

## **Machine learning models for water distribution systems: Best practices and recommendations**

\*Donghwi Jung<sup>1)</sup> and Sanghoon Jun<sup>2)</sup>

1), 2) *School of Civil, Environmental and Architectural Engineering, Korea University, Seoul 02841, Korea*

<sup>1)</sup> [sunnyjung625@korea.ac.kr](mailto:sunnyjung625@korea.ac.kr)

<sup>2)</sup> [sanghoonjun1028@gmail.com](mailto:sanghoonjun1028@gmail.com)

### **ABSTRACT**

Recently, machine learning models have emerged as a superior alternative to traditional statistical or other common methods for addressing various design, operational, and management problems in water distribution system (WDS). While numerous publications exist, there has been limited effort devoted to reviewing them, and comprehensive guidelines for the entire machine learning modelling pipeline (from project setup and data handling to model construction and applications) are lacking. To bridge this gap, this study begins by reviewing state-of-the-art research in WDS, including topics such as pipe burst detection, surrogate modelling for system analysis, and customer demand prediction. Subsequently, we summarize common mistakes, best practices, and recommendations across different stages of machine learning modelling. Through this comprehensive review, we aim to enhance the quality and reliability of WDS modelling efforts.

### **Acknowledgments**

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2021R1A5A1032433).

---

<sup>1)</sup> Associate Professor

<sup>2)</sup> Research Professor