Progressive Education Society's

Modern College of Engineering, Shivajinagar,Pune-05.

Department of Artificial Intelligence and Data Science

Curriculum Booklet

2019 Pattern

Semester: I

Class: BE (Artificial Intelligence and Data Science)

Vision of the Institute

- Creation of a collaborative academic environment to foster professional excellence and ethical values.

Mission of the Institute

- To develop outstanding professionals with high ethical standards capable of creating and managing global enterprises.
- To foster innovation and research by providing a stimulating learning environment.
- To ensure equitable development of students of all ability levels and backgrounds.
- To be responsive to changes in technology, socio-economic and environmental conditions.
- To foster and maintain mutually beneficial partnerships with alumni and industry.

Objectives of the Institute

- To develop infrastructure appropriate for delivering quality education
- To develop the overall personality of students who will be innovators and future leaders capable of prospering in their work environment.
- To inculcate ethical standards and make students aware of their social responsibilities.
- Promote close interaction among industry, faculty and students to enrich the learning process and enhance career opportunities.
- Encourage faculty in continuous professional growth through quality enhancement programs and research and development activities.
- Foster a healthy work environment which allows for freedom of expression and protection of the rights of all stakeholders through open channels of communication

Vision of the Department

To create a collaborative academic environment in the field of Artificial Intelligence and Data Science by imparting required skill sets and interactive industry interface for students and inculcate into them social and ethical values

Mission of the Department

M1: To nurture students with latest technologies in the field of Artificial Intelligence and Data Science.

M2: To build industry academia interface to update the recent trends in field of Artificial Intelligence and Data Science.

M3: To prepare students with professional approach, strong ethical values and research spirit along with leadership skills.

M4: To mentor students in the field of Artificial Intelligence and Data Science research to serve the needs of the society.

Program Educational Objectives (PEOs)

Program Educational Objectives (PEO's)	
A graduate of the Artificial Intelligence and Data Science Program will able to-	
PEO1	Work in the area of Artificial Intelligence and Data Science to design ability of a computer system.
PEO2	Apply analytical skills ,decision making skills, leadership skills and critical thinking skills to solve multidisciplinary problems for the betterment of the society
PEO3	Demonstrate professionalism as a means of lifelong learning in the area of Artificial Intelligence and Data Science with emerging tools and technologies like IoT , Big Data, Cloud Services , Artificial Neural Network.

Program Specific Outcomes (PSOs)

Program Specific Outcomes (PSO's)	
A graduate of the Artificial Intelligence and Data Science Program will demonstrate	
PSO1	**Professional Skills-**The ability to understand, analyze and develop computer programs in the areas of Artificial Intelligence and Data Science for efficient design of computer based systems.
PSO2	**Problem-Solving Skills-** The ability to apply the knowledge of Artificial Intelligence and Data Science to cater the need of industry, academia and society.
PSO3	**Successful Career and Entrepreneurship-** The ability to employ modern computer languages, environments and platforms to become an entrepreneur and to have a zest for higher studies.

Program Outcomes (POs)

1. Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and engineering specialization to the solution of complex engineering problems.

2. Problem analysis: Identify, formulate, research literature, and analyze complex engineering problems to arrive at substantiated conclusions using first principles of mathematics, natural, and engineering sciences.

3. Design/development of solutions: Design solutions for complex engineering problems and design system components, processes to meet the specifications with consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. Conduct investigations of complex problems: Use research based knowledge including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

6. The Engineer and Society: Apply reasoning informed by contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant and cultural issues and the consequent responsibilities relevant to professional engineering practice.

7. Environment and Sustainability: Understand the impact of professional engineering solutions in societal and professional engineering solutions in societal and environmental contexts and demonstrate knowledge of and need for sustainable development.

8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of engineering practice.

9. Individual and Team Work: Function effectively as an individual, and as a member or leader in diverse teams and in multidisciplinary settings.

10. Communication Skills : Communicate effectively on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation make effective presentations and give and receive clear instructions.

11. Project Management and Finance: Demonstrate knowledge and understanding of engineering and management principles and apply these to one’s own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

12. Life-long Learning: Recognize the need for and have the preparation and ability to engage in independent and preparation and ability to engage in independent and lifelong learning in the broadest context of technological change.

Course Structure

Faculty of Science and Technology

Savitribai Phule Pune University

Savitribai Phule Pune University
Fourth Year of Artificial Intelligence and Data Science (2020 Course)
(With effect from Academic Year 2023-24)

Semester-VII

Course Code	Course Name	Teaching Scheme (Hours/Week)			Examination Scheme and Marks						Credit Scheme			
		Lecture	Practical	Tutorial	In-Sem	End-Sem	Term work	Practical	Oral/ Presentation	Total	Lecture	Practical	Tutorial	Total
417521	Machine Learning	03	-	-	30	70	-	-	-	100	03	-	-	03
417522	Data Modeling & Visualization	03	-	-	30	70	-	-	-	100	03	-	-	03
417523	Elective III	03	-	-	30	70	-	-	-	100	03	-	-	03
417524	Elective IV	03	-	-	30	70	-	-	-	100	03	-	-	03
417525	Computer Laboratory I	-	04	-	-	-	50	25	-	75	-	02	-	02
417526	Computer Laboratory II	-	04	-	-	-	50	25	-	75	-	02	-	02
417527	Project Stage I	-	04	-	-	-	50	-	50	100	-	02	-	02
417528	MOOC	-	-	02	-	-	50	-	-	50	-	-	02	02
	Total	12	12	02	120	280	200	50	50	700	12	06	02	20
417529	Audit Course 7													
										Total Credits	**12**	**06**	**02**	**20**

Elective III:
417523(A) Quantum Artificial Intelligence
417523(B) Industrial Internet of Things
417523(C) Enterprise Architecture and Components
417523(D) Bioinformatics

Elective IV:
417524(A) GPU Programming and Architecture
417524(B) Information Retrieval
417524(C) UI/UX Design
417524(D) Optimization Algorithms

Computer Laboratory I:
It is based on two compulsory subjects:
- Machine Learning
- Data Modeling & Visualization

Computer Laboratory II:
It is based on two Elective subjects:
- Elective III
- Elective IV

Audit Course 7:
AC7 – I Block Chain
AC7 – II Entrepreneurship Development
AC7 – III Botnet of Things
AC7 – IV Foreign Language
AC7 – V MOOC-Learn New Skills

Subject 1: Machine Learning

Weekly Work Loads(in Hrs)	Lecture	Tutorial	Practical
	03	-	-

Online/ In-sem	Theory	Practical	Oral	Term-work	Total Marks	Credit
30	70	0	0	0	100	03

Course Objectives:

- Explain the learning paradigms, and models of machine learning
- Apply different regression techniques for making predictions in different applications
- Apply the classification algorithms to classify the data with appropriate labels
- Apply the clustering algorithms to divide the unlabeled data into the similar groups
- Introduce and integrate models in the form of advanced ensembles.
- Explain reinforcement learning and its algorithms

Course Outcomes:

On completion of the course, learner will be able to :

- Describe and compare different models of machine learning
- Design ML models to make predictions by using linear, non-linear and logistic regression techniques
- Implement classification models for two class problems and multiclass problems
- Implement clustering models for unlabeled data
- Integrate multiple machine learning algorithms in the form of ensemble learning
- Apply reinforcement learning and its algorithms for different applications

