Name of Institute: Institute of Engineering and Technology
Name of Faculty: Prof.Bhavin Fataniya

Course code: DS0101
Course name: Distributed Operating System
Pre-requisites: Operating System
Credit points: 5
Offered Semester: 1

Course coordinator (weeks XX - XX)
Full name:Prof.Bhavin Fataniya
Department with siting location: Computer Department 4th Floor Bhanvar Building
Telephone:-
Email:bhavinfataniya.ce@indusuni.ac.in
Consultation times: Monday to Friday **03:pm to 5:00pm**

Course lecturer (weeks xx - XX)
Full name:Prof.Bhavin Fataniya
Department with siting location:Computer Department 4th Floor Bhanvar Building
Telephone:-
Email:bhavinfataniya.ce@indusuni.ac.in
Consultation times: Monday to Friday **03:pm to 5:00pm**

Students will be contacted throughout the session via mail with important information relating to this course.

Course Objective

1. List the principles of distributed systems and describe the problems and challenges associated with these principles.
2. Understand Distributed Computing techniques, Synchronous and Processes.
3. Apply Shared Data access and Files concepts.
4. Design a distributed system that fulfills requirements with regards to key distributed systems
5. properties.
6. Understand Distributed File Systems and Distributed Shared Memory.
7. Apply Distributed web-based system.
8. Understand the importance of security in distributed systems

Course Outcomes (CO)

After successful completion of the course, student will able:

1. List the principles of distributed systems and describe the problems and challenges associated with these principles.
2. Understand Distributed Computing techniques, Synchronous and Processes.
3. Apply Shared Data access and Files concepts.
4. Design a distributed system that fulfills requirements with regards to key distributed systems properties.
5. Understand Distributed File Systems and Distributed Shared Memory.
6. Apply Distributed web-based system.
7. Understand the importance of security in distributed systems

Method of delivery:

Lecture and PPT

Study time

Students must attend three lectures per week.

CO-PO Mapping (PO: Program Outcomes)

Engineering Graduates will be able to:

- **PO1. Engineering knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
- **PO2. Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
- **PO3. Design/development of solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
- **PO4. Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis, and interpretation of data, and synthesis of the information to provide valid conclusions.
- **PO5. 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.
- **PO**6. **The engineer and society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
- **PO7. Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
- **PO8. Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
- **PO9. Individual and team work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

- **PO10. Communication**: 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.
- **PO11. Project management and finance**: Demonstrate knowledge and understanding of the 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.
- **PO12. Life-long learning**: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Mapping of CO with PO's

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO110	PO111	PO112
1	√											
2	√		√	√	√							
3		√	√	√	√							
4		√	√	√	√							
5		√	√	√	√							
6			√	√								
7			√	√	√							

Blooms Taxonomy and Knowledge retention (For reference)

(Blooms taxonomy has been given for reference)

Figure 1: Blooms Taxonomy

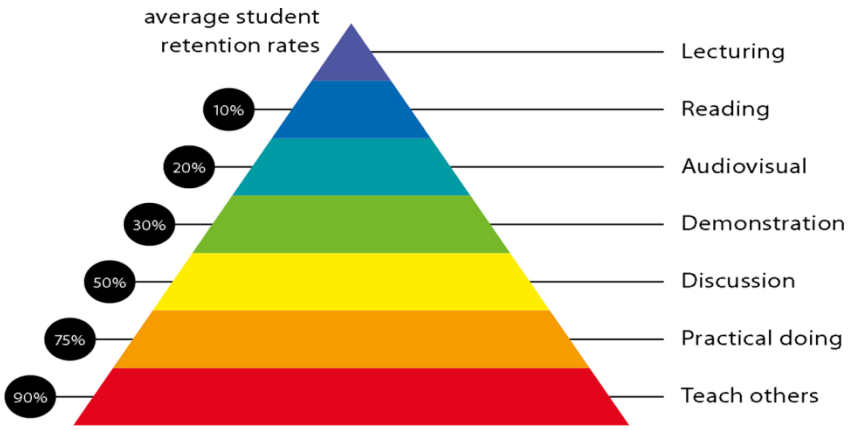


Figure 2: Knowledge retention

Graduate Qualities and Capabilities covered

(Qualities graduates harness crediting this Course)

General Graduate Qualities	Specific Department of _____Graduate Capabilities
Informed Have a sound knowledge of an area of study or profession and understand its current issues, locally and internationally. Know how to apply this knowledge. Understand how an area of study has developed and how it relates to other areas.	**1 Professional knowledge, grounding & awareness**
Independent learners Engage with new ideas and ways of thinking and critically analyze issues. Seek to extend knowledge through ongoing research, inquiry and reflection. Find and evaluate information, using a variety of sources and technologies. Acknowledge the work and ideas of others.	**2 Information literacy, gathering & processing**
Problem solvers Take on challenges and opportunities. Apply creative, logical and critical thinking skills to respond effectively. Make and implement decisions. Be flexible, thorough, innovative and aim for high standards.	**4 Problem-solving skills**
Effective communicators Articulate ideas and convey them effectively using a range of media. Work collaboratively and engage with people in different settings. Recognize how culture can shape communication.	**5 Written communication** **6 Oral communication** **7 Teamwork**

Responsible Understand how decisions can affect others and make ethically informed choices. Appreciate and respect diversity. Act with integrity as part of local, national, global and professional communities.	**10 Sustainability, societal & environmental impact**

Practical work:

Exp No.	Title	Learning Outcomes
1.1	Write a Program to implement Concurrent Echo Client Server Application.	Basic Knowledge of client server model.
2.1	Write the Programs for Remote Procedure call.	Basic Knowledge of RPC.
3.1	Write the Programs for Remote Method Invocation.	Basic Knowledge of RMI.
4.1	Write the Programs for Thread Programming in JAVA	Basic Knowledge of Thread
4.2	Implement CORBA file.	Basic of CORBA
5.1	Write a Program to Increment a Counter in Shared Memory..	Basic Knowledge of Shared Memory
5.2	Implement Network File System (NFS).	Basic Knowledge of Shared Memory
5.3	Creation of a BPEL (Business Process Execution Language) Module and a Composite Application.	Basic Knowledge of Shared Memory
5.4	Study of Grid Services using various Tools.	Basic Knowledge of Shared Memory

Lecture/tutorial times

(Give lecture times in the format below)

Example:

Lecture Tuesday Room LH 30
Lecture/Tutorial Wednesday Room LH 30

Attendance Requirements

The University norms state that it is the responsibility of students to attend all lectures, tutorials, seminars, and practical work as stipulated in the course outline. Minimum attendance requirement as per university norms is compulsory for being eligible for semester examinations.

ASSESSMENT GUIDELINES

Your final course mark will be calculated from the following:

Activity	Marks
Class Test	30
Regularity	10
Presentation	10
Class Attention	10

SUPPLEMENTARY ASSESSMENT

Students who receive an overall mark less than 40% in internal component or less than 40% in the end semester will be considered for supplementary assessment in the respective components (i.e internal component or end semester) of semester concerned. Students must make themselves available during the supplementary examination period to take up the respective components (internal component or end semester) and need to obtain the required minimum 40% marks to clear the concerned components.

Practical Work Report/Laboratory Report:

A report on the practical work is due the subsequent week after completion of the class by each group.

Late Work

Late assignments will not be accepted without supporting documentation. Late submission of the reports will result in a deduction of -% of the maximum mark per calendar day

Format

All assignments must be presented in a neat, legible format with all information sources correctly referenced. **Assignment material handed in throughout the session that is not neat and legible will not be marked and will be returned to the student.**

Retention of Written Work

Written assessment work will be retained by the Course coordinator/lecturer for two weeks after marking to be collected by the students.

University and Faculty Policies

Students should make themselves aware of the University and/or Faculty Policies regarding plagiarism, special consideration, supplementary examinations and other educational issues and student matters.

Plagiarism - Plagiarism is not acceptable and may result in the imposition of severe penalties. Plagiarism is the use of another person's work, or idea, as if it is his or her own - if you have any doubts at all on what constitutes plagiarism, please consult your Course coordinator or lecturer. Plagiarism will be penalized severely.

Do not copy the work of other students.
Do not share your work with other students (except where required for a group activity or assessment)

