Код группы 5091614. Состав (студенты 5 курса группы ИТ-50916: Галисултанов Р.И., Замятин К.А., Коркин И.С., Шаков М.А.)

Применение методов ML к данным из хранилища больших данных

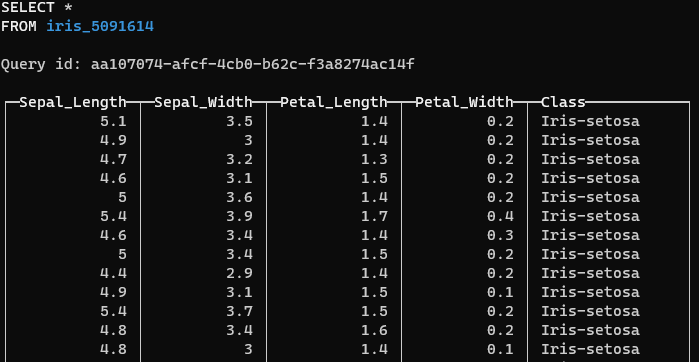
Создадим таблицу для загрузки данных

CREATE TABLE iris\_5091614(Sepal\_Length Float64, Sepal\_Width Float64, Petal\_Length Float64, Petal\_Width Float64, Class String) ENGINE = MergeTree() ORDER BY (Class)

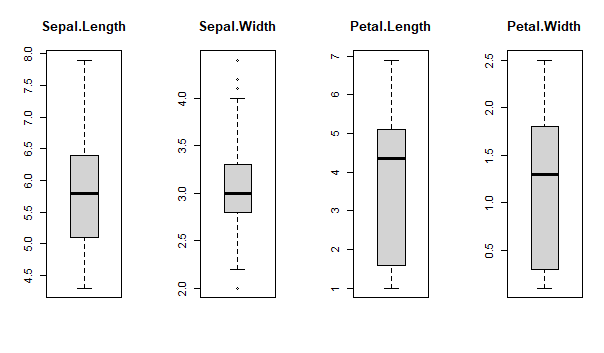
Произведем загрузку данных

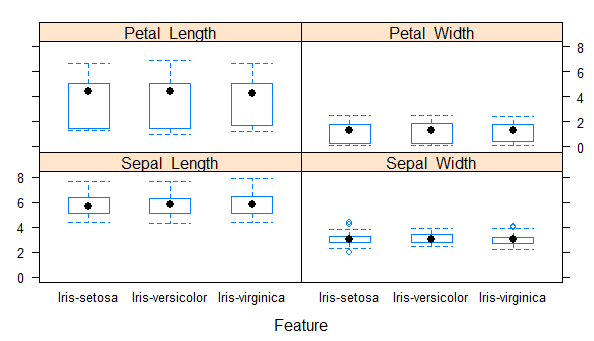
head --lines=150 iris.data | clickhouse-client --user default --password \*\*\*\*\*\*\*\*\*\*\* --query "INSERT INTO db\_5091614.iris\_5091614 FORMAT CSV" --max\_insert\_block\_size=100000

Данные успешно загружены



Перейдем в RStudio





Матрица ошибок

Confusion Matrix and Statistics

Reference

Prediction Iris-setosa Iris-versicolor Iris-virginica

Iris-setosa 10 0 0

Iris-versicolor 0 10 0

Iris-virginica 0 0 10

Overall Statistics

Accuracy : 1

95% CI : (0.8843, 1)

No Information Rate : 0.3333

P-Value [Acc > NIR] : 4.857e-15

Kappa : 1

Mcnemar's Test P-Value : NA

Statistics by Class:

Class: Iris-setosa Class: Iris-versicolor

Sensitivity 1.0000 1.0000

Specificity 1.0000 1.0000

Pos Pred Value 1.0000 1.0000

Neg Pred Value 1.0000 1.0000

Prevalence 0.3333 0.3333

Detection Rate 0.3333 0.3333

Detection Prevalence 0.3333 0.3333

Balanced Accuracy 1.0000 1.0000

Class: Iris-virginica

Sensitivity 1.0000

Specificity 1.0000

Pos Pred Value 1.0000

Neg Pred Value 1.0000

Prevalence 0.3333

Detection Rate 0.3333

Detection Prevalence 0.3333

Balanced Accuracy 1.0000