DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

LIST OF COURSES OFFERED FOR HONOR PROGRAM (R23)

(DATA SCIENCE)

Course code	Course Title	Contact hours/week				Credits
		L	T	P	Total	
23DSH1	Statistics for Data Science	3	0	0	3	3
23DSH2	Principles of Computer vision	3	0	0	3	3
20DSH3	Optimization Techniques for Data Analytics	3	0	0	3	3
20DSH4	Data Visualization and Communication	3	0	0	3	3
23DSH5	Data Science using Python Lab	0	0	3	3	1.5
23DSH6	Data Visualization Lab (using Tableau/Power BI/ Matplotlib)	0	0	3	3	1.5

B.Tech.(CSE)	23DSH1-STATISTICS FOR DATA SCIENCE	L	T	P	Cr.
		3	0	0	3

Pre-requisite : Mathematics.

Course Educational Objective:

- To know basic concepts of statistics like Data Visualization and Estimation.
- To understand Regression Models, Time-Series Data and testing of hypothesis concepts.

Course Outcomes: At the end of this course, the student will be able to

CO1: Analyze the Visualization & distribution of data **(Apply-L3)**

CO2: Apply the inference tests for data **(Apply-L3)**

CO3: Apply the various regression models for analyzing the data **(Apply-L3)**

CO4: Perform various growth curves, trend to measure seasonal indices **(Apply-L3)**

CO5: Understand the fundamentals of classification of Logistic Regression **(Understand-L2)**

UNIT–I: Data Visualization and Distributions

Data Visualization Techniques: Introduction to Statistical methods- Exploratory Data Analysis- Charts (Line, Pie, Bar); Plots (Bubble, Scatter); Maps (Heat, Dot Distribution); Diagrams (Trees and Matrices)-Principal Components Analysis.

Introduction to Data Distributions: Probability Distributions - Discrete (Binomial, Poisson), Continuous Distributions (Normal, Exponential).

UNIT – II: Hypothesis Testing

Introduction to Parametric Estimation - Parametric Confidence Intervals - Choosing a Statistic - Hypothesis Testing - Parametric tests - the t-test - Applications to Hypothesis Testing - Pair wise comparisons.

UNIT-III: Linear Regression and Multiple Regression

Regression: Linear Regression, Curvilinear Regression - Exponential Regression- Polynomial Regression- Power Model.

Practical Examples - The nature of the 'relationship' - Multiple Linear Regression - Important measurements of the regression estimate - Multiple Regression with Categorical Explanatory Variables - Inference in Multiple Regression - Variable Selection.

UNIT – IV: **Time Series**

Time series: Significance of Time series analysis, Components of Time series, Secular trend: Graphic method, Semi-average method, Method of moving averages, **Method of least squares:** straight line and non-linear trends, Logarithmic methods–Exponential trends, Growth curves, **Seasonal Variations:** Method of simple averages, Ratio – to –trend method, Ratio - to-moving averages method, Link relative's method**.**

UNIT – V: Logistic Regression

The classification Problem-Logistic Regression Setup-Interpreting the Results- Comparing Models - Classification using Logistic Regression,

Text Books:

1. Elizabeth Purdom,"Statistical methods for Data science" (08.05.2023)

2. K. Murugesan, P. Gurusamy“Probability,Statistics and Random Processes”, Anuradha Publications (January-2009), ISBN-10(8189638289)/ISBN-13(978-8189638283).

References:

1. Manoj Kumar Srivastava and NamitaSrivastava ,Statistical Inference–Testing of Hypotheses, Prentice Hall of India, 2014.
2. Robert V Hogg, Elliot A Tannis and Dale L.Zimmerman, Probability and Statistical Inference,9thedition, Pearson publishers,2013.
3. Chris Chatfield, “The analysis of time series an introduction,” 5th edition, Chapman &Hall/CRC.
4. Peter J. Brockwell, Richard A.Davis, “Introduction to Time series and Forecasting,” Second edition, Springer.

B. Tech (CSE) **23DSH2- PRINCIPLES OF COMPUTER VISION**

L	T	P	Cr.
3	0	0	3

Pre-requisites: Nil

Course Educational Objectives: To introduce students the fundamentals of image formation; To introduce students the major ideas, methods, and techniques of computer vision and pattern recognition; To develop an appreciation for various issues in the design of computer vision and object recognition systems; and To provide the student with programming experience from implementing computer vision and object recognition applications.

