

# Топологический анализ изображений лиц

Гвоздева Виктория

Соколова Анастасия

Харчевникова Ангелина

# Цель и Задачи

## Цель:

Инструментами Data Mining сравнить топологию фотографий лиц

## Задачи:

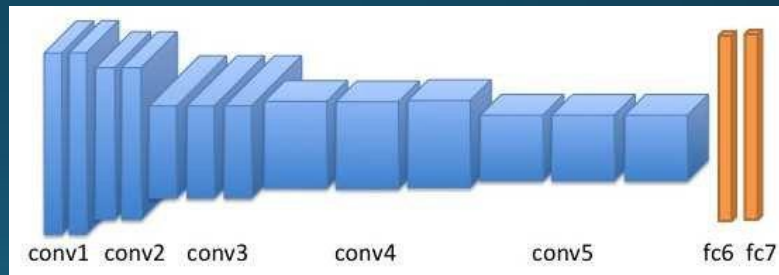
- Выбрать один из существующих публично доступных наборов данных, содержащих фотографии лиц с метками
- Используя глубокие сверточные нейронные сети (подходящие для распознавания лиц), представить фотографии в векторном виде
- Выполнить кластеризацию лиц
- С помощью одной из библиотек для топологического анализа представить полученные данные в виде облака точек
- Интерпретировать результаты

# Инструменты

Набор данных: Labeled Faces in the Wild (LFW)



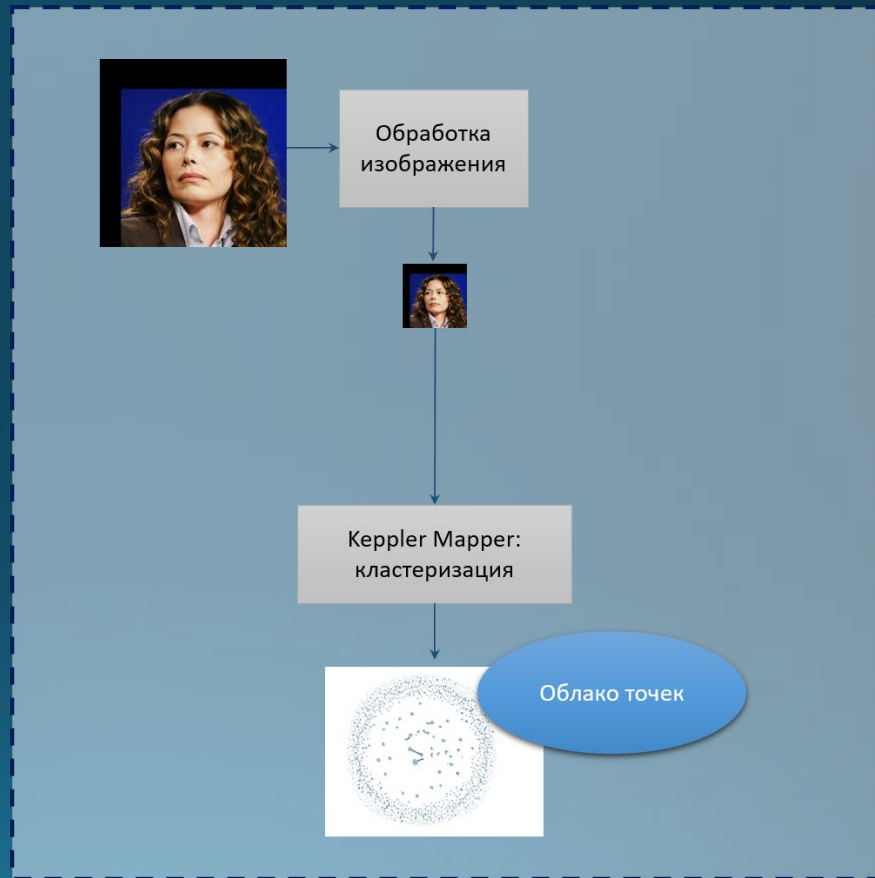
Сверточная сеть: VGGFace2  
(ResNet-50) - вектор признаков из  
2048 элементов