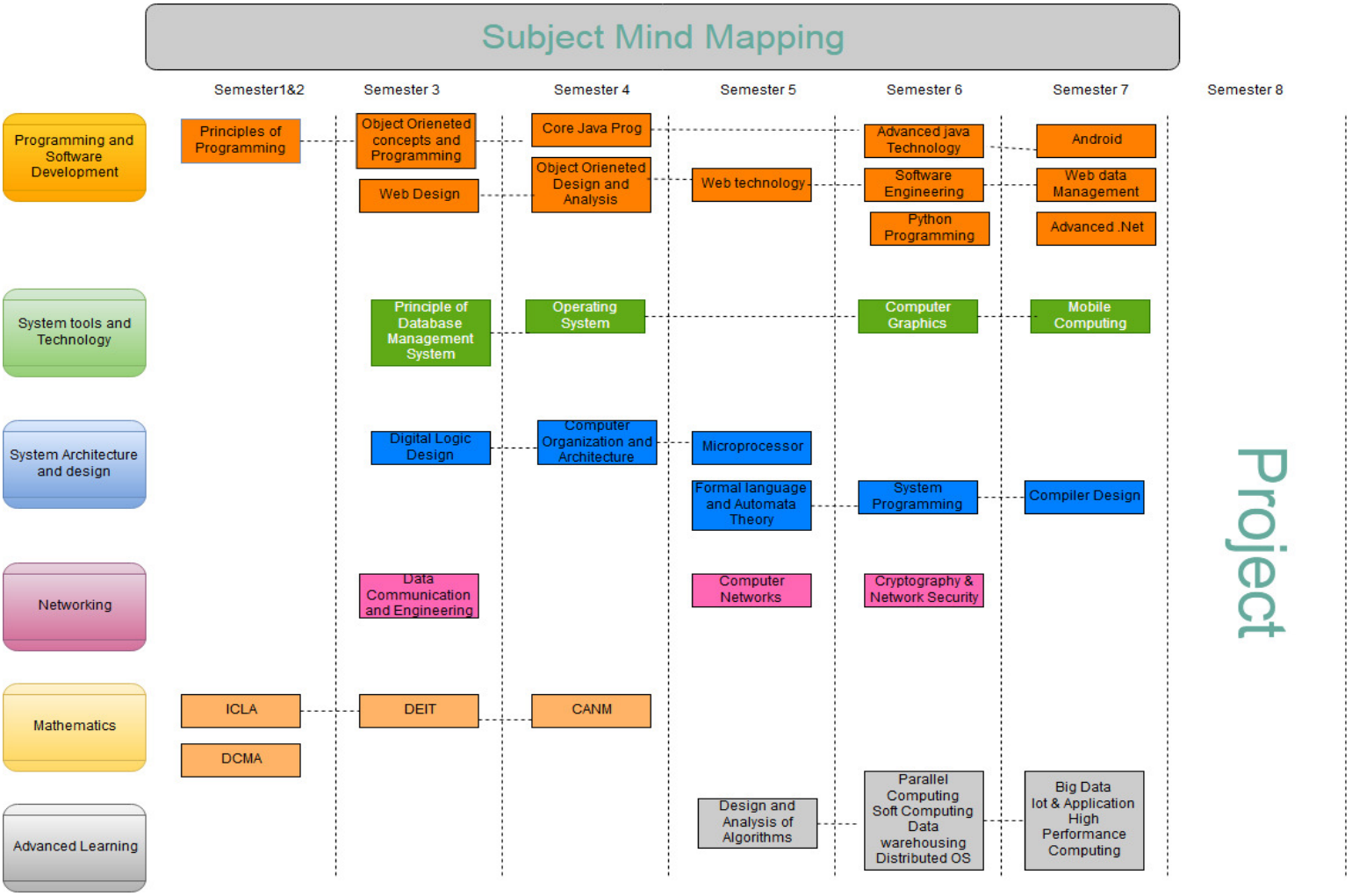
.

Course schedule(subject to change)

(Mention quiz, assignment submission, breaksetcas well in the table under the Teaching Learning Activity Column)

	Week #	Topic & contents	CO Addressed	Teaching Learning Activity (TLA)
	Weeks 1	**Introduction to distributed Systems:** Definition and goals, Hardware and Software concepts, Design issues	1	Lectures
	Weeks 2	**Communication in Distributed System:** Computer Network and Layered protocols, Message passing and related issues, synchronization,	1	Lectures
	Week 3	Client Server model & its implementation, remote procedure call and implementation issues, Case Studies: SUN RPC, DEC RPC	*1*	Lectures
	Week 4	**Synchronization in distributed systems:** Clock synchronization and related algorithms, mutual exclusion, Deadlock in distributed systems	1,2	Lectures,ppt
	Week 5	**Processes and processors in distributed systems:** Threads, system model, processor allocation, scheduling in distributed systems	1,2	Lectures,Assignment,Case study
	Week 6	Load balancing and sharing approach, fault tolerance, Real time distributed systems, Process migration and related issues.	*1,2*	Lectures, Case study
	Week 7	**Distributed File Systems:** Introduction, features & goal of distributed file system, file models, file accessing models, file sharing semantics, file caching scheme, file replication, fault tolerance, trends in distributed file system	4	Lectures
	Week 8	**Distributed Shared Memory:** Introduction, general architecture of DSM systems,	3,4	Lectures
	Week 9	design and implementation issues of DSM, granularity, structure of shared memory space, consistency models, replacement strategy, thrashing	*3,4*	Lectures
	Week 10	**Naming** Overview, Features, Basic concepts, System oriented names, Object locating mechanisms, Issues in designing human oriented names, Name caches, Naming and security, DNS	5,7	Lectures,ppt

	Week 11	**Distributed Web based Systems** Architecture, Processes, Communication, Naming, Synchronization, Consistency and Replication: Web Proxy Caching, Replication for Web Hosting Systems, Replication of Web Applications	5,6	Lectures
	Week 12	**Security**	6,7	Lectures, Assignment,Case study


Subject Mind Mapping
Semester1&2
Semester 3
Semester 4
Semester 5
Semester 6
Semester 7
Semester 8
Programming and Software Development
System tools and Technology
System Architecture and design
Networking
Mathematics
Advanced Learning
Principles of Programming
Object Orieneted concepts and Programming
Web Design
Core Java Prog
Object Orieneted Design and Analysis
Web technology
Advanced java Technology
Software Engineering
Python Programming
Android
Web data Management
Advanced .Net
Principle of Database Management System
Operating System
Computer Graphics
Mobile Computing
Digital Logic Design
Computer Organization and Architecture
Microprocessor
Formal language and Automata Theory
System Programming
Compiler Design
Data Communication and Engineering
Computer Networks
Cryptography & Network Security
ICLA
DCMA
DEIT
CANM
Design and Analysis of Algorithms
Parallel Computing
Soft Computing
Data warehousing
Distributed OS
Big Data
Iot & Application
High Performance Computing
Project

Name of Institute: Indus Institute of Technology & Engineering Name of Faculty: Mr.Hiren V Mer

Course code: DS0102
Course name: Advanced Data Structures (M.Tech)

Pre-requisites:

Student must have basic understanding of Basic Programming and Algorithms

Credit points: 4

Offered Semester: I – Advanced Data Structures

Course coordinator
Full name: Hiren V Mer
Department with sitting location: Bhawar Building,4th Floor
Telephone: +91 9227885688
Email:hirenmer.ce@indusuni.ac.in
Consultation times:
Wednesday: 3:00 PM to 4:00 PM

Course lecturer

Full name: Hiren V Mer
Department with sitting location: Bhawar Building,4th Floor
Telephone: +91 9227885688
Email:hirenmer.ce@indusuni.ac.in
Consultation times:
Wednesday: 3:00 PM to 4:00 PM

Students will be contacted throughout the session via mail with important information relating tothis course.

Course Objectives

1. The student should be able to choose appropriate data structures, understand the ADT/libraries, and use it to design algorithms for a specific problem.

2. Students should be able to understand the necessary mathematical abstraction to solve problems.

3. To familiarize students with advanced paradigms and data structure used to solve algorithmic problems.

4. Student should be able to come up with analysis of efficiency and proofs of correctness.

Course Outcomes (CO)

By participating in and understanding all facets of this course a student will be able to:

CO1: Understand the implementation of symbol table using hashing techniques. [BT-2]

CO2: Develop and analyze algorithms for red-black trees, B-trees and Splay trees. [BT-6]

CO3: Develop algorithms for text processing applications. [BT-6]

CO4: Identify suitable data structures and develop algorithms for computational geometry problems. [BT-1]

CO5: Illustrate and solve Text Processing algorithms [BT-4]

CO6: Understand concept of skip list to search, update and probabilistic analysis. [BT-2]

Program Outcomes (POs)

Engineering Graduates will be able to:

PO1. Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO2. Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3. Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO4. Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO5. 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.

PO6. The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO7. Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO9. Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10. Communication: 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.

PO11. Project management and finance: Demonstrate knowledge and understanding of the 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.

PO12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Course Outline

Course Content

UNIT-I

[7 hours]

Dictionaries: Definition, Dictionary Abstract Data Type, Implementation of Dictionaries. Hashing: Review of Hashing, Hash Function, Collision Resolution Techniques in Hashing, Separate Chaining, Open Addressing, Linear Probing, Quadratic Probing, Double Hashing, Rehashing, Extendible Hashing.

UNIT-II

[14 hours]

Skip Lists: Need for Randomizing Data Structures and Algorithms, Search and Update Operations on Skip Lists, Probabilistic Analysis of Skip Lists, Deterministic Skip Lists. Trees: Binary Search Trees, AVL Trees, Red Black Trees, 2-3 Trees, B-Trees, Splay Trees

UNIT-III

[12 hours]

Text Processing: Sting Operations, Brute-Force Pattern Matching, The Boyer-Moore Algorithm, The Knuth-Morris-Pratt Algorithm, Standard Tries, Compressed Tries, Suffix Tries, The Huffman Coding Algorithm, The Longest Common Subsequence Problem (LCS), Applying Dynamic Programming to the LCS Problem.

UNIT-IV

[15 hours]

Computational Geometry: One Dimensional Range Searching, Two Dimensional Range Searching, Constructing a Priority Search Tree, Searching a Priority Search

Tree, Priority Range Trees, Quadtrees, k-D Trees. Recent Trends in Hashing, Trees, and various computational geometry methods for efficiently solving the new evolving problem.

