

SRR & CVR GOVT. DEGREE COLLEGE(A):: VIJAYAWADA
III BSc (Data Science)
Revised Syllabus 2025-26

Semester	Course Code	Course Title	Hours	Credits
V	DSC M-5305-1	SUPERVISED MACHINE LEARNING WITH PYTHON	45	3

Aim and objectives of Course:

The purpose of this course is to serve as an introduction to Supervised machine learning with Python.

We will explore several classifications, regression algorithms and see how they can help us perform a variety of Supervised machine learning tasks.

Learning outcomes of Course: On successful completion of the course, students will be able to

- understand introduction to machine learning concepts.
- Load datasets, build models and model persistence.
- Understand Feature extraction from data sets.
- do Regression & Classification.
- compare SVM with other classifiers.

UNIT I:

Machine Learning Basics: What is machine learning? Key terminology, Key tasks of machine learning, How to choose the right algorithm, steps in developing machine learning, why python? Getting started with Numpy library, Classifying with k- Nearest Neighbors: The k-Nearest Neighbors classification algorithm, Parsing and importing data from a text file, Creating scatter plots with Matplotlib, Normalizing numeric values.

UNIT II:

Splitting datasets one feature at a time-Decision trees: Introducing decision trees, measuring consistency in a dataset, using recursion to construct a decision tree, plotting trees in Matplotlib.

UNIT III:

Classifying with probability theory-Naïve Bayes: Using probability distributions for classification, learning the naïve Bayes classifier, parsing data from RSS feeds, using naïve Bayes to reveal regional attitudes.

UNIT IV:

Logistic regression: Classification with logistic regression and the sigmoid function, Using optimization to find the best regression coefficients, the gradient descent optimization algorithm, Dealing with missing values in our data.

UNIT V:

Support vector machines: Introducing support vector machines, using the SMO algorithm for optimization, using kernels to “transform” data, Comparing support vector machines with other classifiers.

TEXT BOOK:

1. Machine learning in action, Peter Harrington by Manning publications

SRR & CVR GOVT. DEGREE COLLEGE(A):: VIJAYAWADA
III BSc (Data Science)

Time : 3 hrs

MODEL PAPER

Total : 60 M

SUPERVISED MACHINE LEARNING WITH PYTHON

SECTION – A

ANSWER ANY FIVE QUESTIONS.

5X4=20M

1. What is machine learning?
2. What does the k in k-Nearest Neighbors stand for?
3. What is a decision tree?
4. What is recursion in building a decision tree?
5. What is the main idea of the Naïve Bayes algorithm?
6. How is probability used in Naïve Bayes?
7. What is the sigmoid function used for in logistic regression?
8. What is gradient descent?
9. What is the goal of a support vector machine?
10. What is the use of a kernel in SVM?

SECTION – B

ANSWER ANY FIVE QUESTIONS

5X8=40M

11. (a) Define machine learning in your own words. Explain different types of machine learning?

(OR)

- (b) How to use Matplotlib to create scatter plots for visualizing data?

12. (a) How to measure the consistency or purity of data in decision trees?

(OR)

- (b) Write the steps involved in building a simple decision tree using one feature at a time.

13. (a) Explain how probability distributions are used in Naïve Bayes classification.

(OR)

- (b) What are the advantages of using Naïve Bayes for text classification?

14. (a) What kind of problems can logistic regression solve? Give a simple example.

(OR)

- (b) What are the steps to train a logistic regression model from data?

15. (a) How does SVM find the best line (or hyperplane) to separate data?

(OR)

- (b) List and explain any two advantages of using SVM over other classifiers.

SRR & CVR GOVT. DEGREE COLLEGE(A):: VIJAYAWADA
III BSc (Data Science)
Revised Syllabus 2025-26
SUPERVISED MACHINE LEARNING WITH PYTHON
QUESTION BANK

Unit - I

Short Answer Type Questions

1. What is machine learning?
2. What does the k in k-Nearest Neighbors mean?
3. Why do we use Python for machine learning?
4. What is NumPy used for in machine learning?