SYLLABUS

Unit	Course Content	Hours
I	**Introduction to machine Learning**	06
	Introduction: What is Machine Learning, Definitions and Real-life applications, Comparison of Machine learning with traditional programming, ML vs AI vs Data Science. Learning Paradigms: Learning Tasks- Descriptive and Predictive Tasks, Supervised, Unsupervised,Semi-supervised and Reinforcement Learnings. Models of Machine learning: Geometric model, Probabilistic Models, Logical Models, Grouping and grading models, Parametric and non-parametric models. Feature Transformation: Dimensionality reduction techniques- PCA and LDA	
II	**Regression**	06
	Introduction- Regression, Need of Regression, Difference between Regression and Correlation, Types of Regression: Univariate vs. Multivariate, Linear vs. Nonlinear, Simple Linear vs. Multiple Linear, Bias-Variance tradeoff, Overfitting and Underfitting.Regression Techniques - Polynomial Regression, Stepwise Regression, Decision Tree Regression, Random Forest Regression, Support Vector Regression, Ridge Regression, Lasso Regression, ElasticNet Regression, Bayesian Linear Regression.Evaluation Metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE),R-squared ,Adjusted R-squared.	
III	**Classification**	06
	Introduction: Need of Classification, Types of Classification (Binary and Multiclass), Binary-vs-Multiclass Classification, Balanced and Imbalanced Classification Problems. Binary Classification: Linear Classification model, Performance Evaluation- Confusion Matrix, Accuracy, Precision, Recall, F measures. Multiclass Classification: One-vs-One and One-vs-All classification techniques, Performance Evaluation- Confusion Matrix, Per Class Precision, Per Class Recall Classification Algorithms: K Nearest Neighbor, Linear Support Vector Machines (SVM) – Introduction, Soft Margin SVM, Kernel functions– Radial Basis Kernel, Gaussian, Polynomial, Sigmoid.	
IV	**Clustering**	06
	Introduction: What is clustering, Need of Clustering, Types of Clustering Hierarchical clustering algorithms /connectivity-based clustering): Agglomerative Hierarchical Clustering (AHC) algorithm, Divisive Hierarchical Clustering (DHC) algorithm. Centroid-based clustering algorithms / Partitioning clustering algorithms: K-Means clustering algorithm, Advantages and disadvantages of K-Means	

	clustering algorithm, Elbow method, The Silhouette method, K-Medoids, K-Prototype. Density-based clustering algorithms: DBSCAN algorithm, how it works, Advantages and disadvantages of DBSCAN. Distribution-based clustering algorithms: Gaussian mixture model. Application of Clustering Technique: Market Segmentation, Statistical data analysis, Social network analysis, Image segmentation, Anomaly detection.	
V	**Ensemble Learning**	**06**
	Ensemble Learning: Introduction to Ensemble Learning, Need of Ensemble Learning, Homogeneous and Heterogeneous ensemble methods, Advantages and Limitations of Ensemble methods, Applications of Ensemble Learning. Basic Ensemble Learning Techniques: Voting Ensemble, Types of Voting: Max Voting, Averaging, Weighted Average. Advanced Ensemble Learning Techniques: Bagging: Bootstrapping, Aggregation. Boosting: Adaptive Boosting (AdaBoost), Gradient Boosting, XGBoost . Stacking: Variance Reduction, Blending, Random Forest Ensemble, Advantages of Random Forest.	
VI	**Reinforcement Learning**	**06**
	Reinforcement learning: What is Reinforcement Learning? Need for Reinforcement Learning, Supervised vs Unsupervised vs Reinforcement Learning, Types of Reinforcement, Elements of Reinforcement Learning, Real time applications of Reinforcement learning. Markov's Decision Process: Markov property, Markov chain/process, Markov reward process (MRP), Markov decision process (MDP), Return, Policy, Value functions, Bellman equation. Q Learning: Introduction of Q-Learning, Important terms in Q learning, Q table, Q functions, Q learning algorithm.	

Teaching Plan

Sr. No.	Unit	Topics to be covered	Book Referred	Total Lecture Planned
1	I	Introduction: What is Machine Learning, Definitions and Real-life applications, Comparison of Machine learning with traditional programming, ML vs AI vs Data Science. Learning Paradigms: Learning Tasks- Descriptive and Predictive Tasks, Supervised, Unsupervised, Semi-supervised and Reinforcement Learnings. Models of Machine learning: Geometric model, Probabilistic Models, Logical Models, Grouping and grading models, Parametric and non-parametric models. Feature Transformation: Dimensionality reduction techniques- PCA and LDA	1.1,1.2,2.3,2.4	06
2	II	Introduction- Regression, Need of Regression, Difference between Regression and Correlation, Types of Regression: Univariate vs. Multivariate, Linear vs. Nonlinear, Simple Linear vs. Multiple Linear, Bias-Variance tradeoff, Overfitting and Underfitting. Regression Techniques - Polynomial Regression, Stepwise Regression, Decision Tree Regression, Random Forest Regression, Support Vector Regression, Ridge Regression, Lasso Regression, ElasticNet Regression, Bayesian Linear Regression. Evaluation Metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE),R-squared ,Adjusted R-squared.	1.1,1.2,2.1,2.2	06
3	III	Introduction: Need of Classification, Types of Classification (Binary and Multiclass), Binary-vs-Multiclass Classification, Balanced and Imbalanced Classification Problems. Binary Classification: Linear Classification model, Performance Evaluation- Confusion Matrix, Accuracy, Precision, Recall, F measures. Multiclass Classification: One-vs-One and One-vs-All classification techniques, Performance Evaluation- Confusion Matrix, Per Class Precision, Per Class Recall Classification Algorithms: K Nearest	1.1,1.2,2.1,2.2, 2.3	06

		Neighbor, Linear Support Vector Machines (SVM) – Introduction, Soft Margin SVM, Kernel functions– Radial Basis Kernel, Gaussian, Polynomial,Sigmoid.		
4	IV	Introduction: What is clustering, Need of Clustering, Types of Clustering. Hierarchical clustering algorithms /connectivity-based clustering): Agglomerative Hierarchical Clustering (AHC) algorithm, Divisive Hierarchical Clustering (DHC) algorithm.Centroid-based clustering algorithms / Partitioning clustering algorithms: K-Means clustering algorithm, Advantages and disadvantages of K-Means clustering algorithm, Elbow method, The Silhouette method, K-Medoids, K-Prototype.Density-based clustering algorithms: DBSCAN algorithm, how it works, Advantages and disadvantages of DBSCAN.Distribution-based clustering algorithms: Gaussian mixture model. Application of Clustering Technique: Market Segmentation, Statistical data analysis, Social network analysis, Image segmentation, Anomaly detection.	1.1,1.2,2.3,2.4	06
5	V	Ensemble Learning: Introduction to Ensemble Learning, Need of Ensemble Learning, Homogeneous and Heterogeneous ensemble methods, Advantages and Limitations of Ensemble methods, Applications of Ensemble Learning. Basic Ensemble Learning Techniques: Voting Ensemble, Types of Voting: Max Voting, Averaging, Weighted Average. Advanced Ensemble Learning Techniques: Bagging: Bootstrapping, Aggregation. Boosting: Adaptive Boosting (AdaBoost), Gradient Boosting, XGBoost . Stacking: Variance Reduction, Blending, Random Forest Ensemble, Advantages of Random Forest.	1.1,1.2,2.1,2.2	06
6	VI	Reinforcement learning: What is Reinforcement Learning? Need for Reinforcement Learning, Supervised vs Unsupervised vs Reinforcement Learning, Types of Reinforcement, Elements of Reinforcement Learning, Real time applications of Reinforcement learning. Markov's Decision Process: Markov property, Markov chain/process, Markov reward process (MRP), Markov	1.1,1.2,2.1,2.2, 2.3	06

		decision process (MDP), Return, Policy, Value functions, Bellman equation Q Learning: Introduction of Q-Learning, Important terms in Q learning, Q table, Q functions, Q learning algorithm.		

