README: Scour Prediction Model Using USGS Field Data

Project Overview

This repository provides a machine learning-based model for predicting scour depth at bridge piers using USGS field data. The model leverages XGBoost regression to predict dimensionless scour depth (ys/b) based on key hydraulic and geometric parameters, ensuring conservative design for practical engineering applications.

Key features include:

- - Data preprocessing and feature engineering for dimensionless variables.
- - Training, validation, and testing of the model with performance metrics.
- - SHAP analysis for feature importance and interpretability.
- - Prediction with user-defined input and optional conservatism factors.

Requirements

This project requires Python 3.8+ and the following libraries:

- - xgboost
- - numpy
- - pandas
- - scikit-learn
- - matplotlib
- - seaborn
- - shap

Install the required packages using the following command:

pip install -r requirements.txt

Getting Started

1. Clone the Repository

```*bash*
*git clone https://github.com/<your-username>/scour-prediction-model.git*
*cd scour-prediction-model*
```

2. Prepare Data

Place your input CSV file (e.g., Scour_Field.csv) in the project directory. Ensure it follows the structure:
b, V0, Vc, y0, d50, ys
where:

- b: pier width (ft)
- V0: mean flow velocity (ft/s)
- Vc: critical velocity (ft/s)
- y0: upstream flow depth (ft)
- d50: median particle size (mm)
- ys: observed scour depth (ft)

3. Run the Model

To train and evaluate the model:

```bash
python scour_model.py
```

4. Make Predictions

Provide user-defined inputs to predict scour depth interactively or calculate predictions for a dataset. The code also provides multiplying factors to ensure conservatism in design.

Project Structure

```
.
├── scour_model.py          # Main script for training and evaluating the model
├── Scour_Field7.csv        # Example input data (optional)
├── requirements.txt        # Python dependencies
├── README.md               # Project documentation
├── Total_validation.png    # Example visualization of results
└── shap_result.csv         # Example SHAP results for feature importance
```

Features

Model Training and Evaluation

The model is trained using the following dimensionless variables:
- y/b: Flow depth to pier width ratio.
- b/d50: Pier width to particle size ratio.
- V/Vc: Velocity to critical velocity ratio.
- Frb: Froude number based on pier width.

Evaluation metrics include RMSE, NMSE, IA, SI, NSE, R2, Bias, and Standard Error (SE).

SHAP Analysis

The repository provides interpretability using SHAP (SHapley Additive exPlanations) for:
- Feature importance rankings.
- Waterfall plots for individual predictions.

Conservatism Factors

Multiplying factors can be applied to predictions to ensure safety in design. Users can calculate factors for specific conservatism levels (e.g., 95%).

Results

Example results include:
- Training, validation, and testing performance metrics.
- Plots for measured vs. predicted scour depth with confidence intervals.
- MAPE and conservatism levels for various multiplying factors.

Usage Examples

Interactive Prediction

Run the script and input the following parameters when prompted:
- Pier width (b)
- Pier length (L)
- Skew angle (θ)
- Mean flow velocity (V)
- Upstream flow depth (y)
- Median particle size (d50)
- Conversion constant (Ku, typically 0.3048)

SHAP Feature Importance

Run the script to generate SHAP summary plots and feature importance visualizations:
```*bash*
*python shap_analysis.py*
```

Contact

For questions or collaborations, please reach out via email at tkim5@pknu.ac.kr.