# Department of Computer Science (UG)

Minutes of the meeting of Board of Studies in Computer Science (UG) held on 12-09-2022 at 11:00 AM for B. Sc (MSDS, AIML) programmes.

## LIST OF BOS MEMBERS

| Name of the Member | Role |
|---|---|
| Dr. G. Krishna Mohan, HOD, Dept. of CS, P.B. Siddhartha College of Arts & Science. Mobile: 9440446847, Email: gvlkm@pbsiddhartha.ac.in | Chairman |
| Dr. R. Kiran Kumar, Associate Professor, Department of Computer Science, Krishna University, Machilipatnam. Mobile: 9440872455 | Nominee, Krishna University |
| Dr. Yogesh Kumar Meena<br>Associate Professor, Department of CSE, MNIT Jaipur.<br>Mobile: 7891005056 Email : ymeena.cse@mnit.ac.in | Subject Expert |
| Sri. Prashant R. Nair<br>Associate Professor, Vice- Chairman- IQAC, Dept. of CSE,<br>Amrita Viswa Vidyapeetham, Coimbatore. | Subject Expert |
| Ms. ReemaThareja<br>Professor, Dept. of Computer Science, Shyama Prasad Mukherji College (W), University of Delhi. reemathareja@gmail.com | Subject Expert |
| Bharat Kumar Reddy Gujavarti (MCA, PGDHRM), Hyderabad<br>Founder & CEO, Pragmatiq Systems Inc. Director, Sunblue Technologies Co-founder, Edify Email:bharat@pragmatiq.inMobile: 8978191977 | Industrialist |
| Shankar Lakkaraju, Director, Software Engineer Lowe's India, Bangalore<br>Email: shankar.lakkaraju@gmail.com Mobile: 98851 65651 | Alumni Representative MCA: 1999-2002 |
| Dr. T.S. Ravi Kiran | Member |
| Mr. K. Sridhar | Member |
| Mr. K. Sudhir | Member |
| Dr .K. Udaya Sri | Member |
| Mrs. M. Suneela | Member |

@@@@

| DEPARTMENT OF COMPUTER SCIENCE (UG) | | | | | | |
|---|---|---|---|---|---|---|
| **LIST OF COURSES INTRODUCED IN V SEMESTER 2022 - 2023** | | | | | | |
| **S. No.** | **Title of the Course** | **Course Code** | **Offered in Sem** | **Year of Introduction** | **OBE with BTL** | **Offered to** |
| 1 | Introduction to Machine Learning | DSCSET01 | V | 2022 - 2023 | YES | B. Sc. (MSDS) |
| 2 | Introduction to Machine Learning Lab | DSCSEP01 | V | 2022 - 2023 | YES | B. Sc. (MSDS) |
| 3 | Big Data Technology | DSCSET02 | V | 2022 - 2023 | YES | B. Sc. (MSDS) |
| 4 | Big Data Technology Lab | DSCSEP02 | V | 2022 - 2023 | YES | B. Sc. (MSDS) |
| 5 | Data Mining and Data Analysis | DSCSET03 | V | 2022 - 2023 | YES | B. Sc. (MSDS) |
| 6 | Data Mining and Data Analysis Lab | DSCSEP03 | V | 2022 - 2023 | YES | B. Sc. (MSDS) |
| 7 | Multivariate Technique for Data Analysis | DSCSET04 | V | 2022 - 2023 | YES | B. Sc. (MSDS) |
| 8 | Multivariate Technique for Data Analysis Lab | DSCSEP04 | V | 2022 - 2023 | YES | B. Sc. (MSDS) |
| 9 | Data & Information Security through Python | DSCSET05 | V | 2022 - 2023 | YES | B. Sc. (MSDS) |
| 10 | Data &Information Security through Python Lab | DSCSEP05 | V | 2022 - 2023 | YES | B. Sc. (MSDS) |
| 11 | Spark Programming | DSCSET06 | V | 2022 - 2023 | YES | B. Sc. (MSDS) |
| 12 | Spark Programming Lab | DSCSEP06 | V | 2022 - 2023 | YES | B. Sc. (MSDS) |
| 13 | Introduction to NumPy and Pandas | SDCCSCP | III | 2022 – 2023 | YES | B. Sc. (MSDS) |
| 14 | Foundation of Data science using R | DSCT31B | III | 2022 – 2023 | YES | B. Sc. (MSDS) |
| 15 | Foundation of Data science using R Lab | DSCP31B | III | 2022 – 2023 | YES | B. Sc. (MSDS) |
| 16 | Introduction to Artificial Intelligence | AIMLT31 | III | 2022 - 2023 | YES | B. Sc. (AIML) |
| 17 | Introduction to Artificial Intelligence Lab | AIMLP31 | III | 2022 - 2023 | YES | B. Sc. (AIML) |
| 18 | Document Oriented Database | AIMLT32 | III | 2022 - 2023 | YES | B. Sc. (AIML) |
| 19 | Document Oriented Database Lab | AIMLP32 | III | 2022 - 2023 | YES | B. Sc. (AIML) |

**Resolutions:**

1. It is resolved and recommend to introduce “Introduction to Machine Learning” with course code DSCSET01 in V semester of B. Sc. (MSDS) for the batch of students admitted in 2020-21 and onwards. For the syllabus and model question paper vide page number from 6 to 8.
2. It is resolved and recommend to introduce “Introduction to Machine Learning Lab” with course code DSCSEP01 in V semester of B. Sc. (MSDS) for the batch of students admitted in 2020-21 and onwards. For the syllabus and model question paper vide page number from 9 to 10 .
3. It is resolved and recommend to introduce “Big data Technology” with course code DSCSET02 in V semester of B. Sc. (MSDS) for the batch of students admitted in 2020-21 and onwards. For the syllabus and model question paper vide page number from 11 to 13.
4. It is resolved and recommend to introduce “Big data Technology Lab” with course code DSCSEP02 in V semester of B. Sc. (MSDS) for the batch of students admitted in 2020-21 and onwards. For the syllabus and model question paper vide page number from 14 to 18.
5. It is resolved and recommend to introduce “Data Mining and Data Analysis” with course code DSCSET03 in V semester of B. Sc. (MSDS) for the batch of students admitted in 2020-21 and onwards. For the syllabus and model question paper vide page number from 19 to 22.
6. It is resolved and recommend to introduce “Data Mining and Data Analysis Lab” with course code DSCSEP03 in V semester of B. Sc. (MSDS) for the batch of students admitted in 2020-21 and onwards. For the syllabus and model question paper vide page number from 23 to 24 .
7. It is resolved and recommend to introduce “Multivariate Techniques for Data Analysis” with course code DSCSET04 in V semester of B. Sc. (MSDS) for the batch of students admitted in 2020-21 and onwards. For the syllabus and model question paper vide page number from 25 to 28 .
8. It is resolved and recommend to introduce “Multivariate Techniques for Data Analysis Lab” with course code DSCSEP04 in V semester of B. Sc. (MSDS) for the batch of students admitted in 2020-21 and onwards. For the syllabus and model question paper vide page number from 29 to 30 .

9. It is resolved and recommend to introduce “Data and Information security through Python” with course code DSCSET05 in V semester of B. Sc. (MSDS) for the batch of students admitted in 2020-21 and onwards. For the syllabus and model question paper vide page number from 31 to 33.
10. It is resolved and recommend to introduce “Data and Information Security through Python Lab” with course code DSCSEP05 in V semester of B. Sc. (MSDS) for the batch of students admitted in 2020-21 and onwards. For the syllabus and model question paper vide page number from 34 to 35.
11. It is resolved and recommend to introduce “Spark Programming” with course code DSCSET06 in V semester of B. Sc. (MSDS) for the batch of students admitted in 2020-21 and onwards. For the syllabus and model question paper vide page number from 36 to 39.
12. It is resolved and recommend to introduce “Spark Programming Lab” with course code DSCSEP06 in V semester of B. Sc. (MSDS) for the batch of students admitted in 2020-21 and onwards. For the syllabus and model question paper vide page number from 40 to 41.
13. It is resolved and recommend to introduce “Introduction to NumPy Pandas Lab” with course code SDCCSCP in III semester of B. Sc. (MSDS) for the batch of students admitted in 2021-22 and onwards. For the syllabus and model question paper vide page number from 42 to 43 .
14. It is resolved and recommend to introduce “Foundation of Data Science using R” with course code DSCT31B in III semester of B. Sc. (MSDS) for the batch of students admitted in 2021-22 and onwards. For the syllabus and model question paper vide page number from 44 to 46.
15. It is resolved and recommend to introduce “Foundation of Data Science using RLab” with course code DSCP31B in III semester of B. Sc. (MSDS) for the batch of students admitted in 2021-22 and onwards. For the syllabus and model question paper vide page number 47.
16. It is resolved and recommend to introduce “Introduction to Artificial Intelligence” with course code AIMLT31 in III semester of B. Sc. (AIML) for the batch of students admitted in 2021-22 and onwards. For the syllabus and model question paper vide page number from 48 to 51.

17. It is resolved and recommend to introduce “Introduction to Artificial Intelligence Lab” with course code AIMLT31 in III semester of B. Sc. (AIML) for the batch of students admitted in 2021-22 and onwards. For the syllabus and model question paper vide page number from 52 to 53.
18. It is resolved and recommend to introduce “Document Oriented Database” with course code AIMLT32 in III semester of B. Sc. (AIML) for the batch of students admitted in 2021-22 and onwards. For the syllabus and model question paper vide page number from 54 to 57.
19. It is resolved and recommend to introduce “Document Oriented Database Lab” with course code AIMLP32 in III semester of B. Sc. (AIML) for the batch of students admitted in 2021-22 and onwards. For the syllabus and model question paper vide page number 58.

# P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE

Siddharthanagar, Vijayawada – 520010

*(An Autonomous College under the Jurisdiction of Krishna University)*

*Re– Accredited at 'A$^{+}$' by NAAC–III Cycle*

*College with Potential for Excellence*

*ISO9001 – 2015 Certified*

Course Code: **DSCSET01** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

Max. Marks: **100** (CCIA: 25+ SEE:75) Theory Hrs./Week: 3

## 6A: Introduction to Machine Learning

Type of the Course: Core Credits: **03**

**I. Course Outcomes: Students at the successful completion of the course will be able to:**

1. **Identify** the characteristics of machine learning. **(PO5,PO7)**
2. **Summarize** the Model building and evaluation approaches. **(PO5,PO7)**
3. **Apply** Bayesian learning and regression algorithms for real-world Problems **(PO5, PO6, PO7).**
4. **Apply** supervised learning algorithms to solve real-world Problems. **(CO5, PO7)**.
5. **Apply** unsupervised learning algorithms for the real world data. **(PO5, PO7).**

**II. Syllabus: (Total Theory periods: 45)**

**Unit – 1**

**1. Introduction to Machine Learning and Preparing Model**

**1.1 Introduction to Machine Learning-**

1.1.1 Introduction
1.1.2 What is Human Learning?
1.1.3 Types of Human Learning 1.1.4What is Machine Learning?
1.1.5 Types of Machine Learning,
1.1.6 Problems not to be solved using Machine Learning
1.1.7 Applications of Machine Learning

**1.2 Preparing to Model**

1.2.1 Introduction, Machine Learning Activities
1.2.2 Basic Types of Data in Machine Learning
1.2.3 Exploring Structure of Data
1.2.4 Data Quality and Remediation
1.2.5 Data Pre-Processing

**Unit – II**

**2. Modeling & Evaluation, Basics of Feature Engineering**

**2.1 Modeling & Evaluation**

2.1.1 Introduction,
2.1.2 Selecting a Model,
2.1.3 Training a Model (for Supervised Learning),
2.1.4 Model Representation and Interpretability, Evaluating Performance of a Model.

