ARTIFICIAL INTELLIGNECE & DATA SCIENCE

New Course Introduced under CBCS W.E.F. 2021-22
SCHEME OF INSTRUCTIONS AND EXAMINATIONS

IV - SEMESTER

S.No	Paper Code	Subject	Hours/ Week	No of Credits	Max.Marks Internal assessment	Max. Marks University Exam	Total Marks
1.	C10	Introduction To Big Data	4	3	25	75	100
2.	C11	Data warehousing & Data Mining	4	3	25	75	100
3	C11-P	DM & DW Lab	2	2	-0-	50	50
4.	C12	Data Science	4	3	25	75	100
5.	C13	Artificial Intelligence	4	3	25	75	100
6.	C14	Cloud Computing	4	3	25	75	100
7.	C14-P	Cloud Computing Lab	2	2	-0-	50	50
8.	C15	Data Visualization	4	3	25	75	100
Total			28	22	150	550	700

C 10- INTRODUCTION TO BIG DATA

COURSE OBJECTIVES
- The course gives an overview of the Big Data phenomenon, focusing then on extracting value from the Big Data using predictive analytics techniques
- Students will gain knowledge on analyzing Big Data. It serves as an introductory course for graduate students who are expecting to face Big Data storage, processing, analysis, visualization, and application issues on both workplaces and research environments.
- Gain knowledge on this fast-changing technological direction. Big Data Analytics is probably the fastest evolving issue in the IT world now.
- Get insight on what tools, algorithms, and platforms to use on which types of real world use cases.

COURSE OUTCOMES

Upon completion of this course, the students will be able to
- Outline the importance of Big Data Analytics
- Apply statistical techniques for Big data Analytics.
- Analyze problems appropriate to mining data streams.
- Apply the knowledge of clustering techniques in data mining.
- Use Graph Analytics for Big Data and provide solutions

UNIT1: INTRODUCTION TO BIG DATA

Evolution of Big data - Best Practices for Big data Analytics - Big data characteristics - Validating - The Promotion of the Value of Big Data - Big Data Use Cases-Characteristics of Big Data Applications - Perception and Quantification of Value - Understanding Big Data Storage - Evolution Of Analytic Scalability - Analytic Processes and Tools - Analysis vs Reporting - Modern Data Analytic Tools - Statistical Concepts: Sampling Distributions - Re-Sampling - Statistical Inference - Prediction Error.

UNIT2: DATA ANALYSIS, CLUSTERING AND CLASSIFICATION

Regression Modeling - Multivariate Analysis - Bayesian Modeling - Support Vector and Kernel Methods - Analysis of Time Series: Linear Systems Analysis - Nonlinear Dynamics - Rule Induction. Overview of Clustering - K-means - Use Cases - Overview of the Method - Determining the Number of Clusters - Diagnostics - Reasons to Choose and Cautions .- Classification: Decision Trees - Overview of a Decision Tree - The General Algorithm - Decision Tree Algorithms - Evaluating a Decision Tree - Decision Trees in R - Naïve Bayes - Bayes' Theorem - Naïve Bayes Classifier.

UNIT3: STREAM MEMORY

Introduction To Streams Concepts – Stream Data Model and Architecture - Stream Computing - Sampling Data in a Stream – Filtering Streams – Counting Distinct Elements in a Stream – Estimating Moments – Counting Oneness in a Window – Decaying Window - Real time Analytics Platform(RTAP) Applications - Case Studies

- Real Time Sentiment Analysis, Stock Market Predictions.

UNIT4: ASSOCIATION AND GRAPH MEMORY

Advanced Analytical Theory and Methods: Association Rules - Overview - Apriori Algorithm - Evaluation of Candidate Rules - Applications of Association Rules - Finding Association& finding similarity - Graph Analytics for Big Data:

UNIT5: GRAPH ANALYTICS

The Graph Model - Representation as Triples - Graphs and Network Organization - Choosing Graph Analytics - Graph Analytics Use Cases - Graph Analytics Algorithms and Solution Approaches - Technical Complexity of Analyzing Graphs- Features of a Graph Analytics Platform.

