Curriculum for

M. Tech in Data Science (MDS)

Overview

The Department of Data Science at IIT Palakkad presents the curriculum of a transdisciplinary master program M. Tech degree in Data Science (MDS).

The program is going to be jointly contributed by faculty members from Departments of DS, CSE, EE, MA, CE, ME, HSE, PHY. The curriculum contains core courses and electives along with capstone projects. Along with training on theoretical foundation the curriculum also supports considerable training on hands-on and practical skills through projects, lab courses. Electives will help students to pick appropriate tracks for their specialization. Broad distribution of credits are as follows.

Serial Number	**Category** (Number of items)	**Credits**
1	Core Theory (7)	28
3	Electives (*)	12
4	Projects	20
5	CWC	0
	Total	60

*Can be divided in number of courses with 1 to 4 credits

Semester	**Credits**
First	20
Second	17
Third	13
Fourth	10
Total	60

Curriculum

Semester 1

Sl. No	Course Number	Course Name	Credit	Type
1	MA2040	Probability and Statistics	3-1-0-4	Core
2	EE5007	Linear Algebra	3-0-0-3	Core
3	DS3010	Machine Learning	3-0-3-5	Core
4	DS5003	Data Engineering	3-0-3-5	Core
5	DS5005	Optimization	3-0-0-3	Core
		Total	20	

Semester 2

Sl. No	Course Number	Course Name	Credit	Type
1	DS5102	Big Data Lab	1-0-3-3	Core
2	DS3040	Deep Learning	3-0-3-5	Core
3		Program Major Electives	3	Elective
4		Program Major Electives	3	Elective
5		Program Major Electives	3	Elective
		Total	17	

Summer

Sl. No	Course Number	Course Name	Credit	Type
2	DS5190	Project (phase 0)/ Internship	0	Core
		Total	0	

Semester 3

Sl. No	Course Number	Course Name	Credit	Type
1	GN5001	Communication Skill and Technical Writing	0	CWC
2	DS5110	** Project (phase 1)	10	Core
3		Program Major Electives	3	Elective
		Total	13	

Semester 4

Sl. No	Course Number	Course	Credit	Type
1	GN6001	Research Methodologies and Professional Ethics	0	CWC
2	DS5120	Project (phase 2)	10	Core
		Total	10	

**Students will be allowed to go for industrial projects in S3 and/or S4. They need to follow institute rules for outside projects and have to make sure that all credit requirements are complete, e.g. they can credit extra electives in S2 and GN5001, GN6001 in S1 and S2 respectively.

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Code and Title	EE5007 Linear Algebra for Engineers				
Programme	M.Tech	Year of study		Semester	
Course credit	3-0-0-3				
Course category	PMT				
Prerequisite, if any	No				
Consent of teacher, if required	Required				
Date of proposal		Date of Senate Approval			
Proposing faculty	Lakshmi Narasimhan T				

Course Content

S/N	Topic	Lecture (hours)
1	Vectors and vector spaces : Notion of vector space, subspaces, dimensionality, span, inner product, norms, orthogonality, and geometric examples.	6
2	Linear transformations : Linear transformation of vectors, projection, geometric interpretations, and representation by matrices.	6
3	Bases and matrices : Basis, rank-nullity, orthogonal bases, determinant, projection, symmetric matrices, definiteness, normal matrices, nilpotent matrices and matrix norms.	8
4	Eigen values and eigen vectors : Diagonalization of matrices, eigen values, geometric interpretation, and eigen value problems.	7
5	Singular value decomposition: Existence of singular values, pseudo inverse, condition number and geometric interpretation.	7
6	Matrix decompositions : QR, LDU, Cholesky, Schur decomposition, and Jordan normal form	8
7	Engineering applications and examples	0(include in topics)
	TOTAL	42

Learning Outcomes:
Upon successful completion, the students shall be able to

- Understand basic mathematical tools required for engineering.
- Apply mathematical concepts to research problems in engineering.
- Model engineering systems mathematically.
- Analyze a mathematical model and deduce physical interpretations.

