



SCHOOL  
OF ENERGY  
& POWER ENGINEERING

# Построение ансамблей и поиск лучшего классификатора

Сергей Владимирович Аксёнов,

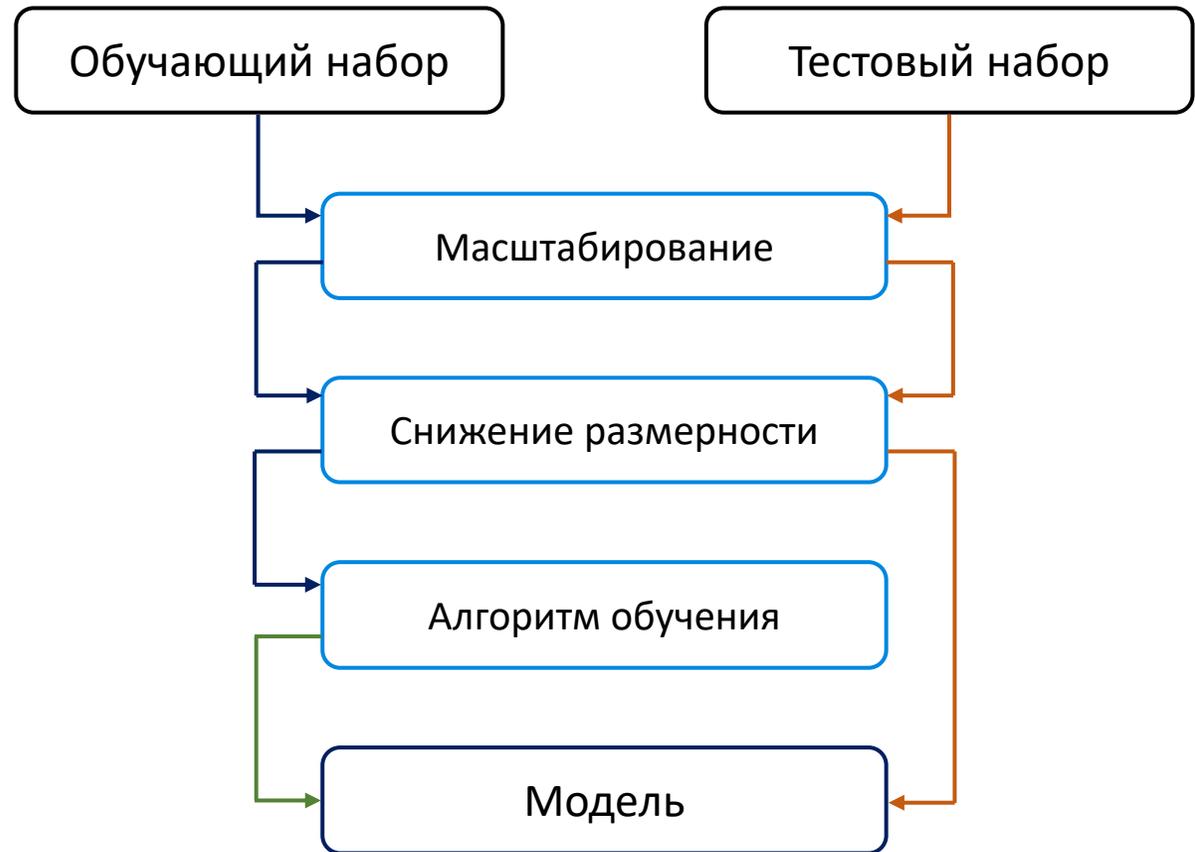
Доцент отделения информационных технологий ИШИТР,

