

**Санкт-Петербургский национальный исследовательский университет  
информационных технологий, механики и оптики**

**Отчет по практической работе №1 по дисциплине:  
«Проектирование интеллектуальных систем в управлении»**

**Подготовила: Лукина А.С.**

**Группа: К3443**

**Преподаватель: Бережков Андрей Вячеславович**

Задание: Построение потока работ, выполняющего решение задачи кластеризации.

Ход работы:

Был построен поток работ, отображающий задачу кластеризации с помощью алгоритма K-means. Построенный поток работ представлен на рисунке 1.

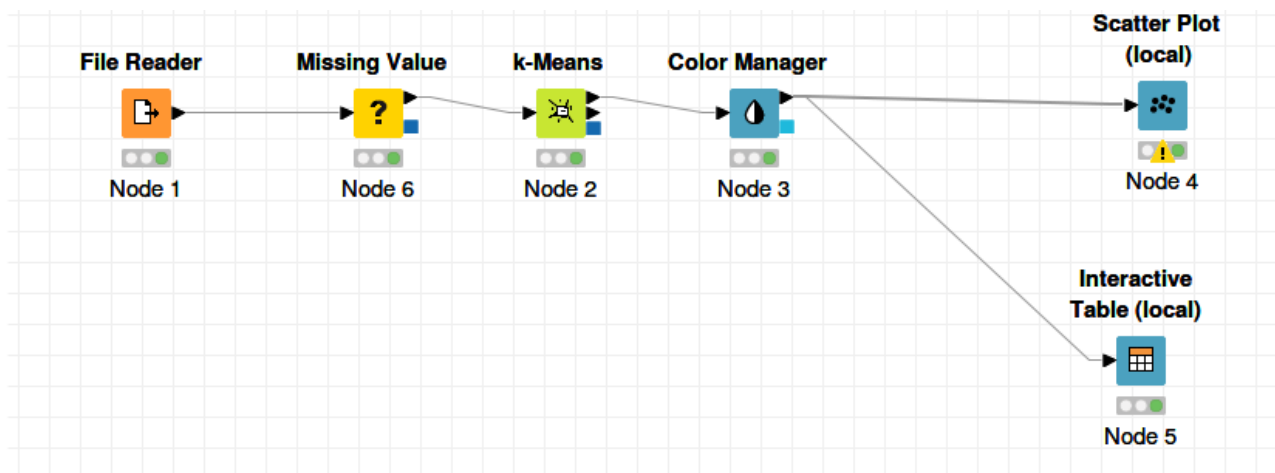


Рисунок 1— Построенный поток работ для кластеризации

В первую очередь был настроен узел «File Reader» и загружены данные для кластеризации. Настройка узла представлена на рисунке 2.

