

| Python for Data Science |  |   |
|-------------------------|--|---|
| Instructors             | Karari E.K   | Email: <a href="mailto:ekarari@cuea.edu">ekarari@cuea.edu</a> |
| Objectives              | At the end of the course, the learner should be able to:<br>1. Identify tools used in Data science<br>2. Install and configure python<br>3. Write programs in python<br>4. Analyze data with python  |   |
| COURSE CONTENT          |  |   |
| Lesson 1                | Overview of Python   |   |
|                         | 1. Definition of terms<br>2. Python and Big Data<br>3. Characteristics of Big Data<br>4. Types of Big data<br>5. Tools used in Big data analysis<br>6. Traditional versus Big data business Approach |   |
| Lesson 2                | Brief review of Hadoop   |   |
|                         | 1. Hadoop and its components<br>2. Relational Databases and MapReduce<br>3. Data Analysis with Hadoop<br>4. Hadoop physical Architecture<br>5. Hadoop limitations                                    |   |
| Lesson 3                | Introduction to python   |   |
|                         | 1. Python and its uses<br>2. Advantages of python<br>3. Python Syntax compared to other languages<br>4. Installing and running python  |   |
| Lesson 4                | Python Basics  |   |
|                         | 1. Indentation and blocks<br>2. Comments<br>3. Variables, constants, classes and objects<br>4. Data types and type casting<br>5. Operators<br>6. Functions   |   |
| Lesson 5                | Python Arrays and related  |   |
|                         | 1. Arrays<br>2. List<br>3. Tuples<br>4. Sets<br>5. Dictionaries  |   |
| Lesson 6                | Python Strings   |   |
|                         | 1. Declaration and definition<br>2. Slicing<br>3. Modifying strings<br>4. Escape characters<br>5. Other string methods   |   |
| Lesson 7                | Python control flow  |   |

|                  |  |
|------------------|--|
|                  | <ol style="list-style-type: none"> <li>1. Decision making</li> <li>2. Looping/ iteration</li> </ol>  |
| Lesson 8         | Python file handling & Modules   |
|                  | <ol style="list-style-type: none"> <li>1. Create, R/W, Delete</li> <li>2. NumPy,Pandas, SciPy,Scikit-learn</li> <li>3. Matplotlib for visualization</li> </ol>   |
| Lesson 9         | Data Analysis using Python   |
|                  | <ol style="list-style-type: none"> <li>1. Linear regression</li> <li>2. Logistic regression</li> <li>3. Correlation</li> </ol>   |
| Lesson 10        | Discovering hidden patterns : Machine learning   |
|                  | <ol style="list-style-type: none"> <li>1. Classification</li> <li>2. Deep learning</li> </ol>  |
| Mode of Delivery | <ol style="list-style-type: none"> <li>1. Lectures</li> <li>2. Discussion Forums</li> <li>3. Presentations</li> <li>4. practical</li> </ol>  |
| Assessment       | Continuous Assessment 40 %    End of Trimester Examination. 60 %   |
| References       | <p>Course Textbook</p> <ol style="list-style-type: none"> <li>1. Introduction to python for Computer Science and Data Science, by Deitel and Deitel</li> </ol>   |
|                  | <p>Other</p> <ol style="list-style-type: none"> <li>1. Big Data from Scratch: Building a 4-nodes Hadoop cluster and use of the Map-Reduce Simple Skyline Algorithm Kindle Edition 2017 by Panagiotis Leliopoulos ASIN: B076WXMKVL</li> <li>2. Programming Hive: Data Warehouse and Query Language for Hadoop 1st Edition 2012 by Edward Capriolo , Dean Wampler , Jason Rutherglen ISBN-10: 1449319335 ISBN-13: 978-1449319335</li> <li>3. Data-Intensive Text Processing with MapReduce (Synthesis Lectures on Human Language Technologies) Paperback – April 30, 2010 by Jimmy Lin , Chris Dyer , Graeme Hirst ISBN-10: 1608453421 ,ISBN-13: 978-1608453429</li> <li>4. MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems 1st Edition 2012 by Donald Miner, Adam Shook ISBN-10: 9781449327170 ISBN-13: 978-1449327170</li> <li>5. Programming Pig: Dataflow Scripting with Hadoop 1st Edition by Alan Gates ISBN-10: 1449302645 ISBN-13: 978-1449302641</li> </ol> |