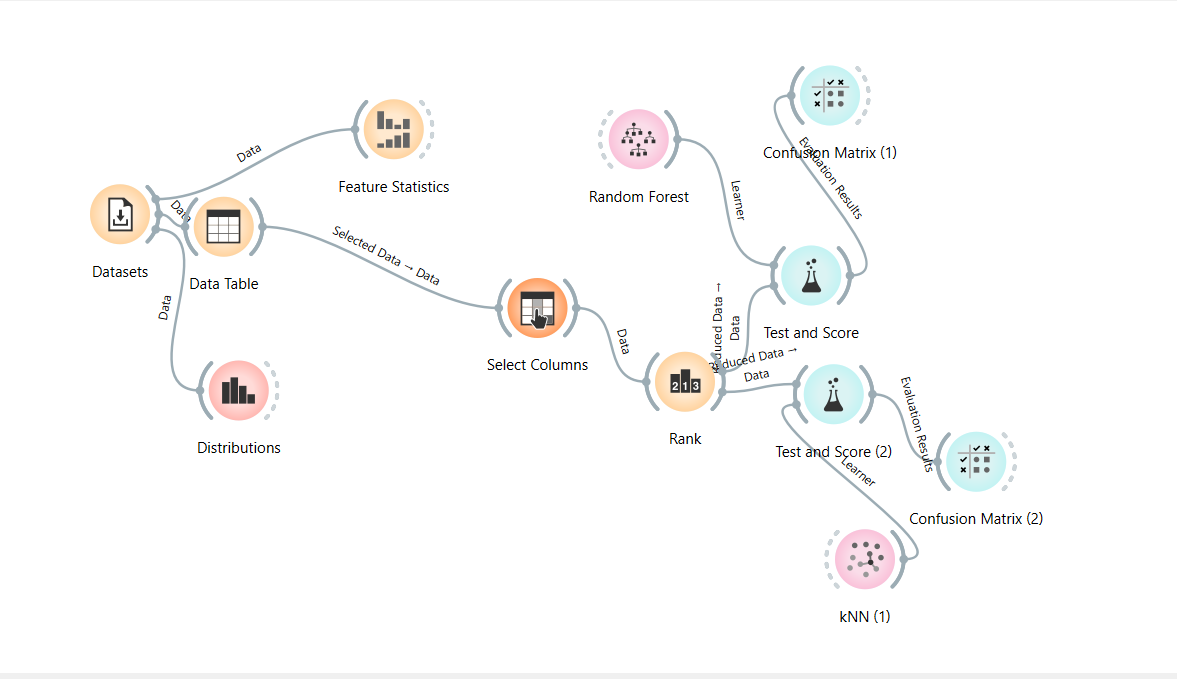
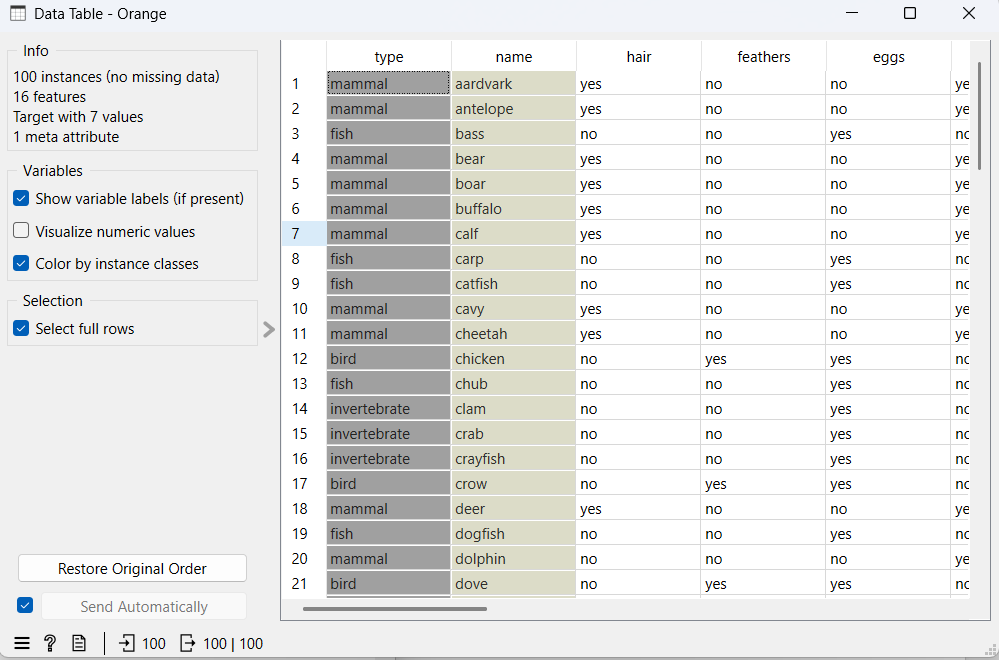
Шумкина Е.О.

