IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)

Department of Computer Science & Engineering (Data Science)

Bachelor of Technology (B.Tech.)

Minor Certification in Computer Science & Engineering (Data Science)

(To be offered to students of other departments excluding CSE (Data Science))

S.No.	Subject Code	Semester	Subject Name	Contact Hours per week			Total Credits
				L	T	P	
1.	MICSDS-501	V	Foundation of Data Science	2	1	2	4
2.	MICSDS-601	VI	Essentials of Data Science with R software	2	1	2	4
3.	MICSDS-701	VII	Database Management System	2	1	2	4
4.	MICSDS-801	VIII	*Operating System	2	1	-	3
			Total	8	4	6	15

1 Hr Lecture 1 Hr Tutorial 2 Hr Practical

1 Credit 1 Credit 1 Credit

Note: *VIII semester subject (Operating System or any other course equivalent to OS) can also be done from MOOC courses (NPTEL etc.) with minimum credit 3.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering (Data Science)

Bachelor of Technology (B.Tech.)
Minor Certification in Computer Science & Engineering (Data Science)

MICSDS-501	**Foundation of Data Science**	**2L : 1T : 2P (5 hrs.)**	**4 credits**

Prerequisite: Engineering Mathematics.

Course Objective:
This course provides a concise introduction to the fundamental concepts of Data Science

Course Contents :(40 hrs)

Module1: **(10 hrs.)**
Introduction: What is Data Science? Big Data and Data Science – Datafication - Current landscape of perspectives - Skill sets needed; Matrices - Matrices to represent relations between data, and necessary linear algebraic operations on matrices -Approximately representing matrices by decompositions (SVD and PCA); Statistics: Descriptive Statistics: distributions and probability - Statistical Inference: Populations and samples - Statistical modeling - probability distributions - fitting a model - Hypothesis Testing

Module 2: **(08 hrs.)**
Data preprocessing: Data cleaning - data integration - Data Reduction Data Transformation and Data Discretization. Evaluation of classification methods – Confusion matrix, Students T-tests and ROC curves-Exploratory Data Analysis - Basic tools (plots, graphs and summary statistics) of EDA, Philosophy of EDA - The Data Science Process.

Module 3: **(10 hrs.)**
Basic Machine Learning Algorithms: Association Rule mining - Linear Regression- Logistic Regression - Classifiers - k-Nearest Neighbors (k-NN), k-means -Decision tree - Naive Bayes- Ensemble Methods - Random Forest. Feature Generation and Feature Selection - Feature Selection algorithms - Filters; Wrappers; Decision Trees; Random Forests.

Module 4: **(08 hrs.)**
Clustering: Choosing distance metrics - Different clustering approaches - hierarchical agglomerative clustering, k-means (Lloyd's algorithm), - DBSCAN - Relative merits of each method - clustering tendency and quality.

Module 5: **(06 hrs**.)
Case Studies of various data science tool their features and usage**.**

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering (Data Science)

Bachelor of Technology (B.Tech.)
Minor Certification in Computer Science & Engineering (Data Science)

Course Outcome:

1. State the overview of the Data Science.
2. Explain the process of Data preprocessing.
3. Discuss the various Machine Learning Algorithms.
4. Explain the clustering techniques.
5. Discuss the various Data Science Tool.

List of Text / Reference Books:

1. Cathy O'Neil and Rachel Schutt, " Doing Data Science, Straight Talk From The Frontline", O'Reilly, 2014.
2. Jiawei Han, Micheline Kamber and Jian Pei, " Data Mining: Concepts and Techniques", Third Edition. ISBN 0123814790, 2011.
3. Mohammed J. Zaki and Wagner Miera Jr, "Data Mining and Analysis: Fundamental Concepts and Algorithms", Cambridge University Press, 2014.
4. Matt Harrison, "Learning the Pandas Library: Python Tools for Data Munging, Analysis, and Visualization , O'Reilly, 2016.
5. Joel Grus, "Data Science from Scratch: First Principles with Python", O'Reilly Media, 2015.
6. Wes McKinney, "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython", O'Reilly Media, 2012.
7. **NPTEL Course Link: https://nptel.ac.in/courses/106/106/106106212/**
8. **NPTEL Course Link: https://nptel.ac.in/courses/106/106/106106179/**

List of Experiments:

Apply pre-processing techniques on the following raw dataset.

