s

Krishna University
కృష్ణా విశ్వవిద్యాలయం
(Established under AP Act No. 29 of 2008)
Machilipatnam, Andhra Pradesh, India - 521004

KRISHNA UNIVERSITY

MACHILIPATNAM

BOARD OF STUDIES AY: 2025-26

B.Sc (Data Science) & B.C.A (Data Science)

V Semester to VIII Semester and Skill Enhancement Courses

MEETING HELD ON 13-05-2025

(ONLINE)

Through Google Meet

S

KRISHNA UNIVERSITY

MACHILIPATNAM

No.KRU/AAC/UG BOS Meetings/to finalise V to VIII semester syllabus/2024-25 Dt:13-03-2025

PROCEEDINGS OF THE VICE – CHANCELLOR

Present: Prof. K. Ramji

Sub: - UG BOS meetings for finalising syllabus for V to VIII semesters in various UG Programs - orders – Issued.

Ref: - 1. The Vice-Chancellors Note Orders Dt.11.03.2025, Computer No. 2726469 2726469, File No.:SCHE-KRU/122/2025-KU-EHE73 SCHE-KRU/122/2025-KU-EHE73.

-oOo-

ORDER:

The Vice-Chancellor is pleased to issue orders for conducting BOS meetings with the members of the Board of studies for various UG programs such as BA, B.Sc, B.Com, BCA, BBA etc., so as to finalise the resolutions, syllabus, model question papers and any other related matters for all UG Programmes, from V to VIII semesters, along with the Skill Enhancement Courses (SECs) in the VII & VIII semesters.

The director, Academic Audit Cell, Krishna University is authorized to communicate with the said BOS members to hold the BOS meetings, in the **online mode** to finalise the resolutions, syllabus, model question papers and any other related matters. The chairpersons of the board of studies may facilitate the BOS meetings by communicating and coordinating with the other members of BOS. After the meeting is over, all the relevant documents with signatures are to be submitted to the office of the academic audit cell, in both hard and soft forms for further processing, along with the filled-in claim forms within **30 days** from the date of this order.

The Vice-Chancellor has also permitted to pay sitting allowance, through online/NEFT payment for all the members of the BOS attending online BOS meeting, from the **"Affiliation Account"** for the financial year 2024-25.

(BY ORDER)

S

SCHE-KRU/122/2025-KU-EHE73 I/3994854/2025

S/d

REGISTRAR (I/C)

Copy to:

1. PA to the Vice Chancellor
2. PA to the Registrar
3. To the Dean of Law
4. File

Digitally signed by
Nekkalapudi Usha
Date: 13-03-2025
16:10:09

S

KRISHNA UNIVERSITY

Dr. P. Veera Bramhachari
REGISTRAR (FAC)
Rudravaram
Machilipatnam-521004.
Mobile : 9542487999
E-mail : registrarku@gmail.com

No.KRU/AAC/Board of Studies Data Science (B.Sc. & B.C.A)/2023 Dated: 20-12-2023

Sub: KU – Academic Audit – The constitution of the UG Board of Studies (BoS) members in Data Science (B.Sc. & B.C.A) for the approval of the Under Graduation (UG) CBCS syllabus w.e.f 2023-24 (APSCHE-AC - Revision of syllabus under CBCS with effect from 2023-24 syllabus) – Reg.

Read: Note order's of the Vice-Chancellor, dt: 20-12-2023

~oOo~

ORDER:

In accordance with the provisions contained by the Krishna University act No:29 of 2008 of A.P. Govt., Hon'ble Vice-Chancellor is pleased to appoint the UG Board of Studies in Faculty of Data Science (B.Sc. & B.C.A) with the following members.

UG BOARD OF STUDIES FOR DATA SCIENCE (B.Sc. & B.C.A)		
S.No.	**Name of the Faculty**	**Designation**
1	Dr. N. Venkatadri Lecturer in Computer Science, Dr. LHR GDC Mylavaram Mobile: 9959108829, E-Mail:venkat.nagala@gmail.com	Chairperson
2	M Varun Rao Department of Data Science, Aditya Degree College, Vijayawada Mobile:9560106720, E-Mail:deanofdatascience@aditya.ac.in	Member
3	N Manikanta Department of Data Science, Aditya Degree College, Vijayawada Mobile:9293959344, E-Mail:manikantanukala1@gmail.com	Member
4	V. Lakshmi Sarvani Lecturer in Computer Science, SRR & CVR Government Degree College (A), Vijayawada Mobile: 91 99856 22936	Member
5	Dr. J. Sarada Lakshmi Lecturer in Computer Science, SRR & CVR Government Degree College (A), Vijayawada Mobile: 91 92466 60822	Member
6	Dr. R. Kiran Kumar Associate Professor, Krishna University Mobile: 9440872455, E-Mail: kirankreddi@gmail.com	University Nominee
7	One Student from Aditya Degree College	Student Member-1
8	One Student from Govt. Degree College, Mylavaram	Student Member-2

The term of office of all the members including the chairman of Board of studies is until further orders with effect from the date of the order or until the new boards of studies constituted.

These orders shall come in to force with immediate effect.

(BY ORDER)

REGISTRAR

REGISTRAR
KRISHNA UNIVERSITY
RUDRAVARAM (V)
MACHILIPATNAM - 521 0
ANDHRA PRADESH, IN

Copy to:
The Chairperson & all members of UG Board of Studies of the above
All Directors and Principals of the University and Constituent Colleges/Units
PA's to Vice-Chancellor/Registrar KRU
Accounts Section/ Establishment Section & File

s

Krishna University: Machilipatnam

Board of Studies Meeting

Minutes of the meeting of the UG Board of Studies in the Subject of

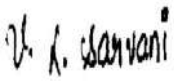
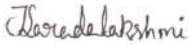
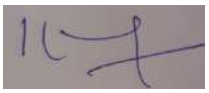
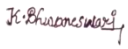
DATA SCIENCE (B.Sc & B.C.A)

The meeting of the UG Board of Studies in the Subject of **DATA SCIENCE** for both B.Sc and B.C.A under Single Major Programme / Minor convened at 11.30 A.M, on 13-05-2025 online (through Google Meet).

The following Members Attended the Meeting:

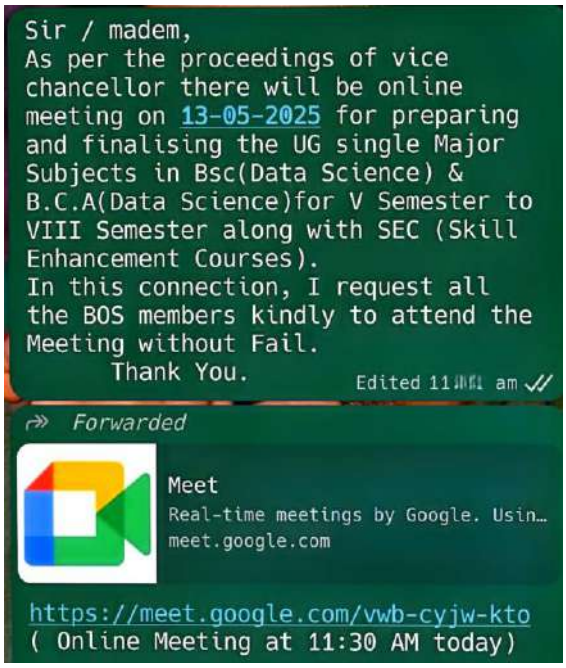
UG BOARD OF STUDIES FOR **DATA SCIENCE(B.Sc& B.C.A)**		
MEMBER	NAME & DESIGNATION	SIGNATURE
CHAIRMAN	Dr.N.Venkatadri, Lecturer in Computer Science, Dr. L.H.R Govt. Degree College, Mylavaram. Mobile: 9959108829,E-Mail:nagala.venkat@gmail.com	
MEMBER	M. Varun Rao, Dept. of Data Science Aditya Degree College, Vijayawada Mobile:9560106720, E-Mail:deanofdatascience@aditya.ac.in	

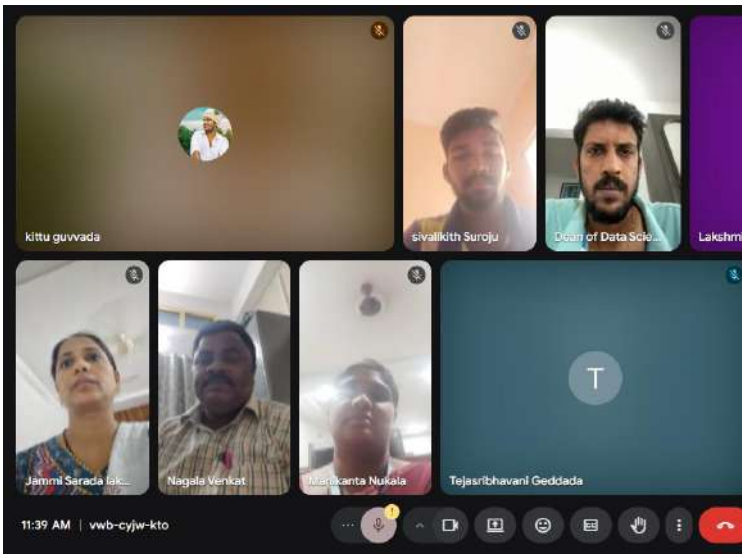
s

MEMBER	N.Manikanta, Dept. of Data Science Aditya Degree College, Vijayawada Mobile:9293959344, E-Mail:manikantanukala1@gmail.com	
MEMBER	V. Lakshmi Sarvani, Lecturer in Computer Science, SRR&CVR Govt. Degree College(A), Vijayawada Mobile:9985622936, E-Mail: sarvani.govt@gmail.com	
MEMBER	Dr. J.Sarada Lakshmi, Lecturer in Computer Science, SRR&CVR Govt. Degree College(A), Vijayawada Mobile:9246660822, E-Mail: sarada.govt@gmail.com	
UNIVERSITY-NOMINEE	Dr. R. Kiran Kumar Associate Professor, Krishna University Mobile: 9440872455, E-Mail: kiranreddi@gmail.com	
STUDENT MEMBER 1	K.SAI PRANEETH, II Year B.C.A (Data Science), Aditya Degree College, Vijayawada Mobile:948574005, E-Mail: praneethdeva72@gmail.com	
STUDENT MEMBER 2	K.BHUVANESWARI, III Year Data Science(B.Sc) Dr. L.H.R Govt. Degree College, Mylavaram. Mobile: 7382112089 E-Mail: bhuvanewarikesireddy@gmail.com	

S

MEETING PHOTOS





s

Agenda:

Item 1: Approval of syllabus for Semester V, VI, VII & VIII for the Academic Year 2025-2026 for **Data Science (B.Sc & B.C.A)**

Item 2: Approval of Blueprint and Model Question Papers

Item 3: Approval of Division of Marks for Internal Assessment.

Item 4: Any other item with the approval of the Chair.

Resolutions:

The members of the Board of Studies for **DATA SCIENCE (B.Sc & B.C.A)** of Krishna University held on 13.05.2025 online made the following resolutions unanimously.

1. It is resolved to adopt and implement the syllabi for the core subjects as appended hereunder for the V,VI,VII & VIII semesters only.

2. The Board of Studies unanimously resolved to have 30 Marks allocated for the Continuous Internal Assessment in the colleges and 70 Marks allocated for semester end examinations to be held by Krishna University.

3. The Board of Studies Unanimously resolved the model question paper with its scheme of distribution of marks as laid down hereunder.

4. The BoS also came out with a model question paper with a question bank of Shorts Answer Questions for 4 Marks & Essay Questions for 10 Marks as envisaged in the attached blueprint of the question paper attached herewith.

ANDHRA PRADESH STATE COUNCIL OF HIGHER EDUCATION

Programme: B.C.A. DATA SCIENCE (Major)
w.e.f. AY 2023-24

COURSE STRUCTURE

Year	Semester	Course	Title of the Course	No. of Hrs /Week	No. of Credits
I	I	1	Fundamentals of Commerce	3+2	4
		2	Business Organization	3+2	4
	II	3	Data Science & R Language	3	3
			Data Science & R Language Lab	2	1
		4	Data Analytics using Excel	3	3
			Data Analytics using Excel Lab	2	1
II	III	5	Statistical Foundation for Data Science	3	3
			Statistical Foundation for Data Science Lab	2	1
		6	Python Programming	3	3
			Python Programming Lab	2	1
		7	Data Mining Techniques using R	3	3
			Data Mining Techniques using R Lab	2	1
		8	Web Technology	3	3
			Web Technology Lab	2	1
	IV	9	Data Visualization using Tableau	3	3
			Data Visualization using Tableau Lab	2	1
		10	Data Visualization using Python	3	3
			Data Visualization using Python Lab	2	1
		11	Introduction to SQL & Advanced Tableau	3	3
			Introduction to SQL & Advanced Tableau Lab	2	1
III	V	12	Supervised Machine Learning with Python	3	3
			Supervised Machine Learning with Python Lab	2	1
		13	Unsupervised Machine Learning with Python	3	3
			Unsupervised Machine Learning with Python Lab	2	1
		14	Cloud Computing (OR) Machine Learning	3	3
			Cloud Computing (OR) Machine Learning Lab	2	1
		15	Software Testing (OR) Advanced Data Analysis	3	3
			Software Testing (OR) Advanced Data Analysis Lab	2	1

	VI	Semester Internship/Apprenticeship with 12 Credits			
Year	**Semester**	**Course**	**Title of the Course**	**No. of Hrs /Week**	**No. of Credits**
IV	VII	16	Big Data Analytics using Spark & Hadoop	3	3
			Big Data Analytics using Spark & Hadoop Lab	2	1
		17	Big Data Security	3	3
			Big Data Security Practical Course	2	1
		18	Introduction to Deep Learning	3	3
			Introduction to Deep Learning Lab	2	1
		SEC			
		19	AI Concepts and Techniques with Python	3	3
			AI Concepts and Techniques with Python Practical Course	2	1
		20	Data and Information Security	3	3
			Data and Information Security Practical Course	2	1
		Syllabus will be available in due course of time			
	VIII	21	Introduction to Neural Networks	3	3
			Introduction to Neural Networks Practical Course	2	1
		22	Natural Language Processing	3	3
			Natural Language Processing Practical Course	2	1
		23	Research Exploration	3	3
			Research Exploration Practical Course	2	1
		SEC			
		24	Computational Data Science	3	3
			Computational Data Science Practical Course	2	1
		25	Computer Vision with Python	3	3
			Computer Vision with Python Practical Course	2	1

SEMESTER-V

COURSE 12: SUPERVISED MACHINE LEARNING WITH PYTHON

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: 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 the 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 Supervised ML with Python Lab

SEMESTER-V

COURSE 12:SUPERVISED MACHINE LEARNING WITH PYTHON LAB

Practical Credits: 1 2 hrs/week

Details of Lab/Practical/Experiments/Tutorials syllabus:

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 naive 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 naive 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 dataset.