Работа с данными датасета «Зоопарк».

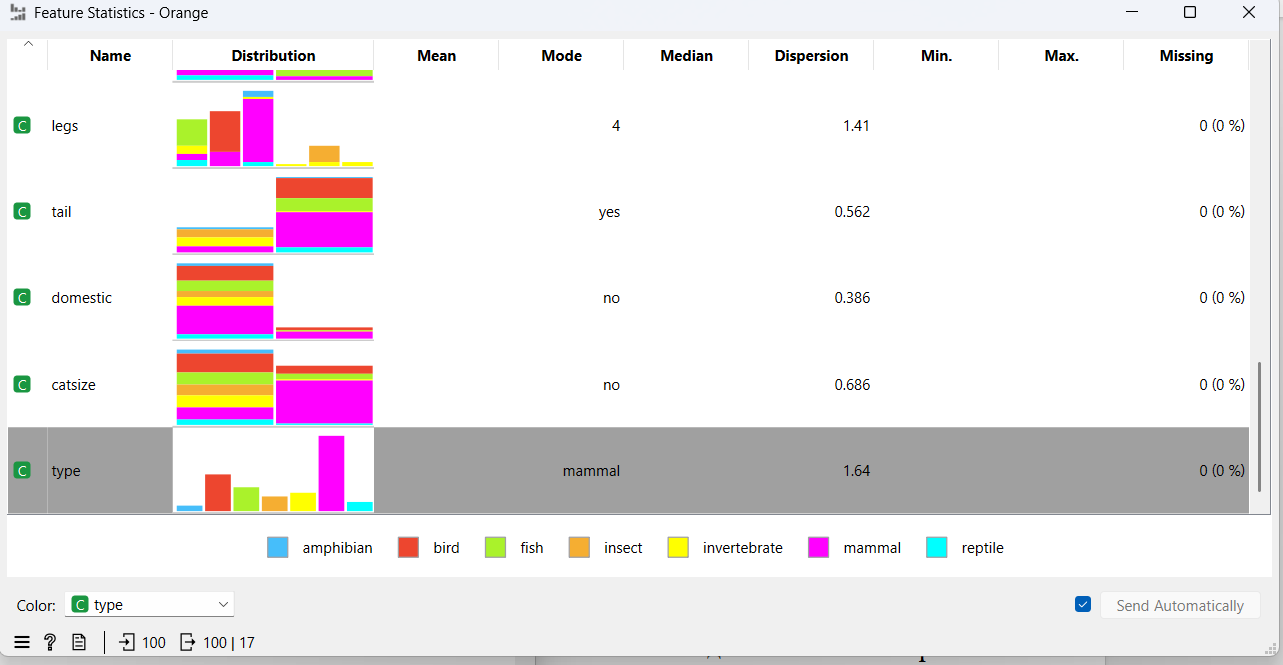


Примером датасета для изучения выбрала «Зоопарк».

Упорядочила данные в таблицу с помощью **Data Table.** Здесь о каждом животном можно узнать подробную информацию по каждому из параметров и категорий.

