

## SUMMARY

The Random Forest classifier demonstrated excellent performance on the HTRU2 pulsar dataset, achieving an overall accuracy of 97.23% and an impressive ROC AUC of 0.968. At the default decision threshold (0.5), the model reached a recall of 0.896, successfully identifying nearly nine out of ten true pulsars.

After performing threshold optimization, an ideal trade-off was found at  $t = 0.66$ , which preserved a strong recall ( $\sim 0.88$ ) while significantly reducing false positives, making the model both sensitive and reliable for practical use in candidate screening pipelines.

For the unsupervised learning phase, the K-Means algorithm ( $k = 3$ ) achieved a silhouette score of 0.388, revealing three moderately distinct pulsar subgroups with cluster sizes of roughly 470, 658, and 511 samples. In comparison, the Self-Organizing Map ( $5 \times 5$  grid) achieved a silhouette score of 0.193, uncovering numerous fine-grained micro-clusters ranging between 31 and 138 samples per node.

These findings suggest that while the eight statistical features in HTRU2 encapsulate some latent structure among pulsars, their overlap limits perfect separability. Future explorations using deep learning architectures on raw pulse-profile and DM-SNR data could uncover richer feature representations and enable more precise categorization of pulsar subtypes.