| DATA SCIENCE SEMESTER – V | | | |
|---|---|---|---|
| Course Code | **M23BCI503** | CIE Marks | **50** |
| Number of Lecture Hours/Week(L: T: P: S) | **(4:0:0:0)** | SEE Marks | **50** |
| Total Number of Lecture Hours | **40 hours Theory** | Total Marks | **100** |
| Credits | **04** | Exam Hours | **03** |

**Course Objectives:**
1. Introduce fundamental concepts and techniques in data science.
2. Equip students with practical skills for data analysis, visualization, and machine learning.
3. Foster an understanding of real-world applications of data science, such as in recommendation systems and predictive modelling.
4. Prepare students for advanced study or careers in data science and related fields.

**Module -1**

**Introduction to Data Science**
**Introduction:** What is Data Science? Big Data and Data Science hype – and getting past the hype, Why now? Datafication, Current landscape of perspectives, Skill sets needed.
**Statistical Inference:** Populations and samples, Statistical modelling, probability distributions, fitting a model.

**Module -2**

**Exploratory Data Analysis and the Data Science Process**
Basic tools (plots, graphs and summary statistics) of EDA, Philosophy of EDA, The Data Science Process, Case Study: Real Direct (online real estate firm). Three Basic Machine Learning Algorithms: Linear Regression, k-Nearest Neighbours (k- NN), k-means.

**Module -3**

**One More Machine Learning Algorithm and Usage in Applications** - Motivating application: Filtering Spam - Why Linear Regression and k-NN are poor choices for Filtering Spam - Naive Bayes and why it works for Filtering Spam - Data Wrangling: APIs and other tools for scrapping the Web.

**Module -4**

**Feature Generation and Feature Selection (Extracting Meaning From Data)** - Motivating application: user (customer) retention - Feature Generation (brainstorming, role of domain expertise, and place for imagination) - Feature Selection algorithms – Filters; Wrappers; Decision Trees; Random Forests.

**Module -5**

**Recommendation Systems:** Building a User-Facing Data Product - Algorithmic ingredients of a Recommendation Engine - Dimensionality Reduction - Singular Value Decomposition - Principal Component Analysis - Exercise: build your own recommendation system
**Mining Social-Network Graphs** - Social networks as graphs - Clustering of graphs - Direct discovery of communities in graphs - Partitioning of graphs - Neighbourhood properties in graphs.

**TEXTBOOKS:**
1. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk FromThe Frontline. O'Reilly. 2014.
2. Data Science for Business: What you need to know about data mining and data-analytic thinking by Foster Provost and Tom Fawcett, O'Reilly Media, 2013.

**REFERENCE BOOKS:**
1. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. Mining of Massive Datasets. v2.1, Cambridge University Press. 2014. (free online).
2. Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. ISBN 0262018020. 2013.
3. Python for Data Analysis by Wes McKinney, O'Reilly Media, 2017.