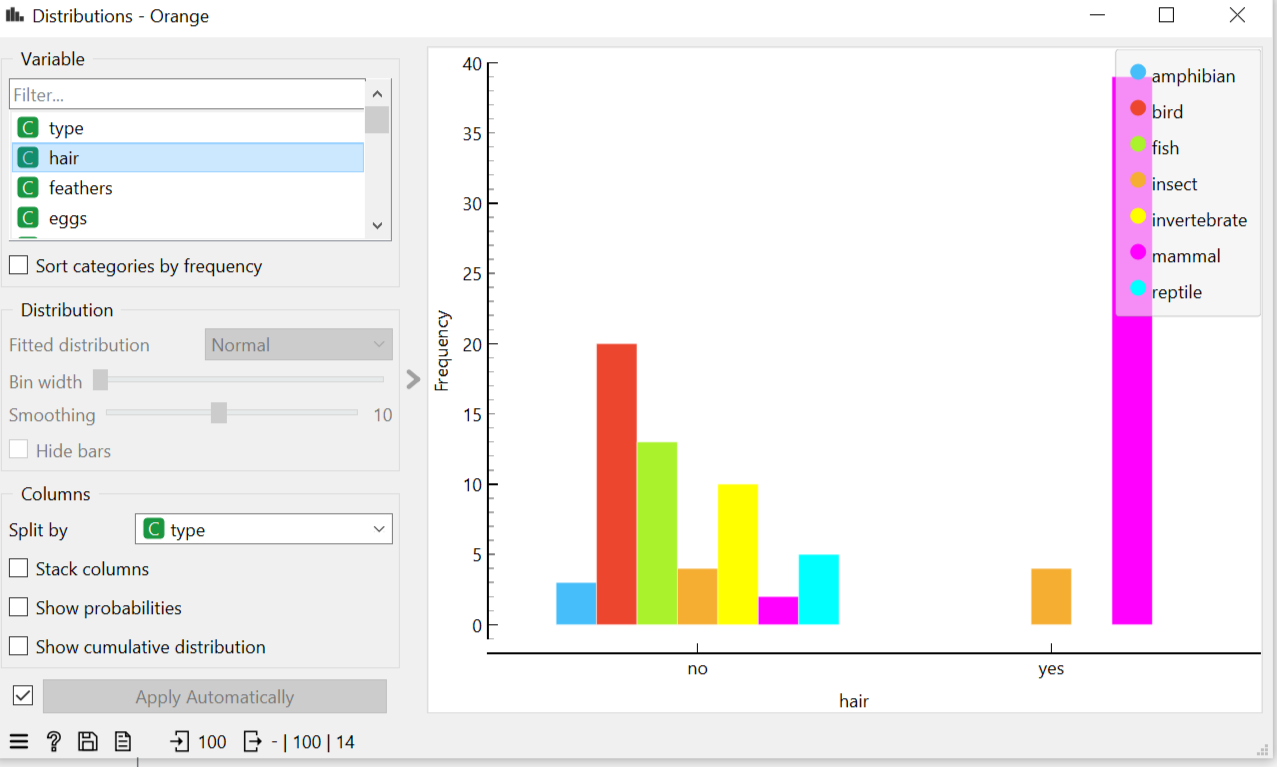
*Выполнила: Чудиновских Юлия*

В DataSets я выбрала **Зоопарк.**

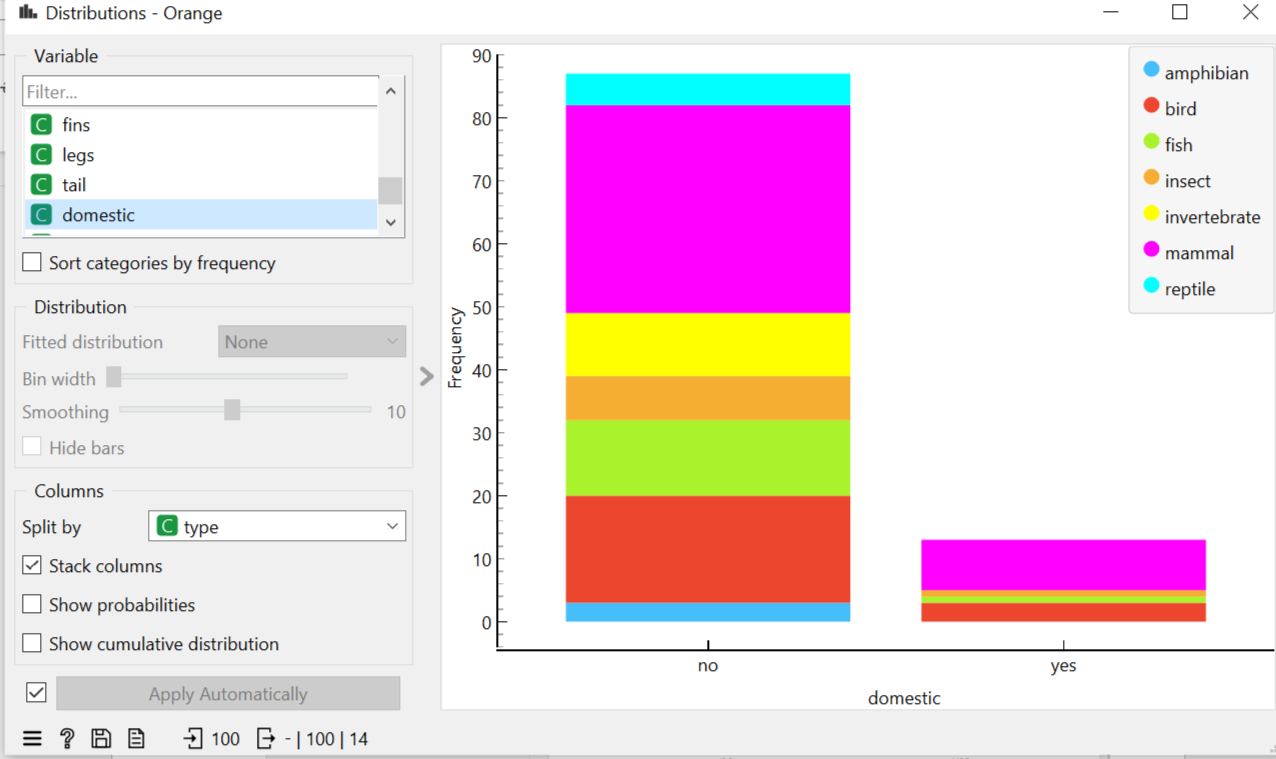
В ***feature statistics*** я увидела, что он имеет 17 параметров (атрибутов) (информацию о наличии шерсти, перьев и зубов, информацию о том, является ли животное водным или воздушным, и тому подобное. Животные имеют названия и подразделяются на семь категорий: амфибии, птицы, рыбы, насекомые, беспозвоночные, млекопитающие и рептилии.)