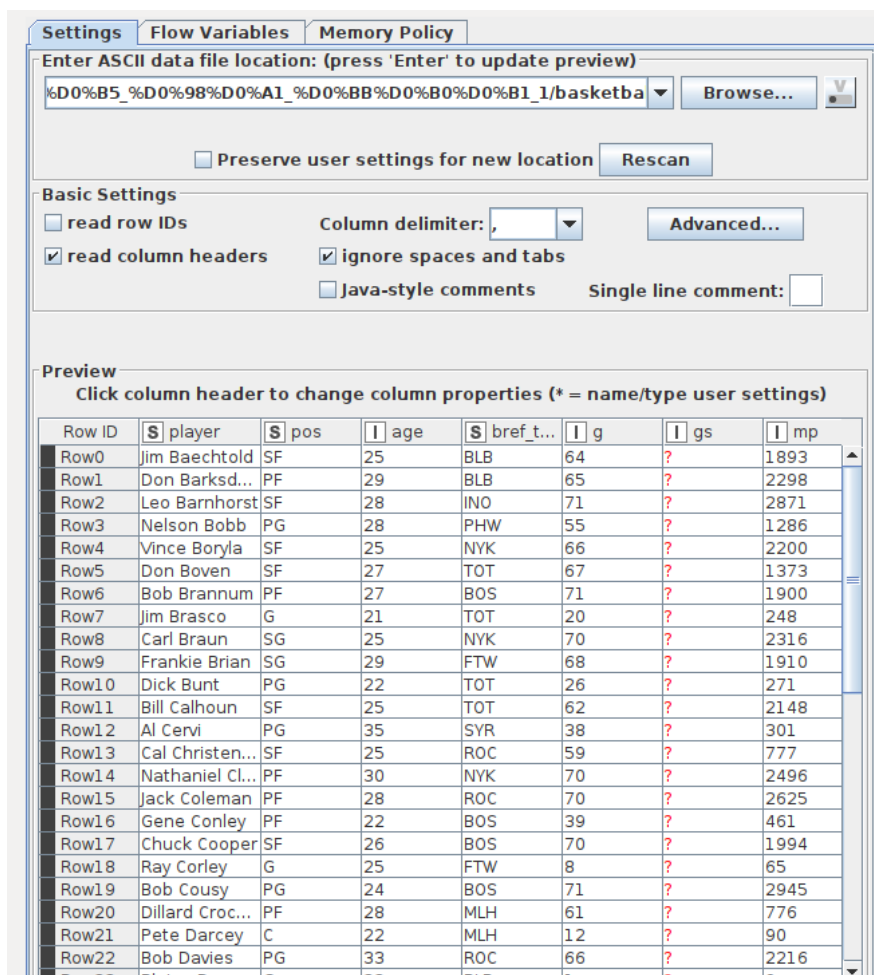


Рисунок 2— Настройка узла "File Reader"

Помимо необходимых по заданию узлов был добавлен узел «Missing Value», который обрабатывает пропущенные значения. Для строкового типа данных — голосование по большинству (самое частое значение), для числового типа данных — среднее по признаку. Настройка данного узла представлена на рисунке 3.

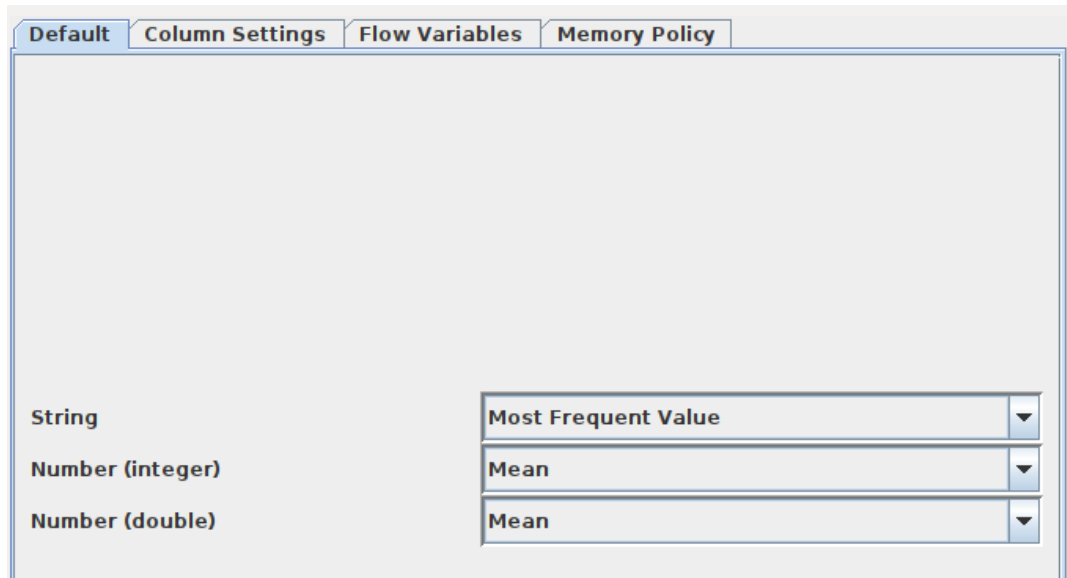


Рисунок 3— Настройка узла "Missing value"

Для настройки узла «k-Means» были установлены различные значения полей числа кластеров «Number of clusters». Настройка данного узла представлена на рисунке 4.

