UNIVERSITY OF MADRAS

M.Sc. DEGREE PROGRAMME IN COMPUTER SCIENCE

SYLLABUS WITH EFFECT FROM 2023-2024

Title of the Paper	**Data Science and Analytics**		
Core–IX - Theory	II Year & III Semester	Credit: 4	536C3A

Objectives:

- Understand the Big Data Platform and its Use cases
- Provide an overview of Apache Hadoop
- Provide HDFS Concepts and Interfacing with HDFS
- Understand Map Reduce Jobs
- Provide hands on Hodoop Eco System

Outcomes:

1.	Identify Big Data and its Business Implications.	K1, K2
2.	List the components of Hadoop and Hadoop Eco-System	K2, K3
3.	Access and Process Data on Distributed File System	K3, K4
4.	Manage Job Execution in Hadoop Environment	K4, K5
5.	Develop Big Data Solutions using Hadoop Eco System	K6
K1-Remember;K2-Understand;K3-Apply;K4-Analyze;K5-Evaluate; K6-Create		

UNIT I:INTRODUCTION TO BIG DATA AND HADOOP

Types of Digital Data, Introduction to Big Data, Big Data Analytics, History of Hadoop, Apache Hadoop, AnalysingData with Unix tools, Analysing Data with Hadoop, Hadoop Streaming, Hadoop Echo System, IBM Big DataStrategy, Introduction to Infosphere BigInsights and Big Sheets.

UNIT II:HDFS(Hadoop Distributed File System)

The Design of HDFS, HDFS Concepts, Command Line Interface, Hadoop file system interfaces, Data flow, DataIngest with Flume and Scoop and Hadoop archives, Hadoop I/O: Compression, Serialization, Avro and File-BasedData structures

UNIT III: Map Reduce: Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Typesand Formats, Map Reduce Features.

UNIT IV: Hadoop Eco System**:** Pig : Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User DefinedFunctions, Data Processing operators.Hive : Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, QueryingData and User Defined Functions.

UNIT V: Hbase: HBasics, Concepts, Clients, Example, Hbase Versus RDBMS.Big SQL: Introduction - Machine Learning: Introduction, Supervised Learning, Unsupervised Learning, Collaborative Filtering.

Text Books

1. Tom White " Hadoop: The Definitive Guide" Third Edit on, O'reily Media, 2012.
2. Seema Acharya, Subhasini Chellappan, "Big Data Analytics" Wiley 2015.

Print to PDF with PDF Writer for Windows 8. This is a free evaluation copy. Buy full version now.

References

1. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007.
2. Jay Liebowitz, "Big Data and Business Analytics" Auerbach Publications, CRC press (2013)
3. Anand Rajaraman and Jefrey David Ulman, "Mining of Massive Datasets", Cambridge University Press,2012.
4. Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced
5. Analytics", John Wiley & sons, 2012.
6. Glen J. Myat, "Making Sense of Data", John Wiley & Sons, 2007
7. Pete Warden, "Big Data Glossary", O'Reily, 2011.
8. Michael Mineli, Michele Chambers, Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley Publications, 2013.
9. ArvindSathi, "BigDataAnalytics: Disruptive Technologies for Changing the Game", MC Press, 2012
10. Paul Zikopoulos ,DirkDeRoos , Krishnan Parasuraman , Thomas Deutsch , James Giles , David Corigan ,
11. "Harness the Power of Big Data The IBM Big Data Platform ", Tata McGraw Hill Publications, 2012.

Related Online Contents

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	S	S	S	S	S	S	S	M	M	S
CO2	S	S	S	S	S	S	S	M	S	S
CO3	S	S	S	S	S	S	S	M	S	S
CO4	S	S	S	S	S	S	S	M	S	S
CO5	S	S	S	S	S	S	S	M	S	S

*S-Strong;M-Medium;L-Low

Print to PDF with PDF Writer for Windows 8. This is a free evaluation copy. Buy full version now.

UNIVERSITY OF MADRAS

M.Sc. DEGREE PROGRAMME IN COMPUTER SCIENCE

SYLLABUS WITH EFFECT FROM 2023-2024

Titleof the Paper	**Machine Learning**		
Core–X - Theory	IIYear&IIISemester	Credit: 4	536C3B

Objectives:

To provide mathematical base for Machine learning

To provide theoretical knowledge on setting hypothesis for pattern recognition.

To impart Knowledge of machine learning techniques for data handling

To provide the skill to evaluate the performance of algorithms and to provide solution for various real-world applications.

