BIJU PATNAIK UNIVERSITY OF TECHNOLOGY, ODISHA
ROURKELA



Curriculum and Syllabus

M.Tech. in Computer Science & Engineering
(Data Science)
for the Admission Batch 2024-25

M.Tech. in Computer Science & Engineering (Data Science)
(Admission Batch : 2024-25)

1st Semester

First Semester					
Theory					
Sl. No.	**Category**	**Course Code**	**Course Title**	**L-T-P**	**Credit**
1	PC	24PC1001	Data Science	3-0-0	3
2	PC	24PC1002	Advanced Data Structures and Algorithms	3-0-0	3
3	PC	24PC1003	Scientific Computing	3-0-0	3
4	PC	24PC1004	Optimization Techniques	3-0-0	3
5	PE	24PE1001 24PE1002 24PE1003 24PE1004	Pervasive Computing Soft Computing and Applications Stochastic Models Cloud Computing	3-0-0	3
Total Credit (Theory)					**15**
Practical					
1	PC	24PC1201	Advanced Data Structure & Algorithms Lab	0-0-6	2
2	PC	24PC1202	Python Programming Lab	0-0-6	2
3	PSI	24PS1201	Seminar and Technical Writing	0-0-3	2
Total Credit (Practical)					**6**
Total Semester Credit					**21**

1st Semester

24PC1001 DATA SCIENCE (3-0-0)

Objectives: The objective of this course to comprehend the idea of Linear Methods for Regression and Classification, Model Assessment and Selection, Additive Models, Support Vector Machines(SVM).

Outcomes: Upon successful completion of this course students will able to

1. Learn the basic idea of regression models and least squares, Multiple regression, Logistic regression.
2. Know Model Assessment and Selection and Apply Boot strap methods, Bayesian approach.
3. Learn Boosting methods-exponential loss, Numerical Optimization via gradient boosting, Cluster analysis.
4. Apply Support Vector Machines (SVM), SVM for regression.

MODULE – I

Linear Methods for Regression and Classification: Overview of supervised learning, Linear regression models and least squares, Multiple regression, Subset selection, Ridge regression, least angle regression and Lasso, Linear Discriminant Analysis, Logistic regression .

MODULE – II

Model Assessment and Selection**:** Bias, Variance and model complexity, Bias-variance trade off, Optimism of the training error rate, Estimate of In-sample prediction error, Effective number of parameters, Bayesian approach and BIC, Cross-validation, Boot strap methods, Confusion matrix and ROC curves.
Dimensionality reduction (PCA, Kernel PCAfeature Selection, Non-negative matrix factorization and collaborative filtering).

MODULE – III

Additive Models, Trees and Boosting: Generalized additive models**,** Regression and classification trees, Boosting methods-exponential loss and AdaBoost, Numerical Optimization via gradient boosting, Examples (Spam data, California housing , New Zealand fish, Demographic data), random forests.
Unsupervised Learning, Cluster analysis (k-means, Hierarchical clustering, spectral clustering).

MODULE – IV

Support Vector Machines(SVM) and K-nearest Neighbor: SVM for classification, Reproducing Kernels, SVM for regression.
K-nearest –Neighbour classifiers(Image Scene Classification), Gaussian mixtures and EM algorithm.

Books Recommended:

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman: The Elements of Statistical Learning-Data Mining, Inference and Prediction, 2nd Edition, Springer Verlag, 2009.
2. G. James, D. Witten, T. Hastie, R. Tibshirani: An Introduction to Statistical Learning with Applications in R, Springer, 2013.

Book for References:

1. C. M. Bishop: Pattern Recognition and Machine Learning, Springer,2006
2. L. Wasserman: All of Statistics.
3. T.M. Mitchell, Machine Learning, Mc. Graw Hill, 1997.

24PC1002 ADVANCED DATA STRUCTURES AND ALGORITHMS (3-0-0)

Objective:

- Understand and apply the Divide-and-Conquer paradigm to solve complex problems, such as Strassen‘s matrix multiplication.
- Analyze the performance of algorithms using probabilistic and randomized techniques.
- Understand the properties of Binary Search Trees (BST) and Red-Black Trees.
- To introduce the concept of parallel algorithms, focusing on how parallelism can be exploited to improve the efficiency of classical algorithms like matrix multiplication and merge sort.
- To prepare students for the challenges of designing and analyzing algorithms for complex computational problems, particularly in parallel and real-time environments.

MODULE – I

Divide-and-Conquer: Strassen‘s algorithm for matrix multiplication. Probabilistic Analysis and Randomized Algorithms: The hiring problem, Indicator random variables, Randomized algorithms, A randomized version of quicksort, Analysis of quicksort.

MODULE – II

Hash Tables: Direct-address tables, Hash tables, Hash functions, Open addressing. Binary Search Trees, Red-Black Trees: Properties of red-black trees, Rotations, Insertion, Deletion. Augmenting Data Structures: Dynamic order statistics, Interval trees, B-Trees.

MODULE – III

Dynamic Programming: Rod cutting, Optimal binary search trees. Amortized Analysis: Aggregate analysis, The accounting method, The potential method, Number-Theoretic Algorithms: Greatest common divisor, Modular arithmetic, The Chinese remainder theorem, The RSA public-key cryptosystem.

MODULE – IV

Parallel Algorithms: Parallel matrix multiplication, Parallel merge sort , Online Algorithms: Waiting for an elevator, Maintaining a search list, Online caching, Machine-Learning Algorithms: Clustering, Multiplicative-weights algorithms, Gradient descent.NP-Completeness: NP-completeness and reducibility, NP-completeness proofs.

Outcome:

1. Understand and apply the Divide-and-Conquer paradigm to solve complex problems, such as Strassen‘s matrix multiplication.
2. Analyze the performance of algorithms using probabilistic and randomized techniques.
3. Develop the ability to design and analyze randomized algorithms, including randomized quicksort,with probabilistic analysis techniques.
4. Understand and apply parallel algorithms, such as parallel matrix multiplication and parallel merge sort, to solve computational problems efficiently.
5. Analyze machine learning algorithms, Understand the theory of NP-completeness, reducibility, and NP-completeness proofs to identify and classify computationally intractable problems.