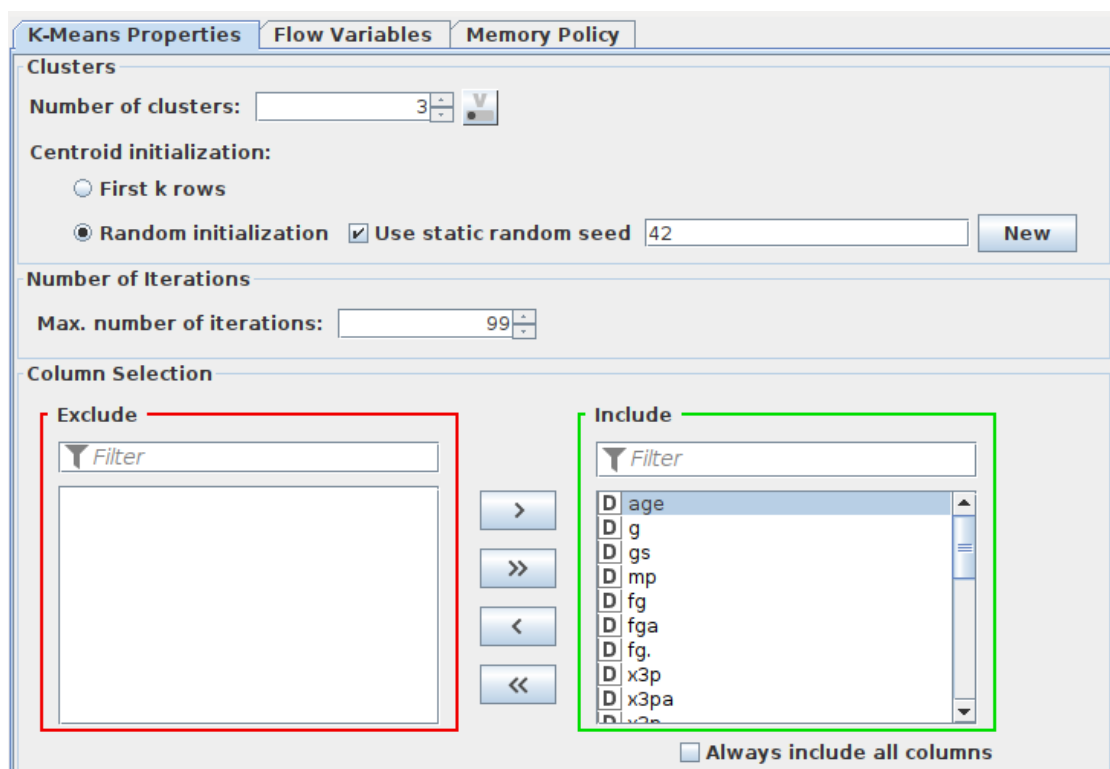


Рисунок 4— Настройка узла "k-Means"

Для того, чтобы настроить цвет для кластеризации были произведены настройки узла «Color Manager», где были выбраны цвета. Настройка данного узла представлена на рисунке 5.

