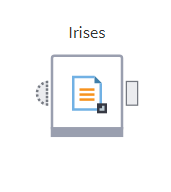
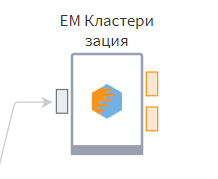
Отчет по практической работе №3. Кластеризация

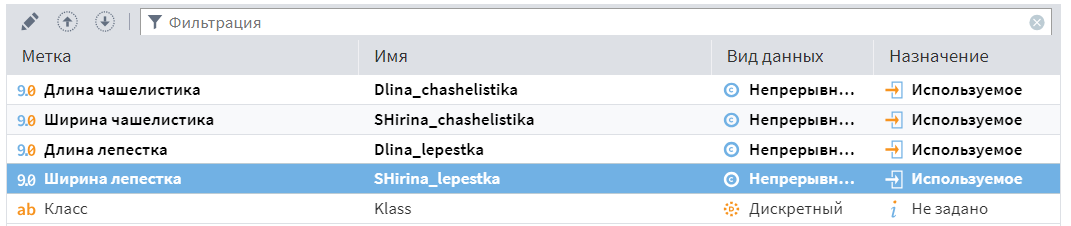
**ЕМ Кластеризация**

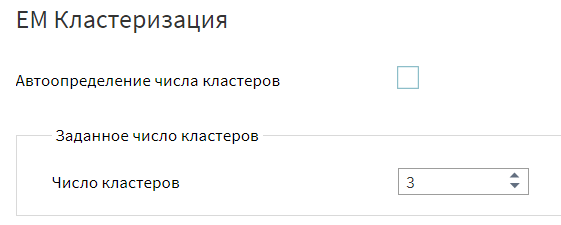
1)Импортируем текстовый документ в наш проект



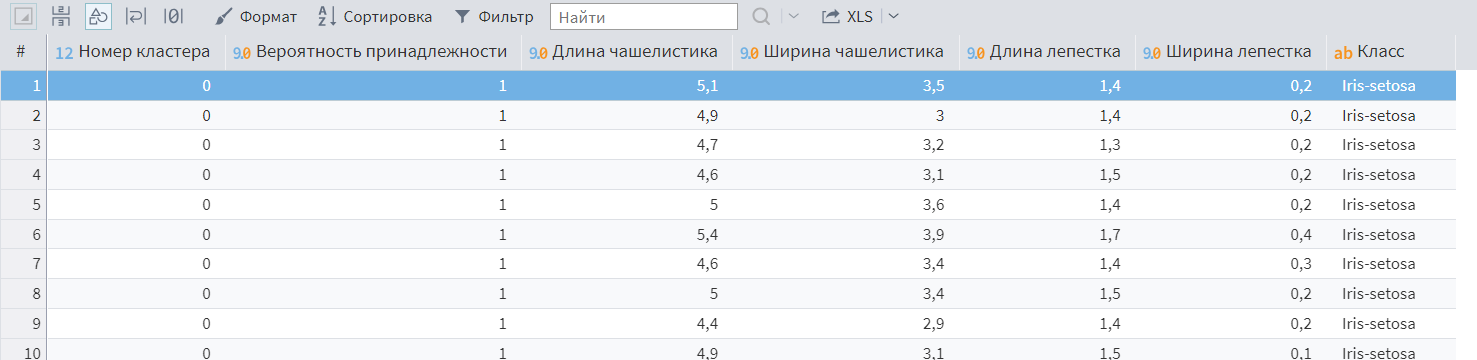
2)Создаем узел EM Кластеризация, и настраиваем его

