

# Pitanja iz oblasti Mašinsko učenje

## Pitanja iz prezentacije “Mašinsko učenje”

1. Kako definišemo mašinsko učenje?
2. U kojim slučajevima (tj. za koje vrste problema) je mašinsko učenje posebno korisno? Ukratko objasniti.
3. Koje su osnovne karakteristike nadgledanog mašinskog učenja?
4. Koje su osnovne karakteristike nenadgledanog mašinskog učenja?
5. Koje su osnovne karakteristike učenja uz podsticaje?
6. Koji su osnovni koraci procesa mašinskog učenja? Za svaki korak, ukratko navesti u čemu se taj korak sastoji.
7. Kojim faktorima je, generalno posmatrano, uslovljen odabir algoritma mašinskom učenja?
8. Koja je uloga validacije u procesu mašinskog učenja?
9. Šta je kros-validacija (*cross-validation*)? Ukratko objasniti kako funkcioniše.
10. Šta označavaju termini *over-fitting* i *under-fitting* (u kontekstu modela mašinskog učenja)? Ukratko objasniti.
11. Šta označavaju termini *bias* i *variance* (u kontekstu modela mašinskog učenja)? Ukratko objasniti.

## Pitanja iz prezentacije “Priprema podataka” (sa vežbi)

12. Šta je stratifikovana kros-validacija? Navesti prednosti / nedostatke u odnosu na klasičan pristup kros validaciji.
13. Šta je normalizacija i koja je njena uloga u postupku pripreme podataka za primenu algoritama mašinskog učenja?
14. Šta je standardizacija i koja je njena uloga u postupku pripreme podataka za primenu algoritama mašinskog učenja?
15. Šta je diskretizacija i koja je njena uloga u postupku pripreme podataka za primenu algoritama mašinskog učenja?
16. Koji su najčešći pristupi diskretizaciji u postupku pripreme podataka za primenu algoritama mašinskog učenja?

17. Šta je diskretizacija sa jednakim širinama opsega (Equal-width binning) u kontekstu pripreme podataka za primenu algoritama mašinskog učenja?
18. Šta je diskretizacija sa jednakim pojavljivanjima u opsezima (Equal-frequency binning) u kontekstu pripreme podataka za primenu algoritama mašinskog učenja?
19. Šta je nadgledana diskretizacija u kontekstu pripreme podataka za primenu algoritama mašinskog učenja?
20. Šta je selekcija atributa i kada se primenjuje u kontekstu pripreme podataka za primenu algoritama mašinskog učenja?
21. Koje su prednosti selekcije atributa u kontekstu pripreme podataka za primenu algoritama mašinskog učenja?
22. Koji su pristupi selekciji atributa u kontekstu pripreme podataka za primenu algoritama mašinskog učenja? Ukratko objasniti sve pristupe.
23. Navesti metode pretrage kod selekcije atributa u kontekstu pripreme podataka za primenu algoritama mašinskog učenja. Objasniti ukratko svaki metod.

### **Pitanja iz prezentacije “Klasifikacija”**

24. Kako definišemo zadatak klasifikacije?
25. Navesti mere koje se tipično koriste za procenu uspešnosti modela klasifikacije.
26. Kako definišemo i izračunavamo Preciznost (*Precision*) i Odziv (*Recall*) kao mere uspešnosti modela klasifikacije?
27. U kakvom su odnosu mere Preciznost (*Precision*) i Odziv (*Recall*)? Ukratko objasniti.
28. U kom slučaju se Tačnost (*Accuracy*) ne može smatrati pouzdanom merom uspešnosti klasifikacionog modela? Ukratko objasniti.
29. Navesti karakteristike F mere (*F measure*). U čemu se ogleda razlika između F mere i F1 mere? Napomena: potrebno je objasniti suštinu ove dve mere, same formule su manje relevantne.
30. Navesti karakteristike mere *AUC* (*Area Under the Curve*).
31. Za dati opis jednog konkretnog zadatka binarne klasifikacije (npr. klasifikacija email poruka na spam i not-spam), interpretirati matricu zabune (*confusion matrix*), tj. objasniti značenje polja matrice (vrednosti TP, TN, FP i FN) u kontekstu datog konkretnog zadatka. Potrebno je takodje, izračunati vrednosti sledećih mera: Tačnost (*Accuracy*), Preciznost (*Precision*) i Odziv (*Recall*).

### **Pitanja iz prezentacije “Naïve Bayes klasifikator”**

32. Napisati i objasniti Bayes-ovo pravilo (potrebno je objasniti značenje svih elemenata pravila).
33. Koje je osnovna pretpostavka Naive Bayes algoritma (zbog koje je nazvan naivnim)? I zašto je bilo potrebno uvesti tu pretpostavku?

34. Navesti osobine Naive Bayes algoritma.

### **Pitanja iz prezentacije “Klasifikacija – Naivni Bajes” (sa vežbi)**

35. Šta je osnovni (baseline) klasifikator?

### **Pitanja iz prezentacije “Stabla odlučivanja”**

36. Objasniti osnovnu ideju na kojoj su zasnovana klasifikaciona stabla

37. Opisati postupak rekurzivne binarne podele prostora atributa na regione (u kontekstu kreiranja klasifikacionih stabala).

38. Koji se kriterijumi često koriste za podelu prostora atributa na regione (u kontekstu kreiranja klasifikacionih stabala)?

39. Šta je to orezivanje stabla odlučivanja (*tree pruning*)? Zašto se radi i na osnovu čega određujemo u kojoj meri je potrebno orezati stablo?

40. Koje su prednosti i nedostaci stabala odlučivanja?

### **Pitanja iz prezentacije “Klasifikacija – Stabla odlučivanja” (sa vežbi)**

41. Napisati formulu za računanje informacione dobiti u kontekstu stabla odlučivanja (zajedno sa legendom pojmova) i objasniti kako se ona koristi prilikom procesa kreiranja stabla.

42. Šta je ARFF fajl i koje elemente sadrži?

### **Pitanja iz prezentacije “Klasterizacija”**

43. Kako definišemo zadatak klasterizacije?

44. Kako se vrši procena sličnosti instanci u postupku klasterizacije?

45. Opisati osnovne korake K-Means algoritma.

46. Koji se kriterijumi tipično koriste za procenu kvaliteta klastera formiranih u procesu klasterizacije?

47. Kako se prevazilazi problem K-Means algoritma uslovljen nasumičnim (*random*) inicijalnim izborom težišta? Ukratko objasniti.

48. Koji su pristupi za procenu broja klastera (tj. parametra K) pri primeni K-Means algoritma?

49. Koja je osnovna karakteristika probablističke klasterizacije?

50. Opisati osnovne korake Expectation Maximization (EM) algoritma.