

## Проект №5

# Картирование различных фациальных зон по данным ГИС

Владислав Трифонов  
Анастасия Волкова  
Ольга Волкова  
Владимир Назаров

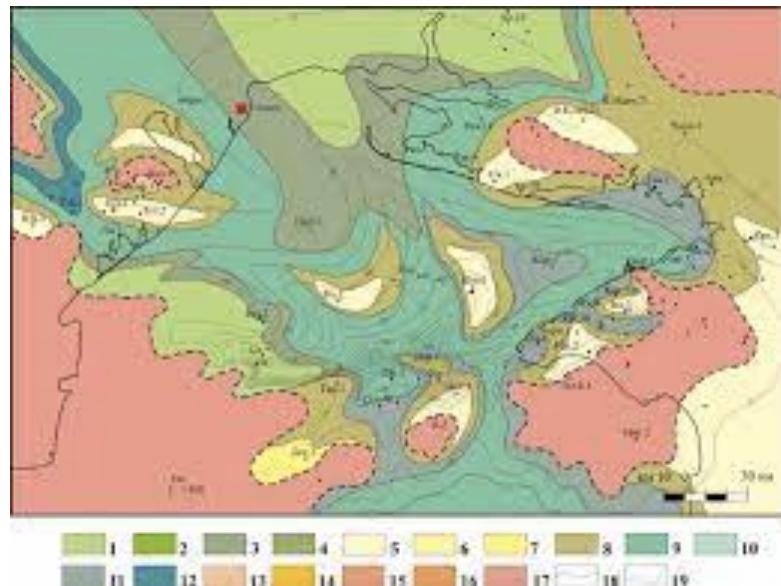
# Содержание



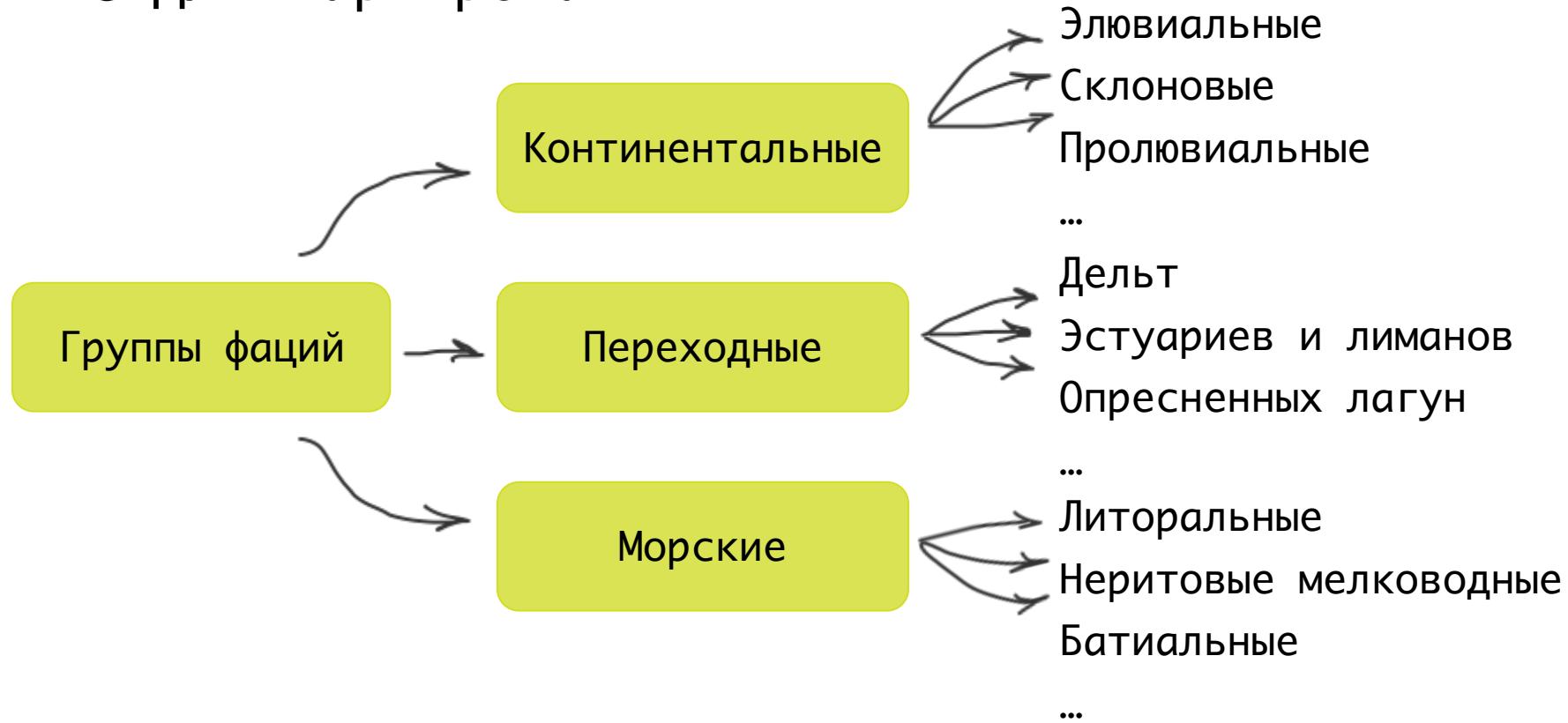
# ГИС для картирования

Использование данных ГИС для локализации и картирования фациальных зон месторождений.

