

Nepal Open University
Manbawan, Lalitpur, Nepal
Faculty of Science, Health and Technology

Master of Philosophy in Information and Communication Technology (MPhil in ICT)

Course Title: Data Science

Credit Hours: 3 (45Hrs.)

Course Code Number: ICT 702

Semester: First

Year: First

Introduction to the Course:

This course is designed with the aims of providing knowledge on data science. Data science is the intersection of computer science and statistics. This course covers the data analysis concepts, advanced statistical knowledge, data mining and machine learning, and relevant tools.

Course Objectives:

The objectives of the course are:

- a) To make students aware about methodology of data science
- b) To familiarize students with the use of statistics
- c) To impart practical knowledge on data mining and machine learning
- d) To have hands on experience in the data analysis tools

Course Contents (Unit title only):

Unit One: Introduction to Data Science

Unit Two: Statistics

Unit Three: Exploratory Data Analysis

Unit Four: Data Mining and Machine Learning

Unit Five: R Programming Language

Contents Detail:

Contents	Hrs
Unit 1: Introduction to Data Science 1.1 Data science concepts 1.3 Importance of data science	10

1.4 Big Data 1.5 Data Extraction and Information Retrieval 1.6 Skills required for Data Science 1.7 Privacy, Security, Ethics in Data Science	
Unit 2: Statistics 2.1 Introduction to Statistics 2.2 Populations and samples 2.3 Advanced Statistical Techniques 2.4 Multi-variate analysis 2.5 Regression	10
Unit 3: Exploratory Data Analysis 3.1 Philosophy of EDA and the Data Science process 3.2 Basic tools (plots, graphs and summary statistics) 3.3 Data visualization and presentation	10
Unit 4 : Data Mining and Machine Learning 4.1 Association Rules (k-NN) 4.2 Clustering techniques (k-means) 4.3 Classification (Naive Bayes) 4.4 Feature Selection algorithms 4.5 Filters, Wrappers, Decision Trees, Random Forest 4.6 Dimensionality Reduction (SVD) 4.7 Principal Component Analysis	10
Unit 5 : R programming 5.1 Basics of R Programming 5.2 Plotting Graphs with R 5.3 Statistical analysis using R	5

Learning Outcomes:

Unit 1: Introduction to Data Science

This unit covers the basics of data science and related fields. This unit also discusses about the security and privacy aspects of Data Science.

Unit 2: Statistics

This unit focuses on statistics including statistical inference, probability distribution, multi-variate analysis etc.

Unit 3: Exploratory Data Analysis

This unit imparts students with the knowledge on exploratory data analysis and data visualization techniques.

Unit 4 : Data Mining and Machine Learning

This unit specifically covers areas of data mining and machine learning which are related to Data Science.

Unit 5 : R programming

This unit provides students with tools required to implement concepts and methods of Data Science using R.

Mode of Delivery:

The mode of course delivery consists of open and distance (online/offline) and face-to-face or both. 25% of the course is designed to be delivered on online mode, 60% on offline, and remaining 15% on face-to-face mode.

Evaluation Scheme:

- In-semester: 40%
- End-semester: 60%

In Semester Evaluation (40%)	End Semester Evaluation (60%)
a) Unit Assignment/Project/Task (20%) b) Critical Comment/Review (15%) c) Interaction with the Tutor (5%)	a) Short answer questions b) Long answer questions or research paper writing (With a focus on higher order thinking skills such as analysis, synthesis, evaluation, etc.)

Semester Guidelines:

- (i) The semester system is not only an examination system. The main objective of this system is to enhance student's knowledge, skill and capacity continuously, extensively and in depth.
- (ii) The normal and maximum duration for obtaining the master's degree is 24 months and 72 months respectively. Students failing to complete the requirements in 72 months have to re-enroll.
- (iii) Students need to maintain 80% online presence (attendance) for both theory and laboratory classes. They should be regular or part time as per the course registration in online classes. They should enter before starting the classes.

References:

1. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk From The Frontline, O'Reilly. 2014.
2. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, An Introduction to Statistical Learning, ISBN 978-1-4614-7138-7. 2013.
3. Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. ISBN 0262018020. 2013.
4. Foster Provost and Tom Fawcett. Data Science for Business: What You Need to Know about Data Mining and Data-analytic Thinking. ISBN 1449361323. 2013.
5. Trevor Hastie, Robert Tibshirani and Jerome Friedman. Elements of Statistical Learning, Second Edition. ISBN 0387952845. 2009.

6. Mohammed J. Zaki and Wagner Miera Jr. Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press. 2014.
7. Jiawei Han, Micheline Kamber and Jian Pei. Data Mining: Concepts and Techniques, Third Edition. ISBN 0123814790. 2011.