CURRICULUM & SYLLABUS

For

B.E(Hons.) Computer Science and Engineering with Specialization in Artificial Intelligence and Data Science

(Choice Based Credit System)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Sri Chandrasekharendra Saraswathi ViswaMahavidyalaya
SCSVMV
(Deemed to be University U/S 3 of UGC Act 1956)
Accredited with "A" Grade by NAAC
Enathur, Kanchipuram – 631 561

B.E(Hons.) Computer Science and Engineering with Specialization in Artificial Intelligence and Data Science

Professional Specialized Courses (PSC)								
SL. No	Semester	Course Category	Course Code	Name of the Course	Hours per week			Credit
					L	T	P	
1.	3	PSC		AI & Machine Learning	3	0	0	3
2.	4	PSC		Introduction to Data Science	3	0	0	3
3.	4	PSC		R Programming Lab	0	0	4	2
4.	5	PSC		Natural Language processing	3	0	0	3
5.	5	PSC		Data Visualization lab	0	0	4	2
6.	6	PSC		Web and Social Media Analytics	3	0	0	3
7.	7	PSC		Mining Massive Dataset	3	0	0	3
Total Credits								**19**

SEMESTER-III

Course Code :	AI & MACHINE LEARNING	L	T	P	C
		3	0	0	3

Objectives:

1. Introduce and define the meaning of Intelligence and explore various paradigms for knowledge encoding in computer systems.
2. Introduce subfields of AI such as NLP, Game Playing, Bayesian Models, etc.
3. Introduce the concept of learning patterns from data and develop a strong theoretical foundation for understanding state of the art Machine Learning algorithms.

Prerequisites:

1. Basic knowledge of Mathematical Logic
2. Basic knowledge of Linear Algebra and Calculus

UNIT - I

Defining Artificial Intelligence, Defining AI techniques - State Space Search and Heuristic Search Techniques - Production systems and characteristics, Hill Climbing, Breadth first and depth first search, Best first search

UNIT - II

Representations and Mappings, Approaches to knowledge representation - Representing simple facts in logic, Computable functions and predicates, Procedural vs Declarative knowledge, Logic Programming, Forward vs backward reasoning- Non-monotonic Reasoning, Logics for non-monotonic reasoning

UNIT - III

Idea of Machines learning from data, Classification of problem – Regression and Classification, Supervised and Unsupervised learning - Model representation for single variable, Single variable Cost Function, Gradient Decent for Linear Regression, Multivariable model representation, Multivariable cost function, Gradient Decent in practice, Normal Equation and non-invertibility

UNIT - IV

Classification, Hypothesis Representation, Decision Boundary, Cost function, Advanced Optimization, Multi-classification (One vs All), Problem of Overfitting, Regularization

UNIT - V

Case Studies: Neural Networks - Support Vector Machines - Recommender Systems

Text Books:

1. Artificial Intelligence: A Modern Approach, Stuart Russel, Peter Norvig
2. Artificial Intelligence, 2nd Edition, Rich and Knight
3. Machine Learning, Tom M. Mitchell
4. Building Machine Learning Systems with Python, Richert & Coelho

SEMESTER-IV

Course Code :	DATA SCIENCE FOR ENGINEERS	L	T	P	C
		3	0	0	3

COURSE OUTCOME:

1. Learn about the Data Evolution and understanding the data
2. Understand the basic concepts of data science.
3. Analyze the basic concepts of Bigdata.
4. Understand the fundamental principles of R.
5. Apply the statistical measures of R in real time environment

UNIT- I

Linear algebra for data science: algebraic view - vectors, matrices, product of matrix & vector, rank, null space, solution of over-determined set of equations and pseudo-inverse.

Linear algebra for data science: geometric view - vectors, distance, projections, eigen value decomposition.

UNIT - II

Statistics: descriptive statistics, notion of probability, distributions, mean, variance, covariance, and covariance matrix.

UNIT - III

Optimization: Typology of data Science problems and a solution framework. Univariate andmultivariate linear regression Model assessment- cross validation.

UNIT - IV

Verifying assumptions used in linear regression, Assessing importance of different variables, subset selection.

UNIT - V

Introduction to classification and classification using logistics regression, Classification using various clustering techniques.

