

Examen Data Mining – Iulie 2016

Nume:.....

Specializare:.....

Notare: 0.5 pcte/ intrebare

Obs: Intrebarile pot avea 1,2 sau mai multe raspunsuri corecte

1. Presupunem ca avem un set de date care contine informatii de logare pe un server web si ne intereseaza identificarea unor profile de utilizatori. Ce tip de prelucrare este adecvata? (a) clasificare; (b) predictie; (c) grupare; (d) reguli de asociere; (e) regresie.
2. Consideram urmatorul set de tranzactii:
T1: {lapte, paine, carne, apa}
T2: {paine, apa}
T3: {paine, unt, carne, apa}
T4: {apa}
si regula: IF paine and carne THEN apa. Calculati valorile pentru suport: si pentru confidenta
3. Consideram un atribut numeric care ia valori in intervalul [a,b). Ne intereseaza o discretizare echidistanta in N valori (v_1, v_2, \dots, v_N). Ce interval de valori este mapat in valoarea v_i ? (a) $[a+i*(b-a)/N, a+(i+1)*(b-a)/N)$; (b) $[a+(i-1)*(b-a)/N, a+(i+1)*(b-a)/N)$; (c) $[a+(i-1)*(b-a)/N, a+i*(b-a)/N)$; (d) $[b-(i+N)*(b-a)/N, b-(i+N-1)*(b-a)/N)$; (e) $[b-(i+N-1)*(b-a)/N, b-(i+N-2)*(b-a)/N)$;
4. Se considera urmatoarea matrice de confuzie corespunzatoare unui clasificator binar

	Predicted C1	Predicted C2
True C1	25	15
True C2	5	55

Calculati acuratetea clasificatorului:

5. Se considera urmatoarele distributii de probabilitate ce corespund unor variante de split (in cazul construirii unui arbore de decizie pentru o clasificare binara): (a) (0.5,0.5); (b) (1,0); (c) (0.25,0.75); (d) (0.75,0.25). Pentru care dintre aceste distributii valoarea entropiei este minima?