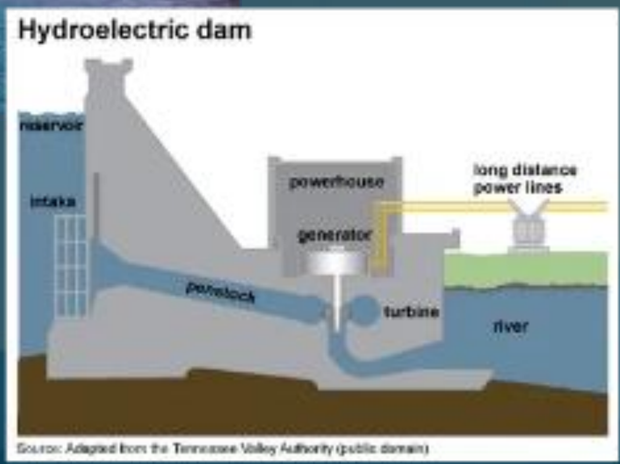
Hydropower Discharge Prediction

Yerin Lee(20240381),
KENTECH, School of energy Engineering



1. Title & Authors

Hydropower Turbine Discharge Prediction Using Machine Learning
Author: YerinLee(20240381)
KENTECH, School of Energy Engineering

2. Motivation & Problem Definition

Hydropower generation depends heavily on how much water flows through the turbine (turbine discharge).
However, turbine discharge is influenced by multiple hydrological and operational factors, and real dam operations are nonlinear and complex.

Goal of this project:
To build a machine learning model that predicts turbine discharge based on real K-water dam data and evaluate which model performs best.

Why it matters:

- Helps understand operational behavior of dam
- Supports hydropower forecasting & operation planning
- Demonstrates how ML handles real-world nonlinear systems

3. Method

1. Idea & Motivation

Hydropower output depends on turbine discharge.

Goal: build a model that predicts turbine discharge using real dam operation data

2. Dataset Construction

Dataset: K-water Daily Hydrological Data (421,451 entries), csv

Preprocessing:
-Remove missing data
-Filter out nonphysical values (≤0)
-Randomly sample 50,000 rows
-Input (X): hydrological & operational variables
-Output (y): turbine discharge

3. Model Development

Compared three models:
-Linear Regression
-Polynomial Regression (degree = 2)
-Random Forest Regressor

4. Result Analysis

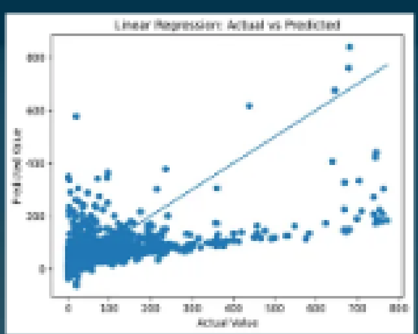
-Trained and tested with 80/20 split
-Evaluated performance using RMSE, MAE, R^2
-Scatter plots (Actual vs Predicted)
-Error comparison charts
Feature importance ranking
-Hyperparameter experiment

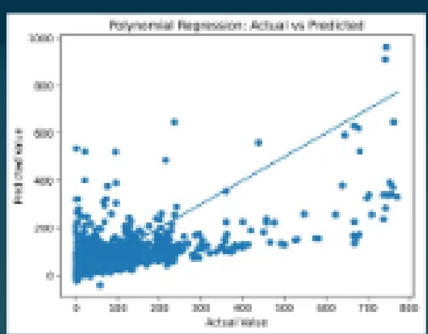
5. Distribution

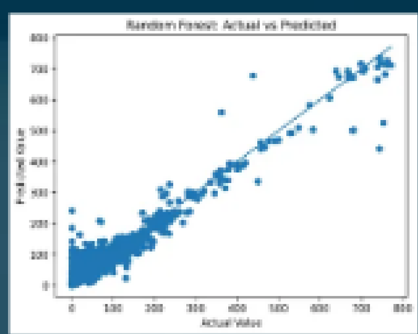
Full code and analysis provided via GitHub

Clear results showing Random Forest achieves $R^2 > 0.95$
Provides foundation for hydropower operation prediction tools

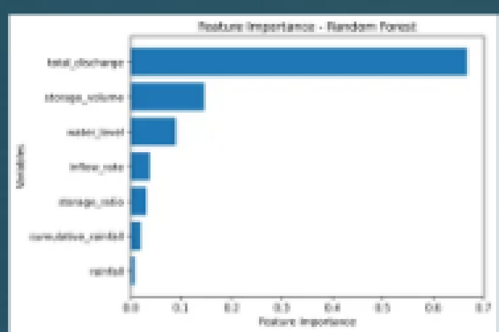
4. Results







	Linear Regression	Polynomial Regression	Random Forest
MSE	1933.244	1593.737	165.101
R^2	0.441	0.539	0.952
MAE	23.058	19.730	5.169
RMSE	43.969	39.922	12.849



- Random Forest achieved $R^2 > 0.95$, demonstrating excellent predictive performance.
- In the hyperparameter experiment, the best configuration was n_estimators = 50 and max_depth = 20, resulting in RMSE = 11.48 and $R^2 = 0.962$.
- In Random Forest model, the model became more accurate as the tree depth increased
- Feature importance analysis showed that total discharge was the most influential variable, followed by storage volume and water level, indicating that both hydrological conditions and reservoir status strongly affect turbine discharge.

5. Discussion & Future work

Key Takeaways

-Linear and polynomial models fail to capture nonlinear dam behavior.
-Random Forest performs the best with very low error.
-Total discharge, storage volume, and water level strongly influence turbine discharge.

Future Improvements

-Include hydraulic head (H)
-Try gradient boosting or neural networks
-Consider seasonal or temporal effects

6. Reference & Acknowledgement

- Dataset: K-water hydrological operation data
한국수자원공사_수문현황정보_일별
https://www.data.go.kr/data/15083335/fileData.do
- scikit-learn documentation
- Image:
https://www.eia.gov/energyexplained/hydropower/
https://water.go.kr/letter/40/sub02.html