TEXT BOOKS:

Jeffrey S.Saltz, Jeffre M. Stanton, "An Introduction to Data Science", Sage Publications,2018

REFERENCES:

1. Nina Zumal, John Mount (2014). Practical Data science in R, Managing Publication Company
2. Bernard Kolman, Robert C. Busby and Sharon Ross (2004). Discrete Mathematical Structures,New Delhi: Prentice Hall
3. V. Bhuvaneswari, T. Devi, (2016). Big Data Analytics: A Practitioner's Approach, Bharathiar University 4. V. Bhuvaneswari (2016). Data Analytics with R, Bharathiar University.
4. https://nptel.ac.in/courses/106/106/106106179/

Course Code :	R PROGRAMMING LAB	L	T	P	C
		0	0	4	2

PRE-REQUISITE:
Basics of Mathematics and Programming Skill

COURSE OBJECTIVES:
Students will be able to:

- Use R for statistical programming, computation, graphics, and modelling.
- Write functions and use R in an efficient way.
- Fit some basic types of statistical models.
- Use R in their own research
- Be able to expand their knowledge of R on their own.

COURSE OUTCOMES:
Course Outcomes:
At the end of the Course, the Student will be able to:

1. Show the installation of R Programming Environment.
2. Utilize and R Data types for developing programs.
3. Make use of different R Data Structures.
4. Develop programming logic using R Packages.
5. Analyze the datasets using R programming capabilities

MAPPING WITH PROGRAMME OUTCOMES:

	PO 01	PO 02	PO 03	PO 04	PO 05	PO 06	PO 07	PO 08	PO 09	PO 10	PO 11	PO 12	PS 01	PS 02	PS 03
CO1	S	M	M					M			L				
CO2		M	M		L		L	M							
CO3									L				S		
C04		M	S	S				L		M					
CO5		S		L						M		L			
S - Strong M – Medium L – Low															

LIST OF EXERCISES

1. Download and install R-Programming environment and install basic packages using

   install.packages() command in R.

2. Learn all the basics of R-Programming (Data types, Variables, Operators etc.)

3. Implement R-Loops with different examples.

4. Learn the basics of functions in R and implement with examples.

5. Implement data frames in R. Write a program to join columns and rows in a data frame

using cbind() and rbind() in R.

6. Implement different String Manipulation functions in R.

7. Implement different data structures in R (Vectors, Lists, Data Frames)

8. Write a program to read a csv file and analyze the data in the file in R

9. Create pie charts and bar charts using R.

10. Create a data set and do statistical analysis on the data using R

REFERENCES:

1. Jared P. Lander, R for Everyone: Advanced Analytics and Graphics, 2nd Edition, Pearson Education, 2018.

2. S. R. Mani Sekhar and T. V. Suresh Kumar, Programming with R, 1st Edition,, CENGAGE, 2017.

PREPARED BY:

Dr. R. POORVADEVI, Assistant Professor/CSE

SEMESTER-V

Course Code :	NATURAL LANGUAGE PROCESSING	L	T	P	C
		3	0	0	3

OBJECTIVES:

- To learn the fundamentals of natural language processing
- To understand the use of CFG and PCFG in NLP
- To understand the role of semantics of sentences and pragmatics
- To apply the NLP techniques to IR applications

OUTCOMES:

Upon completion of the course, the students will be able to:

1. To tag a given text with basic Language features
2. To design an innovative application using NLP components
3. To implement a rule based system to tackle morphology/syntax of a language
4. To design a tag set to be used for statistical processing for real-time applications
5. To compare and contrast the use of different statistical approaches for different types of NLP applications.

UNIT- I

Origins and challenges of NLP – Language Modeling: Grammar-based LM, Statistical LM – Regular Expressions, Finite-State Automata – English Morphology, Transducers for lexicon and rules, Tokenization, Detecting and Correcting Spelling Errors, Minimum Edit Distance

UNIT- II

Unsmoothed N-grams, Evaluating N-grams, Smoothing, Interpolation and Backoff – Word Classes, Part-of-Speech Tagging, Rule-based, Stochastic and Transformation-based tagging, Issues in PoS tagging – Hidden Markov and Maximum Entropy models.

UNIT - III

Context-Free Grammars, Grammar rules for English, Treebanks, Normal Forms for grammar – Dependency Grammar – Syntactic Parsing, Ambiguity, Dynamic Programming parsing – Shallow parsing – Probabilistic CFG, Probabilistic CYK, Probabilistic Lexicalized CFGs – Feature structures, Unification of feature structures.

UNIT - IV

Requirements for representation, First-Order Logic, Description Logics – Syntax-Driven Semantic analysis, Semantic attachments – Word Senses, Relations between Senses, Thematic Roles, selectional restrictions – Word Sense Disambiguation, WSD using Supervised, Dictionary & Thesaurus, Bootstrapping methods – Word Similarity using Thesaurus and Distributional methods.

