|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| https://lh6.googleusercontent.com/QcftzNtI05T0Y6fjdSh1Rr2rt8oqZ1IvnLvbn1jLJ7CCyteVir3k-xBLv4SL1wAgWJsRhmmJSR0UW-RP63_GQenE4vVWv05BRoZTsmIcBccVTnfxwmsnNMvjg599x9SqZd8E3dkd   |  |  |  | | --- | --- | --- | | МИНОБРНАУКИ РОССИИ | | | | Федеральное государственное бюджетное образовательное учреждение  высшего образования  **«МИРЭА – Российский технологический университет»**  **РТУ МИРЭА** | | | |  | | | |  |
|  |  |

**Институт информационных технологий (ИИТ)**

**Кафедра инструментального и прикладного программного обеспечения (ИППО)**

|  |  |
| --- | --- |
| **ОТЧЁТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №1** | |
| **по дисциплине** | |
| «Программное обеспечение интеллектуальных систем» | |
|  | |
| Выполнил студент группы ИКБО-02-17 | *Верба Д.С.* |
| Принял | *Зорина Н.В.* |

|  |  |  |
| --- | --- | --- |
| Работа выполнена | «\_\_»\_\_\_\_\_\_\_2020 г. | *(подпись студента)* |
|  |  |  |
| «Зачтено» | «\_\_»\_\_\_\_\_\_\_2020 г. | *(подпись руководителя)* |

Москва 2020

**Цель работы**

Основной целью работы является познакомиться с мощным графическим инструментом для анализа данных Knime. Работа подразделяется на три цели:

1) Ознакомиться с механизмом поиска шаблонов (правил);

2) Научиться работать с деревом решений;

3) Освоить возможности Knime по кластеризации.

**Ход работы**

Используемый датасет – данные о продажах игр по жанрам.

**Поиск шаблонов**

*Задача*

Построение потока работ, выполняющего решение задачи анализа рыночной корзины и поиска ассоциативных правил. Данный поток должен выполнять следующую последовательность действий: загрузить данные из текстового файла, преобразовать загруженные данные в специализированный тип данных пакета KNIME, найти частые наборы и ассоциативные правила, вывести результаты.

*Выполнение*

Последовательность действий:

1. Создайте поток работ, приведенный на рисунке 1.

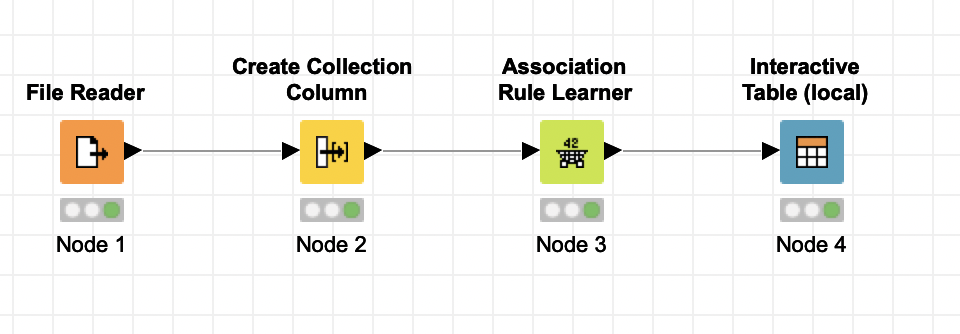


Рисунок 1 – поток данных для работы

1. Выполните настройку узла “File Reader” так, чтобы (рисунок 2):

* в качестве файла исходных данных фигурировал baskets.csv (изменен по рекомендации преподавателя);
* первая строка файла трактовалась как содержащая названия столбцов (“Read column headers”);
* в качестве разделителя столбцов фигурировала запятая;
* узел обрабатывал неполные строки (кнопка “Advanced”, вкладка “Short lines”).