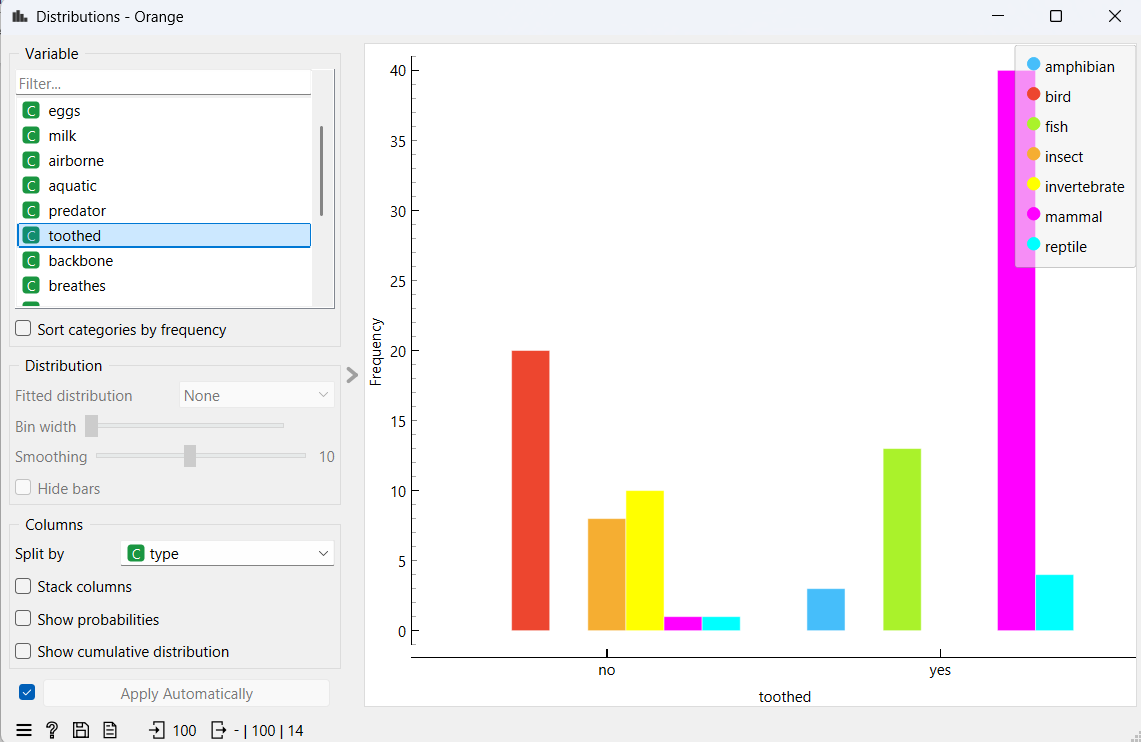


Посмотрела категории животных и данные о них через виджет **Feature Statistics.** Животные распределены по семи категориям (в соответствии с принадлежностью к биологическому виду - амфибии, птицы, рыбы, насекомые, беспозвоночные, рептилии, млекопитающие). Каждое животное изучается по семнадцати параметрам - имеет ли оно перья, несёт ли яйца, есть ли у него зубы, и т.д.

