

HONOURS

S.No	Course Code	Course Title	Scheme of Instructions Hours per Week				Scheme of Examination Maximum Marks		
			L	T	P	C	I	E	Total
1	23HRCSD1	Data Science For Business	3	-	-	3	30	70	100
2	23HRCSD2	Software Project Management Using Agile	3	-	-	3	30	70	100
3	23HRCSD3	Software Defined Data Center	3	-	-	3	30	70	100
4	23HRCSD4	Medical Image Data Processing	3	-	-	3	30	70	100
5	23HRCSD5	Data Analytics	3	-	-	3	30	70	100
6	23HRCSD6	Data Analytics Lab	-	-	3	1.5	30	70	100
7	23HRCSD7	Advanced Python Programming For Data Science Lab	-	-	3	1.5	30	70	100

23HRCSD1	DATA SCIENCE FOR BUSINESS	L	T	P	C
		3	0	0	3

Course Objectives:

- Expose with the basic rudiments of business intelligence system
- Expose with different data analysis tools and techniques

Course Outcomes:

At the end of the course the students will be able to

- Understand the fundamentals of business intelligence.
- Applying link to data mining with business intelligence.
- Apply various modelling techniques.
- Understand the data analysis and knowledge delivery stages.
- Apply business intelligence methods to various situations and decide on appropriate technique.

UNIT I

Lecture 8 Hrs

Introduction – Business problems and Data Science Solutions, Introduction to Predictive modeling: From Correlation to Supervised Segmentation

UNIT II

Lecture 8 Hrs

Fitting the Data- Fitting a Model to Data, Overfitting and its Avoidance

UNIT III

Lecture 9Hrs

Similarity, Neighbors, and Clusters, Decision Analytic Thinking: What is a Good model

UNIT IV

Lecture 8 Hrs

Representing and Mining text, Decision Analytic Thinking II: Toward Analytic Engineering

UNIT V

Lecture 9 Hrs

Other Data Science Tasks and Techniques, Data Science and Business Strategy

Textbooks:

1. Foster Provost and Tom Fawcett, Data Science for Business, O'Reilly, 2013.

Reference Books:

1. Efraim Turban, Ramesh Sharda, DursunDelen, —Decision Support and Business Intelligence Systems, 9 th Edition, Pearson 2013.
2. Larissa T. Moss, S. Atre, —Business Intelligence Roadmap: The Complete ProjectLifecycle of Decision Making, Addison Wesley, 2003.
3. Carlo Vercellis, —Business Intelligence: Data Mining and Optimization for DecisionMaking, Wiley Publications, 2009.
4. David Loshin Morgan, Kaufman, —Business Intelligence: The Savvy Manager'sGuidel, Second Edition, 2012.

Online Learning Resources:

- 1.Edx: IBM Data Warehousing and BI Analytics

23HRCSD2	SOFTWARE PROJECT MANAGEMENT USING AGILE	L	T	P	C
		3	0	0	3

Pre-requisites: Software Engineering Fundamentals**Course Objectives:**

- Teach how to manage a Project
- Discuss Agile method of handling projects

Course Outcomes:

After completion of the course, students will be able to

- Apply Agile methodology for software development
- Critically analyze quality of software
- Estimate the software cost

UNIT I Introduction, The Agile Business Case Lecture 8Hrs

History, Background, and the Manifesto, Traditional Lifecycle, Agile Lifecycle, Scaling for Enterprise Agile, Four Agile Methodologies. The Agile Business Case: The Business Case, Business Value Models, Project Balance Sheet, Building the Business Case by Levels

UNIT II Quality in the Agile Space Lecture 9Hrs

Quality Values and Principles, Thought Leaders and Agile Quality, Sampling for Quality Validation, Agile in the Waterfall: First Principles and Requisite Conditions, The Black Box, Interfaces, and Connectivity, Governing

UNIT III Scope and Requirements Lecture 9Hrs

Developing the Scope and Requirements: Agile Scope, Envisioning, Requirements, Planning at a Distance Planning and Scheduling: Planning in the Enterprise Context, Scheduling, Other Plans in the Enterprise Agile Project

UNIT IV Estimating Cost and Schedule Lecture 8Hrs

The Nature of Estimates, Drivers on Cost and Schedule, Building Estimates Teams Are Everything: The Social Unit, Principle and Values Guide Teams, Teams Are Building Blocks, Some Teams Work; Others Do Not, Matrix Management in the Agile Space

UNIT V Governance, Managing Value Lecture 8Hrs

Governance Is Built on Quality Principles, Governance Verifies Compliance Managing Value: Defining and Accounting for Value, Burn-down Charts and Value Scorecards.