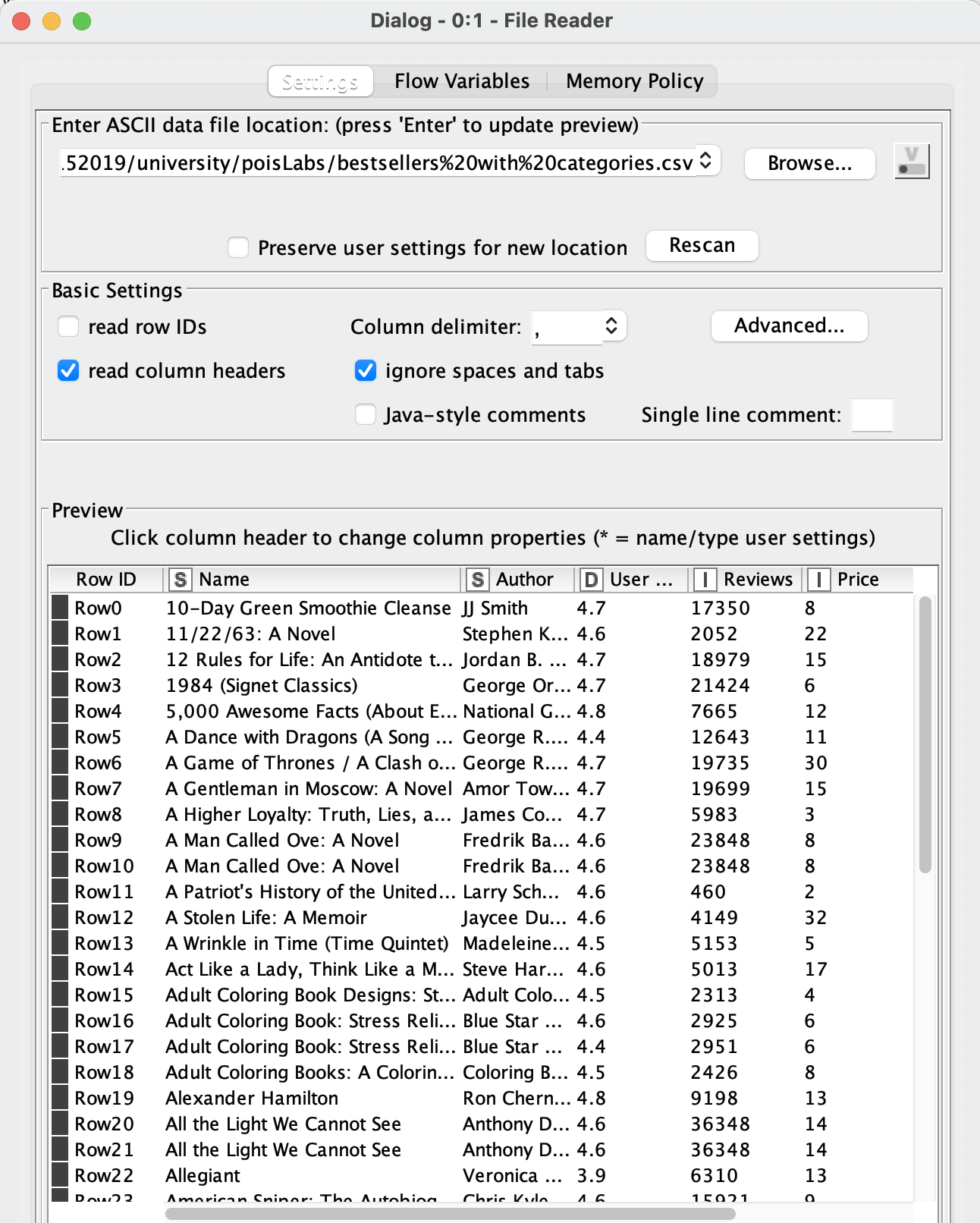


Рисунок 2 – настройка File Reader

1. Выполните настройку узла “Create Collection Column” так, чтобы (рисунок 3):

* все строки исходного файла попали в выходную коллекцию данных;
* установите флаги “Create collection of type set”, “Ignore missing values”, “Remove aggregated columns from table”

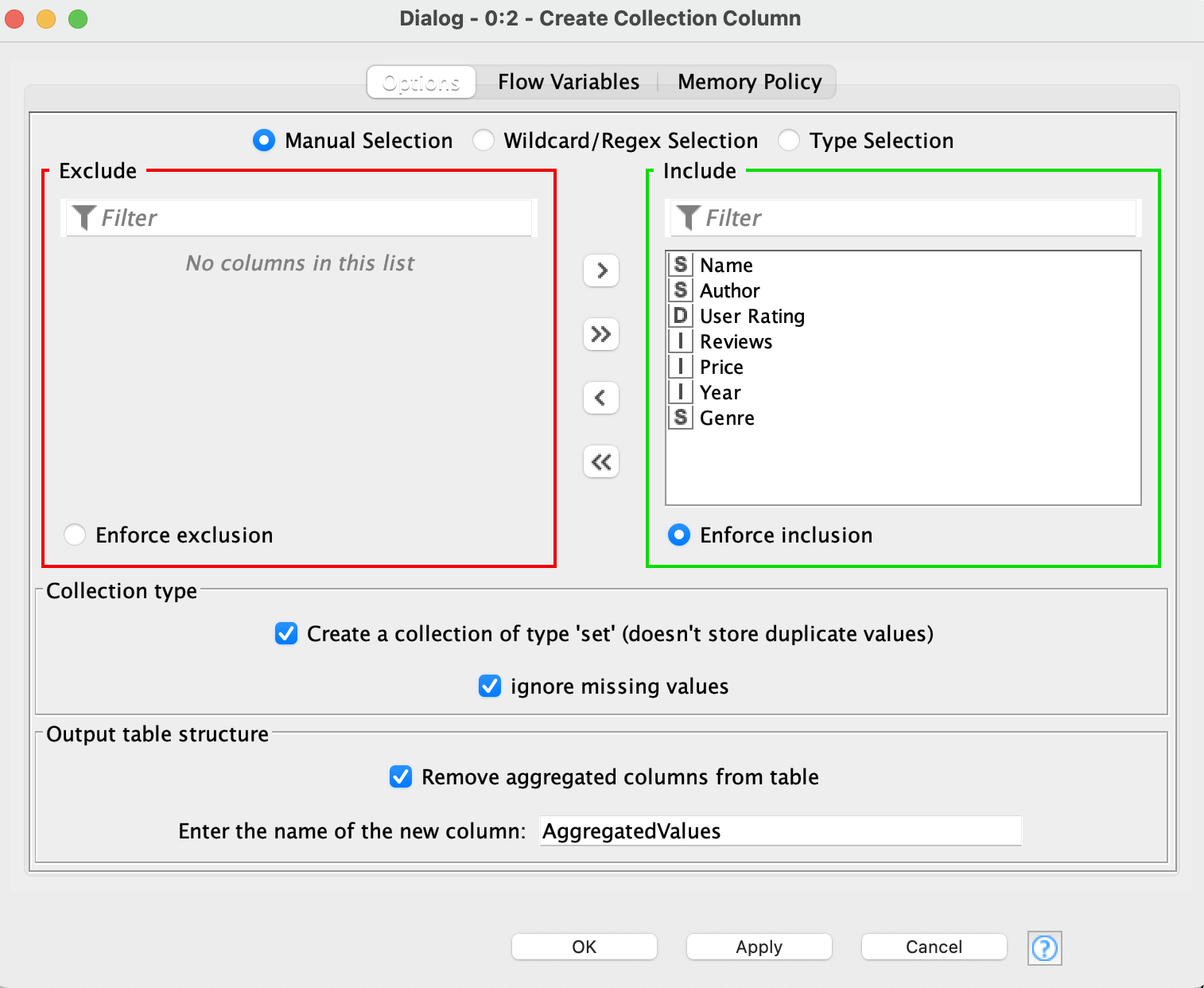


Рисунок 3 – настройка Column Collection

1. Выполните поток работ, предварительно настроив узел “Association Rule Learner”, указав различные значения поддержки (рис. 4 и рис. 5). В таблице иллюстрируется доля содержания определенного значения с исходных данных.

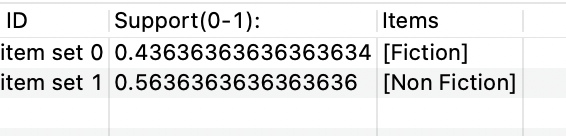


Рисунок 4 – результат работы с указанием minimal support равным 0.03

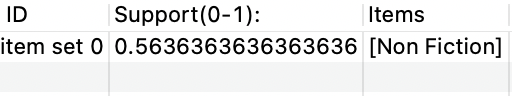


Рисунок 5 – результат работы с указанием minimal support равным 0.05

**Деревья решений**

*Задача*

Построение потока работ, выполняющего решение задачи классификации посредством построения дерева решений. Данный поток должен выполнять следующую последовательность действий: загрузить данные из текстового файла, построить дерево решений, вывести результаты.