UNIT - V

Discourse segmentation, Coherence – Reference Phenomena, Anaphora Resolution using Hobbs and Centering Algorithm – Coreference Resolution – Resources: Porter Stemmer, Lemmatizer, Penn Treebank, Brill's Tagger, WordNet, PropBank, FrameNet, Brown Corpus, British National Corpus (BNC).

TEXT BOOKS:

1. Daniel Jurafsky, James H. Martin—Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, Pearson Publication, 2014.
2. Steven Bird, Ewan Klein and Edward Loper, —Natural Language Processing with Python, First Edition, OReilly Media, 2009.

REFERENCES:

1. Breck Baldwin, —Language Processing with Java and LingPipe Cookbook, Atlantic Publisher, 2015.
2. Richard M Reese, —Natural Language Processing with Java, OReilly Media, 2015.
3. Nitin Indurkhya and Fred J. Damerau, —Handbook of Natural Language Processing, Second Edition, Chapman and Hall/CRC Press, 2010.
4. Tanveer Siddiqui, U.S. Tiwary, —Natural Language Processing and Information Retrieval, Oxford University Press, 2008.

Course Code :	DATA VISUALIZATION LAB	L	T	P	C
		0	0	4	2

PRE-REQUISITE:
Computer Graphics, Image Processing and Data analytics.

COURSE OBJECTIVES:
Students will be able to:
1. Understand the importance of data visualization for business intelligence and decision making.
2. Know approaches to understand visual perception
3. Learn about categories of visualization and application areas
4. Familiarize with the data visualization tools
5. Gain knowledge of effective data visuals to solve workplace problems

COURSE OUTCOMES:
At the end of the course, Students will be able to:
CO1: Use Python, R and Tableau for data visualization **(Understanding, Analysing & Applying)**
CO2: Apply data visuals to convey trends in data over time using tableau (**Applying)**
CO3: Construct effective data visuals to solve workplace problems (**Creating)**
CO4: Explore and work with different plotting libraries (**Applying & Evaluating)**
CO5: Learn and create effective visualizations **(Understanding, remembering & Creating)**

MAPPING WITH PROGRAMME OUTCOMES:

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PS 01	PS 02	PS 03
CO1	**S**						**M**	**M**							
CO2			**M**		**L**			**M**							
CO3					**L**		**S**		**L**						
C04		**M**		**S**		**M**									
CO5					**S**		**S**			**M**					
S - Strong **M – Medium** **L – Low**															

LIST OF EXERCISES

1. Introduction to various Data Visualization tools
2. Basic Visualization in Python
3. Basic Visualization in R
4. Introduction to Tableau and Installation
5. Connecting to Data and preparing data for visualization in Tableau
6. Data Aggregation and Statistical functions in Tableau
7. Creating Scatter Plot maps
8. Data Visualizations in Tableau [Drawing charts, Drawing graphs, data mapping]
9. Basic Dashboards in Tableau
10. Storytelling in Tableau

REFERENCE BOOKS:

1. Data visualization with python: create an impact with meaningful data insights using interactive and engaging visuals, Mario Dobler, Tim Grobmann, Packt Publications, 2019
2. Practical Tableau: 100 Tips, Tutorials, and Strategies from a Tableau Zen Master, Ryan Sleeper, Oreilly Publications, 2018
3. Data Visualization with R: 111 Examples by Thomas Rahlf, Springer, 2020
4. E. Tufte, The Visual Display of Quantitative Information, Graphics Press. 2nd Edition, 2001
5. Alexandru C Telea, Data Visualization: Principles And Practice, 2nd Edition, 2014
6. Wang Kaining, Infographic & Data Visualizations, sew Edition. 2013
7. Andy Krik, Data Visualisation: A Handbook for Data Driven Design, 1st Edition, 2016

PREPARED BY:

Dr. R. POORVADEVI, Assistant Professor/CSE

SEMESTER-VI

Course Code:	WEB AND SOCIAL MEDIA ANALYTICS	L	T	P	C
<PCC-101>		0	0	3	2

Course Objectives:

Exposure to various web and social media analytic techniques.

Course Outcomes:

1. Knowledge on web analytics and its Tools.
2. Apply natural language processing concepts on text analytics.
3. Understand social network concepts
4. Applied Sentimental analysis on Social media data

UNIT – I Introduction to Web Analytics

Definition and History of Web Analytics-Overview in different mediums of Web analytics-Data collection methods in Web Analytics. Metrics used in Web analysis, Pyramid Model of Web Analytics

UNIT – II Web Mining

Overview, Web Content and Web Structure Mining, Search Engines, Search Engine Optimization, Web Usage Mining (Web Analytics), Applications-Web Analytics Maturity Model .