Textbooks:

1. John C. Goodpasture, PMP, —Project Management the Agile Way, Second Edition, J. Ross Publishing 2016.

Reference Books:

1. Kalpesh Ashar, Agile Essentials you always wanted to know, Vibrant publishers, 2020
2. Jutta Eckstein, Agile Software development in the large: Diving into the Deep, Jutta Eckstein Publisher, 2022

Online Learning Resources:

1. Coursera: Agile Project Management offered by Google
2. Coursera: Alex Cowan, Agile Development Specialization

23HRCSD3	SOFTWARE DEFINED DATA CENTER	L	T	P	C
		3	0	0	3

Course Objectives:

- Introduce conventional Data Centers followed by Modern Data Centers
- To discuss various software elements of modern data centers
- Explain Virtualization concepts for Data Centers
- Discuss Compute, Storage and Network virtualization

Course Outcomes:

After completion of the course, students will be able to

- Understanding of difference between Conventional Data Center Vs Modern Data Centers
- Differentiate Cloud computing and Software Defined Data Centers
- Differentiate Virtualization with conventional techniques
- Explore the techniques of Software Defined Compute, Storage and Networking components
- Able Manage Software Defined Data Centers and Develop the techniques for future Data Centers.

UNIT I Introduction

Lecture 12Hrs

Data Center evolution, A history of Modern Data Center, Focus on cost reduction, Focus on

Customer service in the business, Flattening of the IT organization, IT as an operational Expense, Monolithic Storage Array rise and fall, Move From Disk to Flash, Emergence of Convergence, The Role of Cloud computing.

UNIT II Emerging Data Center Trends

Lecture 12Hrs

Emergence of SDCC, Commoditization of Hardware, Software Defined – Compute, Storage,

Networking and Security, Software Defined Storage (SDS), Hyperconvergence, Hyper Converged Infrastructure(HCI) and SDS relationship, Flash in Hyperconvergence, Modern IT business Requirements.

UNIT III Data Center Agility

Lecture 12Hrs

Principles and Strategies, Transform Data Center, Align Data Center and Business Needs, Server virtualization, VDI, Eliminate and Implement Monolithic to Hyperconvergence, Full Stack Management.

UNIT V Hyper converged Infrastructure

Lecture 12Hrs

Software Defined Storage, SDS comparison to Traditional Storage, SDS requirements, SDS in Hyperconverged, Hyperconvergence Design Model, Virtual Storage appliances, Appliance vs. Software/Reference Architecture,

UNIT V Future Data Centers

Lecture 12Hrs

Data growth, Storage capacity, flash storage deployment, Deployment Experiences SDS and HCI, IT transformations- Automation, Orchestration, DevOps, Open Standards and Interoperability, Performance Benchmarking Standards, Future Trends, Containers Instead of virtual machines, Open Source tools, Beyond Today's Flash, Pooling of Resources.

Textbooks:

1. Building a Modern Data Center, Principles and Strategies of Design, Scott D.Lowe, James Green, David Davis. Actual Tech Media, 2016.

Reference Books:

1. Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center, Second Edition, HwaiyuGeng P.E., 2021 John Wiley & Sons.

23HRCSD4	MEDICAL IMAGE DATA PROCESSING	L	T	P	C
		3	0	0	3

Course Objectives:

- Understand the significance of image process in medical industry
- Teach the process of extracting correct information in medical images

Course Outcomes:

The course is designed to

- Analyze medical images
- Apply image processing techniques to medical images

UNIT I Basics of Medical Image Sources Lecture 8 Hrs
Radiology, The Electromagnetic Spectrum, Basic X-Ray Physics, Attenuation and Imaging, Computed Tomography, Magnetic Resonance Tomography, Ultrasound, Nuclear Medicine and Molecular Imaging, Other Imaging Techniques, Radiation Protection and Dosimetry
Image Processing in Clinical Practice: Application Examples, Image Databases, Intensity Operations, Filter Operations, Segmentation, Spatial Transforms, Rendering and Surface Models, Registration, CT Reconstruction

UNIT II Image Representation Lecture 10 Hrs
Pixels and Voxels, Gray Scale and Color Representation, Image File Formats, Dicom, Other Formats – Analyze 7.5, NIFTI And Interfile, Image Quality and The Signal-To-Noise Ratio, Practical Lessons
Operations in Intensity Space: The Intensity Transform Function and The Dynamic Range, Windowing, Histograms and Histogram Operations, Dithering and Depth, Practical Lessons

UNIT III Filtering and Transformations, Segmentation Lecture 8 Hrs
The Filtering Operation, The Fourier Transform, Other Transforms, Practical Lessons
Segmentation: The Segmentation Problem, ROI Definition and Centroids, Thresholding, Region Growing, More Sophisticated Segmentation Methods, Morphological Operations, Evaluation of Segmentation Results