1. Text Books:

1. Ethem Alpaydin, "Introduction to Machine Learning", Publisher: The MIT Press,2014
2. Peter Flach: "Machine Learning: The Art and Science of Algorithms that Make Sense of Data",Cambridge University Press, Edition 2012

2. Reference Books:

1. Ian H Witten, Eibe Frank, Mark A Hall, "Data Mining, Practical Machine Learning Tools andTechniques", Elsevier, 3rd Edition
2. Jiawei Han, Micheline Kamber, and Jian Pie, "Data Mining: Concepts and Techniques", ElsevierPublishers Third Edition, ISBN: 9780123814791, 9780123814807
3. Shalev-Shwartz, Shai, and Shai Ben-David, "Understanding machine learning: From theory toalgorithms", Cambridge university press, 2014
4. McKinney, "Python for Data Analysis O' Reilly media, ISBN : 978-1-449- 31979-3

Reference Web Links/ Research Paper/ Referred Book other than Mention in Syllabus:

- https://timeseriesreasoning.com/
- Reinforcement Learning: https://www.cs.toronto.edu/~urtasun/courses/CSC411_Fall16/19_rl.pdf
- A brief introduction to machine learning for Engineers: https://arxiv.org/pdf/1709.02840.pdf
- Introductory Machine Learning Nodes: http://lcsl.mit.edu/courses/ml/1718/MLNotes.pdf

MOOC Courses:

- Introduction to Machine Learning(IIT kharagpur) : https://nptel.ac.in/courses/106105152
- Introduction to Machine Learning (IIT Madras): https://onlinecourses.nptel.ac.in/noc22_cs29/prevew
- Machine Learning A-Z™: AI, Python & R + ChatGPT Bonus [2023]

https://www.udemy.com/course/machinelearning/

- Machine Learning and Deep Learning A-Z: Hands-On Python https://www.udemy.com/course/machine-learning-and-deep-learning-a-z-hands-on-python/

Unit No: I

Name of the Unit: Introduction to machine Learning

Lecture No.	Details of the Topic to be covered	References
1	**Introduction:** What is Machine Learning, Definitions and Real-life applications, Comparison of Machine learning with traditional programming, ML vs AI vs Data Science.	1.1,1.2,2.4
2	**Learning Paradigms:** Learning Tasks- Descriptive and Predictive Tasks, Supervised, Unsupervised	1.1,1.2
3	Semi-supervised and Reinforcement Learnings. **Models of Machine learning:** Geometric model	1.1,1.2,2.3
4	Probabilistic Models, Logical Models	1.1,1.2,2.3
5	Grouping and grading models, Parametric and non-parametric models.	1.1,1.2,2.3
6	**Feature Transformation**: Dimensionality reduction techniques- PCA and LDA	1.1,1.2,2.4

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	What is the need of Dimensionality Reduction in machine learning? How PCA is used for dimensionality reduction?
Q. 2	Explain role of machine learning the following common un-supervised learning problems: a) Object segmentation b) Similarity detection
Q. 3	Compare Supervised, Unsupervised, Semi-Supervised Learning with examples.
Q. 4	What is reinforcement learning? Elaborate the concept with the help of suitable example & Flow diagram.
Q. 5	What is Machine learning? What is the need of it? Explain four examples of

	machine learning in detail.

Unit No.-II

Name of the Unit: Regression

Lecture No.	Details of the Topic to be covered	References
1	**Introduction-** Regression, Need of Regression, Difference between Regression and Correlation, Types of Regression: Univariate vs. Multivariate, Linear vs. Nonlinear	1.1,1.2,2.1,2.2
2	Simple Linear vs. Multiple Linear, Bias-Variance tradeoff, Overfitting and Underfitting. **Regression Techniques -** Polynomial Regression.	1.1,1.2,
3	Stepwise Regression Decision Tree Regression, Random Forest Regression, Support Vector Regression	1.1,2.1,2.2
4	Ridge Regression, Lasso Regression, ElasticNet Regression, Bayesian Linear Regression.	1.1,1.2,
5	**Evaluation Metrics:** Mean Squared Error (MSE), Mean Absolute Error (MAE),	1.1,1.2,2.1,
6	Root Mean Squared Error (RMSE),R-squared ,Adjusted R-squared.	1.1,1.2,2.2

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	What is a multivariate linear regression?
Q. 2	Discuss the following Regression Models: Ridge, Lasso and ElasticNet. Give a brief comparison of these models.
Q. 3	Explain Bias-Variance tradeoff in regression.
Q. 4	Write a note on : i.Mean Absolute Error (MAE), ii.Root Mean Squared Error (RMSE)
Q. 5	Show that how the values of 0 θ and 1 θ are estimated in a linear regression model of $Y_i' = \theta_0 + \theta 1 . X_i$

Unit No.-III

Name of the Unit: Classification

Lecture No.	Details of the Topic to be covered	References
1	**Introduction:** Need of Classification, Types of Classification (Binary and Multiclass), Binary-vs-Multiclass Classification, Balanced and Imbalanced Classification Problems.	1.1,1.2,2.1
2	**Binary Classification:** Linear Classification model, Performance Evaluation- Confusion Matrix, Accuracy,	1.1,1.2,2.2
3	Precision, Recall, F measures. **Multiclass Classification:** One-vs-One and One-vs-All classification techniques	1.1,1.2,2.1
4	Performance Evaluation- Confusion Matrix, Per Class Precision, Per Class Recall.**Classification Algorithms:** K Nearest Neighbor	1.1,1.2,2.3
5	Linear Support Vector Machines (SVM) – Introduction, Soft Margin SVM,	1.1,1.2,2.3
6	Kernel functions– Radial Basis Kernel, Gaussian, Polynomial, Sigmoid.	1.1,1.2,2.1

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Explain need of classification. Differentiate between Binary-vs-Multiclass Classification
Q. 2	Explain the Multiclass classification techniques in detail
Q. 3	Write a note on SVM
Q. 4	State different kernel functions. Elaborate any two in it.
Q. 5	Write different performance evaluation matrices.

Unit No.-IV

Name of the Unit: Clustering

Lecture No.	Details of the Topic to be covered	References
1	**Introduction:** What is clustering, Need of Clustering, Types of Clustering. Hierarchical clustering algorithms /connectivity-based clustering introduction	1.1, 2.3,2.4
2	Agglomerative Hierarchical Clustering (AHC) algorithm, Divisive Hierarchical Clustering (DHC) algorithm. **Centroid-based clustering algorithms / Partitioning clustering algorithms:** K-Means clustering algorithm,	1.1,1.2,2.3
3	Advantages and disadvantages of K-Means clustering algorithm, Elbow method, The Silhouette method, K-Medoids, K-Prototype.	1.1,1.2,2.3
4	**Density-based clustering algorithms:** DBSCAN algorithm, how it works, Advantages and disadvantages of DBSCAN.	1.1,1.2,2.4
5	**Distribution-based clustering algorithms:** Gaussian mixture model.	1.1,1.2,2.3
6	**Application of Clustering Technique:** Market Segmentation, Statistical data analysis, Social network analysis, Image segmentation, Anomaly detection	1.1,1.2,2.4

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Explain need of clustering. Explain different types in detail
Q. 2	Differentiate between Agglomerative Hierarchical Clustering (AHC) algorithm, Divisive Hierarchical Clustering (DHC) algorithm.
Q. 3	Elaborate on Elbow method in KMeans Clustering. write Advantages and disadvantages of K-Means clustering algorithm
Q. 4	List Application of Clustering Technique and explain one in detail
Q. 5	Write note on : Gaussian mixture model.

Unit No.-V

Name of the Unit: Ensemble Learning

Lecture No.	Details of the Topic to be covered	References
1	**Ensemble Learning:** Introduction to Ensemble Learning, Need of Ensemble Learning, Homogeneous and Heterogeneous ensemble methods	1.1,2.1,2.2
2	Advantages and Limitations of Ensemble methods, Applications of Ensemble Learning.	1.1,1.2
3	**Basic Ensemble Learning Techniques:** Voting Ensemble, Types of Voting: Max Voting, Averaging, Weighted Average.	1.1,1.2
4	**Advanced Ensemble Learning Techniques:** Bagging: Bootstrapping, Aggregation.	1.1, 2.2
5	Boosting: Adaptive Boosting (AdaBoost), Gradient Boosting, XGBoost .	1.1,1.2
6	Stacking: Variance Reduction, Blending, Random Forest Ensemble, Advantages of Random Forest.	1.1,1.2

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	What is Ensemble Learning?Explain Homogeneous and Heterogeneous ensemble methods in detail.
Q. 2	Elaborate on Voting Ensemble along with types of voting
Q. 3	Explain note on advanced ensemble learning technique.
Q. 4	Write a note on boosting techniques
Q. 5	Write a note on : a. Bootstrapping b. Aggregation.

