

Genffice 전문업무APP · 데모 · 2026년 10월

# 데이터 예측 분석

- 관능 · 성분 데이터가 담긴 엑셀 하나를 올리면
- 분포 점검 · 전처리 · 차이검정 · 마커까지,
- 맛 예측식과 제품 단위 교차검증으로

예시 데이터 김치 32제품 × 숙성 3단계 · 60변수



— 무엇을 푸나

## 이 맛은 어떤 성분으로 설명되고, 성분만으로 얼마나 정확히 예측될까

1

### 엑셀이 분석 준비가 안 되어 있다

병합된 머리글, 단위만 다른 중복 열,  
'미검출'과 실제 0이 섞인 값을 사람이  
먼저 찾아야 합니다

2

### 분석 단계가 흩어져 있다

분포 확인, 변환, ANOVA, LDA, 상관 +  
FDR, LASSO, 교차검증을 각기 다른  
도구로 돌리고 다시 모읍니다

3

### 성능이 부풀려지기 쉽다

같은 제품의 다른 시점이 학습과 검증에  
함께 들어가면 예측 성능이 실제보다  
좋게 나옵니다

— 질문을 단계로

## 큰 질문 하나를 다섯 개의 작은 질문으로 나눠 차례로 답합니다

① 데이터는 분석할 준비가 됐나

불러오기 · 분포

② 어떤 변수를 어떻게 다듬어야 하나

전처리

③ 숙성 단계에 따라 무엇이 달라지나

차이검정

④ 맛과 함께 움직이는 성분은 무엇인가

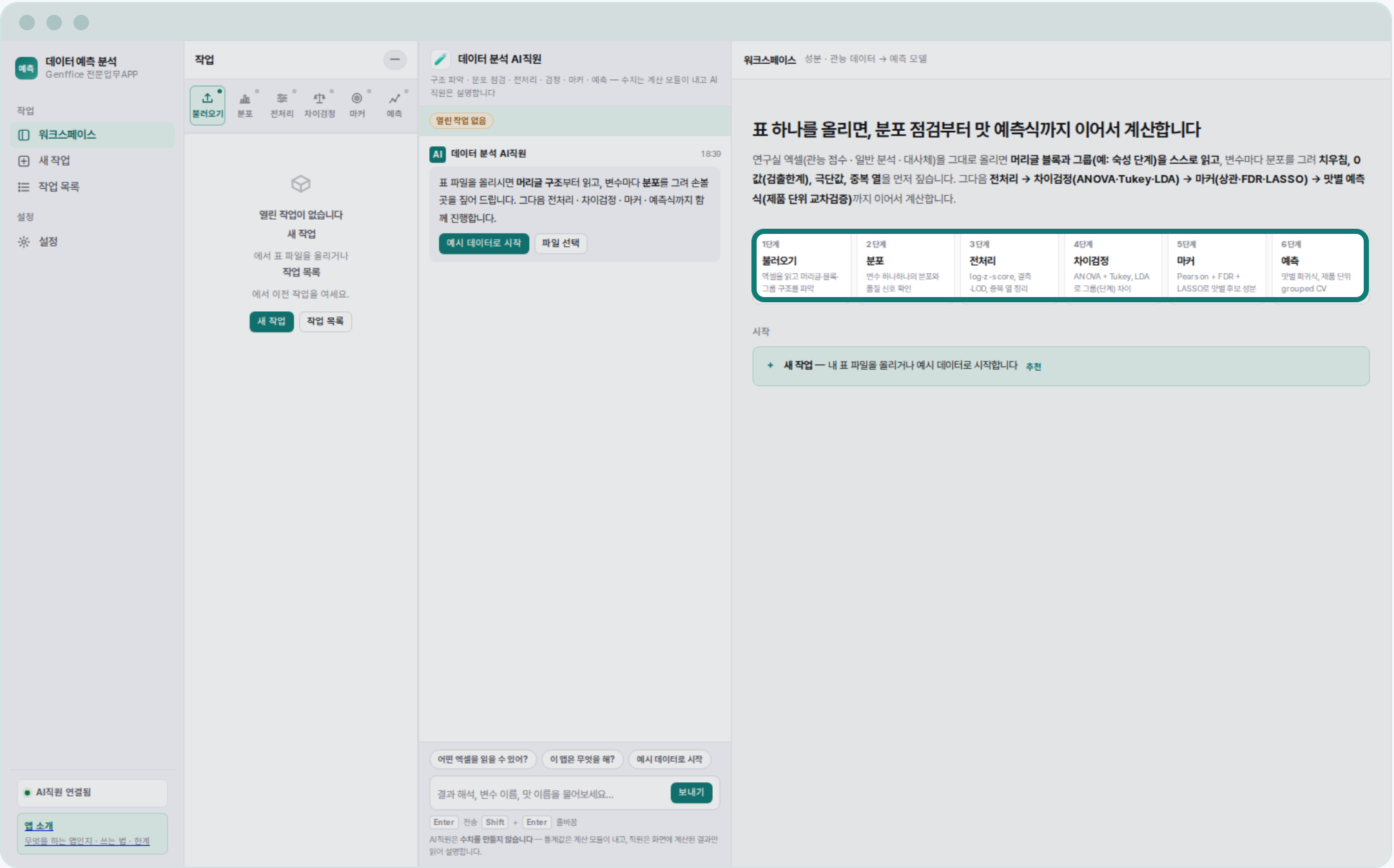
마커

⑤ 성분만으로 맛을 얼마나 맞히나

예측

— 화면 구성

한 화면에서 단계 · AI직원 · 그림과 표를 함께 봅니다



왼쪽

메뉴 · 단계 버튼 · 변수 목록

가운데

AI직원 — 진행과 질문

오른쪽

그림 · 표 · 결과

시작

예시 데이터로 바로 시작

## — 여섯 단계

단계 버튼을 차례로 누르면 결과가 다음 단계로 이어집니다

1

### 불러오기

머리글 · 그룹 ·  
시료명 인식

2

### 분포

히스토그램 ·  
품질 신호

3

### 전처리

제외 · LOD · log ·  
z-score

4

### 차이검정

ANOVA · Tukey  
· LDA

5

### 마커

상관 + FDR ·  
LASSO

6

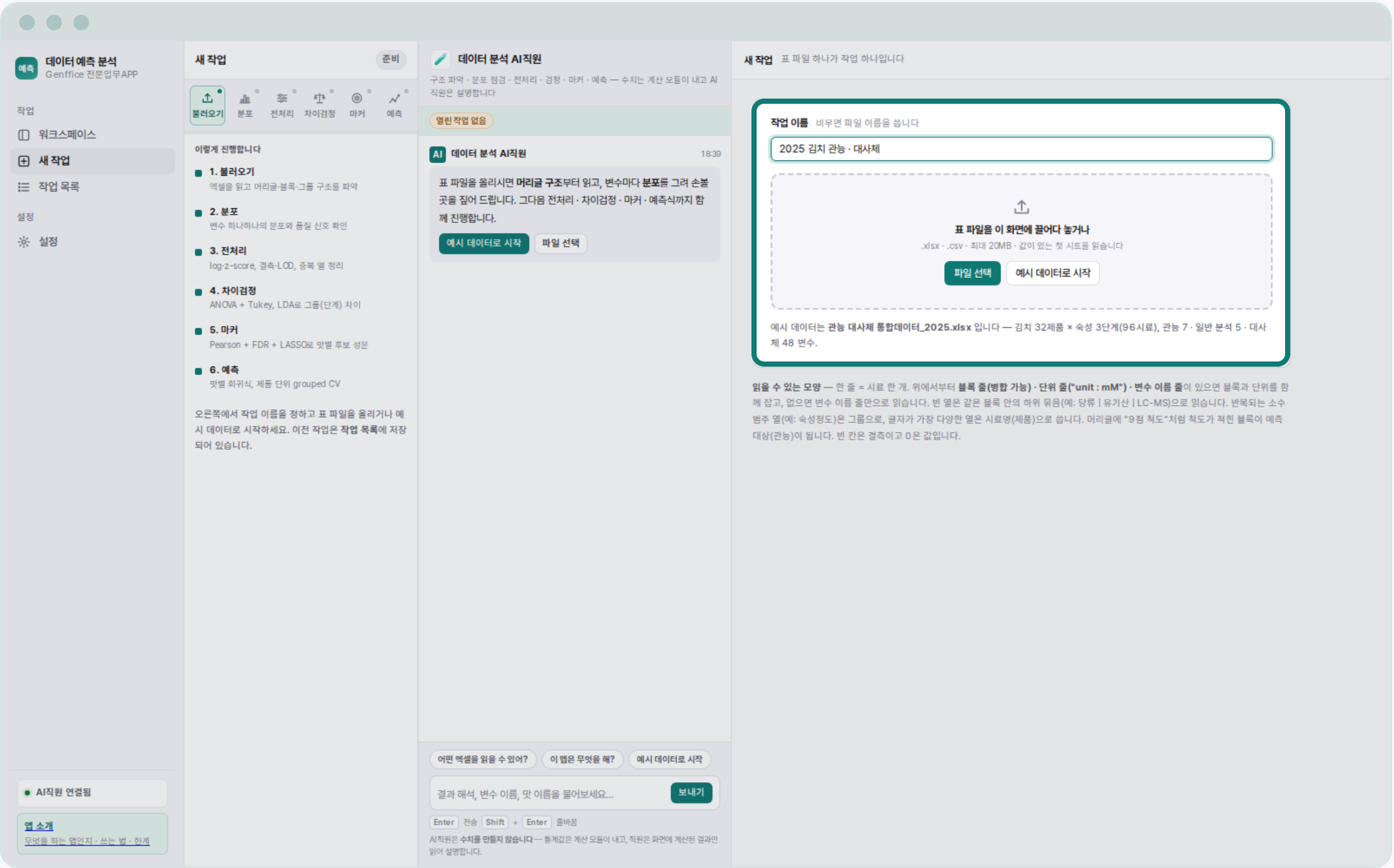
### 예측

회귀식 · 교차검증

전처리 규칙을 고치면 뒤 단계 결과를 지우고 새 규칙으로 다시 계산합니다

— 새 작업

작업 이름을 정하고, 연구실 엑셀을 그대로 올립니다



파일

.xlsx · .csv · 최대 20MB

작업

표 파일 하나 = 작업 하나

예시

김치 관능 · 대사체 데이터

저장

작업 목록에 자동 저장

— STEP 1 — 불러오기

같은 제품을 단계마다 켜 반복측정 구조를 스스로 찾습니다

데이터 예측 분석  
Genffice 전문업무APP

작업  
워크스페이스  
새 작업  
작업 목록  
설정  
설정

1. 불러오기

업로드