UNIT IV Spatial Transforms Lecture 9 Hrs
Discretization – Resolution and Artifacts, Interpolation and Volume Regularization, Translation and Rotation, Reformatting, Tracking and Image-Guided Therapy
Rendering and Surface Models: Visualization, Orthogonal and Perspective Projection, and The Viewpoint, Raycasting, Surface-Based Rendering

UNIT V Registration, CT Reconstruction Lecture 8 Hrs
Fusing Information, Registration Paradigms, Merit Functions, Optimization Strategies, Some General Comments, Camera Calibration, Registration to Physical Space, Evaluation of Registration Results
CT Reconstruction: Introduction, Radon Transform, Algebraic Reconstruction, Some Remarks on Fourier Transform and Filtering, Filtered Back projection

Textbooks:

1. Wolfgang Birkfellner, —Applied Medical Image Processing, Second Edition, CRC Press.

Reference Books:

1. Sinha G.R., Medical Image Processing Concepts and Application, PHI, 2014
2. Geoff Dougherty, Digital Image Processing for Medical Applications, Cambridge university press, 2010

Online Learning Resources: Coursera: Pranav Rajpurkar, AI for Medical Diagnosis

23HRCSD5	DATA ANALYTICS	L	T	P	C
		3	0	0	3

Course Outcomes:

After completion of this course students will be able to

1. Understand the impact of data analytics for business decisions and strategy
2. Carry out data analysis/statistical analysis
3. To carry out standard data visualization and formal inference procedures
4. Design Data Architecture
5. Understand various Data Sources

Unit-1**Data Management:**

Design Data Architecture and manage the data for analysis, understand various sources of Data like Sensors/Signals/GPS etc. Data Management, Data Quality(noise, outliers, missing values, duplicate data) and Data Processing & Processing.

Unit -2**Data Analytics:**

Introduction to Analytics, Introduction to Tools and Environment, Application of Modeling in Business, Databases & Types of Data and variables, Data Modeling Techniques, Missing Imputations etc. Need for Business Modeling.

Unit -3

Regression – Concepts, Blue property assumptions, Least Square Estimation, Variable Rationalization, and Model Building etc. Logistic Regression: Model Theory, Model fit Statistics, Model Construction, Analytics applications to various Business Domains etc.

Unit – 4**Object Segmentation:**

Regression Vs Segmentation – Supervised and Unsupervised Learning, Tree Building – Regression, Classification, Overfitting, Pruning and Complexity, Multiple Decision Trees etc. Time Series Methods: Arima, Measures of Forecast Accuracy, STL approach, Extract features from generated model as Height, Average Energy etc and Analyze for prediction

Unit – 5**Data Visualization:**

Pixel-Oriented Visualization Techniques, Geometric Projection Visualization Techniques, Icon-Based Visualization Techniques, Hierarchical Visualization Techniques, Visualizing Complex Data and Relations.

TEXT BOOKS:

1. Student's Handbook for Associate Analytics – II, III.
2. Data Mining Concepts and Techniques, Han, Kamber, 3rd Edition, Morgan Kaufmann Publishers.

REFERENCE BOOKS:

1. Introduction to Data Mining, Tan, Steinbach and Kumar, AddisonWiseley, 2006.
2. Data Mining Analysis and Concepts, M. Zaki and W. Meira
3. Mining of Massive Datasets, Jure Leskovec Stanford Univ. AnandRajaramanMilliway Labs Jeffrey D Ullman Stanford Univ.

23HRCSD6	DATA ANALYTICS LAB	L	T	P	C
		3	0	0	3

1. **Data Acquisition and Storage**
 - Read and store data from **CSV, JSON, and SQL databases** using pandas.
2. **Handling Missing Data**
 - Identify and handle **missing values** using **imputation techniques** (mean, median, mode).
3. **Data Cleaning and Preprocessing**
 - Detect and remove **duplicates, outliers, and inconsistencies** using numpy and pandas.
4. **Data Transformation and Normalization**
 - Apply **scaling techniques (MinMax, StandardScaler)** using scikit-learn.
5. **Working with Sensor Data**
 - Process **IoT sensor data** from **GPS, signals, and time-series logs**.
6. **xploratory Data Analysis (EDA)**
 - Use matplotlib, seaborn, and pandas-profiling for data visualization and summary statistics.
7. **Feature Engineering**
 - Create **new features** using transformation techniques (log, polynomial, binning).
8. **Implementation of Multiple Linear Regression**
 - Use statsmodels and scikit-learn to fit **regression models** and analyze assumptions.
9. **Polynomial Regression**
 - Implement polynomial regression to capture **non-linearity in data**.
10. **Regularization Techniques**
 - Apply **Lasso and Ridge regression** to prevent overfitting.
11. **Logistic Regression for Classification**
 - Implement and evaluate **logistic regression** on a real dataset.
12. **K-Means Clustering for Image Segmentation**
 - Perform **unsupervised segmentation** on images using OpenCV and scikit-learn.
13. **Decision Tree for Classification**
 - Build a **classification tree** to predict categories from structured data.
14. **Random Forest for Feature Selection**
 - Use **feature importance scores** to identify key predictors.
15. **3D Data Visualization**
 - Use plotly and matplotlib for **3D scatter plots** and **surface plots**.
16. **Hierarchical Data Visualization**
 - Create **tree maps and dendrograms** for hierarchical data.

