

Презентация решения к Заданию AIJ

NoFloodWithAI: паводки на реке Амур

План презентации:

1. Описание задачи.
2. Решение задачи.
3. Оценка получившейся модели.

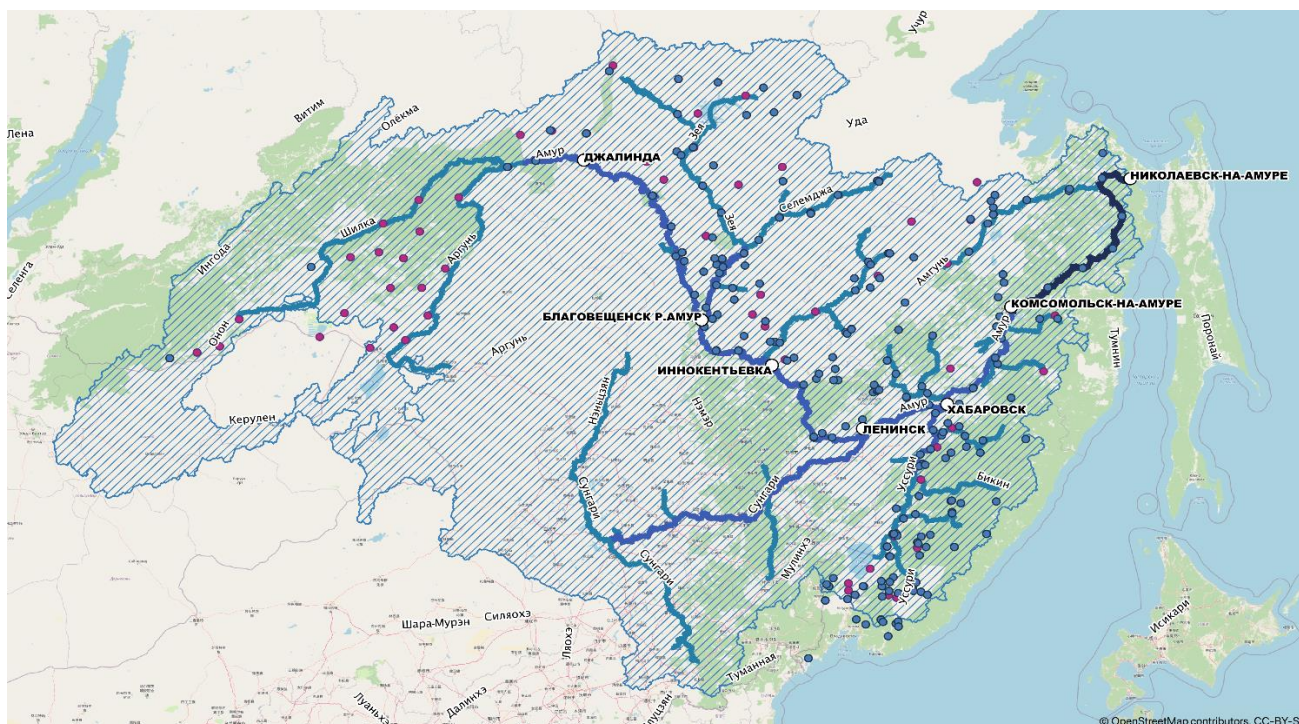
Описание задачи.

Участникам соревнования предлагается разработать алгоритм краткосрочного прогнозирования уровней воды р. прогнозирования уровней воды р. Амур для следующих населенных пунктов (гидропостов): Джалинда, Благовещенск, Иннокентьевка, Ленинское, Хабаровск, Комсомольск-на-Амуре, Николаевск-на-Амуре на 10 дней вперед.

Река Амур является трансграничной рекой, основная часть бассейна которой находится в пределах Российской Федерации. Для Амура характерна низкая водность в зимний период, небольшие половодья весной и неоднократные резкие подъемы воды во второй половине лета и в начале осени. Маловодные периоды сменяются годами большой воды. В многолетнем режиме водного стока Амура отчетливо выражено чередование периодов пониженной и повышенной водности, каждый продолжительностью 12-17 лет. Амур по оценке гидрологов и исходя из истории наблюдений вошел в очередной период высокой водности в конце 2000-х. Исходя из данной гидрологической закономерности режима Амура, в ближайшие 5-7 лет следует ожидать сложную паводковую обстановку в течении Среднего и Нижнего Амура (наиболее сложная обстановка от слияния р. Сунгари и до Комсомольского района включительно).

Наиболее крупномасштабные наводнения произошли в 2013 и 2019 гг. Причиной наводнений стали тропические циклоны, которые несли теплый влажный воздух, вызывали фронтальные разделы и сильные атмосферные осадки. В 2013 на значительной площади за 2-3 мес. сумма выпавших осадков превысила годовую, а местами и полугодовую норму.

