SCTR's PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43

An Autonomous Institute affiliated to the
Savitribai Phule Pune University
Approved by AICTE & Government of Maharashtra,
Accredited by NAAC (A+) & NBA [All eligible UG Programs]

Master of Technology (Data Science)

Department of Computer Engineering

Curriculum Structure

(W.e.f. A.Y. 2024-25)

(Approved by the Board of Studies and Academic Council)

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

Semester - I

Broad Category of Course			Teaching Scheme (Hours/Week)				Credits/ Grades				Examination Scheme and Marks					
	Subject code	Name of subjects	L	P	T	Total	L	P	T	Total	Theory			Practical		Total
											ISE	CIE	ESE	CIE/ TW	ESE (P/OR)	
PCC	MDS1-001	Mathematical Foundations for Computing	2	-	1	3	2	--	1	3	20	30	50	--	--	100
PCC	MDS1-002	Advanced Machine Learning	3	-	-	3	3	--	--	3	20	30	50	--	--	100
PCC	MDS1-003	Advanced Data Management and Warehousing	4	-	-	4	4	--	--	4	20	30	50	--	--	100
HSS	MDS1-004	Research Methodology	3	-	-	3	3	--	--	3	20	30	50	--	--	100
PEC	MDS1-005	Program Elective - I	3	-	-	3	3	--	--	3	20	30	50	--	--	100
VSEC	MDS1-006	Laboratory Proficiency - I	-	4	-	4	--	2	--	2	--	--	--	25	25	50
HSS	MDS1-007	Research and Technical Communication Lab	-	2	-	2	--	1	--	1	--	--	--	25	--	25
VSEC	MDS1-008	Programming Proficiency Lab	-	2	-	2	--	1	--	1	--	--	--	25	--	25
		Total	15	8	1	24	14	4	2	20	100	150	250	75	25	600

MDS1-005: Program Elective I	
MDS1-005A: IoT Architecture & Protocols	MDS1-005B: Information Security
MDS1-005C: Natural Language Processing	MDS1-005D: Distributed Operating System

MDS1-006: Laboratory Proficiency – I
Laboratory Assignments will be based on Program Core and Program Elective Courses.

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

Semester – II

Broad Category of Course			Teaching Scheme (Hours/Week)				Credits/ Grades				Examination Scheme and Marks					
	Subject code	Name of subjects	L	P	T	Total	L	P	T	Total	Theory			Practical		Total
											ISE	CIE	ESE	CIE/ TW	ESE (P/OR)	
PCC	MDS2-009	Data Modelling and Visualization	3	-	-	3	3	-	-	3	20	30	50	-	-	100
PCC	MDS2-010	Generative and Responsible AI	4	-	-	4	4	-	-	4	20	30	50	-	-	100
PEC	MDS2-011	Program Elective - II	3	-	-	3	3	-	-	3	20	30	50	-	-	100
OE	MDS2-012	Open Elective - I(MOOCs)	4	-	-	4	4	-	-	4	20	30	50	-	-	100
VSEC	MDS2-013	Laboratory Proficiency - II	-	6	-	6	-	3	-	3	-	-	-	50	50	100
VSEC	MDS2-014	Skill Enhancement Lab	-	2	-	2	-	1	-	1	-	-	-	25	-	25
AEC	MDS2-015	Seminar-I	-	4	-	4	-	2	-	2	-	-	-	50	25	75
		Total	14	12	0	26	14	6	0	20	80	120	200	125	75	600

Program Elective - II/MOOCs: MDS2-011	
MDS2-011A: Edge Computing and IoT Applications	MDS2-011B: Information Retrieval and Web mining
MDS2-011C: Multimodal Computing	MDS2-011D: Advanced Compilers
MDS2-011E: Applied Security	

Open Elective - I (MOOCs): MDS2-012	**MOOC link**
MDS2-012A: Computer Vision	https://onlinecourses.nptel.ac.in/noc19_cs58/preview
MDS2-012C: Mobile Computing	https://nptel.ac.in/courses/106106147
MDS2-012E: Software Architecture	https://onlinecourses.nptel.ac.in/noc22_cs39/preview

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

Semester – III

Broad Category of Course			Teaching Scheme (Hours/Week)				Credits/ Grades				Examination Scheme and Marks					
	Subject code	Name of subjects	L	P	T	Total	L	P	T	Total	Theory			Practical		Total
											ISE	CIE	ESE	CIE/ TW	ESE (P/OR)	
PEC	MDS3-016	Program Elective - III	4	-	-	4	4	-	-	4	20	30	50	-	-	100
OE	MDS3-017	Open Elective - II (MOOCs)	4	-	-	4	4	--	-	4	20	30	50	-	-	100
IKS	MDS3-018	Indian Knowledge System and Human Values	1	2	-	3	1	1	-	2	-	30	-	20	-	50
ISS	MDS3-019	Internship/Field study	-	8	-	8	-	4	-	4	-	-	-	50	50	100
AEC	MDS3-020	Seminar – II	-	4	-	4	-	2	-	2	-	-	-	50	50	100
AEC	MDS3-021	Dissertation Stage -I	-	8	-	8	-	4	-	4	-	-	-	50	50	100
		Total	9	22	0	31	9	11		20	40	90	100	170	150	550

MDS3-016: Program Elective – III	
MDS3-016A: Applied Deep Learning	MDS3-016B: Pattern Recognition
MDS3-016C: Big Data Analytics	MDS3-016D: Business Intelligence
MDS3-016E: Secure Software Development	

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

MDS3-017 Open Elective - II (MOOCs)	**MOOC link**
MDS3-017A: Industrial IoT	**Introduction to Industry 4.0 and Industrial Internet of Things** By Prof. Sudip Misra | IIT Kharagpur https://onlinecourses.nptel.ac.in/noc20_cs69/preview
MDS3-017B: Bioinformatics	**BioInformatics: Algorithms and Applications** By Prof. Michael Gromiha | IIT Madras https://onlinecourses.nptel.ac.in/noc21_bt06/preview
MDS3-017C: AI in Healthcare	Coursera-AI in Healthcare Specialization [5 courses] (Stanford) | Coursera **NPTEL-Applied Accelerated Artificial Intelligence** Prof. Satyajit Das, Prof. Satyadhyan Chickerur, Prof. Bharatkumar Sharma, Prof. Adesuyi Tosin, Prof.Ashrut Ambasth, IIT Palakkad, KLE Technological University, NVIDIA, NVIDIA https://onlinecourses.nptel.ac.in/noc22_cs83/preview
MDS3-017D: AI for Investments	Artificial Intelligence (AI) for Investments By Prof. Abhinava Tripathi, IIT Kanpur https://onlinecourses.nptel.ac.in/noc23_mg63/preview
MDS3-017E: Responsible and Safe AI	Responsible & Safe AI Systems By Prof. Ponnurangam Kumaraguru, Prof. Balaraman Ravindran, Prof. Arun Rajkumar IIIT Hyderabad, IIT Madras https://onlinecourses.nptel.ac.in/noc24_cs132/preview
MDS3-017F: AI for Economics	Artificial Intelligence for Economics Prof. Adway Mitra, Prof. Dripto Bakshi, Prof. Palash Dey, IIT Kharagpur https://onlinecourses.nptel.ac.in/noc24_cs76/preview
MDS3-017G: ML for Soil and Crop Management	Machine Learning For Soil And Crop Management Prof. Somsubhra Chakraborty, IIT Kharagpur https://onlinecourses.nptel.ac.in/noc22_ag05/preview
MDS3-017H: Robotics	Robotics Prof. Dilip Kumar Pratihar, IIT Kharagpur https://onlinecourses.nptel.ac.in/noc21_me76/preview
MDS3-017I: Ethical Hacking	Ethical Hacking Prof. Indranil Sen Gupta, IIT Kharagpur https://onlinecourses.nptel.ac.in/noc22_cs13/preview
MDS3-017J: AR/VR	Introduction to Virtual Reality Ramesh C Sharma, Dr B R Ambedkar University Delhi, New Delhi onlinecourses.swayam2.ac.in/nou23_ge34/preview

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

Semester – IV

Broad Category of Course			Teaching Scheme (Hours/Week)				Credits/ Grades				Examination Scheme and Marks					
	Subject code	Name of subjects	L	P	T	Total	L	P	T	Total	Theory			Practical		Total
											ISE	CIE	ESE	CIE/ TW	ESE (P/OR)	
AEC	MDS4-022	Seminar – III	-	8	-	8	-	4	-	4	--	--	--	50	50	100
AEC	MDS4-023	Dissertation Stage – II	-	32	-	32	-	16	-	16	--	--	--	100	50	150
		Total	-	40	-	40	-	20	-	20	--	--	--	150	100	250

Guidelines

Program Elective Course
Students may select any one of the courses or NPTEL MOOCs from the list recommended by the department. The total credits earned through MOOCs should be equivalent to the allocated credits for the respective Elective. (One credit will be awarded for a four-week MOOCs). MOOCs list will be updated at the beginning of the academic year.
Open Elective Courses
• Students may select any one of the courses of 4 credits offered by any other department in the institute or Industry supported Course or MOOCs. • The total credits earned through MOOCs should be equivalent to the allocated credits for the respective Open Elective Course. (One credit will be awarded for a four week MOOCs). MOOCs list will be updated at the beginning of the academic year.
Exit Criteria:
• Post graduate diploma will be awarded if a student completes 40 credits and wishes to exit after first year of PG. A student will be allowed to enter/re-enter only at the odd semester and can only exit after the even semester.

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

Abbreviations:

PCC	Program Core Course	**L**	Lecture	**ISE**	In Semester Examination
PEC	Program Elective Course	**P**	Practical	**ESE**	End Semester Examination
OEC	Open Elective Course	**CIE**	Continuous Internal Evaluation	**AEC**	Ability Enhancement Course
VSEC	Vocational Skill Enhancement Course	**TW**	Term Work	**MOOC**	Massive Open Online Course
HSS	Humanities/IKS/Research	**OR**	Oral	**T**	Tutorial

CIE [30]	***Continuous Internal Evaluation:*** (Activity Based Learning Evaluation) The department shall declare the set of all applicable activities such as Problem Based Learning, Quizzes, field work, group discussion, but not limited to etc. The course coordinator, in consultation with course teachers, shall select any of three to four activities suitable for the course from the list declared by the department and get the selected activities approved from HoD. The Course teacher shall get the activities carried out by students, evaluate the student performance based on the prescribed rubrics. Department shall prepare the rubrics for all the activities and display the same before the commencement of academics.
ISE [20]	***In-Semester Examination:*** Written examination shall be conducted for one hour duration on First Module for 20 marks.
ESE [50]	***End-Semester Examination:*** Written examination shall be conducted for three hours duration on Modules II, III, IV for 50 marks.

Semester I

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

Semester I

MDS1-001 Mathematical Foundations for Computing		
Teaching Scheme: L: 02 Hrs./week T: 01 Hr/week	**Credits**: 03	**Examination Scheme:** ISE: 20 marks CIE: 30 marks ESE: 50 marks
Prerequisite: Fundamentals of Mathematics		

Course Objectives:
Course intends to prepare the students

1. To explore basics of discrete mathematics for computing in computer science/data science applications.
2. To explore various data structures and algorithmic solutions for storage and retrieval of data computational problems.
3. To explore statistical measures and hypotheses for computing.
4. To understand the use of probability for solving computer science/data science problems.
5. To understand the use of linear algebra and calculus for solving computational problems.

Course Outcomes:
At the end of the course students will be able to

1. **Apply** discrete mathematics concepts like sets, relations to **solve** computing problems
2. **Choose** suitable tree and graph data structures and its algorithms for solving problems in various domains.
3. **Analyze** various central tendency, dispersion, and relationship measures for preprocessing data and use hypothesis testing.
4. **Select** probability distributions functions and probabilistic models for clustering and classification problems.
5. **Apply** linear algebra methods for reducing dimensionality and evaluating multi variable linear systems.

Course Contents

Module I	**Discrete Mathematics for Computing** **Set Theory:** Elementary Theory of Sets, Set Rules & Sets Combinations. **Relations**: Binary Relation, Equivalence Relation, Composite Relation, Partial Ordering Relation. **Fundamentals of Graph Theory**: Types, graph representation methods, Traveling-salesman problem, applications. **Trees & Search Trees**: balanced search trees, Multiway balanced search trees, splay tree, Trie, KD Tree Heaps: Binomial Heaps, Fibonacci Heap. **Case Study**: Mathematical modeling for applications of Graph Electrical network problem/ Flow graph notation/ Test case generation using graphs/Trie for spell checker/autocomplete string/ KD Tree for geographic information systems (GIS).	**8 Hrs.**
	Statistical Inference Types of Statistical Inference, Descriptive Statistics, Inferential Statistics, Importance of Statistical Inference in computing. Measures of Central Tendency: Mean, Geometric mean, Harmonic mean, Median, Mode, Midrange, Measures of Dispersion: Range, inter	**8 Hrs.**

Module II	quartile range, Variance, Mean Deviation, Standard Deviation. Coefficient of variation: Moments, Skewness, Kurtosis, Pearson Correlation. Measures of relationship: Covariance, Pearson Correlation. Hypothesis testing: **Parametric Tests:** Means, for Differences between Means, for Comparing Two Related Samples of Proportions, for Difference between Proportions, for Comparing a Variance to Some Hypothesized Population Variance. Chi-square Test, and Analysis of variance and covariance. **Nonparametric Tests:** Sign Test, Wilcoxon (Signed Rank). **Case study**: Study datasets like Iris/Diamond/Churn and apply suitable central tendency, dispersion and relationship measures to preprocess the datasets. Use of open-source statistical tools for hypothesis testing.	
Module III	**Probability Theory** Elements of Probability Theory: Basic Concept, Conditional Probability. Random variables: Cumulative Distribution Function, Expectation and Variance of Random variables, Covariance & Correlation. Probability Distributions: Discrete Uniform Distribution, Bernoulli, Binomial, Poisson, Geometric, and Hyper geometric Distributions. Continuous Distribution: Continuous Uniform Distribution, Normal Distribution, Exponential Distribution. Inductive statistics: Maximum Likelihood Estimation. Bayes' Theorem, Joint Probability, Bayesian network. Markov model. **Case Study**: Bayesian Network for Spatial data/Markov model for product lifecycle	**8 Hrs.**
Module IV	**Linear Algebra & Vector Calculus** Matrix and vector algebra, systems of linear equations using matrices, linear independence, Matrix factorization concept/LU decomposition, Eigenvalues and eigenvectors. Understanding of calculus: concept of function and derivative, Multivariate calculus: concept, Partial Derivatives, chain rule. **Case Study**: System of linear equations for Electrical Networks/Eigen values & Eigen vector for Pagerank and PCA algorithm/ Gradient Descent	**8 Hrs.**
	Case Studies of industry relevance/recent trends	**5 Hrs.**

Text Books

1. K.H. Rosen, "Discrete Mathematics and its application," Tata McGraw Hill8th edition, ISBN 978-1-259-67651-2.
2. Thomas H. Cormen, Charles E. Leiserson, Ronald L., Rivest Clifford Stein, " Introduction to Algorithms," The MIT Press 4th Edition, ISBN 9780262046305.
3. Richard A. Johnson, Irwin Miller, John Freund, "Probability and Statistics for Engineers," Pearson Education 9th Edition, ISBN 978-0-321-98624-5.
4. Peter Brass, "Advanced Data Structures," Cambridge University Press, ISBN: 978-1-107-43982-5 ISBN 978-0-511-43685-7.

Reference Books

1. Tremblay and Manohar, "Discrete Mathematical Structures with Applications to Computer Science," Tata McGraw Hill Edition 1997, ISBN-13: 978-0-07-463113-3.

2. Irwin Miller, Marylees Miller, "Mathematical Statistics with Applications," Pearson Education 8th Edition, ISBN -9789332519053.
3. Joe Mott, Abraham Kandel, "Discrete Mathematics and its application," Prentice Hall, 2nd Edition, ISBN 978-81-203-1502-0.
4. David C. Lay, Steven R. Lay, and Judi J. McDonald, "Linear Algebra and its applications," Pearson Publication, 5th Edition, ISBN 978-0-321-98238-4.

