HABIB UNIVERSITY

CS/CE 457/464-L1 Data Science

"Without data you are just another person with an opinion"- W. Edwards Deming

Fall 2024
Monday – Asynchronous session
Friday – Online 8:30AM - 9:45AM

Instructor: Zeehasham Rasheed
Office Hours: After class or by appointment
Course LMS URL: https://hulms.instructure.com/courses/3929
Course Prerequisites: CS 355, MATH 310
Content Area: This course meets requirements for CS Elective. It can be taken as a free elective, SSE elective, or to fulfill the requirements for a CS minor or a major.

I. Rationale:

The requirements for skill set, technology and business are changing and become an essential part in order to be successful. No matter where you are in your career or what field you work in, you will need to understand the language of data. With this Data Science course, you learn hands-on data science today and are able to apply it right away.

II. Course Aims and Outcomes:

Aims
Data science is a cutting-edge field that combines programming, mathematics, statistics and technology to analyze data. Data has transformed every industry, and the ability to identify trends and insights is a highly desirable skill. When you finish this course, you will have the skills and tools to leverage statistics and programming to make predictions, optimize outcomes, and help guide business decisions using data.

I believe data fluency helps people succeed and you have just taken the first step in your journey as a data enthusiast and expert.

III. Computer Science Program's Education Objectives

The Computer Science program at Habib University aims to produce competent computer scientists who

1. Have strong foundational knowledge of mathematics and computer science, and the accompanying skills both in breadth and in depth, to position themselves equally well in the Information Technology industry, as technology entrepreneurs and/or in graduate programs in Computer Science or other technical and scientific fields.
2. Have a hands-on approach to self-learning and research, and will continually update their knowledge, skills and technical know-how.

3. Will be able to assess the societal, cultural, social, religious, legal, environmental, local, and global impact of their actions and will choose an ethical course of action in their professional, personal, and daily lives.
4. Will be able to effectively communicate and collaborate with people from diverse backgrounds and in a variety of settings

IV. Computer Science Program's Learning Outcomes (PLOs)

Upon graduation, students will have the following abilities:

1. Analysis: analyze a given situation and reduce it to one or more problems that can be solved via computer intervention.
2. Design: design one or more computer-based solutions of a given problem and select the solution that is the best under the circumstances.
3. Programming: program a given solution in a variety of programming languages belonging to different paradigms.
4. Implementation: design and implement software systems of varying complexity.
5. Tools: work with the latest tools that support development, e.g., IDEs, version control systems, debuggers, profilers, and continuous build systems.
6. Self-learning: research, learn, and apply requirements needed to implement a solution for a given high level problem description.
7. Ethics and Awareness: foresee both impact and possible ramifications of computing practices
8. Communication and Teamwork: work effectively in inter-disciplinary teams

V. Data Science Learning Outcomes:
By the end of this course, students will:

a) Improve their skills in data manipulation, data visualization, statistics, machine learning and more.
b) Learn how to gain insights into complex data through descriptive and inferential statistics and data visualization.
c) Explore techniques that are useful for drawing conclusions and predicting outcomes.
d) Explore supervised and unsupervised learning algorithms and best practices for applying machine learning.
e) Increase their ability to use Python programming to analyze and interpret complex data.
f) Learn about an ecosystem of useful and powerful tools for doing Data Science with Python
g) Acquire new skills fast in this course that combines video content with immediate hands-on exercises and projects, exploring real data sets covering the entire data science workflow.
h) Get an initial push towards Data Science along with technical skills to launch a successful career in Data Science.

VI. Specific Course Learning Outcomes (CLOs):
By the end of this course, students will:

1. Improve their skills in data manipulation, data visualization, statistics, machine learning and more.
2. Learn how to gain insights into complex data through descriptive and inferential statistics and data visualization.
3. Implement and evaluate Data Science and Machine Learning algorithms that are useful for drawing conclusions and predicting outcomes.

4. Explore Data Science related use cases and best practices for applying machine learning.

Description	Learning Outcomes	Cog Level
CLO 01	Improve their skills in data manipulation, data visualization, statistics, machine learning and more.	1, 3
CLO 02	Learn how to gain insights into complex data through descriptive and inferential statistics and data visualization.	1, 2
CLO 03	Implement and evaluate Data Science and Machine Learning algorithms that are useful for drawing conclusions and predicting outcomes.	4, 5
CLO 04	Explore Data Science related use cases and best practices for applying machine learning.	3, 6

Cognitive Objectives (Cog)
1. Knowledge - Remembering or recalling information.
2. Comprehension - The ability to obtain meaning from information.
3. Application - The ability to use information.
4. Analysis - The ability to break information into parts to understand it better.
5. Synthesis - The ability to put materials together to create something new.
6. Evaluation - The ability to check, judge, and critique materials.

