

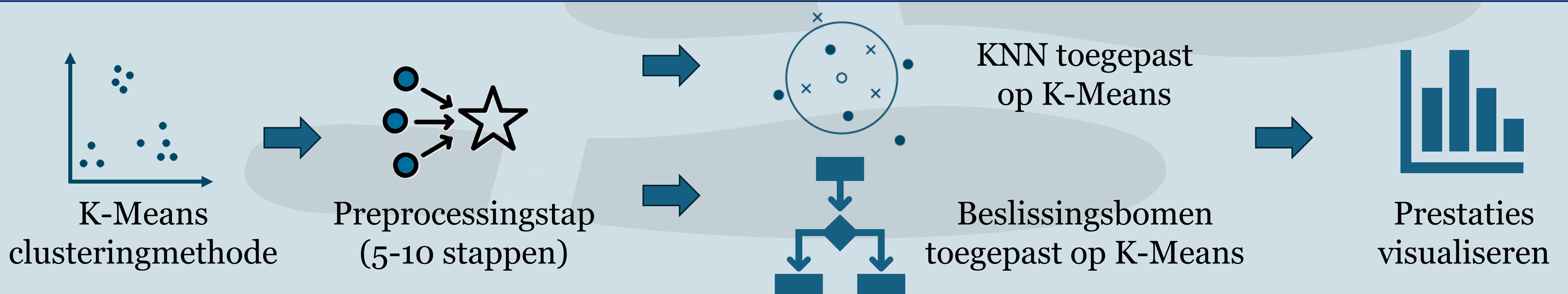
STUDYING THE IMPACT OF NEW FEATURE ENGINEERING METHODS

