

SRI VENKATESWARA UNIVERSITY :: TIRUPATI

B.Sc., Data Science Honours MAJOR w.e.f AY 2025-26 onwards V Semester Course Structure

Year	Semester	Paper	Title of the Course	No. of Hrs./ Week	No. of Credits
3	V	12	Supervised Machine Learning with Python	3	3
			Supervised Machine Learning with Python (P)	2	1
		13	Unsupervised Machine Learning with Python (T)	3	3
			Unsupervised Machine Learning with Python (P)	2	1
		14A	Web Scraping with Python (T)	3	3
			Web Scraping with Python (P)	2	1
			OR		
		14B	Predictive & Advanced Analytics using R (T)	3	3
			Predictive & Advanced Analytics using R (P)	2	1
		15A	Advanced Data Analysis Using Python -(T)	3	3
			Advanced Data Analysis Using Python -(P)	2	1
			OR		
		15B	Data Wrangling with Java Script -(T)	3	3
			Data Wrangling with Java Script -(P)	2	1

Note:

choose either Set-A (14 A and 15 A) (or) Set-B (14 B and 15 B)

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III year V Semester
Course 12: Supervised Machine Learning with Python
(w.e.f. 2025-26)

Theory

Credits: 3

3 hrs/week

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:

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

UNIT I:

Machine Learning Basics: What is machine learning? Key terminology, Key tasks of machine learning, How to choose right algorithm, steps in developing a 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: Tree construction, Plotting trees in Python with Matplotlib annotations, Testing and storing the classifier, Example: using decision trees to predict contact lens type.

UNIT III:

Classifying with probability theory-Naïve Bayes: Classifying with Bayesian decision theory, Conditional probability, Classifying with conditional probabilities, Document classification with naïve Bayes, Classifying text with Python, Example: classifying spam email with naïve Bayes, Example: using naïve Bayes to reveal local attitudes from personal ads.

UNIT IV:

Logistic regression: Classification with logistic regression and the sigmoid function, Using optimization to find the best regression coefficients, Example: estimating horse fatalities from colic.

UNIT V:

Support vector machines: Separating data with the maximum margin, Finding the maximum margin, Efficient optimization with the SMO algorithm, Speeding up optimization with the full Platt SMO, Using kernels for more complex data, Example: revisiting handwriting classification.

TEXT BOOK:

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

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III year V Semester
Course 12: Supervised Machine Learning with Python
(w.e.f. 2025-26)

Practical

Credits: 1

2 hrs/week

Details of Lab/Practical/Experiments/Tutorials syllabus:

S.No Experiment Name

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.

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III year V Semester
Course 12: Supervised Machine Learning with Python
(w.e.f. 2025-26)

MODEL QUESTION PAPER

Time: **3 Hours**

Max. Marks: **75**

PART-A

Answer any **FIVE** of the following. Each Question Carries 5 marks.

(5X5=25)

1. Short answer question from unit-I
2. Short answer question from unit-I
3. Short answer question from unit-II
4. Short answer question from unit-II
5. Short answer question from unit-III
6. Short answer question from unit-III
7. Short answer question from unit-IV
8. Short answer question from unit-IV
9. Short answer question from unit-V
10. Short answer question from unit-V

PART-B

Answer any **FIVE** of the following. Each Question Carries 10 marks.

(5X10=50)

11. a) Essay question from Unit-I
OR
b) Essay question from Unit-I
12. a) Essay question from Unit-II
OR
b) Essay question from Unit-II
13. a) Essay question from Unit-III
OR
b) Essay question from Unit-III
14. a) Essay question from Unit-IV
OR
b) Essay question from Unit-IV
15. a) Essay question from Unit-V
OR
b) Essay question from Unit-V

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III year V Semester
Course 13: Unsupervised Machine Learning with Python
(w.e.f. 2025-26)

Theory

Credits: 3

3 hrs/week

Aim and objectives of Course (Unsupervised ML with Python):

- 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.

Syllabus: (Total Hours: 90 including Teaching, Lab and internal exams, etc.)

UNIT I:

Unsupervised Learning: The k-means clustering algorithm, Improving cluster performance with postprocessing, Bisecting k-means, Example: clustering points on a map.

UNIT II:

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

UNIT III:

Finding frequent item sets: FP-trees: an efficient way to encode a dataset, Build an FP-tree, Mining frequent items from an FP-tree, Example: finding co-occurring words in a Twitter feed, Example: mining a click stream from a news site.

