



**Course Name:** Data Warehousing and Data Mining

**Course Code:** DS-601

**(Semester- VI)**

**Course Broad Category:** Program Core (PC)

.....

**1. Course Prerequisite:** Basic knowledge about Database Management System and knowledge of probability, statistics, linear algebra.

**2. Course Learning Objectives:**

- Understand and implement classical models and algorithms in data warehouses and data mining
- Characterize the kinds of patterns that can be discovered by association rule mining classification and clustering.
- Develop skill in selecting the appropriate data mining algorithm for solving practical problems.

**0. Teaching methodology and evaluation system for the course:**

**Teaching methodology** – Lectures and Presentations, Interactive Discussions and Case Studies.

**Evaluation System –**

- . Mid-Term Exam (40 Marks)- Formative Assessment [Continuous Assessment 1 (CIA-1)]
- B. Internal Assessment (40 Marks)- Formative Assessment [Continuous Assessment 2(CIA-2)]
- C. End-Semester Exam (60 Marks)- Summative Assessment.

**0. Course Content:**

**Course Name:** Data Warehousing and Data Mining

**Course Code:** DS-601

**Hours per Week:** 3L: 0T: 0P

**Credits:** 3

Module	Topics	45L
1.	<b>Data Warehousing:</b> Multidimensional Data Model, Dimension Modelling, OLAP Operations, Slicing and Dicing, Warehouse Schema, Star Schema, Snowflake Schema, Advantages and Disadvantages of Snowflake Schema, Data Warehousing Architecture, Virtual Data Warehouse, Advantages and Disadvantages of Virtual Data Warehouse, Metadata, Types of Metadata, OLAP Engine, Different Options for OLAP Engine, Data Extraction, Data Cleaning, Loading, Refreshing.	6L

Module	Topics	45L
3.	<b>Data Mining:</b> Different Definitions of Data Mining, KDD vs. Data Mining, Stages of KDD, DBMS vs. DM, AI vs. DM, Classifications of Data Mining, Stages of KDD, DM Techniques, Discovery Driven Tasks, Classification, Frequent Episodes, Discovery of Association Rules, Clustering, Deviation Detection, Mining Problems, Applications of DM, Other Mining Problems.	4L
4.	<b>Association Rules:</b> Item set, Support, Confidence, Problem Decomposition, Frequent Item Set, Maximal Frequent Set, Border Set, Applications of Data Mining, Spotting Fraudulent Behaviour, Astronomy etc., Association Rules, Informal a priori Algorithm for Learning Association Rules, Finding Frequent Sets and Association Rules, Formal a priori Algorithm for Association Rule.	5L
5.	<b>Classification</b> – Pattern: Labelled Pattern, Approaches of Classification, Evaluation of Classifiers, Normalized Confusion Matrix, Accuracy, Precision, Recall and F1 score, Cross Validation Technique, Classification Techniques.	6L
6.	<b>Decision Trees:</b> Inductive Learning, ID3 Program, Algorithm for Building Decision Trees, Advantages of Decision Trees for Classification Purpose, Development of Decision Trees for Different Training Data Sets, Rule Extraction from Pattern Set, Covering the instances, Extraction of rules, Instance Space, Covering Algorithm.	5L
7.	<b>Bayesian Belief Nets (DAG):</b> K nearest Neighbour, ANN, learning in ANN, Perceptron as a model of neuron, Single and multiplayer Perceptron for classification and knowledge representation, Back propagation Network, Supervised, Reinforcement and Unsupervised Learning.	5L
8.	<b>Classification (Complex):</b> Support Vector Machine (SVM), Generalization Error, SVM to find out the best classification, Margin.	2L
9.	<b>Clustering:</b> Partitioned and Hierarchical Clustering, k means Clustering, Fast k Means Clustering, Fuzzy K-means Clustering, Hierarchical Clustering, Agglomerative and Divisive Hierarchical Clustering, Single Linkage, Complete Linkage and Average Linkage Clustering.	4L
10.	<b>Clustering (Complex):</b> Outlier Detection, Outlier vs. Cluster, Types of Outliers, Outlier Detection Methodologies, Supervised, Unsupervised and Semi supervised detection, Statistical Approaches, Parametric and Non Parametric Methods, Proximity Based Methods, Clustering Based Methods.	5L
11.	<b>Web Mining:</b> Web Mining Techniques, Web Content Mining, Web Structure Mining, Web Usage Mining, Text Mining.	3L

## 5. Books:

### 1. Data Mining Techniques

*Author:* Arun K Pujari

*Publisher:* Universities Press.

### 2. Data Mining

*Author:* Vikram Pudi, P. Radha Krishna

*Publisher:* Oxford University Press.

### 3. Data Mining

*Author:*J. Han, M. Kamber, J. Pei

*Publisher:* Elsevier.

## 6. COURSE OUTCOMES (COs)

On completion of the course students will be able to

<b>Course Outcomes</b>	<b>Details</b>	<b>Action Verb</b>	<b>Knowledge Level</b>
DS-601.CO1	Describe the fundamental concepts of data warehousing, multidimensional modelling, OLAP operations, ETL processes, and metadata types.	Remember & Understand	Level 1-Remember, Level 2-Understand
DS-601.CO2	Explain and apply data mining concepts including KDD processes, discovery-driven tasks, classification, clustering, association rule mining, and deviation detection.	Explain, Apply	Level 3-Apply
DS-601.CO3	Implement association rule mining techniques, including support, confidence, Apriori algorithm, frequent itemset generation, and rule extraction for various applications.	Apply	Level 3-Apply
DS-601.CO4	Analyze classification models using decision trees, Bayesian networks, KNN, ANN, SVM, and evaluate them using accuracy, precision, recall, and F-score metrics.	Analyze	Level 4-Analyze
DS-601.CO5	Evaluate clustering and outlier detection methodologies such as k-means, hierarchical clustering, fuzzy k-means, statistical and proximity-based methods for real-world datasets.	Evaluate	Level 5-Evaluate
DS-601.CO6	Design and develop integrated data mining solutions using data warehousing, classification, clustering, and web mining techniques to solve real-world analytical problems.	Design, Develop	Level 6-Develop