**2.2 Basics of Feature Engineering**

2.2.1 Introduction,

2.2.2 Feature Transformation,

2.2.3 Feature Subset Selection.

**Unit- III**

**3. Supervised Learning Regression**

**3.1** Introduction

**3.2** Example of Regression

**3.3** Common Regression Algorithms

**3.4** Simple linear Regression, Multiple linear Regression

**3.5** Assumptions in Regression Analysis

**3.6** Main Problems in Regression, Analysis

**3.7** Improving Accuracy of the linear Regression Model

**Unit – IV**

**4. Classification**

**4.1** Naive Bayes model

**4.2** Decision Tree, Tree Terminology

**4.3** Decision Tree learning

**4.4** Decision Boundaries

**4.5** Random Forest

**4.6** Logistic Regression.

**Unit – V**

**5. Unsupervised Learning**

**5.1** Introduction

**5.2** Unsupervised Vs Supervised Learning

**5.3** Applications of Unsupervised Learning

**5.4** Clustering, Clustering as a machine learning task

**5.5** Different types of clustering techniques: K- Means, Density

**5.6** Hierarchical clustering

**5.7** Finding Pattern using Association Rule, Definition of common terms, Association rule

**5.8** The Apriori algorithm for association rule learning, Build the Apriori Principle Rule

**5.9** Introduction to Deep Learning

**III. Text Books:**


1. Machine Learning, 1st Edition by Subramanian Chandramouli, Saikat Dutt, Amit Kumar Das, Pearson Education India.
2. Machine Learning with Python by Abhishek Vijayagria – BPB Publications.


**Reference Books:**


1. Machine Learning, 1997 by Tom M. Mitchell, MGH Publisher
2. Pattern Recognition & Machine Learning , 2006 by Christopher M. Bishop, Newyork Springer Publisher.
3. Understanding Machine Learning : From Theory to Algorithms by Shai S ShaiShalev-Shwartz, ShaiBen David , Cambridge Publisher.
4. Machine Learning in Action,1st edition, 2012 by Peter Harington , Cengage Publisher
5. Introduction to artificial neural systems by J. Zurada, St. Paul , West Publisher.
6. Introduction to machine learning, 2nd Edition by Ethem Alpaydin , The MIT Press


@@@@

# P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE

Siddharthanagar, Vijayawada – 520010

*(An Autonomous College under the Jurisdiction of Krishna University)*

*Re– Accredited at 'A$^{+}$' by NAAC–III Cycle*

*College with Potential for Excellence*

*ISO9001 – 2015 Certified*

Course Code: **DSCSET01** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

## 6A: Introduction to Machine Learning

## SEMESTER END MODEL QUESTON PAPER

**Time: 3 Hrs.** **Max.: 75 Marks** **Min. Pass: 30 Marks**

### SECTION –A

**ANSWER ANY FIVE QUESTIONS** **5 X 5 =25 M.**

1. What is Machine Learning? Briefly explain the types of Machine Learning. {CO1, L2}
2. Differentiate Dimensionality reduction and Feature Selection {CO1, L2}
3. Explain the process of k-fold cross-validation. {CO2, L2}
4. When can a feature be termed as irrelevant? How can it be measured? {CO2 ,L2}
5. What is multiple linear regression? {CO3 ,L2}
6. What are the conditions of a positive and a negative slope in linear regression?{CO3, L2}
7. What is entropy in a decision tree? {CO4, L2}
8. Mention few application areas of unsupervised learning. {CO5, L2}

### SECTION – B

**ANSWER ALL THE QUESTIONS** **5 X 10 =50 M**

9. (A) What is Machine Learning? Explain different perspective and issues in machine learning. {CO1, L2}

   (OR)

   (B) Explain the quantitative and qualitative data in detail. Differentiate between the two. {CO1, L2}

10. (A) Explain, in detail, the process of evaluating the performance of a classification model. {CO2, L2}

(OR)

(B) What are the different techniques for data pre-processing? Explain the filter approach of feature selection. How is the different from wrapper approach? {CO2, L2}

11. (A)Define simple linear regression using a graph explaining slope and intercept?{CO3, L2}

(OR)

(B) What is regression? Explain OLS algorithm with formula for 'a' and 'b'. {CO3, L2}

12. (A)Explain the K – nearest neighbor algorithm. {CO4, L2}

(OR)

(B) What is a decision tree? Discuss the use of decision tree for classification purpose with an example. {CO4, L2}

13. (A) Define Clustering. Explain different types of clustering techniques? {CO5, L2}

(OR)

(B) Explain the Apriori algorithm with suitable example. {CO5, L2}

@@@@

# P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

Course Code: **DSCSET01** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

## 6A: Introduction to Machine Learning Lab

Max. Marks: **50** (CCIA: 10+ SEE: 40) Practical Hrs./Week **: 2**

Type of the Course: **Core Lab** Credits: 02

**I. Course Outcomes: Students at the successful completion of the course will be able to:**

CO1: Apply the appropriate pre-processing techniques on the data set. (PO5, PO7)

CO2: Implement supervised Machine Learning algorithms. (PO5, PO7)

CO3: Implement unsupervised Machine Learning algorithms (PO5, PO7)

**II: Practical (Laboratory) Syllabus: (30 Periods)**

1. Demonstrate the packages Numpy, Pandas, Matplotlib/Seaborn.
2. Scikit Learn Package – Loading, basics of Scikit learn.
3. How to import data, basic operations on data frame: info, shape, head, data types, describe.
4. Checking missing values, Outliers, Unique value, Dropping/Adding Columns, Renaming the Columns.
5. EDA and Data Visualization – Numeric Variables
6. EDA and Data Visualization - Objective Variable
7. Scikit Learn – Splitting data, creating independent and dependent variables
8. Model Building Scikit Learn – Supervised - Classification
9. Model Building Scikit Learn – Supervised - Regression
10. Model Building Scikit Learn – Unsupervised - Clustering

**WEB REFERENCES:**

- **https://towardsdatascience.com/exploratory-data-analysis-eda-python-87178e35b14https://www.analyticsvidhya.com/blog/2020/08/exploratory-data-analysiseda-from-scratch-in-python/**
- **https://www.analyticsvidhya.com/blog/2020/10/feature-selection-techniques-in-machine-learning/**
- https://machinelearningmastery.com/principal-components-analysis-for-dimensionality-reduction-in-python/
- **https://towardsdatascience.com/dimension-reduction-techniques-with-python-f36ca7009e5chttps://www.analyticsvidhya.com/blog/2021/05/machine-learning-model-evaluation/**

# P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

Course Code: **DSCSET01** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

**6A: Introduction to Machine Learning Lab**

Max. Marks: **50** (CCIA: 10+ SEE: 40) Practical Hrs./Week **: 2**

Type of the Course: **Core Lab** Credits: 02

**Model Paper: Practical**

**Time: Three hours** **Max. Marks: 40**

**Section A**

One Major Experiment (Experiment No : ) **15 M**

**Section B**

One Minor Experiment (Experiment No : ) **10 M**

**Section C**

Practical Record **05 M**

**Section D**

Viva Voce **10 M**

***

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**

**Siddharthanagar, Vijayawada – 520010**

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

Course Code: **DSCSET02** | Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** | Semester – **V**

Type of the Course: Core | Credits: **03**

Max. Marks: **100** (CCIA: 25+ SEE: 75) | Theory Hrs. /Week: 3

## 7A: BIG DATA TECHNOLOGY

**Course Objective:**

1. To Understand big data, types of big data and applications in big data
2. To understand the map reduce concepts and scaling map reduce word count program
3. To understand the Concepts of Hadoop Ecosystem
4. To apply file commands in HDFS and concepts of mapper and reducer
5. To analyse Hive queries and concepts of YARN

**Course Outcomes:** At the end of this course the student is able

CO1: Recognize and understand use and applications of big data and analytics. (PO,PO7)

CO2: Learn how to apply Map reduce. (PO1, PO7)

CO3: Understand Hadoop ecosystem components. (PO1, PO7)

CO4: Learn to build and maintain reliable, scalable, distributed systems with Apache Hadoop. (PO1,PO7)

CO5: Able to use Hive as an interface to access data in Hadoop. (PO1, PO7)

**II. Syllabus:** **(Total Theory periods: 45)**

**Unit – 1**

**Introduction to Big data:** What is Big Data, Structuring Big Data-Types of Big Data, Elements of big data- Volume, Velocity, Variety, Veracity, Big Data Analytics-Advantages of Big Data Analytics, Big Data Applications.

**Unit – II**

**Introduction to Hadoop:** What is Hadoop, Understanding distributed systems & Hadoop, Comparing SQL databases and Hadoop, Understanding Map Reduce- scaling word count program manually, scaling word count program in Mapreduce

**Unit – III**

Hadoop Eco System, HDFS-HDFS Architecture, concept of blocks in HDFS- name node, data node, secondary name node, job tracker, task tracker). Introducing HBase-HBase architecture, Regions, storing Big Data with HBase, Why hive, pig, scoop, zookeeper, flume, oozie.

**Unit – IV**

**Working with files in HDFS**-Basic file commands, reading & writing to HDFS programmatically, Anatomy of Map Reduce program-Hadoop data types, Mapper, Reducer, Partitioner, Combiner, word counting with pre-defined mapper and Reducer, Reading & Writing-input format, output format.

**Unit – V**

Background of YARN, limitations of map reduce, advantages of YARN, YARN architecture, working of YARN. Introducing Hive, Hive Services, Hive Variables, Hive Queries, Data types, Hive Built in functions, Hive - DDL, DML, and Data Retrieval Queries.