Long Answer Type Questions

5. Write the basic steps to build a machine learning model.
6. Explain how the k-Nearest Neighbors algorithm works in simple words.
7. How can we read data from a text file using Python?
8. What is a scatter plot, and how is it used in machine learning?

Unit - II

Short Answer Type Questions

9. What is a decision tree in machine learning?
10. What does “splitting a dataset” mean?
11. Why do we use recursion in decision trees?
12. What is meant by a “feature” in a dataset?

Long Answer Type Questions

13. Explain how a decision tree splits data using one feature at a time.
14. What is consistency in a dataset, and why is it important when building a decision tree?
15. Describe the steps to create a decision tree using recursion in simple words.
16. How can we use Matplotlib to show a decision tree visually?

Unit - III

Short Answer Type Questions

17. What is Naïve Bayes used for in machine learning?
18. What is a probability distribution?

19. Why is it called “Naïve” Bayes?
20. What is the purpose of parsing data from RSS feeds?

Long Answer Type Questions

21. Explain how the Naïve Bayes classifier works in simple terms.
22. How can probability be used to classify data? Give an example.
23. Describe the steps to parse data from RSS feeds and use it in classification.
24. How can Naïve Bayes help in finding out people’s attitudes in different regions?

Unit - IV

Short Answer Type Questions

25. What is logistic regression used for?
26. What is the sigmoid function?
27. What does gradient descent help us do?
28. How can we handle missing values in a dataset?

Long Answer Type Questions

29. Explain how logistic regression works for classification problems.
30. What is the role of the sigmoid function in logistic regression?
31. Describe how gradient descent helps to find the best regression line.
32. What are some ways to deal with missing values in data?

Unit - V

Short Answer Type Questions

33. What is a support vector machine (SVM)?
34. What is the purpose of the SMO algorithm in SVM?
35. What is a kernel in SVM?
36. Why do we compare SVM with other classifiers?

Long Answer Type Questions

37. Explain how a support vector machine classifies data in simple terms.
38. What is the SMO algorithm, and how does it help in training an SVM?
39. Describe how kernels are used to transform data in SVM.
40. Compare support vector machines with at least one other classifier you have learned.

SRR&CVR GOVT DEGREE COLLEGE(A):: VIJAYAWADA
III BSc (Data Science)

Revised Syllabus 2025-26

Semester	Course Code	Course Title	Hours	Credits
V	DSC M-5305-1P	SUPERVISED MACHINE LEARNING WITH PYTHON LAB	30	1

List of Experiments:

1. Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a CSV file. For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.
2. Write a program to demonstrate the working of the decision tree based ID3 algorithm.
3. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a CSV file.
4. Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set.

SRR & CVR GOVT. DEGREE COLLEGE(A):: VIJAYAWADA
III BSc (Data Science)
Revised Syllabus 2025-26

Semester	Course Code	Course Title	Hours	Credits
V	DSC M-5305-2	UNSUPERVISED MACHINE LEARNING WITH PYTHON	45	3

Aim and objectives of Course:

- Unsupervised Machine Learning involves finding patterns in datasets.
- The core of this course involves study of Clustering, feature extraction and optimization algorithms.
- The purpose of this course is to serve as an introduction to machine learning with Python.

Learning outcomes of Course:

- Able to do Clustering, feature extraction and optimization.
- Students will be able to understand and implement in Python algorithms of Unsupervised Machine Learning and apply them to real-world datasets.

UNIT I:

Unsupervised Learning: Clustering: k-means clustering algorithm, improving cluster performance with post processing, Bisecting k-means, Example: clustering points on a map.

UNIT II:

Association analysis : Apriori algorithm: Association analysis, The Apriori principle, Finding frequent item sets with the Apriori algorithm, Mining association rules from frequent item sets, uncovering patterns in congressional voting.

UNIT III:

Finding frequent item sets: FP-growth –FP trees, Build FP-tree, mining frequent from an FP- tree, finding co-occurring words in a Twitter feed, mining a click stream from a news site.