## 7. Mapping of course outcomes to module / course content

Module	CO1	CO2	CO3	CO4	CO5	CO6
1	3	2	0	0	0	0
3	0	2	0	2	0	0
4	0	2	3	2	0	0
5	0	0	3	2	0	0
6	0	0	2	1	0	0
7	0	0	3	2	1	0
8	0	0	2	3	1	0
9	0	0	3	3	1	0
10	0	0	2	3	2	0
11	1	2	2	1	1	2

## 8. Mapping of the Course outcomes to Program Outcomes (PO)

Mapping of COs with POs and PSOs (Course Articulation Matrix):

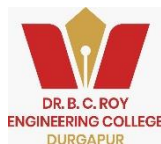
	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
CO1	3	3	2	1	1	2						1
CO2	3	3	3	2	2	2	1	1		1		2
CO3	2	2	3	2	2	2	1	1		1		2
CO4	2	3	3	3	2	2	1	1	1	2	1	2
CO5	2	2	3	3	3	2	1	1	1	2	1	2
CO6	3	3	3	3	3	3	2	2	2	2	2	3

## 9. Mapping to Program Specific Outcome (PSO)

	PSO1	PSO2	PSO3	PSO4
CO1	3	1	2	1
CO2	3	3	3	2
CO3	3	3	3	2
CO4	3	3	3	2
CO5	3	3	3	2
CO6	3	3	3	3

Reference: NIT Durgapur

\*\*\* End of Syllabus\*\*\*



**Course Name:** Fundamentals of Quantum Computing

**Course Code:** DS-606

**(Semester- VI)**

**Course Broad Category:** Professional Elective (PE)

.....

**1. Course Prerequisite:** Linear Algebra, Digital Circuits

**2. Course Learning Objectives:**

This course aims to provide an understanding of the fundamentals of quantum physics from a computer scientist's perspective, with emphasis on how quantum theory describes physical reality and its philosophical implications for computation.

Upon successful completion of the course, students will be able to:

- Understand the mathematical foundations required for quantum computing, including vector spaces, linear operators, and matrices.
- Explain the concept of quantum states, including state vectors and density operators.
- Analyze and represent quantum systems using Dirac notation and matrix formalism.
- Describe the principles of quantum measurement, superposition, and entanglement.
- Apply fundamental quantum concepts to basic quantum information and computation models.

**3. Teaching methodology and evaluation system for the course:**

**Teaching methodology -**

Lectures and Presentations, Interactive Discussions and Case Studies.

**Evaluation System –**

- A. Mid-Term Exam (20 Marks)- Summative Assessment (CIA-1)
- B. Internal Assessment (20 Marks)- Formative Continuous Assessment [Continuous Assessment 1(CIA-2)]
- C. End-Semester Exam (60 Marks)- Summative Assessment.

**4. Course Content:**

**Course Name:** Fundamentals of Quantum Computing (DS-606)

**Course Code:** DS-606

**Hours per Week:** 3L: 0T: 0P

**Credits:** 3

Module	Topics	
1.	<b>Qubit &amp; Quantum States:</b> The Qubit, Vector Spaces. Linear Combination Of Vectors, Uniqueness of a spanning set, basis & dimensions, inner Products, orthonormality, Gram-Schmidt orthogonalization, bra-ket formalism, the Cauchy-Schwarz and triangle Inequalities.	5L
2.	<b>Matrices &amp; Operators:</b> Observables, The Pauli Operators, Outer Products, The Closure Relation, Representation of operators using matrices, outer products & matrix representation, matrix representation of operators in two dimensional spaces, Pauli Matrix, Hermitian unitary and normal operator, Eigen values & Eigen Vectors, Spectral Decomposition, Trace of an operator, important properties of Trace, Expectation Value of Operator, Projection Operator, Positive Operators,	10L
3.	<b>Linear Algebraic Structures in Quantum Mechanics:</b> Commutator Algebra, Heisenberg uncertainty principle, polar decomposition & singular values, Postulates of	5L

	Quantum Mechanics.	
4.	<b>Tensor Products:</b> Representing Composite States in Quantum Mechanics, Computing inner products, Tensor products of column vectors, operators and tensor products of Matrices. Density Operator: Density Operator of Pure & Mix state, Key Properties, Characterizing Mixed State, Practical Trace & Reduce Density Operator, Density Operator & Bloch Vector.	6L
5.	<b>Quantum Measurement Theory:</b> Distinguishing Quantum states & Measures, Projective Measurements, Measurement on Composite systems, Generalized Measurements, Positive Operator- Valued Measures.	8L
6.	<b>Introduction to Error Correction:</b> Fault-tolerance, Simple error correcting codes	4L
7.	<b>Recent trends in Quantum Computing Research:</b> Quantum Computing Applications of Genetic Programming, IBM QX Architecture, NISQ era processors, Quantum advantage claims.	7L
		45L

### 5. Text Book:

1. Quantum Computation and Quantum Information  
*Authors:* Michael A. Nielsen and Isaac L. Chuang  
*Publisher:* Cambridge University Press
2. Quantum Computing: A Gentle Introduction  
*Authors:* Eleanor Rieffel and Wolfgang Polak  
*Publisher:* MIT Press
3. Quantum Mechanics and Path Integrals  
*Authors:* Richard P. Feynman and A. R. Hibbs  
*Publisher:* McGraw-Hill