분포

전처리

차이검정

마커

예측

작업 2025 김치 관능·대사체 이름 바꾸기

관능 대사체 통합데이터\_2025.xlsx 읽음  
Sheet1 · 96시료 × 60변수  
다른 파일로 바꾸기

데이터 구성  
포 데이터  
관능 평가 7 · 일반 분석 5 · 대사체 분석 5 · 대사체 분석 6 · 대사체 분석 37  
그룹 (예: 속성 단계)  
속성정도: 미숙기 (A) → 적숙기 (B) → 과숙기 (C)  
블록 머리글 병합 · 빈 열로 구분  
관능 평가  
7개 · 1-9점 척도  
일반 분석  
5개  
대사체 분석 5  
5개 · mM  
대사체 분석 6  
6개 · mM  
대사체 분석 37  
37개 · uM

AI직원 연결됨  
업 소개  
모의를 하는 열인지 · 쓰는 법 · 함께

2. 분포 보기 →

데이터 분석 AI직원

구조 파악 · 분포 점검 · 전처리 · 검정 · 마커 · 예측 — 수치는 계산 모듈이 내고 AI 직원은 설명합니다

2025 김치 관능·대사체 1. 불러오기 96×60  
행 · 데이터 96줄

데이터 구조  
시료 96개 (미숙기 (A) 32 · 적숙기 (B) 32 · 과숙기 (C) 32)  
변수 60개 · 관능 평가 7 · 일반 분석 5 · 대사체 분석 5 · 대사체 분석 6 · 대사체 분석 37  
반복측정 32제품 × 3단계  
속성정도 · 시료명  
describe\_all 60 변수 기술통계 · 품질 신호 · 열 쌍 상관  
1770개

AI 데이터 분석 AI직원 18:39  
구조를 읽었습니다. 같은 제품을 단계마다 켜 반복측정이라 뒤의 모델 검증은 시료명 기준 grouped CV로 잡아 두겠습니다.  
분포에서 먼저 보실 곳은 이렇습니다.  
치우침 26개 — Sucrose, Taurine, Dihydrocapsalcin 등. log 변환이 필요해 보입니다.  
0 값이 많은 변수 4개 — Mannitol (mM), Mannitol (uM), phenyl\_acetic\_acid. 0이 '미검출'인지 정해야 합니다.  
중복 의심 5쌍 — Citric acid ≈ Citric\_acid, Malic acid ≈ Malic\_acid. 단위만 다른 같은 측정으로 보입니다.  
극단값이 있는 변수 20개 — 원자료 입력 오류인지 대조가 필요합니다.  
속성단계를 가장 잘 가르는 변수는 신맛(η² 0.92), pH(η² 0.89), 산도(η² 0.85) 입니다.  
분포 보기 주의 변수만

치우친 변수 보여줘 0이 많은 변수는? 중복 열 있어?  
속성단계로 잘 갈리는 변수 유산균 수 분포  
결과 해석, 변수 이름, 맛 이름을 물어보세요... 보내기  
Enter 전승 Shift + Enter 종배양  
AI직원은 수치를 만들지 않습니다 — 통계값은 계산 모듈이 내고, 직원은 화면에 계산된 결과만 읽어 설명합니다.