Books Recommended:

1. Introduction to Algorithms, Thomas H.Corman, Charles E.Leiserson, Ronald L.Rivest, Clifford Stein Fourth Edition, The MIT Press 2022.
2. Data Structures and Algorithm Analysis in C++, Mark Allen Weiss, Pearson, 4th Edition (2013)
3. Advanced Data Structures, Peter Brass, Cambridge University Press, 1st Edition (2008)

24PC1003 SCIENTIFIC COMPUTING (3-0-0)

Objective:

1. To understand the principles of Scientific Computing.
2. To be familiar with the various Scientific Computing Models and their applications.

MODULE – I

Introduction, Computer representation of numbers and roundoff error, Solving linear systems of equations.

MODULE – II

Finite difference methods, Solving nonlinear equations, Accuracy in solving linear systems

MODULE – III

Eigenvalues and eigenvectors, Fitting curves to data

MODULE – IV

Numerical integration, Initial value ODEs

Textbook:

Scientific Computing, For Scientists and Engineers, Timo Heister and Leo G. Rebholz
Published by De Gruyter 2023

Outcome:

Technical knowhow of the Data Mining principles and techniques for real time applications.

24PC1004 OPTIMIZATION TECHNIQUES (3-0-0)

Objective:

1. To understand the principles of Optimization Techniques.
2. To be familiar with the different optimization techniques and their Implementations.

MODULE – I

Mathematical preliminaries, Linear algebra and matrices, Vector space, eigen analysis, Elements of probability theory, Elementary multivariable calculus

MODULE – II

Linear Programming: Introduction to linear programming model, Simplex method, DualityKarmarkar's method.

MODULE – III

Unconstrained optimization: One-dimensional search methods, Gradient-based methods, Conjugate direction and quasi-Newton methods

MODULE – IV

Constrained Optimization Lagrange theorem, FONC, SONC, and SOSC conditions, Non-linear problems : Non-linear constrained optimization models, KKT conditions Projection methods

Outcome:

1. Technical knowhow of the optimization techniques for real time applications.

Books Recommended:

1. An introduction to Optimization by Edwin P K Chong, Stainslaw Zak
2. Nonlinear Programming by Dimitri Bertsekas

24PE1001 PERVASIVE COMPUTING (3-0-0)

Objective:

1. To understand the principles of Pervasive Computing.
2. To be familiar with the various Pervasive Computing Models and their applications.

MODULE – I

Pervasive Computing Application - Pervasive Computing Devices and Interfaces - Device technology trends, Connecting issues and protocols. Pervasive Computing and web based Applications - XML and its role in Pervasive Computing.

MODULE – II

Pervasive Computing and web based Applications (continued) - Wireless Application Protocol (WAP) Architecture and Security - Wireless Mark-Up language (WML) – Introduction. Voice Enabling Pervasive Computing - Voice Standards - Speech Applications in Pervasive Computing and security

MODULE – III

PDA in Pervasive Computing – Introduction - PDA software Components, Standards, emerging trends - PDA Device characteristics - PDA Based Access Architecture.

MODULE – IV

User Interface Issues in Pervasive Computing, Architecture - Smart Card- based Authentication Mechanisms - Wearable computing Architecture.

Books Recommended:

1. Jochen Burkhardt, Horst Henn, Stefan Hepper, Thomas Schaec & Klaus Rindtorff., Pervasive Computing Technology and Architecture of Mobile Internet Applications, Addision Wesley, Reading, 2002.
2. Uwe Ha nsman, Lothat Merk, Martin S Nicklous & Thomas Stober, Principles of Mobile Computing, Second Edition, Springer- Verlag, New Delhi, 2003. Reference Books
3. Rahul Banerjee: Internetworking Technologies: An Engineering Perspective, Prentice –Hall of India, New Delhi, 2003. (ISBN 81-203- 2185-5)
4. Rahul Banerjee: Lecture Notes in Pervasive Computing, Outline Notes, BITS-Pilani, 2003.

24PE1002 SOFT COMPUTING AND APPLICATIONS (3-0-0)

Objective:

• To expose the students to the concepts of feed forward neural networks.
• To provide adequate knowledge about feedback neural networks
• To provide adequate knowledge about fuzzy and neuro-fuuzy systems
• To provide comprehensive knowledge of fuzzy logic control to real time systems.
• To provide adequate knowledge of genetic algorithms and its application to economic dispatch and unit commitment problems.

MODULE – I

ARCHITECTURES – ANN:Introduction – Biological neuron – Artificial neuron – Neuron model – Supervised and unsupervised learning- Single layer – Multi layer feed forward network – Learning algorithm- Back propagation network.

MODULE – II

NEURAL NETWORKS FOR CONTROL: Feedback networks – Discrete time Hopfield networks – Transient response of continuous time system – Applications of artificial neural network - Process identification – Neuro controller for inverted pendulum.

MODULE – III

FUZZY SYSTEMS: Classical sets – Fuzzy sets – Fuzzy relations – Fuzzification – Defuzzification – Fuzzy rules - Membership function – Knowledge base – Decision-making logic – Introduction to neuro fuzzy system-Adaptive fuzzy system.

MODULE – IV

APPLICATION OF FUZZY LOGIC SYSTEMS: Fuzzy logic control: Home heating system - liquid level control - aircraft landing- inverted pendulum – fuzzy PID control, Fuzzy based motor control.

GENETIC ALGORITHMS: Introduction-Gradient Search – Non-gradient search – Genetic Algorithms: binary and real representation schemes, selection methods, crossover and mutation operators for binary and real coding - constraint handling methods – applications to economic dispatch and unit commitment problems.

Outcome:

• Ability to understand and apply basic science, circuit theory, Electro-magnetic field theory control theory and apply them to electrical engineering problems.
• To understand and apply computing platform and software for engineering problems.