**Text Books:**

1. Black Book "BIG DATA (covers hadoop2, map reduce, Hive, Yarn, Pig, R and Data Visualization)". Dream Tech Press. **(Units – 1,3,5)**
2. Chuck Lam, "Hadoop in Action", Dream Tech Press**(Units – 2,4)**

**Reference Books:**


1. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, "**Professional Hadoop Solutions**", Wiley,2015.
2. Chris Eaton, Dirk deroos, "**Understanding Big data**", McGraw Hill,2012


**Course Delivery method:** Face-to-face / Blended **Course has focus on:** Skill Development, Employability **Website of Interest:**

https://www.tutorialspoint.com/big_data_tutorials.htm

**P. B. SIDDHARTHACOLLEGE OF ARTS& SCIENCE**

**Siddharthanagar, Vijayawada – 520010**

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

Course Code: **DSCSET02** | Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** | Semester – **V**

Type of the Course: Core | Credits: **03**

Max. Marks: **75** | Time: 3 Hrs.

**7A: BIG DATA TECHNOLOGY**

**SEMESTER END MODEL QUESTON PAPER**

**Section-A**

**ANSWER ANY FIVE QUESTIONS** **5x5M=25M**

1. Write a short note on elements of Big Data.(CO1,L5)
2. Write down the advantages of Big Data Analytics.(CO1,L5)
3. Differentiate between SQL database and Hadoop.(CO2,L2)
4. List and explain the components of Hadoop(CO2,L2)
5. Write a short on Hadoop Ecosystem(CO3,L5)
6. Explain data types in Hadoop.(CO4,L2)
7. Explain advantages of YARN over Map Reduce.(CO5,L2)
8. List out the data types present in hive.(CO5,L1)

**Section-B**

**ANSWER THE FOLLOWING QUESTIONS** **5x10M=50M**

9. (A) Define Big Data? Explain different types of Big Data.(CO1,L2)

   OR

   (B) Discuss the applications of Big Data in detail. (CO1,L4)
10. (A) Define Hadoop? Explain the history of Hadoop in detail.(CO2,L2)

    OR

    (B) Explain Map Reduce word count program with example. (CO2,L2)
11. (A) Define HDFS? Explain different blocks of HDFS in detail.(CO3,L2)

    OR

    (B) Define HBase? Explain the architecture of HBase. (CO3,L2)
12. (A) Explain the following (CO4,L2)

    i) Mapper 5M

    ii) Reducer 5M

    OR

    (B) Illustrate basic file commands in HDFS with examples (CO4,L3)
13. (A) Describe YARN architecture and Working with YARN.(CO5,L2)

    OR

    (B) Explain hive DDL commands with examples. (CO5,L2)

@@@@

# P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

Course Code: **DSCSEP02** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

Max. Marks: **50** (CCIA: 10+ SEE: 40) Practical Hrs./Week **: 3**

Type of the Course: **Skill Enhancement Course** (Elective, Practical) Credits: 02

## Course 7A : BIG DATA TECHNOLOGY LAB

**Experiment – 1:**

Perform setting up and Installing Hadoop in its three operating modes:

- Standalone, □ Pseudo distributed, □ Fully distributed.

**Experiment – 2:**

Implement the file management tasks in Hadoop.

**Experiment – 3:**

Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm.

**Experiment – 4:**

Create a database with name "CustInfo".
Create a customers table with the
following fields Customer Id int
CustomerName String
CustomerAge int
Address String

Perform the following operations on customers table:

1. Describe the customers table
2. Load data of ten customers
3. Alter table name customers to customer details
4. Display the details of all customers
5. Add column name gender to table

Experiment – 5:

Create a database with name "college".
Create a student table with the
following fields Student Id int
Student Name String
Student Age int
Course Name String
Fee double
City String

| | |
|---|---|
| State | String |
| Pincode | int |

Perform the following operations on studenttable:

1. Describe the student table
2. Load data of six students
3. Alter table name to “Student Details”
4. Add column name **year of joining** to table
5. Display the details of all students

Experiment – 6:

Create a database with name “company”.

Create an employee table with the following fields

| | |
|---|---|
| Employee Id | int |
| EmployeeName | String |
| Designation | String |
| Gender | String |
| Salary | double |
| Address | String |
| City | String |
| State | String |
| Pincode | int |

Perform the following operations on employeetable:

1. Describe the employee table
2. Load data of eight employees
3. Alter table name to “EmpDet”
4. Add column name **department** to table
5. Alter the column name address to comaddr
6. Display the details of all employees
7. Add the salary of al lemployees

Experiment – 7:

(A) Create a database with name “retail”. Create a categories table with the following fields

| | |
|---|---|
| CategoryId | int |
| CategoryName | String |

Perform the following operations on student table:

1. Describe the categories table
2. Load data of five categories
3. Alter table name to “Category Details”
4. Display the details of all categories

(B) Create a products table with the following fields

| | |
|---|---|
| CategoryId | int |
| ProductId | int |
| ProductName | String |
| ProductDescription | String |
| ProductPrice | String |
| ProductImage | String |

Perform the following operations on student table:

1. Describe the product stable
2. Load data of eight products
3. Alter table name to “Product Details”
4. Display the details of all products
5. Display the details of products present in second category with category description

Experiment – 8:

Use the database “college”. Create a book information table with the following fields

| | |
|---|---|
| Book Id | int |
| Book ISBNNumber | String |
| BookTitle | String |
| AuthorofBook | String |
| YearofPublication | String |
| PublisherofBook | String |
| EditionofBook | String |
| Book Image | String |
| Priceof Book | int |

Perform the following operations on student table:

1. Describe the book information table
2. Load data of ten books
3. Alter table name to “Book Info”
4. Display the details of all books
5. Display the details of books of specific publisher

Experiment – 9:

Create a database with name “online”.

(A) Create an items table with the following fields

| | |
|---|---|
| ItemId | int |
| ItemName | String |
| ItemDescription | String |
| ItemPrice | String |
| Item Image | String |
| Quantity | int |

Perform the following operations on student table:

1. Describe the items table
2. Load data of ten items
3. Alter table name to “Item Details”
4. Display the details of all items

(B) Create an orders table with the following fields

| | |
|---|---|
| Order Id | int |
| OrderName | String |
| OrderDate | String |
| Amount | double |
| ShippingAddress | String |
| ShippingDate | String |

Perform the following operations on student table:

1. Describe the order stable
2. Load data of five orders
3. Alter table name to “Order Details”
4. Display the details of all orders

## Experiment – 10:

Create a database with name “university”. Create a staff table with the following fields StaffId int

| | |
|---|---|
| StaffName | String |
| Designation | String |
| DepartmentName | String |
| DateofJoining | String |
| Gender | String |
| Basic Salary | double |
| CommunicationAddress | String |
| City | String |
| State | String |
| Pincode | int |
| MobileNumber | String |

Perform the following operations on employee table:

1. Describe the staff table
2. Load data of ten staff members
3. Alter table name to “Staff Details”
4. Add column name **Qualification** to table
5. Get the data of staff members working in Computer Science in department

**********

# P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

Course Code: **DSCSEP02** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

Max. Marks: **50** (CCIA: 10+ SEE: 40) Practical Hrs./Week **: 3**

Type of the Course: **Skill Enhancement Course** (Elective, Practical) Credits: 02

**Title : BIG DATA TECHNOLOGY LAB**

**Model Paper: Practical**

**Time: Three hours** **Max. Marks: 40**

**Section A**

One Major Experiment (Experiment No : ) **15 M**

**Section B**

One Minor Experiment (Experiment No : ) **10 M**

**Section C**

Practical Record **05 M**

**Section D**

Viva Voce **10 M**

***

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

Course Code: **DSCSET03** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

Max. Marks: **100** (CCIA: 25+ SEE: 75) Theory Hrs./Week: 3

**DATA MINING AND DATA ANALYSIS**

Type of the Course: **Skill Enhancement Course** (Elective Theory) Credits: **03**

**I. Course Outcomes: Students at the successful completion of the course will be able to:**

CO1: Understand various Data Mining Processes & Techniques.

CO2: Characterize the various kinds of patterns that can be discovered by association rule mining.

CO3: Evaluate mathematical methods underlying the effective application of data mining.

CO4: Analyse different Classification Techniques

CO5: Apply different clustering models & methods

**II. Syllabus:** **(Total Theory periods: 45)**

Unit-I

1.1 Data mining
1.2 KDD versus data mining
1.3 Stages of the Data Mining Process
1.4 Task primitives
1.5 Data Mining Techniques
1.6 Data mining knowledge representation

Unit-II

2.1 Data mining query languages
2.2 Integration of Data Mining System with a Data Warehouse- Issues
2.3 Data pre-processing
2.4 Data Cleaning.
2.5 Data transformation
2.6 Feature selection
2.7 Dimensionality reduction
2.8 Discretization and generating concept hierarchies
2.9 Mining frequent patterns association
2.10 Correlation.

Unit-III

3.1 Classification
3.1.1 Basic Concepts
3.1.2General Approach to solving a classification problem
3.2 Decision Tree Induction

3.2.1 Working of Decision Tree
3.2.2 Building a decision tree
3.2.3 Methods for expressing an attribute test conditions
3.2.4 Measures for selecting the best split
3.2.5 Algorithm for decision tree induction.

**3.3** Model Over fitting

3.3.1Due to presence of noise
3.3.2 Due to lack of representation samples

3.4 Evaluating the performance of classifier

3.4.1 Holdout method
3.4.2 Random sub sampling
3.4.3 Cross-validation
3.4.4Bootstrap

**Unit-IV**

4.1 Bayesian Classification
4.2 Rule Based Classification
4.3 Classification by back propagation
4.4 Support Vector Machines
4.5 Associative Classification
4.6 Lazy Learners
4.7 Other Classification Methods

Unit-V

5.1 Clustering techniques
5.2 Partitioning methods-k-means
5.3 Hierarchical Methods
5.4 Distance based agglomerative and divisible clustering
5.5 Density
5.6 Based Methods
5.7 Expectation maximization
5.8 Grid Based Methods
5.9 Model
5.10 Based Clustering
5.11 Methods
5.12 Constraint
5.13 Based Cluster Analysis
5.14 Outlier Analysis.

**III. Text Books:**

1. A delchi Azzalini, Bruno Scapa, "Data Analysis and Data mining" , 2nd Ediiton, Oxford University Press Inc., 2012

**Reference Books:**

1. Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques", 3rdEdition, Morgan Kaufmann Publishers, 2011.
2. Alex Berson and Stephen J. Smith, "Data Warehousing, Data Mining & OLAP", 10thEdition, Tata Mc Graw Hill Edition , 2007.
3. G.K. Gupta, "Introduction to Data Mining with Case Studies", 1st Edition, Easter Economy Edition, PHI, 2006.