SEMESTER-V

COURSE 13: U N S U P E R V I S E D M A C H I N E L E A R N I N G W I T H P Y T H O N

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:

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 FPtree, 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

SEMESTER-V

COURSE 13:U N S U P E R V I S E D M A C H I N E L E A R N I N G W I T H P Y T H O N L A B

Practical Credits: 1 2 hrs/week

Details of Lab/Practical/Experiments/Tutorials syllabus:

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.

SEMESTER-V

COURSE 14: CLOUD COMPUTING

Theory Credits: 3 3 hrs/week

Course Objectives:

1. This course introduces the basic principles of cloud computing, Cloud Computing models and Services, Virtualization Techniques, Resource Sharing, Load Balancing and Security issues in Cloud Computing.
2. It will also equip the students to understand major industry players in the public cloud domain.

Course Outcomes: Upon Completion of the course, the students will be able to

1. Understand the basic principles of cloud computing.
2. Analyze different types of cloud services – Cloud Computing Models and Cloud Computing Services.
3. Understand Virtualization techniques
4. Use different Cloud Platforms
5. Learn different File systems and Security issues in Cloud Computing Platforms

Unit-I: Introduction to Cloud Computing, Evaluation and Enabling technologies, Benefits and Challenges, Cloud Computing Model

CASE STUDY: Identify different types of Public Clouds Private Cloud and Hybrid Cloud

Unit-II: Cloud Computing Services, Elements of Cloud Security model, Cloud Security reference model, Examining Cloud Security against traditional computing, Security Policy

CASE STUDY: Identify various Realtime Examples of Cloud Computing Services and work with them

Unit-III: Virtualization basics, Machine or Server Level Virtualization, Major Server virtualization products and vendors, Types of Virtualization, Advantages and Disadvantages of Virtualization, Virtualization Security Threats

CASE STUDY: Install any Virtualization Software and Perform any task

Unit-IV: Resource Pooling and Sharing, Multi-tenancy, Resource Provisioning, Scaling, Scaling strategies in Cloud, Types of Scaling, Load balancing in Cloud, Categories of Load Balancing, Service Oriented Architecture (SOA), Elements of a service, Benefits of SOA

CASE STUDY: Perform different operations using Google Cloud and Amazon Elastic Compute Cloud(EC2)

Unit-V: Cloud Native File System, Storage types, Relational DBMS in Cloud, Non-RelationalDBMS in Cloud, Cloud Security, Treats to cloud security, Infrastructure Security in Cloud, Information Security in Cloud, Identity Management and Access Control

Case Study: Identify and work with Popular cloud storage for developers, Popular General Purpose Cloud Storage

Prescribed Text Books:

1. Cloud Computing: Sandeep Bhowmik , Cambridge University Press
2. CLOUD COMPUTING:Principles and Paradigms by Rajkumar Buyya, James Broberg, Andrzej Goscinski
3. The Definite Guide to Cloud Computing by Dan Sullivan

SEMESTER-V

COURSE 14: CLOUD COMPUTING LAB

Practical Credits: 1 2 hrs/week

Course Outcomes: Upon Completion of the course, the students will be able to

1. Working with Virtual machines.
2. Run simple programs on virtual box.
3. Acquire knowledge in Google App Engine
4. Get practical exposure on CloudSim

Experiments:

1. Install Virtualbox / VMware Workstation with different flavours of linux or windows OS on top of windows10 or 11.
2. Install a C compiler in the virtual machine created using virtual box and execute Simple Programs
3. Install Google App Engine. Create hello world app and other simple web applications using python/java.
4. Use GAE launcher to launch the web applications.
5. Simulate a cloud scenario using CloudSim and run a scheduling algorithm that is not presentin CloudSim.
6. Find a procedure to transfer the files from one virtual machine to another virtual machine.
7. Find a procedure to launch virtual machine using trystack (Online Openstack Demo Version)

Web Resources:

1. Cloud Computing Fundamentals Labs - Cloud Academy

SEMESTER-V

COURSE 14: MACHINE LEARNING

Theory Credits: 3 3 hrs/week

Course Objectives:

1. To learn the basics of machine learning
2. To learn the importance of feature engineering
3. To understand the foundational rules of probability
4. To Study different classification and clustering techniques
5. To understand the basics of neural networks concepts

Course Outcomes: Upon Completion of the course, the students will be able to

1. Develop an appreciation for what is involved in Learning models from data
2. Understand a wide variety of learning algorithms
3. Understand how to evaluate models generated from data
4. Apply the algorithms to a real problem, optimize the models learned and report on the expected accuracy that can be achieved by applying the models
5. Understand different neural network concepts

Unit-I:

Introduction to Machine Learning: Introduction, what is human learning, types of human learning, what is machine learning, types of machine learning, problems not to be solved using machine learning, Applications of machine learning, Tools in machine learning, Issues in machine learning

Preparing to Model: Machine learning activities, Basic types of data in machine learning, Exploring structure of data, Data quality and remediation, Data Preprocessing

Case Study: Machine Learning **Case Study** on Trendyol

Unit-II:

Modelling and Evaluation: Selecting a model, Training a model, Model representation and Interpretability, Evaluating Performance of a model, Improving performance of a model

Basics of Feature Engineering: Introduction, Feature transformation, Feature subset selection

Case Study: Machine Learning **Case Study** On Harley Davidson

Unit-III:

Brief Overview of Probability: Importance of Statistical Tools in Machine Learning, Concept of Probability, Random Variables, Some Common Discrete Distributions, Some Common Continuous Distributions, Multiple Random Variables, Sampling Distributions, Hypothesis Testing

Bayesian Concept Learning: Why Bayesian methods are important, Bayes Theorem, Concept Learning, Bayesian Belief Network

Case Study: Machine Learning **Case Study** on Tesla

Unit-IV:

Supervised Learning : Classification: Example of Supervised Learning, Classification Model, Classification Learning Steps, Common Classification Algorithms

Super vised Learning : Regression: Example of Regression, Common Regression Algorithms

Unsupervised Learning: Unsupervised vs Supervised Learning, Application of Unsupervised Learning, Clustering, Finding Pattern using Association Rule

Case Study: Predicting Heart Failure in Mobile Health

Unit-V:

Basics of Neural Network: Understanding Biological Neuron, Exploring the Artificial Neuron, Types of Activation Functions, Early Implementations of ANN, Architectures of Neural Network, Learning Process in ANN, Back Propagation, Deep Learning

Other Types of Learning: Representation Learning, Active Learning Vs Memory based Learning, Ensemble Learning Algorithm

Case Study: American Cancer Society on Google Cloud ML Engine

Prescribed Text Books:

1. Machine Learning, Pearson by Saikat Dutt, Subramanian Chandramouli, Amit Kumar Das
2. Introduction to Machine Learning with Python:A Guide for Data Scientists by Andreas C.Muller & Sarah Guido
3. Machine Learning for Absolute Beginners, 2nd Edition by Oliver Theobald
4. Machine Learning for Dummies: IBM Limited Edition by Judith Hurwitz and Daniel Kirsch

SEMESTER-V

COURSE 14: MACHINE LEARNING LAB

Practical Credits: 1 2 hrs/week

Course Outcomes: On successful completion of this practical course, student shall be able to:

4. Execute basic programs in either R or Python
5. Gain practical knowledge on different python libraries/packages .
6. Implement different machine learning algorithms

Implement the following using in either R Language or Python

1. **Introduction to basic commands:**
    a) Get and Set Working Directory
    b) See Directory Content
    c) Install and Load Packages
    d) Compile Source File for Execution
    e) Commands for basic user input/output
    f) Basic Data Types and Data Manipulation Functions
2. **Introduction to basic commands Continued:**
    a) Conditional Statements
    b) Loops
3. Data Manipulation Package installation and different operations using installed package
4. Standard Library function to plot the Graphs
5. Basic Data Exploration on any dataset available publicly
6. Starting to Model to find accuracy of the model
7. Learning Algorithms-kNN Linear Regression
8. Unsupervised Algorithm- k-means
9. Supervised Algorithm- Naïve Bayes
10. Implement Decision Tree and Support Vector Machine using Library Functions
11. Implement Neural Network using Library Function

SEMESTER-V

COURSE 15: SOFTWARE TESTING

Theory Credits: 3 3 hrs/week

Course Objectives: Student will gain an understanding of Selenium basics. Student will practice writing tests for a variety of quality intent, including code coverage, defect finding, and statistical testing using Selenium.

Course Outcomes: Upon Completion of the course, the students will be able to

1. Understand the Selenium Basics concepts
2. Know about Selenium Web Driver
3. Understand Selenium Framework
4. Learn how TestNG works
5. Learn Advanced concepts in Selenium

Unit-I:

Selenium Basics:

Software Testing, Automation Testing, Introduction to Selenium and its Components, Selenium IDE Features, Selenium Download and Installation, Creating Scripts using Firebug and Its Installation, Locator Types

Case Study: Automation Testing using Cucumber Tool with Selenium

Unit-II:

Selenium WebDriver:

Selenium WebDriver Installation with Eclipse, Handling Dropdowns, Explicit and Implicit Wait, Handling Alerts/Pop-ups, Handling Web Tables, Frames, Dynamic Elements, Robot API, AutoIT

Case Study: Automation Software Testing using Katalon Studio with Winium

Unit-III:

Selenium Framework:

Test Automation Framework: Introduction, Benefits of Automation Framework, Types of Automation framework

Case Study: Cucumber Based Automation Testing to Validate End-user Experience

Unit-IV:

Introduction to TestNG,TestNG Framework, TestNG installation, TestNG Annotations and Listeners, TestNG Example, TestNG Process Execution: Batch, Controlled Batch & Parallel

Case Study: QA Automation Testing for Media & Entertainment

Unit-V:

Advance Selenium:
Selenium Grid: Introduction, Usage of Selenium Grid, Grid1.0 vs Grid2.0, Selenium Grid architecture, How to setup Selenium Grid using command line, designing test scripts that can run on the Grid, Using DesiredCapabilities Object, Using RemoteWebDriver Object, Running a sample Test Case on the Grid
Case Study: Test Automation Using Selenium Java

Prescribed Text Books:

1. Test Automation using Selenium WebDriver with Java: Step by Step Guide by Navneesh Garg
2. Selenium WebDriver_ From Foundations To Framework by Yuzun Liang & Alex Collins
3. Absolute Beginner Java 4 Selenium Webdriver: Come Learn How to Program for Automation Testing by Rex Allen Jones II

Reference Materials on the Web/web-links:

https://www.softwaretestingmaterial.com/types-test-automation-frameworks/
https://www.guru99.com/introduction-to-selenium-grid.html#6

Faculty & Student Resources:

- YouTube: Selenium Video Tutorials
- YouTube: Selenium Full Course – Simplilearn
- YouTube: Selenium Full Course – Edureka!
- Sample Selenium Assignments/Exercises 1
- Sample Selenium Assignments/Exercises 2
- Sample Selenium Assignments/Exercises 3

SEMESTER-V

COURSE 15: SOFTWARE TESTING LAB

Practical Credits: 1 2 hrs/week

Course Outcomes: On successful completion of this practical course, student shall be able to:

7. Perform automation testing using selenium.
8. Get exposure on Selenium framework.
9. Gain practical knowledge on Selenium web drivers .

Practical (Laboratory) Syllabus: (30 Periods)

1. Write a script to open google.com and verify that title is Google and verify that it is redirected to google.co.in
2. Write a script to open google.co.in using chrome browser (ChromeDriver)
3. Write a script to open google.co.in using internet explorer (InternetExplorerDriver)
4. Write a script to create browser instance based on browser name
5. Write a script to search for specified option in the listbox
6. Write a script to print the content of list in sorted order.
7. Write a script to print all the options. For duplicates add entry only once. Use HashSet.
8. Write a script to close all the browsers without using quit() method.
9. Write generic method in selenium to handle all locators and return web element for any locator.
10. Write generic method in selenium to handle all locators containing dynamic wait and returnweb element for any locator.

