गोंय विद्यापीठ
ताळगांव पठार
गोंय - ४०३ २०६
फोन: +९१-८६६९६०९०४८

भारत 2023 INDIA

(Accredited by NAAC)

Goa University

Taleigao Plateau, Goa - 403 206
Tel : +91-8669609048
Email : registrar@unigoa.ac.in
Website: www.unigoa.ac.in

GU/Acad –PG/BoS -NEP/2023/184/3 Date:04.07.2023

CIRCULAR

The University has notified Ordinance OA-35 governing the **Master of Science in Data Science** Programme offered at the Goa Business School, Goa University Campus for implementation from the Academic year 2023-2024 onwards.

The approved Semester I and II Syllabus of the **Master of Science in Data Science** Programme is attached.

The Dean/ Vice-Deans of the Goa Business School are requested to take note of the above and bring the contents of the Circular to the notice of all concerned.

(Sanket Gaude)
Offg. Assistant Registrar – Academic-PG

To,

1. The Dean, Goa Business School, Goa University.
2. The Vice-Deans, Goa Business School, Goa University.

Copy to:

1. The Chairperson, Board of Studies in Computer Science and Technology (PG).
2. The Programme Director, M.Sc Data Science, Goa University.
3. The Controller of Examinations, Goa University.
4. The Assistant Registrar, PG Examinations, Goa University.
5. Directorate of Internal Quality Assurance, Goa University for uploading the Syllabus on the University website.

Goa University

M.Sc. in Data Science to be effective from Academic Year 2023-24

M.Sc in Data Science Programme is aimed at imparting the following core skills to the students -

1. Core Programming Skills and Techniques, including designing and coding applications, and the important principles of code design and development.
2. Data science tools and techniques, including the principles of data science, data analysis, visualisation and interpretation, and the use of "big data".
3. The application of Data Science knowledge in Research and Industry

Pathway

Fundamentals (Mathematics and Problem Solving, Programming)➔ Core Courses (AI, Machine Learning, Deep Learning, etc.,)➔ Specialization (Natural Language Processing, Computer Vision)➔Research and Dissertation

As per the above pathway vision, the structure for programme has been designed as follows

M.Sc. in Data Science to be effective from Academic Year 2023-24		
SEMESTER I		
Discipline Specific Core(DSC) Courses		
Course Code	**Course Title**	**Credits**
CSD-500	Fundamentals of Data science (Theory)	**2**
CSD-501	Fundamentals of Data science (Practical)	**2**
CSD-502	Machine learning (Theory)	**2**
CSD-503	Machine learning (Practical)	**2**
CSD-504	Mathematical Foundations for Data Science (Theory)	**2**
CSD-505	Mathematical Foundations for Data Science (Practical)	**2**
CSD-506	Fundamentals of Artificial Intelligence (Theory)	**2**
CSD-507	Fundamentals of Artificial Intelligence (Practical)	**2**
	Total Credits for DSC	**16**
Discipline Specific Elective(DSE) Courses – any one to be opted		
Course Code	**Course Title**	**Credits**
CSD-521	Domain specific Predictive Analytics	**4**
CSD-522	Design thinking for Data-Driven App Development	**4**
	Total Credits for DSE	**4**
SEMESTER II		
Discipline Specific Core(DSC) Courses		
Course Code	**Course Title**	**Credits**
CSD-508	Reinforcement learning (Theory)	**2**
CSD-509	Reinforcement learning (Practical)	**2**
CSD-510	Optimization techniques	**4**
CSD-511	MLOps (Theory)	**2**
CSD-512	MLOps (Practical)	**2**
CSD-513	Software Engineering for AI Enabled systems (Theory)	**2**
CSD-514	Software Engineering for AI Enabled systems (Practical)	**2**
	Total Credits for DSC	**16**
Discipline Specific Elective(DSE) Courses – any one to be opted		
Course Code	**Course Title**	**Credits**
CSD-523	Signal processing	**4**
CSD-524	Regression Analytics and Predictive Models	**4**
	Total Credits for DSE	**4**

<table>
<tr><th colspan="3">SEMESTER III</th></tr>
<tr><th colspan="3">Research Specific Elective(RSE) Courses – any two to be opted…</th></tr>
<tr><th>Course Code</th><th>Course Title</th><th>Credits</th></tr>
<tr><td>CSD-600</td><td>Speech Processing</td><td>4</td></tr>
<tr><td>CSD-601</td><td>Natural Language Processing</td><td>4</td></tr>
<tr><td>CSD-602</td><td>Simulation and Modelling</td><td>4</td></tr>
<tr><td>CSD-603</td><td>Deep Learning Models</td><td>4</td></tr>
<tr><td>CSD-604</td><td>Data Engineering</td><td>4</td></tr>
<tr><td>CSD-605</td><td>Sensors , Actuators and Signal Conditioning</td><td>4</td></tr>
<tr><td>CSD-606</td><td>Cloud Computing</td><td>4</td></tr>
<tr><td></td><td>Total Credits for RSE</td><td>8</td></tr>
<tr><th colspan="3">Generic Elective(GE) Courses - total 12 credits to be opted</th></tr>
<tr><th>Course Code</th><th>Course Title</th><th>Credits</th></tr>
<tr><td>CSD-621</td><td>Corporate Skills</td><td>4</td></tr>
<tr><td>CSD-622</td><td>Research Methodology</td><td>4</td></tr>
<tr><td></td><td>To be opted from Courses from other Disciplines</td><td>4</td></tr>
<tr><td></td><td>Total Credits for GE Courses</td><td>12</td></tr>
<tr><th colspan="3">SEMESTER IV</th></tr>
<tr><th colspan="3">One Research Specific Elective(RSE) Course to be opted from the RSE list in consultation with the Mentor. It can be completed in Semester 3.</th></tr>
<tr><th>Course Code</th><th>Course Title</th><th>Credits</th></tr>
<tr><td>CSD-607</td><td>Financial Machine Learning</td><td>4</td></tr>
<tr><td>CSD-608</td><td>Data Science for Atmospheric Science</td><td>4</td></tr>
<tr><td>CSD-609</td><td>Pragmatic AI</td><td>4</td></tr>
<tr><td>CSD-610</td><td>AI for Medical Specialization</td><td>4</td></tr>
<tr><td>CSD-611</td><td>Recommender Systems</td><td>4</td></tr>
<tr><td>CSD-612</td><td>Text Mining and Sentiment Analysis</td><td>4</td></tr>
<tr><td></td><td>Total Credits for RSEC</td><td>4</td></tr>
<tr><th colspan="2">Dissertation Type</th><th>Credits</th></tr>
<tr><td>CSD-651</td><td>Research Project in Academic or Research Institutes</td><td rowspan="2">16</td></tr>
<tr><td>CSA-652</td><td>Industry Internship / Software Project Development</td></tr>
<tr><td colspan="2">Total Credits for Dissertation</td><td>16</td></tr>
</table>

SEMESTER I
Name of the Programme: MSc. in Data Science
Course Code: CSD-500
Title of the Course: Fundamentals of Data Science (Theory)
Number of Credits: 2(2L-0T- 0P)
Effective from AY: 2023-24