Method of delivery

Online Lectures

Study time

3 hours per week

CO-PO Mapping (PO: Program Outcomes)

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
C0 1	3	-	-	-	-	-	-	-	-	-	-	-
C0 2	3	3	3	2	-	-	-	-	1	-	-	-
C0 3	3		3	-	-	-	-	-	1	-	-	-
C0 4	3	3	2	-	-	-	-	-	1	1	-	-
C0 5	2	2	2	2	-	-	-	-	1	1	-	-
C0 6	3	-	-	-	-	-	-	-	-	-	-	-
DS0102	**2.8**	**2.6**	**2.5**	**2**	**-**	**-**	**-**	**-**	**1**	**1**	**-**	**-**

Blooms Taxonomy and Knowledge retention (For reference)

(Blooms taxonomy has been given for reference)

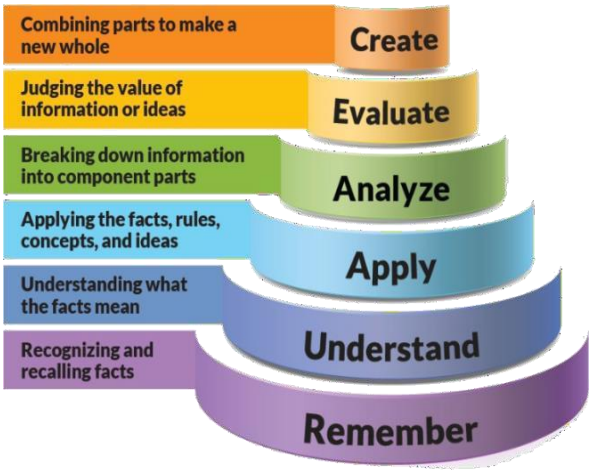


Figure 1: Blooms Taxonomy

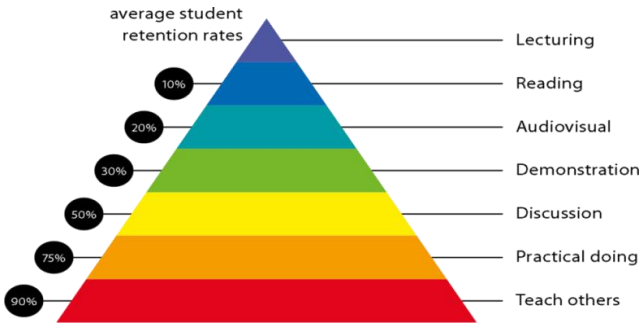


Figure 2: Knowledge retention

Graduate Qualities and Capabilities covered

(Qualities graduates harness crediting this Course)

General Post Graduate Qualities	Department of Computer Engineering Post Graduate Capabilities
Informed Have a sound knowledge of an area of study or profession and understand its current issues, locally and internationally. Know howto apply this knowledge. Understand how an area of study has developed and how it relates to other areas.	**1 Professional knowledge, grounding &awareness**
Independent learners Engage with new ideas and ways of thinking and critically analyze issues. Seek to extend knowledge through ongoing research, enquiry and reflection. Find and evaluate information,using a variety of	**2 Information literacy, gathering &processing**

sources and technologies. Acknowledge the work and ideas of others.	
Problem solvers Take on challenges and opportunities. Apply creative, logical and critical thinking skills to respond effectively. Make and implement decisions. Be flexible, thorough, innovative and aim for high standards.	**4 Problem solving skills**
Effective communicators Articulate ideas and convey them effectively using a range of media. Work collaboratively and engage with people in different settings. Recognize how culture can shape communication.	5 **Written communication** 6 **Oral Communication** 7 **Teamwork**
Responsible Understand how decisions can affect others and make ethically informed choices. Appreciate and respect diversity. Act with integrity as part of local, national, global and professional communities.	**10 Sustainability, societal & environmental impact**

Practical Work:

UNIT	TOPIC/DEFINITION	Learning Outcome
1	WAP for AVL Tree to implement following operations: (For nodes as integers)	Implement the height balancing concept of tree using different methods.
	a. Insertion: Test program for all cases (LL, RR, RL, LR rotation)	
	b. Deletion: Test Program for all cases (R0, R1, R-1, L0, L1, L-1)	
	c. Display: using set notation.	
2	Implement the above program I for nodes as Student structure, with key as Student_ roll_no.	
3	WAP to implement Red-Black trees with insertion and deletion operation for the given input data as Strings	
4	WAP to implement insertion, deletion, display and search operation in m-way B tree (i.e. a non-leaf node can have atmost m children) for the given data as integers (Test the program for m=3, 5, 7).	Implement the concept to be used in file storage structure
5	WAP to implement insertion, deletion and display operation in Min-Max Heap for the given data as integers	To implement data structure which will be used for implementing advanced algorithms..
6	WAP to implement Make_Heap, Insertion, Find_Min, Extract_Min, Union, Decrease_Key and Delete_Key operations in Binomial Heap for the given data as strings.	
7	WAP to implement priority queue.	
8	WAP to implement Make_Set, Find_Set and Union functions for Disjoint Set Data Structure for a given undirected graph G(V,E) using the linked list representation with simple implementation of Union operation.	
9	WAP to implement Dijkstra's algorithm for single-source shortest path in a weighted directed graph using fibonacci heap.	Implement and understand shortest path Data structure that is used in networking applications.
	WAP to store k keys into an array of size n at the location computed using a hash function, loc = key % n, where k<=n and k takes values from [1 to m], m>n. To handle the collisions use the following collision resolution techniques,	Implement the hashing concept that is used in various database application.
	a.Linear probing	
	b. Quadratic probing	
	c. Random probing	
	d. Double hashing/rehashing	

	e. Chaining	

Lecture/ Tutorial times

Monday: 10:20AM to 11:00AM Wednesday: 11:10AM to 12:10PM Thursday: 11:10AM to 12:10PM

Attendance Requirements

The University norms states that it is the responsibility of students to attend all lectures, tutorials, seminars and practical work as stipulated in the course outline. Minimum attendance requirement as per university norms is compulsory for being eligible for semester examinations.

Text books

1. Mark Allen Weiss, Data Structures and Algorithm Analysis in C++, 2nd Edition, Pearson, 2004.

Reference Books:

2. M T Goodrich, Roberto Tamassia, Algorithm Design, John Wiley, 2002.
3. Data Structures and Algorithms in C by Mark Allen Weiss - Pearson
4. The Design and analysis of computer Algorithms by A. V. Aho, R.Sethi and J. D. Ullman, Pearson.
5. Randomized Algorithms by Rajeev Motwani, Prabhakar Raghavan-Cambridge Univ.

Additional Materials

Web Resource
https://nptel.ac.in/courses/106/102/106102064/

https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-851-advanced-data-structures-spring-2012/lecture-videos/

ASSESSMENT GUIDELINES

Your final course mark will be calculated from the following:

CIE-Theory (60 Marks) **Class Regularity – 5 Marks** **Mid Sem- 40 Marks** **Assignments/Quiz/Viva - 15 Marks**	**CIE-Practical (60 Marks)** **Practical Performance + Regularity – 20 Marks** **Quiz/Assignment - 20Marks** **Internal Viva - 20 Marks**
ESE-Theory- 40 Marks	**ESE-Practical-40 Marks**
Total: 200 Marks	

SUPPLEMENTARY ASSESSMENT

Students who receive an overall mark less than 40% in internal component or less than 40% in the end semester will be considered for supplementary assessment in the respective components (i.e internal component or end semester) of semester concerned. Students must make themselves available during the supplementary examination period to take up the respective components (internal component or end semester) and need to obtain the required minimum 40% marks to clear the concerned components.

Practical Work Report/Laboratory Report:

A report on the practical work is due the subsequent week after completion of the class by each group.

Late Work

Late assignments will not be accepted without supporting documentation. Late submission of the reports will result in a deduction of -% of the maximum mark per calendar day

Format

All assignments must be presented in a neat, legible format with all information sources correctly referenced. **Assignment material handed in throughout the session that is not neat and legible will not be marked and will be returned to the student.**

Retention of Written Work

Written assessment work will be retained by the Course coordinator/lecturer for two weeks aftermarking to be collected by the students.

University and Faculty Policies

Students should make themselves aware of the University and/or Faculty Policies regarding plagiarism, special consideration, supplementary examinations and other educational issues and student matters.

Plagiarism - Plagiarism is not acceptable and may result in the imposition of severe penalties. Plagiarism is the use of another person's work, or idea, as if it is his or her own - if you have any doubts at all on what constitutes plagiarism, please consult your Course coordinator or lecturer. Plagiarism will be penalized severely.

Do not copy the work of other students.
Do not share your work with other students (except where required for a group activity orassessment.

Course schedule

(Mention quiz, assignment submission, breaks etc as well in the table under the TeachingLearning Activity Column)

Week No	Topic & Contents	CO Addressed	Teaching Learning Activity
1	Dictionaries: Definition, Dictionary Abstract Data Type, Implementation of Dictionaries. Hashing: Review of Hashing, Hash Function, Collision Resolution Techniques in Hashing, Separate Chaining, Open Addressing,	1	Online Lectures
2	Linear Probing, Quadratic Probing, Double Hashing, Rehashing, Extendible Hashing. Skip Lists: Introduction	1	Online Lectures
3	Skiplist : Need for Randomizing Data Structures and Algorithms Search and Update Operations on Skip Lists, Probabilistic Analysis of Skip Lists	2	Online Lectures
4	Deterministic Skip Lists. Trees: Binary Search Trees,	2	Online Lectures
5	AVL Trees, Red Black Trees,	2	Online Lectures

6	2-3 Trees, B-Trees, Splay Trees	2	Online Lectures
7	Text Processing: Sting Operations, Brute-Force Pattern Matching,	3,5	Online Lectures
8	The Boyer-Moore Algorithm, The Knuth-Morris-Pratt Algorithm, Standard Tries,	4	Online Lectures
9	Compressed Tries, Suffix Tries, The Huffman Coding Algorithm,	4,5	Online Lectures
10	The Longest Common Subsequence Problem (LCS), Applying Dynamic Programming to the LCS Problem.	4,5	Online Lectures
11	Computational Geometry: One Dimensional Range Searching,	4	Online Lectures
12	Two Dimensional Range Searching,	6	Online Lectures
13	Constructing a Priority Search Tree, Searching a Priority Search Tree,	6	Online Lectures
14	Priority Range Trees, Quadtrees, k-D Trees. Recent Trends in Hashing, Trees,	6	Online Lectures
15	and various computational geometry methods for efficiently solving the new evolving problem.	6	Online Lectures
16	Revision of the Syllabus and Doubt solving session	6	Online Lectures

Name of Institute: Indus Institute of Technology & Engineering
Name of Faculty: Pruthvi Patel

Course code: DS0103
Course name: Data Science

Pre-requisites: NA

Credit points: 4
Offered Semester: I

Course co-ordinator:

Full name: Pruthvi Patel
Department with sitting location: Computer Engineering Department (Fourth floor staffroom, Bhanwar building)
Telephone: +91 8866311132
Email: pruthvipatel.ce@indusuni.ac.in
Consultation times:
Monday and Tuesday: 3:30 to 5:00 PM

Course Lecturers:

Full name: Pruthvi Patel
Department with sitting location: Computer Engineering Department (Fourth floor staffroom, Bhanwar building)
Telephone: +91 8866311132
Email: pruthvipatel.ce@indusuni.ac.in
Consultation times:
Monday and Tuesday: 3:30 to 5:00 PM

Students will be contacted throughout the session via mail with important information relating to this course.