불러오기 결과 관능 대사체 통합데이터\_2025.xlsx · Sheet1 2. 분포 보기 →

시료 96  
미숙기 (A) 32 · 적숙기 (B) 32 · 과숙기 (C) 32  
변수 60  
5개 블록  
반복측정 구조 32 × 3  
같은 제품이 단계마다 측정됨  
결측 칸 0  
빈 칸만 생깁니다. 0은 값  
주의 변수 34  
치우침 · 0 다수 · 극단값 · 중복  
검증 설계에 바로 쓰이는 사실입니다. 32개 제품이 3단계에 걸쳐 측정되었습니다. 같은 제품의 다른 시점이 학습-검증에 동시에 들어가면 성능이 부풀러지므로, 6단계 예측 모델은 시료명 을 그룹 키로 한 grouped CV로 같은 제품이 학습과 검증에 나뉘어 들어가지 않게 합니다.

머리글을 이렇게 읽었습니다 틀렸다면 원본 엑셀 머리글을 고쳐 다시 올리세요

|         |      |                             |
|---------|------|-----------------------------|
| 블록 줄    | 2행   | 병합 셀 → 블록 이름 (5개)           |
| 단위 줄    | 3행   | "unit:" 접두어를 떼고 변수 단위로      |
| 변수 이름 줄 | 4행   | 아래 줄부터 데이터 (96줄)            |
| 그룹 열    | 속성정도 | 미숙기 (A) → 적숙기 (B) → 과숙기 (C) |
| 시료명 열   | 시료명  | 고유 32개                      |
| 번호 열    | 연번   | 분석 변수에서 뺍니다                 |

확인이 필요한 것  
변수 이름 "Mannita"이 2번 나옵니다 — 단위를 붙여 구분했습니다

블록

| 블록                        | 변수 | 단위         | 변수 예                                          | 주의 |
|---------------------------|----|------------|-----------------------------------------------|----|
| 관능 평가<br>관능 평가 결과 (9점 척도) | 7  | 변수별 · 1-9점 | 신맛, 짭맛, 단맛, 감칠맛 ...                           | 1  |
| 일반 분석<br>일반 분석 결과         | 5  | 변수별        | pH, 산도, 염도, 당도 ...                            | 1  |
| 대사체 분석 5<br>대사체 분석 결과     | 5  | mM         | Glucose, Fructose, Mannitol (mM), Sucrose ... | 3  |