III. Lab References:

https://www.guru99.com/selenium-tutorial.html
https://www.javatpoint.com/selenium-tutorial
Web-links: **https://youtu.be/Tu61E5mHv18**

SEMESTER-V

COURSE 15: A D V A N C E D D A T A A N A L Y S I S

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”, IGIGlobal.

SEMESTER-V

COURSE 15: A D V A N C E D D A T A A N A L Y S I S L A B

Practical Credits: 1 2 hrs/week

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:

Uni variate Analysis: Frequency, Mean, Median, Mode, Variance, Standard Deviation, Skewness and Kurtosis.

i. Bi variate 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 datasets.

SEMESTER-VII

COURSE 16A: B I G D A T A A N A L Y T I C S U S I N G S P A R K & H A D O O P

Theory Credits: 3 3 hrs/week

Aim and objectives of Course:

- To Understand the Complete Architecture of Spark
- To know the differences between Hadoop and Spark
- To know the concepts of Spark Programming

Learning outcomes of Course:

- Students will get well knowledge of what is
- Big Data Knowledge in Spark Eco System
- Mapping of Data Analytics techniques in Spark
- Application of Spark Programming to Analytics problems

UNIT I:

Introduction to Big Data: What is Big Data-Characteristics, Data in the Warehouse and Data in Hadoop, Why is Big Data Important- When to consider Big Data Solution, Applications. Introduction to Hadoop: Hadoop- definition, Application development in Hadoop. The building blocks of Hadoop, Name Node, Data Node, Secondary Name Node, Job Tracker and Task Tracker.

UNIT II:

Introduction to Spark: What is Apache Spark, Why Spark when Hadoop is there, Spark Features, , Spark components, Spark program flow, Spark Eco System. Differences between implementation of programs in Hadoop and Spark Programming environments.

UNIT III:

Spark Fundamentals- Using spark in action VM, Using Spark Shell and writing first spark program, Basic RDD actions and transformations. Spark SQL-Working with Data Frames, Using SQL Commands, Saving and loading Data Frame.

UNIT IV:

Streaming in Spark- Writing spark streaming applications, Using external data sources, structured streaming. Spark MLlib-Introduction to Machine Learning. Definition of Machine Learning, Machine Learning with Spark.

UNIT V:

Graph Representation in MapReduce: Graph Processing with Spark, Spark GraphX, GraphX features, Graph Examples, Graph algorithms-Shortest Path Algorithm.

TEXT BOOKS:

1. Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data byDirk deRoos, Chris Eaton, George Lapis, Paul Zikopoulos, Tom Deutsch, 1st Edition, TMH,2012.
2. Spark in Action PetarZecevic, markoBonaci Manning Publications-2016.
3. Learning Spark“Holden KarauA. Konwinskietc.,”O’reilly Publications

REFERENCE BOOKS:

1. Hadoop in Action by Chuck Lam, MANNING Publishers.
2. Hadoop: The Definitive Guide by Tom White, 3rd Edition, O’reilly
3. Mining of massive datasets, AnandRajaraman, Jeffrey D Ullman, Wiley Publications.

Student Activities:
Take any dataset and do the following machine learning steps.(https://www.guru99.com/pyspark-tutorial.html)
1. Use basic Operations with PySpark(Spark with Python)
2. Data Pre-processing
3. Build a data processing pipeline
4. Build the classifier
5. Train and evaluate the model
6. Tune the hyper parameter

Continuous assessment:
1. Let the students be tested in the following questions from each unit
2. What is Big Data? Explain the characteristics of it
3. What is Spark? What are the advantages of it over Hadoop
4. Explain Spark SQL
5. Explain Spark Streaming
6. Explain Shortest Path Algorithm.

SEMESTER-VII

COURSE 16A: B I G D A T A A N A L Y T I C S U S I N G S P A R K & H A D O O P L A B

Practicals Credits: 1 2 hrs/week

1. Using Python Implement the following Programs
a) Write Program to implement arithmetic operations
b) Write Program to find the biggest of two numbers
c) Write a program to find the matrix multiplication

2. Install Hadoop

3. Install Spark on top of Hadoop

4. Create and Implement the transformations in RDDs

5. Create a data frame from an existing RDD using Spark Session

6. Execute a Word Count example in Spark Shell by creating RDDs.

7. Implement Spark SQL Queries in Python.

8. Write a Program to implement maximum temperature give the recordings of one year.

9. Write a Program to implement the Pie estimation

10. Write a User Defined Function to convert a given text to Uppercase

SEMESTER-VII

COURSE 16-B: B I G D A T A S E C U R I T Y

Theory Credits: 3 3 hrs/week

Course Objective

With the data generated from electronic devices growing exponentially, the need to analysed data on a large scale is important. Such data are of many types like financial,personal etc. Big data environment also created significant security challenges. When trying to make quick decisions. Data breach poses many complications. This course aims at introducing concepts related to big data security.

Learning Outcomes :

1. Understanding significance of privacy, ethics in big data environment.
2. Analyzing the steps to secure big data.
3. Analyzing data security and event logging.

Unit-I

BIG DATA PRIVACY, ETHICS AND SECURITY: Privacy- Re identification of Anonymous people – Why Big Data Privacy is self regulating? – Ethics – Ownership –Ethical Guidelines - Big Data Security – Organizational Security.

Unit-II

SECUTIY, COMPLIANCE, AUDITING, AND PROTECTION: Steps to secure bigdata – Classifying Data – Protecting – Big Data Compliance – Intellectual Configuration..

Unit-III

HADOOP SECURITY DESIGN: Kerberos – Default Hadoop Model Without security Hadoop Kerberos Security Implementation & Configuration.

Unit-IV

HADOOP ECOSYSTEM SECURITY: Configuring Kerberos for Hadoop ecosystem components – Pig. Hive. Oozie, Flume, HBase, Scoop.

Unit-V

HADOOP ECOSYSTEM SECURITY: Integrating Hadoop with Enterprise SecuritySystemsSecuring Sensitive Data in Hadoop – SIEM System – Setting up audit logging in hadoop cluster.

SEMESTER-VII

COURSE 16-B: B I G D A T A S E C U R I T Y L A B

Practicals Credits: 1 2 hrs/week

1. Learn the basics of Mongo DB.
2. Installation steps for Mongo DB.
3. Use the following commands
(a) DATABASE_NAME.
(b) Drop Database( )
(c) create Collection
(d) insert( )
(e) drop( )
(f) find( )
4. Differentiate between SQL and Mongo DB.
5. Write a program to update a collection in Mongo DB
6. Write a program to remove specific document from Mongo DB.
7. Write a program to implement aggregate function in Mongo DB
8. Apply the Map reduce operation to find total salary of each department assuming employee collection is already exists.

SEMESTER-VII

COURSE 17A: I N T R O D U C T I O N T O D E E P L E A R N I N G

Theory Credits: 3 3 hrs/week

Course Objectives:

1. To introduce basics of linear algebra and probability theory
2. To introduce the fundamental techniques and principles of Neural Networks
3. To familiarize different models in Artificial Neural Networks (ANN) and their applications
4. To familiarize deep learning concepts with Convolutional Neural Network case studies
5. To explain functioning of deep neural networks

Course Outcomes:

After learning the course, the students will be able to:

1. Discuss feed forward networks and their training issues
2. Distinguish different types of ANN architectures
3. Design Feed Forward Neural Network architecture for research problems
4. Apply mathematical concepts such as linear algebra, calculus to solve the research problems.
5. Apply deep learning techniques to practical problems
6. Evaluate model performance and interpret results

Unit 1

Linear Algebra and Probability Theory: Linear Algebra :Scalars, Vectors, Matrices and Tensors, Multiplying Matrices and Vectors, Identity and Inverse Matrices Calculus: Derivatives and Differentiation, Partial Derivatives, Gradients Probability Theory : Basic Probability Theory, Dealing with Multiple Random Variables, Expectation and Variance.

Unit 2

Fundamentals of Neural Networks: Introduction to Neural Network, Model of Artificial Neuron, Learning rules and various activation functions.

Unit 3

Neural Network Architecture: Single layer Feed-forward networks. Multi-layer Feed-forward networks. Recurrent Networks.

Unit 4

Back propagation Networks: Back Propagation networks, Architecture of Back-propagation (BP) Networks, Backpropagation Learning, Variation of Standard Back propagation algorithms.

Unit 5

Deep Neural Networks: Introduction to Deep Neural Networks, training deep models, Training Deep Neural Networks using Back Propagation-Setup and initialization issues, Gradient Descent Strategies, vanishing and exploding Gradient problems, regularization, dropouts.

Text Books:

1. S.Rajasekaran and G.A. Vijayalakshmi Pai, "Neural Networks,Fuzzy Logic and GeneticAlgorithms", PHI Learning Pvt. Ltd., 2003, ISBN:978-81-203-2186-1.
2. Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola, "Dive into Deep Learning", Amazon Science, 2021.

Reference Books:

1. Jacek M. Zurada,"Introduction to artificial neural systems", West Publishing Co., 1992, ISBN: 0-3 14-93391 -

2. Goodfellow I., BengioY., and Courville A., "Deep Learning", MIT Press, 2016, ISBN: 978-0262035613.
3. Bishop C. M., "Pattern Recognition and Machine Learning", Springer, 2006, ISBN: 978-0-387-31073-2.

SEMESTER-VII

COURSE 17A: I N T R O D U C T I O N T O D E E P L E A R N I N G L A B

Practical Credits: 1 2 hrs/week

1. Design and implement a CNN for Image Classification:
a) Select a suitable image classification dataset (medical imaging, agricultural,etc.).
b) Optimized with different hyper-parameters including learning rate, filter size,no. of layers, optimizers, dropouts, etc.

2. Apply a pre-trained network and apply it to a new task using transfer learning
a) Use any three pre-trained models including AlexNet, GoogleNet, VGGNet,MobileNet, ResNet, DenseNet, etc.
b) Fine-tune the hyper-parameters and compare their performance for a suitable application.

3. Design RNN or its variant including LSTM or GRU
a) Select a suitable time series dataset. Example – predict sentiments based on product reviews
b) Apply for prediction

4. build a word2vec model for unstructured data a) Use any unstructured text dataset
b) Convert words into a representative vector of numerical values5.Implement an artificial neural network on GPUs
a) Implement ANN on GPUs.

SEMESTER-VII

COURSE 17B: D E E P L E A R N I N G W I T H P Y T O R C H

Theory Credits: 3 3 hrs/week

Aim and objectives of Course (Predictive and Advanced Analytics):
The course enables students to:

- □ To learn, PyTorch and relates PyTorch to deep learning.
- □ To know about pretrained models to work with data sets
- □ To learn how to work with neural networks in deep learning.

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

- □ Learn about deep learning , PyTorch library and their applications.
- □ How to use Tensors in deep learning Tensor-API
- □ To work with time series in deep learning.

□ To develop a PyTorch Neural Network model using "torch.nn" module.

UNIT I

Introducing deep learning and the PyTorch Library - The Deep Learning revolution -PyTorch for Deep Learning - Why Pytorch - An overview of how PyTorch supports deep learning projects - Hardware and Software requirements - Using Jupyter Notebooks.

UNIT II

Pretrained Networks - A pre-trained that recognizes the subject of an image - A pertained network that describes scenes - Starting with Tensor- Tensors.

UNIT III

Multidimensional arrays - Indexing tensors - Named Tensors - Tensor element types - The tensor API - Tensors: Scenic views of storage.

UNIT IV

Real work data representation using tensors - Working with images - Representing tabular data - Working with time series - Representing text - The mechanics of Learning - Learning is just parameter estimation - Less loss is what we want .

UNIT V

Using a nearly network to fit the data - Artificial neutrons - The PyTorch nn module - Finally a neural network.

TEXTBOOK:

Deep Learning with PyTorch - Eli Stevens, Luca Antiga, Thomas Viehmann - Manning Publications

REFERENCES:

1. Programming PyTorch for Deep Learning: Creating and Deploying Deep LearningApplications by Ian Pointer - Oreilly Publications.

SEMESTER-VII

COURSE 17B: D E E P L E A R N I N G W I T H P Y T O R C H

Practical Credits: 1 2 hrs/week

1. Build a Jupiter Notebook Server for Pytorch.
2. Program On CycleGAN network: Use CycleGAN network model to turn one image into another image. Feed the image of the golden retriever into the horse-to-zebra model. (Refer Prescribed Textbook)
3. Develop a program to convert integers into floating point numbers in pyTorch using tensors.
4. Develop a program on multi-dimensional arrays using TensorFlow.
5. Develop a PyTorch program to perform different Tensor Indexing operations.
6. Develop a PyTorch program on “named tensors”.
7. Develop a program in TensorFlow for representing text. Define a function clean_words, which takes text as input and returns it in lowercase and print individual words in given text.
8. Perform time series operations on bikes data. Use following url for reference (code/p1ch4/4_time_series_bikes.ipynb)
9. Working with images: Take several pictures of red, blue, and green items with your phone or oth-er digital camera . Load each image, and convert it to a tensor. For each imagetensor, use the .mean() method to get a sense of how bright the image is.
10. Develop a PyTorch Neural Network model using “torch.nn” module.

