

# Прогноз самовозгорания угля при открытом хранении

Работу выполнили Тakaев Денис и Лобан  
Филипп



# Какие данные имели?



Fires – возгорания



Supplies - поставки угля



Weather - погода



Temperature –  
температура в  
штабеле

# Проблемы с которыми столкнулись

1. Слишком мало данных о самовозгораниях штабелей

**Решение:** *Размножить данные по каждой дате для каждой кучи, причем кучу считаем по каждой новой поставки supplies*

2. Сильный временной сдвиг ближе к последней дате в обучающей выборке, из-за чего было сложно провалидировать модель

**Решение:** *Взяли более позднюю дату для обучения, чтобы сместить распределение к настоящему. Несмотря на уменьшение числа обучающих данных, результаты стали значительно лучше*

# Архитектура

1. Базы данных(*Sqlite + SQLAlchemy*)

2. Бэкенд(*FastAPI, Python*)

3. Фронтенд(*React + JavaScript*)

4. ML модель(*CatBoostRegressor + CatBoostClassifier*)

## Почему CatBoost?

*CatBoost устойчивый, быстро обучается, хорошо работает с категориальными признаками.*

# Интересные решения

- Учитывать погодные условия за последние 3 дня и за последние 7 дне
- Учитывать «время жизни» кучи
- Добавить новую функцию выявления штабелей, которые, вероятно, загорятся в ближайшие 7 дней
- Учитывать, горела ли куча до этого
- Система весов в классификаторе для баланса классов

# Метрики

*Для регрессора:*

MAE(по дням): 1.5

*Для  
классификатора:*

AUC-ROC: 0.83

Подробнее в приложенном  
ноутбуке...

# Команда



Такаев Денис  
Frontend + DevOps + Data  
Science



Лобан Филипп  
ML + Backend