*Выполнение*

Построим поток работы (рис. 6) и для File Reader зададим в качестве исходных данных файл student\_marks.csv, указав, что нам необходимо трактовать первую строку как содержащую названия столбцов (рис. 7)

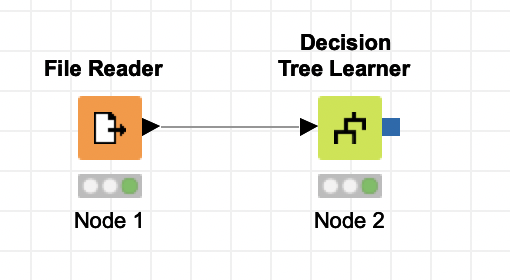


Рисунок 6 – создание потока работ

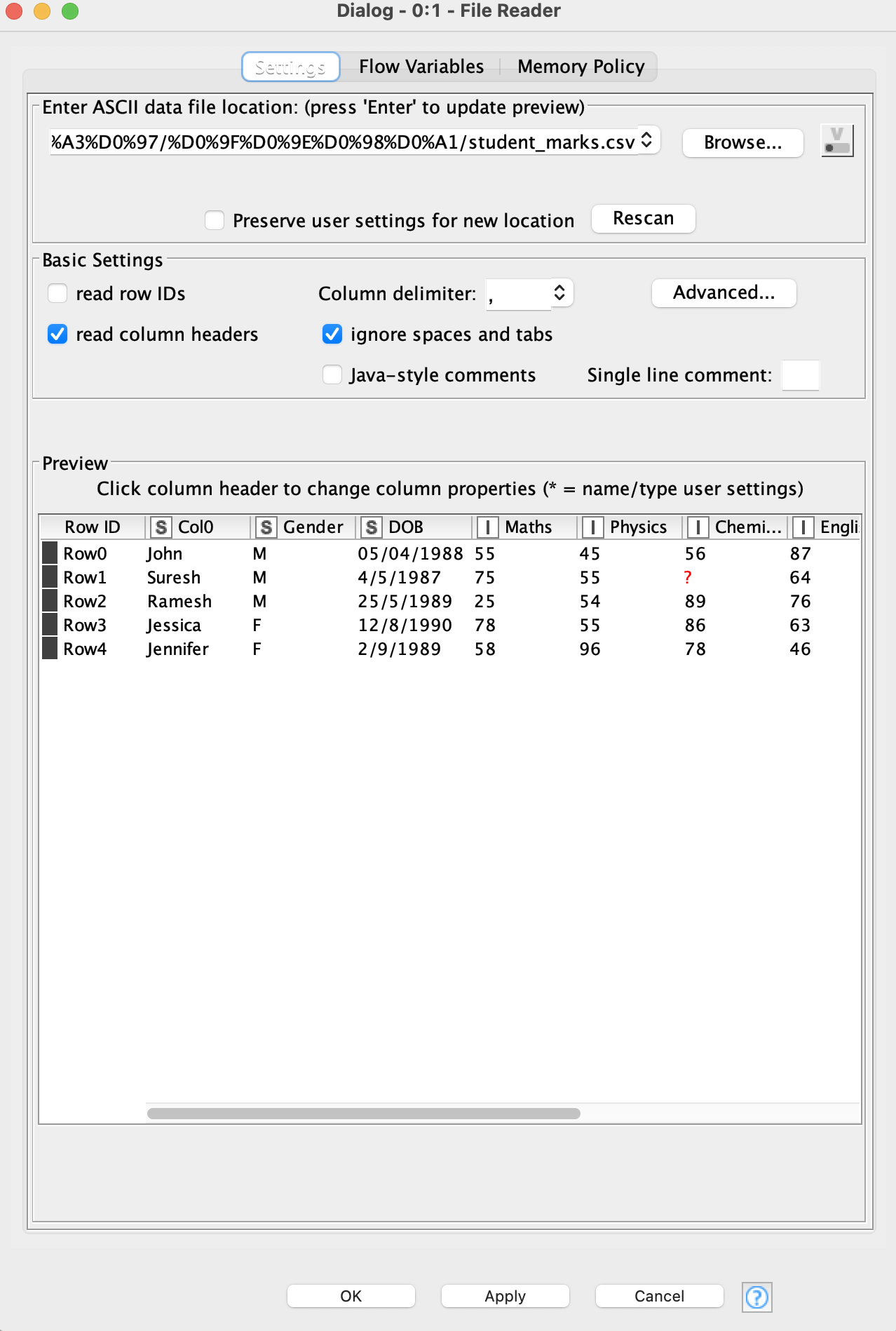


рисунок 7 - набор исходных данных для file reader.

Далее, настроим Decision tree learner и в качестве атрибута классификации выберем Col0 (рис. 8)

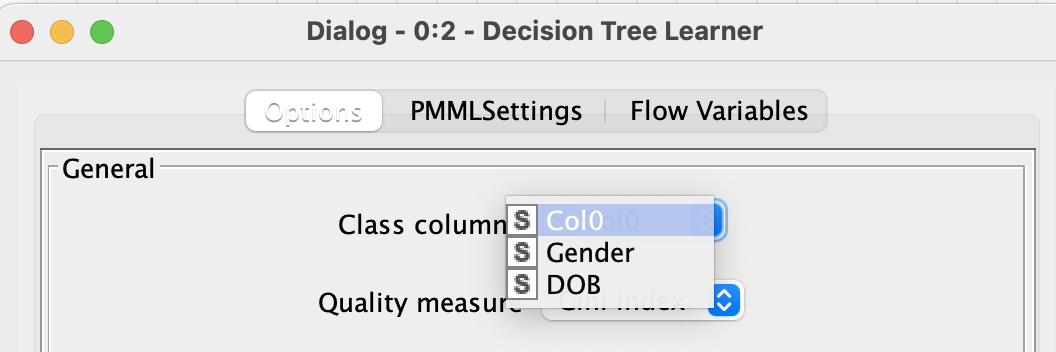


Рисунок 8 – выбор признака класса

Выполним поток работ и просмотрим результат, изображенный в виде дерева (рис. 9)

