JAYARAJ ANNAPACKIAM COLLEGE FOR WOMEN (AUTONOMOUS)

PERIYAKULAM

PG DEPARTMENT OF COMPUTER SCIENCE

POST GRADUATE DIPLOMA IN DATA SCIENCE (PGDDS)

The Post Graduate Diploma in Data Science (PGDDS) is an excellent blend of knowledge and practice in the field of Data Science and its industrial applications. The programme aims to prepare the student for a career as a Data Scientist, in the corporate sector, industries, entrepreneurship, public policy and academia. This programme introduces the learner to business analytics using the most in-demand analytics technologies like R and Python teaches implementation of various data science concepts such as data exploration, visualization, hypothesis testing as well as hands-on laboratory in Data Science. A special focus has been placed on Machine Learning techniques used for regression, classification and clustering. A student can take this programme in addition to her P.G. degree.

Eligibility: A Candidate should possess a Bachelor's Degree with basic mathematical knowledge of any university to get admission for this course **(Full-Time/Part-Time).**

Duration: It is divided into two Semesters of One Year duration.

PATTERN OF EVALUATION FOR PGDDS

For each course, there will be Continuous Internal Assessment (CIA) and Semester Examination (External). The Weightage ratio is

Paper	Internal	External	Total
Theory	25	75	100
Practical	40	60	100

Components for the Continuous Internal Assessment (CIA) Theory:

Component	Marks	Marks
Internal test I	40	Converted to 25
Internal test II	40	
Seminar	10	
Assignment/Case Study	5	
Attendance	5	
Total	**100**	**25**

The Internal Components are:

Practical	
Internal Test (2)	15
Lab Work	10
Record	10
Attendance	05
Total	**40**

Passing Minimum

	Semester Examination
Theory	50% out of 75 Marks (i.e. 37.5 Marks)
Practical	50% out of 60 Marks (i.e. 30 Marks)

Question Patterns for PGDDS

Internal Question Pattern

Time: 2 Hours **Maximum Mark: 40**

Part - A

10 Questions × 1Mark = 10 Marks

Part - B

2 Questions × 5 Marks = 10 Marks

(Internal Choice and One Question from Each Unit)

Part - C

2 Questions × 10 Marks = 20 Marks

(Open Choice, Two Questions out of Three)

External Question Pattern

Time: 3 Hours **Maximum Mark: 75**

Part - A

10 Questions × 1Mark = 10 Marks
(Two Questions from each Unit)

Part - B

5 Questions × 5 Marks = 25 Marks
(Internal Choice and one set of Question from each Unit)

Part - C

5 Questions × 8 Marks = 40 Marks
(Open Choice, Five Questions out of Seven - At least One Question from each Unit)

P.G. DIPLOMA PROGRAMME OUTCOMES

PO. NO.	UPON COMPLETION OF THIS PROGRAMME THE STUDENTS WILL BE ABLE TO
PO-1.	Endow with in-depth knowledge, analyze and apply the understanding of their discipline for the betterment of self and society.
PO-2.	Synthesize ideas from various disciplines, enhance the interdisciplinary knowledge and extend it for research.
PO-3.	Gain confidence and skills to communicate orally/ verbally in research platforms and state clear research finding.
PO-4.	Develop problem solving and computational skills and gain confidence to appear for the competitive examination.
PO-5.	Enhance knowledge regarding research by accumulating practical knowledge in specific areas of research.
PO-6.	Achieve idealistic goals and enrich the values to tackle the societal challenges.

