COURSE TITLE	**DATA SCIENCE & MODELING**
COURSE CODE	**01AD0502**
COURSE CREDITS	**4**

Objective:

1 Learning basic concepts of creating mathematical representations (models) from data to find patterns, make predictions, or gain insights is primary objective of this course. This course specifically make student able to learn mathematical concepts used in data science and modeling techniques for solving real world problems and developing new applications based on further delve into the main statistical problems and solution approaches.

Course Outcomes: After completion of this course, student will be able to:

1 Learn to understand the use of Data Science, Data Analytics tools for the Data Science and career opportunities in Data Science and Modeling.

2 Demonstrate the ability to interpret and analyze descriptive statistical data using probability concepts and tabulation methods.

3 Learn to use R for data science, from data manipulation and machine learning, and thrive as a pro data scientist.

4 Apply statistical methods, regression techniques and related algorithms to both large and small data sets in Python and R Programming for prediction.

5 Demonstrate knowledge of statistical data analysis techniques utilized in decision making.

6 Implement hypothesis testing, various algorithms using various software platforms.

Pre-requisite of course: Basics of computer science including optimization algorithms, Machine Learning algorithms, Probability & statistics, Probability distribution, Fundamentals of R programming language.

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
3	0	2	50	30	20	25	25

DR. MADHU SHAMBHU SHUKLA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Contents : Unit	Topics	Contact Hours
1	**Introduction to Data Science, Data Analytics tools for the Data Science** Background: This foundation of Data Science explains what data science is, its applications across industries, the roles of data professionals that can lead to a career in data science. How Machine Learning and Deep Learning Models are useful for data science?, Describe the Data Scientist's tool kit which includes Data sets, Machine learning models, and Big Data tools, Utilize languages commonly used by data scientists like Python, R. , Demonstrate working knowledge of tools such as Jupyter notebooks and RStudio and utilize their various features. , Create and manage source code for data science using Git repositories and GitHub, Libraries & Packages (Exploratory data analysis (EDA) using Pandas and NumPy, scikit-learn, SciPy, TensorFlow to uncover patterns and insights), Web Scraping (learning about tools, libraries and ethical considerations), Data Visualization using Matplotlib, Seaborn, and Plotly. , R Libraries for Big Data (ggplot2, dplyr, tidyverse, shiny).	5
2	**Probability Concepts, Descriptive and Inferential Statistical** Basic probability, random variables, discrete & continuous distributions functions (Binomial, Poisson, Normal)., Descriptive Statistics (central tendency, variance, standard deviation, covariance, correlation, Law of Large Numbers, Point & Interval Estimation), Inferential Statistics through hypothesis tests (Central limit theorem, hypothesis testing, one-tailed and two-tailed test, and Chi-Square test). , Analysis of Variance, Kruskal-Wallis test, Coefficient of correlation	8
3	**Optimization Approaches, Machine Learning** Introduction to optimization, Constrained optimization, Unconstrained optimization, Linear optimization, Gradient-based methods., Simple and Multiple Linear Regression, Logistic Regression, Bias and Variance Tradeoff in Data, Ridge/Lasso/ElasticNet Regularization Techniques, Data Correlation: Relationship between data and variables (Principal component Analysis, Cluster Analysis, Multiple regressions Analysis).	10
4	**Fundamentals of R and Python Programming** R Data Structures (Vectors, Common Vectorized operations, Matrices, Arrays, Lists, Data Frames, File Handling)., Construct Python programs to clean and prepare data for analysis by addressing missing values, formatting inconsistencies, normalization, and binning, Analyze real-world datasets through exploratory data analysis (EDA),, Apply data operation techniques using dataframes to organize, summarize, and interpret data distributions, correlation analysis, and data pipelines., Apply core machine learning algorithms such as regression, classification, clustering, and dimensionality reduction using Python and scikit-learn. , Use these models to generate predictions and support data-driven decision-making.	9

DR. MADHU SHAMBHU SHUKLA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Contents : Unit	Topics	Contact Hours
5	**Statistical Analysis and Visualizations using R Programming** Creating Datasets in R, Importing Data into R, Data Manipulation Techniques using R, Data Visualization Techniques, Statistical Applications using R (Logical Regression, Hierarchical Clustering, PCA for Dimensionality Reduction)., Libraries, Data Handling capabilities, , Data Modeling Algorithms, Data Visualization, Statistical Analysis	8
6	**Reproducible Research Using R** Reproducible Research using R and Rstudio , knitr, rmarkdown, bookdown, interactive document, shiny presentation, shiny web application	8
Total Hours		**48**

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	**Practical 1** Initial practical will introduce R, Python programming tools and its feature/power of implementation of problems.	4
2	**Practical 2** Demonstrate working knowledge of tools such as Jupyter notebooks and RStudio and utilize their various features. Create and manage source code for data science using Git repositories and GitHub, Libraries & Packages (Exploratory data analysis (EDA) using Pandas and NumPy, scikit-learn, SciPy, TensorFlow to uncover patterns and insights), Web Scraping (learning about tools, libraries and ethical considerations), Data Visualization using Matplotlib, Seaborn, and Plotly. R Libraries for Big Data (ggplot2, dplyr, tidyverse, shiny).	4
3	**Practical 3** Lab on various types of Machine Learning regression models using R and Python.	4
4	**Practical 4** The labs must emphasize on practical aspects of clustering, rule mining, hurdles dusting big data analytics.	4
5	**Practical 5** Practical related to inferential statistics and various hypothetical tests are implemented.	4
6	**Practical 6** Practical related to advanced topics like dimension reduction using Principal Component Analysis and Regularization techniques are implemented with various aspects using R and Python Programming.	4

DR. MADHU SHAMBHU SHUKLA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
7	**Practical 7** Implement practical related to advance techniques of deep learning, reinforcement learning models using TensorFlow and PyTorch.	4
Total Hours		**28**

Textbook :

1 The Elements of Statistical Learning, Trevor Hastie , Robert Tibshirani , Jerome Friedman, Springer New York, NY, 2009

References:

1 Applied statistics and probability for engineers, Applied statistics and probability for engineers, Montgomery, Douglas C., and George C. Runger, John Wiley & Sons, 2010

2 Machine Learning (in Python and R) For Dummies, Machine Learning (in Python and R) For Dummies, John Paul Mueller, Luca Massaron, John Wiley & Sons, 2016

3 Research Methodology Methods & Techniques, Research Methodology Methods & Techniques, C. R. Kothari, New Age International Publisher, 2009

Suggested Theory Distribution:

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery					
Remember / Knowledge	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Higher order Thinking / Creative**
20.00	25.00	25.00	15.00	10.00	5.00

Instructional Method:

1 The course delivery method will depend upon the requirement of content and need of students. The teacher in addition to conventional teaching method by black board, may also use any of tools such as demonstration, role play, Quiz, brainstorming, MOOCs etc.

2 The internal evaluation will be done on the basis of continuous evaluation of students in the laboratory and class-room.

3 Practical examination will be conducted at the end of semester for evaluation of performance of students in laboratory.

4 Students will use resources like online videos, NPTEL course videos, e- courses, Virtual Laboratory

Supplementary Resources:

1 https://onlinecourses.nptel.ac.in/noc20_cs72/course

2 https://onlinecourses.nptel.ac.in/noc25_ma20/preview

DR. MADHU SHAMBHU SHUKLA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Supplementary Resources:

3 https://nptel.ac.in/courses/111104100

4 https://www.coursera.org/specializations/data-science-foundations-r#about

DR. MADHU SHAMBHU SHUKLA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)