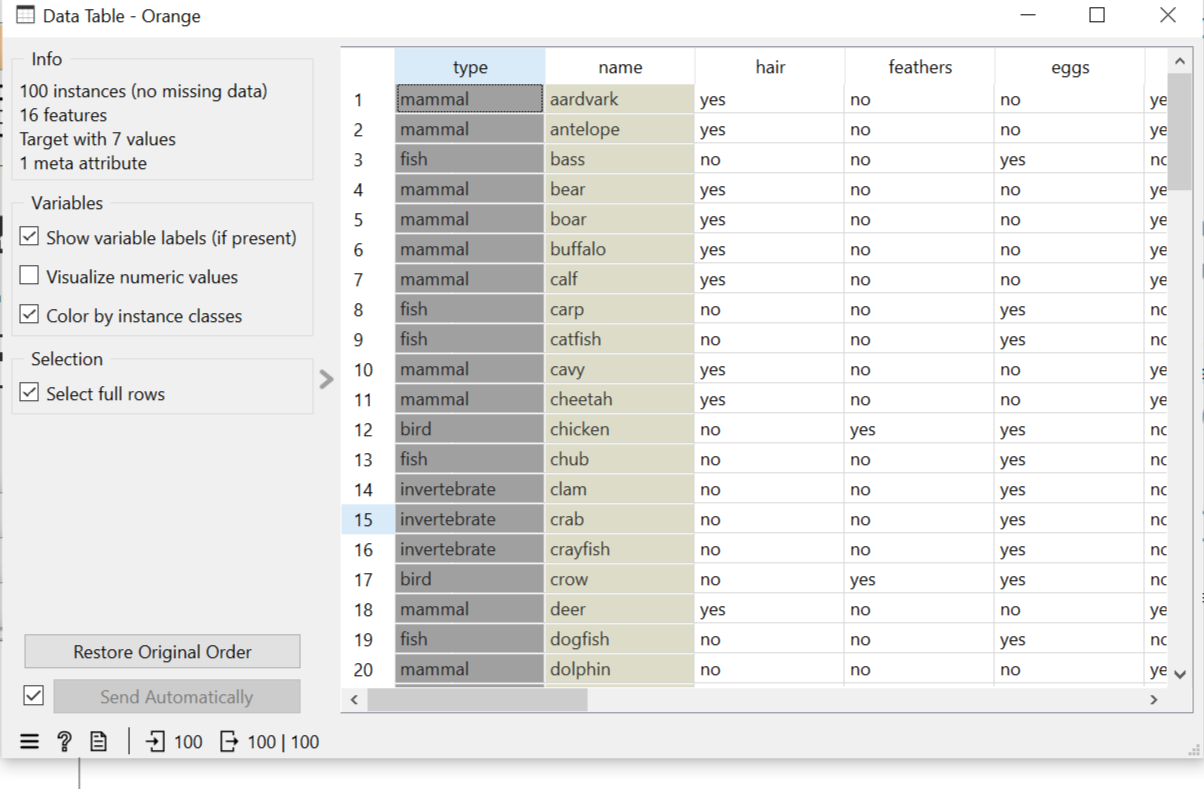
В ***distributions*** можно более подробно проанализировать данные. По столбчатым диаграммам увидеть, например, что млекопитающие имеют больше всего волос



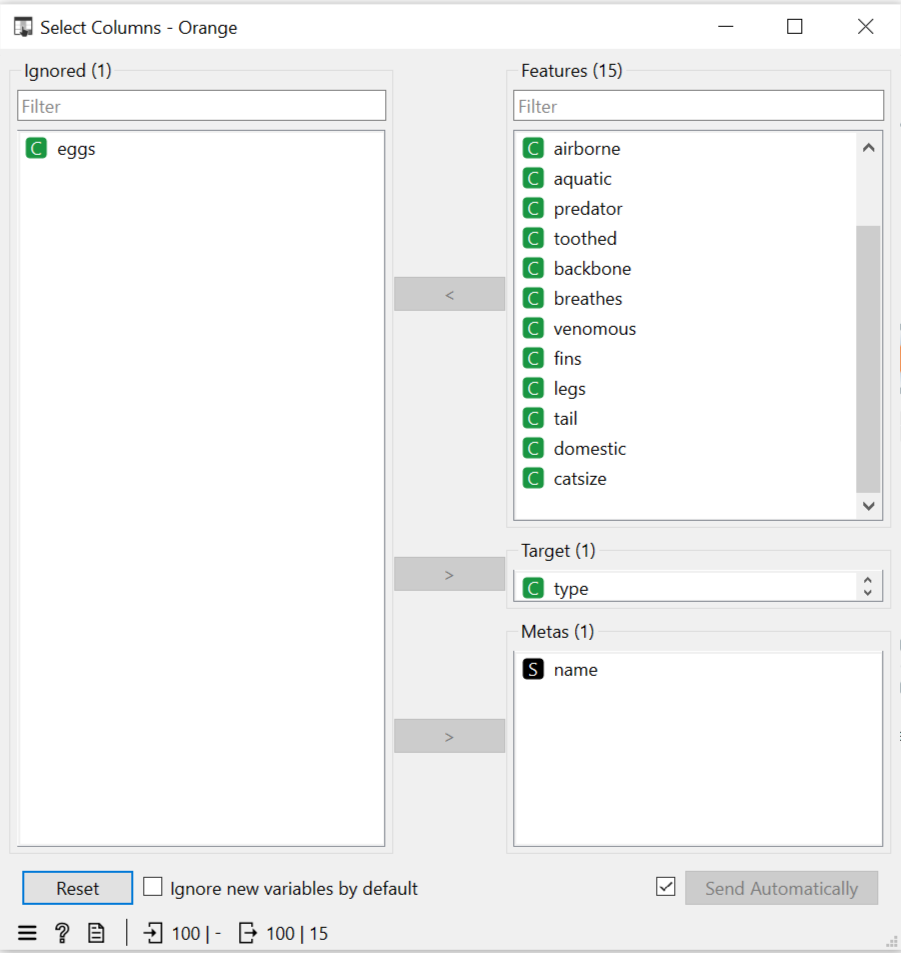
Посмотреть что есть в наборе домашние – это млекопитающие, насекомые, рыбы, птицы, но в этом же наборе есть большинство животных не домашние, те же птицы , рыбы, рептилии и т.д.