**IV. RECOMMENDED CO-CURRICULAR ACTIVITIES:**

(Co-curricular activities shall not promote copying from textbook or from others work and shall encourage self/independent and group learning)

**A. Measurable**

1. Assignments (in writing and doing forms on the aspects of syllabus content and outside the syllabus content. Shall be individual and challenging)
2. Student seminars (on topics of the syllabus and related aspects (individual activity))
3. Quiz (on topics where the content can be compiled by smaller aspects and data (Individuals or groups as teams))
4. Study projects (by very small groups of students on selected local real-time problems pertaining to syllabus or related areas. The individual participation and contribution of students shall be ensured (team activity

**B. General**

1. Group Discussion
2. Try to solve MCQ's available online.
3. Others

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**
Siddharthanagar, Vijayawada – 520010
***(An Autonomous College under the Jurisdiction of Krishna University)***
***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***
***College with Potential for Excellence***
***ISO9001 – 2015 Certified***

## DATA MINING AND DATA ANALYSIS

Course Code: **DSCSET03** Offered to: **B.Sc. (MSDS)**
Domain Subject: **Data Science** Semester – **V**
Type of the Course: **Skill Enhancement Course** (Elective Theory) Credits: **03**
**Time: 3 Hrs.** **Max.: 75 Marks** **Min. Pass: 30 Marks**

**SECTION - A**

**Answer any *five* of the following:** **5 X 5= 25 MARKS**

1. What is Datamining? What kinds of data can be mined? (CO1,L2)
2. Describe classification of data mining systems. (CO1,L2)
3. What is the need of data preprocessing? List major tasks in data preprocessing. (CO2,L2)
4. Write DMQL syntax for specifying task-relevant data. (CO2,L3)
5. Explain market basket analysis in detail. (CO3,L2)
6. List various pre-processing steps may be applied to prepare data for classification and prediction.(CO4,L3)
7. Explain the methods of tree pruning. (CO4,L2)
8. Differentiate between clustering and classification. (CO5,L3)

**SECTION – B**

**Answer *all* the following questions** **5 X 10 = 50 MARKS**

9. (a) Describe data mining functionalities, and thethe kinds of patterns they can discover. (CO1,L2)

OR

(b) Illustrate essential steps in the process of knowledge discovery in databases. (CO1,L3)

10. (a) Explain the methods to handle missing values and smooth noise in data cleaning. (CO2,L2)

OR

(b) What is dimensionality reduction? Explain dimensionality reduction steps in discrete wavelet transform with an example. (CO2,L3)

11. (a) Write an algorithm to generate frequent itemsets using Apriori algorithm. (CO3,L3)

OR

(b) Write an algorithm to generate frequent itemsets using FP-Growth.(CO3,L3)

12. (a) Write an algorithm for classification using Decision tree induction.(CO4,L3)

OR

(b) Explain Bayesian classification in detail. (CO4,L3)

13. (a) Explain categorization of clustering methods in detail. (CO5,L2)

OR

(b) Write K-means clustering algorithm. Explain with suitable example. (CO5,L3)

@@@@

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**
Siddharthanagar, Vijayawada – 520010
***(An Autonomous College under the Jurisdiction of Krishna University)***
***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***
***College with Potential for Excellence***
***ISO9001 – 2015 Certified***

# DATA MINING AND DATA ANALYSIS LAB

Course Code: **DSCSEP03** Offered to: **B.Sc. (MSDS)**
Domain Subject: **Data Science** Semester – **V**
Type of the Course: **Skill Enhancement Course** (Elective Theory) Credits: **02**
Max. Marks: **50** (CCIA: 10+ SEE: 40) Practical Hrs./Week **: 3**

## I. Course Outcomes: Students at the successful completion of the course will be able to:

1. Use Statistical techniques to carry out the analysis of data
2. Gain hands-on skills and experience on data mining tools.

## II: Practical (Laboratory) Syllabus: (30 Periods)

1. Data Analysis – Getting to know the Data (Using ORANGE WEKA)
   - Parametric – Means. T-Test, Correlation
   - Prediction for numerical outcomes – Linear regression
   - Correlation analysis
   - Preparing data for analysis o Pre-Processing techniques

2. Data Mining (Using ORANGE WEKA or any source data mining tool)
   - Implement clustering algorithm
   - Implement classification using oDecision tree oBack Propagation
   - Visualization methods

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**
Siddharthanagar, Vijayawada – 520010
***(An Autonomous College under the Jurisdiction of Krishna University)***
***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***
***College with Potential for Excellence***
***ISO9001 – 2015 Certified***

## DATA MINING AND DATA ANALYSIS LAB

Course Code: **DSCSEP03** Offered to: **B.Sc. (MSDS)**
Domain Subject: **Data Science** Semester – **V**
Type of the Course: **Skill Enhancement Course** (Elective Theory) Credits: **02**
Max. Marks: **40** Practical Hrs./Week **: 3**

**Model Paper: Practical**

**Time: Three hours** **Max. Marks: 40**

**Section A**
One Major Experiment (Experiment No : ) **15 M**
**Section B**
One Minor Experiment (Experiment No : ) **10 M**
**Section C**
Practical Record **05 M**
**Section D**
Viva Voce **10 M**

***

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**
Siddharthanagar, Vijayawada – 520010
***(An Autonomous College under the Jurisdiction of Krishna University)***
***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***
***College with Potential for Excellence***
***ISO9001 – 2015 Certified***

## Multivariate Technique for Data Analysis

Course Code: **DSCSET04** Offered to: **B.Sc. (MSDS)**
Domain Subject: **Data Science** Semester – **V**
Type of the Course: **Skill Enhancement Course** (Elective Theory) Credits: **03**
Max. Marks: **100** (CCIA: 25+ SEE:75) Theory Hrs./Week: **3**

**I. Course Outcomes: Students at the successful completion of the course will be able to:**
CO1: Understanding the usage of multivariate techniques for the problem under the consideration.
CO2: Analyse different Deriving &Interpreting factors
CO3: Classify the Interpretation & validation of profiling the clusters.
CO4: Evaluate Procedure for conducting discriminate analysis
CO5: Apply Graphical & Simplex methods with programming

**II. Syllabus:** **(Total Theory periods: 45)**

Unit-I:

1.1 Introduction to Multivariate Analysis
1.2 Meaning of Multivariate Analysis
1.3 Measurements Scales
1.4 Metric measurement scales
1.5 Non- Metric measurement scales
1.6 classification of multivariate techniques (Dependence Techniques and Inter-dependence Techniques)
1.7Applications of Multivariate Techniques in different disciplines.

Unit-II:

2.1 Factor Analysis
2.1 Factor Analysis
2.2 Meaning
2.3 objectives and Assumptions
2.4 Designing a factor analysis
2.5 Deriving factors and assessing overall factors
2.6 Interpreting the factors
2.7 validation of factor analysis.

Unit-III: Cluster Analysis
3.1 Cluster Analysis
3.2 Objectives and Assumptions
3.3 Research design in cluster analysis

3.4 Deriving clusters and assessing overall fit
3.4.1 Hierarchical Methods
3.4.2 Non Hierarchical Methods
3.4.3 Combinations
3.5 Interpretation of clusters
3.6 validation of profiling of the clusters.

Unit-IV: Discriminate Analysis
4.1 Discriminate Analysis
4.2 Concept
4.3 Objective and applications
4.4 Procedure for conducting discriminate analysis
4.5 Stepwise discriminate analysis
4.6 Mahalanobis procedure
4.7 Logit model.

Unit-V: Linear Programming
5.1 Linear Programming problem
5.2 Formulation
5.3 Graphical method
5.4Simplex method
5.5 Integer Programming
5.6 Transportation
5.7 Assignment problem.

**III. Text Books:**


1. Joseph F Hair, William C Black etal, "Multivariate Data Analysis", Pearson Education, 7th edition, 2013


**Reference Books:**


1. T.W Anderson, " An introduction to Multivariate Statistical Analysis, 3rd Edition", Wiley 2003.
2. William r Dillon, John Wiley & Sons, "Multivariate Analysis Methods and Applications", Wiley, 1984.
3. Naresh K Malhotra, Satyabhusan Dash, "Marketing Research An applied Orientation", Pearson, 2011.
4. Hamdy A Taha, "Operations Research", Pearson, 2012.
5. S R Yaday, A K Malik, "Operations Research", Oxford, 2014.


@@@@

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

**Multivariate Technique for Data Analysis**

Course Code: **DSCSET04** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

Type of the Course: **Skill Enhancement Course** (Elective Theory) Credits: **03**

**Time: 3 Hrs.** **Max.: 75 Marks** **Min. Pass: 30 Marks**

## SECTION - A

**Answer any five of the following:** **5 X 5= 25 MARKS**

1. Define Multivariate Analysis. What are the metrics of Measurement Scale? (CO1,L2)
2. How to build the multivariate model? (CO1,L2)
3. Define Factor Analysis. What are the objectives and assumptions of Factor Analysis?(CO2,L2)
4. How to design a Factor Analysis? (CO2,L3)
5. What are the basic stages in the application of Cluster Analysis? (CO3,L2)
6. What is the purpose of of Cluster Analysis and when should it be used instead of Factor Analysis? (CO4,L3)
7. Define Discriminate Analysis. What are the applications of Discriminate Analysis?(CO4,L2)
8. How to solve a Linear Programming problem? (CO5,L3)

## SECTION – B

**Answer *all* the following questions** **5 X 10 = 50 MARKS**

9.(a) Explain classification of multi variate techniques in detail. (CO1,L2)

OR

(b) What are the applications of Multivariate Techniques in different disciplines? (CO1,L3)

10.(a) How can factor analysis help the researcher improve the results of other mulivarate techniques? (CO2,L2)

OR

(b) How to interpret the factors and Validate of factor analysis? Explain with suitable example. (CO2,L3)

11.(a) How does the researcher know whether to use hierarchical or non hierarchical cluster techniques? Under which conditions would each approach be used? (CO3,L3)

OR

(b) As a data analyst, you could use multiple regressions to predict crop growth. In this example, crop growth is your dependent variable and you want to see how different factors affect it. Your independent variables could be rainfall, temperature, amount of sunlight, and amount of fertilizer added to the soil. A multiple regression model would show you the proportion of variance in crop growth that each independent variable accounts for. (CO3,L3)

12. (a) How to do Discriminate Analysis in stepwise? (CO4,L3)

OR

(b) Explain about Logit model with suitable example. (CO4,L3)

13. (a) Explain about graphical method and simplex method in Linear Programming. (CO5,L2)

OR

(b) Explain about Transportation and Assignment problem with suitable example. (CO5,L3)

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

**Multivariate Technique for Data Analysis Lab**

Course Code: **DSCSEP04** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

Type of the Course: **Skill Enhancement Course** (Elective Theory) Credits: **02**

Max. Marks: **50** (CCIA: 10+ SEE: 40) Practical Hrs./Week **: 3**

**I. Course Outcomes: Students at the successful completion of the course will be able to:**

CO1: Install, Code and Use R Programming Language in R Studio IDE to perform basic tasks on Vectors, Matrices and Data frames.