Paper References:

1. Christopher Krapu, Robert Stewart, and Amy Rose. 2023, "A Review of Bayesian Networks for Spatial Data," in ACM Transaction on Spatial Algorithms System, Vol.9, Issue 1, Article 7 (January 2023), 21 pages. https://doi.org/10.1145/3516523.
2. L. Wan, F. Xia, X. Kong, C. -H. Hsu, R. Huang and J. Ma, "Deep Matrix Factorization for Trust-Aware Recommendation in Social Networks," in IEEE Transactions on Network Science and Engineering, vol. 8, no. 1, pp. 511-528, 1 Jan.-March 2021, doi: 10.1109/TNSE.2020.3044035.
3. Emmert-Streib, Frank, and Matthias Dehmer, "Understanding Statistical Hypothesis Testing: The Logic of Statistical Inference," in MDPI Journal of Machine Learning and Knowledge Extraction, vol. 1, no. 3, pp. 945-961, doi: https://doi.org/10.3390/make1030054.

Relevant MOOCs Course

1. Introduction to Probability Theory and Statistics By Prof. S Dharmaraja, IIT Delhi
   https://onlinecourses.nptel.ac.in/noc22_ma81/preview
2. Discrete Mathematics – IIITB By Prof. Ashish Choudhury IIIT Bangalore
   https://onlinecourses.nptel.ac.in/noc21_cs36/preview
3. Graph Theory By Prof. S.A. Choudum Department of Mathematics, IIT Madras
   https://archive.nptel.ac.in/courses/111/106/111106050/

Other Resources/Links

1. https://ocw.mit.edu/courses/18-06-linear-algebra-spring-2010/video_galleries/video-lectures/

Home

MDS1-002 Advanced Machine Learning

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

Teaching Scheme: **L: 03 Hrs. /week**	**Credits:** 03	**Examination Scheme:** ISE: 20 marks CIE: 30 marks ESE: 50 marks

Prerequisite: Mathematics, Data Mining.

Course Objectives:
Course intends to prepare the students

1. To understand basic concepts and possible applications of machine learning.
2. To study and illustrate various data preprocessing methods and Supervised, Unsupervised machine learning techniques.
3. To study the basic principles of reinforcement learning
4. To understand the basic principles of optimization and its importance in machine learning.

Course Outcomes:
At the end of the course students will be able to

1. **Describe** the basic concepts and possible applications of machine learning.
2. **Evaluate** classification, regression and clustering methods for real world applications for specified data.
3. **Apply** reinforcement algorithms to real world problems.
4. **Solve** Optimization problems in machine learning

Course Contents		
Module I	**Introduction to Machine Learning and Regression**	**8 Hrs.**
	Models of Machine learning: Geometric model, Probabilistic Models, Logical Models, Grouping and grading models, Parametric and non-parametric models **Features:** Concept of Feature, Feature Extraction, Feature Selection **Data Preprocessing and Dimensionality Reduction:**Data integration, data reduction, concept hierarchy generation, PCA, LDA **Regression:** Introduction, Multivariate Linear regression, Logistic regression, Lasso and Ridge Regression **Assessing performance of regression –** Error measures, Overfitting.	
Module II	**Classification and Clustering algorithms**	**8 Hrs.**
	Classification: Classification concepts, Binary and multi-class classification, K nearest neighbor, Decision Tree Representation, Alternative measures for selecting attributes, ID3 algorithm, Naïve Bayes, Support Vector Machines as a linear and non-linear classifier. **Ensemble methods** Introduction, Bagging, Boosting and Adaboost, Random Forest. **Clustering:** Introduction, Distance based clustering- K-means algorithm, Hierarchical Clustering-Agglomerative and Hierarchical, Expectation Maximization methods, Choosing number of clusters- silhouettes. **Performance measure metric:** Mean Square Error, R-squared, Precision, Recall, F1-score, Accuracy.	
	Reinforcement learning	**8 Hrs.**

Module III	Elements of Reinforcement Learning, Model-Based Learning, Dynamic programming, Monte Carlo, Temporal Difference Learning, Generalization. Bandit algorithms: Upper Confidence Bound algorithm. **Case Study:** Incremental Document Classification		
Module IV	**Optimization Algorithms**		**8 Hrs.**
	Convex Functions, First and Second Order Conditions for Optimisations, Convex and Non-Convex Optimisation, problems in Machine Learning. Gradient Descent: math, programming basic optimisation problems and their solutions, Variants of Gradient Descent: Projected, Stochastic, Proximal, Accelerated, Coordinate Descent. **Case Study:** Stochastic Gradient Descent for Employee Attrition		
	Case Studies of industry relevance/recent trends	**5 Hrs.**	

Text Books

1. Ethem Alpaydin, Introduction to Machine Learning, PHI, 2nd edition, 2013, ISBN : 978-0-262-01243-0
2. Kevin Murphy, Machine Learning: A Probabilistic Approach, MIT Press, 1st Edition, 2012, ISBN: 978-0262-30616-4
3. C. M. Bishop, Pattern Recognition and Machine learning, Springer, 1st Edition, 2013, ISBN.: 978-81-322-0906-5
4. Josh Patterson, Adam Gibson, "Deep Learning: A Practitioner's Approach," O'Reilly Media, Inc, 2017, ISBN No.: 9781491914250

Reference Books

1. Jiawei Han, Micheline Kamber, and Jian Pie, "Data Mining: Concepts and Techniques," Elsevier Publishers Third Edition, ISBN: 9780123814791, 9780123814807
2. Hastie, Trevor, et al., —"The elements of statistical learning: Data mining, inference, and prediction," Vol. 2. New York: springer, 2009.
3. Peter Flach, "Machine Learning: The Art and Science of Algorithms that make sense of data," Cambridge University Press, 1st Edition, 2012, ISBN: 978-1-316-50611-0
4. Hastie, Tibshirani, Friedman, "Introduction to statistical machine learning with applications in R," Springer, 2nd Edition, 2013, ISBN: 978-1-4614-7138-7
5. Tom Mitchell, "Machine Learning," McGraw Hill, 1997, ISBN: 0-07-042807-7
6. Parag Kulkarni, "Reinforcement and Systemic Machine learning for Decision Making," Wiley-IEEE Press, 2012, ISBN: 978-0-470-91999-6
7. S. Boyd and L. Vandenberghe, Convex optimization, Cambridge university press, 2004.

Paper References

1. Gou, Jianping, Wenmo Qiu, Zhang Yi, Yong Xu, Qirong Mao, and Yongzhao Zhan. "A local mean representation-based K-nearest neighbor classifier." ACM Transactions on Intelligent Systems and Technology (TIST) 10, no. 3 (2019): 1-25.
2. Mudumba, Bharadwaj, and Md Faisal Kabir. "Mine-first association rule mining: An integration of independent frequent patterns in distributed environments." Decision Analytics Journal (2024): 100434.

3. Taye, Mohammad Mustafa. 2023. "Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions" Computation 11, no. 3: 52. https://doi.org/10.3390/computation11030052

Relevant MOOCs

1. Introduction to Machine Learning by Sudeshna Sarkar IIT Kharagpur
   https://nptel.ac.in/courses/106105152
2. Introduction to Machine Learning by Dr. Balaraman Ravindran, IIT Madras
   https://nptel.ac.in/courses/106106139
3. Machine Learning with Python by IBM
   https://cognitiveclass.ai/courses/machine-learning-with-python
4. Deep Learning by Prof. Prabir Kumar Biswas,IIT Kharagpur
   https://onlinecourses.nptel.ac.in/noc20_cs62/preview
5. Optimisation for Machine Learning : Theory and Implementation, IIT Delhi
   https://nptel.ac.in/courses/106106245

Home

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

MDS1-003 Advanced Data Management and Warehousing		
Teaching Scheme: L: 04 Hrs./Week	**Credits:** 04	**Examination Scheme:** ISE: 20 marks CIE: 30 marks ESE: 50 marks
Prerequisite: Discrete Mathematics, Database Management System		

Course Objectives:

Course intends to prepare the students

1. To understand the database environment and development process.
2. To study database design and its types.
3. To understand data warehouse architecture and enterprise-wide applications.
4. To understand advances in Data warehouse.

Course Outcomes:

At the end of the course students will be able to

1. **Design** a suitable database in a parallel and distributed environment.
2. **Apply** appropriate normal forms and implement queries for given applications.
3. **Design** a data warehouse model and apply big data descriptive analytics for real-world application.
4. **Illustrate** recent developments in the Data warehouse.

Course Contents

Module I	**The Database Environment and development process** Basic concepts: Data, Database System Structure, Types of databases like RDBMS, OORDBMS, NoSQL: document-oriented, graph. The Database approach: Data models, The components of database environment. The database development process: System development life cycle. Evolution of database system. Database application: Personal, multitier and Enterprise applications. Case study: Design and Develop data modeling for any suitable application.	**9 Hrs.**
Module II	**Database design** Enhanced ER model, universal data model, Relational Database design, Normalization during logical database design and reverse engineering Advanced Normal Forms. Implementation: Basic and Advanced SQL Parallel Databases: I/O Parallelism, Inter and Intra Query Parallelism, Inter and Intra Operation Parallelism. Distributed DBMS: Location transparency, synchronous and asynchronous, local and global transaction, commit protocol	**10 Hrs.**
	Big Data Descriptive Analytics	

Module III	Master Data Management, Need for Data warehousing: Escalating need for strategic information, failure of past system, operational versus decision support systems.Data Integration for Data warehousing :ETL process, Data Transformation. Descriptive analytics using data warehouse, Online analytical processing tools, SQL OLAP queries, MOLAP and ROLAP. Predictive and prescriptive analytics Case Study- Design a data warehouse for a suitable system using an open-source tool.	**9 Hrs.**
Module IV	**Advanced topics in Data Warehouse** Spatial and temporal Data Warehouse: General concepts of spatial databases, Conceptual modeling of spatial data warehouse, Manipulating temporal information in SQL, conceptual design of temporal data warehouse Graph data warehouse: Graph Data Models, Property Graph database system. Case Studies of industry relevance/recent trends	**9 Hrs.**

Text Books

1. Abraham Silberschatz, Henry F. Korth, and S. Sudarshan,Database System Concepts, McGraw-Hill Education, 2010, 6th Edition, ISBN:9332901384
2. Jeff Hoffer, Ramesh Venkatraman, Heikki Topi, Modern Database Management, Twelfth Edition, Perason, 2016, ISBN: 10-1-292-10185-7
3. Jiawei Han & Micheline Kamber,"Data Mining–Concepts and Techniques," Morgan Kaufmann Publishers, Elsevier,2006.2nd Edition ISBN: 9780123814791
4. Paulraj Ponnaiah, "Data Warehousing Fundamentals" ,Wiley, 2nd Edition, ISBN: 8126537299

Reference Books

1. Connolly, T. M., & Begg, C. E. (2005). "Database systems: a practical approach to design, implementation, and management." Pearson Education, ISBN: 9788131720257
2. Inmon, William, "Building the data warehouse," H.. John wiley& sons, 2005,3rd Edition, ISBN: 0-471-08130-2
3. Vaisman, A., & Zimányi, E. (2014). Data Warehouse Systems: Design and Implementation, Springer, ISBN: 978-3-642-54655-6

Paper References:

1. L. Sautot, S. Bimonte and L. Journaux, "A Semi-Automatic Design Methodology for (Big) Data Warehouse Transforming Facts into Dimensions," in IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 1, pp. 28-42, 1 Jan. 2021, doi: 10.1109/TKDE.2019.2925621.
2. H. Akid, G. Frey, M. B. Ayed and N. Lachiche, "Performance of NoSQL Graph Implementations of Star vs. Snowflake Schemas," in IEEE Access, vol. 10, pp. 48603-48614, 2022, doi: 10.1109/ACCESS.2022.3171256.
3. D. Wang, L. Liu and Y. Liu, "Normalized Storage Model Construction and Query Optimization of Book Multi-Source Heterogeneous Massive Data," in IEEE Access, vol. 11, pp. 96543-96553, 2023, doi: 10.1109/ACCESS.2023.3301134.

Relevant MOOCs Course

1. NPTEL 12 weeks course on "Database Management System" by By Prof. Partha Pratim Das, Prof. Samiran Chattopadhyay | IIT Kharagpur Name: Dr. Parth Das https://onlinecourses.nptel.ac.in/noc19_cs46/
2. Coursera 4 weeks course on " Data Warehouse Concepts, Design, and Data Integration" by University of Colorado System Instructor Name : Michael Mannino https://www.coursera.org/learn/dwdesign
3. Coursera 4 weeks course on " Relational Database Support for Data Warehouses," offered by University of Colorado System Instructor Name : Michael Mannino https://www.coursera.org/learn/dwrelational
4. NPTEL course on "Data Mining"offered by IIT KharagpurInstructor Name: Prof.PabitraMitra https://nptel.ac.in/courses/106105174/

Other Resources/Links

1. https://nifi.apache.org/
2. https://aws.amazon.com/redshift/
3. https://www.sas.com/
4. https://www.talend.com/

Home

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

MDS1-004 Research Methodology		
Teaching Scheme: L: 03 Hrs./week	**Credits:** 03	**Examination Scheme:** ISE: 20 marks CIE: 30 marks ESE: 50 marks
Prerequisite: Mathematical foundations		

Course Objectives:
Course intends to prepare the students

1. To understand the philosophy of research in general.
2. To understand diverse aspects of research methodologies.
3. To enhance proficiency in both verbal and written presentation abilities.
4. To get familiar with Intellectual Property Rights (IPR) and plagiarism.

Course Outcomes:
At the end of the Course Students will be able to

1. **Develop** the significance of research and methodologies to collect and analyze data.
2. **Formulate** hypotheses through experimentation.
3. **Design** a comprehensive research report.
4. **Distinguish** various forms of intellectual properties.

Course Contents

Module I	**Introduction to RM, defining the research problem and research design** Why do research? Meaning, objectives, Research types: computer science and engineering applications specific to nature of research in system building, research process, Criteria of good research. Formulating a research problem: Selecting a research problem, Writing research Objectives Conceptualizing a research design: Meaning, Need and Basic principles of research design.	**8 Hrs.**
Module II	**Literature survey, Data Collection and Sampling methods** Literature Survey: Importance, planning, locating relevant literature, managing information, and writing a review. Methods for primary and secondary data collection, data processing operations, measures for analyzing data, basics of sampling theory, concept of standard error, estimations on populations, determining sample size.	**8 Hrs.**
Module III	**Technical Content Writing** **Report writing** significance and steps. **Types of technical content publishing:** White paper, journal, conference, poster, short paper etc. **Writing various technical papers:** survey paper, journal paper, indexing agencies, COPE.	**8 Hrs.**
Module IV	**Intellectual Property Rights** **Intellectual property**: types of IP, IPR in India and abroad.	**8 Hrs.**

	Steps in patenting: Searching prior art, preparing the patent application, claims, filing, prosecution, objections, appeal, issuance or rejection of patent. Plagiarism and Research ethics. **Research Tools:** Plagiarism checking tools, Grammer checker tools	
	Discussion on research by eminent researchers and their research work with citation analysis and related patents and IPR	**5 Hrs.**

Text Books

1. Ranjit Kumar, Research Methodology- a step-by-step guide for beginners, SAGE Publication,4th edition, ISBN: 9789351501336, 2014
2. C. R. Kothari, "Research Methodology- Methods and Techniques," New Age International Publishers, 4th edition, ISBN: 9789386649225, 2019.