Томский политехнический университет

Томск-2023

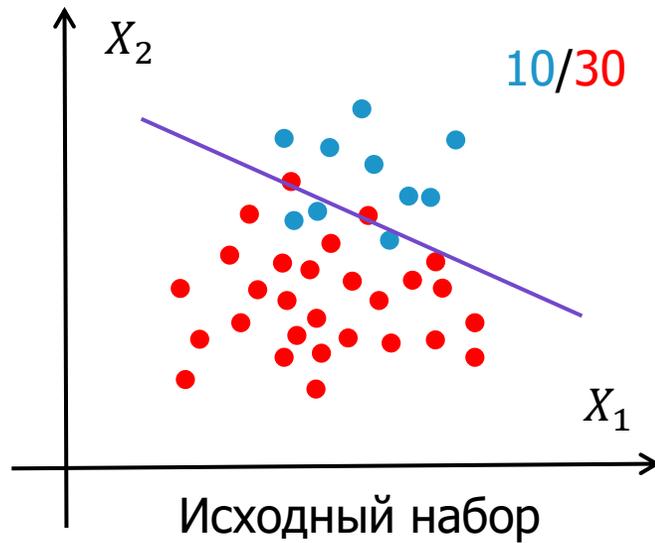
# Конвейер

---

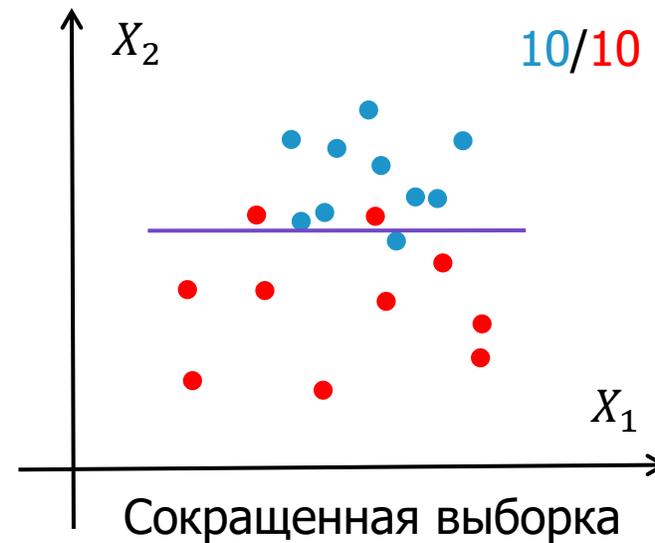
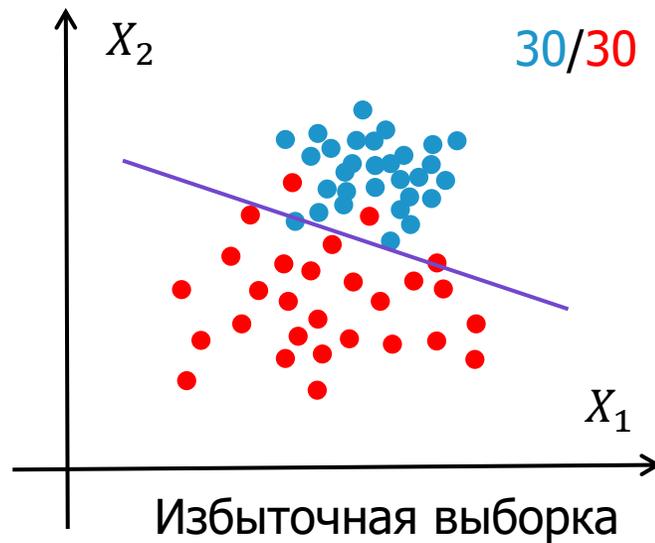