CO2: Describe key terminologies, concepts and techniques employed in Statistical Analysis.

CO3: Define, Creating and manipulating network objects.

**II: Practical (Laboratory) Syllabus: (30 Periods)**

1. Navigating the basic operating environment of 'R'
2. Importing network data.
3. Creating and manipulating network objects.
4. Plotting Network Graphs.
5. Network Descriptive Statistics.
6. Hypothesis Testing.

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

**Multivariate Technique for Data Analysis Lab**

Course Code: **DSCSEP04** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

Type of the Course: **Skill Enhancement Course** (Elective Theory) Credits: **02**

Max. Marks: **40** Time **: 3 Hrs.**

**Model Paper: Practical**

**Time: Three hours** **Max. Marks: 40**

**Section A**

One Major Experiment (Experiment No : ) **15 M**

**Section B**

One Minor Experiment (Experiment No : ) **10 M**

**Section C**

Practical Record **05 M**

**Section D**

Viva Voce **10 M**

***

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

Course Code: **DSCCSET05** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

Max. Marks: **100** (CCIA: 25+ SEE:75) Theory Hrs./Week: 3

Type of the Course: **Skill Enhancement Course** (Elective Theory) Credits: **03**

## Data and Information Security through Python

**I. Course Outcomes: Students at the successful completion of the course will be able to:**

**CO1:** Examine and apply the fundamental techniques of computer security.
**CO2:** Identify and explain risk and potential security issues.
**CO3:** Demonstrate responsible computer use as it deals with Cryptographic algorithms.
**CO4:** Applying Digital Signatures for information security/assurance within the organization.
**CO5:** Demonstrate knowledge of security objectives and policy development

**II. Syllabus:** **(Total Theory periods: 45)**

**Unit -I**

**1. Overview of Security**

1.1Protection versus security
1.2 Aspects of security
1.3 Data integrity
1.4 Data availability
1.5 Privacy
1.6 Security problems
1.7 User authentication
1.8 Orange Book.

**Unit -II**

**2. Security Threats**

2.1Program threats
2.2 Worms
2.3 Viruses
2.4 Trojan horse
2.5 Trap door
2.6 stack and buffer overflow
2.7 System threats
2.8 Intruders
2.9 Communication threats
2.10 Tapping and piracy
2.11 Denial of Services (DoS)

**Unit -III**

**3. Cryptography**

3.1Substitution
3.2 Transposition ciphers
3.3 Symmetric
3.4 key algorithms
3.5 Data Encryption Standard
3.6 Advanced encryption standards
3.7 Public key encryption
3.8 RSA
3.9 Diffie-Hellman key exchange
3.10 ECC cryptography
3.11 Message Authentication
3.12 MAC& Hash functions.

**Unit -IV**

**4. Digital Signatures**

4.1Symmetric key signatures
4.2 public key signatures
4.3 Message digests
4.4 Public key infrastructures.
4.5 Firewalls
4.6 kerberos
4.7 session hierarchy

**Unit -V**

**5. Security Mechanism**

**5.1** Intrusion detection
5.2 Auditing and logging
5.3 Tripwire
5.4 System
5.5 Call monitoring.

**III. Text Books:**

1. W. Stallings, Cryptography and Network Security Principles and Practices (4th ed.,), Prentice – Hall of India, 2006 http://uru.ac.in/uruonlinelibrary/Cyber_Security/Cryptography_and_Network_Security.pdf
2. Behrouz A. Forouzan and Debdeep Mukhopadhyay, "Cryptography & Network Security", 3rd Edition, TMH.

**References:**

1. C. Pfleeger and SL Pfleeger, Security in Computing (3rd ed., ), Prentice- Hall of India, 2007.
2. D. Gollamann, Computer Security, John Wiley and Sons, Ny, 2002.
3. J. Piwprzyk, T. Hardjono and J. Seberry, Fundamentals of Computer Security, Springer-VerlagBerling, 2003.
4. J.M. Kizza, Computer Netwrok Security, Springer, 2007
5. M. Merkow and J. Breithaupt, Information Security: Principles and Practices, Pearson Education, 2006.

@@@@

**P. B. SIDDHARTHACOLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

Course Code: **DSCCSET05** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

Max. Marks: **75** Time: 3 Hrs

Type of the Course: **Skill Enhancement Course** (Elective Theory) Credits: **03**

**Data and Information Security through Python**

**SECTION - A**

**Answer any *five* of the following: 5 X 5= 25 MARKS**

1. Explain Data Integrity. (CO1,L2)
2. Explin Intruders in Network Security. (CO2,L2)
3. What is the difference between Private Key and Public Key encryption.(CO3,L2)
4. Breifly Explain Symmetric key Signatures.(CO4,L3)
5. Write a short note on Auditing and logging.(CO5,L2)
6. What is the importance of Digital Signatures.(CO4, L3)
7. Explain Security Mechanisms.(CO5,L2)
8. Discuss Secure hash Algorithm. (CO3,L3)

**SECTION – B**

**Answer *all* the following questions 5 X 10 = 50 MARKS**

9. (a)List and briefly define categories of passive and active security attacks(CO1,L2)

   OR

   (b) List and briefly define categories of security services. (CO1,L3)
10. (a) Explain the different types of threats in security. (CO2,L2)

    OR

    (b) What is inforamation security. Discuss Program threats and system threats. (CO2,L3)
11. (a)What is the role of a compression function in a hash function? (CO3,L3)

    OR

    (b) What are three broad categories of applications of public-key cryptosystems (CO3,L3)
12. (a) What are some threats associated with a direct digital signature scheme? (CO4, L3)

    OR

    (b) What are the properties a digital signature should have? (CO4, L3)
13. (a) Explain Relationship between Security Services and Mechanisms(CO5, L2)

    OR

    (b) Write about Specific Security Mechanisms (CO5, L3)

@@@@

**P. B. SIDDHARTHACOLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

Course Code: **DSCCSEP05** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

Max. Marks: **50** (CCIA: 10+ SEE: 40) Practical Hrs./Week **: 3**

Type of the Course: **Skill Enhancement Course** (Elective, Practical) Credits: 02

## DATA & INFORMATION SECURITY THROUGH PYTHON LAB

### I. Course Outcomes: Students at the successful completion of the course will be able to:

Develop and implement Python Programming for encryption and decryption algorithms i.e., DES, MD5 and RSA algorithms

### II: Practical (Laboratory) Syllabus: (30 Periods)

1. Implement Ceiser Cipher encryption in Python.
2. Implement Ceiser Cipher decryption in Python.
3. Implement Transposition technique encryption in Python.
4. Implement Substitution cipher encryption in Python
5. Implement Substitution cipher decryption in Python.
6. Implement one time Pad cipher in Python.
7. Implement DES encryption in Python.
8. Implement RSA Public Key encryption in Python.

@@@@

**P. B. SIDDHARTHACOLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

Course Code: **DSCCSEP05** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

Max. Marks: **40** Time **: 3 Hrs.**

Type of the Course: **Skill Enhancement Course** (Elective, Practical) Credits: 02

## DATA & INFORMATION SECURITY THROUGH PYTHON LAB

**Model Paper: Practical**

**Time: Three hours** **Max. Marks: 40**

**Section A**

One Major Experiment (Experiment No : ) **15 M**

**Section B**

One Minor Experiment (Experiment No : ) **10 M**

**Section C**

Practical Record **05 M**

**Section D**

Viva Voce **10 M**

***

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

Course Code: **DSCSET06** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

Max. Marks: **100** (CCIA: 25+ SEE:75) Theory Hrs./Week: 3

Type of the Course: **Skill Enhancement Course** (Elective Theory) Credits: **03**

## Spark Programming

**I. Course Outcomes: Students at the successful completion of the course will be able to:**

CO1: Understand the architecture of data, layers & Spark libraries

CO2: Analysing the Twitter data

CO3: To learn about Spark Algorithms

CO4: Understand the Supervised machine learning workflows

CO5: Building a fault tolerance, reliable and scalable streaming app

**II. Syllabus:** **(Total Theory periods: 45)**

**Unit - I**

1.1 Setting Up a Spark Virtual Environment
1.2 Understanding the architecture of data
1.3 Intensive applications
1.4 Infrastructure layer
1.5 Persistence layer
1.6 Integration layer
1.7 Analytics layer
1.8 Engagement layer
1.9 Understanding Spark
1.10 Spark libraries
1.11 PySpark in action
1.12 The Resilient distributed Dataset
1.13 Understanding Anaconda

**Unit – II**

2.1 Building Batch and Streaming Apps with Spark
2.2 Architecting data-intensive apps
2.3 Processing data at rest
2.4 Processing data in motion
2.5 Exploring data interactively
2.6 Connecting to social networks
2.7 Getting Twitter data
2.8 Analysing the data
2.9 Discovering the anatomy of tweets

**Unit – III**

3.1 Learning from Data Using Spark
3.2 Contextualizing Spark Mllib in the app architecture
3.3 Classifying Spark Mllib algorithms
3.4 Supervised and unsupervised learning
3.5 Additional learning algorithms
3.6 Spark Mllib data types

**Unit – IV**

4.1 Machine learning workflows and data flows
4.2 Supervised machine learning workflows
4.3 Unsupervised machine learning workflows
4.4 Clustering the Twitter dataset
4.5 Applying Scikit
4.6 Learn on the Twitter dataset
4.7 Pre-processing the dataset
4.8 Running the clustering algorithm
4.9 Evaluating the model and the results
4.10 Building machine learning pipelines

**Unit - V**

5.1 Streaming Live Data with Spark
5.2 Laying the foundations of streaming architecture
5.3 Spark Streaming inner working
5.4 Going under the hood of Spark Streaming
5.5 Building in fault tolerance
5.6 Processing live data with TCP sockets
5.7 Setting up TCP sockets, Processing live data
5.8 Manipulating Twitter data in real time
5.9 Processing Tweets in real time from the Twitter firehose
5.10 Building a reliable and scalable streaming app

**III. Text Books:**

1. Spark for Python developers, Amit Nandi, Orielly Publishing,2015.

**Reference Books:**

2. Interactive Spark using PySpark, Benjamin Bengfort& Jenny Kim, Orielly.
3. Spark: The Definitive Guide: Big Data Processing Made Simple, 1st Edition by Bill Chambers &MateiZaharia

**https://www.pdfdrive.com/spark-for-python-developers-d158022250.html**


Spark for Python Developers ( PDFDri

## IV. RECOMMENDED CO-CURRICULAR ACTIVITIES:

(Co-curricular activities shall not promote copying from textbook or from others work and shall encourage self/independent and group learning)

**A. Measurable**

1. Assignments (in writing and doing forms on the aspects of syllabus content and outside the syllabus content. Shall be individual and challenging)
2. Student seminars (on topics of the syllabus and related aspects (individual activity))
3. Quiz (on topics where the content can be compiled by smaller aspects and data (Individuals or groups as teams))
4. Study projects (by very small groups of students on selected local real-time problems pertaining to syllabus or related areas. The individual participation and contribution of students shall be ensured (team activity