시료

96개

변수

60개 · 5블록

반복측정

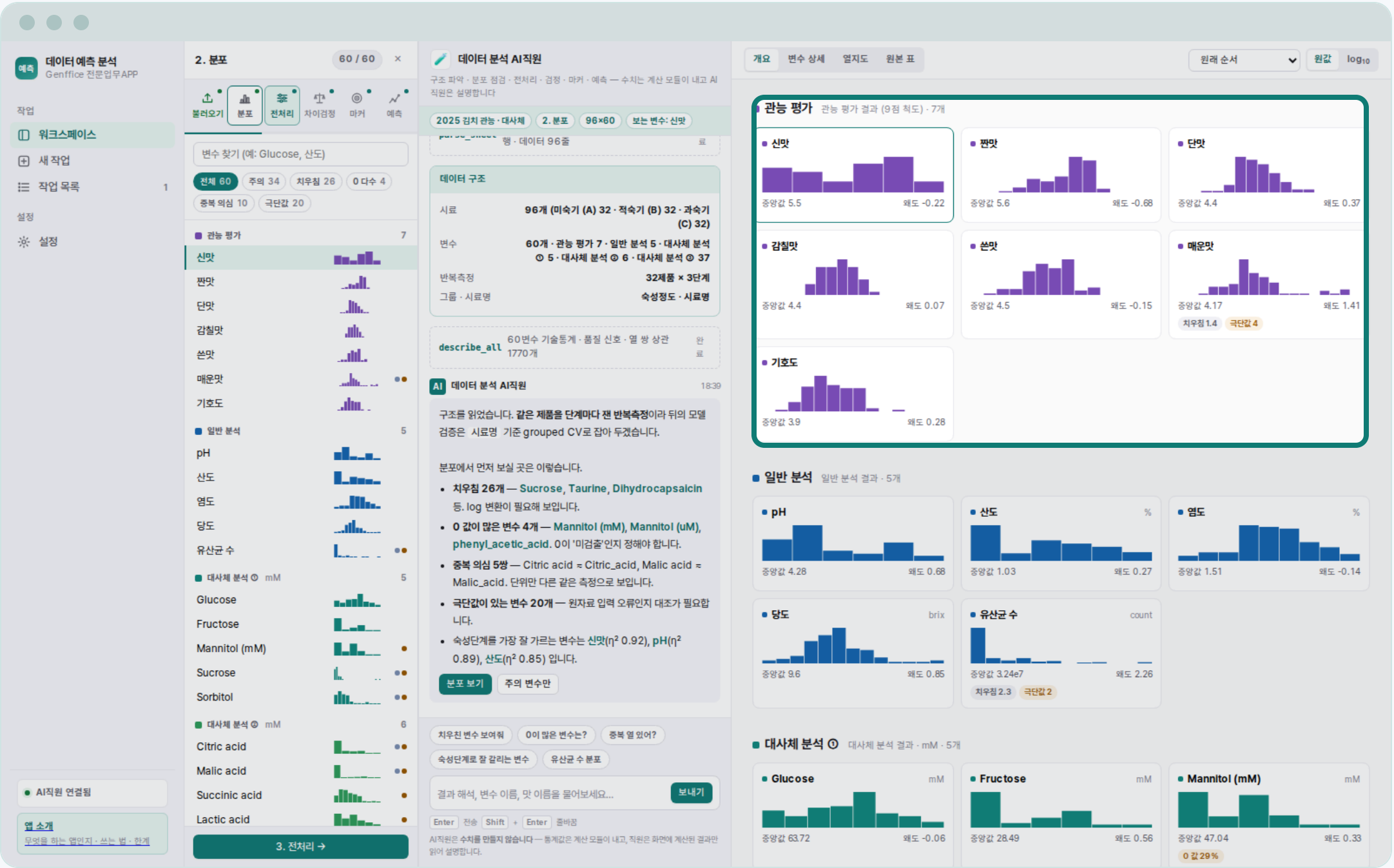
32제품 × 3단계

중복 의심

단위만 다른 5쌍

— STEP 2 — 분포

60개 변수의 분포와 품질 신호를 한 번에 훑습니다



치우침  
26개

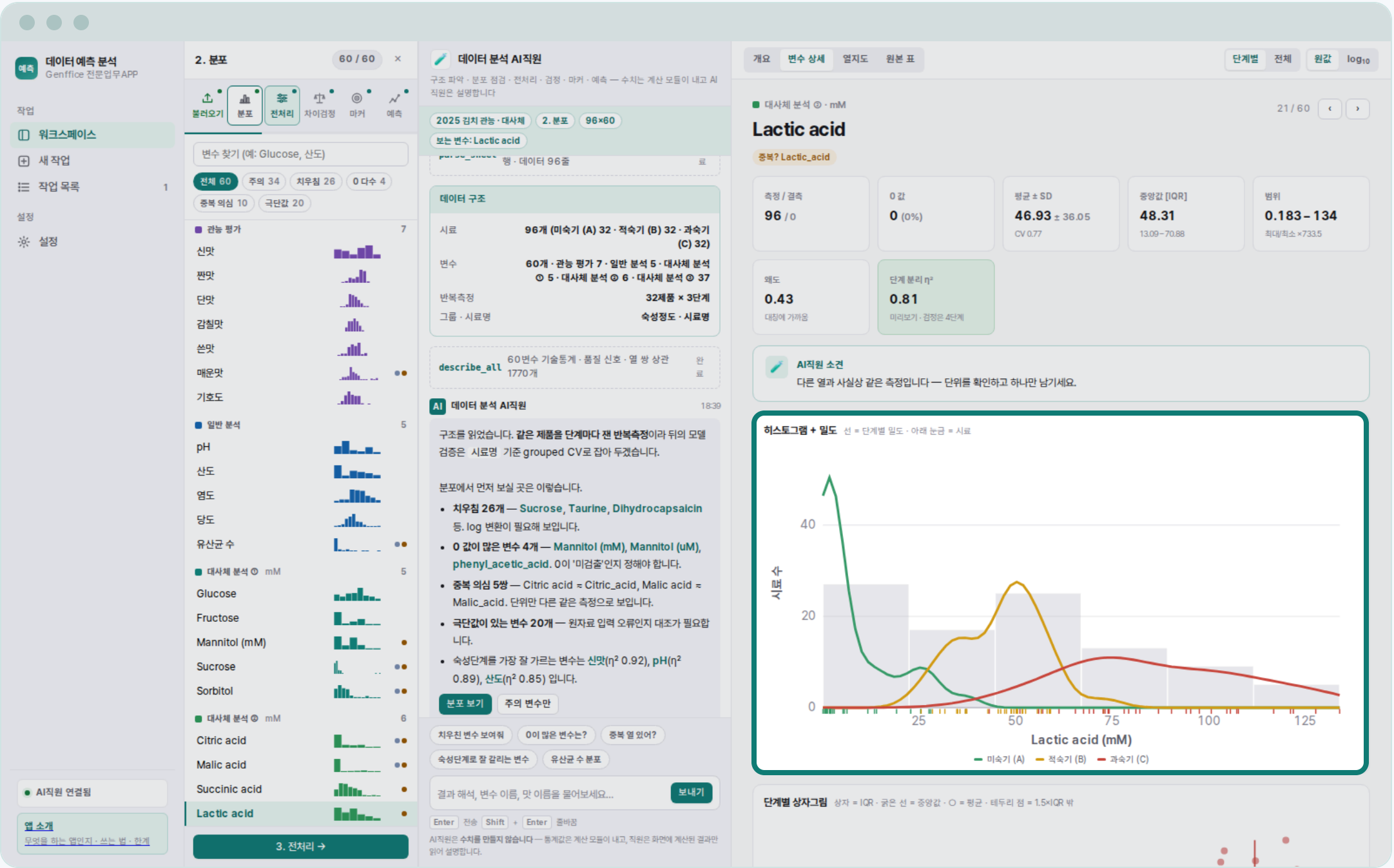
0 값 다수  
4개

극단값  
20개

중복 의심  
10열

— STEP 2 — 변수 상세

변수 하나를 히스토그램 · 상자그림 · Q-Q로 나란히 봅니다



예  
Lactic acid

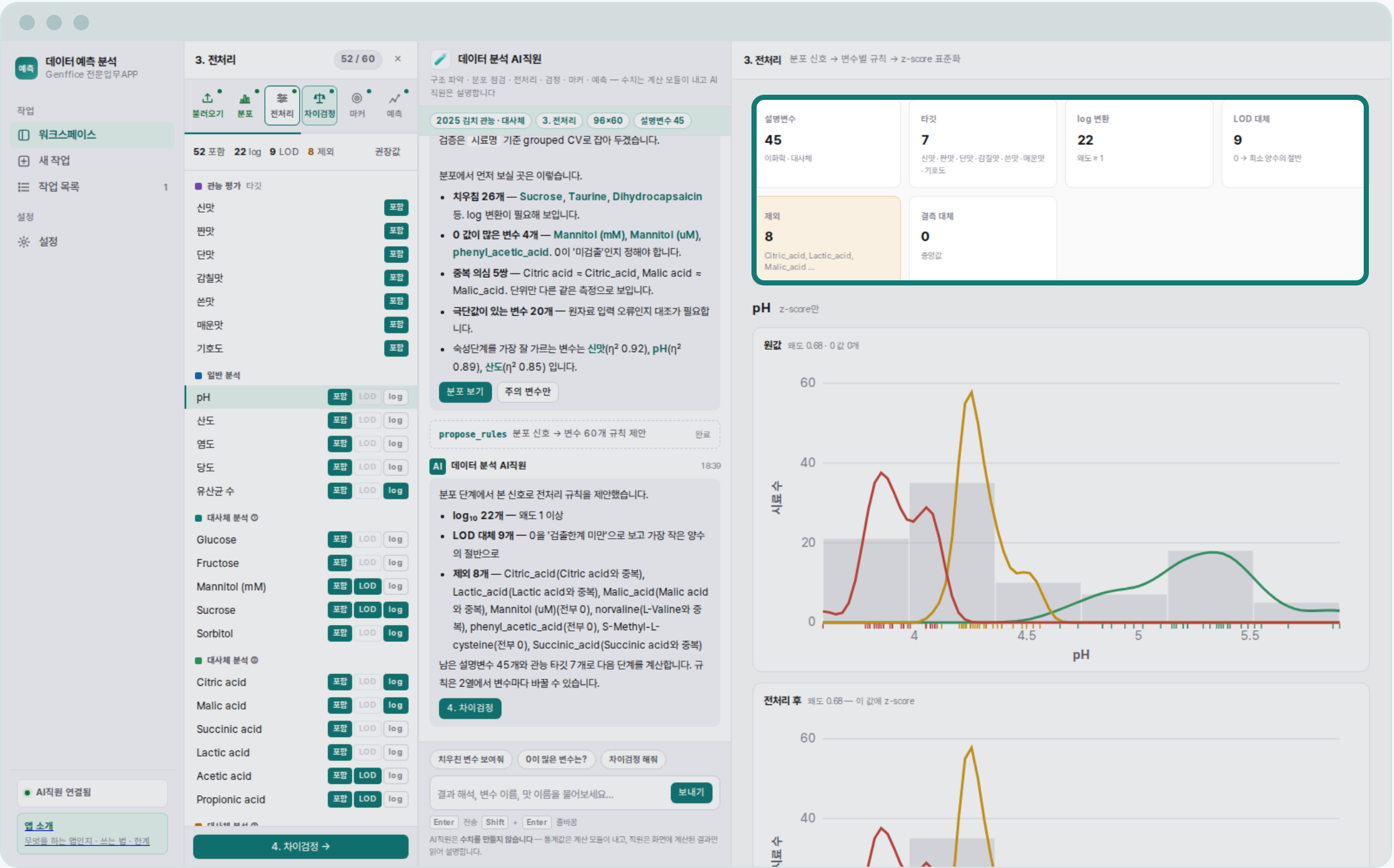
단계 분리  
 $\eta^2$  0.81

그림  
