COURSE DESCRIPTION FORM

INSTITUTION National University of Computer & Emerging Sciences (FAST-NUCES) Karachi

PROGRAM (S) TO BE EVALUATED MS (Data Science) Fall 2023

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	**DS5003 (Old Code: DS501)**
Course Title	Statistical and Mathematical Methods for Data Science
Credit Hours	3
Prerequisites by Course(s) and Topics	None
Assessment Instruments with Weights (homework, quizzes, midterms, final, programming assignments, lab work, etc.)	Quizzes + Assignments 20% Mid-term Exam 20% Project 10% End-term Exam (Final) 50%
Course Coordinator	Fahad Samad (fahad.samad@nu.edu.pk)
URL (if any)	https://classroom.google.com/c/NjE3NzM1MDA4OTkz
Current Catalog Description	This course is designed to teach the core mathematical concepts required for data science in detail that will later help these graduate students for their research. This course includes some topics from advanced linear algebra, probability and statistics, calculus, optimization methods etc. that are used in data science and machine learning research applications
Textbook (or **Laboratory Manual** for Laboratory Courses)	No specific book because this course is a mix of many diverse topics.
Reference Material	1. Probability and Statistics for Computer Scientists, Michael Baron, 3rd Edition, CRC Press, 2019 2. Probability and Statistics for Data Science Math + R + Data, Latest Edition, Norman Matloff, CRC Press 3. Essential Statistics: Exploring the World through Data, Global Edition, by Robert Gould, Colleen Ryan, and Reecca Wong, Pearson Education

	Limited, 2020 4. Data Science from Scratch First Principles with Python, Joel Grus, 2nd Edition, O'reilly, 2019 5. Basics of Linear Algebra for Machine Learning: Discover the Mathematical Language of Data in Python, Latest Edition, Jason Brownlee 6. Linear Algebra and its Applications, Global Edition, David C. Lay and Steven R. Lay, 6th Edition, Pearson Education, 2022 7. Essential Math for Data Science: Take Control of Your Data with Fundamental Linear Algebra, Probability, and Statistics, Thomas Nield, First Edition, O'reilly, 2022 8. Practical Linear Algebra for Data Science From Core Concepts to Applications Using Python (Mike X Cohen), 1st Edition, O'reilly, 2022 9. Data Science and Machine Learning: Mathematical and Statistical Methods by Dirk P. Kroese, Zdravko I. Botev, Thomas Taimre and Radislav Vaisman 7. Information Theory, A tutorial Introduction, James V. Stone, 1st Edition, Sebtel Press, 2015 8. Information Theory, Inference, and Learning Algorithms, David J. C. MacKay, 4th Printing, Cambridge University Press, 2003 9. Convex Optimization, Boyd and Vandenberghe, Cambridge University Press, 2004 10. Mathematics for Machine Learning, Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong, Cambridge University Press, 2020 ISBN-13: 978-1108455145 11. Tutorials, Handouts, and Scientific Research Papers
Course Goals	The course offers an introduction to the fundamental mathematical practices used in data science. The main goal of the course is to build a good practical knowledge and a mathematical understanding of the methods that are used to analyze modern datasets. .
	A. Course Learning Outcomes (CLOs) After completion of the course, the students shall be able to: On successful completion of this course students will be able to: 1. Demonstrate understanding of basic mathematical concepts in data science, relating to linear algebra, probability, statistics, calculus and optimization. **(Bloom's Taxonomy – C2 Understand)** 2. Employ methods related to these concepts in a variety of data science applications. **(Bloom's Taxonomy – C3 Apply)** 3. Apply logical thinking to problem-solving in context and use appropriate technology to aid problem-solving and data analysis. They will be able to analyze datasets using a modern programming language

such as Python. **(Bloom's Taxonomy – C3 Apply)**

B. Program Learning Outcomes (Graduating Attributes)

For each attribute below, indicate whether this attribute is covered in this course or not. Leave the cell blank if the enablement is little or non-existent.

