

# Project

## AI and SEM Master Modules:

**Data mining (AI),  
Data mining & Data  
warehousing (SEM)**

# Schedule

- ◆ Topic selection: Week 2
- ◆ Presentations: Weeks 3-11. PPT
- ◆ Algorithm implementation: Weeks 3-11
- ◆ Project presentation with final results: Weeks 13-14

# Possible topics (algorithms):

1. Apriori
2. Patricia Mine
3. FP-growth
4. ID3
5. C4.5
6. CART (Classification and regression trees)
7. CHAID decision tree
8. Multivariate adaptive regression splines (MARS)

# Possible topics (algorithms):

9. Naïve-Bayes
10. SLIQ algorithm
11. Neural networks
12. Support vector machines
13. k-Means clustering
14. k-Means++
15. k-Medoids
16. BIRCH

# Possible topics (algorithms):

17. CURE data clustering algorithm
18. O-cluster
19. Fuzzy C-Means
20. Self organizing maps
21. k Nearest Neighbors
22. Nearest-neighbor chain algorithm
23. EM algorithm
24. Social Networks Analysis

# Possible topics (algorithms):

25. Page Rank and other ranking algorithms
26. Bagging and boosting (AdaBoosting)
27. Multidimensional scaling algorithms
28. Silhouette (interpretation and validation of clusters of data)
29. Collaborative filtering and recommender systems
30. Information integration algorithms/methods
31. Sequence mining algorithm
32. Web log mining algorithm

# Possible topics (algorithms):

- 33. Eclat (depth-first search algorithm)
- 34. GSP (Generalized Sequential Pattern algorithm)
- 35. Random forest (classification)
- 36. DBSCAN (clustering)
- 37. OPTICS (Ordering Points To Identify the Clustering Structure)
- 38. CLARANS: clustering objects for spatial data mining

# Possible topics (algorithms):

- 39. GEP (Gene Expression Programming - classification)
- 40. Learning Vector Quantization (LVQ - classification)
- 41. Kriging (interpolation)
- 42. Imputation (preprocessing: dealing with missing data)

Other DM algorithm may be used after approval



# Deliverables

- ◆ Printed document including:
  - ◆ Importance and practical applications of the algorithm
  - ◆ Algorithm general presentation
  - ◆ Known results and issues
  - ◆ Datasets used (names and samples)
  - ◆ Results (including samples) and evaluation
  - ◆ References
- ◆ Archive with printed document (electronic form), PPT presentation, sample datasets and results must be uploaded on `cs.curs.pub.ro` before project delivery.