Prerequisites for the course	Statistics and probability theory and python programming	
Objectives	To get started with basics of data science and learn all aspects of data science in its entirety	
Content	Introduction: Typology of problems - Data science in a big data world: Benefits and uses of data science and big data-Facets of data-The data science process-The big data ecosystem and data science-The data science process: Overview of the data science process- Defining research goals and creating a project charter- Retrieving data-Cleansing, integrating, and transforming data-Exploratory data analysis-Build the models- Presenting findings and building applications on top of them.	4 hours
	Mathematics for Data Science – A quick Review: Importance of linear algebra, statistics and optimization from a data science perspective; Structured thinking for solving data science problems. Linear Algebra: Matrices and their properties (determinants, traces, rank, nullity, etc.); Eigenvalues and eigenvectors; Matrix factorizations; Inner products; Distance measures; Projections; Notion of hyperplanes; half-planes. Probability, Statistics and Random Processes: Probability theory and axioms; Random variables; Probability distributions and density functions (univariate and multivariate); Expectations and moments; Covariance and correlation; Statistics and sampling distributions; Hypothesis testing of means, proportions, variances and correlations; Confidence (statistical) intervals; Correlation functions; White-noise process. Data clearing (EDA)	6 hours
	Introduction to Data Science Methods: Linear regression as an exemplar function approximation problem; Linear classification problems-PCA	4 hours
	Handling large data on a single computer - The problems you face when handling large data-General techniques for handling large volumes of data-General programming tips for dealing with large data sets - Case study 1: Predicting malicious URLs - First steps in big data-Distributing data storage and processing with frameworks	4 hours
	Introduction to NoSQL The rise of graph databases Introducing connected data and graph databases Introducing Neo4j: a graph database	4 hours
	Data visualization to the end user	4 hours
	Data visualization options Cross filter, the JavaScript MapReduce library Creating an interactive dashboard with dc.js Dashboard development tools	4 hours
Pedagogy	Lectures/ Tutorials/Hands-on assignments/Self-study	
References / Readings	1. Practical statistics for data science by peter bruce and andrew bruce 2. Naked statistics by charles wheelon 3. Business data science by matt taddy 4. Elements of statistical learning by Trevor Hastie, Robert and jerome 5. Python for data analysis 6. Data science and big data analytics -EMC2	

Course Outcomes	1. Understanding of fundamental concepts and techniques in data science. 2. Proficiency in data manipulation, analysis, and visualization using tools like Python or R. 3. Introduction to machine learning algorithms and evaluation methods. 4. Awareness of ethical considerations and responsible practices in data science.

(Back to top)

Name of the Programme: M.Sc. in Data Science
Course code:CSD-501
Title of course: Fundamentals of Data Science (Practical)
Number of credits: 2(0L-0T-2P)
Effective from AY: 2023-24

Prerequisites for the course	Basic programming skills, Statistics	
Objectives	• To introduce Basic process of data science, Python and Jupyter notebooks. • To understanding how to manipulate and analyse uncurated datasets • To learn basic statistical analysis and machine learning methods and effectively visualize results	
Content	Jupyter and Numpy: Jupyter notebooks are one of the most commonly used tools in data science as they allow you to combine your research notes with the code for the analysis. After getting started in Jupyter, we'll learn how to use numpy for data analysis. numpy offers many useful functions for processing data as well as data structures which are time and space efficient.	10 hours
	Pandas: Pandas, built on top of numpy, adds data frames which offer critical data analysis functionality and features.	10 hours
	Visualization: When working with large datasets, you often need to visualize your data to gain a better understanding of it. Also, when you reach conclusions about the data, you'll often wish to use visualizations to present your results.	10 hours
	Mini Project: With the tools of Jupyter notebooks, numpy, pandas, and Visualization, you're ready to do sophisticated analysis on your own. You'll pick a dataset we've worked with already and perform an analysis for this first project.	10 hours
	Machine Learning: To take your data analysis skills one step further, we'll introduce you to the basics of machine learning and how to use sci-kit learn - a powerful library for machine learning.	10 hours
	Working with Text and Databases: You'll find yourself often working with text data or data from databases. This week will give you the skills to access that data. For text data, we'll also give you a preview of how to analyse text data using ideas from the field of Natural Language Processing and how to apply those ideas using the Natural Language Processing Toolkit (NLTK) library.	5 hours
	Final Project: These weeks let you showcase all your new skills in an end-to-end data analysis project. You'll pick the dataset, do the data munging, ask the research questions, visualize the data, draw conclusions, and present your results.	5 hours
Pedagogy	Tutorials/ Lab assignments/ Project work	
References/ Readings	1. Practical statistics for data science by Peter bruce and andrew bruce 2. Naked statistics by charles wheelon 3. Business data science by matt taddy 4. Elements of statistical learning by Trevor Hastie, Robert and jerome 5. Python for data analysis 6. Data science and big data analytics -EMC2	
Course Outcomes	1. Practical data analysis skills using data science tools. 2. Hands-on experience with real-world data projects. 3. Collaboration and teamwork in interdisciplinary settings. 4. Ethical considerations and responsible practices in data science Experimentation and evaluation of data science techniques.	

(Back to top)

Name of the Programme: M.Sc. in Data Science
Course Code:CSD-502
Title of the Course: Machine Learning (Theory)
Number of Credits: 2(2L-0T-0P)
Effective from AY: 2023-24

Prerequisites for the course:	Familiarity with linear algebra, statistics & probability theory	
Objectives:	This course provides students with • In-depth introduction to three main areas of Machine Learning: supervised and unsupervised and reinforcement learning. • This course will cover some of the main models and algorithms for regression, classification, clustering and Markov decision processes. Topics will include linear and logistic regression, regularisation, SVMs and kernel methods, ANNs, clustering, and dimensionality reduction ,sequential learning Like HMM and deep learning CNN and RNN	
Content:	**1. Introduction:** well posed learning problem, designing a learning system, perspectives and issues in machine learning- types of learning - supervised, unsupervised and reinforcement learning	3 hours
	2. Concept learning: concept learning task , notation, inductive learning hypothesis, concept learning as search, version space and candidate elimination algorithm, decision tree, random forest.	3 hours
	3. Linear regression: logistic regression-Support vector machine kernel, Model selection and feature selection-Ensemble methods: Bagging, boosting, Evaluating and debugging learning algorithms.	3 hours
	4. Continuous Latent Variables: Principal Component Analysis, Maximum variance formulation, Minimum error formulation, Applications of PCA, PCA for high-dimensional data.	3 hours
	5. Neural Networks: -Feed-forward Network, Functions, perceptron, -Weight-space symmetries, Network Training, Parameter optimization, Local quadratic approximation, Use of gradient information, Gradient descent optimization, Error Backpropagation, Evaluation of error-function derivatives, Efficiency of backpropagation.	3 hours
	6. Deep learning: Deep Feedforward Networks, Gradient-Based Learning, Hidden Units, -Architecture Design, CNN and RNN (simple RNN and LSTM).	4 hours
	7. Unsupervised learning; Clustering, K-means, EM.Mixture of Gaussians.	4 hours
	8. Sequential Data: Markov Models, Hidden Markov Models, Maximum likelihood for the HMM, The forward-backward algorithm, The sum-product algorithm for the HMM, Scaling factors, -The Viterbi algorithm.	4 hours
	9. Reinforcement learning: introduction- learning task-Q learning, non-deterministic rewards and actions-temporal difference learning.	3 hours
Pedagogy:	lectures/ tutorials/assignments/self-study/lab assignment/ project work	
References/ Readings	Main Reading:- 1. James, Gareth, et al. An introduction to statistical learning. Vol. 112. New York: springer, 2013. 2. Alpaydin, Ethem. Introduction to machine learning. MIT press, 2020. 3. Hart, Peter E., David G. Stork, and Richard O. Duda. Pattern classification. Hoboken: Wiley, 2000. 4. Flach, Peter. Machine learning: the art and science of algorithms that make sense of data. Cambridge University Press, 2012.	

	5. Bishop, Christopher M. "Pattern recognition and machine learning: springer New York." (2006). 6. Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016. 7. Mitchell, Tom, and Machine Learning McGraw-Hill. "Edition." (1997). 8. machine learning and AI online google course by cassie kozyrkov
Course Outcomes	1.Develop an appreciation for what is involved in learning from data. 2.Understand a wide variety of learning algorithms. 3.Understand how to apply a variety of learning algorithms to data. 4.Understand how to perform evaluation of learning algorithms and model selection and Have a basic understanding of deep learning.

