| Logo Fast | **NATIONAL UNIVERSITY**  **of Computer & Emerging Sciences, Lahore** |
| --- | --- |

Department of Computer Science

**CS4048 – Data Science**

**Spring 2022**

**Instructor Name:**  Amir Iqbal **TA Name (if any):**

**Email address:** amir.iqbal@nu.edu.pk **TA Email address:**

**Office Location/Number:** C-136 **Office Location/Number:**

**Office Hours:** Tue - Thur 3:00-4:00 PM  **Office Hours:**

**TA Information:**

**8A - Abdul Manan Khan l181066@lhr.nu.edu.pk**

**8B - Kamran Ahmad l181120@lhr.nu.edu.pk**

**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)**: Google Classroom

**Class Meeting Time:**

**Section B**: Mon, Wed 11:30 AM– 1:00 PM, **Section A**: Mon, Wed 1:00 PM – 2:30 PM

**Class Venue:** E&M 3

**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, students 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:**

Geron Aurelien. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems

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.

Joel Grus. Data Science from Scratch, 2nd Edition

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

**Tentative Weekly Schedule**

| Week | Topics to be covered | Readings | Assignments/Projects? |
| --- | --- | --- | --- |
| 1 | Introduction to Data Science  Statistical inference  Modeling  The data science process |  |  |
| 2-3 | Exploratory data analysis  Data Visualization  Basic tools (plots, graphs and summary statistics) of EDA  Data Wrangling  Data cleaning, data reshaping, data preprocessing |  |  |
| 4 | Data Collection  Measurement scales, missing values, outliers |  |  |
| 5 | Gradient Descent - Linear Regression - single variable |  |  |
| 6 | Mid 1 |  |  |
| 7-9 | Gradient Descent - Linear Regression - multiple variable  Logistic Regression, Regularization  Neural Networks, Representation & Backpropagation  Advice for applying ML , ML System Design, model selection, parameters optimization, learning curves, model’s underfitting and overfitting detection and solution, evaluation measures  Support Vector Machines  Non-linear classification with SVM Kernels |  |  |
| 10-11 | Features engineering, Dimensionality Reduction (Principal Component Analysis)  Feature Selection  Recommendation engines  Anomaly detection |  |  |
| 12 | Mid 2 |  |  |
| 13 | Time Series Forecasting |  |  |
| 14-15 | Big Data  Map Reduce (Hadoop)  Spark |  |  |
| 15- 16 | Deep Learning topics  Training Deep Neural Networks  Convolution Neural Networks |  |  |

**Grading Criteria**

Quizzes/*Assignments/Homeworks/Project* 40%

*Midterms* 20%

*Final Exam* 40%

**Course Policies**

* *Course outline may change as we proceed in the semester*
* ***Grading scheme: Absolute***

**Attendance**

All students are expected to attend all lectures from beginning to end. Partial or full absence from a lecture may hamper chances of securing good grades.

**Exams**

Exams will be closed-book and closed-notes. The syllabus for the final exam will be comprehensive.

**Office Hours**

Students are encouraged to take full advantage of the instructor's office hours. Any doubts regarding concepts covered in class or any questions regarding quizzes, projects, etc. may be clarified during office hours. In case a student is not able to make it during office hours, he/she may schedule an appointment with the instructor for another time slot.

**Quizzes**

Quizzes may be announced or unannounced. A quiz will usually be about 5 – 10 minutes long and it may be given anytime during the lecture.

**Reading Material**

Students are encouraged to finish the assigned readings BEFORE the lecture. This is likely to improve lecture comprehension and class participation.

**Revision of Grades**

Students can contest their grades on quizzes and project deliverables ONLY within a week of the release of grades. Exams will be available for review according to university policies.

**Unfair Means**

Students are expected to demonstrate the highest degree of moral and ethical conduct. Any student caught cheating, copying, plagiarizing, or using any other unfair means will be strictly dealt with per university policies.