

# *Ансамбли моделей ч.2*

---

КУХАЛЬСКИЙ НИКОЛАЙ ГЕННАДЬЕВИЧ

# *Вопросы занятия*

1. Композиция моделей разных классов.

## *В конце занятия научимся:*

- строить композиции моделей разных классов.

# Особо случайные деревья

Можно сделать деревья еще более случайными за счет применения случайных порогов для каждого признака вместо поиска наилучших возможных порогов (подобно тому, как поступают обыкновенные деревья принятия решений).

```
from sklearn.ensemble import ExtraTreesClassifier (ExtraTreesRegressor)
```

- более высокое смещение обменивается на низкую дисперсию;
- обучаются намного быстрее, чем обыкновенные случайные леса, поскольку нахождение наилучшего возможного порога для каждого признака в каждом узле является одной из самых затратных в плане времени задач по выращиванию дерева;
- имеет такой же API-интерфейс, как у класса RandomForestClassifier (RandomForestRegressor).

# ***ПРАКТИКА***

***Доработать код с  
использованием особо  
случайных деревьев***

# ПРАКТИКА

## Композиции моделей разных классов

Линейной комбинацией альтернатив  $P$  и  $Q$  называется распределение  $R$  для которого справедливо:

$$R = \alpha \cdot P + (1 - \alpha) \cdot Q, \text{ где } 0 \leq \alpha \leq 1.$$

Линейная комбинация моделей разного типа - смешаем дерево и логистическую регрессию.

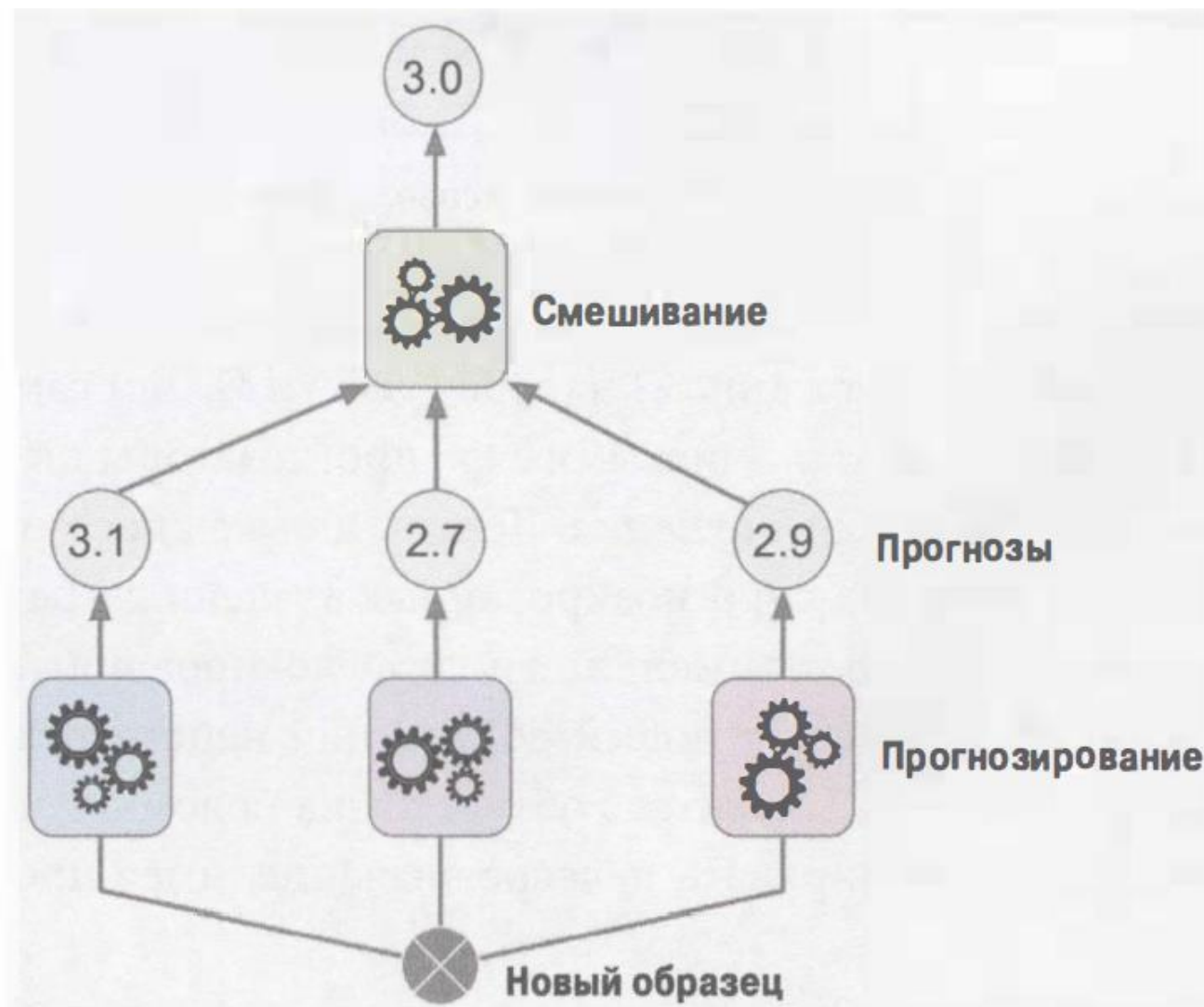
Построить линейную комбинацию вида:

$$y = \alpha y_1 + (1 - \alpha) y_2$$

Параметр  $\alpha$  переберем по сетке от 0 до 1, оценивая качество на тестовой выборке

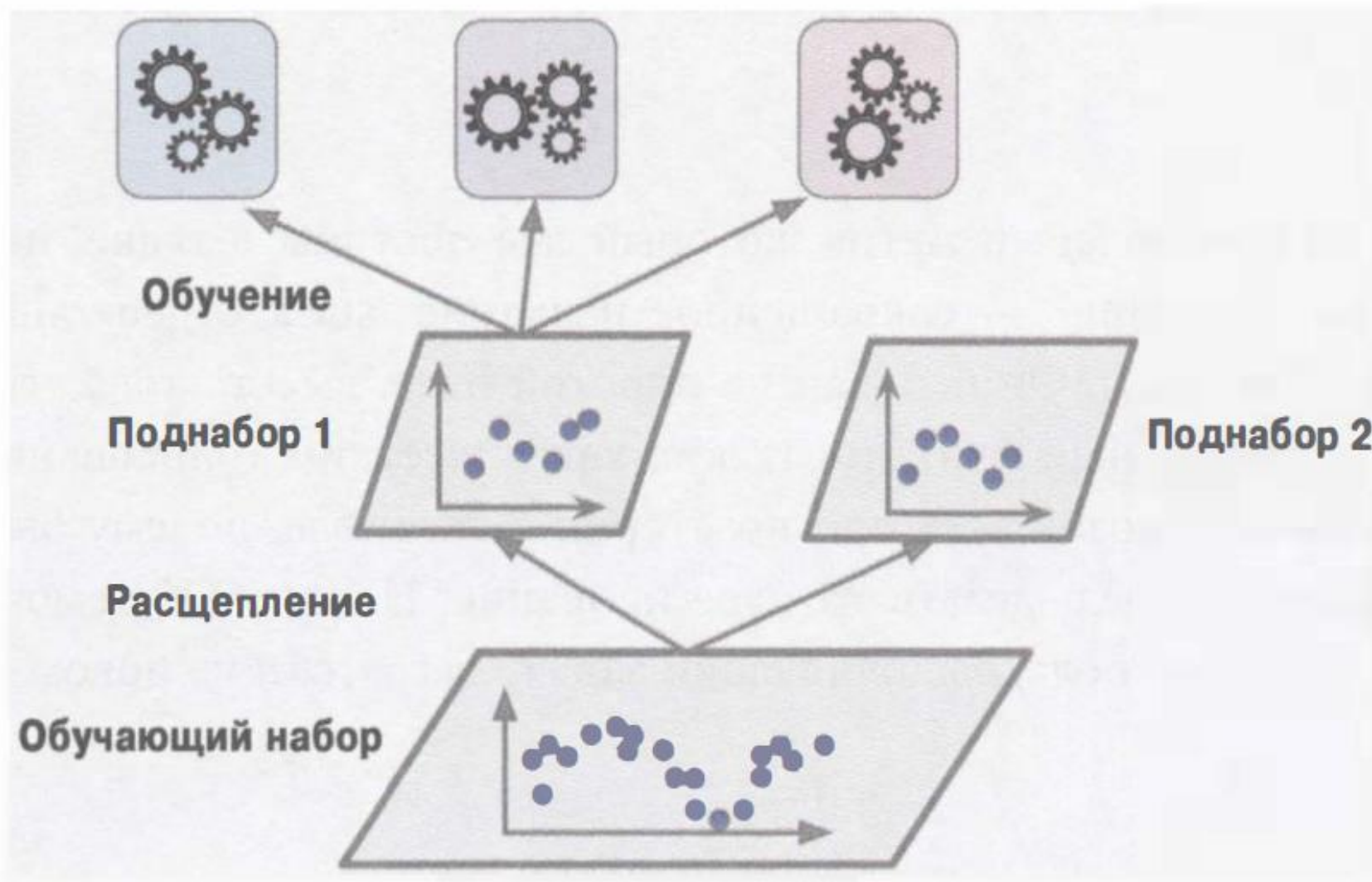
# *СТЕКИНГ*

## Композиции моделей разных классов





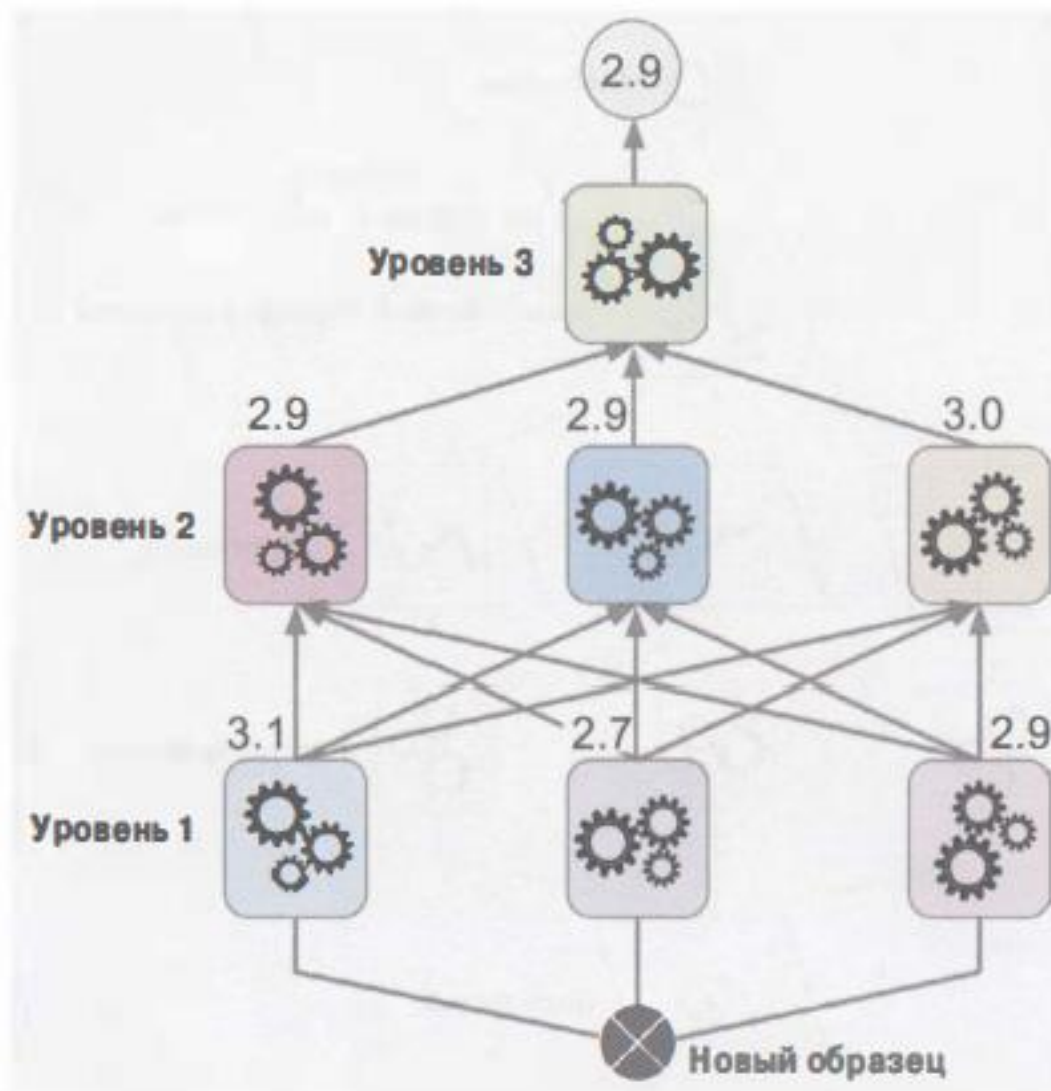
## Композиции моделей разных классов



# Композиции моделей разных классов

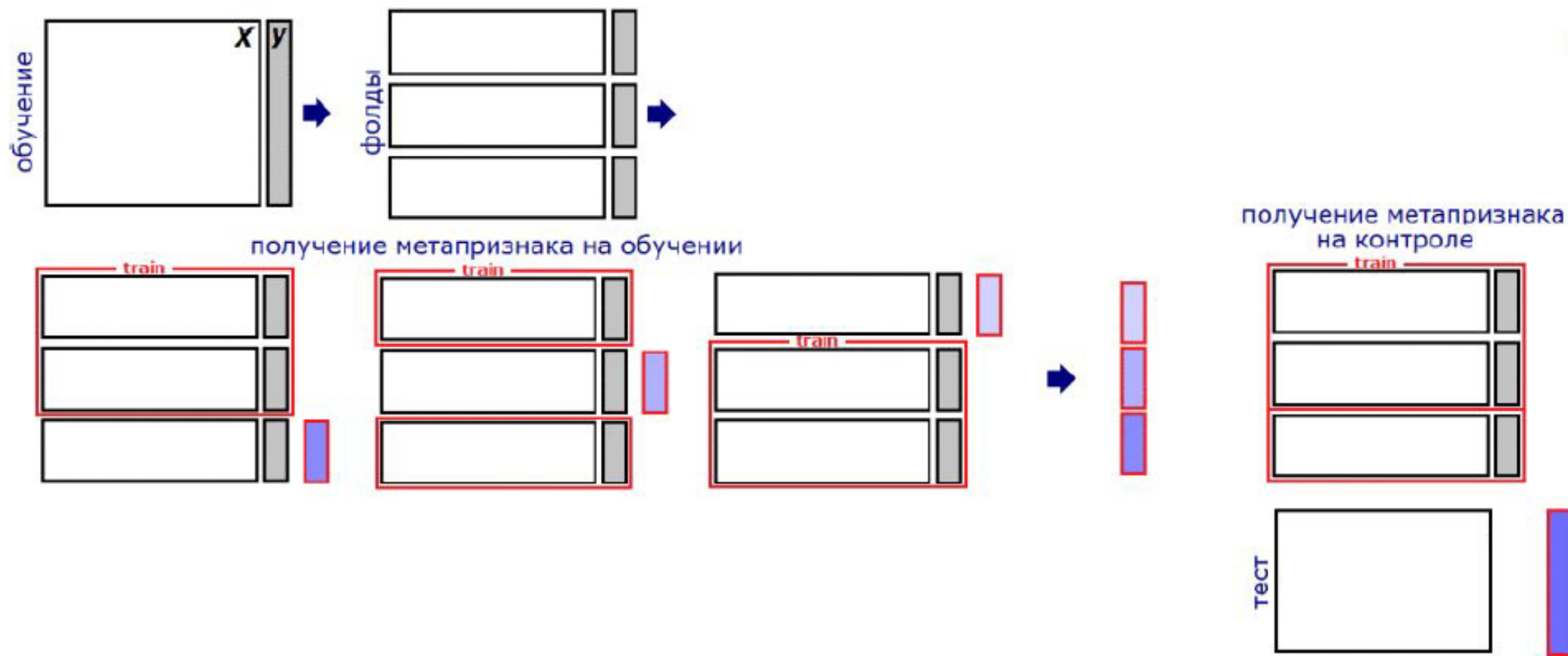


# Композиции моделей разных классов



# Композиции моделей разных классов

## АЛГОРИТМ



## *Композиции моделей разных классов*

### **АЛГОРИТМ**

1. Разбиваем обучающую выборку на  **$K$**  фолдов по кросс-валидации
2. Обучаем алгоритм на  **$K-1$**  фолдах и предсказываем на оставшемся. Повторяем для каждого фолда.
3. Полученные out-of-fold предсказания используем в качестве новых признаков - **мета-признаков**
4. Выполняем пункты 1-3 для разных ML моделей

## *Композиции моделей разных классов*

### **АЛГОРИТМ**

4. Получаем мета-признаки для тестовой выборки:
  - i. Обучаем алгоритм на всей обучающей выборке, предсказываем на тестовой
  - ii. Усредняем предсказания на тесте ***K*** моделей, обученных на шаге 2.
5. Обучаем алгоритм 2-го уровня на мета-признаках

