ST. JOSEPH'S COLLEGE (AUTONOMOUS)

BENGALURU-27

Re-accredited with **'A++' GRADE with 3.79/4 CGPA** by
NAAC Recognized by UGC as College of Excellence

ST. JOSEPH'S INSTITUTE OF INFORMATION TECHNOLOGY

DEPARTMENT OF ADVANCED COMPUTING

SYLLABUS FOR UNDERGRADUATE PROGRAMME

For Batch 2021-2024

SUMMARY OF CREDITS IN BCA(Data Analytics)
Revision Year - 2021

OPEN ELECTIVES

GENERIC ELECTIVE COURSES(GSE)/CAN INCLUDE OPEN ELECTIVE OFFERED	
Course Title	Code Number
Basics of Data Science(For First Semester)	
Python Programming(For First Semester)	
Digital Design using HDL(For Second Semester)	
Machine Learning Using R Programming(For Second Semester)	

Semester	I
Paper Code	
Paper Title	**BASICS OF DATA SCIENCE**
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE DESCRIPTION:

To provide strong foundation for data science and application area related to it and understand the underlying core concepts and emerging technologies in data science.

COURSE OBJECTIVES:

To make the students learn the process of working with data in large scale. Make the student understand the existence of data with its wilderness and make use of it.

COURSE OUTCOMES:

CO1: Understand the fundamental concepts of data.

CO2: Understand the fundamental concepts of data science process.

CO3: Understand the fundamental concepts of Machine Learning

CO4: Fundamental concepts of large data & Data Visualization

CO5: To implement the aspects of Data Science through case studies.

UNIT 1: PREPARING AND GATHERING DATA AND KNOWLEDGE (9 Hrs)

Philosophies of data science - Data science in a big data world - Benefits and uses of data science and big data - facts of data: Structured data , Unstructured data, Natural Language, Machine generated data, Audio, Image and video streaming data - The Big data Eco system: Distributed file system, Distributed Programming framework, Data Integration frame work, Machine learning Framework, NoSQL Databases, Scheduling tools, Benchmarking Tools, System Deployment, Service programming and Security.

UNIT 2: THE DATA SCIENCE PROCESS (9 Hrs)

Overview of the data science process- Retrieving data –Data Preparation: Cleansing, integrating, and transforming data - Exploratory data analysis – Data Modeling: Model and variable selection, Model execution, Model diagnostic and model comparison - Presentation and automation: Presenting data, Automating data analysis

UNIT 3: MACHINE LEARNING (9 Hrs)

Application for machine learning in data science- Tools used in machine learning- Modeling Process – Training model – Validating model – Predicting new observations –Types of machine learning Algorithm : Supervised learning algorithms, Unsupervised learning algorithms.

UNIT 4: VISUALIZATION (9 Hrs)

Introduction to data visualization – Data visualization options – Filters – MapReduce – Dashboard development tools.

UNIT 5: CASE STUDIES (9 Hrs)

Distributing data storage and processing with frameworks - Case study: e.g, Assessing risk when lending money.

TEXT BOOKS:

1. Introducing Data Science, Davy Cielen, Arno D. B. Meysman and Mohamed Ali, Manning Publications, 2016.
2. Think Like a Data Scientist, Brian Godsey, Manning Publications, 2017.

REFERENCE BOOKS:

1. Doing Data Science, Straight Talk from the Frontline, Cathy O'Neil, Rachel Schutt, O' Reilly, 1st edition, 2013.
2. Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman, Cambridge University Press, 2nd edition, 2014
3. An Introduction to Statistical Learning: with Applications in R, Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer, 1st edition, 2013

Code number:

Title of the paper: **BASICS OF DATA SCIENCE**

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit 1	09	15
Unit 2	09	25
Unit 3	09	25
Unit 4	09	25
Unit 5	09	25
TOTAL	**45**	**115**
Maximum marks for the paper (Excluding bonus question) = 70		

Semester	I
Paper Code	
Paper Title	**PYTHON PROGRAMMING**
Number of teaching hrs per week	3 Hrs

Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

The course is designed to provide Basic knowledge of Python. Python programming is intended for software engineers, system analysts, program managers and user support personnel who wish to learn the Python programming language.