Reference Books

1. David V. Thiel, Research Methods for Engineers, Cambridge University Press., 2nd edition, ISBN: 9781139542326, 2017
2. C.G. Thomas, "Research methodology and scientific writing," Springer, 2nd edition, ISBN: 978-3-030-64864-0, 2021
3. Professional Programme Intellectual Property Rights, Law and Practice, The Institute of Company Secretaries of India, Statutory Body Under an Act of Parliament, September 2013: https://www.icsi.edu/media/webmodules/IPRLP_NOV29.pdf

Paper References:

1. Bayoudh, K., Knani, R., Hamdaoui, F. *et al.* A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. The Visual Computer, 38(8), 2939–2970 (2022). https://doi.org/10.1007/s00371-021-02166-7
2. Taherdoost, H. (2016). Sampling methods in research methodology; how to choose a sampling technique for research. How to Choose a Sampling Technique for Research (April 10, 2016).
3. Romer, Paul. When should we use intellectual property rights?. American Economic Review 92.2 (2002): 213-216.
4. Wang, Yuntao, et al. A survey on ChatGPT: AI-generated contents, challenges, and solutions. IEEE Open Journal of the Computer Society (2023).

Relevant MOOCs

1. Research Methodology :Prof. Edamana Prasad, Prof. Prathap Haridoss, IIT Madras
   https://onlinecourses.nptel.ac.in/noc24_ge21/preview
   Research Methodology :Prof. Soumitro Banerjee, IISER Kolkata
   https://onlinecourses.nptel.ac.in/noc22_ge08/preview
2. Introduction to Statistical Hypothesis Testing: Dr. Arun Tangirala IIT Madras
   https://nptel.ac.in/courses/103106120
3. Roadmap for patent creation: Dr. Gouri Gargate IIT Kharagpur
   https://nptel.ac.in/courses/127105008

Other Resources/Links

1. The world intellectual Property Organization: https://www.wipo.int/portal/en/index.html
2. IP India: http://www.ipindia.nic.in/
3. Cell For IPR Promotion and Management : http://cipam.gov.in/
4. Draft patent rules: http://cipam.gov.in/wp-content/uploads/2018/12/Draft-Patent-Rules-2018.pdf

5. Manual of Patent Office Practice and Procedure: http://www.ipindia.nic.in/writereaddata/Portal/Images/pdf/Manual_for_Patent_Office_Practice_and_Procedure_.pdf
6. WIPO IPR Resources: https://www.wipo.int/reference/en/

Home

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

Program Elective - I

MDS1-005A IoT Architecture and Protocols		
Teaching Scheme: L : 03 Hrs. / Week	**Credits**: 03	**Examination Scheme:** ISE: 20 marks CIE: 30 marks ESE: 50 marks
Prerequisite: Computer Networks, Data Communication, Wireless Sensor Network		

Course Objectives:
Course intends to prepare the students

1. To understand the fundamentals of embedded systems and IoT.
2. To learn methodologies for IoT application development.
3. To study architecture and protocol standards of IoT systems.
4. To learn real world application scenarios of IoT using real time example case studies.

Course Outcomes:
At the end of the course students will be able to

1. **Identify** the need of IoT and embedded systems.
2. **Illustrate** integration of sensors with embedded devices like Raspberry Pi/BBB/Arduino.
3. **Distinguish** different IoT application layer protocols.
4. **Design** Cloud based IoT applications using embedded devices.

Course Contents

Module I	**Introduction to embedded systems and IoT** Introduction to Embedded systems: Introduction to Embedded systems, Sensors and actuators, Challenges. IoT: Introduction, characteristics, application areas, Structure of IoT Applications and Functional blocks of IoT IoT enabling technologies: Cloud computing, Big data analytics, WSN, advantages and disadvantages of IoT with challenges.	**8 Hrs.**
Module II	**IoT Architectures and Components** IoT Architectures: Introduction, Functional View, Information View, Deployment and Operational View, Other Relevant architectural views. Hardware and software Components of IoT: Introduction to Arduino, Raspberry pi, NodeMCU, BBB and interfacing. Designing IoT applications using embedded devices.	**8 Hrs.**
Module III	**IoT Data Link Layer and Network Layer protocols** IoT reference model, IoT Levels and Deployment templates. Physical/Link Layer, IEEE802.15.4 and ZigBee, Low-power Wi-Fi, Bluetooth and BLE. Network Layer IP Based Protocols, The 6LoWPAN Adaptation Layer and RPL.	**8 Hrs.**
Module IV	**IoT application layer protocols**	

	Application Layer protocols: MQTT, CoAP, AMQP, XMPP, WebSocket. Integration of IoT with cloud and REST Architecture. IoT Security challenges, Privacy Issues in the IoT , The Role of Authorization, IoT-OAS: Delegation-based Authorization for the Internet of Things, MQTTs, HTTPs. **Case Study** : Home automation system/Smart irrigation system/Remote health monitoring using IoT	**8 Hrs.**
	Case Studies of industry relevance/recent trends	**5 Hrs.**

Text Books

1. Cirani, Simone, Gianluigi Ferrari, Marco Picone, and Luca Veltri "Internet of Things: Architectures, Protocols and Standards" John Wiley & Sons, 2018. DOI: 10.1002/9781119359715,ISBN: 9781119359678
2. Arshdeep Bahga and Vijay Madisetti "Internet of Things: A Hands-On Approach" January 2014,Edition: 1,Publisher:Orient Blackswan Private Limited Publication - New Delhi.ISBN:978-0996025515
3. Jeeva Jose "Internet of Things," Khanna Publishing House ISBN: 9788195123162, Edition: First, 2022,Pages: 344

Reference Books

1. Hersent, Olivier, David Boswarthick, and Omar Elloumi, second edition, "The Internet of Things: Key Applications and Protocols" ISBN: 978-1-119-99435-0 February 2012, 376 Pages John Wiley & Sons, 2011
2. Bradbury, Alex, and Ben Everard, "Learning Python with Raspberry Pi" John Wiley & Sons, 2014. Raj Kamal "Internet of Things," Second edition, 2022 TMH

Paper References:

1. Lata, N. ,Kumar, R. (2020, November). Internet of Things: A Review of Architecture and Protocols. In 2020 International Conference on Decision Aid Sciences and Application (DASA) (pp. 1027-1031).IEEE.
2. Castellani, Angelo P., et al. "Architecture and protocols for the internet of things: A case study,"2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops). IEEE, 2010.

Relevant MOOCs

1. Swayam NPTEL "Introduction To Internet of Things," by Sudip Mishra, IIT Kharagpur
2. Udemy: A Complete Course on an IOT system - Design and Development
3. Udemy: The Ultimate Guide to IoT with Raspberry Pi and Python -2024

Other Resources/Links:

1. https://www.udemy.com/course/a-complete-course-on-an-iot-system-design-and-development/?couponCode=JUST4U02223
2. https://www.udemy.com/course/internet-of-things-using-raspberry-pi-and-python-2023/?couponCode=JUST4U02223

Home

MDS1-005B Information Security		
Teaching Scheme: L: 03 Hrs./week	**Credits:** 03	**Examination Scheme:** ISE: 20 marks CIE: 30 marks ESE: 50 marks
Prerequisite: Basic Mathematics, Matrix Inverse, Networking Protocols, Sniffing Tools		

Course Objectives:
Course intends to prepare the students

1. To classify threats to assess damages to information systems.
2. To acquire the knowledge of mathematics for cryptography, understand the concepts of cryptography.
3. To examine firewalls in the context of intrusion detection systems (IDS).
4. To learn the strengths & weaknesses of the Indian IT Act along with the amendment to the Act.

Course Outcomes:
At the end of the course students will be able to

1. **Apply** the Euclidean algorithm, Fermat's theorem and Euler's theorem for numerical problems.
2. **Apply** appropriate cryptographic techniques such as symmetric & asymmetric key cryptography.
3. **Evaluate** information security threats & vulnerabilities in information systems using security measures.
4. **Comprehend** the authentication services Kerberos & X.509 directory services.

Course Contents

Module I	**Mathematical Foundations for Cryptography**	**08 Hrs.**
	Modular Arithmetic: Notations & Operations, Divisibility & Division Algorithm, Euler's totient theorem, General formula Phi function, Euclid's method of finding GCD, Extended Euclid's algorithm. Computing Galois field, Factoring, Prime numbers, relative prime, Fermat's theorem, Application of Fermat's little theorem & congruence. Primality testing & its types, Recommended primality test. The Chinese Remainder Theorem, Discrete Logarithm. Applications of mathematics in cryptography.	
Module II	**Elementary Cryptography**	**08 Hrs.**
	Model of Network Security, Network Access Security Model, Security Attacks, Security Services. CIA (Confidentiality Integrity Availability). Classical Encryption Techniques & their cryptanalysis: Stream Ciphers, Substitution Techniques: Caesar, Mono alphabetic Ciphers, Playfair Cipher, Hill Cipher, Polyalphabetic Ciphers, Transposition Techniques, Block Ciphers, Data Encryption Standards (DES), 3DES, Advanced Encryption standard. Public key Cryptography & RSA, Mutual trust: Key Management & distribution, Diffie–Hellman key exchange, Elliptic curve cryptography. Example for a given scenario to identify & justify levels of compromise in CIA Open Source / Trial version tools: cryptool version 2. Study– Sniffing tool, Cain & Abel, iptables, Suricata. Survey papers e.g. Fully Homomorphic Encryption (FHE)	

Module III	**Security to Authenticity and Integrity**	**08 Hrs.**
	Authentication: Requirements, Protocols: One–way Authentication, Mutual Authentication, Centralized Authentication, The Needham–Schroeder Protocol versions, Applications: Kerberos, X.509 Directory Services, Biometrics. Cryptographic Hash functions: SHA3, MD5	
Module IV	**Cybercrimes and Cybersecurity: The Legal Perspective**	**08 Hrs.**
	Wrestling between safeguard & attack. Vulnerabilities: Technology weaknesses, Configuration weaknesses, Security policy weaknesses. Attacks: Reconnaissance, Access, DoS (Denial of Service), Worms, viruses & Trojan horse. Non–Cryptographic Protocol Vulnerabilities: DDoS (Distributed DoS), Session Hijacking & Spoofing, Wireless LAN (Local Area Network) vulnerabilities. IEEE 802.11 wireless LAN security. Cyber Crime, Classification of Cybercrimes, Cyber stalking, The Indian IT Act, Psychology, Personality disorder & its causes. Cybercrime & Cyberterrorism: Social, Political, Ethical & Psychological mindset Dimensions, Cyber offenses: How criminals plan them. Intellectual Property in Cyberspace. Case study: Configure filtering on router / Survey on Data Recovery Tools / Obfuscation & write your own wrappers for it. Cybercrime: Illustrations, Mini–cases.	
	Case Studies of industry relevance/recent trends	**5 Hrs.**

Text Books:

1. William Stallings, "Cryptography and Network Security, Principles and Practice," Pearson, 7th edition, ISBN: 978–1–292–15858.
2. Behrouz A. Forouzan, Debdeep Mukhopadhyay, "Cryptography and Network Security," 2nd edition, TMH, ISBN: 978–00–707–0208–0.
3. Nina Godbole, Sunita Belapure, "Cyber Security - Understanding Cyber Crimes, Computer, Forensics and Legal Perspectives," Wiley India, 1st edition 2011, ISBN: 978–81–265–2179–1.
4. C. K. Shyamala, N Harini, Dr. T. R. Padmanabhan, "Cryptography and Security," Wiley India. ISBN: 978–81–245–2285–9.

Reference Books:

1. Bernard Menezes, "Network Security and Cryptography," Cengage Learning India, 2014, ISBN: 8131513491.
2. Atul Kahate, "Cryptography and Network Security," 3rd edition, McGraw Hill Education.
3. V. K. Pachghare, "Cryptography and Information Security," 2nd edition, PHI Learning, ISBN: 978–81–203–5082–3.
4. Bruice Schneier, "Applied Cryptography– Protocols, Algorithms and Source code in C," 2nd edition, Wiley Indian Edition, India Pvt. Ltd, ISBN 978–81–265–1368–0.
5. Hoffstein, Jeffrey, Pipher, Jill, Silverman, J.H., "An Introduction to Mathematical Cryptography"

6. Vivek Sood, "Cyber Law Simplified,"McGrawHill Education, ISBN: 0070435065.

Paper References

1. Kundan Munjal, Rekha Bhatia; 'A systematic review of homomorphic encryption and its contributions in healthcare industry', Springer Nature, Published: 03 May 2022 Volume 9, pp. 3759–3786, 2023 DOI https://doi.org/10.1007/s40747-022-00756-z
2. Whitfield Diffie and Martin E. Hellman, member IEEE: New Directions in Cryptography, IEEE transactions on Information Theory, Vol. IT-22, no. 6th November 1976 pp. 644-654
3. Alexander Viand, Christian Knabenhans, Anwar Hithnawi; 'Verifiable Fully Homomorphic Encryption', 11th February 2023, arXiv, Cornell University, pp. 1-18, DOI: https://doi.org/10.48550/arXiv.2301.07041

Relevant MOOCs

1. Cryptograpghy and Nework Security Prof. Sourav Mukhopadhyay, IIT Kharagpur https://onlinecourses.nptel.ac.in/noc20_cs21/preview
2. Forensic Linguistics Prof. Deepak Mashru, National Forensic Sciences University (INI) https://onlinecourses.nptel.ac.in/noc24_hs88/preview

Other Resources/Links

1. https://ocw.mit.edu/courses/

Home

MDS1-005C Natural Language Processing		
Teaching Scheme: L: 03 Hrs./ Week	**Credits:** 03	**Examination Scheme:** ISE: 20 marks CIE: 30 marks ESE: 50 marks
Prerequisite: Theory of Computation, Elementary Probability and Statistics, Basic understanding of Machine Learning		

Course Objectives:
Course intends to prepare the students

1. To be familiar with fundamental concepts of Computational Linguistics.
2. To understand various text pre-processing methods like tagging, parsing and semantic analysis.
3. To study neural network models for processing of natural language.
4. To get acquainted with modern approaches and applications in Computational Linguistics.

Course Outcomes
At the end of the course students will be able to

1. **Apply** basic text processing techniques for given input.
2. **Analyze** syntax and semantics of natural language data.
3. **Design** neural language models for NLP application.
4. **Design** machine translation and dialogue system.

