

Subiect 1

1. De pe site-ul <http://www.amazon.com/> , alegeti 10 produse din aceeaasi categorie si extrageti:
 - numele
 - pretul
 - descrierea
 - cati oameni au gasi util al doilea comentariu afisat
2. De pe site-ul <http://www.crunchbase.com/> , pentru o companie oarecare, extrageti locatia companiei, numărul de angajați, website-ul și totalul fondurilor primite.
3. De pe pagina de wikipedia a unei monede naționale extrageți utilizatorii monedei, simbolul, subdiviziunile și valoarea inflației
4. De pe pagina de amazon.com a unei cărți extrageți titlul cărții, autorul și anul de publicare.
5. De pe site-ul <http://www.amazon.com/> , alegeti 10 produse din aceeaasi categorie si extrageti:
 - numele
 - pretul
 - descrierea
 - cati oameni au gasi util al doilea comentariu afisat
6. De pe pagina de metacritic a unui film, extrageti regizorul, scorul mediu acordat de critici, scorul mediu acordat de utilizatori și durata filmului.
7. De pe 10 pagini de www.themoviedb.org ale unor filme, extrageti regizorul, durata si principalii actori.

Subiect 2

1. Se da setul de date: <http://archive.ics.uci.edu/ml/datasets/Iris>. Sa se construiasca arborele de clasificare corespunzator. Utilizati primele 90% de inregistrari pentru construirea unui arbore de decizie. Se va testa arborele pe inregistrarile 90-100.
2. Se da setul de date: <http://archive.ics.uci.edu/ml/datasets/Iris>. Sa se utilizeze un mecanism de clustering. În ce măsură reflectă rezultatele obținute tipul de plantă?
3. Se da setul de date: http://archive.ics.uci.edu/ml/datasets/Dresses_Attribute_Sales. Să se aplice un algoritm de clustering și să se afișeze care intrare aparține cărui cluster.
4. Se da setul de date: <http://archive.ics.uci.edu/ml/datasets/User+Knowledge+Modeling#>. Să se aplice un algoritm de clustering și să se afișeze care intrare aparține cărui cluster.
5. Se da setul de date: <http://archive.ics.uci.edu/ml/datasets/Congressional+Voting+Records>. Sa se clasifice

printr-un arbore de decizie. Ce informatii suplimentarea capatam alegand metoda respectiva? Ultimele 10% dintre randuri vor fi utilizate ca exemple de test.

6. Se da setul de date: <http://archive.ics.uci.edu/ml/machine-learning-databases/mushroom/>. Să se construiască un mecanism de predicție pentru ciuperci otrăvitoare. Ultimele 10% dintre randuri vor fi utilizate ca exemple de test.
7. Se da setul de date: <http://archive.ics.uci.edu/ml/datasets/Bank+Marketing#>. Sa se antreneze un mecanism care sa prezica daca un client va face un depozit pe termen lung. Ultimele 10% dintre randuri vor fi utilizate ca exemple de test.
8. Se da setul de date <http://archive.ics.uci.edu/ml/datasets/Car+Evaluation>. Sa se construiasca un mecanism de predictie a pretului (pretul este reprezentat ca o categorie). Ultimele 10% dintre randuri vor fi utilizate ca exemple de test.
9. Se da setul de date: <http://archive.ics.uci.edu/ml/datasets/Syskill+and+Webert+Web+Page+Ratings>. Sa se construiasca un mecanism de predictie a ratingului (ratingul este reprezentat ca o categorie). Ultimele 10% dintre randuri vor fi utilizate ca exemple de test.
10. Se da setul de date: <http://archive.ics.uci.edu/ml/machine-learning-databases/tic-tac-toe/>. Sa se construiasca un arbore de decizie pentru rezolvarea jocului.

Subject 3

1. Se da setul de date: <http://archive.ics.uci.edu/ml/datasets/Iris>. Sa se construiasca arborele de clasificare corespunzator. Utilizati primele 90% de inregistrari pentru construirea unui arbore de decizie. Se va testa arborele pe inregistrarile 90-100.
2. Se da setul de date: <http://archive.ics.uci.edu/ml/datasets/Iris>. Sa se utilizeze un mecanism de clustering. În ce măsură reflectă rezultatele obținute tipul de plantă?
3. Se da setul de date: http://archive.ics.uci.edu/ml/datasets/Dresses_Attribute_Sales. Să se aplice un algoritm de clustering și să se afișeze care intrare aparține cărui cluster.
4. Se da setul de date: <http://archive.ics.uci.edu/ml/datasets/User+Knowledge+Modeling#>. Să se aplice un algoritm de clustering și să se afișeze care intrare aparține cărui cluster.
5. Se da setul de date: <http://archive.ics.uci.edu/ml/datasets/Congressional+Voting+Records>. Sa se clasifice printr-un arbore de decizie. Ce informatii suplimentarea capatam alegand metoda respectiva? Ultimele 10% dintre randuri vor fi utilizate ca exemple de test.
6. Se da setul de date: <http://archive.ics.uci.edu/ml/machine-learning-databases/mushroom/>. Să se construiască un mecanism de predicție pentru ciuperci otrăvitoare. Ultimele 10% dintre randuri vor fi utilizate ca exemple de test.

7. Se da setul de date: <http://archive.ics.uci.edu/ml/datasets/Bank+Marketing#>. Sa se antreneze un mecanism care sa prezica daca un client va face un depozit pe termen lung. Ultimele 10% dintre randuri vor fi utilizate ca exemple de test.
8. Se da setul de date <http://archive.ics.uci.edu/ml/datasets/Car+Evaluation>. Sa se construiasca un mecanism de predictie a pretului (pretul este reprezentat ca o categorie). Ultimele 10% dintre randuri vor fi utilizate ca exemple de test.
9. Se da setul de date:
<http://archive.ics.uci.edu/ml/datasets/Syskill+and+Webert+Web+Page+Ratings>. Sa se construiasca un mecanism de predictie a ratingului (ratingul este reprezentat ca o categorie). Ultimele 10% dintre randuri vor fi utilizate ca exemple de test.
10. Se da setul de date:
<http://archive.ics.uci.edu/ml/machine-learning-databases/tic-tac-toe/>. Sa se construiasca un arbore de decizie pentru rezolvarea jocului.