

2025 데이터 문제해결은행 데이터레시피 작성 예시

현황 보고



과학기술정보통신부



한국데이터산업진흥원

데이터레시피 작성 예시 (1/6)



레시피 분류정보

레시피명

자동화된 데이터 처리를 통해 대량의 제조 데이터를 활용한 이상치 탐지

데이터 활용목적 대분류

고객예측

데이터 활용목적 중분류

고객·시장 수요 예측

표준산업 대분류

가구 내 고용활동 및 달리 분류되지 않은 ...

표준산업 중분류

가구 내 고용활동

표준산업 소분류

가구 내 고용활동

표준산업 세분류

가구 내 고용활동

표준산업 세세분류

가구 내 고용활동

산업분야

고용

기업유형

1인창조기업

기업소재지

서울특별시

분석모델

이상치 탐지

알고리즘

Autoencoder

키워드

이상치 탐지, 제조 데이터, 자동화

키워드는 쉼표(,) 구분하여 입력해 주세요.

레시피 분류 정보 작성 가이드

• 레시피명

분석 목적, 특징, 데이터, 방법론(모델) 등이 잘 드러날 수 있도록 작성

• 데이터 활용 목적

분석 주제와 가장 유사한 활용 목적을 선택

• 키워드

활용목적, 표준산업분류, 데이터 등 키워드 작성

• 분석모델/알고리즘

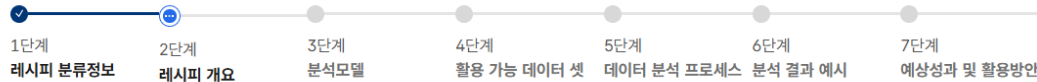
활용한 분석 모델 및 세부 알고리즘을 지정

레시피 분류 정보

활용목적	서비스 개발	기업소재지	경상남도
업종	제조업 > 전자 부품, 컴퓨터, 영상, 음향 및 통신장비 제조업 > 전자 부품 제조업 > 인쇄회로기판 및 전자부품 실장기판 제조업 > 인쇄회로기판용 적응판 제조업	기업유형	소기업
분야	제조	사업수행년도	2024

예시

데이터레시피 작성 예시 (2/6)



레시피 개요

- 제조 공정에서 발생하는 이상치를 탐지하여 품질 관리와 생산 지표를 향상시키고자 함.
- 자동화된 데이터 처리를 통해 대량의 제조 데이터를 활용한 이상치 탐지를 수행함.
- 이상치 탐지를 통해 불량률을 낮추고, 재료 낭비를 최소화함.
- 제조 데이터를 활용한 이상치 탐지

레시피 개요

(분석 배경) 홈인테리어 시장의 온라인화가 급속히 진행되면서, 고객의 취향과 수요에 따른 맞춤형 제안의 중요성이 커지고 있음. 현재 상품 노출과 프로모션이 획일적이어서, 이탈률 감소 및 전환을 제고를 위한 데이터 기반 전략이 필요함

(데이터 및 분석 필요성) 고객의 검색 이력, 구매 로그, 장바구니 데이터 등 다양한 행동 데이터를 활용하여 군집분석을 수행함으로써, 고객을 세분화하고 각 그룹의 특성을 반영한 전략 수립이 필요함

(사업/과제 개요) 고객 구매패턴 분석 기반 가구 기획전 추천시스템 고도화와 고객의 검색 이력, 구매 로그, 장바구니 데이터 등을 분석하여 취향군을 세분화하고, 각 군별로 반응률이 높은 상품 구성 및 추천 알고리즘을 구축하고자 함

(활용방안 및 기대효과) 군집별 소비 특성에 따라 맞춤형 추천·프로모션 전략을 적용하고, 분석 결과를 기획전 운영·상품 개발·공급 전략에 연계하여 전사 마케팅 고도화를 지원할 수 있음