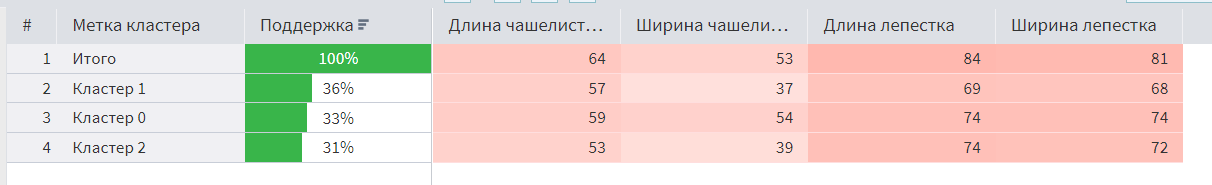


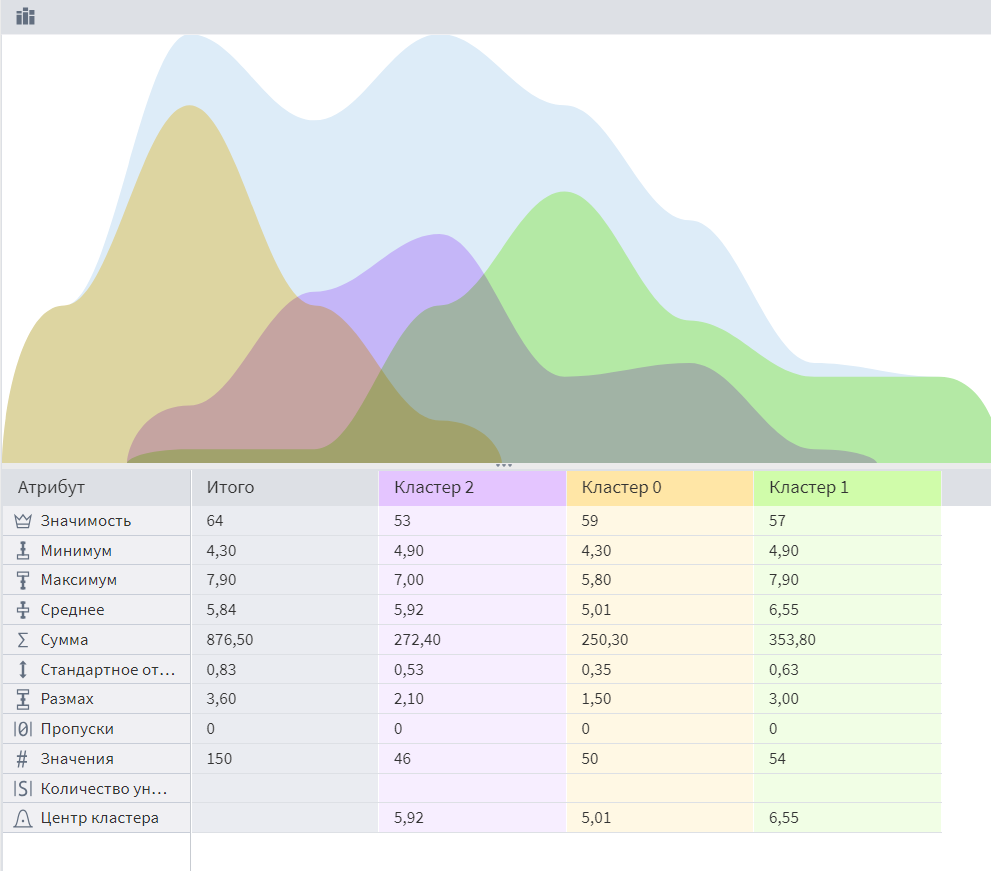




3)Обучаем модель, и визуализируем полученные результаты в виде таблицы







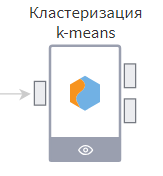
Вывод: Алгоритм выделил 3 кластера, которые

совпадают с количеством исходных классов и примерно равны, что говорит о

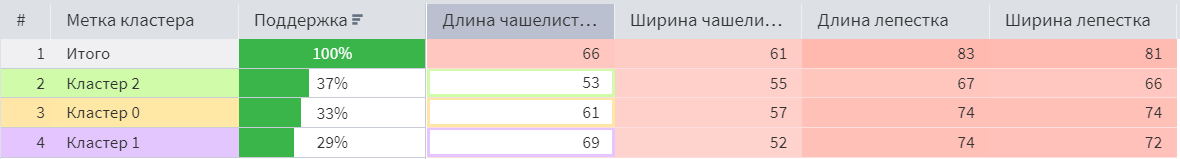
хорошей работе алгоритма EM кластеризации.

