Syllabus

First Semester Courses in M.Sc. Big Data Analytics 2023-2024

Contents:

- **Syllabus for Core Courses:**
  - PSBDA6001CR1: Linear Algebra and Linear Programming
  - PSBDA6002CR1: Probability and Stochastic Processes
  - PSBDA6003CR1: Data Base Management -Relational
  - PSBDA6004CR1: Basics of Programming
- Evaluation and Assessment guidelines

APPROVED SYLLABUS

M.Sc-I(Big Data Analytics)

Course Code: PSBDA6001CR1

Title: Linear Algebra and Linear Programming

Credits :4 (Theory -3 -Total 45 hr and Practical 1-Total -30hr)

Course Objectives:

1. Understanding the fundamental concepts: The course aims to develop a deep understanding of the basic concepts of linear algebra, such as vectors, matrices, systems of linear equations, vector spaces, linear transformations, and eigenvectors/eigenvalues.
2. Performing computations: Students learn how to perform various computations and manipulations involving vectors and matrices, vector spaces operations, solving systems of linear equations, finding inverses, and calculating determinants.
3. Developing geometric intuition: Linear algebra often involves geometric interpretations, and the course objectives may include developing geometric intuition for vector spaces, subspaces, linear transformations, and their relationships.

Course outcomes:

On completing the course, the student will be able to:

1. Use the basic tools of matrices to understand their linkage to the real-world problems.
2. Recognize consistent and inconsistent systems of linear equations by the row echelon form of the augmented matrix, using rank
3. Find eigenvalues and corresponding eigenvectors for a square matrix.

Unit 1 15

Linear Algebra:

Linear equations and matrices, matrix operations, solving system of linear equations, Gauss-Jordan method, Concept & Computation of determinant and inverse of matrix.

Unit 2 15

Eigen Values and Eigen Vectors:

Eigen values and eigen vectors, Illustrations of the methods, Positive semi definite and position definite matrices, illustrations.

Unit 3 15

Linear Programming:

Definition of the problem, convex sets, corner points, feasibility, basic feasible solutions, Simplex method

List of Recommended Reference Books

1. Strang, Gilbert. Linear Algebra and Its Applications. 2011., Academic Press
2. Vinod, Hrishikesh D. Hands-on Matrix Algebra Using R. World Scientific Publishing Company, 28 Mar. 2011.
3. Hadley, G. (1962). Linear Programming. India: Addison-Wesley Publishing Company.

ASSESSMENT:
THEORY:
CIA I: Written test **for 20 marks**
CIA II: Assignments / Project / Presentation / Case Study/ Written Test **for 20 marks**
End semester Exam :60 marks
Practical Exam:50 marks
Template for the Core course End Semester examination in Semester I

UNITS	KNOWLEDGE	UNDERSTANDING	APPLICATION and ANALYSES	TOTAL MARKS- Per unit
1	5	5	10	20
2	5	10	5	20
3	5	5	10	20
-TOTAL - Per objective	15	20	25	60
% WEIGHTAGE	25%	33%	42%	100%

M.Sc-I(Big Data Analytics)

Course Code: PSBDA6002CR1

Title: Probability and Stochastic Processes
Credits :4 (Theory -3 -Total 45 hr and Practical 1-Total -30hr)

Course Objective:
To build a solid understanding of fundamental probability concepts, including random variables, probability distributions, and events.

Course Outcomes:
On completing the course, the student will be able to:

1. Acquire an understanding of a random experiment
2. Acquire the knowledge to conceptualize the probabilities of events.
3. Apply the notion of conditional probability including the concept of Bayes' Theorem.
4. Understand the concept of Markov Chains and their relevance in modeling various processes.
5. Solve real-world problems using these discrete and continuous probability distributions

Unit 1 15

Probability:

Concepts of experiments, Outcomes, Sample space, Events, Combinatorial probability, Birthday paradox, Principle of inclusion & exclusion, Conditional probability, Independence, Bayes Theorem.

Unit 2 15

Random Variables and Standard Discrete Distribution

Random Variables: discrete and continuous, introduction of bivariate distribution, some standard discrete probability distributions: Binomial ,Negative Binomial distribution, Poisson, Geometric, Hypergeometric. expectation, variance and other properties of the distribution.

Unit 3 15

Standard Continuous Distribution and Markov Chains

Uniform, Normal, exponential, expectation, variance and other properties of the distribution.

Markov Chains, Classification of states, Stationery distribution, limit theorems, Poisson process, illustrations and applications.

List of Recommended Reference Books

1. Ross, S. M. (2010). A first course in probability. United Kingdom: Pearson Prentice Hall.
2. Introduction to Stochastic Processes. (1986). (n.p.): Waveland Press.
3. Shumway, R. H., Stoffer, D. S. (2017). Time Series Analysis and Its Applications: With R Examples. Switzerland: Springer International Publishing.

