

SYLLABUS**DS3104PC: INTRODUCTION TO DATA SCIENCE**

III-I:CSE(DS)								
Course Code	Category	Hours/Week			Credits	Max Marks		
DS3104PC	Core	L	T	P	C	CIE	SEE	Total
		3	0	0	3	25	75	100
Contact Classes:45	Tutorial classes:15	Practical classes: Nil				Total Classes:60		
Prerequisites: None								

Course Objectives**Course Objectives:**

- Learn data science project concepts
- Learn to collect data and process
- Learn to visualize data

Course Outcomes:

- Able to collect data from various resources and process data
- Able to collect data and apply different techniques on data
- Able to plot data using various methods
- Able to Model and develop the dataset
- Able to visualize data from various data sources

Unit – I:**Introduction**

Introduction to Data Science – Evolution of Data Science – Data Science Roles – Stages in a Data Science Project – Applications of Data Science in various fields – Data Security Issues.

Unit – II:**Data Collection and Data Pre-Processing**

Data Collection Strategies – Data Pre-Processing Overview – Data Cleaning – Data Integration and Transformation – Data Reduction – Data Discretization.

Unit – III:**Exploratory Data Analytics**

Descriptive Statistics – Mean, Standard Deviation, Skewness and Kurtosis – Box Plots –

INTRODUCTION TO DATA SCIENCE

Pivot Table – Heat Map – Correlation Statistics – ANOVA.

Unit – IV:

Model Development

Simple and Multiple Regression – Model Evaluation using Visualization – Residual Plot

Distribution Plot – Polynomial Regression and Pipelines – Measures for In-sample

Evaluation – Prediction and Decision Making.

Unit – V

Model Evaluation

Generalization Error – Out-of-Sample Evaluation Metrics – Cross Validation – Over fitting Under Fitting and Model Selection – Prediction by using Ridge Regression – Testing multiple Parameters by using Grid Search.

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 Data Analytics”, IGI Global.



NIRCM

YOUR TROTS TO SUCCESS