

6	ADVANCED TOPICS Mobile Databases, Distributed Object Management, Multi-databases	4	
---	--	---	--

Text book and Reference books:

1. Principles of Distributed Database Systems, M.T. Ozsu and PValduriez, Prentice-Hall, 1991.
2. Distributed Database Systems, D. Bell and J. Grimson, Addison-Wesley, 1992.

Data Mining

Code: PECAIML-601B

Contacts: 3L

Name of the Course:	Data Mining and prediction by machines		
Course Code PEC-AIML601B	Semester: VI		
Duration: 6 months	Maximum Marks: 100		
Teaching Scheme		Examination Scheme	
Theory:3 hrs./week		Mid Semester exam: 15	
Tutorial: NIL		Assignment and Quiz: 10 marks	
		Attendance: 5 marks	
Practical: NIL		End Semester Exam:70 Marks	
Credit Points:	3		

Unit	Content	Hrs/Unit	Marks/Unit
	Unit 1:		
1	Overview of data mining and predictive analytics. Where does it apply and where does it not apply. The emerging interdisciplinary field of Data Science – what on Earth is it? The potential pitfalls of analytics including big bad data and the problem of local sparsity in large data sets - big never guarantees sufficient. Brief discussion of Career Opportunities including an overview of the UNH MS Analytics program.	4	
	Unit 2:		
2	Data preprocessing and cleanup including informative missing values and imputation.	3	

3	Unit 3: Unsupervised learning: Exploring data with visualization (primarily JMP Pro and Enterprise Guide), Principal Components, Cluster Analysis, Variables Clustering, and Market Basket analysis (association analysis). The problem of explanatory (traditional) vs predictive modeling and why it matters.	8	
4.	Unit 4: The under and overfitting dilemma of predictive modeling. Includes a discussion of measures of overfitting and underfitting such as AICc, BIC, and the very new ERIC.	6	
5	Unit 5: Validation strategies to assess model predictive behavior and predictive inference	2	
6	Unit 6: Supervised learning for classification: k-nearest neighbors, Decision Trees and Random Forests, Naïve Bayes, Neural Nets, Logistic Regression, Generalized Regression, Support Vector Machines, Discriminant Analysis. Topics include boosted neural and tree models.	5	
7	Unit 7: Supervised learning for prediction: review of multiple linear regression and related topics like influence and multi-collinearity, PCR, Neural Nets, Generalized Regression including the LASSO (adaptive), LARS, Ridge, and Elastic Net (adaptive). Traditional variable Selection strategies such as Forward Selection and All Possible Models will also be covered.	6	
8	Unit 8: Model assessment measures for predictive and classification models: model scoring, prediction error analysis, ROC and Lift curves, profit matrices for	6	

classification, various model comparison criteria. Ensemble Modeling: combining predictive models to create even more powerful models; includes boosting and bagging strategies.		
---	--	--

Text book and Reference books:

1. Data Mining for Business Intelligence: Concepts, Techniques and Applications with JMP Pro; Shmueli, Bruce, Stephens, Patel 2017, Wiley & Sons
2. Preparing Data for Analysis with JMP by Robert Carver
3. Introduction to Statistical Learning, sixth printing, by Gareth, Tibshirani, Hastie, and Whitten

Database Management Systems**Code: OECAIML-601A****Contact: 3L**

Name of the Course:	Database Management Systems		
Course Code: OECAIML-601A	Semester: VI		
Duration:6 months	Maximum Marks:100		
Teaching Scheme		Examination Scheme	
Theory:3 hrs./week		Mid Semester exam: 15	
Tutorial: NIL		Assignment and Quiz: 10 marks	
		Attendance: 5 marks	
Practical: hrs./week		End Semester Exam:70 Marks	
Credit Points:	3		

Unit	Content	Hrs/Unit	Marks/Unit
1	Database system architecture: Data Abstraction, Data Independence,Data Definition Language(DDL),Data ManipulationLanguage(DML). Data models: Entity-relationshipmodel, network model, relational and object oriented data models, integrity constraints, data manipulation operations.	9	