(Back to top)

Name of the Programme: M.Sc. in Data Science
Course Code:CSD-503
Title of the Course: Machine Learning(Practical)
Number of Credits: 2(0L-0T-2P)
Effective from AY: 2023-24

Prerequisites for the course:	Machine learning theory and programming in python	
Objective:	This course provides students with Aimed at imparting implementation of machine learning algorithms using python and its APIs	
Content:	**Suggested Lab assignments/work with respect to the following using python (scikit /keras libraries) /amazon sage maker/matlab toolbox - each assignment with duration of 4 hrs. and 8 hrs. for project work**	
	1. Write a program to implement version space.	5 hours
	2. Write a program to implement a decision tree for given data.	5 hours
	3. Write a program to implement linear regression for given data.	5 hours
	4. Write a program to implement logistic regression.	5 hours
	5. Write a program to implement SVM.	5 hours
	6. Write a program to implement perceptron.	5 hours
	7. Write a program to implement a multilayer perceptron.	5 hours
	8. Write a program to implement RNN.	5 hours
	9. Write a program to implement CNN.	5 hours
	10. Write a program to implement HMM.	5 hours
	Capstone Mini Project work to assess the overall learning.	10 hours
Pedagogy:	Lab Assignments / Mini Project	
References/ Readings	Main Reading:- 1. James, Gareth, et al. An introduction to statistical learning. Vol. 112. New York: springer, 2013. 2. Alpaydin, Ethem. Introduction to machine learning. MIT press, 2020. 3. Hart, Peter E., David G. Stork, and Richard O. Duda. Pattern classification. Hoboken: Wiley, 2000. 4. Flach, Peter. Machine learning: the art and science of algorithms that make sense of data. Cambridge University Press, 2012. 5. Bishop, Christopher M. "Pattern recognition and machine learning: springer New York." (2006). 6. Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016. 7. Mitchell, Tom, and Machine Learning McGraw-Hill. "Edition." (1997). 8. machine learning and AI online google course by cassie kozyrkov	
Course Outcomes	1. Practical implementation skills of machine learning algorithms. 2. Model development, evaluation, and feature engineering techniques. 3. Interpretability and explainability of machine learning models. 4. Awareness of ethical considerations in machine learning.	

(Back to top)

Name of the Programme: M.Sc. in Data Science
Course code: CSD-504
Title of course: Mathematics foundation for Data Science (Theory)
Number of credits: 2 (2L-0T-0P)
Effective from AY: 2023-24

Prerequisites for the course	Basic mathematics	
Objectives	• To build a strong foundation in maths required for learning computer science/data science subjects. • To understand fundamental concepts and tools in calculus, linear algebra etc. with emphasis on their applications to computer science in particular data science/machine learning	
Content	**Introduction** Importance of mathematics and their applications for computer science/machine learning/data science/deep learning Functions, variables, equations, graphs revision	5 hours
	Probability and Statistics: Probability Rules & Axioms, Bayes' Theorem, Random Variables, Variance and Expectation, Conditional and Joint Distributions, Standard Distributions (Bernoulli, Binomial, Multinomial, Uniform and Gaussian), Moment Generating Functions, Maximum Likelihood Estimation (MLE), Prior and Posterior, Maximum a Posteriori Estimation (MAP) and Sampling Methods-confidence intervals, Hypothesis testing, p-values, A/B testing-ANOVA, t-test, Linear regression, regularization	5 hours
	Calculus Overview of Differential and Integral Calculus, Partial Derivatives Product and chain rule-Taylor's series, infinite series summation/integration concepts-Fundamental and mean value-theorems of integral calculus, evaluation of definite and improper integrals-Beta and Gamma functions, Functions of multiple variables, limit, continuity, partial derivatives-Basics of ordinary and partial differential equations -Applications of Calculus	5 hours
	Linear Algebra: Systems of Linear Equations-Matrices-Solving Systems of Linear Equations-Vector Spaces-Linear Independence-Basis and Rank-Linear Mappings Affine Spaces	5 hours
	Analytic Geometry Norms-(Inner Products-Lengths and Distances Angles and Orthogonality-Orthonormal Basis Orthogonal Complement-Inner Product of Functions-Orthogonal Projections-Rotations) - Eigen value decomposition and SVD	5 hours
	Optimization Differentiation of Univariate Functions-Partial Differentiation and Gradients-Gradients of Vector-Valued Functions-Gradients of Matrices Useful Identities for ComputingGradients-Backpropagation and Automatic Differentiation-Higher-Order Derivatives-Linearization and Multivariate Taylor Series-Gradient Descent-Constrained Optimization - Lagrange Multipliers-Convex Optimization,	5 hours
Pedagogy	Problem solving approach and carrying out small project work using matlab tools	
References/	1. Statistics Written, Robert S. Witte and John S. Witte	

Readings	2. Barron's AP Statistics, 8th Edition, Martin Sternstein, PhD. 3. Statistics for Business and Economics by- James T. McClave, P. George Benson and Terry T Sincich 4. Naked Statistics: Stripping the Dread from the Data, Charles Wheelan 5. Introduction to Linear Algebra, Gilbert Strang 6. Linear Algebra and Its Applications, David C. Lay 7. No bullshit guide to Linear algebra, Ivon Savov 8. Functions and Graphs by I M Gelfand 9. Cartoon guide to calculus, Larry Gonick 10. Optimization Methods in Business Analytics — edX, MIT
Course Outcomes	1. Strong understanding of mathematical concepts relevant to data science, including linear algebra, calculus, probability theory, and statistics. 2. Ability to apply mathematical principles to solve data science problems, such as dimensionality reduction, optimization, and uncertainty modeling. 3. Proficiency in mathematical modeling techniques and algorithms used in data science, such as regression, clustering, and classification. 4. Development of mathematical reasoning and problem-solving skills for analyzing and interpreting data, formulating mathematical solutions, and communicating results.

(Back to top)

Name of the Programme: M.Sc. in Data Science
Course code: CSD-505
Title of course: Mathematical foundation for Data Science (Practical)
Number of credits: 2 (0L-0T-2P)
Effective from AY: 2023-24

Prerequisites for the course	Mathematical foundation theory and programming background	
Objectives	The lab assignment are aimed at demonstration of the following regarding statistics	
Content	Recap of following – A. NumPy is a third-party library for numerical computing, optimized for working with single- and multi-dimensional arrays. Its primary type is the array type called ndarray. This library contains many routines for statistical analysis. B. SciPy is a third-party library for scientific computing based on NumPy. It offers additional functionality compared to NumPy, including scipy.stats for statistical analysis. C. Pandas is a third-party library for numerical computing based on NumPy. It excels in handling labelled one-dimensional (1D) data with Series objects and two-dimensional (2D) data with Data Frame objects. D. Matplotlib is a third-party library for data visualization. It works well in combination with NumPy, SciPy, and Pandas.	3 hours
	Assignment 1 - Write program to implement the EDA concepts using python libraries -Numpy,Pandas, matplotlib, seaborn,scipy, scrapy and beautiful soup, and tensor flow ,keras and pytorch etc.	3 hours
	Assignment -2 - Sampling, Variables in Statistics, Frequency Distributions. Generate frequency distribution tables, Generate grouped frequency distribution tables and -Visualizing Frequency Distributions -Generate bar plots, pie charts, and histograms ,Employ bar plots, pie charts and histograms.	6 hours
	Assignment-3-Comparing Frequency Distributions -grouped bar plots-step-type histogram-kernel density estimate plots- strip plots and box plots	6 hours
	Assignment-4 -Multidimensional image operations, Solving differential equations and the Fourier transform using scipy	6 hours
	Assignment-5 -Optimization algorithms using scipy.	6 hours
	Assignment -6 -Linear algebra using scipy	6 hours
	Assignment- 7-Program in python to implement the concepts such as Vector space, subspace, span, column space, row space, null space, left-null space, rank, basis, orthogonal matrix, symmetric matrix.	6 hours
	Assignment -8 – Implement Eigen value decomposition in python.	6 hours
	Assignment-9 – implement SVD using python.	6 hours
	Assignment -10 – implements some of optimization algorithm using the python library.	6 hours
Pedagogy	lab assignments /Project	

References/ Readings	1. Statistics Written, Robert S. Witte and John S. Witte 2. Barron's AP Statistics, 8th Edition, Martin Sternstein, PhD. 3. Statistics for Business and Economics by- James T. McClave, P. George Benson and Terry T Sincich 4. Naked Statistics: Stripping the Dread from the Data, Charles Wheelan 5. Introduction to Linear Algebra, Gilbert Strang 6. Linear Algebra and Its Applications, David C. Lay 7. No bullshit guide to Linear algebra, Ivon Savov 8. Functions and Graphs by I M Gelfand 9. Cartoon guide to calculus, Larry Gonick 10. Optimization Methods in Business Analytics — edX, MIT
Course Outcomes	1. Practical application of mathematical concepts in data science. 2. Proficiency in using mathematical software and tools for data analysis. 3. Hands-on experience in data analysis and modeling using mathematical techniques. 4. Collaborative teamwork on data science projects involving mathematical foundations.

