

ANNA UNIVERSITY, CHENNAI
NON - AUTONOMOUS COLLEGES AFFILIATED ANNA UNIVERSITY
M.E. BIG DATA ANALYTICS
REGULATIONS – 2021
CHOICE BASED CREDIT SYSTEM

1. PROGRAMME EDUCATIONAL OBJECTIVES (PEOs):

- I. Apply quantitative modeling and data analysis techniques to the solution of real world business problems, communicate findings, and effectively present results using data visualization techniques
- II. Apply principles of Data Science to the analysis of business problems.
- III. Employ cutting edge tools and technologies to analyze big data.
- IV. Adopt ethical practices, collaborate with team members as a member or leader, and engage in constant updation of tools and techniques.
- V. Strongly focus on ingenious ideas and critical analysis to serve the society, locally and internationally as entrepreneurs using the principles of data analytics.

2. PROGRAM SPECIFIC OUTCOMES (PSOs) :

1. Understand the impact of big data for business decisions and strategy.
2. Gain hands-on experience on large-scale analytics tools to solve some open big data problems.
3. Understand the concept and challenge of big data and the inadequacy of existing technologies to analyze big data.

ANNA UNIVERSITY, CHENNAI
NON - AUTONOMOUS COLLEGES AFFILIATED ANNA UNIVERSITY
M.E. BIG DATA ANALYTICS
REGULATIONS – 2021
CHOICE BASED CREDIT SYSTEM
I TO IV SEMESTERS CURRICULA AND 1st SEMESTER SYLLABI
SEMESTER I

S. NO.	COURSE CODE	COURSE TITLE	CATE- GORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
THEORY								
1.	MA4109	Applied Probability and Statistics	FC	4	0	0	4	4
2.	RM4151	Research Methodology and IPR	RMC	2	0	0	2	2
3.	CP4151	Advanced Data Structures and Algorithms	PCC	3	0	0	3	3
4.	BD4151	Foundations of Data Science	PCC	3	0	0	3	3
5.	CP4252	Machine Learning	PCC	3	0	2	5	4
6.	CP4152	Database Practices	PCC	3	0	2	5	4
7.		Audit Course – I*	AC	2	0	0	2	0
PRACTICALS								
8.	CP4161	Advanced Data Structures and Algorithms Laboratory	PCC	0	0	4	4	2
9.	BD4111	Big Data Computing Laboratory	PCC	0	0	2	2	1
TOTAL				20	0	10	30	23

*Audit course is optional

SEMESTER II

S. NO.	COURSE CODE	COURSE TITLE	CATE- GORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
THEORY								
1.	BD4252	Big Data Security	PCC	3	0	0	3	3
2.	BD4251	Big Data Mining and Analytics	PCC	3	0	0	3	3
3.	MP4253	Cloud Computing Technologies	PCC	3	0	0	3	3
4.	BD4201	Information Storage Management	PCC	3	0	0	3	3
5.	MP4151	Embedded Systems and IIOT	PCC	3	0	2	5	4
6.		Professional Elective I	PEC	3	0	0	3	3
7.		Audit Course – II*	ACC	2	0	0	2	0
PRACTICALS								
8.	BD4211	Big Data Mining And Analytics Laboratory	PCC	0	0	4	4	2
9.	BD4212	Term Paper and Seminar	EEC	0	0	2	2	1
TOTAL				20	0	8	28	22

*Audit course is optional

SEMESTER III

S. NO.	COURSE CODE	COURSE TITLE	CATE- GORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
THEORY								
1.		Professional Elective II	PEC	3	0	0	3	3
2.		Professional Elective III	PEC	3	0	0	3	3
3.		Professional Elective IV	PEC	3	0	0	3	3
4.		Open Elective	OEC	3	0	0	3	3
PRACTICALS								
5.	BD4311	Project Work I	EEC	0	0	12	12	6
TOTAL				12	0	12	24	18

SEMESTER IV

S. NO.	COURSE CODE	COURSE TITLE	CATE- GORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
PRACTICALS								
1.	BD4411	Project Work II	EEC	0	0	24	24	12
TOTAL				0	0	24	24	12

TOTAL NO. OF CREDITS: 75

PROGRESS THROUGH KNOWLEDGE

PROFESSIONAL ELECTIVES**SEMESTER II, ELECTIVE I**

S. NO.	COURSE CODE	COURSE TITLE	CATEGORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
1.	BD4071	High Performance Computing for Big Data	PEC	3	0	0	3	3
2.	MP4075	Web Services and API Design	PEC	3	0	0	3	3
3.	CP4076	Information Retrieval Techniques	PEC	3	0	0	3	3
4.	CP4074	Data Visualization Techniques	PEC	3	0	0	3	3
5.	BD4001	Principles of Supply Chain Management	PEC	3	0	0	3	3
6.	BD4002	Computational Geometry	PEC	3	0	0	3	3

SEMESTER III, ELECTIVE II

S. NO.	COURSE CODE	COURSE TITLE	CATEGORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
1.	MP4074	Soft Computing Techniques	PEC	3	0	0	3	3
2.	MU4152	Multimedia Communication Networks	PEC	3	0	0	3	3
3.	BD4003	Parallel and Distributed Computing	PEC	3	0	0	3	3
4.	BD4072	Predictive Modeling	PEC	3	0	0	3	3
5.	BD4004	Image Processing and Analysis	PEC	3	0	0	3	3

SEMESTER III, ELECTIVE III

S. NO.	COURSE CODE	COURSE TITLE	CATEGORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
1.	MP4071	Cognitive computing	PEC	3	0	0	3	3
2.	IF4079	Social Network analysis	PEC	3	0	0	3	3
3.	BD4005	Virtualization techniques and applications	PEC	3	0	0	3	3
4.	ML4251	Natural Language Processing	PEC	2	0	2	4	3

