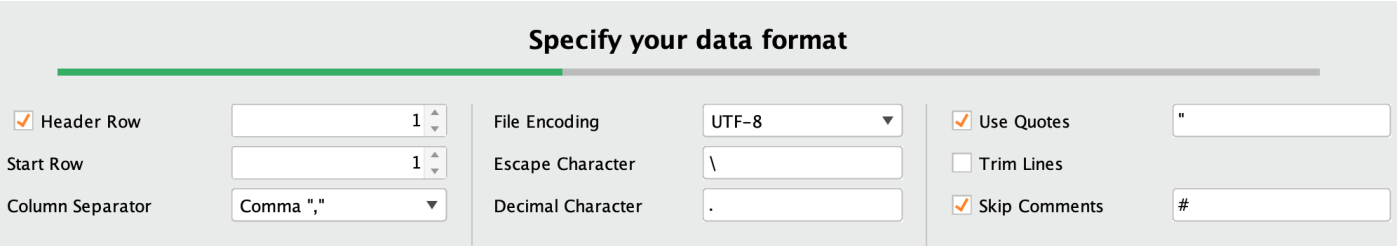
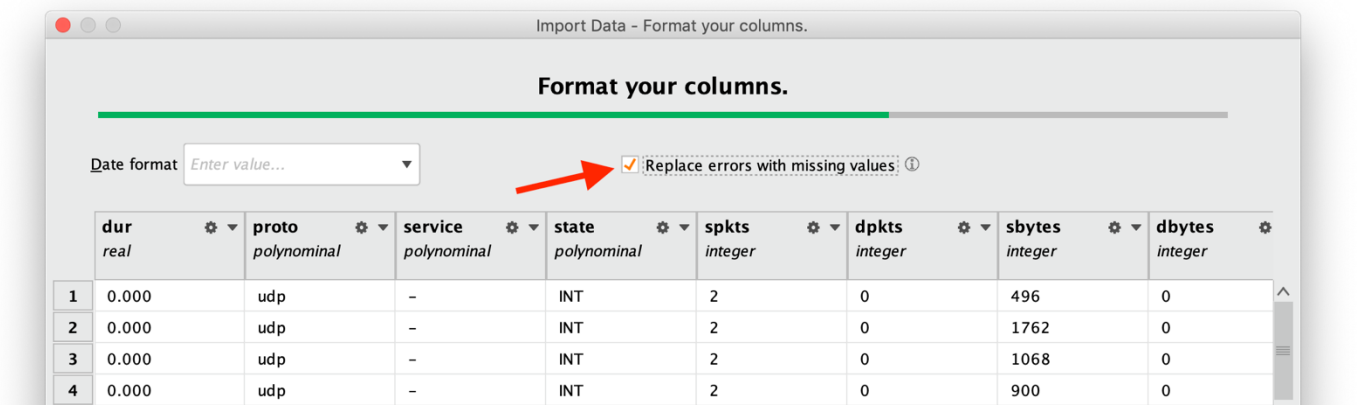
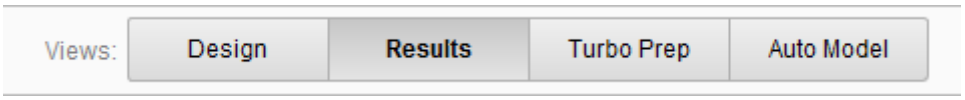
1. Импортируем большой датасет Data в RapidMiner Studio (File -> Import Data -> My Computer).
2. Формат данных оставляем по умолчанию.



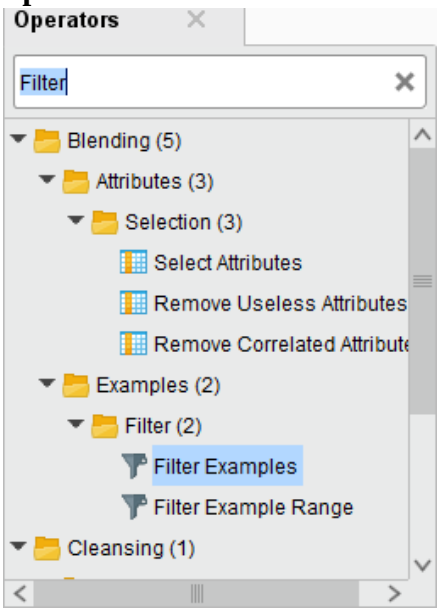
3. На следующем шаге, не забываем поставить галочку:



4. Перетаскиваем загруженный датасет в рабочую область, перейдя в режим Design



5. Добавляем оператор **Filter Examples**



6. Добавляем фильтр полю *is_anomally* – “is not missing” (рисунок 1). Необходимо для того, чтобы на всякий случай исключить отсутствующие значения.