Course Outcomes: At the end of the course, students will be able to,

CO1: Identify basic concepts, terminology, theories, models and methods in the field of computer vision. **(Understand-L2)**
CO2: Understand various feature extraction methods and its significance. **(Understand-L2)**
CO3: Describe basic methods of computer vision related to multi-scale representation **(Apply-L3)**
CO4: Describe basic methods of computer vision related to segmentation **(Understand-L2)**
CO5: Understand basic methods of computer vision related to object recognition **(Understand-L2)**

UNIT I

Image Formation Models: Monocular imaging system, Orthographic& Perspective Projection, Camera model and Camera calibration, Binocular imaging systems

UNIT II

Image Processing and Feature Extraction: Image representations (continuous and discrete), Edge detection

UNIT III

Motion Estimation: Regularization theory, Optical computation, Stereo Vision, Motion estimation, Structure from motion

UNIT IV

Shape Representation and Segmentation: Deformable curves and surfaces, Snakes and active contours, Level set representations, Fourier and wavelet descriptors, Medial representations, Multiresolution analysis

UNIT V

Object recognition: Hough transforms and other simple object recognition methods, Shape correspondence and shape matching, Principal Component analysis, Shape priors for recognition.

TEXT BOOKS:

1. Computer Vision - A modern approach, by D. Forsyth and J. Ponce, Prentice Hall Robot Vision, by B. K. P. Horn, McGraw-Hill.

REFERENCE BOOKS:

1. Richard Szeliksy "Computer Vision: Algorithms and Applications" (http://szeliski.org/Book/)
2. Haralick& Shapiro, "Computer and Robot Vision", Vol II
3. G_erardMedioni and Sing Bing Kang "Emerging topics in computer vision"
4. Emanuele Trucco and AllessandroVerri "Introductory Techniques for 3-D Computer Vision", Prentice Hall, 1998.

B.Tech. (CSE)	23DSH3- OPTIMIZATION TECHNIQUES FOR DATA ANALYSIS	L	T	P	Cr.
		3	0	0	3

Course Outcomes: On completion of the course the student should be able to
CO1. Understand the concept of optimality criteria for various types of optimization problem. **(Understand-L2)**
CO2. Analyze optimization algorithms for Linear Programming (**Analyze-L4)**
CO3. Solve various constrained and unconstrained nonlinear Programming problems. **(Apply-L3)**
CO4. Apply the modern optimization methods to provide optimal solution for a given problem. **(Apply-L3)**

UNIT I:
Introduction to Optimization: Introduction, Historical Development, Engineering Applications of Optimization, Statement of an Optimization Problem, Classification of Optimization Problems.
UNIT II:
Classical Optimization Techniques: Single-Variable Optimization, Multivariable Optimization with No Constraints, Multivariable Optimization with Equality Constraints, Multivariable Optimization with Inequality Constraints.
UNIT III:
Linear Programming
Introduction, Applications of Linear Programming, Standard Form of a Linear Programming Problem, Geometry of Linear Programming Problems, Solution of a System of Linear Simultaneous Equations, Pivotal Reduction of a General System of Equations Simplex Method: Motivation of the Simplex Method, Simplex Algorithm, Improving a Nonoptimal Basic Feasible Solution, Two Phases of the Simplex Method.
UNIT IV:
Nonlinear Programming Algorithms: Unconstrained Algorithms – Direct Search Method, Gradient method, Constrained Algorithms - Separable Programming, Quadratic Programming, Chance- Constrained Programming, Linear Combinations method, SUMT Algorithm.
Case Study 1: Chance Constrained Problem
UNIT V:
Modern Methods of Optimization
Introduction, Genetic Algorithms, Simulated Annealing, Particle Swarm Optimization, Ant Colony Optimization, Optimization of Fuzzy Systems, Neural Network-Based Optimization
Case Study 2: Travelling Salesperson Problem

Text books:

1. Singiresu S Rao, “Engineering Optimization Theory and Practice", John Wiley and sons, 4th Edition, 2009.
2. HamdyA.Taha, “OprationResearch : An Introduction”, 8th Edition, Pearson Prentice Hall, 2007.
3. Paulo Cortez, “Modern Optimization with R”, Springer series, 2014.