Библиотека: <https://github.com/scikit-tda/kepler-mapper>

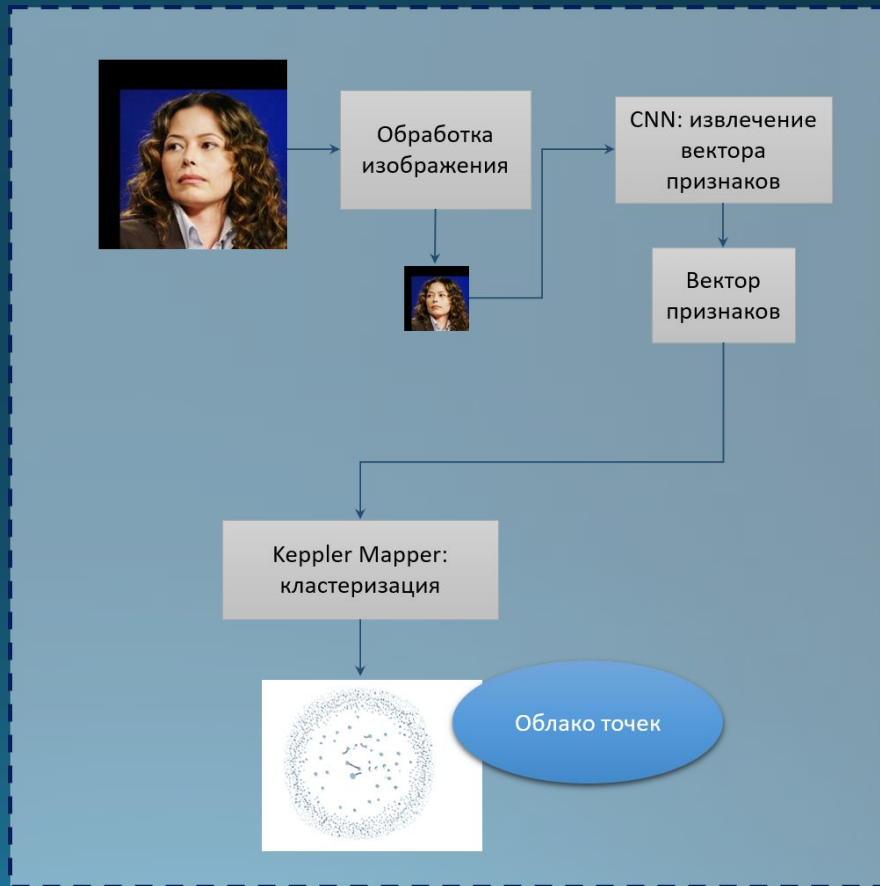
# Подход 1

1. Предобработка изображений
2. Визуализация кластеризации с помощью Kepler Mapper



# Подход 2

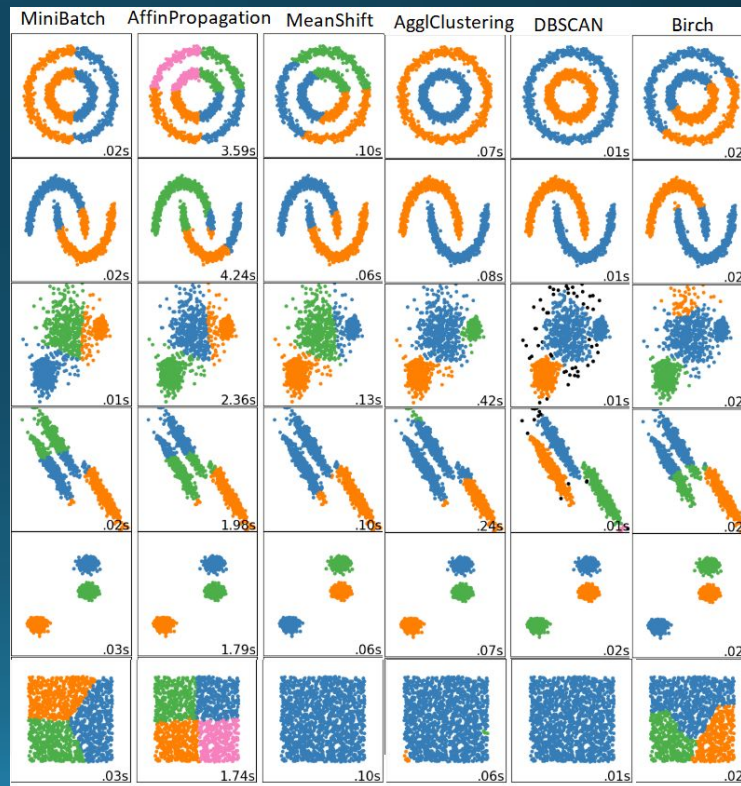
1. Предобработка изображений
2. Подача на вход CNN
3. Получение вектора признаков
4. Визуализация кластеризации с помощью Kepler Mapper