1. Titanic Data Set.
2. Boston Housing Data Set.
3. Walmart Sales Forecasting Data Set.
4. Iris Data Set.
5. Netflix Movies and TV Shows dataset.
6. IPL Data Set.
7. Twitter Data Set.
8. Haberman Cancer Survival data set.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering (Data Science)

Bachelor of Technology (B.Tech.)
Minor Certification in Computer Science & Engineering (Data Science)

MICSDS-601	**Essentials of data science with R software**	**2L: 1T: 2P (5 hrs.)**	**Credits:04**

Prerequisite: Engineering Mathematics.

Course Objective:

Course Objective: The main objective of this course is to provide students with the foundations of probabilistic and statistical analysis using R Programming which is mostly used in varied applications in engineering

Module 1: **(08 hrs.)**

Introduction to basic concepts of R programming language, Introduction to Probability , Sample Space and Events , Set Theory and Events using Venn Diagrams , Relative Frequency and Probability , Probability and Relative Frequency - An Example , Axiomatic Definition of Probability , Some Rules of Probability , Basic Principles of Counting- Ordered Set, Unordered Set, and Permutations , Basic Principles of Counting- Combination , Conditional Probability , Multiplication Theorem of Probability, Bayes' Theorem , Independent Events ..

Module 2: **(10 hrs.)**

Random Variables - Discrete and Continuous , Cumulative Distribution and Probability Density Function , Discrete Random Variables, Probability Mass Function and Cumulative Distribution Function , Expectation of Variables, Moments and Variance , Skewness and Kurtosis , Quantiles and Tschebyschevâ€™s Inequality .

Module 3: **(08 hrs.)**

Degenerate and Discrete Uniform Distributions, Bernoulli and Binomial Distribution, Poisson Distribution , Geometric Distribution , Continuous Random Variables and Uniform Distribution, Normal Distribution, Exponential Distribution , Bivariate Probability Distribution for Discrete Random Variables, Bivariate Probability Distribution for Continuous Random Variables, Examples in Bivariate Probability Distribution Functions .

Module 4: **(10 hrs.)**

Covariance and Correlation , Bivariate Normal Distribution, Chi square Distribution, t - Distribution , F – Distribution, Distribution of Sample Mean, Convergence in Probability and Weak Law of Large Numbers , Central Limit Theorem , Needs for Drawing Statistical Inferences , Unbiased Estimators , Efficiency of Estimators , CramÃ©râ€“Rao Lower Bound and Efficiency of Estimators , Consistency and Sufficiency of Estimators, Method of Moments, Method of Maximum Likelihood and Rao Blackwell Theorem.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering (Data Science)

Bachelor of Technology (B.Tech.)
Minor Certification in Computer Science & Engineering (Data Science)

Module 5: **(08 hrs.)**

Basic Concepts of Confidence Interval Estimation ,Confidence Interval for Mean in One Sample with Known Variance ,Confidence Interval for Mean and Variance, Basics of Tests of Hypothesis and Decision Rules ,Test Procedures for One Sample Test for Mean with Known Variance ,One Sample Test for Mean with Unknown Variance ,Two Sample Test for Mean with Known and Unknown Variances ,Test of Hypothesis for Variance in One and Two Samples.

List of Text Books / Reference Books:

1. Introduction to Statistics and Data Analysis With Exercises, Solutions and Applications in R Authors:Heumann, Christian, Schomaker, Michael, Shalabh, Publisher" Springer 2016
2. Applied Statistics and Probability for Engineers, Douglas C. Montgomery, George C. Runger, 2018,Wiley (Low price edition available)
3. Introduction to. Mathematical. Statistics. Robert V. Hogg. Allen T. Craig,, Low price Indian edition byPearson Education
4. Probability and Statistics for Engineers. Richard A. Johnson, Irwin Miller, John Freund
5. Mathematical Statistics with Applications. Irwin Miller, Marylees Miller, Pearson Education
6. The R Software-Fundamentals of Programming and Statistical Analysis -Pierre Lafaye de Micheaux,Rémy Drouilhet, Benoit Liquet, Springer 2013
7. A Beginner's Guide to R (Use R) By Alain F. Zuur, Elena N. Ieno, Erik H.W.G. Meesters, Springer2009.