(Back to top)

Name of the Programme: M.Sc. in Data Science
Course code: CSD-506
Title of course: Fundamentals of AI (Theory)
Number of credits: 2(2L-0T-0P)
Effective from AY: 2023-24

Prerequisites for the course	Programming back programming and probability and statistics and linear algebra	
Objectives	To develop a basic understanding of • Problem solving • Knowledge representation • Reasoning and learning methods of AI.	
Content	Artificial Intelligence Introduction -Intelligent Agents, Problem-solving Solving Problems by Searching -Search in Complex Environments - Adversarial Search and Games- Constraint Satisfaction Problems Knowledge, reasoning, and planning Knowledge Representation-First-Order Predicate Logic - Unification Forward and Backward Chaining - Resolution - Ontological Engineering Categories and Objects - Events-Mental Events and Mental Objects - Reasoning Systems for Categories - Reasoning with Default Information Uncertain knowledge and reasoning	5 hours
	Quantifying Uncertainty - Probabilistic Reasoning - Probabilistic Reasoning over Time Probabilistic Programming -Making Simple Decisions - Making Complex Decisions -Multiagent Decision Making	4 hours
	Machine Learning, Learning from Examples - Learning Probabilistic Models - Deep Learning - Reinforcement Learning Communicating, perceiving, and acting	4 hours
	Natural Language Processing - Deep Learning for Natural Language Processing - Computer Vision - Robotics.	5 hours
	Artificial Intelligence applications Language Models - Information Retrieval - Information Extraction	4 hours
	Natural Language Processing - Machine Translation - Speech Recognition	4 hours
	Robotics-Hardware and Software for Robots - Planning and Perception Explainable AI - Definitions and concepts such as black-box models, transparency, interpretable machine learning and explanations. - Decision-making and decision support, Human-Computer Interaction (HCI) and AI. - Explainable AI. - Methods for Explainable AI. - Applications and examples. - Trust and acceptance.-Evaluation methods and metrics. - Ethical, legal and social issues of explainable AI. Contemporary issues in AI- Philosophy, Ethics, and Safety of AI - The Future of AI	4 hours
Pedagogy	Tutorials / Hands-on-assignments / Self-study	
References/ Readings	1. A Classical Approach to Artificial Intelligence, M.C. Trivedi, Khanna Book Publishing, 2019. 2. Artificial Intelligence: A modern approach by Stuart Russel, Pearson Education, 2010. 3. Artificial Intelligence by Rich and Knight, The McGraw Hill, 2017. 4. Artificial Intelligence: A new synthesis by Nils and Nilson, Elsevier, 1997. 5. Artificial Intelligence by Luger, Pearson Education, 2002. 6. Artificial Intelligence by Padhy, Oxford Press, 2005.	

	7. https://www.edx.org/course/artificial-intelligence-ai 8. https://www.udemy.com/course/artificial-intelligence-az
Course Outcomes	1. Understand the basic concepts and techniques of Artificial Intelligence. 2. Apply AI algorithms for solving practical problems. 3. Apply basics of Fuzzy logic and neural networks. 4. Explain Expert System and implementation.

(Back to top)

Name of the Programme: M.Sc. in Data Science
Course Code: CSD-507
Title of the Course: Fundamentals of Artificial Intelligence (Practical)
Number of Credits: 2 (0L-0T-2P)
Effective from AY: 2023-24

Prerequisites for the course:	Artificial Intelligence theory, probability and statistics , linear algebra and Python programming	
Objectives:	To develop a basic understanding of • Problem solving • Knowledge representation • Reasoning and learning methods of AI • Implement AI algorithms	
Content:	Assignment-1 -Real-world path planning for pedestrians. In the first part, students implement A* over a map that includes roads/paths as well as elevations. In the second part, students collect actual data through walking around the real world, and the cost model is then learned via regression techniques.	10 hours
	Assignment-2 -Solve maze via search -this assignment involves formulating maze-solving as a search problem, image processing (via OpenCV) as a step in maze-solving, as well as guided performance/quality analysis of representational parameters	10 hours
	Assignment 3-Within the context of an artificial intelligence course, students are taught to identify ethical issues within technical projects and to engage in moral problem solving with regard to such issues.	10 hours
	Assignment 4-Neural network for face recognition using tensor flow -build feedforward neural networks for face recognition using TensorFlow. Students then visualize the weights of the neural networks they train. The visualization allows students to understand feedforward one-hidden layer neural networks in terms of template matching, and allows students to explore overfitting.	10 hours
	Assignment -5 -Organic path finding -Students develop a "human-like" pathfinding technique by specializing a generic search algorithm with custom action cost and heuristic cost functions. Students apply classical search algorithms and reflect on example organic paths to achieve "human-like" pathfinding.	10 hours
	Assignment - 6 -Implement a genetic algorithm in Python to evolve strategies for Robby the Robot to collect empty soda cans that lie scattered around his rectangular grid world. And also Compare the performances of a brute-force search and a search employing the Minimum Remaining Values (MRV) heuristic in solving Sudoku puzzles.	10 hours
Pedagogy:	lectures/practical/ tutorials/assignments/self-study	

References /Readings:	1. A Classical Approach to Artificial Intelligence, M.C. Trivedi, Khanna Book Publishing, 2019. 2. Artificial Intelligence: A modern approach by Stuart Russel, Pearson Education, 2010. 3. Artificial Intelligence by Rich and Knight, The McGraw Hill, 2017. 4. Artificial Intelligence: A new synthesis by Nils and Nilson, Elsevier, 1997. 5. Artificial Intelligence by Luger, Pearson Education, 2002. 6. Artificial Intelligence by Padhy, Oxford Press, 2005. 7. https://www.edx.org/course/artificial-intelligence-ai 8. https://www.udemy.com/course/artificial-intelligence-az/
Course Outcomes:	1.The students need to understand and extend an existing implementation of the back-propagation algorithm and use it to recognize static hand gestures in images. 2.Students learn about feedforward neural networks and the backpropagation algorithm by implementing a perceptron network for AND and XOR Boolean functions 3. given an implementation of a feedforward network, learn digit recognition using the MNIST data set. 4.In this assignment students extend a Tic Tac Toe program to Ultimate Tic Tac Toe and implement a different search strategy than the example code.

(Back to top)

Semester II
Name of the Programme: M.Sc. in Data Science
Course Code: CSD-508
Title of the Course: Reinforcement Learning(Theory)
Number of Credits: 2(2L-0T-0P)
Effective from AY: 2023-24

Prerequisites for the course	Linear algebra, multivariable calculus Basic machine learning knowledge	
Objectives	To enable the student to understand • The reinforcement learning paradigm • Identify when an RL formulation is appropriate • Understand the basic solution approaches in RL • Implement and evaluate various RL algorithms.	
Content	Review of ML fundamentals – Classification, Regression. Review of probability theory and optimization concepts.	2 hours
	RL Framework; Supervised learning vs. RL; Explore-Exploit Dilemma; Examples.	2 hours
	MAB: Definition, Uses, Algorithms, Contextual Bandits, Transition to full RL, Intro to full RL problem	2 hours
	Intro to MDPs: Definitions, Returns, Value function, Q-function.	2 hours
	Bellman Equation, DP, Value Iteration, Policy Iteration, Generalized Policy Iteration.	2 hours
	Evaluation and Control: TD learning, SARSA, Q-learning, Monte Carlo, TD Lambda, Eligibility Traces.	2 hours
	Maximization-Bias & Representations: Double Q learning, Tabular learning vs. Parameterized, Q-learning with NNs	2 hours
	Function approximation: Semi-gradient methods, SGD, DQNs, Replay Buffer.	2 hours
	Policy Gradients: Introduction, Motivation, REINFORCE, PG theorem, Introduction to AC methods	3 hours
	Actor-Critic Methods, Baselines, Advantage AC, A3C Advanced Value-Based Methods: Double DQN, Prioritized Experience Replay, Dueling Architectures, Expected SARSA.	3 hours
	Advanced PG/A-C methods: Deterministic PG and DDPG, Soft Actor-Critic (SAC) HRL: Introduction to hierarchies, types of optimality, SMDPs, Options, HRL algorithms POMDPS: Intro, Definitions, Belief states, Solution Methods; History-based methods, LSTMS, Q-MDPs, Direct Solutions, PSR.	4 hours
	Model-Based RL: Introduction, Motivation, Connections to Planning, Types of MBRL, Benefits, RL with a Learnt Model, Dyna-style models, Latent variable models, Examples, Implicit MBRL. Case study on design of RL solution for real-world problems.	4 hours
Pedagogy	Hands-on assignments / tutorials / peer-teaching / flip classroom/ presentations.	
References/ Readings	1. Reinforcement learning -Introduction by Richard sutton and Andrew barto, 2nd edition, MIT press. 2. Algorithms for reinforcement learning by Csaba Szepesvari, Ronald Brachman, et al,2010.	
Course Outcomes	1. Understanding of fundamental concepts and algorithms in reinforcement learning. 2. Proficiency in implementing and evaluating reinforcement learning algorithms. 3. Application of reinforcement learning to real-world problems. 4. Critical analysis and research skills in the field of reinforcement learning.	