- Настройка/Преобразование
- Преобразование
- Обучение
- Активация

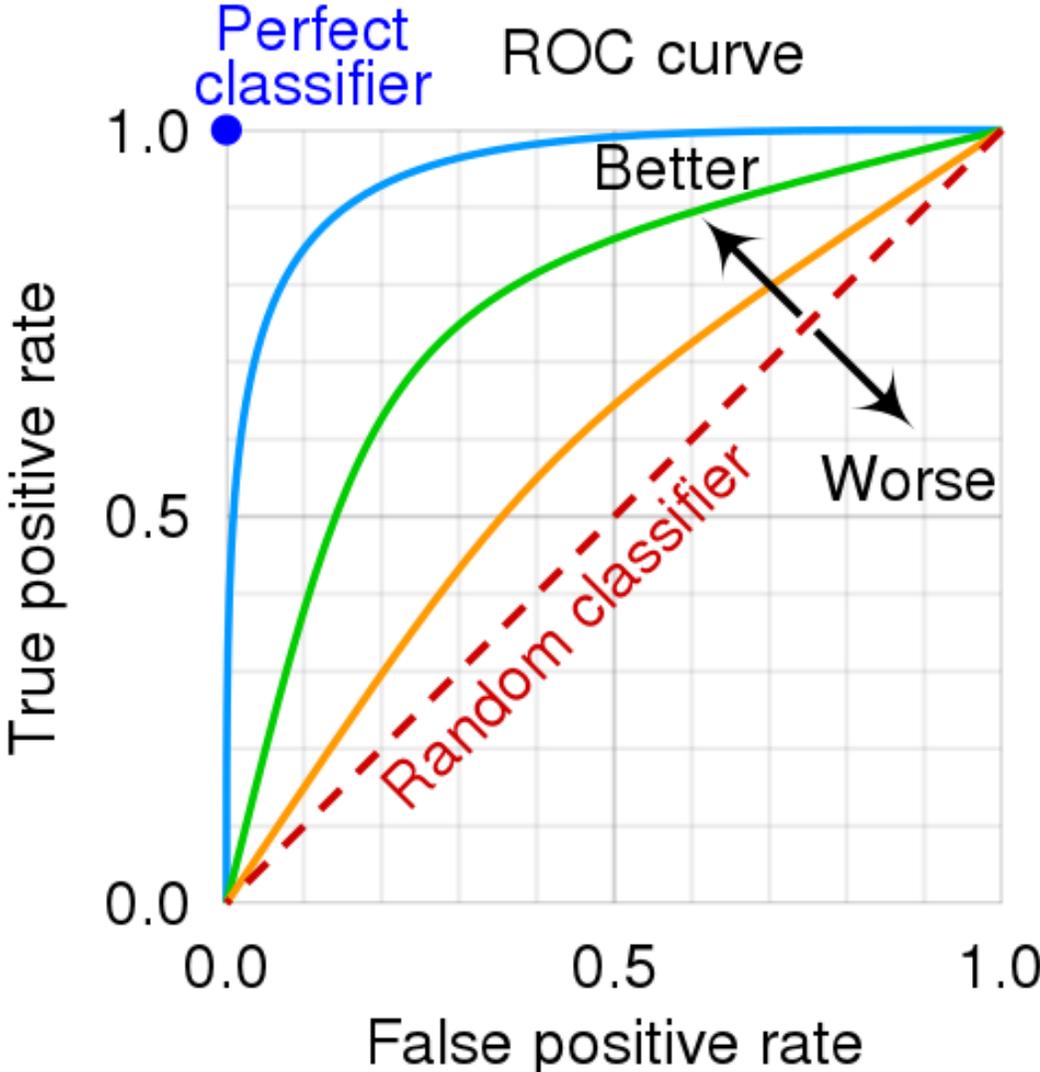
# Дисбаланс классов



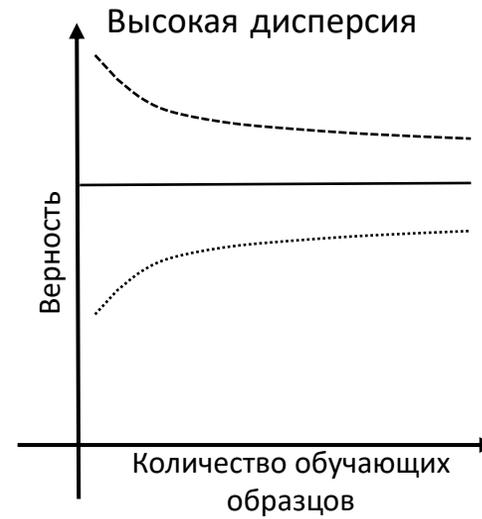
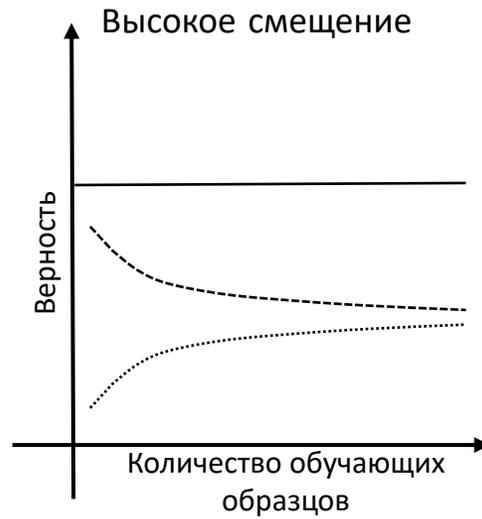
- Класс А
- Класс В
- Граница решения



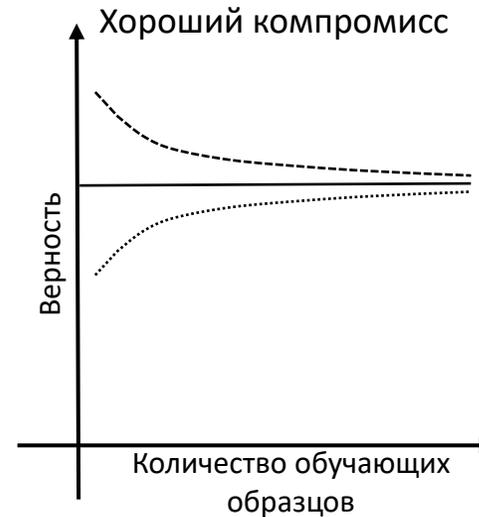
# ROC-кривая



# Кривые обучения и валидации



- Требуемая верность
- - - Верность на обучающем наборе
- ..... Верность на тестовом наборе



