Master of Engineering - ME (Big Data Analytics)

Course File

Course Name	:	Machine Learning for Big Data
Course Code	:	BDA 5201
Academic Year	:	2024 - 25
Semester	:	II
Name of the Course Coordinator	:	Dr. AROCKIARAJ S
Name of the Program Coordinator	:	Dr. PRATHVIRAJ N

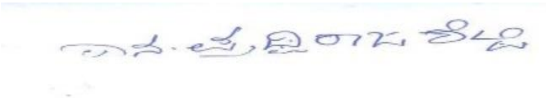	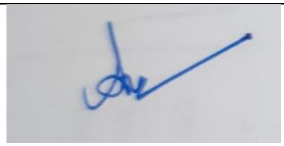
Signature of Program Coordinator with Date	**Signature of Course Coordinator with Date**

Table of Contents

Program Education Objectives (PEOs)

The overall objectives of the Learning Outcomes-based Curriculum Framework (LOCF) for **ME (Big Data Analytics)**, program are as follows.

PEO No.	Education Objective
PEO 1	Develop in depth understanding of the key technologies in data engineering, data science and business analytics.
PEO 2	Practice problem analysis and decision-making using machine learning techniques.
PEO 3	Gain practical, hands-on experience with statistics, programming languages and big data tools through coursework and applied research experiences.

Program Outcomes (POs)

By the end of the postgraduate program in **ME (Big Data Analytics)**, graduates will be able to:

PO1	Independently carry out research /investigation and development work to solve practical problems.
PO2	Write and present a substantial technical report/document.
PO3	Demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program.
PO4	Develop and implement big data analysis strategies based on theoretical principles, ethical considerations, and detailed knowledge of the underlying data.
PO5	Demonstrate knowledge of the underlying principles and evaluation methods for analyzing data for decision-making.

1. Course Plan

1.1 Primary Information

Course Name	**:**	Machine Learning for Big Data [BDA 5201]
L-T-P-C	**:**	3-0-0-3
Contact Hours	**:**	36 Hours
Pre-requisite	**:**	Basic concepts of Machine Learning
Core/ PE/OE	**:**	Core

1.2 Course Outcomes (COs), Program outcomes (POs) and Bloom's Taxonomy Mapping

CO	At the end of this course, the student should be able to:	No. of Contact Hours	Program Outcomes (PO's)	BL
CO1	Apply Artificial Neural Network, Clustering, Support Vector Machine, Deep Neural Network and Reinforcement Learning models.	10	PO3	3
CO2	Analyze the performance of single layer, multilayer, and deep neural networks.	10	PO4	4
CO3	Compare the performance of different clustering algorithms.	8	PO4	4
CO4	Evaluate the performance of different types of artificial neural network models, clustering models, deep neural network models, and reinforcement learning models	8	PO5	5

1.3 Assessment Plan

Components	Midterm	Flexible Assessments	End semester/ Makeup examination
Duration	90 minutes	3 written assignments	180 minutes
Weightage	0.3	0.2	0.5
Typology of questions	Applying; Analyzing.	Applying, analyzing and evaluating.	Applying, analyzing and evaluating.
Pattern	Answer all 5 questions of 10 marks each. Each question may have 2 to 3 parts of 3/4/5/6/7 marks.	Written assignments	Answer all 10 full questions of 10 marks each. Each question may have 2 to 3 parts of 3/4/5/6/7 marks.
Schedule	As per academic calendar.	Assignment submission on LMS: 5th April 2025	As per academic calendar.
Topics covered	Artificial Neural Networks – Clustering. Support Vector Machines.		Comprehensive examination covering the full syllabus. Students are expected to answer all questions.

1.4 Lesson Plan

L. No.	TOPICS	Course Outcome Addressed
L0	Course delivery plan, Course assessment plan, Course outcomes, Program outcomes, CO-PO mapping, reference books	---
L1	Artificial Neural Networks: Neurons and biological motivation	CO1
L2	Activation functions and threshold units	CO1
L3	Supervised and unsupervised learning	CO1
L4	Perceptron Model: representational limitation and gradient descent training	CO2
L5	Multilayer networks	CO2
L6	Back propagation	CO2
L7	Back propagation	CO4
L8	Overfitting	CO2
L9	Learning from unclassified data - Clustering	CO1
L10	Hierarchical Agglomerative Clustering	CO3
L11	Non- Hierarchical Clustering - k-means partitional clustering	CO1
L12	Non- Hierarchical Clustering - k-means partitional clustering	CO3
L13	Expectation Maximization (EM) for soft clustering	CO1
L14	Semi-supervised learning with EM using labelled and unlabelled data.	CO3