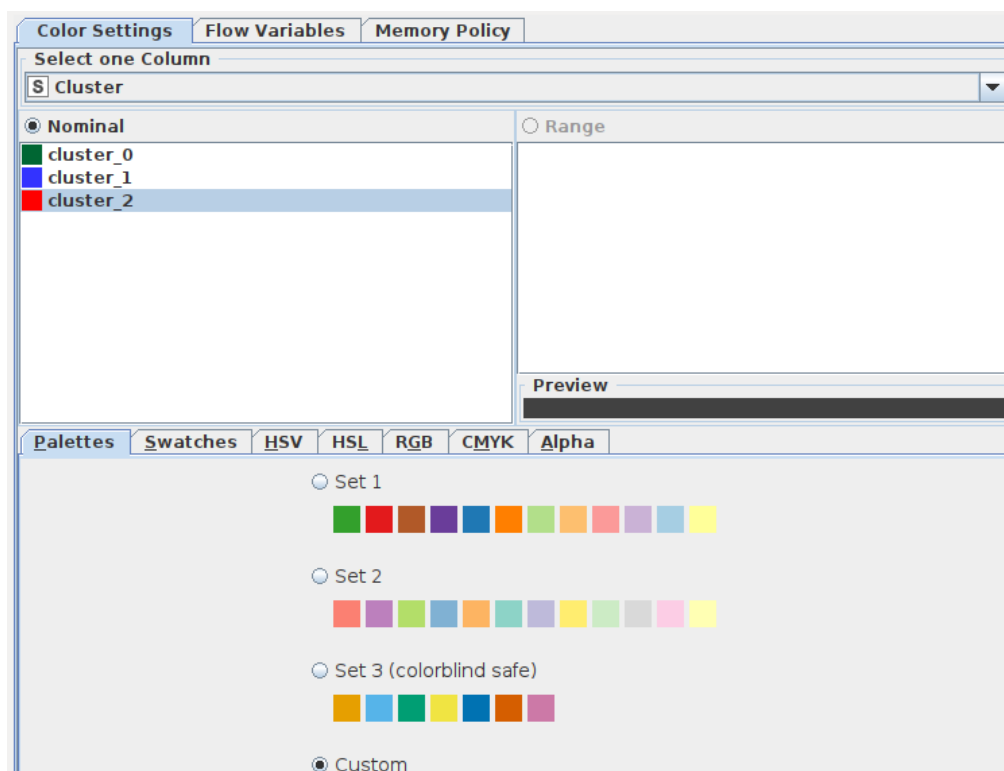


Рисунок 5— Настройка узла «Color Manager»

После настройки всех узлов и выполнения потока работ через узел «Scatter plot» была произведена визуализация кластеризации. На рисунках 6-8 представлены результаты для кластеризации с количеством кластеров — 3, 5 и 7 и разными осями x и y.

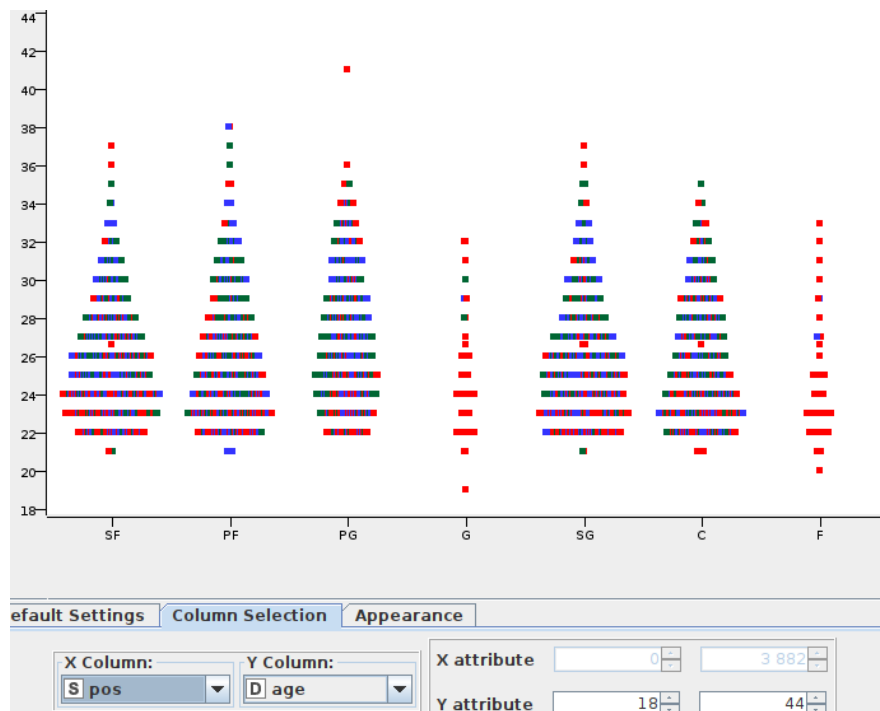


Рисунок 6— Результат кластеризации для параметров pos и age с количеством кластеров — 3