P.G. DIPLOMA IN DATA SCIENCE PROGRAM SPECIFIC OUTCOMES

PSO	UPON COMPLETION OF THIS PROGRAMME THE STUDENTS WILL BE ABLE TO	PO MAPPED
PSO-1	**Develop Domain Expertise:** Understand and analyse the principles applied to contexts and environments of data science.	PO-1
PSO-2	**Acquire Computing Skills and Ethics:** Design effective advanced analytics models, simulations for decision making and data management for the betterment of society.	PO-6
PSO-3	**Inculcate Research:** Create integrated views of data collected from multiple sources of an enterprise and identify, analyze and synthesize knowledge behind these data using computing analysis with life-long learning.	PO-2, 5
PSO-4	**Develop Leadership Qualities:** Communicate effectively, to function efficiently on teams to accomplish shared computing models and analysis or implementation through case studies.	PO-3
PSO-5	**Produce Efficient Techno-Entrepreneurs:** To inculcate initiative, Innovative and professional outlook skilled elite for better industry business developments and apply their computing expertise to promote entrepreneurship.	PO-4

PGDDS COURSE PATTERN

(Affiliated to Mother Teresa University, Kodaikanal)

Sem.	Code	Title of the Paper	Hours	Credits
I	25PGDDS01	Fundamentals of Data Science	6	4
	25PGDDS02	Descriptive Statistics and Probability	6	4
	25PGDDS03	Data Analysis using Python	6	4
	25PGDDSP1	Data Analysis using Python – Lab	6	3
	25PGDDSP2	Data Analysis Using Spreadsheet - Lab	6	3
			30	**18**
II	25PGDDS04	Machine Learning Algorithms	6	4
	25PGDDS05	Data Analytics Using R	6	4
	25PGDDS06	Internet of Things	6	4
	25PGDDSP3	Data Analytics Using R – Lab	6	3
	25PGDDSP4	Internet of Things - Lab	6	3
			30	**18**
TOTAL			**60**	**36**

FUNDAMENTALS OF DATA SCIENCE

Semester : I **Hours: 6**

Code : 25PGDDS01 **Credits: 4**

Upon completion of this course students will be able to

- Understand the fundamental concepts of data science and implement machine learning models, including regression, classification, clustering, and ensemble methods, to solve real-world problems.
- Develop and implement data analysis strategies in the large data sets based on theoretical principles, and detailed knowledge of the underlying data.
- Familiar with the Concepts, Terminology, Characteristics and explain the use of Big Data in Business Context.

UNIT I

Introduction: Basics of Data Science. **Decision Theory**: Descriptive Analytics - Diagnostic Analytics - Predictive Analytics - Prescriptive Analytics. **Estimation Theory:** Basic Terminologies - Good Estimator. **Coordinate Systems:** Geographic Coordinate System - Projected Coordinate System - Local Coordinate System- Projections. **Linear Transformations:** Steps in Linear Transformation - Common Transformations. **Graph Theory:** Fundamentals of Graphs - Types of Graphs. **Algorithms:** Regression Algorithms - K-Nearest Neighbors (KNN) Algorithm - Clustering Algorithms - Artificial Neural Networks. **Machine Learning:** Categories of Machine Learning - Common Machine Learning Frameworks. **(18 Hours)**

UNIT II

Data Collection, Modelling, and Compilation: Data Collection - Cleaning and Organizing Data - Modeling Machine Learning. **Data Analysis:** Data Analysis Methods. **Data Presentation and Visualization**: Types of Data Presentation Frequency Distribution - Data Visualization in R. **Data Science Software Tools**: RapidMiner. **Programming Languages for Data Science**: R for Data Science - Python for Data Science. **Applications of Data Science. (18 Hours)**

UNIT III

Understanding Big Data: Concepts and Terminology - Big Data Characteristics - Different Types of Data. **Business Motivations and Drivers for Big Data Adoption:** Marketplace Dynamics - Business Architecture - Business Process Management - Information and Communications Technology - Internet of Everything (loE). **(18 Hours)**

UNIT IV

Big Data Adoption and Planning Considerations: Organization Prerequisites - Data Procurement – Privacy – Security – Provenance - Limited Realtime Support - Distinct Performance Challenges - Distinct Governance Requirements - Distinct Methodology – Clouds - Big Data Analytics Lifecycle. **Enterprise Technologies and Big Data Business Intelligence:** Online Transaction Processing (OLTP) - Online Analytical Processing (OLAP) – Extract Transform Load (ETL) - Data Warehouse's – Data Marts – Traditional Bl – Big Data Bl. **(18 Hours)**

