를 기반으로 연속 또는 범주형 대상을 예측



☺ **장점**
: 강력한 예측 성능

☹ **단점**
: Black Box
(결과해석 어려움)

데이터 파티션

훈련 데이터 : 검정 데이터 = 50 : 50 비율로 파티션

파티션

생성(G) 미리보기(P)

설정 주석

파티션 필드: 파티션

파티션: ☒ 학습 및 테스트(T) ☐ 학습, 테스트 및 검증(V)

트레이닝 파티션 크기: 50 레이블: 훈련 값 = "1_훈련"

테스트 파티션 크기: 50 레이블: 검정 값 = "2_검정"

확인 파티션 크기: 0 레이블: 확인 값 = "3_확인"

총 크기: 100%

값: ☐ 시스템 정의 값 사용("1", "2" 및 "3")
☒ 시스템 정의 값에 레이블 추가
☐ 레이블을 값으로 사용

☒ 반복 가능한 파티션 할당

시드: 1234567 생성

☐ 고유 필드를 사용하여 파티션 지정:

확인 취소 적용(A) 재설정(R)

테이블 (21개 필드, 14개 레코드)

파일(F) 편집(E) 생성(G)

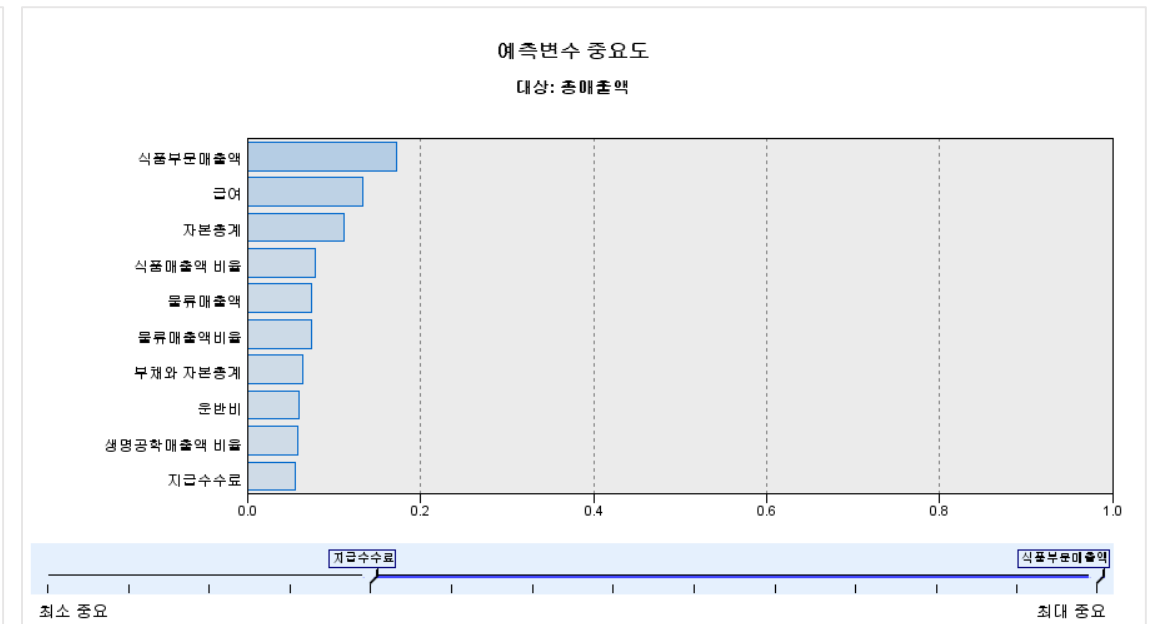
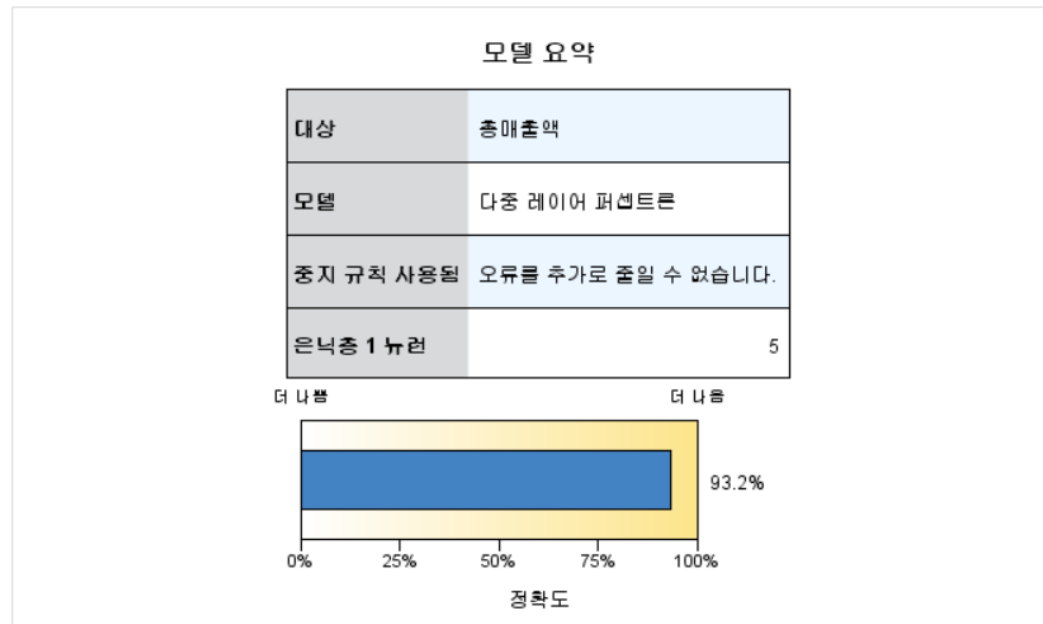
테이블 주석