Books Recommended:

1. Introduction to Algorithms, Thomas H.Corman, Charles E.Leiserson, Ronald L.Rivest, Second Edition, PHI 2003.
2. Data structures and Algorithm Analysis in C++, Mark Allen Weiss, PearsonEducation, 3rd Ed, 2007.
3. Online Computation and Competitive Analysis - A. Borodin and R. El-Yaniv,Cambridge Univ. Press, 1998.
4. Approximation Algorithms - Vijay V. Vazirani, Springer Verlag, 2003.
5. LauranceFausett, Englewood cliffs, N.J., 'Fundamentals of Neural Networks',Pearson Education, 1992.
6. Timothy J. Ross, 'Fuzzy Logic with Engineering Applications', Tata McGraw Hill, 1997.
7. S.N.Sivanandam and S.N.Deepa, Principles of Soft computing, Wiley India Edition, 2nd Edition, 2013.
8. Simon Haykin, 'Neural Networks', Pearson Education, 2003.
9. John Yen & Reza Langari, 'Fuzzy Logic – Intelligence Control & Information', Pearson Education, New Delhi, 2003.

10. M.Gen and R,Cheng, Genetic algorithms and Optimization, Wiley Series in Engineering Design and Automation, 2000.
11. Hagan, Demuth, Beale, " Neural Network Design", Cengage Learning, 2012.
12. N.P.Padhy, " Artificial Intelligence and Intelligent Systems", Oxford, 2013.
13. William S.Levine, "Control System Advanced Methods," The Control Handbook CRC Press, 2011.

24PE1003 STOCHASTIC MODELS (3-0-0)

Objective:

Course description Mathematical models based on probability theory prove to be extremely useful in describing and analyzing complex systems that exhibit random components. The goal of this course is to introduce several classes of stochastic processes, analyze their behavior over a finite or infinite time horizon, and help students enhance their problem solving skills.

The course combines classic topics such as martingales, Markov chains, renewal processes, and queuing systems with approaches based on Stein's method and on concentration inequalities.

The course focuses mostly on discrete-time models and explores a number of applications in operations research, finance, and engineering.

The course is directed towards graduate students who have a mathematically rigorous interest in stochastic processes and stochastic modelling.

MODULE – I

Preliminaries

- Expectation and integration
- Almost sure convergence and the dominated convergence theorem
- Convergence in probability and in distribution
- The law of large number and the ergodic theorem

Stein's method and central limit theorems

- Coupling
- Poisson approximation and Le Cam's theorem
- The Stein-Chen method
- Stein's method for the geometric and the normal distribution

MODULE – II

Conditional expectation and martingales

- Conditional expectation
- Martingales
- The martingale stopping theorem
- The Hoeffding-Azuma inequality
- The martingale convergence theorem
- Uniform integrability

Probability inequalities

- Jensen's inequality
- Probability bounds via the importance sampling identity
- Chernoff bounds
- Second moment and conditional expectation inequalities

MODULE – III

Discrete-time Markov chains

- Chapman-Kolmogorov equations and classification of states
- The strong Markov property
- Stationary and limiting distributions
- Transition among classes, the gambler's ruin problem, and mean times in transient states
- Branching processes
- Time reversibility

Renewal theory

- Limit theorems
- Renewal reward processes
- Blackwell's theorem

MODULE – IV

Queuing theory

- The Poisson process
- Queuing systems: M/M/1, M/G/1, G/M/1, and G/G/1

Outcome:

Students would acquire a rigorous understanding of basic concepts in probability theory. They would learn some important concepts concerning multiple random variables such as Bayes rule for random variables, conditional expectation and its uses etc. They would also learn stochastic processes, including Markov Chains and Poisson Processes. The course would provide the background needed to study topics such as Machine Learning, Adaptive Signal Processing, Estimation Theory etc.

Books Recommended:

1. Introduction to Algorithms, Thomas H.Corman, Charles E.Leiserson, Ronald L.Rivest, Second Edition, PHI 2003.
2. Data structures and Algorithm Analysis in C++, Mark Allen Weiss, PearsonEducation, 3rd Ed, 2007.
3. Online Computation and Competitive Analysis - A. Borodin and R. El-Yaniv,Cambridge Univ. Press, 1998.
4. Approximation Algorithms - Vijay V. Vazirani, Springer Verlag, 2003.
5. S.M. Ross and E.A. Peköz, A second course in probability, www.probabilitybookstore.com,Boston, MA, 2007.
6. S.M. Ross, Stochastic processes, John Wiley & Sons, Inc., New York, second edition, 1996
7. G.R. Grimmett and D.R. Stirzaker, Probability and random processes, Oxford University Press,New York, third edition, 2001, (same level).
8. G.F. Lawler, Introduction to stochastic processes, Chapman & Hall/CRC, Boca Raton, FL,second edition, 2006, (same level).
9. D. Williams, Probability with martingales, Cambridge University Press, Cambridge, 1991.

24PE1004 CLOUD COMPUTING (3-0-0)

Objective:

1. To understand the principles of Cloud Computing.
2. To be familiar with the Cloud Computing and their Implementation.

Module-I: Cloud Computing Overview

Origins of Cloud computing – Cloud components - Essential characteristics – On-demand selfservice, Broad network access, Location independent resource pooling ,Rapid elasticity , Measured service, Comparing cloud providers with traditional IT service providers, Roots of cloud computing.

Module-II: Cloud Insights

Architectural influences – High-performance computing, Utility and Enterprise grid computing, Cloud scenarios – Benefits: scalability ,simplicity ,vendors ,security, Limitations – Sensitive information - Application development- security level of third party - security benefits, Regularity issues: Government policies.

Module-III: Cloud Architecture- Layers and Models

Layers in cloud architecture, Software as a Service (SaaS), features of SaaS and benefits, Platform as a Service (PaaS), features of PaaS and benefits, Infrastructure as a Service (IaaS), features of IaaS and benefits, Service providers, challenges and risks in cloud adoption. Cloud deployment model: Public clouds – Private clouds – Community clouds - Hybrid clouds - Advantages of Cloud computing.

