COURSE DESCRIPTION FORM

INSTITUTION FAST University of Computer and Emerging Sciences

BS(CS) / BS (SE) / BS(CY) / BS(AI)

PROGRAM (S) TO BE EVALUATED

A. Course Description

(Fill out the following table for each course in your computer science curriculum. A filled out form should not be more than 2-3 pages.)

Course Code	CS-4048 / EE 4047
Course Title	Data Science / Data Science for Engineers
Credit Hours	3
Prerequisites by Course(s) and Topics	Data Structures
Assessment Instruments with Weights (homework, quizzes, midterms, final, programming assignments, lab work, etc.)	Assignments: 12 Quizzes: 8 Project: 5 Mid Term I: 12.5 Mid Term II: 12.5 Final: 50
Course Coordinator	Dr. Rabia Tabassum
URL (if any)	
Current Catalog Description	Data Science is a dynamic and fast-growing field at the interface of Statistics and Computer Science. It is an interdisciplinary field about processes and systems to extract knowledge or insights from data in various forms. This course will introduce students to this rapidly growing field and equip them with some of its basic principles and tools including data collection and integration, data cleaning, data analysis using machine learning, visualization and effective communication. The main focus of these topics with be on understanding and integration of concepts and their application to solving problems.

Textbook (or **Laboratory Manual** for Laboratory Courses)	Lecture Notes Davy Cielen, Arno D. B. Meysman, and Mohamed Ali, Introducing Data Science, Big data, and more, using Python tools, May 2016
Reference Material	Journals: Machine Learning, Pattern Recognition Conferences: ICPR, ICDM, ICML, KDD www.datacamp.com (Students are given free access on many tutorials)
Program Learning Outcomes	

PLO1	Computing Knowledge	Apply knowledge of mathematics, natural sciences, computing fundamentals, and a computing specialization to the solution of complex computing problems.	✔
PLO2	Problem Analysis	Identify, formulate, research literature, and analyse complex computing problems, reaching substantiated conclusions using first principles of mathematics, natural sciences, and computing sciences.	✔
PLO3	Design/Develop Solutions	Design solutions for complex computing problems and design systems, components, and processes that meet specified needs with appropriate consideration for public health and safety, cultural, societal, and environmental considerations.	✔
PLO4	Investigation & Experimentation	Conduct investigation of complex computing problems using research based knowledge and research based methods	✔
PLO5	Modern Tool Usage	Create, select, and apply appropriate techniques, resources and modern computing tools, including prediction and modelling for complex computing problems.	✔
PLO6	Society Responsibility	Apply reasoning informed by contextual knowledge to assess societal, health, safety, legal, and cultural issues relevant to context of complex computing problems..	✔

PLO7	ment and Sustainability	Understand and evaluate sustainability and impact of professional computing work in the solution of complex computing problems	✔
PLO8	Ethics	Apply ethical principles and commit to professional ethics and responsibilities and norms of computing practice.	✔
PLO9	Individual and Team Work	Function effectively as an individual, and as a member or leader in diverse teams and in multi-disciplinary settings.	✔
PLO10	Communication	Communicate effectively on complex computing activities with the computing community and with society at large.	✔
PLO11	Project Managment and Finance	Demonstrate knowledge and understanding of management principles and economic decision making and apply these to one's own work as a member or a team.	✔
PLO12	Life Long Learning	Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological changes.	✔

Course Goals	***Course Learning Objectives:*** should be able to articulate a comprehensive understanding of Data Science, encompassing its fundamental concepts, methodologies, and practical applications. **CLO 2**: Students should be able to design and implement data pre-processing pipelines to clean, transform, and prepare raw data for analysis, ensuring data quality and integrity. **CLO 3**: Students should be able to apply programming languages and open-source tools to manipulate data, develop predictive models, and visualize results effectively. **CLO 4**: Students should be able to evaluate and select appropriate algorithms and tools for specific analytical tasks, considering factors such as scalability, accuracy, and interpretability of results. **CLO 5**: Students should be able to work as a team while integrating important components in data science *Student Outcomes Addressed by the Course (From ABET)* (a) An ability to apply knowledge of computing and mathematics appropriate to the program's student outcomes and to the discipline (b) An ability to analyze a problem, and identify and define the computing requirements appropriate to its solution (f) An ability to communicate effectively with a range of audiences (j) An ability to apply mathematical foundations, algorithmic principles, and computer science theory in the modeling and design of computer-based systems in a way that demonstrates comprehension of the tradeoffs involved in design choices [CS] (k) An ability to apply design and development principles in the construction of software systems of varying complexity [CS]
Relation between CLOs and PLOs	

CLO No.	Domain	Taxonomy level	PLO
1	Cognitive	3	1
2	Cognitive	4	2
3	Cognitive	5	2
4	Cognitive	4	5
5	Cognitive	5	9

Topics Covered in the Course, with Number of Lectures on Each Top
(assume 15-week instruction and one-hour lectures)

1. Topics to be covered:

f Topics	Week	Contact Hours	CLO
Basics of Data Science, Big data, Types of Data, Business Intelligence (BI) vs. Data Science, Motivating Examples, Data Science process	**1**	**2**	**1, 4**
Lab 1: Introduction to Python	**1**	**1**	**4**
Data Preparation: Data Cleaning EDA: Compute Simple Statistics, Simple Visualization, Case Studies	**2**	**2**	**1, 2, 4**
Lab 2: Practical Examples	**2**	**1**	**4**
Introduction to Machine Learning: Basic Concepts of Confusion Matrix for Multi-class	**3**	**3**	**3, 4**
Supervised Classification (kNN, Naïve Bayes (also Naïve Bayes for Text Classification)	**4,5**	**3**	**3, 4**
Mid 1 exam (6th week)			
Decision Trees, Logistic Regression, Neural Networks, Overfitting Reduction, SVM Bias-variance Tradeoff, Methods for Cross Validation (LOOCV, Hold-out, x-fold cross validation)	**7,8**	**6**	**3, 4**
Unsupervised Learning & Feature Extraction (K-means, PCA, SVD, LDA) Associative Learning (Apriori + PCY, multi-hash) Recommendations Systems (Collaborative Filtering: User and Item-based)	**9,10,11**	**9**	**3, 4**

	Mid II exam (12th week)			
	Regression Analysis	**13**	**3**	**2, 3, 4**
	Lab: Data Visualization	**13**	**1**	**4**
	Google Page Rank, LSH and Hyperloglog for counting unique users in large streaming applications	**14**	**3**	**2, 3, 4**
	Ensemble Classifiers	**14**	**3**	**3, 4**
	Introduction to Graph Analytics / Visualization, Hadoop Map Reduce **Data Science Ethics**	**15**	**3**	**2, 3, 4**
	Project Presentations	**16**	**3**	**5**
	Total	**16**	**48**	

Laboratory Projects/Experiments Done in the Course	Yes. Project and Regular Lab Classes once a week			
Programmin Assignments Done in the Course	practically covered through programming assignments.			
Class Time Spent on (in credit hours)	**Theory**	**Problem Analysis**	**Solution Design**	**Social and Ethical Issues**
	40 %	25 %	25%	10%
Oral and Written Communications	Every student is required to submit at least __1___ written reports of typically _8____ pages and to make __1___ oral presentations of typically __10___ minute's duration. Include only material that is graded for grammar, spelling, style, and so forth, as well as for technical content, completeness, and accuracy.			

Instructor Name ____Dr Rabia Tabassum _____

Instructor Signature _____

Date _____17/01/2026_____