Curriculum for BS Programme in Data Science & Engineering				
Sem	**Course No.**	**Course Name**	**Credits**	**Total**
I	CHM 101	General Chemistry	3	21
	MTH 101	Calculus of One Variable	3	
	PHY 101	Mechanics	3	
	EES 101	Introduction to Earth Sciences	3	
	CHE 103	Engineering Design And Drawing	3	
	HSS 101	English for Communication	2	
	PHY 103	General Physics Laboratory-I	1	
	BIO 101* Or ECO 101*	BIOLOGY I: Biomolecules or Principles of Economics-I	3	
II	CHM 112	Basic Organic Chemistry-I	3	19
	MTH 102	Linear Algebra	3	
	PHY 102	Modern Physics	3	
	EES 102	Introduction to Environmental Sciences	3	
	ECS 102	Introduction to Programming	3	
	CHM 114	Chemistry Laboratory-I	1	
	BIO 102* Or ECO 102*	Biology II: Fundamental of Cell Biology or Principles of Economics-II	3	
Students must either do both BIO101 and BIO102 in their 1st year, or both ECO101 and ECO102				
III	ECS 201	Discrete Mathematics -I	3	19
	ECO 201	Econometrics-I	4	
	MTH 201	Multivariable Calculus	3	
	*** ***	3 Open Electives	3*3=9	
IV	ECS 202	Data Structure and Algorithms	3	18
	ECS 204	Signals And Systems	3	
	MTH 202	Probability and Statistics	3	
	*** ***	3 Open Electives	3*3=9	

V	DSE 310	Fundamentals of Database systems	4	20
	DSE 312	Computer Vision	4	
	DSE 315	Data Science in Practise	4	
	DSE 317	Machine Learning	4	
	*** ***	Open elective	4	
VI	DSE 304	Algorithms	4	20
	DSE 311	Applied Optimization	4	
	DSE 313	Artificial Intelligence	4	
	DSE 316	Deep Learning	4	
	*** ***	Open elective	4	
VII	DSE ***	DSE courses at 400 level (3 courses)	3 *4=12	24
	*** ***	Open electives (3 courses)	3*4=12	
VIII	DSE 400	Project work	18	18
Curriculum for additional one year to obtain BS-MS				
IX	#ECO 500	Law Related to Intellectual Property and Patents	1	19
	DSE ***	DSE courses (2 courses)	2*4 = 8	
	DSE 501	MS Thesis	10	
X	DSE 502	MS Thesis	20	20

Minor is DSE: Students must take 18 credits of DSE courses.
Currently, there are no mandatory DSE courses for **minors**.

#ECO 500 can be taken in any semester throughout the program.

List of non-exhaustive electives

Course no	Title
ECO 307 / DSE 305	Game Theory
MTH 308/DSE 306	Combinatorics and Graph Theory
HSS 322 / DSE 308	Computational Linguistics
BIO 402/DSE 402	Bioinformatics
BIO 407/DSE 401	Biostatistics

EES 407 / DSE 405	Data Analysis and Statistics for Geosciences
DSE 404	Network Science: Theory and Applications
DSE 406/606	Spatial Data Science and applications
DSE 407/607	Natural Language Processing
DSE 410/ 610	Transfer Learning in Computer vision
DSE 412/612	Machine Learning for BioPharma
DSE 314	Reinforcement Learning
DSE 303/ ECO 305	Econometrics II
DSE 409/609	Digital Image Processing and Applications in Bioimage Analysis
DSE 418/618	Advanced Natural Language Processing
DSE 420/620	Biomedical Text Mining
	Biometrics: An Introduction to Research
DSE 333/533	Internet of Things

Course No. DSE310

Course Name Fundamentals of Database Systems

Content

Introduction: Database applications, purpose, accessing and modifying databases, need for transactions, architecture - users and administrators, data mining, information retrieval.

Relational Databases: relational model, database schema, keys, relational query languages, algebra, tuple, and domain calculus example queries, (optional: equivalence of relational calculus and relational algebra).

SQL: Data definition, basic SQL query structure, set operations, nested subqueries, aggregation, null values, database modification, join expressions, views.

Storage and File Structure: Overview of secondary storage, RAID, and flash storage. Storing tables: row-wise, column database, database buffer. Indexing: concepts, clustered and non-clustered indices, B+-tree indices, multiple key access, hashed files, linear hash files, bitmap indices, Index definition in SQL, ++R-trees.

Query Processing: Overview, measures of query cost, selection, sorting, join processing algorithms-nested loops, merge-sort, hash join, aggregation.

Query Optimization: purpose, transformation of relational expressions, estimating cost and statistics of expression, choosing evaluation plans, linear and bushy plans, dynamic programming algorithms. Transactions: Concept and purpose, ACID properties and their necessity, transactions in SQL. Problems with full isolation and levels of isolation.