**B. General**

1. Group Discussion
2. Try to solve MCQ's available online.
3. Others

@@@@

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

Course Code: **DSCSET06** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

**Max.: 75 Marks** **Time: 3 Hrs.**

Type of the Course: **Skill Enhancement Course** (Elective Theory) Credits: **03**

**Spark Programming**

**Model Question Paper**

**SECTION – A**

**Answer any Five questions.** **5X5=25 Marks**

1. Explain the Spark Libraries. (CO1,L2)
2. Why Anaconda is used in Python?(CO1,L2)
3. How to Analyze data in Spark. (CO2,L3)
4. What are the MLlib data types in Spark.(CO3,L2)
5. How to build a machine learning Pipelines. (CO4, L2)
6. How evaluating the model and the result in Spark.(CO4, L2)
7. How to manipulating Twitter data in real time. (CO5, L3)
8. Explain Spark streaming inner working.(CO5, L3)

**SECTION- B**

**Answer all questions.** **5X10 = 50 Marks**

**9.** A) Explain the architecture of data-intensive applications in Spark. (CO1,L2)

OR

B) What are the applications of Spark.(CO1,L2)

**10.** A)Data Intensive App Frame work.(CO2,L3)

OR

B) Explain the stages in batch and streaming data processing by Spark.(CO2,L3)

**11.** A) Classifying Spark MLlib Algorithms.(CO3,L2)

OR

B) Explain Supervised and Unsupervised Learning Algorithms.(CO3,L2)

**12.** A) Explain Machine Learning Workflows and Data flows. (CO4, L2)

OR

B) How to Apply Scikit-Learn on the Twitter dataset.(CO4, L3)

**13.** A) Explain the foundations of streaming architecture. (CO5, L3)

OR

B) How to Processing live data with TCP sockets. (CO5, L3)

@@@@

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

Course Code: **DSCSEP06** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

Max. Marks: **50** (CCIA: 10+ SEE: 40) Practical Hrs./Week **: 3**

Type of the Course: **Skill Enhancement Course** (Elective, Practical) Credits: 02

## SPARK PROGRAMMING LAB

**I. Course Outcomes: Students at the successful completion of the course will be able to:**

**CO1:** Prepare data sets and write spark programming in python

**CO2:** Apply basic operation required for data analytics in spark

CO3: Analyse the streaming process in real time scenario

**II: Practical (Laboratory) Syllabus: (30 Periods)**

1. To work with Ipython Notebook for a friendlier user experience than the console.
2. Word count of manual script using Pyspark. Create tuple (Count, word) and sort in descending, take top 20 words by frequently.
3. Create a function for histogram of most frequent words. Visualize the most frequent words by plotting them.
4. Create AWS EC2 key pair via the AWS console. ( http://aws.amazon,com.console/)
5. Create the Python twitter API class and its base methods for authentication, searching and parsing the results. Self. auth gets the credentials from Twitter.
6. Create Python program to call the Meet up API using an authentication token. Retrieve the profile of the meet up members in order to analyse their participation in similar meet up groups.
7. Write a python class that manages to persist data in CSV format and read from a given CSV.
8. Create a Pandas Data frame by reading the parsed tweets saved in a CSV file, on tweet data

@@@@

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

Course Code: **DSCSEP06** Offered to: **B.Sc. (MSDS)**

Domain Subject: **Data Science** Semester – **V**

Max. Marks: 40 Time: 3 Hrs.

Type of the Course: **Skill Enhancement Course** (Elective, Practical) Credits: 02

## SPARK PROGRAMMING LAB

**Model Paper: Practical**

**Time: Three hours** **Max. Marks: 40**

**Section A**

One Major Experiment (Experiment No : ) **15 M**

**Section B**

One Minor Experiment (Experiment No : ) **10 M**

**Section C**

Practical Record **05 M**

**Section D**

Viva Voce **10 M**

***

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**
Siddharthanagar, Vijayawada – 520010
***(An Autonomous College under the Jurisdiction of Krishna University)***
***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***
***College with Potential for Excellence***
***ISO9001 – 2015 Certified***

Course Code: **SDCCSCP** Offered to: **B.Sc. (MSDS)**
Domain Subject: **Data Science** Semester: III
Max. Marks: **50** (CCIA: 10+ SEE: 40) Practical Hrs./Week **: 2**

**Course Prerequisites (if any):** Basics of Python, Linear Equation
**Course Description:** Make the Students to understand Numpy and Pandas.

**Course Objective:**

1. Understand the Numpy Library files.
2. Understand the Pandas Library files.
3. Generates the Random numbers, Statistical Functions.
4. Create and cleaning the data.
5. Data Visualization

**Course Outcomes:** At the end of this course, students should be able to:

CO1: Understand the Array concepts, Statistical functions.(PO5, PO7)
CO2: Implement Vector and Matrix, Statistical functions using Numpy (PO5, PO6, PO7)
CO3: Understand and Implement the Data Frames (PO5, PO7)
CO4: Cleaning the data (PO5, PO6, PO7)
CO5: Visualizing the data (PO5, PO6, PO7)

## List of Experiments

1. Implement Vector and matrix using Numpy.
2. Implement Statistical functions.
3. Generate Random number using Numpy
4. Creation of Series and Data Frames
5. Importing data, Exporting data, Data selection, Adding Columns and rows
6. Rows to a Data Frame, dropping rows and columns in a Data Frame.
7. Grouping data, Sorting a Data Frame, Binning numerical features, Creating dummy variables.
8. Implement Descriptive Statistics, Accessing elements in series.
9. Data Cleaning: Missing values, Outliers, duplicates, detection and implication. Split-Apply- Combine operators
10. Data Visualization with Pandas for Uni- Variate, Bi-Variate data.

**Recommended book:**

**1. *Jake VanderPlas,* "Python Data Science Handbook" by** 2nd Edition, December 2022, ISBN: 9781098121204, O'Reilly

Reference Books:

1. *Fabio Nelli*, "Python Data Analytics With Pandas NumPy and Matplotlib",A press.

**Course Delivery method:** Face-to-face / Blended **Course has focus on : S**kill Development **Websites of Interest :**

1. **https://jakevdp.github.io/PythonDataScienceHandbook/**
2. **https://github.com/jakevdp/PythonDataScienceHandbook**
3. **https://indianpdf.com/python-data-analytics-pdf/**
4. **https://www.hackerearth.com/practice/machine-learning/data-manipulation-visualisation-r-python/tutorial-data-manipulation-numpy-pandas-python/tutorial/**
5. **https://cloudxlab.com/blog/numpy-pandas-introduction/**
6. **https://github.com/PacktPublishing/Hands-On-Data-Analysis-with-NumPy-and-pandas**

**Co-curricular Activities:** Hands on Exercises

@@@@

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**
Siddharthanagar, Vijayawada – 520010
***(An Autonomous College under the Jurisdiction of Krishna University)***
***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***
***College with Potential for Excellence***
***ISO9001 – 2015 Certified***

Course Code: **DSCT31B** Offered to: **B.Sc. (MSDS)**
Domain Subject: **Data Science** Semester: III
**SEMESTER: III** **Credits: 4**

## FOUNDATION OF DATA SCIENCE USING R

**I. Type of the Course: Skill Enhancement Course (Elective Theory)**
**Course Outcomes**: On successful completion of the course, students will be able to do following:
CO1: Able to load data into R and spot problems in data loaded. (PO5, PO7)
CO2: Understand basics of R and control structures in R. (PO5)
CO3: Load data into R objects and manipulate them as needed. (PO5)
CO4: Create and edit visualizations with R (PO7)
CO5: Document and transfer the results and communicate the findings using visualization techniques. (PO5, PO7)

**II. Syllabus** **(Total Theory Hours: 60)**

**UNIT-I** **(12 hours)**

1. **Introduction to Data Science**
1.1 Data science process
    1.1.1. Roles
    1.1.2. Stages in data science project
1.2 Loading data into R:
    1.2.1. Working with data from files
    1.2.2. Working with relational databases
1.3 Exploring data
    1.3.1. Using summary statistics to spot problems
    1.3.2. Spotting problems using graphics and visualization
1.4 Managing data
    1.4.1. Cleaning
1.4.2. Sampling for modelling and Validation

**UNIT-II** **(12 hours)**

**2. Introduction to R and getting started with R**
    2.1. What is R? Why R?
    2.2. Advantages of R over other programming languages
    2.3. Data types in R - logical, numeric, integer, character, double
    2.4. Complex, raw
    2.5. Coercion, ls () command
    2.5. Expressions, Variables and functions, control structures
    2.6. Array, Matrix,Vectors, Factors, R packages

**UNIT-III** **(12 hours)**

**3. Exploring data in R**
    3.1. Data frames-data frame access, ordering data frames
    3.2. R functions for data frames -

dim(), nrow(), ncol(), str(), summary(), names(), head(), tail(), edit()

3.3. Load data frames—reading from .CSV files

3.4. Sub setting data frames, reading from tab separated value files

3.5. Reading from tables, merging data frames

**UNIT-IV** **(12 hours)**

**4. Data Visualization using R**

4.1. Reading and getting data into R (External Data),Using CSV files

4.2. XML files, Web Data, JSON files, Databases, Excel files

4.3. Working with R Charts and Graphs:
Histograms, Boxplots, Bar Charts, Line Graphs, Scatterplots, Pie Chart

**UNIT-V** **(12 hours)**

**5. Delivering Results**

5.1 Displaying multivariate data

5.2 Plot () function

5.3 Matrix plots

5.4 Multiple plots in one window

5.5 Exporting graph

5.6 using graphics parameters

**III. TEXT BOOKS:**

1. Nina Zumel, John Mount, "Practical Data Science with R", Manning Publications, 2014.(UNIT I)

2. Seema Acharya--Data Analytics using R, McGraw Hill education (India) Private Limited. (UNIT II,III,IV).

3. Mark Gardener, "Beginning R - The Statistical Programming Language", John Wiley & Sons, Inc., 2012.(UNIT V)

**Reference Books**:

1. An Introduction to R, Notes on R: A Programming Environment for Data Analysis and Graphics. W. N. Venables, D.M. Smith and the R Development Core Team

**Websites of References:**

https://nbisweden.github.io/workshop-r/2011/slide_elements_1.pdf
https://www.guru99.com/r-tutorial.html
https://www.javatpoint.com/r-tutorial
https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

@@@@

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**
Siddharthanagar, Vijayawada – 520010
***(An Autonomous College under the Jurisdiction of Krishna University)***
***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***
***College with Potential for Excellence***
***ISO9001 – 2015 Certified***