		장비임차료	지급수수료	총매출액	C18	C19	C20	파티션
1	000	343181427.000	218515767.0...	4348588146...	\$n...	\$n...	\$n...	1_훈련
2	000	369230803.000	226689352.0...	8802336934...	\$n...	\$n...	\$n...	1_훈련
3	000	432804433.000	237103384.0...	1374791178...	\$n...	\$n...	\$n...	2_검정
4	000	0.000	470080726.0...	1867005998...	\$n...	\$n...	\$n...	2_검정
5	000	404384954.000	211819684.0...	1482983234...	\$n...	\$n...	\$n...	1_훈련
6	0.0...	447248519.000	198328681.0...	8802336934...	\$n...	\$n...	\$n...	1_훈련
7	1.0...	376259647.000	265260819.0...	1374791178...	\$n...	\$n...	\$n...	2_검정
8	000	1400540223.0...	487377532.0...	1867005998...	\$n...	\$n...	\$n...	1_훈련
9	0.0...	0.000	303251365.0...	5017753362...	\$n...	\$n...	\$n...	1_훈련
10	0.0...	0.000	176671689.0...	1053306357...	\$n...	\$n...	\$n...	1_훈련
11	0.0...	0.000	291374493.0...	1639117481...	\$n...	\$n...	\$n...	2_검정
12	1.0...	0.000	1061040894...	2235246229...	\$n...	\$n...	\$n...	2_검정
13	2.0...	0.000	273204165.0...	5830942047...	\$n...	\$n...	\$n...	2_검정
14	2.0...	\$null\$	286672974.0...	1175182140...	\$n...	\$n...	\$n...	1_훈련