# Алгоритмы кластеризации

1. Агломеративная кластеризация
2. MiniBatchKMeans
3. DBSCAN
4. Birch
5. AffinityPropagation
6. MeanShift

<https://scikit-learn.org/stable/modules/clustering.html>

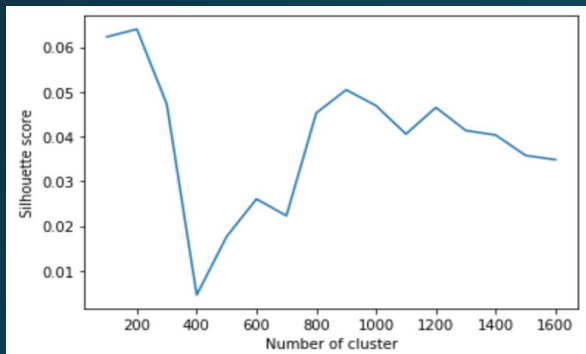


# Метрики

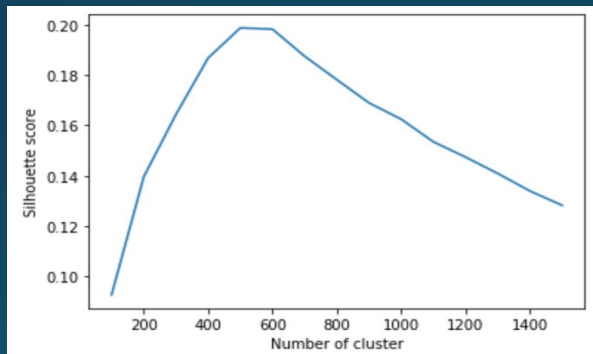
1. Adjusted Rand Index (ARI)
2. Adjusted Mutual Information (AMI)
3. Homogeneity
4. Completeness
5. V-measure
6. Fowlkes-Mallows Scores (FMS)