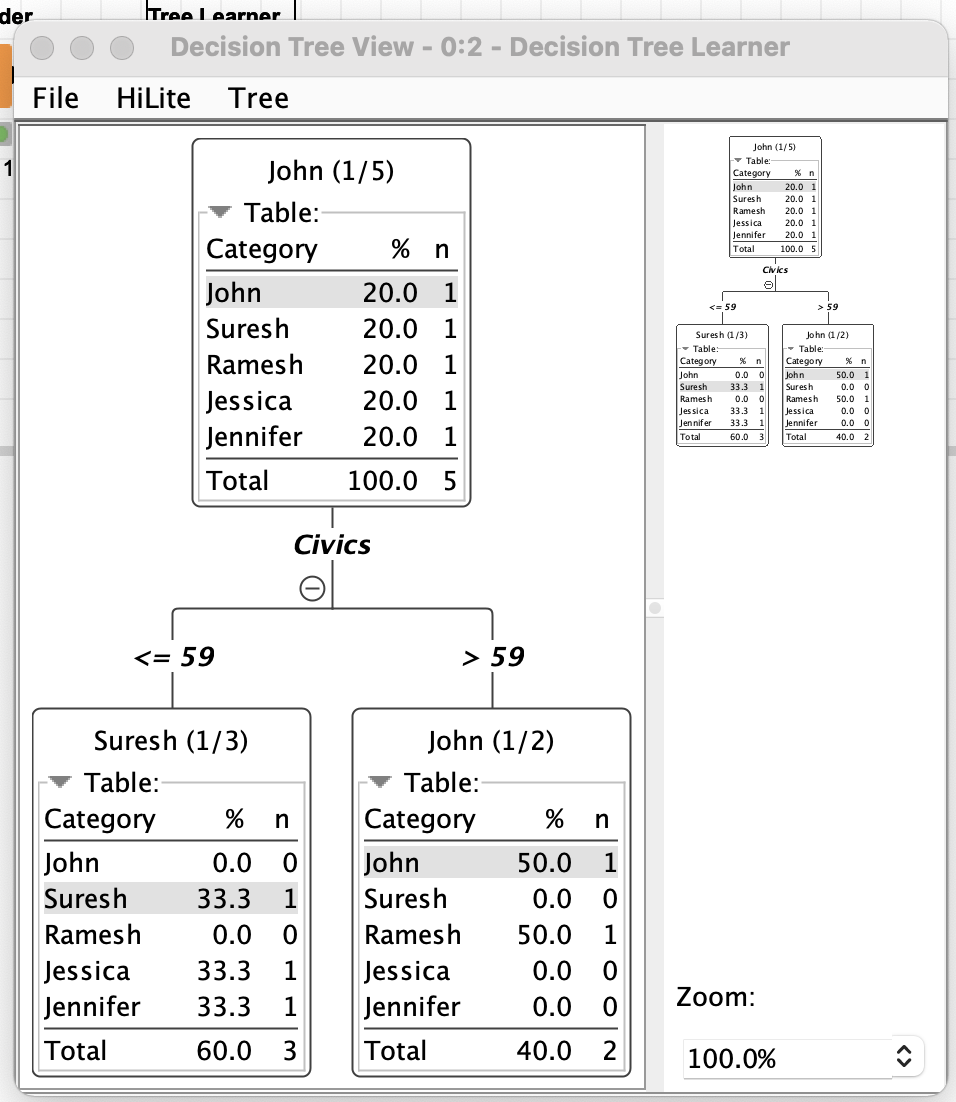


Рисунок 9 - просмотр результатов дерева решений

Перейдем к решению факультативного задания, заключающегося в настройке цвета для выбранного поля в Color Manager. Первым шагом добавим в поток работа Color manager, затем настроим его таким образом, чтобы col0 соответствовал определенный набор цветов (рис. 10)