Concurrency Control: lock-based protocols, 2-phase locking, deadlock handling, multiple granularity, timestamp-based protocols, index locking.

Recovery: Failures and their classification, recovery and atomicity, recovery algorithms, Undo Redo with write-ahead logging, no Undo no Redo and other combinations, buffer management. Hadoop and Spark.

Course No. DSE312

Course Name Computer Vision

Content

Introduction to Computer Vision, Camera geometry and camera calibration, Review of Digital Image Processing, Edge Detection and Hough Transforms, Image Segmentation, Feature Point Detection - Harris, SIFT, HOG, LBP, STIP, Feature Detection, and Description - Bag Of Words, VLAD, Object Recognition - SVMs, Detection - Viola-Jones Object detector, Convolutional Neural Networks and Applications, Optical Flow, KLT based object tracking, Linear Algebra review, Projective Geometry - Basics and 2D transformations (Euclidean, Similarity, Affine, and Projective), Epipolar Geometry - Fundamental and Essential Matrix, Least Squares and Robust Estimation (RANSAC), Stereo reconstruction, SfM and Bundle Adjustment, Homography and panorama creation, Recent Progress in Computer Vision.

Course No. DSE315/DSE615

Course Name Data Science in Practice

Content

Basic concepts in Data Science:

Overview of bar chart, pi chart, histogram, exploratory data analysis Multivariate statistics: Variance, covariance, the correlation for ML models Concept of statistical significance of scores of different models Concept of outlier detection, missing value handling techniques Programming Components: The concept of array, string, list and dictionary Plotting and Python graphics Function, class, templates, and library File handling, handling data frames Sorting and searching Exception handling Overview of Numpy, Scipy, Matplotlib, Scikit-learn, Tensorflow, Keras, Pytorch How to analyze performance of different predictive models? How to generate confusion matrix, AUC curve, etc.

How to generate pipelines using different predictive models? Projects: Problem statements from basic science, engineering, and social science.

Course No. DSE317/DSE617

Course Name Machine Learning

Content

What is machine learning? The idea of supervised and unsupervised learning, regression vs classification, training and test set, classification vs clustering, and significance of feature engineering.

Overview of decision tree learning. Different splitting and pruning techniques. Different decision tree algorithms.

Concept of minimum distance classifier and k-Nearest Neighbor decision rule. Cross-validation

and parameter tuning. Different performance evaluation functions.

Bayes decision rule: Bayes theorem, Bayes classifier, and error rate of Bayes classifier. Naive Bayes classifier: Gaussian model, Multinomial model, Bernoulli model.

Concept of linear regression, Logistic regression, maximum margin classifier, discriminative and generative models.

Support vector machine (SVM) classification method, kernels.

Ensemble of Classifiers: bagging and boosting techniques, AdaBoost, Random Forest classifier.

Overview of artificial neural network (ANN), perceptron learning algorithm, multilayer perceptron learning algorithm, and backpropagation technique.

Data clustering techniques, cluster validity index: f-measure, normalized mutual information.

Partitional clustering methods e.g., k-means, Hierarchical clustering techniques, density-based clustering techniques e.g., DBSCAN.

Feature engineering: an overview of feature selection, supervised and unsupervised feature selection techniques.

Overview of principal component analysis (PCA) for feature extraction.

Course No. DSE 314

Course Name Reinforcement Learning

Content

The Reinforcement Learning Problem; Multi-arm Bandits; Finite Markov Decision Processes; Dynamic Programming; Monte Carlo Methods; Temporal-Difference Learning; On-policy Approximation of Action Values; Policy Approximation; Deep reinforcement learning.

Course No. DSE313

Course Name Artificial Intelligence

Content

Introduction; philosophy of AI, agents; informed search, uninformed search; local search heuristic search; evolutionary algorithms; adversarial search; min-max algorithm. Constraint satisfaction, Logic, and satisfiability; decision theory; Bayesian networks representation, inference, and learning. Learning agents, Cost optimization, Learning template, Discussions on a few algorithms (Regression, Neural Networks).