L15	Support Vector Machines (SMV): Maximum margin linear separators	CO1
L16	Quadratic programming solution to finding maximum margin separators	CO1
L17	Quadratic programming solution to finding maximum margin separators	CO4
L18	Kernels for learning non-linear functions	CO1
L19	Kernels for learning non-linear functions	CO4
L20	Kernels for learning non-linear functions	CO4
L21	Varying length pattern classification using SVM	CO1
L22	Varying length pattern classification using SVM	CO4
MTE	Midterm	CO1, CO2, CO3, & CO4
L23	Introduction to Deep Learning	CO1
L24	Introduction to convolutional Neural Network (CNN)	CO1
L25	CNN Architecture and layers	CO1
L26	CNN Architecture and layers	CO2
L27	Building simple CNN model for classification	CO1
L28	Building simple CNN model for classification	CO2
L29	Training and testing the CNN model	CO4
L30	Training and testing the CNN model	CO4
L31	Reinforcement Learning: Characteristics	CO1
L32	N-arm Bandit Problem	CO2
L33	Calculating the Value Function, Associative Learning – Adding States	CO1
L34	Calculating the Value Function, Associative Learning – Adding States	CO4

L35	The Markov Property & Markov Decision Process.	CO1
L36	The Markov Property & Markov Decision Process.	CO4
ESE	End semester	CO1, CO2, CO3, & CO4

1.5 References

1. Machine Learning, T. Mitchell, McGraw-Hill, 1997
2. Machine Learning, E. Alpaydin, MIT Press, 2010
3. Pattern Recognition and Machine Learning, C. Bishop, Springer, 2006
4. Pattern Classification, R. Duda, E. Hart, and D. Stork, Wiley-Interscience, 2000
5. Neural Networks - A Class Room Approach, Satish Kumar, Second Edition, Tata McGraw-Hill, 2013
6. The Elements of Statistical Learning: Data Mining, Inference and Prediction, T. Hastie, R. Tibshirani and J. Friedman, Springer, 2nd Edition, 2009
7. Machine Learning for Big Data, Jason Bell, Wiley Big Data Series
8. Kernel Methods for Pattern Analysis, J. Shawe-Taylor and N. Cristianini, Cambridge University Press, 2004
9. Neural Networks and Learning Machines, S. Haykin, Prentice Hall of India, 2010
10. Multidimensional Neural Networks Unified Theory, Rama Murthy G
11. F. Camastra and A. Vinciarelli, Machine Learning for Audio, Image and Video Analysis – Theory and Applications, Springer, 2008.
12. https://www.coursera.org/specializations/deep-learning
13. MOOC: Machine Learning | Courserahttps://www.coursera.org/specializations/deep-learning

1.6 Other Resources (Online, Text, Multimedia, etc.)

1. Web Resources: Blog, Online tools and cloud resources.
2. Journal Articles.

1.7 Course Timetable

2nd Semester Big Data Analytics				Room: LG1 LH 3				
	9-10	10-11	11-12	12-1	1-2	2-3	3-4	4-5
MON								
TUE		MLBD						
WED								
THU		MLBD						
FRI		MLBD						
SAT								

1.8 Assessment Plan

COs		Marks & Weightage			
CO No.	CO Name	Midterm (Max. 50)	Assignment (Max. 20)	End Semester (Max. 100)	CO wise Weightage
CO1	Apply Artificial Neural Network, Clustering, Support Vector Machine, Deep Neural Network and Reinforcement Learning models.	20	5	30	**0.32**
CO2	Analyze the performance of single layer, multilayer, and deep neural networks.	10	5	30	**0.26**
CO3	Compare the performance of different clustering algorithms.	10	5	20	**0.21**
CO4	Evaluate the performance of different types of artificial neural network models, clustering models, deep neural network models, and reinforcement learning models	10	5	20	**0.21**
	Marks (weightage)	**0.3**	**0.2**	**0.5**	**1.0**

Note:

- In-semester Assessment is considered as the Internal Assessment (IA) in this course for 50 marks, which includes the performances in class participation, assignment work, class tests, mid-term tests, quizzes etc.
- End-semester examination (ESE) for this course is conducted for a maximum of 100 and the same will be scaled down to 50.
- End-semester marks for a maximum of 50 and IA marks for a maximum of 50 are added for a maximum of 100 marks to decide upon the grade in this course.

Weightage for CO1 = = (mid semester marks for CO1 / 1.6666 + Assignment marks for CO1/1.0 + ESE marks for CO1 / 2)/100

= (20/1.666 + 5 + 30/ 2)/100 = 0.32

1.9 Assessment Details

The assessment tools to be used for the Current Academic Year (CAY) are as follows:

SI. No.	Tools	Weightage	Frequency	Details of Measurement (Weightage/Rubrics/Duration, etc.)
1	Midterm	0.3	1	• Performance is measured using internal test attainment level. • Reference: question paper and answer scheme. • Midterm test is assessed for a maximum of 50 marks and scaled down to 30 marks.
2	Assignments	0.2	1	• Performance is measured using assignments attainment level. • Assignments are evaluated for a maximum of 20 marks.
3	ESE	0.5	1	• Performance is measured using ESE attainment level. • Reference: question paper and answer scheme. • ESE is assessed for a maximum of 100 marks and scaled down to 50 marks.

1.10 Course Articulation Matrix

CO	PO1	PO2	PO3	PO4	PO5
CO1			**Y**		
CO2				**Y**	
CO3				**Y**	
CO4					**Y**
Average Articulation Level			*	*	*