Course Contents

Module I	**Introduction** Introduction: Computing, Natural Language Processing and Linguistics, Stages of NLP Text Pre-processing: sentence segmentation, Tokenization, feature extraction, issues in tokenization for different languages, word segmentation, text segmentation, normalization, case folding, Morphology, Stemming: Porters Algorithm, lemmatization. Spelling correction - dynamic programming approach for finding edit distance, N-gram Language Modeling- context sensitive spelling correction probabilistic language model, auto completion prediction.	**08 Hrs.**
Module II	**Tagging, Syntax and Semantics** Sequence Labeling for Parts of Speech, HMM for speech tagging, Viterbi Algorithm, Conditional Random Fields.	**8 Hrs.**

	Syntax: Constituency and dependency parsing, Constituency parser -Syntactic structure, Parsing methodology, Different parsing algorithms, Parsing in case of ambiguity, Probabilistic parsing, CKY algorithm, Issues in parsing, Dependency parsing- Syntactic structure, Parsing methodology, Transition-Based Dependency Parsing, Graph-Based dependency parsing, Evaluation, Co-reference resolution Semantics: Word Senses, Word relations, Word similarity and thesaurus methods, Name Entity Recognition, Word sense disambiguation, WordNet. Lexical and Distributional Semantics - Introduction, models of semantics, applications.	
Module III	**Neural Language Models, Large Language Models** Overview of Machine Learning- Basics of Neural Networks, Feedforward neural Network for NLP classification task, word embeddings, Recurrent Neural Networks (RNNs) as language model, Long short-term memory (LSTMs) network Transformers-Self attention network, Large Language Models: with transformers, Generation by sampling, training transformer. Fine-tuning and Masked language models: Bidirectional transformer encoders, training bidirectional encoders, fine tuning language models	**08 Hrs.**
Module IV	**Applications - Machine Translation and Dialogue System** Language Divergences and Typology, Machine Translation using encoder decoder, Back-Translation, MT Evaluation Chatbots and Dialogue System: Properties of Human Conversation, Frame-Based Dialogue system, Chatbot, Dialogue system design Transfer learning via Prompting, Chain of Thought Prompting, Tree of Thought Prompting and Instruction Tuning, Retrieval Augmented Generation (RAG) techniques, Using vector datastore for RAG	**08 Hrs.**
	Case Studies of industry relevance/recent trends	**5 Hrs.**

Text Books

1. Daniel Jurafsky and James H. Martin, "Speech and Language Processing," 2nd Edition, Prentice Hall, 2008, ISBN: 978-0131873216
2. Allen James, "Natural Language Understanding," 2nd Edition, Benjamin/Cumming, 1994, ISBN: 978-0805303346
3. Chris Manning and Hinrich Schuetze, "Foundations of Statistical Natural Language Processing," MIT Press, ISBN: 978-0262133609
4. Jacob Perkins, "Python 3 text processing with NLTK 3 cookbook," Packt Publishing Ltd., 2014, ISBN-13: 9781782167853

Reference Books

1. Jacob Eisenstein. "Natural Language Processing," The MIT Press, 2019.ISBN: 9780262042840

Paper References:

1. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin, "Attention is all you need," in Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17),pp. 6000–6010,2017,doi:10.5555/3295222.3295349
2. Pradhan, R., Sharma, D.K., "An ensemble deep learning classifier for sentiment analysis on code-mix Hindi–English data," in *Soft Comput* **27**, 11053 (2023),doi:10.1007/s00500-022-07091-y
3. Jessica López Espejel, El Hassane Ettifouri, Mahaman Sanoussi Yahaya Alassan, El Mehdi Chouham, Walid Dahhane, "GPT-3.5, GPT-4, or BARD? Evaluating LLMs reasoning ability in zero-shot setting and performance boosting through prompts," in Natural Language Processing Journal,Volume 5,2023,doi:10.1016/j.nlp.2023.100032
4. Lee, Jooyoung, Fan Yang, Thanh Tran, Qian Hu, Emre Barut, Kai-Wei Chang and Chengwei Su., "Can Small Language Models Help Large Language Models Reason Better?: LM-Guided Chain-of-Thought," *ArXiv* abs/2404.03414 (2024)
5. Ramchandra Joshi, Rusbabh Karnavat, Kaustubh Jirapure, and Raviraj Joshi. 2020. Domain Adaptation of NMT models for English-Hindi Machine Translation Task : AdapMT Shared Task ICON 2020. In Proceedings of the 17th International Conference on Natural Language Processing (ICON), pages 11–16, Patna, India. NLP Association of India (NLPAI).
6. R. Joshi, "L3cube-mahacorpus and mahabert: Marathi monolingual corpus, marathi bert language models, and resources," in Proceedings of the WILDRE-6 Workshop within the 13th Language Resources and Evaluation Conference, 2022, pp. 97–101
7. Facundo Carrillo, Guillermo A. Cecchi, Mariano Sigman, Diego Fernández Slezak, "Fast Distributed Dynamics of Semantic Networks via Social Media,," Computational Intelligence and Neuroscience, vol. 2015,2015,doi: 10.1155/2015/712835

Relevant MOOCs

1. Course Title: Natural Language Processing Specialization offered by Kaiser, Coursera.
   Link: https://www.coursera.org/specializations/natural-language-processing
2. Course Title: Applied Natural Language Processing offered by Swayam, NPTEL
   Link: https://onlinecourses.nptel.ac.in/noc20_cs87/preview

Other Resources/Links

1. https://www.ibm.com/topics/instruction-tuning
2. https://www.ibm.com/products/watsonx-ai

Home

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

MDS1-005D Distributed Operating System		
Teaching Scheme: L: 03 Hrs./week	**Credits:** 03	**Examination Scheme:** ISE: 20 marks CIE: 30 marks ESE: 50 marks
Prerequisite: Fundamentals of Operating System		

Course Objectives:
Course intends to prepare the students
1. To understand Characteristics and Challenges of distributed systems.
2. To learn process management in distributed operating systems.
3. To study various strategies of Shared Memory and Inter Process Communication.
4. To understand Design Principles of Distributed File Management.

Course Outcomes:
At the end of the course students will be able to
1. **Describe** the challenges in distributed OS.
2. **Demonstrate** the process of communication and synchronization.
3. **Design** application to retrieve the data stored in distributed Memory.
4. **Compare** different File Systems.

Course Contents

Module I	**Introduction to Distributed Operating System** Fundamentals: Characteristics and challenges of distributed systems. Design issues in distributed operating systems; Architectural models, Desirable features of good global scheduling algorithms, Task Assignment Approach, Load-Balancing Approach, Load-Sharing Approach.	**8 Hrs.**
Module II	**Process Management and Synchronization** Threads and Process, Process States, Scheduling in DOS, Process Migration. Synchronization: Clock Synchronization, Event Ordering, Mutual Exclusion, Deadlock, Election Algorithms.	**8 Hrs.**
Module III	**Distributed Shared Memory** General Architecture of DSM Systems, Design and Implementation issues in DSM, Consistency Models, Implementing Sequential Consistency Model, Page based distributed shared memory, shared – variable distributed shared memory, object-based distributed shared memory. Replacement Strategy, Thrashing, Heterogeneous DSM, Advantages of DSM.	**8 Hrs.**
Module IV	**Distributed File System and Naming** File-Accessing Models, File-Sharing Semantics, File-caching Schemes, File Replication, Fault Tolerance, Atomic Transactions, Design Principles, Naming: Fundamental Terminologies and Concepts, System-Oriented names, Object-Locating Mechanisms, Human-Oriented names, Name cache, Naming and Security. Case Study : File Management in Linux/Windows/Android (EoS)/Andrew File System File Handling System Call implementation in Rust. Real Time Linux(RT Linux).	**8 Hrs.**
	Case Studies of industry relevance/recent trends	**5 Hrs.**

Text Books
1. Sinha P. K., "Distributed Operating Systems Concepts and Design," PHI, (1997). ISBN-13: 978-9332550254 2. Mukesh Singhal and Niranjan Shivaratri, "Advanced Concepts In Operating Systems" , (July 2017) ,ISBN-13: 978-007047268
Reference Books
1. Tanenbaum A. S., "Distributed Operating Systems," Pearson Education India, 1995,ISBN-13: 9781556353505 2. Ajay D Kshemkalyani, and Mukesh Singhal "Distributed Computing , Principles , Algorithms and systems Computing," (Reprint -2020) , Paperback,ISBN-13: 9780521876346
Paper References:
1. Chia-Han Yang, Valeriy Vyatkin, and Cheng Pang, "Model-Driven Development of Control Software for Distributed Automation: A Survey and an Approach," IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 44, No. 3, March 2014 2. Dacheng Wen, Yupeng Li, Francis C. M. Lau, "A Survey of Machine Learning-Based Ride-Hailing Planning," IEEE Transactions on Intelligent Transportation Systems, DOI: 10.1109/TITS.2023.3345174
Relevant MOOCs
1. Cloud Computing and Distributed Systems, Dr. Rajiv Mishra IIT Patna https://nptel.ac.in/courses/106104182 2. Distributed Systems Dr. Rajiv Mishra IIT Patna https://nptel.ac.in/courses/106106168
Other Resources/Links
1. https://www.rust-lang.org/

Home

MDS1-006 Laboratory Proficiency -I

Teaching Scheme: P: 04 Hrs./Wee	**Credits**: 02	**Examination Scheme:** CIE/TW: 25 Marks ESE(P/OR): 25 Marks

Prerequisite: Knowledge of programming languages, Basics of Python/R/Java

All assignments are compulsory. Each student should implement the assignments individually. Laboratory teachers should make sure that the dataset/code/write up is not the same.

Course Objectives:
Course intends to prepare the students

1. To explore graph theory in computational problem-solving.
2. To use descriptive statistical measures for conducting statistical inference.
3. To explore data design and skills for data warehouse development phases.
4. To examine NoSQL databases for their functionality and applications.
5. To develop IoT applications utilizing embedded devices.
6. To engineer security solutions tailored for real-world applications.
7. To Explore different approaches to Shared Memory and Inter-process Communication.
8. To Comprehend the fundamental principles behind designing Distributed File Management systems.

Course Outcomes:
At the end of the course students will be able to

1. **Apply** Graph Theory for problem modeling and resolution.
2. **Analyze** population characteristics through central tendency, dispersion, and relationships.
3. **Design** schema and execute database queries using NoSQL.
4. **Design** & store data in a data warehouse using open-source tool.
5. **Develop** IoT applications leveraging embedded devices.
6. **Apply** cryptographic techniques and tools for problem-solving applications.
7. **Implement** process and file management system calls using C/Rust.

All assignments are compulsory. Each student should implement the assignment individually. Every student should use a different data set.

Mathematical Foundations for Computing

1. a. Write a program to represent user's information in social networks using graphs. Find the number of associations of each user.

   OR

   b. Design an algorithm to organize the sequence of courses within engineering programs, leveraging graph theory to account for the prerequisite dependencies among courses.
2. Write a program, in the language of your choice, to calculate measures of central tendency, dispersion and relationship of given data.

Advanced Machine Learning

3. Download Email spam classification dataset from Kaggle https://www.kaggle.com/datasets/balaka18/email-spam-classification-dataset-csv.
Email Spam detection has two states: a) Normal State – Not Spam, b) Abnormal State – Spam. Use K-Nearest Neighbors and Support Vector Machine to classify the email. Extract confusion matrix from the test results and compare the performance of both.

4. Download Housing Rent Prediction Dataset from Kaggle and predict the house rent using regression. https://www.kaggle.com/datasets/iamsouravbanerjee/house-rent-prediction-dataset/

Advanced Data Management and Warehousing

5. For an organization of your choice, choose a set of business processes.
   a. Design star/ snowflake schemas for analyzing these processes. Create a fact constellation schema by combining them. Eg. University database, Manufacturing Industry.
   b. Use an open source ETL tool to extract data from different data sources, apply suitable transformations and load into destination tables. Eg. Manufacturing Industry- data of sales, products, purchase.

6. Implement the following NoSQL queries for the product review database.
   c. Create a collection product review database.
   d. Insert 5 documents in the collection.
   e. Find the product list sorted by price of the product from lowest to highest. Skip first two documents.
   f. Delete product review document from the collection where the rating is less than two and product type is 'camera'

IoT Architecture & Protocols

7. Design and develop a home automation security system using embedded devices.
8. Design and develop a remote health monitoring system using IoT devices and ThingSpeak cloud.

Information security

9. Implement the Diffie-Hellman Key Exchange mechanism using HTML & JavaScript. Consider the end user as one of the parties (Alice) & the JavaScript application as an other party (bob). Also demonstrate MITM (man in the middle attack) on it.

10. a. Write a program for Elliptic curve cryptography.

OR

b. Write a program to apply Advanced Encryption Standard (AES) Algorithm for a practical application like URL Encryption

Natural Language Processing

11. Implement complete natural language processing pipeline for Indic language text input.
12. Implement Conversational interface (chatbot) using python.

Distributed Operating System
13. Develop a program in C or Rust to implement Process Management System Calls, which enable the creation, manipulation, and control of processes within an operating system environment. 14. Design and implement a Distributed File System (DFS) like Hadoop Distributed File System (HDFS), capable of storing and managing large volumes of data across multiple nodes in a distributed environment.
15. Mini Project (Based on Core or Elective subjects)

Home

MDS1-007 Research and Technical Communication Laboratory

Teaching Scheme: P: 02 Hrs./Week	**Credits:** 01	**Examination Scheme:** CIE/TW: 25 Marks

Prerequisite: Basic statistics

Course Objectives:
Course intends to prepare the students

1. To understand research papers.
2. To formulate research questions, hypotheses and derive research objectives.
3. To enhance proficiency in both verbal and written presentation abilities.
4. To get familiar with research proposals and patent writing.

Course Outcomes:
At the end of the course students will be able to

1. **Review** the research paper for survey, methodology and conclusion.
2. **Analyze** research data using relevant statistical techniques.
3. **Interpret** the research findings.
4. **Communicate** research findings through written reports and oral presentations.

1. Reviewing Research papers
a. Write a set of research questions for the case study of your interested topic. Construct the set of research objectives from the set of questions.
b. Utilize academic search tools to extract pertinent journal papers. Engage in comprehensive reading of the paper, focusing on the following aspects: major contribution, datasets utilized, methodology employed, analysis and outcomes, conclusions drawn, and any limitations identified. Post-review, articulate the research question believed to have been addressed by the author. Evaluate whether the paper effectively substantiates its conclusions in addressing the aforementioned question.

2. Literature Review Quality and research proposal
a) Select a journal paper within your engineering discipline and provide a detailed summary of its literature review section, comprising 400 to 500 words. Evaluate the quality and relevance of the papers cited within the review, assessing the commentary on their contributions to the broader field. Additionally, analyze any notable omissions of papers that hold significant importance within the field.
b) Craft a new research proposal based on insights gleaned from a chosen published journal paper. Focus specifically on scrutinizing the discussion and conclusion sections of the paper to identify suggestions for future research endeavors.

3. Data modeling and statistical review
a) Download a set of weather data from the Internet covering the temperature and atmospheric pressure over five days period. Present the data using 2D and 3D plots, and so deduce if the weather conditions are trending either higher or lower over these five days period.

b) Numerical modeling: Find a research paper that employs numerical modeling to validate experimental findings. Compare and contrast the differences observed between the experimental and modeling results. Evaluate whether the authors have provided insights into the accuracy of both the experimental and modeling methodologies. Provide recommendations for enhancing the quality of the modeling techniques presented in the paper.

c) Statistical review: Within your engineering domain, examine a published paper incorporating statistical analysis. Draft a concise report delineating the statistical methods employed. Propose enhancements to the statistical analysis. Recommend additional parameters that could have been captured during data acquisition and elucidate how to analyze the comprehensive dataset to ascertain the influence and statistical significance of these supplementary measurements.

4.Research Paper analysis

1. Note the keywords and type them into one of the web-based academic search engines (e.g. googlescholar.com).
    a. Does the original article appear in the search results?
    b. Compare the citations of this paper with those from the most highly cited paper in the search results.
    c. If this paper was published before your original article, is it cited in your article? Do you think this high-cited paper should have been listed as a reference in your original article? Give reasons for your decision.
2. How many citations does this article have?
3. Have the same authors published further work in this field?
    a. How many citations does this highly cited article have?

5. Research proposal

Generate a novel research proposal based on an existing published journal paper by following these steps:

1. Begin with an introduction that outlines the research problem and objectives.
2. Provide background information and rationale to contextualize the proposed research.
3. Detail the methodology, including research design, data collection methods, and analytical techniques.
4. Write a research plan and budget requirements.
5. Summarize the conclusions drawn from the existing paper and propose areas for future research.
6. Compile a bibliography listing all the references cited in the proposal.

6. Patent claims identification

Examine a journal article within your field that was published around five years ago. Analyze the paper's key findings and identify significant outcomes. Structure these outcomes in a format akin to patent claims, ensuring they represent novel advancements not readily deduced from prior research.

Home

MDS1-008 Programming Proficiency Laboratory

Teaching Scheme: P: 02 Hrs./Week	**Credits**: 01	**Examination Scheme:** CIE/TW: 25 Marks

Prerequisite: Basics of Programming

Course Objectives:
Course intends to prepare the students

1. To study Python/R/Java programming.
2. To study use of Python/R/Java programming in various applications such as File Handling, Web Scraping, Data Preprocessing, Data Visualization, Natural Language Processing, Machine Learning, Graphical User Interface.
3. To explore functions in Python/R/Java for evaluating performance of the designed model.

Course Outcomes:
At the end of the course students will be able to

1. **Write** a program in Python/R/Java language.
2. **Implement** File Handling, Web Scraping using Python/R/Java language.
3. **Write** a program for Data Preprocessing, Data Visualization using Python/R language.
4. **Develop** Graphical User Interface in Python/R/Java language.
5. **Analyze** performance of the designed model by using functions in Python/R/Java.