Процесс можно автоматизировать с помощью машинного обучения



# ГИС для картирования



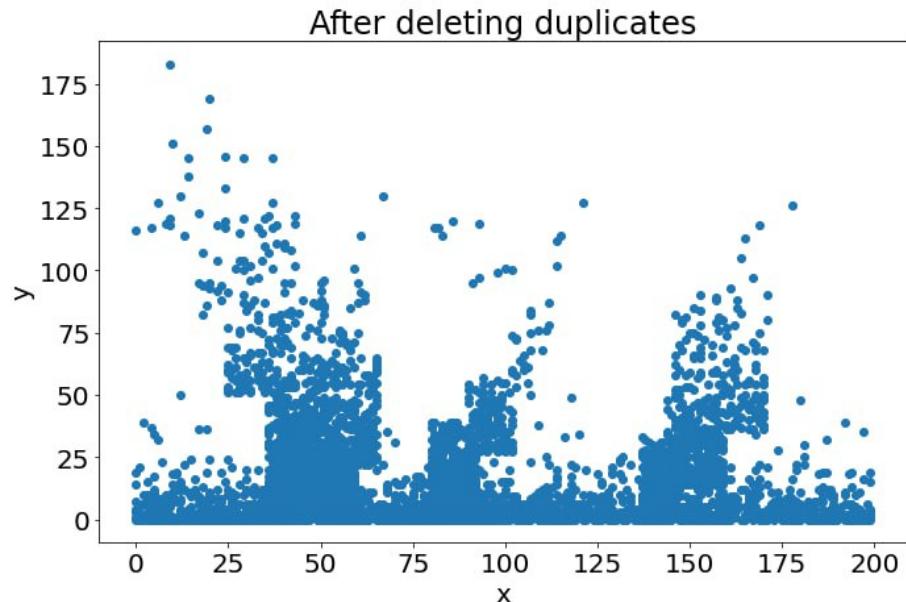
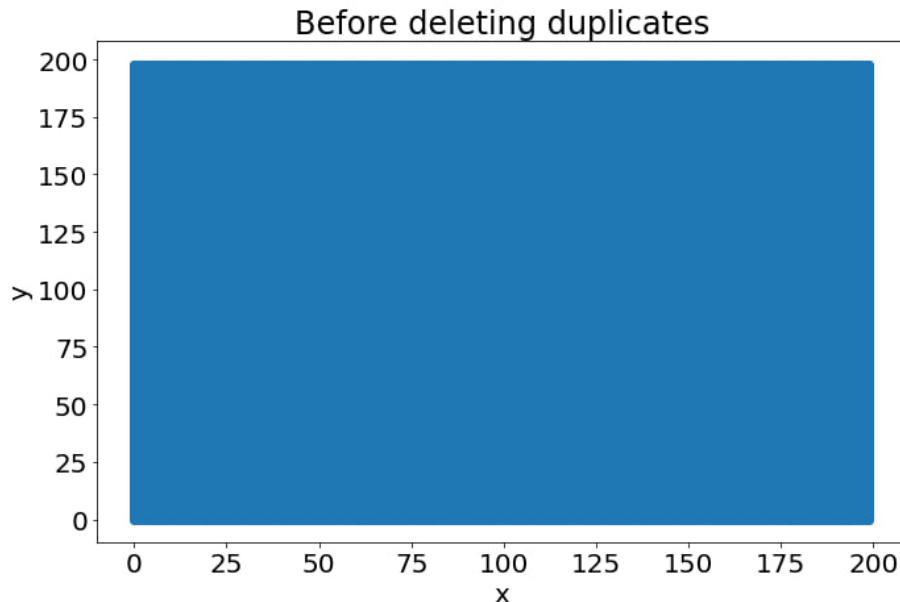
# Анализ входных данных

Количество скважин	Тип ГИС	Представление
39800	Альфа ПС	Временной ряд

Номер скважины	ГИС	Координаты
0 well_21889	uwi	sp x y
1 well_21052	[0.55285714, 0.5158285696, 0.4787999992, 0.441...	0 0
2 well_21108	[0.40714286, 0.40714286, 0.40714286, 0.4071428...	1 0
3 well_21583	[0.5128930786643355, 0.5256043462865803, 0.463...	2 0
4 well_21784	[0.5941460545447186, 0.49742460330502825, 0.43...	3 0
		4 0

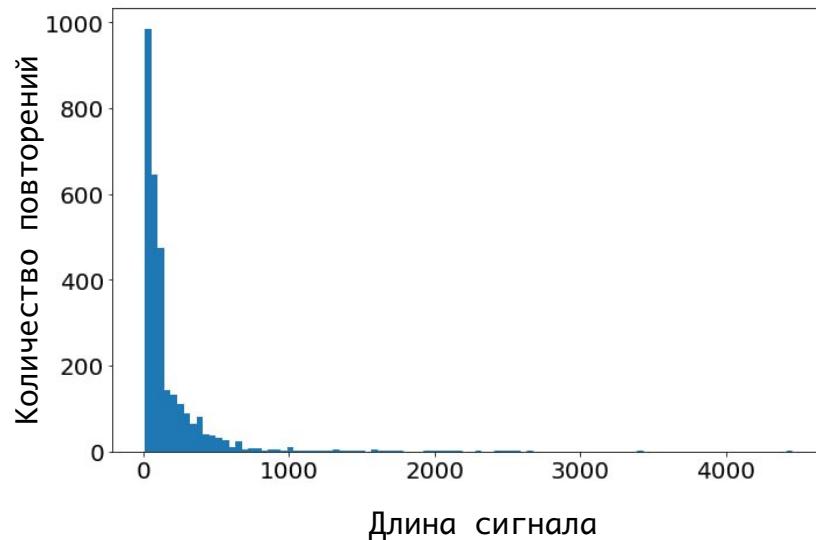
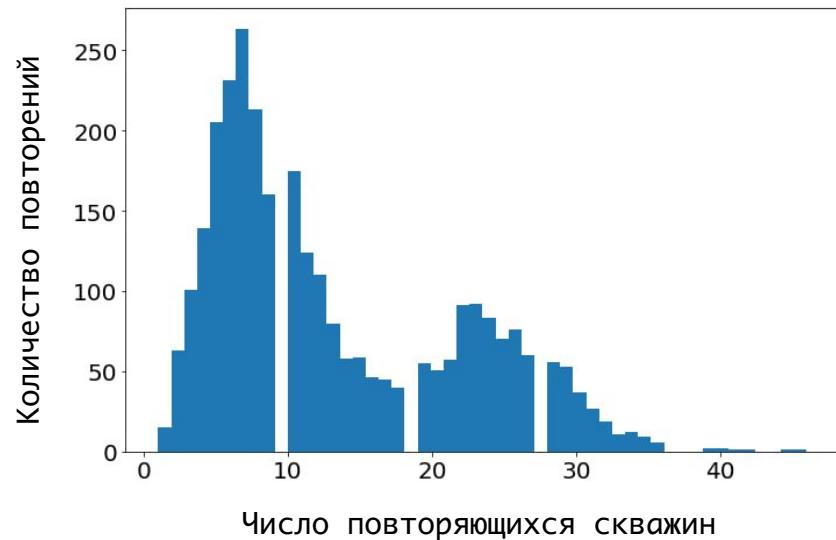
# Анализ входных данных