List of Experiments:

1. Installation and Working with R.
2. Installation and Working with R Studio.
3. Calculations with R as a Calculator, Calculations with Data Vectors.
4. Built-in Commands and Bivariate Plots.
5. Logical Operators and Selection of Sample.
6. Computation of Probability using R
7. Data Based Moments and Variance in R Software.
8. Discrete Uniform Distribution in R.
9. Binomial Distribution in R.
10. Poisson distribution in R.
11. Geometric Distribution in R.
12. Normal Distribution in R.
13. Chi square Distribution in R
14. Bivariate Probability Distribution in R Software.
15. Covariance and Correlation in R.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering (Data Science)

Bachelor of Technology (B.Tech.)
Minor Certification in Computer Science & Engineering (Data Science)

MICS-701	**Database Management System**	**2L: 1T: 2P (5 hrs.)**	**4 credits**

Course Objective:

The main objective of this course is to understand fundamental of database management system.

Course Contents: (45 hrs.)

Module 1: (06 hrs.)

DBMS Concepts and architecture Introduction, Database approach v/s Traditional file accessing approach, Advantages, of database systems, Data models, Schemas and instances, Data independence, Data Base Language and interfaces, Overall Database Structure, Functions of DBA and designer, ER data model: Entitles and attributes, Entity types, Defining the E-R diagram, Concept of Generalization, Aggregation and Specialization. transforming ER diagram into the tables. Various other data models object- oriented data Model, Network data model, and Relational data model, Comparison between the three types of models.

Module 2: (08hrs.)

Relational Data models: Domains, Tuples, Attributes, Relations, Characteristics of relations, Keys, Key attributes of relation, Relational database, Schemas, Integrity constraints. Referential integrity, Relational Query languages: SQL-DDL, DML, integrity con straints, Complex queries, various joins, Relational algebra and relational calculus, Relational algebra operations like select, Project, Join, Division, outer union. Types of relational calculus i.e. Tuple oriented and domain oriented relational calculus and its operations.

Module 3: (14 hrs.)

Data Base Design: Introduction to normalization, Normal forms, Functional dependency, Decomposition, Dependency preservation and lossless join, problems with null valued and dangling tuples, multivalued dependencies. Query Optimization : Introduction, steps of optimization, various algorithms to implement select, project and join operations of relational algebra, optimization methods: heuristic based, cost estimation based.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering (Data Science)

Bachelor of Technology (B.Tech.)
Minor Certification in Computer Science & Engineering (Data Science)

Module 4: **(09 hrs.)**

Transaction Processing Concepts: -Transaction s ystem , Testing of Serializability, Serializability of schedules, conflict & view serializable schedule, recoverability, Recovery from transaction failures. Log based recovery. Checkpoints deadlock handling. Concurrency

Control Techniques: Concurrency Control, locking Techniques for concurrency control, time stamping protocols for concurrency control, validation based protocol, multiple granularity. Multi version schemes, Recovery with concurrent transaction.

Module 5: **(08 hrs.)**

Study of Relational Database Management Systems through Oracle/PL SQL

QL/MySQL: Architecture, physical files, memory structures, background process. Concept of table spaces, segments, extents and block. Dedicated server, multi threaded server. SQL queries, Data extraction from single, multiple tables equi- join, non equi-join, self - join, outer join. Usage of like, any, all, exists, in Special operators. Cursor management: nested and parameterized cursors, Oracle exception handling mechanism. Stored procedures, in, out, in out type parameters, usage of parameters in procedures. User defined functions their limitations. Triggers, mutating errors, instead of triggers.

Course Outcomes:

1. Describe basic concepts of DBMS and Explain ER model.
2. Solve queries using Relational Algebra, Relational Calculus and SQL.
3. Explain database schema and discuss the Query optimization methods.
4. Describe transaction processing, concurrency control and recovery technique.
5. Analyze the Various DBMS software like Oracle, SQL/PL SQL etc.

List of Text / Reference Books:

1. Date C J, “An Introduction to Database System”, Pearson Educations, 8th Edition,2003.
2. Korth, Silbertz,Sudarshan, “Fundamental of Database System”, McGraw Hill,5th Edition,2006.
3. Peter Rob, “ Data Base System:Design Implementation & Management”, Cengage Learninig 4th Edition,2000.
4. Elmasri, Navathe, “Fundamentals of Database Systems”, Pearson Educations,7th Edition 2017.
5 . Atul Kahate ,“ Introduction to Database Management System”, Pearson Educations,2004.
6. Oracle 9i Database Administration Fundamental-I, Volume I, Oracle Press, TMH.
7. Paneerselvam,”DataBase Management System”, PHI Learning,3rd Edition,2018.
8. J. D. Ullman, “Principles of Database and Knowledge – Base Systems”, Computer Science Press,2nd Edition 1988.
9. Serge Abiteboul, Richard Hull, Victor Vianu,“Foundations of Databases”, Addison-Wesley,1995.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering (Data Science)

