

POST GRADUATE DIPLOMA IN DATA SCIENCE AND STATISTICS (PGDDSS)

PROGRAMME CURRICULUM

Semester I

1. Basics of Statistics
2. Introduction to Data Science
3. Data Structures and Algorithms
4. Introduction to R Programming

Basics of Statistics

1. Introduction to Statistics
2. Collection and Scrutiny of Data
3. Classification and Tabulation of Data
4. Diagrammatic Presentation of Data
5. Graphical Presentation of Data
6. Measures of Central Tendency (Mean, Median and Mode)
7. Measures of Dispersion
8. Correlation Coefficient
9. Rank Correlation
10. Intra-Class Correlation

Introduction to Data Science

1. Basics of Data
2. What is Data Science?
3. Data Science and Ethical Issues
4. Big Data and Data Science Hype, Datafication
5. Understand Data Science Pipeline – Data Wrangling, Exploratory Analysis, Modeling
6. Getting and Cleaning Data
7. Visualising the Data
8. The Data Scientist's Toolbox
9. Applications of Data Science in Business and Industry
10. Case Study

Data Structures and Algorithms

1. Programming Fundamentals
2. Data types
3. Arrays and Pointers
4. Functions
5. Stack and Queues
6. Linked List
7. Trees
8. Searching Algorithms

9. Sorting Algorithms
10. Graphs
11. Dynamic Programming

Introduction to R Programming

1. Introduction to R
2. Install R, RStudio, R Package
3. Vectors
4. Matrices
5. Lists
6. Data Frames
7. Basic Expressions – Conditional, Loop
8. Functions in R
9. Doing Maths in R
10. Input/output and Object Oriented Programming
11. Graphics
12. Debugging
13. Parallel R and String Manipulation

Semester II

1. Python Programming
2. Machine Learning with R and Python
3. Advanced Statistics with Data Analysis
4. Emerging Trends in Data Science

Python Programming

1. Introduction to Python
2. Install Python
3. Variables, expressions and statements
4. Functions
5. Conditionals, recursion and iteration
6. Strings
7. Lists
8. Tuples
9. Dictionaries
10. Object Oriented Programming
11. Files and Error Handling
12. Testing, Debugging and Profiling
13. Data Structures - Linked lists, Stacks, Queues and Trees

Machine Learning with R and Python

1. Basics of Machine Learning
2. Supervised Machine Learning – K-NN, Naïve Bayes, Decision tree, SVM
3. Unsupervised Machine Learning – K means, Apriori algorithms

4. Regression Models
5. Clustering Models
6. knitr, RPub, R Markdown, swirl, ggplot2
7. Computation with Python – NumPy, SciPy
8. Data Manipulation in Python- Pandas
9. Understanding DataFrame
10. Data Visualisation in Python - matplotlib
11. Introduction to Scikit – Machine learning
12. Web Scraping in Python – BeautifulSoup
13. Integration using PySpark, Hadoop, MapReduce

Advanced Statistics with Data Analysis

1. Multiple Correlation
2. Partial Correlation
3. Introduction to Probability
4. Different Approaches to Probability Theory
5. Laws of Probability
6. Linear Regression
7. Multiple Regression
8. Introduction to Sample Surveys
9. Simple Random Sampling
10. Stratified Random Sampling, Other Sampling Schemes
11. Analysis of Variance
12. Data sanitisation
13. Data analysis using s/w packages like MS Excel, SPSS etc

Emerging Trends in Data Science

1. Big Data
2. Apache Spark and Scala
3. Deep Learning
4. Artificial Intelligence
5. Business Intelligence
6. Natural Language Processing
7. Types of Analytics – Descriptive, Predictive, Prescriptive
8. Business Analytics
9. Web Analytics
10. Next-generation data analytics
11. Case Study

Semester III

1. Case Study
2. Project