Карта скважин



# Анализ входных данных

Удаление дубликатов, добавление длины сигнала

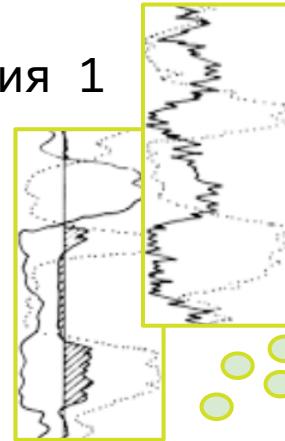


# Кластеризация ГИС

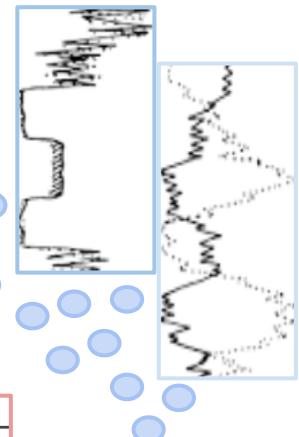
Данные ГИС

кластеризация

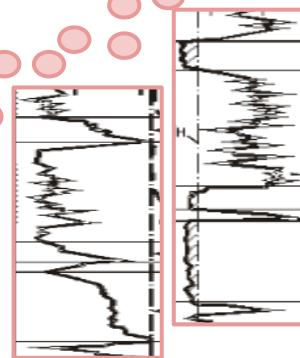
фация 1



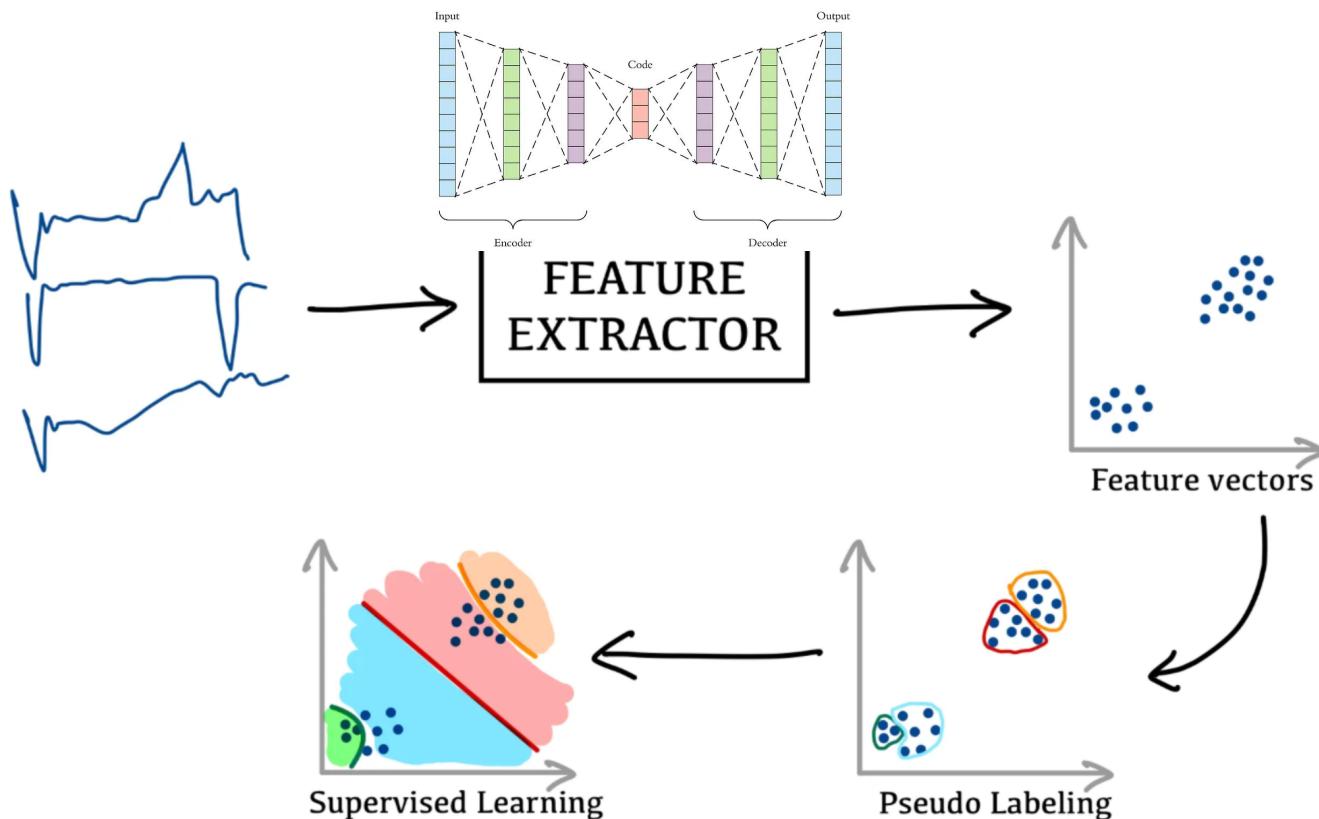
фация 2



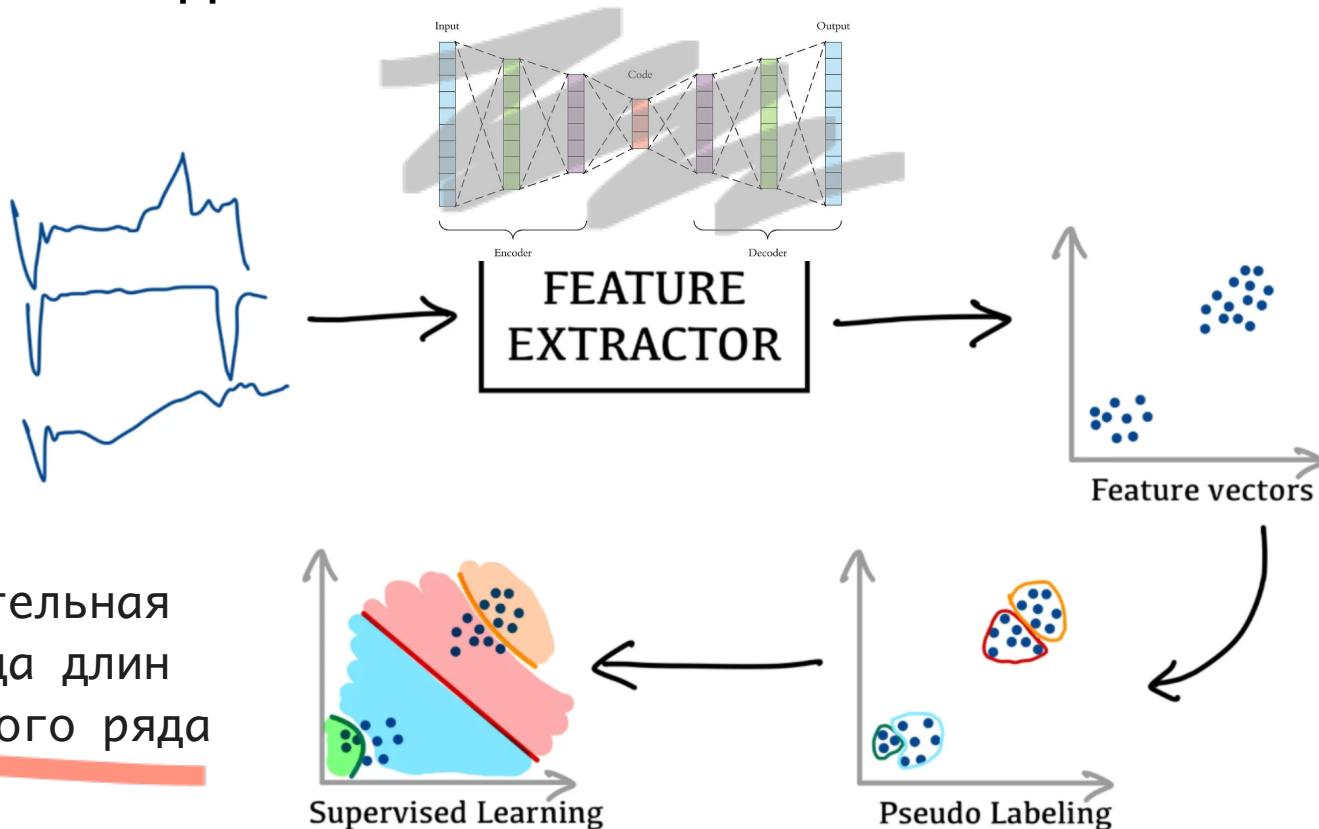
фация 3



# Обработка данных



# Обработка данных



# Обработка данных

Работа с временными рядами

