

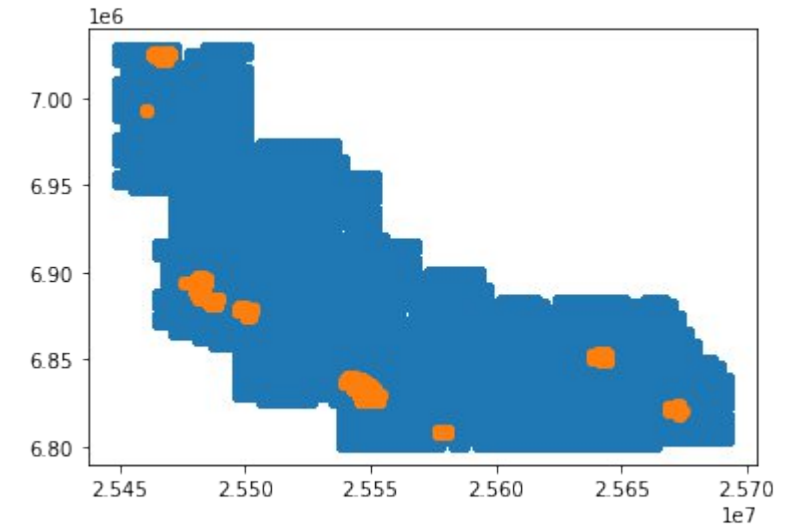
# Преимущества ML подхода.

- Высокая скорость получения результатов на новых данных.
- Нахождение наиболее важных признаков.
- Электронный помощник специалиста.
- Результаты станут более точными после дообучения на новых данных.



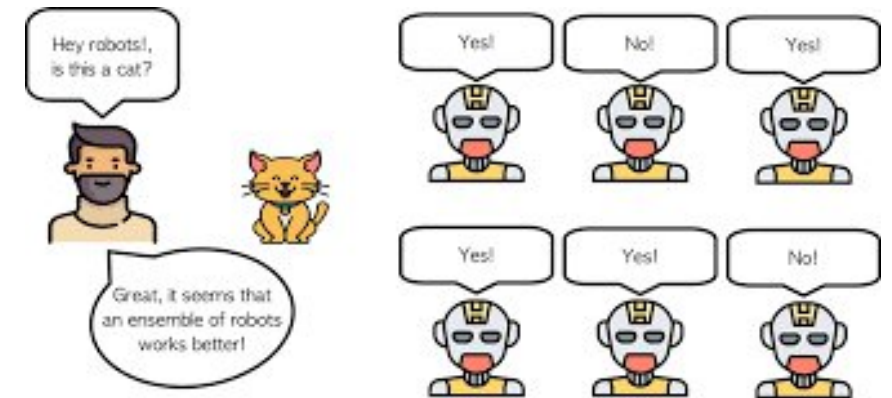
# Задача.

- 5 целевых переменных (Intr, Dep\_labels, F, Atk\_outline, MF).
- 25 признаков.
- **Задача сегментации:** для каждой точки определить наличие или отсутствие целевой переменной.



# Подход.

- Можно использовать методы работы с табличными данными.  
SOTA: **catboost**.
- Можно использовать методы работы с изображениями.  
SOTA: **DeepLabV3Plus с ResNet backbone**.
- Объединение 3-5 упомянутых выше моделей с помощью **блендинга** (голосования).



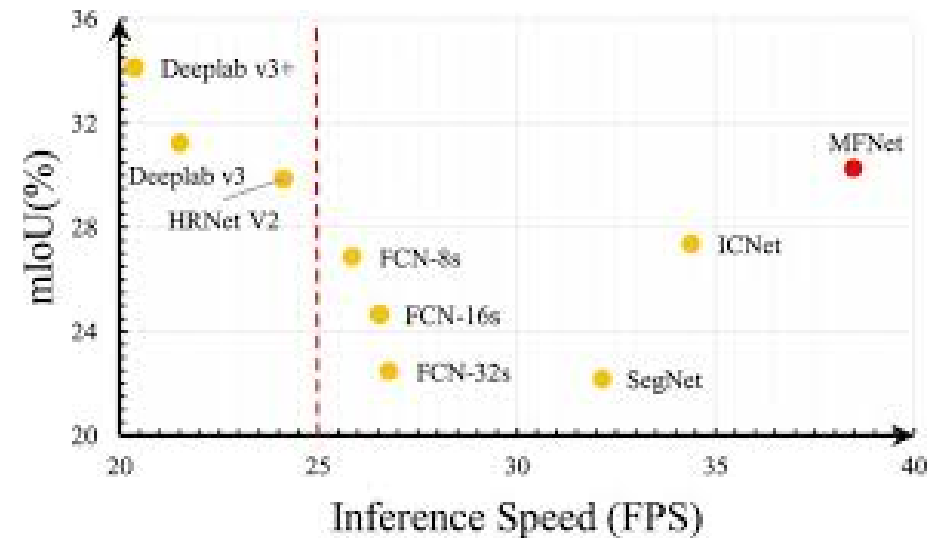
# Catboost.

- Решение от Яндекса, лучшее, что есть, для табличных данных.
- Хорошо работает “из коробки” с минимальным тюнингом.
- Последовательное улучшение множества слабых решений, каждое следующее учится на ошибках предыдущих, модель учится работать в слабых местах.

	GBM	CatBoost	LightGBM	XGBoost
Baseline	0.9455	0.9534	0.9453	0.9438
Fashion	0.9826	0.9833	0.9815	0.9814
Retail	0.9614	0.9617	0.96	0.96
Digital goods	0.8739	0.8767	0.8732	0.8751

# DeepLabV3Plus.

- Лучшее решение для задачи сегментации изображений.
- Модель обучается 8 часов на GPU.
- Модель имеет несколько десятков слоёв, каждый из которых отвечает за работу со своими признаками изображения.
- Модель может работать с признаками на разных масштабах.



# Обзор существующих решений.

Есть десятки работ по использованию машинного обучения в геологии. Основные методы:

- Методы работы с табличными данными.
- Методы решения задачи классификации изображений.
- Методы решения задачи сегментации изображений.
- Методы обнаружения объектов на изображениях.

Наша работа объединяет в себе **лучшие** методы из выделённых зелёным категорий.