UNIT V

Big Data Storage Concepts: Clusters – File Systems and Distributed File Systems – NoSQL – Sharding – Replication - Sharding and Replication – CAP Theorem – ACID – BASE. **Big Data Processing Concepts:** Parallel Data Processing - Distributed Data Processing – Hadoop – Processing Workloads – Cluster – Processing in Batch Mode - Processing in Realtime Mode. **(18 Hours)**

BOOKS FOR STUDY

1. **"Fundamentals of Data Science – Take the First Step to Become a Data Scientist",** Samuel Burns, Amazon KDP Printing and Publishing, First Edition, 2019.

   UNIT I : Chapters : 1 – 8

   UNIT II : Chapters : 9 – 14

2. **"Big Data Fundamentals – Concepts, Drivers & Techniques",** Thomas Erl, Wajid Khattak and Paul Buhler, Pearson India Education Pvt. Ltd., First Edition, 2016.

   UNIT III : Chapters : 1, 2

   UNIT IV : Chapters : 3, 4

   UNIT V : Chapters : 5, 6

BOOKS FOR REFERENCE

1. **"Fundamentals of Data Science",** Dr. B. Srikanth, Dr. Swarajya Lakshmi V Papineni and Dr. Syed Khasim, Walnut Publication, First Edition, March 2022.
2. **"Fundamentals of Data Science",** Dr. P. M. Ashok Kumar, Bluerose Publishers, First Edition, 2022.

DESCRIPTIVE STATISTICS AND PROBABILITY

Semester : I **Hours: 6**

Code : 25PGDDS02 **Credits: 4**

Upon completion of this course students will be able to

- Acquire the basic knowledge of statistics history and analyze statistical data graphically using frequency distributions.
- Understand the Measures of central tendency, Dispersion, Skewness and Kurtosis.
- Analyse the concept of correlation and regression for relating two or more related variables and impart the skills to develop Association of Attributes.

UNIT I

Introduction (Meaning and Scope): Origin and Development of Statistics – Statistics Defined - Importance and Scope of Statistics - Limitations of Statistics - Distrust of Statistics. **Descriptive Measures:** Introduction - Frequency Distributions - Graphic Representation of a Frequency Distribution. **(18 Hours)**

UNIT II

Descriptive Measures: Averages or Measures of Central Tendency - Arithmetic Mean - Median - Mode - Geometric Mean - Harmonic Mean - Selection of an Average - Partition Values - Dispersion - Measures of Dispersion - Coefficient of Dispersion - Moments - Skewness – Kurtosis. **(18 Hours)**

UNIT III

Theory of Probability: Introduction - Basic Terminology - Mathematical or Classical Probability - Statistical or Empirical Probability – Subjective Probability - Mathematical Tools - Axiomatic Approach to Probability – Some Theorems on Probability - Conditional Probability – Multiplication Theorem of Probability - Independent Events - Baye's Theorem. **(18 Hours)**

UNIT IV

Correlation: Meaning of Correlation - Scatter Diagram - Karl Pearson Coefficient of Correlation. **Curve Fitting and Regression Analysis:** Linear Regression – Curvilinear Regression - Regression Curves. **Multiple and Partial Correlation and Regression Analysis:** Multiple and Partial Correlation and Regression – Plane of Regression – Coefficient of Multiple Correlation – Coefficient of Partial Correlation. **(18 Hours)**

UNIT V

Theory of Attributes: Notations and Terminology - Classes and Class Frequencies - Consistency of Data – Independence of Attributes - Association of Attributes. **Large Sample Theory:** Types of Sampling – Parameter and Statistic – Test of Significance – Procedure for Testing of Hypothesis – Test of Significance for Large Samples - Sampling of Attributes – Sampling of Variables. **(18 Hours)**

BOOK FOR STUDY

1. **"Fundamentals of Mathematical Statistics",** S. C. Gupta and V. K. Kapoor, Sultan Chand & Sons, Twelfth Edition, New Delhi, 2020.