SEMESTER-VII

COURSE 18A: A I C O N C E P T S A N D T E C H N I Q U E S W I T H P Y T H O N

Theory Credits: 3 3 hrs/week

Aim and objectives of Course :

 This course provides an introduction to the fundamentals of artificial intelligence.
 Demonstrates fundamental understanding of the history of artificial intelligence (AI)
 and its foundations.
 Apply basic principles of AI in solutions that require problem solving, inference,
 perception, knowledge representation, and learning.
 Demonstrates awareness and a fundamental understanding of various applications of AI
 techniques in intelligent Agents.

Learning outcomes of Course:

 List the objectives and functions of modern Artificial Intelligence.
 Categorize an AI problem based on its characteristics and its constraints.
 Understand and implement search algorithms.
 Learn how to analyze the complexity of a given problem and come with suitable
 optimizations.
 Demonstrate practical experience by implementing and experimenting with the learnt
 Algorithms

UNIT I:

Problems and Search: What is Artificial Intelligence, The AI Problems, and Underlying Assumption, what is an AI Technique. Problems, Problems Spaces, and Search: Defining the problem as a state space search,production systems, problems characteristics, issues in the design of search programs.

UNIT II:

Heuristic Search Techniques: Generate-and-test, Hill Climbing, Best-First Search, Problem Reduction, Constraint Satisfaction, Means-Ends Analysis

UNIT III:

Knowledge Representation Issues: Representations and Mapping, Approaches to Knowledge Representation, The frame problem. Using Predicate Logic: Representing simple facts in logic, Representing Isa relationships, predicates, Resolution

UNIT IV:

Representing Knowledge using Rules: Procedural Vs Declarative knowledge, Logic Programming, Forward Vs Backward Reasoning, Matching, Control Knowledge

UNIT V:

Symbolic Reasoning under Uncertainty: Introduction to Non-monotonic Reasoning, Logics for Non-monotonic Reasoning, Implementation issues, Augmenting a Problem solver, implementation: DFS, BFS. Statistical Reasoning: Probability and Bayes Theorem, Certainty Factors and Rule-Based Systems, Bayesian Networks, Dempster-Shafer Theory.

TEXT BOOK:

1. Artificial Intelligence, Second Edition, Elaine Rich, Kevin Knight, Tata McGraw-Hill

Edition.

REFERENCES BOOK:

1. Russell, S., &Norvig, P. Artificial intelligence: a modern approach. Third Edition.Pearson new International edition. 2014

SEMESTER-VII

COURSE 18A: A I C O N C E P T S A N D T E C H N I Q U E S W I T H P Y T H O N L A B

Practicals Credits: 1 2 hrs/week

1. Write a Program to Implement Breadth First Search using Python.

2. Write a Program to Implement Depth First Search using Python.

3. Write a Program to Implement Tic-Tac-Toe game using Python.

4. Write a Program to implement 8-Puzzle problem using Python.

5. Write a Program to Implement Water-Jug problem using Python.

6. Write a Program to Implement Travelling Salesman problem using Python.

7. Write a Program to Implement Towers of Hanoi problem using Python.

8. Write a Program to implement 8-Queens problem using Python.

SEMESTER-VII

COURSE 18B: D A T A & I N F O R M A T I O N S E C U R I T Y

Theory Credits: 3 3 hrs/week

Unit-I

Overview of Security: Protection versus security; aspects of security – data integrity, data availability, privacy; security problems, user authentication, Orange Book.

Unit -II

Security Threats: Program threats, worms, viruses, Trojan horse, trap door, stack and buffer overflow; system threats- intrudes; communication threats- tapping and piracy.

Unit -III

Cryptography: Substitution, transposition ciphers, symmetric-key algorithms – Data Encryption Standard, advanced encryption standards, public key encryption – RSA; DiffieHellman key exchange, ECC cryptography, Message Authentication – MAC, has functions.

Unit -IV

Digital Signatures: Symmetric key signatures, public key signatures, message digests,public key infrastructures.

Unit -V

Security Mechanism: Intrusion detection, auditing and logging, tripwire, system –call monitoring.

Text Books:

1. W. Stallings, Cryptography and Network Security Principles and Practices (4th ed.,), Prentice – Hall of India, 2006.
2. C. Pfleeger and SL Pfleeger, Security in Computing (3rd ed.,), Prentice- Hall of India, 2007.

Reference Books:

1. D. Gollamann, Computer Security, John Wiley and Sons, Ny, 2002.
2. J. Piwprzyk, T. Hardjono and J. Seberry, Fundamentals of Computer Security,
3. Springer- Verlag Berling, 2003.
4. J.M. Kizza, Computer Netwrok Security, Springer, 2007
5. M. Merkow and J. Breithaupt, Information Security: Principles and Practices, Pearson
6. Education, 2006.

Student Activity

Case Study I: Transform Data from one format to another format using Cryptography.Case Study II: How mails are hacked.

SEMESTER-VII

COURSE 18B: D A T A & I N F O R M A T I O N S E C U R I T Y L A B

Practical Credits: 1 1 hrs/week

1. Implement Caesar Cipher encryption in Python.

2. Implement Caesar 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

SEMESTER-VII

COURSE 19: I N T R O D U C T I O N T O N E U R A L N E T W O R K S

Theory Credits: 3 3 hrs/week

Learning Objectives:

1. The main objective of Neural Network Techniques to Improve Data Analysis Solutions is to strengthen the dialogue between the statistics and soft computing research communities in order to cross-pollinate both fields and generate mutual improvement activities.
2. Also introduce the neural networks for classification, regression and to give design methodologies for artificial neural networks.

Learning Outcome:

On successful completion of this course, student will be able to

1. Obtain the fundamentals and types of neural networks. The student will have a broad knowledge in developing the different algorithms for neural networks.
2. Analyze neural controllers
3. Have a broad knowledge in Fuzzy logic principles and will be able to determine different methods of Deffuzification

UNIT 1

Introduction to Artificial Intelligence System: Neural Network, Fuzzy logic, Genetic Algorithm. Fundamentals of Neural Networks: What is Neural Network, Model of Artificial Neuron, Learning rules and various activation functions.

Unit 2

Neural Network Architecture: Single layer Feed-forward networks. Multilayer Feed-forward networks. Recurrent Networks. Back propagation Networks: Back Propagation networks, Architecture of Back-propagation(BP) Networks, Back-propagation Learning, Variation of Standard Back propagation algorithms.

Unit 3

Introduction about Fuzzy set theory: Fuzzy versus Crisp, Crisp and fuzzy sets, Crisp and Fuzzy relations. Fuzzy Systems: Crisp Logic, Predicate Logic, Fuzzy logic, Fuzzy rule based system, Defuzzification Methods, Applications.

Unit 4

Integration of Neural Network, Fuzzy logic and Genetic Algorithm: Hybrid system. Neural Networks, Fuzzy logic, and Genetic Algorithm Hybrids.

Unit 5

Associative Memory: Autocorrelators, Heterocorrelators, Wang et al's Multiple Training Encoding Strategy, Exponential BAM, Associative Memory for Real coded pattern pairs, Applications.

Text book:

1. Neural Networks a Comprehensive Foundations, Simon S Haykin, PHI Ed.,.
2. Introduction to Artificial Neural Systems Jacek M. Zurada, JAICO Publishing HouseEd. 2006.
3. Neural Networks in Computer Inteligance, Li Min Fu TMH 2003 Neuro-FuzzySystems, Chin Teng Lin, C. S. George Lee, PHI.
4. Build_Neural_Network_With_MS_Excel_sample by Joe choong.

SEMESTER-VII

COURSE 19: I N T R O D U C T I O N T O N E U R A L N E T W O R K S L A B

Practicals Credits: 1 2 hrs/week

1. Create a perception with appropriate no. of inputs and outputs. Train it using fixed increment learning algorithm until no change in weights is required. Output the finalweights.
2 Create a simple ADALINE network with appropriate no. of input and output nodes. Train it using delta learning rule until no change in weights is required. Output the final weights.
3 Train the autocorrelator by given patterns: A1=(-1,1,-1,1), A2=(1,1,1,-1), A3=(-1, -1, -1,1). Test it using patterns: Ax=(-1,1,-1,1), Ay=(1,1,1,1), Az=(-1,-1,-1,-1).
4. To Implement Convolution Neural network for Text classification or Image Classification
5. Implementation of Naïve Bayes/SVM/SGD/SVM classifier on text and image
6. To study Word Embedding techniques : Word2vec,doc2vec,Glove
7. Implement Linear/Logistic regression

SEMESTER-VII

COURSE 20: N A T U R A L L A N G U A G E P R O C E S S I N G

Theory Credits: 3 3 hrs/week

COURSE OBJECTIVE:

The basic objectives of natural language processing course are the following:

1. Learn the basics of natural language processing and understand various steps in it.
2. To introduce the fundamentals of language processing from the algorithmic viewpoint.
3. To discuss various issues that make natural language processing a hard task.
4. To discuss some well-known applications of natural language processing

LEARNING OUTCOME:

At the end of the course, the student should be able to:

1. Appreciate the fundamental concepts of natural language processing.
2. Design algorithms for natural language processing tasks.
3. Develop useful systems for language processing and related tasks involving text processing.

UNIT I :

INTRODUCTION :Natural Language Processing tasks in syntax, semantics, and pragmatics – Issues – Applications – The role of machine learning – Probability Basics –Information theory – Collocations -N-gram Language Models – Estimating parameters and smoothing – Evaluating language models.

UNIT II :

WORD LEVEL AND SYNTACTIC ANALYSIS :Word Level Analysis: Regular Expressions Finite-State Automata-Morphological Parsing-Spelling Error Detection and correction-Words and Word classes-Part-of Speech Tagging. Syntactic Analysis: Context free GrammarConstituency-Parsing-Probabilistic Parsing.

UNIT III:

SEMANTIC ANALYSIS AND DISCOURSE PROCESSING :Semantic Analysis: Meaning Representation-Lexical Semantics- Ambiguity-Word Sense Disambiguation. Discourse Processing: cohesion-Reference Resolution- Discourse Coherence and Structure.

UNIT IV :

NATURAL LANGUAGE GENERATION AND MACHINE TRANSLATION : Natural Language Generation: Architecture of NLG Systems- Generation Tasks and Representations Application of NLG. Machine Translation: Problems in Machine Translation- Characteristics of Indian Languages- Machine Translation Approaches-Translation involving Indian Languages.

UNIT V :

INFORMATION RETRIEVAL AND LEXICAL RESOURCES :Information Retrieval: Design features of Information Retrieval Systems-Classical, Non-classical, Alternative Models of Information Retrieval – valuation Lexical Resources: World Net-Frame NetStemmers-POS Tagger- Research Corpora

Text Books:

1. Natural Language Processing with Python. Steven Bird, Ewan Klein, and Edward Lope, O'Reily, 2009
2. Natural Language Processing Recipes: Unlocking Text Data with Machine Learning andDeep Learning using Python. Akshay Kulkarni, Adarsha Shivananda, Apress, 2019

Suggested Reading:

References Books:

3. Allen James, Natural Language Understanding, Benjamin/Cumming,1995.
4. Charniack, Eugene, Statistical Language Learning, MIT Press, 1993.

SEMESTER-VII

COURSE 20: N A T U R A L L A N G U A G E P R O C E S S I N G

Practicals Credits: 1 2 hrs/week

1. Text segmentation: Segment a text into linguistically meaningful units, such as paragraphs, sentences, or words. Write programs to segment text (in different formats) into tokens (words and word-like units) using regular expressions. Compare an automatic tokenization with a gold standard
2. Part-of-speech tagging: Label words (tokens) with parts of speech such as noun, adjective, and verb using a variety of tagging methods, e.g., default tagger, regular expression tagger, unigram tagger, and n-gram taggers.
3. Text classification: Categorize text documents into predefined classes using Naïve BayesClassifier and the Perceptron model
4. Chunk extraction, or partial parsing: Extract short phrases from a part-of-speech tagged sentence. This is different from full parsing in that we're interested in standalone chunks, or phrases, instead of full parse trees
5. Parsing: parsing specific kinds of data, focusing primarily on dates, times, and HTML. Make use of the following preprocessing libraries:
 dateutil which provides datetime parsing and timezone conversion
 lxml and BeautifulSoup which can parse, clean, and convert HTML
 charade and UnicodeDammit which can detect and convert text character encoding
6. Sentiment Analysis: Using Libraries TextBlob and nltk, give the sentiment of a document

SEMESTER-VIII

COURSE 21 A: R E S E A R C H E X P L O R A T I O N

Theory Credits: 3 3 hrs/week

Course Objective: The purpose of a research methodology is to explain the reasoning behind your approach to your research - you'll need to support your collection methods, methods of analysis, and other key points of your work. Think of it like writing a plan or an outline for you what you intend to do.

UNIT I – Foundations of Research: Meaning, Objectives, Motivation, Utility. Concept of theory, empiricism, deductive and inductive theory. Characteristics of scientific method – Understanding the language of research – Concept, Construct, Definition, Variable. Research Process

UNIT II – Problem Identification & Formulation – Research Question – Investigation Question
– Measurement Issues – Hypothesis – Qualities of a good Hypothesis –Null Hypothesis & Alternative Hypothesis. Hypothesis Testing – Logic & Importance

UNIT III – Research Design: Concept and Importance in Research – Features of a good research design – Exploratory Research Design – concept, types and uses, Descriptive Research Designs – concept, types and uses. Experimental Design: Concept of Independent & Dependent variables.

UNIT IV – Qualitative and Quantitative Research: Qualitative research – Quantitative research
– Concept of measurement, causality, generalization, replication. Merging the two approaches. Measurement: Concept of measurement– what is measured? Problems in measurement in research – Validity and Reliability. Levels of measurement – Nominal, Ordinal, Interval, Ratio.