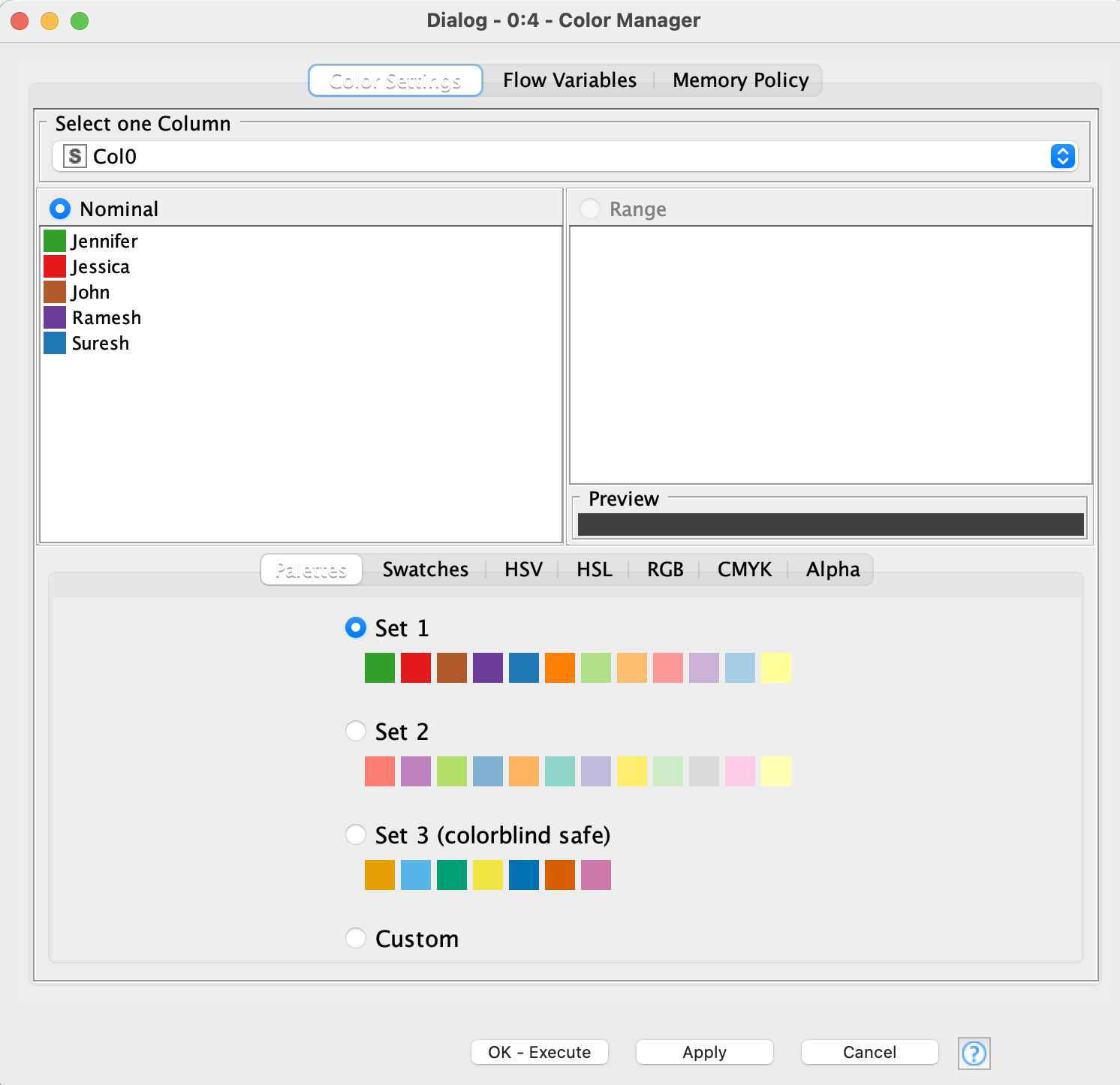


Рисунок 10 - настройка цвета для col0

Затем просмотрим результат Decision Tree Learner в двух вариантах после настроек цвета (рис. 11 и рис. 12)

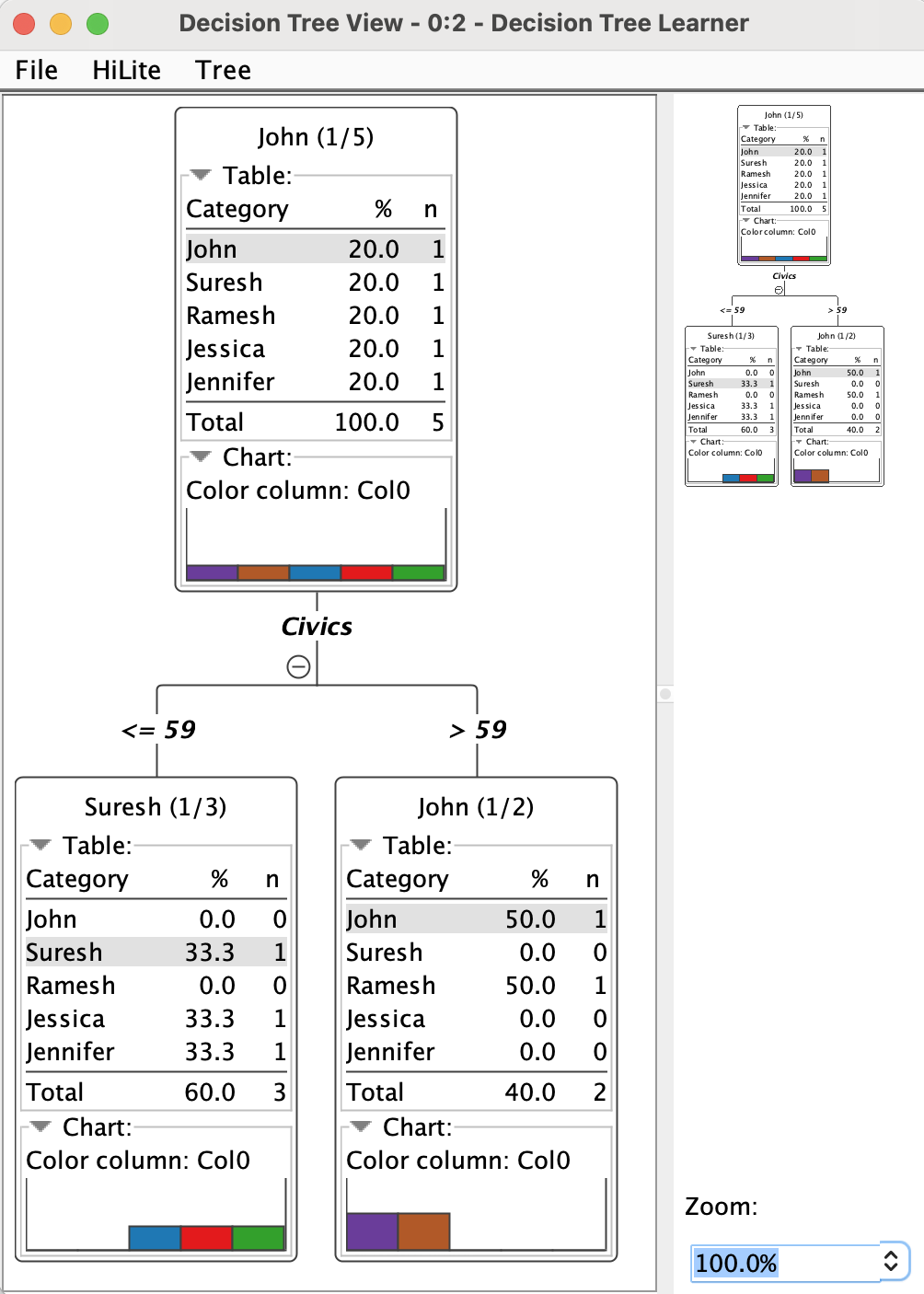


рисунок 11 - Просмотр View Decision Tree Learner

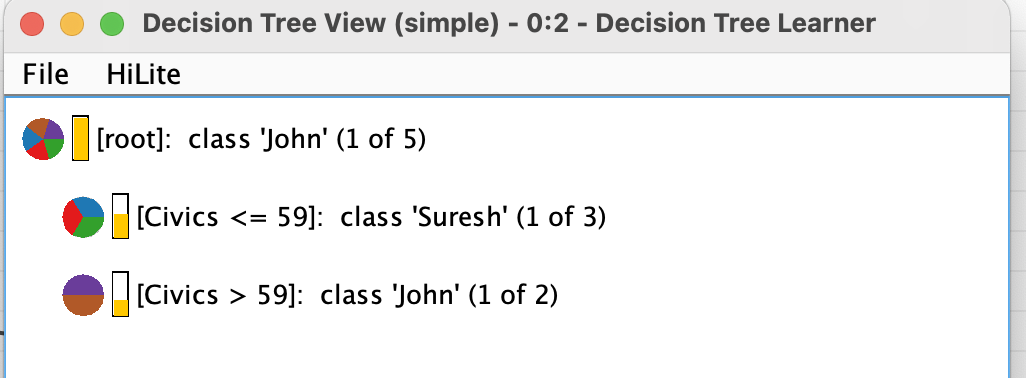


Рисунок 12 - просмотр Simple View Decision Tree Learner

**Кластеризация**

*Задача*

Построение потока работ, выполняющего решение задачи кластеризации. Данный поток должен выполнять следующую последовательность действий: загрузить данные из текстового файла, построить дерево решений, вывести результаты.