예시

레시피 개요 작성 가이드

• 분석 배경

분석의 배경이 되는 비즈니스 이슈, 조직 내외의 현황 등을 자세히 작성

• 사업 및 과제 개요

분석의 배경이 되는 사업 및 과제의 세부 사항을 구체적으로 작성하여 과업의 목적을 정리

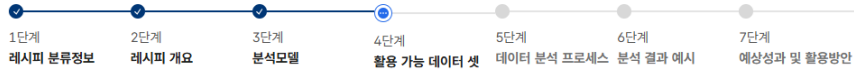
• 데이터 및 분석 필요성

사용하는 데이터를 명시하고, 구체적인 분석 방법, 분석 프로세스와 함께 분석 필요성을 작성

• 활용방안 및 기대효과

본 분석 결과의 활용 방안 및 기대 효과를 구체적으로 작성

데이터레시피 작성 예시 (3/6)



활용 가능 데이터 셋

데이터 구분

데이터 출처

데이터 테이블명

데이터 테이블 컬럼정의 파일

[⊕ 내 PC에서 파일을 직접 선택하거나, 파일을 끌어오세요.](#)

보유데이터 활용가이드 파일

[⊕ 내 PC에서 파일을 직접 선택하거나, 파일을 끌어오세요.](#)

데이터셋 파일

[⊕ 내 PC에서 파일을 직접 선택하거나, 파일을 끌어오세요.](#)

데이터 변수 가이드라인

데이터 변수 가이드라인을 입력해주세요

데이터 설명

데이터 설명을 입력해주세요

활용 가능 데이터 셋 작성 가이드

• 데이터 구분/출처

재현 데이터 혹은 공개 데이터 등 해당되는 데이터 선택 후 데이터의 출처에 대해 작성

• 데이터 테이블명

해당 데이터를 가장 잘 설명할 수 있는 테이블 명칭 작성

• 테이블 컬럼정의 파일

데이터 컬럼별 타입(type), 컬럼명, 데이터 특징 등이 작성된 테이블 정의서 업로드

활용 데이터 테이블 열릴 정보

변수 번호	타겟 변수 여부	데이터 타입	데이터 활용 명	비고
1	X	Float64	X_ActualPosition	부품의 실제 x 위치
2	X	Float64	X_ActualVelocity	부품의 실제 x 속도
3	X	Float64	X_ActualAcceleration	부품의 실제 x 가속도

예시

• 데이터 활용가이드 파일

컬럼별 데이터 특성 및 활용 안내사항이 작성된 가이드 파일 업로드

• 데이터 변수 가이드라인 / 데이터 설명

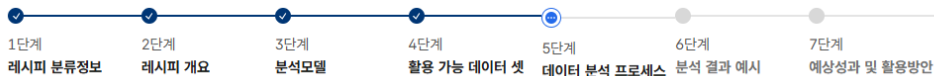
데이터 컬럼별 주요 특성 및 데이터 개요 작성

▪ 데이터 변수 가이드라인

- 카테고리, 성별, 주문날짜(월, 요일), 가격 정보를 기반으로 데이터 간 유사도를 계산하고, 비지도 학습 알고리즘을 활용해 그룹 도출

예시

데이터레시피 작성 예시 (4/6)



데이터 분석 프로세스

regular

- 15 +

I

U

↶

■ ▼

[↻](#)



⋮ ▼

≡

≡

≡

≡

내용을 입력해주세요.

- 분석 프로세스

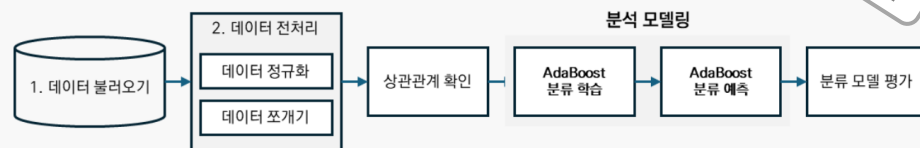
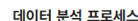
데이터 불러오기, EDA, 전처리, 모델링 등 데이터 분석 전체 과정에 대한 프로세스를 이미지로 삽입

- 세부 프로세스 단계별 설명

세부 분석 프로세스별 역할 및 주요 활동을
구체적으로 작성하여 설명

• 단계별 세부 특징 설명

본 분석에서 차별적인 단계 별특징 설명



1. 데이터 불러오기
분류 모델에 활용할 데이터 불러오기
데이터 구분자를 알맞게 설정하여 데이터 명확하게 읽기
컬럼 이름을 요구사항에 맞는 형태로 변경하기