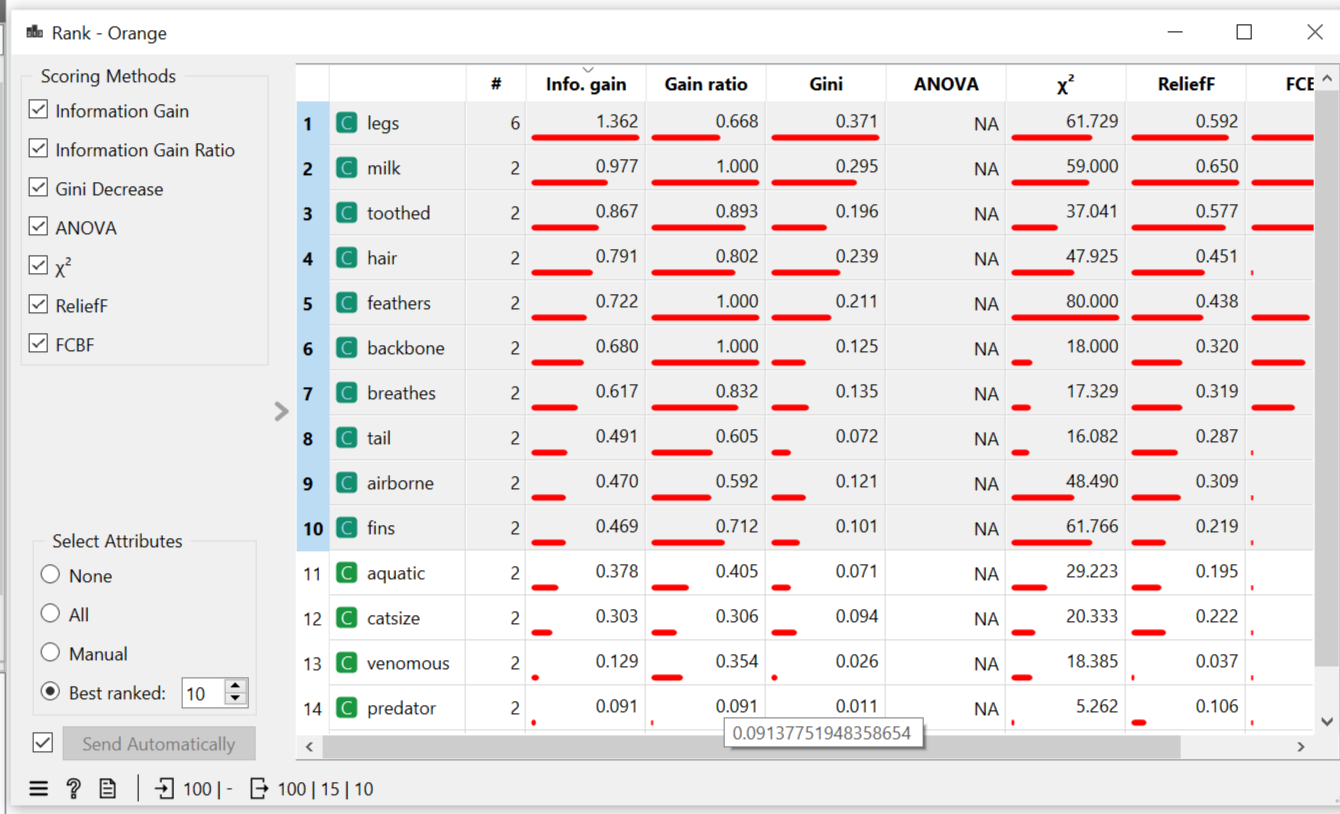
Посторив ***data table*** я наконец-то более подробно посмотрела на данные увидела, что за животные представлены в данном наборе, к какому типу они относятся, хищники они, ядовитые, есть ли шерсть и т.д.