UNIT IV:

Principal component analysis: Dimensionality reduction techniques, using PCA to reduce the dimensionality of semiconductor manufacturing data.

UNIT V:

Singular value decomposition: Applications of the SVD, Matrix factorization, SVD in Python, Collaborative filtering–based recommendation engines, a restaurant dish recommendation engine.

TEXT BOOK:

1. Machine learning in action, Peter Harrington by Manning publications Unsupervised ML with Python Lab.

SRR & CVR GOVT. DEGREE COLLEGE(A):: VIJAYAWADA
III BSc (Data Science)

Time : 3 hrs

MODEL PAPER

Total : 60 M

UNSUPERVISED MACHINE LEARNING WITH PYTHON

SECTION – A

ANSWER ANY FIVE QUESTIONS.

5X4=20M

1. What is k-means clustering and how does it group data?
2. What is bisecting k-means and how is it different from basic k-means?
3. What is the Apriori principle in association analysis?
4. What is a frequent item set in the Apriori algorithm?
5. What is an FP-tree used for in the FP-growth algorithm?
6. What kind of data can be mined using the FP-growth method?
7. What is dimensionality reduction?
8. How does PCA help in reducing the number of features in data?
9. What is Singular Value Decomposition (SVD)?
10. How is SVD used in recommendation systems?

SECTION – B

ANSWER ANY FIVE QUESTIONS

5X8=40M

11. (a). What is clustering? Write the steps of the k-means algorithm in simple words.
(OR)
(b). How can we use clustering to group locations on a map? Give a simple example.
12. (a). What is association analysis? Give some examples where it can be used.
(OR)
(b). How does the Apriori algorithm help in finding hidden patterns in data?
13. (a). What are the steps to build an FP-tree in the FP-growth method?
(OR)
(b). How can FP-growth find common items from user clicks or tweets?
14. (a). Why do we reduce the number of features in data? What are the benefits?
(OR)
(b). How is PCA used to simplify data in manufacturing?
15. (a). What is matrix factorization? How is it useful in machine learning?
(OR)
(b). How does SVD help in recommending food items in a restaurant app?

SRR & CVR GOVT. DEGREE COLLEGE(A):: VIJAYAWADA

III BSc (Data Science)

Revised Syllabus 2025-26

UNSUPERVISED MACHINE LEARNING WITH PYTHON

QUESTION BANK

Unit - I

Short Answer Type Questions

1. What does k-means clustering do?
2. What is a “cluster” in data?
3. What does the “k” mean in k-means?
4. Why do we sometimes need to change the clusters after k-means finishes?

Long Answer Type Questions

5. Explain the steps of how k-means groups data points.
6. What is bisecting k-means, and how is it different from normal k-means?
7. How can clustering help us find groups of places on a map?
8. Why do we use Python to do clustering in the lab?

Unit - II

Short Answer Type Questions

9. What is association analysis used for?
10. What does the Apriori principle help us do?
11. What are “frequent item sets” in the Apriori algorithm?
12. Why do we want to find association rules?

Long Answer Type Questions

13. Explain the Apriori algorithm in simple steps.
14. How does the Apriori principle make finding patterns easier?
15. Describe how association rules are created from frequent item sets.
16. Give an example of how association analysis can be used to understand voting patterns.

Unit - III

Short Answer Type Questions

17. What is the purpose of the FP-growth algorithm?

18. What is an FP-tree?
19. How does FP-growth find frequent item sets faster than Apriori?
20. What kind of data can we analyze using FP-growth?

Long Answer Type Questions

21. Explain how an FP-tree is built in the FP-growth algorithm.
22. Describe how FP-growth mines frequent item sets from an FP-tree.
23. How can FP-growth be used to find common words in Twitter posts?
24. Explain how mining clickstream data from a news site can help improve user experience.

