Semester IV

Theory of Data Science (21M22MA213)

Volume, velocity, variety, machine learning, supervised and unsupervised learning, predictions and forecasts, innovation and experimentation, the dark side, big errors, privacy, example, polynomial curve fitting, probability theory, model selection, the curse of dimensionality, decision theory, information theory, regularization and stability, VC dimension., linear models for regression, parameter estimation methods - maximum likelihood method and maximum a posteriori method, regularization, ridge regression, lasso, bias-variance decomposition, bayesian linear regression, mixture models expectation-maximization method for parameter estimation, naive Bayes classifier, Non-parametric techniques for density estimation, Parzen-window method, k-nearest neighbors method, logistic regression, perceptron, Hidden Markov models (HMMS) for sequential pattern classification discrete HMMS and continuous density HMMS, Support vector machine, decision trees, bagging, boosting, gradient boosting, Principal component analysis, partial least squares, factor analysis, fisher discriminant analysis, multiple discriminant analysis, extracting data from web sources using APIs, text classification, metrics, grading text, text summarization.

Course Description

Course Code	21M22MA213	Semester	Even	Semester IV Session- 2023- 2024
				Month from Jan -June
Course Name	Theory of Data Science			
Credits	3		Contact Hours	3-0-0
Faculty (Names)	Coordinator(s)			
	Teacher(s) (Alphabetically)			
COURSE OUTCOMES: After the successful completion of this course, the student will be able to				COGNITIVE LEVELS
C235.1	explain important terms related to the art of data science.			Understanding Level (C2)
C235.2	make use of various regression techniques for data modeling.			Applying Level (C3)
C235.3	analyze different classification techniques for various datasets.			Analyzing Level (C4)
C235.4	judge quality of dataset based on available information.			Evaluating Level (C5)
Module No.	Title of the Module	Topics in the Module		No. of Lectures for the module
1.	The art of data Science	Volume, velocity, variety, machine learning, supervised and unsupervised learning, predictions and forecasts, innovation and experimentation,		6

		the dark side, big errors, privacy, example, polynomial curve fitting, probability theory, model selection, the curse of dimensionality, decision theory, information theory, regularization and stability, VC dimension.	
2.	Methods for function approximation:	linear models for regression, parameter estimation methods - maximum likelihood method and maximum a posteriori method, regularization, ridge regression, lasso, bias-variance decomposition, bayesian linear regression	7
3	Classification based on Bayesian decision theory	Bayesian decision theory, Bayes classifier, minimum error-rate classification, normal (Gaussian) density discriminant functions, decision surfaces, maximum-likelihood estimation, maximum a posteriori estimation, Gaussian mixture models expectation-maximization method for parameter estimation, naive Bayes classifier.	6
4	Classification based on non-parametric techniques	Non-parametric techniques for density estimation, Parzen-window method, k-nearest neighbors method, logistic regression, perceptron,	5
5	Sequential pattern classification	Hidden Markov models (HMMS) for sequential pattern classification discrete HMMS and continuous density HMMS	5
6	Boosting of classifiers	Support vector machine, decision trees, bagging, boosting, gradient boosting	5
7.	Dimensionality reduction	Principal component analysis, partial least squares, factor analysis, fisher discriminant analysis, multiple discriminant analysis.	4
8.	Extracting information from news	Algorithms, extracting data from web sources using APIs, text classification, metrics, grading text, text summarization.	4
		Total number of lectures	**42**

Evaluation Criteria

Components	Maximum Marks
T1	20
T2	20
End Semester Examination	35
TA	25 (Quiz, Assignments, Tutorials,Project)
Total	**100**

Project based learning: Students in a small group will collect sample data set and make classification models. They will validate the model by various selection and assessment methods. By this student will be able to make classification models and validate it

	Recommended Reading material: Author(s), Title, Edition, Publisher, Year of Publication etc. (Text books, Reference Books, Journals, Reports, Websites etc. in the IEEE format)
1.	**E. Alpaydin,**Introduction to Machine Learning, 2nd Ed., PHI Learning 2012.
2.	**C. M. Bishop,**Pattern Recognition and Machine Learning, Springer 2013.
3.	**T. Hastie, R. Tibshirani and J. Friedman,**The Elements of Statistical Learning, 2nd Ed., Springer 2008
4.	**S. R. Das,**Data Science Theories, Models, Algorithms, and Analytics, Apache License, 2016
5.	**S. S. Shwartz and S. B.David,** Understanding Machine Learning: from Theory to Algorithms, Cambridge University Press, 2014
6.	**R.O.Duda, P.E.Hart and D.G.Stork,** Pattern Classification, John Wiley, 2001

CO-PO and CO-PSO Mapping:

COs	PO1	PO2	PO3	PSO1
C235.1	2	2	-	2
C235.2	3	3	-	3
C235.3	3	3	-	3
C235.4	3	3	-	3

Dissertation (19M27MA211)

Course Description

Course Code	19M27MA211	Semester Even	Semester IV Session- 2023-2024
			Month from Jan -June
Course Name	Dissertation		
Credits	10	**Contact Hours**	
Faculty (Names)	**Coordinator(s)**		
	Teacher(s) (Alphabetically)		

COURSE OUTCOMES: After completion of the dissertation, student will be able to		COGNITIVE LEVELS
C250.1	understand the research-oriented problems and related areas.	Understanding Level (C2)
C250.2	organize the literature to form a problem in said area of study.	Applying Level (C3)
C250.3	develop the solution of the problem.	Applying Level (C3)

C250.4	analyze findings in terms of a report.	Analyzing Level (C4)

Module No.	Topics in module
1	Identification of the dissertation problem and literature review in the related field and explore experimental and theoretical tools/ techniques/software/hardware.
2	Acquire knowledge and analyze various methods/techniques to be used in solving the defined problem and find a suitable methodology.
3	Utilize latest techniques/software/hardware tools to achieve the proposed objectives and obtain results. Evaluation/analysis of the obtained results and their interpretation.

Evaluation Criteria

Components	Maximum Marks
Day to Day Evaluation	40 (To be awarded by supervisor)
End Term Evaluation	50 (To be awarded by a panel of 3 examiners)
Special Contribution	10 (To be awarded by a panel of 3 examiners)
Total	**100**

Employability: In this course, the students will be working on research problems in various fields of pure and applied Mathematics as per their specialization. The students will be able to learn to use the latest methods/techniques/tools/softwares to achieve the defined objectives of their dissertation. This will help the students to develop mathematical and scientific research temperament which will be beneficial for their future academics and research endeavors.

CO-PO-PSO Mapping

	PO1	PO2	PO3	PSO1
CO1	2	2	-	2
CO2	2	3	-	3
CO3	2	3	-	3
CO4	2	2	2	3