Learning Objective:

- To introduce the fundamentals of mathematical concepts required for engineers.
- To study and acquire the knowledge of various mathematical tools which are used in engineering applications.
- To mathematically model and analyze practical engineering systems using mathematical tools and concepts.
- To solve linear systems of the form Ax = b

Teaching Methodology **:** Lecture based

Assessment Methods **:** Exam based

Text Books

1. G. Strang, “Introduction to linear algebra”, Fifth edition, Wellesley-Cambridge Press, ISBN-13: 978-0980232776.
2. S. Axler, “Linear Algebra Done Right”, Ninth edition, Pearson Education, ISBN-13: 978-0387982588.

References

This course has previously been approved as EE5004 - Graduate Engineering Mathematics.

Changes from previous courses are

1. Removed: "Linear Programming" and "Statistics" components.
2. Added topics: symmetric matrices, definitiness, normal matrices, nilpotent matrices, matrix norms, and condition number.
3. Added components: "Matrix Decompositions" - LDU, Cholesky, Schur and Jordan canonical form

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Code and Title	DS5101 ~~CS5101~~: Machine Learning Lab				
Programme	M.Tech/MS/PhD	Year of study		Semester	
Course credit	0-0-3-2				
Course category	PMT				
Prerequisite, if any	co-requisite CS5512				
Consent of teacher, if required	required				
Date of proposal		Date of Senate Approval			
Proposing faculty	Sahely Bhadra				

Course Content

S/N	Topic	Lab (hours)
1	Introduction to NumPy	3
2	Regression: linear regression, ridge regression using scipy	6
3	Introduction to Matplotlib	3
4	Gradient descent method for optimization	3
5	Various classification methods using scikitlearn	6
6	Principal component analysis, Canonical correlation analysis	3
7	Ensemble methods: boosting, bagging, random forests.	3
8	Clustering using scikitlearn	6
9	Sequential Learning : hidden Markov model	3
10	Feed forward NN : Tensorflow	6
	TOTAL	42

Learning Outcomes:

- Given a task, derive a learning model by defining appropriate loss function, regulariser, optimization problem and stating the best possible solution.
- Analyse and compare models and algorithms with respect to their complexity, performance and applicability
- Develop models/algorithms with small modifications of existing standard techniques for a modification of known task

Learning Objectives:

- To introduce classical and foundational concepts, results, methodologies and applications in machine learning
- To develop abilities for developing a solution for a given problem starting from problem and data to presenting results

Teaching Methodology **:** Lab based

Assessment Methods **:** Exam based

Text Books

1. Richard Duda, Peter Hart, David Stork, Pattern Classification, 2nd Ed, John Wiley & Sons, 2001. ISBN 9788126511167
2. Christopher Bishop. Pattern Recognition and Machine Learning. ISBN 0387310738.
3. Trevor Hastie, Robert Tibshirani, Jerome Friedman. Elements of Statistical Learning. ISBN 0387952845.

References

1. Tom Mitchell. Machine Learning. McGraw-Hill. ISBN 0070428077.
2. Shai Shalev-Shwartz, and Shai Ben-David, Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press, 2014. ISBN 978-1-107-05713-5.

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Code and Title	DS5003 ~~CS5015~~: Data Engineering				
Programme	M.Tech/MS/PhD	Year of study		Semester	
Course credit	3-0-0-3				
Course category	PMT				
Prerequisite, if any	Familiarity with Algorithms, Probability, Linear Algebra, Programming				
Consent of teacher, if required	required				
Date of proposal		Date of Senate Approval			
Proposing faculty	Mrinal Kanti Das				