밀도 · 상자 · Q-Q · 누적

전환  
원값  $\leftrightarrow \log_{10}$  · 단계별  $\leftrightarrow$  전체

— STEP 3 — 전처리

분포 신호에서 규칙을 제안하고, 사람이 변수마다 고칩니다



설명변수  
45개 남김

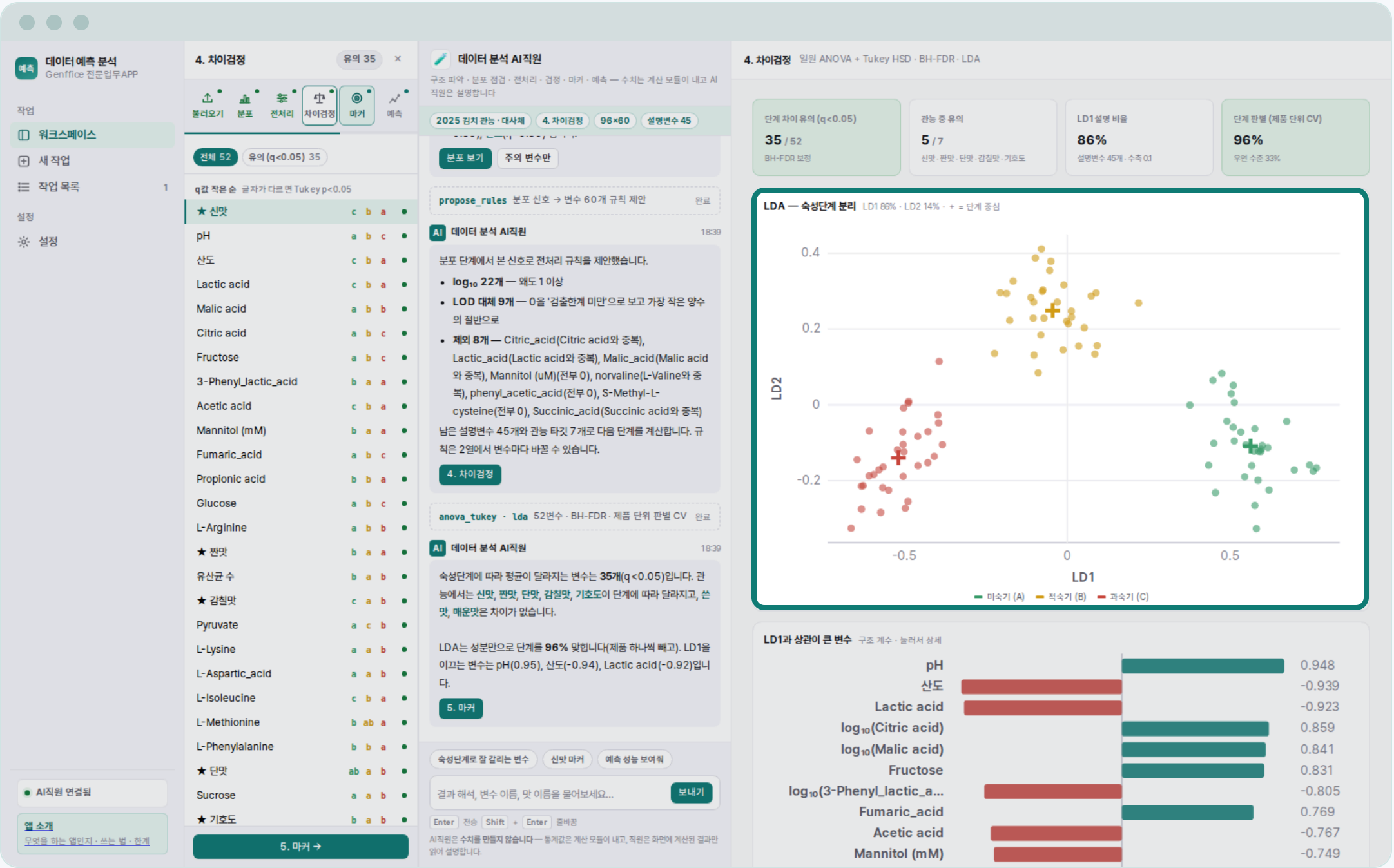
log<sub>10</sub> 변환  
22개

LOD 대체  
9개

제외  
8개 · 중복 · 상수

— STEP 4 — 차이검정

숙성 단계에 따라 무엇이 달라지는지, 성분만으로 단계를 가를 수 있는지



단계 차이 유의

35 / 52 변수

관능 중 유의

5 / 7

LDA 단계 판별