SEMESTER III, ELECTIVE IV

S. NO.	COURSE CODE	COURSE TITLE	CATEGORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
1.	BD4006	Data Intensive Computing	PEC	3	0	0	3	3
2.	BD4007	R Language for Mining	PEC	3	0	0	3	3
3.	CP4081	Web analytics	PEC	3	0	0	3	3
4.	MP4072	Healthcare Analytics	PEC	3	0	0	3	3
5.	BD4008	Statistics for Business Analytics	PEC	3	0	0	3	3

AUDIT COURSES (AC)

Registration for any of these courses is optional to students

SL. NO	COURSE CODE	COURSE TITLE	PERIODS PER WEEK			CREDITS
			L	T	P	
1.	AX4091	English for Research Paper Writing	2	0	0	0
2.	AX4092	Disaster Management	2	0	0	0
3.	AX4093	Constitution of India	2	0	0	0
4.	AX4094	நற்றமிழ் இலக்கியம்	2	0	0	0

PROGRESS THROUGH KNOWLEDGE

COURSE OBJECTIVES:

- To compute probabilities and moments of standard distributions.
- To gain knowledge about regression and correlation.
- To provide the most appropriate estimator of the parameter in statistical inference.
- To decide whether to accept or reject specific values of a parameter.
- To understand many real-world problems fall naturally within the framework of multivariate normal theory.

UNIT I PROBABILITY AND RANDOM VARIABLES**12**

Probability – Axioms of probability – Conditional probability – Bayes theorem - Random variables - Probability function – Moments – Moment generating functions and their properties – Binomial, Poisson, Geometric, Uniform, Exponential, Gamma and Normal distributions – Function of a random variable.

UNIT II TWO DIMENSIONAL RANDOM VARIABLES**12**

Joint distributions – Marginal and conditional distributions – Functions of two dimensional random variables – Regression curve – Correlation.

UNIT III ESTIMATION THEORY**12**

Unbiased estimators – Method of moments – Maximum likelihood estimation - Curve fitting by principle of least squares – Regression lines.

UNIT IV TESTING OF HYPOTHESIS**12**

Sampling distributions – Type I and Type II errors – Small and large samples – Tests based on Normal, t, Chi square and F distributions for testing of mean, variance and proportions – Tests for independence of attributes and goodness of fit.

UNIT V MULTIVARIATE ANALYSIS**12**

Random vectors and matrices – Mean vectors and covariance matrices – Multivariate normal density and its properties – Principal components - Population principal components – Principal components from standardized variables

TOTAL : 60 PERIODS**COURSE OUTCOMES :**

After completing this course, students should demonstrate competency in the following topics:

- Basic probability axioms and rules and the moments of discrete and continuous random variables.
- To deal with problems involving two dimensional random variables.

- Consistency, efficiency and unbiasedness of estimators, method of maximum likelihood estimation and Central Limit Theorem.
- Use statistical tests in testing hypotheses on data.
- Perform exploratory analysis of multivariate data, such as multivariate normal density, calculating descriptive statistics, testing for multivariate normality.

REFERENCES :

1. Devore, J. L., "Probability and Statistics for Engineering and the Sciences", 8th Edition, Cengage Learning, 2014.
2. Dallas E. Johnson, "Applied Multivariate Methods for Data Analysis", Thomson and Duxbury press, 1998.
3. Gupta S.C. and Kapoor V.K., "Fundamentals of Mathematical Statistics", 12th Edition, Sultan and Sons, New Delhi, 2020.
4. Johnson, R.A., Miller, I and Freund J., "Miller and Freund's Probability and Statistics for Engineers ", 9th Edition, Pearson Education, Asia, 2016.
5. Richard A. Johnson and Dean W. Wichern, "Applied Multivariate Statistical Analysis", 6th Edition, Pearson Education, Asia, 2012.

www.binils.com

RM4151

RESEARCH METHODOLOGY AND IPR

L T P C

2 0 0 2

UNIT I RESEARCH DESIGN

6

Overview of research process and design, Use of Secondary and exploratory data to answer the research question, Qualitative research, Observation studies, Experiments and Surveys.

UNIT II DATA COLLECTION AND SOURCES

6

Measurements, Measurement Scales, Questionnaires and Instruments, Sampling and methods. Data - Preparing, Exploring, examining and displaying.

UNIT III DATA ANALYSIS AND REPORTING

6

Overview of Multivariate analysis, Hypotheses testing and Measures of Association. Presenting Insights and findings using written reports and oral presentation.

UNIT IV INTELLECTUAL PROPERTY RIGHTS

6

Intellectual Property – The concept of IPR, Evolution and development of concept of IPR, IPR development process, Trade secrets, utility Models, IPR & Biodiversity, Role of WIPO and WTO in

IPR establishments, Right of Property, Common rules of IPR practices, Types and Features of IPR Agreement, Trademark, Functions of UNESCO in IPR maintenance.

UNIT V PATENTS

6

Patents – objectives and benefits of patent, Concept, features of patent, Inventive step, Specification, Types of patent application, process E-filing, Examination of patent, Grant of patent, Revocation, Equitable Assignments, Licences, Licensing of related patents, patent agents, Registration of patent agents.

