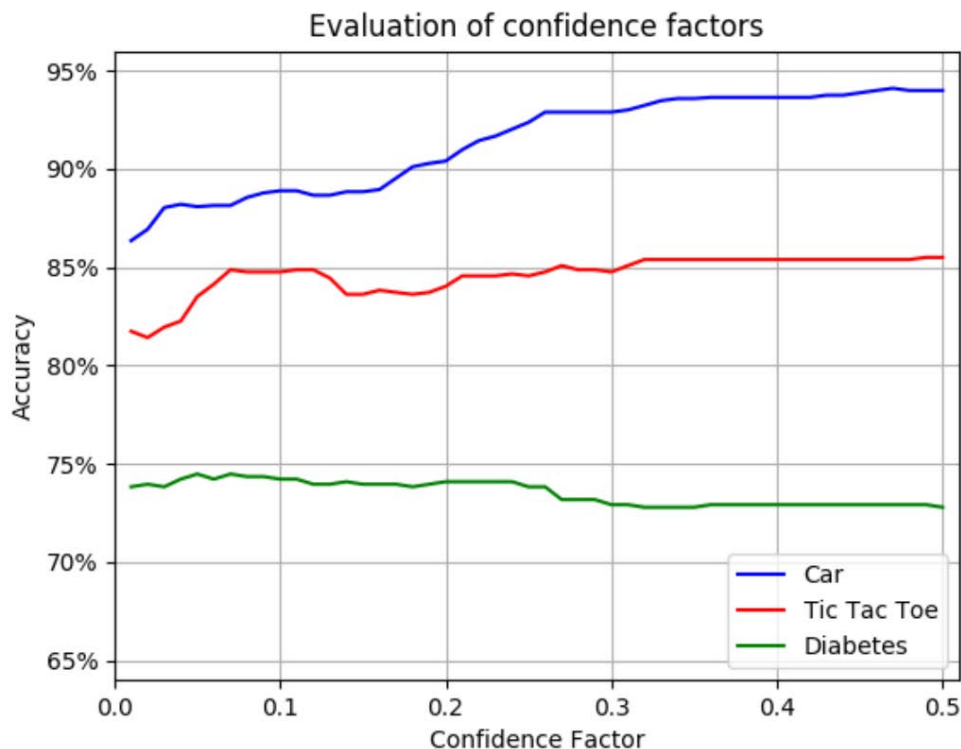


# Aufgabe 1:

**WEKA-Test:** J48 -C 0.01 0.5 50 (in 0.01 Schritten bis 0.5)

Mit dem J48 (C4.5) Algorithmus ergaben sich bei der Evaluierung, je nach Datensatz, ganz unterschiedlich Werte.

Bei dem Car und TTT (Tic-Tac-Toe) Datensatz ist bis C 0.5 eine Steigung zu beobachten, während es beim Diabetes Datensatz ab ca. C 0.25 sinkt.



Die besten Werte sind:

- **Car:** 94.1% mit C 0.47
- **TTT:** 85.5% mit C 0.5
- **Diabetes:** 74.48% mit C 0.05

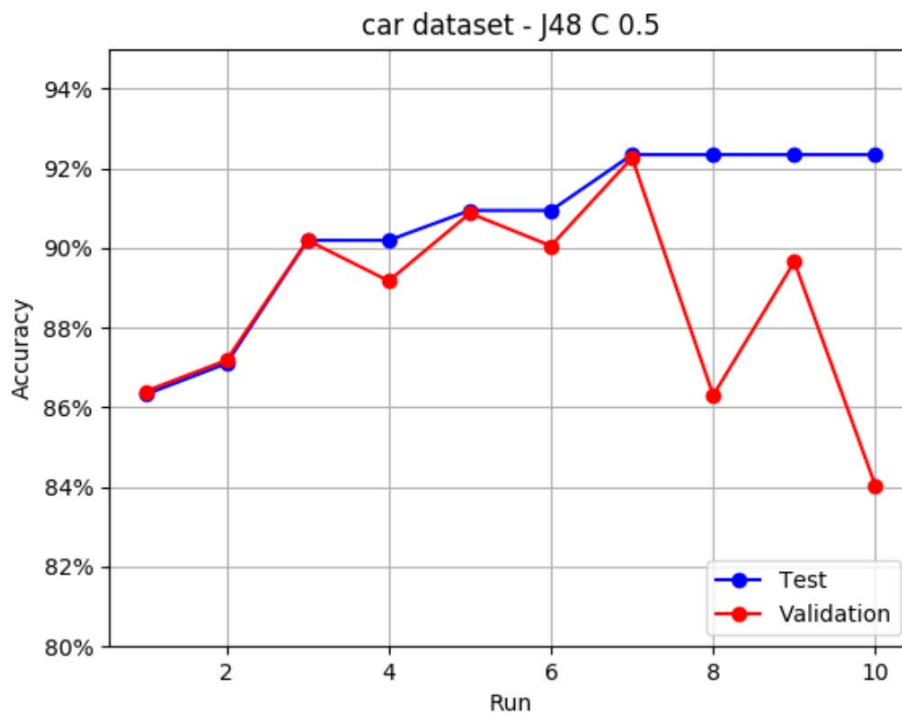
Anschließend wurden die Datensätze mit dem eigenen Classifier „OptimalJ48“ Evaluert. Für jeden Datensatz wurde ein 10-maliger Testlauf durchgeführt. Wobei bei jedem Testdurchlauf der optimal gefundene Parameter anschließend mit dem Validierungsset geprüft wurde.

Die Datensätze wurden in Train: 60%, Test: 20% und Validation: 20% aufgeteilt.

