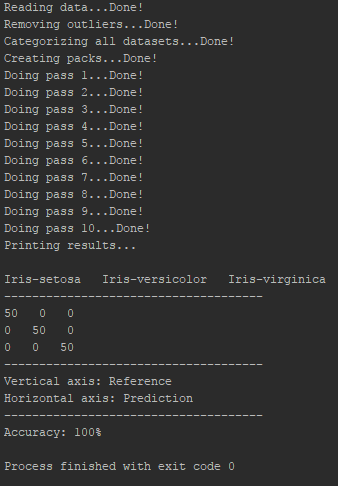
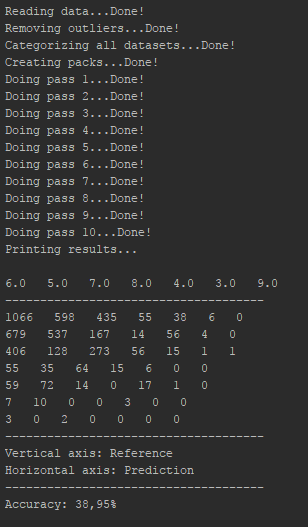
# Dokumentation K-Nearest-Neighbour Algorithmus

Wir haben uns entschlossen ein Framework zu bauen welches man für sämtliche Datensätze verwenden kann. Ich werde nun die Confusion-Matrix von dem Wein-Datensatz und für den IRIS-Datensatz. Zusätzlich haben wir entdeckt, dass der Algorithmus mit Java 8 ungefähr um das 10-fache schneller arbeitet als unter Java 9. Daher sind folgende Auswertungen mit Java 8 ausgewertet.

Man sieht, dass es beim Wein sehr schwierig ist zu Klassifizieren. Wenn man sich nun die Confusion-Matrix für die IRIS-Blüten (unten rechts) ansieht, erkennt man das die Accurancy auf 100 % ist. Für uns bedeutet es, dass die Daten vom Wein sehr schwierig auszuwerten sind.



https://puu.sh/zK8q1/abd9c2b5f1.pngNun haben wir mit dem Algorithmus die Zeit für 1000, 10 000, 100 000 Klassifizierungen gemessen und sind auf folgendes Ergebnis für die IRIS – Blüten gekommen:

https://puu.sh/zK8r9/7b1c37428c.pnghttps://puu.sh/zK8sB/dfdac523b1.png

Die Werte für den Wein-Datensatz für 1000, 10 000, 100 000 Datensätze ist folgender:

https://puu.sh/zK8uy/80e7a8d8af.pnghttps://puu.sh/zK8vF/e1d07d4e45.pnghttps://puu.sh/zK8Eu/dc40eab6ce.png