### 6. COURSE OUTCOMES (COs)

On completion of the course students will be able to

Course Outcomes	Details	Action Verb	Knowledge Level
<b>DS-606.CO1</b>	Understand and explain the mathematical foundations of qubits and quantum states.	Understand, Apply	Level 2-Understand, Level 3-Apply
<b>DS-606.CO2</b>	Analyze and apply matrix representations of quantum operators.	Apply, Analyze	Level 3-Apply, Level 4- Analyze
<b>DS-606.CO3</b>	Evaluate composite quantum systems using tensor products.	Analyze, Apply	Level 3-Apply, Level 4- Analyze,
<b>DS-606.CO4</b>	Apply and interpret density operators for pure and mixed states.	Analyze, Evaluate	Level 4-Analyze, Level 5-Evaluate
<b>DS-606.CO5</b>	Explain and implement quantum measurement theory.	Understand, Apply, Analyze	Level 2-Understand, Level 3-Apply, Level 4-Analyze
<b>DS-606.CO6</b>	Analyze quantum error-correction concepts and recent trends in quantum computing.	Analyze, Evaluate	Level 4-Analyze, Level 5-Evaluate

## 7. Mapping of course outcomes to module / course content

Course Outcome (CO)	Mapped Course Content / Module Topics
CO1	Qubit, Vector Spaces, Linear Combination of Vectors, Spanning Set, Basis & Dimensions, Inner Products, Orthonormality, Gram–Schmidt, Bra–ket Formalism, Cauchy–Schwarz & Triangle Inequalities
CO2	Observables, Pauli Operators, Outer Products, Closure Relation, Matrix Representation of Operators, Hermitian/Unitary/Normal Operators, Eigenvalues & Eigenvectors, Spectral Decomposition, Trace & Properties, Expectation Value, Projection & Positive Operators, Commutator Algebra, Heisenberg Principle, Polar Decomposition, Singular Values, Postulates
CO3	Composite State Representation, Inner Products in Composite Systems, Tensor Product of Vectors, Tensor Product of Operators, Tensor Product of Matrices
CO4	Density Operator (Pure & Mixed States), Key Properties, Characterizing Mixed States, Trace & Partial Trace, Reduced Density Operators, Bloch Vector Representation
CO5	Distinguishing Quantum States, Projective Measurements, Measurement on Composite Systems, POVMs, Generalized Measurements
CO6	Fault Tolerance, Simple Error-Correcting Codes, Quantum Genetic Programming, IBM QX Architecture, NISQ Processors, Quantum Advantage Claims

## 8. Mapping of the Course outcomes to Program Outcomes (PO) (1 = Low, 2 = Moderate, 3 = High)

### Mapping of COs with POs and PSOs (Course Articulation Matrix):

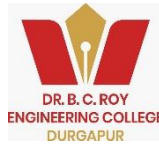
CO / PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
CO1	3	2	1	1	1	0	0	0	0	0	0	0
CO2	3	3	2	2	2	0	0	0	0	0	0	0
CO3	3	3	2	2	2	0	0	0	0	0	0	0
CO4	2	3	2	3	2	0	0	0	0	0	0	0
CO5	2	3	2	3	2	0	0	0	0	0	0	0
CO6	2	3	1	2	3	0	0	0	0	0	0	1

## 9. Mapping to Program Specific Outcome (PSO)

CO / PSO	PSO1	PSO2	PSO3	PSO4
CO1	2	1	2	1
CO2	2	1	3	1
CO3	2	1	3	1
CO4	2	1	3	1
CO5	1	1	2	2
CO6	1	1	2	2

**Reference: NIT Trichy, IIT Madras, and IISc Bangalore**

\*\*\*\* End of Syllabus\*\*\*\*



**Course Name:** Advanced Statistical Analysis for Data Science

**Course Code: DS 625**

**(Semester- VI)**

**Course Broad Category:** Open Elective (OE)

**1. Course Prerequisite:** Probability and Statistics for Data Science

Course Code: BSC-M 301 (Semester III)

**2. Course Learning Objectives:**

This course aims to provide an understanding of the fundamentals of quantum physics from a computer scientist's perspective, with emphasis on how quantum theory describes physical reality and its philosophical implications for computation.

Upon successful completion of the course, students will be able to:

- Understand the mathematical foundations required for quantum computing, including vector spaces, linear operators, and matrices.
- Explain the concept of quantum states, including state vectors and density operators.
- Analyze and represent quantum systems using Dirac notation and matrix formalism.
- Describe the principles of quantum measurement, superposition, and entanglement.
- Apply fundamental quantum concepts to basic quantum information and computation models.

**3. Teaching methodology and evaluation system for the course:**

**Teaching methodology -**

Lectures and Presentations, Interactive Discussions and Case Studies.

**Evaluation System –**

- A. Mid-Term Exam (20 Marks)- Summative Assessment (CIA-1)
- B. Internal Assessment (20 Marks)- Formative Continuous Assessment [Continuous Assessment 1(CIA-2)]
- C. End-Semester Exam (60 Marks)- Summative Assessment.