A. Mapping Assessments to CLOs

Assessment	CLO 01	CLO 02	CLO 03	CLO 04
HW 01	X			
HW 02	X	X		
HW 03	X	X		
HW 04		X		
HW 05		X		
HW 06	X	X		
HW 07			X	X
HW 08			X	X
HW 09			X	X
HW 10			X	X
HW 11			X	X
HW 12			X	X
HW 13			X	X
Midterm	X	X	X	X
Final	X	X	X	X

B. Mapping CLOs to PLOs

	PLO 01	PLO 02	PLO 03	PLO 04	PLO 05	PLO 06
CLO 01	X					
CLO 02		X				X
CLO 03			X	X	X	
CLO 04						X

III. Format and Procedures:
a) Synchronous session of 1 Hour 15 mins (on Zoom) and 1 Hour 15 minutes recorded (asynchronous session) shared via Canvas per week.
b) Students will complete learning assignments and homework assigned to them each week.
c) There will be one remote class per week between the instructor and the enrolled students.
d) Students are expected to work independently with high degree of self motivation. The University's rules on attendance, conduct, and academic honesty apply.

IV. Course Requirements:
Prior experience in basic Python programming and basic Statistics is required for understanding the material and hands-on exercises.

1. **Class participation policy:**
   Students are encouraged to keep contact with instructor on a regular basis, particularly as they go through the early weekly assignments. Even if students do not require any help related to content or assignments, they are encouraged to post their progress updates. Students are also welcomed to send their suggestions, likes and dislikes about this course.

2. **Course readings:**
   a) Required text: Weekly assigned video lectures and exercise accessible through HU portal and DataCamp which is an online course repository.
   b) Any online material related to course content or topics covered. Instructor will provide supplementary material if needed.

V. Grading Procedures: Grades will be based on:

GRADING SCALE		
LETTER GRADE	GPA POINTS	PERCENTAGE
A+	4.00	[95, 100]
A	4.00	[90, 95)
A-	3.67	[85, 90)
B+	3.33	[80, 85)
B	3.00	[75, 80)
B-	2.67	[70, 75)
C+	2.33	[67, 70)
C	2.00	[63, 67)
C-	1.67	[60, 63)
F	0.00	[0, 60)

Weekly Homework Assignments (50%)	Total 13 assignments. One assignment with lowest score will be dropped (if needed). All weekly assignments are due next Friday 11:59pm. Late submission after 24 hours will not be accepted. You will loose 0.1 points every hour after due date.
Midterm Exam (20%)	This will be a lab based practical exam in the class. Duration will be 2 hours.
Final Exam (25%)	This will be a lab based practical exam in the class. Duration will be 2 hours.
Class Participation (5%)	Based on Attendance and additional weekly tasks. Attendance will be marked during Online class at any time.

VI. Attendance Policy:

Habib University requires that all freshmen and sophomores must maintain at least 85% attendance and all juniors and seniors must maintain at least 75% attendance for each class in which they are registered. Non-compliance with minimum attendance requirements will result in <u>automatic failure</u> of the course and may require the student to repeat the course when next offered. This policy is at a minimum. Departments, schools, and individual faculty members <u>may alter this policy to include stronger attendance requirements</u> and/or implement them for all levels of students. It is the responsibility of the student to keep track of their own attendance and speak with their faculty member or the Office of the Registrar for any clarification.

VII. Accommodations for students with disabilities

In compliance with the Habib University policy and equal access laws, I am available to discuss appropriate academic accommodations that may be required for student with disabilities. Requests for academic accommodations are to be made during the first two weeks of the semester, except for unusual circumstances, so arrangements can be made. Students are encouraged to register with the Office of Academic Performance to verify their eligibility for appropriate accommodations.