UNIT IV:

Principal component analysis: Dimensionality reduction techniques, Principal component analysis, Example: 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

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III year V Semester
Course 13: Unsupervised Machine Learning with Python (w.e.f. 2025-26)

Practical	Credits: 1	2 hrs/week
------------------	-------------------	-------------------

S.No	Name of the Experiment
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.

SRI VENKATESWARA UNIVERSITY: : TIRUPATI
B.Sc. Data Science Honours
III year V Semester
Course 13: Unsupervised Machine Learning with Python

(w.e.f. 2025-26)
MODEL QUESTION PAPER

Time: **3 Hours**

Max. Marks: **75**

PART-A

Answer any **FIVE** of the following. Each Question Carries 5 marks.

(5X5=25)

1. Short answer question from unit-I
2. Short answer question from unit-I
3. Short answer question from unit-II
4. Short answer question from unit-II
5. Short answer question from unit-III
6. Short answer question from unit-III
7. Short answer question from unit-IV
8. Short answer question from unit-IV
9. Short answer question from unit-V
10. Short answer question from unit-V

PART-B

Answer any **FIVE** of the following. Each Question Carries 10 marks.

(5X10=50)

11. a) Essay question from Unit-I
OR
b) Essay question from Unit-I
12. a) Essay question from Unit-II
OR
b) Essay question from Unit-II
13. a) Essay question from Unit-III
OR
b) Essay question from Unit-III
14. a) Essay question from Unit-IV
OR
b) Essay question from Unit-IV
15. a) Essay question from Unit-V
OR
b) Essay question from Unit-V

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III year V Semester
Course 14 A: Web Scraping With Python
(w.e.f. 2025-26)

Theory

Credits: 3

3 hrs/week

Learning Outcomes

Students at the successful completion of the course will be able to:

1. Parse complicated HTML pages
2. Develop crawlers with the Scrapy framework
3. Learn methods to store data you scrape
4. Read and extract data from documents
5. Clean and normalize badly formatted data
6. Read and write natural languages
7. Crawl through forms and logins
8. Scrape JavaScript and crawl through APIs
9. Use and write image-to-text software
10. Avoid scraping traps and bot blockers
11. Use scrapers to test your website

Detailed Syllabus: (Five units with each unit having 12 hours of class work)

UNIT-1

Building Scrapers: First Web Scraper: Connecting, An Introduction to BeautifulSoup, **Advanced HTML Parsing:** Another Serving of BeautifulSoup, Regular Expressions, Regular Expressions and BeautifulSoup, Accessing Attributes, Lambda Expressions, **Writing Web Crawlers:** Traversing a Single Domain, Crawling an Entire Site, Crawling Across the Internet.

UNIT-II

Web Crawling Models: Planning and Defining Objects, Dealing with Different Website Layouts, Structuring Crawlers. **Scrapy:** Installing Scrapy, Initializing a New Spider, Writing a Simple Scraper, Spidering with Rules, Creating Items, Outputting Items, The Item Pipeline, Logging with Scrapy. **Storing Data:** Media Files, Storing Data to CSV, MySQL: Integrating with Python, Database Techniques and Good Practice, Six Degrees in MySQL, Email

UNIT-III

Advanced Scraping:Reading Documents: Document Encoding, Text, CSV, PDF, Microsoft Word and .docx,**Reading and Writing Natural Languages:** Summarizing Data, Markov Models, Natural Language Toolkit. **Crawling Through Forms and Logins:** Python Requests Library, Submitting a Basic Form, Radio Buttons, Checkboxes, and Other Inputs, Submitting Files and Images, Handling Logins and Cookies.

UNIT-IV

Crawling Through APIs: A Brief Introduction to APIs, Parsing JSON, Undocumented APIs, Finding Undocumented APIs, Documenting Undocumented APIs, Finding and Documenting APIs Automatically.

Image Processing and Text Recognition: Overview of Libraries, Pillow, Processing Well-Formatted Text, Reading CAPTCHAs and Training Tesseract, Retrieving CAPTCHAs and Submitting Solutions

UNIT-V

Avoiding Scraping Traps: A Note on Ethics, Looking Like a Human, Common Form Security Features, The Human Checklist Testing Your Website with Scrapers: An Introduction to Testing, Python unit test, Testing with Selenium, unittest or Selenium.

Text Books:

1. “Web Scraping with Python, by Ryan Mitchell”, 2nd Edition, O'Reilly Media, Inc., 2018.
2. “Hands-On Web Scraping with Python”, by Anish Chapagain, Packt Publishing, 2019.