TOTAL : 30 PERIODS

REFERENCES:

1. Cooper Donald R, Schindler Pamela S and Sharma JK, “Business Research Methods”, Tata McGraw Hill Education, 11e (2012).
2. Catherine J. Holland, “Intellectual property: Patents, Trademarks, Copyrights, Trade Secrets”, Entrepreneur Press, 2007.
3. David Hunt, Long Nguyen, Matthew Rodgers, “Patent searching: tools & techniques”, Wiley, 2007.
4. The Institute of Company Secretaries of India, Statutory body under an Act of parliament, “Professional Programme Intellectual Property Rights, Law and practice”, September 2013.

www.binils.com

CP4151

ADVANCED DATA STRUCTURES AND ALGORITHMS

L T P C

3 0 0 3

COURSE OBJECTIVES:

- To understand the usage of algorithms in computing
- To learn and use hierarchical data structures and its operations
- To learn the usage of graphs and its applications
- To select and design data structures and algorithms that is appropriate for problems
- To study about NP Completeness of problems.

UNIT I ROLE OF ALGORITHMS IN COMPUTING & COMPLEXITY ANALYSIS

9

Algorithms – Algorithms as a Technology -Time and Space complexity of algorithms- Asymptotic analysis-Average and worst-case analysis-Asymptotic notation-Importance of efficient algorithms- Program performance measurement - Recurrences: The Substitution Method – The Recursion-Tree Method- Data structures and algorithms.

UNIT II HIERARCHICAL DATA STRUCTURES

9

Binary Search Trees: Basics – Querying a Binary search tree – Insertion and Deletion- Red Black trees: Properties of Red-Black Trees – Rotations – Insertion – Deletion -B-Trees: Definition of B -

REFERENCES:

1. S.Sridhar," Design and Analysis of Algorithms", Oxford University Press, 1st Edition, 2014.
2. Adam Drozdex, "Data Structures and algorithms in C++", Cengage Learning, 4th Edition, 2013.
3. T.H. Cormen, C.E.Leiserson, R.L. Rivest and C.Stein, "Introduction to Algorithms", Prentice Hall of India, 3rd Edition, 2012.
4. Mark Allen Weiss, "Data Structures and Algorithms in C++", Pearson Education, 3rd Edition, 2009.
5. E. Horowitz, S. Sahni and S. Rajasekaran, "Fundamentals of Computer Algorithms", University Press, 2nd Edition, 2008.
6. Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman, "Data Structures and Algorithms", Pearson Education, Reprint 2006.

BD4151

FOUNDATIONS OF DATA SCIENCE

L T P C

3 0 0 3

COURSE OBJECTIVES:

- To apply fundamental algorithms to process data.
- Learn to apply hypotheses and data into actionable predictions.
- Document and transfer the results and effectively communicate the findings using visualization techniques.
- To learn statistical methods and machine learning algorithms required for Data Science.
- To develop the fundamental knowledge and understand concepts to become a data science professional.

UNIT I INTRODUCTION TO DATA SCIENCE

9

Data science process – roles, stages in data science project – working with data from files – working with relational databases – exploring data – managing data – cleaning and sampling for modeling and validation – introduction to NoSQL.

UNIT II MODELING METHODS

9

Choosing and evaluating models – mapping problems to machine learning, evaluating clustering models, validating models – cluster analysis – K-means algorithm, Naïve Bayes – Memorization Methods – Linear and logistic regression – unsupervised methods.

UNIT III INTRODUCTION TO R

9

Reading and getting data into R – ordered and unordered factors – arrays and matrices – lists and data frames – reading data from files – probability distributions – statistical models in R -

manipulating objects – data distribution.

UNIT IV MAP REDUCE

9

Introduction – distributed file system – algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce – Hadoop - Understanding the Map Reduce architecture - Writing Hadoop MapReduce Programs - Loading data into HDFS - Executing the Map phase - Shuffling and sorting - Reducing phase execution.

UNIT V DATA VISUALIZATION

9

Documentation and deployment – producing effective presentations – Introduction to graphical analysis – plot() function – displaying multivariate data – matrix plots – multiple plots in one window - exporting graph using graphics parameters - Case studies.

TOTAL : 45 PERIODS

COURSE OUTCOMES:

CO1: Obtain, clean/process and transform data.

CO2: Analyze and interpret data using an ethically responsible approach.

CO3: Use appropriate models of analysis, assess the quality of input, derive insight from results, and investigate potential issues.

CO4: Apply computing theory, languages and algorithms, as well as mathematical and statistical models, and the principles of optimization to appropriately formulate and use data analyses.

CO5: Formulate and use appropriate models of data analysis to solve business-related challenges.

REFERENCES

1. Nina Zumel, John Mount, "Practical Data Science with R", Manning Publications, 2014.
2. Mark Gardener, "Beginning R - The Statistical Programming Language", John Wiley & Sons, Inc., 2012.
3. W. N. Venables, D. M. Smith and the R Core Team, "An Introduction to R", 2013.
4. Tony Ojeda, Sean Patrick Murphy, Benjamin Bengfort, Abhijit Dasgupta, "Practical Data Science Cookbook", Packt Publishing Ltd., 2014.
5. Nathan Yau, "Visualize This: The FlowingData Guide to Design, Visualization, and
6. Statistics", Wiley, 2011.
7. Boris Lublinsky, Kevin T. Smith, Alexey Yakubovich, "Professional Hadoop Solutions", John Wiley & Sons Inc., 2013.