Course Objectives

1. Learn the fundamentals of data analytics and the data science pipeline
2. Learn how to scope the resources required for a data science project
3. Apply principles of Data Science to the analysis of business problems.
4. Apply data mining software to solve real-world problems.
5. Employ cutting edge tools and technologies to analyze Big Data

Course Outcomes (CO)

After successful completion of the course:

- Students will demonstrate knowledge of big data analytics.
- Students will demonstrate the ability to think critically in making decisions based on data
- Students will able to interpret data, extract meaningful information, and assess findings.
- Students will identify and analyze social, legal, and ethical issues in data science.
- Students will be able to choose and apply tools and methodologies to solve data science tasks.

Course Outline

UNIT-I	[12 hours]
Introduction	
Defining Data Science, what do data science people do? Current landscape of perspectives, Data Science in Business, Use Cases for Data Science	
UNIT-II	**[12 hours]**
Statistical Inference	
Statistical modeling, probability distributions, fitting a model, Intro to R	
Descriptive Statistics	
Introduction to the course, Descriptive Statistics, Probability Distributions	
Inferential Statistics	
Inferential Statistics through hypothesis tests, Permutation & Randomization Test	
UNIT-III	**[12 hours]**
Machine Learning Introduction and Concepts	
Differentiating algorithmic and model based frameworks, Regression: Ordinary Least Squares, Ridge Regression, Lasso Regression, K Nearest Neighbors, Regression & Classification	
Supervised Learning with Regression and Classification techniques	
Bias-Variance Dichotomy, Model Validation Approaches, Logistic Regression, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Regression and Classification Trees, Support Vector Machines, Ensemble Methods: Random Forest, Neural Networks, Deep learning	
Unsupervised Learning and data modelling	
Clustering, Associative Rule Mining, Logical Modelling: Converting a conceptual model to logical model , Integrity constraints, Normalization.	
UNIT-IV	**[12 hours]**
Data Visualization	
Basic principles, ideas and tools for data visualization	
Data Science and Ethical Issues	
Discussions on privacy, security, ethics	

Method of delivery

Chalk and Board, PowerPoint presentation, Practical Demonstrations

Study time

3 hrs theory, 2 hrs practical

CO-PO Mapping (PO: Program Outcomes)

Program Outcomes:

1. An understanding of the theoretical foundations and the limits of computing.
2. An ability to adapt existing models, techniques, algorithms, data structures, etc. for efficiently solving problems.
3. An ability to design, develop and evaluate new computer based systems for novel applications which meet the desired needs of industry and society.
4. Understanding and ability to use advanced computing techniques and tools.
5. An ability to undertake original research at the cutting edge of computer science & its related areas.
6. An ability to function effectively individually or as a part of a team to accomplish a stated goal.
7. An understanding of professional and ethical responsibility.
8. An ability to communicate effectively with a wide range of audience.
9. An ability to learn independently and engage in lifelong learning.
10. An understanding of the impact of IT related solutions in an economic, social and environment context.

CO-PO MAPPING	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	1	1	3	-	-	-	1	-	-	-
CO2	2	2	3	-	-	-	-	-	-	-
CO3	2	2	3	-	-	2	-	3	-	-
CO4	1	1	2	-	1	-	-	-	-	-
CO5	2	2	3	-	2	-	2	-	-	-

Blooms Taxonomy and Knowledge retention (For reference)

(Blooms taxonomy has been given for reference)

Figure 1: Blooms Taxonomy

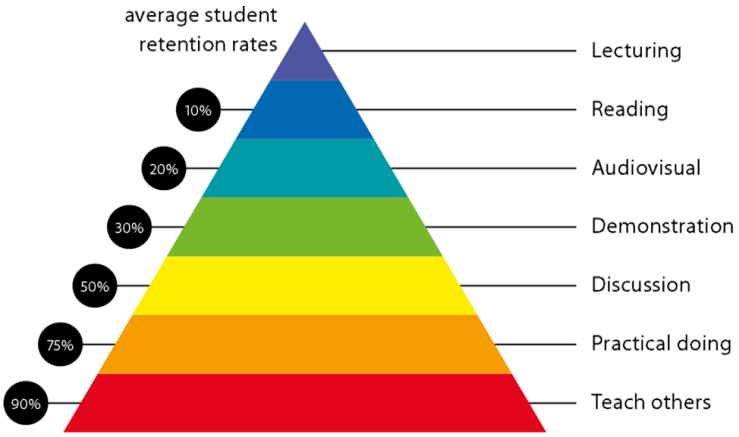


Figure 2: Knowledge retention

Graduate Qualities and Capabilities covered

General Graduate Qualities	Specific Department of _____Graduate Capabilities
Informed Have a sound knowledge of an area of study or profession and understand its current issues, locally and internationally. Know how to apply this knowledge. Understand how an area of study has developed and how it relates to other areas.	**1 Professional knowledge, grounding & awareness**
Independent learners Engage with new ideas and ways of thinking and critically analyze issues. Seek to extend knowledge through ongoing research, enquiry and reflection. Find and evaluate information, using a variety of sources and technologies. Acknowledge the work and ideas of others.	**2 Information literacy, gathering & processing**
Problem solvers Take on challenges and opportunities. Apply creative, logical and critical thinking skills to	**4 Problem solving skills**

respond effectively. Make and implement decisions. Be flexible, thorough, innovative and aim for high standards.	
Effective communicators Articulate ideas and convey them effectively using a range of media. Work collaboratively and engage with people in different settings. Recognize how culture can shape communication.	**5 Written communication** **6 Oral communication** **7 Teamwork**
Responsible Understand how decisions can affect others and make ethically informed choices. Appreciate and respect diversity. Act with integrity as part of local, national, global and professional communities.	**10 Sustainability, societal & environmental impact**

Practical work:

WkNo.	Class Activity	List of Practical
01	Lab 1	Introduction to R studio
02	Lab 2	Introduction to Python, Google Colaboratory
03	Lab 3	Performing Linear Regression in python.
04	Lab 4	Performing Linear Regression in R.
05	Lab 5	Performing Logistic Regression in Python.
06	Lab 6	Performing Lasso Regression in Python.
07	Lab 7	Performing Ridge Regression in Python.
08	Lab 8	Performing KNN in Python.
09	Lab 9	Performing Decision Tree in Python.
10	Lab 10	Performing SVM in Python.

Lecture/tutorial times

Attendance Requirements

The University norms states that it is the responsibility of students to attend all lectures, tutorials, seminars and practical work as stipulated in the course outline. Minimum attendance requirement as per university norms is compulsory for being eligible for semester examinations.

Text books

1. Applied statistics and probability for engineers, Montgomery, Douglas C., George C. Runger, John Wiley & Sons, 2010
2. Mining of Massive Datasets , Anand Rajaraman and Jeffrey David Ullman 2012
3. Doing Data Science, Straight Talk From The Frontline., Cathy O'Neil and Rachel Schutt , O'Reilly. 2014

Reference Book:

1. Machine Learning: A Probabilistic Perspective. Kevin P. Murphy.

ASSESSMENT GUIDELINES

CIE-Theory (60 Marks): **Class Test [30 Marks]** **Assignments [30 Marks]**	**CIE-Practical (60 Marks):** **Lab manual + Regularity [30 Marks]** **Case Study [30 Marks]**
ESE-Theory- 40 Marks	**ESE-Practical- 40 Marks**
Total: 200 Marks	

Your final course mark will be calculated from the following:

SUPPLEMENTARY ASSESSMENT

Students who receive an overall mark less than 40% in internal component or less than 40% in the end semester will be considered for supplementary assessment in the respective components (i.e internal component or end semester) of semester concerned. Students must make themselves available during the supplementary examination period to take up the respective components (internal component or end semester) and need to obtain the required minimum 40% marks to clear the concerned components.

Practical Work Report/Laboratory Report:

A report on the practical work is due the subsequent week after completion of the class by each group.

Late Work

Late assignments will not be accepted without supporting documentation. Late submission of the reports will result in a deduction of -% of the maximum mark per calendar day

Format

All assignments must be presented in a neat, legible format with all information sources correctly referenced. **Assignment material handed in throughout the session that is not neat and legible will not be marked and will be returned to the student.**

Retention of Written Work

Written assessment work will be retained by the Course coordinator/lecturer for two weeks after marking to be collected by the students.

University and Faculty Policies

Students should make themselves aware of the University and/or Faculty Policies regarding plagiarism, special consideration, supplementary examinations and other educational issues and student matters.

