

UOE042: Introduction to Data Analytics							
University Open Elective-IV (CSE & AIML)							
Lect.	Tut.	Pract.	Credits	Evaluation Scheme			
				Component	Exam	Weightage %	Pass %
2	-	-	2	Theory 100 Marks	FA	50	40
					SA	50	40

### Course Objectives

To provide strong foundation for data analytics and application area related to it and understand the underlying core concepts and emerging technologies in data analytics.

### Course Outcomes

**CO1:** Explore the fundamental concepts of data analytics

**CO2:** Understand data analysis techniques for applications handling large data

**CO3:** Understand various machine learning algorithms used in data analytics process

**CO4:** Visualize and present the inference using various tools

**CO5:** Learn to think through the ethics surrounding privacy, data sharing and algorithmic decision-making

### Unit-1

**Teaching Hours: 7**

#### INTRODUCTION

Data Analytics - Types – Phases - Quality and Quantity of data – Measurement - Exploratory data analysis - Business Intelligence.

### Unit-2

**Teaching Hours: 7**

#### BIG DATA

Big Data and Cloud technologies - Introduction to HADOOP: Big Data, Apache Hadoop, MapReduce - Data Serialization - Data Extraction - Stacking Data - Dealing with data.

### Unit-3

**Teaching Hours: 7**

#### DATA VISUALIZATION

Introduction to data visualization – Data visualization options – Filters – Dashboard development tools – Creating an interactive dashboard with dc.js - summary.

### Unit-4

**Teaching Hours: 7**

#### ANALYTICS AND MACHINE LEARNING

Machine learning – Modeling Process – Training model – Validating model – Predicting new observations –Supervised learning algorithms – Unsupervised learning algorithms.

**ETHICS AND RECENT TRENDS**

Data Science Ethics – Doing good data science – Owners of the data - Valuing different aspects of privacy - Getting informed consent - The Five Cs – Diversity – Inclusion – Future Trends.

**Books:**

- [1] Davy Cielen, Arno D. B. Meysman, Mohamed Ali, Introducing Data Science, ManningPublications Co., 1<sup>st</sup> edition, 2016.
- [2] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, An Introduction to Statistical Learning: with Applications in R, Springer, 1<sup>st</sup> edition, 2013.
- [3] Bart Baesens, Analytics in a Big Data World: The Essential Guide to Data Science and its Applications, Wiley.
- [4] D J Patil, Hilary Mason, Mike Loukides, Ethics and Data Science, O’ Reilly, 1<sup>st</sup> edition, 2018.

**References:**

- [1] Dr Anil Maheshwari, Data Analytics Made Accessible, Publisher: Amazon.com ServicesLLC.
- [2] Joel Grus, Data Science from Scratch: First Principles with Python, O’Reilly, 1<sup>st</sup> edition, 2015.
- [3] Cathy O’Neil, Rachel Schutt, Doing Data Science, Straight Talk from the Frontline, O’Reilly, 1<sup>st</sup> edition, 2013.
- [4] Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press, 2<sup>nd</sup> edition, 2014.
- [5] Eric Siegel, Predictive Analytics The Power to Predict Who Will Click, Buy, Lie, or Die, 2<sup>nd</sup> Ed., Wiley.