To impart the knowledge of identifying similarities and differences in various patterns of data

Outcomes:

1.	Recognize the characteristics of machine learning strategies. Apply various supervised learning methods to appropriate problems.	K1, K2
2.	Identify and integrate more than one technique to enhance the performance of learning.	K2, K3
3.	Analyze the co-occurrence of data to find interesting frequent patterns.	K3, K4
4.	Preprocess the data before applying to any real-world problem and can evaluate its performance.	K4, K5
5.	Create probabilistic and unsupervised learning models for handling unknown pattern.	K5, K6
K1-Remember;K2-Understand;K3-Apply;K4-Analyze;K5-Evaluate; K6-Create		

Unit I: BASIC MATHEMATICS FOR MACHINE LEARNING: Regression Correlation and Regression, types of correlation – Pearson's, Spearman's correlations –Ordinary Least Squares, Fitting a regression line, logistic regression, Rank Correlation Partial and Multiple correlation-Multiple regression, multicollinearity. Gradient descent methods, Newton method, interior point methods, active set, proximity methods, accelerated gradient methods, coordinate descent, cutting planes, stochastic gradient descent. Discriminant analysis, Principal component analysis, Factor analysis, k means.

Unit II: INTRODUCTION TO MACHINE LEARNING: Introduction, Examples of various Learning Paradigms, Perspectives and Issues, Version Spaces, Finite and Infinite Hypothesis Spaces, PAC Learning, VC Dimension.

Unit III: SUPERVISED LEARNING ALGORITHMS Learning a Class from Examples, Linear, Non-linear, Multi-class and Multi-label classification, Decision Trees: ID3, Classification and Regression Trees (CART), Regression: Linear Regression, Multiple Linear Regression, Logistic Regression. Neural Networks: Introduction, Perceptron, Multilayer Perceptron, Support vector machines: Linear and Nonlinear, Kernel Functions, K-Nearest Neighbors

Unit IV: ENSEMBLE LEARNING: Ensemble Learning Model Combination Schemes, Voting, Error-Correcting Output Codes, Bagging: RandomForest Trees, Boosting: Adaboost, Stacking:

Print to PDF with PDF Writer for Windows 8. This is a free evaluation copy. Buy full version now.

UNSUPERVISED LEARNING: Introduction to clustering, Hierarchical: AGNES, DIANA, Partitional: K-means clustering, K-Mode Clustering, Self Organizing Map, Expectation Maximization, Gaussian Mixture Models, Principal Component Analysis (PCA), Locally Linear Embedding (LLE), Factor Analysis

Unit V: PROBABILISTIC LEARNING: Bayesian Learning, Bayes Optimal Classifier, Naïve Bayes Classifier, Bayesian Belief Networks, Mining Frequent Patterns: MACHINE LEARNING IN PRACTICE: Design, Analysis and Evaluation of Machine Learning Experiments, Other Issues: Handling imbalanced data sets

Recommended Texts:

1. EthemAlpaydin, "Introduction to Machine Learning", MIT Press, Prentice Hall of India, Third Edition 2014.
2. MehryarMohri, Afshin Rostamizadeh, Ameet Talwalkar "Foundations of Machine Learning", MIT Press, 2012.

Reference Books:

1. Tom Mitchell, "Machine Learning", McGraw Hill, 3 rd Edition,1997.
2. Charu C. Aggarwal, "Data Classification Algorithms and Applications", CRC Press, 2014.
3. Stephen Marsland, "Machine Learning – An Algorithmic Perspective", 2 nd Edition, CRC Press, 2015.
4. Kevin P. Murphy "Machine Learning: A Probabilistic Perspective", The MIT Press, 2012
5. Jiawei Han and MichelineKambers and Jian Pei, "Data Mining –Concepts and Techniques", 3 rdEdition,Morgan Kaufman Publications, 2012.
6. Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong, "Mathematics for Machine Learning", Cambridge University Press, 2019.

Web References:

1. https://www.youtube.com/watch?v=r4sgKrRL2Ys&list=PL1xHD4vteKYVpaIiy295pg6_SY5qznc77

Mapping with Programme Outcomes:

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10
CO 1	S	S	M	S	M	L	M	S	L	L
CO 2	S	M	S	L	S	L	M	L	M	S
CO 3	M	S	L	M	M	S	L	S	L	S
CO 4	L	S	S	L	S	M	S	L	S	M
CO 5	S	L	M	S	L	L	M	S	M	S

S-Strong M-Medium L-Low

Print to PDF with PDF Writer for Windows 8. This is a free evaluation copy. Buy full version now.