*Выполнение*

Был создан поток работы, изображенный на рисунке 13.

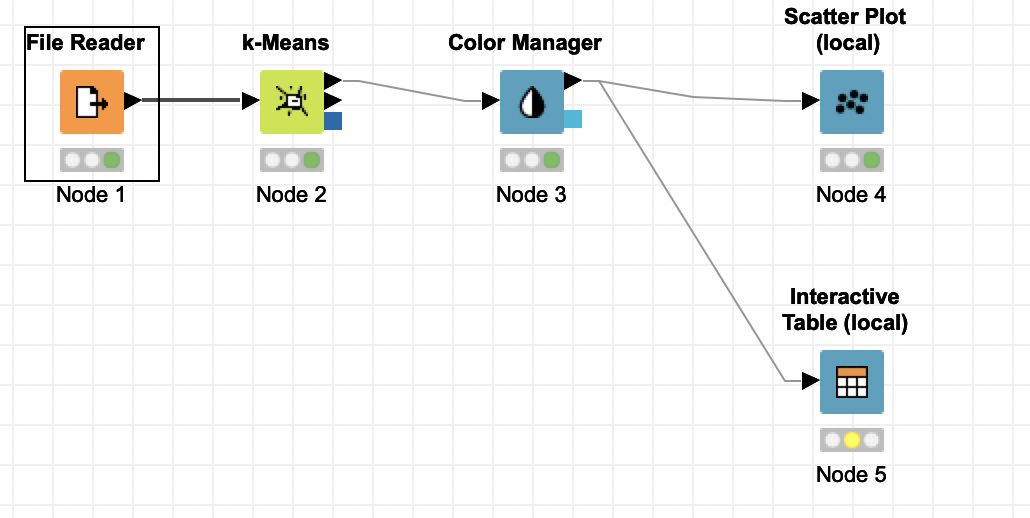


Рисунок 13 – искомый поток работ

Настроим узел k-Means как показано на рисунке 14

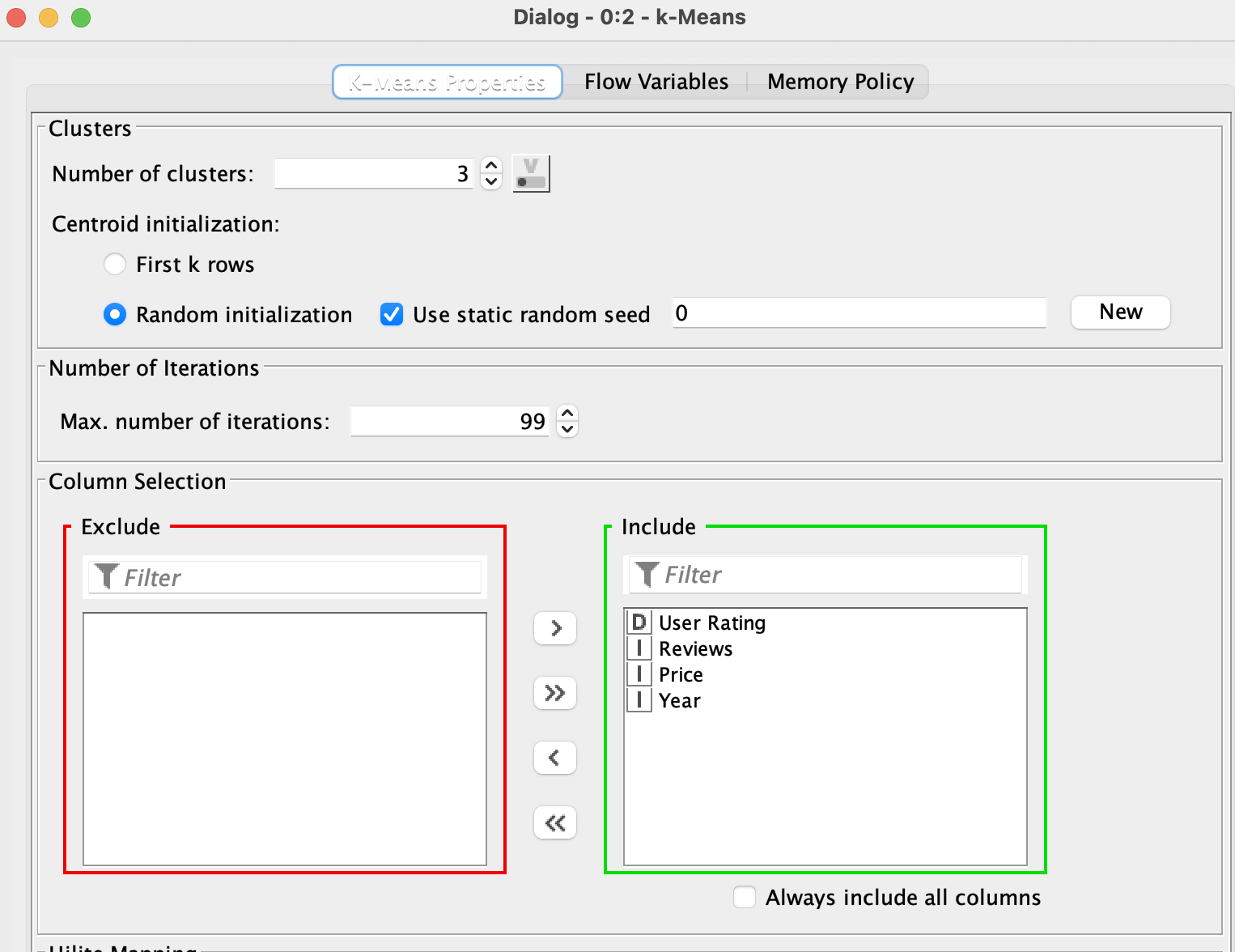
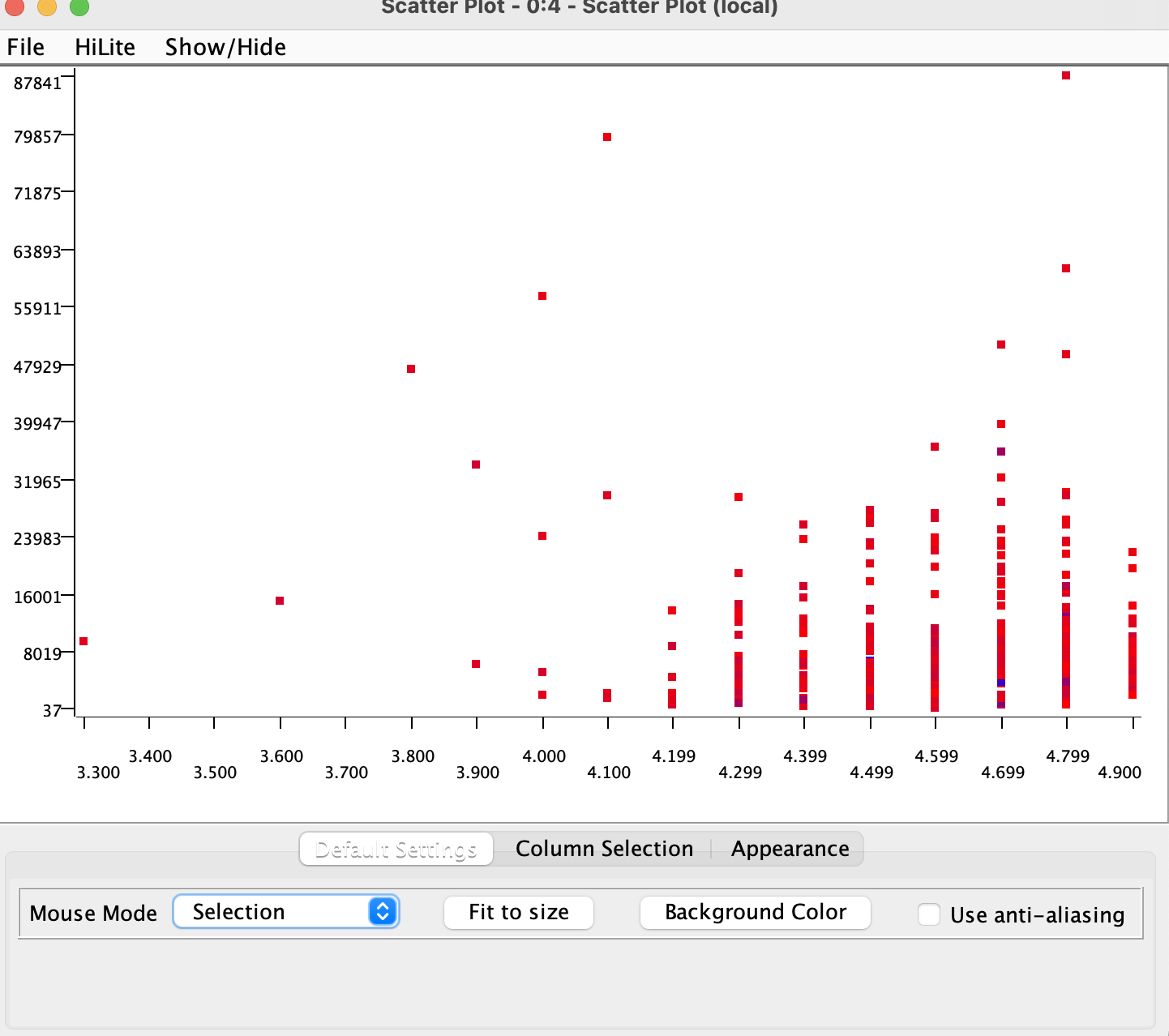


Рисунок 14 – настройка k-Means

На рисунке 15 представлена получившаяся диаграмма кластеров, показанная в зависимости от Price, где в каждой строке показана зависимость от кластера с помощью цвета.

Рисунок 15 - полученная диаграммы кластеров для Price

Выберем другой цвет и другое поле Reviews. Таким образом, задаем зависимость новых кластеров от цвета, как показано на рисунке 16.