В целях минимизации экономического ущерба необходимо создать инструмент для прогнозирования паводковых волн на Амуре и его притоках.



На карте выше представлен бассейн реки Амур и его основные притоки с градацией по среднему расходу воды. Синие точки — все гидропосты из датасета, белые — интересные гидропосты, розовые — метеостанции.

Решение задачи.

За основу было взято baseline решение от организаторов размещенное: https://github.com/sberbank-ai/no_flood_with_ai_aij2020. К имеющимся признакам были добавлены:

1. `day_temperature_air_90_10` (средняя дневная температура воздуха за последние 90 дней, 10 дней назад) и `night_temperature_air_90_10` (средняя ночная температура воздуха за последние 90 дней, 10 дней назад) - наличие весной продолжительных ночных заморозков, сопровождающееся дневными плюсовыми температурами, способствует снижению уровня воды;
2. количество рассматриваемых лет увеличено до 30 (от `past_1_0` до `past_30_10`) – в резких наводнениях р. Амур во всех источниках отмечается цикл, не превышающий 30 лет.

Так как в дальнейшем использую библиотеку `heapy`, чтобы избежать обучение на данных из будущего, беру данные об уровнях реки от 11 дня (с `ts_11` по `ts_20`). Отсюда получаю признак `ts_raise` (`ts_20/ ts_11`), показывающий прирост/снижение уровня воды за последние 10 дней.

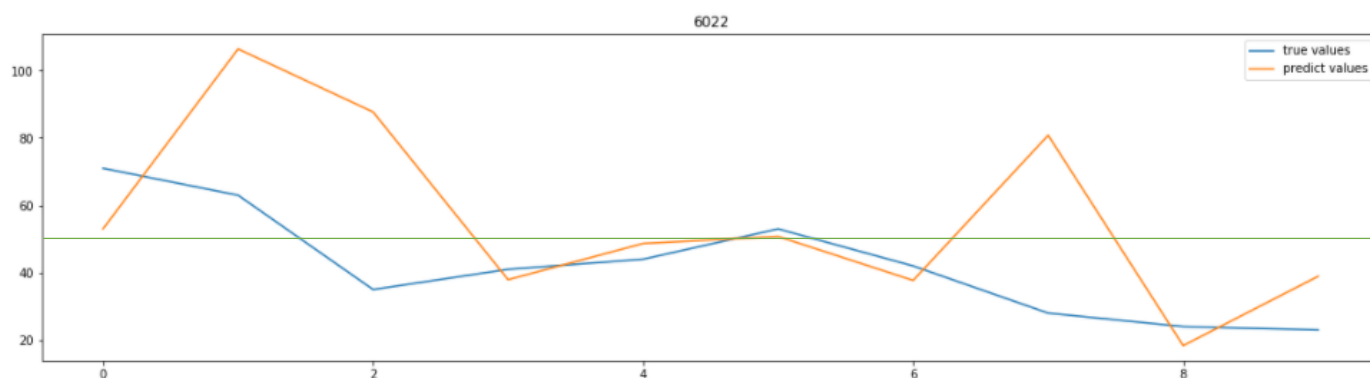
Для получения дополнительных признаков использую модель `Prophet`. В итоге получаем признаки: `'yhat'` (уровень воды в реке по данным модели), `'trend'` (тренд изменения уровня воды), `'daily'` (дневные значения уровня воды по данным модели).

Как уже упоминалось, для обучения используем библиотеку `heapy`, а точнее, одну из схем стекинга – блендинг. В качестве базовых алгоритмов используем `CatBoostRegressor` и `XGBRegressor`. Прогнав модели через блендинг в пропорции 20/80 (test/train), получаем метапризнаки, которые подаем на вход метаалгоритму `ExtraTreesRegressor` после чего получаем ответ.

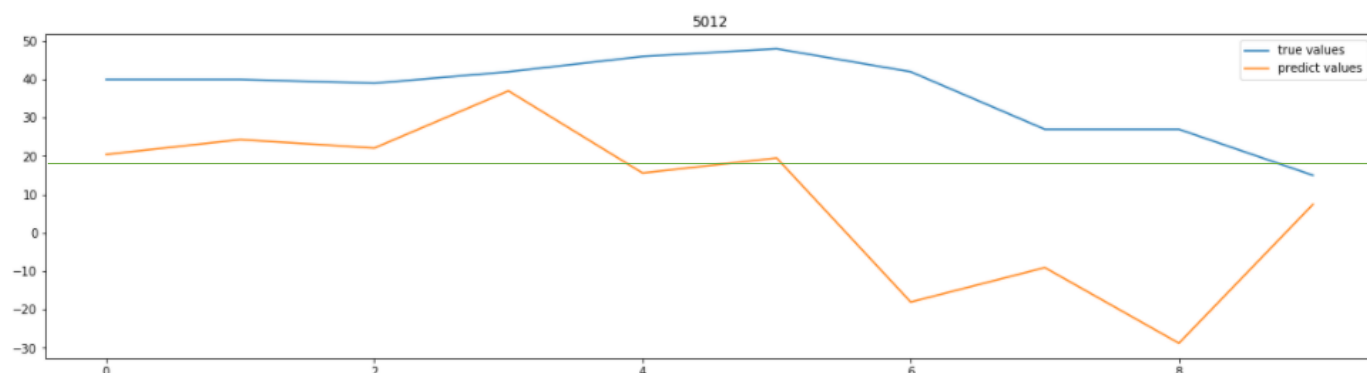
Оценка получившейся модели.