UNIT V – Sampling: Concepts of Statistical Population, Sample, Sampling Frame, Sampling Error, Sample Size, Non Response. Characteristics of a good sample. Probability Sample – Simple Ran-dom Sample, Systematic Sample, Stratified Random Sample & Multi-stage sampling. Determining size of the sample – Practical considerations in sampling and sample size

Text Books Recommended:-

Business Research Methods – Donald Cooper & Pamela Schindler, TMGH, 9th edition
Business Research Methods – Alan Bryman & Emma Bell, Oxford University Press.
Research Methodology – C.R.Kothari

SEMESTER-VIII

COURSE 21 A: R E S E A R C H E X P L O R A T I O N L A B

Practical Credits: 1 2 hrs/week

Research Exploration: Case Studies

The purpose of a research methodology is to explain the reasoning behind your approach to your research - you'll need to support your collection methods, methods of analysis, and other key points of your work. Think of it like writing a plan or an outline for you what you intend to do.

1. Apply the research concepts and tools to identify a research problem, sub-statements of the problem to clarify the direction of research effort?
2. Construct a research framework showing the variables of the study relevant for a specific area of inquiry
3. Design a research methodology by classifying the research design, population and sampling techniques suitable for a research problem.
4. Develop a research proposal by using principles and frameworks of research methodology to justify an ethical, practical and effective research project.

SEMESTER-VIII

COURSE 21 B: C O M P U T A T I O N A L D A T A S C I E N C E

Theory Credits: 3 3 hrs/week

Aim and objectives of Course (Predictive and Advanced Analytics):
The course enables students to:
1. To learn descriptive statistics using computational tools
2. To learn distributions in probability.
3. To work with different types of dispersion measures.

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

- To understand different computational tools and statistical techniques.
- To learn about curve fitting in data science.
- Able to perform visual diagnostics.

UNIT I

Data Science Computational Tools - Statistical Techniques -Plotting the Classics - Literary Characters - Causality and Experiments - Tables -Randomness - Conditional Statements - Iteration - Simulation .

UNIT II

Finding Probabilities - Sampling and Empirical Distributions - Empirical Distributions - Sampling from a Population - Empirical Distribution of a Statistic - Testing Hypotheses - Assessing Models - Multiple Categories - Decisions and Uncertainty - Error Probabilities.

UNIT III

Comparing Two Samples - A/B Testing –Deflategate - Causality - Estimation – Importance of Mean- Properties of the Mean - Variability - The SD and the Normal Curve.

UNIT IV

The Central Limit Theorem - The Variability of the Sample Mean - Choosing a Sample Size - Prediction - Correlation - The Regression Line - The Method of Least Squares - Least Squares Regression

UNIT V

Visual Diagnostics -Numerical Diagnostics. Inference for Regression - A Regression Model -Inference for the True Slope - Prediction Intervals

TEXT BOOK:

1. Ani Adhikari and John DeNero, David Wagner "Computational and Inferential Thinking: TheFoundations of Data Science", 2019.
2. Web link: https://inferentialthinking.com/chapters/intro.html

REFERENCES:

1. The Art of Data Science: A Guide for Anyone Who Works with Data, Roger D. Peng, andEliz-abeth Matsui, ISBN: 9781365061462, 2018.
2. The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios.by Steve Wexler, Jeffrey Shaffer, Andy Cotgreave, ISBN: 1119282713,2017

SEMESTER-VIII

COURSE 21 B: C O M P U T A T I O N A L D A T A S C I E N C E L A B

Practical Credits: 1 1 hrs/week

1. Project Gutenberg is a website that publishes public domain books online. Using Python,load the text of two classic novels directly from the web.
2. Create a Table, add new columns, add data, move data to a table using csv file, access the data in a column and choose sets of columns.
3. Perform scatter Plots and Line Graphs using Python(https://inferentialthinking.com/chapters/07/Visualization.html)
4. Write a Program that simulates Multiple values
5. Based on the top_movies_2017.csv data set,draw some samples, create a deterministic sample,find probability sample and find systematic sample.(https://inferentialthinking.com/chapters/10/Sampling_and_Empirical_Distributions.html)
6. Simulate one value, multiple value of the statistic and visualize the results in an empirical history ram circuits.(https://inferentialthinking.com/chapters/11/1/Assessing_a_Model.html)
7. Under A/B Testing, observed difference, Predict the static under null hypothesis, and perform permutations test.
8. Using Central Limit theorem, find out the average flight delay from the database 'united_summer2015.csv.'

SEMESTER-VIII

COURSE 22 A: C O M P U T E R V I S I O N W I T H P Y T H O N

Theory Credits: 3 3 hrs/week

Course Outcomes:
This course enables the learners to understand the advanced concepts in computer vision. The course covers the basics of image processing, imagingg eometry, image segmentation, feature extraction, object recognition and classification and common applications of computer vision.
This course helps the students to design solutions for complex real-life problems.

Unit – 1 (Image Formation and Processing) Image formation and Image model- Components of a vision system- Cameras- camera model and camera calibration-Radiometry- Light in space- Light in surface - Sources, shadows and shading. Fundamentals of Image processing: Basic steps of Image processing system sampling and quantization of an Image – Basic relationship between pixels.

Unit - 2(Feature Extraction) Points and Patches – Feature detectors,feature descriptors, feature matching, feature tracking. Edges – edge detection, edge linking. Lines - Successive approximation, Hough transforms, Vanishing points.

Unit - 3 (Image Segmentation) Classification of segmentation techniques, Edge detection, Edge linking, Thresholding, Region growing, Region splitting and merging, Watershed based segmentation. Shadow detection and removal. Image processing using OpenCV - blending, smoothing, and reshaping.

Unit - 4 (Image Recognition and Classification) Shape based object classification, Motion based object classification, Viola Jones Object Detection Framework, Object classification using CNNs, use of RCNN for object classification.

Unit - 5 (Applications) Speech and Handwriting Recognition, Automatic Face Recognition, Video Segmentation and Keyframe Extraction, Real-Time Hand Pose Recognition.
Text Books:
1. David A. Forsyth & Jean Ponce, Computer vision – A Modern Approach, Prentice Hall, 2002.
2. Richard Szeliski, Computer Vision: Algorithms and Applications, Springer.
3. Maheshkumar H Kolekar, “Intelligent Video Surveillance Systems:An Algorithmic Approach”, CRC Press. ARTIFICIAL INTELLIGENCE AND DATA SCIENCE
4. Francesco Camastra, Alessandro Vinciarelli, “Machine Learning for Audio, Image and Video Analysis: Theory and Applications”, Springer 2015.
Reference Books 1. Reinhard Klette, “Concise Computer Vision: An Introduction into Theory and Algorithms”, Springer London, 2014.
2. Olivier Faugeras, “Three-Dimensional Computer Vision”, The MITPress, 1993.

SEMESTER-VIII

COURSE 22 A: C O M P U T E R V I S I O N W I T H P Y T H O N L A B

Practical Credits: 1 2 hrs/week

1. Reducing the Number of Intensity Levels in an Image.
2. Zooming and Shrinking Images by Pixel Replication.
3. Zooming and Shrinking Images by Bilinear Interpolation.
4. Arithmetic Operations.
5. Image Enhancement Using Intensity Transformations.
6. Histogram Equalization.
7. Spatial Filtering.
8. Enhancement Using the Laplacian.
9. Unsharp Masking

SEMESTER-VIII

COURSE 22 B: D A T A W R A N G L I N G W I T H J A V A S C R I P T

Theory Credits: 3 3 hrs/week

Course Outcomes:

Aim and objectives of Course (Predictive and Advanced Analytics):The course enables students to:

To learn how to use REST APIs.

To learn different types of data formats.

To learn connectivity of different databases like MYSQL , MongoDB.

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

To learn working with NodeJS

To learn how to host a server and run a server in local host.

Importing and exporting data from different data formats.

Develop a web visualization application

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 dataParsing CSV text data-Importing data from databases - Importing data from MongoDBImporting 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 - Output-ting a new CSV file

UNIT V

Exploratory coding in the browser - 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

SEMESTER-VIII

COURSE 22 B: D A T A W R A N G L I N G W I T H J A V A S C R I P T L A B

Practical Credits: 1 2 hrs/week

1. Install Node.js and install dependencies
2. Run one web application and get the data
3. Install “Express Node.js framework” . Create a simple Web application to print yourcollege name on Web Page.
4. Create a simple web server and add static files to it
5. Add a REST API to web server to dynamically generate a report.
6. 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 CSVfile.
7. Retrieve data from REST API
8. Import Earthquakes data from REST API
9. Import data from MySql
10. Use HTTP GET to retrieve the data from your CSV file.
11. Develop a program to clean your data by rewriting rows to fix bad data.
12. Develop a web visualization application by importing data from Excel file.

SEMESTER-VIII

COURSE 23 A: S O C I A L M E D I A A N A L Y T I C S

Theory Credits: 3 3 hrs/week

OBJECTIVES

To Understand the Complete Architecture of Social Media Analytics

 To know Web analytics tools

 To know the social media of ongoing campaigns

OUTCOMES

 Students will get well knowledge of what is Social Media AnalyticsKnowledge in Web analytics tools & Mapping of Network Analysis. (LinkedIn, Instagram, YouTube Twitter etc.)

UNIT-I

Introduction to Social Media Analytics (SMA): Social media landscape, Need for SMA; SMA in Small organizations; SMA in large organizations; Application of SMA in different areas Network fundamentals and models: The social networks perspective - nodes, ties and influencers, Social network and web data and methods.

UNIT-II

Graphs and Matrices- Basic measures for individuals and networks. Information visualization Making connections: Link analysis. Random graphs and network evolution. Social contexts:Affiliation and identity. Web analytics tools and techniques: Click stream analysis, A/B testing,online surveys.

UNIT-III

Web crawling and Indexing; Natural Language Processing Techniques for Micro-text Analysis Facebook Analytics: Introduction, parameters, demographics. Analyzing page audience. Reachand Engagement analysis.

UNIT-IV

Processing and Visualizing Data, Influence Maximization, Link Prediction, Collective Classification. Applications in Advertising and Game Analytics (Use of tools like Unity30 / PyCharm).Introduction to Python Programming, Collecting and analyzing social media data; visualization and exploration

UNIT-V

Social campaigns. Measuring and Analyzing social campaigns, defining goals and evaluating outcomes, Network Analysis. (LinkedIn, Instagram, YouTube Twitter etc.) Use of Google Analytics. Post Performance on FB, Use of Facebook Business Manager

Text Books:

1.Mathew Ganis, Avinash Koihrkar Social Media Analytics IBM Press 2015 / 1st2 Jim Sterne Social Media Metrics Wiley Latest

3 Oliver Blanchard Social Media ROI QuePublishing Latest

4 Marshall Sponder, Gorah F. KhanDigital Analytics for Marketing Routledge 2017 / 1st

Reference books:

1. Marshall Sponder Social Media Analytics McGraw Hill Latest

2. Tracy L. Tuten, Michael R. Solomon Social Media Marketing Sage 2018 / 3rd

3. Gohar F. Khan Creating Value With Social Media AnalyticsCreateSpaceIndependentPublishing 2018 1st

4. Alex Gonsalves Social Media Analytics Strategy Appress 2017 / 1st

SEMESTER-VIII

COURSE 23 A: S O C I A L M E D I A A N A L Y T I C S L A B

Practical Credits: 1 2 hrs/week

Using Python Implement the following Programs

1.Managing text data

2.Syntactical analysis

3.Vector semantics and latent semantic analysis (LSA)

4.Clustering and topic modeling:
PLSA, LDA, SLDA, ...

5.Text classification: naive Bayesian, maximum entropy, SVM, ...

6. Clustering Physician Reviews

7. Discovering Topics on Twitter

8.Deep Learning with Text

9.Students should analyses the social media of any ongoing campaigns and present the findings

SEMESTER-VIII

COURSE 23 B: M A J O R P Y S P A R K E S S E N T I A L S F O R D A T A

Theory Credits: 3 3 hrs/week

Course Outcomes:

- □ To learn PYSPARK, and to develop models
- □ To know the use of PYSPARK to automate model selection.
- □ To advice on when and how to use each model. Also learn how to combine two or more Models.

UNIT-I

Data Engineering
Distributed Computing Primer: Technical requirements, Distributed Computing, Distributed Computing with Apache Spark, Big data processing with Spark SQL and Data Frames.
Data Ingestion: Technical requirements, Introduction to Enterprise Decision Support Systems,ingesting data from data sources, ingesting data into data sinks, Using file formats for data storage in data lakes, Building data ingestion pipelines in batch and real time, Unifying batch and real time using Lambda Architecture.

UNIT-II
Data Cleansing and Integration: Technical requirements, transforming raw data into enriched meaningful data, building analytical data stores using cloud data lakes, consolidating data using data integration, Making raw data analytics-ready using data cleansing.
Real-Time Data Analytics: Technical requirements, Real-time analytics systems architecture, Stream processing engines, Real-time analytics industry use cases, Simplifying the Lambda Architecture using Delta Lake, Change Data Capture, Handling late-arriving data, Multi-hop pipelines.