2. 데이터 전처리

데이터 정규화를 통해 각 특성 값들이 일정 범위에 포함되도록 하여 학습 과정에서 특성들 간의 불균형을 방지할 수 있음

학습 데이터를 훈련용과 테스트용 데이터로 나누는 과정이며, 훈련용 데이터는 모델 학습에 사용되고, 테스트 데이터는 학습된 모델을 평가하는 데 사용함

3. 상관관계 확인

모델을 만들기 전에 데이터의 특성들 간 상관관계를 확인함

상관관계가 높은 변수는 다중공선성 문제를 일으켜 모델의 성능 저하를 기할 수 있어 주의해야 함

상관관계 매트릭스나 같은 도구를 사용하여 상관관계를 확인하고, 상관관계가 높은 특성들을 적절히 조정하거나 제거하여 모델의 효율성을 높임

4. AdaBoost 분류 학습
AdaBoost는 약한 학습자를 여러 번 학습시켜 강한 학습자를 만드는 부스팅 모델
기본적으로 약한 학습자는 결정 트리나 같은 간단한 모델이며, 반복적으로 학습을 진행하며 잘못 예측된 샘플에 더 높은 가중치를 부여하여
가중치를 통해 각 학습 단계에서 이전에 틀린 예측을 더 잘 처리할 수 있도록 학습

5. AdaBoost 분류 예측

학습된 AdaBoost 모델을 사용하여 새로운 데이터를 분류하는 단계

테스트 데이터를 모델에 넣으면 각 약한 학습자가 예측을 수행하며, 최종적으로 모든 학습자의 결과가 합쳐져 최종 예측값을 도출함

약한 학습자들이 각각 다른 가중치를 가지고 예측에 참여하기 때문에 각 학습자의 역할이 결합되어 보다 정확한 예측을 가능하게 함

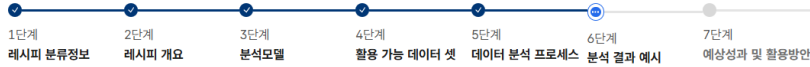
6. 분류 모델 평가

모델의 성능을 평가하기 위해 여러 평가 지표를 활용하여 성능을 다각도로 확인

혼동 행렬(Confusion Matrix)을 활용하여 모델의 예측 결과를 비율을 통해 시각적으로 확인할 수 있음

정확도(Accuracy)는 전체 데이터 중 올바르게 예측한 비율을 의미하고, 분류 모델의 가장 기본적인 성능 평가 지표로 활용됨

데이터레시피 작성 예시 (5/6)



분석 결과 예시

분석 결과 예시1

분석 결과 예시

④ 내 PC에서 파일을 직접 선택하거나, 파일을 끌어오세요.

분석 결과 예시

분석 결과 예시를 입력해주세요

레시피 개요 작성 가이드

• 분석 결과 업로드(이미지)

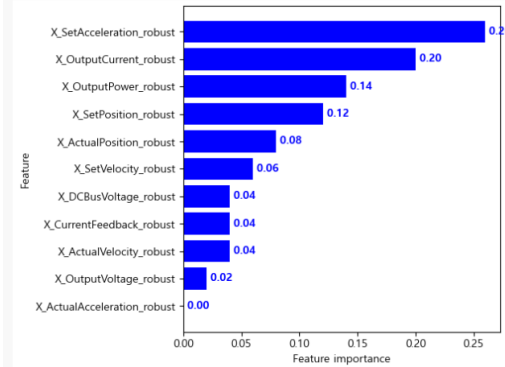
분석을 통해 도출된 결과를 시각적으로 표현하여 작성 모델 및 분석 핵심 지표 등을 제시

• 분석 결과 해석

분석 목적을 고려하여 분석 결과에 대해 해석을 작성하고 설명 문구 기입

분석결과와 결과해석은 한 쌍으로 작성하고,
이미지는 최대 5장까지 삽입 가능

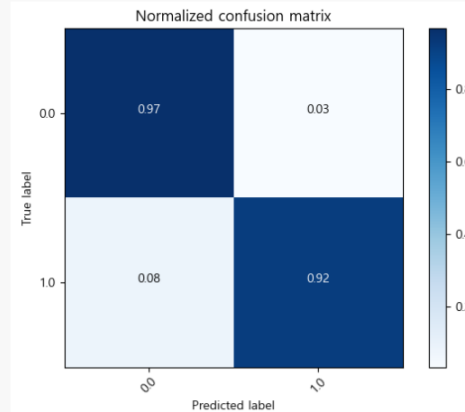
분석결과 예시 2



변수별 중요도(Feature Importance)

