Title: Analyzing Wastewater Treatment Plant Data for Energy Consumption and Environmental Factors

Abstract

This study analyzes data from a wastewater treatment plant, focusing on energy consumption and its correlation with environmental and operational parameters. Using machine learning models and statistical visualizations, we identify key trends in temperature, humidity, atmospheric pressure, and pollutant levels. The research provides insights into optimizing plant operations while maintaining sustainability.

1. Introduction

Wastewater treatment plants (WWTPs) play a critical role in ensuring environmental sustainability. Optimizing their energy consumption is vital for reducing operational costs and environmental impact. This study examines various environmental factors and their influence on energy consumption using statistical analysis and machine learning models.

2. Data Overview

The dataset comprises multiple features, including:

- **Operational Factors:** Energy consumption, inflow/outflow levels, ammonia concentration, biological oxygen demand, and chemical oxygen demand.
- **Environmental Parameters:** Temperature, humidity, atmospheric pressure, wind speed, and rainfall.
- **Time Factors:** Year, month, and day.

Preliminary data inspection included:

- Checking for missing values.
- Statistical distribution analysis.
- Identifying outliers and trends.
- Data visualizations using histograms, KDE plots, and box plots to detect anomalies and understand variable distributions.

3. Exploratory Data Analysis (EDA)

3.1 Feature Distributions

Several histograms and KDE plots were used to analyze the distribution of numerical variables:

- **Temperature Distribution:** Skewed towards moderate temperatures with occasional high values.
- **Humidity Distribution:** Normally distributed, peaking at around 60%–70%.
- **Atmospheric Pressure:** Shows extreme values, requiring further investigation.
- **Total Rainfall & Visibility:** Skewed towards lower values, indicating minimal heavy rainfall events.

3.2 Correlation Analysis

- **Energy Consumption vs. Environmental Factors:**
    - Higher energy consumption correlates with extreme temperature fluctuations.
    - Humidity and rainfall appear to have a negligible effect.
    - Ammonia and biological oxygen demand (BOD) showed significant correlations with energy consumption, suggesting operational dependencies.
- **Outlier Detection:** A box plot revealed that energy consumption had significantly higher values compared to other features, indicating potential inefficiencies or peak operational loads.

4. Machine Learning Models

Several models were applied to predict energy consumption:

4.1 Regression Models

- **Linear Regression:** Provided a baseline model with limited predictive power.
- **Decision Tree & Random Forest Regressor:** Captured non-linear dependencies effectively and improved prediction accuracy.
- **Gradient Boosting Regressor:** Outperformed others in prediction accuracy, demonstrating strong learning capabilities from the given data.

4.2 Deep Learning Models

- **LSTM & GRU Networks:** These time-series models captured seasonal variations in energy consumption more effectively than traditional regression models.
- **Results:** The deep learning models outperformed other approaches in predictive accuracy, confirming the importance of temporal dependencies in wastewater treatment operations.

4.3 Model Comparison

A comparative analysis of different models was performed based on **Mean Squared Error (MSE) and Root Mean Squared Error (RMSE):**

- **Neural Network:**
  - Mean Squared Error (MSE): **9144.96**
  - R-squared (R2) Score: **0.54**
- **Feedforward Neural Network (FNN):**
  - RMSE on test set: **148.35**
- **Convolutional Neural Network (CNN):**
  - RMSE on test set: **124.51**
- **Recurrent Neural Network (RNN):**
  - RMSE on test set: **111.22**
- **Simple Feedforward Neural Network Model:**
  - RMSE on test set: **90.83**

This comparison highlights that the **Simple Feedforward Neural Network Model achieved the lowest RMSE**, making it the most effective predictive model for energy consumption in this study.

5. Discussion and Key Findings

- **Energy Consumption Trends:** Fluctuations align with seasonal variations and external environmental conditions.
- **Feature Importance:** Decision tree-based models highlighted key contributing factors like inflow levels, ammonia levels, and maximum temperature.
- **Potential Optimization Strategies:**
  - Implementing predictive energy management based on weather forecasts.
  - Adjusting operational settings dynamically to reduce excess energy consumption.
  - Using AI-based automation to enhance efficiency and reduce unnecessary power use.

6. Conclusion

This research demonstrates the impact of environmental factors on wastewater treatment plant energy consumption. Machine learning models, particularly gradient boosting and deep learning networks, provide effective predictive capabilities. The study highlights the importance of temperature fluctuations, ammonia levels, and BOD in driving energy consumption. Future work should explore real-time monitoring systems for more dynamic optimization.