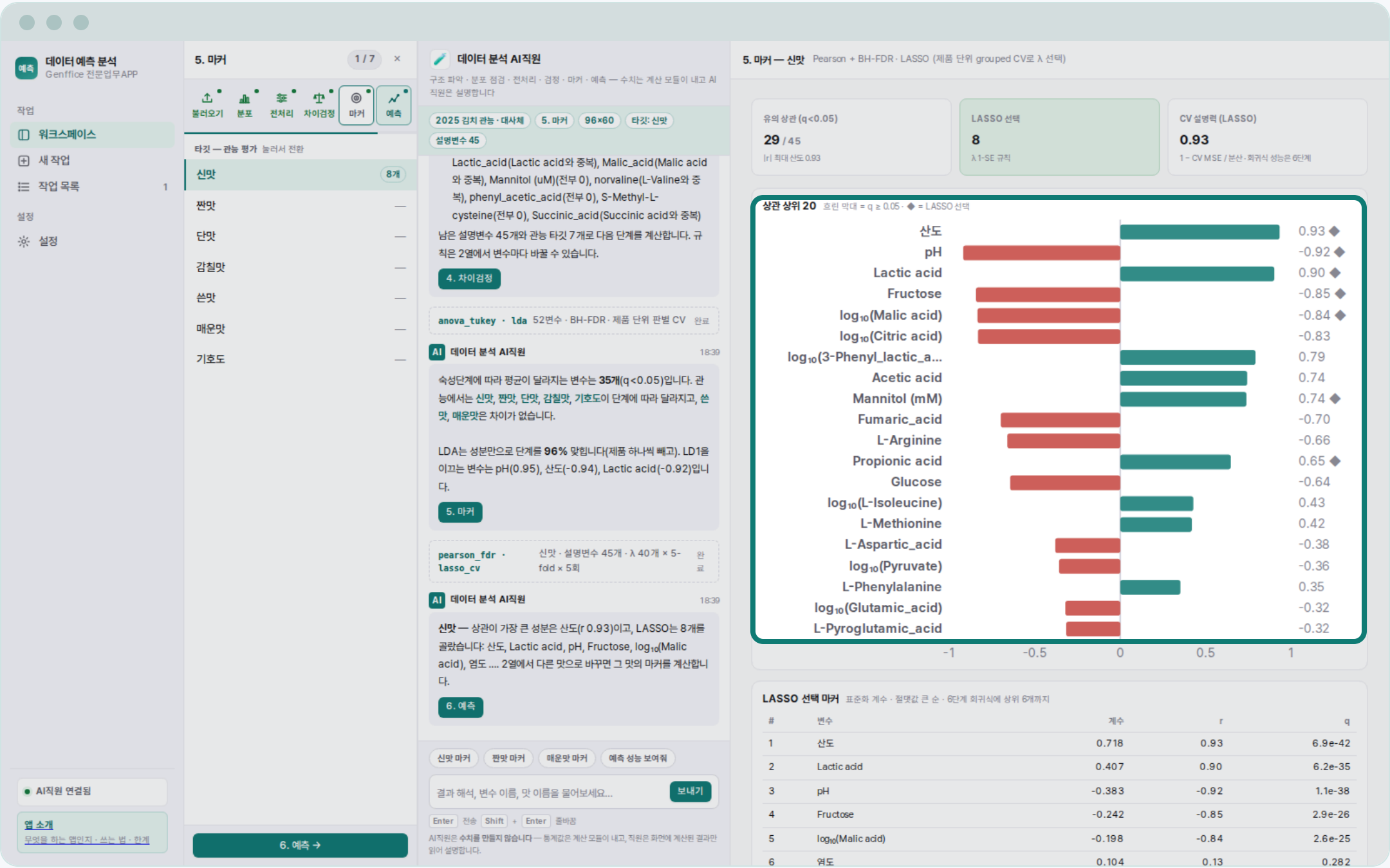
96% · 제품 하나씩 빼고

가르는 성분

pH · 산도 · Lactic acid

— STEP 5 — 마커

맛마다 함께 움직이는 성분을 찾고, 겹치지 않는 핵심만 고릅니다



신맛 유의 상관

29 / 45

LASSO 선택

8개

상관 1위

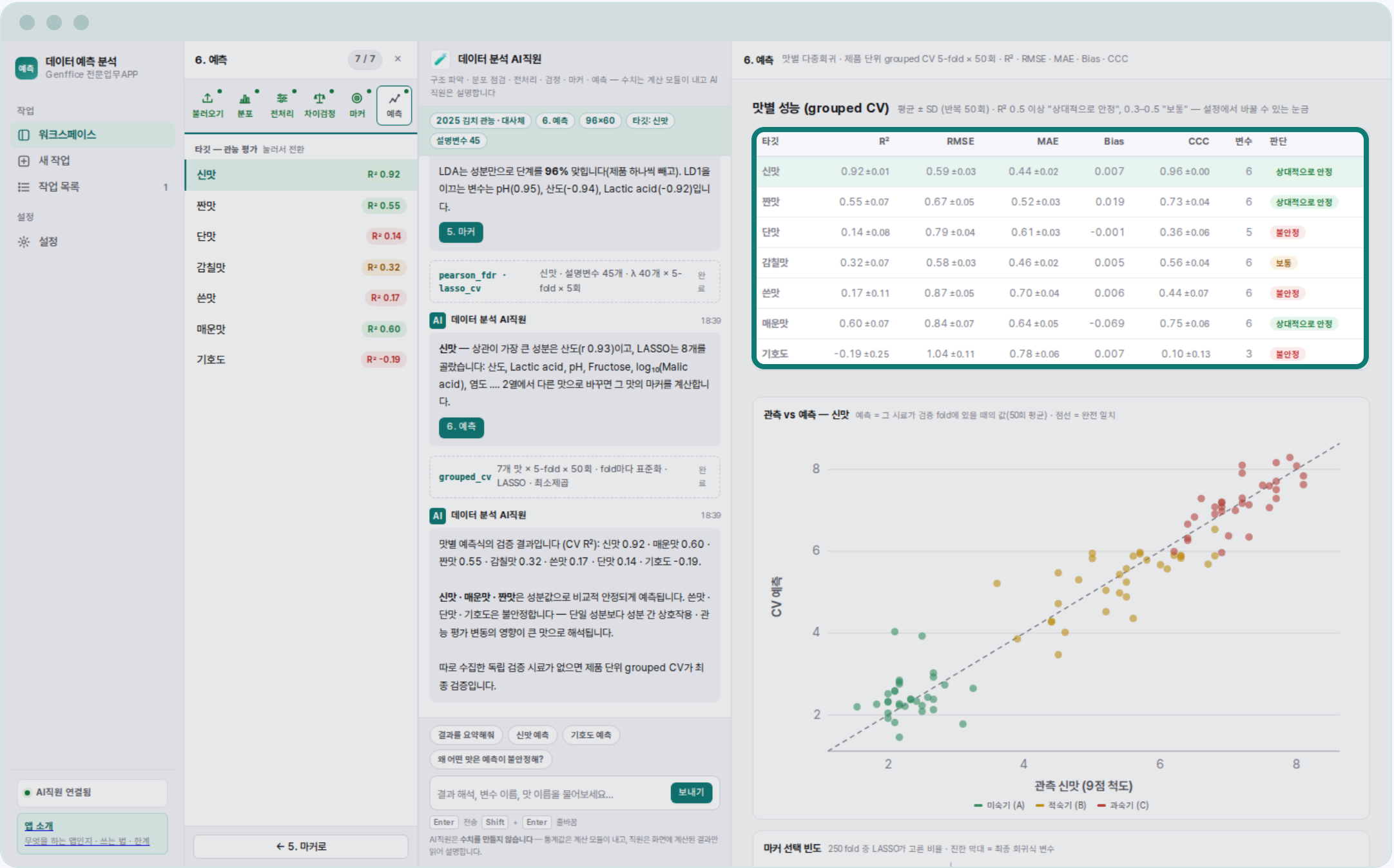
산도 r 0.93

겹침

pH와 산도는 정보가 겹침

— STEP 6 — 예측

같은 제품은 같은 접기에 묶어, 맛별 예측 성능을 정직하게 잽니다



6. 예측 맛별 다중회귀 · 제품 단위 grouped CV 5-fold × 50회 · R<sup>2</sup> · RMSE · MAE · Bias · CCC