Module-IV: Cloud Simulators- CloudSim and GreenCloud

Introduction to Simulator, understanding CloudSim simulator, CloudSim Architecture(User code, CloudSim, GridSim, SimJava) Understanding Working platform for CloudSim, Introduction to GreenCloud. Introduction to VMWare Simulator Basics of VMWare, advantages of VMware virtualization, using Vmware workstation, creating virtual machines-understanding virtual machines, create a new virtual machine on local host, cloning virtual machines, virtualize a physical machine, starting and stopping a virtual machine.

Outcome:

1. Technical knowhow of the Cloud Computing techniques for real time applications.

Books Recommended:

1. Cloud computing a practical approach - Anthony T.Velte , Toby J. Velte Robert Elsenpeter, TATA McGraw- Hill , New Delhi – 2010
2. Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online - Michael Miller - Que 2008
3. Cloud computing for dummies- Judith Hurwitz , Robin Bloor , Marcia Kaufman ,Fern Halper, Wiley Publishing, Inc, 2010
4. Cloud Computing (Principles and Paradigms), Edited by Rajkumar Buyya, James Broberg, Andrzej Goscinski, John Wiley & Sons, Inc. 2011

24PC1201 ADVANCED DATA STRUCTURE & ALGORITHMS LAB

1. Understand the main features and importance of the Advance Java.
2. Apply working knowledge of Advance Java package to simulate and solve numerical problems and Applications.
3. Solve, Simulate and Analyse various Data base connectivity Techniques.
4. Solve, Simulate and Analyse various Evolutionary computation Techniques.
5. Solve, Simulate and Analyse various Internet and web programming techniques.

24PC1202 PYTHON PROGRAMMING LAB

1. Krishna and his five friends have decided to go for an industrial visit by sharing the expenses of the fuel equally. Write a Python program to calculate the amount (in Rs) each of them needs to put in for the complete journey. The program should also display True, if the amount to be paid by each person is divisible by 3, otherwise it should display False.
   Hint: Use the relational operators in print statement.
   Assumptions: Assume that mileage of the vehicle, amount per liter of fuel and distance for one way are given.

2. Write a program that prints a user's grade given a percent of points achieved in the class.
   The program should prompt the user to enter his/her percent of points. It should then print a letter grade A, A-, B+, B, B-, C+, C, C-, D+, D, D- and F respectively as shown in the table below.

3. The Twelve Days of Christmas is a repetitive song that describes an increasingly long list of gifts sent to one's true love on each of 12 days. A single gift is sent on the first day. A new gift is added to the collection on each additional day, and then the complete collection is sent.
   Write a program that displays the complete lyrics for The Twelve Days of Christmas.
   Write a function that takes the verse number as it's only parameter and displays the specified verse of the song. Then call that function 12 times with integers that increase from 1 to 12.

4. Write a function that generates a random password. The password should have a random length of between 7 and 10 characters. Each character should be randomly selected from positions 33 to 126 in the ASCII table. Your function will not take any parameters. It will return the randomly generated password as its only result. Also, write a function that determines whether or not a password is good. A good password is the one, that is at least 8 characters long and contains at least one uppercase letter, at least one lowercase letter, and at least one number. Your function should return Grade If Greater Than or Equal to
   A 93.33
   A- 90
   B+ 86.67
   B 83.33
   B- 80
   C+ 76.67
   C 73.33
   C- 70
   D+ 66.67
   D 63.33
   D- 60
   F 0
   true if the password passed to it as its only parameter is good. Otherwise it should return false.
   Hint: Use chr( ) function.

5. A teacher is conducting a camp for a group of five children. Based on their performance and behavior during the camp, the teacher rewards them with chocolates.

   Write a Python function to
   1. Find the total number of chocolates received by all the children put together. Assume that each child is identified by an id and it is stored in a tuple and the number of chocolates given to each child is stored in a list.
   2. The teacher also rewards a child with few extra chocolates for his/her best conduct during the camp.
      - If the number of extra chocolates is less than 1, an error message "Extra chocolates is less than 1", should be displayed.
      - If the given child Id is invalid, an error message "Child id is invalid" should be displayed. Otherwise, the extra chocolates provided for the child must be added to his/her existing number of chocolates and display the list containing the total number of chocolates received by each child.

6. A vendor at a food court is in the process of automating his order management system.
   The vendor serves the following menu – Veg Roll, Noodles, Fried Rice and Soup and also maintains the quantity available for each item. The customer can order any combination of items. The customer is provided the item if the requested quantity of item is available with the vendor.
   Write a python program which implements the following functions.
   place_order (*item_tuple): This function accepts the order placed by the customer.
   Consider it to be a variable length argument as each customer may have a
   different order. The function should check whether the items requested are present in
   the vendor's menu and if so, it should check whether the requested quantity is
   available for each by invoking the check_quantity_available() method. The function
   should display appropriate messages for each item in the order for the below scenarios:
   1. When the requested item is not available in vendor's menu, display <ItemName> is not available.
   2. When the quantity requested by the customer is not available, display <ItemName> stock is over.
   3. When the requested quantity of the item is available with the vendor, display <ItemName> is available.

7. A spell checker can be a helpful tool for people who struggle to spell words correctly.
   In this exercise, you will write a program that reads a file and displays all of the words in it that are misspelled. Misspelled words will be identified by checking each word in the file against a list of known words. Any words in the user's file that do not appear in the list of known words will be reported as spelling mistakes. The user will provide the name of the file to check for spelling mistakes as a command line parameter. Your program should display an appropriate error message if the command line parameter is missing. An error message should also be displayed if your program is unable to open the user's file. Words followed by a comma, period or other punctuation mark are not reported as spelling mistakes. Ignore the capitalization of the words when checking their spelling.

8. Write a python program to handle the following cases using re module functionalities:
   1. Retrieve all the lines that contain "This" in the beginning of the line.
   2. Repeat Q1 but retrieve both upper- and lower-case letters.
   3. Retrieve all lines that contain consecutive te's.
   4. Retrieve lines that contain word of any length starting with s &ending with e.
   5. Retrieve all lines with a date in the form of 1 or 2 digits, a dot, 1 or 2 digits, a dot, two digits.