COURSE OBJECTIVES:

- To understand the concepts and mathematical foundations of machine learning and types of problems tackled by machine learning
- To explore the different supervised learning techniques including ensemble methods
- To outline different aspects of unsupervised learning and reinforcement learning
- To outline the role of probabilistic methods for machine learning
- To understand the basic concepts of neural networks and deep learning

UNIT I INTRODUCTION AND MATHEMATICAL FOUNDATIONS 9

What is Machine Learning? Need –History – Definitions – Applications - Advantages, Disadvantages & Challenges -Types of Machine Learning Problems – Mathematical Foundations - Linear Algebra & Analytical Geometry -Probability and Statistics -Vector Calculus & Optimization -Information theory

UNIT II SUPERVISED LEARNING 9

Introduction-Discriminative and Generative Models -Linear Regression -Least Squares -Under fitting / Overfitting -Cross-Validation – Lasso Regression-Classification -Logistic Regression-Gradient Linear Models -Support Vector Machines -Kernel Methods -Instance based Methods - K-Nearest Neighbours - Tree based Methods –Decision Trees –ID3 – CART - Ensemble Methods –Random Forest - Evaluation of Classification Algorithms

UNIT III UNSUPERVISED LEARNING AND REINFORCEMENT LEARNING 9

Introduction - Clustering Algorithms -K – Means – Hierarchical Clustering - Cluster Validity - Dimensionality Reduction –Introduction -Principal Component Analysis – Recommendation Systems - EM algorithm. Reinforcement Learning – Elements -Model based Learning – Temporal Difference Learning

UNIT IV PROBABILISTIC METHODS FOR LEARNING 9

Introduction -Naïve Bayes Algorithm -Maximum Likelihood -Maximum Apriori -Bayesian Belief Networks -Probabilistic Modelling of Problems -Inference in Bayesian Belief Networks – Probability Density Estimation - Sequence Models – Markov Models – Hidden Markov Models

UNIT V NEURAL NETWORKS AND DEEP LEARNING 9

Neural Networks – Biological Motivation- Perceptron – Multi-layer Perceptron – Feed Forward Network – BackPropagation-Activation and Loss Functions- Limitations of Machine Learning – Deep Learning – introduction – Convolution Neural Networks – Recurrent Neural Networks – LSTM- Use cases

SUGGESTED ACTIVITIES:

1. Give a new example from our daily life for each type of Machine Learning problem.
2. Study at least 3 open source tools/frameworks available for Machine Learning and discuss the pros and cons of each tool/framework.
3. Take an example of a classification problem. Implement a Random Forest and visualize the individual trees to understand how the model works.
4. Examine the various cross disciplinary use cases of Machine Learning. Eg Machine Learning in Bioinformatics, Climate Science, Economics, etc.
5. Outline 10 machine learning applications in healthcare.
6. Discuss the recent advancements in Reinforcement Learning and why research in Reinforcement Learning is hard.
7. Discuss case studies on the ethical issues that have gained traction in recent years due to bulk collection of data. Eg. Racial profiling, Cambridge Analytica.
8. Give 5 examples where sequential models are suitable.
9. Discuss recent CNN architectures.

COURSE OUTCOMES:

CO1: Understand and outline problems for each type of machine learning

CO2: Design a Decision tree and Random forest for an application

CO3: Implement Probabilistic Discriminative and Generative algorithms for an application and analyze the results.

CO4: Use a tool to implement typical Clustering algorithms for different types of applications.

CO5: Design and implement an HMM for a Sequence Model type of application.

CO6: Identify applications suitable for different types of Machine Learning with suitable justification.

PRACTICAL EXERCISES:

TOTAL : 30 PERIODS

1. Implement a Linear Regression with a Real Dataset (<https://www.kaggle.com/harrywang/housing>). Experiment with different features in building a model. Tune the model's hyperparameters.
2. Implement a binary classification model. That is, answers a binary question such as "Are houses in this neighborhood above a certain price?" (use data from exercise 1). Modify the classification threshold and determine how that modification influences the model. Experiment with different classification metrics to determine your model's effectiveness.
3. Classification with Nearest Neighbours. In this question, you will use the scikit-learn's KNN classifier to classify real vs. fake news headlines. The aim of this question is for you to read

the scikit-learn API and get comfortable with training/validation splits. Use California Housing Dataset

4. In this exercise, you'll experiment with validation sets and test sets using the dataset. Split a training set into a smaller training set and a validation set. Analyze deltas between training set and validation set results. Test the trained model with a test set to determine whether your trained model is overfitting. Detecting and fixing a common training problem.

5. Implement the k-means algorithm using

<https://archive.ics.uci.edu/ml/datasets/Codon+usage> dataset

6. Implement the Naïve Bayes Classifier using

<https://archive.ics.uci.edu/ml/datasets/Gait+Classification> dataset

7. Project - (in Pairs) Your project must implement one or more machine learning algorithms and apply them to some data.

a. Your project may be a comparison of several existing algorithms, or it may propose a new algorithm in which case you still must compare it to at least one other approach.

b. You can either pick a project of your own design, or you can choose from the set of pre defined projects.

c. You are free to use any third-party ideas or code that you wish as long as it is publicly available.

d. You must properly provide references to any work that is not your own in the write-up.

e. Project proposal You must turn in a brief project proposal. Your project proposal should describe the idea behind your project. You should also briefly describe software you will need to write, and papers (2-3) you plan to read.

List of Projects (datasets available)

1. Sentiment Analysis of Product Reviews

2. Stock Prediction

3. Sales Forecasting

4. Music Recommendation

5. Handwritten Digit Classification

6. Fake News Detection

7. Sports Prediction

8. Object Detection

9. Disease Prediction

Hardware/Software requirements

1. Python 3.x

2. JupyterLab

3. Scientific Computing Libraries: Numpy, JAX, Matplotlib

4. Machine Learning Libraries: Scikit-Learning, Turi Create
5. Deep Learning Libraries: Pytorch 1.0, Tensorflow 2.0, TRAX, DyNet
6. Weka, Wekatorator
7. Cloud (for Deep Learning): Google Colab, Paperspace Gradient
8. Intel Core i7 9700K or Ryzen 7 5800X CPU, with minimum 16GB RAM
9. Etc.

TOTAL : 75 PERIODS

COURSE OUTCOMES:

CO1: Understand and outline problems for each type of machine learning

CO2: Design a Decision tree and Random forest for an application

CO3: Implement Probabilistic Discriminative and Generative algorithms for an application and analyze the results.

