

# 비점오염원관리지역의 머신러닝 기법을 통한 수질 예측 가능성 연구

## Study on Water Quality Predictability through Machine Learning Techniques in Non-point Pollutant Management Area

유나영\*, 신민환\*\*, 금동혁\*\*\*, 임경재\*\*\*\*, 김종건\*\*\*\*\*

Yeong Na Yu, Min Hwan Shin, Dong Hyuk Kum, Kyoung Jae Lim, Jong Gun Kim

### 요 지

강우에 의해 발생하는 비점오염물질의 수질 데이터가 충분하지 않아 비점오염원이 문제가 되고 있는 유역의 수질개선을 위한 대책마련이 어려운 실정이다. 기존에 환경부에서 운영하고 있는 자동측정망은 1시간 간격으로 데이터를 축적하고 있으나, 비점오염원이 문제가 되는 유역에 설치되어 있지 않거나 수온, DO, pH 등 현장항목만을 측정하고 있어 하천의 수질오염을 대표할 수 있는 T-P나 SS 등의 수질분석 항목의 부재하다. 이로 인해 유역의 수질개선 대책을 수립하기 위한 오염원의 현황을 파악하기 어려운 실정이다. 따라서, 본 연구에서는 비점오염원관리지역 중 골지천 유역을 대상으로 수질항목별 상관성을 분석하고, 실측자료를 기반으로 DT, MLP, SVM, RF, GB, XGB 등의 머신러닝 기법을 통해 수질 예측 가능성을 연구하였다.

상관관계 분석결과 입력변수인 탁도 항목이 예측 수질과 뚜렷한 상관관계를 보이는 것으로 나타났다. 그 외 항목에서는 약한 상관관계를 보이거나 상관관계가 없는 것으로 나타났다. 머신러닝 기법을 활용한 수질 예측 분석 결과, 검무교와 태봉2교, 제1여량교는 RF 기법에서 결정계수 ( $R^2$ ) 0.57~0.86, RMSE 16.49~175.60으로 예측성이 우수한 것으로 나타났다. 관말교는 SVM 기법에서  $R^2$  0.65, RMSE 57.69로, 송계교는 XGB 기법에서  $R^2$  0.74, RMSE 282.86으로 가장 예측성이 우수한 것으로 나타났다. 분석결과와 같이 머신러닝 기법을 활용한 수질 예측은 가능하나, 예측성이 우수한 머신러닝 기법의  $R^2$  비교 결과, 유역면적이 큰 제1여량교와 작은 관말교에서 0.57과 0.65로 다른 지점에 비해 낮은 것으로 나타났다. RMSE 비교 결과, 상류 산간지역에 발생한 국지성 호우의 영향으로 흙탕물이 가장 자주 발생하는 태봉2교 지점과 우선관리지역이 합류되는 송계교 지점에서 175.60과 282.86으로 예측값과 실측값의 오차가 큰 것으로 나타났다.

연구결과와 같이 하천 수질을 예측하기 위해서는 유역면적 혹은 유역특성과 관련한 기초자료를 추가로 적용하여 머신러닝 기법을 적용 해야할 것으로 판단된다. 또한, 본 연구에서 예측한 수질 항목 이외에 입력변수를 추가로 확보하여 수질의 예측 가능성을 검토해야 할 것으로 보여진다.

**핵심용어** : 머신러닝, 상관관계, 결정계수( $R^2$ ), RMSE, 수질예측

\* 정회원 · (주)이엠연구소 부장 · E-mail : [yiny129@naver.com](mailto:yiny129@naver.com)

\*\* 정회원 · (주)이엠연구소 대표이사 · E-mail : [uv2000wind@nate.com](mailto:uv2000wind@nate.com)

\*\*\* 정회원 · (주)이엠연구소 이사 · E-mail : [kumdong@hotmail.com](mailto:kumdong@hotmail.com)

\*\*\*\* 정회원 · 강원대학교 농업생명과학대학 지역건설공학과 교수 · E-mail : [kyoungjaelim@gmail.com](mailto:kyoungjaelim@gmail.com)

\*\*\*\*\* 정회원 · 강원대학교 농업생명과학대학 지역건설공학과 교수 · E-mail : [kimjg23@gmail.com](mailto:kimjg23@gmail.com)