

Data Science					
Credit Hours:	3	Course Code:	COSC-3109	Prerequisites:	None
Course Learning Outcomes (CLOs):					
At the end of the course, the students will be able to:				Domain	BT Level*
1. Understand data science and basic concepts related to data science				C	1
2. Identify or Explain tools/techniques or algorithms to solve the data science problems.				C	2
3. Manage the data science cycle including getting data, cleaning data, data preprocessing, and data visualization.				C	3
4. Practice data science issues in statistical computing using programming language such as R and Python.				C	4
* BT= Bloom's Taxonomy, C=Cognitive Domain, P=Psychomotor Domain, A=Affective Domain					
Course Contents:					
Introduction to Data Science, Data Science Life cycle & Process (Asking Right Questions, Obtaining Data, Understanding Data, Building Predictive Models, Generating Visualizations), For Building Data Products, Introduction to Data (Types of Data and Data Sets), Data Quality (Measurement and Data Collection Issues), Data pre-processing Stages (Aggregation, Sampling, Dimensionality Reduction, Feature subset selection, Feature creation etc.), Algebraic & Probabilistic View of Data, Introduction to Python Data Science Stack (Python, Numpy, Pandas, Matplotlib), Relational Algebra & SQL, Scraping & Data Wrangling (assessing, structuring, cleaning & munging of data), Basic Descriptive & Exploratory Data Analysis using Plotly and Matplotlib, Introduction to Text Analysis (Stemming, Lemmatization, Bag of Words, TF-IDF), Introduction to Prediction and Inference (Supervised & Unsupervised) Algorithms, Introduction to Scikit Learn, Bias-Variance Tradeoff, Model Evaluation & Performance Metrics (Accuracy, Contingency Matrix, Precision-Recall, F1-Score, etc.), Introduction to Map-Reduce paradigm, Introduction to R, Reading data into R, Data frames, Basic and advanced plots for Data Visualization.					
Teaching Methodology:					
Lectures, Written Assignments, Practical labs, Semester Project, Presentations.					
Course Assessment:					
Mid Term Exam, Home Assignments, Quizzes, Project, Presentations, Final Exam.					
Reference Material:					
1. Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications by Laura Igual and Santo Segui, Springer, 1st Edition, 2017; ISBN 978-3-319-50016-4 2. Python Data Science Cookbook by Gobi Subramanian, Packt Publishing, 1st Edition, 2015; ISBN 978-1-78439-640-4 3. Data Science from Scratch by Joel Grus, O'Reilly Media, 1st Edition, 2015; ISBN 978-1-491-90142-7 4. Data Mining and Analysis: Fundamental Concepts and Algorithms by Muhammad J. Zaki and Wagner Meira, Jr, Cambridge University Press, 1st Edition, 2014; ISBN 978-0-521-76633-3					