Bachelor of Technology (B.Tech.)
Minor Certification in Computer Science & Engineering (Data Science)

List of Experiments

1. Introduction to Oracle and SQL
2. Write the queries for Data Definition language (DDL)
3. Write the queries for Data manipulation language (DML)
4. Use of various types of Integrity Constraints
5. Write the queries for Data Control language (DCL)
6. Use of SELECT command with different clauses.
7. Write SQL queries using logical operation (AND, OR, NOT)
8. Write SQL queries for aggregate functions (Max, Min, Sum, Avg, and Count)
9. Write SQL queries for group by and having
10. Write SQL queries for sub queries and nested queries
11. Write an SQL query to implement JOINS
12. Write SQL queries to create views
13. Write program by the use of PL/SQL
14. Design and implementation of any Data base system (like Banking, University etc).

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering (Data Science)

Bachelor of Technology (B.Tech.)
Minor Certification in Computer Science & Engineering (Data Science)

MICSDS-801	**Operating System**	**2L: 1T: 0P (3 hrs.)**	**3 credits**

Course Objective:
This Course provides a comprehensive introduction of Operating System, Process Management, Memory Management, File Management and I/O management.

Course Contents: (40 hrs.)

Module 1: **(06 hrs.)**
Introduction to Operating Systems: Function, Evolution, Different Types, Desirable Characteristics and features of an O/S, Operating Systems Services: Types of Services, Different ways of providing these Services – Utility Programs, System Calls, Operating System Structure, and Spooling & Buffering.

Module 2: **(11 hrs.)**
CPU Scheduling : Process Concept, Scheduling Concepts, Types of Schedulers, Scheduling Criteria, Process State Diagram, Scheduling Algorithms, Operation on Process, Algorithms Evaluation, System calls for Process Management; Multiple Processor Scheduling; Concept of Threads. Concurrent Processes : Real and Virtual Concurrency, Mutual Exclusion, Synchronization, Inter- Process Communication, Critical Section Problem, Solution to Critical Section Problem : Semaphores – Binary and Counting Semaphores, WAIT & SIGNAL Operations and their implementation. Deadlocks: Deadlock Problems, Characterization, Prevention, Avoidance, Recovery.

Module 3: **(11 hrs.)**
Memory Management: Different Memory Management Techniques – Partitioning, Swapping, Segmentation, Paging, Paged Segmentation, Comparison of these techniques, Techniques for supporting the execution of large programs: Overlay, Dynamic Linking and Loading, Virtual Memory – Concept, Implementation by Demand Paging etc., Page replacement algorithms.

Module 4: **(06 hrs.)**
File Systems: File Concept, User's and System Programmer's view of File System, Disk Organization, Tape Organization, Different Modules of a File System, Disk Space Allocation Methods – Contiguous, Linked and Indexed. Directory Structures, File Protection, System Calls for File Management, Disk Scheduling Algorithms.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering (Data Science)

Bachelor of Technology (B.Tech.)
Minor Certification in Computer Science & Engineering (Data Science)

Module 5: **(06 hrs.)**

Introduction to Network, Distributed and Multiprocessor Operating Systems. Case tudies: Unix/Linux, WINDOWS and other Contemporary Operating Systems.

Course Outcomes:

1. State the core concepts of operating system, evolution and types of operating system.
2. Illustrate various input output concepts, interprocess communication and deadlock
3. Illustrate process scheduling and memory management techniques.
4. Describe the concept of file and disk management.
5. State the core concepts of network, distributed and multiprocessor operating system.

List of Text / Reference Books:

1. Avi Silberschatz, Peter Galvin, Greg Gagne, "Operating System Concepts Essentials", Wiley Asia Student Edition, 10th Edition, 2018.
2. William Stallings, "Operating Systems: Internals and Design Principles", Prentice Hallof India, 5th Edition, 2005.
3. Charles Crowley, "Operating System: A Design-oriented Approach", Irwin Publishing, 1st Edition.
4. Gary J. Nutt, "Operating Systems: A Modern Perspective", Addison-Wesley, 2nd Edition.
5. Maurice Bach, "Design of the Unix Operating Systems", Prentice-Hall of India, 8th Edition.
6. Daniel P. Bovet, Marco Cesati, "Understanding the Linux Kernel", O'Reillyand Associates, 3rd Edition.
7. Andrew S. Tanenbaum, "Modern Operating Systems", Prentice Hall, 3rd Edition,2007.
8. Bovet & Cesati, "Understanding the Linux Kernel", O'Reily, 3rd Edition.