VIII. Inclusivity Statement

We understand that our members represent a rich variety of backgrounds and perspectives. Habib University is committed to providing an atmosphere for learning that respects diversity. While working together to build this community we ask all members to:
- share their unique experiences, values and beliefs
- be open to the views of others
- honor the uniqueness of their colleagues
- appreciate the opportunity that we have to learn from each other in this community
- value each other's opinions and communicate in a respectful manner
- keep confidential discussions that the community has of a personal (or professional) nature
- use this opportunity together to discuss ways in which we can create an inclusive environment in this course and across the Habib community

IX. Office hours:

Office hours have been scheduled, circulated, and posted. During these hours the course instructor will be available to answer questions or provide additional help. Every student enrolled in this course <u>must meet individually with the course instructor during course office hours</u> at least once during the semester. The first meeting should happen within the first five weeks of the semester but must occur before midterms. Any student who does not meet with the instructor may face a grade reduction or other penalties at the discretion of the instructor and will have an academic hold placed by the Registrar's Office.

X. Academic Integrity

Each student in this course is expected to abide by the Habib University Student Honor Code of Academic Integrity. Any work submitted by a student in this course for academic credit will be the student's own work.

[*Optional:* For this course, collaboration is allowed in the following instances: *list instances.*]

Scholastic dishonesty shall be considered a serious violation of these rules and regulations and is subject to strict disciplinary action as prescribed by Habib University regulations and policies. Scholastic dishonesty includes, but is not limited to, cheating on exams, plagiarism on assignments, and collusion.

PLAGIARISM: Plagiarism is the act of taking the work created by another person or entity and presenting it as one's own for the purpose of personal gain or of obtaining academic credit. As per University policy, plagiarism includes the submission of or incorporation of the work of others without acknowledging its provenance or giving due credit according to established academic practices. This includes the submission of material that has been appropriated, bought, received as a gift, downloaded, or obtained by any other means. Students must not, unless they have been granted permission from all faculty members concerned, submit the same assignment or project for academic credit for different courses.

CHEATING: The term cheating shall refer to the use of or obtaining of unauthorized information in order to obtain personal benefit or academic credit.

COLLUSION: Collusion is the act of providing unauthorized assistance to one or more person or of not taking the appropriate precautions against doing so. All violations of academic integrity will also be immediately reported to the Student Conduct Office.

You are encouraged to study together and to discuss information and concepts covered in lecture and the sections with other students. You can give "consulting" help to or receive "consulting" help from such students. However, this permissible cooperation should never involve one student having possession of a copy of all or part of work done by someone else, in the form of an e-mail, an e-mail attachment file, a diskette, or a hard copy.

Should copying occur, the student who copied work from another student and the student who gave material to be copied will both be in violation of the Student Code of Conduct.

During examinations, you must do your own work. Talking or discussion is not permitted during the examinations, nor may you compare papers, copy from others, or collaborate in any way. Any collaborative behavior during the examinations will result in failure of the exam, and may lead to failure of the course and University disciplinary action. Penalty for violation of this Code can also be extended to include failure of the course and University disciplinary action.

XII. Week-wise Schedule (Tentative and subject to change)

Week	Topic(s)	Assignment(s)	Remarks
Week 1 **August** **19 – 23**	Introduction to Data Science Pandas Library in Python		**First day of Classes:** **August 19** Sessions
Week 2 **August** **26– 30**	Exploratory Data Analysis (EDA)		Sessions
Week 3 **September** **02 – 06**	Foundation of Inference Statistics		Sessions
Week 4 **September** **09 – 13**	SQL Data Analysis		Sessions
Week 5 **September** **16 – 20**	NoSQL Data Analysis		Sessions
Week 6 **September** **23 – 27**	Introduction to Machine Learning Supervised Learning using Regression Analysis		Sessions
Week 7 **Sept-Oct** **30 – 04**	Midterm Exam		
Week 8 **October** **07 – 11**	Introduction to Machine Learning Supervised Learning using Regression Analysis		Sessions
Week 9 **October** **14 – 18**	Supervised Learning using Classification Analysis		No Class
Week 10 **October** **21 – 25**	Unsupervised Learning using Cluster Analysis		Sessions
Week 11 **Oct - Nov** **28 – 01**	Time Series Analysis		Sessions
Week 12 **November** **04 – 08**	Natural Language Processing (NLP)		Sessions

Week	Topic		Sessions
Week 13 **November** **11 – 15**	Introduction to Deep Learning		Sessions
Week 14 **November** **18 – 22**	Big Data Processing using Apache Spark		Sessions
Week 15 **November** **25 – 29**	Dash Python		Sessions
Week 16 **December** **02 – 06**	Reading Days		
Week 17 **December** **09 – 13**	Final Exam		Sessions

XI. Additional Resource Readings

Additional material will be provided by the instructor if needed.