Master of Engineering - ME (Big Data Analytics)

Course File

Course Name	:	Fundamentals of Machine Learning
Course Code	:	BDA 5103
Academic Year	:	2024 - 25
Semester	:	I
Name of the Course Coordinator	:	Dr. Arockiaraj S
Name of the Program Coordinator	:	Dr. Prathviraj N

	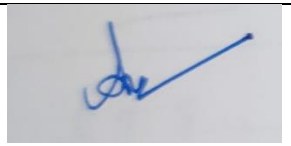
Signature of Program Coordinator with Date	**Signature of Course Coordinator with Date**

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	**:**	Fundamentals of Machine Learning [BDA 5103]
L-T-P-C	**:**	3-0-0-3
Contact Hours	**:**	36 Hours
Pre-requisite	**:**	Basic Programming with Python
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** different types of supervised and unsupervised machine learning algorithms to practical problems.	8	PO3	3
CO2	**Analyse** different types of machine learning paradigms.	12	PO4	4
CO3	**Evaluate** the performance of machine learning algorithms.	16	PO5	5

1.3 Assessment Plan

Components	Mid semester	Flexible Assessments (2 – 3 in number)	End semester/ Makeup examination
Duration	90 minutes	To be decided by the faculty.	180 minutes
Weightage	0.3	0.2	0.5
Typology of questions	Applying; Analyzing and Evaluating.	Applying; Analyzing. Evaluating.	Applying; Analyzing; 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.	**Assignment:** Solving problems by applying, analyzing and evaluating Generative AI use cases.	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:** November 2024	As per academic calendar.
Topics covered	Introduction to Machine Learning; Decision Trees- Linear Model: K-nearest Neighbours Algorithm- Cross-validation - Dimension Reduction.	Generative AI use cases, project lifecycle, and model pre-training - Fine-tuning and evaluating large language models.	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	Overview of Supervised (regression and classification), unsupervised (clustering and dimensionality reduction), semi-supervised, and reinforcement learning with practical examples.	CO1
L2	Machine learning nomenclature: raw data, types of features and outputs, feature vector.	CO1
L3	Decision tree model of learning. Classification and regression using decision trees.	CO1
L4	Splitting criteria: entropy, information gain, Gini impurity.	CO1
L5	Splitting criteria: entropy, information gain, Gini impurity.	CO1
L6	Splitting criteria: Gini impurity.	CO1
L7	Overfitting in decision trees.	CO1
L8	Pruning in decision trees.	CO1
L9	Linear regression: model, estimation, and interpretation of coefficients.	CO2
L10	Introduction to bias/variance trade-off.	CO2
L11	Regularized linear regression.	CO2
L12	K-nearest neighbours algorithm.	CO2
L13	K-nearest neighbours algorithm.	CO2

L14	Cross-validation	CO2
L15	Dimension reduction using principal component analysis (PCA)	CO2
L16	Dimension reduction using principal component analysis (PCA)	CO2
L17	Dimension reduction using multidimensional scaling	CO2
L18	Dimension reduction using multidimensional scaling	CO2
L19	Dimension reduction using t-SNE (t-distributed Stochastic Neighbour Embedding).	CO2
L20	Dimension reduction using t-SNE (t-distributed Stochastic Neighbour Embedding).	CO2
IT1	Internal test 1	CO1 & CO2
L21	Bagging: classification using random forest.	CO3
L22	Boosting.	CO3
L23	Probabilistic modelling of data using parameters - Introduction to maximum likelihood estimation (MLE) of parameters.	CO3
L24	Probabilistic modelling of data using parameters - Introduction to maximum likelihood estimation (MLE) of parameters.	CO3
L25	Naive Bayes model for classification.	CO3
L26	Naive Bayes model for classification.	CO3
L27	Logistic regression for binary classification.	CO3
L28	Logistic regression for binary classification.	CO3
L29	Model performance metrics.	CO3
L30	Cross-validation.	CO3
L31	Generative AI use cases.	CO3

L32	Generative AI use cases.	CO3
L33	Project lifecycle, and model pre-training.	CO3
L34	Project lifecycle, and model pre-training.	CO3
L35	Fine-tuning and evaluating large language models.	CO3
L36	Fine-tuning and evaluating large language models.	CO3

1.5 References

1. Module: Introduction to Machine Learning (https://www.intel.com/content/www/us/en/developer/tools/oneapi/training/academic-program/educators/intro-machine-learning-training-kit.html)
2. Module: Get started with AI on Azure (https://learn.microsoft.com/en-us/training/modules/get-started-ai-fundamentals/)
3. Module: Microsoft Azure AI Fundamentals: Get started with artificial intelligence (https://learn.microsoft.com/en-us/training/paths/get-started-with-artificial-intelligence-on-azure/)
4. Learning path: Understand data science for machine learning (https://learn.microsoft.com/en-us/training/paths/understand-machine-learning/)
5. Module: Generative AI with Large Language Models (https://www.coursera.org/learn/generative-ai-with-llms)
6. Grokking Machine Learning, Luis G. Serrano, Manning Publications; 1st Edition, 2019 - Online resource from Manning Publications available at https://www.manning.com/books/grokking-machine-learning
7. A Course in Machine Learning, Hal Daumé III – Online resource available at http://ciml.info/
8. An Introduction to Statistical Learning: with Applications in Python (Springer Texts in Statistics), Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor, 1st Edition, 2023 – Online resource available at https://www.statlearning.com/

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

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

1.7 Course Timetable

1st Semester Big Data Analytics				Lecture Hall:				
	9-10	10-11	11-12	12-1	1-2	2-3	3-4	4-5
MON								
TUE	FML							
WED								
THU	FML							
FRI								
SAT	FML							

1.8 Assessment Plan

COs		Marks & Weightage			
CO No.	CO Name	Mid semester (Max. 50)	Assignment (Max. 20)	End Semester (Max. 100)	CO wise Weightage
CO1	Apply different types of supervised and unsupervised machine learning algorithms to practical problems.	25	5	20	**0.30**
CO2	Analyse different types of machine learning paradigms.	25	5	40	**0.40**
CO3	Evaluate the performance of machine learning algorithms.	-	10	40	**0.30**
	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

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	Internal Test	0.3	1	• Performance is measured using internal test attainment level. • Reference: question paper and answer scheme. • Each internal test is assessed for a maximum of 50 marks and scaled down to 40 marks.
2	Assignments	0.2	2	• Performance is measured using assignments/quiz attainment level. • Assignments/quiz are evaluated for a maximum of 10 marks.
3	End semester	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
Average Articulation Level			Y	Y	Y