# Бутстреп-выборки

#	A1	A2	A3	A4	A5
1					
2					
3					
4					
5					

Исходный набор:  
признаки

Выборка 1

#	A2	A4
1		
3		
4		

Выборка 2

#	A1	A3	A4
3			
5			

Выборка 3

#	A1	A5
1		
2		
3		

Выборка 4

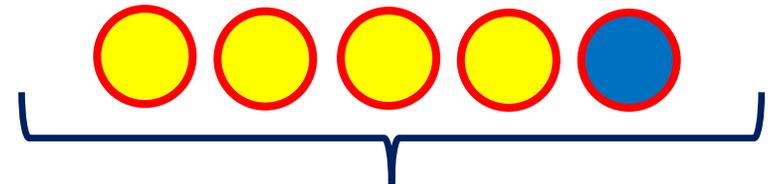
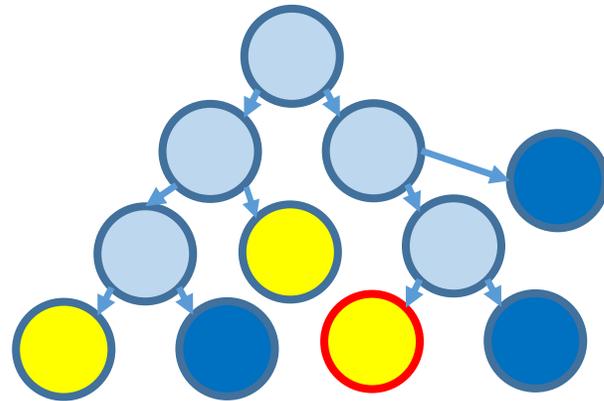
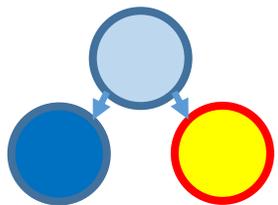
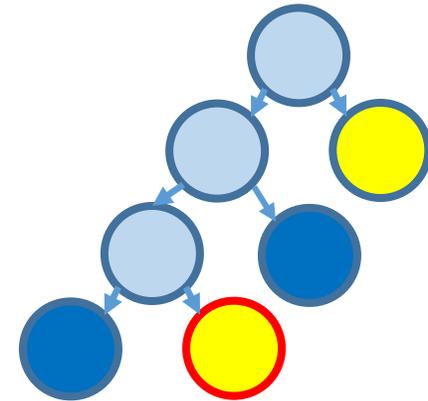
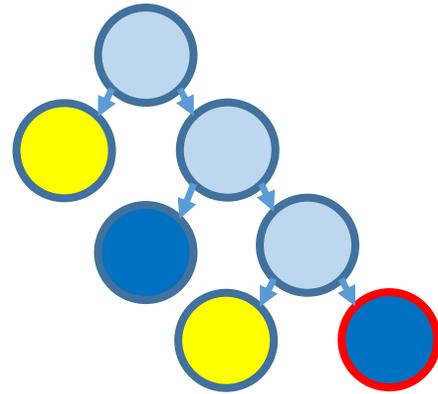
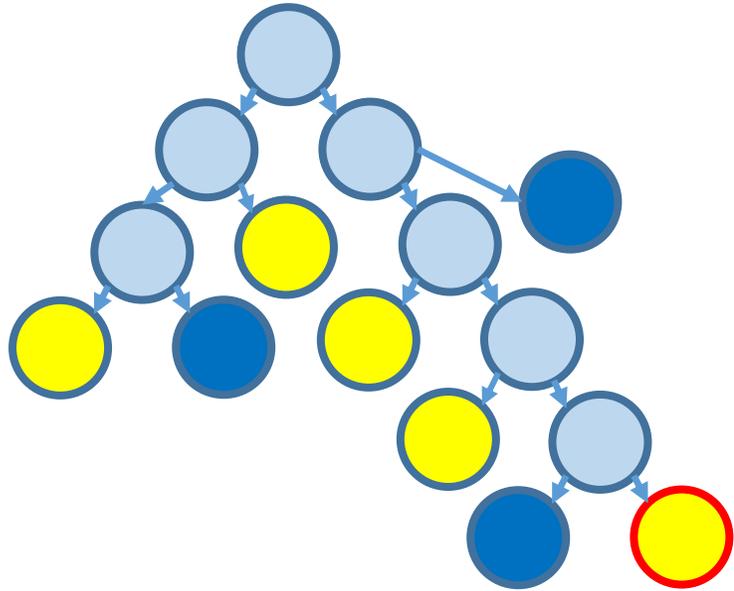
#	A2	A3	A5
2			
4			
5			