UNIT I **:** Chapters : 1 (1.1 - 1.5), 2 (2.1 – 2.3)

UNIT II **:** Chapter : 2 (2.5 – 2.17)

UNIT III **:** Chapters : 3 (3.1 – 3.12), 4 (4.2)

UNIT IV **:** Chapters : 10 (10.1 – 10.4), 11(11.1 – 11.4), 12 (12.1 - 12.3, 12.5, 12.6)

UNIT V **:** Chapters : 13, 14

BOOKS FOR REFERENCE

1. **"Basic Statistics",** Agarwal B.L, New Age International (P) Limited Publishers, Sixth Edition, 2018.
2. **"Applied Statistics and Probability for Engineers",** Montgomery D.C and Runger G.C, Wiley India, New Delhi, Seventh Edition, 2018.
3. **"Statistics",** S. Arumugam, A. Issac, New Gamma Publication House, Palayamkottai, 2013.

DATA ANALYSIS USING PYTHON

Semester : I **Hours:6**

Subject Code : 25PGDDS03 **Credits:4**

Upon completion of this course students will be able to

- Outline the need for Python programming language and understand basic programming concepts.
- Interpret Python syntax and semantics and be fluent in the use of Python programming statements.
- Implement data handling and visualization techniques using NumPy & Pandas modules

UNIT I

Introduction: Software Development - History of Python Programming Language - Thrust Areas of Python - Installing Anaconda Python Distribution - Installing PyCharm IDE to Set Up a Python Development Environment - Creating and Running Your First Python Project - Installing and Using Jupyter Notebook. **Parts of Python Programming Language:** Identifiers - Keywords – Statements and Expressions - Variables – Operators – Precedence and Associativity - Data Types - Indentation - Comments - Reading Input - Print Output - Type Conversions. **(18 Hours)**

UNIT II

Control Flow Statements: The *if* Decision Control Flow Statement- The *if…else* Decision Control Flow Statement - The *if…elif…else* Decision Control Statement - Nested *if* Statement - The *while* Loop - The *for* Loop - The *continue* and *break* Statements - Catching Exceptions Using try and except Statement. **Functions:** Built-In Functions - Commonly Used Modules - Function Definition and Calling the Function - The return Statement and void Function - Scope and Lifetime of Variables - Default Parameters - Keyword Arguments - *args and **kwargs - Command Line Arguments. **(18 Hours)**

UNIT III

Strings: Creating and Storing Strings - Basic String Operations - Accessing Characters in String by Index Number - String Slicing and Joining - String Methods - Formatting Strings. **Lists:** Creating Lists - Basic List Operations - Indexing and Slicing in Lists - Built-In Functions Used on Lists - List Methods - The

del Statement. **Dictionaries:** Creating Dictionary - Accessing and Modifying key:value Pairs in Dictionaries - Built-In Functions Used on Dictionaries - Dictionary Methods - The del Statement. **(18 Hours)**

UNIT IV

Tuples and Sets: Creating Tuples - Basic Tuple Operations - Indexing and Slicing in Tuples - Built-In Functions Used on Tuples - Relation between Tuples and Lists - Relation between Tuples and Dictionaries - Tuple Methods - Using *zip()* Function - Sets - Set Methods – Frozenset. **(18 Hours)**

UNIT V

Introduction to Data Science: Functional Programming - JSON and XML in Python - NumPy with Python – Pandas. **Visualization with Matplotlib:** General Matplotlib Tips - Two Interfaces for the Price of One - Simple Line Plots - Visualizing Errors - Density and Contour Plots - Histograms, Binnings, and Density, Customizing Matplotlib: Configurations and Stylesheets - Three-Dimensional Plotting in Matplotlib - Geographic Data with Basemap - Visualization with Seaborn. **(18 Hours)**

BOOKS FOR STUDY

1. **"Introduction to Python programming",** Gowrishankar S, Veena A, CRC Press, First Edition, 2019.