UNIT-III
Data Science
Scalable Machine Learning with PySpark: Technical requirements, ML overview, Scaling out machine learning, Data wrangling with Apache Spark and MLlib.
Feature Engineering – Extraction, Transformation, and Selection: Technical requirements, The machine learning process, Feature extraction, Feature transformation, Feature selection, Featurestore as a central feature repository, Delta Lake as an offline feature store

UNIT-IV
Data Analysis
Data Visualization with PySpark:, Technical requirements, Importance of data visualization, Techniques for visualizing data using PySpark, Considerations for PySpark to pandas conversion.
Spark SQL Primer: Technical requirements, Introduction to SQL, Introduction to Spark SQL, Spark SQL language reference, Optimizing Spark SQL performance.

UNIT-V
Integrating External Tools with Spark SQL: Technical requirements, Apache Spark as a distributed SQL engine, Spark connectivity to SQL analysis tools, Spark connectivity to BI tools, Connecting Python applications to Spark SQL using Pyodbc The Data Lakehouse: Moving from BI to AI, The data lakehouse paradigm, Advantages of data lakehouses.

Text Books:

1. Essential PySpark for Scalable Data Analytics, by Sreeram Nudurupati, Packt Publishing, 2021.
2. Applied Data Science Using Pyspark, by Ramcharan Kakarla, Sundar Krishnan, Sridhar Alla,Apress, Springer India
Reference Books:
1. Machine Learning with Pyspark, by Pramod Singh, Apress India
2. Learning Pyspark, by Tomasz Drabas Denny Lee, Packt Publications, 2017.PYSPARK ESSENTIALS FOR DATA SCIENCE LAB

SEMESTER-VIII

COURSE 23 B: M A J O R P Y S P A R K E S S E N T I A L S F O R D A T A L A B

Practical Credits: 1 2 hrs/week

1. Installing PySpark
2. Demonstrate on Big Data concepts in PythonLambda Functions
Filter(), map(), amd reduce()Sets.
3. Hello World in PySpark
4. PySpark API and Data Strtuctures
5. Running PySpark Programs in
    - Jupyter Notebook
    - Command-Line Interface
    - Cluster
    - PySpark Shell
6. Combining PySpark with other Tools
7. Demonstrate on Data Cleaning with PySpark
8. Demonstrate on Data wrangling with Apache Spark and MLlib
9. Demonstrate Data Visualization with PySpark
10. Demonstrate Connecting Python applications to Spark SQL using Pyodbc

SEMESTER-VIII

COURSE 24: M A J O R P Y S P A R K E S S E N T I A L S F O R D A T A

Theory Credits: 3 3 hrs/week

OBJECTIVES:

- The student should be made to be familiar with the Business intelligence architectures most fundamental Graphs and results.
- Be exposed to the techniques of proofs and analysis.

Detailed Syllabus:

UNIT-I

BUSINESS INTELLIGENCE: Effective and timely decisions - Data, information, and knowledge - Role of mathematical models - Business intelligence architectures: Cycle of a business intelligence analysis - Enabling factors in business intelligence projects

UNIT-II

KNOWLEDGE DELIVERY: The business intelligence user types. Standard reports. Interactive Analysis. and AdHoc Querying, Parameterized Reports: and self-Service Reporting,dimensional analysis Analysis/notifications, Visualizations: Charts, Graphs, Widgets, Scorecards, and Dashboard.

UNIT- III

EFFICIENCY: Efficiency measures -The CCR model: Definition of target objectives- peer groups - Identification of good operating practices; cross-efficiency analysis - virtual inputs and outputs - Other models. Pattern matching -cluster analysis, outlier analysis

UNIT-IV

BUSINESS INTELLIGENCE APPLICATIONS: Marketing model- Logistic and Production models -Case studies. Development of a business intelligence system -Ethics and business intelligence. Geographic Visualization, Integrated Analytics, Considerations: Optimizing the Presentation for the Right Message.

UNIT- V

FUTU RE OF BUSINESS INTELLIGENCE: Future of business intelligence - Emerging Technologies, Machine Learning, Predicting the Future, BI Search & Text Analytics - Advanced Visualization -Rich Report, Future beyond Technology.

TEXT BOOK:

I. Efraim Turban, Ramesh Sharda, Dursun Deleo, "Decision Support and Business IntelligenceSystems". 9th Edition, Pearson 20 13.

.

REFERENCES:

I. Larissa T. Moss, S. Acre, "Business Intelligence Roadmap: The Complete Project Lifecycle of Decision Making", Addison Wesley,2003.
2. Carlo Vercellis, "Business Intelligence: Data Mining and Optimization for Decision Making ", Wiley Publications,2009.
3. David Loshin Morgan, Kaufman. "Business Intelligence: The Savvy Managers Guide", Second Edition,20 12.
4. Cindi Howson, "Successful Business Intelligence: Secret< to Making Bl " Killer App".McGraw-Hill,2007.

5. Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker, "The DataWarehouse Lifecycle Toolkit", Wiley publication lnc.,2007.

SEMESTER-VIII

COURSE 24: M A J O R P Y S P A R K E S S E N T I A L S F O R D A T A L A B

Practical Credits: 1 2 hrs/week

Practical 1: Import the legacy data from different sources such as (Excel, SqlServer, Oracle etc.) and load in the target system.
Practical 2: Perform the Extraction Transformation and Loading (ETL) process to construct the database in the Sqlserver / Power BI.
Practical 3: Data Visualization from ETL Process Power BI Desktop Practical 4: Creating a Cube in SQL server 2012
Practical 5: Apply the what – if Analysis for data visualization. Design and generate necessary reports based on the data warehouse data.
Practical 6: Implementation of Classification algorithm in R Programming.
Practical 7: Practical Implementation of Decision Tree using R Tool
Practical 8: k-means clustering using R
Practical 9: Prediction Using Linear Regression Practical 10: Data Analysis using Time Series Analysis
Practical 11: Data Modelling and Analytics with Pivot Table in Excel
Practical 12: Data Analysis and Visualization using Advanced Excel

SEMESTER –VIII

COURSE 25: DATA VISUALIZATION USING JAVA SCRIPT

Theory **Credits: 3** **3 hrs/week**

Objectives:

The main objectives of D**ata Visualization using JavaScript** are to effectively communicate information, enable user interaction, and support decision-making through visual representation.

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

Unit 1: Introduction to Data Visualization and JavaScript Basics: Importance & principles of data visualization, Types of charts and graphs (bar, pie, line, scatter, etc.). Overview of JavaScript for data visualization, JavaScript essentials (ES6+ syntax, arrays, objects, functions) HTML & CSS basics (DOM manipulation) Working with datasets (JSON, CSV) Fetching and parsing data using fetch() API

Unit II: Chart Creation Using Chart.js: Introduction to Chart.js Setting up Chart.js in a web project Creating basic chart types: Bar, Line, Pie, Radar. Chart customization: Labels, colors, axes, legends, and tooltips. Responsive design & real-time updates Adding interactivity with JavaScript events. Learners can create and customize static and dynamic charts using Chart.js.

Unit III: Interactive Visualizations with D3.js – Basics: Introduction to D3.js and the data binding concept Creating SVG elements with D3 Selections, data joins, and binding Scales, axes, and color schemes Creating basic visualizations: Bar chart Line chart Scatter plot

Unit IV: Advanced Visualizations and Interactivity with D3.js : Enter-update-exit pattern Transitions and animations Tooltips and hover interactions Dynamic & real-time data handling Building composite visuals (multiple charts in one view) Mini-project: Interactive data story using D3.js

Unit V: Capstone Project and Geospatial Visualizations: Introduction to map visualizations Plotting GeoJSON data on maps D3 with maps and external APIs Overview of tools: Leaflet.js, Google Maps API Final Project: Build an interactive dashboard or visualization project Deployment: Hosting on GitHub Pages, Netlify, or Vercel Documentation and presentation

Text Book:

1. Scott Murray, "Interactive Data Visualization for the Web", 2nd Edition, O'Reilly Media 2017.

Reference Books:

1. Nick Qi Zhu, "**Data Visualization with D3 4.x Cookbook",** 2nd Edition Packt Publishing.

2. Ændrew Rininsland & Swizec Teller, "Learning D3.js Data Visualization", 2nd Edition Packt Publishing.

3. Stephen A. Thomas, "Data Visualization with JavaScript", No Starch Press **ISBN**: 978-1593275204.

SEMESTER – VIII

COURSE 25: DATA VISUALIZATION USING JAVA SCRIPT

Practical	Credits: 1	2 hrs/week

Practical 1: Fetching and Displaying JSON Data

Practical 2: Creating a Static Bar Chart Using Chart.js

Practical 3: Building a Responsive Line Chart Using Chart.js

Practical 4: Pie and Doughnut Charts for Categorical Data

Practical 5: Creating a Basic Bar Chart Using D3.js

Practical 6: Drawing Axes and Scales in D3.js

Practical 7: Interactive Tooltip with D3.js

Practical 8: Real-Time Chart Update Using JavaScript Events

Practical 9: Creating a Choropleth Map Using D3.js and GeoJSON

Practical 10: Mini Project – Data Dashboard with Multiple Visualizations

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA (Data Science)

Semester - 5

MACHINE LEARNING [Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION-A

(Short Answer Type Questions) Marks: 5 x 4 M= 20M

Answer any FIVE questions out of the following EIGHT questions

1. What are the different types of human learning, and how do they relate to machine learning paradigms?
2. Discuss common issues in machine learning and suggest ways to remediate poor data quality.
3. Explain the concept of feature transformation with examples.
4. How can you improve the performance of a machine learning model?
5. Define a random variable and differentiate between discrete and continuous distributions with examples.
6. What is the significance of Bayes Theorem in machine learning?
7. Compare and contrast supervised and unsupervised learning with suitable examples.
8. Compare and contrast supervised and unsupervised learning with suitable examples.

SECTION-B

(Long Answer Type Questions) Marks: 5x10M= 50M

Answer any **FIVE** out of the following **TEN** questions

9. a) What is machine learning? Describe different types of machine learning and provide examples where ML should and should not be applied.

OR

b) Analyze the **Trendyol** case study and explain how machine learning is used in online retail platforms. Mention the steps from data collection to prediction.

10. a) Explain the steps involved in training and evaluating a machine learning model. Include aspects of model selection and interpretability

OR

b) Discuss how **Harley Davidson** utilizes machine learning for business operations or customer analytics. Relate this to performance metrics and feature engineering.

11. a) Explain the role of probability in machine learning. Describe any two probability distributions (one discrete, one continuous) commonly used.

OR

b) Elaborate on the **Tesla** case study. How might Bayesian learning and belief networks assist Tesla in autonomous vehicle decision-making?

12. a) Discuss classification and regression models with examples of algorithms used. Highlight the major differences in their learning goals.

OR

b) Explain how machine learning is used to **predict heart failure** in a mobile health setting. Describe data preprocessing, model selection, and evaluation.

13. a) What is deep learning? Explain the architecture and learning process of a neural network with reference to backpropagation.

OR

b) Discuss how **American Express (or American Airlines)** applies neural networks or other machine learning models to detect fraud, manage customer satisfaction, or enhance personalization.

BLUE PRINT

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA Data Science

Semester - 5

Machine Learning

Unit	S.A.Q **Section :A** (including choice) 4 Marks	E.Q **Section: B** (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	2	2	28
III	1	2	24
IV	2	2	28
V	1	2	24
Total	8	10	132

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA (Data Science)

Semester - 5

CLOUD COMPUTING [Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION-A

(Short Answer Type Questions) Marks: 5 x 4 M= 20M

Answer any FIVE questions out of the following EIGHT questions

1. What are the key differences between public, private, and hybrid cloud deployment models?
2. List any four major benefits and four major challenges of adopting cloud computing.
3. What are the core elements of a Cloud Security Model? Briefly explain each.
4. How does cloud security differ from traditional on-premises computing security?
5. Define virtualization. What are the major types of virtualization?
6. What are some common security threats associated with virtualization?
7. Define **multi-tenancy** in the context of cloud computing. What are its advantages?
8. Briefly explain any two types of scaling and when they are used in cloud environments.

SECTION-B

(Long Answer Type Questions) Marks: 5x10M= 50M

Answer any **FIVE** out of the following **TEN** questions

9. a) Explain the concept of cloud computing. Discuss the enabling technologies and the evolution that led to modern cloud systems.

OR

b) Compare and contrast public, private, and hybrid clouds. Identify popular real-time services that use each model (e.g., AWS, Azure Stack, VMware).

10. a) Describe the cloud security reference model in detail. How does it ensure data protection in cloud environment?

OR

b) Explore and describe **three real-time cloud services** (like Google Cloud Functions, AWS S3, or Azure Blob Storage). Discuss how they are used in real-world applications.

11. a) Explain machine/server-level virtualization with the help of a diagram. Mention top virtualization vendors and products.

OR

b) Install and demonstrate any virtualization software (like VMware Workstation, VirtualBox, or Hyper-V). Describe the task performed, such as installing an OS.

12. a) Analyze the advantages and disadvantages of virtualization. In what scenerios does it outperform traditional server management

OR

b) What is load balancing in cloud computing? Explain its types and how it supports scaling and performance.

13. a) Describe the concept of Service-Oriented Architecture (SOA). Discuss its elements and benefits in a cloud environment.

OR

b) Discuss resource pooling and sharing in the cloud. How does it enable efficient resource provisioning in a multi-tenant system?