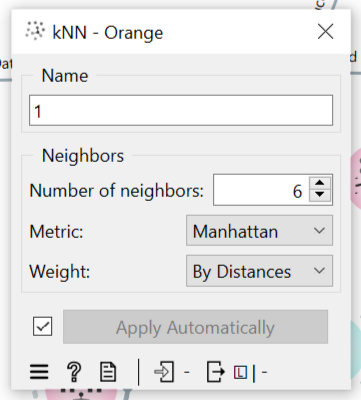
В **Select Columns** я попробовала убрать 1 категорию из таблицы.



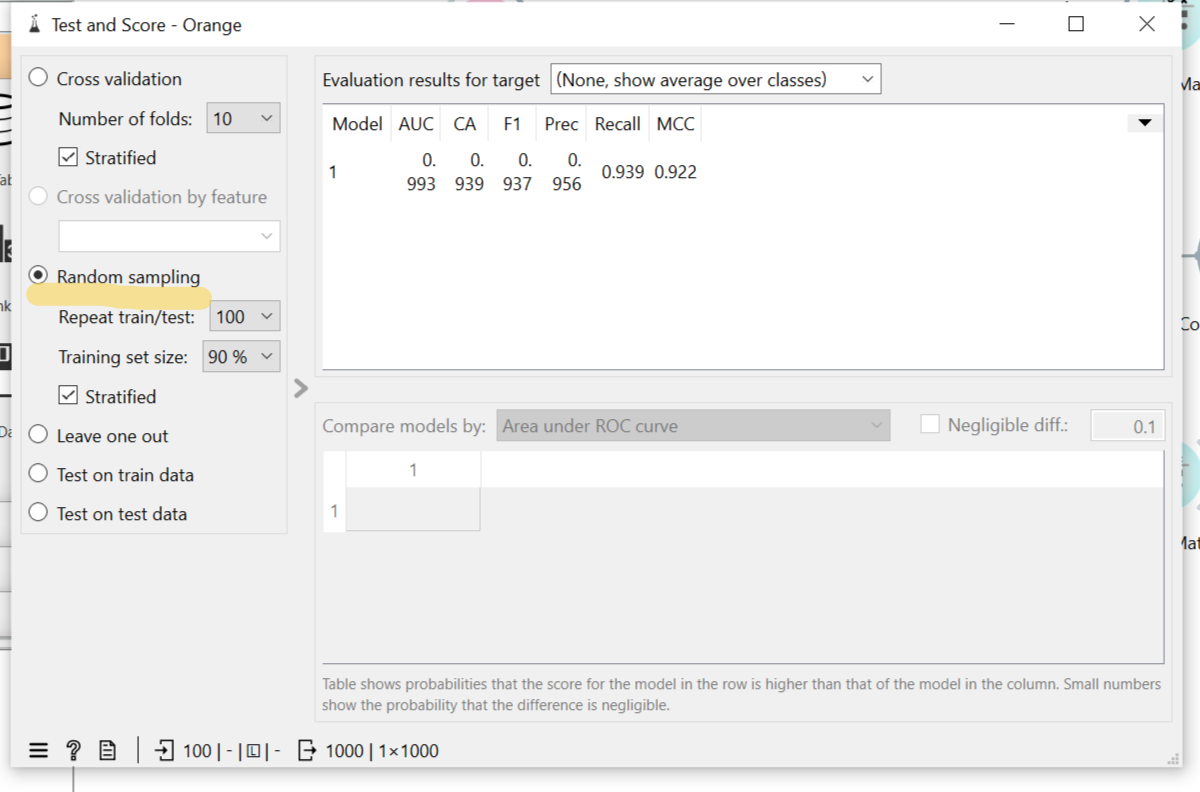
В **Rank** я увидела, что наибольшее влияние оказывает ноги, молоко, зубы.



