



Department of Data Science

Introduction to Data Science

FALL 20XX

Instructor Name: Muhammad Saif ul Islam

TA Name (if any):

Email address:

Email address:

Office Location/Number:

Office Location/Number:

Office Hours:

Office Hours:

Course Information

Program: BSDS

Credit Hours: 3

Type: Core

Pre-requisites (if any): Programming competence,

Course Website (if any) :

Class Meeting Time:

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:

- Jure Leskovek, AnandRajaraman 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
1	<ul style="list-style-type: none">• Introduction to Data Science<ul style="list-style-type: none">o Role of a data scientisto Data Science pipelineo Applications
2-3	<ul style="list-style-type: none">• Introduction to Python<ul style="list-style-type: none">o Python Basics: basic data types, container types, variables, control structures, functions & procedures, classes & instances, modules & packages, exceptions, files & standard library, etc.o Data Manipulation using Numpy• Other popular Data Science modules: Pandas, Matplotlib, Tools etc,
4	<ul style="list-style-type: none">• Data Collectionand Representation<ul style="list-style-type: none">o Collecting publically available dataseto Creating own dataset (Web scrapping)o Types of features (categorical, nominal, continuous, etc.)o Introduction to indexing and slicing in data frames (Pandas)
5	<ul style="list-style-type: none">• Data Wrangling<ul style="list-style-type: none">o Working with duplicate, missing data, and outlierso Working with data typeso Label Encoding, Features encoding and Conversiono Data cleaning, data reshaping
6	<ul style="list-style-type: none">• Statistical foundations for Data Analysis<ul style="list-style-type: none">o Descriptive statistics, Descriptive analysiso Statistical inferenceo Distributions, Variance, co-variance, correlation
7-8	<ul style="list-style-type: none">• Introduction to Exploratory Data Analysis (EDA)<ul style="list-style-type: none">o Basic tools (plots, graphs and summary statistics) of EDA



	<ul style="list-style-type: none"> o Grouping and merging in data frames o Plotting single/multiple line charts, subplots o Distributions and frequency plotting (histograms and bar chart) o Correlation analysis (Scatter charts) o Data Spread, Range, and outlier analysis (Box-and-whisker plots) o Model transformations
9	<ul style="list-style-type: none"> • Spatial Data, geospatial data, time-oriented data, textual data.
10-11	<ul style="list-style-type: none"> • Machine Learning (Regression) <ul style="list-style-type: none"> o Introduction to Machine Learning and its types o Linear Regression/Multiple Variable Linear Regression o R-Squared; Coefficient of determination, Mean Squared Error
12- 13	<ul style="list-style-type: none"> • Machine Learning (Classification) <ul style="list-style-type: none"> o k-Nearest Neighbors (k-NN) o Basic idea of perceptron learning. o Advice for applying ML , ML System Design, model selection, parameters optimization, learning curves, model's underfitting and overfitting detection and solution, evaluation measures, o Clustering Algorithms
14	<ul style="list-style-type: none"> • Feature Generation and Feature selection <ul style="list-style-type: none"> o Overview of Feature Generation (Feature extraction), o Overview of Feature Selection approaches o Feature Engineering o Introduction to Filter-based and Wrapper-based approaches;
15	<ul style="list-style-type: none"> • Mining Social Network Graphs • Data Science and Ethical Issues • Advanced Topics

(Tentative) Grading Criteria*

Quizzes	10%	<i>Assignments/Project</i>	20%
Midterms	30%	<i>Final Exam</i>	40%

* Tentative. Subject to change.

Course Policies

- *Grading Scheme: **Absolute***
- *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***



Edit with WPS Office

committee.

- *Attendance MUST be ensured according to the University policy to avoid disqualification.*



Edit with WPS Office