COURSE OUTCOMES:

CO1: To understand the basic concepts in Python programming.

CO2: Learn how to write, debug and execute Python program.

CO3. Understand and demonstrate the use of Branching and Looping Structures.

CO4: To get insight knowledge related to advanced data types such as lists, tuples, dictionaries.

CO5: Acquire the basic knowledge of Object-Oriented Programming Concept and Exception Handling.

UNIT 1: INTRODUCTION (9 Hrs)

Introduction to Python Programming, History of Python, its features, Scope of Python, Downloading and installing Python, Python code execution process, run a simple program on Python interpreter and IDLE.

UNIT 2: DATA TYPES AND OPERATORS (9 Hrs)

The concept of data types; variables, assignments; immutable variables; numerical types; arithmetic operators and expressions; comments in the program; understanding error messages; Illustrative programs.

UNIT 3: BRANCHING AND LOOPING (10 Hrs)

Conditions, Boolean logic, logical operators; ranges; Control statements: if-else, loops (for, while); short-circuit (lazy) evaluation. Illustrative programs

UNIT 4: LISTS, TUPLES AND DICTONARIES (10 Hrs)

Lists, tuples, and dictionaries; basic list operators, replacing, inserting, removing an element; searching and sorting lists; dictionary literals, adding and removing keys, accessing and replacing values, Illustrative programs

UNIT 5: OOPS AND EXCEPTION HANDLING (7 Hrs)

Classes and OOP: classes, objects, attributes and methods; defining classes; design with classes, exception handling, Illustrative programs

TEXTBOOKS:

1. Python in easy steps - Mike McGrath, In Easy Steps Limited, Second Edition

2. "Hello World" - Computer Programming for Kids and other Beginners - Warren and Carter, Manning Publications, 2014

SUGGESSTED BOOKS:

1. Python3 Tutorial – Tutorialspoint
2. Mark Lutz, Programming Python, O`Reilly, 4th Edition, 2010

Code number:

Title of the paper: **PYTHON PROGRAMMING**

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit 1	09	15
Unit 2	09	25
Unit 3	09	25
Unit 4	09	25
Unit 5	09	25
TOTAL	**45**	**115**
Maximum marks for the paper (Excluding bonus question) = 70		

Semester	II
Paper Code	
Paper Title	**DIGITAL DESIGN USING HDL**
Number of teaching hrs per week	3 Hrs

Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

- Learn different Verilog HDL constructs.
- Familiarize the different levels of abstraction in Verilog.
- Understand Verilog Tasks and Directives.
- Understand timing and delay Simulation.

COURSE OUTCOMES:

CO1: Design and analyze combinational & sequential circuits
CO2: Understand different design methodologies
CO3: Write Verilog programs in gate, dataflow (RTL), behavioral and switch modeling levels of Abstraction.
CO4: Write the programs more effectively using Verilog tasks and directives
CO5: To learn and implement the Behavioral Modeling Concepts

UNIT 1: DESIGN OF COMBINATIONAL LOGIC & SEQUENTIAL LOGIC(9 Hrs)

Decoders, Encoders, Digital multiplexers, Adders and subtractors, Look ahead carry, Binary comparators. Latches, Flipflops, Counters, Design of sequential counters, state machine & State diagrams.

UNIT 2: OVERVIEW OF DIGITAL DESIGN WITH VERILOG HDL (9 Hrs)

Evolution of CAD, emergence of HDLs, typical HDL-flow, why Verilog HDL?, trends in HDLs. Hierarchical Modeling Concepts : Top-down and bottom-up design methodology, differences between modules and module instances, parts of a simulation, design block, stimulus block.

UNIT 3: VERILOG BASIC CONCEPTS & MODULES & PORTS (9 Hrs)

Lexical conventions, data types, system tasks, compiler directives. Module definition, port declaration, connecting ports, hierarchical name referencing.