Onderzoekscontext



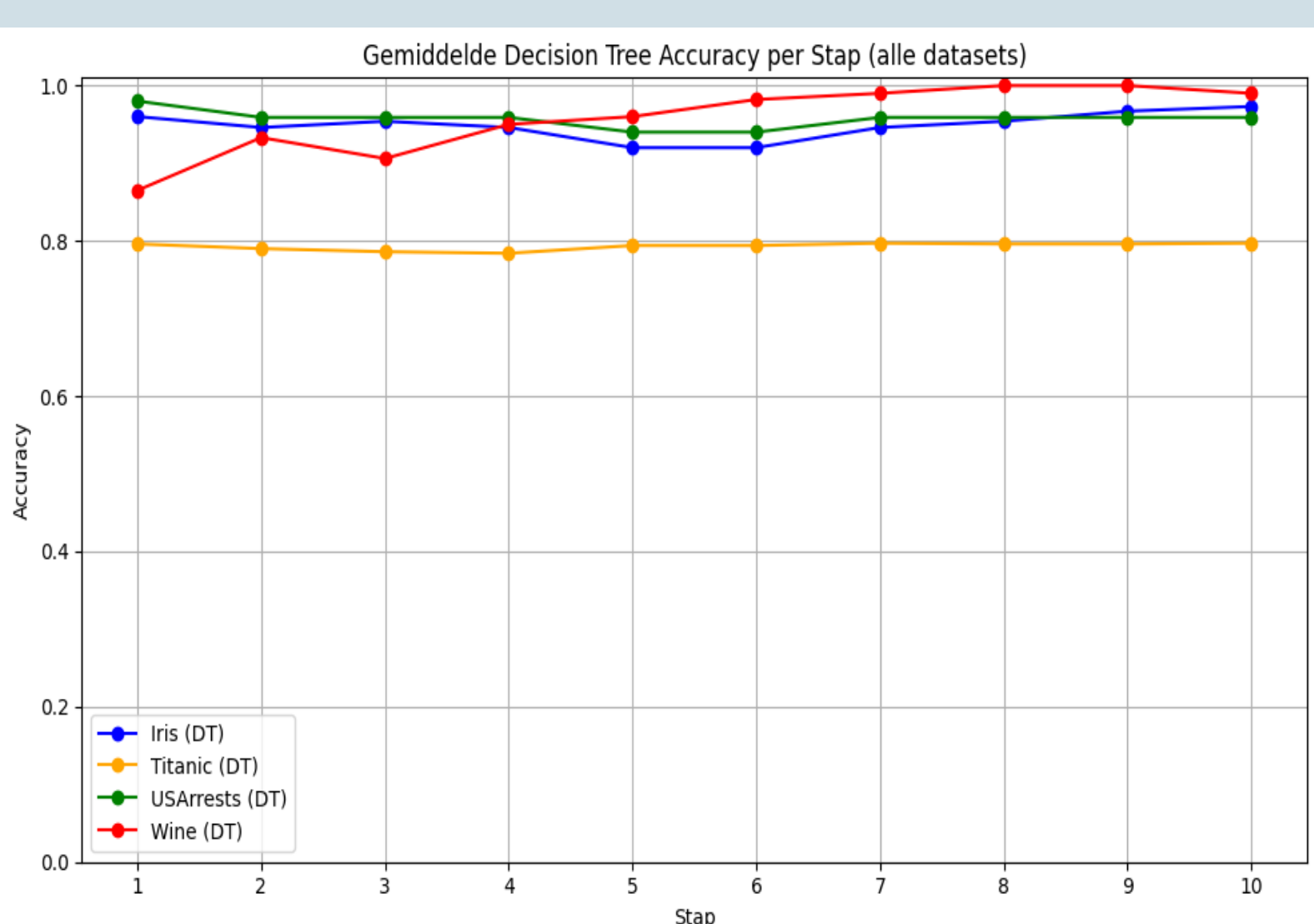
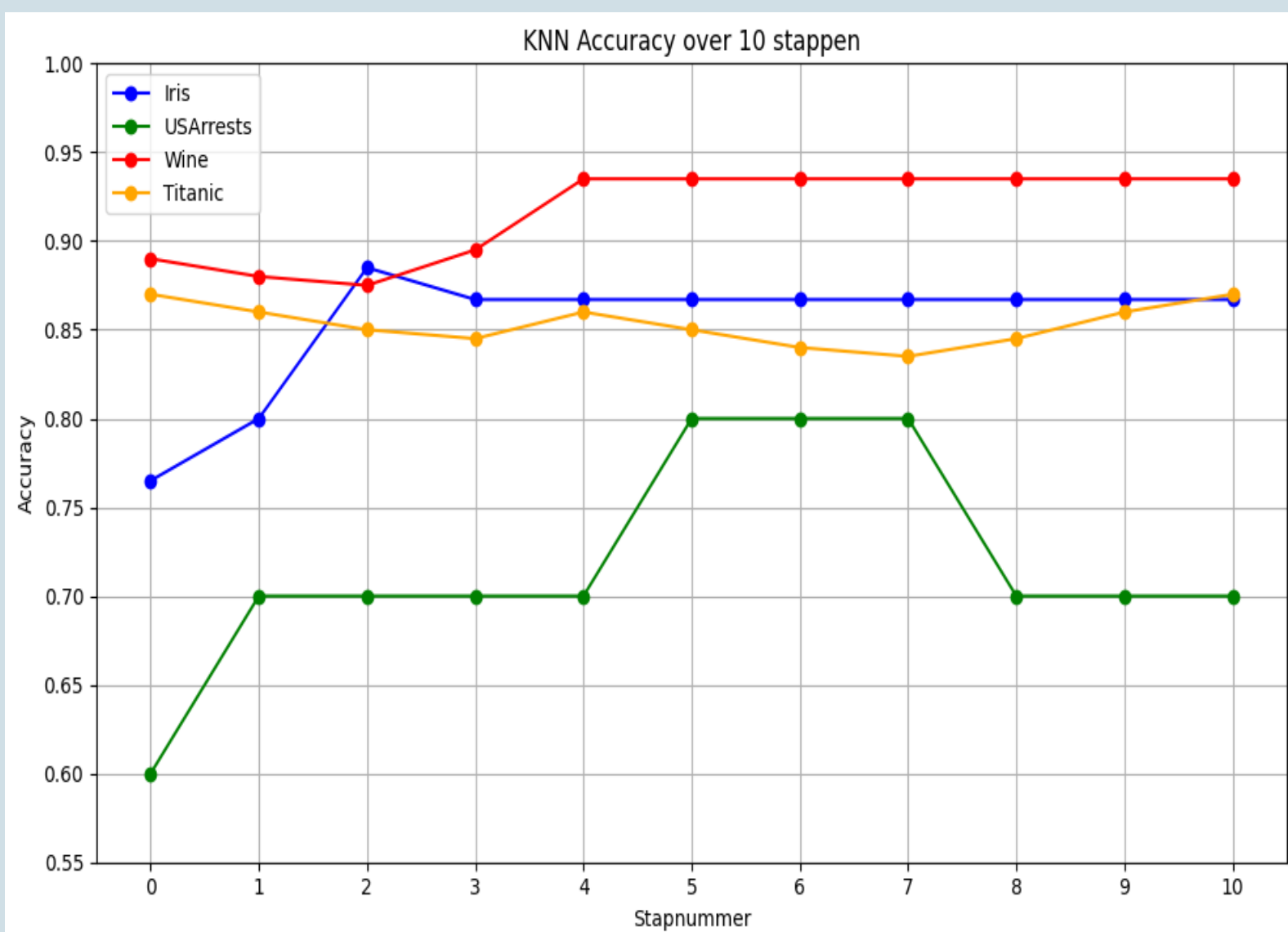
Methodologie