Name of the Programme: M.Sc. in Data Science
Course Code: CSD-509
Title of the Course: Reinforcement Learning(Practical)
Number of Credits: 2 (0L-0T-2P)
Effective from AY: 2023-24

Prerequisites for the course	Linear algebra, multivariable calculus, Basic machine learning knowledge and programming background.	
Objectives	To understand the theory by carrying out the lab assignment based on the key ideas of reinforcement learning.	
Content	1. RL task formulation (action space, state space, environment definition)	7 hours
	2. Tabular based solutions (dynamic programming, Monte Carlo, temporal-difference)	7 hours
	3. Function approximation solutions (Deep Q-networks)	7 hours
	4. Policy gradient from basic (REINFORCE) towards advanced topics (proximal policy optimization, deep deterministic policy gradient, etc.)	7 hours
	5. Model-based reinforcement learning	7 hours
	6. Imitation learning (behavioral cloning, inverse RL, generative adversarial imitation learning)	7 hours
	7. Meta-learning	8 hours
	8. Multi-agent learning, partial observable environments	10 hours
Pedagogy	Lab assignments/ mini project	
References/ Readings	1. Richard S. Sutton and Andrew G. Barto, "Reinforcement learning: An introduction", Second Edition, MIT Press, 2019. 2. Li, Yuxi. "Deep reinforcement learning." arXiv preprint arXiv:1810.06339 (2018). 3. Wiering, Marco, and Martijn Van Otterlo. "Reinforcement learning." Adaptation, learning, and optimization 12 (2012): 3. 4. Russell, Stuart J., and Peter Norvig. "Artificial intelligence: a modern approach."Pearson Education Limited, 2016. 5. Goodfellow, Ian, YoshuaBengio, and Aaron Courville. "Deep learning." MIT press, 2016. 6. David Silver's course on Reinforcement Learning (link).	
Course Outcomes	1. Practical implementation of reinforcement learning algorithms in lab exercises. 2. Experimental evaluation and analysis of reinforcement learning algorithms. 3. Application of reinforcement learning techniques to real-world problems. 4. Systematic problem-solving approach in reinforcement learning.	

(Back to top)

Name of the Programme: M.Sc. in Data Science
Course Code:CSD-510
Title of the Course: Optimization Techniques
Number of Credits: 4(2L-2T-0P)
Effective from AY: 2023-24

Prerequisites for the course	NIL	
Objectives	• To familiarize the students with some basic concepts of optimization techniques and approaches. • To formulate a real-world problem as a mathematical programming model. • To develop the model formulation and applications are used in solving decision problems. • To solve specialized linear programming problems like the transportation and assignment problems.	
Content:	**Introduction to Operations Research** Introduction-Mathematical models of Operation Research - Scope and applications of Operation Research - Phases of Operation Research study - Characteristics of Operation Research - Limitations of Operation Research.	4 hours
	Linear Programming Introduction –Properties of Linear Programming-Basic assumptions-Mathematical formulation of Linear Programming-Limitations or constraints-Methods for the solution of LP Problem-Graphical analysis of LP-Graphical LP Maximization problem-Graphical LP Minimization problem.	4 hours
	Linear Programming Models Simplex Method-Basics of Simplex Method - Formulating the Simplex Method-Simplex Method with two variables - Simplex Method with more than two variables - Big M Method.	4 hours
	Dual Linear Programming Introduction- Primal and Dual problem - Dual problem properties-Solution techniques of Dual problem - Dual Simplex method-Relations between direct and dual problem-Economic interpretation of Duality.	4 hours
	Transportation and Assignment Models Introduction: Transportation problem - Balanced - Unbalanced - Methods of basic feasible solution Optimal solution-MODI method. Assignment problem-Hungarian Method.	4 hours
	Network Analysis Basic concepts-Construction of Network-Rules and precautions-CPM and PERT Networks Obtaining of critical path. Probability and cost consideration. Advantages of Network.	5 hours
	Theory of Games Introduction-Terminology-Two Person Zero-Sum game-Solution of games with saddle points and without saddle points-2X2 games-dominance principle – mX2 and 2Xn games-Graphical method.	5 hours
Tutorial Sessions	Case Studies and Mini Projects based on concepts covered during theory lectures	2 * 15 = 30 hours
Pedagogy:	Assignment / Quiz / invited talks on current issues/ Research and Analytical problems on various applications of the industrial issues.	
References/ Readings	**Text Book(s)** • HamdyTaha, Operations Research, 10th edition, Prentice Hall India, 2019. • P. K. Gupta and D. S. Hira, Operations Research, S. Chand & co., 2007.2	

	Reference Books • S.D. Sharma (2000), Operations Research, Nath& Co., Meerut. • Maurice Solient, Arthur Yaspen, Lawrence Fridman, (2003), OR methods and Problems, New Age International Edition. • J K Sharma (2007), Operations Research Theory & Applications, 3e, Macmillan India Ltd. • P. SankaraIyer, (2008), Operations Research, Tata McGraw-Hill. • A Ravindran, Don T Philips and James J Solberg, Operations Research: Principles and practice, 2nd edition, John Wiley and sons, 2007
Course Outcomes	1.Apply operations research techniques like linear programming problem in industrial optimization problems. 2.Solve allocation problems using various OR methods. 3.Understand the characteristics of different types of decision making environment and the appropriate decision making approaches and tools to be used in each type. 4.Recognize competitive forces in the marketplace and develop appropriate reactions based on existing constraints and resources.

(Back to top)

Name of the Programme: M.Sc. in Data Science
Course Code: CSD-511
Title of the Course: MLOps(Theory)
Number of Credits: 2(2L-0T-0P)
Effective from AY: 2023-24