맛별 성능 (grouped CV) 평균 ± SD (반복 50회) · R<sup>2</sup> 0.5 이상 "상대적으로 안정", 0.3~0.5 "보통" — 설정에서 바꿀 수 있는 눈금

| 타겟  | R <sup>2</sup> | RMSE        | MAE         | Bias   | CCC         | 변수 | 판단       |
|-----|----------------|-------------|-------------|--------|-------------|----|----------|
| 신맛  | 0.92 ± 0.01    | 0.59 ± 0.03 | 0.44 ± 0.02 | 0.007  | 0.96 ± 0.00 | 6  | 상대적으로 안정 |
| 짭맛  | 0.55 ± 0.07    | 0.67 ± 0.05 | 0.52 ± 0.03 | 0.019  | 0.73 ± 0.04 | 6  | 상대적으로 안정 |
| 단맛  | 0.14 ± 0.08    | 0.79 ± 0.04 | 0.61 ± 0.03 | -0.001 | 0.36 ± 0.06 | 5  | 불안정      |
| 감칠맛 | 0.32 ± 0.07    | 0.58 ± 0.03 | 0.46 ± 0.02 | 0.005  | 0.56 ± 0.04 | 6  | 보통       |
| 쓴맛  | 0.17 ± 0.11    | 0.87 ± 0.05 | 0.70 ± 0.04 | 0.006  | 0.44 ± 0.07 | 6  | 불안정      |
| 매운맛 | 0.60 ± 0.07    | 0.84 ± 0.07 | 0.64 ± 0.05 | -0.069 | 0.75 ± 0.06 | 6  | 상대적으로 안정 |
| 기호도 | -0.19 ± 0.25   | 1.04 ± 0.11 | 0.78 ± 0.06 | 0.007  | 0.10 ± 0.13 | 3  | 불안정      |

관측 vs 예측 — 신맛

예측 = 그 시료가 검증 fold에 있을 때의 값(50회 평균) · 점선 = 완전 일치

마커 선택 빈도 250 fold 중 LASSO가 고른 비율 · 진한 막대 = 최종 최귀식 변수

신맛

R<sup>2</sup> 0.92 · 상대적으로 안정

매운맛 · 짭맛

R<sup>2</sup> 0.60 · 0.55

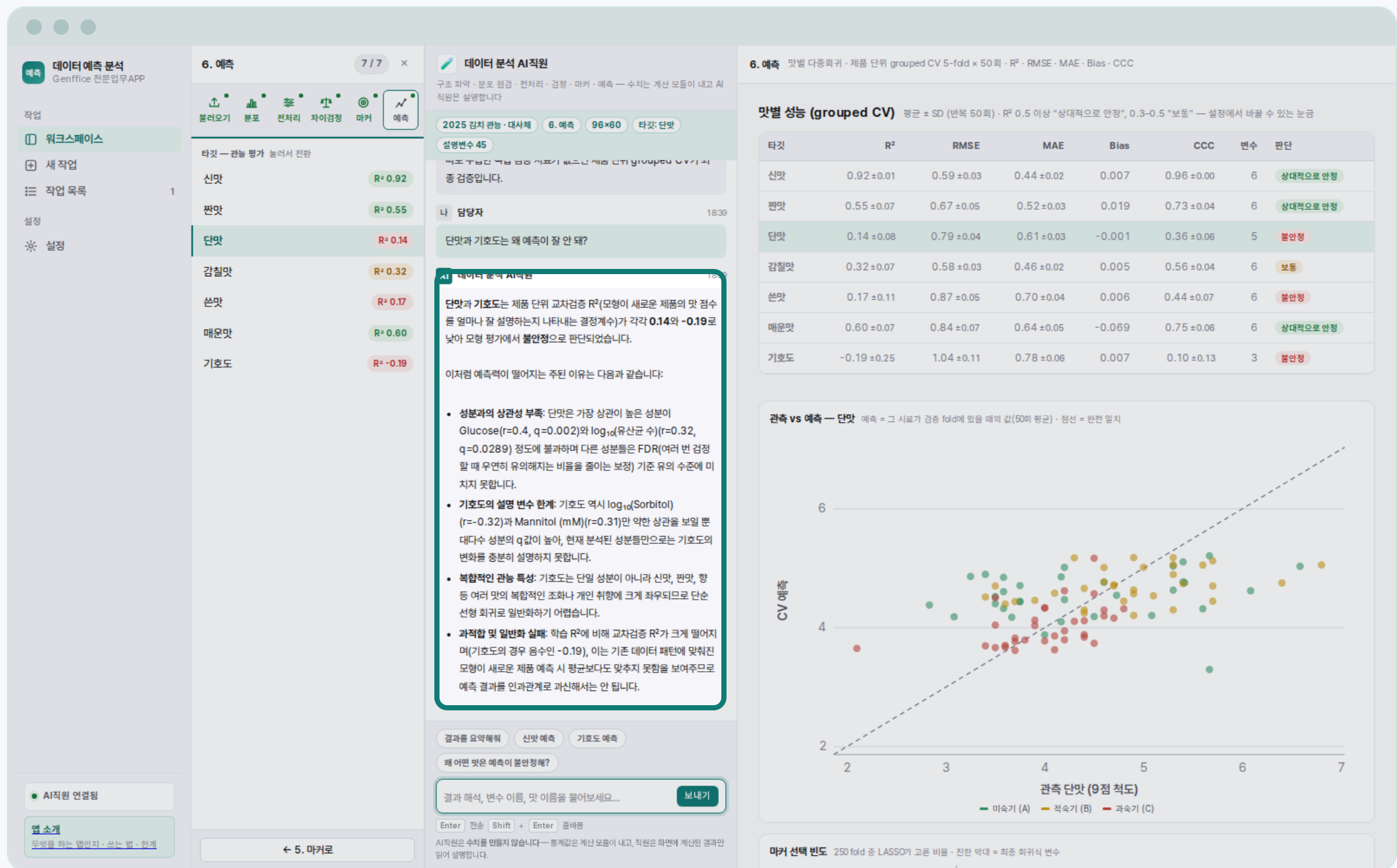
감칠맛

R<sup>2</sup> 0.32 · 보통

단맛 · 쓴맛 · 기호도

불안정

# 말로 물으면, 지금 계산된 값만 인용해 설명합니다



## 질문

단맛과 기호도는 왜 예측이 잘 안 돼?