References:

1. “Getting Structured Data from the Internet”, by Jay M. Patel, Apress, 2020
2. “Python Web Scraping Cookbook”, by Michael Heydt, Packt Publishing, 2018.
3. “Web Scraping with Python, by Richard Lawson”, Packt Publishing, 2015.

CO-CURRICULAR ACTIVITIES:

- Quiz Competition, Expert Lectures, Seminars

EXTRA CURRICULAR ACTIVITIES:

- Formal Examination, Lab Practical, Presentation , Simple Projects

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III year V Semester
Course 14 A: Web Scraping With Python
(w.e.f. 2025-26)

Practical	Credits: 1	2 hrs/week
------------------	-------------------	-------------------

S.No	Name of the Experiment
1.	Write a Python program to test if a given page is found or not on the server.
2.	Write a Python program to download and display the content of robot.txt for en.wikipedia.org.
3.	Write a Python program to get the number of datasets currently listed on data.gov.
4.	Write a Python program to convert an address into geographic coordinates (like latitude and longitude).
5.	Write a Python program to display the name of the most recently added dataset on data.gov.
6.	Write a Python program to extract h1 tag from example.com.
7.	Write a Python program to extract and display all the header tags from en.wikipedia.org/wiki/Main_Page
8.	Write a Python program to extract and display all the image links from a website.
9.	Write a Python program to get 90 days of visits broken down by browser for all sites on data.gov.
10.	Write a Python program to that retrieves an arbitrary Wikipedia page of "Python" and creates a list of links on that page.
11.	Write a Python program to check whether a page contains a title or not.
12.	Write a Python program to list all language names and number of related articles in the order they appear in wikipedia.org.
13.	Write a Python program to get the number of followers of a given twitter account.
14.	Write a Python program to find the live weather report (temperature, wind speed, description and weather) of a given city.
15.	Write a Python program to display the date, days, title, city, country of next 25 events.
16.	Write a Python program to download IMDB's Top 250 data (movie name, Initial release, director name and stars).
17.	Write a Python program to get movie name, year and a brief summary of the top 10 random movies

SRI VENKATESWARA UNIVERSITY: : TIRUPATI

**B.Sc. Data Science Honours
III year V Semester
Course 14 A: Web Scrapping With Python
(w.e.f. 2025-26)
MODEL QUESTION PAPER**

Time: **3 Hours**

Max. Marks: **75**

PART-A

Answer any **FIVE** of the following. Each Question Carries 5 marks.

(5X5=25)

1. Short answer question from unit-I
2. Short answer question from unit-I
3. Short answer question from unit-II
4. Short answer question from unit-II
5. Short answer question from unit-III
6. Short answer question from unit-III
7. Short answer question from unit-IV
8. Short answer question from unit-IV
9. Short answer question from unit-V
10. Short answer question from unit-V

PART-B

Answer any **FIVE** of the following. Each Question Carries 10 marks.

(5X10=50)

11. a) Essay question from Unit-I
OR
b) Essay question from Unit-I
12. a) Essay question from Unit-II
OR
b) Essay question from Unit-II
13. a) Essay question from Unit-III
OR
b) Essay question from Unit-III
14. a) Essay question from Unit-IV
OR
b) Essay question from Unit-IV
15. a) Essay question from Unit-V
OR
b) Essay question from Unit-V

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III year V Semester
Course 14 B: Predictive And Advanced Analytics Using R
(w.e.f. 2025-26)

Theory

Credits: 3

3 hrs/week

Aim and objectives of Course (Predictive and Advanced Analytics):

The course enables students to:

- To learn, how to develop models to predict categorical and continuous outcomes, using such techniques as neural networks, decision trees, logistic regression, support vector machines and Bayesian network models.
- To know the use of the binary classifier and numeric predictor nodes to automate model selection.
- To advice on when and how to use each model. Also learn how to combine two or more models to improve prediction

Learning outcomes of Course (In consonance with the Bloom's Taxonomy): The students will be able to:

- Understand the process of formulating business objectives, data selection/collection, preparation and process to successfully design, build, evaluate and implement predictive models for a various business application.
- Compare the underlying predictive modeling techniques.
- Select appropriate predictive modeling approaches to identify cases to progress with.
- Apply predictive modeling approaches using a suitable package such as SPSS Modeler

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 Book:

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.