**4. Course Content:**

**Course Name:** Advanced Statistical Analysis for Data Science (DS-625)

**Course Code:** DS-625

**Hours per Week:** 3L: 1T: 0P

**Credits:** 3

Module	Topics	
1.	<b>Introduction to Multivariate Analysis:</b> Multivariate data structures; data visualization using heat maps, pair plots; covariance structure; correlation matrix; multicollinearity; Mahalanobis distance; missing data handling; outlier detection; standardization and normalization techniques.	6L
2.	<b>Principal Component Analysis (PCA) and Factor Analysis:</b> Eigen-decomposition; PCA for feature reduction; reconstruction error; kernel PCA; factor analysis models; rotation techniques; factor loading interpretation; applications in machine learning preprocessing.	6 L
3	<b>Multivariate Regression, MANOVA, and Advanced Techniques:</b> Multivariate linear models using matrix algebra; multivariate multiple regression; collinearity detection using VIF and condition index; MANOVA tests (Wilk's Lambda, Pillai's Trace); Canonical Correlation Analysis (CCA); Partial Least Squares (PLS).	8 L

4.	<b>Applied Statistical Inference:</b> Basic concepts of sampling theory; probability and non-probability sampling methods; simple random, stratified and cluster sampling; sampling distributions and standard errors; bootstrap and jackknife resampling; permutation tests; cross-validation; bias-variance trade-off; model selection and assessment; statistical learning perspective; introduction to Bayesian inference for data science.	10L
5.	<b>Time Series Analysis:</b> Stationarity assessment using the Augmented Dickey–Fuller (ADF) test; components of time series decomposition; forecasting with Autoregressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA) models; trend and seasonality estimation using exponential smoothing methods such as Holt and Winters techniques; fundamentals of machine-learning-based forecasting; evaluation of predictive performance using metrics like Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE); and an introduction to multivariate time-series modeling through Vector Auto Regression (VAR) models.	10L
6.	<b>Index Numbers:</b> Advanced concepts: chain indices, seasonal indices; Fisher’s ideal index; tests of consistency; consumer price index (CPI), wholesale price index (WPI); use of index numbers in business intelligence and inflation analysis.	5L
		45L

### 5. Text Book:

1. Johnson, R.A. & Wichern, D.W. – Applied Multivariate Statistical Analysis, Pearson.
2. Anderson, T.W. – An Introduction to Multivariate Statistical Analysis, Wiley.

### Reference Books

1. Montgomery, D.C., Jennings, C.L. & Kulahci, M. – Introduction to Time Series Analysis and Forecasting, Wiley.
2. Gujarati, D.N. – Basic Econometrics, McGraw-Hill.
3. Sharma, S. – Applied Multivariate Techniques, Wiley.

### 6. COURSE OUTCOMES (COs)

On completion of the course students will be able to

Course Outcomes	Details	Action Verb	Knowledge Level
<b>DS-625.CO1</b>	Understand and analyze multivariate data structures and inter-variable relationships.	Understand, Apply	Level 2-Understand, Level 3-Apply
<b>DS-625.CO2</b>	Apply dimension reduction and factor extraction techniques for high-dimensional data.	Apply, Analyze	Level 3-Apply, Level 4-Analyze
<b>DS-625.CO3</b>	Analyze and evaluate multivariate predictive models for classification, regression, and clustering using appropriate performance measures.	Analyze, Evaluate	Level 4-Analyze, Level 5-Evaluate
<b>DS-625.CO4</b>	Understand, apply, and analyze sampling theory, resampling techniques, and model validation methods in data science using real-world datasets.	Understand, Apply, Analyze	Level 2-Understand, Level 2-Apply, Level 4- Analyze
<b>DS-625.CO5</b>	Analyze and apply time series models for forecasting real-world temporal data.	Analyze, Apply,	Level 3-Apply, Level 4-Analyze
<b>DS-625.CO6</b>	Analyze, evaluate, and develop index numbers for business, economic, and market analytics applications.	Analyze, Evaluate	Level 4-Analyze, Level 5-Evaluate, Level 6- Develop

## 7. Mapping of course outcomes to module / course content

Course Outcome (CO)	Mapped Course Content / Module Topics
CO1	<b>Introduction to Multivariate Analysis:</b> Multivariate data structures; data visualization using heatmaps, pair plots; covariance structure; correlation matrix; multicollinearity; Mahalanobis distance; missing data handling; outlier detection; standardization and normalization techniques.
CO2	<b>Principal Component Analysis (PCA) and Factor Analysis:</b> Eigen-decomposition; PCA for feature reduction; reconstruction error; kernel PCA; factor analysis models; rotation techniques; factor loading interpretation; applications in machine learning preprocessing.
CO3	<b>Multivariate Regression, MANOVA, and Advanced Techniques:</b> Multivariate linear models using matrix algebra; multivariate multiple regression; collinearity detection using VIF and condition index; MANOVA tests (Wilk's Lambda, Pillai's Trace); Canonical Correlation Analysis (CCA); Partial Least Squares (PLS).
CO4	<b>Applied Statistical Inference:</b> Basic concepts of sampling theory; probability and non-probability sampling methods; simple random, stratified and cluster sampling; sampling distributions and standard errors; bootstrap and jackknife resampling; permutation tests; cross-validation; bias-variance trade-off; model selection and assessment; statistical learning perspective; introduction to Bayesian inference for data science.
CO5	<b>Time Series Analysis:</b> Stationarity assessment using the Augmented Dickey–Fuller (ADF) test; components of time series decomposition; forecasting with Autoregressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA) models; trend and seasonality estimation using exponential smoothing methods such as Holt and Winters techniques; fundamentals of machine-learning-based forecasting; evaluation of predictive performance using metrics like Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE); and an introduction to multivariate time-series modeling through Vector Auto Regression (VAR) models.
CO6	<b>Index Numbers:</b> Advanced concepts: chain indices, seasonal indices; Fisher's ideal index; tests of consistency; consumer price index (CPI), wholesale price index (WPI); use of index numbers in business intelligence and inflation analysis.

