

Рачунарство високих перформанси у информационом инжењерингу

Опште упутство за практични део

Потребно је анализирати један велики скуп података спроводећи следеће активности:

- поређење перформанси пет или више класификационих модела који су истог типа, за иста предикторска и исто циљно обележје, али су креирани за различите вредности параметара
 - креирање моделâ, као и одређивање вредности за различите нумеричке показатеље перформанси моделâ применом валидационог приступа [5 поена]
 - визуализовање и тумачење односа између вредности параметара и перформанси у класификацији [5 поена]
- поређење перформанси три или више типова класификационих модела, за иста предикторска и исто циљно обележје (резултати могу бити дати у табеларном облику)
 - креирање моделâ, као и одређивање вредности за различите нумеричке показатеље перформанси моделâ применом валидационог приступа [5 поена]
 - одређивање перформанси за сваки тип класификационог модела применом k -тоструке унакрсне валидације [5 поена]
- кластеризовање података извршено за различите вредности параметара, као и анализа и тумачење добијених кластера [5 поена]
- формирање посебног аналитичког извештаја који ће обухватити опис анализираних података, преглед свих радњи над подацима, приказ свих добијених резултата и тумачење резултата [5 поена]

Очекивано је да креирање класификационих модела, оцењивање перформанси класификационих модела и кластеризовање података буду изведени применом система *Apache Spark*.