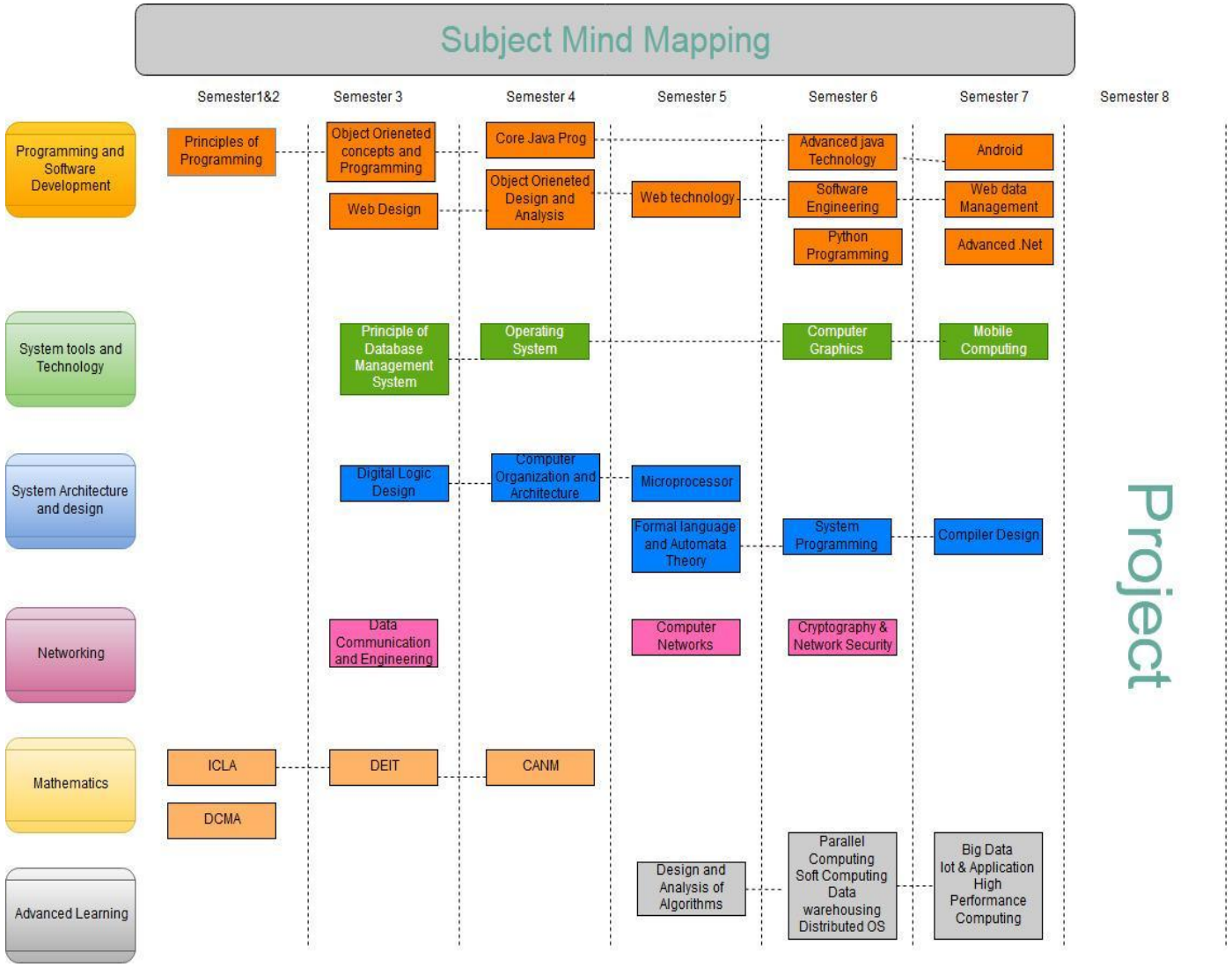
Plagiarism - Plagiarism is not acceptable and may result in the imposition of severe penalties. Plagiarism is the use of another person's work, or idea, as if it is his or her own - if you have any doubts at all on what constitutes plagiarism, please consult your Course coordinator or lecturer. Plagiarism will be penalized severely.

Do not copy the work of other students.
Do not share your work with other students (except where required for a group activity or assessment.

Course schedule (subject to change)

	Week #	Topic & contents	CO Addressed	Teaching Learning Activity (TLA)
	Week 1	Defining Data Science, what do data science people do? Current landscape of perspectives,	I	Chalk & Board, Discussion
	Week 2	Data Science in Business, Use Cases for Data Science	II	Chalk & Board, Demonstration
	Week 3	**Statistical Inference** Statistical modeling, probability distributions, fitting a model, Intro to R	II	Chalk & Board, Demonstration
	Week 4	**Descriptive Statistics** Introduction to the course, Descriptive Statistics, Probability Distributions	II	Chalk & Board, Demonstration
	Week 5	**Inferential Statistics** Inferential Statistics through hypothesis tests, Permutation & Randomization Test	II	Chalk & Board, Demonstration
	Week 6	**Machine Learning Introduction and Concepts** Differentiating algorithmic and model based frameworks	III	Chalk & Board (Class Test)
	Week 7	Regression: Ordinary Least Squares, Ridge Regression, Lasso Regression,	III	Chalk & Board, Demonstration
	Week 8	K Nearest Neighbors, Regression & Classification	I, IV	Chalk & Board, Demonstration

	Week 9	**Supervised Learning with Regression and Classification techniques** Bias-Variance Dichotomy, Model Validation Approaches,	I, IV	Chalk & Board Demonstration
	Week 10	Logistic Regression, Linear Discriminant Analysis, Quadratic Discriminant Analysis, 2,5	I, IV	Chalk & Board Demonstration
	Week 11	Regression and Classification Trees, Support Vector Machines,	I, IV	Chalk & Board Demonstration
	Week 12	Ensemble Methods: Random Forest, Neural Networks, Deep learning	IV, V	Chalk & Board Demonstration
	Week 13	**Unsupervised Learning and data modelling** Clustering, Associative Rule Mining, Logical Modelling: Converting a conceptual model to logical model , Integrity constraints, Normalization.	V	Chalk & Board Demonstration
	Week 14	**Data Visualization** Basic principles, ideas and tools for data visualization	I, V	Chalk & Board Demonstration
	Week 15	**Data Science and Ethical Issues** Discussions on privacy, security, ethics	I, V	Chalk & Board Demonstration


Subject Mind Mapping
Semester1&2
Semester 3
Semester 4
Semester 5
Semester 6
Semester 7
Semester 8
Programming and Software Development
System tools and Technology
System Architecture and design
Networking
Mathematics
Advanced Learning
Principles of Programming
Object Orieneted concepts and Programming
Web Design
Core Java Prog
Object Orieneted Design and Analysis
Web technology
Advanced java Technology
Software Engineering
Python Programming
Android
Web data Management
Advanced .Net
Principle of Database Management System
Operating System
Computer Graphics
Mobile Computing
Digital Logic Design
Computer Organization and Architecture
Microprocessor
Formal language and Automata Theory
System Programming
Compiler Design
Data Communication and Engineering
Computer Networks
Cryptography & Network Security
ICLA
DCMA
DEIT
CANM
Design and Analysis of Algorithms
Parallel Computing
Soft Computing
Data warehousing
Distributed OS
Big Data
Iot & Application
High Performance Computing
Project

Name of Institute: INDUS INSTITUTE OF TECHNOLOGY & ENGINEERING
(DEPARTMENT OF COMPUTER ENGINEERING)

Name of Faculty: Manisha Valera

Course code: DS0106

Course name: Data Preparation and Analysis (Department Elective - I)

Pre-requisites: Data Mining
Credit points: 4

L	T	P	C
3	0	2	4

Offered Semester: 1st Sem

Course Coordinator (weeks 01 - 15)
Full Name: Manisha Valera
Department with sitting location: 4th Floor, Faculty Wing, Bhanwar Building, IITE - IU
Telephone:9714960628
Email: manishavalera.ce@indusuni.ac.in

Consultation times: 04.30 PM – 04:50 PM (Monday- Friday)
09.00 AM – 10.00 AM (Working Saturdays)

Course Lecturer (weeks 01 - 15)
Full Name: Manisha Valera
Department with sitting location: 4th Floor, Faculty Wing, Bhanwar Building, IITE - IU
Telephone:9714960628
Email: manishavalera.ce@indusuni.ac.in
Consultation times: 04.30 PM – 04:50 PM (Monday- Friday)
09.00 AM – 10.00 AM (Working Saturdays)

Students will be contacted throughout the Session via Mail with important information relating to this Course.

Course Objectives

1. To learn how to gather and analyze large sets of data to gain useful business understanding and how to produce a quantitative analysis report/memo with the necessary information to make Decisions.
2. To survey industrial and scientific applications of Data Analytics, with case studies.
3. To prepare data for analytics and perform exploratory data analysis.
4. To develop meaningful data visualizations.
5. To perform cleaning and reformatting real world data for analysis.
6. To apply summary statistics techniques over datasets.

Course Outcomes (CO)

At the end of this subject, students should be able to:

1. Extract the data for performing the Analysis.
2. To perform cleaning and reformatting real world data for analysis.
3. Develop meaningful Data Visualizations.
4. Understand Ethics in the Profession.

Course Outline

(Key in topics to be dealt)

Subject: Data Preparation and Analysis (Department Elective - I)								
Program: M.Tech. in Data Science & Analysis			Subject Code: DS0106				Semester: I	
Teaching Scheme				**Examination Evaluation Scheme**				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)- Theory**	**Continuous Internal Evaluation (CIE)- Practical**	**Total**
3	0	2	4	40	40	60	60	200

UNIT-I

Introduction to Data Analysis **[10 hours]**

Defining data analysis problems: Knowing the client, Understanding the questions; Data Gathering and Preparation: Data formats, parsing and transformation, Scalability and real-time issues

UNIT-II

Exploratory Analysis **[12 hours]**

Data Cleaning: Consistency Checking, Heterogeneous and Missing data, Data Transformation and Segmentation; Exploratory Analysis: Descriptive and Comparative Analysis, Clustering and Association, Hypothesis Generation

UNIT-III

Visualization **[08 hours]**

Designing Visualizations, Time Series, Geolocated Data, Correlations and Connections, Hierarchies and Networks, Interactivity

UNIT-IV

[06 hours]

Ethics in the Profession

Cases in Computing, Statistics and Communication, Professional ethics codes: ACM, IEEE, AM Stat. Assoc.

TEXT BOOKS

1. Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining, by Glenn J. Myatt

REFERENCES

1. Data Analytics by Anil Maheshwari – McGraw Hill
2. Data Preparation for Data Mining by Dorian Pyle – Morgan Kaufmann Series
3. Data Preparation in the Big Data by Federico Castanedo – O'Reilly

Method of delivery

Chalk, Board

(Face to face lectures, self study material, Active Learning Techniques)

Chalk

Study time

3 hours –Lectures

(How many hours per week including class attendance)

CO-PO Mapping (PO: Program Outcomes)

Program Outcomes (PO's)

Engineering Graduates will be able to:

1. An understanding of the theoretical foundations and the limits of computing.
2. An ability to adapt existing models, techniques, algorithms, data structures, etc. for efficiently solving problems.
3. An ability to design, develop and evaluate new computer based systems for novel applications which meet the desired needs of industry and society.
4. Understanding and ability to use advanced computing techniques and tools.
5. An ability to undertake original research at the cutting edge of computer science & its related areas.
6. An ability to function effectively individually or as a part of a team to accomplish a stated goal.
7. An understanding of professional and ethical responsibility.
8. An ability to communicate effectively with a wide range of audience.

