CSE 462 - Introduction to Data Science
Syllabus

General Information

Course Number	**CSE 462**
Credit Hours	3 (Theory Credit Hour = 3)
Prerequisite	1. Probability and Statistics 2. Databases 3. Basic Programming
Course Coordinator	Not Specified

Course Objectives

Data Science is the study of the generic extraction of knowledge from data. Being a data scientist requires an integrated skill set spanning mathematics, statistics, machine learning, databases and other branches of computer science along with a good understanding of the craft of problem formulation to engineer effective solutions. This course will introduce students to this rapidly growing field and equip them with some of its basic principles and tools as well as its general mindset. Students will learn concepts, techniques and tools they need to deal with various facets of data science practice, including data collection and integration, exploratory data analysis, predictive modeling, descriptive modeling, data product creation, evaluation, and effective communication. The focus in the treatment of these topics will be on breadth, rather than depth, and emphasis will be placed on integration and synthesis of concepts and their application to solving problems. To make the learning contextual, market oriented content, course content will be more practically rather than theoretical.

Catalog Description

CSE 462

Course Content

Session No.	Week No.	Topics	Suggested Readings (Chapters)
01-02	1	Introduction to Data Science (What & Why) Applications - Data Science Prerequisites of Data Science Data Scientist's Work Life Cycle - Data Science Python (Why?)	Chap 1 [Davy Cielen]
03-04	2	Basics of Python Python Data Structures (Lists, Dictionaries, Tuples, Sets) Python Numpy	Handouts
05-06	3	Visualization with Matplotlib - Line Charts, Bar Charts, Pie Charts Data Manipulation using Pandas	Handouts
07-08	4	Data Preprocessing techniques Feature Transformations Missing Values in python - Discovering what's missing - Filling in missing data - Counting missing values - Filtering out missing values	Handouts
09-10	5	Data Visualization Exploratory Data Analysis	Handouts
First Mid Exams			

13-14	7	Descriptive Statistics - Mean, Mode, Median - Standard Deviation, Variance	Handouts
15-16	8	Machine Learning - What is ML? Why ML? - Supervised vs. Unsupervised - ML Applications / Examples - Overfitting and Under-fitting	Chap: 5 [Jake]
17-18	9	Importing Data (Practical) - Plain Text Files - CSV, TSV Files - Excel Files - Scrapping web data using Beautiful Soup Library - Automate download files in python using https requests - Using Twitter API	Handouts
19-20-21-22-23-24	10-11-12	Linear Regression - Equation, Slope, Intercept - Intro Gradient Descent (What & Why) - Calculating RSS, RSE, MSE - R^2 value calculating - Live Demo (Python) Logistic Regression - Introduction - Live Demo using Iris Dataset (Python) k-Nearest Neighbor - What is kNN? - Industrial Applications - How things are predicted using kNN algo? - How to choose value of k? - Live Demo (Python)	Chap 12, 16 [Joel]
Second Mid Exams			
27-28	13	Chi-square - What is Chi-square? - Why do we use it? - What does it show? - How do we calculate and interpret it? - Class Activity Pearson's r correlation - What is Pearson's r? - Why we calculate? - How we calculate?	[Handouts]
29-30	14	Decision Tree - What is DT? - DT Terminologies (Root Node, Leaf Node, Splitting, Branches, Pruning, Parent/Child Node) - How does a tree decide, where to split? - Entropy (What and how to calculate) - Information Gain - Pruning – Reducing the complexity	Chap 17 [Joel]
31-32	16	Project Demonstration Future Directions	
Final Exams			

Text Book

1. Data Science from Scratch FIRST PRINCIPLES WITH PYTHON by Joel Grus
2. Python Data Science Handbook ESSENTIAL TOOLS FOR WORKING WITH DATA by Jake VanderPlas

Reference Material

1. Introducing Data Science by Davy Cielen Arno D. B. Meysman Mohamed Ali
2. https://www.edx.org/course/introduction-python-data-science-3
3. https://www.khanacademy.org/math/statistics-probability

Course Learning Outcomes

	Course Learning Outcomes (CLO)
1	Understanding basic concepts and process of Data Science and Programming Constructs.
2	Apply Statistical and ML techniques to design solutions to real time problems.

CLO-SO Map

	SO IDs										
CLO ID	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**
CLO 1	1	0	0	0	0	0	0	0	0	0	0
CLO 2	0	1	0	0	1	0	0	0	0	0	0

Approvals

Prepared By	Saif Hassan
Approved By	Not Specified
Last Update	15/01/2019