Reference Books:

1. S.Rao, “Engineering optimization: Theory and practice”, 4th Edition, New Age International, 2009.

2. Edwin K. P. Chong and Stanislaw. Zak "An Introduction to Optimization", John Wiley and sons, 2nd Edition 2001.
3. Andreas Antoniou, "Practical Optimization Algorithms and Engineering Applications",
4. An Introduction to Optimization Edwin K., P. Chong & Stanislaw h. Zak.
5. Andreas Antoniou. "Practical Optimization Algorithms and Engineering Applications", Springer Series, 2007

B.Tech. (CSE) 23DSH4- DATA VISUALIZATION AND COMMUNICATION

L	T	P	Cr.
3	0	0	3

Course Objectives:

1. To introduce the fundamental principles and importance of data visualization in analytics and decision-making.
2. To equip students with the skills to choose appropriate visualization methods for different types of data.
3. To enable learners to utilize modern visualization tools and libraries for data exploration and presentation.
4. To develop students' ability to communicate insights effectively using data storytelling techniques.
5. To apply design thinking and ethical practices in developing clear, accurate, and compelling visual representations of data.

Course Outcomes:

CO1: Understand the basic concepts, types, and significance of data visualization**. (Understand-L2)**

CO2: Apply effective design principles and appropriate chart types to visualize data accurately. **(Apply-L3)**

CO3: Utilize tools such as Excel, Tableau, Power BI, and Python libraries (Matplotlib, Seaborn, Plotly) to create visualizations**.(Apply-L3)**

CO4: Construct coherent data stories and present insights tailored to different audience types. **.(Apply-L3)**

CO5: Evaluate and apply advanced visualization techniques with a focus on ethics and clarity in communication. **.(Apply-L3)**

UNIT I: Introduction to Data Visualization

Importance and purpose of data visualization; Types of data: categorical, ordinal, quantitative, time series; Visual perception and cognitive principles; Chart types overview: bar, line, scatter, pie, histogram; Data-Ink ratio and Tufte's principles.

UNIT II: Principles of Effective Visualization Design

Choosing the right graph; Design for clarity: scale, axis, and labels; Visual hierarchy and Pre attentive attributes; Color theory in visualization; Avoiding misleading visuals and chartjunk.

UNIT III: Tools and Technologies for Visualization

Overview of popular visualization tools: Python (Matplotlib, Seaborn, Plotly); Visualization in Python: basic usage and examples; Creating and customizing visualizations using open-source tools.

UNIT IV: Data Storytelling and Communication

Data narrative techniques; Framing your story: beginning, middle, end; Annotating for insight; Audience-centric communication; Presentation dos and don'ts.

UNIT V: Advanced Visualization & Real-World Applications

Complex visualizations: maps, heatmaps, treemaps, network graphs; Visualizing uncertainty and variability; Ethics in data visualization; Case studies from business, journalism, healthcare, etc.

Text Books:

1. Storytelling with Data: A Data Visualization Guide for Business Professionals, Cole Nussbaumer Knaflic, Wiley, 2015.

Reference Books:

1. Data Points: Visualization That Means Something, Nathan Yau, Wiley, 2013.
2. The Visual Display of Quantitative Information, Edward R. Tufte, Graphics Press, 2001.
3. The Truthful Art: Data, Charts, and Maps for Communication, Alberto Cairo, New Riders, 2016.
4. Visualizing Data: Exploring and Explaining Data with the Processing Environment, Ben Fry, O'Reilly, 2008.

		L	T	P	Cr.
B.Tech.(CSE)	23DSH5-DATA SCIENCE USING PYTHON LAB	0	0	3	1.5

Prerequisite:

1. Knowledge in Computer Programming.
2. Background in linear algebra ,data structures and algorithms.

Course Educational Objective : The main objective of the course is to inculcate the basic understanding of Data Science and its practical implementation using Python.

Course Outcomes: At the end of this course ,the student will be able to

CO1: Apply the basic concepts of Python libraries.**(Apply-L3)**

CO2: Perform various operations using Numpy & Pandas.**(Apply-L3)**

CO3: Apply the Advanced packages like NLTK,SciKit-Learn and Spicy.**(Apply-L3)**

CO4: Improve individual/teamwork skills, communicate & report writing skills with Ethical values.

List of Experiments:

1. Creating a NumPy Array
   a. Basic ndarray
   b. Array of zeros
   c. Array of ones
   d. Random numbers in ndarray
   e. An array of your choice
   f. Imatrix in NumPy
   g. Evenly spaced ndarray
2. The Shape and Reshaping of NumPy Array
   a. Dimensions of NumPy array
   b. Shape of NumPy array
   c. Size of NumPy array
   d. Reshaping a NumPy array
   e. Flattening a NumPy array
   f. Transpose of a NumPy array
3. Expanding and Squeezing a NumPy Array
   a. Expanding a NumPy array
   b. Squeezing a NumPy array
   c. Sorting in NumPy Arrays
4. Indexing and Slicing of NumPy Array
   a. Slicing 1-D NumPy arrays
   b. Slicing 2-D NumPy arrays
   c. Slicing 3-D NumPy arrays
   d. Negative slicing of NumPy arrays
5. Stacking and Concatenating Numpy Arrays
   a. Stacking ndarrays
   b. Concatenating ndarrays
   c. Broadcasting in Numpy Arrays