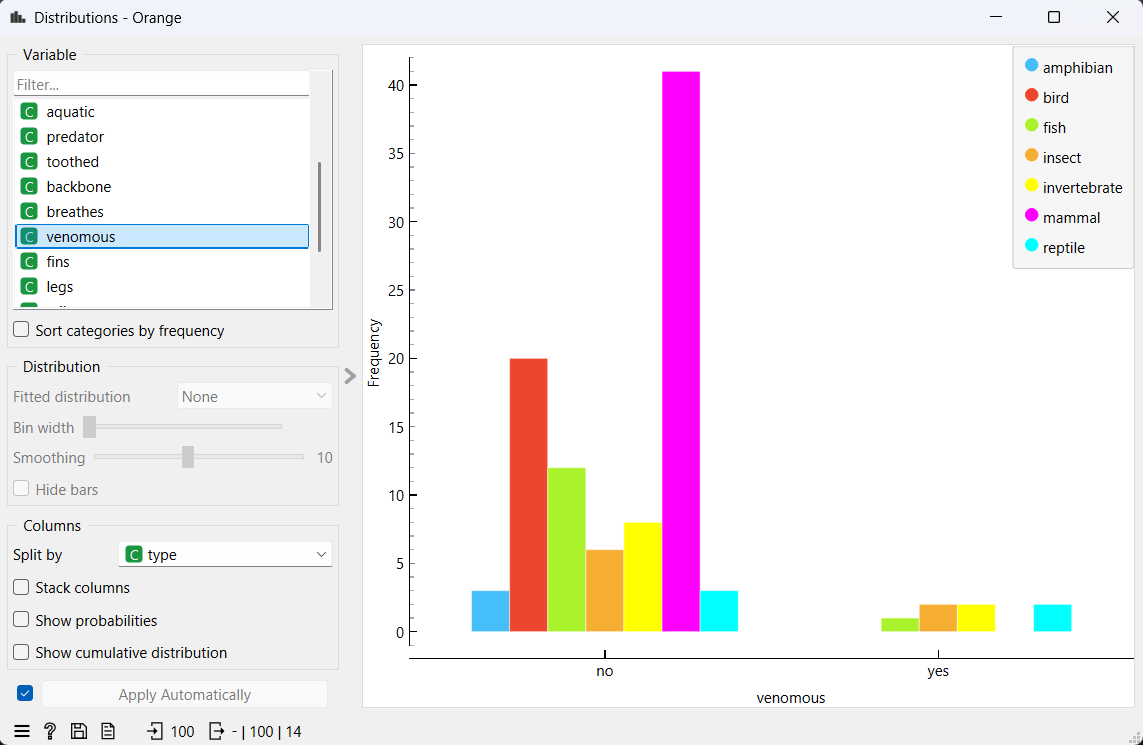


Каждый вид по параметрам можно более подробно изучить в distributions, ориентируясь на столбцы диаграммы.

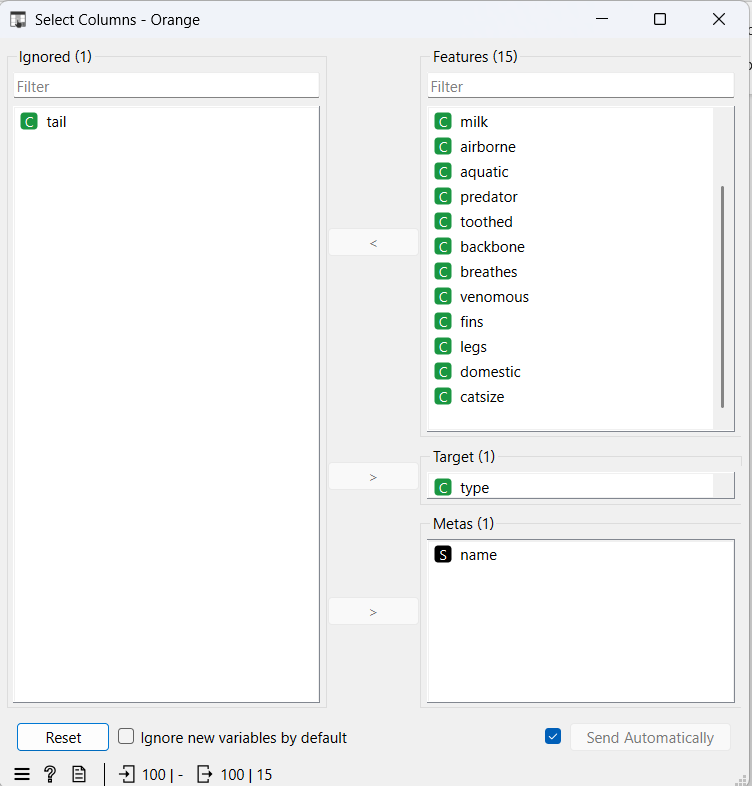
Так, например, среди тех, у кого есть зубы, больше всего зубов у млекопитающих.



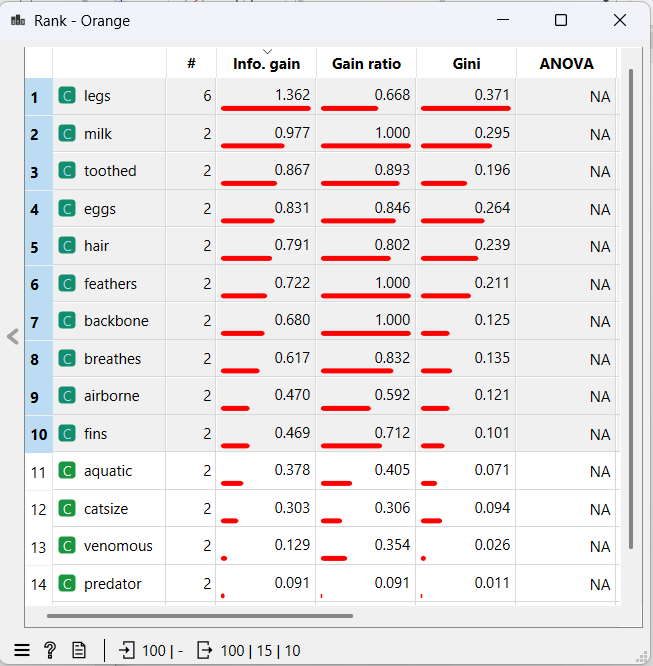
Здесь же можно посмотреть, есть ли в зоопарке ядовитые животные. И если есть, то среди каких видов - ядовитые животные есть, но их мало - некоторые рыбы, насекомые, беспозвоночные, рептилии.