Course Content

S/N	Topic	Lecture (hours)
1	**Data Collection:** Various sources and types of data: text, video, audio, biology etc	3
2	**Data Preprocessing:** Cleaning data, missing data imputation, noise elimination, feature selection and dimensionality reduction, normalization	6
3	**Data Storage:** Database, Schema, ER diagram, SQL, functions, stored procedures, indexing B+tree, MongoDB, Client-Server Architecture [3	9
4	**Information Retrieval:** index construction, scoring models, complete search engine mechanism, evaluation methods.	6
5	**Data Processing:** *Data structures.* Stack, Queue, Linked List, Associated memory, Graphs. *Algorithms***.** Searching, Sorting, Graph traversal, Complexity	9
6	**Data Analysis:** regression, principal component analysis, canonical correlation analysis, analysis of variance	6
7	**Data Visualization:** table, graph, histogram, pie-chart, area-plot, box-plot, scatter-plot, bubble-plot, waffle charts, word clouds.	3
	TOTAL	42

Learning Outcomes: To be able to state and analyse

- Preprocessing techniques for various datasets,
- Standard database systems concepts like tables, relations, query
- Information retrieval techniques such as indexing, scoring, ranking, evaluation
- Data processing algorithms and data structures
- Visualization techniques

Learning Objectives:

- To be able to learn about the entire pipeline of a typical system involving data, collection, preprocessing, storage, retrieval, processing, analysis, and visualization.

Teaching Methodology **:** Lecture based

Assessment Methods **:** Exam based

Text Books

1. Introduction to Algorithms. Cormen, Leiserson, Rivest, Stein. MIT Press 3ed. ISBN-13: 978-0262533058
2. Database System Concepts. Silberschatz, Korth, Sudarshan. McGraw Hill Education; Sixth edition.ISBN-13: 978-9332901384
3. Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools. Cielen, Meysman,Ali. Dreamtech Press. ISBN-13: 978-9351199373

References

1. Data Engineering: A Novel Approach to Data Design. Brian Shive. Technics Publications. ISBN-13: 978-1935504603
2. Python Data Science Handbook: Essential Tools for Working with Data. Joel Grus. O'Reilly. ISBN-13**:** 978-9352134915

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Code and Title	DS5103 ~~CS5103~~: Data Engineering Lab				
Programme	M.Tech/MS/PhD	Year of study		Semester	
Course credit	0-0-3-2				
Course category	PMT				
Prerequisite, if any	Co-requisite of Data Engineering				
Consent of teacher, if required	required				
Date of proposal		Date of Senate Approval			
Proposing faculty	Mrinal Kanti Das				

Course Content

S/N	Topic	Lab (hours)
1	**Data Collection:** Various sources and types of data: text, video, audio, biology etc	3
2	**Data Preprocessing:** Cleaning data, missing data imputation, noise elimination, feature selection and dimensionality reduction, normalization	3
3	**Data Storage:** Database, Schema, ER diagram, SQL, functions, stored procedures, indexing B+tree, MongoDB, Client-Server Architecture [3	9
4	**Information Retrieval:** index construction, scoring models, complete search engine mechanism, evaluation methods.	6
5	**Data Processing:** *Data structures.* Stack, Queue, Linked List, Associated memory, Graphs. *Algorithms***.** Searching, Sorting, Graph traversal, Complexity	12
6	**Data Analysis:** regression, principal component analysis, canonical correlation analysis, analysis of variance	6
7	**Data Visualization:** table, graph, histogram, pie-chart, area-plot, box-plot, scatter-plot, bubble-plot, waffle charts, word clouds.	3
	TOTAL	42

Learning Outcomes: To be able to state and analyse

- Preprocessing techniques for various datasets,
- Standard database systems concepts like tables, relations, query
- Information retrieval techniques such as indexing, scoring, ranking, evaluation
- Data processing algorithms and data structures
- Visualization techniques

Learning Objectives:

- To be able to learn about the entire pipeline of a typical system involving data, collection, preprocessing, storage, retrieval, processing, analysis, and visualization.

