ITEC 313 (Introduction to Data Science)

General Information

Course Code	ITEC 313	Required (R)/Selected Elective (SE)		R
Credit Hours	Theory	2	Lab	1	Total	3
Prerequisites	NIL			
Course Coordinator	Dr. Nadim Rana			

Course Description

Data Science is the study of the generalizable extraction of knowledge from data. Being a data scientist requires an integrated skill set spanning mathematics, statistics, 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, real datasets from a variety of disciplines will be used.

Course Objectives

- Describe Data Science and the skill sets needed to be a data scientist.
- Understand the Data Science Process and how its components interact.
- Carry out basic statistical modeling and analysis.
- Explain the significance of exploratory data analysis (EDA) in data science.
- Apply basic tools (plots, graphs, summary statistics) to carry out EDA.
- Use APIs and other tools to scrap the Web and collect data.

Course Contents

List of Topics	Weeks
Chapter 1: Introduction to Data Science: Introduction to Data Science, Importance of Data Science, Where Do We See Data Science?, How Does Data Science Relate to Other Fields?, Skills for Data Science A, Data Science Profile, Data Scientist's tasks, Tools for Data Science, Issues of Ethics, Bias, and Privacy in Data Science	*1,2,3*
Chapter 2 Data: Data Types, Structured Data, Unstructured Data, Challenges with Unstructured Data, Data Collections, Open Data, Social Media Data, Multimodal Data, Data Storage and Presentation, Data Pre-processing, Data Cleaning, Data Integration, Data Transformation, Data Reduction, Data Discretization	*4, 5, 6*
Chapter 3 Data Analysis: Data Analysis and Data Analytics, Descriptive Analysis, Variables, Frequency Distribution, Measures of Centrality, Dispersion of a Distribution	*7, 8*
Chapter 4 Data Analytics: Diagnostic Analytics, Predictive Analytics, Prescriptive Analytics, Exploratory Analysis, Mechanistic Analysis	*9, 10*
Chapter 5 Machine Learning: What Is Machine Learning?, Machine Learning Examples, Machine Learning Applications, Machine Learning Algorithms, Regression, Error Function or Cost Function, Gradient Descent, Choosing ML Algorithm, Choosing the right estimator	*11, 12, 13*
Chapter 6 Data Collection, Experimentation and Evaluation Data Collection Methods: Picking Data Collection and Analysis Methods, Evaluation, Comparing Models, Hand-on-Example: Collecting and Analyzing Twitter Data	*14, 15*

Textbook

- Chirag Shah, "A Hands-On Introduction to Data Science", 1st Edition, © 2020 | ISBN-10:1108472443, ISBN-13:978-1108472449, Cambridge University Press

Reference Materials

- Foster Provost and Tom Fawcett, "Data Science for Business: What You Need to Know about Data Mining and Data-analytic Thinking", 2013, ISBN 1449361323.
- Cathy O'Neil and Rachel Schutt, "Doing Data Science, Straight Talk From The Frontline", O'Reilly, 2014, ISBN: 978-1-449-35865-5.
- Trevor Hastie, Robert Tibshirani and Jerome Friedman, "Elements of Statistical Learning", Second Edition, 2009, ISBN 0387952845. (free online)

Course Learning Outcomes

CLO-IDs	Course Learning Objective (CLOs)	Level of Learning (LoL)	Mapped PIs
CLO-1	**Define** the basic concepts and terminologies of the data science process	Knowledge	PI 1.1, PI 2.1
CLO-2	**Identify** the critical methods and techniques for data preprocessing used to solve a data problem	Comprehension	PI 1.3
CLO-3	**Demonstrate** proficiency with the tools and techniques for obtaining, organizing, exploring, and analyzing data	Applying	PI 2.3
CLO-4	**Analyze** and **compare** appropriate solutions for a problem by applying basic data analysis and statistical modeling tools to carry out Exploratory Data Analysis (EDA)	Analyzing	PI 1.4
CLO-5	**Discover** data-driven insights and **show** their analysis results effectively through written reports.	Applying	PI 3.1, PI 3.3

CLO-SO-PI Mapping

CLO IDs	SO-IDs					
	SO-1	**SO-2**	**SO-3**	**SO-4**	**SO-5**	**SO-6**
CLO-1	PI 1.1	PI 2.1	-	-	-	-
CLO-2	PI 1.3	-	-	-	-	-
CLO-3	-	PI 2.3	-	-	-	-
CLO-4	PI 1.4	-	-	-	-	-
CLO-5	-	-	PI 3.1, PI 3.3	-	-	-

Approvals

Prepared by CC	Dr. Nadim Rana
Approved by TL	Dr. John Martin
Last updated	05-10-2024