1.
A. Design a program that retrieves and lists the names of all regional languages along with the number of related articles in the order they are presented on Wikipedia.org
B. Develop a tool to extract and organize images from a specific web page URL, providing options for filtering, sorting, and downloading the images based on various criteria such as resolution, file type, or metadata

2.
A. Write a program to perform various data preprocessing operations such as handling missing values, one hot encoding, label encoding & data normalization on the following dataset. Dataset: https://rb.gy/ppmnxv
B. Write a program to reduce the number of features in the following dataset while preserving most of the important information. Use techniques such Principal Component Analysis (PCA), t-distributed Stochastic Neighbor Embedding (t-SNE), or feature selection methods like Recursive Feature Elimination (RFE) for this purpose. Dataset: https://shorturl.at/aejsM.

3.
A. Write a program to visualize educational data such as student performance, attendance, and learning outcomes to identify areas for improvement, track progress over time, and personalize learning experiences.
B. Design a data visualization dashboard to analyze the monthly sales performance of a retail store over the past year. The dashboard should include visualizations of total sales, average sales per day, and sales trends over time. Additionally, it should identify the best and worst performing months and provide insights into factors affecting sales fluctuations.

4.
A.Write a program to explore different tokenization techniques such as whitespace tokenization, word tokenization, or sentence tokenization.
B.Write a program to implement the Porter Stemmer algorithm for Stemming.

5.
A. Write a program to predict the house price using a regression algorithm in machine learning. Evaluate the performance of the model using various performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Huber Loss & R2 score. Link to dataset: https://rb.gy/ppmnxv
B. Write a program to split the Iris flower dataset into 80% train data and 20% test data. Fit the data into the Naive Bayes model for training and testing. Plot the Confusion Matrix. Evaluate the performance using various performance metrics such as Accuracy, Precision, Recall & F1 Score

6.
A.Write a program to design a BMI (Body Mass Index) calculator application using tkinter library in Python / shiny library in R where users can input their height and weight, and the application calculates and displays their BMI along with a corresponding interpretation (e.g., underweight, normal weight, overweight).
B.Build a simple calculator application using the tkinter library in Python / shiny library in R that performs basic arithmetic operations such as addition, subtraction, multiplication, and division. The application should have a user interface with buttons for numeric input and operations. (Graphical User Interface)

Relevant MOOCs Courses:
1. SWAYAM - Programming in Python by Dr. Rizwan Rehman, Dibrugarh University.
https://onlinecourses.swayam2.ac.in/cec22_cs20/preview
2. NPTEL - The Joy of Computing using Python by Prof. Sudarshan Iyengar, IIT Ropar
https://onlinecourses.nptel.ac.in/noc24_cs57/preview
3. IBM - Data Visualization using python.
https://cognitiveclass.ai/courses/data-visualization-python
4. NPTEL - Introduction to R Software by Prof. Shalabh , IIT Kanpur
https://onlinecourses.nptel.ac.in/noc19_ma33/preview

Text Books:
1. Reema Thareja, "Python Programming Using Problem Solving Approach," Oxford University Press, ISBN 13: 978-0-19-948017-6
2. R. Nageswara Rao, "Core Python Programming," Dreamtech Press; Second edition ISBN-13: 978-9386052308
3. Andrie de Vries, "R For Dummies," 2nd edition (21 July 2015), ISBN-101119055806

Reference Books:
1. Martin C. Brown, "Python: The Complete Reference," McGraw Hill Education, ISBN-13: 978-9387572942
2. Romano Fabrizio, "Learning Python," Packt Publishing Limited, ISBN: 9781783551712, 1783551712
3. Paul Barry, "Head First Python- A Brain Friendly Guide," SPD O'Reilly, 2nd Edition, ISBN:978-93-5213-482-3

Home

Semester II

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

Semester II

MDS2-009 Data Modelling and Visualization		
Teaching Scheme: L: 03 Hrs./Week	**Credits**: 03	**Examination Scheme:** ISE: 20 marks CIE: 30 marks ESE: 50 marks
Prerequisite: Mathematical Foundations for Computing		

Course Objectives:

Course intends to prepare the students

1. To understand Data Analytics Life cycle.
2. To learn different types of Data Models and visualization.
3. To study quantitative and non-quantitative data visualization.
4. To select the appropriate software and chart types to model and visualize data.
5. To prepare visual reports, interactive charts and dashboards.

Course Outcomes:

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

1. **Summarize** data using Preprocessing Techniques and visualization.
2. **Evaluate** different visualizations for numeric and non-numeric data.
3. **Recommend** a method for visualizing high dimensional Data.
4. **Create** a story using data, design principles and tools for communication.

Course Contents

Module I	**Introduction to Data Science, Big Data and Data Modeling** Need and Applications of Data Science and Big Data, **Data Analytics Life Cycle**, Phase 1: Discovery, Phase 2: Data Preparation, Phase 3: Model Planning, Phase 4: Model Building, Phase 5: Communication results, Phase 6: Operationalize. Introduction to Data Modeling, Modeling, Data Hybrid ,Object-Oriented, Prototyping , and Agile Methodologies, Data and Process Modeling, Unified Modeling Language (UML).	**8 Hrs.**
Module II	**Visualization of Numeric and Non Numeric data.** Data Visualization, Types of data visualization, Data Visualization Techniques, Data Transformations and Visualization, Basic graphs, charts, and plots, histogram, pie charts, bar plots, Box plots, Visualization of Non-Numeric Data : Networks, Hierarchies, Symbol And Shaded Maps, Tree Map. Case Study: Tools used in Visualization, Create visualizations using d3.js.	**8 Hrs.**
Module III	**High dimensional data Visualization** Data Analysis, Mapping of high dimensional data into suitable visualization method, Dimensionality Reduction, Principal component analysis, multidimensional clustering, Mosaic Plots, Trellis Displays, Parallel coordinate plots	**8 Hrs.**
Module IV	**Big data Modeling and Visualization** Big data modeling: Introduction to big data modeling, Graph data models, Knowledge graph, modeling of streaming data using real-world datasets.	**8 Hrs.**

	Case Study: Visualizing And Analyzing Data in Tableau. Creating Dashboard, Data Stories with Tableau.	
	Case Studies of industry relevance/recent trends	**5 Hrs.**

Text Books:

1. Andy Oppel, "Data Modeling, A Beginner's Guide," McGraw-Hill, 2010.ISBN-13: 9780071623988
2. Colin Ware, "Information visualization perception for design," MK publication 2004.ISBN-13: 9780123814647
3. DT Editorial Services, "Big Data, Black Book," 2016 Edition. ISBN 13: 9789351199311
4. James Lee, Tao Wei, Suresh Kumar Mukhiya, "Hands-On Big Data Modeling,"PaperBack, 2018.

Reference Books:

1. Knaflic Cole, "Storey Telling with Data," 2015, ISBN-13: 9781788620901
2. Andy kirk, "Handbook for visualizing: a handbook for data driven design," Paperback 2016, ISBN-13: 9781473912144
3. Sleeper Ryan, "Practical Tableau," Kindle Edition, ISBN-13: 9781491977316
4. Ashley Ohmann and Matt FloydCT, "Creating Data Stories with Tableau Public," 2015. ISBN-13: 9781849694773
5. Chun-houh Chen, "HandBook of Data Visualization," Springer 2008.ISBN-13: 9783540330363
6. Jiawei Han, Micheline Kambar, Jian Pei, "Data Mining, Concepts and Techniques," MK ,2012, ISBN-13: 9780123814807

Paper References:

1. Jingyi Zhang, Jiaxing Huang, Sheng Jin and Shijian Lu," Vision-Language Models for Vision Tasks: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, DOI 10.1109/TPAMI.2024.3369699
2. Shreya Ghosh, Abhinav Dhall, et al., "Automatic Gaze Analysis: A Survey of Deep Learning Based Approaches," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 46, No. 1, January 2024
3. Junpeng Wang, Shixia Liu, and Wei Zhang, "Visual Analytics for Machine Learning: A Data Perspective Survey," IEEE Transactions on Visualization and Computer Graphics. DOI 10.1109/TVCG.2024.3357065

Relevant MOOCs

1. Introduction to Data Analytics, Prof. Nandan Sudarsanam, Dr. Balaraman RavindranIIT Madras https://nptel.ac.in/courses/110106064
2. Business analytics and data mining Modeling using R,Prof. Gaurav Dixit, IIT Roorkee. https://onlinecourses.nptel.ac.in/noc20_mg24/preview

Other Resources/Links

1. https://online.hbs.edu/blog/post/data-visualization-tools
2. https://powerbi.microsoft.com/en-us/
3. https://public.tableau.com/en-us/s/about
4. https://public.tableau.com/en-us/s/resources

Home

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

MDS2-010 Generative and Responsible AI		
Teaching Scheme: L: 04 Hrs./Week	**Credits**: 04	**Examination Scheme:** ISE: 20 marks CIE: 30 marks ESE: 50 marks
Prerequisite: Data structures and Algorithms, Discrete Mathematics		
Course Objectives: **Course intends to prepare the students** 1. To learn the concepts of artificial intelligence and methods. 2. To understand the concepts of Generative AI and Responsible AI 3. To build models for Generative AI 4. To design and solve real world problems using AI approaches.		
Course Outcomes: **At the end of the course students will be able to** 1. **Apply** the concepts of intelligent agents for different AI techniques. 2. **Understand** the concepts of Generative AI 3. **Illustrate** generative AI models for different applications 4. **Identify** Responsible AI Features for AI applications		
Course Contents		
Module I	**Problem solving using AI techniques.** Introduction to AI and applications, The Structure of Agents, Informed (Heuristic) Search Strategies, Unification and First-Order Inference, Forward Chaining, Backward Chaining, Resolution, Algorithms for classical planning, Goal stack planning, Heuristics for planning. Case study of Route planning and navigation in GPS systems and mapping applications	**8 Hrs.**
Module II	**Introduction to Generative AI** Generative AI: Introduction, Difference between Generative and discriminative models, History and evolution of Generative AI, Neural Networks for Generative AI, GenAI Project Life Cycle, ML Model Building Process, Data Collection and Preprocessing, Training and Evaluation Overview, Deep Learning Techniques	**8 Hrs.**
Module III	**Generative AI Models** Language Models, Transfer Learning and Pre-trained Models, Autoencoders, Variational Autoencoders, Advanced Generative AI Models, GAN Training Techniques, GAN Evaluation Techniques, Flow based models, Introduction to Generative AI Creativity Tools and Examples.	**8 Hrs.**
Module IV	**Responsible AI** Responsible AI: Introduction, Principles and pillars of Responsible AI, Bias & Fairness: Explainability & Interpretability, Safety, Security, and Privacy. Metrics and Tools for RAI , Adversarial Testing, Explanations for Lime, SHAP, GradCam, Problems with generative models – Ethical Alignment Hallucination, Factual Correctness, Prompt Injection, Data Leakage, Deep	**8 Hrs.**

	Fakes, Copyright Infringement.	
	Case Studies of industry relevance/recent trends	**5 Hrs.**

Text Books

1. Russell S. and Norvig P. "Artificial Intelligence: A Modern Approach," Pearson Education, 4th edition, 2022, ISBN: 978-9356063570.
2. Elaine Rich, Kevin Knight and Nair, "Artificial Intelligence," TMH, 2017 ISBN- 978-0070087705.
3. David Foster, "Generative Deep Learning," 2nd Edition April 2023, O'Reilly Media, Inc.ISBN: 9781098134181
4. Avinash Manure, Shaleen Bengani, Saravanan S, "Introduction to Responsible AI: Implement Ethical AI Using Python" November 2023, ISBN-10 : 1484299817

Reference Books

1. Nilsson Nils J, "Artificial Intelligence: A new Synthesis," Morgan Kaufmann Publishers Inc. San Francisco, CA, ISBN: 978-1-55-860467-4
2. Patrick Henry Winston, "Artificial Intelligence," Addison-Wesley Publishing Company, ISBN: 0- 201-53377-4
3. Dr. Lavika Goel, "Artificial Intelligence: Concepts and Applications," Wiley publication, ISBN: 9788126519934
4. Dr. Nilakshi Jain, "Artificial Intelligence, as per AICTE: Making a System Intelligent,"Wiley publication, ISBN: 9788126579945
5. Deepak Khemani, "A First Course in Artificial Intelligence," McGraw Hill Education, ISBN: 9781259029981
6. David M. Patel , Artificial Intelligence & Generative AI for Beginners: Kindle Edition
7. Numa Dhamani and Maggie Engler, "Introduction to Generative AI" January 2024, ISBN 9781633437197
8. Virginia Dignum," Responsible Artificial Intelligence How to Develop and Use AI in a Responsible Way",Springer https://link.springer.com/book/10.1007/978-3-030-30371-6 ,ISBN 978-3-030-30371-6

Paper References:

1. Edward J Hu, Moksh Jain, Eric Elmoznino, Younesse Kaddar, Guillaume Lajoie, Yoshua Bengio, Nikolay Malkin, "Amortizing Intractable Inference in Large Language Models" International Conference on Learning Representations (ICLR) 2024 https://arxiv.org/pdf/2310.04363
2. Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, Furu Wei,"Multilingual E5 Text Embeddings: A Technical Report" Microsoft Corporation,https://arxiv.org/pdf/2402.05672
3. Christopher Archibald, Delma Nieves-Rivera , "Estimating Agent Skill in Continuous Action Domains," Journal of Artificial Intelligence Research 80 (2024) 27-86,https://www.jair.org/index.php/jair/article/view/15326/27033
4. Qiang Yang, "Toward Responsible AI: An Overview of Federated Learning for User-centered Privacy-preserving Computing," ACM Transactions on Interactive Intelligent Systems Volume 11 Issue 3-4 Article No.: 32pp 1–22
5. Francisco Rodríguez-Gómez, José del Campo-Ávila ∗, Llanos Mora-López, "A novel clustering based method for characterizing , household electricity consumption profiles," Engineering Applications of Artificial Intelligence 129 (2024) 107653
6. Bengesi, Staphord, Hoda El-Sayed, Md Kamruzzaman Sarker, Yao Houkpati, John Irungu, and Timothy Oladunni. "Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers." IEEE Access (2024).
7. Xia, Boming, Qinghua Lu, Liming Zhu, Sung Une Lee, Yue Liu, and Zhenchang Xing. "Towards a Responsible AI Metrics Catalogue: A Collection of Metrics for AI Accountability." In Proceedings of

the IEEE/ACM 3rd International Conference on AI Engineering-Software Engineering for AI, pp. 100-111. 2024.

Relevant MOOC

1. An Introduction to AI Prof. Mausam, IIT Delhi https://nptel.ac.in/courses/106/102/106102220/
2. Artificial Intelligence Prof Anupam Basu, Prof. S. Sarkar, IIT Kharagpur https://nptel.ac.in/courses/106/105/106105077/
3. Artificial Intelligence :Knowledge Representation and Reasoning, Prof Kemani IIT Madras https://nptel.ac.in/courses/106/106/106106140/
4. Responsible & Safe AI Systems Prof. Ponnurangam Kumaraguru, Prof. Balaraman Ravindran, IIIT Hyderabad, IIT Madras https://onlinecourses.nptel.ac.in/noc24_cs132/preview
5. Responsible AI in theGenerative AI Era https://www.coursera.org/learn/responsible-ai-in-generative-ai
6. Generative AI with Large Language Models https://www.coursera.org/learn/generative-ai-with-llms

Other Resources/Links

1. https://en.wikipedia.org/wiki/Artificial_intelligence
2. https://indiaai.in/
3. https://www.nvidia.com/en-us/glossary/generative-ai/
4. https://www.oracle.com/in/artificial-intelligence/generative-ai/what-is-generative-ai/
5. https://en.wikipedia.org/wiki/Generative_artificial_intelligence
6. https://research.ibm.com/blog/what-is-generative-AI
7. https://ai.google/responsibility/responsible-ai-practices/
8. https://learn.microsoft.com/en-us/azure/machine-learning/concept-responsible-ai?view=azureml-api-2
9. https://www.cloudskillsboost.google/course_templates/554

Home

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

MDS2-011 Program Electives II

MDS2-011A Edge Computing and IoT Applications		
Teaching Scheme: L: 03 Hrs./week	**Credits**: 03	**Examination Scheme:** ISE: 20 marks CIE: 30 marks ESE: 50 marks
Prerequisite: Computer Networks, Internet of Things (IoT)		

Course Objectives:
Course intends to prepare the students

1. To understand and compare working principles of cloud, edge and fog computing.
2. To study edge computing architectures and enabling technologies.
3. To study real-world use cases of edge computing in IoT including autonomous vehicles, industrial automation, and smart cities.