ASSESSMENT:
THEORY:
CIA I: Written test **for 20 marks**
CIA II: Assignments / Project / Presentation / Case Study/ Written Test **for 20 marks**
End semester Exam :60 marks
Practical Exam:50 marks
Template for the Core course End Semester examination in Semester I

UNITS	KNOWLEDGE	UNDERSTANDING	APPLICATION and ANALYSES	TOTAL MARKS- Per unit
1	5	5	10	20
2	5	10	5	20
3	5	5	10	20
-TOTAL - Per objective	15	20	25	60
% WEIGHTAGE	25%	33%	42%	100%

M.Sc-I(Big Data Analytics)

Course Code: PSBDA6003CR1

Title: Data Base Management -Relational
Credits :4 (Theory -3 -Total 45 hr and practical 1-Total -30hr)

Course Objective:
1. Understand the basic concepts of database management systems.
2. Acquire the skill of using a structured query language.

Course outcome:
On completing the course, the student will be able to:
1. Acquire a solid understanding of fundamental database concepts, including data models, schemas, and relationships.
2. Acquire proficiency in working with relational databases, including creating tables, defining schemas, and enforcing data integrity
3. Understand the fundamental concepts and principles of aggregate Functions and Subqueries
4. Develop the skill of data base programming language

Unit 1	**15**
Basic Concepts: Introduction to databases, Different data models, ER and EER diagram, schema, table, DML, DDL	
Unit 2	15
Constraints and types of constraints, Joins, Databases: Structure, various operations, Normalization, views.	
Unit 3	15
Aggregate functions, Group by clause, Order by clause, Nested Sub-Queries, correlated Sub -Queries Implementation: ORACLE SQL/MS SQL/MySQL, the concept of database security.	

List of Recommended Reference Books

1. Silberschatz, A., Korth, H. F., Sudarshan, S. (2011). Database System Concepts. United States: McGraw-Hill.
2. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. (2015). Germany: Wiley.
3. Ramakrishnan, R., Gehrke, J. (2000). Database Management Systems. United States: McGraw-Hill.

ASSESSMENT:
THEORY:
CIA I: Written test **for 20 marks**
CIA II: Assignments / Project / Presentation / Case Study/ Written Test **for 20 marks**
End semester Exam :60 marks
Practical Exam:50 marks
Template for the Core course End Semester examination in Semester I

UNITS	KNOWLEDGE	UNDERSTANDING	APPLICATION and ANALYSES	TOTAL MARKS- Per unit
1	5	5	10	20
2	5	10	5	20
3	5	5	10	20
-TOTAL - Per objective	15	20	25	60
% WEIGHTAGE	25%	33%	42%	100%

M.Sc-I(Big Data Analytics)

Course Code: PSBDA6004CR1

Title: Basics of Programming
Credits :4 (Theory -1-Total 15 hr and Practical 1-Total -30hr)

Course Objectives:
Develop the skill to create basic applications using Java.

Course outcomes:
On completing the course, the student will be able to:

1. Learn basic programming fundamentals using Java.
2. Understand various data structures in java along with learning object-oriented programming through java.
3. Explain OOP principles in Java, including classes, objects, inheritance, and polymorphism.

Unit 1 15

Overview of Java: OOPS fundamentals, Interface and Package Overview of Java Difference between C++ and Java. Installation of JDK, Features JDK. Difference between JDK and JRE. Architecture of Java, portability and Features of Java.

Variables and Datatypes: Variables and datatype in Java, scope and lifetime of variables, Arrays in Java-1D and 2D, Arithmetic operator, Boolean operators, assignment operators, operator, Strings.

Control statements: while, do-while, for, if-else, switch.

Unit 2 15

OOPS fundamentals: What is class and objects, Meaning of Object oriented and its Features? Assigning Object Reference Variables, Methods, Passing different parameter to method with different return type,

Constructors: What is constructor, types of constructors, this and super keyword.

Inheritance: Derived Class Objects, Inheritance and Access Control, Default Base Class Constructors, Method overloading Vs Method overriding.

Interface and Packages: Packages, Importing Packages, Interfaces, Defining an Interface Implementing Interfaces, Nested Interfaces, Extending Interfaces.

List of Recommended Reference Books

1. Programming with JAVA - A Primer: Third Edition. (2014). India: McGraw-Hill Professionals.
2. Schildt, H. (2018). Java: The Complete Reference, Eleventh Edition. United States: McGraw Hill LLC.

