FUNDAMENTALS OF DATA SCIENCE

Course Code: 20CD1101

L	T	P	C
3	0	0	3

Course Outcomes:

At the end of the course, a student will be able to

CO1: Describe the significance of data science and understand the Data Science process. (L2)

CO2: Explain how data is collected, managed and stored for data science.(L2)

CO3: Build, and prepare data for use with a variety of statistical methods and models (L3)

CO4: Analyze Data using various Visualization techniques. (L4)

CO5: Choose contemporary models, such as machine learning, AI, techniques to solve practical problems (L4)

UNIT – I — 10 Lectures

Introduction To Data Science: Definition, Big Data and Data Science Hype, Datafication , Data Science Profile, Meta-Definition, Data Scientist, Statistical Inference, Populations and Samples, Populations and Samples of Big Data, Big Data Can Mean Big Assumptions, Modeling, Philosophy of Exploratory Data Analysis, The Data Science Process , A Data Scientist's Role in this Process
Case Study: RealDirect.(Text Book 2)

Learning Outcomes: At the end of the module the student will be able to

1. Understand the basics of data science(L1)
2. Summarize testable predictions for real-time data(L2)
3. Understand Data Scientist's Role in the analysis Process (L2)

UNIT –II — 10 Lectures

Mathematical Preliminaries: Probability,Descriptive Statistics, Correlation Analysis.(TextBook 1)

Data Munging: Properties of Data, Languages for Data Science, Collecting Data, Cleaning Data, Crowdsourcing. (Text Book 1)

Learning Outcomes: At the end of the module the student will be able to

1. Understand the concepts of Data collection and management (L1)
2. Establish sources of data ((L2)
3. Explain various mathematical concepts for Data Science (L2)

UNIT – III 8 Lectures

Scores and Rankings: Developing Scoring Systems, Z-scores and Normalization, Advanced Ranking Techniques

Statistical Analysis: Sampling from Distributions, Statistical Distributions, Statistical Significance, Permutation Tests and P-values(Text Book 1)

Learning Outcomes: At the end of the module the student will be able to

1. Use the concepts of statistics. (L3)
2. Identify distribution properties of data using statistical concepts. (L3)

UNIT- IV 10 Lectures

Visualizing Data: Exploratory Data Analysis, Developing a Visualization Aesthetic, Chart Types, Great Visualizations

Mathematical Models: Philosophies of Modeling, A Taxonomy of Models, Baseline Models, Evaluating Models, Evaluation Environment.(Text Book 1)

Learning Outcomes: At the end of the module the student will be able to

1. Understand types of data Visualization techniques (L1)
2. Use the measures for model evaluation(L3)
3. Evaluate models for multiple environments. (L4)

UNIT-V 12 Lectures

Supervised Learning: Linear Regression, Better Regression Models, Regression as Parameter Fitting, Simplifying Models through Regularization Classification and Logistic Regression, Issues in Logistic Classification, Naive Bayes, Decision Trees Classifiers (Text Book 1)

Learning Outcomes: At the end of the module, the student will be able to:

1. Understand regression techniques (L1)
2. Compare multiple classification techniques (L3)
3. Interpret multiple techniques for solving Data science applications .(L4)

TEXT BOOKS:

1. Steven S. Skiena, “The Data Science Design Manual”, Springer 2017.
2. Rachel Schutt & O’neil, “Doing Data Science”, Straight Talk from The Frontline O’REILLY, ISBN:978-1-449-35865-5, 1st edition, October 2013.

REFERENCE BOOKS

1. Joel Grus," Data Science from Scratch" First Edition, April 2015
2. Gareth James, Daniela Witten, Trevor Hatie, RoberstTibhirani , "An Introduction to Statistical Learning-with Applications in R", 2013
3. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. Mining of Massive Datasets. v2.1, Cambridge University Press. 2 edition (30 September 2014)
4. R Programming for Data Science, Roger D. Peng, LeanPub, 2015.

WEB REFERENCES:

1. "Data science for engineers" https://nptel.ac.in/noc/courses/noc20/SEM1/noc20-cs28/