TEXT BOOKS

1.David Loshin, "Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph", 2013. 2. AnandRajaraman and Jeffrey David Ullman, "Mining of Massive Datasets", Cambridge University Press, 2012 3. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007.

REFERENCE BOOKS

1. EMC Education Services, "Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data", Wiley publishers, 2015.
2. Bart Baesens, "Analytics in a Big Data World: The Essential Guide to Data Science and its Applications", Wiley Publishers, 2015.
3. Kim H. Pries and Robert Dunnigan, "Big Data Analytics: A Practical Guide for Managers " CRC Press, 2015

C11: DATA WAREHOUSING AND DATA MINING

OBJECTIVES:
- Study data warehouse principles and its working
- Learn Data mining concepts and understand Association Rule Mining
- Study Classification Algorithms 4. Gain knowledge of how data is grouped using clustering techniques.

OUTCOMES:
- Comparison of functional differences between data warehouse and database systems.
- Ability to perform the pre-processing of data and apply mining techniques on it.
- Capability to identify the association rules, classification and clusters in large data sets.
- Skills to solve real world problems in business and scientific information using data mining.

UNIT-I Data warehouse:

Introduction to Data warehouse, Difference between operational database systems and data warehouses, Data warehouse Characteristics, Data warehouse Architecture and its Components, Extraction-Transformation-Loading, Logical(Multi-Dimensional), Data Modeling, Schema Design, Star and Snow-Flake Schema, Fact Constellation, Fact Table, Fully Addictive, Semi-Addictive, Non Addictive Measures; Fact-Less-Facts, Dimension Table Characteristics; OLAP Cube, OLAP Operations, OLAP Server Architecture-ROLAP, MOLAP and HOLAP.

UNIT-II Introduction:

Fundamentals of data mining, Data Mining Functionalities, Classification of Data Mining systems, Data Mining Task Primitives, Integration of a Data Mining System with a Database or Data Warehouse System, Major issues in Data Mining. Data Preprocessing: Need for Preprocessing the Data, Data Cleaning, Data Integration &Transformation, Data Reduction, Discretization and Concept Hierarchy Generation.

UNIT-III Association Rules:

Problem Definition, Frequent Item Set Generation, The APRIORI Principle, Support and Confidence Measures, Association Rule Generation; APRIOIRI Algorithm, The Partition Algorithms, FP-Growth Algorithms, Compact Representation of Frequent Item Set- Maximal Frequent Item Set, Closed Frequent Item Set.

UNIT-IV Classification:

Problem Definition, General Approaches to solving a classification problem, Evaluation of Classifiers, Classification techniques, Decision Trees-Decision tree Construction, Methods for Expressing attribute test conditions, Measures for Selecting the Best Split, Algorithm for Decision tree Induction; Naive-Bayes Classifier, Bayesian Belief Networks; K- Nearest neighbor classification-Algorithm and Characteristics. Prediction: Accuracy and Error measures, Evaluating the accuracy of classifier or a

predictor, Ensemble methods

UNIT-V Clustering:

Clustering Overview, A Categorization of Major Clustering Methods, Partitioning Methods, Hierarchical Methods, Partitioning Clustering-K-Means Algorithm, PAM Algorithm; Hierarchical Clustering-Agglomerative Methods and divisive methods, Basic Agglomerative Hierarchical Clustering Algorithm, Key Issues in Hierarchical Clustering, Strengths and Weakness, Outlier Detection.

TEXT BOOKS:

1) Data Mining- Concepts and -1.chniques- Jiawei Han, Micheline Kamber, Morgan Kaufmann Publishers, Elsevier, 2 Edition, 2006.
2) Introduction to Data Mining, Psng-Ning Tan, Vipin Kumar, Michael Steinbanch, Pearson Educatior.