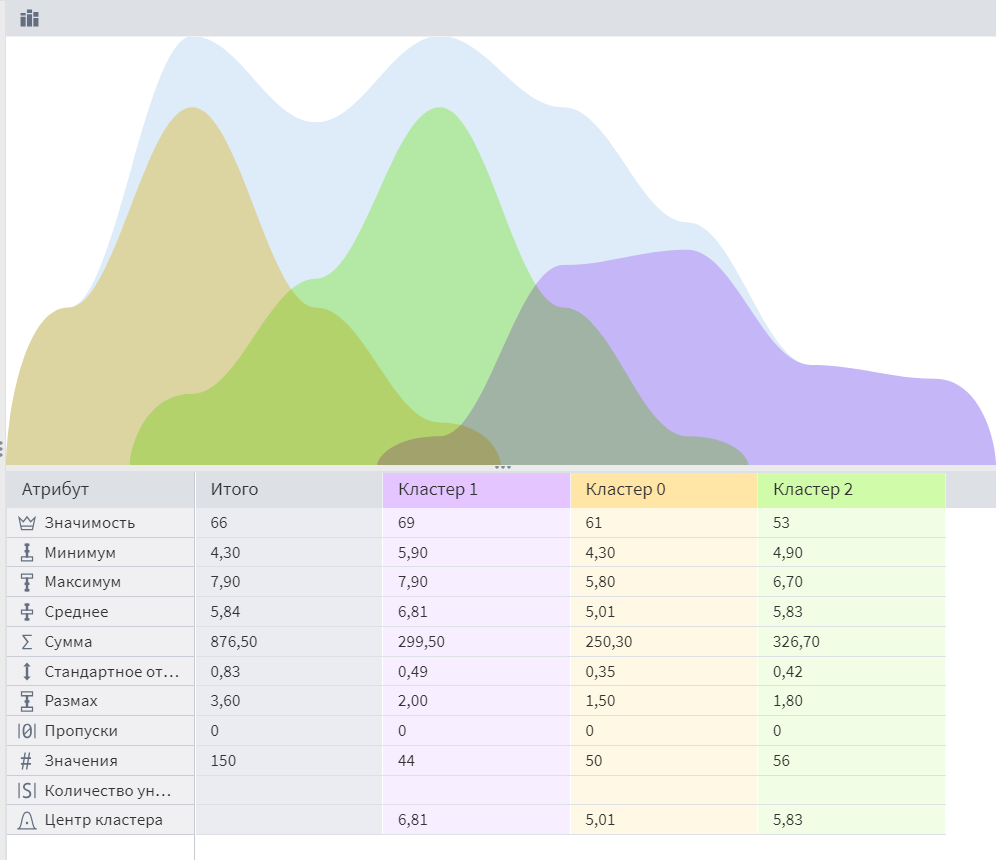
**Кластеризация k-means**

4)По тому же принципу создаем узел Кластеризация, называем k-means и применяем к нему те же настройки, что и в узле EM Кластеризация



6)Обучаем узел, и проверяем результаты

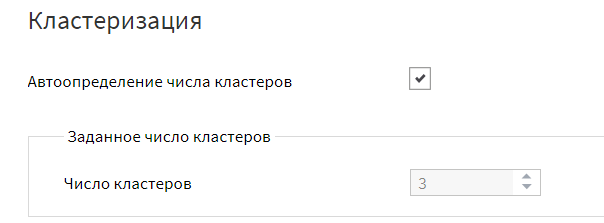




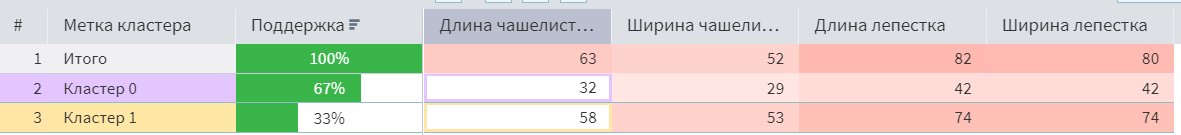
Вывод: Алгоритм выделил 3 кластера, которые совпадают с количеством исходных классов, и примерно равны, тем не менее, результат работы этого алгоритма отличается от результата предыдущего алгоритма