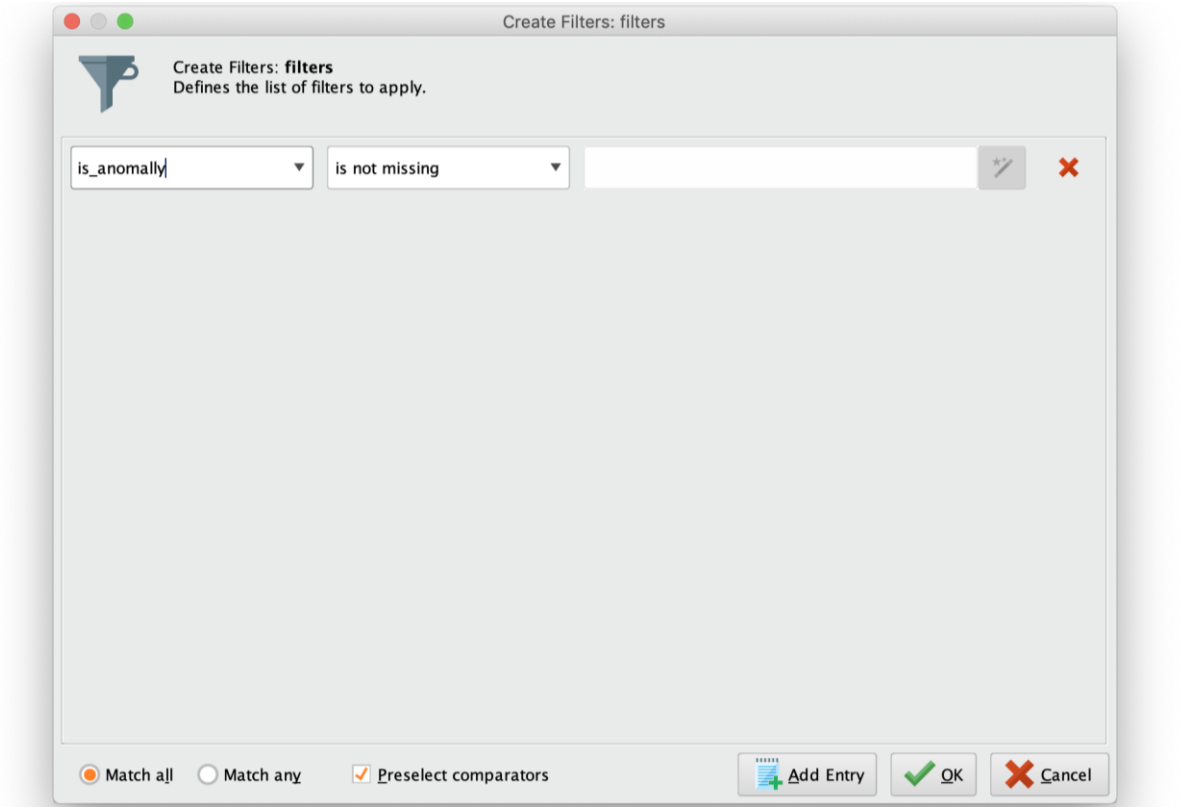
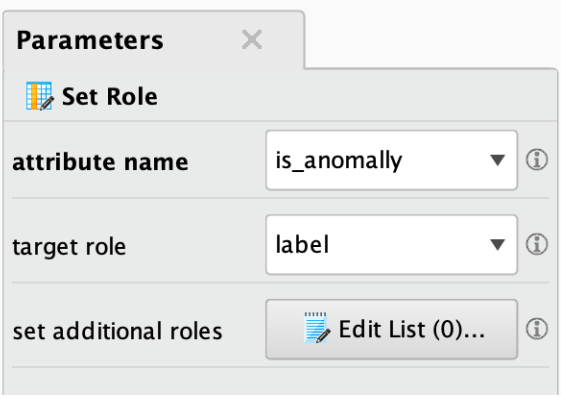


Рисунок 1

7. Далее с помощью оператора **Set Role** выбираем нужное поля для предсказания. В нашем случае – *is_anomally*, target role помечаем как label.



8. Добавляем элемент **Nominal to Numerical** (нужен для избежания ошибок на этапе предсказания). Соединяем линии так, как указано на рисунке 2.