## 근거

교차검증 R<sup>2</sup> 0.14 · -0.19

## 원칙

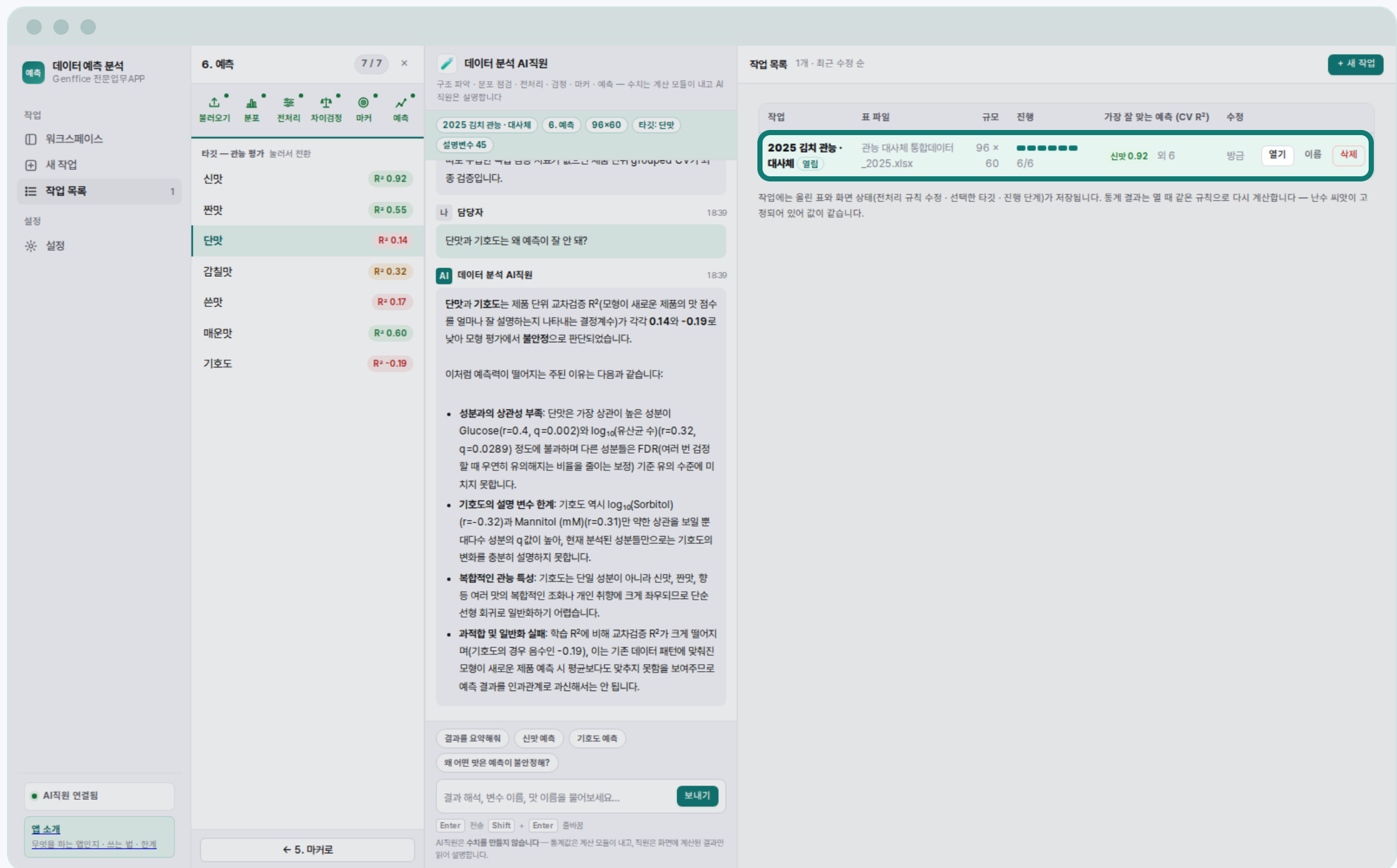
새 수치를 만들지 않음

## 이동

변수 · 맛 이름으로 화면 이동

— 작업 목록

# 다시 열면 보던 단계와 규칙 그대로 돌아옵니다



## 저장

표 · 규칙 수정 · 선택한 맛 · 단계

## 한눈에

진행 6/6 · 신맛 R<sup>2</sup> 0.92

## 관리

열기 · 이름 바꾸기 · 삭제

## 재현

같은 표면 같은 결과

**분석 기준은 설정에서 바꾸고, 저장하면 지금 작업을 다시 계산합니다**

## 분포

## 치우침 1·0 다수 20%

## 차이검정

유의수준 0.05

## 마커

**λ 5접기 × 5회 · 1-SE**

## 예측

5접기 × 50회 · 최대 6변수

— 믿을 수 있게 만든 원칙

## 부풀려진 성능을 가장 경계했습니다



**원본은 바꾸지 않는다**

변환은 계산용 사본에만



**틀릴 수 있는 곳은 드러낸다**

머리글 추정 · 0 값 · 중복



**누수를 막는다**

같은 제품은 같은 접기



**같은 표면 같은 결과**

난수 씨앗 고정



**AI는 수치를 만들지 않는다**

계산 요약만 인용



**상관 · 예측 · 인과 구분**

마커는 후보일 뿐

— 예시 데이터와 계산

통계는 브라우저 안에서 계산하고, AI직원은 그 결과를 설명만 합니다

96개

시료

60개

변수 · 5블록

32제품

교차검증 묶음

250회

맛마다 검증 적합

모든 통계 계산은 AI 없이 동작합니다 — AI직원은 계산된 결과를 읽고 설명하는 역할입니다

— 지금 상태

## 화면만 만든 목업이 아니라, 실제로 계산하는 시제품입니다

1

### 어떤 표든 다시 계산

같은 형식의 다른 엑셀을 올리면 머리글  
해석부터 예측까지 새로 계산

2

### 작업 저장과 복원

작업을 서버에 저장하고, 다시 열면 보던  
단계와 규칙으로 돌아옴

3

### AI직원 연결

계산 결과를 근거로 질문에 답하고,  
단계가 끝날 때마다 요약을 보고

— 아직 채워야 할 것

# 보고 · 출판에 쓰기 전에 채워야 할 것이 있습니다

|         |          |          |          |        |         |
|---------|----------|----------|----------|--------|---------|
| 독립 검증   | 별도 시료 없음 | 반복측정 검정  | 일원 ANOVA | 사진 데이터 | 자리만     |
| 미생물 데이터 | 자리만      | 보고서 내보내기 | 아직 없음    | 큰 데이터  | 브라우저 계산 |
| 접근 · 보안 | 로그인 없음   | 전문가 검증   | 미 실시     | 플랫폼 연결 | 단독 앱    |

— 다음 단계 제안

## 검증부터, 세 단계로 넓혀 가겠습니다

1

### 1단계 · 검증 보강

새 시료에 회귀식을 그대로 적용하는  
독립 검증 화면, 제품을 임의효과로  
둔 혼합모형

2

### 2단계 · 보고서

논문 · 보고서용 표와 그림 묶음 내보내기,  
분석 조건 기록

3

### 3단계 · 다중모달

사진 · 미생물 데이터를 더해 성능이  
실제로 오르는지 확인

# 데이터 예측 분석

데모

[dataprediction.app.genver.online/](http://dataprediction.app.genver.online/)

소개 문서

[dataprediction.app.genver.online/intro/](http://dataprediction.app.genver.online/intro/)

소개 영상

[dataprediction.app.genver.online/video/](http://dataprediction.app.genver.online/video/)

결과는 탐색을 돕는 참고 자료입니다 — 마커와 예측식은 원인을  
증명하지 않으며, 보고 전에는 원자료 대조와 전문가 검토가 필요합니다