#	A1	A2
1		
2		

Выборка 5

# Случайный лес

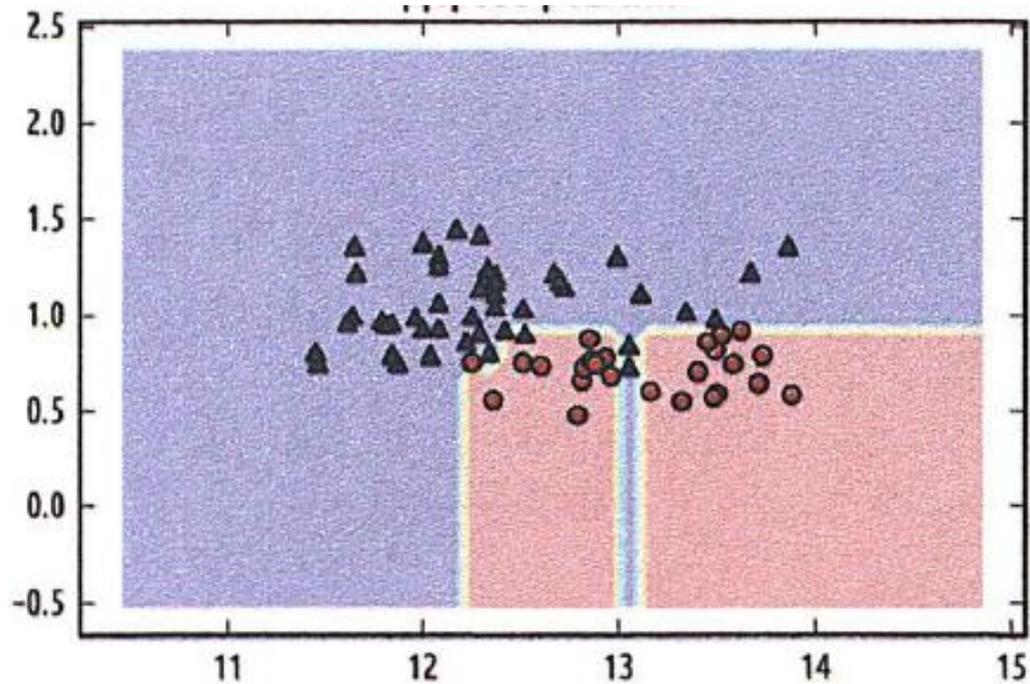
Объединение работы нескольких деревьев.  
Мажоритарное голосование



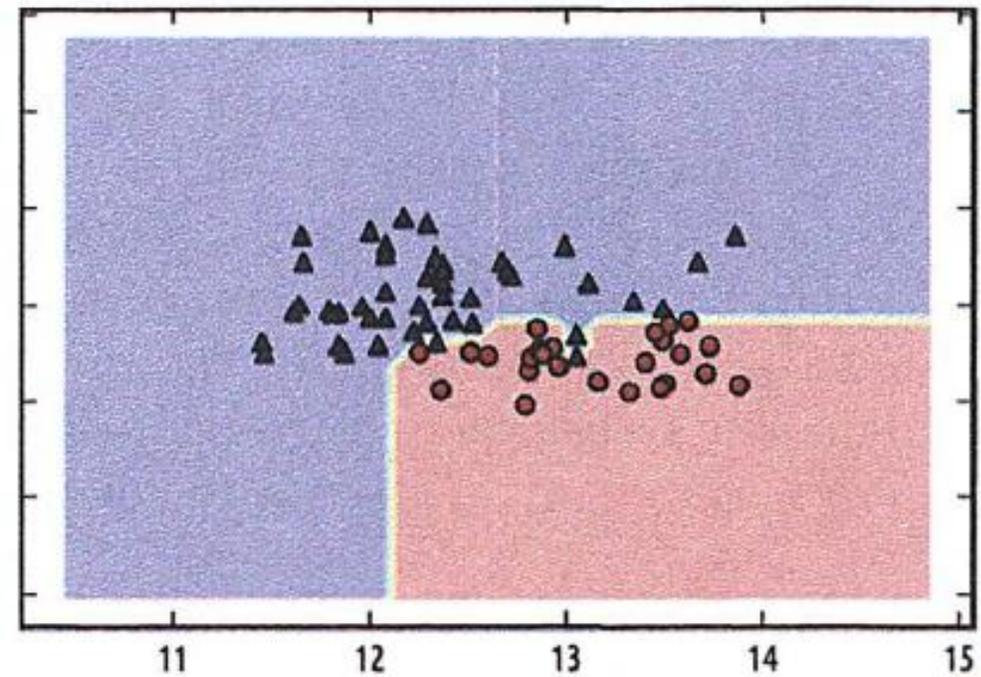
 Класс, выбранный моделью

Итог леса

# Сравнение случайного леса и дерева решений



Дерево решений



Случайный лес

# Поиск по сетке параметров

---

Пример:

исследуемые параметры дерева решений



Глубина дерева



Критерий разделения



Минимальное количество наблюдений в  
неоднородном узле, который нужно разделить

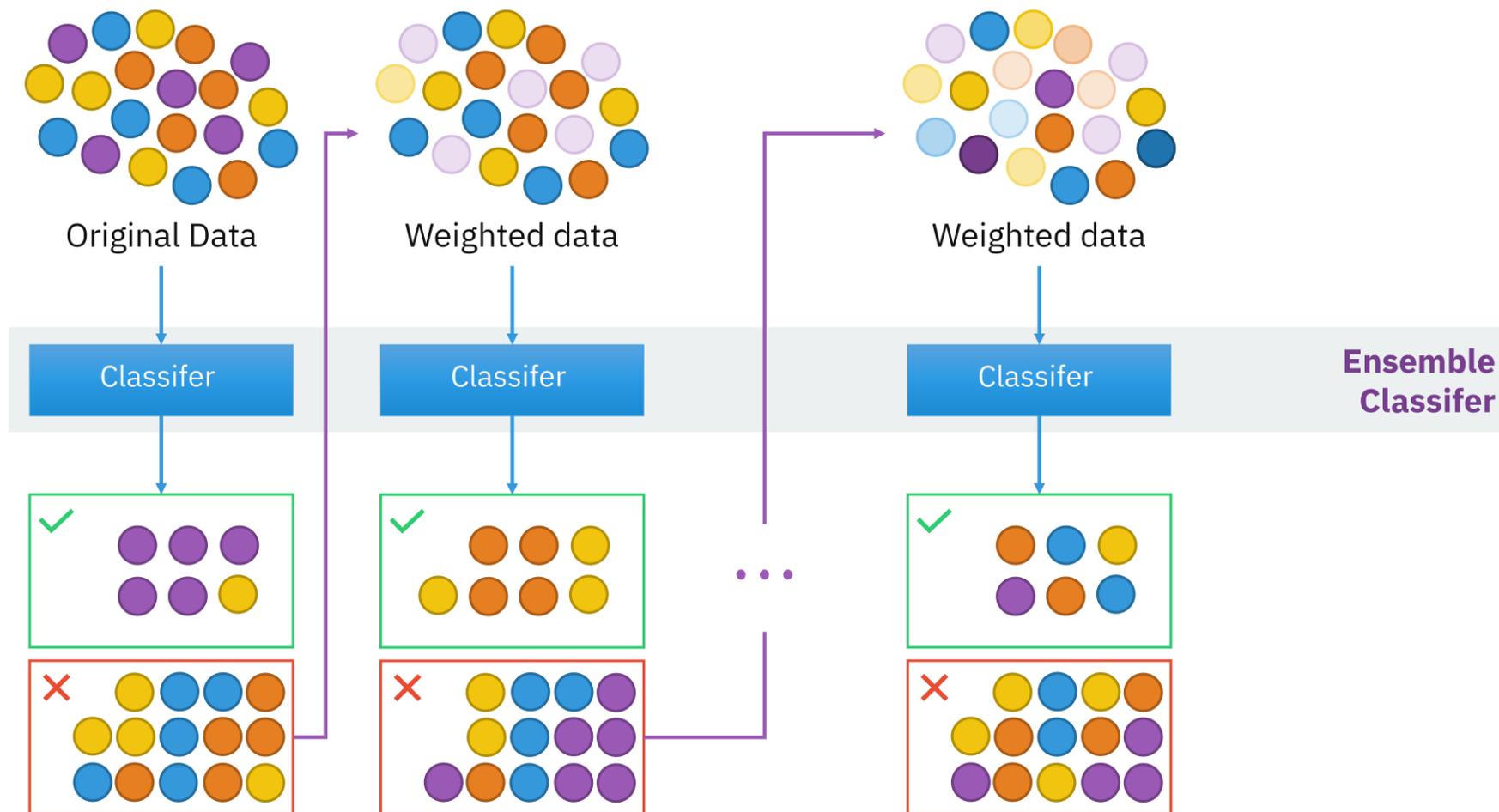


Степень неоднородности, который указывает, что  
узел нужно разделить

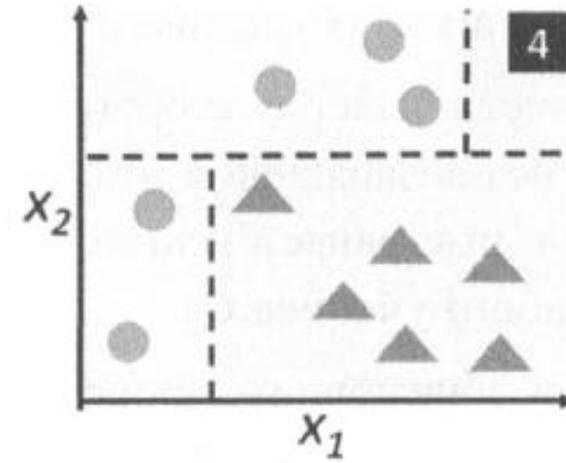
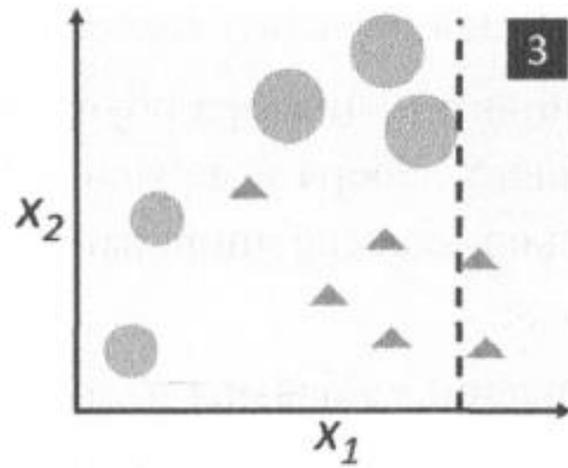
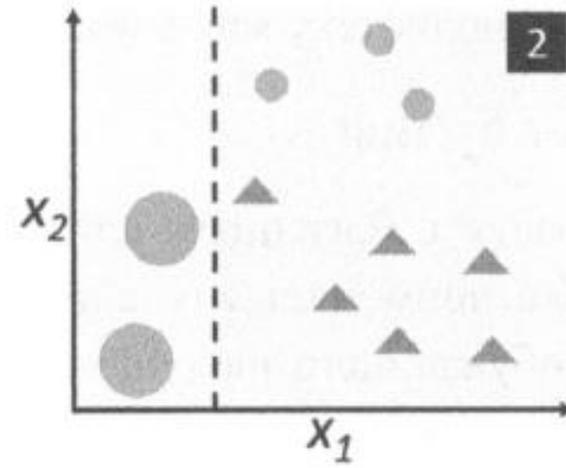
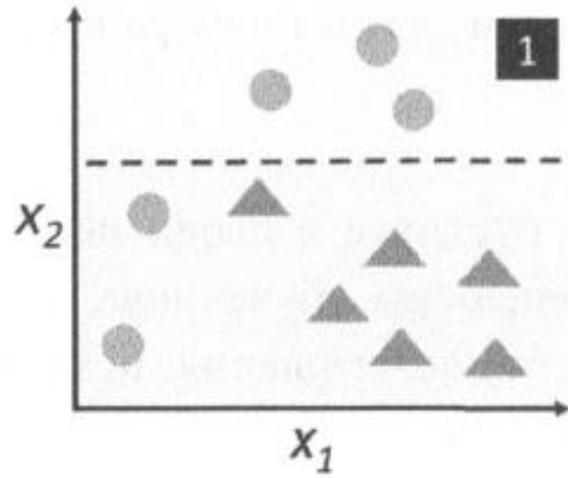


