

# 하이브리드 Auto-sklearn 앙상블 모델을 이용한 댐 유입량 예측 및 평가

## Dam Inflow Prediction and Evaluation Using Hybrid Auto-sklearn Ensemble Model

이서로\* 배주현\*\*, 이관재\*\*\* 양동석\*\*\*\* 홍지영\*\*\*\*\*,김종건\*\*\*\*\* 임경재†

Seoro Lee, Joo Hyun Bae, Gwanjae Lee, Dongseok Yang, Jiyeong Hong,  
Jonggun Kim, Kyoung Jae Lim

### Abstract

최근 기후변화와 댐 상류 토지이용 변화 등과 같은 다양한 원인에 의해 댐 유입량의 변동성이 증가하면서 댐 관리 및 운영조작 의사 결정에 어려움이 발생하고 있다. 따라서 이러한 댐 유입량의 변동 특성을 반영하여 댐 유입량을 정확하고 효율적으로 예측할 수 있는 방안이 필요한 실정이다. 머신러닝 기술이 발전하면서 Auto-ML(Automated Machine Learning)이 다양한 분야에서 활용되고 있다. Auto-ML은 데이터 전처리, 최적 알고리즘 선택, 하이퍼파라미터 튜닝, 모델 학습 및 평가 등의 모든 과정을 자동화하는 기술이다. 그러나 아직까지 수문 분야에서 댐 유입량을 예측하기 위한 모델을 개발하는데 있어서 Auto-ML을 활용한 사례는 부족하고, 특히 댐 유입량의 예측 정확성을 확보하기 위해 High-inflow and low-inflow의 변동 특성을 고려한 하이브리드 결합 방식을 통해 Auto-ML 기반 앙상블 모델을 개발하고 평가한 연구는 없다. 본 연구에서는 Auto-ML의 패키지 중 Auto-sklearn을 통해 홍수기, 비홍수기 유입량 변동 특성을 반영한 하이브리드 앙상블 댐 유입량 예측 모델을 개발하였다. 소양강댐을 대상으로 적용한 결과, 하이브리드 Auto-sklearn 앙상블 모델의 댐 유입량 예측 성능은  $R^2$  0.868, RMSE 66.23  $m^3/s$ , MAE 16.45  $m^3/s$ 로 단일 Auto-sklearn을 통해 구축된 앙상블 모델보다 전반적으로 우수한 것으로 나타났다. 특히 FDC (Flow Duration Curve)의 저수기, 갈수기 구간에서 두 모델의 유입량 예측 경향은 큰 차이를 보였으며, 하이브리드 Auto-sklearn 모델의 예측 값이 관측 값과 더욱 유사한 것으로 나타났다. 이는 홍수기, 비홍수기 구간에 대한 앙상블 모델이 독립적으로 구축되는 과정에서 각 모델에 대한 하이퍼파라미터가 최적화되었기 때문이라 판단된다. 향후 본 연구의 방법론은 보다 정확한 댐 유입량 예측 자료를 생성하기 위한 방안 수립뿐만 아니라 다양한 분야의 불균형한 데이터셋을 이용한 앙상블 모델을 구축하는데도 유용하게 활용될 수 있을 것으로 사료된다.

**핵심용어** : 댐 유입량, Auto-ML, 하이브리드 Auto-sklearn, 앙상블

### 감사의 글

본 결과물은 환경부의 재원으로 한국환경산업기술원의 수생태계 건강성 확보 기술개발사업의 지원을 받아 연구되었습니다.(2020003030004)

\* 정회원 · 강원대학교 지역건설공학과 박사과정 수료 · E-mail : seorolee91@gmail.com

\*\* 정회원 · 한국수계환경연구소 박사 후 연구원 · E-mail : baegop@pusan.ac.kr

\*\*\* 정회원 · 강원대학교 지역건설공학과 박사과정 수료 · E-mail : gwanjae2@gmail.com

\*\*\*\* 정회원 · 강원대학교 지역건설공학과 박사과정 수료 · E-mail : dsyang1024@gmail.com

\*\*\*\*\* 정회원 Department of Earth and Environment, Boston University · E-mail : jyhong@bu.edu

\*\*\*\*\* 정회원 · 강원대학교 지역건설공학과 교수 · E-mail : kimjg23@gmail.com

† 정회원 · 강원대학교 지역건설공학과 교수 · E-mail : kjlim@kangwon.ac.kr