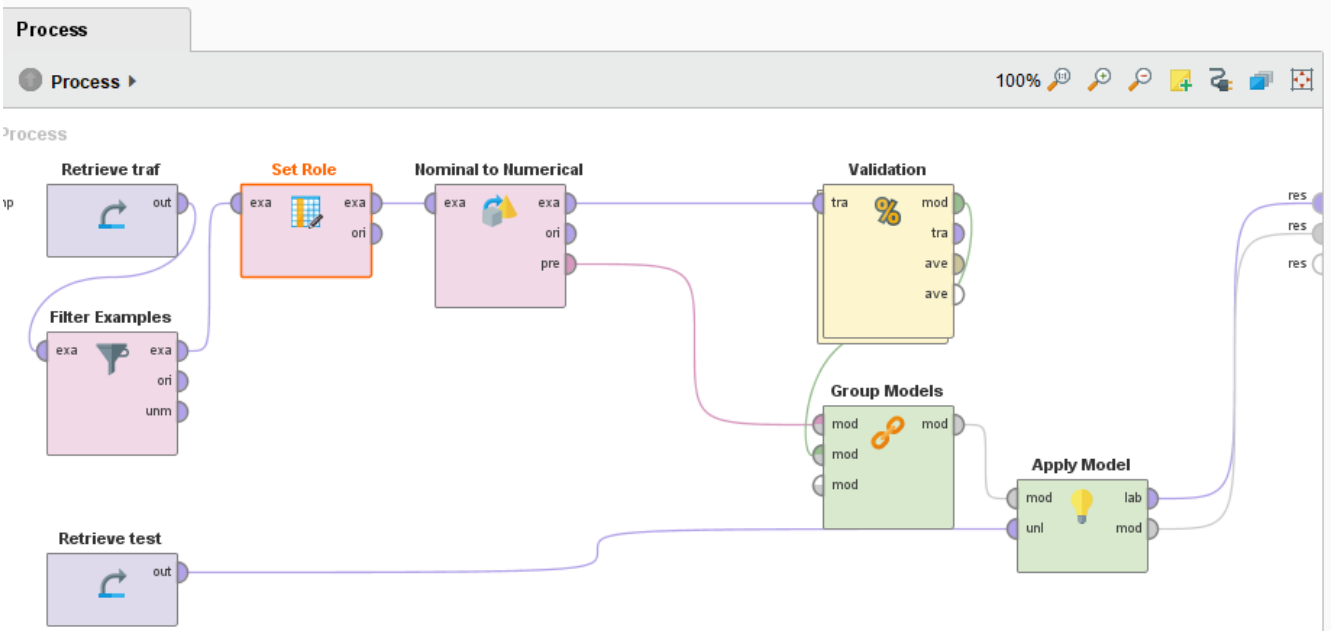
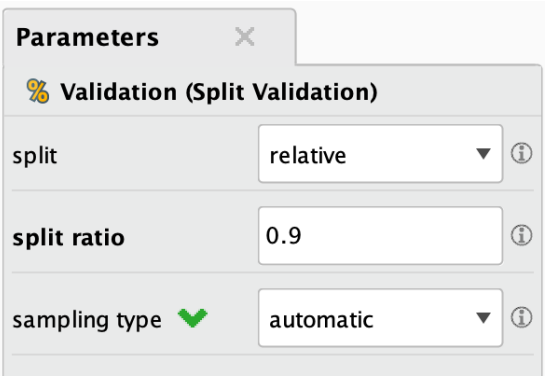


Рисунок 2

9. Добавляем оператор **Split Validation**, внутри него организовываем разбивку датасета на Training и Testing в пропорциях 90% и 10% соответственно.



10. Открываем оператор **Validation** двойным кликом и вставляем операторы **Decision Tree, Apply Model, Performance,** соединяем как указано на рисунке 3.

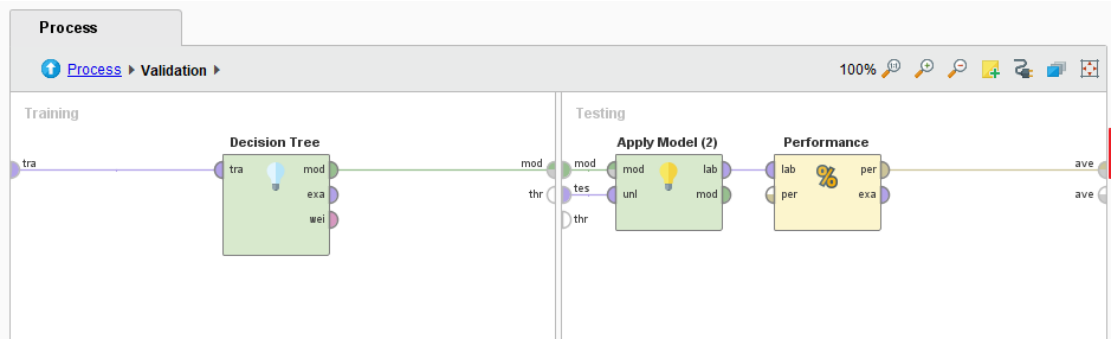


Рисунок 3

11. Выходим из **Split Validation**, далее в общую схему добавляем оператор **Group Models** (необходим для классификации новых данных)**,** также добавляем ещё один оператор **Apply Model**. Импортируем новые данные (датасет без указания, аномальный ли трафик) также с помощью Import Data, добавляем его в схему и соединяем линиями так, как это указано на рисунке 4, должна получится подобная схема.