Unit - IV

Short Answer Type Questions

25. What is dimensionality reduction?
26. Why do we use Principal Component Analysis (PCA)?
27. What does PCA do to the data?
28. Give an example of a field where PCA can be used.

Long Answer Type Questions

29. Explain in simple terms how PCA reduces the number of features in data.
30. Why is reducing dimensionality important in machine learning?
31. Describe how PCA can be applied to semiconductor manufacturing data.
32. What are some benefits and challenges of using PCA?

Unit - V

Short Answer Type Questions

33. What is Singular Value Decomposition (SVD)?
34. What does matrix factorization mean?
35. How is SVD used in recommendation engines?
36. Give an example of a recommendation system that can use SVD.

Long Answer Type Questions

37. Explain the basic idea of Singular Value Decomposition in simple terms.
38. Describe how matrix factorization helps in collaborative filtering.
39. How can SVD be used to build a recommendation engine for restaurant dishes?
40. What are some practical applications of SVD besides recommendation systems?

SRR&CVR GOVT DEGREE COLLEGE(A):: VIJAYAWADA
III BSc (Data Science)

Revised Syllabus 2025-26

Semester	Course Code	Course Title	Hours	Credits
V	DSC M-5305-2P	UNSUPERVISED MACHINE LEARNING WITH PYTHON LAB	30	1

List of Experiments:

1. Implementation of K-Means Clustering
2. Implement the bisecting k-means clustering algorithm
3. Implement Apriori algorithm
4. Implement Association rule-generation functions
5. Implement FP-tree creation
6. Write a function to find all paths ending with a given item.
7. Implement Code to access the Twitter Python library
8. Implement the PCA algorithm
9. Write a program to find Rating estimation by using the SVD
10. Implement Image-compression functions using SVD.

SRR&CVR GOVT DEGREE COLLEGE(A):: VIJAYAWADA
III BSc (Data Science)

Revised Syllabus 2025-26

Semester	Course Code	Course Title	Hours	Credits
V	DSCS M-5305-3B	PREDICTIVE AND ADVANCED ANALYTICS USING R	45	3

Course Objectives

To build and apply predictive models using R and to understand how to choose, combine, and improve models for better prediction and decision-making.

Course Outcomes

Upon successful completion of the course, the students will be able to

1. Understand the process of formulating business objectives, data selection/collection and preparation.
2. Understand the process to successfully design, build, evaluate and implement predictive models for various business applications.
3. Compare the underlying predictive modelling techniques.
4. Select appropriate predictive modelling approaches to identify cases to progress with.
5. Apply predictive modelling approaches using a suitable package

UNIT – I

Introduction to Data Mining: Introduction, what is Data Mining? Concepts of Data mining, Technologies Used, Data Mining Process, KDD Process Model, CRISP – DM, Mining on various kinds of data, Applications of Data Mining, Challenges of Data Mining.

UNIT – II

Data Understanding and Preparation: Introduction, Reading data from various sources, Data visualization, Distributions and summary statistics, Relationships among variables, Extent of Missing Data. Segmentation, Outlier detection

UNIT – III

Model development & techniques: Data Partitioning, Model selection, Model Development Techniques, Neural networks, Decision trees, Logistic regression, Discriminant analysis, Support vector machine, Bayesian Networks, Linear Regression, Cox Regression, Association rules.

UNIT – IV

Automated Data Preparation, Combining data files, Aggregate Data, Duplicate Removal, Sampling DATA, Data Caching, Partitioning data, Missing Values. Model Evaluation and Deployment Introduction, Model Validation, Rule Induction Using CHAID

UNIT – V

Automating Models for Categorical and Continuous targets, Comparing and Combining Models, Evaluation Charts for Model Comparison, Deploying Model, Assessing Model Performance, Updating a Model.

Recommended Text Books:

1. Predictive & Advanced Analytics (IBM ICE Publication)
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman , The Elements of Statistical Learning-Data Mining, Inference, and Prediction ,Second Edition , Springer Verlag,2009.