9. An ability to learn independently and engage in lifelong learning.
10. An understanding of the impact of IT related solutions in an economic, social and environment context

PO1. **Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
PO2. **Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
PO3. **Design/development of solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
PO4. **Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
PO5. **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.
PO6. **The engineer and society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
PO7. **Environment and sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
PO8. **Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
PO9. **Individual and team work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
PO10. **Communication**: 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.
PO11. **Project management and finance**: Demonstrate knowledge and understanding of the 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.
PO12. Life-long learning: Recognize the need for and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Mapping CO's with PO's

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	-	2	3	-	-	-	-	-	-	-	-	-
CO2	-	2	3	-	-	-	-	-	-	-	-	-
CO3	-	-	-	3	-	-	-	-	-	-	-	-
CO4	-	-	-	-	-	-	3	-	-	3	-	-

1-Lightly Mapped 2- Moderately Mapped 3- Highly Mapped

7.2 Mapping of CO's with PSO's

	PSO 1	PSO 2
CO1	-	2
CO2	-	-
CO3	-	-
CO4	-	-

1-Lightly Mapped 2- Moderately Mapped 3- Highly Mapped

Blooms Taxonomy and Knowledge retention (For reference)

(Blooms taxonomy has been given for reference)

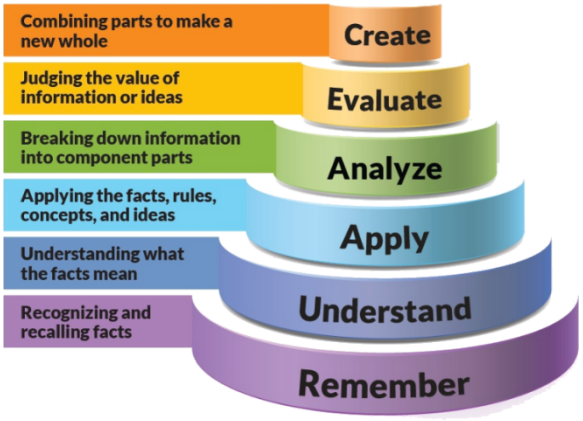


Figure 1: Blooms Taxonomy

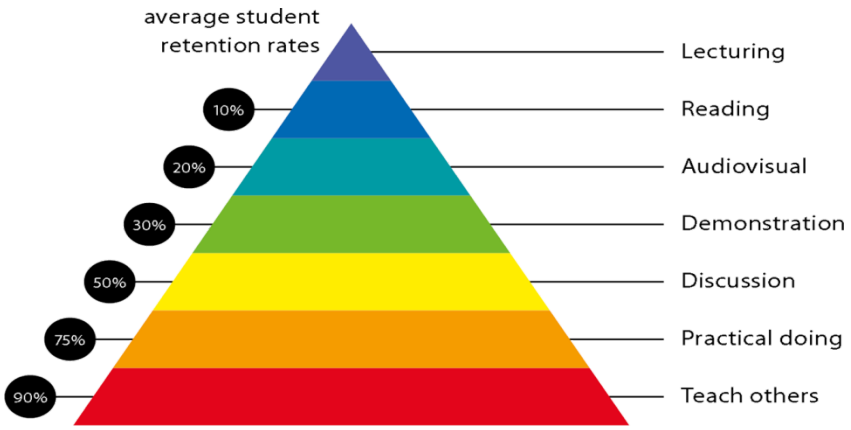


Figure 2: Knowledge retention

Lecture/tutorial times

Example:
Lecture- Monday – 1.30 pm
Lecture- Wednesday – 11.00 am
Lecture-Friday – 9.55 am

Lab- Thursday 9.00 am

(Give lecture times in the format below)

Attendance Requirements

The University norms states that it is the responsibility of students to attend all lectures, tutorials, seminars and practical work as stipulated in the Course outline. Minimum

attendance requirement as per university norms is compulsory for being eligible for mid and end semester examinations.

Details of referencing system to be used in written work

Text books

1. Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining, by Glenn J. Myatt
2. Data Analytics by Anil Maheshwari – McGraw Hill
3. Data Preparation for Data Mining by Dorian Pyle – Morgan Kaufmann Series
4. Data Preparation in the Big Data by Federico Castanedo – O'Reilly

Additional Materials

ASSESSMENT GUIDELINES

Your final course mark will be calculated from the following:

Example:

Quiz 1	5% (week 4)	Objective (1-3)
Quiz II	5% (week 8)	Objective (1-4)
Mid semester	30% (due week 10)	Objectives (2-5)
Final exam (*closed book*)	60%	Objectives (1-5)

Internal Evaluation – Theory Component	
Parameters	**Percentage input**

Continuous Internal Evaluation: (60 Marks): **1.** Class TEST (40 Marks) 2. Internal Evaluation (20 Marks): a. Assignment (5 Marks) b. Quiz / Student Involvement/ Behaviour (05 Marks)	**100 %** 75 % 25 % 12.5 % 12.5 %
Internal Evaluation- Practical Component	
Parameters	**Percentage input**
Continuous Internal Evaluation: (60 Marks)	100 %
Note: Lab Marks foil attached	

University End Sem Exam – Theory Component	
Parameters	**Percentage input**
End Semester Examination (40 Marks)	100 %
University End Sem Exam – Practical Component	
Parameters	**Percentage input**
End Semester Examination (40 Marks)	NA

SUPPL EMENTARY ASSESSMENT

Students who receive an overall mark less than 40% in mid semester or end semester will be considered for supplementary assessment in the respective components (i.e mid semester or end semester) of semester concerned. Students must make themselves available during the supplementary examination period to take up the respective components (mid semester or end semester) and need to obtain the required minimum 40% marks to clear the concerned components.

Practical Work Report/Laboratory Report:

A report on the practical work is due the subsequent week after completion of the class by each group.

Late Work

Late assignments will not be accepted without supporting documentation. Late submission of the reports will result in a deduction of 20% of the maximum mark per calendar day

Format

All assignments must be presented in a neat, legible format with all information sources correctly referenced. **Assignment material handed in throughout the session that is not neat and legible will not be marked and will be returned to the student.**

Retention of Written Work

Written assessment work will be retained by the Course coordinator/lecturer for two weeks after marking to be collected by the students.

University and Faculty Policies

Students should make themselves aware of the University and/or Faculty Policies regarding plagiarism, special consideration, supplementary examinations and other educational issues and student matters.

Plagiarism - Plagiarism is not acceptable and may result in the imposition of severe penalties. Plagiarism is the use of another person's work, or idea, as if it is his or her own - if you have any doubts at all on what constitutes plagiarism, please consult your Course coordinator or lecturer. Plagiarism will be penalized severely.

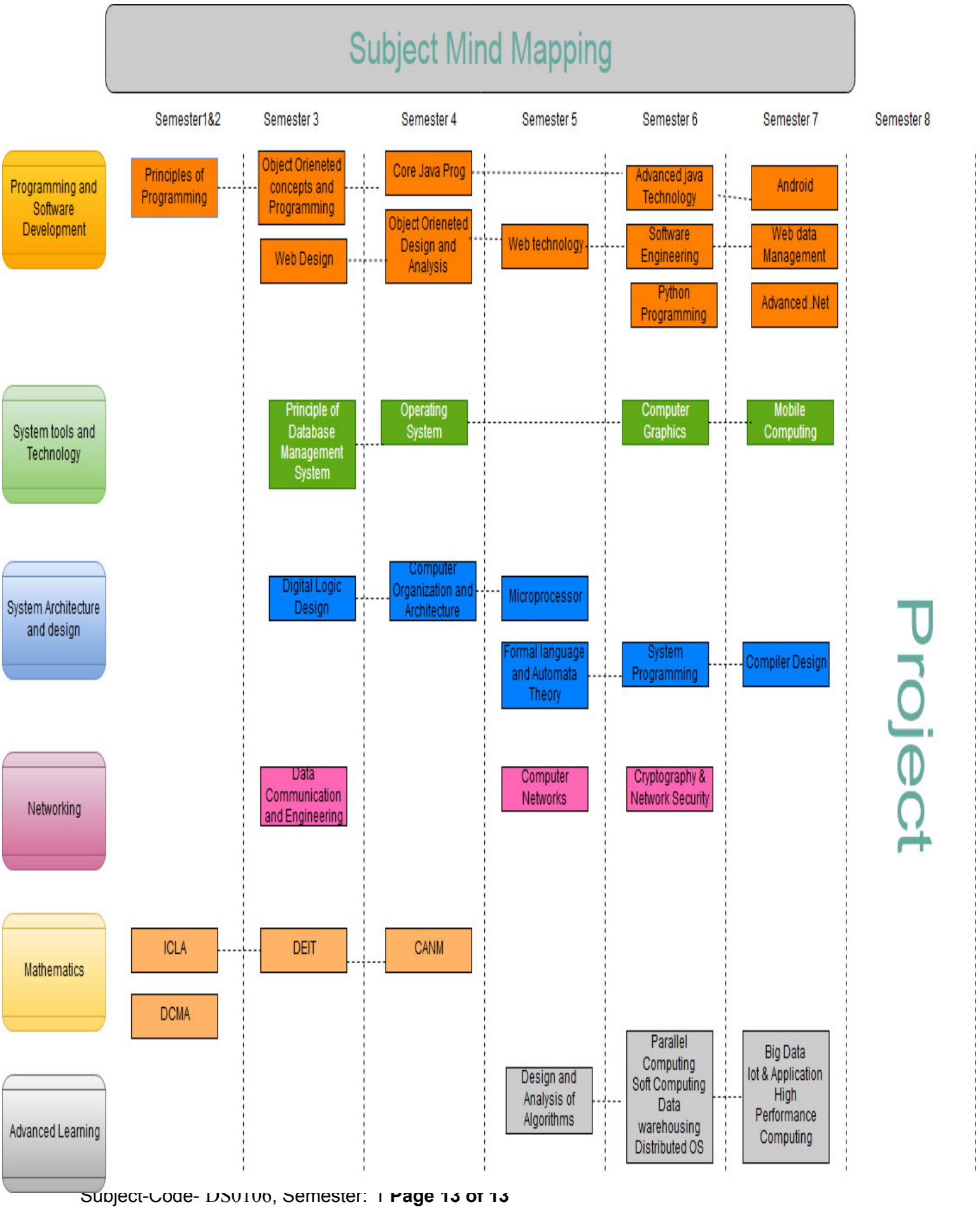
Do not copy the work of other students.
Do not share your work with other students (except where required for a group activity or assessment)

.