Course Outcomes:
At the end of the course students will be able to

1. **Differentiate** between edge, fog and cloud computing in the context of IoT applications.
2. **Design** the architecture of edge-enabled systems for IoT applications.
3. **Design** applications in cloud-edge-IoT environments.
4. **Illustrate** different deployment models, such as fog computing, mobile edge computing.

Course Contents

Module I	**Fundamentals of Edge Computing** Overview of cloud, edge and fog computing, cloud computing architectures, SaaS, PaaS, EaaS, limitations of cloud supported IoT applications, advantages and applications of edge computing, edge computing architectures, challenges and limitations of edge computing, recent trends in edge computing, Cloud-Edge-Fog comparison.	**8 Hrs.**
Module II	**Edge and Fog Computing Architectures** Evolution of IoT edge computing capabilities, edge-IoT integration, mapping of Edge Computing Architectures(ECA) to IoT layer models, ECA and IoT limitations and challenges, introduction to edge devices, overview of fog computing, fog computing architectures, characteristics of fog computing, applications, limitations, and challenges of fog computing, fog computing platforms.	**8 Hrs.**
Module III	**Enabling Technologies for Edge Computing** Introduction to MQTT and Kafka for end-to-end edge pipeline, containerization at the edge, cloudlet, mobile edge computing, machine learning for edge: sensor data in predictive maintenance, image classifier and self-driving cars.	**8 Hrs.**
Module IV	**Advanced Topics in Edge Computing and IoT** Security concerns in edge computing, lightweight edge clouds, Cloud-Fog-Edge enabled analytics, edge and fog intelligence, cognitive IoT applications using edge and fog computing for vision, speech, and text.	**8 Hrs.**
	Case Studies of industry relevance/recent trends	**5 Hrs.**

Text Books

1. Perry Lea, "IoT and Edge Computing for Architects," Second Edition, Packt Publishing, 2020, ISBN: 9781839214806.
2. K. Anitha Kumari, G. Sudha Sadasivam, D. Dharani, M. Niranjanamurthy, "Edge Computing: Fundamentals, Advances and Applications," CRC Press, 2021, ISBN: 9781000483598, 1000483592.
3. David Jensen, "Beginning Azure IoT Edge Computing: Extending the Cloud to the Intelligent Edge," Microsoft Azure, ISBN: 978-1484245354.

Reference Books

1. Rajkumar Buyya (Editor), Satish Narayana Srirama (Editor), "Fog and Edge Computing: Principles and Paradigms," Wiley, 2019, ISBN: 978-1-119-52498-4.
2. Anwesha Mukherjee (editor) Debashis De (editor) Rajkumar Buyya (editor), "Mobile Edge Computing," Springer, 2021, ISBN: 978-3-030-69892-8.

Paper References

1. Bourechak, Amira, Ouarda Zedadra, Mohamed Nadjib Kouahla, Antonio Guerrieri, Hamid Seridi, and Giancarlo Fortino. "At the confluence of artificial intelligence and edge computing in iot-based applications: A review and new perspectives." Sensors 23, no. 3 (2023): 1639.
2. Kong, Xiangjie, Yuhan Wu, Hui Wang, and Feng Xia. "Edge computing for internet of everything: A survey." IEEE Internet of Things Journal 9, no. 23 (2022): 23472-23485.
3. Laroui, Mohammed, Boubakr Nour, Hassine Moungla, Moussa A. Cherif, Hossam Afifi, and Mohsen Guizani. "Edge and fog computing for IoT: A survey on current research activities & future directions." Computer Communications 180 (2021): 210-231.
4. Bermejo, Belen, and Carlos Juiz. "Improving cloud/edge sustainability through artificial intelligence: A systematic review." Journal of Parallel and Distributed Computing 176 (2023): 41-54.

Relevant MOOCs

1. Edge Computing, Prof. Rajiv Misra , IIT Patna
   https://onlinecourses.nptel.ac.in/noc24_cs66/preview

Other Resources/Links

1. Edge computing technologies for Internet of Things: a primer - ScienceDirect
2. Edge-Computing Architectures for Internet of Things Applications: A Survey - PMC (nih.gov)
3. Edge computing technologies for Internet of Things: a primer (sciencedirectassets.com)
4. Shaping the Future of IoT with Edge Intelligence | How Edge Computing (taylorfrancis.com)
5. Fog computing security challenges and future directions- IEEE Xplore Full-Text PDF
6. A Review on Fog Computing: Architecture, Fog with IoT, Algorithms and Research Challenges - ScienceDirect
7. Edge-Computing Architectures for Internet of Things Applications: A Survey - PMC (nih.gov)

Home

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

MDS2-011B Information Retrieval and Web Mining		
Teaching Scheme: L: 03 Hrs./Week	**Credits**: 03	**Examination Scheme:** ISE: 20 marks CIE: 30 marks ESE: 50 marks
Prerequisite: Database Management System, Data Structures and Algorithms, Probability and Statistics		

Course Objectives:
Course intends to prepare the students

1. To explore text documents representation and retrieval methods.
2. To understand query formulation, evaluation and feedback relevance.
3. To study Supervised and Unsupervised learning Method for text analysis and retrieval.
4. To understand behavioral patterns and profiles of web users.

Course Outcomes:
At the end of the course students will be able to

1. **Analyze** text documents representation and retrieval methods.
2. **Analyze** query performance for information retrieval systems.
3. **Evaluate** the performance of classification and clustering methods on text documents.
4. **Assess** behavioral patterns and profiles of the web by studying web user transactions.

Course Contents

Module I	**Introduction** Overview of text retrieval systems: Boolean retrieval, the term vocabulary and postings list, Dictionaries and tolerant retrieval, Index construction and compression, Retrieval models and implementation: Vector Space Models, Vector Space Model, TF-IDF Weight.	**8 Hrs.**
Module II	**Language Models, Query Expansion and feedback evaluation** Probabilistic models: statistical language models, Okapi/BM25, Language models for information retrieval, KL-divergence, Smoothing, Query expansion and feedback: Relevance feedback, pseudo relevance feedback, Query Reformulation, Computing scores in a complete search system, Evaluation in information retrieval.	**8 Hrs.**
Module III	**Text classification and Text clustering** **Text classification & Text clustering**: The text classification problem, Naive Bayes text classification, Vector-space clustering: Partition based clustering algorithm, K-means algorithm, Hierarchical clustering, EM algorithm, Performance analysis. **Ethical Issues in IR:** Privacy, Fairness, Fake news and disinformation, Filter bubble, Viewpoint diversity, fostering extremism, Internet addiction.	**8 Hrs.**
Module IV	**Social Network Analysis and Web usage Mining** **Social Network Analysis**: Centrality, Prestige, Co-Citation and Bibliographic Coupling: Co-Citation, Bibliographic Coupling, Page Ranking: PageRank Algorithm, Strengths and Weaknesses of PageRank, Timed PageRank and Recency Search. **Web Usage Mining**: Key Elements of Web Usage Data Pre-Processing, Data Modeling for Web Usage Mining, Discovery and Analysis of Web Usage Patterns: Session and Visitor Analysis, Cluster Analysis and Visitor Segmentation, Association and Correlation	**8 Hrs.**

	Analysis, Analysis of Sequential and Navigational, Classification and Prediction based on Web User Transactions. web usage mining tool.	
	Case Studies of industry relevance/recent trends	**5 Hrs.**

Text Books

1. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, "Introduction to Information Retrieval," Cambridge University Press. 2008. ISBN: 0521865719. http://nlp.stanford.edu/IR-book/information-retrieval-book.html
2. Ricardo Baeza -Yates and Berthier Ribeiro – Neto, "Modern Information Retrieval: The Concepts and Technology behind Search" 2nd Edition, ACM Press Books 2011.
3. Bing Liu, "Web data Mining: exploring hyperlinks, contents and usage data," Springer, second edition-2011 ISBN 978-3642194597

Reference Books

1. Cheng Xiang Zhai, "Statistical Language Models for Information Retrieval (Synthesis Lectures Series on Human Language Technologies)," Morgan & Claypool Publishers, 2008.
2. W. Bruce Croft, Donald Metzler, and Trevor Strohman. "Search Engines - Information Retrieval in Practice," Cambridge University Press, 2009. ISBN 9781598295900
3. G. Kowalski, M.T. Maybury, "Information storage and Retrieval System theory and implementation," Kluwer Academic Publishers 2nd edition, 2002, ISBN: 0-306-47031-4.

Paper References:

1. T. Shaik, Thanveer and Tao, Xiaohui and Li, Yan and Dann, Christopher and McDonald, Jacquie and Redmond, Petrea and Galligan, Linda, "A Review of the Trends and Challenges in Adopting Natural Language Processing Methods for Education Feedback Analysis," in *IEEE Access*, May 2022, vol. 10, pp. 56720-56739. doi: 10.1109/ACCESS.2022.3177752.
2. Wayne Xin Zhao, Jing Liu, Ruiyang Ren, and Ji-Rong Wen. 2024. "Dense Text Retrieval Based on Pretrained Language Models: A Survey". ACM Trans. Inf. Syst. 42, 4, Article 89 (July 2024), 60 pages. https://doi.org/10.1145/3637870.
3. Edward Kai Fung Dang, Robert Wing Pong Luk, and James Allan. 2021. "A Comparison between Term-Independence Retrieval Models for Ad Hoc Retrieval". ACM Trans. Inf. Syst. 40, 3, Article 62 (July 2022), 37 pages. https://doi.org/10.1145/3483612.

Relevant MOOCs

1. Business Analytics & Text Mining Modeling Using Python by Dr. Gaurav Dixit IIT Roorkee https://onlinecourses.nptel.ac.in/noc19_mg47/preview
2. Natural Language processing :Prof. Pushpak Bhattacharyya IIT Bombay https://nptel.ac.in/courses/106101007

Other Resources/Links

1. C.J. Rijsbergen, "Information Retrieval," (http://www.dcs.gla.ac.uk/Keith/Preface.html)
2. Special interest group on Information retrieval (https://dl.acm.org/sig/sigir)
3. https://www.cs.cmu.edu/~dst/WordEmbeddingDemo/tutorial.html
4. https://www.searchenginejournal.com/semantic-search-how-it-works-who-its-for/438960/
5. https://aws.amazon.com/what-is/retrieval-augmented-generation

Home

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

MDS2-011C Multimodal Computing		
Teaching Scheme: L: 03 Hrs./week	**Credits**: 03	**Examination Scheme:** ISE: 20 marks CIE: 30 marks ESE: 50 marks

Prerequisite:

Course Objectives:

Course intends to prepare the students

1. To introduce fundamentals of multimodal computing.
2. To elaborate different multimodal learning paradigms.
3. To learn and analyze different multimodal learning strategies.

Course Outcomes:

At the end of the course students will be able to

1. **Comprehend** complex concepts better when exposed to multiple modalities.
2. **Apply** knowledge across contexts with deeper understanding.
3. **Express** themselves using various mediums, improving communication skills.

Course Contents

Module I	**Introduction to Multimodal computing** Multimodality introduction, Overview of human communication practices from a multimodal perspective, need of multimodal computing, Need of Multimodal Behaviors & signals, Modes of Multimodal Learning (speech, audio, written and print, illustrations), Multimodal Machine learning and Artificial Intelligence, Applications of MM, Limitations of MM.	**8 Hrs.**
Module II	**Text Processing in Multimodal computing** Overview of text processing techniques-tokenization, stemming, lemmatization, handle noisy and unstructured text.NLP fundamentals, large language models (BERT, GPT), Multimodal fusion methods, Visual text analysis, transfer learning across modalities, Applications.	**8 Hrs.**
Module III	**Speech Processing in Multimodal computing** Overview of speech processing, Digital Signal processing basics, STFT (Short time Fourier transform). **Acoustic Phonetics and Articulatory Phonetics,** Speech Prosody, Speech Prosody Modeling (Fujisaki Model) **Automatic speech recognition- Hidden Markov models, Speech synthesis -text to speech and speech to text, applications.**	**8 Hrs.**
Module IV	**Image and Video processing in Multimodal computing** Overview of multimodal learning and visual data, Image Representation ways-pixel values, histograms, deep features, Extraction of features from images using CNN activations, texture descriptors, applications, limitations. Overview of video processing in MM, Video compression, motion estimation, temporal modeling, video denoising, stabilization, summarization techniques, applications, case studies.	**8 Hrs.**
	Case Studies of industry relevance/recent trends	**5 Hrs.**

Text Books
1. Multimedia Computing, Gerald Friedland, Ramesh Jain, Cambridge University Press, 2014, ISBN 0521764513, 9780521764513 2. Multimedia : Computing, Communications and Applications, Ralf Steinmetz, Klara NaHrs.tedt, Pearson, 2012
Reference Books
1. Akkus, Cem, et al. "Multimodal deep learning." *arXiv preprint arXiv:2301.04856* (2023).
Paper References
1. Bayoudh, K., Knani, R., Hamdaoui, F. et al. A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. Vis Computing 38, 2939–2970 (2022). https://doi.org/10.1007/s00371-021-02166-7 2. P. Xu, X. Zhu and D. Clifton, "Multimodal Learning With Transformers: A Survey" in IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 45, no. 10, pp. 12113-12132, 2023. 3. Ngiam, Jiquan, et al. "Multimodal deep learning." Proceedings of the 28th international conference on machine learning (ICML-11). 2011.
Relevant MOOCs
1. Digital Speech Processing - Course (nptel.ac.in) Prof. Shyamal Kumar Das Mandal ,IIT Kharagpur 2. Multimodal Literacies: Communication and Learning in the Era of Digital Media | Coursera
Other Resources/Links
1. An Introduction to Multimodal Models - Comet 2. A Gentle Introduction to Multiple-Model Machine Learning - MachineLearningMastery.com 3. Chapter 3 Multimodal architectures | Multimodal Deep Learning (slds-lmu.github.io) 4. Multimodal Learning | SpringerLink 5. Multimodal interaction: A review - ScienceDirect

Home

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

MDS2-011D Advanced Compilers		
Teaching Scheme: L: 03 Hrs./week	**Credits**: 03	**Examination Scheme:** ISE: 20 marks CIE: 30 marks ESE: 50 marks
Prerequisite: Knowledge of Automata Theory and Languages.		

Course Objectives:

Course intends to prepare the students

1. To understand the structure of compilers.
2. To learn different ways of intermediate representation and Code generations.
3. To learn the optimization methods and LLVM framework.

