POST GRADUATE DIPLOMA IN DATA SCIENCE (PGDDS)

PROGRAMME CURRICULUM

Semester - I

Basics of Statistics

1. Basics of Statistics
2. Data Collection and Measurement
3. Data Presentation
4. Data Processing and Analysis
5. Measures of Central Tendency (Mean, Median and Mode)
6. Measures of Dispersion
7. Correlation

Introduction to Data Science

1. Basics of Data
2. Basics of Data Science
3. Big Data, Datafication & its impact on Data Science
4. Data Science Pipeline, EDA & Data Preparation
5. Data Scientist Toolbox, Applications & Case Studies

Data Structures and Algorithms

1. Programming Fundamentals
2. Control Flow
3. Arrays and Pointers
4. Functions
6. Stacks and Queues
7. Linked Lists
8. Trees
9. Searching Algorithms
10. Sorting Algorithms
11. Graphs

Introduction to R Programming

1. Introduction to R
2. Data Types and Data Structures
3. Loops and Functions in R
4. Mathematics in R
5. Graphs
6. String Manipulation and Input/output
7. Object Oriented Programming – I
8. Object Oriented Programming – II
9. Debugging and Condition Handling
10. Introduction to Parallel Computing in R

Semester II

Big data with Data Warehousing and Data Mining

1. Introduction to Data Warehouse
2. Data Warehouse Architecture
3. Dimensional Modelling
4. Data Warehouse Implementation
5. Data Warehouse and OLAP Technologies
6. Introduction to Data Mining
7. Mining Association Rules
8. Classification and Prediction
9. Mining Complex Types of Data
10. Data Mining Applications

Advanced Statistics

1. Sampling and Sampling Technique
2. Probability
3. Normal Distribution
4. Linear Regression
5. Multiple Linear Regression
6. Random Variables

Python Programming

1. Introduction to Python
2. Variables, expressions and statements
3. Control Structures, Data structures- Arrays and Linked lists, Queues
4. Functions
5. Conditionals, recursion and iteration
6. Strings
7. Lists and Tuples
8. Dictionaries
9. Object Oriented Programming
10. Files and Error Handling
11. Testing, Debugging and Profiling
12. Handling data with Python
13. Python Graphical User Interface Development

Submission I

In Semester II students are required to submit a submission as per guidelines given by SCDL.

Semester III

NoSQL Databases

1. Introduction to NoSQL
2. Basics of NoSQL
3. Replication and Sharding
4. Key-Value Databases
5. Document Databases
6. Column-Oriented Databases
7. Graph Databases
8. Advanced NoSQL

Data Visualisation

1. Introduction to Data Visualisation
2. Visualisation of Numerical Data
3. Visualisation of Non-numerical Data
4. Common Visualisation Idioms
5. Visualisation of Spatial Data, Networks and Trees
6. Data Reduction
7. Introduction to Tableau
8. Data Visualisation with SPSS

Machine Learning with R and Python

1. Basics of Machine Learning
2. Supervised Machine Learning
3. Unsupervised Learning
4. Regression Algorithms
5. Clustering Models
6. R Markdown, KNITR, RPUBS
7. GGPLOT2
8. Computation with Python – NumPy, SciPy
9. Pandas
10. Aggregating and Analysing Data with DPLYR
11. Data Visualisation in Python – Matplotlib
12. Introduction to scikit-learn
13. Web Scraping in Python – Beautiful Soup
14. Introduction to (PY) Spark

Ethical and Legal Issues in Data Science

1. What are Ethics?
2. Some Ethical concern of Data Science
3. History, Concept of Informed Consent
4. Data Ownership
5. Privacy, Anonymity, Data Validity
6. Algorithmic Fairness
7. Societal Consequences
8. Code of Ethics

Semester IV

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. Data Analytics
8. Web Analytics
9. Case Study

Submission II

In Semester IV students are required to submit a submission as per guidelines given by SCDL.

Project

Student should choose a technical or Techno-business topic of his/her interest and is required to develop the Project based on the provided guidelines.