Course Code: **DSCT31B** Offered to: **B.Sc. (MSDS)**
Domain Subject: **Data Science** Semester: III
**SEMESTER: III** **Credits: 4**

**FOUNDATION OF DATA SCIENCE USING R**

**Max Marks: 75** **Time: 3 Hrs.**

**SECTION – A**

**Short Answer Questions** **(Total: 5 x 5 = 25 Marks)**

**Answer any *five* questions**

1. List and explain the roles in a data science project. (CO1, L4)
2. Explain ls () command in R. (CO2, L2)
3. Write about the control structures in R with examples. (CO2, L1)
4. Develop R script to load data into data frames from files. (CO3, L6)
5. Write a short note on charts. (CO4, L1)
6. Develop bar chart in R. (CO4, L6)
7. Write about the plot () function in R. (CO5, L1)
8. Explain about Multiple plots in one window. (CO5, L2)

**SECTION B**

**Answer all questions** **(Total: 5 x 10 = 50 Marks)**

9. A) List different stages of the Data Science Project and explain them. (CO1, L1)
(OR)
B) What are the various problems encountered when using data summaries? (CO1, L1)

10. A) Write about the Data types in R Explain with examples. (CO2, L1)
(OR)
B) Construct a Vector in R and explain various operations on it. (CO2, L3)

11. A) What are the data frames? Write its significance in R-Language. (CO3, L1)
(OR)
B) Demonstrate various functions used in data frames. (CO3, L2)

12. A) Build code in R for reading and getting data into R from databases. (CO4, L6)
(OR)
B) Develop any four plots in R. (CO4, L6)

13. A) Develop a matrix plot and explain it. (CO5, L3)
(OR)
B) Explain about exploring graphs. (CO5, L2)

@@@@

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**
Siddharthanagar, Vijayawada – 520010
***(An Autonomous College under the Jurisdiction of Krishna University)***
***Re– Accredited at 'A$^+$' by NAAC–III Cycle***
***College with Potential for Excellence***
***ISO9001 – 2015 Certified***

Course Code: **DSCP31B** Offered to: **B.Sc. (MSDS)**
Domain Subject: **Data Science** Semester: III
**SEMESTER: III** **Credits: 1**

# FOUNDATION OF DATA SCIENCE USING R LAB

## PRACTICAL SYLLABUS

**I. Type of the Course: Skill Enhancement Course (Elective, Practical)**
**Course Outcomes**: At the end of this course, students should be able to:
CO1: Implement simple scripts or programs in R.(PO5)
CO2: Access online resources for R and import new function packages into the R workspace. (PO5, PO7)
CO3: Import, review, manipulate and summarize data-sets in R (PO5, PO7)
CO4: Explore data-sets to create testable hypotheses and identify appropriate statistical tests. (PO5, PO7)
CO5: Create and edit visualizations with R. (PO5, PO7)

**II: Practical (Laboratory) Syllabus: (30 Hours)**

1. Create a vector in R and perform operations on it (arithmetic operations, combining vectors, retrieving elements of vector, assign names to vector elements).
2. Create integer, complex, logical, character data type objects in R and print their values and their class using print and class functions.
3. Create a matrix of values in R and extract data from matrix. (Ex. Second row third etc) Find transpose of matrix and combine two matrices using Rbind and Cbind functions.
4. Create a list in R and perform operations on it like list slicing, sum and mean functions, Head and tail functions and finally delete list using rm() function.
5. Create data frame in R and perform operations on it
6. Write code in R to find out whether a number is prime or not.
7. Print numbers from 1 to 100 using while loop and for loop in R.
8. Find the factorial of a number using recursion in R.
9. Perform arithmetic operations in R using switch case
10. Write a code in R to find out whether the number is Armstrong or not.
11. Program to find Multiplication table from 1 to 10 number input by user.
12. Import data into R from text and excel files using read.table() and read.csv() function.
13. Create a dataset and draw different types of graphics using plot, box plot, histogram, pair plot functions.
14. Create a dataset and draw different types of graphs using bar charts, pie chart functions.
15. Create custom contingency in R and perform operations on it.

*******

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^+$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

| COMPUTER SCIENCE | AMLT31 | 2022 - 2023 | B.Sc. (AI &ML) |
|---|---|---|---|

**SEMESTER – III** **Credits – 4**

**Introduction to Artificial Intelligence**

**Total: 60 Prds.**

**Course Objective:**

The objective of this course is to educate students in basic Artificial Intelligence concepts and provide insights of solving problems using AI. This course also aims to educate students in basics of practical natural language processing and robotics.

**Course Outcomes:**

| **COURSE OUTCOME NO** | Upon successful completion of this course, students should have the knowledge and skills to: | **PROGRAM OUTCOME NO** |
|---|---|---|
| CO1 | Understand the need of AI and generic Search Techniques. | PO5, PO7 |
| $CO_2$ | Understand Different Search Methods. | PO5, PO7 |
| $CO_3$ | Gain knowledge about learning Gaming Theory | PO5, PO7 |
| $CO_4$ | Learn the knowledge representation and reasoning | PO5, PO7 |
| $CO_5$ | Gain knowledge about machine learning process | PO5, PO7 |

**Unit- I : Artificial Intelligence** **12 prds**

Introduction: Artificial Intelligence –Historical Backdrop – What is Intelligence? – The Turing Test – Topics in AI.-State Space Search – Generate and Test – Graph Search – Generic Search Algorithm - Depth First Search (DFS) – Breadth First Search(BFS) – Iterative Deepening – Uniform cost Search – Bidirectional Search.

**Unit-II Problem Solving using Search** **12 Prds**

Heuristic Search –Heuristic Functions – Heuristics and AI –Problem Relaxation – Admissible Heuristics – Dominance – Composite Heuristics – Consistent Heuristics - Informed Search Heuristic Search – Types of Heuristic Search – BFS – A* Algorithm – Local Search – Hill

Climbing – Simulated Annealing – Constraint Satisfaction Problems – N-queens Problem – Varieties of CSP – Backtracking Search – Strategies – AND-OR Graph – AO* Algorithm.

**Unit-III Game Play** **12 prds**

Introduction – Grundy's Game – Game Trees – Minimax – Tic-Tac-Toe – Minimax – Alpha-beta Pruning - Planning – Forward State Space Planning – Backwards State Space Planning – Goal Stack Planning – Plan Space Planning – A Unified Framework for Planning.

**Unit-IV Knowledge Representation** **12 prds**

Introduction – Logic for KR – Propositional Logic – First Order Logic – Inference of first order Logic – Answer Extraction – Reasoning with uncertainty – Bayesian Network – Decision Network – Fuzzy Reasoning – Sequential Decision Problems.

**Unit-V Machine Learning** **12 Prds**

Introduction – Decision Tree –Linear Regression - Support Vector Machine – Unsupervised Learning – Reinforcement Learning – Introduction to Neural Network – Introduction to Deep Learning

**TEXT BOOKS**

Stuart Russell, Peter Norvig: "Artificial Intelligence: A Modern Approach",7th Edition, Pearson Education, 2007

**REFERENCES**

1. Artificial Neural Networks B. YagnaNarayana, PHI
2. Artificial Intelligence , 2nd Edition, E.Rich and K.Knight (TMH).
3. Artificial Intelligence and Expert Systems – Patterson PHI.
4. Expert Systems: Principles and Programming- Fourth Edn, Giarrantana/ Riley, Thomson.
5. PROLOG Programming for Artificial Intelligence. Ivan Bratka- Third Edition – PearsonEducation.
6. Neural Networks Simon Haykin PHI

**Web Resources:**

https://onlinecourses.nptel.ac.in/noc22_ge29/preview
**https://onlinecourses.swayam2.ac.in/aic20_sp06/preview**
**https://onlinecourses.swayam2.ac.in/arp19_ap79/preview**

**Recommended Co – Curricular Activities:**

(Co-curricular activities shall not promote copying from textbook or from others work and shall encourage self/independent and group learning)

**A: Measurable**

1. Assignments (in writing and doing forms on the aspects of syllabus content and outside the syllabus content. Shall be individual and challenging)
2. Student seminars (on topics of the syllabus and related aspects (individual activity))
3. Quiz (on topics where the content can be compiled by smaller aspects and data (Individuals or groups as teams))
4. Study projects (by very small groups of students on selected local real-time problems pertaining to syllabus or related areas. The individual participation and contribution of students shall be ensured (team activity))

**B: General**

1. Group Discussion
2. Others

**RECOMMENDED CONTINUOUS ASSESSMENT METHODS:**

Some of the following suggested assessment methodologies could be adopted;

1. The oral and written examinations (Scheduled and surprise tests),
2. Closed-book and open-book tests,
3. Programming exercises,
4. Practical assignments and laboratory reports,
5. Observation of practical skills,
6. Individual and group project reports.
7. Efficient delivery using seminar presentations,
8. Viva voce interviews.
9. Computerized adaptive testing, literature surveys and evaluations,
10. Peers and self-assessment, outputs form individual and collaborative work.

@@@@

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

**TITLE: Introduction to Artificial Intelligence**

**COURSE CODE: AML31** **Max.Marks:75M**

**CLASS: B.Sc. (AI & ML)** **Semester III**

## Section-A

**Answer any FIVE questions.** **5 X 5=25M**

1. Explain in detail the applications of Artificial Intelligence.
2. Design perceptrons for AND and NOT boolean functions.
3. Explain the syntax and semantics of propositional logic.
4. Discuss A* algorithm in detail.
5. Discuss categorization of intelligent systems.
6. Compare Fuzzy logic with traditional logic.
7. Solve the water-jug problem by writing the production rules.
8. Explain about certainty factor theory

## Section-B

**Answer ALL questions.** **10 X 5 = 50M.**

9. (a). Explain about Tic-Tac-Toe game problem by assuming one player is X the other one can be either human or a computer by taking 3X3 grid space.

   Or

   (b) Explain Multi layerperceptron( MLP ) with back propagation with schematic block diagram.
10. (a) What is Inference Engine? Describe Backward and Forward chaining mechanism used by an inference engine?

    or

    (b)Explain the forward-chaining algorithm for propositional logic
11. (a) Describe the mathematical model of perceptron with example

    OR

    (b)Explain about Extended semantic networks for KR
12. (a) Explain about hill climbing heuristic search technique.

    or

    (b) Write a short note on Bayesian networks?
13. (a)Describe case based reasoning and learning.

    or

    (b)What is machine learning? Differentiate between supervised learning and unsupervised learning

@@@@

**P.B.SIDDHARTHACOLLEGE OF ARTS& SCIENCE**
**Siddharthanagar, Vijayawada – 520010**
**(An Autonomous College under the Jurisdiction of Krishna University)**
**Re– Accredited at 'A$^{+}$' by NAAC–III Cycle**
**College with Potential for Excellence**
**ISO9001 – 2015 Certified**

| COMPUTER SCIENCE | AMLP31 | 2022 - 2023 | B.Sc. (AI &ML) |
|---|---|---|---|

**SEMESTER – III** **Credits – 2**

**Artificial Intelligence Lab**

**Total: 30 Periods.**

**Course Objective:**

The objective of this course is to enable students to analyse various AI related problems and develop a solution using Python programming language.