UNIVERSITY OF MADRAS

M.Sc. DEGREE PROGRAMME IN COMPUTER SCIENCE

SYLLABUS WITH EFFECT FROM 2023-2024

Title of the Paper	**Theory of Computation**		
Core–XI - Theory	II Year & III Semester	Credit: 3	536C3C

Objectives:

To give an overview of the theoretical foundations of computer science from the perspective of formal languages

To illustrate finite state machines to solve problems in computing

To explain the hierarchy of problems arising in the computer sciences.

To familiarize Regular grammars, context frees grammar.

To use basic concepts of formal languages of finite automata techniques

Outcomes:

1.	Use the concepts and techniques of discrete mathematics for theoretical computer science	K1
2.	Design Finite Automata for different Regular Expressions and Languages	K2
3.	Identify and use different formal languages and their relationship.	K3,K4
4.	To solve various problems of applying normal form techniques, push down automata and Turing Machines	K4,K5
5.	Analyze various concepts of undecidability and Computable Function and Discuss analytically and intuitively for problem-solving situation	K6
K1-Remember;K2-Understand;K3-Apply;K4-Analyze;K5-Evaluate; K6-Create		

Unit I: Review of Mathematical Theory

Combinatorics: Review of Permutation and Combination - Mathematical Induction - Pigeon hole principle - Principle of Inclusion and Exclusion - generating function - Recurrence relations. Statements – Connectives – Truth Tables – Normal forms – Predicate calculus – Inference – Theory for Statement Calculus and Predicate Calculus

Unit-II: Regular Languages and Finite Automata

Regular Expressions, Regular Languages, Application of Finite Automata, Automata with output - Moore machine & Mealy machine, Finite Automata, Memory requirement in a recognizer, Definitions, union- intersection and complement of regular languages, Non Deterministic Finite Automata, Conversion from NFA to FA, ??- Non Deterministic Finite Automata, Conversion of NFA- ? to NFA, Kleene's Theorem, Minimization of Finite automata, Regular And Non Regular Languages – pumping lemma.

Unit-III: Context free grammar (CFG)

Definitions and Examples, Unions Concatenations And Kleene's of Context free language, Regular Grammar for Regular Language, Derivations and Ambiguity , Unambiguous CFG and Algebraic Expressions, Backus Naur Form (BNF), Normal Form – CNF.

Print to PDF with PDF Writer for Windows 8. This is a free evaluation copy. Buy full version now.

Unit-IV: Pushdown Automata, CFL And NCFL
Definitions, Deterministic PDA, Equivalence of CFG and PDA & Conversion, Pumping lemma for CFL, Intersections and Complements of CFL, Non-CFL.

Unit-V: Turing Machine (TM)
TM Definition, Model Of Computation, Turing Machine as Language Acceptor, TM that Compute Partial Function, Church Turing Thesis, Combining TM, Variations Of TM, Non Deterministic TM, Universal TM, Recursively and Enumerable Languages, Context sensitive languages and Chomsky hierarchy.

Recommended Texts:

1. John E. Hopcroft, Rajeev Motwani, Jeffrey D. Ullman ; Introduction to Automata Theory Languages andComputation; Pearson Education, India; 3rd edition;2008
2. KENNETH H. ROSEN ; Discrete Mathematics and Its Applications (SIE) 8th Edition ;2021

Reference Books:

1. K. L. P Mishra, N. Chandrashekaran (2003), Theory of Computer Science Automata Languages and Computation, 2nd edition, Prentice Hall of India, India.

Web References:

1. https://www.youtube.com/playlist?list=PLbtzT1TYeoMjNOGEiaRmm_vMIwUAidnQz
2. https://nptel.ac.in/courses/106106049

Mapping with Programme Outcomes:

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10
CO 1	S	M	M	S	M	L	S	S	L	M
CO 2	S	M	S	L	S	L	M	L	M	S
CO 3	M	S	L	M	M	S	L	S	L	S
CO 4	L	S	S	L	S	M	S	L	S	M
CO 5	S	L	M	S	L	L	M	S	M	S

S-Strong M-Medium L-Low

Print to PDF with PDF Writer for Windows 8. This is a free evaluation copy. Buy full version now.

UNIVERSITY OF MADRAS

M.Sc. DEGREE PROGRAMME IN COMPUTER SCIENCE

SYLLABUS WITH EFFECT FROM 2023-2024

Title of the Paper	**Machine Learning Practical**		
Core–XIII - Practical	II Year & III Semester	Credit:3	536C3E

Objectives:

Make use of Data sets in implementing the machine learning algorithms

Implement the machine learning concepts and algorithms in any suitable language of choice.

The programs can be implemented in either JAVA or Python.