BLUE PRINT

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA Data Science

Semester - 5

Cloud Computing

Unit	S.A.Q **Section :A** (including choice) 4 Marks	E.Q **Section: B** (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	2	2	28
III	1	2	24
IV	2	2	28
V	1	2	24
Total	8	10	132

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA (Data Science)

Semester - 5

UNSUPERVISED MACHINE LEARNING WITH PYTHON

[Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION-A

(Short Answer Type Questions) Marks: 5 x 4 M= 20M

Answer any FIVE questions out of the following EIGHT questions

1. What is the k-means clustering algorithm? Describe its main steps.
2. How does the bisecting k-means algorithm differ from standard k-means?
3. Explain the Apriori principal in the context of Association Analysis?
4. What are frequent item sets, and how are they used to generate association rules?
5. What is an FP-tree, and how is it constructed?
6. How can frequent pattern growth (FP-Growth) be applied to analyze Twitter data?
7. What is Principal Component Analysis (PCA), and how does it help in dimensionality reduction?
8. List two real-world applications of Singular Value Decomposition (SVD).

SECTION-B

(Long Answer Type Questions) *Marks*: 5x10M= 50M

Answer any **FIVE** out of the following **TEN** questions

9. a) Describe the k-means clustering algorithm. Discuss how clustering can be applied to points on a geographical map, and explain the concept of post-processing in improving cluster performance.

OR

b) Compare k-means and bisecting k-means clustering methods. Illustrate with an example when bisecting k-means is preferred.

10. a) Explain the Apriori algorithm for mining association rules. How does the Apriori principle help reduce the search space?

OR

b) Discuss the process of uncovering patterns using association rules in a dataset like congressional voting records. Mention the steps involved in extracting and interpreting rules.

11. a) What is the FP-Growth algorithm? Describe the steps to build an FP-tree and use it to mine frequent item sets efficiently.

OR

b) Describe how FP-Growth can be applied to mine clickstream data from a news website. How does it help in recommending content?

12. a)Explain Principal Component Analysis (PCA). Discuss its mathematical intuition and how it can be used to reduce the dimensionality of a high-dimensional dataset like semiconductor manufacturing data

OR

b) What are the key differences between PCA and other dimensionality reduction techniques? How do you interpret the principal components?

13. a) What is Singular Value Decomposition (SVD)? Explain its use in collaborative filtering with an example of a restaurant recommendation engine.

OR

b) Describe the concept of matrix factorization in SVD and its role in building recommendation systems. Include implementation hints in Python.

BLUE PRINT

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA Data Science

Semester - 5

Unsupervised machine learning with python

Unit	S.A.Q **Section :A** (including choice) 4 Marks	E.Q **Section: B** (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	2	2	28
III	1	2	24
IV	2	2	28
V	1	2	24
Total	8	10	132

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA (Data Science)

Semester - 5

SUPERVISED MACHINE LEARNING WITH PYTHON

[Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION-A

(Short Answer Type Questions) Marks: 5 x 4 M= 20M

Answer any FIVE questions out of the following EIGHT questions

1. What are the key tasks in machine learning, and how do they differ from each other?
2. Explain the purpose of normalizing numeric values before applying k-Nearest Neighbours.
3. Describe how consistency in a dataset is measured when building a decision tree.
4. What is recursion, and how is it used to construct a decision tree?
5. Write a short note on how the naïve Bayes classifier uses probability distributions for classification?
6. What role does the sigmoid function play in logistic regression?
7. Describe how the gradient descent algorithm is used to optimize regression coefficients.
8. What is a kernel in Support Vector Machines, and how does it help in transforming data?

SECTION-B

(Long Answer Type Questions) *Marks*: 5x10M= 50M

Answer any **FIVE** out of the following **TEN** questions

9. a) Define machine learning. Explain the steps involved in developing a machine learning application, and discuss why Python is a preferred language for ML.

OR

b) Explain the working of the k-Nearest Neighbors (k-NN) algorithm. Illustrate its implementation using NumPy and Matplotlib for a classification task.

10. a) Describe the process of building the decision tree classifier. Explain how decision tree use feature splitting and give a example using simple dataset?

OR

b) How can decision trees be visualized using Matplotlib? Discuss its importance in model interpretability with a use case.

11. a) Explain the Naïve Bayes classification technique with an example. How can RSS feeds be used to extract data for classification?

OR

b) Discuss how Naïve Bayes can reveal regional attitudes using text classification. Highlight preprocessing and feature extraction steps.

12. a) What is logistic regression? Discuss its mathematical formulation and how optimization is used to find the best regression coefficients.

OR

b) How does the gradient descent algorithm work? Write the steps involved and discuss how missing values in the data are handled during logistic regression.

13. a) What are Support Vector Machines (SVMs)? Explain the working of SVMs with the SMO optimization algorithm.

OR

b) Compare Support Vector Machines with other classifiers like k-NN, Decision Trees, and Naïve Bayes in terms of accuracy, complexity, and interpretability.

BLUE PRINT

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA Data Science

Semester - 5

Supervised learning with python

Unit	S.A.Q **Section :A** (including choice) 4 Marks	E.Q **Section: B** (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	2	2	28
III	1	2	24
IV	2	2	28
V	1	2	24
Total	8	10	132

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA (Data Science)

Semester - 5

ADVANCED DATA ANALYSIS [Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION-A

(Short Answer Type Questions) Marks: 5 x 4 M= 20M

Answer any FIVE questions out of the following EIGHT questions

1. What are the key stages in a Data Science project life cycle? Briefly describe each.
2. List and explain any three major applications of data science in real-world domains.
3. Explain the difference between data cleaning and data transformation.
4. What is data discretization? Provide a scenario where it might be useful.
5. What are skewness and kurtosis? How do they help in understanding data distribution?
6. Describe the use of heat maps in exploratory data analysis (EDA).
7. Differentiate between simple and multiple regression with examples.
8. What is cross-validation? Why is it important in model evaluation?

SECTION-B

(Long Answer Type Questions) *Marks*: 5x10M= 50M

*Answer any **FIVE** out of the following **TEN** questions*

9. a) Define Data Science. Explain its evolution over time and describe different roles in a data science team.

OR

b) Discuss major data security issues associated with data science workflows. How can these be addressed?

10. a) Describe the complete process of data pre-processing. Explain the significance of each step.

OR

b) With examples, explain various strategies used in data collection and the impact of poor data collection.

11. a) Explain the use of descriptive statistics in exploratory data analysis. Include examples of mean, standard deviation, and correlation.

OR

b) What is ANOVA? How does it help in hypothesis testing across multiple groups?

12. a) Describe polynomial regression and how pipelines can simplify the model building process

OR

b) Explain the concept of residual and distribution plots. How are they used in model evaluation?

13. a) What is overfitting and underfitting? Explain with examples how model selection helps to avoid them.

OR

b) Discuss Ridge Regression and Grid Search. How are they used for prediction and parameter tuning?

BLUE PRINT

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA Data Science

Semester - 5

Advanced Data Analysis

Unit	S.A.Q **Section :A** (including choice) 4 Marks	E.Q **Section: B** (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	2	2	28
III	1	2	24
IV	2	2	28
V	1	2	24
Total	8	10	132

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA (Data Science)

Semester - 5

SOFTWARE TESTING [Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION-A

(Short Answer Type Questions) Marks: 5 x 4 M= 20M

Answer any FIVE questions out of the following EIGHT questions

1. What is software testing
2. List the selenium component.
3. Explain about frames.
4. Explain handling alerts/pop-ups.
5. What is automation framework
6. What are the uses of TestNG
7. Write the uses of Selenium Grid.
8. Write about Grid 1.0 Vs Grid 2.0

SECTION-B

(Long Answer Type Questions) Marks: 5x10M= 50M

Answer any **FIVE** out of the following **TEN** questions

9. a) Explain about selenium IDE and its features?

OR

b) Explain about creating scripts using firebug.

10. a) Explain about selenium webdriver installation with eclipse.

OR

b) Explain about Robot API and AutoIT

11. a) Explain about benefits of Automation Framework.

OR

b) Explain about different types of Automation Framework

12. a) Explain about TestNG and write an example?

OR

b) Explain about TestNG process execution

13. a) Explain about selenium grid architecture

OR

b) Define RemoteWebDriver Object and write about its applications?

BLUE PRINT

BCA Data Science

Semester - 5

SOFTWARE TESTING

Unit	S.A.Q **Section :A** (including choice) 4 Marks	E.Q **Section: B** (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	2	2	28
III	1	2	24
IV	2	2	28
V	1	2	24
Total	8	10	132

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA (Data Science)

Semester - VII

BIG DATA SECURITY

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION-A

(Short Answer Type Questions) Marks: 5 x 4 M= 20M

Answer any FIVE questions out of the following EIGHT questions

14. What is Big Data Security?

15. Explain about Classification of Data?

16. What is Intellectual Configuration?

17. What is Hadoop?

18. What is kerberos?

19. What are the components of Hadoop?

20. What is Scoop?

21. What is SIEM system?

SECTION-B

(Long Answer Type Questions) Marks: 5x10M = 50M

Answer any FIVE out of the following TEN questions

22. a) Explain about Big data Privacy, Ethics and Security?

OR

b) Explain about Ethical Guideline of Big Data?

23. a) Explain about Big Data compliance?

OR

b) Explain about Steps to secure big Data?

24. a) Explain about Hadoop security Design?

OR

b) Explain about Kerberos security implementation and Configuration?

25. a) What is Hadoop Ecosystem Security?

OR

b) Explain key features and functions OOzie?

26. a) Explain about Setting up audit logging in Hadoop cluster?

OR

b) Explain about Integrating Hadoop with Enterprise Security System ?

BLUE PRINT

BCA Data Science

Semester - 7

BIG DATA SECURITY

Unit	S.A.Q Section :A (including choice) 4 Marks	E.Q Section: B (including choice) 10 Marks	Marks Allotted
I	1	2	24
II	2	2	28
III	2	2	28
IV	2	2	28
V	1	2	24
Total	8	10	132

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA (Data Science)

Semester - VII

BIG DATA SECURITY

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION-A

(Short Answer Type Questions) Marks: 5 x 4 M= 20M

Answer any FIVE questions out of the following EIGHT questions

1. What is Big Data?
2. What is Hadoop?
3. What is Apache Spark Features?
4. What is Spark Ecosystem?
5. Write about Saving and loading Data frames?
6. Definition of Machine Learning?
7. Explain about Streaming in Spark?
8. Explain about Graph X features?

SECTION-B

(Long Answer Type Questions) Marks: 5x10M = 50M

Answer any FIVE out of the following TEN questions

9. a) Explain the Applications of Hadoop ?

OR

b) Explain about the Building Blocks of Hadoop?

10. a) Difference between Hadoop and Spark?

OR

b) Explain about Spark Components?

11. a) Explain about RDD actions and Transformations?

OR

b) Explain about Spark SQl?

12. a) Explain about Spark Streaming Applications?

OR

b) Explain about Structured Streaming?

13. a) Explain about Graph Processing with Spark

OR

b) Explain about Shortest path Algorithm

BLUE PRINT

BCA Data Science

Semester - 7

BIG DATA ANALYTICS USING SPARK &HADOOP l [Major / Minor]

Unit	S.A.Q Section :A (including choice) 4 Marks	E.Q Section: B (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	2	2	28
III	1	2	24
IV	2	2	28
V	1	2	24
Total	8	10	132

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA (Data Science)

Semester - VII

INTRODUCTION TO DEEP LEARNING

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION-A

(Short Answer Type Questions) Marks: 5 x 4 M= 20M

Answer any FIVE questions out of the following EIGHT questions

1. What is linear algebra?

2. What is Basic probability Theory?

3. What is Neural Network?

4. What id Recurrent Network?

5. What are Neural Network applications?

6. What is Back Propagation Network?

7. What is Deep Rural Network?

8. What are Gradient problems?

SECTION-B

(Long Answer Type Questions) Marks: 5x10M = 50M

Answer any FIVE out of the following TEN questions

9. a) Explain about Inverse Matrices calculus?

OR

b) Describe about the Dealing with Multiple Random Variables?

10. a) Explain about the model of Artificial Neuron?

OR

b) Write about the Rules and various activation functions?

11. a) Explain about Single layer feed -forward network?

OR

b) Explain about Multi layer feed -forward network?

12. a) What is Architecture Back Propagation Network?

OR

b) Explain about variation standard Back Propagation and Network?

13. a) Explain about Deep Neural Network using Back Propagation Network?

OR

b) Explain about Gradient Descent Strategies?

BLUE PRINT

BCA Data Science

Semester – 7

INTRODUCTION TO DEEP LEARNING

[Major / Minor]

Unit	S.A.Q Section :A (including choice) 4 Marks	E.Q Section: B (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	1	2	24
III	2	2	28
IV	1	2	24
V	2	2	28
Total	8	10	132

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA (Data Science)

Semester - VII

AI CONCEPTS AND TECHNIQUES WITH PYTHON

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION-A

(Short Answer Type Questions) Marks: 5 x 4 M= 20M

Answer any FIVE questions out of the following EIGHT questions

1. What is Artificial Intelligence?

2. What are the characteristics of problems in AI?

3. What is constraint satisfaction?

4. What is knowledge?

5. What is Predicate Logic?

6. What is Logic Programming?

7. What is Bayes Theorem?

8. What is a Rule-Based System?

SECTION-B

(Long Answer Type Questions) Marks: 5x10M = 50M

Answer any FIVE out of the following TEN questions

9. a) Explain about AI Techniques?

OR

b) What is Problem, Problem Spaces and Search in AI?

10. a) Explain about Hill Climbing in AI?

OR

b) Explain about Best-First Search?