23HRCSD7	ADVANCED PYTHON PROGRAMMING FOR DATA SCIENCE LAB	L	T	P	C
		3	0	0	3

Course Objectives:

- Understand the python Programming Language libraries.
- Exposure on Solving of data science problems.
- Understand The classification and Regression Model.

Course Outcomes (CO):

After completion of the course, students will be able to

- Apply principles and techniques for optimizing the performance of Python numeric applications
- Implement parallel computing applications using Python
- Develop GPU accelerated Python applications

List of Experiments:

W-1 The number of birds banded at a series of sampling sites has been counted by your field crew and entered into the following list. The first item in each sublist is an alphanumeric code for the site and the second value is the number of birds banded. Cut and paste the list into your assignment and then answer the following questions by printing them to the screen.

```
data = [['A1', 28], ['A2', 32], ['A3', 1], ['A4', 0],
        ['A5', 10], ['A6', 22], ['A7', 30], ['A8', 19],
        ['B1', 145], ['B2', 27], ['B3', 36], ['B4', 25],
        ['B5', 9], ['B6', 38], ['B7', 21], ['B8', 12],
        ['C1', 122], ['C2', 87], ['C3', 36], ['C4', 3],
        ['D1', 0], ['D2', 5], ['D3', 55], ['D4', 62],
        ['D5', 98], ['D6', 32]]
```

1. How many sites are there?
2. How many birds were counted at the 7th site?
3. How many birds were counted at the last site?
4. What is the total number of birds counted across all sites?
5. What is the average number of birds seen on a site?

- 6. What is the total number of birds counted on sites with codes beginning with C? (don't just identify this site by eye, in the real world there could be hundreds or thousands of sites)
- W-2 1. Multiplication of two Matrices in Single line using Numpy in Python
 - 2. Transpose a matrix in Single line using Python
 - 3. Python program to print checkerboard pattern of nxn using numpy
- W-3 Reading different types of data sets (.txt, .csv) from Web and disk and writing in file in specific disk location.

Reading Excel data sheet

Reading XML dataset

- W-4 1. Find the data distributions using box and scatter plot.
 - 2. Find the outliers using plot.
 - 3. Plot the histogram, bar chart and pie chart on sample data
- W-5 1. Find the correlation matrix.
 - 2. Plot the correlation plot on dataset and visualize giving an overview of relationships among data on iris data.
 - 3. Analysis of covariance: variance (ANOVA), if data have categorical variables on iris data.
- W-6 Import a data from web storage. Name the dataset and now do LogisticRegression to find out relation between variables that are affecting the admission of a student in a institute based on his or her GRE score, GPA obtained and rank of the student. Also check the model is fit or not. Require (foreign), require (MASS).
- w-7 Decision Tree Classification, attribute selection measures, and how to build and optimize Decision Tree Classifier using Python Scikit-learn
- W-8 Apply multiple regressions, if data have a continuous independent variable. Apply on above dataset.
- W-9 Apply regression Model techniques to predict the data
- W-10 1. Install relevant package for classification.
 - 2. Choose classifier for classification problem.
 - 3. Evaluate the performance of classifier.
- W-11 Clustering algorithms for unsupervised classification.

Plot the cluster data using python with Matplotlib visualizations.
- W-12 **Case Study: Data Science in Education**

Data Science has also changed the way in which students interact with teachers and evaluate their performance. Instructors can use data science to

analyse the feedback received from the students and use it to improve their teaching.

Use Predictive modeling Data Science that can predict the drop-out rate of students based on their performance and inform the instructors to take necessary precautions.

References:

1. <https://www.w3schools.com/datascience/>
2. <https://data-flair.training/blogs/data-science-tutorials-home/>
3. <https://www.javatpoint.com/data-science>
4. https://www.tutorialspoint.com/python_data_science/index.htm
5. <https://intellipaat.com/blog/tutorial/data-science-tutorial/>

Online learning Resources/Virtual labs <https://www.vlab.co.in>