SRR&CVR GOVT DEGREE COLLEGE(A):: VIJAYAWADA
III BSc (Data Science)

Time : 3 hrs

MODEL PAPER

Total : 60 M

PREDICTIVE AND ADVANCED ANALYTICS USING R

SECTION – A

Answer any five of the following questions

5 X 4 = 20 M

1. What is data mining?
2. What are the applications of data mining?
3. What are outliers?
4. What is segmentation?
5. Discuss about Decision Trees
6. What is a neural network?
7. Explain the purpose of splitting data into training, validation, and test sets.
8. What is model selection and why is it important in predictive modeling?
9. Name and briefly describe two common evaluation charts used for model comparison.
10. Distinguish between categorical and continuous targets in model automation.

SECTION – B

Answer the following questions

5 X 8 = 40 M

11. a) Explain the stages of KDD process.
(OR)
b) Explain about data mining technologies.
12. a) Describe different types of data attributes and basic statistical measures used to analyze them.
(OR)
b) Explain the types of data visualizations in R.
13. a) Explain about linear, multiple, and logistic regression techniques.
(OR)
b) Explain Support Vector Machines and Bayesian networks.
14. a) What is rule induction and how is CHAID used for it?
(OR)
b) Describe the importance of data cleaning.
15. a) Discuss ensemble methods for combining models.
(OR)
b) Explain how model performance is assessed?

SRR&CVR GOVT DEGREE COLLEGE(A):: VIJAYAWADA

III BSc (Data Science)

Revised Syllabus 2025-26

PREDICTIVE AND ADVANCED ANALYTICS USING R

Question Bank

UNIT -I

Short answer questions

1. What is data mining?
2. What are the applications of data mining?
3. Mention key challenges in data mining.

Long Answer Questions

1. Explain about data mining technologies
2. What are the key steps in the KDD (Knowledge Discovery in Databases) process?
3. What are the different kinds of data on which data mining can be performed?

UNIT -II

Short answer questions

1. Write a short note on distributions and summary statistics
2. Discuss about outlier detection?
3. What is segmentation?

Long Answer Questions

1. Describe different types of data attributes and basic statistical measures used to analyze them.
2. Explain the types of data visualizations in R
3. Explain methods to explore relationships among variables

UNIT -III

Short answer questions

1. What is Cox Regression and where is it used?
2. Discuss about Decision Trees
3. What is a neural network?

Long Answer Questions

1. Discuss association rule mining
2. Explain about linear, multiple, and logistic regression techniques
3. Explain Support Vector Machines and Bayesian networks

UNIT -IV

Short answer questions

1. Explain the purpose of splitting data into training, validation, and test sets.
2. What is model selection and why is it important in predictive modeling?
3. Explain how data aggregation and combining files are crucial steps

Long Answer Questions

1. Describe the importance of data cleaning

2. What are the methods used for partitioning data and caching during modeling?
3. What is rule induction and how is CHAID used for it?

UNIT -V

Short answer questions

1. Name and briefly describe two common evaluation charts used for model comparison.
2. Discuss how models are compared.
3. Distinguish between categorical and continuous targets in model automation.

Long answer questions

1. Discuss ensemble methods for combining models
2. Explain how model performance is accessed?
3. What are the different methods used to update a predictive model when new data becomes available?

SRR&CVR GOVT DEGREE COLLEGE(A):: VIJAYAWADA
III BSc (Data Science)

Revised Syllabus 2025-26

Semester	Course Code	Course Title	Hours	Credits
V	DSCS M-5305-3BP	PREDICTIVE AND ADVANCED ANALYTICS USING R LAB	30	1

List of Experiments:

Implementation of following methods using R or Matlab

1. Simple and multiple linear regression
2. Logistic regression
3. Linear discriminant analysis
4. Ridge regression
5. Cross-validation and boot strap
6. Fitting classification and regression trees
7. K-nearest neighbors
8. Principal component analysis
9. K-means clustering

SRR&CVR GOVT DEGREE COLLEGE(A):: VIJAYAWADA
III BSc (Data Science)

Revised Syllabus 2025-26

Semester	Course Code	Course Title	Hours	Credits
V	DSCS M-5305-4B	DATA WRANGLING WITH JAVA SCRIPT	45	3

Course Objectives

To build and apply predictive models using R and to understand how to choose, combine, and improve models for better prediction and decision-making.