For Problems 1 to 6 and 10, programs are to be developed without using the builtin classes or APIs of Java/Python.

Data sets can be taken from standard repositories (https://archive.ics.uci.edu/ml/datasets.html) or constructed by the students.

Outcomes:

1.	Understand the implementation procedures for the machine learning algorithms.	K1, K2
2.	Design Java/Python programs for various Learning algorithms.	K2, K3
3.	Apply appropriate data sets to the Machine Learning algorithms.	K3, K4
4.	Identify and apply Machine Learning algorithms to solve real world problems.	K4, K5
5.	be capable of confidently applying common Machine Learning algorithms in practice and implementing their own;	K5, K6
K1-Remember;K2-Understand;K3-Apply;K4-Analyze;K5-Evaluate; K6-Create		

1. Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file
2. For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.
3. Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.
4. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.
5. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.
6. Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set.

Print to PDF with PDF Writer for Windows 8. This is a free evaluation copy. Buy full version now.

7. Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. You can use Java/Python ML library classes/API.
8. Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering. You can add Java/Python ML library classes/API in the program.
9. Write a program to implement k-Nearest Neighbor algorithm to classify the iris data set. Print both correct and wrong predictions. Java/Python ML library classes can be used for this problem.
10. Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.

Recommended Texts:

1. Dr. Kamlesh Namdev, LAP LAMBERT ; Lab manual of Machine Learning: Machine Learning Practicals in Python; Academic Publishing; 2021

Reference Books:

1. Introduction to Machine Learning with Python by Andreas C. Müller, Sarah Guido ReleasedOctober 2016 Publisher(s): O'Reilly Media, Inc. ISBN: 9781449369415

Web References:

1. https://www.youtube.com/watch?v=RnFGwxJwx-0

Mapping with Programme Outcomes:

Mapping with Programmers outcomes*										
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	S	M	S	L	S	M	L	M	L	L
CO2	S	M	S	S	S	M	M	L	M	L
CO3	M	S	M	M	S	S	L	L	L	M
CO4	S	L	M	M	M	L	M	L	M	S
CO5	S	S	M	S	L	M	M	L	M	L

S-Strong M-Medium L-Low

Print to PDF with PDF Writer for Windows 8. This is a free evaluation copy. Buy full version now.

UNIVERSITY OF MADRAS

M.Sc. DEGREE PROGRAMME IN COMPUTER SCIENCE

SYLLABUS WITH EFFECT FROM 2023-2024

Title of the Paper	**Data Science and Analytics Practical**		
Core–XII - Practical	II Year & III Semester	Credit:3	536C3D

Objectives:

- Understand Hadoop Distributed File System and examine MapReduce Programming.
- Explore Hadoop tools and manage Hadoop with Ambari.
- Appraise the role of Business intelligence and its applications across industries.
- Assess core data mining techniques for data analytics
- Identify various Text Mining techniques

Outcomes:

1.	To Describe the key issues in Big Data Management and experiment with the Hadoop framework.	K1, K2
2.	To Explain the structure and unstructured data by using NoSQL commands.	K2, K3
3.	To Apply scientific computing algorithms for finding similar items and clustering.	K3, K4
4.	To Test fundamental enabling techniques and scalable algorithms for data stream mining.	K4, K5
5.	To Develop problem solving and critical thinking skills in fundamental enable techniques like Hadoop & MapReduce	K6
K1-Remember;K2-Understand;K3-Apply;K4-Analyze;K5-Evaluate; K6-Create		

.

LISTOF PROGRAMS

1. Installation of Hadoop Framework, it‘s components and study the HADOOP ecosystem
2. Write a program to implement word count program using MapReduce
3. Experiment on Hadoop Map-Reduce / PySpark: -Implementing simple algorithms in Map-Reduce: Matrix multiplication
4. Install and configure MongoDB/ Cassandra/ HBase/ Hypertable to execute NoSQL Commands.
5. Implementing DGIM algorithm using any Programming Language/ Implement Bloom Filter using any programming language
6. Implement and Perform Streaming Data Analysis using flume for data capture, PYSpark / HIVE for data analysis of twitter data, chat data, weblog analysis etc.
7. Implement any one Clustering algorithm (K-Means/CURE) using Map-Reduce.
8. Implement Page Rank Algorithm using Map-Reduce.

Print to PDF with PDF Writer for Windows 8. This is a free evaluation copy. Buy full version now.