Unit No.-VI

Name of the Unit: Reinforcement Learning

Lecture No.	Details of the Topic to be covered	References
1	**Reinforcement learning:** What is Reinforcement Learning? Need for Reinforcement Learning, Supervised vs Unsupervised vs Reinforcement Learning	1.1,1.2,2.1,2.3
2	Types of Reinforcement, Elements of Reinforcement Learning, Real time applications of Reinforcement learning	1.1,1.2,2.1,2.3
3	**Markov's Decision Process:** Markov property, Markov chain/process, Markov reward process (MRP),	1.1, 2.1,2.2
4	Markov decision process (MDP), Return, Policy, Value functions, Bellman equation	1.1,1.2,
5	**Q Learning:** Introduction of Q-Learning, Important terms in Q learning	1.1,2.2,2.3
6	Q table, Q functions, Q learning algorithm.	1.1,1.2,2.3

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Explain **Markov's Decision Process(MDP)** in detail
Q. 2	Write a note on Q Learning process
Q. 3	What is Reinforcement Learning? What is need for Reinforcement Learning
Q. 4	Supervised vs Unsupervised vs Reinforcement Learning
Q. 5	Elaborate on real time applications of Reinforcement learning

Subject 2: Data Modeling and Visualization

Weekly Work Loads(in Hrs)	Lecture	Tutorial	Practical
	03	-	-

Online/ In-sem	Theory	Practical	Oral	Term-work	Total Marks	Credit
30	70	0	0	0	100	03

Course Objectives:

- Creating an emerging data model for the data to be stored in a database
- Conceptualized representation of Data objects
- Create associations between different data objects, and the rules
- Organize data description, data semantics, and consistency constraints of data
- Identifying data trends
- Incorporate data visualization tools and reap transformative benefits in their critical areas ofoperations

Course Outcomes:

After completion of the course, learners should be able to-

- Summarize data analysis and visualization in the field of exploratory data science
- Analyze the characteristics and requirements of data and select an appropriate data model
- Describe to load, clean, transform, merge and reshape data
- Design a probabilistic data modeling, interpretation, and analysis
- Evaluate time series data
- Integrate real world data analysis problems

Syllabus

Unit	Course Content	Hours
I	**Introduction to Data Modeling**	06
	Basic probability: Discrete and continuous random variables, independence, covariance, central limit theorem, Chebyshev inequality, diverse continuous and discrete distributions. Statistics, Parameter Estimation, and Fitting a Distribution: Descriptive statistics, graphical statistics, method of moments, maximum likelihood estimation Data Modeling Concepts • Understand and model subtypes and supertypes • Understand and model hierarchical data • Understand and model recursive relationships • Understand and model historical data	
II	**Testing and Data Modeling**	06
	Random Numbers and Simulation: Sampling of continuous distributions, Monte Carlo methods Hypothesis Testing: Type I and II errors, rejection regions; Z-test, T-test, F-test, Chi-Square test, Bayesian test Stochastic Processes and Data Modeling: Markov process, Hidden Markov Models, Poisson Process, Gaussian Processes, Auto-Regressive and Moving average processes, Bayesian Network, Regression, Queuing systems	
III	**Basics of Data Visualization**	06
	Computational Statistics and Data Visualization, Types of Data Visualization, Presentation and Exploratory Graphics, Graphics and Computing, Statistical Historiography, Scientific Design Choices in Data Visualization, Higher-dimensional Displays and Special Structures, Static Graphics: Complete Plots, Customization, Extensibility, Other Issues: 3-D Plots, Speed, Output Formats, Data Handling	
IV	**Data Visualization and Data Wrangling**	06
	Data Wrangling: Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting. Data Visualization matplotlib: Basics of matplotlib, plotting with pandas and seaborn, other python visualization tools. Data Visualization Through Their Graph Representations: Data and Graphs Graph Layout Techniques, Force-directed Techniques Multidimensional Scaling, The Pulling Under Constraints Model, Bipartite Graphs	
V	**Data Aggregation and Analysis**	06
	Data Aggregation and Group operations: Group by Mechanics, Data aggregation, General split- apply-combine, Pivot tables and cross tabulation 67 Time Series Data Analysis: Date and Time Data Types and Tools, Time series Basics, date Ranges, Frequencies and Shifting, Time Zone Handling, Periods and Periods Arithmetic, Resampling and Frequency conversion,	

	Moving Window Functions.	
VI	**Data Analysis of Visualization and Modelling**	**06**
	Reconstruction, Visualization and Analysis of Medical Images Introduction: - PET Images, Ultrasound Images, Magnetic Resonance Images, Conclusion and Discussion, Case Study: ER/Studio, Erwin data modeler, DbSchema Pro, Archi, SQL Database Modeler, LucidChart, Pgmodeler	

Teaching Plan

Sr. No.	Unit	Topics to be covered	Book Referred	Total Lecture Planned
1	I	**Basic probability:** Discrete and continuous random variables, independence, covariance, central limit theorem, Chebyshev inequality, diverse continuous and discrete distributions. **Statistics,** Parameter Estimation, and Fitting a Distribution: Descriptive statistics, graphical statistics, method of moments, maximum likelihood estimation **Data Modeling Concepts** • Understand and model subtypes and supertypes • Understand and model hierarchical data • Understand and model recursive relationships • Understand and model historical data	2.5,2.6	7
2	II	**Random Numbers and Simulation:** Sampling of continuous distributions, Monte Carlo methods **Hypothesis Testing:** Type I and II errors, rejection regions; Z-test, T-test, F-test, Chi-Square test, Bayesian test **Stochastic Processes and Data Modeling:** Markov process, Hidden Markov Models, Poisson Process, Gaussian Processes, Auto-Regressive and Moving average processes, Bayesian Network, Regression, Queuing systems	1.6,2.5,2.6	6
3	III	**Computational Statistics and Data Visualization,** Types of Data Visualization, Presentation and Exploratory Graphics, Graphics and Computing, Statistical Historiography, Scientific **Design Choices in Data Visualization,** Higher-dimensional Displays and Special Structures, **Static Graphics:** Complete Plots, Customization, Extensibility, **Other Issues:** 3-D Plots, Speed, Output Formats, Data Handling	2.5,2.6	8
4	IV	**Data Wrangling:** Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting. Data Visualization matplotlib: Basics of matplotlib, plotting with pandas and seaborn, other python visualization tools **Data Visualization Through Their Graph**	1.4,1.5	7

		Representations: Data and Graphs Graph Layout Techniques, Force-directed Techniques Multidimensional Scaling, The Pulling Under Constraints Model, Bipartite Graphs		
5	V	**Data Aggregation and Group operations:** Group by Mechanics, Data aggregation, General split- apply-combine, Pivot tables and cross tabulation 67 Time Series **Data Analysis:** Date and Time Data Types and Tools, Time series Basics, date Ranges, Frequencies and Shifting, Time Zone Handling, Periods and Periods Arithmetic, Resampling and Frequency conversion, Moving Window Functions.	1.5,1.6	8
6	VI	Reconstruction, Visualization and Analysis of Medical Images Introduction: - PET Images, Ultrasound Images, Magnetic Resonance Images, Conclusion and Discussion, Case Study: ER/Studio, Erwin data modeler, Db Schema Pro, Archi, SQL Database Modeler, LucidChart, Pgmodeler	1.4,1.5, 1.6	6

1. Text Books:

1. Chun-houh Chen, Wolfgang Härdle, Antony Unwin, “Handbook of Data Visualization”, Springer

2. Ben Fry, “Visualizing Data”, O’Reilly Media

3. Clous O.Wilke, “Fundamentals of Data Visualization - A Primer on Making Informative and Compelling Figures”, O’Reilly Media, Inc.