CO4: Use a tool to implement typical Clustering algorithms for different types of applications.

CO5: Design and implement an HMM for a Sequence Model type of application.

CO6: Identify applications suitable for different types of Machine Learning with suitable justification.

REFERENCES:

1. Probabilistic Machine Learning: An Introduction by Kevin Murphy, MIT Press 2022. <https://probml.github.io/pml-book/book1.html>
2. Deep Learning by Ian Goodfellow, Yoshua Bengio and Aaron Courville, MIT Press 2016
3. Dive into Deep Learning by Aston Zhang, Zack C. Lipton, Mu Li and Alex J. Smola <https://d2l.ai/>
4. Kevin Murphy, "Machine Learning: A Probabilistic Perspective", MIT Press, 2012
5. Ethem Alpaydin, "Introduction to Machine Learning", Third Edition, Adaptive Computation and Machine Learning Series, MIT Press, 2014
6. Tom M Mitchell, "Machine Learning", McGraw Hill Education, 2013
7. Peter Flach, "Machine Learning: The Art and Science of Algorithms that Make Sense of Data", First Edition, Cambridge University Press, 2012.
8. Shai Shalev-Shwartz and Shai Ben-David, "Understanding Machine Learning: From Theory to Algorithms", Cambridge University Press, 2015
9. Christopher Bishop, "Pattern Recognition and Machine Learning", Springer, 2007.
10. Hal Daumé III, "A Course in Machine Learning", 2017 (freely available online)
11. Trevor Hastie, Robert Tibshirani, Jerome Friedman, "The Elements of Statistical Learning", Springer, Second Edition, 2009. (freely available online).
12. Stephen Marsland, "Machine Learning - An Algorithmic Perspective", Chapman and Hall/CRC Press, Second Edition, 2014

COURSE OBJECTIVES

- Describe the fundamental elements of relational database management systems
- Explain the basic concepts of relational data model, entity-relationship model, relational database design, relational algebra and SQL.
- Understand query processing in a distributed database system
- Understand the basics of XML and create well-formed and valid XML documents.
- Distinguish the different types of NoSQL databases
- To understand the different models involved in database security and their applications in real time world to protect the database and information associated with them.

UNIT I RELATIONAL DATA MODEL**12**

Entity Relationship Model – Relational Data Model – Mapping Entity Relationship Model to Relational Model – Relational Algebra – Structured Query Language – Database Normalization.

Suggested Activities:**Data Definition Language**

- Create, Alter and Drop
- Enforce Primary Key, Foreign Key, Check, Unique and Not Null Constraints
- Creating Views

Data Manipulation Language

- Insert, Delete, Update
- Cartesian Product, Equi Join, Left Outer Join, Right Outer Join and Full Outer Join
- Aggregate Functions
- Set Operations
- Nested Queries

Transaction Control Language

- Commit, Rollback and Save Points

UNIT II DISTRIBUTED DATABASES, ACTIVE DATABASES AND OPEN DATABASE CONNECTIVITY**12**

Distributed Database Architecture – Distributed Data Storage – Distributed Transactions – Distributed Query Processing – Distributed Transaction Management – Event Condition Action Model – Design and Implementation Issues for Active Databases – Open Database Connectivity.

Suggested Activities:

- Distributed Database Design and Implementation
- Row Level and Statement Level Triggers
- Accessing a Relational Database using PHP, Python and R

UNIT III XML DATABASES

12

Structured, Semi structured, and Unstructured Data – XML Hierarchical Data Model – XML Documents – Document Type Definition – XML Schema – XML Documents and Databases – XML Querying – XPath – XQuery

Suggested Activities:

- Creating XML Documents, Document Type Definition and XML Schema
- Using a Relational Database to store the XML documents as text
- Using a Relational Database to store the XML documents as data elements
- Creating or publishing customized XML documents from pre-existing relational databases
- Extracting XML Documents from Relational Databases
- XML Querying

UNIT IV NOSQL DATABASES AND BIG DATA STORAGE SYSTEMS

12

NoSQL – Categories of NoSQL Systems – CAP Theorem – Document-Based NoSQL Systems and MongoDB – MongoDB Data Model – MongoDB Distributed Systems Characteristics – NoSQL Key-Value Stores – DynamoDB Overview – Voldemort Key-Value Distributed Data Store – Wide Column NoSQL Systems – Hbase Data Model – Hbase Crud Operations – Hbase Storage and Distributed System Concepts – NoSQL Graph Databases and Neo4j – Cypher Query Language of Neo4j – Big Data – MapReduce – Hadoop – YARN.

Suggested Activities:

- Creating Databases using MongoDB, DynamoDB, Voldemort Key-Value Distributed Data Store Hbase and Neo4j.
- Writing simple queries to access databases created using MongoDB, DynamoDB, Voldemort Key-Value Distributed Data Store Hbase and Neo4j.

UNIT V DATABASE SECURITY

12

Database Security Issues – Discretionary Access Control Based on Granting and Revoking Privileges – Mandatory Access Control and Role-Based Access Control for Multilevel Security – SQL Injection – Statistical Database Security – Flow Control – Encryption and Public Key Infrastructures – Preserving Data Privacy – Challenges to Maintaining Database Security – Database Survivability – Oracle Label-Based Security.

Suggested Activities:

Implementing Access Control in Relational Databases

TOTAL : 75 PERIODS

COURSE OUTCOMES

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

- Convert the ER-model to relational tables, populate relational database and formulate SQL queries on data.