   UNIT I : Chapters : 1 (1.3 – 1.9), 2 (2.1 - 2.12)

   UNIT II : Chapters : 3, 4

   UNIT III : Chapters : 5 - 7

   UNIT IV : Chapter :8

   UNIT V : Chapter : 12

2. **"Python Data Science Handbook - Essential Tools for Working with Data",** JakeVanderPlas, O'Reilly Media, Inc., First Edition, 2017.

   UNIT V : Chapter : 4

BOOKS OF REFERENCE

1. **"Introduction to computing & Problem Solving with Python ",** Jeeva Jose, P. SojanLal, Khanna Book Publishing Co. (P) LTD., 2020.

2. **"Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython",** William McKinney, O'Reilly, Second Edition, 2017.

DATA ANALYSIS USING PYTHON - LAB

Semester: I **Hours: 6**

Code : 25PGDDSP1 **Credits: 3**

Upon completion of this course students will be able to

- Implement fundamental data structures like lists, tuples, and dictionaries, and use their associated methods for manipulation.
- Design and implement Python programs to address real-world challenges across various domains.
- Demonstrate proficiency in using Python libraries such as NumPy, Pandas, and Matplotlib for data analysis and visualization.

LIST OF PRACTICALS

1. Simple Python Programs
2. Program using Control Flow Statements
3. Program using functions
4. Create a list and perform the following methods
   i) insert() ii) remove() iii) append()
   iv) len() v) pop() vi) clear()
5. Create a dictionary and apply the following methods
   i)Print the dictionary items ii) access items
   iii) useget() iv) change values v)use len()
6. Create tuple and perform the following methods
   i) Add items ii) len()
   iii) check for item in tuple iv)Access items
7. Program using lambda(), filter() and map()
8. Program using the calendar module
9. Program using numpy module
10. Program to read a csv file using pandas and matplot modules
11. Import a CSV file and perform various Statistical and Comparison operations

DATA ANALYSIS USING SPREADSHEET – LAB

Semester : I **Hours: 6**

Code : 25PGDDSP2 **Credit: 3**

Upon completion of this course students will be able to

- Understand the basic features of spreadsheet software, including cell operations, formatting, and navigation.
- Apply data validation techniques to ensure accuracy and consistency in datasets.
- Visualize data using charts, graphs, and pivot tables to derive insights and present findings effectively.

LIST OF PRACTICALS

Introduction to Spreadsheets

1. Reading data into Excel using various formats
2. Basic functions in Excel (Arithmetic, Logical functions)
3. Using formulas in Excel and their copy and paste using absolute and relative referencing

Spreadsheet Functions to Organize Data

4. IF and the nested IF functions
5. VLOOKUP and HLOOKUP
6. The RANDBETWEEN function

Mathematical and Statistical Calculations

7. Measures of central tendency (mean, median, and mode) using Excel functions
8. Measures of dispersion (range, standard deviation, quartile deviation) using Excel functions
9. Correlation and Correlation matrix
10. Regression analysis and their significance
11. Calculation of probabilities of different events using probability properties

Introduction to Filtering and Pivot Tables

12. Data filtering in Excel
13. Use of Pivot tables with categorical as well as numerical data

Advanced Graphing and Charting

14. Line, Bar and Pie charts
15. Pivot charts
16. Scatter plots
17. Histograms

MACHINE LEARNING ALGORITHMS

Semester : II **Hours: 6**

Code : 25PGDDS04 **Credits: 4**

Upon completion of this course students will be able to

- Acquire the basic knowledge of a variety of Machine learning and its applications.
- Apply the Bayesian analysis with dimensionality reduction on real-time datasets.
- Analyze the dataset using regression algorithm and neural networks for learning linear and non-linear functions.