4. Kieran Healy, “Data Visualization - A Practical Introduction”

5. McKinney W., “Python for Data Analysis: Data Wrangling with Pandas, NumPy and IPython”. 2nd edition, O’Reilly Media

6. Gelman, Andrew, and Jennifer Hill, “Data Analysis Using Regression and Multilevel /Hierarchical Models”. 1st edition, Cambridge, UK: Cambridge University Press, 2006, ISBN: 9780521867061

7. Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin, “Bayesian Data

Analysis", 2 nd edition, New York, NY: Chapman & Hall, 2003, ISBN: 9781584883883

2. Reference Books:

1. David Dietrich, Barry Hiller, "Data Science and Big Data Analytics", EMC education services, Wiley publication, 2012, ISBN: 0-07-120413-X

2. Trent Hauk, "Scikit-learn Cookbook", Packt Publishing, ISBN: 9781787286382

3. Chirag Shah, "A Hands-On Introduction to Data Science", Cambridge University Press, 2020, ISBN: 978-1-108-47244-9

4. S. C. Gupta, V. K. Kapoor, "Fundamentals of Mathematics Statistics (A Modern Approach) "Sultan Chand & Sons Educational Publishers, Tenth revised edition, ISBN: 81-7014-791-3

5. Medhi , "Statistical Methods: An Introductory Text", 2 nd Edition, New Age International Ltd., ISBN:8122419577

Reference Web Links/ Research Paper/ Referred Book other than Mention in Syllabus:

e-Resources :

1. An Introduction to Statistical Learning by Gareth James

https://www.ime.unicamp.br/~dias/Intoduction%20to%20Statistical%20Learning.pdf

2. Python Data Science Handbook by Jake VanderPlas

https://tanthiamhuat.files.wordpress.com/2018/04/pythondatasciencehandbook.pdf

3. Elements of Statistical Learning: data mining, inference, and prediction, 2nd Edition.

(su.domains)

MOOC Courses :

1. https://www.youtube.com/watch?v=WSNqcYqByFk

2. https://www.youtube.com/watch?v=eFByJkA3ti4

3. Computer Science and Engineering - NOC:Data Science for Engineers

4. Computer Science and Engineering - NOC:Python for Data Science

5. Introduction to Data Analytics: https://nptel.ac.in/courses/110106072

Unit No.-I

Name of the Unit :Introduction to Data Modeling

Lecture No.	Details of the Topic to be covered	References
1	**Basic probability:** Discrete and continuous random variables, independence, covariance, central limit theorem,	2.5,2.6
2	Chebyshev inequality, diverse continuous and discrete distributions.	
3	**Statistics,** Parameter Estimation, and Fitting a Distribution: Descriptive statistics, graphical statistics,	
4	method of moments, maximum likelihood estimation	
5	**Data Modeling Concepts** Understand and model subtypes and supertypes , Understand and model hierarchical data	
6	Understand and model recursive relationships	
7	Understand and model historical data	

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Define descriptive statistics. Provide examples of common descriptive statistics measures and explain their significance in data analysis.
Q. 2	Formulate the Chebyshev Inequality and explain its use in probability theory.
Q. 3	State the Central Limit Theorem. Discuss its significance in probability theory and practical applications.
Q. 4	What is maximum likelihood estimation (MLE)? How does it differ from the method

	of moments? Provide an example of MLE in action.

Q. 5	Discuss hierarchical data modeling. How is hierarchical data represented in databases or other data storage systems?

Unit No.-II

Name of the Unit : Testing and Data Modeling

Lecture No.	Details of the Topic to be covered	References
1	**Random Numbers and Simulation:** Sampling of continuous distributions, Monte Carlo methods	1.6,2.5,2.6
2	**Hypothesis Testing:** Type I and II errors, rejection regions; Z-test, T-test, F-test, Chi-Square test, Bayesian test	
3	**Stochastic Processes and Data Modeling:** Markov process, Hidden Markov Models, Poisson Process,	
4	Gaussian Processes, Auto-Regressive and Moving average processes,	
5	Bayesian Network, Regression	
6	Queuing systems	

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	What are Monte Carlo methods? Provide an example of a problem that can be solved using Monte Carlo simulation.

Q. 2	Describe the Z-test, T-test, F-test, and Chi-Square test. Provide examples of situations where each test is appropriate.

Q. 3	Define Type I and Type II errors in hypothesis testing. How are these errors related to the concepts of significance level and power?

Q. 4	What are Hidden Markov Models (HMMs)? Explain the structure and applications of HMMs in data modeling.

Q. 5	Explain Gaussian processes. How are Gaussian processes used in regression and machine learning?

Unit No.-III

Name of the Unit: Basics of Data Visualization

Lecture No.	Details of the Topic to be covered	References
1	**Computational Statistics and Data Visualization,** Types of Data Visualization,	2.5,2.6
2	Presentation and Exploratory Graphics, Graphics and Computing,	
3	Statistical Historiography, Scientific	
4	**Design Choices in Data Visualization,** Higher-dimensional Displays and Special Structures,	
5	**Static Graphics:** Complete Plots, Customization, Extensibility,	
6	**Other Issues:** 3-D Plots, Speed, Output Formats, Data Handling	

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	How do different types of data visualization techniques (e.g., scatter plots, histograms, heatmaps) contribute to understanding different aspects of data?

Q. 2	What is computational statistics, and how does it differ from traditional statistics? Provide examples of computational techniques used in statistical analysis.

Q. 3	Discuss the role of programming languages and libraries (e.g., R, Python, matplotlib, ggplot2) in creating advanced data visualizations.

Q. 4	Discuss the extensibility of static graphics and techniques for combining multiple plots into a cohesive visual narrative.

Unit No.-IV

Name of the Unit : Data Visualization and Data Wrangling

Lecture No.	Details of the Topic to be covered	References
1	**Data Wrangling:** Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting.	1.4,1.5
2	Data Visualization matplotlib: Basics of matplotlib,	
3	plotting with pandas and seaborn, other python visualization tools	
4	**Data Visualization Through Their Graph Representations:** Data and Graphs Graph Layout	
5	Techniques, Force-directed Techniques Multidimensional Scaling,	
6	The Pulling Under Constraints Model, Bipartite Graphs	

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Discuss the process of combining and merging datasets in data wrangling. Provide examples of when and how different types of joins (e.g., inner join, outer join) are used.

Q. 2	What is reshaping and pivoting in data wrangling? How can reshaping and pivoting be used to transform data into different formats for analysis?

Q. 3	How can Matplotlib be used to create various types of plots, such as line plots, scatter plots, histograms, and bar plots?

Q. 4	What are bipartite graphs, and how are they used in data visualization? Provide examples of real-world applications of bipartite graphs.

Unit No-V

Name of the Unit : Data Aggregation and Analysis

Lecture No.	Details of the Topic to be covered	References
1	**Data Aggregation and Group operations:** Group by Mechanics	1.5,1.6
2	Data aggregation, General split-apply-combine,	
3	Pivot tables and cross tabulation 67 Time Series	
4	**Data Analysis:** Date and Time Data Types and Tools,	
5	Time series Basics, date Ranges	
6	Frequencies Moving Window Functions and Shifting,	
7	Time Zone Handling, Periods and Periods Arithmetic, Resampling and Frequency conversion,	

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Discuss the concept of "split-apply-combine" in data analysis. How does it work, and what are its advantages?

Q. 2	Describe the use of pivot tables and cross-tabulation in data analysis. How do pivot tables and cross-tabulation help summarize and analyze data?

Q. 3	Explain the basics of time series analysis. What are the key components of a time series?

Q. 4	Discuss the process of resampling and frequency conversion in time series analysis. What are the different resampling methods, and when are they used?

Unit No.-VI

Name of the Unit : Data Analysis of Visualization and Modeling

Lecture No.	Details of the Topic to be covered	References
1	Reconstruction, Visualization and Analysis of Medical Images	1.4,1.5, 1.6
2	Introduction: - PET Images	
3	Ultrasound Images,	
4	Magnetic Resonance Images, Conclusion and Discussion,	
5	Case Study: ER/Studio, Erwin data modeler, Db Schema Pro, Archi, SQL Database Modeler, LucidChart, Pgmodeler	

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Explain the principles behind PET (Positron Emission Tomography), ultrasound, and Magnetic Resonance Imaging (MRI). What are the key differences between these imaging modalities?