2x stappen herhalen → 1 100% van de dataset + unsupervised K- Means 2 80% van de dataset + supervised K- Means

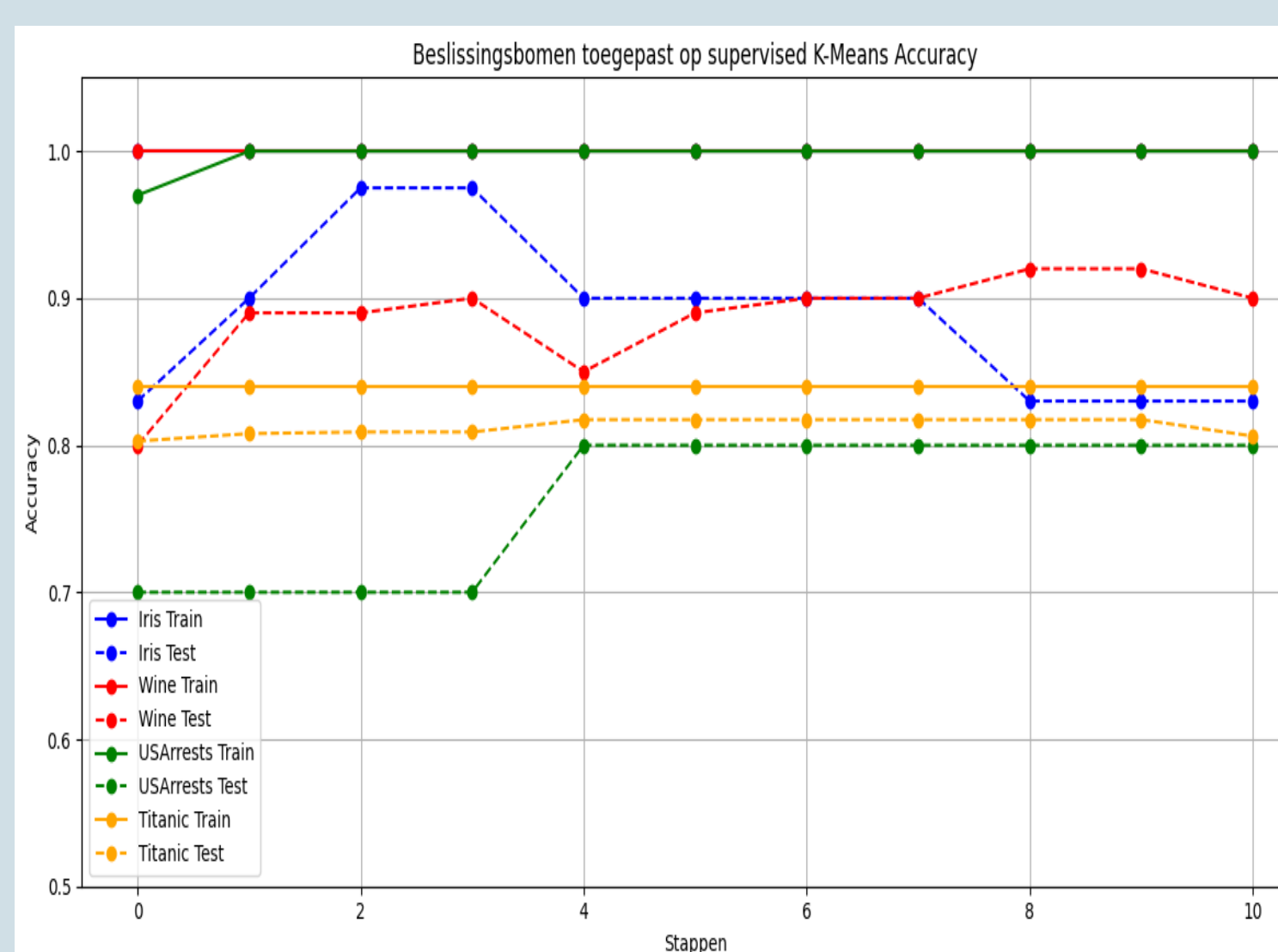
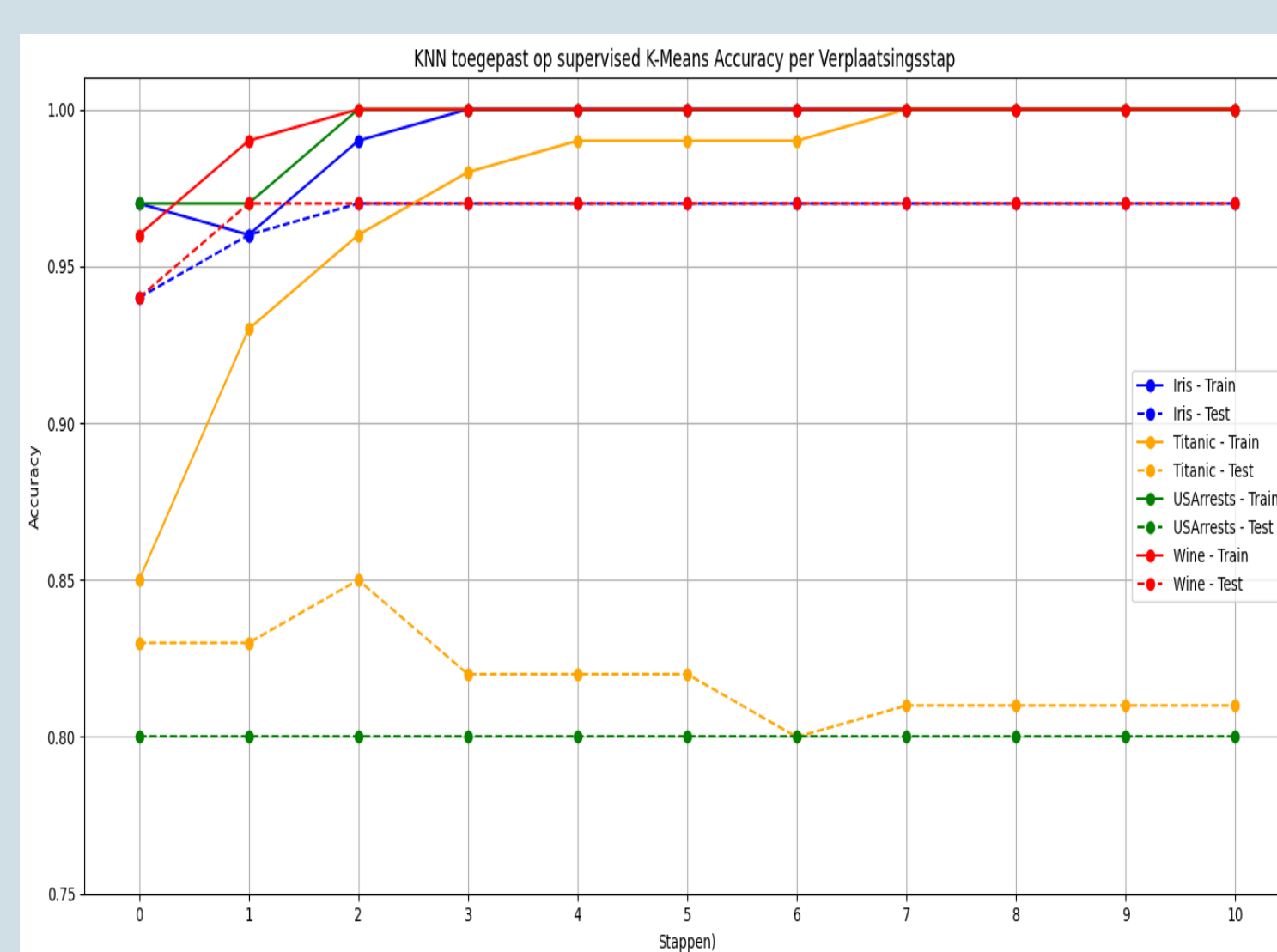


Resultaten

Metriek	Scenario clusteraantal*		
	1	2	3
Silhouette	↗	↗	↗
Standaard afwijking	↘	↘	↘
WSS	↘	↘	↘



Metriek	Scenario clusteraantal*		
	1	2	3
Silhouette	↗	↗ →	↗ →
Standaard afwijking	↘	↘	↘
WSS	↘	↘	↘



* Scenario 1: optimale clusteraantal 2 of 3
Scenario 2: verhoging naar 4 of 6
Scenario 3: verhoging naar 6 of 9

Conclusie

1) K-Means clustering

- Betere clusterdefinitie
- Compactere clusters
- Minder variatie

2) KNN

- Verder van scheidingslijn liggen
- Scheidingslijn lineair
- Accuracy ≈

3) Beslissingsbomen

- Minder splitsingen
- Minder brede beslissingen
- Accuracy ≈