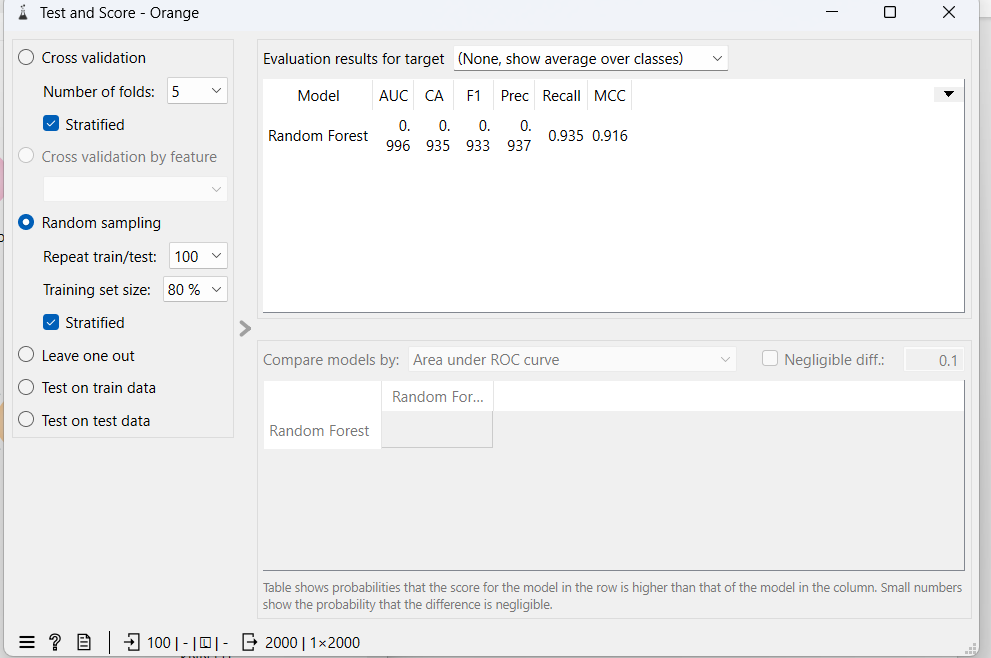
Исключила из выборки тех, кто имеет хвост через **Select Columns**



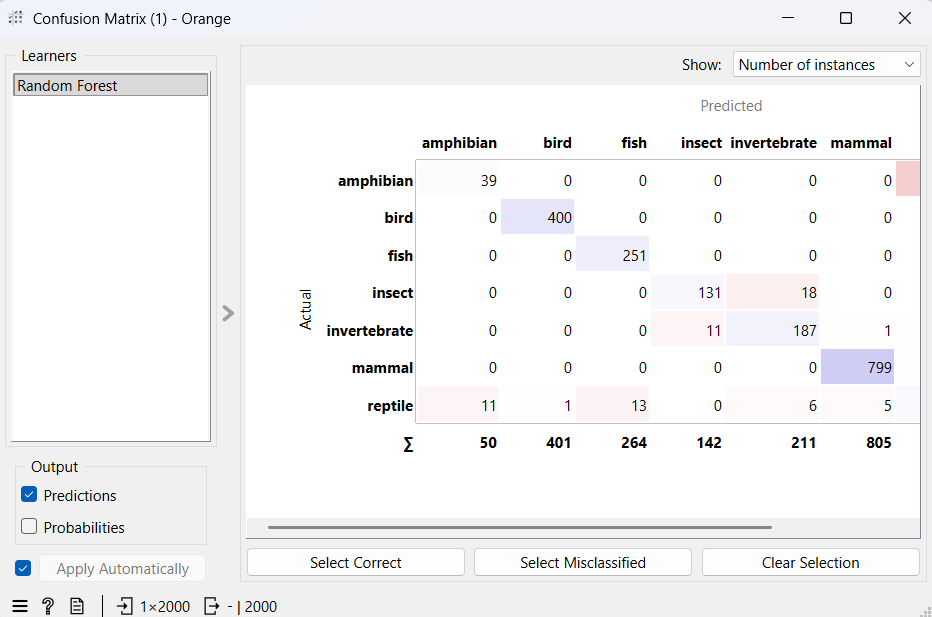
По **Rank** ограничила выборку 10 первыми позициями. Наибольшее значение имеют такие параметры как наличие ног и зубов, способность давать молоко.