UNIT – III Tools and Techniques

Web analytics tools and techniques: Click stream analysis, A/B testing, online surveys, Use of Google Analytics; Web crawling and Indexing; Natural Language Processing Techniques for Micro-text Analysis

UNIT –IV Social Media Analytics

Introduction to Social Media Analytics (SMA): Social media landscape, Need for SMA; SMA in Small organizations; SMA in large organizations; Application of SMA in different areas. Network fundamentals and models: The social networks perspective - nodes, ties and Influencers, Social network and web data and methods. Graphs and Matrices- Basic Measures for individuals and networks. Information visualization.

Unit V APPLICATIONS

Google Analytics, Email analytics , Facebook analytics, Sentimental analysis on Social media data, Twitter data.

TEXT BOOK:

1. Web Analytics 2.0 The Art of Online Accountability and Science of Customer Centricity by Avinash Kaushik
2. Mathew Ganis, Avinash Koihrkar Social Media Analytics IBM Press 2015 / 1st Edition

REFERENCE BOOKS:

1. TakeshiMoriguchi, Web Analytics Consultant Official Textbook, 7th Edition, 2016
2. Marshall Sponder Social Media Analytics McGraw Hill 2014
3. Reza Zafarani Mohammad Ali Abbasi Huan Liu, Social Media Mining, Cambridge University Press, ISBN: 10: 1107018854

Prepared By
Dr.M.Saraswathi,Assistant Professor,CSE

SEMESTER-VII

Course Code:	MINING MASSIVE DATASET	L	T	P	C

PREREQUISITES:

Students should be familiar with Data mining, algorithms, basic probability theory and discrete math.

COURSE OBJECTIVES:

- This course will cover practical algorithms for solving key problems in mining of massive datasets.
- This course focuses on parallel algorithmic techniques that are used for large datasets.
- This course will cover stream processing algorithms for data streams that arrive constantly.
- page ranking algorithms for web search.
- online advertisement systems that are studied in detail.

COURSE OUTCOMES:

1. Handle massive data using Map Reduce.
2. Develop and implement algorithms for massive data sets and methodologies in the context of data mining.
3. Understand the algorithms for extracting models and information from large datasets
4. Develop recommendation systems.
5. Gain experience in matching various algorithms for particular classes of problems.

MAPPING WITH PROGRAMME OUTCOMES:

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PS0 1	PS0 2	PS0 3
CO1	M				S							M			
CO2				S				L							
CO3		M					L			M					
C04			S			M			L						
CO5				M							M				
S - Strong M – Medium L – Low															

UNIT - I

Data Mining-Introduction-Definition of Data Mining-Statistical Limits on Data Mining, MapReduce and the New Software Stack-Distributed File Systems, MapReduce, Algorithms Using MapReduce.

UNIT - II

Similarity Search: Finding Similar Items-Applications of Near-Neighbor Search, Shingling of Documents, Similarity-Preserving Summaries of Sets, Distance Measures. Streaming Data: Mining Data Streams-The Stream Data Model, Sampling Data in a Stream, Filtering Streams.

UNIT - III

Link Analysis-PageRank, Efficient Computation of PageRank, Link Spam. Frequent Itemsets - Handling Larger Datasets in Main Memory, Limited-Pass Algorithms, Counting Frequent Items in a Stream. Clustering-The CURE Algorithm, Clustering in Non-Euclidean Spaces, Clustering for Streams and Parallelism.

UNIT - IV

Advertising on the Web-Issues in On-Line Advertising, On-Line Algorithms, The Matching Problem, The Adwords Problem, Adwords Implementation. Recommendation Systems - A Model for Recommendation Systems, Content-Based Recommendations, Collaborative Filtering, Dimensionality Reduction, The NetFlix Challenge.

UNIT - V

Mining Social-Network Graphs-Social Networks as Graphs, Clustering of Social-Network Graphs, Partitioning of Graphs, Simrank, Counting Triangles.

TEXT BOOKS:

1. Jure Leskovec, Anand Rajaraman, Jeff Ullman, Mining of Massive Datasets, 3rd Edition.

REFERENCE BOOKS:

1. Jiawei Han & Micheline Kamber, Data Mining – Concepts and Techniques 3rd Edition Elsevier.
2. Margaret H Dunham, Data Mining Introductory and Advanced topics, PEA.
3. Ian H. Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques,Morgan Kaufmann.