

Seminarska naloga ITAP - 2.del

Vito Rozman

12. junij 2022

Krčenje razsežnosti Pri manjšanju razsežnosti podatkov sem uporabil štiri pristope, varianca, t-test, relief in pomembnost *rf*. Za vsako od teh metod sem razvrstil spremenljivke od najpomembnejše do najmanj pomembne. Nato sem testiral natančnost na testni množici, v odvisnosti od števila spremenljivk. Model, ki sem ga uporabil za učenje je bil *rf*, z nastavitvijo parametra *ntree* = 100. Poskusil sem tudi metodo *pca*.

Za vsako napovedovanje območja (gozdovi, pozidano) sem natreniral model z vsemi značilkami in preveril natančnost na testni množici, tako sem dobil željeno natančnost, ki sem jo hotel doseči z čim manj značilkami. Nato sem na podlagi zgoraj izbraih metod primerjal natančnost modelov, kjer sem dodajal vedno več značil (seveda od najbolj pomembne do najmanj). Tista metoda, ki je z najmanj značilk dosegla mejno natančnost, je bila izbrana za najboljšo.

Izbira metode Izbiro najpomembnejših spremenljivk pri napovedovanju gozdov sem naredil na podlagi *relief* metode, pri napovedovanju pozidanega območja pa z *t-test*-om in sicer glede na vredost statistike.

Glavne komponente

- Napovedovanje gozdov: B02, B03, B12, B08, NDWI, NDBI, NDVI
- Napovedovanje pozidanega območja: B09, B8A, NDVI