UNIT I

Introduction to Machine Learning: Types of Machine Learning **-** Application of Machine Learning - Hypothesis Space - Inductive Bias - Evaluation and Cross Validation. **(18 Hours)**

UNIT II

Basic Machine Learning Algorithms: Linear Regression - Decision Tree - Basic Decision Tree Learning Algorithm - K-nearest Neighbour - Collaborative Filtering - Overfitting. **(18 Hours)**

UNIT III

Dimensionality Reduction: Introduction of Dimensionality Reduction- Feature and Feature Engineering-Feature Transformation - Feature Subset Selection- **Bayesian Concept of Learning:** Importance of Bayesian Methods - Bayes Theorem – Bayes' Theorem and Concept Learning- Bayesian Belief Network. **(18 Hours)**

UNIT IV

Logistic Regression and Support Vector Machine - Logistic Regression - Introduction to Support Vector Machine - Kernel Methods for Non-linearity. **Basics of Neural Network:** Introduction to Neural Network- Biological Neurons – Architecture of Neural Network - Implementation of ANN - Backpropagation Algorithm - Deep Learning. **(18 Hours)**

UNIT V

Computation and Ensemble Learning: Introduction to Computation Learning - Sample Complexity: Finite Hypothesis Space - Introduction to Ensembles -

Basic Concepts of Clustering: Introduction to Clustering - Hierarchical Clustering- Agglomerative Hierarchical Clustering. **(18 Hours)**

BOOK FOR STUDY

1. **"Introduction to Machine Learning",** I. A. Dhotre, Technical Publications, First Edition, 2021.

UNIT I	:	Chapter	: 1
UNIT II	:	Chapter	: 2
UNIT III	:	Chapters	: 3, 4
UNIT IV	:	Chapters	: 5, 6
UNIT V	:	Chapters	: 7, 8

BOOKS FOR REFERENCE

1. **"Introduction to Machine Learning"**, E. Alpaydin, MIT Press, Third Edition, 2015.
2. **"Machine Learning: A Probabilistic Perspective",** K. P. Murphy, MIT Press, 2012.
3. **"Machine Learning in Action",** P. Harrington, Manning Publications, 2012.

DATA ANALYTICS USING R

Semester : II **Hours: 6**

Code : 25PGDDS05 **Credits: 4**

Upon completion of this course students will be able to

- Understand, Analyse, Interpret Correlation and Regression to analyse the underlying relationships between different variables.
- Apply analytical knowledge with the R interface and language for different fields.
- Cultivating cognitive skills acquired on existing data and perform all conventional statistical analysis tests using R knowledge on data management.

UNIT I:

Introduction to R: R and RStudio Environment - RStudio Environment – Four Windows in RStudio. **Basics of R:** Set Working Directory in R – Comment Statements in R – Variables in R – Data Types in R – Operators in R – Functions in R – Vectors in R – Lists in R – Data Frames in R – Packages in R. **(18 Hours)**

UNIT II

Exploratory Data Analysis: Steps in Data Preprocessing - Understanding Data - Looking at the Data – Visualizing Data – Dealing with Outliers - Dealing with Missing Values - Standardizing Data. **(18 Hours)**

UNIT III

Data Visualization: Introduction - Importance of Data Visualization - Data Visualization for Machine Learning - Data Visualization Techniques. **Simple Data Visualization Using R:** Generic Plot - Scatter Plot - Strip Chart - Stacked Bar Plot - Grouped Bar Plot - Pie Chart - Kernel Density Plot. **Data Visualization Using Ggplots inR:** Scatter Plot - Line Plot - Boxplot - Violin Plot - Ridge Plot.

(18 Hours)

UNIT IV

Dimensionality Reduction Techniques: Dimensionality Reduction – Independent and Dependent Variables. **Relationship between Variables: Correlation:** Application of Factor Analysis using R Programming – Multicollinearity**. Factor Analysis:** Eigen Value – Scree Plot – Unrotated Factor Matrix – Rotated Factor Matrix. **Unsupervised Learning Algorithms:** Introduction – **Association Rule Mining**: Transaction Dataset – Support – Confidence – Lift – Apriori Algorithm – Association Rule – Plotting of Rules.