9. The novel "Gadsby" is over 50,000 words in length. While 50,000 words aren't normally remarkable for a novel, it is in this case because none of the words in the book use the letter "e". This is particularly noteworthy when one considers that "e" is the most common letter in English. Write a program that reads a list of words from a file and determines what proportion of the words use each letter of the alphabet. Display the result for all 26 letters. Include an additional message identifying the letter that is used in the smallest proportion of the words.

NOTE: Your program should ignore any punctuation marks and it should treat uppercase and lowercase letters as equivalent.

10. Write a python program that takes a number N from command line and creates an N x N multiplication table in an Excel Spreadsheet. Also, extend the program to accept two integers N and M indicating the rows and columns along with file name as a string argument in the command line. Finally display the result in the excel sheet in the inverted form. (Inverse the row and columns cells).

M.Tech. in Computer Science & Engineering (Data Science)
(Admission Batch : 2024-25)

2nd Semester

Second Semester					
Theory					
Sl. No.	**Category**	**Course Code**	**Course Title**	**L-T-P**	**Credit**
1	PC	24PC1005	Machine Learning	3-0-0	3
2	PC	24PC1006	Research Methodology, Ethics and IPR	3-0-0	3
3	PC	24PC1007	Big Data Analytics	3-0-0	3
4	PC	24PC1008	Advanced Data Visualization	3-0-0	3
5	PE	24PE1005 24PE1006 24PE1007 24PE1008	Natural Language Processing Image Analysis Business Analytics Data Privacy and Security	3-0-0	3
Total Credit (Theory)					**15**
Practical					
1	PC	24PC1203	Data Visualization Lab	0-0-6	2
2	PC	24PC1204	Machine Learning Lab	0-0-6	2
3	PSI	24PS1201	Mini Project with Seminar	0-0-6	2
Total Credit (Practical)					**6**
Total Semester Credit					**21**

2nd Semester

MACHINE LEARNING (3-0-0)

Objective:

1. To understand the principles of Machine Learning.
2. To be familiar with the various Machine Learning Models and their applications.

MODULE – I

Introduction toML: Motivation and role of machine learning in computer science and problem solving. Representation (features), line art transformations, Appreciate linear transformations and matrix vector operations in the context of data and representation. Problem formulations (classification and regression). Appreciate the probability distributions in the context ofdata, Prior probabilities and Bayes Rule. Introduce paradigms of Learning (primarily supervised and unsupervised. Also a brief overview of others)

MODULE – II

Fundamentals of ML: PCA and Dimensionality Reduction, Nearest Neighbours and KNN. Linear Regression, Decision Tree Classifiers, Notion of Generalization and concern of Over fitting, Notion of Training, Validation and Testing; Connect to generalization and over fitting.

MODULE – III

Ensembling and RF, Linear SVM,K Means, Logistic Regression, Naive Bayes, Role of Loss Functions and Optimization, Gradient Descent and Perceptron/ Delta Learning, MLP, Back propagation MLP for Classification and Regression, Regularisation, Early Stopping, Introduction to Deep Learning, CNNs

MODULE – IV

Kernels (with SVM), Bayesian Methods, Generative Methods, HMM, EM, PAC learning, Popular CNN, Architectures, RNNs, GANS and Generative Models, Advances in Back propagation and Optimization for Neural Networks, Adversarial Learning

Outcome:

1. Technical knowhow of the Machine Learning techniques for real time applications.

Books Recommended:

1. Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong, Mathematics for Machine Learning, Cambridge University Press (23 April 2020)
1. Tom M. Mitchell- Machine Learning - McGraw Hill Education, International Edition
2. AurélienGéron Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'ReillyMedia, Inc. 2nd Edition
3. Ian Goodfellow, YoshouaBengio, and Aaron Courville Deep Learning MIT Press Ltd,Illustrated edition
4. Christopher M. Bishop Pattern Recognition and Machine Learning - Springer, 2nd edition
   Trevor Hastie, Robert Tibshirani, and Jerome Friedman - The Elements ofStatistical Learning: Data Mining, Inference, and Prediction - Springer, 2nd edition

RESEARCH METHODOLOGY, ETHICS AND IPR (3-0-0)

Objective:

1. To understand the principles of Research Methodology.
2. To be familiar with the Research Methodology and their Implementation.
3. To understand the principles of Intellectual Property Rights.
4. To be familiar with the Intellectual Property Rightsand their Implementation.

MODULE – I

Introduction to RM: Meaning and significance of research. Importance of scientific research in decision making. Types of research and research process. Identification of research problem and formulation of hypothesis. Research Designs. Measurement and Data Collection. Primary data, Secondary data, Design of questionnaire; Sampling fundamentals and sample designs. Measurement and Scaling Techniques, Data Processing.

MODULE – II

Data Analysis – I: Hypothesis testing; Z-test, t-test, F-test, Chi-square test. Analysis of variance. Non-parametric Test – Sign Test, Run test, Krushall – Wallis test. Data Analysis – II: Factor analysis, Multiple Regressions Analysis. Discriminant Analysis, Use of SPS Package.

MODULE – III

Intellectual property: meaning, nature and significance, need for intellectual property Right (IPR), IPR in India – Genesis and development, IPR in abroad. What is a patent, What can be protected by a patent, Why should I apply for a patent? Patent Law, Patentability requirements, Non-Patentable subject matters, Layout of the Patents. Procedure for domestic and international filing of applications, Restoration, Surrender and Revocations of Patents, Rights of Patentee and Working of Patent, Licensing and Enforcing Intellectual Property.

MODULE – IV

Copyright: meaning, scope; What is covered by copyright? How long does copyright last? Why protect copyright? Related rights, Rights covered by copyright. Ownership: Duration, Division, Transfer and Termination of Transfers. Literal and non-literal infringement, Role of claims, Doctrines on infringement: Equivalent doctrine, Pith and Marrow doctrine, Comparative test. Defenses: Gillette Defense, General grounds, Patents granted with conditions, Parallel import. Remedies: Civil, Administrative. STATE LAW: TRADE SECRET, CONTRACT, MISAPPROPRIATION, RIGHT OF PUBLICITY.Trademarks, Trade Secret - Overview, Requirements, Misappropriation of Trade Secret, Departing Employees, Remedies, Criminal Liability, Misappropriation, Clickwrap Agreements, Idea Submissions; Right of Publicity, Federal Preemption, Review.