Preprocessing

Fast Fourier Transform



Clustering

Dynamic Time Warping

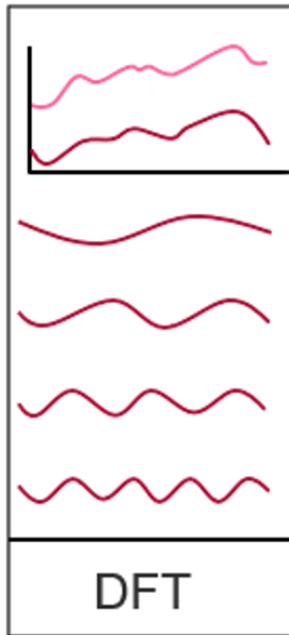


Clustering

Euclidian

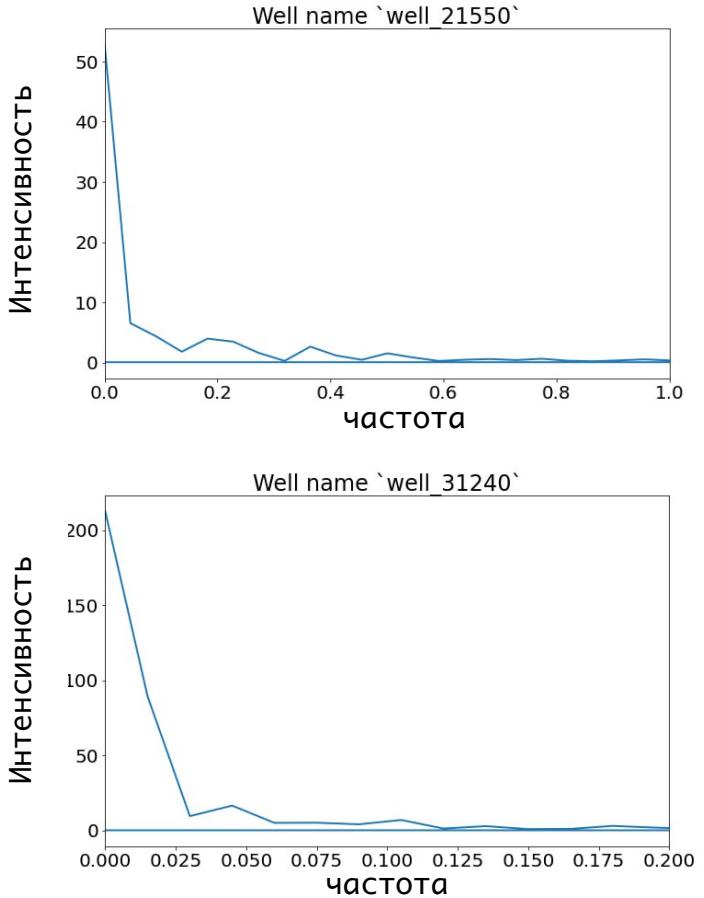


# Обработка данных



Дискретное  
преобразование Фурье

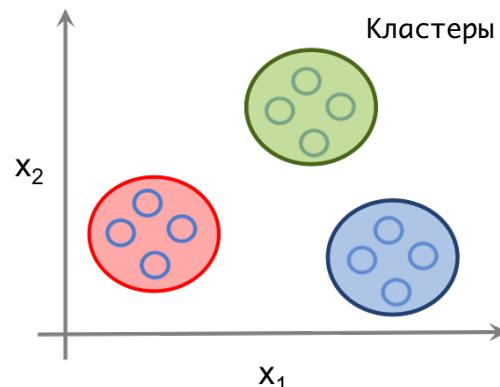
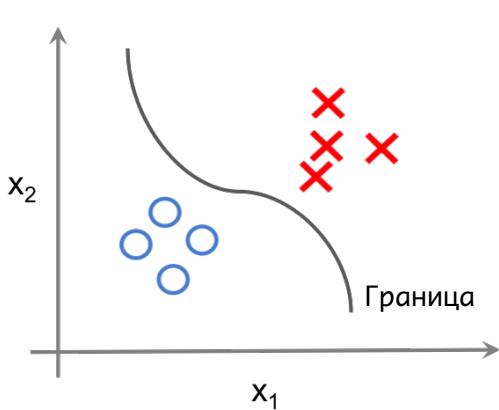
↓  
Поиск значимых  
частот и амплитуд



# Кластеризация

Supervised  
(с учителем)

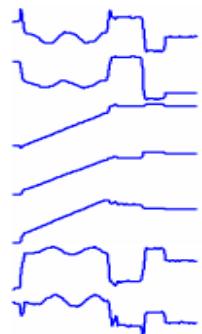
Unsupervised  
(без учителя)