- Understand and write well-formed XML documents
- Be able to apply methods and techniques for distributed query processing.
- Design and Implement secure database systems.
- Use the data control, definition, and manipulation languages of the NoSQL databases

REFERENCES:

1. R. Elmasri, S.B. Navathe, "Fundamentals of Database Systems", Seventh Edition, Pearson Education 2016.
2. Henry F. Korth, Abraham Silberschatz, S. Sudharshan, "Database System Concepts", Seventh Edition, McGraw Hill, 2019.
3. C.J.Date, A.Kannan, S.Swamynathan, "An Introduction to Database Systems, Eighth Edition, Pearson Education, 2006
4. Raghu Ramakrishnan, Johannes Gehrke "Database Management Systems", Fourth Edition, McGraw Hill Education, 2015.
5. Harrison, Guy, "Next Generation Databases, NoSQL and Big Data", First Edition, Apress publishers, 2015
6. Thomas Cannolly and Carolyn Begg, "Database Systems, A Practical Approach to Design, Implementation and Management", Sixth Edition, Pearson Education, 2015

www.binils.com

CP4161

**ADVANCED DATA STRUCTURES AND ALGORITHMS
LABORATORY**

**L T P C
0 0 4 2**

OBJECTIVES:

- To acquire the knowledge of using advanced tree structures
- To learn the usage of heap structures
- To understand the usage of graph structures and spanning trees
- To understand the problems such as matrix chain multiplication, activity selection and Huffman coding
- To understand the necessary mathematical abstraction to solve problems.

LIST OF EXPERIMENTS:

- 1: Implementation of recursive function for tree traversal and Fibonacci
- 2: Implementation of iteration function for tree traversal and Fibonacci
- 3: Implementation of Merge Sort and Quick Sort
- 4: Implementation of a Binary Search Tree
- 5: Red-Black Tree Implementation
- 6: Heap Implementation

- 7: Fibonacci Heap Implementation
- 8: Graph Traversals
- 9: Spanning Tree Implementation
- 10: Shortest Path Algorithms (Dijkstra's algorithm, Bellman Ford Algorithm)
- 11: Implementation of Matrix Chain Multiplication
- 12: Activity Selection and Huffman Coding Implementation

HARDWARE/SOFTWARE REQUIREMENTS

- 1: 64-bit Open source Linux or its derivative
- 2: Open Source C++ Programming tool like G++/GCC

TOTAL : 60 PERIODS

COURSE OUTCOMES:

CO1: Design and implement basic and advanced data structures extensively

CO2: Design algorithms using graph structures

CO3: Design and develop efficient algorithms with minimum complexity using design techniques

CO4: Develop programs using various algorithms.

CO5: Choose appropriate data structures and algorithms, understand the ADT/libraries, and use it to design algorithms for a specific problem.

REFERENCES:

1. Lipschutz Seymour, "Data Structures Schaum's Outlines Series", Tata McGraw Hill, 3rd Edition, 2014.
2. Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman, "Data Structures and Algorithms", Pearson Education, Reprint 2006.
3. <http://www.coursera.org/specializations/data-structures-algorithms>
4. http://www.tutorialspoint.com/data_structures_algorithms
5. <http://www.geeksforgeeks.org/data-structures/>

BD4111

BIG DATA COMPUTING LABORATORY

L T P C

0 0 2 1

COURSE OBJECTIVES:

- To set up single and multi-node Hadoop Clusters.
- To solve Big Data problems using Map Reduce Technique.
- To learn NoSQL queries.
- To design algorithms that uses Map Reduce Technique to apply on Unstructured and structured data.
- To learn Scalable machine learning using Mahout.

LIST OF EXPERIMENTS:

1. Set up a pseudo-distributed, single-node Hadoop cluster backed by the Hadoop Distributed File System, running on Ubuntu Linux. After successful installation on one node, configuration of a multi-node Hadoop cluster (one master and multiple slaves).
2. MapReduce application for word counting on Hadoop cluster.
3. Unstructured data into NoSQL data and do all operations such as NoSQL query with API.
4. K-means clustering using map reduce.
5. Page Rank Computation.
6. Mahout machine learning library to facilitate the knowledge build up in big data analysis.
7. Application of Recommendation Systems using Hadoop/mahout libraries.

HARDWARE/SOFTWARE REQUIREMENTS

- 1: Java
- 2: Hadoop
- 3: Mahout
- 4: HBase/MongoDB

COURSE OUTCOMES:

CO1: Set up single and multi-node Hadoop Clusters.

CO2: Apply Map Reduce technique for various algorithms.

CO3: Design new algorithms that use Map Reduce to apply on Unstructured and structured data.

CO4: Develop Scalable machine learning algorithms for various Big data applications using Mahout.

CO5: Represent NoSQL data.

TOTAL: 30 PERIODS

REFERENCES:

1. Kristina Chodorow, "MongoDB: The Definitive Guide – Powerful and Scalable Data Storage", O'Reilly, 3rd Edition, 2019.
2. Lars George, "HBase: The Definitive Guide", O'Reilly, 2015.
3. Tom White, "Hadoop: The Definitive Guide – Storage and Analysis at Internet Scale", O'Reilly, 4th Edition, 2015.
4. Robin Anil, Sean Owen, Ellen G. Friedman, Ted Dunning, "Mahout in Action", Manning Publications, 2011.