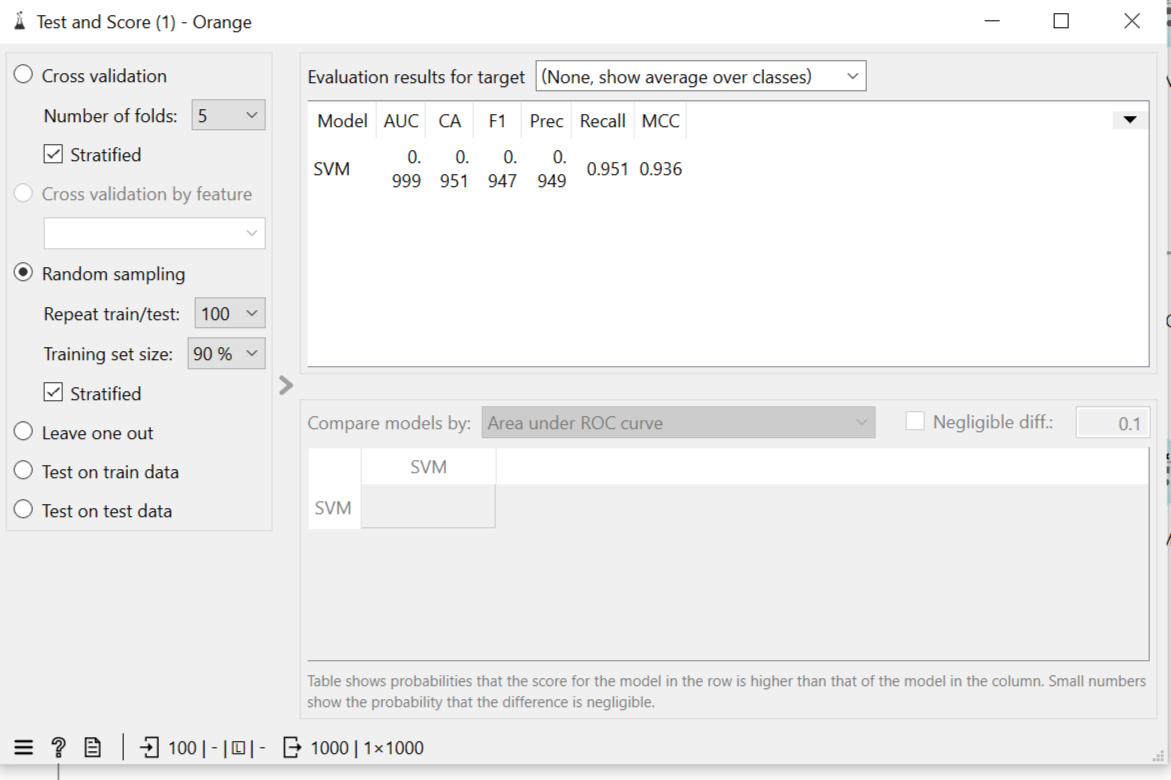
Запускаем обучение по модели **kNN** (меняла параметры, чтобы повысить точность)