Outcome:

1. Technical knowhow of the Research Methodology for real time applications.
2. Technical knowhow of the Intellectual Property Rights for real time applications.

Books Recommended:

1. Research Methodology, Chawla and Sondhi, Vikas
2. Research Methodology, Paneersevam, PHI
3. W. R. Cornish and D. Llewellyn, Intellectual Property: Patents, Copyrights, Trade Marks and Allied Rights, Sweet & Maxwell.
4. Lionel Bently and Brad Sherman, Intellectual Property Law, Oxford University Press.
5. P. Narayanan, Intellectual Property Law, Eastern Law House
6. B. L. Wadehra, Law Relating to Intellectual Property, Universal Law Publishing Co.
7. V. K. Ahuja, Law Relating to Intellectual Property Rights, LexisNexis.
8. AjitParulekar and Sarita D'Souza, Indian Patents Law – Legal & Business Implications;Macmillan India ltd, 2006

9. P. Narayanan; Law of Copyright and Industrial Designs; Eastern law House, Delhi, 2010.
10. Academic Syllabus for M. Tech in Computer Science, 2016 – 2017

BIG DATA ANALYSIS (3-0-0)

Objective:

- To familiarize students with big data analysis as a tool for analysing large complex dataset.
- To learn to use various techniques for mining data stream.
- Understand the applications using Map Reduce Concepts
- Provide hands on Hodoop Eco System
- To introduce programming tools PIG & HIVE in Hadoop echo system

MODULE – I

Introduction To Big Data, Data Storage and Analysis - Characteristics of Big Data – Big Data Analytics - Typical Analytical Architecture – Requirement for new analytical architecture – Challenges in Big Data Analytics – Need of big data frameworks

MODULE – II

NoSQL Database: NoSQL Databases - Schema less Models, Increasing Flexibility for Data Manipulation-Key Value Stores, Document Stores, Tabular Stores, Object Data Stores – GraphDatabases, Big data for twitter, Big data for E-Commerce blogs.

MODULE – III

Big Data: Evolution of Big data, Best Practices for Big data Analytics - Big data characteristics -
Big Data Use Cases, Characteristics of Big Data Applications, Big Data Modelling, HDFS performance and tuning, Map reduce algorithm, Hadoop Eco system Pig : Introduction to PIG,Execution Modes of Pig, Grunt, Pig Latin, User Defined Functions, Data Processing operators.

MODULE – IV

Hive: Hive Shell, Hive Services, HiveQL, Tables, Querying Data and User Defined Functions. Hbase :HBasics, Concepts, Clients, Example, Spark Mining Data Streams: Introduction to Streams Concepts, Stream Data Model and Architecture -Sampling Data in a Stream, Filtering Streams, Counting Distinct Elements in a Stream –Real time Analytics Platform (RTAP) applications, Case Studies, Real Time Sentiment Analysis- Stock Market Predictions.

Outcomes:

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

- Process data in Big Data platform and explore the big data analytics techniques for business applications
- Analyse Map Reduce technologies in big data analytics
- Develop Big Data solutions using Hadoop Eco System
- Design efficient algorithms for stream data mining on big data platform

Books Recommended:

1. Jure Leskovec, AnandRajaraman and Jeffrey David Ullman, "Mining of Massive Datasets",Cambridge University Press, 2014.
2. Tom White ,Hadoop: The Definitive Guide, 4th edition O'Reily Publications, 2015
3. Judith Hurwitz, Alan Nugent, Dr. Fern Halper, and Marcia Kaufman, "Big data for dummies" A wiley brand publications.
4. Holden Harau, "Learning Spark: Lightning-Fast Big Data Analysis", O-Reilly Publications
5. David Loshin, "Big Data Analytics: From Strategic Planning to Enterprise Integration withTools, Techniques, NoSQL, and Graph", 2013.

6. BartBaesens, "Analytics in a Big Data World: The Essential Guide to Data Science and itsApplications", Wiley Publishers, 2015.
7. Kim H. Pries and Robert Dunnigan, "Big Data Analytics: A Practical Guide for Managers "CRC Press, 2015.
8. EMC Education Services, "Data Science and Big Data Analytics: Discovering, Analyzing,Visualizing and Presenting Data", Wiley publishers, 2015.
9. DietmarJannach, Markus Zanker, Alexander Felfernig and Gerhard Friedrich"Recommender Systems: An Introduction", Cambridge University Press, 2010.
10. Jimmy Lin, Chris Dyer and Graeme Hirst, "Data-Intensive Text Processing withMapReduce", Synthesis Lectures on Human Language Technologies, Vol. 3, No. 1, Pages1-177, Morgan Claypool publishers, 2010.

ADVANCED DATA VISUALIZATION (3-0-0)

Objective:

- To extend student's knowledge in the area of Data Science with emphasis on Predictions utilizing associated statistical methods and software tools.

MODULE – I

Introduction to Data Acquisition – Applications –Process- Data Extraction- Data Cleaning and Annotation- Data Integration -Data Reduction- Data Transformation –Visualization-Introduction -Terminology- Basic Charts and Plots- Multivariate Data Visualization- Data Visualization Techniques– Pixel-Oriented Visualization Techniques- Geometric Projection Visualization
Techniques- Icon-Based Visualization Techniques- Hierarchical Visualization.

MODULE – II

Techniques Visualizing Complex Data and Relations Data Visualization Tools– Rank Analysis Tools- Trend Analysis Tools- Multivariate Analysis Tools- Distribution Analysis Tools- Correlation Analysis Tools- Geographical Analysis Tools.

MODULE – III

Regression model building framework: Problem definition, Data pre-processing; Model building; Diagnostics and validation Simple Linear Regression: Coefficient of determination, Significance tests, Residual analysis, Confidence and Prediction intervals.