ASSESSMENT:
THEORY:
CIA I: Written test **for 20 marks**
OR
Assignments / Project / Presentation / Case Study **for 20 marks**

End semester Practical Exam: 30 Marks

Syllabus

Second Semester Courses in M.Sc. Big Data Analytics 2023-2024

Contents:

APPROVED SYLLABUS

M.Sc-I (Big Data Analytics)

Course Code: PSBDA6005CR1

Title: FOUNDATIONS OF DATA SCIENCE
Credits :4 (Theory -3 -Total 45 hr and Practical -1-Total -30hr)

Course objectives:
Apply graph theory concepts to solve practical problems and implement graph algorithms using Java and Python

Course outcomes:
On completing the course, the student will be able to:
1. Understand the Fundamentals of Graph Theory:
2. Explore High-Dimensional Space:
3. Analyze Random Graphs:
4. Master Singular Value Decomposition (SVD):
5. Tackle Massive Data Problems

Unit 1 **15**

Graph Theory:
Basic Concepts, Algorithms for connectedness, shortest path, Minimum Sampling Tree, Lab: Graph Databases, Java/Python Programming

High Dimensional Space:
Properties, Law of large numbers, Sphere and cube in high dimension, Generating points on the surface of a sphere, Gaussians in High dimension, Random projection, Applications.

Unit 2 **15**

Random Graphs: Large graphs, G(n,p) model, Giant Component, Connectivity, Cycles, Non-Uniform models, Applications.

Singular Value Decomposition (SVD): Best rank k approximation, Power method for computing the SVD, Applications.

Unit 3 **15**

Random Walks:
Reflection Principle, Long leads, Changes of Sign, Illustrations.

Algorithm for Massive Data Problems: Frequency Moments of data streams, matrix algorithms.

SUGGESTED BOOK:

1. Blum, A., Hopcroft, J., Kannan, R. (2020). Foundations of Data Science. India: Cambridge University Press.

ASSESSMENT:
THEORY:
CIA I: Written test **for 20 marks**
CIA II: Assignments / Project / Presentation / Case Study/ Written Test **for 20 marks**
End semester Exam: 60 marks
Practical Exam:50 marks
Template for the Core course End Semester examination in Semester II

UNITS	KNOWLEDGE	UNDERSTANDING	APPLICATION and ANALYSES	TOTAL MARKS- Per unit
1	5	5	10	20
2	5	10	5	20
3	5	5	10	20
-TOTAL - Per objective	15	20	25	60
% WEIGHTAGE	25%	33%	42%	100%

M.Sc-I (Big Data Analytics)

Course Code: PSBDA6006CR1

Title: Advanced Statistical Methods
Credits :4 (Theory -3 -Total 45 hr and Practical -1-Total -30hr)
Course objectives:
1. Develop good estimators for population characteristics using different estimation techniques.
2. Understand formulation of Statistical hypothesis in real-life situations
3. Perform ANOVA techniques.
4. Learn regression analysis and study variable selection criteria to choose the best model.

Course outcomes: On completing the course, the student will be able to:
1. Apply suitable estimation techniques that meet the constraints of the given data scenario
2. Formulate Statistical hypothesis & apply appropriate tests to validate them
3. Perform One Way & Two Way ANOVA and interpret the results.
4. Estimate the parameters of the linear models, perform validation of a regression model and interpret the results in practical examples

Unit 1:
Estimation: Unbiasedness, Consistency, Sufficiency and CRLB, UMVUE, Maximum likelihood estimates.
Unit 2:
Test of Hypotheses & Analysis of Variance (ANOVA) : Two types of errors, test statistic, parametric tests for equality of means & variances, One Way & Two Way ANOVA
Unit 3:
Linear Model & Regression Analysis: Gauss Markov Model, least square estimators, Multiple linear regression, all possible, forward, backward & stepwise regression, Logistic Regression.

SUGGESTED BOOKS:

1. Statistical Inference: P. J. Bickel and K. A. Docksum, 2nd Edition, Prentice Hall.
2. Montgomery, D. C., Peck, E. A., Vining, G. G. (2021). Introduction to Linear Regression Analysis. United Kingdom: Wiley.

ASSESSMENT:

ASSESSMENT:
THEORY:
CIA I: Written test **for 20 marks**
CIA II: Assignments / Project / Presentation / Case Study/ Written Test **for 20 marks**
End semester Exam :60 marks
Practical Exam:50 marks
Template for the Core course End Semester examination in Semester II

UNITS	KNOWLEDGE	UNDERSTANDING	APPLICATION and ANALYSES	TOTAL MARKS- Per unit
1	5	5	10	20
2	5	10	5	20
3	5	5	10	20
-TOTAL - Per objective	15	20	25	60
% WEIGHTAGE	25%	33%	42%	100%

M.Sc-I (Big Data Analytics)

Course Code: PSBDA6007CR1

Title: Machine Learning- 1
Credits :4 (Theory -3 -Total 45 hr and Practical- 1-Total -30hr)

Course objectives:

1. Understand and apply linear regression techniques with multiple variables.
2. Explore real-world applications where linear regression is useful.
3. Gain a comprehensive understanding of representation learning and neural network architectures.
4. Implement single and multi-layer perceptrons.
5. Learn the backpropagation algorithm for training neural networks.
6. Apply neural networks to solve various real-world problems.
7. Evaluate the performance of a learning algorithm, considering metrics and methodologies.