Conjoint Analysis: Full and Fractional Factorial Design – Choice Cards – Attribute Importance. **(18 Hours)**

UNIT V :

Supervised Learning Algorithms: Decision Tree and Random Forest: Decision Tree – Tree Structure – Criteria for Splitting Decision Node. **Classification and Regression Technique:** Control Parameters – Pruning the Tree – Model Performance Measures – Insights from Decision Rules. **Random Forest:** Control Parameters – Out of Bag Error Rate – Tuning the Random Forest – Variable Importance Plot – Model Performance Measures. **Supervised Learning Algorithm: K-Nearest Neighbors:** Similarity Based on Distance Function – Select Appropriate K Value – KNN Model Building – Model Performance Measures. **Naive Bayes Algorithm:** Types of Naïve Bayes Theorem – Building Naïve Bayes Classifier – Model Performance Measures.

(18 Hours)

BOOK FOR STUDY

1. **"Introduction to Data Science Practical Approach with R and Python",** B. Uma Maheshwari and R. Sujatha, Wiley India Pvt. Ltd., First Edition, 2021.

   Unit I : Chapter : 2 (2.1 - 2.3)
   Unit II : Chapter : 3 (3.1 - 3.8)
   Unit III : Chapter : 4 (4.1 - 4.5)
   Unit IV : Chapters : 5 (5.1 - 5.5), 7 (7.1 - 7.3)
   Unit V : Chapters : 10 (10.1 - 10.3), 11 (11.1, 11.2)

BOOKS FOR REFERENCE

1. **"R for Everyone Advanced Analytics and Graphics",** Jared P. Lander, Pearson Education Inc., Second Impression, 2016.
2. **"R for Data Science",** Hadley Wickham and Garrett Grolemund, O'Reilly, First Edition, 2017.
3. **"Data Analysis and Graphics Using R - An Example - Based Approach**", John Maindonald & W. John Braun, Cambridge University Press, Third Edition, 2010.

INTERNET OF THINGS

Semester: II **Hours: 6**

Code : 25PGDDS06 **Credits: 4**

Upon completion of this course students will be able to

- Understand the fundamentals of the Internet of Things ecosystem.
- Design and develop IoT platform design Methodology.
- Apply things in the creation of smart world scenario.

UNIT I

Introduction to Internet of Things (IoT): Introduction and Definition of Internet of Things – IoT Growth – A Statistical View – Application Areas of IoT – Characteristics of IoT – Things in IoT – IoT Stack – Enabling Technologies – IoT Challenges – IoT Levels – Cyber Physical System versus IoT – Wireless Sensor Network versus IoT. **Introduction to Sensors, Microcontrollers and their Interfacing:** Introduction to Sensors Interfacing – Types of Sensors – Controlling Sensors through Webpages. **Microcontrollers**: A Quick Walkthrough – Advanced RISC Machine. **(18 Hours)**

UNIT II

PROTOCOLS FOR IOT: Messaging and Transport: Introduction – Messaging protocols – XMPP and DDS Protocols - Transport Protocols. **Protocols for IoT - Addressing and Identification:** Introduction – Internet Protocol version 4 (IPv4) – Internet Protocol Version 6 (IPv6) – IPv6- A Quick Overview – Internet Protocol Version 5 (IPv5) - Uniform Resource Identifier (URI). **(18 Hours)**

UNIT III

Cloud for IoT: Introduction – IoT with Cloud – Challenges – Selection of Cloud Service Provider: An Overview – Introduction to Fog Computing – Cloud Computing: Security Aspects. **(18 Hours)**

UNIT IV

DATA ANALYTICS - Visualising the power of Data from IoT: Introduction – Data Analysis – Machine Learning – Types of Machine Learning Models – Model Building Process – Modelling Algorithms – Model Performance - Big Data Platform – Big Data Pipeline – Real Life Projects – Recommendations in IoT Gadgets.

