

Lecture 11: Feature Selection - Discretization

Project 10

Χρησιμοποιώντας το dataset [HTRU2](#), και σε συνδυασμό με έναν αλγόριθμο μηχανικής μάθησης της επιλογής σας (ο οποίος θα προβλέπει εάν ένα παράδειγμα του dataset είναι ένα αστέρι pulsar ή όχι), να εφαρμόσετε 2 μεθόδους Feature Importance.

Να παραδοθεί ο κώδικας, ο οποίος θα εμφανίζει:

- τις επιδόσεις του αντίστοιχου αλγορίθμου (*Accuracy*, *Precision*, *Recall*, *F1*, *γράφημα AUC*) πριν και μετά τη χρήση της μεθόδου PCA για $n=4$
- μια λίστα με τα πιο σημαντικά χαρακτηριστικά βάσει του μοντέλου, σε φθίνουσα σειρά (δηλαδή απ'το πιο σημαντικό στο λιγότερο σημαντικό)
- τις επιδόσεις του αντίστοιχου αλγορίθμου ο οποίος θα χρησιμοποιεί μόνο τα 4 πιο σημαντικά χαρακτηριστικά που βρέθηκαν προηγουμένως. Να συγκριθούν στη συνέχεια με τις επιδόσεις που βρέθηκαν στην πρώτη περίπτωση.