UNIT 4: GATE LEVEL & DATA FLOW MODELLING (9 Hrs)

Gate-Level Modeling

Modeling using basic Verilog gate primitives, description of and/or and buf/not type gates, rise, fall and turn-off delays, min, max, and typical delays.

Dataflow Modeling

Continuous assignments, delay specification, expressions, operators, operands, operator types.

UNIT 5: BEHAVIORAL MODELING (9 Hrs)

Structured procedures, initial and always, blocking and non-blocking statements, delay control, generate statement, event control, conditional statements, structural modeling.

TEXTBOOK:

1. Samir Palnitkar, "Verilog HDL: A Guide to Digital Design and Synthesis", Pearson Education, Second Edition.
2. John M Yarbrough,-Digital Logic Applications and Design, Thomson Learning,2001

SUGGESTED BOOK:

1. Donald E. Thomas, Philip R. Moorby, "The Verilog Hardware Description Language", Springer Science+Business Media, LLC, Fifth edition.
2. Michael D. Ciletti, "Advanced Digital Design with the Verilog HDL" Pearson (Prentice Hall), Second edition.

Code number:

Title of the paper: **DIGITAL DESIGN USING HDL**

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit 1	09	15
Unit 2	09	25
Unit 3	09	25
Unit 4	09	25
Unit 5	09	25
TOTAL	**45**	**115**
Maximum marks for the paper (Excluding bonus question) = 70		

Semester	II
Paper Code	
Paper Title	**MACHINE LEARNING USING R PROGRAMMING**
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

To make the students learn the statistics & mathematical concepts, Hypothesis & Dimension Reduction Technique, R Programming Concepts and Machine Learning.

COURSE OUTCOMES:

CO1: Understand the fundamental concepts of Statistics & Mathematics

CO2: Understand Hypothesis & Dimension Reduction Techniques

CO3: Hands on Experience in R Programming

CO4: Understand Machine Learning Concepts using R

CO5: To have basic knowledge of various predictive models.

UNIT 1: STATISTICS & MATHEMATICAL ESSENTIALS (9 Hrs)

Measure of Central Tendency - Mean, Median, Mode - Dispersion Technique - Range Inter Quartile Range - Variance, Standard Deviation - Mean Square Error & Root Mean Square - Probability Distribution.

UNIT 2: HYPOTHESIS AND DIMENSION REDUCTION TECHNIQUE (9 Hrs)

Types of Hypothesis - Sample testing - T-test - Z-test - Chi-square test - Anova test -. One Way Anova. Two Way Anova - Principle component analysis - Collinearity and multicollinearity

UNIT 3: R PROGRAMMING CONCEPTS (9 Hrs)

The Data types in R & its uses -Build in functions in R- Data Manipulation - Data import Techniques – Exploratory Data Analysis – Data Visualization.

UNIT 4: MACHINE LEARNING (9 Hrs)

ML Fundamental & common use cases - Approach to Machine Learning - Understanding Supervised learning technique - Unsupervised learning technique

UNIT 5: PREDICTIVE MODELLING IN R (9 Hrs)

Introduction to predictive modeling - Regression Problem - Classification Problem - Linear Regression - Logistic Regression – Clustering - Distance measure types- K means clustering – Decision Tree Classifier – Random Forest Classifier – Support Vector Machine.

TEXTBOOK:

1. Introducing Data Science, Davy Cielen, Arno D. B. Meysman and Mohamed Ali, Manning Publications, 2016.
2. Think Like a Data Scientist, Brian Godsey, Manning Publications, 2017.

SUGGESTED BOOK:

1. Doing Data Science, Straight Talk from the Frontline, Cathy O'Neil, Rachel Schutt, O' Reilly, 1st edition, 2013.
2. Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman, Cambridge University Press, 2nd edition, 2014
3. An Introduction to Statistical Learning: with Applications in R, Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer, 1st edition, 2013

Code number:

Title of the paper: **MACHINE LEARNING USING R PROGRAMMING**

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit 1	09	15
Unit 2	09	25
Unit 3	09	25
Unit 4	09	25
Unit 5	09	25

TOTAL	**45**	**115**
Maximum marks for the paper (Excluding bonus question) = 70		