

**UNIVERSITY OF MADRAS**  
**BACHELOR OF COMPUTER APPLICATIONS (BCA)**  
**DEGREE PROGRAMME**  
**SYLLABUS WITH EFFECT FROM 2023-2024**

**Year: III**

**Semester: V**

<b>Introduction To Data Science</b> Common for B.C.A. , B.Sc.-SA , B.Sc.-CSc	<b>320E5F</b>
<b>Credits 3</b>	<b>Lecture Hours:4 per week</b>
<b>Learning Objectives:</b> (for teachers: what they have to do in the class/lab/field) <ul style="list-style-type: none"> <li>• An understanding of the data operations</li> <li>• An overview of simple statistical models and the basics of machine learning techniques of regression.</li> <li>• An understanding good practices of data science</li> <li>• Skills in the use of tools such as python, IDE</li> <li>• Understanding of the basics of the Supervised learning</li> </ul>	
<b>Course Outcomes:</b> (for students: To know what they are going to learn) <ol style="list-style-type: none"> <li>1. Clean and reshape messy datasets</li> <li>2. Use exploratory tools such as clustering and visualization tools to analyze data</li> <li>3. Perform linear regression analysis</li> <li>4. Use methods such as logistic regression, nearest neighbours, decision trees, support vector machines, and neural networks to build a classifier</li> <li>5. Apply dimensionality reduction tools such as principal component analysis</li> </ol>	

Units	Contents
<b>I</b>	Introduction: Introduction to Data Science – Evolution of Data Science – Data Science Roles – Stages in aData Science Project – Applications of Data Science in various fields – Data Security Issues.
<b>II</b>	Data Collection and Data Pre-Processing: Data Collection Strategies – Data Pre-Processing Overview – Data Cleaning – DataIntegration and Transformation – Data Reduction – Data Discretization.
<b>III</b>	Exploratory Data Analytics: Descriptive Statistics – Mean, Standard Deviation, Skewness and Kurtosis – Box Plots –Pivot Table – Heat Map – Correlation Statistics – ANOVA.
<b>IV</b>	Model Development: Simple and Multiple Regression – Model Evaluation using Visualization – Residual Plot –Distribution Plot – Polynomial Regression and Pipelines – Measures for In-sampleEvaluation – Prediction and Decision Making.
<b>V</b>	Model Evaluation: Generalization Error – Out-of-Sample Evaluation Metrics – Cross Validation – Overfitting –Under Fitting and Model Selection – Prediction by using Ridge Regression – TestingMultiple Parameters by using Grid Search

**Books for References**

1. Jojo Moolayil, “Smarter Decisions : The Intersection of IoT and Data Science”,PACKT, 2016.
2. Cathy O’Neil and Rachel Schutt , “Doing Data Science”, O'Reilly, 2015.
3. David Dietrich, Barry Heller, Beibei Yang, “Data Science and Big data Analytics”,EMC 2013
4. Raj, Pethuru, “Handbook of Research on Cloud Infrastructures for Big DataAnalytics”, IGI Global.