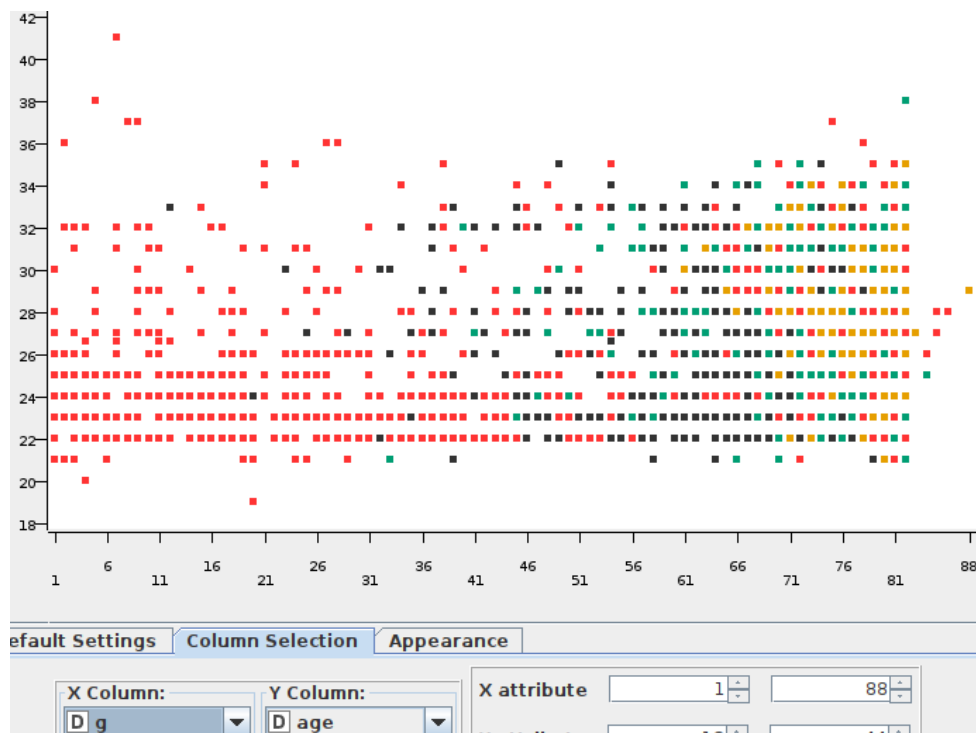


Рисунок 7— Результат кластеризации для параметров g и age с количеством кластеров — 5

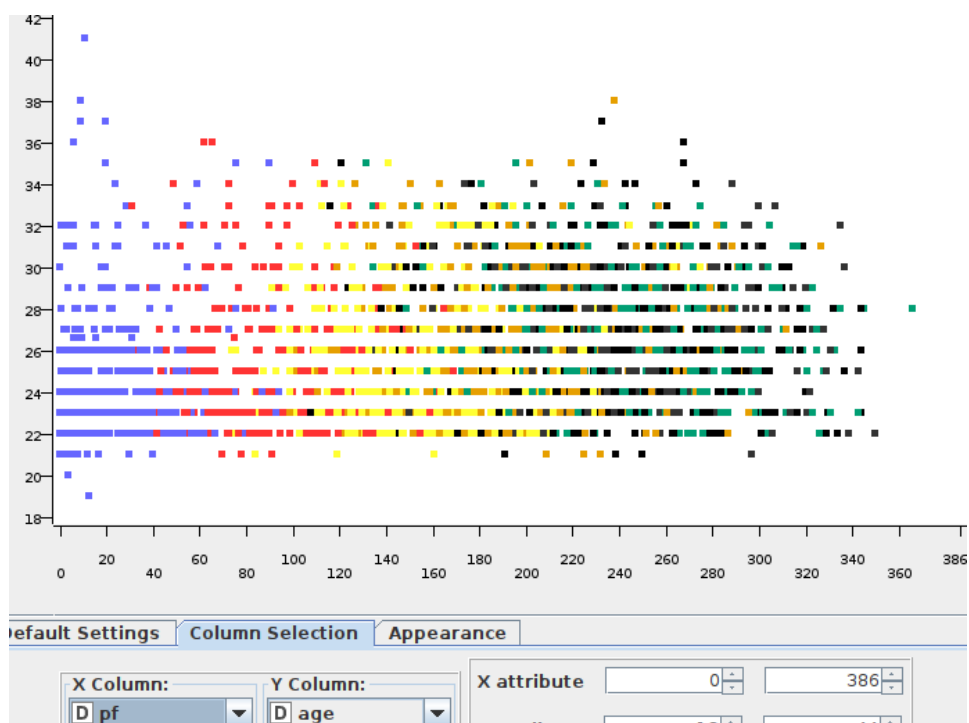


Рисунок 8— Результат кластеризации для параметров pf и age с количеством кластеров - 7