1. Dopunite sliku jednostavnog modela DM procesa. Protumačite je.

SQL upiti »flat« fajlovi

Operaciona BP

....................................

............

............

*Data mining*

1. Objasnite razliku između:
   1. KDD i DM
   2. strategije učenja, metoda/tehnike rudarenja podataka, DM zadatka
2. Šta podrazmevamo pod selekcijom podataka za DM?
3. Objasnite šta obuhvata postupak preprocesiranja podataka za DM.
4. Koji su koraci CRISP-DM metodologije?
5. Najbolji opis *data mining*-a je:
   1. identifikacija paterna sadržanih u podacima
   2. dedukcija relacija sadržanih u podacima
   3. reprezentacija podataka
   4. simulacija trendova sadržanih u podacima.
6. Podatke koje koristimo za izgradnju *data mining* modela nazivamo:

a. validacionim podacima

b. podacima za obučavanje

c. test podacima

d. skrivenim podacima

1. Nadgledano učenje se razlikuje od nenadgledanog klasterovanja po tome što nadgledano učenje zahteva:
2. najmanje jedan ulazni atribut
3. da ulazni atributi budu kategorički
4. najmanje jedan izlazni atribut
5. da izlazni atributi budu kategorički
6. Koje tvrđenje o problemu predikcije je istinito?
7. Izlazni atribut mora biti kategorički.
8. Izlazni atribut mora biti numerički.
9. Rezultujući model je dizajniran da odredi buduće ishode.
10. Rezultujući model je dizajniran da klasifikuje trenutno ponašanje.
11. Za svaki od navedenih scenarija odredite da li je za njegovo rešenje potrebno automatizovano (nenadgledano / nadgledano) učenje ili klasičan upit u bazu podataka:
12. Koja je prosečna nedeljna zarada svih zaposlenih ženskog pola mlađih od 40 godina?
13. Odrediti karakteristike uspešnog trgovca polovnim automobilima.
14. Kada posetioci dođu na moj sajt, koji su ti proizvodi koje će najverovatnije kupiti zajedno?
15. Koji procenat mojih radnika ne dolazi jedan ili dva dana mesečno na posao?
16. Koji od ovih iskaza su netačni i zašto?
    1. k-means algoritam je pogodan za nadgledano učenje.
    2. Rezultati koje daje k-means algoritam nad skupom podataka su jednoznačni.
    3. k-means algoritam je moguće primeniti na sve tipove podataka
    4. k-means algoritam garantovano konvergira optimalnom rešenju.
17. Ako je data jednostavna transakciona baza :

TID Artikli

* + 1. A,B,C,D
    2. A,C,D
    3. C,D,E,A
    4. A,D,B
    5. B,C,G
    6. D,G
    7. A,B
    8. C,D,E

koristeći prag od 55% za pokrivanje, nađite »veliki skup artikala« L2 za ovu bazu. Koliko je poverenje pravila: Ako je kupljen artikal A, tada će kupci kupiti i artikal D.

1. Izračunajte poboljšanje u klasifikaciji (“lift”) nastalo upotrebom modela X sa matricom konfuzije:

|  |  |  |
| --- | --- | --- |
| Model X | Model: DA | Model: NE |
| DA | 45 | 55 |
| NE | 1955 | 7945 |

i **protumačite rezultat**!

1. Koje od navedenih tvrđenja, vezano za datu matricu konfuzije, su tačna?

|  |  |  |
| --- | --- | --- |
|  | Savet: prihvati | Savet: odbaci |
| Prihvaćeni | 600 | 25 |
| Odbačeni | 75 | 300 |

* + - * 1. Tačnost klasifikacije modela je 90%
        2. Stepen greške klasifikacionog modela je 90%
        3. Tačnost klasifikacije prihvaćenih entiteta u modelu je 30%
        4. Stepen greške klasifikacionog modela za klasu odbačenih entiteta je 25:325

1. Koji ulazni atribut Indexi i=1,2,3 je najbitniji, a koji najmanje važan, za određivanje vrednosti zavisne promenljive y u sledećoj jednačini linearne regresije:

y= 0.0001 Index1 + 23.456 Index2 –567.78 Index3 –11.22 Protumačite.

1. Koje su osnovne komponente modela neurona?
2. Zbog čega su kontinualne funkcije kao što je sigmoidna ili hiperbolički tangens često korištene funkcije aktivacije u aplikacijama neuronskih mreža na stvarne probleme?
3. Tumačite Delta pravilo učenja:



1. Objasnite iskaze:

Funkcionisanje neuronske mreže je crna kutija.

1. Obučili smo neuronsku mrežu da predviđa broj prevezenih putnika. U prošlom mesecu se broj putnika kretao između 2300 i 2600 putnika dnevno.
   1. Konvertujte 2450 prevezenih putnika u vrednost u intervalu od 0 do 1.
   2. Pretpostavimo da je na izlazu dobijen broj 0,7. Konvertujte ovu vrednost u razumljiv broj prevezenih putnika.
2. Dopunite reči koje nedostaju:

NM su .............................. koji oponaša ljudski mozak. Znanje je predstavljeno putem konekcija sa pridruženim ........................... U neuron dolaze ulazisa prethodnih neurona, kombinuju se u *funkciju ...............* i izračunava se izlaz iz neurona *funkcijom ..................*.

1. Objasnite pojmove:
2. izračunati atributi
3. normalizacija
4. Na primeru objasnite načine kodiranja podataka sistemom 1-od-N
5. Na primeru objasnite načine kodiranja podataka termometar sistemom.

1. Dopunite reči koje nedostaju: (tekst se odnosi na nadgedano učenje)
2. Ako se izlaz NM razlikuje od potrebnog, tada podešavamo ................. ........... tako da dobijemo izlaz sličniji željenom. Ulazi u NM su raniji slučajevi ili primeri iz problemskog domena i ........................ . Izlaz iz NM je realan broj iz ....................... . U fazi ..................... mreža ne uči.
3. Izračunajte izlaz iz neurona feed forward NM za ulaze: x1=0.1, x2=0.3, x3=-0.2, w1=0.5, w2=-0.4, w3=0.7 i F(x) = 1/(1+e-x).