# Выбор оптимального числа кластеров

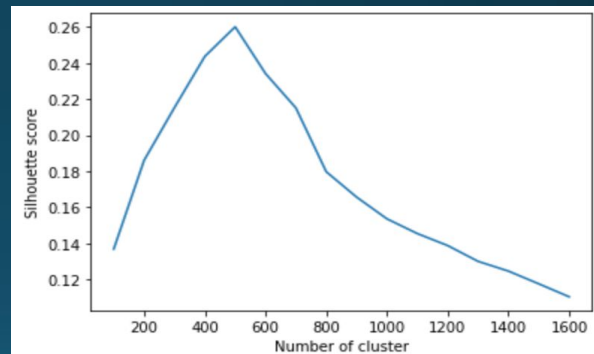
## Silhouette score



MiniBatchKMeans



Birch



AgglomerativeClustering

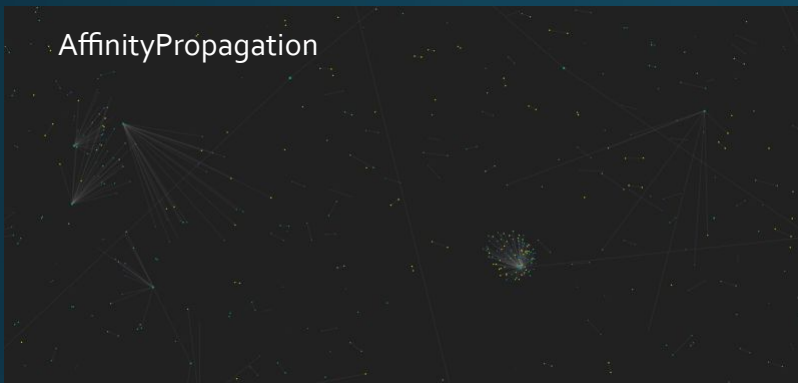


# Результаты

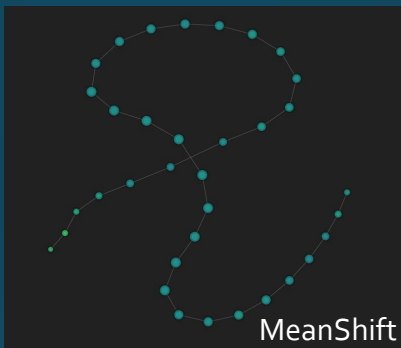
Name of algorithm	Clusters	ARI	AMI	Homogeneity	Completeness	V-measure	FMS
AgglomerativeClustering	500	0.8002	0.7401	0.9852	0.9825	0.9839	0.9246
	500	<b>0.9936</b>	<b>0.9519</b>	<b>0.9957</b>	0.9769	<b>0.9862</b>	<b>0.9937</b>
MiniBatchKMeans	200	0.1987	0.5012	0.9012	0.7234	0.8123	0.3725
	200	0.3468	0.6038	0.8993	0.7619	0.8249	0.3649
DBSCAN	3628	0.0	0.4e-12	0.5396	0.7248	0.6322	0.0
	3739	0.0	1.2e-11	0.6517	<b>1.0</b>	0.7891	0.0
Birch	500	0.2513	0.2103	0.9203	0.9775	0.9480	0.4329
	500	0.3794	0.8283	0.9203	0.9775	0.9480	0.4585
AffinityPropagation	455	0.3008	0.8039	0.9120	0.9222	0.9171	0.4001
	455	0.3336	0.8036	0.9077	0.9478	0.9273	0.4337
MeanShift	3628	0.0	1.7e-11	0.6729	0.9834	0.8282	0.0
	3739	0.0	1.2e-11	0.6517	<b>1.0</b>	0.7891	0.0