Teaching Methodology **:** Lab based

Assessment Methods **:** Exam based

Text Books

1. Introduction to Algorithms. Cormen, Leiserson, Rivest, Stein. MIT Press 3ed. ISBN-13: 978-0262533058
2. Database System Concepts. Silberschatz, Korth, Sudarshan. McGraw Hill Education; Sixth edition.ISBN-13: 978-9332901384
3. Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools. Cielen, Meysman,Ali. Dreamtech Press. ISBN-13: 978-9351199373

References

1. Data Engineering: A Novel Approach to Data Design. Brian Shive. Technics Publications. ISBN-13: 978-1935504603
2. Python Data Science Handbook: Essential Tools for Working with Data. Joel Grus. O'Reilly. ISBN-13**:** 978-9352134915

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Code and Title	DS5005 ~~CS5011~~: Optimisation				
Programme	B.Tech/M.Tech/MS/PhD	Year of study		Semester	
Course credit	3-0-0-3				
Course category	PMT				
Prerequisite, if any					
Consent of teacher, if required	required				
Date of proposal		Date of Senate Approval			
Proposing faculty	Chandra Sekhar Lakshminarayanan				

Course Content

S/N	Topic	Lecture (hours)
1	Introduction: Motivation and examples	3
2	Basics: R^d, vectors, matrices, norm, sequences & convergence, functions in one and several variables, Taylor series, derivatives, gradient, sub-gradient, Hessian, properties of symmetric operators, contours, affine functions, hyper-planes, convex functions, minima: local and global, subspaces, affine spaces, half-spaces, convex sets	9
3	Unconstrained Optimisation: gradient descent, line search, rates for various classes of convex functions, steepest descent, Newton's method, conjugate gradient method, quasi-Newton method, linear least-squares regression: rates	9
4	Constrained Optimisation: linear and convex constraint sets, linear programming, simplex, interior-point method, duality theory: primal/dual programs, weak, strong duality, KKT conditions	12
5	Stochastic Optimisation: stochastic gradient descent, step-size conditions, Keifer-Wolwowitz method, simultaneous perturbation stochastic approximation (SPSA) method, smoothed functional method	9
	TOTAL	42

Learning Outcomes: As a result of this course, the student should be able to

- Pose a given optimisation problem by identifying the objective and constraints.
- Draw level sets and graphs of functions, identify constraint regions described by set of function.
- Choose stepsize and identify rates of convergence of optimisation algorithms.
- Use stochastic optimisation techniques to derive data driven algorithms.

Learning Objectives: To

- Look at the regions and functions in `d' dimensions.
- Build algorithms that find minima using first and second order information.
- Look at constraint optimisation and duality theory with linear programming as a special case.
- Introduce basics of stochastic optimisation

Teaching Methodology **:** Lecture based

Assessment Methods **:** Exam based

Text Books

1. Introduction to Optimization: Edwin K. P. Chong and Stanislaw H. Zak, Wiley-Interscience Series in Discrete Mathematics and Optimization (ISBN-13: 978-0471089490, ISBN-10: 0471089494).

References

1. Boyd, Stephen, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004. (ISBN-13: 978-0521833783, ISBN-10: 0521833787)
2. Fletcher, Roger. *Practical methods of optimization*. John Wiley & Sons, 2013. (ISBN: 978-0-471-49463-8)

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Code and Title	DS5102 ~~CS5104~~: Big Data Lab				
Programme	B.Tech/M.Tech/MS/PhD	Year of study		Semester	
Course credit	1-0-3-3				
Course category	PMT				
Prerequisite, if any					
Consent of teacher, if required	required				
Date of proposal		Date of Senate Approval			
Proposing faculty	Satyajit Das				

Course Content

S/N	Topic	Lecture (hours)	Lab (hours)
1	**Lab on set up :** amp; manipulating files in HDFS	2	9
2	**Basic programs of Hadoop MapReduce:** Driver code, Mapper code, Reducer code, RecordReader, Combiner, Partitioner	4	9
3	**Pig :** Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators	4	12
4	**Big data analytics in Spark using PySpark:** Installing Apache Spark, Spark Ecosystem, Resilient Distributed Dataset (RDD) in Spark, building machine learning model using PySpark	4	12
	TOTAL	14	42

Learning Outcomes:

- Preparing for data summarization, query, and analysis.
- Applying data modelling techniques to large data sets
- Creating applications for Big Data analytics
- Building a complete business data analytic solution

Learning Objectives:

- The primary objective of this course is to optimize business decisions and create a competitive advantage with Big Data analytics. This course will introduce the basics required to develop map reduce programs, derive business benefit from unstructured data. This course will also give an overview of the architectural concepts of Hadoop and introducing map reduce paradigm. Another objective of this course is to introduce programming tools PIG & HIVE in Hadoop ecosystem.

