CSE462-IntroductiontoDataScience Syllabus

GeneralInformation

CourseNumber	**CSE462**
CreditHours	3 (Theory Credit Hour = 3)
Prerequisite	1. Probabilityand Statistics 2. Databases 3. BasicProgramming
CourseInstructor	

CourseObjectives

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 ofcomputer science along with a good understanding of the craft of problem formulation to engineer effectivesolutions. This course will introduce students to this rapidly growing field and equip them with some of itsbasic principles and tools as well as its general mindset. Students will learn concepts, techniques and toolsthey 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, andeffective 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 solvingproblems. To make the learning contextual, market oriented content, course content will be more practicallyratherthan theoretical.

CatalogDescription

CSE462

CourseContent

SessionNo.	WeekNo.	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[DavyCielen]
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	DataVisualization Exploratory Data Analysis	Handouts

11-12	6	Descriptive Statistics - Mean, Mode, Median - Standard Deviation, Variance	Handouts
13-14	7	MachineLearning - What is ML? Why ML? - Supervised vs. Unsupervised - ML Applications/Examples - Overfitting and Under-fitting	Chap:5[Jake]
15-16	8	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
	9	Midterm Examination	
17-18-19-20-21-22	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)	Chap12,16[Joel]
23-24	13-15	Chi-square - What is Chi-square? - Why do we use it? - What does it show? - How do we calculate and interpret it? - ClassActivity Pearson'sr correlation - What is Pearson'sr? - Why we calculate? - How we calculate?	[Handouts]
	16	Project Demonstration Future Directions	
		FinalExamination	

TextBook

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 Vander Plas

ReferenceMaterial

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

CourseLearningOutcomes

	CourseLearningOutcomes(CLO)
1	UnderstandingbasicconceptsandprocessofData ScienceandProgrammingConstructs.
2	ApplyStatistical andMLtechniques todesign solutionsto realtimeproblems.

CLO-SOMap

	SOIDs										
CLO ID	**A**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**
CLO1	1	0	0	0	0	0	0	0	0	0	0
CLO2	0	1	0	0	1	0	0	0	0	0	0

Approvals

Prepared By	
ApprovedBy	
LastUpdate	