MODULE – IV

Multiple Linear Regression: Coefficient of multiple coefficient of determination, Interpretation of regression coefficients, Categorical variables, Heteroscedasticity, Multi-co linearity, outliers, Auto regression and transformation of variables, Regression model building.

Outcome:

- Ability to apply specific statistical and regression analysis methods applicable to predictive analytics to identify new trends and patterns, uncover relationships, create forecasts, predict likelihoods, and test predictive hypotheses.
- Ability to develop and use various quantitative and classification predictive models based on various regression and decision tree methods.

Books Recommended:

1. Andy Kirk, Data Visualization A Handbook for Data Driven Design, Sage Publications, 2016
2. Philipp K. Janert, Gnuplot in Action, Understanding Data with Graphs, Manning Publications, 2010.
3. Alberto Cordoba, "Understanding the Predictive Analytics Lifecycle", Wiley, 2014.
4. Eric Siegel, Thomas H. Davenport, "Predictive Analytics: The Power to Predict Who Will Click,Buy, Lie, or Die", Wiley, 2013.
5. James R Evans, "Business Analytics – Methods, Models and Decisions", Pearson 2013.
6. R. N. Prasad, SeemaAcharya, "Fundamentals of Business Analytics", Wiley, 2015.

NATURAL LANGUAGE PROCESSING (3-0-0)

Objective:

1. To understand the principles of Natural Language Processing.
2. To be familiar with the Natural Language Processing algorithms and their Implementation.

MODULE – I

NLP tasks in syntax, semantics and pragmatics; Applications such as information extraction, question answering, and machine translation, The problem of ambiguity, The role of machine learning.

MODULE – II

Brief history of the field POS-tagging, POS-tagging perspective, POS tagging and HMM, Hidden Markov models (Forward and Viterbi algorithm and EM training), POS-tag set, Machine translation, Parsing algorithms, Probabilistic parsing, Parser Comparison Grammar, constituency and dependency, CYK algorithm, Parse tree construction, Semantics.

MODULE – III

Word sense disambiguation Knowledge based and supervised WSD, Unsupervised EM based WSD, Multilingual Resource constrained WSD Linear and logistic Regression, Machine translation.

MODULE – IV

Statistical Machine translation, Binding Theory and Merger, X-bar theory.

Outcome:

1. Technical knowhow of the Natural Language Processing for real time applications.

Books Recommended:

1. James Allen, “Natural Language Understanding” .
2. Benjamin/Cummins E. Charniack, “Statistical Language Learning”, MIT Press .
3. Daniel Jurafsky and J.H. Martin, “Speech and Language Processing”, Prentice Hall.
4. H. Lane, H. Hapke, C. Howard, “Natural language processing in Action: Understanding, analyzing, and generating text with Python”, Manning publications.
5. B. Bengfort, R. Bilbro, “Applied Text Analysis with Python: Enabling Language Aware
6. Data Products with Machine Learning”, O”Reilly

IMAGE ANALYSIS (3-0-0)

Objective:

1. To understand the principles of Image Analysis.
2. To be familiar with the Image Analysis algorithms and their Implementation.

MODULE – I

Light and Electromagnetic spectrum, Components of Image processing system, Image formation and digitization concepts, Neighbours of pixel adjacency connectivity, regions and boundaries, Distance measures, Applications.

MODULE – II

Image Enhancements:

In spatial domain: Basic gray level transformations, Histogram processing, Using arithmetic/Logic operations, smoothing spatial filters, Sharpening spatial filters. In Frequency domain: Introduction to the Fourier transform and frequency domain concepts, smoothing frequency-domain filters, Sharpening frequency domain filters.

MODULE – III

Image Restoration:

Various noise models, image restoration using spatial domain filtering, image restoration using frequency domain filtering, Estimating the degradation function, Inverse filtering.

MODULE – IV

Colour Image processing:

Colour fundamentals, Colour models, Colour transformation, Smoothing and Sharpening, Colour segmentation. Wavelet and Multi-resolution processing: Image pyramids, Multi-resolution expansion, wavelet transform. Image compression: Introduction, Image compression model, Error-free compression, Lossy compression.
Image segmentation: Detection of discontinuities, Edge linking and boundary detection, thresholding.

Outcome:

1. Technical knowhow of the Image Analysis techniques for real time applications.

Books Recommended:

1. Digital Image Processing, Second Edition by Rafel C. Gonzalez and Richard E. Woods, PearsonEducation
2. Digital Image Processing by Bhabatosh Chanda and Dwijesh Majumder, PHI
3. Fundamentals of Digital Image Processing by Anil K Jain, PHI
4. Digital Image Processing Using Matlab, Rafel C. Gonzalez and Richard E. Woods, PearsonEducation

Course Outcome:

After learning the course the students should be able to:

1. Understand the basic image enhancement techniques in spatial & frequency domains
2. Understand the various kind of noise present in the image and how to restore the noisy image.
3. Understand the basic multi-resolution techniques and segmentation methods.
4. To apply this concepts for image handling in various fields.

BUSINESS ANALYTICS (3-0-0)

Objective:

1. To understand the principles of Business Analytics.
2. To be familiar with the Business Analyticsalgorithms and their Implementation.

MODULE – I

Describing and Summarizing Data:

• Recognize trends in data and detect outliers
• Summarize data sets concisely
• Analyze relationships between variables

MODULE – II

Sampling and Estimation:

• Create representative samples and draw conclusions about the larger population
• Craft sound survey questions

MODULE – III

Hypothesis Testing:

• Quantify the evidence in favor of or against your hypothesis in order to make managerial decisions

Single Variable Linear Regression:

• Analyze the relationship between two variables and develop forecasts for values outside the data set

MODULE – IV

Multiple Regression

• Identify relationships among three or more variables to improve understanding of data and provide better forecasts

Outcome:

1. Technical knowhow of the Business Analytics for real time applications.

Books Recommended:

1. Microsoft Excel Data Analysis and Business Modeling by Wayne L. Winston
2. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython by Wes McKinney
3. Business Analytics by Jay Liebowitz, Publisher(s): Auerbach Publications, O'Reilly.