Course No. DSE316

Course Name Deep Learning

Content

Neural networks: Backpropagation, DL Optimizers (SGD, MBGD, AdaGrad, Adam) and Regularization, Initialization Methods

DL Models: Autoencoder, Convolutional Neural Networks, Recurrent Neural Networks, LSTM, Network

Architecture Search (NAS)

Deep Generative Models: Deep Belief Networks, Variational Autoencoders, Generative Adversarial, Diffusion models

Networks, Deep Convolutional GAN

Representation learning: Unsupervised Pretraining, Transfer learning, and Domain adaptation, Few Shot, Zero-Shot Algorithm model, Distributed representation, Discovering underlying causes

Laboratory

Autoencoder, CNN, LSTM, DBM, GANs (variants), Transfer Learning, NLM, Graph NN, Adversarial losses

Course No. DSE311

Course Name Applied Optimization

Content

Introduction to optimization problems: Constraints, Critical Points, Conditions for Local Minima, Contour Plots

Derivatives and gradients: Derivatives in Multiple Dimensions, Numerical Differentiation, Automatic Differentiation

Search methods: Unimodality, Finding an Initial Bracket, Fibonacci Search, Golden Section Search, Quadratic Fit Search, Shubert-Piyavskii Method, Bisection Method

Local descent methods: Descent Direction Iteration, Line Search, Approximate Line Search, Trust Region Methods, Termination Conditions

First-order methods: Gradient Descent, Conjugate Gradient, Momentum, Nesterov Momentum, Adagrad, RMSProp, Adadelta, Adam, Hypergradient Descent

Second-order methods: Newton's Method, Secant Method, Quasi-Newton Methods Linear

constrained optimization: Problem Formulation, Simplex Algorithm, Dual Certificates"

Course No. DSE407/DSE607

Course Name Natural Language Processing

Content

Introduction

Basic Text Processing

N-gram Language Models

Naive Bayes and Sentiment Classification

Logistic Regression

Vector Semantics and Embeddings

Neural Networks and Neural Language Models

RNN, LSTM, Transformers

HMM and CRF Parts of Speech and Named Entities

Machine Translation and Encoder-Decoder Model

Modern Language Models

Course No. DSE410/DSE610

Course Name Transfer Learning in Computer Vision

Content

Basic of Transfer Learning: Limitations of deep learning models, the effect of train/test data distributions, the effect of availability of training data, Knowledge distillation Out of distribution detection.

Domain Adaptation: Discrepancy measure of data distributions, Adversarial learning, Openset and close-set domain adaptation, Domain Generalizations, Source data free adaptation, Active learning, entropy-based sample selection,

Incremental Learning: Attention method, Self-attention, and primer of transformers.

Multimodal Learning: Modality fusion, cross model, and joint model.

Recent Advance Models: Basics of attention, Transformers, Vision Transformers.

Course No. DSE412/DSE612

Course Name Machine Learning for BioPharma

Content

The course will have four components.

1. Predictive modeling, which involves bioactivity, toxicity, and pharmacokinetic property prediction
2. Generative modeling, which involves generating new drugs using GenAI 3. Language models for unstructured biopharma data
4. Leveraging Knowledge graphs for structured biopharma data

Course No. DSE503/DSE603

Course Name Applied Accelerated Artificial Intelligence

Content

Introduction to AI System Hardware CPU, RAM, GPU, Interconnects, Storage, Network Controller; Introduction to AI Accelerators GPUs (Lecture); Introduction to System Software Operating System, Virtualization, Cloud; (Lecture)

Introduction to Containers and IDE (Jupyter Demo) (Lecture + Demo); Scheduling and Resource Management Introduction to schedulers and orchestration tools (Lecture); DeepOps: Deep-dive into Kubernetes with deployment of various AI-based services (Lecture + Demo)

DeepOps (contd) (Lecture + Demo); Design principles for building High Performance compute clusters for AI (Lecture); Implementation details for building High Performance compute clusters for AI (contd) (Lecture)

Frameworks for Accelerated Deep Learning Workloads - PyTorch (Lecture); Frameworks for Accelerated Deep Learning Workloads - PyTorch (contd) (Lecture + Demo); Accelerated PyTorch (Lecture + Demo)

Frameworks for Accelerated Deep Learning Workloads - TensorFlow (Lecture); Frameworks for Accelerated Deep Learning Workloads - TensorFlow (contd) (Lecture + Demo); Accelerated TensorFlow (Lecture + Demo)

Optimizing Deep Learning Training: Automated Mixed Precision (Lecture + Demo); Optimizing Deep Learning Training: Transfer Learning (Lecture + Demo)

Fundamentals of Distributed AI Computing: Multi-GPU and multi-node implementation (MPI, NCCL, RDMA) (Lecture); Distributed AI Computing: Horovod (Lecture + Demo)

Challenges with Distributed Deep Learning Training Convergence (Lecture + Demo); Fundamentals of Accelerating Deployment (Lecture + Demo)

Accelerating neural network inference in PyTorch and TensorFlow (Lecture + Demo); Accelerated Data Analytics (Lecture + Demo); Accelerated Machine Learning (Lecture + Demo)

Scale Out with DASK; Web visualizations to GPU accelerated crossfiltering (Lecture + Demo); Accelerated ETL Pipeline with SPARK

Applied AI: Smart City (Intelligent Video Analytics); Applied AI: Smart City (Intelligent Video Analytics) (Contd.)

