

AI & 데이터 기반 스마트팜 프로젝트 설계

(Hugging Face 기반 모델 융합 실습 중심)

데이터셋

[농촌진흥데이터셋 바로가기](#)

원본 데이터 파일명:

- 2021_cultInfo.xlsx
- 2021_growth_strawberry.xlsx
- 2021_env.xlsx
- 2021_sale.xlsx

데이터 구성 요약

데이터	주요 변수 (예시)
재배정보	연도, 작기, 지역, 농가명, 온실종류, 품목, 품종, 총면적, 식부면적 등
생육정보	조사일자, 초장, 엽수, 엽장, 엽폭, 관부직경, 화방번호, 착과수 등
환경정보	측정시간, 외부/내부 온·습도, CO ₂ , 토양온도, 일사량 등
판매정보	출하일자, 총출하량, 판매금액

가능한 프로젝트 아이디어

- 작물 수확량 예측 ★
(환경, 재배, 생육정보 → 총 출하량)
- 생육 상태 평가
(정상/이상 탐지, 생육정보·환경정보 활용)
- 판매가 예측
(환경, 재배, 생육정보 → 판매금액)
- 환경 최적화 권고 시스템 ★
(실시간 환경 분석 → 작물별 최적 환경 조건 제안)

활용 가능한 Hugging Face 모델

1. Time Series Transformer

- 대표 예시: [Informer](#)
- 시계열 데이터 기반 장기 예측에 강점
- 농업 데이터에 효과적으로 활용

| 활용 가능한 Hugging Face 모델

2. Tabular Transformer (SAINT 등)

- 구조화/표 데이터에 적합
- 수확량, 판매가 등 예측에 효율적
- [SAINT 모델 검색](#)



Task 1: 작물 수확량 예측

데이터 프레임과 주요 변수

구분	주요 변수
재배정보	연도, 작기, 지역, 농가명, 온실종류, 품종, 총면적, 식부면적 등
환경정보	측정시간, 온도 _ 내부, 상대습도 _ 내부, 잔존CO ₂ , 토양온도, 일사량 _ 외부
생육정보	조사일자, 초장, 엽수, 엽장, 관부직경, 화방번호, 착과수 등
판매정보(Target)	출하일자, 총출하량

- **환경·생육·재배정보**가 수확량에 영향을 주는가?
- 다양한 변수 결합 및 특징학습(Feature Engineering)이 핵심



Task 2: 환경 최적화 권고 시스템

구분	주요 변수
환경정보	측정시간, 온도 _ 내부, 습도 _ 내부, 잔존CO ₂ , 토양온도, 일사량 _ 외부
생육정보	조사일자, 초장, 엽수, 관부직경, 착과수 등

- **생육 지표별 최적 환경** 도출
- 환경·생육 데이터의 관계 분석 및 최적 조건 추천

두 Task 비교

구분	작물 수확량 예측	환경 최적화 권고
목적	미래 생산량 예측	최적 환경조건 추천
타겟	총출하량	초장, 엽수 등 생육지표
주요 모델	Informer, SAINT 등 회귀	회귀, 분류, 최적화 모델
데이터 활용 초점	장기 환경·생육 영향력 분석	실시간 환경-생육 관계 분석

데이터 처리 및 분석 프로토타입 계획

1. 데이터 전처리/병합

- 환경·생육·재배정보 병합 (Key: 농가명, 날짜 등)
- 결측치/이상치 처리, 정규화

2. EDA(탐색적 데이터 분석)

- 변수 간 상관관계 파악
- 데이터 분포, 특성 시각화

3. 모델 구축/훈련

- Hugging Face 모델 fine-tuning
- 회귀, 분류 Task별 출력 조정

4. 모델 성능평가

- MSE, RMSE, MAE 등 평가지표
- 예측 결과 시각화

5. 프로토타입 배포

- [Streamlit](#) 또는 [Gradio](#)로 간단 웹 데모 구현

실습 참고: Colab & 노트북 링크

- [데이터 필터링 \(딸기\)](#)
- [특징 탐색 데이터 분석](#)

Hugging Face AI 융합 설계 (색다른 접근법)

A. 전통적/표준적 방법

- **Informer, SAINT** 등 시계열/표 구조 모델 활용
- 환경·생육·재배정보 → 예측 Task 수행

B. 실험적/융합적 접근

- **멀티모달 융합**
(사진 + 센서 + 텍스트 등 동시 입력 → CLIP, BLIP 등)
- **LLM + Tabular 모델**
(LLM이 텍스트 특성, Tabular가 수치 → 앙상블/결합)
- **AutoML 기반 융합**
(AutoTrain Tabular로 다양한 모델 자동 조합)

융합 프로젝트의 교육적·학술적 강점

- 정형 + 비정형 + 시계열 데이터 융합 실험
- 모델별 비교·실험 구조 설계
- 데이터 결합/피처 엔지니어링/앙상블 등 실무 능력 향상

실습형 과제 제안 (예시 흐름)

1. 환경정보로 생육정보 예측

- 온실 온도/습도/토양온도 → 관부직경, 엽수 등

2. 생육정보로 수확량 예측

- 관부직경, 엽수, 식부면적 등 → 총출하량

3. 환경 + 생육 통합 → 수확량 직접 예측 (end-to-end)

4. 멀티모달 정보 융합 (도전 과제)

- 농장 설명, 현장 사진 등 추가

5. 모델 성능 비교, 실무 적용성 분석 및 발표

Q&A 및 토론

질문/아이디어 환영!

- Task별로 적합한 모델을 어떻게 선정할까?
- 멀티모달 융합에 도전한다면 준비할 것은?
- 실질적 교육·현장 적용을 위해 더 필요한 데이터/피처는?