AUDIT COURSES

AX4091

ENGLISH FOR RESEARCH PAPER WRITING

L T P C

2 0 0 0

COURSE OBJECTIVES:

- Teach how to improve writing skills and level of readability
- Tell about what to write in each section
- Summarize the skills needed when writing a Title
- Infer the skills needed when writing the Conclusion
- Ensure the quality of paper at very first-time submission

UNIT I INTRODUCTION TO RESEARCH PAPER WRITING 6

Planning and Preparation, Word Order, Breaking up long sentences, Structuring Paragraphs and Sentences, Being Concise and Removing Redundancy, Avoiding Ambiguity and Vagueness

UNIT II PRESENTATION SKILLS 6

Clarifying Who Did What, Highlighting Your Findings, Hedging and Criticizing, Paraphrasing and Plagiarism, Sections of a Paper, Abstracts, Introduction

UNIT III TITLE WRITING SKILLS 6

Key skills are needed when writing a Title, key skills are needed when writing an Abstract, key skills are needed when writing an Introduction, skills needed when writing a Review of the Literature, Methods, Results, Discussion, Conclusions, The Final Check

UNIT IV RESULT WRITING SKILLS 6

Skills are needed when writing the Methods, skills needed when writing the Results, skills are needed when writing the Discussion, skills are needed when writing the Conclusions

UNIT V VERIFICATION SKILLS 6

Useful phrases, checking Plagiarism, how to ensure paper is as good as it could possibly be the first- time submission

TOTAL: 30 PERIODS

COURSE OUTCOMES:

- CO1 –Understand that how to improve your writing skills and level of readability
- CO2 – Learn about what to write in each section
- CO3 – Understand the skills needed when writing a Title
- CO4 – Understand the skills needed when writing the Conclusion
- CO5 – Ensure the good quality of paper at very first-time submission

REFERENCES:

1. Adrian Wallwork , English for Writing Research Papers, Springer New York Dordrecht Heidelberg London, 2011
2. Day R How to Write and Publish a Scientific Paper, Cambridge University Press 2006
3. Goldbort R Writing for Science, Yale University Press (available on Google Books) 2006
4. Highman N, Handbook of Writing for the Mathematical Sciences, SIAM. Highman's book 1998.

AX4092

DISASTER MANAGEMENT

L T P C

2 0 0 0

COURSE OBJECTIVES:

- Summarize basics of disaster
- Explain a critical understanding of key concepts in disaster risk reduction and humanitarian response.
- Illustrate disaster risk reduction and humanitarian response policy and practice from multiple perspectives.
- Describe an understanding of standards of humanitarian response and practical relevance in specific types of disasters and conflict situations.
- Develop the strengths and weaknesses of disaster management approaches

UNIT I INTRODUCTION

6

Disaster: Definition, Factors and Significance; Difference between Hazard And Disaster; Natural and Manmade Disasters: Difference, Nature, Types and Magnitude.

UNIT II REPERCUSSIONS OF DISASTERS AND HAZARDS

6

Economic Damage, Loss of Human and Animal Life, Destruction Of Ecosystem. Natural Disasters: Earthquakes, Volcanisms, Cyclones, Tsunamis, Floods, Droughts And Famines, Landslides And Avalanches, Man-made disaster: Nuclear Reactor Meltdown, Industrial Accidents, Oil Slicks And Spills, Outbreaks Of Disease And Epidemics, War And Conflicts.

UNIT III DISASTER PRONE AREAS IN INDIA

6

Study of Seismic Zones; Areas Prone To Floods and Droughts, Landslides And Avalanches; Areas Prone To Cyclonic and Coastal Hazards with Special Reference To Tsunami; Post-Disaster Diseases and Epidemics

UNIT IV DISASTER PREPAREDNESS AND MANAGEMENT

6

Preparedness: Monitoring Of Phenomena Triggering a Disaster or Hazard; Evaluation of Risk: Application of Remote Sensing, Data from Meteorological And Other Agencies, Media Reports: Governmental and Community Preparedness.

UNIT V RISK ASSESSMENT

6

Disaster Risk: Concept and Elements, Disaster Risk Reduction, Global and National Disaster Risk Situation. Techniques of Risk Assessment, Global Co-Operation in Risk Assessment and Warning, People's Participation in Risk Assessment. Strategies for Survival

TOTAL : 30 PERIODS

COURSE OUTCOMES:

CO1: Ability to summarize basics of disaster

CO2: Ability to explain a critical understanding of key concepts in disaster risk reduction and humanitarian response.

CO3: Ability to illustrate disaster risk reduction and humanitarian response policy and practice from multiple perspectives.

CO4: Ability to describe an understanding of standards of humanitarian response and practical relevance in specific types of disasters and conflict situations.

CO5: Ability to develop the strengths and weaknesses of disaster management approaches

REFERENCES:

1. Goel S. L., Disaster Administration And Management Text And Case Studies", Deep & Deep Publication Pvt. Ltd., New Delhi, 2009.
2. Nishitha Rai, Singh AK, "Disaster Management in India: Perspectives, issues and strategies " New Royal book Company, 2007.
3. Sahni, Pradeep Et. Al. , " Disaster Mitigation Experiences And Reflections", Prentice Hall Of India, New Delhi, 2001.

AX4093

CONSTITUTION OF INDIA

L T P C

2 0 0 0

COURSE OBJECTIVES:

Students will be able to:

- Understand the premises informing the twin themes of liberty and freedom from a civil rights perspective.
- To address the growth of Indian opinion regarding modern Indian intellectuals' constitutional
- Role and entitlement to civil and economic rights as well as the emergence of nationhood in the early years of Indian nationalism.
- To address the role of socialism in India after the commencement of the Bolshevik Revolution 1917 And its impact on the initial drafting of the Indian Constitution.

UNIT I HISTORY OF MAKING OF THE INDIAN CONSTITUTION

History, Drafting Committee, (Composition & Working)

UNIT II PHILOSOPHY OF THE INDIAN CONSTITUTION

Preamble, Salient Features

UNIT III CONTOURS OF CONSTITUTIONAL RIGHTS AND DUTIES

Fundamental Rights, Right to Equality, Right to Freedom, Right against Exploitation, Right to Freedom of Religion, Cultural and Educational Rights, Right to Constitutional Remedies, Directive Principles of State Policy, Fundamental Duties.