# Кластеризация

Supervised  
(с учителем)

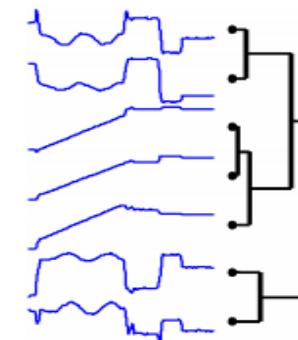
Unsupervised  
(без учителя)



Входные данные

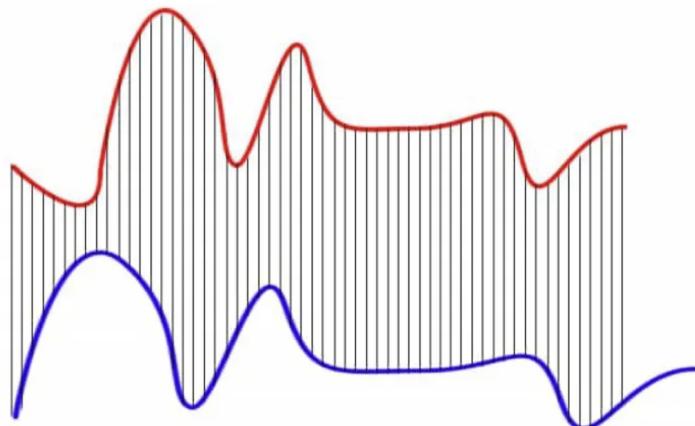
Обработка данных  
(FFT, длина сигнала)

Без предобработки

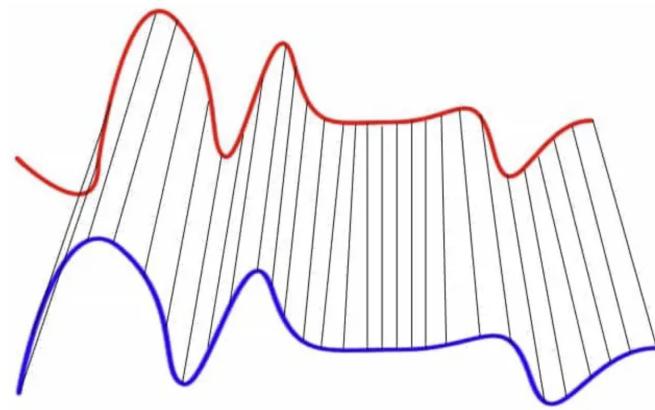


Кластеризация

# Дистанции для кластеризации



Euclidean Matching



Dynamic Time Warping Matching

# Используемые модели

Данные без  
предобработки



KMeans  
KMedoids  
LinkageTree

Обработанные  
данные

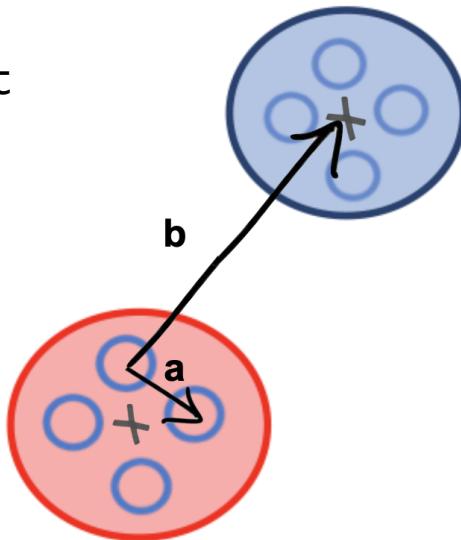


KMeans  
Gaussian Mixture  
Agglomerative Clustering  
Birch

# Оценка качества моделей

Silhouette Coefficient

$$S = \frac{b - a}{\max(a, b)}$$



$S$  - Silhouette Score

$a$  - среднее внутрикластерное расстояние

$b$  - среднее межкластерное расстояние

$$S = 1$$

клusterы далеко,  
четко различимы

$$S = 0$$

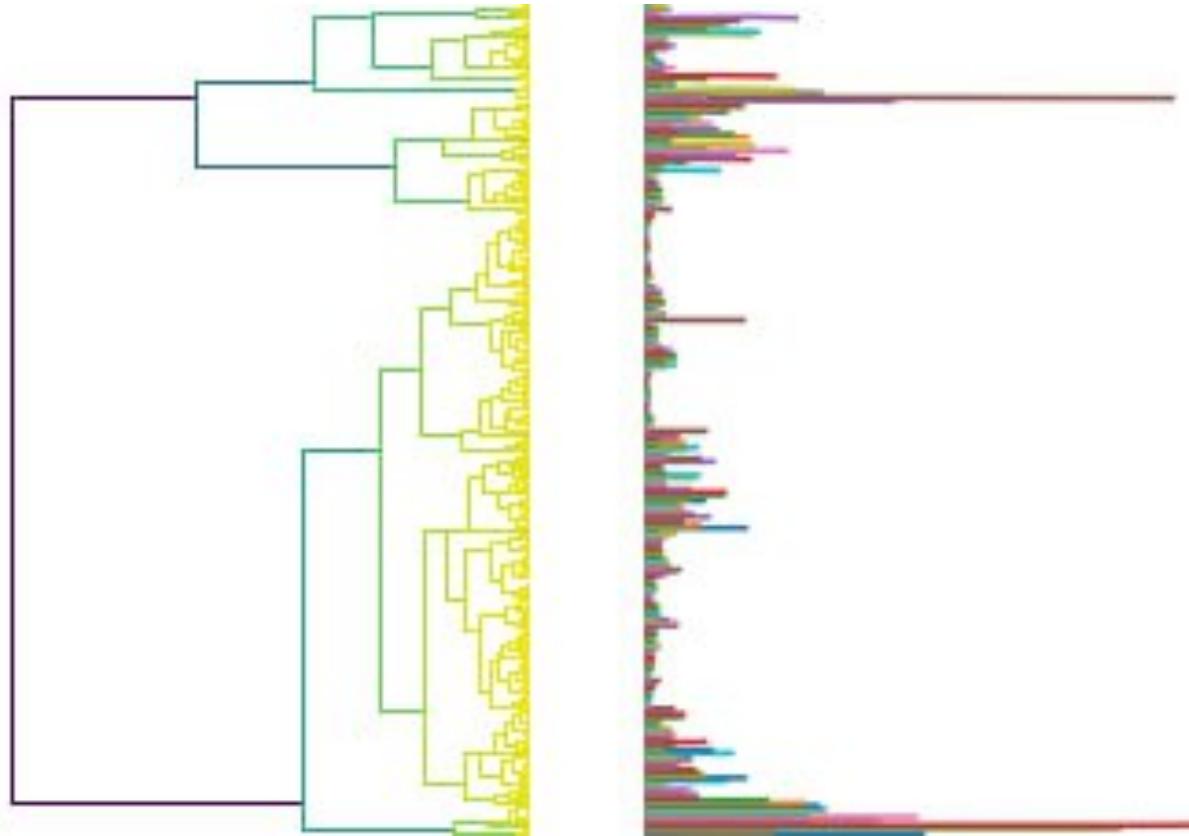
клusterы индифферентны,  
расстояние незначительно