# Результаты кластеризации для изображений

AffinityPropagation



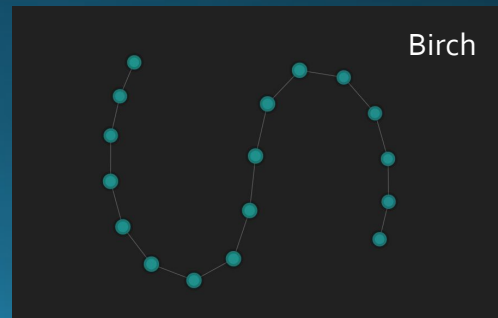
MeanShift



AgglomerativeClustering



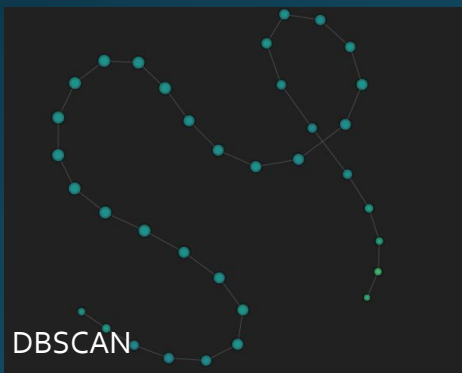
Birch



MiniBatchKMeans



DBSCAN



# Результаты кластеризации для векторов признаков

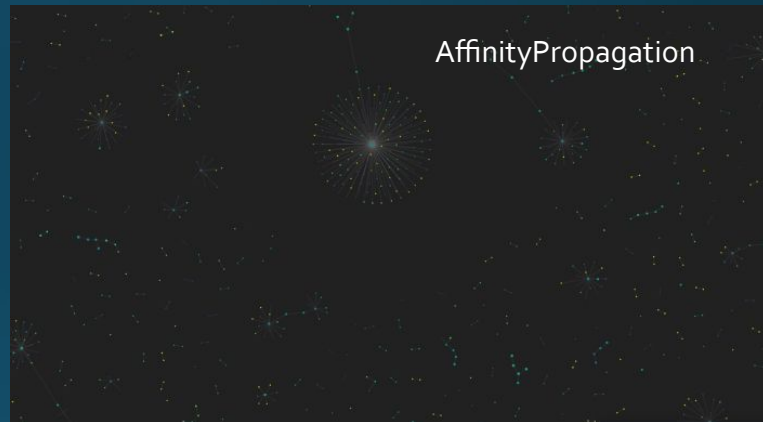
Birch



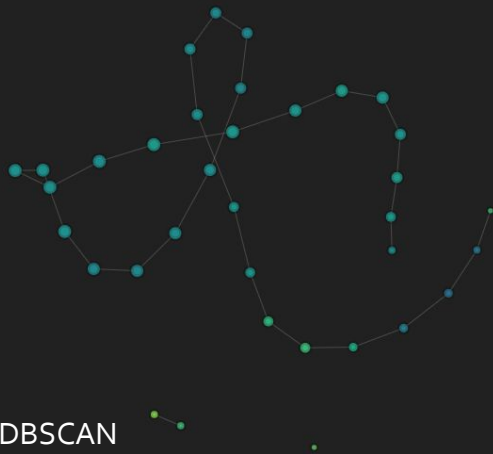
MeanShift



AffinityPropagation



DBSCAN



AgglomerativeClustering



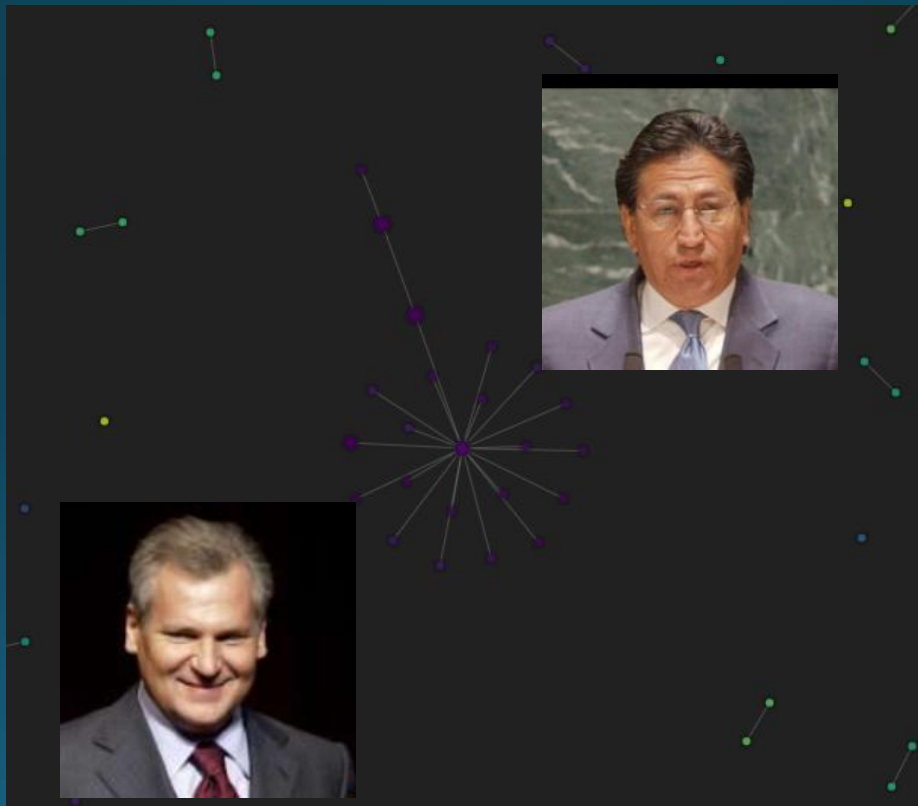
MiniBatchKMeans



# Результаты агломеративной кластеризации для изображений



# Результаты агломеративной кластеризации для векторов признаков



# Заключение

Поставленная цель достигнута и все задачи выполнены

Результаты экспериментального исследования:

- Использование векторов признаков демонстрирует более высокие показатели по точности кластеризации и производительности по сравнению с рассмотрением изображений as-is
- Наиболее показательные результаты демонстрирует алгоритм Агломеративная Кластеризация
- С помощью библиотеки Kerplur Mapper визуализированы результаты кластеризации лиц: алгоритмы с высокими показателями точности продемонстрировали более точное объединение фотографий лиц одного и того же человека

Спасибо за внимание!