UNIT IV ORGANS OF GOVERNANCE

Parliament, Composition, Qualifications and Disqualifications, Powers and Functions, Executive, President, Governor, Council of Ministers, Judiciary, Appointment and Transfer of Judges, Qualifications, Powers and Functions.

UNIT V LOCAL ADMINISTRATION

District's Administration head: Role and Importance, □Municipalities: Introduction, Mayor and role of Elected Representative, CEO, Municipal Corporation. Pachayati raj: Introduction, PRI: Zila Panchayat. Elected officials and their roles, CEO Zila Pachayat: Position and role. Block level: Organizational Hierarchy(Different departments), Village level:Role of Elected and Appointed officials, Importance of grass root democracy.

UNIT VI ELECTION COMMISSION

Election Commission: Role and Functioning. Chief Election Commissioner and Election Commissioners - Institute and Bodies for the welfare of SC/ST/OBC and women.

TOTAL: 30 PERIODS

COURSE OUTCOMES:

Students will be able to:

1. Discuss the growth of the demand for civil rights in India for the bulk of Indians before the arrival of Gandhi in Indian politics.
2. Discuss the intellectual origins of the framework of argument that informed the conceptualization of social reforms leading to revolution in India.
3. Discuss the circumstances surrounding the foundation of the Congress Socialist Party[CSP] under the leadership of Jawaharlal Nehru and the eventual failure of the proposal of direct elections through adult suffrage in the Indian Constitution.
4. Discuss the passage of the Hindu Code Bill of 1956.

SUGGESTED READING

1. The Constitution of India,1950(Bare Act),Government Publication.
2. Dr.S.N.Busi, Dr.B. R.Ambedkar framing of Indian Constitution,1st Edition, 2015.
3. M.P. Jain, Indian Constitution Law, 7th Edn., LexisNexis,2014.
4. D.D. Basu, Introduction to the Constitution of India, LexisNexis, 2015.

UNIT I

சங்க இலக்கியம்

1.தமிழின் துவக்க நூல் தொல்காப்பியம்

- எழுத்து, சொல், பொருள்

2.அகநானூறு (82)

- இயற்கை இன்னிசை அரங்கம்

3.குறிஞ்சிப் பாட்டின் மலர்க்காட்சி

4.புறநானூறு (95,195)

- போரை நிறுத்திய ஔவையார்

UNIT II

அறநெறித் தமிழ்

6

1. அறநெறி வகுத்த திருவள்ளுவர்

- அறம் வலியுறுத்தல், அன்புடைமை, ஒப்புறவு அறிதல், ஈகை, புகழ்

2. பிற அறநூல்கள் - இலக்கிய மருந்து

- ஏலாதி, சிறுபஞ்சமூலம், திரிகடுகம், ஆசாரக்கோவை (தூய்மையை வலியுறுத்தும் நூல்)

UNIT III

இரட்டைக் காப்பியங்கள்

6

1. கண்ணகியின் புரட்சி

- சிலப்பதிகார வழக்குரை காதை

2. சமூகசேவை இலக்கியம் மணிமேகலை

- சிறைக்கோட்டம் அறக்கோட்டமாகிய காதை

UNIT IV

அருள்நெறித் தமிழ்

6

1. சிறுபாணாற்றுப்படை

- பாரி முல்லைக்குத் தேர் கொடுத்தது, பேகன் மயிலுக்குப்

போர்வை கொடுத்தது, அதியமான் ஔவைக்கு நெல்லிக்கனி கொடுத்தது, அரசர் பண்புகள்

2. நற்றிணை

- அன்னைக்குரிய புன்னை சிறப்பு

3. திருமந்திரம் (617, 618)

- இயமம் நியமம் விதிகள்

4. தர்மச்சாலையை நிறுவிய வள்ளலார்

5. புறநானூறு

- சிறுவனே வள்ளலானான்

6. அகநானூறு (4) - வண்டு
நற்றிணை (11) - நண்டு
கலித்தொகை (11) - யானை, புறா
ஐந்திணை 50 (27) - மான்

ஆகியவை பற்றிய செய்திகள்

UNIT V நவீன தமிழ் இலக்கியம்

6

1. உரைநடைத் தமிழ்,
 - தமிழின் முதல் புதினம்,
 - தமிழின் முதல் சிறுகதை,
 - கட்டுரை இலக்கியம்,
 - பயண இலக்கியம்,
 - நாடகம்,
2. நாட்டு விடுதலை போராட்டமும் தமிழ் இலக்கியமும்,
3. சமுதாய விடுதலையும் தமிழ் இலக்கியமும்,
4. பெண் விடுதலையும் விளிம்பு நிலையினரின் மேம்பாட்டில் தமிழ் இலக்கியமும்,
5. அறிவியல் தமிழ்,
6. இணையத்தில் தமிழ்,
7. சுற்றுச்சூழல் மேம்பாட்டில் தமிழ் இலக்கியம்.

TOTAL: 30 PERIODS

தமிழ் இலக்கிய வெளியீடுகள் / புத்தகங்கள்

1. தமிழ் இணைய கல்விக்கழகம் (Tamil Virtual University)
 - www.tamilvu.org
2. தமிழ் விக்கிப்பீடியா (Tamil Wikipedia)
 - <https://ta.wikipedia.org>
3. தர்மபுர ஆதின வெளியீடு
4. வாழ்வியல் களஞ்சியம்
 - தமிழ்ப் பல்கலைக்கழகம், தஞ்சாவூர்
5. தமிழ்கலைக் களஞ்சியம்
 - தமிழ் வளர்ச்சித் துறை (thamilvalarchithurai.com)
6. அறிவியல் களஞ்சியம்
 - தமிழ்ப் பல்கலைக்கழகம், தஞ்சாவூர்