Course outcomes:

On completing the course, the student will be able to:

1. Analyze diverse datasets and select appropriate machine learning techniques for specific tasks.
2. Implement neural networks, including single and multi-layer perceptrons, for various applications.
3. Identify and address challenges and limitations in real-world machine learning applications.
4. Apply machine learning algorithms and concepts to solve real-world problems
5. Analyze diverse datasets and select appropriate machine learning techniques for specific tasks.

Unit 1 15

Linear Regression: Linear Regression with Multiple variables, applications.
Logistic Regression: Model, Classification, Problem of over-fitting, Applications.

Unit 2 15

Neural Networks: Representation Learning, Different Models like single and multi-layer perceptron, back propagation, Application.
Machine Learning System Design: Evaluating a learning algorithm, handling skewed data, using large data sets

Unit 3 15

Support Vector Machines: Model, Large Margin Classification, Kernels, SVMs in practice, Unsupervised Learning, Dimensionality Reduction, Anomaly Detection.

SUGGESTED BOOKS:

1. Mitchell, T. M. (1997). Machine Learning. Germany: McGraw-Hill Education.
2. Machine Learning: An Artificial Intelligence Approach. (2013). Germany: Springer Berlin Heidelberg.

ASSESSMENT:
THEORY:
CIA I: Written test **for 20 marks**
CIA II: Assignments / Project / Presentation / Case Study/ Written Test **for 20 marks**
End semester Exam: 60 marks
Practical Exam: 50 marks
Template for the Core course End Semester examination in Semester II

UNITS	KNOWLEDGE	UNDERSTANDING	APPLICATION and ANALYSES	TOTAL MARKS- Per unit
1	5	5	10	20
2	5	10	5	20
3	5	5	10	20
-TOTAL - Per objective	15	20	25	60
% WEIGHTAGE	25%	33%	42%	100%

M.Sc-I (Big Data Analytics)

Course Code: PSBDA6008CR1

Title: Enabling Technologies for Data Science I
Credits: 4 (Theory -3 -Total 45 hr and Practical -1-Total -30hr)

Course objective:

1. Understand Data Mining:
2. Introduce students to the concepts of data mining and knowledge discovery from databases.
3. Provide a foundation in data warehousing principles, including modeling, design, implementation, and optimization.
4. Teach data preprocessing techniques, sub-sampling, and feature selection.
5. Enable students to clean and transform raw data into a suitable format for analysis.

Course outcomes:

On completing the course, the student will be able to:

1. Understand the fundamental concepts of data mining and its practical applications.
2. Grasp the principles of data warehousing, including modeling and design.
3. Perform data preprocessing, sub-sampling, and feature selection.
4. Prepare raw data for analysis by cleaning and transforming it effectively.
5. Implement classification and prediction algorithms such as Bayes learning, decision trees, and ensemble learning

Unit 1 **15**

Introduction: Data mining, Knowledge discovery from databases, scalability issues, application of data mining

***Data Warehousing*:** General principles, modeling, design, implementation and optimization, OLAP Vs OLTP, ETL, dimensional modelling, Snowflake Vs Star Schema.

Unit 2 **15**

Data Preparation*:* Pre-processing, sub-sampling, feature selection.

Classification and Prediction: Bayes learning, decision trees, CART, associations, ensemble learning dependence analysis, rule generation

Unit 3 **15**

Cluster Analysis and Deviation Detection:

Partitioning algorithms, Density bases algorithm, Grid based algorithm.

SUGGESTED BOOKS

1. Pujari, A. K. (2001). Data Mining Techniques. India: Universities Press.
2. Mastering Data Mining: The Art And Science Of Customer Relationship Management. (2008). India: Wiley India Pvt. Limited.
3. Rud, O. P. (2001). Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management. Germany: Wiley.

ASSESSMENT:
THEORY:
CIA I: Written test **for 20 marks**
CIA II: Assignments / Project / Presentation / Case Study/ Written Test **for 20 marks**
End semester Exam: 60 marks
Practical Exam:50 marks
Template for the Core course End Semester examination in Semester II

UNITS	KNOWLEDGE	UNDERSTANDING	APPLICATION and ANALYSES	TOTAL MARKS- Per unit
1	5	5	10	20
2	5	10	5	20
3	5	5	10	20
-TOTAL - Per objective	15	20	25	60
% WEIGHTAGE	25%	33%	42%	100%