| | | |
|---|---|---|
| **Program** | BS Geography | |
| **Course Code** | GEOG-312 | |
| **Course Title** | Introduction to Data Science | |
| **Credit Hours** | **Theory** | **Lab** |
| | 2 | 1 |
| **Lecture Duration** | 90 minutes (1 Hours), 2 lectures per week, 3 hours lab session per week | |
| **Semester** | 5 | |
| **Pre-requisites** | **Courses** | **Knowledge** |
| | | Nil |
| **Follow Up Courses** | Data Mining, Data Warehousing and Business Intelligence | |

**Course Learning Outcomes (CLOs)**

| CLO No | Course Learning Outcome | Bloom Taxonomy |
|---|---|---|
| CLO-1 | Describe what Data Science is and the skill sets needed to be a data scientist. | C2 (Understand) |
| CLO-2 | Apply EDA and the Data Science process in a case study. | C3 (Apply) |
| CLO-3 | Comprehend the fundamental constructs of Python programming language. | C2 (Understand) |
| CLO-4 | Apply basic machine learning algorithms to solve real world problems of moderate complexity. | C3 (Apply) |

| | |
|---|---|
| **Aims and Objectives** | 1. Data Science is the study of the generalizable extraction of knowledge from data.<br>2. 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. |
| | 3. The aim of this course is to: Introduce students to this rapidly growing field and equip them with some of its basic principles and tools as well as its general mindset.<br>4. Explain the significance of exploratory data analysis in data science. Identify common approaches used for Feature Generation as well as Feature Selection, and finally discuss the Ethical and Privacy issues.<br>5. Programming language Python has been proposed for the practical work of this course. |
| **Learning Outcomes** | • Describe what Data Science is and the skill sets needed to be a data scientist.<br>• Apply EDA and the Data Science process in a case study.<br>• Comprehend the fundamental constructs of Python programming language.<br>• Apply basic machine learning algorithms to solve real world problems of moderate complexity. |

| | |
|---|---|
| **Syllabus** | Introduction: What is Data Science? Big Data and Data Science hype, Datafication, Current landscape of perspectives, Skill sets needed; Statistical Inference: Populations and samples, Statistical modeling, probability distributions, fitting a model, Intro to Python; Exploratory Data Analysis and the Data Science Process; Basic Machine Learning Algorithms: Linear Regression, k-Nearest Neighbors (k-NN), k-means, Naive Bayes; Feature Generation and Feature Selection; Dimensionality Reduction: Singular Value Decomposition, Principal Component Analysis; Mining Social-Network Graphs: Social networks as graphs, Clustering of graphs, Direct discovery of communities in graphs, Partitioning of graphs, Neighborhood properties in graphs; Data Visualization: Basic principles, |
| | ideas and tools for data visualization; Data Science and Ethical Issues: Discussions on privacy, security, ethics, Next-generation data scientists. |
| **Contents** | 1. Introduction: What is Data Science? Big Data and Data Science hype, Datafication, Current landscape of perspectives, Skill sets needed;<br>2. Statistical Inference: Populations and samples, Statistical modeling, probability distributions, fitting a model;<br>3. Intro to Python;<br>4. Exploratory Data Analysis and the Data Science Process;<br>5. Basic Machine Learning Algorithms: Linear Regression, k-Nearest Neighbors (k-NN), k-means, Naive Bayes;<br>6. Feature Generation and Feature Selection;<br>7. Dimensionality Reduction: Singular Value Decomposition, Principal Component Analysis;<br>8. Mining Social-Network Graphs: Social networks as graphs, Clustering of graphs, Direct discovery of communities in graphs, Partitioning of graphs, Neighborhood properties in graphs;<br>9. Data Visualization: Basic principles, ideas and tools for data visualization;<br>10. Data Science and Ethical Issues: Discussions on privacy, security, ethics, Next-generation data scientists. |
| **Teaching-learning Strategies** | • Interactive class session<br>• Hands on practices in class<br>• Brainstorming and Group discussion sessions<br>• Coding in LABS |
| **Assignments** | • Paper based written assignments 4<br>• Coding assignments in Python 6 |
| **Textbooks** | • An Introduction to Data Science, Jeffrey S. Saltz, Jeffrey M. Stanton, SAGE Publications, 2017.<br>• Python for everybody: Exploring data using Python 3, Severance, C.R., CreateSpace Independent Pub Platform. 2016. |

| | |
|---|---|
| **Reference Material/Suggested Readings** | • Foundations of data science, Blum, A., Hopcroft, J., & Kannan, R., Vorabversion eines Lehrbuchs, 2016.<br>• Doing Data Science, Straight Talk from the Frontline, Cathy O'Neil and Rachel Schutt, O'Reilly. 2014.<br>• Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, EMC Education Services, John Wiley & Sons, 2015. |
| **Notes** | • Academic integrity is expected of all students. Plagiarism or cheating in any assessment will result in at least an F grade in the course, and possibly more severe penalties.<br>• You bear all the responsibility for protecting your assignments from plagiarism. If anyone else submits your assignment or uses your code in his/her assignment, you will be considered equally responsible.<br>• The instructor reserves the right to modify the grading scheme/marks division and course outline during the semester.<br>• Introductory knowledge of using the computers is assumed for this course. All code written in quizzes, assignments, homework's, and exams must be in JavaScript. Code must be intelligently documented (commented). Undocumented code may not be given any credit.<br>• The IDE use is not allowed, Notepad++ has to be used for coding.<br>• There is no makeup for a missed sessional grading instruments like quizzes, assignments, and homework's. |