11. a) Explain knowledge Representation Issues?

OR

b) What is Resolution in AI?

12. a) Explain about Procedural Vs Declarative Knowledge?

OR

b) Explain about Forward Vs Backward reasoning?

13. a) Explain about BFS and DFS?

OR

b) Explain Bayes Probabilistic interferences and Dempster-Shafer Theory?

BLUE PRINT

BCA Data Science

Semester – 7

AI CONCEPTS AND TECHNIQUES WITH PYTHON

[Major / Minor]

Unit	S.A.Q Section :A (including choice) 4 Marks	E.Q Section: B (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	1	2	24
III	2	2	28
IV	1	2	24
V	2	2	28
Total	8	10	132

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA (Data Science)

Semester - VII

DATA & INFORMATION SECURITY

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION-A

(Short Answer Type Questions) Marks: 5 x 4 M= 20M

Answer any FIVE questions out of the following EIGHT questions

1. What is a Security?
2. What is Orange Book?
3. What is a Worm?
4. What is Public key and Private key?
5. What is Diffie-Hellman key exchange?
6. What are the benefits of digital signatures?
7. What is Intrusion detection?
8. What is a Tripwire?

SECTION-B

(Long Answer Type Questions) Marks: 5x10M = 50M

Answer any FIVE out of the following TEN questions

9. a) Explain about Protection versus Security?

OR

b) Explain about Security problems?

10. a) What is a virus? Explain different types of viruses?

OR

b) What is Tapping and Piracy in Security threats?

11. a) Explain about cryptography?

OR

b) Explain about RSA algorithm?

12. a) Explain about Digital Signatures?

OR

b) Explain about Public Key Infrastructures?

13. a) Explain about security mechanism?

OR

b) Explain about System – call monitoring?

BLUE PRINT

BCA Data Science

Semester – 7

DATA & INFORMATION SECURITY

[Major / Minor]

Unit	S.A.Q Section :A (including choice) 4 Marks	E.Q Section: B (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	1	2	24
III	2	2	28
IV	1	2	24
V	2	2	28
Total	8	10	132

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA Data Science

Semester - VIII

INRRODUCTION TO NEURAL NETWORKS

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024

ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION–A

(Short Answer Type Questions) Marks: 5 x 4M = 20M

Answer any five out of the following eight questions

1. What is a Neural Network and how does it function?
2. Define Back Propagation and its importance in training neural networks.
3. What is the difference between Crisp and Fuzzy Sets?
4. What are the applications of Fuzzy Rule-Based Systems?
5. Define Hybrid System in AI.
6. What are Genetic Algorithms used for in AI systems?
7. What is Associative Memory and its significance?
8. Explain Exponential BAM briefly.

SECTION–B

(Long Answer Type Questions) Marks: 5 x 10M = 50M

Answer any five out of the following ten questions

9. a) Explain Neural Networks, Fuzzy Logic, and Genetic Algorithms in the context of

Artificial Intelligence Systems.

OR

b) Describe the model of an Artificial Neuron along with learning rules and activation functions.

10. a) Explain the architecture and functioning of Single Layer and Multilayer Feed forward Neural Networks.

OR

b) What is a Back Propagation Network? Explain its architecture and variations of the algorithm.

11. a) Differentiate between Crisp Sets and Fuzzy Sets with suitable examples.

OR

b) Explain Fuzzy Logic and its implementation in a Fuzzy Rule-Based System.

12. a) Describe how Neural Networks, Fuzzy Logic, and Genetic Algorithms can be integrated into a Hybrid AI System.

OR

b) What are the advantages and use cases of Hybrid Systems in Artificial Intelligence?

13. a) What is Associative Memory? Explain Wang et. all's Multiple Training Encoding Strategy in detail.

OR

b) Write short notes on Autocorrelations, Heterocorrelations, and Exponential BAM.

BLUE PRINT

BCA Data Science

Semester - VIII

INTRODUCTION TO NEURAL NETWORKS

Unit	S.A.Q **Section :A** (including choice) 4 Marks	E.Q **Section: B** (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	1	2	24
III	2	2	28
IV	1	2	24
V	2	2	28
Total	8	10	132

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA Data Science

Semester - VIII

NATURAL LANGUAGE PROCESSING

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024

ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION–A

(Short Answer Type Questions) Marks: 5 x 4M = 20M

Answer any five out of the following eight questions

1. Define N-gram Language Models with an example.
2. What is the role of Machine Learning in NLP?
3. What are Regular Expressions and how are they used in NLP?
4. What is POS tagging?
5. Define Lexical Semantics.
6. What is Word Sense Disambiguation?
7. List the major challenges in Machine Translation.
8. What is WordNet?

SECTION–B

(Long Answer Type Questions) Marks: 5 x 10M = 50M

Answer any five out of the following ten questions

9. a) Explain the applications and tasks in syntax, semantics, and pragmatics in NLP.

OR

b) Describe Collocations, Information Theory, and their role in NLP.

10. a) Explain Morphological Parsing and Spelling Error Detection.

OR

b) Describe Context-Free Grammar and Probabilistic Parsing.

11. a) Discuss Ambiguity and Word Sense Disambiguation in detail.

OR

b) Explain Discourse Coherence and Reference Resolution.

12. a) Explain the architecture and tasks of Natural Language Generation.

OR

b) Describe the various approaches in Machine Translation, especially for Indian languages.

13. a) Explain different models of Information Retrieval systems.

OR

b) Describe major Lexical Resources used in NLP like Word Net, Frame Net, and POS Taggers.

BLUE PRINT

BCA Data Science

Semester - VIII

NATURAL LANGUAGE PROCESSING

Unit	S.A.Q **Section :A** (including choice) 4 Marks	E.Q **Section: B** (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	2	2	28
III	2	2	28
IV	1	2	24
V	1	2	24
Total	8	10	132

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA Data Science

Semester - VIII

RESEARCH EXPLORATION

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024

ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION–A

(Short Answer Type Questions) Marks: 5 x 4M = 20M

Answer any five out of the following eight questions

1. Define Empiricism and its role in research.
2. What are the characteristics of a scientific method?
3. Define Null and Alternative Hypothesis.
4. What are the types of Descriptive Research Design?
5. What is the difference between Independent and Dependent Variables?
6. What are the problems in measurement in research?
7. Define Sampling Frame and Sampling Error.
8. What are the characteristics of a good sample?

SECTION–B

(Long Answer Type Questions) Marks: 5 x 10M = 50M

Answer any five out of the following ten questions

9. a) Explain the foundations of research and its objectives, motivation, and utility.

OR

b) Describe the concepts of theory, deductive and inductive theory, and variables.

10. a) What is Hypothesis Testing? Explain its logic and importance.

OR

b) What are the qualities of a good hypothesis? Differentiate between Null and Alternative Hypothesis.

11. a) Explain the importance and features of a good research design.

OR

b) Describe Exploratory Research Design and Descriptive Research Design with examples.

12. a) Discuss the differences between Qualitative and Quantitative research.

OR

b) Explain the concept of measurement and its levels: Nominal, Ordinal, Interval, and Ratio.

13. a) Define and differentiate types of probability sampling methods.

OR

b) Explain practical considerations in determining sample size in research.

BLUE PRINT

BCA Data Science

Semester - VIII

RESEARCH EXPLORATION

Unit	S.A.Q **Section :A** (including choice) 4 Marks	E.Q **Section: B** (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	1	2	24
III	2	2	28
IV	1	2	24
V	2	2	28
Total	8	10	132

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA Data Science

Semester - VIII

COMPUTATIONAL DATA SCIENCE

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024

ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION–A

(Short Answer Type Questions) Marks: 5 x 4M = 20M

Answer any five out of the following eight questions

1. What are the key computational tools used in Data Science?
2. Explain the concept of Empirical Distribution.
3. Define A/B Testing and its importance.
4. What is the Central Limit Theorem?
5. Explain the properties of the Mean.
6. What is the Method of Least Squares?
7. Define Prediction Interval in regression analysis.
8. What are Visual Diagnostics in data science?

SECTION–B

(Long Answer Type Questions) Marks: 5 x 10M = 50M

Answer any five out of the following ten questions

9. a) Discuss the role of statistical techniques and simulation in computational data science.

OR

b) Explain Conditional Statements and Iteration with examples in data science.

10. a) Explain Sampling and Hypothesis Testing in the context of empirical distributions.

OR

b) Describe how error probabilities affect decisions under uncertainty.

11. a) Describe the significance of Mean and Standard Deviation in comparing two samples.

OR

b) Explain the concept of causality and its application in A/B testing.

12. a) Elaborate on the Central Limit Theorem and its implications on choosing sample size.

OR

b) Discuss correlation and the Regression Line with the Method of Least Squares.

13. a) Explain Inference for Regression including the true slope estimation.

OR

b) Discuss Visual and Numerical Diagnostics in regression models.

BLUE PRINT

BCA Data Science

Semester - VIII

COMPUTATIONAL DATA SCIENCE

Unit	S.A.Q **Section :A** (including choice) 4 Marks	E.Q **Section: B** (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	2	2	28
III	1	2	24
IV	2	2	28
V	1	2	24
Total	8	10	132

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA Data Science

Semester - VIII

COMPUTER VISION WITH PYTHON

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024

ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION–A

(Short Answer Type Questions) Marks: 5 x 4M = 20M

Answer any five out of the following eight questions

1. What are the basic components of a vision system?
2. Define Sampling and Quantization in image processing.
3. What is a Feature Descriptor in image processing?
4. Explain Hough Transform for line detection.
5. What is Region Growing in image segmentation?
6. Define Shadow Detection and its significance.
7. What is the Viola Jones Object Detection Framework?
8. List any two real-time applications of Computer Vision.

SECTION–B

(Long Answer Type Questions) Marks: 5 x 10M = 50M

Answer any five out of the following ten questions

9. a) Explain image formation and camera calibration in vision systems.

OR

b) Discuss radiometry and the interaction of light with surfaces including shading and shadows.

10. a) Explain different types of feature detectors, descriptors, and feature matching techniques.

OR

b) what is Hough Transform? Describe its use in detecting lines and vanishing points.

11. a) Explain image segmentation techniques: Thresholding, Region Growing, Splitting and Merging.

OR

b) Describe how OpenCV is used for image blending, smoothing, and reshaping with examples.

12. a) Discuss various object classification techniques based on shape and motion.

OR

b) Explain object detection using CNN and RCNN frameworks.

13. a) Describe automatic face recognition and key frame extraction in video segmentation.

OR

b) Explain real-time hand poses recognition and its applications.

BLUE PRINT

BCA Data Science

Semester - VIII

COMPUTER VISION WITH PYTHON

Unit	S.A.Q **Section :A** (including choice) 4 Marks	E.Q **Section: B** (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	2	2	28
III	2	2	28
IV	1	2	24
V	1	2	24
Total	8	10	132

KRISHNA UNIVERSITY: MACHILIPATNAM - 521004

BCA (Data Science)

Semester - VIII

DATA VISUALIZATION USING JAVASCRIPT

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

Suggested Question Paper Pattern

Time: 3 hrs Max marks: 70

SECTION-A

(Short Answer Type Questions) Marks: 5 x 4 M= 20M

Answer any FIVE questions out of the following EIGHT questions

1. Explain any two benefits of using JavaScript for data visualization.

2 Write a JavaScript function to fetch data from a public API and log the response.

3. List and describe any four types of charts supported by Chart.js.

4. Write a sample code snippet to create a bar chart using Chart.js.

5. What is data binding in D3.js? Explain with a simple example.

6. Differentiate between SVG and Canvas in the context of D3.js.

7. What is the enter-update-exit pattern in D3.js? Why is it important?

8. How can D3.js be used to display geographical data? Name one external data format and one visualization type.

SECTION-B

(Long Answer Type Questions) Marks: 5x10M= 50M

Answer any **FIVE** questions out of the following **TEN** questions

9 a) Discuss the importance of data visualization in web applications. Explain how

JavaScript plays a crucial role in building interactive data visualizations..

OR

b) Write a JavaScript program that fetches data from a JSON API and displays it on a web page. Explain each step in your code.

10. a) Create a full web page that uses Chart.js to display a line chart representing monthly temperature data. Include customization for labels, colors, and tooltips.

OR

b) Explain the structure of a Chart.js configuration object. Describe how different components like datasets, labels, and options are used to control the appearance and behavior of charts..

11. a) Explain the process of creating a bar chart in D3.js using dynamic data. Write a sample script and describe each part of the code.

OR

b) Describe how D3.js uses selections and data joins to manipulate the DOM. Explain with an example how enter(), update(), and exit() selections work.

12. a) Explain the concept of transitions in D3.js. Write a code snippet that animates changes in a bar chart when new data is loaded.

OR

b) Design an interactive dashboard component using D3.js that updates charts based on user input (e.g., dropdown selection or checkbox). Explain the logic and interactivity.

13. a) Describe the steps to create a choropleth map using D3.js and GeoJSON data.

and Discuss how color scales are used to represent different data ranges.

OR

b) Outline the complete process of building and deploying a data visualization project. Include steps such as data collection, design, development, testing, and hosting.

BLUE PRINT

BCA Data Science

Semester - VIII

DATA VISUALIZATION USING JAVASCRIPT

Unit	S.A.Q **Section :A** (including choice) 4 Marks	E.Q **Section: B** (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	1	2	24
III	2	2	28
IV	1	2	24
V	2	2	28
Total	**8**	**10**	**132**