## 8. Mapping of the Course outcomes to Program Outcomes (PO) (1 = Low, 2 = Moderate, 3 = High)

### Mapping of COs with POs and PSOs (Course Articulation Matrix):

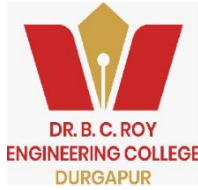
CO / PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
CO1	3	3	2	2	1	0	0	0	0	0	0	1
CO2	3	2	3	2	2	1	0	0	0	0	0	2
CO3	2	3	3	3	2	1	0	0	0	0	0	2
CO4	2	0	3	3	3	2	0	0	0	0	0	2
CO5	1	2	2	3	3	1	0	0	0	0	0	1
CO6	1	1	2	2	2	1	0	0	0	0	0	3

## 9. Mapping to Program Specific Outcome (PSO)

CO / PSO	PSO1	PSO2	PSO3	PSO4
CO1	2	1	1	1
CO2	1	2	2	2
CO3	1	1	2	1
CO4	2	1	2	1
CO5	3	2	3	2
CO6	1	1	2	1

**Reference: IIT (ISM) Dhanbad, IIT, Kharagpur**

\*\*\*\* End of Syllabus\*\*\*\*



**Course Name:** Deep Learning  
**Course Code:** DS-626  
**(Semester- VI)**  
**Course Broad Category:** Program Core (PC)

.....

**1. Course Prerequisite:** Linear algebra, Calculus, Probability and Statistics, Machine Learning

**2. Course Learning Objectives:**

- i. To develop a comprehensive understanding of the theoretical concepts underlying deep learning, including neural network architectures, activation functions, loss functions, optimization techniques, and regularization methods.
- ii. To gain the ability to design, train, and evaluate various deep learning architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers using frameworks like TensorFlow or PyTorch to solve real-world problems.
- iii. To acquire the skills to critically analyze model behavior, interpret results, fine-tune hyperparameters, and apply techniques such as transfer learning and model deployment to improve accuracy and generalization

**3. Teaching methodology and evaluation system for the course:**

**Teaching methodology** – Lectures and Presentations, Interactive Discussions and Case Studies.

**Evaluation System –**

- A. Mid-Term Exam (40 Marks)- Formative Assessment [Continuous Assessment 1 (CIA-1)]
- B. Internal Assessment (40 Marks)- Formative Assessment [Continuous Assessment 2(CIA-2)]
- C. End-Semester Exam (60 Marks)- Summative Assessment.

**4. Course Content:**

**Course Name:** Deep Learning  
**Course Code:** DS-626

**Hours per Week:** 3L: 0T: 0P

**Credits:** 3

Module	Topics	45L
1.	<b>Machine Learning Basics:</b> Extracting meaning from data, expert system, learning algorithms, overfitting and underfitting, regularization, hyperparameters and validation sets, estimator, bias and variance, ML estimation, Bayesian statistics, supervised learning, unsupervised learning, Stochastic Gradient Descent, building a machine learning algorithm, challenges motivating Deep Learning.	8L
2.	<b>Fundamentals of Feedforward Networks:</b> Single-layer and multilayer feedforward networks, Neural Network Graphs, activation functions, deep feedforward networks, hidden units, Learning XOR, gradient-based learning, Back-propagation algorithm and other differentiation algorithms.	5L
3.	<b>Regularization for Deep Learning:</b> Parameter Norm Penalties, Norm Penalties as Constrained Optimization, Regularization and Under-Constrained Problems, Dataset Augmentation, Early Stopping, Sparse Representations, Dropout.	5L
4.	<b>Optimization for Training Deep Models:</b> How Learning Differs from Pure Optimization, Challenges in Neural Network Optimization, Basic Algorithms, Parameter Initialization Strategies, Algorithms with Adaptive Learning Rates, Approximate Second-Order Methods, Batch Normalization.	5L
5.	<b>Convolutional Networks:</b> The Convolution Operation, Pooling, Variants of the Basic Convolution Function, Structured Outputs, Structured outputs and datatypes.	6L
6.	<b>Sequence Modelling, Recurrent Neural Networks (RNN):</b> Unfolding Computational Graphs, RNNs, Bidirectional RNNs, LSTM.	5L
7.	<b>Autoencoders:</b> Undercomplete Autoencoders, Regularized Autoencoders, Stochastic Encoders and Decoders, Denoising Autoencoders, Contractive Autoencoders.	5L
8.	<b>Some Popular Deep Networks and Applications:</b> Generative Adversarial Networks, VGG net, ResNet, Inception Net. Applications of deep learning.	6L

## 5.

### Text Book:

1. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, The MIT Press, 2017
2. Charu C. Aggarwal, *Neural Networks and Deep Learning*, Springer, 2018.
3. F. Chollet, *Deep Learning with Python*, Manning Publications Co., 2018.

### Reference Books:

1. F. Chollet, M. Watson, *Deep Learning with Python (3rd Edition)*, Manning Publications Co., 2025.
2. N. Buduma, *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*, O'REILLY, 2017

## 6. COURSE OUTCOMES (COs)

On completion of the course students will be able to

Course Outcomes	Details	Action Verb	Knowledge Level
<b>DS-626.CO1</b>	Understand the foundational concepts of machine learning that motivate deep learning, including generalization, regularization, bias–variance, and gradient-based learning.	Understand	L2 – Understand
<b>DS-626.CO2</b>	Explain and implement feedforward neural networks, activation functions, and backpropagation for supervised learning tasks.	Implement	L3 – Apply
<b>DS-626.CO3</b>	Apply regularization and optimization strategies to train deep neural networks effectively and evaluate their influence on convergence and generalization.	Apply/ Evaluate	L5 – Evaluate
<b>DS-626.CO4</b>	Design and analyse convolutional neural networks (CNNs) and sequence models such as recurrent neural networks (RNNs) and long short-term memories (LSTMs) for vision and sequence-based applications.	Design	L6 – Create
<b>DS-626.CO5</b>	Develop and compare different autoencoder architectures and evaluate their representational capabilities for dimensionality reduction and denoising tasks.	Develop	L6 – Create
<b>DS-626.CO6</b>	Explain the architectures of modern deep networks and apply deep learning techniques to real-world applications in computer vision, natural language processing (NLP), and generative modelling.	Apply/ Explain	L3 – Apply