Q. 2	Summarize the key challenges and limitations associated with each imaging modality.

Q. 3	Discuss a case study where one or more of these tools were used to design and implement a database schema for a healthcare application.

Q. 4	Provide an overview of ER/Studio, Erwin Data Modeler, DbSchema Pro, Archi, SQL Database Modeler, LucidChart, and Pgmodeler. What are the main features and functionalities of each tool?

Subject 3: Elective III Bioinformatics

Weekly Work Loads(in Hrs)	Lecture	Tutorial	Practical
	03	-	-

Online/ In-sem	Theory	Practical	Oral	Term-work	Total Marks	Credit
30	70	0	0	0	100	03

Course Objectives

- To study fundamental concepts of bioinformatics
- To study knowledge from basic to advanced level
- To refer appropriate, suitable datasets
- To study appropriate Bioinformatics tools
- To visualize and analyse recent research

Course Outcomes

At the end of the Course, Students will be able to,

- Become aware of concept of bioinformatics
- Apply Knowledge from basic to advanced level in bioinformatics
- Learn major topics of Bioinformatics
- Demonstrate different biological suitable datasets
- Demonstrate appropriate Bioinformatics tools
- Connect and integrate the knowledge obtained for applications related to Bioinformatics, their tools and database

Syllabus

Unit	Course Content	Hours
I	**Introduction to Bioinformatics and Molecular Biology**	06
	Introduction to bioinformatics and its importance. Overview of molecular biology and biotechnology. Understanding biological molecules and cellular processes. DNA, RNA, protein and their functions. Genome wide Association Studies	
II	**Computational Genomics and Transcriptomics**	06
	DNA sequence analysis. Intron, Exon, Microarray, RNAseq. Genome annotation and gene prediction. RNA sequencing and analysis. Differential gene expression analysis. NCBI datasets, repositories	
III	**Structural Bioinformatics and Drug Discovery**	06
	Protein structure prediction. Homology modeling and threading. Protein-ligand interactions and molecular docking. Computer-aided drug design. Data Science for Medical Image analysis	
IV	**Systems Biology and Network Analysis**	06
	Overview of systems biology. Regulatory networks. Metabolic networks. Network analysis and visualization tools.	
V	**Data Mining and Machine Learning in Bioinformatics**	06
	Introduction to data mining and machine learning. Clustering and classification. Feature selection and dimensionality reduction. Deep learning in bioinformatics. Machine Learning, Deep Learning, Convolution Neural Network Application for Gene Networks	
VI	**Applications of Bioinformatics**	06
	Personalized medicine and pharmacogenomics. Disease diagnosis and treatment. Agricultural and environmental biotechnology. Current research and future directions. Implementation use cases on recent research through Researchgate, GitHub.	

Teaching Plan

Sr. No.	Unit	Topics to be covered	Book Referred	Total Lecture Planned
1	I	Introduction to bioinformatics and its importance. Overview of molecular biology and biotechnology. Understanding biological molecules and cellular processes. DNA, RNA, protein and their functions. Genome wide Association Studies	1.1,1.2	6
2	II	DNA sequence analysis. Intron, Exon, Microarray, RNAseq. Genome annotation and gene prediction. RNA sequencing and analysis. Differential gene expression analysis. NCBI datasets, repositories.	1.1,2.1	6
3	III	Protein structure prediction. Homology modeling and threading. Protein-ligand interactions and molecular docking. Computer-aided drug design. Data Science for Medical Image analysis.	1.1,1.2,1.3	6
4	IV	Overview of systems biology. Regulatory networks. Metabolic networks. Network analysis and visualization tools.	1.1,1.2,1.3,2.1, 2.2	6
5	V	Introduction to data mining and machine learning. Clustering and classification. Feature selection and dimensionality reduction. Deep learning in bioinformatics. Machine Learning, Deep Learning, Convolution Neural Network Application for Gene Networks	2.1, 2.2	6
6	VI	Personalized medicine and pharmacogenomics. Disease diagnosis and treatment. Agricultural and environmental biotechnology. Current research and future directions. Implementation use cases on recent research through Researchgate, GitHub.	2.1, 2.2	6

1. Text Books:

1. S.C. Rastogi & others, "Bioinformatics- Concepts, Skills, and Applications", 2 nd edition, CBS Publishing, 2016, ISBN: 9788123914824
2. Cynthia Gibas and Per Jambeck, "Developing Bioinformatics Computer Skills" O'Reilly press, Shorff Publishers and Distributors Pvt. Ltd.,.
3. Mario Cannataro, Pietro Hiram Guzzi, et.al., "Artificial Intelligence in Bioinformatics", Elsevier Science
4. Bourne PE, Weissig H., "Structural Bioinformatics" (Methods of Biochemical Analysis, V. 44). Wiley-Liss, 2003

2. Reference Books:

1. Basant K. Tiwary, "Bioinformatics and Computational Biology : A Primer for Biologists", Springer Nature, 2021
1. Zoe'Lacroix and critchlow, "Bioinformatics: Managing scientific data", Morgan Kaufmann, 2004
2. Campbell AM and Heyer LJm, "Discovering Genomics, Proteomics and Bioinformatics", Pearson Education
3. S.C., Rastogi, Parag, Mendiratta, Namita, "Bioinformatics: Methods and Applications: Genomics, Proteomics and Drug Discovery", 5 th edition, PHI Learning, 2022

Reference Web Links/ Research Paper/ Referred Book other than Mention in Syllabus:

- **e-Resources:**

1. http://www.bioinformatics.org/
2. https://ocw.mit.edu/ans7870/6/6.047/f15/MIT6_047F15_Compiled.pdf

- **MOOC Courses:**

1. ***https://archive.nptel.ac.in/courses/102/106/102106065/https://nptel.ac.in/courses/106/102/106102065/***
2. ***https://www.udemy.com/course/genetics-and-next-generation-sequencing-for-bioinformatics/***
3. ***https://www.coursera.org/specializations/bioinformatics***

Unit No.-I-

Name of the Unit: Introduction to Bioinformatics and Molecular Biology

Lecture No.	Details of the Topic to be covered	References
1	Introduction to bioinformatics and its importance.	1.1,1.2
2	Overview of molecular biology and biotechnology	1.1,1.2
3	Understanding biological molecules and cellular processes.	1.1,1.2
4	DNA, RNA	1.1,1.2
5	Protein and their functions.	1.1,1.2
6	Genome wide Association Studies	1.1,1.2

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Define Bioinformatics and its importance.
Q. 2	What is DNA, RNA, protein and their functions?
Q. 3	Explain genome-wide association studies (GWAS)

Unit No.-II-

Name of the Unit: Computational Genomics and Transcriptomics

Lecture No.	Details of the Topic to be covered	References
1	DNA sequence analysis.	1.1,2.1
2	Intron, Exon, Microarray, RNA sequence	1.1,2.1
3	Genome annotation and gene prediction.	1.1,2.1
4	RNA sequencing and analysis	1.1,2.1
5	Differential gene expression analysis.	1.1,2.1

6	NCBI datasets, repositories	1.1,2.1

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Explain Bioinformatics application related to the Genome Annotation
Q. 2	What is the difference between DNA sequence and RNA sequence?
Q. 3	What is NCBI datasets

Unit No.-III

Name of the Unit: Structural Bioinformatics and Drug Discovery

Lecture No.	Details of the Topic to be covered	References
1	Protein structure prediction.	1.1,1.2,1.3
2	Homology modeling and threading	1.1,1.2,1.3
3	Protein-ligand interactions	1.1,1.2,1.3
4	molecular docking.	1.1,1.2,1.3
5	Computer-aided drug design	1.1,1.2,1.3
6	Data Science for Medical Image analysis.	1.1,1.2,1.3

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Distinguish between Homology modeling and threading
Q. 2	Describe Protein structure prediction in detail.
Q. 3	Explain Data Science for Medical Image analysis.
Q. 4	Distinguish between Protein-ligand interactions and molecular docking
Q. 5	Define the terms: Structural Bioinformatics and Drug Discovery

Q. 6	What is Computer-aided drug design?

Unit No.-IV

Name of the Unit: Systems Biology and Network Analysis

Lecture No.	Details of the Topic to be covered	References
1	Introduction	1.1, 1.2, 1.3, 2.1,2.2
2	Overview of systems biology.	1.1, 1.2, 1.3, 2.1,2.2
3	Regulatory networks.	1.1, 1.2, 1.3, 2.1,2.2
4	Metabolic networks.	1.1, 1.2, 1.3, 2.1,2.2
5	Network analysis	1.1, 1.2, 1.3, 2.1,2.2
6	visualization tools.	1.1, 1.2, 1.3, 2.1,2.2

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Distinguish between Regulatory networks and Metabolic networks
Q. 2	Describe Network analysis in detail.
Q. 3	Explain systems biology.
Q. 4	Distinguish between Network analysis and visualization tools.
Q. 5	Define the term: Metabolic networks
Q. 6	What are network biology methods?

