

Влияние статических признаков на предсказание речного стока методами ML

Проблема эквивиальности в моделировании гидрологических процессов

Эльдар Хузин

МФТИ

1 апреля 2025 г.

Содержание

- 1 Введение
- 2 Статус и планы
- 3 Конец

Постановка задачи

- Данные:
 - временные ряды метеорологических данных (температура, осадки),
 - временные ряды стока и уровня воды,
 - статические характеристики водосборов (лесистость, глубина водоностного слоя).
- Целевые переменные: сток и уровень воды.
 - Зная данные за год, предсказать следующую неделю.
- Оптимизационная задача:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\% \quad (\text{MAPE})$$

Модели и эквивалентность

- Quantile Regression Forests
- Gradient Boosting (CatBoost)
- Recurrent Neural Networks (RNN)
- Long Short-Term Memory (LSTM)
- Потенциально фантазия на тему transformer'ов

Статус

- Написаны "сборники" датасетов.
- Catboost обучался, сейчас — переработка под SHAP.

Планы

- Завершить SHAP-адаптацию и переобучение.
- Запустить новые модели для сравнения.

Конец