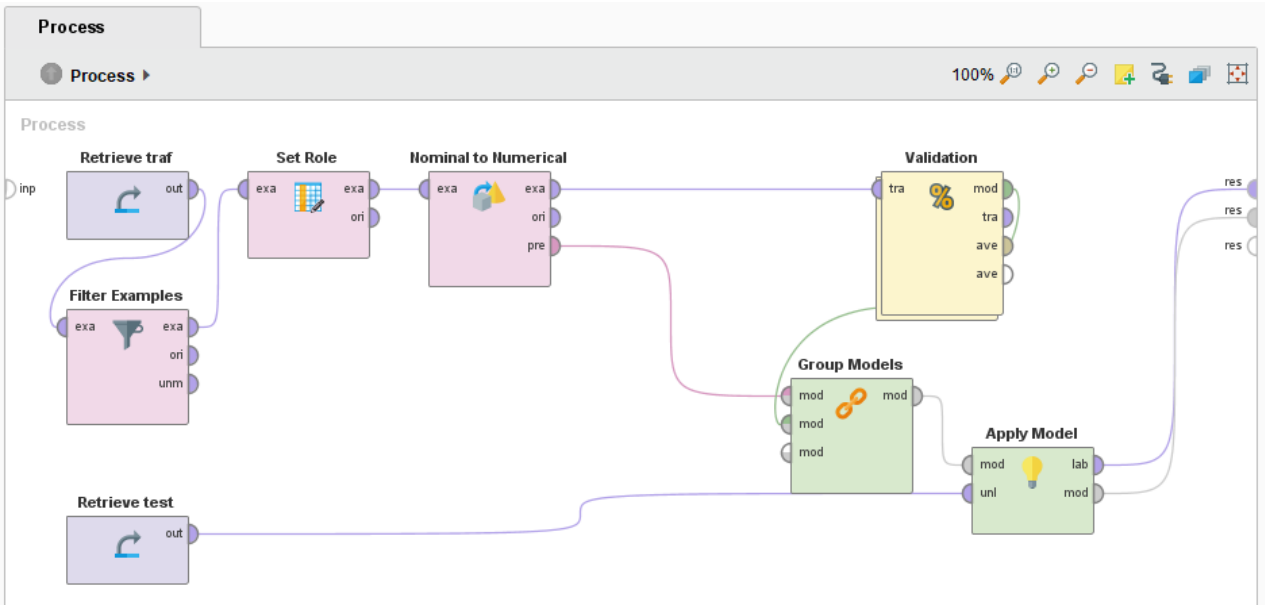


Рисунок 4

12. Далее нажимаем на Play (начать классификацию), открываем результаты и смотрим предсказание по трафику добавленным новым данным (рисунок 5).

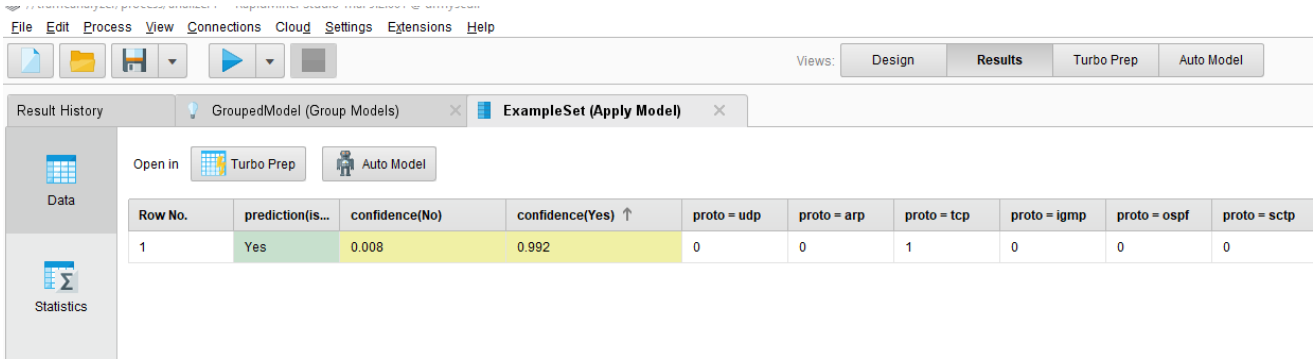


Рисунок 5