На период с 01.11.2019 по 11.11.2019 имеем следующие предсказания по городам: Благовещенск (6022), Хабаровск (5012).

6022 MAE: 20.28 MEDIAN: 49.715



5012 MAE: 27.549 MEDIAN: 17.555

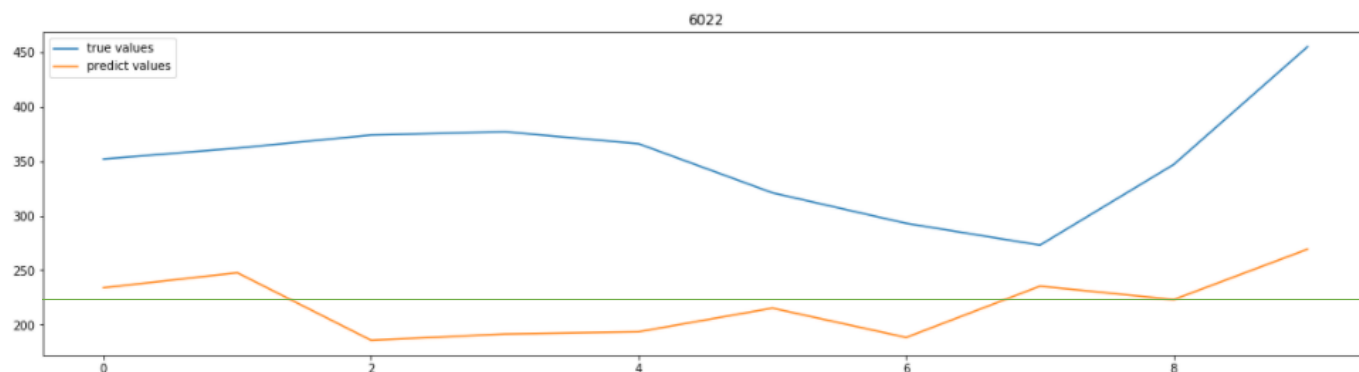


Несмотря на высокое значение MAE в целом мы наблюдаем несущественное отклонение от медианы предсказанного массива фактических показателей уровня воды.

Но стоит рассмотреть получившуюся модель на более экстремальных периодах для данных городов: для Благовещенска (6022) - с 13.07.2013 по 23.07.2013 и с 15.07.2013 по 25.07.2013, для Хабаровска (5012) – с 03.08.2013 по 13.08.2013 и с 05.08.2013 по 15.08.2013.