Course schedule (subject to change)

(Mention quiz, assignment submission, breaks etc as well in the table under the Teaching Learning Activity Column)

	Week #	Topic & contents	CO Addressed	Teaching Learning Activity (TLA)
	Weeks 1	**Introduction to Data Analysis** Defining data analysis problems: Knowing the client, Understanding the questions;	1	Chalk,Board, Projector
	Weeks 2	Data Gathering and Preparation	2	Chalk,Board, Projector
	Week 3	Data formats, parsing and transformation	2	Chalk,Board, Projector
	Week 4	Scalability , real-time issues	1	Chalk,Board, Projector
	Week 5	**Exploratory Analysis** Data Cleaning: Consistency Checking	2	Chalk,Board, Projector
	Week 6	Heterogeneous and Missing data, Data Transformation and Segmentation;	2	Chalk,Board, Projector
	Week 7	Exploratory Analysis: Descriptive and Comparative Analysis	1	Chalk,Board, Projector
	Week 8	Clustering and Association, Hypothesis Generation	1	Chalk,Board, Projector
	Week 9	**Visualization** Designing Visualizations, Time Series,	3	Chalk,Board, Projector
	Week 10	Geolocated Data, Correlations and Connections	3	Chalk,Board, Projector
	Week 11	Hierarchies and Networks, Interactivity	3	Chalk,Board, Projector
	Week 12	**Ethics in the Profession** Cases in Computing, Statistics and Communication, Professional ethics codes: ACM, IEEE, AM Stat. Assoc.	4	Chalk,Board, Projector


Subject Mind Mapping
Semester1&2
Semester 3
Semester 4
Semester 5
Semester 6
Semester 7
Semester 8
Programming and Software Development
Principles of Programming
Object Orieneted concepts and Programming
Web Design
Core Java Prog
Object Orieneted Design and Analysis
Web technology
Advanced java Technology
Software Engineering
Python Programming
Android
Web data Management
Advanced .Net
System tools and Technology
Principle of Database Management System
Operating System
Computer Graphics
Mobile Computing
System Architecture and design
Digital Logic Design
Computer Organization and Architecture
Microprocessor
Formal language and Automata Theory
System Programming
Compiler Design
Networking
Data Communication and Engineering
Computer Networks
Cryptography & Network Security
Mathematics
ICLA
DCMA
DEIT
CANM
Advanced Learning
Design and Analysis of Algorithms
Parallel Computing Soft Computing Data warehousing Distributed OS
Big Data Iot & Application High Performance Computing
Project

Name of Institute: Indus Institute of Technology & Engineering
Name of Faculty: Kavita Pandya

Course code: DS0108
Course name: Machine Learning (Department Elective - II)

Pre-requisites:
Required background knowledge includes probability theory, linear algebra, continuous mathematics, multivariate calculus and multivariate probability theory, as well as good programming skills. Students are required to have taken the **Machine Learning** course.

Credit points: 4
Offered Semester: I

Course co-ordinator:

Full name: Kavita Pandya
Department with sitting location: Computer Engineering Department (Fourth floor staffroom, Bhanwar building)
Telephone: +91 9624342414
Email: kavitapandya.ce@indusuni.ac.in
Consultation times:
4:30 to 5:00 PM

Course Lecturers:

Full name: Kavita Pandya
Department with sitting location: Computer Engineering Department (Fourth floor staffroom, Bhanwar building)
Telephone: +91 9624342414
Email: kavitapandya.ce@indusuni.ac.in
Consultation times:
4:30 to 5:00 PM

Students will be contacted throughout the session via mail with important information relating to this course.

Course Objectives

1. Machine learning techniques enable us to automatically extract features from data so as to solve predictive tasks, such as forecasting, missing data imputation, anomaly detection, classification, ranking, control and decision making.

2. The course will introduce the mathematical definitions of the relevant machine learning models and derive their associated optimization algorithms.

Course Outcomes (CO)

After successful completion of the course, student will able:

- To provide deep understanding on various mathematical models and algorithms.
- Learn modeling skills, optimization algorithms, and theories commonly in machine learning and data analytics.
- Apply advanced statistical machine learning techniques
- Understand the caveats of real-world data.
- Apply modern machine learning methods in enterprise.

Course Outline

UNIT-I	[12 hours]
Stochastic gradient descent: the workhorse of machine learning	
Gradient descent, Stochastic gradient descent, SGD for convex objectives, The effect of choosing the step size/learning rate, Mini-batching and batch size, Overfitting in neural nets: Backpropagation, Generalization error, Regularization, Conjugate gradient and Early stopping	
Initialization and momentum in deep learning	
Momentum and acceleration, Momentum for quadratic optimization, Momentum for principle component analysis (PCA), The kernel trick, Gram matrix versus feature extraction: systems tradeoffs, Adaptive/data-dependent feature mappings	
UNIT-II	**[12 hours]**
Random search for hyper-parameter optimization	
Online versus offline learning, Variance reduction, SVRG, Fast linear rates for convex objectives, Metaparameter optimization, Assigning parameters from folklore, Random search over parameters	
Bayesian optimization of machine learning algorithms	
Non-convex stochastic gradient descent, Weakness of theoretical guarantees, stochastic power iteration, Deep learning as non-convex optimization	
UNIT-III	**[12 hours]**
Parallelism in Machine Learning	
Major bottleneck for ML systems: parallelism, Asynchronous execution, Major bottleneck for ML systems: memory bandwidth and locality, Low precision computation, Vector computation, Scan orders	

Deep Compression
Stochastic coordinate descent, Markov chain Monte Carlo and Gibbs sampling, Contrastive divergence, Derivative free optimization, Compressing deep neural networks with pruning, trained quantization and Huffman coding
UNIT-IV **[12 hours]**
Performance analysis of a tensor processing unit
Hardware for machine learning, The dominance of GPUs, Accelerators for machine learning
Machine learning frameworks and cluster parallelism
TensorFlow, SciKit, PyTorch

Method of delivery

Chalk and Board, PowerPoint presentation, Practical Demonstrations

Study time

3 hrs theory, 2 hrs practical

CO-PO Mapping (PO: Program Outcomes)

CO-PO MAPPING	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	1	1	3	-	-	-	-	-	-	-
CO2	2	2	3	-	-	-	-	-	-	-
CO3	2	2	3	-	-	-	-	-	-	-
CO4	1	1	2	-	-	-	-	-	-	-
CO5	2	2	3	-	-	-	-	-	-	-

Blooms Taxonomy and Knowledge retention (For reference)

(Blooms taxonomy has been given for reference)

Figure 1: Blooms Taxonomy

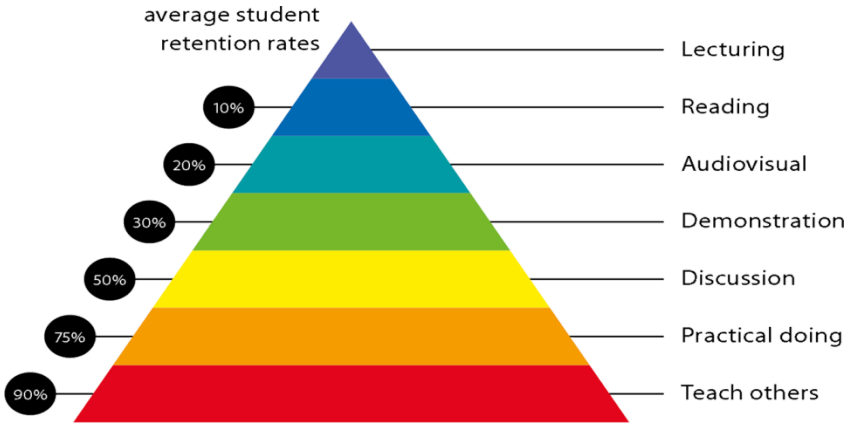


Figure 2: Knowledge retention

Graduate Qualities and Capabilities covered

General Graduate Qualities	Specific Department of _____Graduate Capabilities
Informed Have a sound knowledge of an area of study or profession and understand its current issues, locally and internationally. Know how to apply this knowledge. Understand how an area of study has developed and how it relates to other areas.	**1 Professional knowledge, grounding & awareness**
Independent learners Engage with new ideas and ways of thinking and critically analyze issues. Seek to extend knowledge through ongoing research, enquiry and reflection. Find and evaluate information, using a variety of sources and technologies. Acknowledge the work and ideas of others.	**2 Information literacy, gathering & processing**
Problem solvers Take on challenges and opportunities. Apply creative, logical and critical thinking skills to respond effectively. Make and implement decisions. Be flexible, thorough, innovative and aim for high standards.	**4 Problem solving skills**
Effective communicators Articulate ideas and convey them effectively using a range of media. Work collaboratively and engage with people in different settings. Recognize how culture can shape communication.	**5 Written communication** **6 Oral communication** **7 Teamwork**
Responsible Understand how decisions can affect others and make ethically informed choices. Appreciate and respect diversity. Act with	**10 Sustainability, societal & environmental impact**

integrity as part of local, national, global and professional communities.	