## 7. Mapping of course outcomes to module / course content

Module	CO1	CO2	CO3	CO4	CO5
1	3	1	2	1	1
2	1	3	1	2	1
3	2	1	3	1	2
4	2	1	3	1	1
5	1	1	1	3	1
6	1	1	1	3	1
7	1	1	1	1	3
8	2	1	1	2	1

## 8. Mapping of the Course outcomes to Program Outcomes (PO)

### Mapping of COs with POs and PSOs (Course Articulation Matrix):

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
CO1	3	2	2	2	2	1	1	--	--	2	1	1
CO2	3	3	3	2	3	1	1	--	1	2	1	1
CO3	3	3	3	2	3	1	1	--	1	2	1	1
CO4	3	3	3	2	3	1	1	--	1	2	1	1
CO5	3	2	3	2	3	1	1	--	1	2	1	1
CO6	3	2	3	2	3	1	1	--	1	3	1	1

## 9. Mapping to Program Specific Outcome (PSO)

	PSO1	PSO2	PSO3	PSO4
CO1	3	2	2	1
CO2	3	3	3	1
CO3	3	3	3	1
CO4	3	3	3	1
CO5	3	3	3	1
CO6	3	3	3	1

Reference: NIT Durgapur

\*\*\*\* End of Syllabus\*\*\*\*



**Course Name:** Natural Language Processing

**Course Code:** DS-633

**(Semester- VI)**

**Course Broad Category:** Open Elective (OEC)

**1. Course Prerequisite:** Linear Algebra, Introduction to Probability, Python coding, Theory of Computation, Compiler Design

**2. Course Learning Objectives:**

The objective of this course is to develop a strong understanding of Natural Language Processing concepts, linguistic foundations, and text preprocessing techniques.

Upon successful completion of the course, students will be able to:

- Students will learn and apply statistical, machine learning, and deep learning models—including transformer-based architecture
- To analyze and generate natural language.
- By the end of the course, students will be able to build NLP applications, evaluate their performance, and understand ethical considerations and responsible AI deployment.

**3. Teaching methodology and evaluation system for the course:**

**Teaching methodology -**

Lectures and Presentations, Interactive Discussions and Case Studies.

**Evaluation System –**

- A. Mid-Term Exam (20 Marks)- Summative Assessment (CIA-1)
- B. Internal Assessment (20 Marks)- Formative Continuous Assessment [Continuous Assessment 1(CIA-2)]
- C. End-Semester Exam (60 Marks)- Summative Assessment.

**4. Course Content:**

**Course Name:** Natural Language Processing (DS-633)

**Course Code:** DS-633

**Hours per Week:** 3L: 0T: 0P

**Credits:** 3

Module	Topics	
1.	<p><b>Introduction to NLP and Regular Expressions and Automata Recap:</b> Introduction to Natural Language (NL) Why is it hard to process a natural language? Levels of Language Processing, Linguistic Fundamentals for NLP. Regular Expression, Finite State Automata.</p> <p><b>Fundamentals of Text Mining :</b> Data extraction, Introduction to pre-process text, Steps in text pre- processing, Tokenization, Normalization, Stop word removal, Stemming, lemmatization, Morphological Analysis &amp; Finite State Transducers (Code demo: Tokenization, Stop-Word-Removel, Stemming, Lemmatization), Text vectorization and DTM, TFIDF and Topic modeling, Text visualization</p>	5L

2.	<p><b>Statistical NLP:</b> Probabilistic language models, N-gram language models, Smoothing techniques, Maximum Likelihood Estimation (MLE), Hidden Markov Models (HMMs), Viterbi Algorithm, Part-of-Speech tagging with HMMs, MEMM and CRF</p> <p><b>Machine Learning for NLP:</b> Supervised and unsupervised learning in NLP, Classification Models: Naive Bayes, Decision Trees, Maximum Entropy, Support Vector Machines, Neural Networks</p>	10L
3.	<p><b>Syntax and Semantics:</b> Context-Free Grammars, Constituency Parsing, Dependency Parsing, Semantic Parsing, Semantic Role Labeling, Parsing Algorithms Meaning Representation: Logical Semantics, Word Sense Disambiguation, Distributional Semantics, Semantic Role Labelling, n-gram and Word2Vec, GloVe; Discourse Processing: Anaphora and Coreference Resolution and Discourse Connectives. Machine Translation</p> <p><b>Word Embedding:</b> Introduction to Word embedding, Limitations of count vectorizers, Cosine Similarity, Co-occurrence Matrix, Pre-trained word embedding, Applications of word embedding, Vector space models, Manipulating words in Vector spaces.</p>	10L
4.	<p><b>Neural Network for NLP:</b> Recurrent neural networks, LSTMs/GRUs, Neural Sequence Models, Contextualized Word Embeddings: TagLM, ELMO, ULMFIT, etc., Attention Mechanism</p> <p><b>Transformers for NLP:</b> Transformers, Self-attention Mechanism, Sub-word tokenization, Positional encoding, Pre-trained Language Models (PLMs): BERT, GPT, etc. Fine-tuning and transfer learning.</p>	8L
5.	<p><b>Natural Language Generation:</b> Natural Language Generation, Decoding schemes: greedy, Random sampling, Top-k, Top-p, Speculative sampling, etc.</p> <p><b>Large Language Models (LLMs) with Hugging face:</b> Parameter Efficient Fine Tuning: Prefix-coding, LORA, QLORA, etc. Emergent Behavior: In-context learning, Instruction Fine Tuning, RLHF, DPO.</p>	7L
6.	<p><b>Applications of NLP:</b> Information Retrieval, Text Classification, Sentiment Analysis, Machine Translation, Speech Recognition, Summarization, NLI, Fact-checking, etc. Retrieval Augmented generation (RAG).</p> <p><b>Model Explainability and Ethics in NLP:</b> Attention maps, Attention-flow/rollout, Integrated gradients, etc. Ethical and cultural considerations and biases</p>	5L
		45L