Unit No.-V

Name of the Unit: Data Mining and Machine Learning in Bioinformatics

Lecture No.	Details of the Topic to be covered	References
1	Introduction to data mining and machine learning.	2.1, 2.2
2	Clustering and classification.	2.1, 2.2
3	Feature selection and dimensionality reduction	2.1, 2.2
4	Deep learning in bioinformatics	2.1, 2.2
5	Machine Learning, Deep Learning,	2.1, 2.2
6	Convolution Neural Network Application for Gene Networks	2.1, 2.2

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Explain clustering requires merging approach?
Q. 2	Which is the most direct application of neural networks?
Q. 3	what is single layer associative neural networks?

Unit No.-VI-

Name of the Unit: Introduction to Animation and Gaming

Lecture No.	Details of the Topic to be covered	References
1	Personalized medicine and pharmacogenomics.	2.1, 2.2
2	Disease diagnosis and treatment.	2.1, 2.2
3	Agricultural and environmental biotechnology.	2.1, 2.2
4	Current research and future directions	2.1, 2.2

5	Implementation use cases on recent research through Researchgate, GitHub.	2.1, 2.2
6	Applications of Bioinformatics	2.1, 2.2

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Describe Personalized medicine and pharmacogenomics?
Q. 2	What is Agricultural and environmental biotechnology
Q. 3	Explain the term: a) Current research and future directions b)Disease diagnosis and treatment
Q. 4	Implementation use cases on recent research through Researchgate, GitHub.

Subject 4: Elective IV Information Retrieval

Weekly Work Loads(in Hrs)	Lecture	Tutorial	Practical
	03	-	-

Online/ In-sem	Theory	Practical	Oral	Term-work	Total Marks	Credit
30	70	0	0	0	100	03

Course Objectives:

- To understand the basics of Information Retrieval
- To understand the concepts of Indexing & Query Processing for Information Retrieval
- To provide comprehensive details about various Evaluation methods
- To understand the different methods of Text Classification and Clustering
- To understand various search engine system operations and web structures
- To understand various applications of Information Retrieval

Course Outcomes:

At the end of the Course, Students will be able to,

- Understand the concept of Information Retrieval
- Use an indexing approach for retrieval of documents
- Evaluate and analyze the retrieved information
- Apply appropriate method of Text Classification and Clustering
- Design and implement innovative features in search engines
- Analyze different real-life application of Information Retrieval

Syllabus

Unit	Course Content	Hours
I	**Introduction**	06
	Introduction to information retrieval, Major challenges in IR, Features of an IR system, components of an IR model, IR system block diagram, Boolean retrieval, Information versus Data Retrieval, Text categorization, IR processes and fields, Vector Model, Probabilistic Model and Latent Semantic Indexing Model. **Scan conversion:** Line drawing algorithms: Digital Differential Analyzer (DDA), Bresenham. Circle drawing algorithms: DDA, Bresenham, and Midpoint.	
II	**Dictionaries and Query Processing**	06
	Components of Index, Index Life Cycle, Static Inverted Index, Dictionaries-Types (Sort Based, Hash Based, Interleaving & Posting Lists), Index Construction (In memory, Sort Based, Merge Based, Disk Based Index Construction), Dynamic Indexing, Query Processing for Ranked Retrieval, Document at a Time Query Processing, Term at a Time Query Processing, Pre-computing Score Contributions, Impact Ordering, Query Optimization.	
III	**Probabilistic Retrieval and Language Modelling related methods**	06
	Probabilistic Retrieval: Review of Basic Probability Theory; The Probability Ranking Principle: The 1/0 loss case, the PRP with retrieval costs; The Binary Independence Model; Term Frequency; An appraisal and some extensions: An appraisal of probabilistic models, tree-structured dependencies between terms, Okapi BM25: a non-binary model, Bayesian network approaches to IR, Relevance Feedback, Field Weights:BM25F. Language models for information retrieval: generating queries from documents; Language models: finite automata and language models; types of language models; multinomial distributions over words; Ranking with Language Models; Divergence from Randomness, Passage Retrieval, and Ranking.	
IV	**Text classification & Text clustering**	06
	Text Classification: Introduction to Text Classification, Naiyes Bayes Model, K Nearest neighbor, spam filtering, Support Vector Machine Classifier, Vector Space classification using hyperplanes, kernel function. Text Clustering: Clustering vs Classification,	

	partitioning methods. Clustering Algorithms: k-means clustering, Agglomerative hierarchical clustering, Expectation Maximization, Mixture of Gaussians Model	
V	**Web Retrieval and Web Crawling**	06
	Parallel information retrieval: Parallel query processing, Mapreduce Web Retrieval: Search Engine Architectures, Cluster based Architecture, Distributed Architectures, Search Engine Ranking. Link based Ranking, Page Ranking Algorithm, Simple Ranking Functions and Evaluations. Web Crawler: Web Crawler structure, Web crawler libraries, Python Scrapy, Beautiful Soup, Applications	
VI	**IR applications**	06
	Multimedia Information Retrieval: Spoken Language Audio Retrieval, Non-Speech Audio Retrieval, Graph Retrieval, Imagery Retrieval, Video Retrieval Recommender System: Collaborative Filtering, Content Based Recommendation, Knowledge Based Recommendation Information Extraction and Integration: Extracting Data from Text. Semantic Web, Collecting and Integrating Specialized Information on the web.	

Teaching Plan

Sr. No.	Unit	Topics to be covered	Book Referred	Total Lecture Planned
1	I	Introduction to information retrieval, Major challenges in IR, Features of an IR system, components of an IR model, IR system block diagram, Boolean retrieval, Information versus Data Retrieval, Text categorization, IR processes and fields, Vector Model, Probabilistic Model and Latent Semantic Indexing Model.	C. Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval"	06
2	II	Components of Index, Index Life Cycle, Static Inverted Index, Dictionaries-Types (Sort Based, Hash Based, Interleaving & Posting Lists), Index Construction (In memory, Sort Based, Merge Based, Disk Based Index Construction), Dynamic Indexing, Query Processing for Ranked Retrieval, Document at a Time Query Processing, Term at a Time Query Processing, Pre-computing Score Contributions, Impact Ordering, Query Optimization	G. Kowalski, M.T. Maybury, "Information storage and Retrieval System"	06
3	III	Probabilistic Retrieval: Review of Basic Probability Theory; The Probability Ranking Principle: The 1/0 loss case, the PRP with retrieval costs; The Binary Independence Model; Term Frequency; An appraisal and some extensions: An appraisal of probabilistic models, tree-structured dependencies between terms, Okapi BM25: a non-binary model, Bayesian network approaches to IR, Relevance Feedback, Field Weights:BM25F. Language models for information retrieval: generating queries from documents; Language models: finite automata and language models; types of language models; multinomial distributions over words;Ranking with Language Models; Divergence from Randomness, Passage Retrieval, and Ranking	W.B. Croft, J. Lafferty, "Language Modeling for Information Retrieval"	06
4	IV	Text Classification: Introduction to Text Classification, Naiyes Bayes Model, K Nearest	C. Manning, P. Raghavan, and H.	06

