# Use of big data analytics in machine learning to improve management of water distribution system with smart water grid in Korea

**KANG MIN KOO**[1], KANG MIN KOO[1], GYUMIN LEE[3], KYUNG TAEK YUM[4], KYUNG SOO JUN[1], KYUNG SOO JUN[*2]

*[1]Graduate School of Water Resources, Sungkyunkwan University, [2]Sungkyunkwan University, [3] Construction and Environmental Research Center, Sungkyunkwan University, [4]Smart Water Grid Research Group, Sungkyunkwan University*

**(a) Purpose of study or research hypothesis**

In order to increase the efficiency of water management in the Water Distribution System (WDS), the big data analysis techniques of machine learning were applied to the water consumption data in the Smart Water Grid (SWG). Detailed study contents include the pre-processing technique for raw data, analysis of water consumption patterns, and forecast of water demand.

**(b) Key issue(s) or problem(s) addressed**

Nowadays, water demand is rapidly increasing due to population growth and urbanization, and the efficiency of the WDS decreases as the aging progresses. By combining conventional water management techniques with the SWG using ICT (sensors, communication devices, servers, etc.), it is possible to collect water consumption data in real-time. However, it is necessary to develop the management process and operation technology to efficiently manage the WDS using big data.

**(c) Methodology or approach used**

The process of applying the big data analysis technique of machine learning for water consumption data is divided into three stages. First, correct missing and outlier (data pre-processing) using Lagged k-NN and Fast Fourier transform. Second, estimate the water consumption pattern by applying k-means clustering. Finally, forecast water demand using Long-Short Term Memory.

**(d) Results or conclusions derived from the project**

First, if the lagged k-NN and Fast Fourier transform are applied for missing and outlier, there is an advantage that daily data can be corrected even if all hourly water usage data is missing.

Second, it was possible to estimate consumption patterns according to usage, pipe diameter, and day of the week using k-means clustering, helping to understand consumer characteristics.

Finally, long-short term memory, one of the deep neural networks, can be used for water demand forecasting more precisely.

However, machine learning relies on big data, the error may increase if the reliability of the data is low or the amount of data is small. In addition, only water use data is used for water demand forecasting, it is necessary to further study other factors affecting water demand.

**(e) Implications of the project relevant to congress themes**

This project is closely related to smart and ICT among the themes of the congress. Because it aims to maximize water management efficiency through the development of process and operational technologies that use machine learning technology for water consumption data in the smart water grid.

**Keywords** : Smart water grid, Water distribution networks, Machine learning, Big data, Water consumption