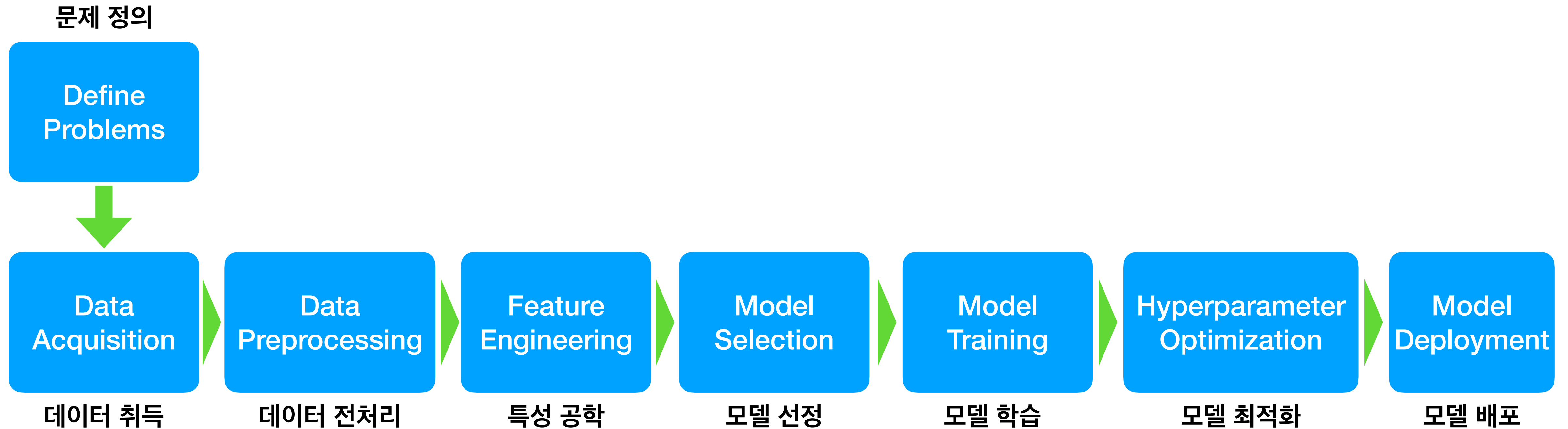


Chapter 12. Pycaret & Autogluon

머신 러닝 모델 개발 프로세스

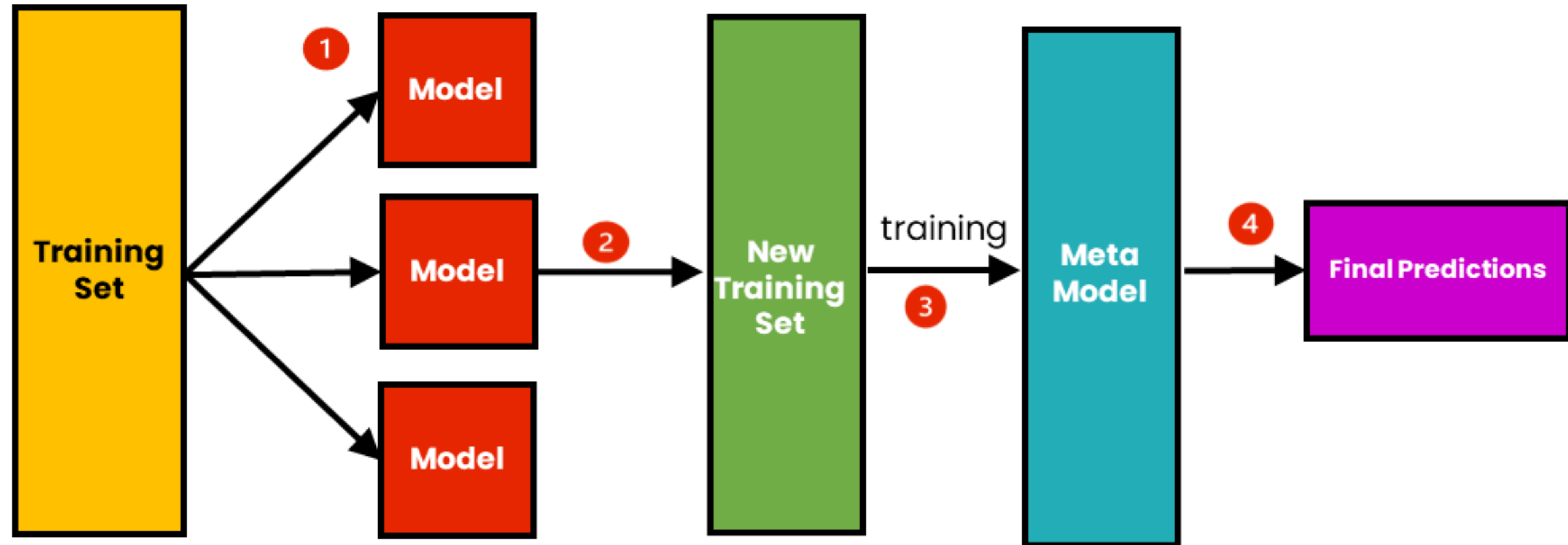


AutoML이란?

- 머신러닝의 전 과정을 자동화하는 기술
- 주요 자동화 대상:
 - 전처리
 - 모델 선택 및 튜닝
 - 앙상블
 - 성능 평가

Stacking

The Process of Stacking



Pycaret

- PyCaret: Python 기반 low-code AutoML 라이브러리
 - scikit-learn 기반의 low-code AutoML 라이브러리
 - Classification, Regression, Clustering, NLP 등 다양한 task 지원
 - 셋업부터 비교, 튜닝, 배포까지 일관성

Pycaret

- `setup()`: 데이터 자동 전처리 및 환경 설정
- `compare_models()`: 여러 모델을 자동 비교
- `tune_model()`: 하이퍼파라미터 튜닝
- `blend_models()`: 앙상블 구성
- `interpret_model()`: SHAP 기반 해석 가능

Autogluon

- Amazon에서 개발한 범용 AutoML 프레임워크
- Tabular, Text, Image 모두 지원
- 다양한 모델을 스택킹 기반으로 앙상블
- GPU 활용 가능, 딥러닝 기반 모델 포함

Autogluon - TabularPredictor

- TabularPredictor 주요 기능
- 다양한 모델(LGBM, CatBoost, NN 등)을 자동 탐색
- Feature preprocessing, selection 자동화
- Early stopping, 모델 앙상블 자동 적용
- 모델 성능 leaderboard 제공

Autogluon - 주요 메서드

- `fit()`: 데이터를 학습하여 예측 모델을 자동으로 생성. 다양한 모델을 훈련하고 앙상블
- `predict()`: 학습된 모델로 새로운 데이터에 대한 예측 수행2.
- `evaluate()`: 테스트 데이터에 대한 모델 성능 평가. 다양한 지표(정확도, RMSE 등) 사용 가능
- `feature_importance()`: 각 피처(변수)의 중요도를 계산하여 반환
- `leaderboard()`: 다양한 모델의 성능을 표로 정리하여 보여줌. 모델 비교에 활용