		neighbor, spam filtering, Support Vector Machine Classifier, Vector Space classification using hyperplanes, kernel function. Text Clustering: Clustering vs Classification, partitioning methods. Clustering Algorithms: k-means clustering, Agglomerative hierarchical clustering, Expectation Maximization, Mixture of Gaussians Model	Schütze, "Introduction to Information Retrieval"	
5	V	Parallel information retrieval: Parallel query processing, Mapreduce Web Retrieval: Search Engine Architectures, Cluster based Architecture, Distributed Architectures, Search Engine Ranking. Link based Ranking, Page Ranking Algorithm, Simple Ranking Functions and Evaluations. Web Crawler: Web Crawler structure, Web crawler libraries, Python Scrapy, Beautiful Soup, Applications	S. Buttcher, C. Clarke, and G. Cormack, "Information Retrieval: Implementing and Evaluating Search Engines"	06
6	VI	Multimedia Information Retrieval: Spoken Language Audio Retrieval, Non-Speech Audio Retrieval, Graph Retrieval, Imagery Retrieval, Video Retrieval Recommender System: Collaborative Filtering, Content Based Recommendation, Knowledge Based Recommendation Information Extraction and Integration: Extracting Data from Text. Semantic Web, Collecting and Integrating Specialized Information on the web.	Jannach D., Zanker M. and FelFering A., "Recommender Systems: An Introduction"	06

Text Books:

1. C. Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval", Cambridge University Press, 2008, ISBN: 9780521865715
2. S. Buttcher, C. Clarke, and G. Cormack, "Information Retrieval: Implementing and Evaluating Search Engines" MIT Press, 2010, ISBN: 0-408-70929-4
3. Bruce Croft, Donald Metzler and Trevor Strohman, "Search Engines: Information Retrieval in Practice", 1st edition, Addison Wesley, 2009, ISBN: 9780135756324
4. Jannach D., Zanker M. and FelFering A., "Recommender Systems: An Introduction", 1st edition, Cambridge University Press

Reference Books:

5. Manouselis N., Drachsler H., Verbert K., Duval E., "Recommender Systems For

Learning", 1st edition, Springer, 2013

6. G. Kowalski, M.T. Maybury, "Information storage and Retrieval System", Springer, 2005
7. W.B. Croft, J. Lafferty, "Language Modeling for Information Retrieval", Springer, 2003

Reference Web Links/ Research Paper/ Referred Book other than Mention in Syllabus:

- Information Retrieval- http://www.informationretrieval.org
- https://www.youtube.com/watch?v=fFxpSmyICwI
- https://www.youtube.com/watch?v=X5GvBh4qY0s

Unit No: I Introduction

Lecture No.	Details of the Topic to be covered	References
1	Introduction to information retrieval, Major challenges in IR, Features of an IR system	1.1
2	components of an IR model, IR system block diagram	1.1
3	Boolean retrieval, Information versus Data Retrieval	1.1
4	Text categorization, IR processes and fields	1.1
5	Vector Model, Probabilistic Model	1.1
6	Latent Semantic Indexing Model	1.1

Question Bank: Theory & Numerical Mapped to Course Outcome:

Q. 1	Explain major challenges in IR.

Q. 2	Explain Components of IR
Q. 3	Explain IR system with neat block diagram.
Q. 4	Differentiate between Information retrieval and data retrieval
Q. 5	What is Latent Semantic Indexing Model?

Unit No.-II Dictionaries and Query Processing

Lecture No.	Details of the Topic to be covered	References
1	Components of Index, Index Life Cycle, Static Inverted Index	1.1, 2.2
2	Dictionaries-Types (Sort Based, Hash Based, Interleaving & Posting Lists)	1.1, 2.2
3	Index Construction (In memory, Sort Based, Merge Based, Disk Based Index Construction)	1.1, 2.2
4	Dynamic Indexing, Query Processing for Ranked Retrieval	1.1, 2.2
5	Document at a Time Query Processing, Term at a Time Query Processing, Pre-computing Score Contributions, Impact Ordering	1.1, 2.2
6	Query Optimization, Review of Basic Probability Theory; The Probability Ranking Principle	1.1, 2.2

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Explain Index life cycle.
Q. 2	Explain types of dictionary.
Q. 3	Explain process of index construction.

Q. 4	Explain Query Processing for Ranked Retrieval

Q. 5	What is query optimization?

Unit No.-III Probabilistic Retrieval and Language Modeling related methods

Lecture No.	Details of the Topic to be covered	References
1	Review of Basic Probability Theory; The Probability Ranking Principle: The 1/0 loss case, the PRP with retrieval costs; The Binary Independence Model; Term Frequency	2.3
2	An appraisal and some extensions: An appraisal of probabilistic models	2.3
3	tree-structured dependencies between terms, Okapi BM25: a non-binary model, generating queries from documents;	2.3
4	Bayesian network approaches to IR, Relevance Feedback, Field Weights:BM25F	2.3
5	Language models: finite automata and language models; types of language models	2.3
6	Multinomial distributions over words; Ranking with Language Models; Divergence from Randomness, Passage Retrieval and Ranking.	2.3

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Explain the Probability Ranking Principle.

Q. 2	Explain Okapi BM25: a non-binary model.

Q. 3	Enlist and explain Bayesian network approaches to IR
Q. 4	Enlist and explain types of language models
Q. 5	What is passage Retrieval and Ranking?

Unit No.-IV Text classification & Text clustering

Lecture No.	Details of the Topic to be covered	References
1	Introduction to Text Classification, Naiyes Bayes Model,K Nearest neighbor	1.1, 1.2, 2.2
2	spam filtering, Support Vector Machine Classifier	1.1, 1.2, 2.2
3	Vector Space classification using hyperplanes, kernel function	1.1, 1.2, 2.2
4	Clustering vs Classification, partitioning methods, k-means clustering	1.1, 1.2, 2.2
5	Agglomerative hierarchical clustering,	1.1, 1.2, 2.2
6	Expectation Maximization, Mixture of Gaussians Model	1.1, 1.2, 2.2

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Differentiate between Clustering and Classification
Q. 2	Explain spam filtering and it’s methods.
Q. 3	Explain Vector Space classification using hyperplane.
Q. 4	Explain types of clustering.
Q. 5	What is expectation maximization?

Unit No.-V Web Retrieval and Web Crawling

Lecture No.	Details of the Topic to be covered	References
1	Parallel query processing, Mapreduce	1.3
2	Search Engine Architectures, Cluster based Architecture	1.3
3	Distributed Architectures, Search Engine Ranking	1.3
4	Link based Ranking, Page Ranking Algorithm, Simple Ranking Functions and Evaluations	1.3
5	Web Crawler structure, Web crawler libraries	1.3
6	Python Scrapy, Beautiful Soup, Applications	1.3

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	What is parallel query processing?
Q. 2	Explain page ranking algorithm.
Q. 3	Explain ranking functions and their evaluation.
Q. 4	Enlist and explain web crawler libraries.
Q. 5	What are the applications of Beautiful Soup

Unit No.-VI IR applications

Lecture No.	Details of the Topic to be covered	References
1	Spoken Language Audio Retrieval, Non-Speech Audio Retrieval	1.4, 2.1, 2.2
2	Graph Retrieval, Imagery Retrieval, Video Retrieval	1.4, 2.1, 2.2
3	Collaborative Filtering, Content Based Recommendation	1.4, 2.1, 2.2
4	Knowledge Based Recommendation Information Extraction and Integration	1.4, 2.1, 2.2
5	Extracting Data from Text, Semantic Web	1.4, 2.1, 2.2
6	Collecting and Integrating Specialized Information on the web	1.4, 2.1, 2.2

Question Bank: Theory & Numerical
Mapped to Course Outcome:

Q. 1	Differentiate between Spoken Language Audio Retrieval and Non-Speech Audio Retrieval.
Q. 2	Explain types of retrieval of information
Q. 3	What is collaborative filtering?
Q. 4	How can extract data from text?
Q. 5	How we collect and integrate the information on the web?