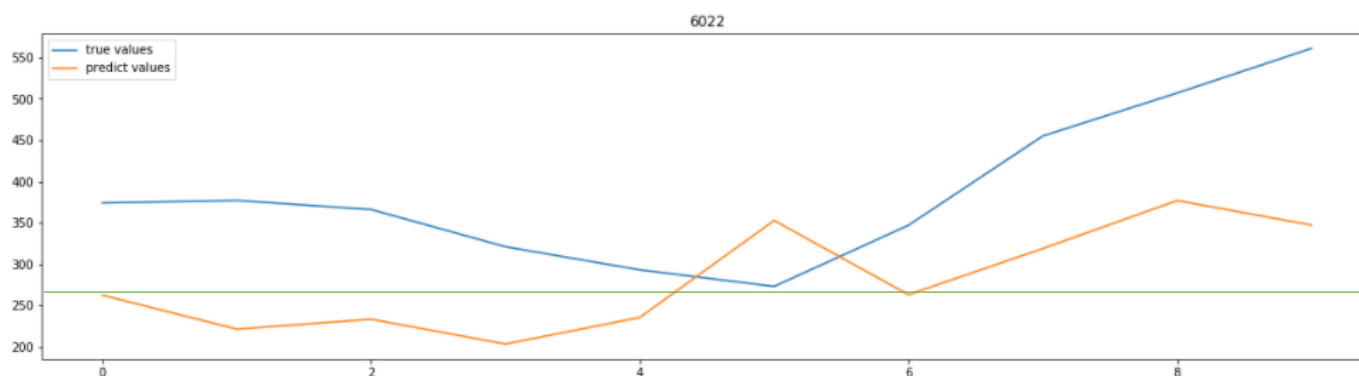
с 13.07.2013 по 23.07.2013

6022 MAE: 133.6635000000003 MEDIAN: 219.12



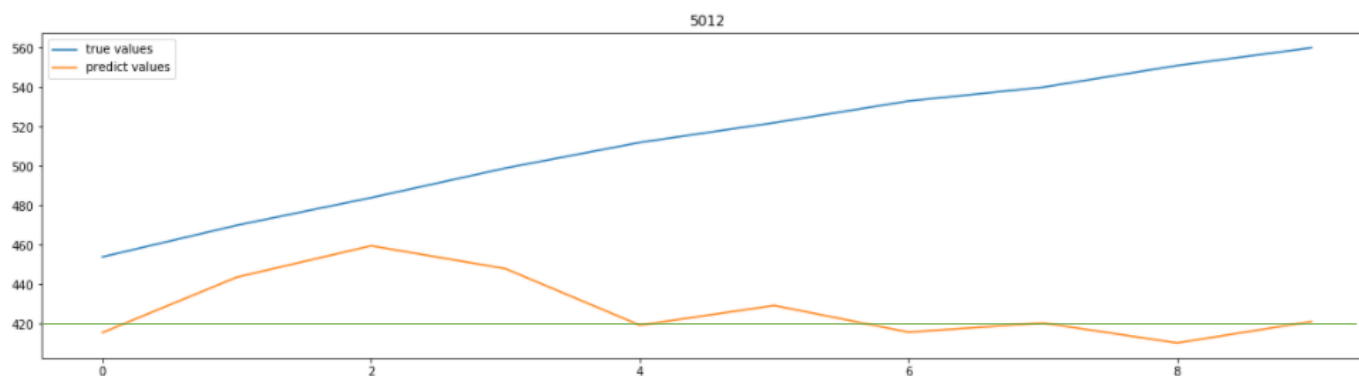
с 15.07.2013 по 25.07.2013

6022 MAE: 121.915 MEDIAN: 262.685



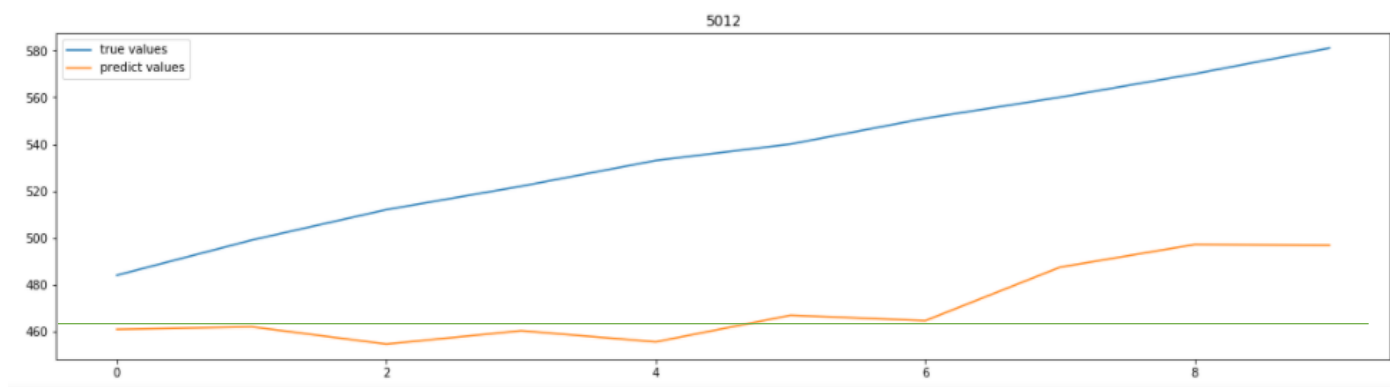
с 03.08.2013 по 13.08.2013

5012 MAE: 84.155 MEDIAN: 420.855



с 05.08.2013 по 15.08.2013

5012 MAE: 64.576 MEDIAN: 463.3



К сожалению, в такие периоды модель не способна давать приближенно точный прогноз об уровне воды в реке. Тренд на увеличение уровня воды прослеживается не явно и здесь лучше смотреть на изменение медианы за несколько периодов. В целом можно отметить, что в период паводка ошибка предсказания модели уровня воды в отдельные дни не превышала 2 метров, для решения реальных практических задач это много, но тем не менее это дает возможность полагать, что модель имеет предсказательную способность даже в таких чрезвычайных ситуациях.

Таким образом, данную модель можно рекомендовать к использованию не отдельно, а с уже имеющимися средствами прогнозирования паводков на р. Амур.