PLO 1	Academic Education	To prepare graduates as computing professionals
PLO 2	Knowledge for Solving Computing and Data Science Problems	Apply knowledge of computing fundamentals, knowledge of a computing specialization, and mathematics, science, and domain knowledge appropriate for the computing and data science specialization to the abstraction and conceptualization of computing and data science models from defined problems and requirements.
PLO 3	Problem Analysis	Identify, formulate, research literature, and solve complex computing problems reaching substantiated conclusions using fundamental principles of mathematics, computing sciences, and data science disciplines.
PLO 4	Design/ Develop Solutions	Design and evaluate solutions for complex computing and data science problems, and design and evaluate systems, components, or processes that meet specified needs with appropriate consideration for public health and safety, cultural, societal, and environmental considerations.
PLO 5	Modern Tool Usage	Create, select, adapt and apply appropriate techniques, resources, and modern computing and data science tools to complex computing activities, with an understanding of the limitations.
PLO 6	Individual and Team Work	Function effectively as an individual, and as a member or leader in diverse teams and in multi-disciplinary settings.
PLO 7	Communicatio n	Communicate effectively with the computing and data science

		community and with society at large about complex computing and data science activities by being able to comprehend and write effective reports, design documentation, make effective presentations, and give and understand clear instructions.
PLO 8	Computing Professionalism and Society	Understand and assess societal, health, safety, legal, and cultural issues within local and global contexts, and the consequential responsibilities relevant to professional computing and data science practices.
PLO 9	Ethics	Understand and commit to professional ethics, responsibilities, and norms of professional computing and data science practice.
PLO 10	Life Long Learning	Recognize the need, and have the ability, to engage in independent learning for continual development as a computing and data science professional.

C. Mapping of CLOs on PLOs

(CLO: Course Learning Outcome, PLOs: Program Learning Outcomes)

		PLOs									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
CLOs	1										
	2										
	3										

<table>
<tr><td rowspan="16">Topics Covered in the Course, with Number of Lectures on Each Topic (assume 15-week instruction and three-hour lecture per week)</td><td>Week #</td><td>Topics</td></tr>
<tr><td>1</td><td>Introduction: Need for Statistics and Mathematics for Data Science</td></tr>
<tr><td>2-3</td><td>Probability: Basic Overview, axioms of probability, conditional probability, Random variables, multivariate random variables</td></tr>
<tr><td>4-5</td><td>Expectation, Independence, Random Processes, maximum likelihood estimation;</td></tr>
<tr><td>6-7</td><td>Statistics: Concentration Bounds, Law of large numbers, Central limit theorem, Minimum Mean Square Error Estimation, Confidence intervals</td></tr>
<tr><td>8</td><td>Midterm Exam</td></tr>
<tr><td>9-10</td><td>Hypothesis Testing</td></tr>
<tr><td>11</td><td>Linear Algebra: Vector Spaces, Projections, Least Regression, Linear Transformations</td></tr>
<tr><td>12</td><td>Eigen-decomposition, Power Method</td></tr>
<tr><td>13</td><td>Principal Component Analysis and Singular Value Decomposition</td></tr>
<tr><td>14</td><td>Linear functions and Least squares</td></tr>
<tr><td>15</td><td>Calculus/Optimization: Introduction to Optimization, Optimization Techniques, Gradient Descent and Coordinate Descent</td></tr>
<tr><td>16</td><td>Calculus/Optimization: Matrix Calculus with Langrage Multipliers and Convex Optimization</td></tr>
<tr><td>17</td><td>Final Exam</td></tr>
<tr><td>Lecture and Attendance Policy</td><td colspan="4">Content will be covered via class lectures, home assignments, and project. Students are expected to attend all of the classes, be ready to begin the class on time and not leave before the designated time. However, there is a margin of 20% in attendance for emergency reasons.</td></tr>
<tr><td>Laboratory Projects/Experiments Done in the Course</td><td colspan="4">Students will be given assignments related to the theory concepts they learn in classroom lectures. A project for the application of studied concepts will also be assigned.</td></tr>
<tr><td>Programming Assignments Done in the Course</td><td colspan="4">A few programming labs are optionally given to apply the key concepts of mathematics in data science.</td></tr>
<tr><td rowspan="2">Class Time Spent on (in % of credit hours)</td><td>Theory</td><td>Problem Analysis</td><td>Solution Design</td><td>Social and Ethical Issues</td></tr>
<tr><td>20%</td><td>40%</td><td>40%</td><td>Not Applicable</td></tr>
</table>

Oral and Written Communications	Every student group is required to submit at least 01 written report of typically 08 to 10 pages and to make 01 oral presentations of typically 15 minute's durations. Include only material that is graded for grammar, spelling, style, and so forth, as well as for technical content, completeness, and accuracy.
Late Submission Policy	Deadlines are meant to be strictly followed. Any late submission (without any valid reason and justification/ evidence) will be penalized. The penalty will be 50%. Any delay of more than a week would mean ZERO credit in that particular assessment (assignments, labs, project).
Plagiarism Policy	Any copied/plagiarized work (assignment, lab or project) will get ZERO credit. In documents where you need to refer to other's work, please give them due credit by citing the original work.

Instructor Name ___**M. Shahid Ashraf**___

Instructor Signature _____

Date ____**29/08/2023**_____