Course Outcomes:

At the end of the course students will be able to

1. **Compare** different parsing techniques
2. **Apply** optimization techniques to enhance the performance of generated code.
3. **Evaluate** the effectiveness of different compiler optimization strategies in terms of code size, speed, and memory usage.
4. **Illustrate** the basic workflows of LLVM framework

Course Contents

Module I	**Introduction** Overview of compilation. Phases of Compiler, Lexical Analysis, Syntax analysis, Different types of Parsers. Predictive Parsing, LR parsing, Using Parser Generators, Parser error recovery. Semantic analysis: Type checking,Type checking Declaration and Expressions. Syntax-directed translation.	**8 Hrs.**
Module II	**Intermediate Representations and Code Generation** Introduction, three address code generation, Control Flow Graphs, Intermediate representations trees, Translation into trees, declaration, Building a Simple Compiler Front end. Code generation: Introduction, Target Machine Description, instruction selection, pipelining and instruction scheduling. Peep-Hole Optimization.	**8 Hrs.**
Module III	**Code Optimization** Intro to code optimization. DFA basics. Program representations for optimization. SSA form. SSA-enabled optimizations: conditional constant propagation, value numbering, partial redundancy elimination. SSA construction and destruction algorithms.	**8 Hrs.**
Module IV	**Types of Compilation and LLVM** Just in-time Compilation, Garbage collection in Compiler, Parallelizing Compilers, Domain Specific Language Compilation. Intro to LLVM framework, Hands on Exercise with LLVM, Custom Compiler development using LLVM. Case study: Compiler design for high performance computing and message passing machines and scalable shared memory machines	**8 Hrs.**
	Case Studies of industry relevance/recent trends	**5 Hrs.**

Text Books
1. Dick Grune, Kees van Reeuwijk, Henri E. Bal, Ceriel J.H. Jacobs, Koen Langenhoven, "Modern Compiler Design," 2016 , ISBN:9781493944729, 149394472X 2. Ken Kennedy, Randy Allen," Optimizing Compilers for Modern Architectures: A Dependence-based Approach,"2001 3. A V Aho, R Sethi, J D Ullman, "Compilers: Principles, Techniques, and Tools," Pearson Edition, ISBN 81-7758-590- 4. Michael Wolfe, "High-Performance Compilers for Parallel Computing," Pearson Edition 1995,ISBN-13: 978-0805327304
Reference Books
1. Steven S. Muchnick. Advanced Compiler Design and Implementation. Harcourt Asia Private Ltd, 2000. 2. Y. N. Srikant and P. Shankar (Ed.) The Compiler Design Handbook: Optimizations and Machine Code Generation. CRC Press, 2002.
Paper References
1. David Kaeli ACM Transaction on Architecture and Code Optimization 21, 2 (June 2024). https://dl.acm.org/toc/taco/2024/21/2
Relevant MOOCs
1. Compiler design, Prof. Santanu Chattopadhyay, IIT Kharagpur https://onlinecourses.nptel.ac.in/noc21_cs07/preview

Home

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

MDS2-011E Applied Security		
Teaching Scheme: L: 03 Hrs./week	**Credits**: 03	**Examination Scheme:** ISE: 20Marks CIE:30 Marks ESE:50 Marks
Prerequisite: Basics of computer Network Model, Information Security		

Course Objectives :
Course intends to prepare the students

1. To State controls to protect against program flaws in execution– operating system support & administrative control
2. To learn about security threats posed by e–mails.
3. To learn Tools used in Computer Forensics and Cyber Applications.

Course Outcomes:
At the end of the Course Students will be able to:

1. **Summarize** countermeasures to protect from the threats posed by e–mail systems & understand Internet Security
2. **Write** preventive measures by identifying different cyber attacks
3. **Indicate** usage of forensic linguistic in solving cases

Course Contents

Securing Computer Network		
Module I	Taxonomies of vulnerabilities, attacks & mitigation for: Physical / DLL (Data Link Layer) /Network / Transport / Application layer: viz; Security at Network Layer (IPSec = IP Security), Security at Transport Layer (SSL = Secure Sockets Layer & TLS = Transport Layer Security), Security at Application Layer (PGP = Pretty Good Privacy, S/MIME = Secure/Multipurpose Internet Mail) Electronic Mail Security & e–mail (electronic mail) crime investigation RFC2822. IDS (Intrusion Detection System), IPS (Intrusion Prevention System). Router security: default setting Case studies: Fingerprints & Iris Scans. fail2ban, Snort / design honeypot.	**9 Hrs**
Security in Computing		
Module II	Program Security: Secure Programs– Secure coding ISO standard 5000, Non–malicious Program Errors, Targeted Malicious Code, Control against program threats. Trusted Operating System Design & Access Control in OS, Database Security: security requirements & sensitive data	**8 Hrs**
Administering Security		
Module III	Security Planning, Risk Analysis, Organizational Security Policies, Physical Security, Legal, Privacy & Ethical issues in computer security. Cost of Cybercrimes & IPR (Intellectual Property Rights) issues– Lessons for Organizations. Issues in Security Management. Computer Management in a practical way. Exemplar: Case studies of Ethics.	**8 Hrs**
Information Security in AI		
Module IV	Data Security in AI: Data Security: Protecting data used for AI training and operation from unauthorized access, tampering, and theft. - Encryption: Ensuring data is encrypted both in transit and at rest to prevent unauthorized access. - Access Control: Implementing strict access controls to limit who can view or	**7 Hrs**

	manipulate data. Privacy in AI - Privacy: Safeguarding the privacy of individuals whose data is used by AI systems. - Data Minimization: Using only the minimum amount of data necessary for AI training and operation. - Differential Privacy: Applying techniques to ensure that individual data points cannot be inferred from the output of an AI system.	
	Case Studies on applications of Blockchain Technology	**5 Hrs**

Text Books

1. Charles P.Pfleeger, Shari Lawrence Pfleeger, 'Security in Computing', Pearson Education, 3rd edition,
2. William Stallings, 'Cryptography & Network Security, Principles & Practice', Pearson, 7th edition, ISBN: 978–1–292–15858
3. Nina Godbole, Sunita Belapure, "Cyber Security– Understanding Cyber Crimes, Computer Forensics & Legal Perspectives," Wiley, India, First edition = 2011, ISBN: 978–81–265–2179–1
4. Bill Nelson, Amelia Phillips & Christopher Steuart, "Guide to Computer Forensics & Investigations," 4th edition, Cengage Learning. ISBN: 978–81–315–1946–2.
5. Martin Quest, "Blockchain Dynamics: A Quick Beginner's Guide on Understanding the Foundations of Bitcoin and Other Crypto currencies," Create Space Independent Publishing Platform, 2018, ISBN-1719197091
6. Imran Bashir, "Mastering Blockchain: Distributed Ledger Technology, Decentralization and Smart Contracts Explained," 2nd Edition, Packt Publishing, 2018. ISBN-1788839048

Reference Books

1. Jan L. Harrington, 'Network Security A Practical Approach', Elsevier
2. Bruice Schneier , Applied Cryptography– Protocols, Algorithms and Source code in C, 2nd edition, Wiley Indian Edition, India Pvt Ltd, ISBN 978–81–265–1368–0.
3. Nina Godbole, 'Information Systems Security– Security Management, Metrics, Frameworks & Best Practices', Wiley India, ISBN: 978–81–265–1692–6
4. Antoon W. Rufi, 'Network Security 1 & 2 Companion Guide', Cisco Networking Academy Program, Low Price Edition, Pearson Education. ISBN: 81–317–0892–6
5. Bikramaditya Singhal, Gautam Dhameja, Priyansu Sekhar Panda, "Beginning Blockchain A Beginner's Guide to Building Blockchain Solutions," 2018
6. Chris Dannen, "Introducing Ethereum and Solidity," Foundations of Crypto currency and Blockchain Programming for Beginners.

Paper References

1. Abebe Abeshu and Naveen Chilamkurti; Deep learning: The frontier for distributed attack detection in fog–to–things computing. IEEE Communications Magazine, 56(2):169–175, 2018.
2. B. Li, P. Liu and L. Lin, 'A Cluster-Based Intrusion Detection Framework for Monitoring the Traffic of Cloud Environments', 2016 IEEE 3rd International Conference on Cyber Security and Cloud Computing (CSCloud), Beijing, China, 2016, pp. 42-45, doi: 10.1109/CSCloud.2016.43

3. Geet Shingi, Preeti A. Jain, Harsh Saglani; 'Segmented Federated Learning for Adaptive Intrusion Detection System', July 2021, DoI: https://doi.org/10.48550/arXiv.2107.00881

Relevant MOOC

1. Ethical Hacking Prof. Indranil Sen Gupta, IIT Kharagpur https://onlinecourses.nptel.ac.in/noc19_cs68/preview
2. Cyber Security and Privacy Prof. Saji K Mathew, IIT Madras https://onlinecourses.nptel.ac.in/noc23_cs127/preview
3. Blockchain and its Applications Prof. Sandip Chakraborty, Prof. Shamik Sural , IIT Kharagpur https://onlinecourses.nptel.ac.in/noc22_cs44/preview

Other Useful Resources/Links

1. https://ocw.mit.edu/courses/
2. https://users.cs.fiu.edu/~prabakar/cen5079/Common/textbooks/Mastering_Blockchain_2nd_Edition.pdf
3. https://www.lopp.net/pdf/princeton_bitcoin_book.pdf
4. https://www.blockchainexpert.uk/book/blockchain-book.pdf

Home

MDS2-013 Laboratory Proficiency - II

Teaching Scheme: P: 06 Hrs./Week	**Credits**: 03	**Examination Scheme:** CIE/TW: 25 Marks ESE(P/OR): 50 Marks

Prerequisite Courses: Knowledge of programming languages, Basics of Python/R/Java

All assignments are compulsory. Each student should implement the assignments individually. Laboratory teachers should make sure that the dataset/code/write up is not the same.

Course Objectives:
Course intends to prepare the students

1. To learn different types of Data Models and visualization.
2. To think visually and prepare visual reports, interactive charts and dashboards.
3. To design and solve real world problems using AI approaches.
4. To explore cloud-edge-IoT environment and the architecture of edge-enabled systems for IoT applications.
5. To explore methods for representing and retrieving text documents.
6. To gain insight into the behavioral patterns and profiles of web users.
7. To design software architecture for a selected software system.
8. To learn different phases of compiler and LLVM framework.
9. To learn preventive measures by identifying different cyber-attacks.
10. To understand the need and feasibility of using blockchain technology in real-world applications.

Course Outcomes:
At the end of the course students will be able to

1. **Construct** different visualizations for numeric and non-numeric data and data story.
2. **Apply** AI techniques to solve given problems.
3. **Develop** architectural designs for edge-enabled systems in IoT.
4. **Evaluate** methods for representing and retrieving text documents.
5. **Implement** loop restructuring and code optimisation techniques in the compiler.
6. **Formulate** preventive measures against various cyber-attacks through proactive identification.
7. **Assess** the necessity and viability of integrating blockchain technology into practical applications.

Data Modelling and Visualization

1. Application development using Plots/Graphs, Hierarchies, Symbol and Shaded Maps, Tree Map Python/PowerBI/Tableau.
2. Creating Data stories in Tableau for a Business Application

Generative and Responsible AI

1. Write a program to implement 8 puzzle problems /Blocks world problems using any informed search algorithm.
2. Implement generative models using TensorFlow or PyTorch.

Edge Computing & IoT Applications

1. a) Setup edge-cloud-IoT environment and develop a data analytics application in the above set-up using open IoT/IIoT datasets.

   OR

   b) Setup edge-cloud-IoT environment and develop an application using lightweight machine learning models for edge enabled IoT systems.
2. Develop an edge-enabled application for real-time monitoring, video analytics,or sensor data processing.

Information Retrieval & Web mining

1. Consider 10 documents. Construct a count of each document term and return the top 10 frequently occurring terms in the document collection using the TF-IDF score. Find similarity between documents using different similarity measures. Construct a graph for zips and heaps law. Evaluate the performance of similarity measures using precision, recall and F measure.
2. Develop a program to construct a social network graph and calculate the PageRank scores for each node in a given graph and display the score of each node for each iteration of the PageRank algorithm computation.

Multi Modal Computing

1. a) Develop a customer service chatbot using speech recognition and text analysis to provide a more natural and efficient user experience. OR

   b) Develop a robust speech emotion recognition system using a short time Fourier transform to accurately classify emotions from spoken audio in a noisy environment.
2. Design and develop a system that can recognise traffic signs such as speed limit, signals and directional indicators using Convolutional Neural network.

Advanced Compilers

1. Write a program to demonstrate loop restructuring in C/C++.
2. Write a program to implement code optimization.

Applied Security

1. a) Develop data visualization of last few years breach record

   Or

   b) Write a program to analyze e–mail header
2. Develop a decentralized app for real-world application using blockchain

MDS2-014 Skill Enhancement Lab		
Teaching Scheme: P: 02 Hrs./week	**Credits:** 01	**Examination Scheme:** CIE/TW: 25 Marks

Prerequisite:

Course Objectives:

Course intends to prepare the students

1. To solve and devise solutions to a range of elementary real-world problems.
2. To explore and apply key concepts in logical thinking to business problems.
3. To enable students to critically analyze material (information) to evaluate evidence, construct reasoned arguments, and communicate inferences and conclusions.

Course Outcomes:

At the end of the course students will be able to

1. **Develop** problem-solving skills and critical thinking abilities in the context of recruitment aptitude tests.
2. Illustrate self-efficacy through verbal and nonverbal communication behaviors.

Guidelines for conduction and evaluation of the laboratory sessions:

Students will be encouraged to study and prepare for the contents relevant to the modules- Quantitative Aptitude, Logical Reasoning and Technical communication.

Evaluation can be in terms of practice tests/ mock tests/ MCQ based tests, Group discussion, Mock interview, and presentation.

Course Contents

Module I	**Quantitative Aptitude** Percentage; Profit/ Loss; Simplification and Approximation: Decimal fractions, Square Roots & Cube Roots; Average; Ratio and Proportion: Partnerships; Age; Mixture and alligation; Number Series; Arithmetic: Time and work, Time, Distance and Speed, Boats and streams; Data Interpretation: Pie chart, Bar chart, Table chart.	**6 Hrs.**
Module II	**Logical Reasoning** Distance and direction; Blood relation: Linear inequalities; Ranking arrangement; Coding decoding: Chinese/ Symbolic, Number, Symbolic, Miscellaneous, Letter; Syllogism; Seating arrangement; Puzzle; Input- Output; Alphabetic series; Statements and assumptions; Statements and conclusions.	**6 Hrs.**
	Technical Communication	
Module III	Vocabulary building; Oral Communication and Speaking Techniques; Group Discussions– dynamics of group discussion, Lateral thinking, Brainstorming and Negotiation skills; Meetings– making meeting effective, chairing a meeting, decision making, seeking opinions, interrupting, and handling interruptions, clarifications, closure Agenda, Minute writing; Presentation skills; Interview skills– formal & informal interviews, concept and process, pre-interview planning, opening strategies, answering strategies, interview through tele and video-conferencing.	**6 Hrs.**

Text Books
1. R. S. Aggarwal, Quantitative Aptitude (Fully solved), Reprint 2016, S. Chand Publishing.
Reference Books
1. Simon Sweeny, "Communicating in Business," Second Edition, Cambridge University Press 2. M. Ashraf Rizvi, "Effective Technical Communication," Tata McGraw-Hill Publishing Company Ltd. 2005. 3. Andrea J. Rutherford, "Basic Communication Skills for Technology," 2nd Edition, Pearson Education, 2007. 4. Meenakshi Raman & Sangeeta Sharma, "Technical Communication," Oxford University Press, 2011.
Relevant MOOC
1. English Language for Competitive Exams, IIT Madras https://nptel.ac.in/courses/109106116 2. Employment Communication A Lab based course, IIT Kharagpur https://nptel.ac.in/courses/109105144
Other Resources/Links

Home

MDS2-015 Seminar I		
Teaching Scheme: P: 04 Hrs./week	**Credit**: 02	**Examination Scheme:** CIE/TW: 50 Marks ESE(OR): 50 Marks

Course Objectives:
Course intends to prepare the students

1. To deepen students' understanding of advanced topics, emerging trends, and recent developments in computer engineering.
2. To enhance students' research skills, including literature review, experimental design, data collection and analysis, and interpretation of results, in the context of computer engineering research.
3. To improve students' ability to communicate technical concepts effectively through written reports, oral presentations, and technical documentation, targeting both technical and non-technical audiences.

Course Outcomes:
At the end of the course students will be able to

1. **Conduct** thorough literature surveys confined to the domain of choice.
2. **Develop** presentation skills to deliver the technical contents.
3. **Prepare** the report of the technical research domain.