Teaching Methodology **:** Lecture and Lab based

Assessment Methods **:** Exam based

Text Books

1. Big Java 4th Edition, Cay Horstmann, Wiley John Wiley & Sons, INC, ISBN: 9780470509487
2. Hadoop: The Definitive Guide by Tom White, 3 rd Edition, O'reilly, ISBN: 9781449328917

References

1. Hadoop MapReduce Cookbook,Srinath Perera, Thilina Gunarathne, O'reilly, ISBN: 9781849517287
2. Hadoop for Dummies by Dirk deRoos, Paul C.Zikopoulos, Roman B.Melnyk,Bruce Brown, Rafael Coss, John Wiley & Sons, 2014, ISBN: 1118607554
3. Hadoop in Practice by Alex Holmes, MANNING Publication, ISBN: 9351197425

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Code and Title	CS5012 : AI for Cyber Security				
Programme	B.Tech/M.Tech/MS/PhD	Year of study		Semester	
Course credit	3-0-0-3				
Course category	PMT				
Prerequisite, if any	Familiarity with Probability, Machine Learning				
Consent of teacher, if required	Required				
Date of proposal		Date of Senate Approval			
Proposing faculty	Vivek Chaturvedi				

Course Content

S/N	Topic	Lecture (hours)
1	Overview on Machine Learning with use cases from cybersecurity, classification of threats, attacks, vulnerabilities, malware, trojans etc.	6
2	Classification of malware using supervised/unsupervised learning based on signatures and profiling. Decision Tree and context based malicious event detection	9
3	Time Series Analysis and Ensemble modelling to detect deviation from normal behaviour, case studies in Reconnaissance detection	9
4	Efficient Network Anomaly detection; familiarize with various stages of network attack and address using deep neural networks, develop intrusion detection systems	9
5	Adversarial attacks on ML systems, model poisoning, black box attacks, white box attacks, state-of-art research paper reading on deep learning systems	9
	TOTAL	42

Learning Outcomes:

- Students will be able to develop ML models to classify malwares.
- Able to implement simple intrusion detection systems using deep neural networks.
- They will be able to demonstrate the vulnerabilities in ML systems and state methods to address adversarial attacks.

Learning Objectives:

- Machine Learning (ML) is increasingly used in sensitive and time-critical systems such as autonomous driving, cyber physical systems etc. to deliver higher performance and protect the confidentiality of the systems. Though ML based systems can be used to classify various malware attacks and develop intrusion detection systems, these systems are also susceptible to several adversarial attacks. This course covers a systematic approach on developing ML based cybersecurity methodologies. It will also cover adversarial attacks which intentionally forces ML systems to behave unexpectedly.

Teaching Methodology **:** Lecture based

Assessment Methods **:** Exam based

Text Books

1. A. Hands-on Machine Learning for Cyber Security by Soma Halder, ISBN139781788992282

References

1. Machine Learning and Security by David Freeman, Clarence Chio Publisher: O'Reilly Media, Inc. Release Date: February 2018 ISBN: 9781491979891

2. Malware Data Science by Joshua Saxe with Hillary Sanders, ISBN-10: 1-59327-859-4 ISBN-13: 978-1-59327-859-5 Publisher: William Pollock

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Code and Title	HS5004: Econometrics				
Programme	B.Tech/M.Tech/MS/PhD	Year of study		Semester	
Course credit	3-0-0-3				
Course category	ERC				
Prerequisite, if any	Familiarity with Statistics				
Consent of teacher, if required	Required				
Date of proposal		Date of Senate Approval			
Proposing faculty	Amrita Roy				