Course Outcomes

Upon successful completion of the course, the students will be able to

1. Understand the process of formulating business objectives, data selection/collection and preparation.
2. Understand the process to successfully design, build, evaluate and implement predictive models for various business applications.
3. Compare the underlying predictive modelling techniques.
4. Select appropriate predictive modelling approaches to identify cases to progress with.
5. Apply predictive modelling approaches using a suitable package

UNIT – I

Getting started: establishing your data pipeline - Why data wrangling-What's data wrangling -Why use JavaScript for data wrangling- Is JavaScript appropriate for data analysis? Navigating the JavaScript ecosystem - Establishing your data pipeline

UNIT - II

Getting started with Node.js - Building a simple reporting system -Getting the code and data - Viewing the code - Installing Node.js - Running Node.js code -Running a web application - Getting the data - Checking your Node.js version -Working with Node.js Creating a Node.js project - Creating a command-line application - Creating a code library -Creating a simple web server - Asynchronous coding - Loading a single file - Loading multiple files – Error handling - Asynchronous coding with promises - Wrapping asynchronous operations.

UNIT - III

Acquisition, storage, and retrieval -Getting the code and data -The core data representation- Loading data from text files - Loading data from a REST API-Parsing JSON text data-Parsing CSV text data- Importing data from databases - Importing data from MongoDB-Importing data from MySQL - Exporting data - Exporting data to text files - Exporting data to JSON text files - Exporting data to CSV text files- Exporting data to a database - Exporting data to MongoDB- Exporting data to MySQL.

UNIT - IV

Exploratory coding - Iteration and your feedback loop - A first pass at understanding your data - Working with a reduced data sample - Prototyping with Excel - Exploratory coding with Node.js -Using Nodemon -Exploring your data - Using Data-Forge - Computing the trend column - Outputting a new CSV file - Exploratory coding in the browser .

UNIT - V

Clean and prepare - The need for data cleanup and preparation - Where does broken data come from? - How does data cleanup fit into the pipeline- Identifying bad data – Techniques for fixing bad data - Cleaning our data set -Preparing our data for effective use.

TEXTBOOK

1. Data Wrangling with JavaScript - Ashley Davis - Manning Publication

REFERENCE TEXTBOOKS:

1. Principles of Data Wrangling - Practical Techniques for Data Preparation - Tye Rattenbury, Joseph M. Hellerstein, Jeffrey Heer, Sean Kandel, Connor Carreras - Oreilly Publication.
2. Data Wrangling with Python - Jacqueline Kazil, Katharine Jarmul 2016-Oreilly Publication.

SRR&CVR GOVT DEGREE COLLEGE(A):: VIJAYAWADA
III BSc (Data Science)

Time : 3 hrs

MODEL PAPER

Total : 60 M

DATA WRANGLING WITH JAVA SCRIPT

SECTION – A

Answer any FIVE of the following questions

5 X 4 = 20 M

1. What is Data Wrangling?
2. Explain about java script objects.
3. How to create and run a webserver
4. How does Promise.then() work? Explain with an example?
5. How do you load JSON data from a REST API in Node.js?
6. Explain the concept of the Core Data Representation (CDR) in a data pipeline
7. What is a feedback loop in exploratory coding?
8. How can Excel be used for data prototyping?
9. List main sources of broken data and give one example for each.
10. Where does data cleanup fit into a data pipeline?