Practical work:

1. **Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file.** 2. **For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.** 3. **Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.** 4. **Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.** 5. **Study parallelism in Machine Learning: GPUs, CUDA, and Practical Applications.** 6. **Study parallel processing in python** 7. **Write a program to count how many numbers exist between a given range in each row with and without parallelization** 8. **Study how to parallelize a Pandas Data Frame** 9. **Implement multi-layer fully connected network in Keras**

Lecture/tutorial times

Attendance Requirements

The University norms states that it is the responsibility of students to attend all lectures, tutorials, seminars and practical work as stipulated in the course outline. Minimum attendance requirement as per university norms is compulsory for being eligible for semester examinations.

Text books

1. Machine Learning Tom Mitchell; McGraw-Hill Science
2. Hands-On GPU Programming with Python and CUDA: Boost your application's performance and productivity with CUDA Brian Tuomanen

Reference Books:

1. Parallel Machine Learning Using Concurrency Control Xinghao Pan Pattern Classification Richard O. Duda, Peter E. Hart, David G. Stork
2. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods Nello Cristianini, John Shawe-Taylor; Cambridge University Press
3. Rajjan Shinghal, Pattern Recognition, Techniques and Applications, OXFORD

Additional Materials

Web Resource

- https://professional.mit.edu/programs/short-programs/advanced-machine-learning

ASSESSMENT GUIDELINES

CIE-Theory (60 Marks): **Class Test [30 Marks]** **Assignments [20 Marks]** **Attendance [10 Marks]**	**CIE-Practical (60 Marks):** **Lab manual + Regularity [20 Marks]** **Practical performance [20 Marks]** **Internal Practical Exam [20 Marks]**
ESE-Theory- 40 Marks	**ESE-Practical- 40 Marks**
Total: 200 Marks	

Your final course mark will be calculated from the following:

SUPPLEMENTARY ASSESSMENT

Students who receive an overall mark less than 40% in internal component or less than 40% in the end semester will be considered for supplementary assessment in the respective components (i.e internal component or end semester) of semester concerned. Students must make themselves available during the supplementary examination period to take up the respective components (internal component or end semester) and need to obtain the required minimum 40% marks to clear the concerned components.

Practical Work Report/Laboratory Report:

A report on the practical work is due the subsequent week after completion of the class by each group.

Late Work

Late assignments will not be accepted without supporting documentation. Late submission of the reports will result in a deduction of -% of the maximum mark per calendar day

Format

All assignments must be presented in a neat, legible format with all information sources correctly referenced. **Assignment material handed in throughout the session that is not neat and legible will not be marked and will be returned to the student.**

Retention of Written Work

Written assessment work will be retained by the Course coordinator/lecturer for two weeks after marking to be collected by the students.

University and Faculty Policies

Students should make themselves aware of the University and/or Faculty Policies regarding plagiarism, special consideration, supplementary examinations and other educational issues and student matters.

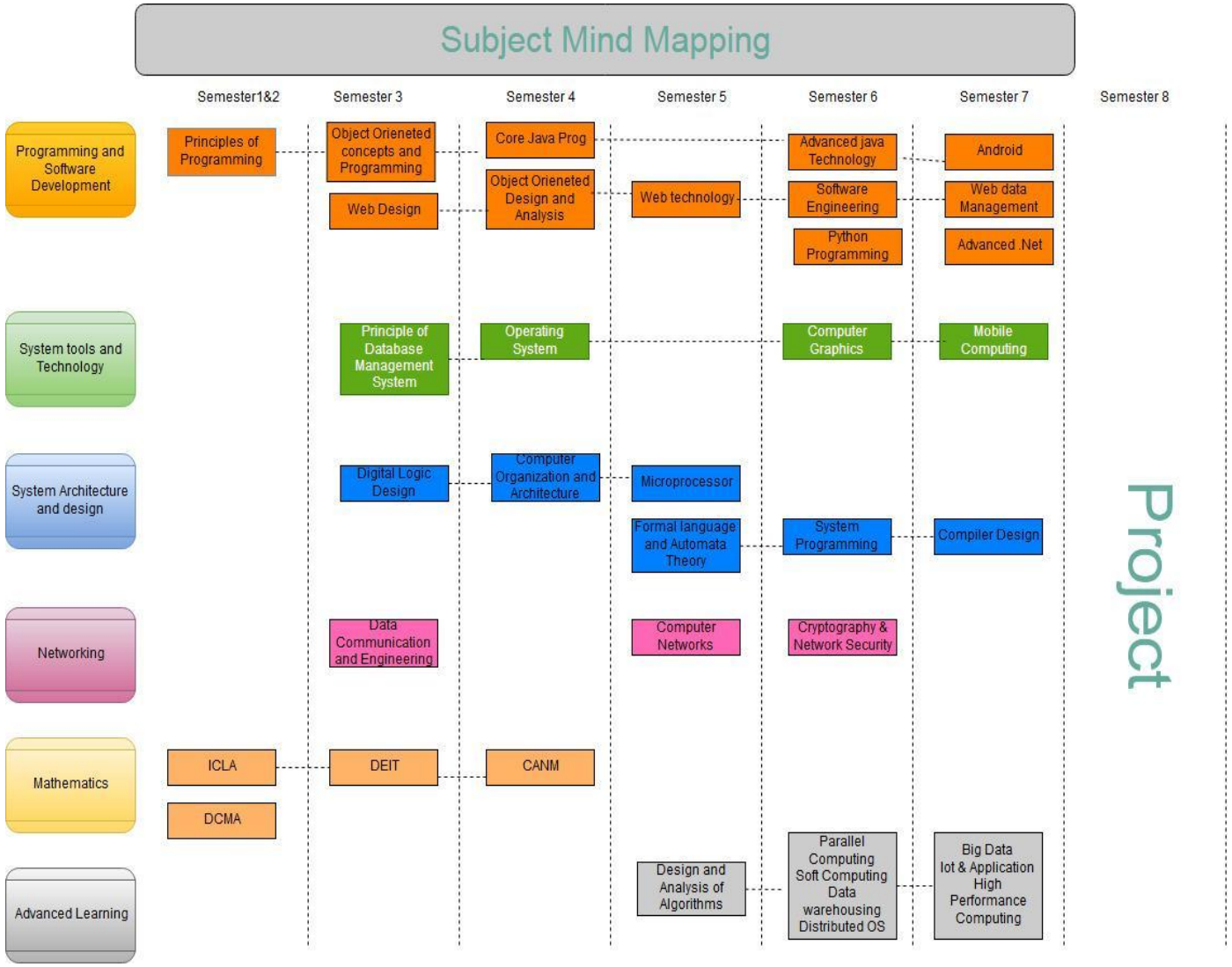
Plagiarism - Plagiarism is not acceptable and may result in the imposition of severe penalties. Plagiarism is the use of another person's work, or idea, as if it is his or her own - if you have any doubts at all on what constitutes plagiarism, please consult your Course coordinator or lecturer. Plagiarism will be penalized severely.

Do not copy the work of other students.
Do not share your work with other students (except where required for a group activity or assessment.

Course schedule (subject to change)

	Week #	Topic & contents	CO Addressed	Teaching Learning Activity (TLA)
	Week 1	Gradient descent, Stochastic gradient descent, SGD for convex objectives, The effect of choosing the step size/learning rate	I	Chalk & Board, Discussion
	Week 2	Mini-batching and batch size, Overfitting in neural nets: Backpropagation, Generalization error, Regularization, Conjugate gradient and Early stopping	II	Chalk & Board, Demonstration
	Week 3	Momentum and acceleration, Momentum for quadratic optimization, Momentum for principle component analysis (PCA)	II	Chalk & Board, Demonstration
	Week 4	The kernel trick, Gram matrix versus feature extraction: systems tradeoffs, Adaptive/data-dependent feature mappings 2	II	Chalk & Board, Demonstration
	Week 5	Online versus offline learning, Variance reduction, SVRG, Fast linear rates for convex objectives	II	Chalk & Board, Demonstration
	Week 6	Meta parameter optimization, Assigning parameters from folklore, Random search over parameters	III	Chalk & Board (Class Test)

		3		
	Week 7	Non-convex stochastic gradient descent, Weakness of theoretical guarantees, stochastic power iteration	III	Chalk & Board, Demonstration
	Week 8	Deep learning as non-convex optimization	I, IV	Chalk & Board, Demonstration (Presentation)
	Week 9	Major bottleneck for ML systems: parallelism, Asynchronous execution	I, IV	Chalk & Board Demonstration
	Week 10	Major bottleneck for ML systems: memory bandwidth and locality, Low precision computation, Vector computation, Scan orders	I, IV	Chalk & Board Demonstration
	Week 11	Stochastic coordinate descent, Markov chain Monte Carlo and Gibbs sampling, Contrastive divergence	I, IV	Chalk & Board Demonstration (Quiz)
	Week 12	Derivative free optimization, Compressing deep neural networks with pruning, trained quantization and Huffman coding. 2,5	VI	Chalk & Board Demonstration
	Week 13	Hardware for machine learning, The dominance of GPUs, Accelerators for machine learning	VI	Chalk & Board Demonstration
	Week 14	TensorFlow, SciKit, PyTorch	I, VI	Chalk & Board Demonstration
	Week 15	Applications	I, VI	Chalk & Board Demonstration


Subject Mind Mapping
Semester1&2
Semester 3
Semester 4
Semester 5
Semester 6
Semester 7
Semester 8
Programming and Software Development
System tools and Technology
System Architecture and design
Networking
Mathematics
Advanced Learning
Principles of Programming
Object Orieneted concepts and Programming
Web Design
Core Java Prog
Object Orieneted Design and Analysis
Web technology
Advanced java Technology
Software Engineering
Python Programming
Android
Web data Management
Advanced .Net
Principle of Database Management System
Operating System
Computer Graphics
Mobile Computing
Digital Logic Design
Computer Organization and Architecture
Microprocessor
Formal language and Automata Theory
System Programming
Compiler Design
Data Communication and Engineering
Computer Networks
Cryptography & Network Security
ICLA
DCMA
DEIT
CANM
Design and Analysis of Algorithms
Parallel Computing Soft Computing Data warehousing Distributed OS
Big Data Iot & Application High Performance Computing
Project