...

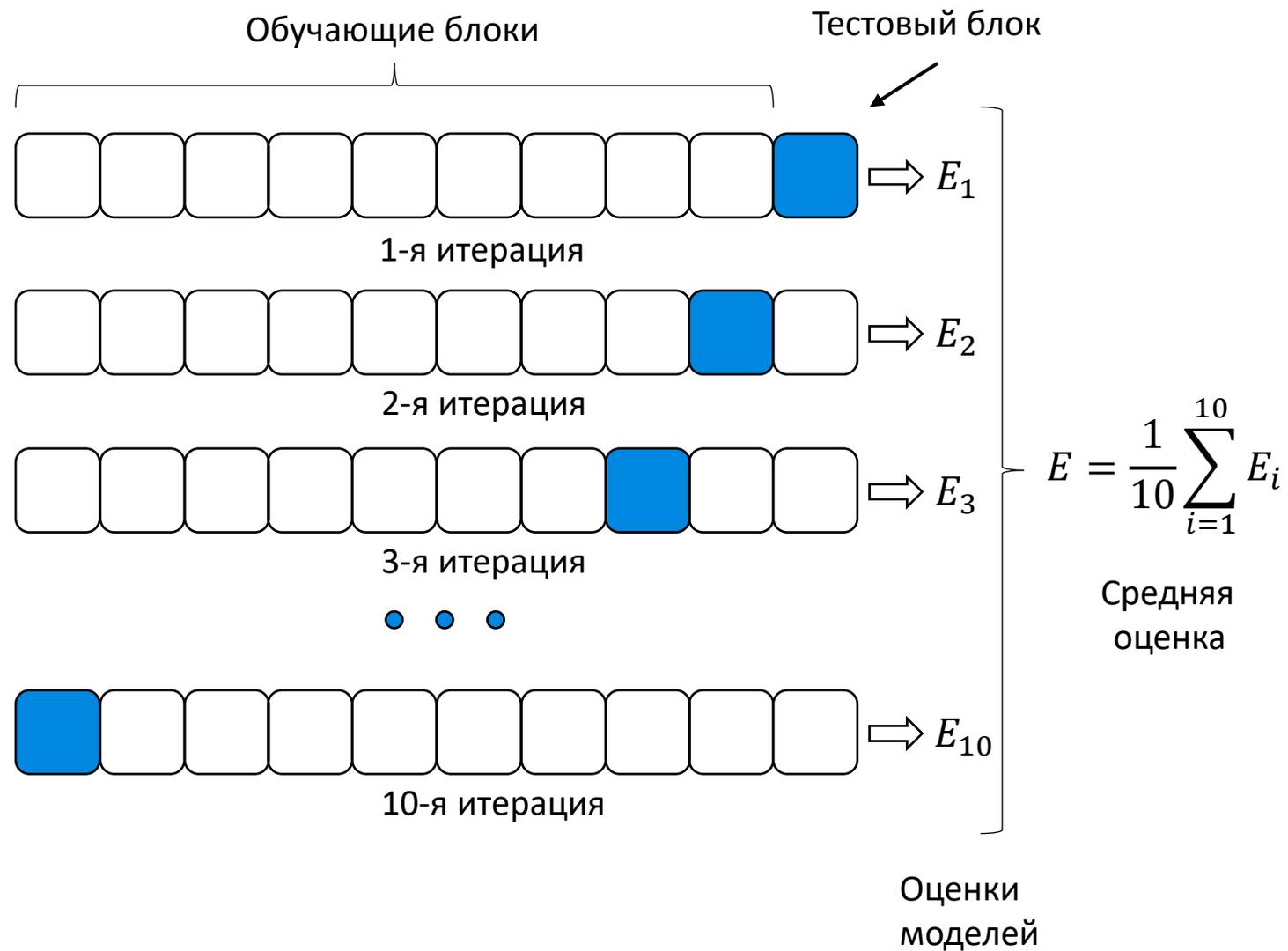
# Бустинг



# AdaBoost



# K-блочная перекрестная проверка

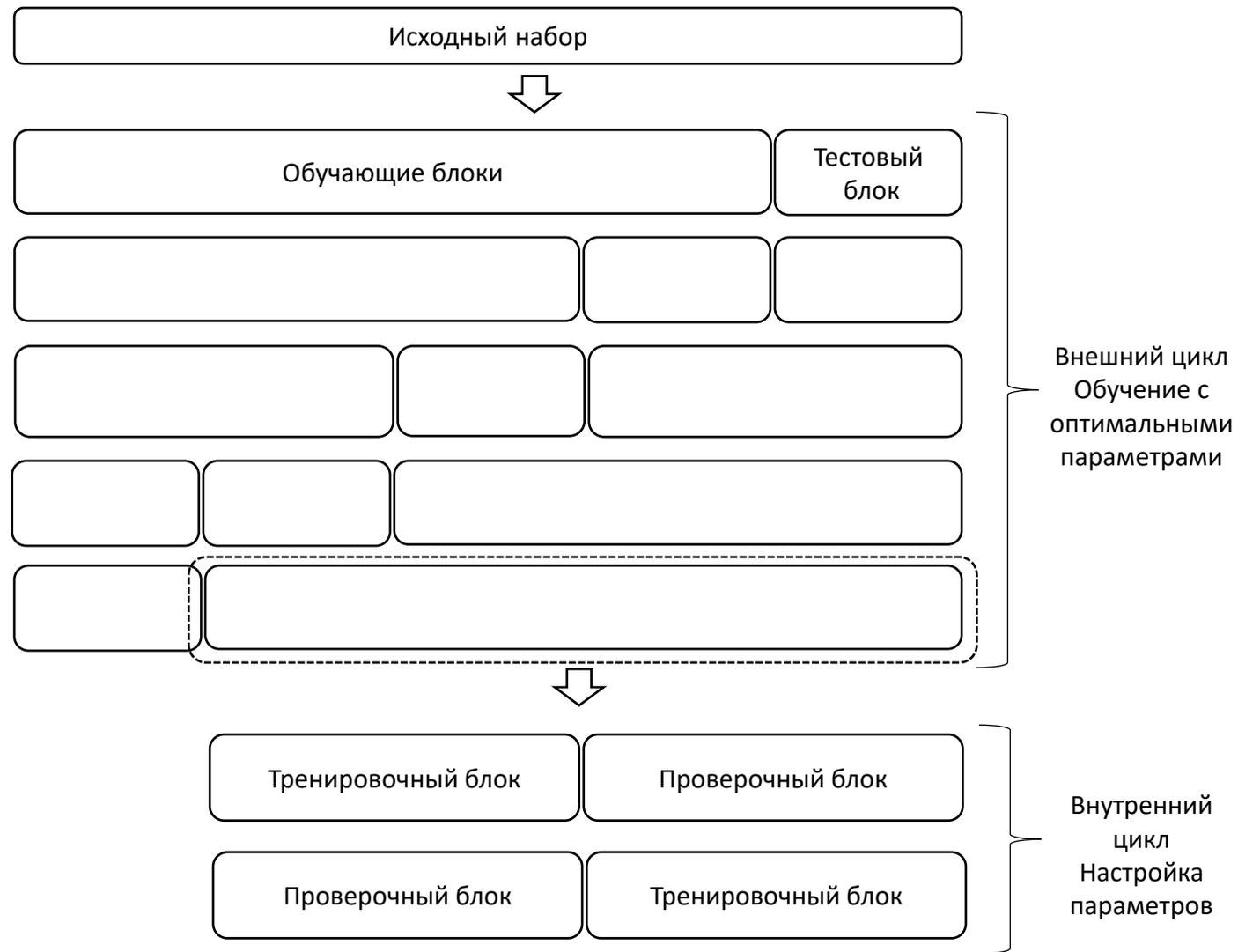


# Перекрестная проверка с удержанием

---



# Перекрестная проверка 5x2



# Ансамбли моделей

---



## Мажоритарное голосование/Стекинг

Решение задачи обучения с учителем путем голосования независимых оценщиков