SECTION – B

Answer the following questions

5 X 8 = 40 M

11. a) Why is Data Wrangling important?
(OR)
b) Why is it important to establish a Data Pipeline For data wrangling?
12. a) What are the main components of Node.js project? Explain.
(OR)
b) How are reusable functions created and then imported into a Node.js application?
13. a) Explain how to read data from a .csv file and store it in MongoDB Atlas.
(OR)
b) Explain how to read JSON data from a REST API in Node.js and convert it into JavaScript objects.
14. a) Describe the process of exploratory coding using Node.js with suitable examples.
(OR)
b) Describe the process of exploratory coding in the browser.
15. a) Explain why data cleanup and preparation is important?
(OR)
b) Explain techniques used to identify bad data in a dataset and techniques for fixing bad data.

SRR&CVR GOVT DEGREE COLLEGE(A):: VIJAYAWADA

III BSc (Data Science)

Revised Syllabus 2025-26

DATA WRANGLING WITH JAVA SCRIPT

UNIT - I

Short answer questions

1. What is Data Wrangling?
2. Why is it important to understand how to "Navigate the JavaScript Ecosystem"?
3. Explain about java script objects.

Long answer questions

1. Why is Data Wrangling important?
2. Why use JavaScript for Data Wrangling?
3. Why is it important to establish a Data Pipeline For data wrangling?

UNIT - II

Short answer questions

1. How do you add and manage third-party dependencies in a Node.js project using npm?
2. How does Promise.then() work? Explain with an example?
3. What is the role of .catch() in Promise chains?
4. How to create and run a webserver

Long Answer Questions

1. What are the main components of Node.js project? Explain.
2. How are reusable functions created and then imported into a Node.js application?
3. What are Promises? Differentiate between Promise and Asynchronous functions

UNIT III

Short answer questions

1. Explain the concept of the Core Data Representation (CDR) in a data pipeline
2. Explain how to load data from a csv file in Node.js.
3. Explain how to read JSON data from a REST API in Node.js and convert it into JavaScript objects.

Long Answer Questions

1. Explain how to load a csv file using a Promise-based function in Node.js.
2. Explain how to import data from MongoDB Atlas using Node.js and Mongoose
3. Explain how to import data from MYSQL

UNIT IV

Short Answer Questions

1. What is a feedback loop in exploratory coding?
2. Explain how to reduce a data sample and why it's useful in data exploration.
3. How can Excel be used for data prototyping?

Long Answer Questions

1. Describe the process of exploratory coding in the browser
2. Describe the process of exploratory coding using Node.js with suitable examples.
3. Explain what is the use of nodemon and liveserver

UNIT V

Short Answer Questions

1. List main sources of broken data and give one example for each.
2. Where does data cleanup fit into a data pipeline?

Long Answer Questions

1. Explain why data cleanup and preparation is important?
2. Discuss the techniques for preparing data after cleaning
3. Explain about the Techniques for Cleaning a Data Set
4. Explain techniques used to identify bad data in a dataset and techniques for fixing bad data

SRR&CVR GOVT DEGREE COLLEGE(A):: VIJAYAWADA
III BSc (Data Science)

Revised Syllabus 2025-26

Semester	Course Code	Course Title	Hours	Credits
V	DSCS M-5305-4BP	DATA WRANGLING WITH JAVA SCRIPT LAB	30	1

List of Experiments:

1. Installing steps for Node.js, Installing npm dependencies, printing version of Node.js.
Developing a Simple program to print “Hello World” on console.
2. Create a Code library to a command-line application in Node.js to produce a report from data.
3. Install “Express Node.js framework” . Create a simple Web application to print your college name on Web Page.
4. Develop an Web application to add static files to web server.
5. Develop an application to import & export data using MongoDB.
6. Develop an application to import & export data using MYSQL.
7. Loading your input CSV file and printing its contents to the console. Apply slicing operation on data and print on Console. Print datatype of each column(attribute) of CSV file.
8. List column names and data types of column names of a CSV file using Data-Forge (An open source data-wrangling toolkit for JavaScript).
9. Develop a web visualization application by importing data from Excel file.
10. Develop an application to do Excel data analysis in Node.js using Formulas.
11. Develop a program to clean your data by rewriting rows to fix bad data.