$$S = -1$$

клusterы назначены  
неправильно

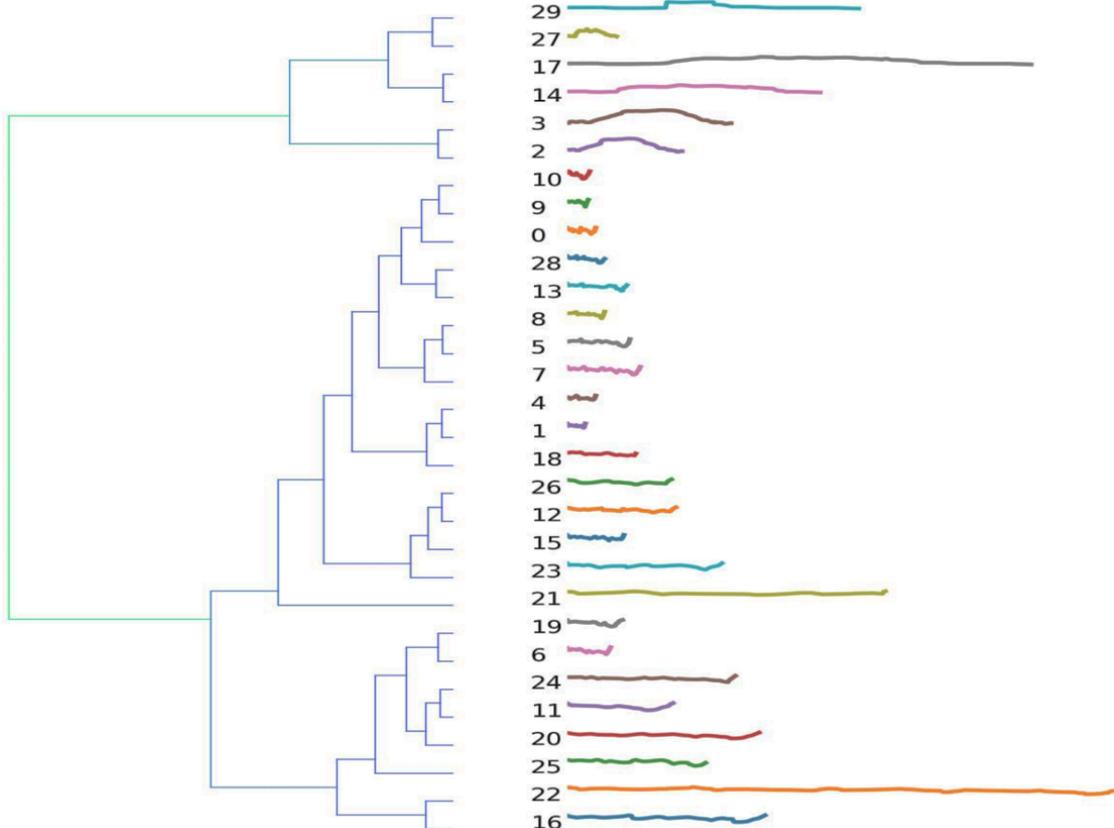
# Результаты

Дендрограмма  
LinkageTree



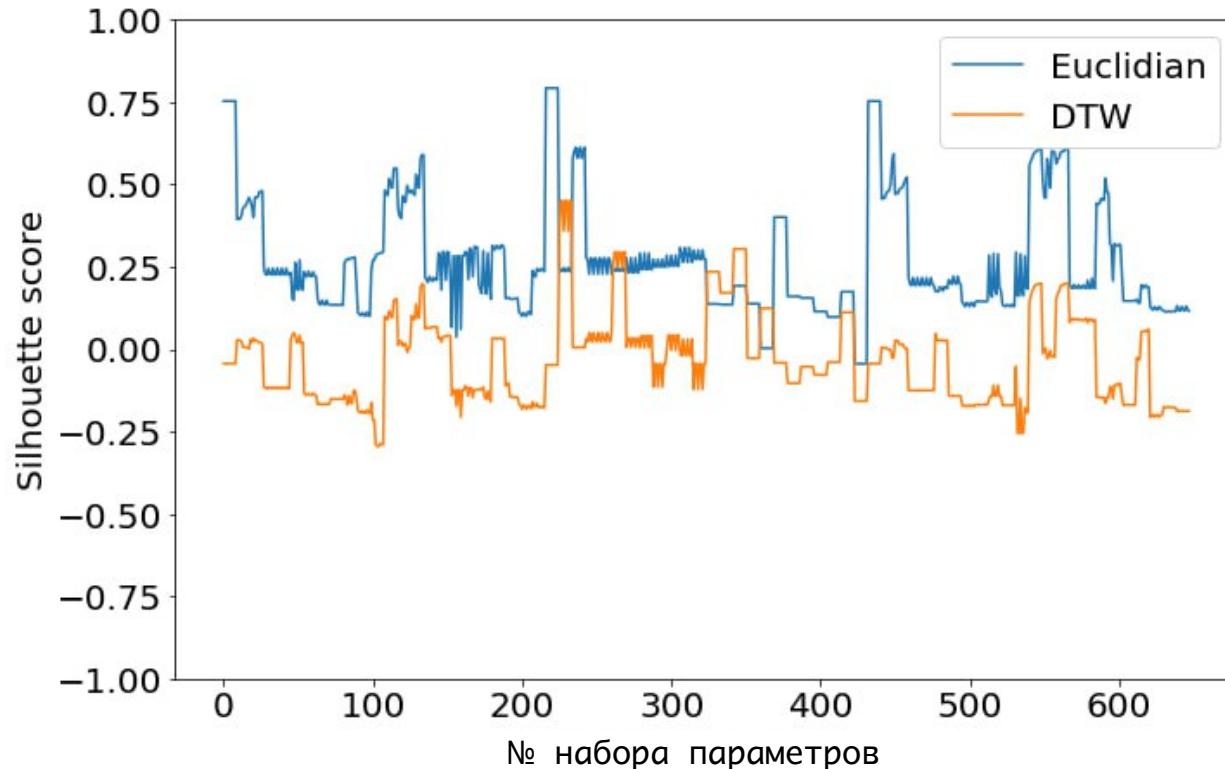
# Результаты

Дендрограмма  
LinkageTree



# Результаты

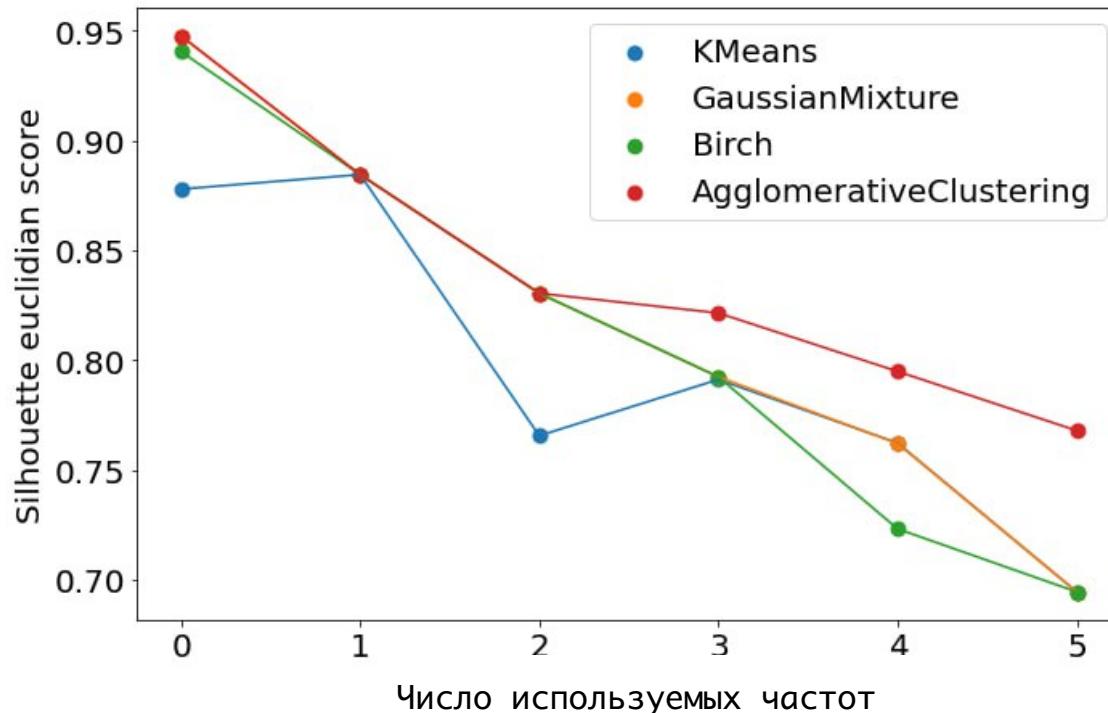
Обработанные данные



GaussianMixture  
3 частоты

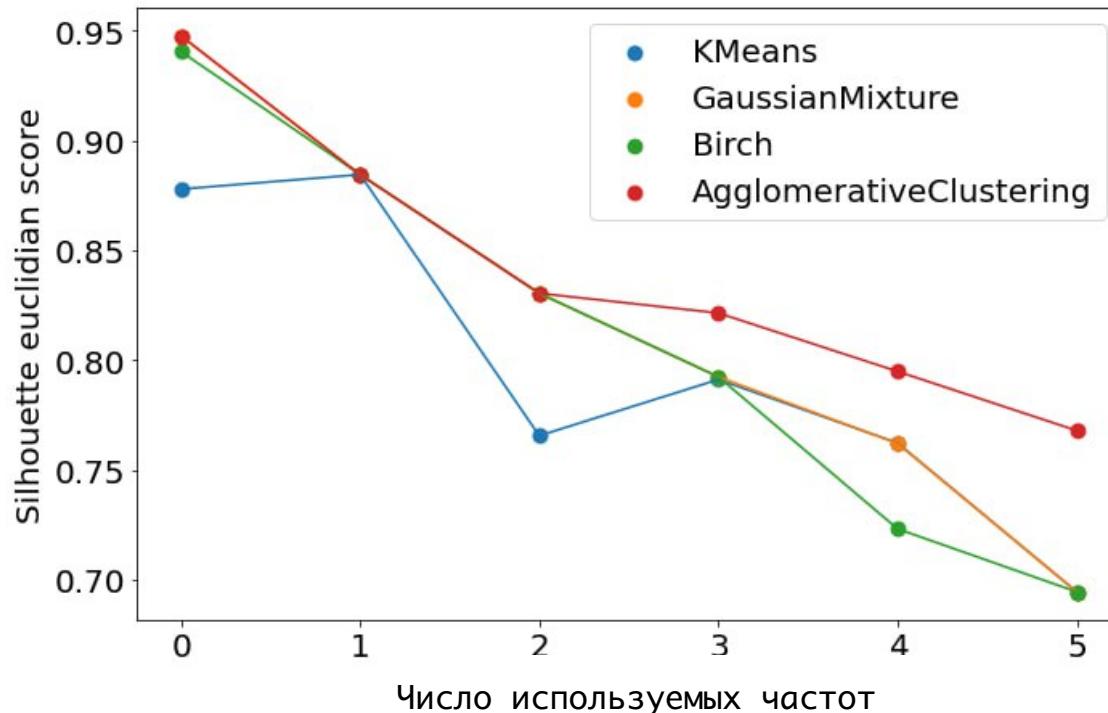
# Результаты

Обработанные данные



# Результаты

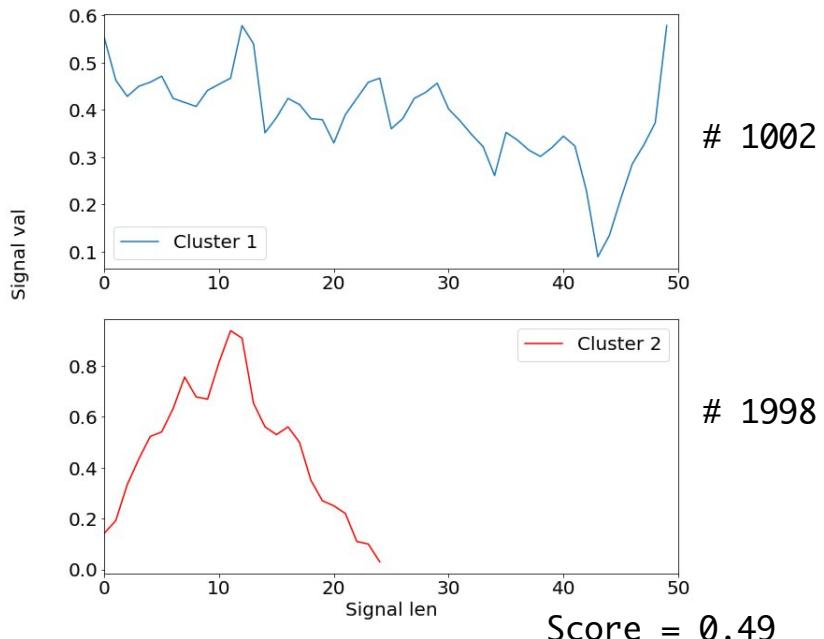
Обработанные данные



Cluster 1: # 2999  
Cluster 2: # 1

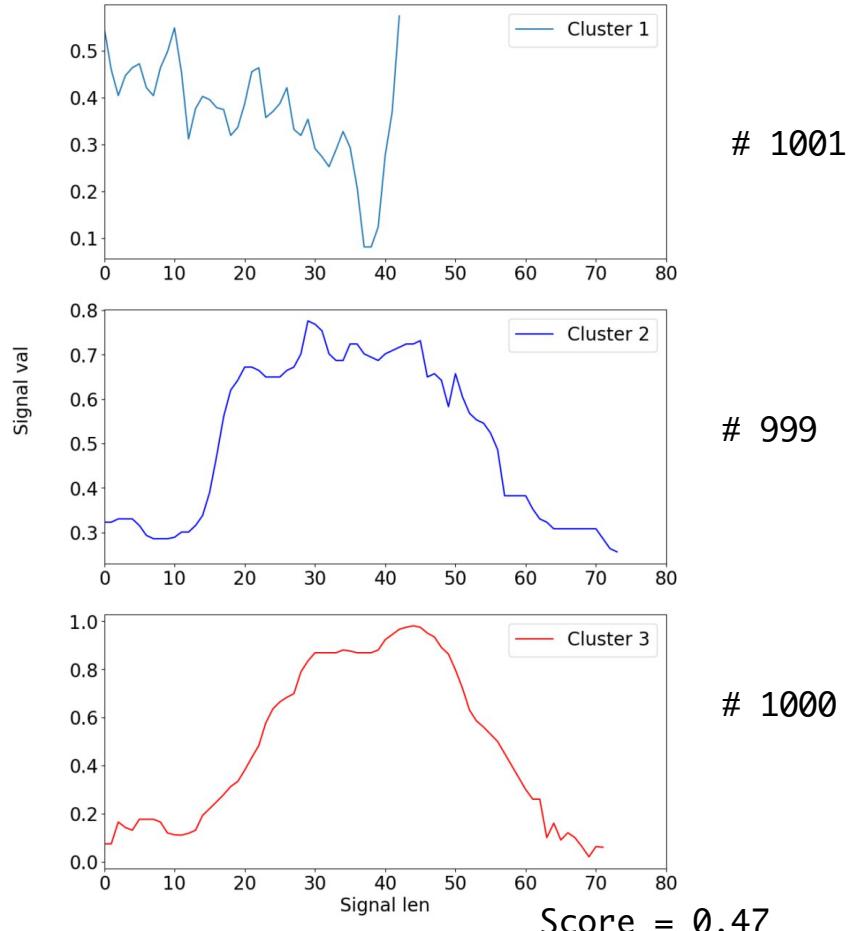