Applied AI: Healthcare (Federated Learning, AI Assisted Annotation); Applied AI: Healthcare (Federated Learning, AI Assisted Annotation)

Course No. DSE505/DSE605

Course Name Cloud Computing

Content

Introduction to Cloud Computing

Cloud Computing Architecture

Service Management in Cloud Computing Data Management in Cloud Computing

Resource Management in Cloud

Cloud Security

Open Source and Commercial Clouds, Cloud Simulator Research trend in Cloud Computing, Fog Computing VM Resource Allocation, Management, and Monitoring Cloud-Fog-Edge enabled Analytics

Serverless Computing and FaaS Model

Case Studies and Recent Advancements

Course No. DSE 418/618

Course Name Advanced Natural Language Processing

Content

- Current Language Models and their philosophies
  - ELMO
  - BERT
  - RoBERTa
  - XLNet
  - GPT-1,2,3
- Transfer learning in NLP
- Knowledge Graph: current trends
- Question answering: current trends
- Dialogue systems: current trends
- Coreference Resolution and their applications
- Discourse Coherence
- Constituency Grammars, Constituency Parsing and their applications

Course No. DSE 420/620

Course Name Biomedical Text Mining

Content

Overview and importance of biomedical text. Ethics and principles of working with biomedical texts.

Basics of text mining and text preprocessing techniques. Regular expressions in Python to extract relevant snippets from unstructured text.

Word sense disambiguation, negation detection, and the NegEx tool for biomedical text mining.

Biomedical ontologies and databases - UMLS, SNOMED CT, MIMIC, PhenomeNET, MONDO, UK BioBank, i2b2.

Biomedical text mining systems - Metamap, cTAKES, CogStack, MedCAT, Komenti. Biomedical text classification to categorize relevant information.

Information extraction from biomedical text, de-identification of protected health information from clinical notes.

Course project to address an open problem or to participate in an international biomedical shared task related to biomedical text mining.

Course No. DSE 422/622

Course Name Biometrics: An Introduction to Research

Content

• Introduction to Biometrics: Understand the definition and history of biometrics recognition. Why, how, and where the biometrics recognition is used and how it is impacting our day-to-day lives.

• Properties and Different Types of Biometric Modalities: Learn different types of biometrics modalities such as the face, iris, and fingerprint and what the properties are because of which they are popular, and where they are effectively deployed including the Aadhar project.

• Algorithms for Biometrics Recognition: Understand and learn different forms of machine learning and deep learning algorithms developed for different biometric modalities including face, speech, and fingerprint, and implement them.

• Understanding Current Research in Biometrics: Learn how to conduct research in biometrics recognition which can result in a possible paper submission.

Solve several assignments, and evaluations through multiple quizzes and multiple projects to get practical experience in the biometrics field.

Course No. DSE406/ DSE606

Course Name Spatial Data Science and Application

Content

Introduction to remote sensing.

Development of AI techniques to analyze remotely sensed data.

Acquisition and preparation of 2D and 3D (Point Cloud, CAD, etc.) spatial data. Development of AI techniques used in acquiring and preparing 2D/3D data. Various 3D data spatial data preparation (LiDAR, Stereo Vision, etc...)

Machine Learning for 2D/3D data preparation (NeRF..).

Introduction to geodesy

Defining spatial coordinate systems

GNSS and GPS

Development of AI techniques for using GNSS and GPS data in various studies, e.g., automatizing 3D modeling, soil moisture estimation, finding outliers, etc.

2D and 3D Spatial data types, storage manipulation, and visualization.

Data formats, Spatial databases querying for information extraction, etc. Development of Geographic Information Systems (GIS)

Analysis of 2D and 3D spatial data using AI techniques.

Spatial Trees (2D/3D)

Generation of Synthetic Spatial Data.

Physics-based game engines

AI-based data generation.

Integration of the above techniques for solving real-world problems.

Course No. DSE409

Course Name Digital Image Processing and Applications in Bioimage Analysis

Content

Fundamentals of Digital image processing: Basic steps and components Mathematical tools in image processing

Image filtering: spatial domain and frequency domain

Image restorations: noise models

Image transforms

Morphological image processing

Programming application in Matlab

Image segmentation (2D and 3D)

Feature extraction

Image pattern classification

Biological image applications and tools: microscopy

Image segmentation, tracking and feature extraction in cell and developmental biology

Examples from research works

Non-image-based pattern recognition methods: brief overview (motivation and future)

Computer vision in biological research (motivation and future)