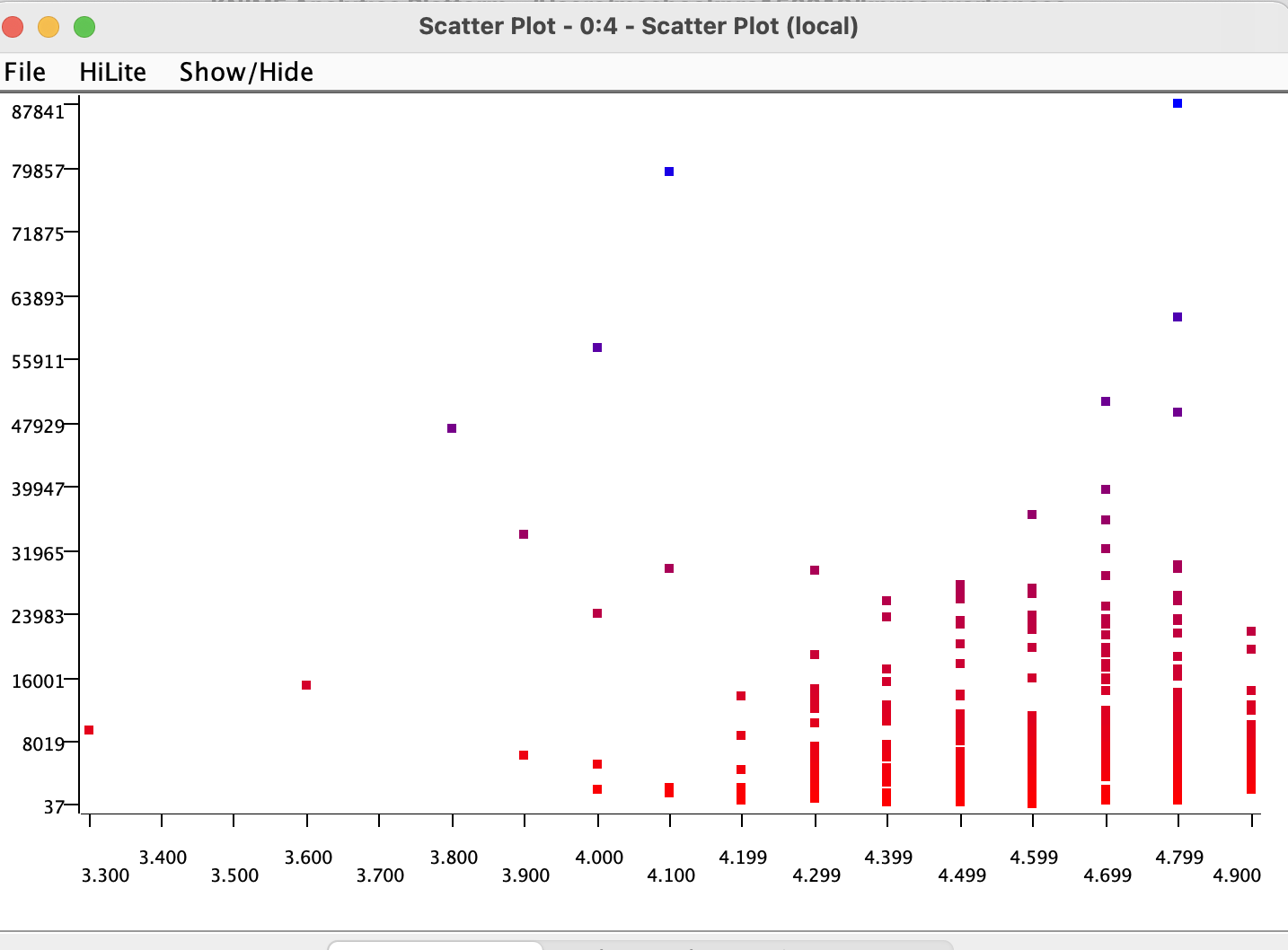


рисунок 16 - Диаграмма зависимости кластеров от Reviews

Также посмотрим на таблицу исходных данных, полученную в узле Interactive Table (рис. 17)

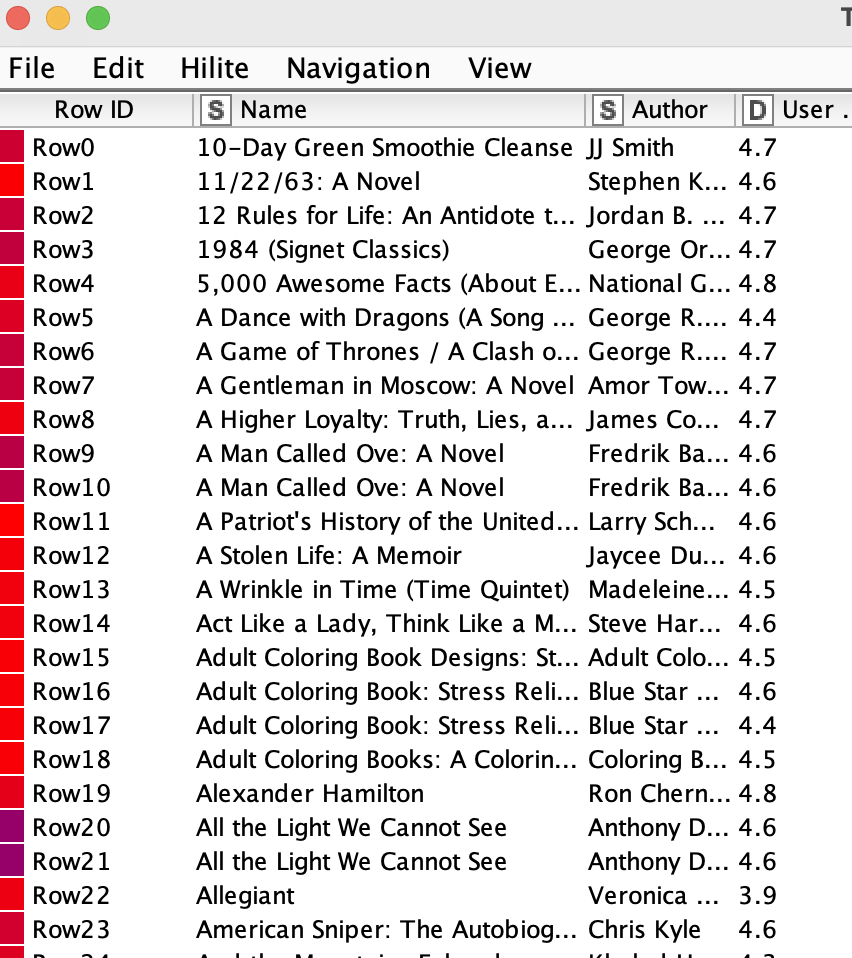


рисунок 17 - фрагмент таблицы исходных данных

Ссылка на GitHub: https://github.com/GooDiVer/sem7/tree/main/pois/lab1/knime

**Вывод**

В ходе выполнения данной лабораторной работы было произведено знакомство с программным обеспечением KNIME и некоторыми базовыми элементами, используемыми для работы с данными. Были выполнены поставленные задачи путем освоения механизмов кластеризации, деревьев решений и поиска шаблонов программного обеспечения KNIME.

**Список источников**

1. Обзор KNIME Analytics Platform – open source системы для анализа данных [Электронный ресурс]. – Режим доступа: https://habr.com/ru/post/320500/ . – (Дата обращения: 05.10.2020).
2. Association Rule Learner [Электронный ресурс]. – Режим доступа: https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.mine.subgroupminer.SubgroupMinerFactory2 . – (Дата обращения: 05.10.2020).
3. Create Collection Column [Электронный ресурс]. – Режим доступа: https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.collection.list.create2.CollectionCreate2NodeFactory . – (Дата обращения: 05.10.2020).
4. Knime [Электронный ресурс]. – Режим доступа: https://www.knime.com/ . – (Дата обращения: 05 .10.2020).