Cocurricular Activities:

- Quiz Competition, Expert Lectures, Seminars

Assessment Methods:

- Formal Examination, Lab Practical, Presentation, Simple Projects

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III year V Semester
Course 14 B: Predictive And Advanced Analytics Using R
(w.e.f. 2025-26)

Practical

Credits: 1

2 hrs/week

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

SRI VENKATESWARA UNIVERSITY: : TIRUPATI
B.Sc. Data Science Honours
III year V Semester
Course 14 B: Predictive And Advanced Analytics Using R
(w.e.f. 2025-26)
MODEL QUESTION PAPER

Time: **3 Hours**

Max. Marks: **75**

PART-A

Answer any **FIVE** of the following. Each Question Carries 5 marks.

(5X5=25)

1. Short answer question from unit-I
2. Short answer question from unit-I
3. Short answer question from unit-II
4. Short answer question from unit-II
5. Short answer question from unit-III
6. Short answer question from unit-III
7. Short answer question from unit-IV
8. Short answer question from unit-IV
9. Short answer question from unit-V
10. Short answer question from unit-V

PART-B

Answer any **FIVE** of the following. Each Question Carries 10 marks.

(5X10=50)

11. a) Essay question from Unit-I
OR
b) Essay question from Unit-I
12. a) Essay question from Unit-II
OR
b) Essay question from Unit-II
13. a) Essay question from Unit-III
OR
b) Essay question from Unit-III
14. a) Essay question from Unit-IV
OR
b) Essay question from Unit-IV
15. a) Essay question from Unit-V
OR
b) Essay question from Unit-V

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III year V Semester
Course 15A: Advanced Data Analysis Using Python
(w.e.f. 2025-26)

Theory

Credits: 3

3 hrs/week

Course outcomes:

This course will enable the student to:

- Present an overview data science and applications.
- Plan the methods of data collection.
- Describe the statistical methods in EDA.
- Apply statistical methods to develop and evaluate the models.
- Becoming an expert in decision making for complex projects.

UNIT I – INTRODUCTION

Introduction to Data Science – Evolution of Data Science – Data Science Roles – Stages in a Data Science Project – Applications of Data Science in various fields – Data Security Issues.

UNIT II – DATA COLLECTION AND PRE-PROCESSING

Data Collection Strategies – Data Pre-Processing Overview – Data Cleaning – Data Integration and Transformation – Data Reduction – Data Discretization.

UNIT III – EXPLORATORY DATA ANALYTICS

Descriptive Statistics – Mean, Standard Deviation, Skewness and Kurtosis – Box Plots – Pivot Table – Heat Map – Correlation Statistics – ANOVA.

UNIT IV – MODEL DEVELOPMENT

Simple and Multiple Regression – Model Evaluation using Visualization – Residual Plot – Distribution Plot – Polynomial Regression and Pipelines – Measures for In-sample Evaluation – Prediction and Decision Making.

UNIT V – MODEL EVALUATION

Generalization Error – Out-of-Sample Evaluation Metrics – Cross Validation – Overfitting – Under Fitting and Model Selection – Prediction by using Ridge Regression – Testing Multiple Parameters by using Grid Search.

Text Books:

1. Jojo Moolayil, “Smarter Decisions: The Intersection of IoT and Data Science”, PACKT, 2016.
2. Cathy O’Neil and Rachel Schutt , “Doing Data Science”, O'Reilly, 2015.
3. David Dietrich, Barry Heller, Beibei Yang, “Data Science and Big data Analytics”, EMC 2013
4. Raj, Pethuru, “Handbook of Research on Cloud Infrastructures for Big Data Analytics”, IGI Global.

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III year V Semester
Course 15A: Advanced Data Analysis Using Python
(w.e.f. 2025-26)
Credits: 1

Practical

2 hrs/week

List of Experiments:

1. Creating a Data Frame and Matrix-like Operations on a Data Frame, Merging two Data Frames
2. Applying functions to Data Frames, import of external data in various file formats, statistical functions, compilation of data.
3. Using Functions with Factors
4. Accessing the Internet
5. Visualization Effects
6. Plotting with Layers
7. Overriding Aesthetics
8. Histograms and Density Charts
9. Simple Linear Regression – Fitting, Evaluation and Visualization
10. Multiple Linear Regression, Lasso and Ridge Regression
11. Use the following scenarios:
12. Use the Diabetes data set from UCI and Pima Indians Diabetes data set for performing the following:
 - a. Univariate Analysis: Frequency, Mean, Median, Mode, Variance, Standard Deviation, Skewness and Kurtosis.
 - i. Bivariate Analysis: Linear and logistic regression modeling.
 - ii. Multiple Regression Analysis
 - iii. Also Compare the results of the above analysis for the two data sets.
 - b. Data Modelling
 - i. Apply Bayesian and SVM techniques on Iris and Diabetes data set.
 - ii. Apply and explore various plotting functions on UCI data sets

SRI VENKATESWARA UNIVERSITY: : TIRUPATI
B.Sc. Data Science Honours
III year V Semester
Course 15A: Advanced Data Analysis Using Python
(w.e.f. 2025-26) (w.e.f. 2025-26)
MODEL QUESTION PAPER

Time: **3 Hours**

Max. Marks: **75**

PART-A

Answer any **FIVE** of the following. Each Question Carries 5 marks.