UNIVERSITY OF MADRAS

M.Sc. DEGREE PROGRAMME IN COMPUTER SCIENCE

SYLLABUS WITH EFFECT FROM 2023-2024

Title of the Paper	**Cryptography**		
Elective –V - Theory	II Year & III Semester	Credit:3	536E3B

Objectives:

To learn the emerging concepts of cryptography and algorithms

To defend the security attacks on information systems using secure algorithms and Authentication process

To categorize and analyze the key concepts in network and wireless security

To Infer the need of security to introduce strong cryptosystems.

Outcomes:

1.	Analyze the cryptographic algorithms for information security.	K1, K2
2.	Identify the authentication schemes for membership authorization.	K2, K3
3.	Identify computer and network security threats, classify the threats and develop a security model for detect and mitigate the attacks.	K3, K4
4.	Identify the requirements for secure communication and challenges related to the secure web services	K4, K5
5.	Ability to identify the need of ethical and professional practices, risk management Using emerging security solutions.	K5, K6
K1-Remember;K2-Understand;K3-Apply;K4-Analyze;K5-Evaluate; K6-Create		

Unit I: Introduction and Symmetric Key Cryptographic Systems : Introduction to Cryptography, Types of Attacks, Symmetric Key Cryptography, Data Encryption Standard (DES), Differential and Linear cryptanalysis, Advanced Encryption Standard(AES), Modes of operation, Stream Ciphers: Feedback shift registers, Stream ciphers based on LFSRs.

Unit II: Asymmetric Key Cryptosystems: Applications of asymmetric Cryptosystems – RSA Rabin, Elgamal, Probabilistic Cryptosystems, Elliptic Curve Cryptography (ECC), Diffie-Hellman key exchange protocol, Chinese Remainder Theorem (CRT).

Unit III: Data Integrity and Authentication: Message Authentication Code (MAC), Hash function properties, General model for iterated hash functions -MD5, Secure Hash algorithms, HMAC, Attacks on hash functions.

Unit IV: Digital Signature algorithm, Public key infrastructure: X. 509 digital certificate, Kerberos, Zero-Knowledge Protocol.

Unit V: Advanced Cryptographic Techniques: Multiparty Computation and Secret Sharing, Introduction - Indistinguishability - Secret - Sharing Simulation - Based Security-Security against Active Corruption-BGW Protocol (Active, Honest Majority)- Homomorphic Encryption-Lattice Cryptography

Print to PDF with PDF Writer for Windows 8. This is a free evaluation copy. Buy full version now.

Recommended Texts:

1. J. Katz and Y. Lindell, Introduction to Modern Cryptography. Chapman & Hall/CRC Press, 2014
2. W. Stallings, Cryptography and Network Security: Principles and Practice, 7th Ed. Pearson Publishers, 2017.
3. C. Paar and J Pelzl, Understanding Cryptography, Springer, 2010
4. Behrouz A. Forouzan, Cryptography and Network Security:6th Ed. McGraw Hill,2017
5. Dan Boneh and Victor Shoup, A Graduate Course in Applied Cryptography, Jan 2020

Reference Books:

1. Kaufman, Perlman and Speciner. Network Security: Private Communication in a Public World., 2 nd edition,2002 , Pearson Publishers (ISBN No.:978-01-3-04601- 96)
2. Alfred J. Menezes, Paul C. van Oorschot and Scott A. Vanstone,Handbook of Applied Cryptography,5th edition,2001,CRC Press,(ISBN No:0-8493-8523-7)
3. D. R. Stinson, Cryptography: Theory and Practice, 3 rd Ed. Boca Raton, FL: Chapman &Hall/CRC, 2005. (ISBN No.:978-1-58-488508-5)J. H. Silverman, A Friendly
4. Introduction to Number Theory, 4th Ed. Boston: Pearson, 2012. (ISBN No.:978-0- 321-81619-1)
5. Ronald Cramer, Ivan BjerreDamgård, JesperBuus Nielsen, “Secure Multiparty Computation and Secret Sharing”, ISBN 9781107043053, Cambridge University Press, 2015
6. Philip N. Klein, “A Cryptography Primer-Secrets and Promises”, ISBN 9781107603455, Cambridge University Press, 2014

Web References:

1. https://www.youtube.com/watch?v=iTVyKbDCJrA&list=PLgMDNELGJ1CbdGLyn7OrVAP-IKg-0q2U2

Mapping with Programme Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	M	S	M	M	S	M	L	M	L	S
CO2	S	S	M	L	M	L	M	L	M	S
CO3	S	S	L	M	S	S	L	M	L	M
CO4	M	L	S	M	M	L	S	L	M	S
CO5	S	S	L	S	L	M	L	M	M	L

S-Strong M-Medium L-Low

Print to PDF with PDF Writer for Windows 8. This is a free evaluation copy. Buy full version now.