AdaBoost 모델에서 도출된 변수별 중요도를 나타내는 그래프로, 모델의 예측에 각 변수가 기여하는 정도를 나타냄
X_SetAcceleration_robust 변수는 가장 높은 중요도를 보이며, 분류 모델에서 가장 중요한 변수로 작용함
상위 중요도를 가진 변수들은 모델 성능을 좌우하는 주요 변수이며, 특히 가속도 설정 값과 출력 전류는 매우 중요한 요인임을 알 수 있음

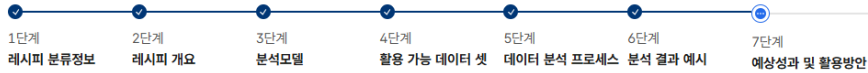
분석결과 예시 3



혼동 행렬(Confusion Matrix)

AdaBoost 모델의 성능을 평가한 정규화된 혼동행렬로, 모델의 예측 결과와 실제 레이블을 비교하여 분류 모델의 성능 평가 가능
좌상단은 실제 값이 0이고, 모델이 0으로 예측한 비율이 0.97로 양품을 정확하게 예측한 비율을 의미함
우상단은 실제 값이 양품인데 모델이 불량품으로 예측한 비율이 0.03로 모델이 잘못 분류한 비율을 나타냄

데이터레시피 작성 예시 (6/6)



예상성과 및 활용방안

성과 지표

성과지표를 입력해주세요

분석결과 활용방안

분석결과 활용방안을 입력해주세요

☐ [필수] 데이터 수집 및 활용 동의

예상성과 및 활용방안

성과지표

(정량적 성과)

고객별 맞춤 추천 알고리즘을 통해 유입된 고객 이탈률 감소, 주요 카테고리 전환율 증가 등 판매 전환 KPI 지표 개선
취향군별 반응을 높은 상품 구성으로 구매 확률 상승, 평균 장바구니 금액 증가 등 매출 향상

(정성적 성과)

데이터 분석을 통한 맞춤형 제안 체계 확립 및 개인화된 고객 접점 관리 체계 구축
취향군 세분화 결과를 바탕으로 기획된 운영 및 전략 수립, 상품 개발 방향성 제시에 기여
고객 취향에 맞춘 홈인테리어 스타일링 제공, 브랜드 충성도와 고객 만족도 제고 및 시장 경쟁력 강화
분석 결과의 전사적 활용으로 마케팅 고도화, 상품 공급 전략 최적화 및 효과적인 시장 대응 체계 확립

분석결과 활용방안

고객의 세분화된 취향군을 기반으로 맞춤형 기획전을 구성함으로써, 개인화된 쇼핑 경험을 제공하고 소비자 만족도를 증대시키는 전략
각 고객군의 반응 패턴 분석을 통해 상품 공급의 효율성을 극대화하고, 재고 관리와 마케팅 비용을 절감하는 운영 효율성 확보
다양한 품목으로의 추천 시스템 확장을 통해 고객의 다양한 니즈를 충족하고, 크로스 세일 기회를 높이는 수익 증대 가능성 모색
데이터 협업을 통해 더욱 풍부하고 정확한 인사이트를 도출, 타 가구 플랫폼과 연계하여 시장 분석 및 예측력을 강화하는 기회가 됨

예시

레시피 개요 작성 가이드

• 성과 지표

분석 결과를 활용하여 창출하고 싶은 예상성과에 대해서 구체적으로 작성

• 성과지표(세부)

정량적 성과는 KPI 기반 수치로 작성하고, 정성적 성과는 전략·의사결정 활용 중심으로 구분하여 작성

• 분석결과 활용방안

본 분석을 통해 도출된 분석 결과를 실제 업무·현장에 어떻게 적용할지, 구체적인 방안을 작성하고, 단기적 활용과 장기적 활용에 대해 작성

감사합니다



과학기술정보통신부



한국데이터산업진흥원