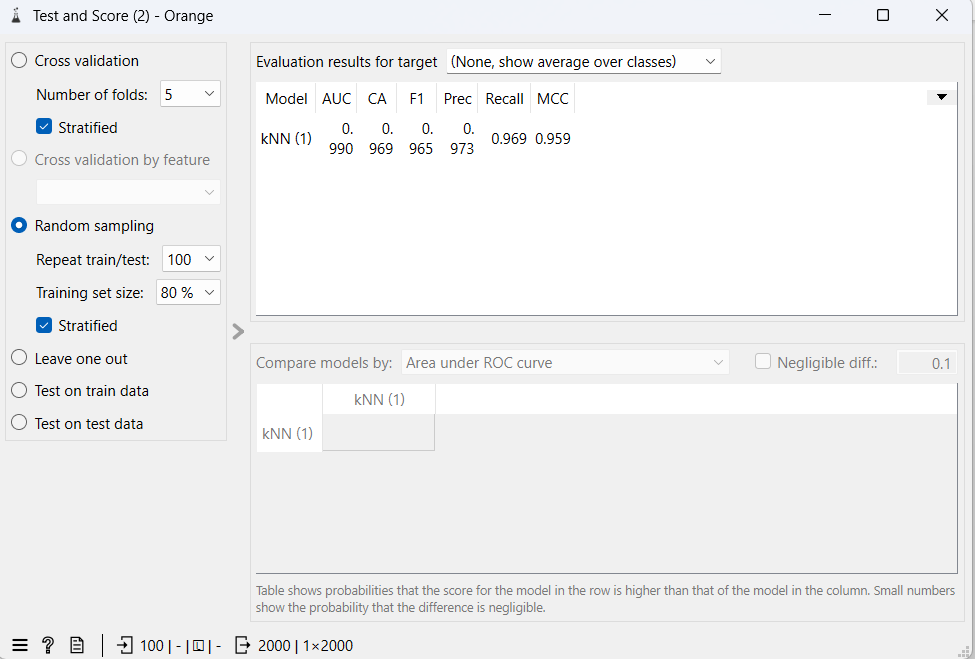
Запустила обучение по виджету **Random Forest.** Самая высокая точность при выборе Random sampling и параметрах 100 и 80%. Точность - 99,6 %.



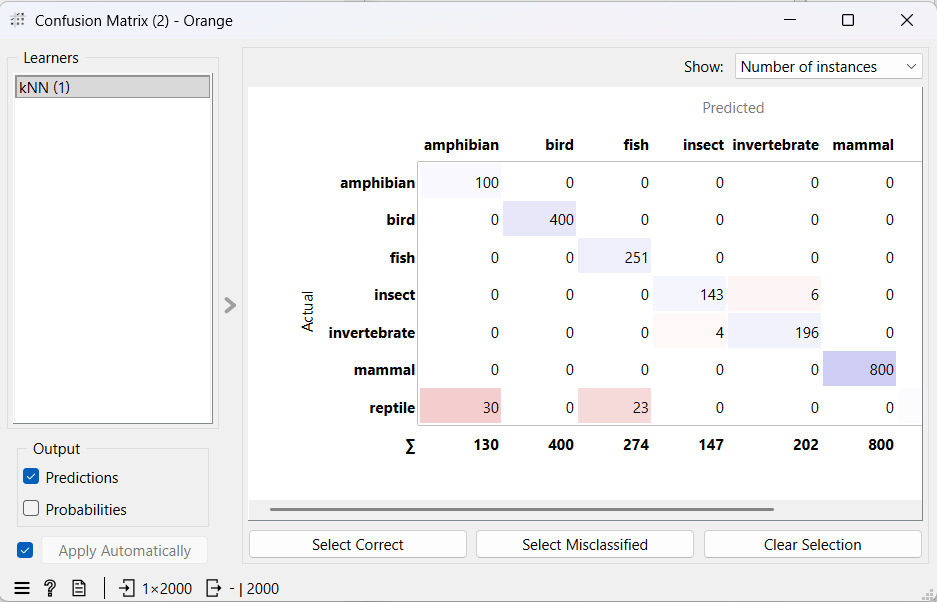
Матрица по этой модели:



Запустила обучение по виджету **kNN.** Самая высокая точность при выборе Random sampling и параметрах 100 и 80%. Точность - 99 %.



Матрица по этой модели:



В данном случае ошибок при обучении по kNN меньше, точность чуть больше.