NATIONAL UNIVERSITY
of Computer & Emerging Sciences, Lahore

Department of Computer Science

CS481 – Data Science

FALL 2020

Instructor Name: Dr. Irfan Younas **TA Name (if any):**
Email address: irfan.younas@nu.edu.pk **Email address:**
Office Location/Number: C-146 **Office Location/Number:**
Office Hours: Wed 3:00-4:00 PM, Friday 2:00 to 3:00pm **Office Hours:**

Course Information

Program: BS **Credit Hours:** 3 **Type:** Elective
Pre-requisites (if any): Programming competence, Discrete Maths, Linear Algebra, Probabilty& Statistics

Course Website (if any) : SLATE
Class Meeting Time:
Section B: Mon, Wed 11:00 AM– 12:20 PM, **Section A**: Mon, Wed 12:30 PM – 1:50 PM
Class Venue: Online/CS-06

Course Description/Objectives/Goals:

Data Science is the study of the generalizable extraction of knowledge from data. Being a data scientist requires an integrated skill set spanning computer science, mathematics, statistics, and domain expertise along with a good understanding of the art of problem formulation to engineer effective solutions. The goal of this course is to teach students to answer questions with data. To do this, we will learn the necessary skills to manage and analyze data with case studies. In this course student learn concepts such as data collection and integration, exploratory data analysis, statistical inference and modeling, machine learning, and high-dimensional data analysis.

Course Learning Outcomes (CLOs):		
At the end of the course students will be able to:	**Domain**	**BT* Level**
understand the basics of Data Science,		
prepare and wrangle the data for analysis,		
perform exploratory data analysis to investigate data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations,		

understand and apply machine learning algorithms to gain insight from the data		

* BT= Bloom's Taxonomy, C=Cognitive domain, P=Psychomotor domain, A= Affective domain.

Bloom's taxonomy Levels: 1. Knowledge, 2. Comprehension, 3. Application, 4. Analysis, 5. Synthesis, 6. Evaluation

Textbook(s) /Supplementary Readings:

There is no standard one "textbook" for this course. The following book will be used as a primary text to guide some of the discussions, but it will be heavily supplemented with lecture notes and reading assignments from other sources.

Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk From The Frontline. O'Reilly. 2014. ISBN 978-1-449-35865-5.

Additional references and books related to the course:

Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. Mining of Massive Datasets. v2.1, Cambridge University Press. 2014. (Free online.)

Jiawei Han, Micheline Kamber and Jian Pei. Data Mining: Concepts and Techniques, Third Edition. Morgan Kaufmann Publishers. 2012. ISBN 978-0-12-381479-1.

Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. MIT Press. 2013. ISBN 0262018020. (Online info available here.)

Foster Provost and Tom Fawcett. Data Science for Business: What You Need to Know about Data Mining and Data-analytic Thinking. O'Reilly 2013. ISBN 978-1-449-36132-7.

Tentative Weekly Schedule

Week	Topics to be covered	Readings	Assignments/ Projects?
1	• Introduction to Data Science • Introduction to Machine Learning: Supervised, unsupervised learning • Intro to Linear Regression: Model Representation for single variable Linear Regression,		Home work-1
2	• Gradient Descent • Multiple Variable Linear Regression and Polynomial Regression, Normal Equation,		Quiz 1, Assignment 1, Project

	Logistic Regression		
3	• Logistic Regression • Regularization		HOMEWORK 2, Assignment 2
4 -5	• Advice for applying ML , ML System Design, model selection, parameters optimization, learning curves, model's underfitting and overfitting detection and solution, evaluation measures		Quiz 2
5-6	• Support Vector Machines (linear, SVM with Kernels) • Clustering Algorithms		Assignment# 3
7-8	• Features engineering, Dimensionality Reduction (Principal Component Analysis)		
8-10	• Deep Learning topics o Neural networks o Object detection algorithms o Other advanced topics		Quiz 3, Assignment 4,
11-12	• Introduction to Statistics basics, descriptive stat, statistical inference • Data Wrangling o Data cleaning, data reshaping, data preprocessing		Assignment 3,
13	• Exploratory Data Analysis o Basic tools (plots, graphs and summary statistics) of EDA • Statistical measures for analysis		Homework 3,
13-14	• Exploratory Data Analysis o Use of R-squared and other measures for analysis o Model transformations		Quiz 4
15	• Data visualization		Homework 4

(Tentative) Grading Criteria

Quizzes	10%
Assignments/Homeworks/Project	*15* - 25%
Midterms	25-30%
Final Exam	40 - 45%
Total:	**100 %**

Course Policies

- *Course outline may change 10-20% as we proceed in the semester*
- ***Grading scheme: Relative***
- Depending on the situation of COVID 19, this weightage of midterms can be reduced and added in assignments/homeworks/project.
- *Weightage of other evaluations can also be adjusted if needed.*
- *Assignment deadlines for assignment and Project are hard.*
- *NO Cell Phone usage in class, they must be turned off at all times.*
- *There will be no retake of quizzes or exams.*
- ***Integrity in the assignments/quizzes is expected; otherwise result would be an F grade in the course or may be the case is forwarded to Disciplinary committee.***
- *Attendance MUST be ensured according to the University policy to avoid disqualification.*