Prerequisites for the course	Familiarity with linear algebra, probability theory, machine learning , familiarity with python.	
Objectives	This course is aimed at anyone who wishes to • Explore deep learning from scratch. • This course offers a practical hand on exploration of deep learning, avoiding mathematical notation, preferring instead to explain quantitative concepts through programming using python API	
Content	Introduction to MLOps Rise of the Machine Learning Engineer and MLOps-What Is MLOps?-DevOps and MLOps-An MLOps Hierarchy of Needs-Implementing DevOps-Configuring-Continuous Integration with GitHub Actions-DataOps and Data Engineering-Platform Automation-MLOps	3 hours
	MLOps Foundations-Bash and the Linux Command Line-Cloud Shell Development Environments-Bash Shell and Commands-List Files Run Commands Files and Navigation-Input/output-Configuration-Writing a Script-Cloud Computing Foundations and Building Blocks-Getting Started with Cloud Computing- minimalistic python revision-Descriptive Statistics and Normal Distributions-Optimization-Machine Learning Key Concepts-Doing Data Science-Build an MLOps Pipeline from Zero	3 hours
	MLOps for Containers and Edge Devices Containers-Container Runtime-Creating a Container Running a Container-Best Practices-Serving a Trained Model Over HTTP-Edge Devices-Coral Azure Percept-TFHub-Porting Over Non-TPU Models-Containers for Managed ML Systems-Containers in Monetizing MLOps-Build Once, Run Many MLOps Workflow	3 hours
	Continuous Delivery for Machine Learning Models-Packaging for ML Models-Infrastructure as Code for Continuous Delivery of ML Models-Using Cloud Pipelines-Controlled Rollout of Models-Testing Techniques for Model Deployment	3 hours
	AutoML and Kaizen ML-Auto ML-MLOps Industrial Revolution-Kaizen Versus Kaizen ML-Feature Stores-Apple's Ecosystem-Apple's AutoML: Create ML-Apple's Core ML Tools or Google's AutoML and Edge Computer Vision or Azure's AutoML or AWS AutoML-Open Source AutoML Solutions-Ludwig-FLAML-Model Explainability	3 hours
	Monitoring and Logging-Observability for Cloud MLOps-Introduction to Logging-Logging in Python-Modifying Log Levels-Logging Different Applications-Monitoring and Observability-Basics of Model Monitoring-Monitoring Drift with AWS SageMaker-Monitoring Drift with Azure ML	3 hours
	MLOps for AWS-Introduction to AWS-Getting Started with AWS Services-MLOps on AWS-MLOps Cookbook on AWS-CLI Tools-Flask Microservice-AWS Lambda Recipes-AWS Lambda-SAM Local-AWS Lambda-SAM Containerized Deploy-Applying AWS Machine Learning to the Real World	3 hours
	Machine Learning Interoperability-Why Interoperability Is Critical-ONNX: Open Neural Network Exchange-ONNX Model Zoo-Convert PyTorch into ONNX -Convert TensorFlow into ONNX-Deploy ONNX to Azure-Apple Core ML-Edge Integration.	3 hours

	Building MLOps Command Line Tools and Microservices-Python Packaging-The Requirements File-Command Line Tools-Creating a Dataset Linter Modularizing a Command Line Tool-Microservices-Creating a Serverless Function-Authenticating to Cloud Functions-Building a Cloud-Based CLI-Machine Learning CLI Workflows	3 hours
	Machine Learning Engineering and MLOps Case StudiesUnlikely Benefits of Ignorance in Building Machine Learning Models-MLOps Projects at Sqor Sports Social Network-Mechanical Turk Data Labeling-Influencer Rank-Athlete intelligence (AI product)-The perfect techniques versus the real world-critical challenges in MLops- Ethical and unintended consequences-lack of operational excellences- focus on prediction accuracy vs the big picture	3 hours
Pedagogy	Lectures/ tutorials/lab assignments/self-study	
References/ Readings	Main Reading :- • Practical MLops – Noah Gift and AlfredoDeza • Introduction to MLOps – Noah Gift and AlfredoDeza	
Course Outcomes	1. Integration of machine learning and software engineering for production systems. 2. Automation of model development, training, and deployment processes. 3. Scalable and reliable infrastructure design for machine learning applications. 4. Monitoring and maintenance of deployed machine learning systems.	

(Back to top)

Name of the Programme: M.Sc. in Data Science
Course Code: CSD-512
Title of the Course: MLOps(Practical)
Number of Credits: 2(0L-0T-2P)
Effective from AY: 2023-24

Prerequisites for the course	Machine Learning and programming	
Objectives	Aimed at imparting the knowledge required to deploy ML models	
Content	• Perfect Project Structure – Cookiecutter& readme.so	6 hours
	• Speed Exploratory Data Analysis to Minutes – Pandas Profiling, SweetViz	6 hours
	• Track Data Science Projects with CI, CD, CT, CM –Data Version Control (DVC)	6 hours
	• Explainable AI / XAI – SHAP, LIME, SHAPASH	6 hours
	• Deploy ML Projects in minutes – Docker, FastAPI	6 hours
	• End to End Machine Learning – MLflow	6 hours
	• Building Production Ready ML Pipelines - Model Registry, Feature Store (Feast, ButterFlow)	6 hours
	• Big Data using Python, instead of PySpark – DASK	6 hours
	• Build a Chat bot and Deploy it (open-source)	6 hours
	• FaaS Framework implementation – Apache OpenWhisk, OpenFaas	6 hours
Pedagogy	Lab Assignments / mini project	
References/ Readings	1. Machine Learning Engineering By AndriyBurkov Publisher : True Positive Inc. (8 September 2020) 2. ML Ops: Operationalizing Data Science By David Sweenor, DevKannabiran, Thomas Hill, Steven Hillion, Dan Rope and Michael O'Connell-O'Reilly 3. Building Machine Learning Pipelines By HannesHapke, Catherine Nelson 4. Practical MLOps by Noah Gift, Alfredo Deza. O'Reilly 5. Introducing MLOps By Mark Treveil&Dataiku Team 6. Beginning MLOps with MLFlow: Deploy Models in AWS SageMaker, Google Cloud, and Microsoft AzureBy Sridhar Alla, SumanKalyanAdari, O'Reilly	
Course Outcomes	1. Hands-on experience with MLOps tools and technologies. 2. Building end-to-end machine learning pipelines. 3. Deployment and management of infrastructure for machine learning models. 4. Collaboration and adoption of DevOps practices in MLOps.	

(Back to top)

Name of the Programme: M.Sc. in Data Science
Course code: CSD-513
Title of course: Software Engineering for AI Enabled Systems (Theory)
Number of credits: 2 (2L-0T-0P)
Effective from AY: 2023-24

Prerequisites for the course	Programming & Data Structures, Python	
Objectives	• Gain an in-depth understanding of Software Engineering including its importance. • Learn Scrum, Kanban, Agile, Waterfall, Prototyping, Incremental, RAD and Spiral Software Process Models. • Learn to perform systematic Software Requirement Engineering. • Applying SE approach to developing AI solutions	
Content	Software Engineering: Software Processes, SDLC, agile approaches to SE	5 hours
	Requirements Engineering: elicitation techniques, specification. SCRUM and user stories.	5 hours
	Test Driven Development: Refactoring and Unit testing	5 hours
	Use of frameworks and APIS and handling of big data	5 hours
	Configuration management, continuous integration, and automated software engineering	5 hours
	Cloud based software development, DevOps	5 hours
Pedagogy	Classroom/hands-on instructions, assignments, mini projects	
References/ Readings	1. Hands-On Software Engineering with Python: Move beyond basic programming and construct reliable and efficient software with complex code, Brian Allbee, Packt Publishing. 2. A concise Introduction to Software Engineering, Pankaj Jalote-2008n- Springer. 3. Agile Estimating and Scrum, Mike Cohn, Prentice Hall.	
Course Outcomes	1. Application of SE principles for AI and Data Science projects 2. How to work in self organizing teams 3. Use of tools and techniques for automating 4. managing software development	

(Back to top)

Name of the Programme: M.Sc. in Data Science
Course code: CSD-514
Title of course: Software Engineering for AI Enabled Systems(Practical)
Number of credits: 2 (0L-0T-2P)
Effective from AY: 2023-24

Prerequisites for the course	Programming & Data Structures, Python	
Objectives	Applying SE approach to developing AI solutions Use of modern software engineering tools and frameworks	
Content	1 Version Control Tools- Git and Github	12 hours
	2 TDD –Unit testing and refactoring with Python	12 hours
	3 Working with Python libraries and frameworks	12 hours
	4 Use of testing tools- selenium, Jmeter	12 hours
	5 Cloud based software development &DevOps	12 hours
Pedagogy	Lab sessions and projects	
References/ Readings	1. Hands-On Software Engineering with Python: Move beyond basic programming and construct reliable and efficient software with complex code, Brian Allbee, Packt Publishing. 2. A concise Introduction to Software Engineering, Pankaj Jalote-2008n- Springer. 3. Agile Estimating and Scrum, Mike Cohn, Prentice Hall.	
Course Outcomes	1. Application of SE principles for AI and Data Science projects 2. How to work in self-organizing teams 3. Use of tools and techniques for automating 4. managing software development	

(Back to top)

Semester I
Name of the Programme: M.Sc. in Data Science
Course Code: CSD-521
Title of the Course: Domain Specific Predictive Analytics
Number of Credits: 4(2L-2T-0P)
Effective from AY: 2023-24

