Krčenje podatkov

Najboljša izbira sprem

Pri manjšanju razsežnosti podatkov sem uporabil štiri načine. Varianca, t-test, relief in pomembnost rf. Za vsako od teh metod sem razvrstil spremenljivke od naj pomembnejše do najmanj pomembne. Nato sem testiral natančnost na testni množici, v odvisnosti od števila spremenljivk. Model, ki sem ga uporabil za učenje je rf, z nastavitvijo parametra ntree=100. Poskusil sem tudi metodo pca, ki pa se ni izkazalo za preveč učinkovito.

Izbiro najpomebnejših sprmenljivk sem naredil na podlagi t-testa in sicer z razvrstitvijo spremenljivk, glede na parameter statistika, ki jo vrne »ttest$statistic«. Razvrstil sem po padajočih vrednostih.

Glavne komponente so: …..

Grafi….