## Бэггинг

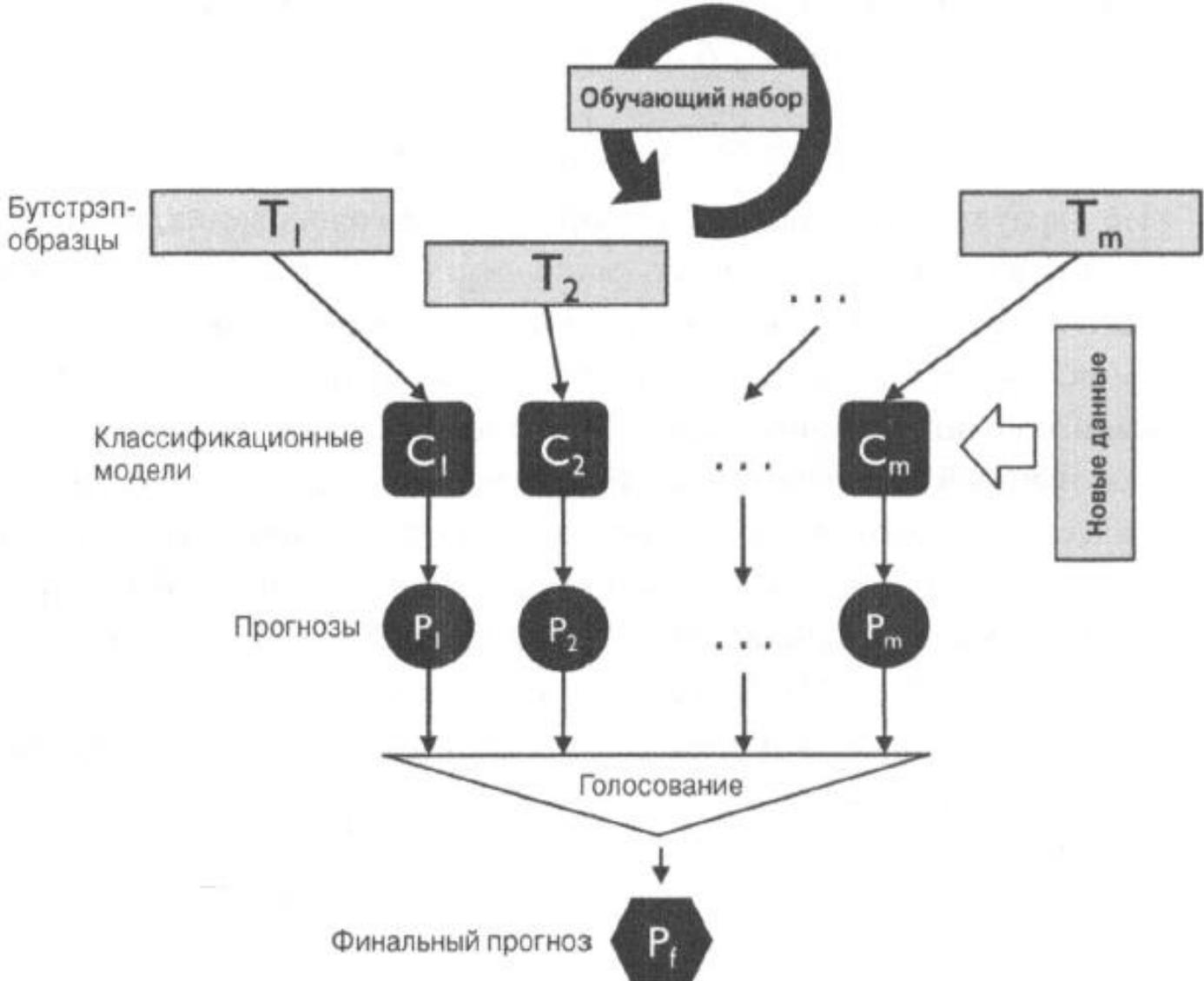
Ансамблевое обучение классификаторов, обученных на бутстрап-выборках



## Бустинг

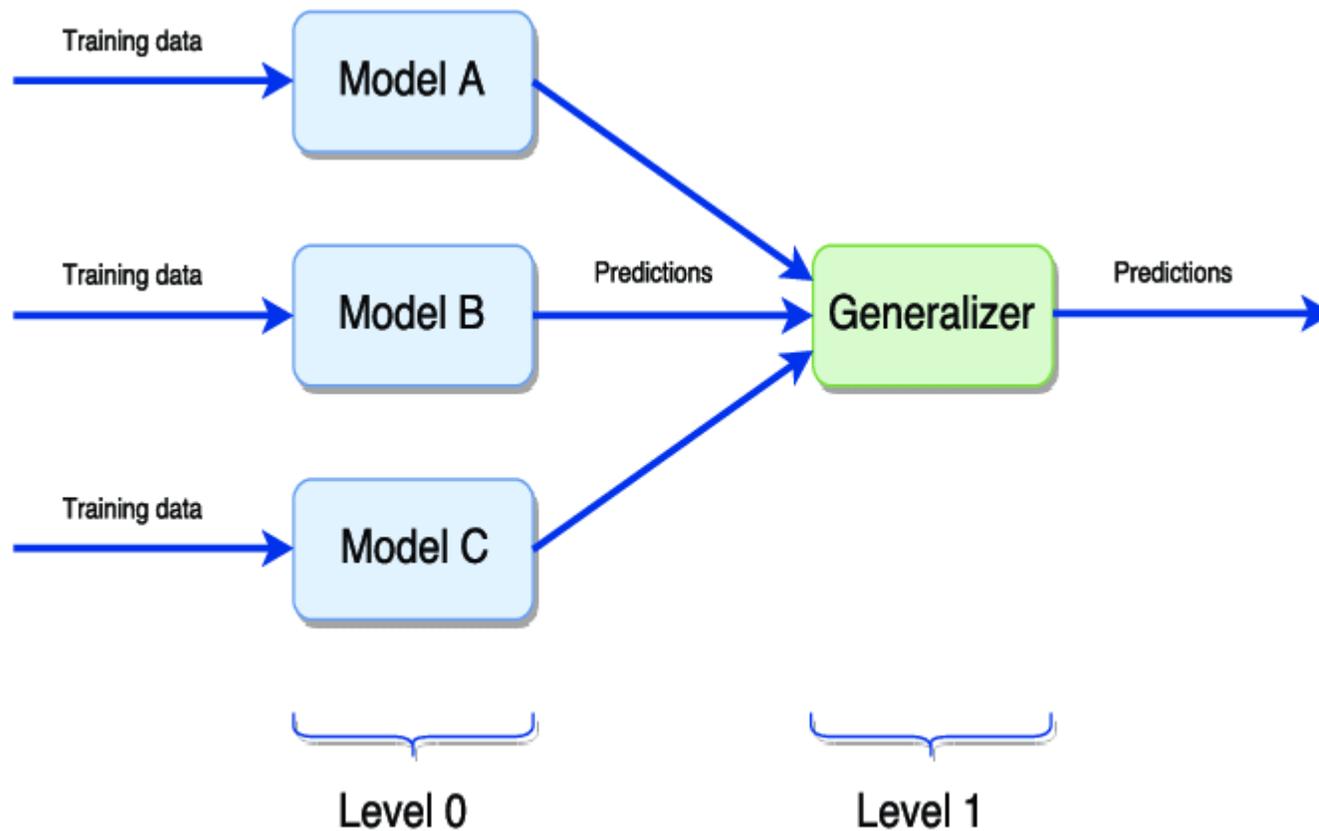
Последовательная настройка ансамбля, в которой каждая последующая модель настраивается на ошибках предыдущей

# Бэггинг



# Стекинг

---



# Автоматическое машинное обучение (AutoML)

---

