

SPO – projektni zadatak za 5. Ishod učenja

Zadatak je potrebno riješiti u Jupyter notebooku .

1. Kontrola kvalitete podataka

- Prepoznavanje nedostajućih vrijednosti
- Rekodiranje nedostajućih vrijednosti npr .null. -> „missing“
- Prepoznavanje uniformnih ili nelogičnih distribucija
- Provjeriti / redizajnirati /odrediti ciljnu varijablu i njenu kvalitetu

2. Statističke analize uzorka podataka

- Za svaku kategoričku varijablu prikazati distribucija frekvencije u grafu
- Za svaku kontinuiranu varijablu izračunati srednju vrijednost, z vrijednosti, zaobljenost distribucije, nagnutost distribucije
- Izračunati korelacije svake varijable sa ciljnom varijablom
- Izračunati korelacije svake varijable izražene kroz Z vrijednost i kategorizirane u razrede <-3 , $(-3 \text{ do } 3)$, >3 s ciljnom varijablom
- Objasniti što ste spoznali vezano uz promatrani skup podataka temeljem analiza iz točke 2
- Sve mora biti vidljivo u Jupyter Notebooku, kod i rezultati, kao i komentari

3. IV, WoE izračuni

- Izračunati IV, WoE **za svaku varijablu** i prikazati u Jupyter Notebooku
- Izdvojiti varijable sa najjačim IV-om i prikazati grafički WoE graf za sve najjače varijable
- Grupirati razrede po kriteriju WoE za najjače varijable
- Objasniti što ste spoznali temeljem ove analize (profili, uzroci, uzročno posljedične veze)
- Predložite kako biste riješili problem vezano uz vaš skup podataka temeljem IV, WoE analize (što bi trebalo mijenjati, raditi drugačije, gdje je problem uzrok problema, koji su rizični profili ...)
- Sve mora biti vidljivo u Jupyter Notebooku, kod i rezultati, kao i komentari