REFERENCE BOOKS:

1) Data Mining Techniques, Arun KPujari, 3rd Edition, Universities Press.
2) Data Warehousing Fundament's, Pualraj Ponnaiah, Wiley Student Edition.
3) The Data Warehouse Life CycleToolkit — Ralph Kimball, Wiley Student Edition.
4) Data Mining, Vikaram Pudi, P Rddha Krishna, Oxford University Press

DM & DW USING R

1. Get and Clean data using swirl exercises.(Use 'swirl' package, library andinstall that topic from swirl).

2. Visualize all Statistical measures(Mean ,Mode, Median, Range, Inter QuartileRange etc., using Histograms, Boxplots and Scatter Plots).

3. Create a data frame with the following structure.

EMP ID	EMP NAME	SALARY	START DATE
1	Satish	5000	01-11-2013
2	Vani	7500	05-06-2011
3	Ramesh	10000	21-09-1999
4	Praveen	9500	13-09-2005
5	Pallavi	4500	23-10-2000

a. Extract two column names using column name.

b. Extract the first two rows and then all columns.

c. Extract 3rd and 5th row with 2nd and 4th column.

4. Create a data frame with 10 observations and 3 variables and add new rows and columns to it using 'rbind' and 'cbind' function.

5. Create a function to discretize a numeric variable into 3 quantiles and label them as low, medium, and high. Apply it on each attribute of any dataset to create a new data frame. 'discrete' with Categorical variables and the class label.

6. Create a simple scatter plot using any dataset using 'dplyr' library. Use the same data to indicate distribution densities using boxwhiskers.

7. Write R Programs to implement k-means clustering, k-medoids clustering and density based clustering on any datasets.

8. Write a R Program to implement decision trees using 'readingSkills' dataset.

9. Implement decision trees using any dataset using package party and 'rpart'.

10. Train SVM Model by taking any dataset.

C 12: INTRODUCTION TO DATASCIENCE

OBJECTIVES:

The objective of the course is to help students to:

Overview of the interdisciplinary concepts of Data Science in the real time

- applications. The importance of maths and statistics in data science.
- Role of machine learning techniques in data science.
- Integrated role of computers and its components.
- Flow and process model of data science project management.

OUTCOMES:

After completing this course, students will be able to Distinguish between data Science terminologies and business analytics.

- Develop mathematical and statistical models for Data science applications.
- Apply appropriate machine learning technique for the data science application.
- Select appropriate software tools
- Install and configure the software tool and the data science application project.

UNIT I DATASCIENCE -AN OVERVIEW

Introduction to Data Science, Definition and description of Data Science, history and development of Data Science, terminologies related with Data Science, basic framework and architecture, difference between Data Science and business analytics, importance of Data Science in today's business world, primary components of Data Science, users of Data Science and its hierarchy, overview of different Data Science techniques, challenges and opportunities in business analytics, different industrial application of Data Science technique

UNIT II MATHEMATICS ANDSTATISTICSIN DATASCIENCE

Role of mathematics in Data Science, importance of probability and statistics in Data Science, important types of statistical measures in Data Science Descriptive, Predictive and prescriptive statistics, introduction to statistical inference and its usage in Data Science, application of statistical techniques in Data Science overview of linear algebra: matrix and vector theory, role of linear algebra in Data Science, exploratory data analysis and visualization techniques, difference between exploratory and descriptive statistics, EDA and visualization as key component of Data Science.

UNIT III MACHINELEARNING INDATASCIENCE

Role of machine learning in Data Science, different types of machine learning techniques and its broad scope in Data Science: Supervised, unsupervised, reinforcement and deep learning, difference between different machine learning techniques, brief introduction to machine learning algorithms, importance of machine learning in today's business, difference between machine learning classification and prediction.