## 5. Text Book:

1. Daniel Jurafsky and James H Martin. Speech and Language Processing, 2e, Pearson Education, 2024
2. Lewis Tunstall, Leandro Von Werra, Thomas Wolf. Natural Language Processing with Transformers, O'Reilly, 2023
3. Sowmya Vajjala, Bodhisattwa Majumder, Anju Gupta, Harshit Surana. Practical Natural Language Processing, O'Reilly, 2021
4. James Allen. Natural language Understanding 2e, Pearson Education, 2007
5. Akshar Bharati, Vineet Chaitanya, Rajeev Sangal. Natural Language Processing: A Paninian Perspective, PHI, 2023
6. Tanveer Siddiqui., U.S. Tiwary. Natural Language Processing and Information Retrieval, OUP, 2008

## 6. COURSE OUTCOMES (COs)

On completion of the course students will be able to

Course Outcomes	Details	Action Verb	Knowledge Level
DS-633.CO1	Understand and explain the linguistic foundations, text preprocessing, and NLP pipeline.	Understand, Explain	Level 2- Understand
DS-633.CO2	Apply statistical and machine learning-based NLP models such as N-gram, HMM, CRFs, and classification algorithms.	Apply	Level 3-Apply
DS-633.CO3	Analyze syntactic and semantic processing techniques such as parsing, word embeddings, and semantic representations.	Analyze	Level 4-Analyze
DS-633.CO4	Implement and evaluate transformer-based architectures, deep learning models, and contextual embeddings.	Implement, Evaluate	Level 3-Apply Level 5-Evaluate
DS-633.CO5	Build NLP applications such as text classification, sentiment analysis, MT, RAG-based systems, and generation models.	Build, Create	Level 6-Create
DS-633.CO6	Evaluate ethical considerations, explainability, and responsible deployment of NLP and LLM systems.	Evaluate, Explain	Level 5-Evaluate, Level 2-Understand

## 7. Mapping of course outcomes to module / course content

Course Outcome (CO)	Mapped Course Content / Module Topics
DS-633.CO1	<b>Module 1:</b> Introduction to NLP & Regular Expressions, Text Preprocessing — Introduction to NLP, Levels of Language Processing, Linguistic Fundamentals, Regular Expressions, Finite State Automata, Tokenization, Normalization, Stop Word Removal, Stemming, Lemmatization, Morphological Analysis, Text Vectorization, TF-IDF, Topic Modeling, Text Visualization
DS-633.CO2	<b>Module 2:</b> Statistical NLP & Machine Learning for NLP — Probabilistic models, N-grams, Smoothing, MLE, HMM, Viterbi Algorithm, POS tagging, MEMM, CRF, Naïve Bayes, Decision Trees, Maximum Entropy, SVM, Neural Networks
DS-633.CO3	<b>Module 3:</b> Syntax and Semantics & Word Embedding — Constituency and Dependency Parsing, Semantic Role Labelling, Meaning Representation, WSD, Distributional Semantics, Word2Vec, GloVe, Cosine Similarity, Vector Space Models, Discourse Processing, Machine Translation

DS-633.CO4	<b>Module 4:</b> Neural Network for NLP & Transformers — RNNs, LSTM, GRUs, Attention Mechanism, Contextualized Embeddings (ELMO, ULMFit, TagLM), Transformer Architecture, Pretrained Language Models (BERT, GPT), Fine-Tuning, Transfer Learning
DS-633.CO5	<b>Module 5:</b> Natural Language Generation & LLM Techniques — NLG techniques, Decoding schemes (Greedy, Top-K, Top-P, Sampling), PEFT (LoRA, QLoRA, Prefix Tuning), In-Context Learning, Instruction Tuning, RLHF, HuggingFace Models
DS-633.CO6	<b>Module 6:</b> NLP Applications & Ethics — Applications: IR, Sentiment Analysis, Text Classification, Speech Recognition, Summarization, RAG Systems, Explainability Techniques, Bias, Ethics, Responsible AI

**8. Mapping of the Course outcomes to Program Outcomes (PO) (1 = Low, 2 = Moderate, 3 = High)**

**Mapping of COs with POs and PSOs (Course Articulation Matrix):**

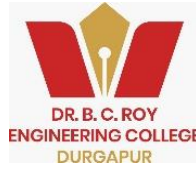
CO / PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
DS-633.CO1	3	2	1	1	1	0	0	0	0	0	0	0
DS-633.CO2	3	3	2	2	2	0	0	0	0	0	0	0
DS-633.CO3	3	3	2	2	2	0	0	0	0	0	0	0
DS-633.CO4	2	3	2	3	2	0	0	0	0	0	0	0
DS-633.CO5	2	3	2	3	2	0	0	0	0	0	0	0
DS-633.CO6	2	3	1	2	3	0	0	0	0	0	0	1

**9. Mapping to Program Specific Outcome (PSO)**

CO / PSO	PSO1	PSO2	PSO3	PSO4
DS-633.CO1	2	1	2	1
DS-633.CO2	2	1	3	1
DS-633.CO3	2	1	3	1
DS-633.CO4	2	1	3	1
DS-633.CO5	1	1	2	2
DS-633.CO6	1	1	2	2

**Reference: IIT Madras, and IIT Dharwad**

\*\*\*\*\* End of Syllabus\*\*\*\*\*



**Course Name: Data Mining Lab**

**Course Code: DS-691**

**(Semester– VI)**

**Course Broad Category: Program Core**

.....  
**1. Course Prerequisite:**

- Basic knowledge of Python Programming.

**2. Course Learning Objectives:**

- Illustrate techniques such as classification, clustering, and other data analysis methods on large datasets.
- Expertise in leveraging data mining for practical applications.