(5X5=25)

11. Short answer question from unit-I
12. Short answer question from unit-I
13. Short answer question from unit-II
14. Short answer question from unit-II
15. Short answer question from unit-III
16. Short answer question from unit-III
17. Short answer question from unit-IV
18. Short answer question from unit-IV
19. Short answer question from unit-V
20. Short answer question from unit-V

PART-B

Answer any **FIVE** of the following. Each Question Carries 10 marks.

(5X10=50)

11. a) Essay question from Unit-I
OR
b) Essay question from Unit-I
12. a) Essay question from Unit-II
OR
b) Essay question from Unit-II
13. a) Essay question from Unit-III
OR
b) Essay question from Unit-III
14. a) Essay question from Unit-IV
OR
b) Essay question from Unit-IV
15. a) Essay question from Unit-V
OR
b) Essay question from Unit-V

SRI VENKATESWARA UNIVERSITY::TIRUPATI

B.Sc. Data Science Honours

III year V Semester

Course 15 B : Data Wrangling With Java Script

(w.e.f. 2025-26)

Theory

Credits: 3

3 hrs/week

Course Objectives

By the end of this course, students will:

1. Understand the importance and role of **data wrangling** in modern data pipelines.
2. Gain familiarity with the **JavaScript ecosystem**, especially in the context of **data manipulation and processing**.
3. Learn to set up and use **Node.js** for building command-line tools, libraries, and simple web applications.
4. Develop the ability to **acquire, store, retrieve, and export data** using various file formats and databases (e.g., MongoDB, MySQL).
5. Practice **exploratory coding** to understand data patterns, trends, and insights using JavaScript tools like **Data-Forge**.
6. Understand the significance of **data cleaning and preparation**, and apply techniques to identify and fix common data issues.

Learning Outcomes

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

- Define and explain **data wrangling** and its relevance in data workflows.
- Navigate the **JavaScript environment** and use it effectively for data wrangling tasks.
- Build and run **Node.js applications**, including web servers and command-line utilities.
- Handle **asynchronous operations** in JavaScript using **callbacks, promises, and asynchronous functions**.
- Load, parse, and manipulate data from various sources, including **text files, REST APIs, and databases**.
- Export data to different formats such as **JSON, CSV**, and databases like **MongoDB and MySQL**.
- Perform **exploratory data analysis (EDA)** using JavaScript, both on the server and in the browser.
- Apply data cleaning techniques to remove errors and inconsistencies, and prepare datasets for meaningful analysis and usage.

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 5

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

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III year V Semester
Course 15 B : Data Wrangling With Java Script
(w.e.f. 2025-26)

Practical

Credits: 1

2 hrs/week

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 Formulajs.
11. Develop a program to clean your data by rewriting rows to fix bad data.

SRI VENKATESWARA UNIVERSITY: : TIRUPATI
B.Sc. Data Science Honours
III year V Semester
Course 15 B : Data Wrangling With Java Script
(w.e.f. 2025-26)
MODEL QUESTION PAPER

Time: **3 Hours**

Max. Marks: **75**

PART-A

Answer any **FIVE** of the following. Each Question Carries 5 marks.

(5X5=25)

1. Short answer question from unit-I
2. Short answer question from unit-I
3. Short answer question from unit-II
4. Short answer question from unit-II
5. Short answer question from unit-III
6. Short answer question from unit-III
7. Short answer question from unit-IV
8. Short answer question from unit-IV
9. Short answer question from unit-V
10. Short answer question from unit-V

PART-B

Answer any **FIVE** of the following. Each Question Carries 10 marks.

(5X10=50)

11. a) Essay question from Unit-I
OR
b) Essay question from Unit-I
12. a) Essay question from Unit-II
OR
b) Essay question from Unit-II
13. a) Essay question from Unit-III
OR
b) Essay question from Unit-III
14. a) Essay question from Unit-IV
OR
b) Essay question from Unit-IV
15. a) Essay question from Unit-V
OR
b) Essay question from Unit-V