DATA PRIVACY AND SECURITY (3-0-0)

Objective:

1. To understand the principles of Business Analytics.
2. To be familiar with the Business Analytics algorithms and their Implementation.

MODULE – I

Passwords, security questions, challenge-response, Cryptographic hash functions, Biometrics, Phishing.

MODULE – II

Web security model, Web authentication and session management, Cross-site request forgery, SQL injection, cross-site scripting, Logic flaws in Web applications, Click jacking.

MODULE – III

Online tracking, Symmetric encryption, Kerberos, Memory corruption attacks and defenses, Viruses and rootkits.

MODULE – IV

Spam, Attacks on TCP/IP, DNS, BGP. Denial of service, Worms and botnets, Advance Persistent, Threats, Firewall and intrusion detection, Public Key Cryptography, SSL and certificates, Anonymity networks, Side channel attacks: acoustics and reflections.

Outcome:

1. Technical knowhow of the Data Privacy and Security for real time applications.

Books Recommended:

1. Network Security (2nd edition) by Kaufman, Perlman, and Speciner -- required textbook! Security Engineering by Anderson
2. The Art of Intrusion by Mitnick and Simon
3. The Shellcoder's Handbook by Koziol et al.
4. Network Security Essentials by Stallings

DATA VISUALIZATION LAB (0-0-6)

1. Defining data visualization; Visualization workflow: describing data visualization workflow, process in practice;
2. Data Representation: chart types: categorical, hierarchical, relational, temporal & spatial;
3. 2-D: bar charts, Clustered bar charts, dot plots, connected dot plots, pictograms, proportional shape charts, bubble charts, radar charts, polar charts, Range chart, Box-and-whiskerplots, univariate scatter plots, histograms word cloud, pie chart, waffle chart, stacked bar chart,
4. Back-to-back bar chart, treemap and all relevant 2-D charts. 3-D: surfaces, contours,
5. Hidden surfaces, pm3d coloring,
6. 3D mapping; multi-dimensional data visualization; manifold visualization;
7. Graph data visualization;
8. Annotation.

MACHINE LEARNING LAB (0-0-6)

1. Experiments based on validation of models, regression and classification.
2. Experiments on Gradient descent and stochastic gradient descent.
3. Case studies using SVN and Multiclass SVM
4. Experiments on Back propagation
5. Implementations of Dimension reduction, EM algorithm and HMM
6. Implementation of MAP, PCA, LDA, Kernel methods.
7. Apply of machine learning methods to real word applications (Case studies)
8. Experiments on building federated learning models

MINI PROJECT WITH SEMINAR (0-0-6)

M.Tech. in Computer Science & Engineering (Data Science)
(Admission Batch : 2024-25)

3rd Semester

Third Semester					
Theory					
Sl. No.	**Category**	**Course Code**	**Course Title**	**L-T-P**	**Credit**
1	PC	24PC2001	Deep Learning	3-0-0	3
2	PC	24PC2002	Data Fusion	3-0-0	3
Total Credit (Theory)					**06**
Practical					
1	PC	24PC2201	Pre Dissertation		12
2	PSI	24PS2201	Comprehensive Viva-Voce		2
Total Credit (Practical)					**14**
Total Semester Credit					**20**

3rd Semester

DEEP LEARNING (3-0-0)

Objective:

1. To understand the principles of Deep Learning.
2. To be familiar with the Deep Learning algorithms and their Implementation.

MODULE – I

Introduction to Deep Learning, Bayesian Learning, Decision Surfaces, Linear Classifiers, Linear Machines with Hinge Loss, Optimization Techniques, Gradient Descent, Batch Optimization

MODULE – II

Introduction to Neural Network, Multilayer Perception, Back Propagation Learning, Unsupervised Learning with Deep Network, Auto encoders, Convolution Neural Network, Building blocks of CNN, Transfer Learning

MODULE – III

Revisiting Gradient Descent, Momentum Optimizer, RMSProp, Adam, Effective training in Deep Net- early stopping, Dropout, Batch Normalization, Instance Normalization, Group Normalization, Recent Trends in Deep Learning Architectures, Residual Network, Skip Connection Network, Fully Connected CNN etc.

MODULE – IV

Classical Supervised Tasks with Deep Learning, Image Denoising, Semanticd Segmentation, Object Detection etc., LSTM Networks, Generative Modelling with DL, Variation Auto encoder, Generative Adversarial Network Revisiting Gradient Descent, Momentum

Outcome:

1. Technical knowhow of the Deep Learning for real time applications.

Books Recommended:

1. Deep Learning- Ian Goodfelllow, Yoshua Benjio, Aaron Courville, The MIT Press
2. Pattern Classification- Richard O. Duda, Peter E. Hart, David G. Stork, John Wiley & Sons Inc.

DATA FUSION (3-0-0)

Objective:

1. To understand the principles of Data Fusion.
2. To be familiar with the Data Fusion algorithms and their Implementation.

MODULE – I

Fusion Models, Sensors and Intelligence, Approaches to handle uncertainty, Neuro-Probabilist Approach, Neo-Calculist Approach, Neo-Logistic Approach, Neo-Possibilist Approach.

MODULE – II

Target Tracking, Single Sensor Single Target Tracking, Multi Sensor Single Target Tracking, Multi Sensor Multi Target Tracking, Interacting Multiple Models.

MODULE – III

Target Classification, Target Aggregation, Model based Situation Assessment – Bayesian Belief Network. Model based Situation Assessment.

MODULE – IV

Handling Non Linear and Hybrid Models, Decision Support, Fusion Models, Cognitive Agents for Data Fusion, Distributed Fusion

Outcome:

1. Technical knowhow of the Data Fusion for real time applications.

Recommended Books:

1. High Level Data Fusion, Subrata Das, Barnes & Noble

M.Tech. in Computer Science & Engineering (Data Science)
(Admission Batch : 2024-25)

4th Semester

Fourth Semester					
Practical					
1	PC	24PC2202	Dissertation		18
Total Credit (Practical)					**18**
Total Semester Credit					**18**