Bei der Evaluierung der Datensätze ergaben sich folgende Werte:

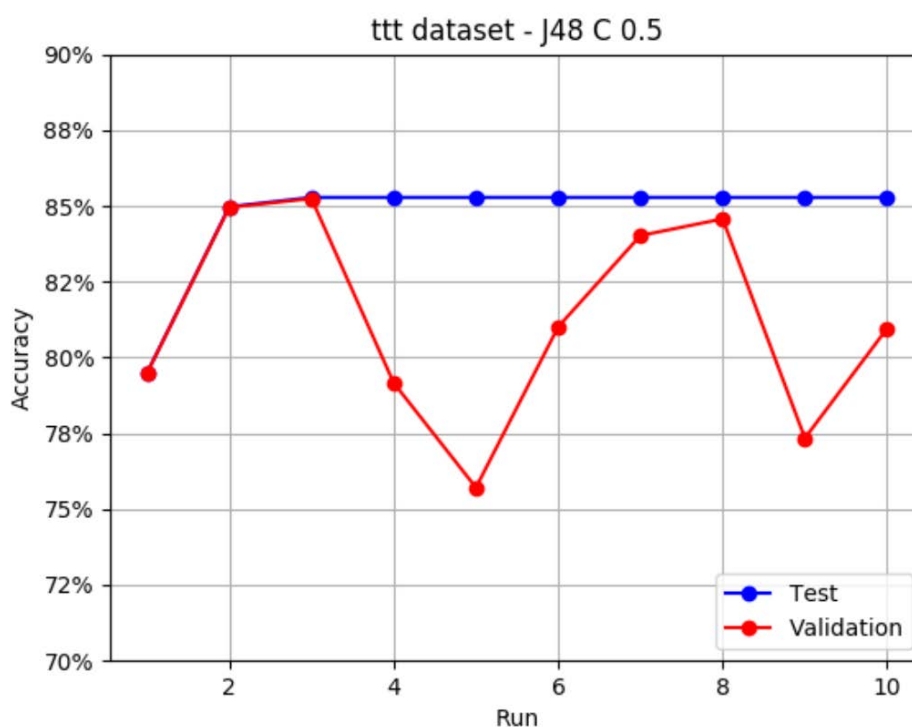
➤ **Car – bester Parameter C 0.5**

- Min: 86.32% Test; Abweichung Validierung 0%
- Max: 92.34% Test; Abweichung Validierung 8.31%



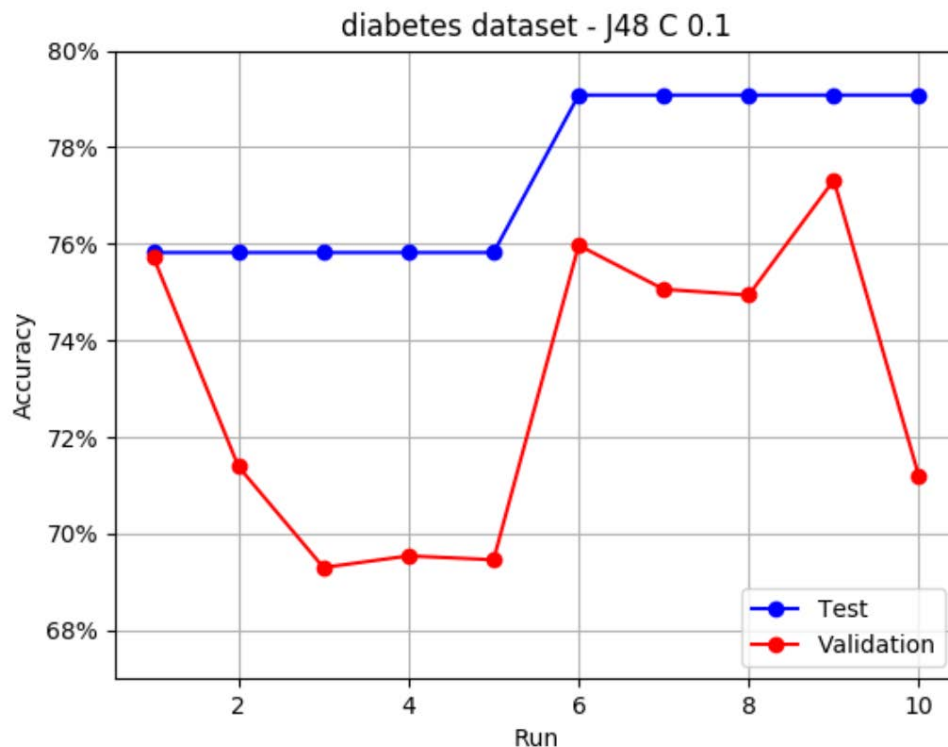
➤ **TTT – bester Parameter C 0.5**

- Min: 79.49% Test; Abweichung Validierung 0.01%
- Max: 85.28% Test; Abweichung Validierung 9.57%



➤ **Diabetes – bester Parameter C 0.1**

- Min: 75.82% Test; Abweichung Validierung 0.09%
- Max: 79.08% Test; Abweichung Validierung 7.89%



Wie man sehen kann entsprechen die gefunden Parameter fast denen aus dem beobachteten Weka-Test. Auch die Accuracy wird erreicht, natürlich durch die Randomisierung mit Schwankungen. Bei der Validierung kann Overfitting beobachtet werden, auch wenn diese eher gering ausfällt. Die Ergebnisse der Validierung liegen stellenweise sogar fast gleich mit dem Testergebnis, aber sie sind nie besser.