

Department	International College of Liberal Arts		
Semester	Spring 2023	Year Offered (Odd/Even/Every Year)	Every Year
Course Number	DATA350		
Course Title	Machine Learning		
Prerequisites	DATA160 Coding Bootcamp: Python AND DATA240 Data Visualization techniques in Python AND DATA250 Mathematics for Data Science AND DATA260 Coding Bootcamp: Applied Probability and Statistics		
Course Instructor	PARIDA Abhishek	Year Available (Grade Level)	3
Subject Area	Data Science	Number of Credits	3
Class Style	Lecture	Class Methods	Face to face

(NOTE 1) Class Methods are subject to change

(NOTE 2) Depending on the class size and the capacity of the facility, we may not be able to accommodate all students who wish to register for the course"

Course Description	This course introduces various Machine Learning algorithms (theory) and their implementations using Sklearn in Python (practical programming). It provides essential training on data literacy (reading, handling, and explaining the data) and discusses the latest trends in utilizing Data Science and Artificial Intelligence. Each section follows several guided case studies analyzing real-world data sets and practice problems.
Class plan based on course evaluation from previous academic year	None
Course related to the instructor's practical experience (Summary of experience)	None
Learning Goals	<p>Course Objectives</p> <p>A student taking this course would be considered a prospective Data Engineer who would already be familiarized with statistical notions and clearly understand the schemes needed to reach this point. This module's key takeaway is the rigorous theory sections that build the concepts and the hands-on learning that can be readily used in a Data Science professional's everyday work once mastered. Students would develop the skills needed to master mathematics for Data Science and Artificial Intelligence.</p>

iCLA Diploma Policy	DP1/DP2
---------------------	---------

iCLA Diploma Policy

(DP1) To Value Knowledge – Having high oral and written communication skills to be able to both comprehend and transfer knowledge

(DP2) To Be Able to Adapt to a Changing World – Having critical, creative, problem-solving, intercultural skills, global and independent mindset to adopt to a changing world

(DP3) To Believe in Collaboration – Having a disposition to work effectively and inclusively in teams

(DP4) To Act from a Sense of Personal and Social Responsibility – Having good ethical and moral values to make positive impacts in the world

Active Learning Methods	Students are advised to take handwritten notes; this will drastically increase their ability to retain the information; further, they are expected to practice regularly.
Use of ICT in Class	None

Use of ICT outside Class	None
Expected study hours outside class	One to two hours of study is required before the class preparation, and an equal amount of practice is needed after each lecture.
Feedback Methods	The best way to correspond during the course is through the UNIPA system or direct emails. Please check the UNIPA account regularly for updates related to classes. To have a better grade, be regular in the study, active and attentive in class, do a revision of classwork regularly, and participate in-class quizzes.

Grading Criteria		
Grading Methods	Grading Weights	Grading Content
Assignments	40%	
Quizzes	60%	

Required Textbook(s)	Aurélien Géron- Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (O'Reilly) Andreas C. Mueller, Sarah Guido- Introduction to Machine Learning with Python: A Guide for Data Scientists"
Other Reading Materials/URL	None
Plagiarism Policy	Plagiarism is the dishonest presentation of others' work as if it were one's own. Duplicate submission is also treated as plagiarism. Depending on the nature of plagiarism, one may fail the assignment or the course. The repeated act of plagiarism will be reported to the University, which may apply additional penalties.
Other Additional Notes	None

(NOTE 3) Class schedule is subject to change

Class Schedule	
Class Number	Content
Class 1	Module 1: Machine Learning Overview and Data Pre-processing Overview: Supervised versus unsupervised learning, Regression versus Classification problems
Class 2	Data cleaning with Pandas: Take care of missing data, encode categorical data; Introduction to sklearn, feature scaling, train-test split
Class 3	Module 2: Regression models Simple Linear Regression, Multiple Linear regression,

Class 4	Polynomial regression and practice exercises
Class 5	Support Vector Machine Regression
Class 6	Support Vector Machine Regression (continued)
Class 7	Decision Tree Regression
Class 8	Practice exercises – Case Studies
Class 9	Module 3: Evaluating Regression Models Mean Square Error, R-Square, Adjusted R-Square, Bias-Variance trade-off
Class 10	Practice exercises – Case Studies
Class 11	Module 4: Classification models Logistic Regression, Importance of sigmoid function
Class 12	K- Nearest Neighbors
Class 13	Module 5: Evaluating Classification Models False positives and false negatives, Confusion Matrix, Accuracy, Precision, and Recall, ROC curve
Class 14	Practice exercises – Case Studies
Class 15	Module 6: Feature Selection and Feature Engineering Regularization
Class 16	Outlier handling, Pipeline
Class 17	Practice exercises – Case Studies
Class 18	Module 7: Unsupervised Learning K- Means Clustering, Hierarchical Clustering

Class 19	Practice exercises – Case Studies
Class 20	Module 8: Dimensionality Reduction Techniques Principal Component Analysis, Linear Discriminant Analysis, Case Studies
Class 21	Practice exercises – Case Studies
Class 22	Module 9: Model Selection and Boosting Model Selection: k- fold cross-validation, grid search; Boosting: XGBoost
Class 23	Practice exercises – Case Studies
Class 24	Practice exercises – Case Studies
Class 25	Practice exercises – Case Studies
Class 26	Module 10: Natural Language Processing Regular Expressions; Introduction to NLTK and Spacy
Class 27	Practice exercises – Case Studies
Class 28	Tokenization, Stemming, Lemmatization, Stop words; Text classification; Sentiment analysis
Class 29	Project: Involving web scraping, data cleaning, and model training/ evaluation (Extra)
Class 30	Project: Involving web scraping, data cleaning, and model training/ evaluation (continued)