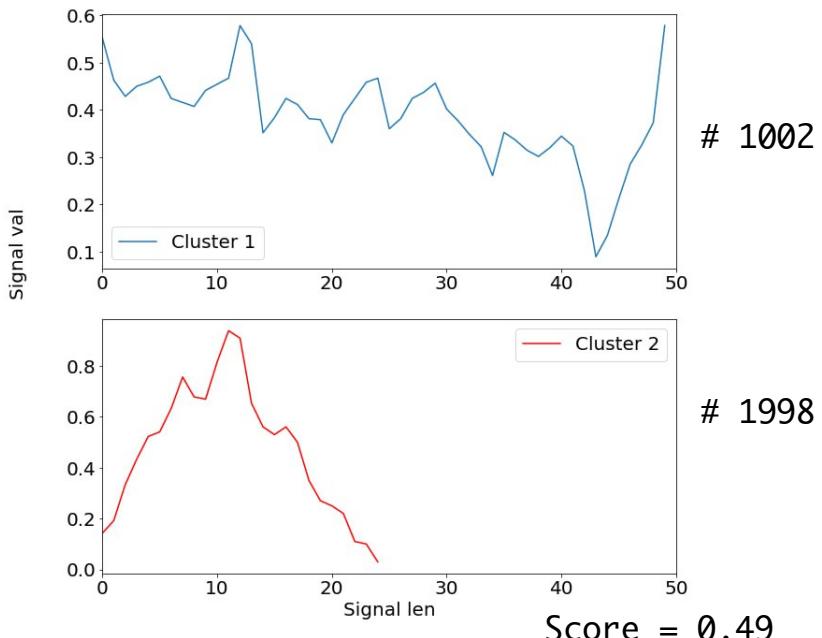
# Результаты

Данные без предобработки



# Результаты

Данные без предобработки



# Выводы

- Подтверждена гипотеза о возможности выполнения задачи кластеризации по данным ГИС на синтетических данных
- Частоты разложения сигнала с FFT не являются значимыми признаками
- Кластеризация необработанных временных рядов дает положительный результат
- Кластеризация необработанных временных рядов на основании дистанции DTW дает высокую точность разбиения на кластеры

# Следующие шаги

- Работа с большим количеством ГИС
- Использование спектрограммы в качестве признака
- Проверка полученных результатов на реальных данных
- Построение фациальной модели месторождения

# Команда



Владислав Трифонов

Data Scientist  
Skoltech

Vladislav.Trifonov@skoltech.ru



Анастасия Волкова

Engineer, Data Scientist Jr  
Skoltech

Anastasia.Volkova@skoltech.ru



Ольга Волкова

Petroleum Engineer, Data Scientist Jr  
Skoltech

Olga.Volkova2@skoltech.ru



Владимир Назаров

Data Scientist  
ITMO

fiji8800@gmail.com



СВЯЗАТЬСЯ С НАМИ



ПОДДЕРЖАТЬ НАС

Thx