


Создание модели для предсказания погодных условий

Выполнила команда: Easy_hack

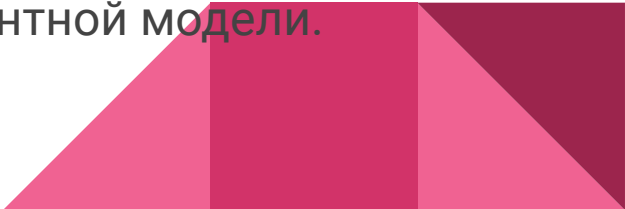
Постановка задачи

- 1) Имеется датасет с погодными показателями за 43 часа наблюдений, сделать предсказание этих же параметров на следующие 5 часов.
- 2) Загрузить результаты в Telegram bot, получить оценку качества. Команды соревнуются в точности предсказаний (можно делать неограниченное число попыток).

Данные имеют вид замеров различных показателей в точках (latitude, longitude) на сетке 30x30 с шагом около 5 км. Все массивы кроме высот в качестве первой размерности имеют шкалу времени с шагом в 1 час.



Идеи для реализации

- 1) Коэффициентная модель - создает коэффициенты между вчерашним и сегодняшним днем для каждой точки на карте, затем умножает вчерашние вечерние погодные параметры на эти коэффициенты.
 - 2) Catboost. Для Catboost были созданы фичи в виде: количество часов в течение суток, разница параметров погоды за последние один, два ..., пять часов.
 - 3) Elastic (не используется). Фичи использовались такие же, как и в catboost.
 - 4) Mix модель - объединение Catboost и коэффициентной модели.
- 

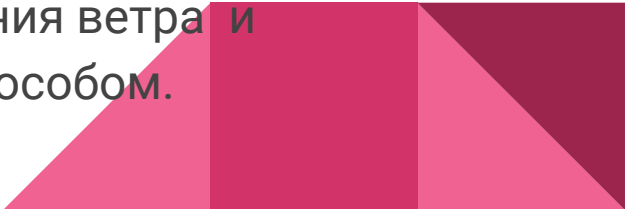
Mix модель

Данная модель представляет собой мерж предсказаний от коэффициентной модели и catboost.

Зачем нужен mix? На тестовом датасете было выявлено, что ошибка у catboost меньше для давления, направления ветра и облачности.

Соответственно коэффициентная модель предсказывала температуру, влажность и скорость ветра. Было принято решение создать mix, которая включает в себя две модели.


Mix модель применяет коэффициенты для направления ветра и облачности. Они были рассчитаны эмпирическим способом.



Коэффициентная модель

Под капотом модель делит значения параметров погоды за вчера на значения за сегодня, затем получившиеся коэффициенты усредняются за 19 часов, и получается множитель для конкретной ячейки на карте. Чтобы получить предсказание, достаточно рассчитанные коэффициенты умножить на значения параметров погоды с 19 по 23 часов.

Данная модель предсказывает параметры погоды (MAPE на submission):


- 1) температура (4.81%);
 - 2) влажность (10.8%);
 - 3) скорость ветра (47.3%).
- 

Catboost

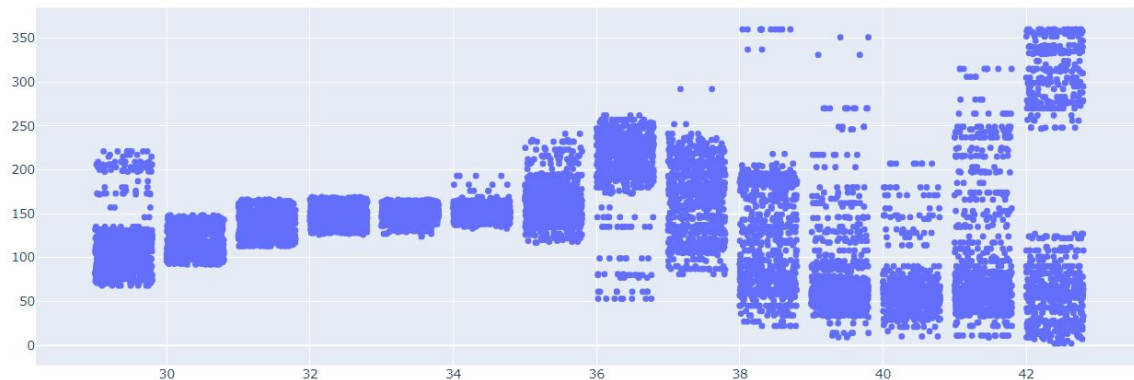
При работе с Catboost основной упор был сделан на создание новых фичей. Были добавлены фичи в виде времени в течение суток и производные параметров погоды с шагом от 1 до 5 часов.

Был обучен набор моделей, каждая из которых предназначена для предсказания определенного параметра погоды на определенный период.

Данная модель предсказывает параметры погоды (MAPE на submission):

- 1) давление (0.1%);
 - 2) направление ветра (99.8%);
 - 3) облачность (57.8%).
- 

О чем нельзя молчать!



Здесь представлен график направления ветра от времени. Можно заметить, что график раздвоился на 42 часе, что выглядит очень странно. Было принято решение брать значение параметра погоды менее 100 градусов и только последние 5 часов обучающего датасета. Благодаря этому ошибка уменьшилась почти на 100%.

Итоги

Удалось создать `mix` модель, которая отлично предсказывает значения давления, температуры и влажности, но имеет большую ошибку ($\sim 100\%$) при предсказании направления ветра.

Средняя ошибка по всем погодным показателям составила 36.8%.

Данная модель включает в себя преимущества сразу двух подходов, метрический расчет коэффициентов пропорциональности и модель машинного обучения в виде Catboost.