## *Композиции моделей разных классов*

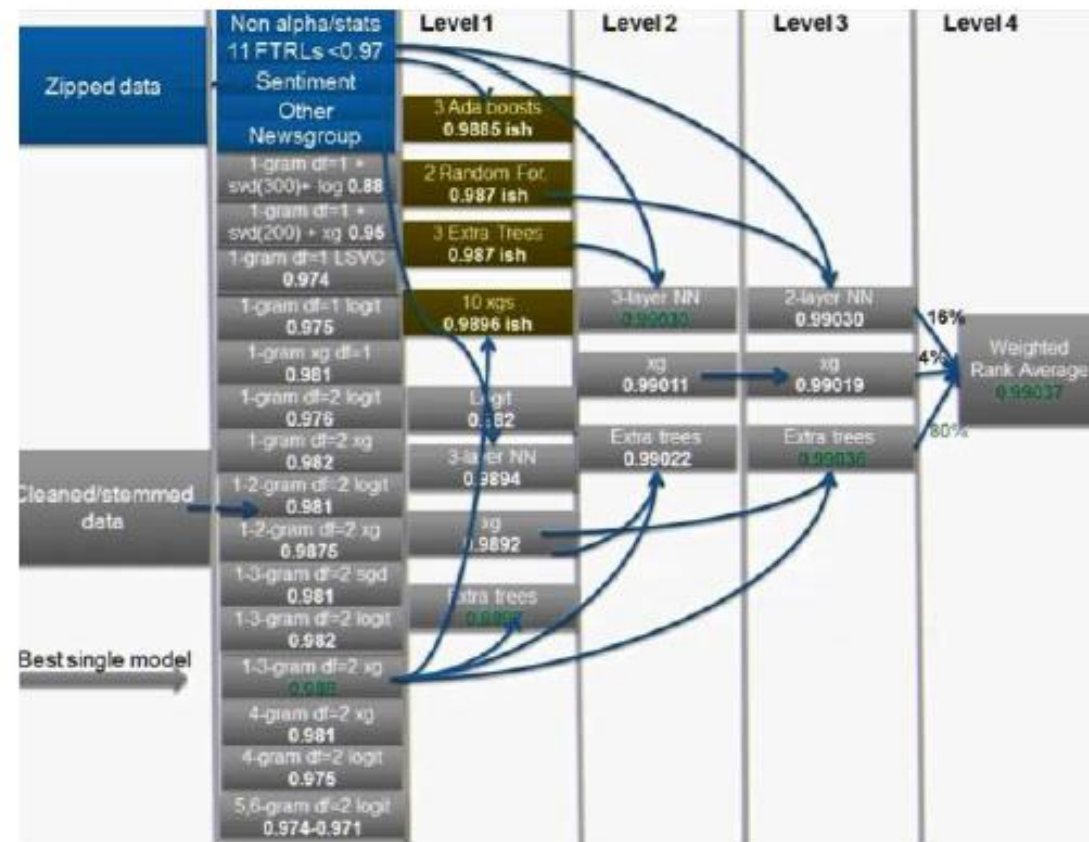
### **ВАЖНО**

- не допустить утечки информации от целевой переменной в признаки;
- мета-признаки часто сильно коррелированы, нужно использовать регуляризованные модели на втором уровне;
- использовать разные по природе модели, причем качество отдельной модели может быть небольшим, но давать большой прирост в стекинге.



# Композиции моделей разных классов

## KAGGLE: MONSTER ENSEMBLE





# *Композиции моделей разных классов*

## **ПЛЮСЫ**

- повышает качество модели, когда уже ничего не помогает;
- позволяет эффективно смешивать модели разных классов, позволяя объединить их сильные стороны и убрать слабые;
- поможет выиграть “золото” на kaggle.

## *Композиции моделей разных классов*

### **МИНУСЫ**

- сложность интерпретации;
- легко можно переобучиться, допустив утечку информации;
- хорошо работает только на достаточно больших выборках;
- большая вычислительная сложность, редко применяется в production системах.

# ***ПРАКТИКА***

## ***Многоуровневые модели***

## *ЧТО МЫ СЕГОДНЯ УЗНАЛИ*

- как строятся композиция моделей разных классов.

# *Ансамбли моделей ч.2*

---

КУХАЛЬСКИЙ НИКОЛАЙ ГЕННАДЬЕВИЧ