**3. Teaching methodology and evaluation system for the course:**

**Teaching methodology: Instruction:** This method recognizes that students have different learning styles, abilities, and backgrounds, and aims to create a learning environment that accommodates these differences.

**Evaluation System –**

**Internal Assessment (60 Marks)-** Formative Continuous Assessment [Continuous Assessment; Note Book (30 Marks), Viva Voce (20 Marks), Attendance (10 Marks)]

**End-Semester Exam (40 Marks)-** Summative Assessment.

**4. Course Content:**

**Course Name: Data Mining Lab**

**Course Code: DS-691**

**Hours per Week: 0L: 1T: 2P**

**Credits: 2**

<b>Experiment Number</b>	<b>Title</b>
1	Pre-process a given dataset based on the following: a. Attribute Selection b. Handling Missing Values
2	Pre-process a given dataset based on the following: a. Discretization b. Eliminating Outliers
3	Build a Decision Tree using ID3 algorithm

4	Demonstrate classification process on a given dataset using Naïve Bayesian Classifier.
5	Demonstrate classification process on a given dataset using Nearest neighbor Classifier.
6	Implement Linear Regression and Logistic Regression.
7	Implement SVM on a given dataset.
8	Cluster the given dataset by using the k-Means algorithm and visualize the cluster mean values and standard deviation of dataset attributes.
9	Cluster the given dataset using a hierarchical clustering algorithm.
10	Cluster the given dataset using the DBSCAN algorithm.
11	Detect anomalies using any clustering algorithm.
12	Implement density based outlier detection.
13	Implement association rule mining techniques such as Apriori algorithm

### 5. Books:

1. "Data Mining: Concepts and Techniques", 3rd Edition

*Author:* Jiawei Han, Micheline Kamber,

*Publisher:* Elsevier India Private Limited

2. "Data Mining: Introductory and Advanced Topics",

*Author:* Margaret H. Dunham,

*Publisher:* Pearson Education, 2012.

### 6. Course Outcomes (CO):

**After going through this course the Students will be able to:**

Course Outcomes	Details	Action Verb	Knowledge Level
DS-691.CO1	Understand and apply data preprocessing techniques, including attribute selection, handling missing values, discretization, and outlier elimination, to prepare datasets for analysis.	Apply, Understand	Level 2- Understand, Level 3- Apply

<b>DS-691.CO2</b>	Construct and evaluate classification models using Decision Tree (ID3), Naïve Bayes, and Nearest Neighbor algorithms, and interpret their performance on real datasets.	Apply, Analyze	Level 3- Level 4- Apply, Level 5- Analyze
<b>DS-691.CO3</b>	Implement regression models, including Linear and Logistic Regression, and interpret the results for predictive analytics tasks.	Apply, Analyze	Level 3- Apply, Level 4- Analyze
<b>DS-691.CO4</b>	Develop and apply Support Vector Machine (SVM) models for classification problems and evaluate their performance.	Apply	Level 3- Apply
<b>DS-691.CO5</b>	Perform clustering using k-Means, hierarchical, and DBSCAN algorithms, analyze cluster structures, visualize results, and calculate statistical measures like mean and standard deviation.	Apply, Analyze	Level 3- Apply Level 4- Analyze
<b>DS-691.CO6</b>	Apply unsupervised learning techniques for anomaly detection and association rule mining (Apriori algorithm), and interpret the discovered patterns for data-driven decision making.	Apply, Analyze, Evaluate	Level 3- Apply Level 4- Analyze, Level 5- Evaluate

## 7. Mapping of course outcomes to Experiments

Experiment No.	Experiment Title	CO1	CO2	CO3	CO4	CO5	CO6
1	Pre-process: Attribute Selection & Handling Missing Values	3	1	1	1	1	1
2	Pre-process: Discretization & Eliminating Outliers	3	1	1	1	1	1
3	Build Decision Tree using ID3	1	3	1	1	1	1
4	Classification using Naïve Bayesian Classifier	1	3	1	1	1	1
5	Classification using Nearest Neighbor Classifier	1	3	1	1	1	1

6	Implement Linear & Logistic Regression	1	1	3	1	1	1
7	Implement SVM	1	1	1	3	1	1
8	Cluster using k-Means; visualize mean & SD	1	1	1	1	3	1
9	Cluster using Hierarchical algorithm	1	1	1	1	3	1
10	Cluster using DBSCAN	1	1	1	1	3	1
11	Anomaly detection using clustering	1	1	1	1	2	3
12	Density-based outlier detection	1	1	1	1	2	3
13	Association rule mining (Apriori)	1	1	1	1	1	3

### 8. Mapping of the Course outcomes to Program Outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
<b>CO1</b>	3	3	2	2	1	1	1	1	1	2	2	2
<b>CO2</b>	3	3	3	3	2	1	1	1	1	3	2	2
<b>CO3</b>	3	3	3	3	2	1	1	1	1	3	2	2
<b>CO4</b>	3	3	3	3	2	1	1	1	1	2	2	2
<b>CO5</b>	3	3	3	2	3	1	1	2	1	3	2	3
<b>CO6</b>	3	3	3	3	2	1	1	2	1	3	2	3

### 9. Mapping to Program Specific Outcome (PSO)

Course Outcomes (COs)	PSO1	PSO2	PSO3	PSO4
<b>CO1</b>	<b>3</b>	2	2	1
<b>CO2</b>	<b>3</b>	<b>3</b>	2	1
<b>CO3</b>	<b>3</b>	<b>3</b>	2	1
<b>CO4</b>	<b>3</b>	<b>3</b>	2	1
<b>CO5</b>	<b>3</b>	2	<b>3</b>	1
<b>CO6</b>	<b>3</b>	<b>3</b>	<b>3</b>	1

Reference: NIT Trichy

\*\*\* End of the Syllabus \*\*\*