Prerequisites for the course	Data science fundamentals and programming background	
Objectives	• It introduces theoretical foundations • Algorithms, Methodologies for analysing data in various domains such Retail, Finance, Risk and Healthcare.	
Content for Theory	**Retail Analytics** Understanding Customer: Profiling and Segmentation, Modelling Churn. Modelling Lifetime Value, Modelling Risk, Market Basket Analysis.	4 hours
	Risk Analytics Risk Management and Operational Hedging: An Overview, Supply Chain Risk Management, A Bayesian Framework for Supply Chain Risk Management, Credit Scoring and Bankruptcy Prediction	4 hours
	Financial Data Analytics Financial News analytics: Framework, techniques, and metrics, News events impact market sentiment, Relating news analytics to stock returns	4 hours
	Financial Time Series Analytics Financial Time Series and Their Characteristics, Common Financial Time Series models, Autoregressive models, Markov chain models, Time series models with leading indicators, Long term forecasting	4 hours
	Introduction Healthcare Analytics An Introduction to Healthcare Data Analytics, Electronic Health Records, Privacy-Preserving Data Publishing Methods in Healthcare, Clinical Decision Support Systems	4 hours
	Healthcare Data Analytics Natural Language Processing and Data Mining for Clinical Text: Core NLP Components, Information Extraction and Named Entity Recognition, Social Media Analytics for Healthcare: Tracking of Infectious Disease Outbreaks, Readmission risk Prediction	4 hours
	Genomic Data Analytics Microarray Data, Microarray Data Analysis , Genomic Data Analysis for Personalized Medicine , Patient Survival Prediction from Gene Expression Data , Genome Sequence Analysis	6 hours
Content for Practical to be carried out during the Tutorial Slots	• Finance: a) Stock Market Prediction: Develop a predictive model to forecast stock prices based on historical data, using techniques such as time series analysis and machine learning algorithms.	3 hours
	b) Credit Risk Assessment: Build a model to predict the creditworthiness of individuals or businesses, incorporating relevant financial and non-financial factors to assess default probabilities.	3 hours
	c) Fraud Detection: Create an algorithm to identify fraudulent transactions or activities in financial systems by analysing patterns, anomalies, and historical data.	3 hours

	• Medical Science: a) Disease Diagnosis: Develop a predictive model to diagnose diseases based on patient symptoms, medical history, and test results, using techniques like classification algorithms and medical data analysis.	3 hours
	b) Patient Readmission Prediction: Build a model to predict the likelihood of a patient being readmitted to the hospital within a certain time frame, considering factors such as demographics, medical conditions, and treatment history.	3 hours
	c) Drug Effectiveness Prediction: Create a model to predict the effectiveness of a particular drug for a specific patient or group of patients, utilizing genetic information, clinical data, and treatment outcomes.	3 hours
	• Genomic Science: Predictive analytics in the domain of genomics can be highly beneficial for various applications, such as disease prediction, drug discovery, personalized medicine, and genetic engineering. Here are a few examples of predictive analytics techniques that can be applied in genomics a) Disease Risk Prediction: By analyzing an individual's genomic data, predictive analytics can be used to assess the risk of developing specific diseases. Machine learning algorithms can identify patterns and genetic markers associated with various diseases, allowing for early detection and preventive measures. For example, predictive models can be built to predict the risk of developing conditions like cancer, cardiovascular diseases, or genetic disorders.	3 hours
	b) Pharmacogenomics: Predictive analytics can aid in predicting an individual's response to specific drugs based on their genetic makeup. By analyzing genomic data along with clinical information, machine learning models can predict drug efficacy, potential side effects, and optimal dosage. This information can be used to develop personalized treatment plans and improve patient outcomes.	3 hours
	c) Genomic Variant Interpretation: Genomic variants play a crucial role in determining an individual's susceptibility to diseases. Predictive analytics can be used to interpret the functional consequences of these variants. Machine learning algorithms can predict the impact of genetic mutations on protein structure and function, helping researchers and clinicians understand the underlying mechanisms of diseases and develop targeted therapies.	3 hours
	d) Gene Expression Analysis: Predictive analytics can analyze gene expression data to identify patterns and correlations between genes and specific traits or diseases. By using machine learning algorithms, it is possible to predict gene expression levels based on genomic features and environmental factors. This can provide valuable insights into gene regulatory networks and help in understanding disease mechanisms and identifying potential therapeutic	3 hours
Pedagogy	Lectures/ tutorials/assignments/self-study	
References/ Readings	1. Chris Chapman, Elea McDonnell Feit "R for Marketing Research and Analytics", Springer, 2015.	

	2. Olivia Parr Rud "Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management", Wiley, 2001. 3. Chandan K. Reddy, Charu C. Aggarwal "Healthcare Data Analytics", CRC Press, 2015. 4. Rene Carmona "Statistical Analysis of Financial Data in R", Springer, 2014. 5. James B. Ayers "Handbook Of Supply Chain Management" Auerbach Publications, 2006. 6. PanosKouvelis, Lingxiu Dong, OnurBoyabatli, Rong Li "The Handbook of Integrated Risk Management in Global Supply Chains", Wiley, 2012.
Course Outcomes	1.Retail Analytics and Risk Analytics 2. Financial Data Analytics, Financial Time Series Analytics, 3.Healthcare Analytics, Healthcare Data Analytics and 4. Genomic Data Analytics.

(Back to top)

Name of the Programme: M.Sc. in Data Science
Course Code: CSD-522
Title of the Course: Design Thinking for Data-Driven App Development
Number of Credits: 4(4L-0T-0P)
Effective from AY: 2023-24

Prerequisites for the course	None	
Objectives	This course helps you learn • The basics of Design Thinking in an experiential way. • This course aims at an empathy-led data-driven app development approach for data scientists. • The learners will launch a fully functioning app in a real app store at the end of the course.	
Content	Introduction to Design Thinking – Course outline and projects, Intro to the Design of Everyday Things, Intro to Design Thinking in software apps, Project management. Empathize phase (Iteration #1)-- Emotional and intellectual map of the user stories from interviews, User story creation and Customer Journey Mapping	15 hours
	Analyse phase (Iteration #1) - Stated needs and unsaid/latent needs, Root cause analysis, Multiple perspectives of customers and manufacturers, Frame conflicts from popular movies. Solve phase (Iteration #1)Structured and unstructured creativity, Dynamics of group thinking, Optimal conditions of creativity, Natural creativity, Concept creation via group activities, Silent brainstorming, inventive principles and concept consolidation	15 hours
	Test phase (Iteration #1)/ Empathize phase (Iteration #2) - Basics of prototyping, Assumptions in creation of new concepts, Features rather than ideas. Basics of Digital Marketing, User Experience Design, Website Development	15 hours
	Analyse phase (Iteration #2) Solve phase (Iteration #2) - Introduced problems via the solution from iteration #1, the subsequent ideation process in iteration #2, apply solutioning and analysis tools in iteration #2, subsequent testing and field trial skills required for iteration #3, analytical tools and data oriented tools on iteration #3. Test (Iteration #2) / Empathize (Iteration #3) - Basics of obtaining insights from feedback from a live audience. Analyse (Iteration #3). Test phase (Iteration #3) - Launch of the App.	15 hours
Pedagogy	Hands-on assignments / Tutorials / Peer-teaching /Presentations	
References/ Readings	1. Design of everyday things by Don A. Norman, 2013. 2. This is Service Design thinking- basics, tools and cases by Marc Stickdorn, 1st edition, John Wiley & Sons Inc., 2012.	
Course Outcomes	1.Recall the basics of Design Thinking 2.Apply Agile method to developing software 3.Design an App using the principles of Design Thinking 4.Develop an App for Android and Collaborate with other developers using git version control method	

(Back to top)

Semester II
Name of the Programme: M.Sc. in Data Science
Course Code: CSD-523
Title of the Course: Signal Processing
Number of Credits: 4(2L-2T-0P)
Effective from AY: 2023-24