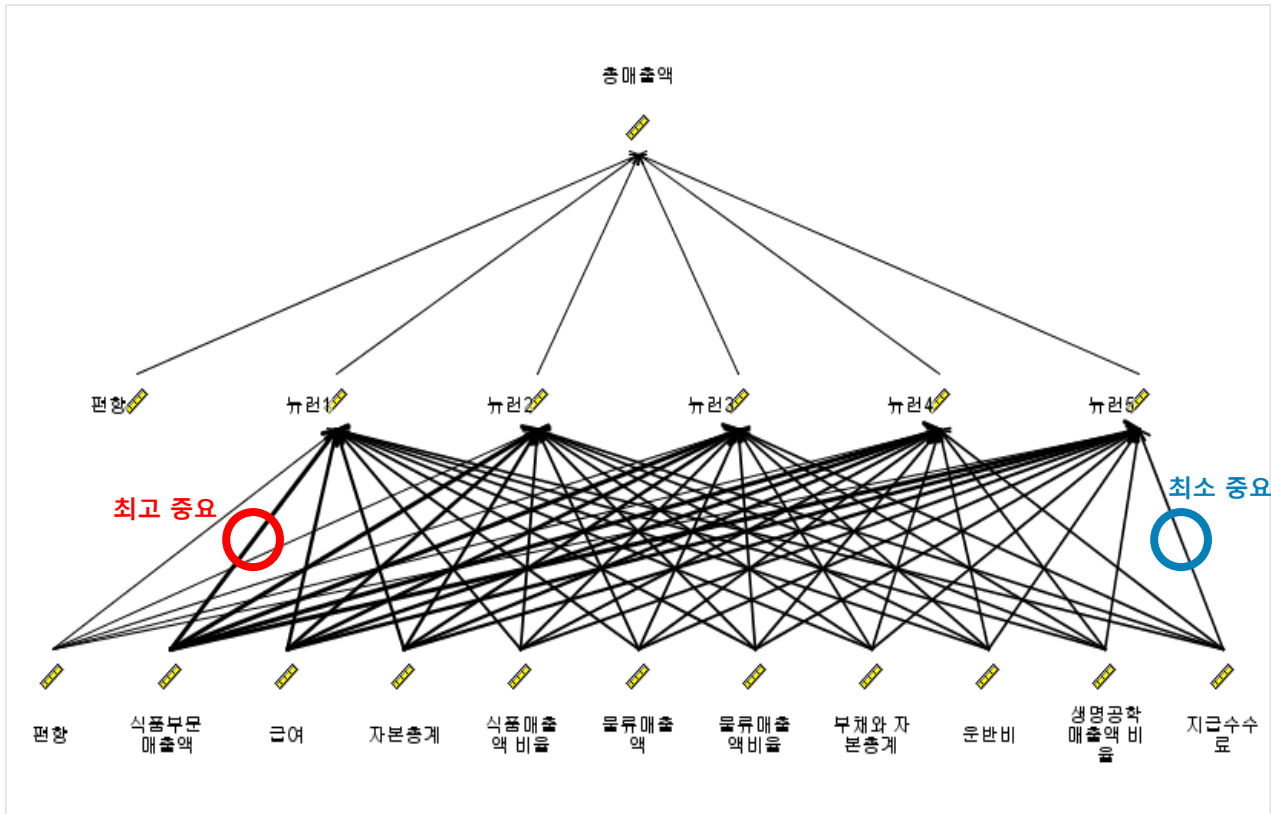
확인

분석 결과

- 모델의 정확도가 93.2%로 높게 나타남
- 예측변수 중요도 식품부문매출액 > 급여 > 자본총계 순으로 높음



분석 결과



- 은닉층 1 뉴런의 수 = 5개
- 선이 두꺼울수록 높은 중요도를 가짐
- 최고 중요 : 식품부문매출액
- 최소 중요 : 지급수수료

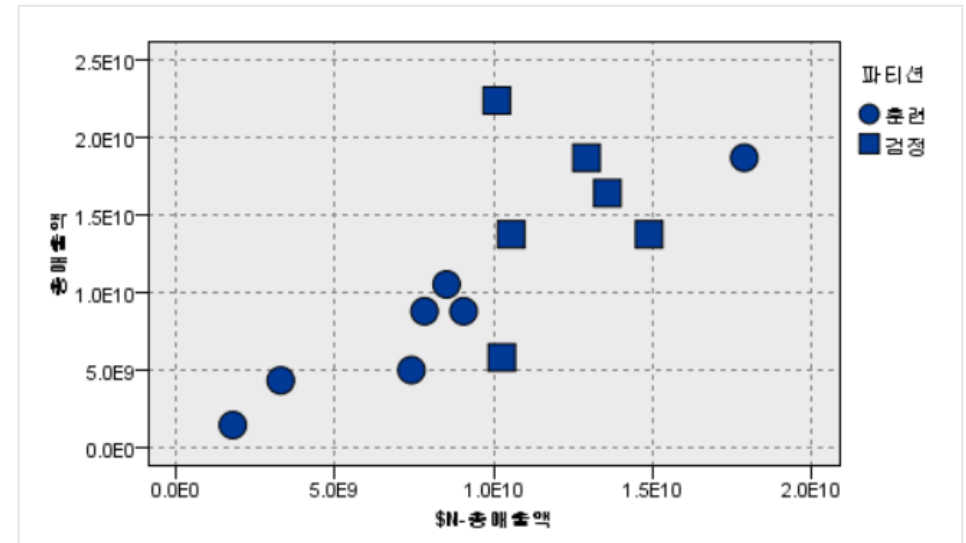
분석 결과 - 학습/검정데이터 비교

- 검정데이터보다 훈련데이터에서 더 작은 오차 및 표준편차 보임
→ 데이터 크기의 한계로 인한 과적합 현상

출력 필드 총매출액의 결과

\$N-총매출액과(와) 총매출액 비교

'파티션'	1_훈련	2_검정
최소 오류	-2388601483.053	-4418549935.711
최대 오류	2024864556.07	12264219881.919
평균 오류	273741001.443	3084091527.162
평균 절대 오차	1113257817.573	4933289718.888
표준 편차	1422321155.091	5753988822.618
선형 상관관계	0.968	0.101
발생	8	6



결론 및 한계

- 분석 결과 비교

	선행회귀분석	의사결정나무 (CART, CHAID)	인공신경망분석
모델의 정확도	<u>96.0%</u>	-	93.2%
주요예측변수	<u>식품부문매출액</u> 운반비 지급수수료	생명공학매출액 단기차입금 무형자산상각비 급여	<u>식품부문매출액</u> 급여 자본총계

➡ 적극적인 식품 R&D 및 신제품 개발 등의 활동을 통해
식품시장에서 우위를 점하는 것이 중요

결론 및 한계

- 한계점

1. 데이터 크기의 한계

- 2017 1Q~2020 2Q (총 14개 데이터)
- 예측의 정확도 감소, 데이터 과적합 등의 문제 발생 우려

2. 데이터 다양성의 한계

- 재무상태표, 손익계산서

감사합니다 😊