UNIT IV COMPUTERSINDATASCIENCE

Role of computer science in Data Science, various components of computer science being used for Data Science, role of relation data base systems in Data Science: SQL, No SQL, role of data warehousing in Data Science, terms related with data warehousing techniques, importance of operating concepts and memory management, various freely available software tools used in Data Science

UNIT V SOFTWARETOOLSANDAPPLICATIONS

R, Python, important proprietary and open source software tools, different business intelligence tools and its crucial role in Data Science project presentation

TEXT BOOKS:

1. Data Smart: Using Data Science to Transform Information into Insight 1st Edition by John W. Foreman. (2015) Wiley Publication.
2. Data Science from Scratch: First Principles with Python 1st Edition by Joel Grus.

REFERENCES:

1. Data Science For Dummies by Lillian Pierson(2015)
2. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking by Foster Provost, TomFawcett.

C 13: ARTIFICIAL INTELLIGENCE

COURSE OBJECTIVES:

This course is designed to:

- Define Artificial Intelligence and establish the cultural background for study
- Understand various learning algorithms
- Explore the searching and optimization techniques for problem solving
- Provide basic knowledge on Natural Language Processing and Robotics

COURSE OUTCOMES:

Upon completion of the course, the students should be able to:

- Apply searching techniques for solving a problem (L3)
- Design Intelligent Agents (L6)
- Develop Natural Language Interface for Machines (L6)
- Design mini robots (L6)
- Summarize past, present and future of Artificial Intelligence (L5)

Unit – I

Introduction: What is AI, Foundations of AI, History of AI, The State of Art.

Intelligent Agents: Agents and Environments, Good Behavior: The Concept of Rationality, The Nature of Environments, The Structure of Agents.

Unit – II

Solving Problems by searching: Problem Solving Agents, Example problems: Toy problem and Real world problem, Search Strategies: Breadth first search, Uniform cost Search, Depth First search, Depth Limited Search.

Informed search strategies: Best first search, Heuristic Functions.

Unit – III

Reinforcement Learning: Introduction, Passive Reinforcement Learning-Naïve updating, Adaptive Dynamic Programming, Temporal difference learning. Passive learning in an unknown environment. Active Reinforcement Learning, Generalization in Reinforcement Learning-Application to game playing, Application to robot control. Applications of Reinforcement Learning, policy search.

Unit-IV

Natural Language Processing: Language Models, Text Classification, Information Retrieval, Information Extraction.

Natural Language for Communication: Phrase structure grammars, Syntactic Analysis, Augmented Grammars and semantic Interpretation, Machine Translation,

Speech Recognition

Unit-V

Perception: Image Formation-Pihole camera, lens system, photometry and spectrophotometry of image formation. Early Image Processing Operations, Extracting 3D information using vision, vision for Manipulation and Navigation. Object Representation and Recognition, speech recognition.

Robotics: Introduction, Robot Hardware, Robotic Perception.

TEXTBOOK:

1. Stuart J.Russell, Peter Norvig, "Artificial Intelligence A Modern Approach", 3rd Edition, Pearson Education, 2019.

REFERENCES:

1. Nilsson, Nils J., and Nils Johan Nilsson. Artificial intelligence: a new synthesis. Morgan Kaufmann, 1998.

2. Johnson, Benny G., Fred Phillips, and Linda G. Chase. "An intelligent tutoring system for the accounting cycle: Enhancing textbook homework with artificial intelligence." Journal of Accounting Education 27.1 (2009): 30-39.

ARTIFICIAL INTELLIGENCE LAB

1. Write a program to implement DFS

2. Write a program to implement BFS.

3. Write a program to implement BST.

4. Write a program to implement to Towers of Hanoi problem.

5. Write a Program to find the solution for travelling salesman Problem

6. Write a program to implement 8 puzzle problem.

7. Write a program to implement A* Algorithm

8. Write a program to implement Hill Climbing Algorithm

9. Write a program to implement Simulated Annealing Algorithm

10. Build a Chatbot using AWS Lex, Pandora bots.

11. Build a Chatbot using Whatsapp.

12. Build a bot which provides all the information related to your college.

C14: CLOUD COMPUTING

OBJECTIVES:

- Cloud computing is a colloquial expression used to describe a variety of different computing concepts that involve a large number of computers involves a large number of computers that are connected through a real-time communication network.
- In science, cloud computing is a synonym for distributed computing over a network and means the ability to run a program on many connected computers at the same time.
- This course covers basic concepts of cloud types, services and security etc.