(18 Hours)

UNIT V

Application Building with IoT: Introduction – Smart Perishable Tracking with IoT and Sensors – Smart Healthcare-Elderly Fall Detection with IoT and Sensors – Smart

Inflight Lavatory Maintenance with IoT – IoT-Based Application to Monitor Water Quality – Smart Warehouse Monitoring - Smart Retail – IoT Possibilities in the Retail Sector – Prevention of Drowsiness of Drivers by IoT – Based Smart Driver Assistant Systems – System to Measure Collision Impact in an Accident with IoT – Integrated Vehicle Health Management. **(18 Hours)**

BOOK FOR STUDY

1. **"Internet of Things",** Shriram K Vasudevan, Abhishek S Nagarajan, RMD Sundaram", Second Edition, Wiley India Pvt. Ltd., 2020.

Unit I	:	Chapters	:	1, 2
Unit II	:	Chapters	:	3, 4
Unit III	:	Chapter	:	5
Unit IV	:	Chapter	:	6
Unit V	:	Chapter	:	7

BOOKS FOR REFERENCE

1. "**Internet of Things Architecture, Implementation and Security**", Mayur Ramgir, Pearson Education, First Impression, 2020.
2. **"IoT Fundamentals",** David Hanes, Gonzalo Salgueiro, Patrick Grossetete, Rob Barton, Jerome Henry, CisCo, Pearson India Education Services Pvt. Ltd, 2018.
3. **"Internet of Things: A Hands-On Approach",** Arshdeep Bahga, Vijay Madisetti, Universities Press, 2015.

DATA ANALYTICS USING R - LAB

Semester : II **Hours: 6**

Code : 25PGDDSP3 **Credits: 3**

Upon completion of this course students will be able to

- Understand the basic syntax and programming concepts in R, including data types, variables, and operators.
- Visualize data using R libraries such as ggplot2, creating bar charts, scatter plots, histograms, and line graphs.
- Solve real-world problems by working with structured and unstructured datasets in R.

LIST OF PRACTICALS

1. Demonstrate the usage of Numbers and Vectors in R.
2. Implement the concepts of Arrays and Matrices
3. Implement the concept of grouping and conditional execution on Data Framesand Lists
4. Import CSV file into Data Frame in R and perform data manipulations.
5. Use a Dataset to handle
    a. the Categorical and numerical data
    b. the Bi-variate categorical data
    c. the Multivariate categorical data
6. Demonstrate the usage of apply() functions.
7. Implement the usage of dplyr package
8. Utilize a lattice and ggplot2 packages to plot 1D, 2D and 3D plots for a givendataset.
9. Demonstrate correlation and rank correlation.
10. Create a data set and do classification & association algorithms to analyse the data using R.

INTERNET OF THINGS - LAB

Semester : II **Hours: 6**

Code : 25PGDDSP4 **Credits: 3**

Upon completion of this course students will be able to

- Understand the basic concepts and architecture of IoT, including hardware components, sensors, and actuators.
- Establish communication between IoT devices and cloud services using protocols such as MQTT and HTTP
- Develop end-to-end IoT applications that integrate sensing, processing, and communication.

LIST OF PRACTICALS

1. Familiarization with Arduino/Raspberry Pi and perform necessary softwareinstallation.
2. IoT program to turn ON/OFF LED light (3.3V)
3. Smart light using LED/Buzzer with Arduino to turn ON LED for 1 sec after every 2 seconds.
4. IoT program using IR sensor
5. IoT program using Humidity and Temperature Sensor (Weather Monitoring)
6. IoT program using Soil Moisture Sensor
7. IoT program using Ultrasonic Sensor (Distance Measurement)
8. Real-time IoT program using Relay Module (Smart Home Automation with 230V)
9. IoT program for Fire Detection (Home, Industry, etc.)
10. IoT program for Gas Leakage detection (Home, Industry, etc.)
11. IoT program using Heartbeat Sensor
12. IoT web server program for local hosting
13. To Write an IoT Program to upload and retrieve temperature and humidity data using Thingspeak cloud.