The student shall have to deliver the seminar I in semester II on a topic approved by guide and authorities. It is recommended to allot a guide to the student since the commencement of semester I. The guide allotment preferably needs to be carried out in synchronization with mutual domains of interest. It is recommended that seminar shall be on the topic relevant to latest trends in the field of concerned branch, preferably on the topic of specialization based on the electives selected or domain of interest.
It is appreciated and strongly recommended that the student will select the domain of his/her dissertation and identify the literature confined to the domain. Thorough literature study based on the broad identified topic has to be carried out. This practice will eventually lead to convergence of the efforts for the dissertation.
The relevant literature then be explored as state-of-the-art, exotic, recent technological advancement, future trend, application and research & innovation. Multidisciplinary topics are encouraged. The student shall submit the duly approved and certified seminar report in standard format, for satisfactory completion of the work by the concerned Guide and head of the department/institute. The student will be assessed based on his/her presentation and preparations by the panel of examiners out of them one has to be an external examiner.
The students are expected to validate their study undertaken by publishing it at standard platforms.
The student has to exhibit the continuous progress through regular reporting and presentations and proper documentation of the frequency of the activities in the sole discretion of the PG coordination.
The continuous assessment of the progress needs to be documented unambiguously. For standardization and documentation, follow the guidelines circulated / as in the seminar logbook approved by the Board of Studies.

Home

Semester III

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

Semester – III

MDS3-018 Indian Knowledge System and Human Values		
Teaching Scheme: L: 01 Hrs./week P: 02 Hrs./week	**Credits**: 02	**Examination Scheme:** CIE: 30 marks CIE/TW: 20 marks

Prerequisite:

Course Objectives:
Course intends to prepare the students
Study/Explore/Understand the significance of IKS for a sustainable environment, love for nature, holistic ways of living, and health conscious that ultimately leads to quality life for everyone in the plane.

1. To preserve and showcase the depth and breadth of Indian knowledge for a sustainable society and future generations.
2. To explore research in the area of Mindfulness for Sustainable Health.
3. To help the students to ensure sustained happiness and prosperity which are the core aspirations of all human beings.
4. To facilitate the development of a Holistic perspective among students towards life and profession as well as towards happiness and prosperity.

Course Outcomes:
At the end of the course students will be able to

1. **Differentiate** between harmonious and disharmonious living
2. **Contribute** in developing solutions to social problems leading to peaceful life.
3. **Demonstrate** Holistic perspective towards life.

Course Contents

Module I	**Indian Knowledge System – An Introduction:**	**02 Hrs.**
	What is IKS? Why do we need IKS? Organization of IKS Historicity of IKS Some salient aspects of IKS.	
Module II	**A. Students can select any one from A/B/C**	**05 Hrs.**
	Introduction to Linguistics, Phonetics, Word generation, Computational aspects, Mnemonics Recursive operations, Rule-based operations, Sentence formation, Verbs and prefixes, Role of Sanskrit in natural language processing.	
	B. Health Wellness and Psychology	

	Introduction to health, Āyurveda: approach to health, Sapta-dhātavaḥ : seven-tissues, Role of agni in health, Tri-doṣas , Āyurveda: definition of health, Psychological aspects of health ,Disease management elements , Dinacaryā: daily regimen for health & wellness, Importance of sleep, Food intake methods and drugs , Approach to lead a healthy life, Indian approach to psychology, The tri guṇa system & holistic picture of the individual, The Nature of Consciousness, Consciousness studies and issues	
	C. essentials for development	
	Indian scheme of knowledge, The knowledge triangle , Prameya – A vaiśeṣikan approach to physical reality , Dravyas – the constituents of the physical reality , Attributes – the properties of substances and Action – the driver of conjunction and disjunction , Sāmānya, viśēṣa, samavāya , Pramāṇa – the means of valid knowledge , Saṃśaya – ambiguities in existing knowledge, Framework for establishing valid knowledge , Deductive or inductive logic framework , Potential fallacies in the reasoning process , Siddhanta: established tenets in a field of study. Case Study : **Societal Problem Solving :** Smart City**,** Town Planning, Health Care Systems, etc.	
Module III	**Value Education and Harmony in Human being**	**06 Hrs .**
	Understanding Value Education, Self-exploration as the Process for Value Education, Continuous Happiness and Prosperity – the Basic Human Aspirations, Right Understanding, Relationship and Physical Facility, Happiness and Prosperity – Current Scenario, Method to fulfill the Basic Human Aspirations Understanding Human being as the Co-existence of the Self and the Body, distinguishing between the Needs of the Self and the Body, The Body as an Instrument of the Self, Understanding Harmony in the Self, Harmony of the Self with the Body, Program to ensure self-regulation and Health.	
Module IV	**Harmony in the Family, Society and Nature**	
	The Basic Unit of Human Interaction, Values in Human–to–Human Relationship, Nine universal values in relationships viz. Trust, Respect, Affection, Care, Guidance, Reverence, Glory, Gratitude, Love. Understanding Harmony in Society, Vision for the Universal Human Order, Human Order Five Dimension.. Understanding Harmony in Nature, self–regulation & mutual fulfillment among the Four orders of Nature, Realizing Existence as coexistence at all levels holistic perception of harmony in existence.	**05 Hrs.**
	Case Studies of industry relevance/recent trends	**5 Hrs.**

Text Books
1. Kapur K and Singh A. K (Eds) 2005). Indian Knowledge Systems, Vol. 1. Indian Institute of Advanced Study, Shimla. 2. Introduction to IKS: Concepts and Applications by Prof. B Mahadevan, IIM Bengaluru 3. AK Pathak, Science and Technology in India, Anshika prakashan pratapgarh, 2016 4. R P Kulkarni, Glimpse of Indian Engineering and Technology (Ancient & Medieval period, Munshiram Manoharlal Publishers Pvt. Ltd. 2018. 5. Human Values and Professional Ethics by R R Gaur, R Sangal, G P Bagaria, Excel Books, New Delhi, 2010 6. A Foundation Course in Human Values and Professional Ethics, R R Gaur, R Asthana, G P Bagaria, 2nd Revised Edition, Excel Books, New Delhi, 2019. ISBN 978–93–87034–47–1
Reference Books
1. Jeevan Vidya: Ek Parichaya, A Nagaraj, Jeevan Vidya Prakashan, Amarkantak, 1999. 2. Human Values, A.N. Tripathi, New Age Intl. Publishers, New Delhi, 2004. 3. The Story of My Experiments with Truth - by Mohandas Karamchand Gandhi 4. On Education – J Krishnamurthy 5. Rediscovering India – by Dharampal Hind Swaraj or Indian Home Rule – by Mohandas K. Gandhi.
MOOC Courses:
1. Indian Knowledge System(IKS): Concepts and Applications in Engineering By Prof. B. Mahadevan, Dr. Vinayak Rajat Bhat, Dr. R Venkata Raghavan | Indian Institute of Management Bangalore (IIMB), Chanakya University, Bangalore Link of the Course: https://onlinecourses.swayam2.ac.in/imb23_mg53/preview 2. Indian Knowledge System(IKS): Humanities and Social Sciences By Prof. B. Mahadevan, Dr. Vinayak Rajat Bhat, Dr. R Venkata Raghavan | Indian Institute of Management Bangalore (IIMB), Chanakya University, Bangalore. https://onlinecourses.swayam2.ac.in/imb23_mg55/preview
ACTIVITY BASED LEARNING (SUGGESTED ACTIVITIES IN CLASS)
1. Presentation Participation individually and in teams. 2. Extempore, Impromptu small talks 3. Quizzes/Assignment /Seminars/Group discussions.
Paper References
1. Prof. Archana Saxena, Tanu Sharma; 'Universal Human Values for Scientific & Technical Development', IJCRT international open access peer reviewed refereed journal, Vol: 20, Issue 4th April 2022, ISSN: 2320-2882 2. Dr. Mahesh T. Kolte, Gauri Shriram Narshihna; 'Universal Human Values for Consciousness Development as Value Education in Technical Education', JETIR August 2018, Vol: 5, Issue 8, ISSN-2349-5162

Home

MDS3-019 Internship/Field Study		
Teaching Scheme: P: 04 Hrs./week	**Credit**: 02	**Examination Scheme:** CIE/TW: 50 Marks ESE(OR): 50 Marks

Course Objectives:

Course intends to prepare the students

1. To identify the domain of interest.
2. To learn to communicate in a scientific language through collaboration with a guide.
3. To categorize the research material confined to the domain of choice.
4. To work in professional environment

Course Outcomes:

At the end of the course students will be able to

1. **Conduc**t literature survey confined to the domain of choice
2. **Develo**p presentation skills to deliver the technical contents
3. **Create** report of the technical research domain
4. **Analyze** the findings and work of various authors confined to the chosen domain

Conduction guidelines

Industry or research internship should include partial/complete project implementation. The preferences/choices of the domain should be taken from the students. The guide needs to be allocated based on the preference/choices. Students should be allocated to the research guide in the first semester itself and the same guide should be continued for the**:** Industry Internship-I/ In house Research Project –I. In case of Industry Internship-I, the assigned guide from college has to monitor and evaluate the progress of the student. The student has to exhibit continuous progress through regular reporting and presentations and proper documentation. The continuous assessment of the progress needs to be documented unambiguously.

Home

MDS3-020 Seminar - II

Teaching Scheme: Practical: 04 Hrs./week	**Credit**: 02	**Examination Scheme:** CIE/TW: 50 Marks ESE(P/OR): 50 Marks

Course Objectives:

Course intends to prepare the students

1. To develop the ability to critically analyze complex problems in computer engineering and devise innovative solutions by applying theoretical knowledge and practical skills.
2. To enhance research skills, including literature review, experimental design, data collection and analysis, and interpretation of results, in the context of computer engineering research.
3. To improve ability to communicate technical concepts effectively through written reports, oral presentations, and technical documentation, targeting both technical and non-technical audiences.

Course Outcomes:

At the end of the course students will be able to

1. **Use** multiple thinking strategies to examine multidisciplinary domains.
2. **Identify** research findings of literature survey.
3. **Demonstrate** the findings and work of various authors confined to the chosen domain
4. **Furnish** the report of the technical research domain

The student shall have to deliver the seminar II in semester III on a topic approved by guide and authorities. It is appreciated if a student has already selected the domain of his/her dissertation work and identified the literature confined to the domain and thorough literature study based on the identified topic has been carried out. This practice will eventually lead to convergence of the efforts for the dissertation work. The meticulous analyses of the literature can be part of the seminar.

The relevant literature then be explored as state-of-the-art, exotic, recent technological advancements, future trends, applications and research & innovations. The student shall submit the duly approved and certified seminar report in standard format, for satisfactory completion of the work by the concerned Guide and head of the department/institute.

The student will be assessed based on his/her presentation and preparations by the panel of examiners out of them one has to be an external examiner. The students are expected to validate their study undertaken by publishing it at standard platforms. The student has to exhibit the continuous progress through regular reporting and presentations and proper documentation of the frequency of the activities in the sole discretion of the PG coordination.

The continuous assessment of the progress needs to be documented unambiguously. For standardization and documentation, follow the guidelines circulated / as in the seminar logbook approved by the Board of Studies.

Home

MDS3-021 Dissertation Stage -I		
Teaching Scheme: P: 08 Hrs./week	**Credit**: 04	**Examination Scheme:** CIE/TW: 50 Marks ESE(P/OR): 50 Marks

Course Objectives:
Course intends to prepare the students

1. To identify the domain of research.
2. To formulate research problems with the help of the guide/mentor elaborating the research.
3. To acquire information and identify scope for the dissertation work.

Course Outcomes:
At the end of the course students will be able to

1. **Conduct** thorough literature surveys confined to the domain of choice.
2. **Identify** scope of the dissertation work by **analyzing** the findings, work of various authors confined to the chosen domain.
3. **Design** the system and **prepare** the technical report of the dissertation work.
4. **Develop** presentation skills to deliver the technical contents.

Guidelines

Dissertation Stage – I is an integral part of the Dissertation work. In this, the student shall complete the partial work of the Dissertation which will consist of problem statement, literature review, design, scheme of implementation (Mathematical Model/SRS/UML Diagrams /ERD/block diagram/ PERT chart) and Layout & Design of the Set-up.

The student is expected to complete the dissertation at least up to the design phase. As a part of the progress report of Dissertation work Stage-I, the candidate shall deliver a presentation on the advancement in Technology pertaining to the selected dissertation topic. The student shall submit approved and certified Dissertation Stage-I report in standard format for satisfactory completion of the work duly signed by the concerned guide and head of the Department/Institute.

The dissertation stage - I work will be assessed by a panel of examiners of which one is necessarily an external examiner. The assessment will be broadly based on literature study, work undergone, content delivery, presentation skills, documentation and report. The students are expected to validate their study undertaken by publishing it at standard platforms. The investigations and findings need to be validated appropriately at standard platforms – conference and/or peer reviewed journals.

The student has to exhibit the continuous progress through regular reporting and presentations and proper documentation of the frequency of the activities in the sole discretion of the PG coordination. The continuous assessment of the progress needs to be documented unambiguously. For standardization and documentation, it is recommended to follow the formats and guidelines in the dissertation workbook approved by the department.

Home

Semester IV

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25

Semester – IV

MDS4-022 Seminar III		
Teaching Scheme: P: 08 Hrs./week	**Credit**: 04	**Examination Scheme:** TW: 50 Marks OR: 50 marks

Course Objectives:

Course intends to prepare the students

1. To identify the domain of interest.
2. To learn to communicate in a scientific language through collaboration with a guide.
3. To categorize the research material confined to the domain of choice
4. To work in professional environment

Course Outcomes:

At the end of the course students will be able to

1. **Conduct** thorough literature surveys confined to the domain of choice.
2. **Develop** presentation skills to deliver the technical contents
3. **Furnish** the report of the technical research domain.
4. **Analyze** the findings and work of various authors confined to the chosen domain

Conduction guidelines

Industry or research internship should include partial/complete project implementation. The preferences/choices of the domain should be taken from the students. The guide needs to be allocated based on the preference/choices. Students should be allocated to the research guide in the first semester itself and the same guide should be continued for the**:** Industry Internship-I/ In house Research Project –I. In case of Industry Internship-I, the assigned guide from college has to monitor and evaluate the progress of the student. The student has to exhibit continuous progress through regular reporting and presentations and proper documentation. The continuous assessment of the progress needs to be documented unambiguously.

Home

MDS4-023 Dissertation Stage -II

Teaching Scheme: Practical:32 Hrs./week	**Credit**: 16	**Examination Scheme:** TW : 100 Marks P/OR :50 marks

Course Objectives:
Course intends to prepare the students

1. To follow SDLC meticulously and meet the objectives of proposed work.
2. To test rigorously before deployment of the system.
3. To validate the work undertaken.
4. To consolidate the work as a furnished report.

Course Outcomes:
At the end of the course students will be able to

1. **Demonstrate** a depth knowledge of the domain of choice.
2. **Analyze** findings, evaluate and present the results and their interpretation.
3. **Prepare** an independent dissertation report, resulting in publication.
4. **Demonstrate** an ability to present and defend dissertation work to a panel of experts.

Guidelines

In Dissertation Stage–II, the student shall consolidate and complete the remaining part of the dissertation which will consist of selection of technology, installations, implementations, testing, results, measuring performance, discussions using data tables as per parameter considered for the improvement with existing/known algorithms/systems, comparative analysis, validation of results and conclusions. The student shall prepare certified final report of Dissertation in standard format for satisfactory completion of the work duly signed by the concerned guide and head of the Department/Institute.

The dissertation stage - II work will be assessed by a panel of examiners of which one is necessarily an external examiner. The students are expected to validate their study undertaken by publishing it at standard platforms. The investigations and findings need to be validated appropriately at standard platforms – conference and/or peer reviewed journals.

The student has to exhibit the continuous progress through regular reporting, presentations, and proper documentation of the frequency of the activities in the sole discretion of the PG coordination. The continuous assessment of the progress needs to be documented unambiguously. It is recommended to continue with guidelines and formats as mentioned in the Dissertation Workbook approved by the department.

Home

SCTR'S PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE-43
M. Tech. (Data Science) Curriculum Structure
With effect from the A.Y. 2024-25