OUTCOMES: At the end of the course, the students will be able to:

- Learn the underlying principles of Cloud Technology and various types of cloud computing architecture and types.
- Evaluate between different cloud solutions offered by various providers based on their merits and demerits.
- Understand the Cloud Cost Management and Selection of Cloud Provider.
- Understand the IT governance in cloud computing.
- Track the Ten cloud do an do nots.:

UNIT I: Introduction

Introduction to Cloud Computing, History and Evolution of Cloud Computing, Types of clouds, Private Public and hybrid clouds, Cloud Computing architecture, Cloud computing infrastructure, Merits of Cloud computing, , Cloud computing delivery models and services (IaaS, PaaS, SaaS), obstacles for cloud technology, Cloud vulnerabilities, Cloud challenges, Practical applications of cloud computing.

UNIT II: Cloud Computing Companies and Migrating to Cloud

Web-based business services, Delivering Business Processes from the Cloud: Business process examples, Broad Approaches to Migrating into the Cloud, The Seven-Step Model of Migration into a Cloud, Efficient Steps for migrating to cloud., Risks: Measuring and assessment of risks, Company concerns Risk Mitigation methodology for Cloud computing, Case Studies

UNIT III: Cloud Cost Management and Selection of Cloud Provider

Assessing the Cloud: software Evaluation, System Testing, Seasonal or peak loading,

Cost cutting and cost-benefit analysis, selecting the right scalable application. Considerations for selecting cloud solution. Understanding Best Practices used in selection of Cloud service and providers, Clouding the Standards and Best Practices Issue: Interoperability, Portability, Integration, Security, Standards Organizations and Groups associated with Cloud Computing, Commercial and Business Consideration

UNIT IV: Governance in the Cloud

Industry Standards Organizations and Groups associated with Cloud Computing, Need for IT governance in cloud computing, Cloud Governance Solution: Access Controls, Financial Controls, Key Management and Encryption, Logging and Auditing, API integration. Legal Issues: Data Privacy and Security Issues, Cloud Contracting models, Jurisdictional Issues Raised by Virtualization and Data Location, Legal issues in Commercial and Business Considerations

UNIT V: Ten cloud do an do nots

Don't be reactive, do consider the cloud a financial issue, don't go alone, do think about your architecture, don't neglect governance, don't forget about business purpose, do make security the centerpiece of your strategy, don't apply the cloud to everything, don't forget about Service Management, do start with a pilot project.

TEXT BOOKS:

1. Cloud Computing: Principles and Paradigms, Rajkumar Buyya, James Broberg, Andrzej M. Goscinski,, John Wiley and Sons Publications, 2011

REFERENCES:

1. Brief Guide to Cloud Computing, Christopher Barnett, Constable & Robinson Limited, 2010
2. Handbook on Cloud Computing, Borivoje Furht, Armando Escalante, Springer, 2010

LIST OF PROGRAMS:

Study the basic cloud architecture and represent it using a case study

1. Enlist Major difference between SAAS PAAS & Iaas also submit a research done on various companies in cloud business and the corresponding services provided by them , tag them under SAAS , Paas & Iaas.
2. Study and present a report on Jolly cloud.
3. Present a report on obstacles and vulnerabilities in cloud computing on generic level
4. Present a report on Amazon cloud services.

5. Present a report on Microsoft cloud services.

6. Present a report on cost management on cloud

7. Enlist and explain legal issues involved in the cloud with the lelp of a case study

8. Explain the process of migrating to cloud with a case study.

9. Present a report on google cloud and cloud services

10. Create a scenario based on real time domain

.