**Course Outcomes:**

| **COURSE OUTCOME NO** | Upon successful completion of this course, students should have the knowledge and skills to: | **PROGRAM OUTCOME NO** |
|---|---|---|
| CO1 | Develop various basic python programs. | PO5, PO7 |
| $CO_2$ | Analyse and develop solutions for various problems like water jug, Tic – Tack – Toe, etc. | PO5, PO7 |
| $CO_3$ | Develop programs using DFS, BFS, A* and hill climbing algorithms. | PO5, PO7 |
| $CO_4$ | Develop python programs for analysing given data set. | PO5, PO7 |
| $CO_5$ | Develop python programs for implementing Bayes Classification. | PO5, PO7 |

**Lab List**

1. Basic Programs in Python.
2. Program implementing list , Vector.
3. Program implementing Matrix and Array.
4. Write a Program to Implement Breadth First Search using Python.
5. Write a Program to Implement Depth First Search using Python.
6. Write a Program to Implement Tic-Tac-Toe game using Python.
7. Write a Program to Implement 8-Puzzle problem using Python.
8. Write a Program to Implement Water-Jug problem using Python.
9. Write a Program to Implement Tower of Hanoi using Python.
10. Write a Program to Implement N-Queens Problem using Python.

11. Develop Python code for Mini-max algorithm.
12. Develop Python for Implementing Grundy's Game
13. Write a Program to Implement 8-Puzzle problem using A* algorithm.
**14.** Write a program to implement alpha-beta pruning algorithm.

***@@@@***

**P.B.SIDDHARTHA COLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

**SEMESTER III** **Credits:4**

**Total: 60 Periods**

## Document Oriented Databases

**Course Objective:**

To educate student regarding databases and how to manage databases.

To handle the large amount of data handling demands of business

To implement a data store that provides high performance, high availability, and automatic scaling

To Process an immense diversity of data that needs to be stored and processed.

To make use of features and functionalities to work on NO SQL Data Base Mongo DB

| **COURSE OUTCOME NO** | Upon successful completion of this course, students should have the knowledge and skills to: | **PROGRAM OUTCOME NO** |
|---|---|---|
| CO1 | Have knowledge about database and DBMS Architecture | PO5, PO7 |
| CO2 | Able to know No SQL databases, various features of Mongo DB, the installation procedure, and how to interact with MongoDB. | PO5, PO7 |
| CO3 | Able to work on Mongo DB's rich query language to support create, read, update, and delete (CRUD) operations. | PO5, PO7 |
| CO4 | Anslyses the aggregation framework to perform aggregation operations. | PO5, PO7 |
| CO5 | Able to work on indexes, types of index, index properties, and the various indexing strategies to be considered. Indexes are used to improve the performance of a query. | PO5, PO7 |

**Unit –I** **12periods**

**Overview of Database Management Systems:**

Introduction to Data, information, data vs. information –database and DBMS Role and advantages of DBMS – types of databases –problems with file system data management.

**Data Models:**

The importance of Data models –The evolution of Data Models-Degrees of data abstraction Introduction to Sql-Data Definition Commands – Data Types - Creating Table Structures - advanced data definition commands - alter – drop

**Unit-II** **12periods**

Entity Super types and Subtypes- Specialization and Generalization -entity integrity - selecting primary keys - Natural Keys and Primary Keys - Primary Key Guidelines - The need for normalization – The normalization process – converting to first normal form – conversion to second normal form – conversion to third normal form – higher level normal forms.

SQL Constraints Adding Table Rows Saving Table Changes - Updating Table Rows - Restoring Table Contents - Deleting Table Rows

**Unit – III** **12 Periods**

**Data Manipulation Language:**

**Select Queries:** Selecting Rows with Conditional Restrictions – operators - advanced select queries – virtual tables – joining database tables – sub queries – SQL functions

Mongo DB Features and Installation, The Need for No SQL Databases, What Are No SQL Databases?

CAP Theorem, BASE Approach, Types of NoSQL Databases, Mongo DB Features, Document Database

Mongo DB Is Schema less Mongo DB Uses BSON, Rich Query Language, Aggregation Framework

Indexing, Grid FS, Replication, Sharing The mongo Shell , Terms Used in Mongo DB, Data Types in Mongo DB, Working with Database Commands, Create Database, Drop Database , Display List of Databases, Display the Version of Mongo DB, Display a List of Commands

**Unit IV** **12 Periods**

Mongo DB CRUD Operations, Collections, Create a Collection, Create Capped Collections, Create Operations, Insert Documents, Read Operations, Query Documents, Update Operations, Update Documents, Delete Operations, Delete Documents, Working with Arrays, Working with Arrays, Query for Null or Missing Fields, To Query Null or Missing Fields, Working with the limit() and skip() Methods, limit() and skip() Methods

**Unit V** **12 Periods**

Data Modelling and Aggregation, Data Models, Embedded Data Models, Normalized Data Models

Data Model Relationship Between Documents, Data Model Using an Embedded Document, Data Model Using Document References, SQL Aggregation Terms and Corresponding Mongo DB, Aggregation Operators, Matching SQL Aggregation to Mongo DB, Aggregation Operations .

**Text Book:**

1. "Fundamentals of Database Systems" by R. Elmasri and S.Navathe
2. "Database System Concepts" by Abraham Silberschatz, Henry Korth, and S. Sudarshan, Mc Grawhill, 2010.
3. Mongo DB Recipes: With Data Modeling and Query Building Strategies By Subhashini Chellappan, Dharanitharan Ganesan ,Publisher : Apress

**Reference Book:**

1. "Database Management Systems" by Raghu Ramakrishnan, NcGrawhill,2002
2. "Principles of Database Systems" by J.D.Ullman
3. Mongo DB Basics 1st ed. Edition ,by Peter Membrey (Author) Publisher **Apress :**Web Resources

**Web Links:**

1. https://docs.mongodb.com/manual/tutorial/getting-started
2. https://www.tutorialspoint.com/mongodb/index.htm

**Recommended Co – Curricular Activities:**

**A. Measurable**

1. Assignments (in writing and doing forms on the aspects of syllabus content and outside the syllabus content. Shall be individual and challenging)
2. Student seminars (on topics of the syllabus and related aspects (individual activity))
3. Quiz (on topics where the content can be compiled by smaller aspects and data (Individuals or groups as teams))
4. Study projects (by very small groups of students on selected local real-time problems pertaining to syllabus or related areas. The individual participation and contribution of students shall be ensured (team activity))

**B. General**

1. Group Discussion
2. Others

**RECOMMENDED CONTINUOUS ASSESSMENT METHODS:**

1. Programming exercises,
2. Practical assignments and laboratory reports,
3. Observation of practical skills,
4. Individual and group project reports.
5. Efficient delivery using seminar presentations,
6. Viva voce interviews.
7. Computerized adaptive testing, literature surveys and evaluations,
8. Peers and self-assessment, outputs form individual and collaborative work.

@@@@

**P.B.SIDDHARTHA COLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

**TITLE: Document Oriented Databases**

**COURSE CODE: AMLT32** **Max.Marks:75M**

**CLASS: B.Sc. (AI & ML)** **Semester III**

**Section-A**

**Answer any FIVE questions.** **5 X 5=25M**

1. How dependency preservation can be achieved?
2. Write about the usability of 'group by' and 'having' clauses in SQL.
3. Drop Database() Method
4. Explain any four SQL Aggregate operators with an example.
5. Advantages of Mongo DB over RDBMS
6. The pretty() Method
7. List out set operators can be used in SQL?
8. Differentiate File systems from DBMS.

**Section-B**

**Answer ALL questions.** **10 X 5 = 50M.**

9. (a) Explain the architecture of DBMS with a neat sketch.

   Or

   (b) Compare and contrast various Data Models
10. (a)Explain BCNF and the properties of decompositions

    Or

    (b) Briefly Explain about Mongo DB Data types
11. (a) Explain about Mongo DB Replication

    Or

    (b) Some considerations while designing Schema in Mongo DB
12. (a) Write any three data base applications with their functionalities.

    Or

    (b)Explain FOURTH and THIRD normal forms with examples.
13. (a) Elaborate the importance of computing closure of functional dependencies. Explain the procedure with an example.

    Or

    (b) Explain the Limit () Method
14. (a) What is the utilisation of Mongo DB Aggregation

    Or

    (b) List out any four operations on relational algebra. Explain.

@@@@

**P.B.SIDDHARTHA COLLEGE OF ARTS& SCIENCE**

Siddharthanagar, Vijayawada – 520010

***(An Autonomous College under the Jurisdiction of Krishna University)***

***Re– Accredited at 'A$^{+}$' by NAAC–III Cycle***

***College with Potential for Excellence***

***ISO9001 – 2015 Certified***

Course Code: AMLP32 Offeredto: **II B.Sc(AI & ML)**

Domain Subject: **COMPUTER SCIENCE** Semester:**III**

Max. Marks: **50**(CCIA:10+SEE:40) Practical Hrs./Week**:2**

**Credits:2**

## Document Oriented Databases LAB

**Course Objective:**

The objective of this course is to enable student to implement database related queries using Mongo DB.

| COURSE OUTCOME NO | UPON SUCCESSFUL COMPLETION OF THIS COURSE, STUDENTS SHOULD HAVE THE KNOWLEDGE AND SKILLS TO: | PROGRAM OUTCOME NO |
|---|---|---|
| CO1 | Familiarity of the Database Structures | PO5, PO7 |
| CO2 | DDL, DML commands and Query execution | PO5, PO7 |
| CO3 | Installation of mongo db ,configuring, running mongo db | PO5, PO7 |
| CO4 | Implementation of crud operations | PO5, PO7 |
| CO5 | Implementing multiple document transactions in mongo db | PO5, PO7 |

WEEK 1: Introduction to Database. Environment of Oracle / MySQL
Week 2: Creating the database with DDL Commands CRUD Operations
Week 3: Working with Queries
Week 4: Relationship among tables and queries
Week 5: Installation of Mongo db
Week 6: Difference between SQL and NOSQL.
Week 7: Creating Database structure in Mongo db
Week 8: CRUD Operations in Mongo db
Week 9: CRUD Operations in Mongo db
Week 10: Implement Aggregation in Mongo db
Week 11: Implement different functions in Mongo db
Week 12: Implement different functions in Mongo db
Week 13: Identifying the data storage in SQL and NOSQL
Week 14: Working with CaseStudy1
Week 15: Working with CaseStudy2

@@@@

**END OF THE DOCUMENT**