Course Content

S/N	Topic	Lecture (hours)
1	Introduction to econometrics	3
2	Simple Linear Regression Model: Two Variable Case, Estimation of model by the method of ordinary least squares (OLS), Properties of OLS, Gauss-Markov Assumptions, Hypothesis testing with OLS	6
3	Multiple Regression Analysis: Estimation, Partial regression coefficients, Interpretations of OLS estimates, Goodness-of-fit and selection of regressor, Multiple Regression Analysis with qualitative (dummy) independent variables	6
4	Violations of Classical Assumptions: Consequences, Detection and Remedies, Multicollinearity, Heteroscedasticity, Serial Correlation	12
5	Specification Analysis: Omission of a relevant variable, Inclusion of irrelevant variable, Tests of specification	12
	TOTAL	39

Learning Outcomes:

- The student will be able to apply the tool econometrics to study our world.
- Students will be able to construct, estimate an econometric model and interpret the results for our application.

Learning Objectives:

- Principal objective of the course is to provide basic understanding of Econometrics as it is used as a tool in various fields of study like economics and other social sciences. This course emphasises on understanding the basic techniques and its application in the empirical aspect of our analysis.

Teaching Methodology **:** Lecture based

Assessment Methods **:** Exam based

Text Books

1. C. Dougherty, Introduction to Econometrics, 4th edition, Oxford, OUP, ISBN-13: 978-0-19-965050-7
2. J.H. Stock and M.W. Watson, *Introduction to Econometrics* (first edition), Addison-Wesley, 2003, ISBN-13: 978-0133486872

References

1. J.M. Wooldridge, *Introductory Econometrics* (fifth edition), South-Western College Publishing, ISBN-13: 978-1-111-53104-1
2. D. N. Gujarati, Basic Econometrics (fourth edition), McGrawHill, ISBN:978-0-07-112342-6

Response to Reviewers Comment:

Course Title: MDS curriculum along with following course

- DS xxxx : Machine Learning Lab, Proposing faculty Dr. Sahely
- EE xxxx : Linear Algebra, Proposing faculty Dr. Lakshmi Narasimhan
- DS xxxx : Data Engineering, Proposing faculty Dr. Mrinal
- DS xxxx : Data Engineering Lab, Proposing faculty Dr. Mrinal
- DS xxxx : Big Data Lab, Proposing faculty Dr. Satyajit
- CS xxxx : AI for cyber security, Proposing faculty Dr. Vivek

Reviewer 1: Prof B. Ravindran, IIT Madras (Chairman of Curriculum Committee for MTech Data Science)

1. I am fine with the courses. Two concerns: the data engineering and ML lab syllabus is not well defined. In fact there is no mention of lab hours or exercises in the ML course. The cyber security course syllabus seems more like a sketch than the complete syllabus. The curriculum is ok.

Response:

1. We have put hourly details for ML Lab. Then received Okay from him.
2. We have put hourly details for the Data Engineering Lab. Then received Okay from him.
3. We have worked on the Cyber Security Course. Then we received "Looks good. Perhaps the cybersecurity course can be renamed as AI for cybersecurity?"
4. We have then change the name as proposed

Reviewer 2: Prof Chiranjib Bhattacharyya, IISc Bangalore (External member of Curriculum Committee for MTech Data Science)

1. The courses are well structured and extremely suited for a MTech program in Data Science. This curriculum has my full support. best

Response: NA

Reviewer 3:

1.

Response:

List of possible electives (tentative)

Semester-2	Semester-3
Reinforcement Learning	Digital Signal Processing
Information Theory	Causal Reasoning
Econometrics	Natural Language Processing
Bayesian Models	Computer Vision
Kernel Methods	Foundation of ML
Detection Theory	
AI on Edge	

Approval status of courses

Probability and Statistics	Will be through Mathematics Department
Linear Algebra	Proposed here
Machine Learning	Approved as CS 5512: Machine Learning
Machine Learning Lab	Proposed here
Data Engineering	Proposed here
Data Engineering Lab	Proposed here
Optimization	Proposed here
Big Data Lab	Proposed here
Deep Learning	Approved as CS5007**:** Deep Learning
Cyber Security	Proposed here
Econometrics	Proposed here