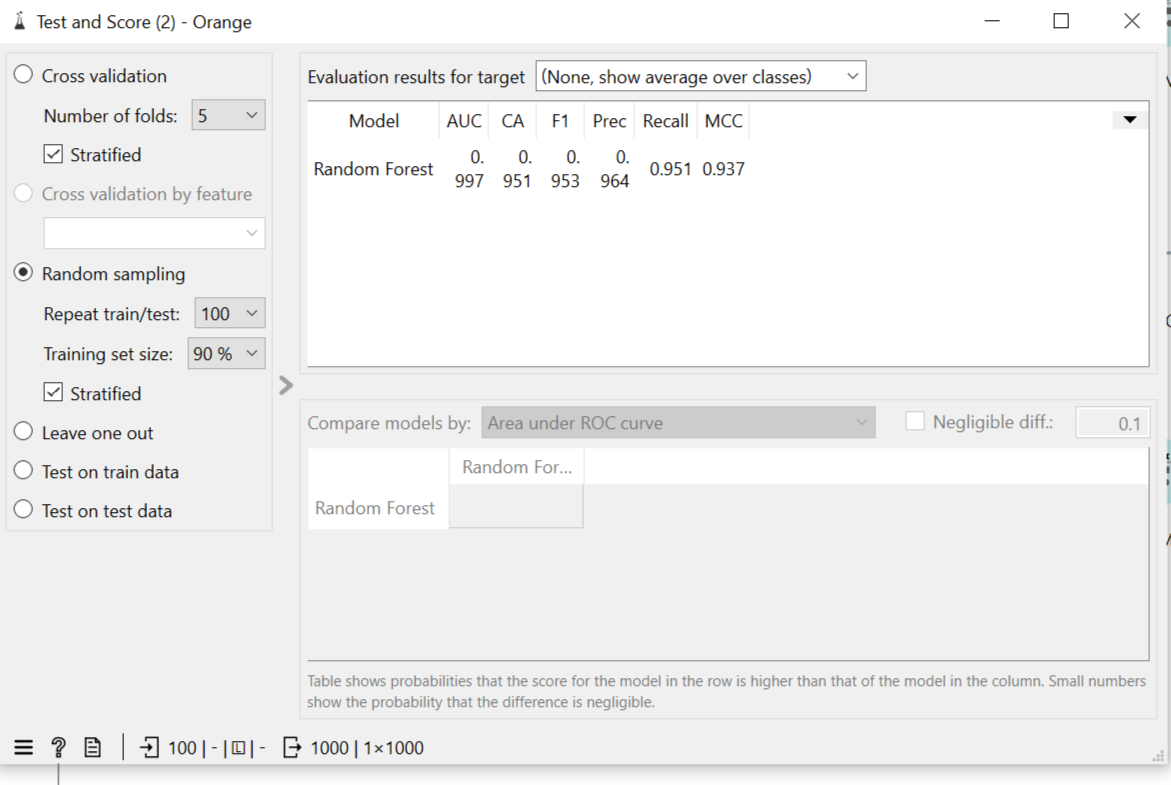
Обучение модели через **Random sampling. Точность 99,3%** (Циклов обучений 100, данных используем 90%, случайным образом делим дата сет на части)



Запускаем обучение по **модели SVM** (не меняла параметры, переключила только на random sampling)**. Точность 99,9%** (

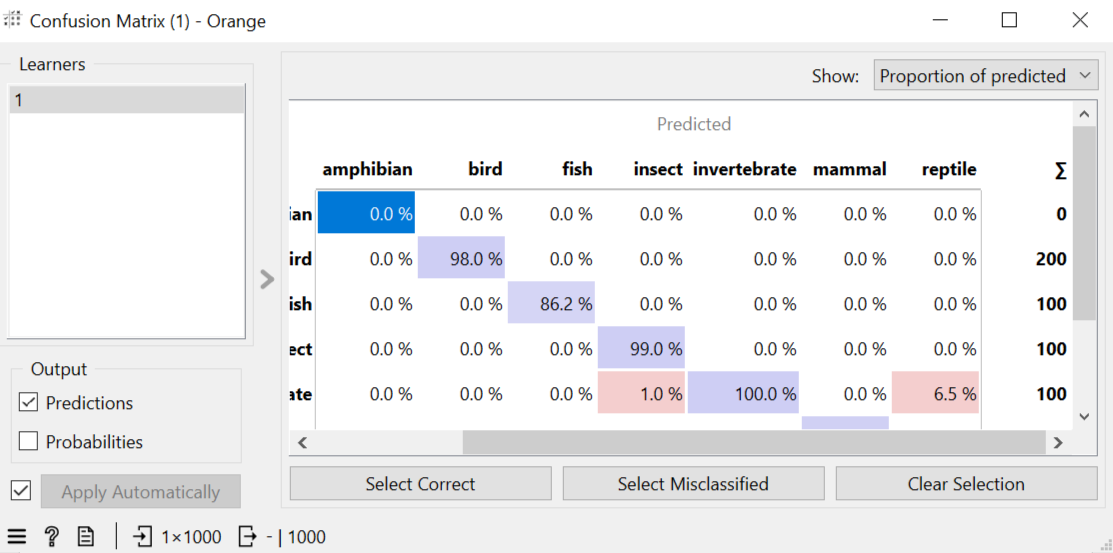
****

Запускаем обучение по модели **Random Forest** (не меняла параметры, переключила только на random sampling)**. Точность 99,7%**