Prerequisites for the course	1. Linear algebra, 2. Calculus and multivariable calculus, 3. At least high school math on trigonometry, 4. Complex number 5. A little bit familiarity with programming, especially for numerical computation, such as GNU Octave.	
Objectives	1. To study various types of signals and its characteristics. 2. To study various operations on the signals. 3. To analyse the signals using Fourier transform and Laplace Transform. 4. To learn the fundamentals of robotics and sensor technology. 5. To understand the controlling applications of robotics using sensor responses.	
Content for Theory	**Module:1** Introduction to Signals Continuous-time and Discrete-time Signals: Representation of signals, Signal classification, Types of Signals, Operations on signals - Scaling, Shifting	4 hours
	Module: 2 Fourier Analysis of Continuous-time Signals Introduction to Fourier series, Gibbs Phenomenon, and Continuous-time Fourier transform (CTFT), Existence, Magnitude and phase response, Parseval's theorem, Inverse Fourier transform. Relation between Laplace and Fourier transforms, Laplace Transform, Magnitude and phase response	4 hours
	Module: 3 signal conditioning Sensing - Pre-processing – Noise reduction, enhancement of details. Signal Conversion –Sampling, Quantization, Encoding	4 hours
	Module:4 Data Acquisition and sensing in Robotics Data Acquisition: Analogy and digital data acquisition, single channel and multi-channel data acquisition Image processing in Robotics: Vision sensor, Introduction to computer vision, Point operators, Linear Filters, More neighbourhood operators, Fourier transforms, Pyramids and wavelets, Geometric transformations.	4 hours
	Module: 5 Fundamentals of Robotics Basic components of robotic system. Basic terminology- Accuracy, Repeatability, Resolution, Degree of freedom. Mechanisms and transmission, End effectors, Grippers-different methods of gripping, Mechanical grippers-Slider crank mechanism, Screw type, Rotary actuators, Cam type gripper, Magnetic grippers, Vacuum grippers, Air operated grippers; Specifications of robot.	4 hours
	Module: 6 Drive Systems and Sensors in Robotics Drive system-hydraulic, pneumatic and electric systems. Sensors in robot – Touch sensors, Tactile sensor, Proximity and range sensors, Robotic vision sensor, Force sensor, Light sensors, and Pressure sensors.	4 hours
	Module: 7 Signal processing application in Robotics Robot applications: Application of robots in surgery, Manufacturing industries, space and underwater. Humanoid robots, Micro robots, Social issues and Future of robotics.	6 hours

Content for Practicals during Tutorial:	• To find Discrete Fourier Transform and Inverse Discrete Fourier Transform of given digital signal using MATLAB software.	3 hours
	• To obtain Linear Convolution of two finite length sequences using MATLAB software.	3 hours
	• To compute auto correlation between two sequences using MATLAB software.	3 hours
	• AIM: To find frequency response of a given system in differential equation form using MATLAB software.	3 hours
	• AIM: To find the FFT of a given sequence using MATLAB software.	3 hours
	• Determination of Power Spectrum of a given signal using MATLAB software.	3 hours
	• To implement LP FIR filter for a given sequence using MATLAB software.	6 hours
	• To implement HP FIR filter for a given sequence using MATLAB software.	6 hours
Pedagogy	ISA/Assignments/seminar	
References/ Readings	**Text Book(s)** 1. Signals and Systems, second edition-P. Rama Krishna Rao and Shankar Prakriya-Mc-Graw Hill, 2013. 2. Groover. M.P. Industrial Robotics, technology, programming and application Mc-Graw Hill 2012. 3. S. R.Deb, "Robotics technology and flexible automation", Tata McGraw Hill publishing company limited, 1994. **Reference Books** 1. Signals and systems, second edition-Alan. V. Oppenheim, Alan. S. Willsk,S. Hamid Nawab, PHI learning Pvt ltd, 1997 2. Signals and systems, second edition - Simon Haykin, Barry VanVeen, Wiley, Wiley India, 2007. 3. S. K. Saha, "Introduction to Robotics", Tata McGraw-Hill Publishing Company Ltd. (2008). 4. Ramon Pallas-Areny, John G. Webster, "Sensors and Signal Conditioning", 2nd Edition, Wiley India Pvt.Ltd., India, 2012. Mode of Evaluation: Assignments / Assignments / Quiz	
Course Outcomes	1.To differentiate continuous and discrete time signals and To analyse the sensor response using Fourier transform 2.To analyse the trajectory of sensor signal using Laplace transform and To understand the signal conditioning and acquisition mechanism 3. To learn the fundamentals and peripherals of robots and To explore sensor responses in controlling robots 4.. To explore various real-time application of sensor signal in robotics	

(Back to top)

Name of the Programme: M.Sc. in Data Science
Course Code: CSD-524
Title of the Course: Regression Analytics and Predictive Models
Number of Credits: 2 (2L-2T-0P)
Effective from AY: 2023-24

Prerequisites for the course	Probability Theory and Distributions	
Objectives	• Develop an understanding of regression analysis and model building. • Provide the ability to develop relationship between variables • Investigate possible diagnostics in regression techniques • Formulate feasible solutions using a regression model for real-life problems.	
Content (Theory)	**Simple Regression Analysis** Introduction to a linear and nonlinear model. Ordinary Least Square methods. Simple linear regression model, using simple regression to describe a linear relationship. Fitting a linear trend to time series data, Validating simple regression model using t, F and p test. Developing confidence interval. Precautions in interpreting regression results.	4 hours
	Multiple Regression Analysis Concept of Multiple regression model to describe a linear relationship, Assessing the fit of the regression line, inferences from multiple regression analysis, problem of over fitting of a model, comparing two regression model, prediction with multiple regression equation.	4 hours
	Fitting Curves and Model Adequacy Checking Introduction, fitting curvilinear relationship, residual analysis, PRESS statistics, detection and treatment of outliers, lack of fit of the regression model, test of lack of fit, Problem of autocorrelation and heteroscedasticity. Estimation of pure errors from near neighbors.	4 hours
	Transformation techniques Introduction, variance stabilizing transformations, transformations to linearize the model, Box Cox methods, transformations on the repressors variables, Generalized and weighted least squares, Some practical applications.	4 hours
	Multicollinearity Introduction, sources of multicollinearity, effects of multicollinearity. Multicollinearity diagnostics: examination of correlation matrix, variance Inflation factors (VIF), Eigen system analysis of X1X. Methods of dealing with Multicollinearity: collecting additional data, model re-specification, and ridge regression.	4 hours
	Generalized Linear Models Generalized linear model: link functions and linear predictors, parameter estimation and inference in the GLM, prediction and estimation with the GLM, Residual Analysis, and concept of over dispersion.	4 hours
	Model building and Nonlinear Regression Variable selection, model building, model misspecification. Model validation techniques: Analysis of model coefficients, and predicted values, data splitting method. Nonlinear regression model, nonlinear least squares, transformation to linear model, parameter estimation in nonlinear system, statistical inference in nonlinear regression.	6 hours
Content for	1. Linear Regression	2 hours

Practicals during the Tutorial Slots:	2. Minimum Least Square Method	2 hours
	3. Calculating coefficients values	2 hours
	4. Ascombe's Quartet	2 hours
	5. Regression Equations- x on y & y on x	2 hours
	6. Predicting mom's height based on daughter's height	2 hours
	7. Regression-Solved problem-2	2 hours
	8. Probable Error- Calculating correlation coefficient of POPULATION	2 hours
	9. Predictive modelling project for credit card fraud detection	4 hours
	Any two Projects from below -	
	10. Predictive modelling project for customer value prediction	5 hours
	11. Predictive modelling project for stock market forecasting	5 hours
	12. Predictive modelling project for corporate bankruptcy prediction	5 hours
Pedagogy	Lectures/ tutorials/assignments/self-study	
References/ Readings	1. Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, Introduction to Linear Regression Analysis, Third Ed., Wiley India Pvt. Ltd., 2016. Norman R. 2. Draper, Harry Smith; Applied Regression Analysis, WILEY India Pvt. Ltd. New Delhi; Third Edition, 2015. 3. Johnson, R A., Wichern, D. W., Applied Multivariate Statistical Analysis, Sixth Ed., PHI learning Pvt., Ltd., 2013. 4. Iain Pardoe, Applied Regression Modeling, John Wiley and Sons, Inc, 2012.	
Course Outcomes	1.Develop in-depth understanding of the linear and nonlinear regression model. 2.Demonstrate the knowledge of regression modelling and model selection techniques. 3.Examine the relationships between dependent and independent variables. 4.Estimate the parameters and fit a model.	

(Back to top)