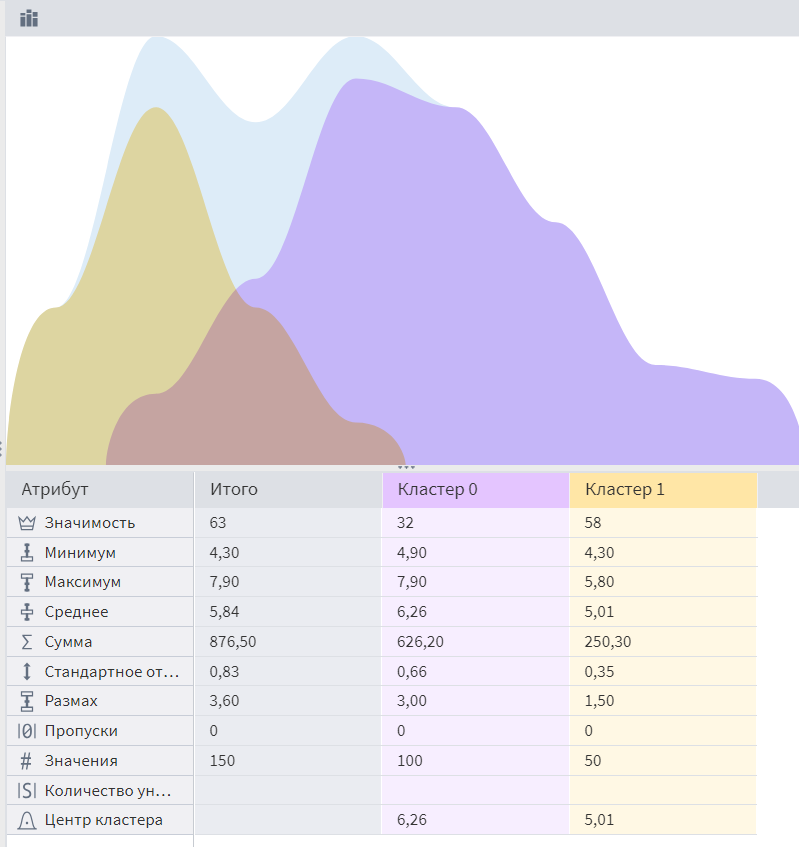
**Кластеризация g-means**

7)Создаем еще один узел кластеризации, и называем его g-means. Применяем к нему те же настройки, что и в предыдущих узлах, за исключением одного параметра, устанавливаем автоопределение числа кластеров



8)Обучаем модель, и проверяем результаты





Вывод: Алгоритм выделил 2 кластера, количество которых не совпадает с исходным количеством данных, что говорит о не точной работе алгоритма.

**Общий вывод**: В результате проделанной работы можно сказать, что наиболее точными методы кластеризации оказались k-means и EM Кластеризация.

**Вопросы:**

1. Какой метод дал наиболее четкое разделение на кластеры?

**Ответ**: Наиболее четкое разделение на кластеры дали методы: k-means и EM Кластеризация

1. Совпадает ли число кластеров, найденное g-means, с исходным классами ирисов?

**Ответ**: Число кластеров, найденное методом g-means не совпадает с исходным количеством классов

1. Какие признаки (длина/ширина чашелистика/лепестка) наиболее значимы для кластеризации?

**Ответ**: Во всех трех методах наиболее значимый признак для кластеризации это – Длинна лепестка