6. Perform following operations usingpandas
   a. Creating data frame
   b. concat()
   c. Setting conditions
   d. Adding a new column
7. Perform following operations using pandas
   a. Filling NaN with string
   b. Sorting based on column values
   c. groupby()
8. Read the following file formats using pandas
   a. Text files
   b. CSV files
   c. Excel files
   d. JSON files
9. Read the following file formats
   a. Pickle files
   b. Image files using PIL
   c. Multiple files using Glob
   d. Importing data from database
10. Demonstrate web scraping using python
11. Perform following preprocessing techniques on loan prediction dataset
    a. Feature Scaling
    b. Feature Standardization
    c. Label Encoding
    d. One Hot Encoding
12. Perform following visualizations using matplotlib
    a. Bar Graph
    b. Pie Chart
    c. Box Plot
    d. Histogram
    e. Line Chart and Subplots
    f. Scatter Plot
13. Getting started with NLTK, install NLTK using PIP
14. Python program to implement with Python SciKit-Learn & NLTK
15. Python program to implement with Python NLTK/Spicy/PyNLPI

Web References:

1. https://www.analyticsvidhya.com/blog/2020/04/the-ultimate-numpy-tutorial-for-data-science-beginners/
2. https://www.analyticsvidhya.com/blog/2021/07/data-science-with-pandas-2-minutes-guide-to-key-concepts/
3. https://www.analyticsvidhya.com/blog/2020/04/how-to-read-common-file-formats-python/
4. https://www.analyticsvidhya.com/blog/2016/07/practical-guide-data-preprocessing-python-scikit-learn/
5. https://www.analyticsvidhya.com/blog/2020/02/beginner-guide-matplotlib-data-visualization-exploration-python/6.
6. https://www.nltk.org/book/ch01.html

B. Tech (CSE) **23DSH6-DATA VISUALIZATION LAB**

L	T	P	Cr.
0	0	3	1.5

Course Objectives:

- To introduce various types of data visualizations using open-source tools.
- To implement clear, well-labeled visualizations using Python libraries.
- To develop skills in interpreting and customizing visual plots.
- To apply visual communication principles for presenting insights effectively.

Course Outcomes (COs)

CO1: Understand and apply different types of data visualizations using open-source libraries.
CO2: Create clear and well-structured visualizations using Matplotlib and Seaborn.
CO3: Develop interactive and advanced plots using Plotly.
CO4: Demonstrate effective communication of insights through labeled, customized, and annotated plots.

List of Lab Experiments

1. Introduction to Python visualization stack: Matplotlib, Seaborn, Plotly
2. Creating basic plots with Matplotlib: line, bar, pie, histogram
3. Customizing plots: titles, labels, legends, colors, gridlines, figure size
4. Subplots and figure layout adjustments in Matplotlib
5. Introduction to Seaborn: statistical plots (countplot, distplot, boxplot, violin plot)
6. Multivariate and categorical visualization using Seaborn: scatterplot, pairplot, heatmap
7. Customizing Seaborn plots: palettes, styles, themes, and annotations
8. Introduction to interactive visualizations using Plotly (bar, line, scatter)
9. Advanced Plotly visualizations: bubble chart, area chart, subplots, and multiple axes
10. Time series visualization using Matplotlib and Plotly
11. Creating and visualizing grouped and stacked charts
12. Annotating visualizations: highlighting key data points, trends, and narratives
13. Comparative visualizations: combining plots to compare variables or categories
14. Mini-project: Visualize a real-world dataset using multiple chart types
15. Project Presentation & Viva

Text Books:

1. Storytelling with Data: A Data Visualization Guide for Business Professionals, Cole Nussbaumer Knaflic, Wiley, 2015.

Reference Books:

1. Data Points: Visualization That Means Something, Nathan Yau, Wiley, 2013.
2. The Visual Display of Quantitative Information, Edward R. Tufte, Graphics Press, 2001.
3. The Truthful Art: Data, Charts, and Maps for Communication, Alberto Cairo, New Riders, 2016.
4. Visualizing Data: Exploring and Explaining Data with the Processing Environment, Ben Fry, O'Reilly, 2008.