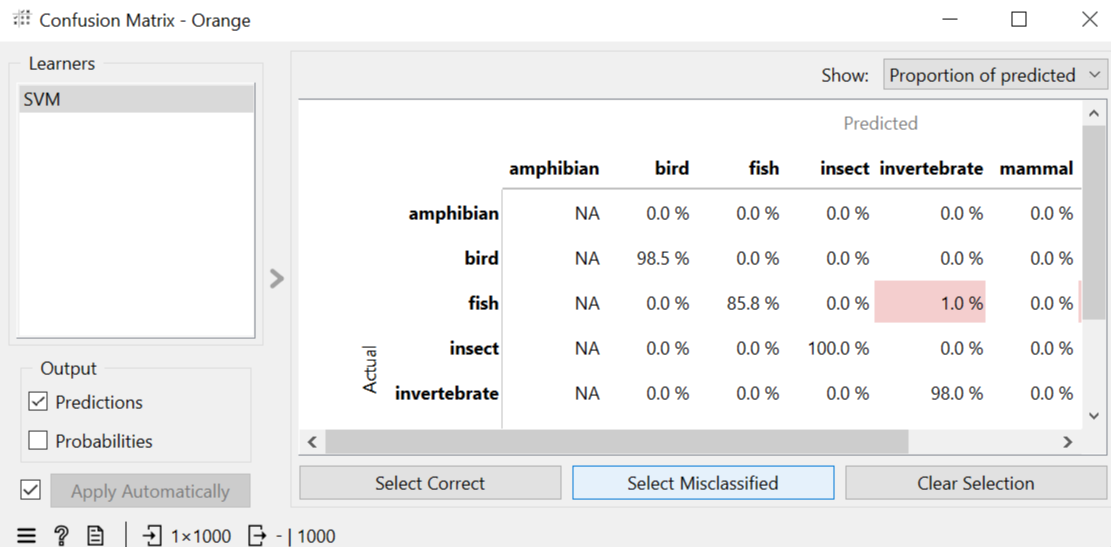


**Построила матрицы для каждой модели.**

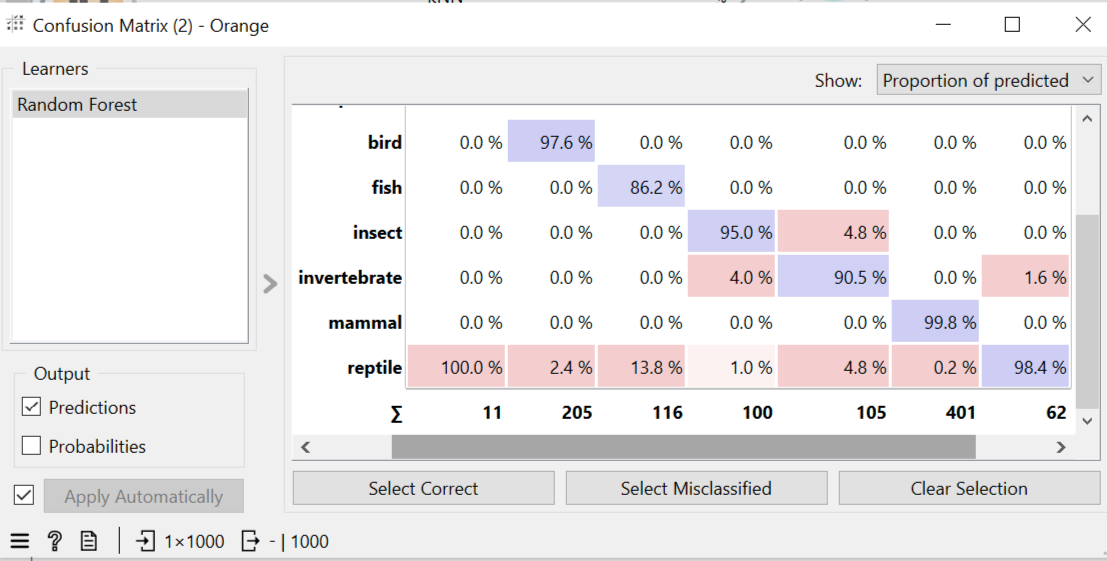
*KNN*



*SVM*



*Random Forest*



У этой модели как будто бы меньше ошибок, чем у предыдущих ошибок, хотя точность не такая высокая по сравнению с имющимися.