C 15 :DATA VISUALIZATION

OBJECTIVES:

The Objective of the course is to assist the student in understanding the visualizing the data in a real-world environment with many different functionalities. Visualization tool like Tableau will help the student to visually represent data.

COURSE OUTCOMES

1. Ability to collect and arrange data in a specific manner using visualization tool

2. Ability to analyze data visually with the concepts like filter, sort etc.

3. Able to differentiate different comparison methodologies

4. Able to understand various types of charts in Tableau

5. Ability to generate reports with functions, joins, unions etc.

UNIT-1:

Data Visualization: Definition, importance of data visualization in data science, various visualization tools, History of Visualization, developing a visualization aesthetic – maximizing data link ratio, proper scaling and labeling, effective use of color and shading, the power of repetition.

UNIT-2:

Basics of Tableau:

Introduction, Advantages of Tableau, Tableau products, file types in Tableau: Tableau Workbook, Tableau Packaged Workbook, Tableau data source, Tableau Packaged data source, Tableau Data Extract, Installing Tableau software, Tableau Creator, Tableau prep, Tableau Explorer, Tableau Server.

UNIT-3:

Tableau Desktop UI: The Tableau product line, Desktop Architecture, Tableau environment, Data source Page, Workspace, workbooks and sheets, Visual cues and Icons in Tableau, fields in data pane.

Connect to files: Connect to Excel or Text file, Connect to MS-Access, Connect to RData files

Connect to Server: Connect to Oracle, Connect to SQL Server, Connect to MySQL

UNIT-4:

Tableau Charts:

Understanding dimensions and measures, Chart Types: Area Chart, bar Chart, Bubble chart, Funnel Chart, Gantt chart, Waterfall chart, Geographic map, Sorting: Sorting on

Axis, Sorting on specific fields, Filtering Data,

UNIT-5:

Tableau Reports:

Joins: Inner, Left and Right joins, Unions, Functions used in Tableau: Number functions, logic functions, string functions, aggregate functions.

Text Books:

1. Pro Tableau – A Step-by-Step guide – Seema Acharya, Subhashini Chellappan – Apress

2. Learning Tableau 2019 Tools for Business Intelligence, data prep, and visual analytics – Joshua N. Milligan - Packt

3. Tableau for Dummies – Molly Monsey and Paul Sochan - John Wiley & Sons, Inc.,

TABLEAU FOR DATA VISUALIZATION LAB

1. Demonstration of installation of Tableau desktop software

2. Demonstration of start page of Tableau environment.

3. Bringing data from a text file by connecting from Tableau

4. Bringing data from a Excel file by connecting from Tableau

5. Bringing data from a Oracle Database by connecting from Tableau

6. Bringing data from a R-Data file by connecting from Tableau

7. Applying filters on data

8. Joining the data from different data sources

9. Performing union operations on the data source

10. Creating and changing the measures and dimensions of various charts

SRI VENKATESWARA UNIVERSITY

BCA DEGREE COURSE IN ARTIFICIAL INTELLIGNECE &DATA SCIENCE

SECOND YEAR – FOURTH SEMESTER
(Syllabus under CBCS w.e.f. 2021-22)
MODEL QUESTION PAPER (for all papers)

Time: 3 hours   Marks: 75 marks

Note: This question paper contains two parts A and B.

Part A is compulsory which carries 25 marks. Answer any five of the following questions in Part A.

Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks

PART – A

Answer any _Five_ of the following question.   **(5X5=25M)**

1	
2	
3	
4	
5	
6	
7	
8	

PART – B
Answer All The Questions. Each question carries 10 marks   **(5X10= 50M)**

9	(A)	
		OR
	(B)	
10	(A)	
		OR
	(B)	
11	(A)	
		OR
	(B)	
12	(A)	
		OR
	(B)	
13	(A)	
		OR
	(B)	