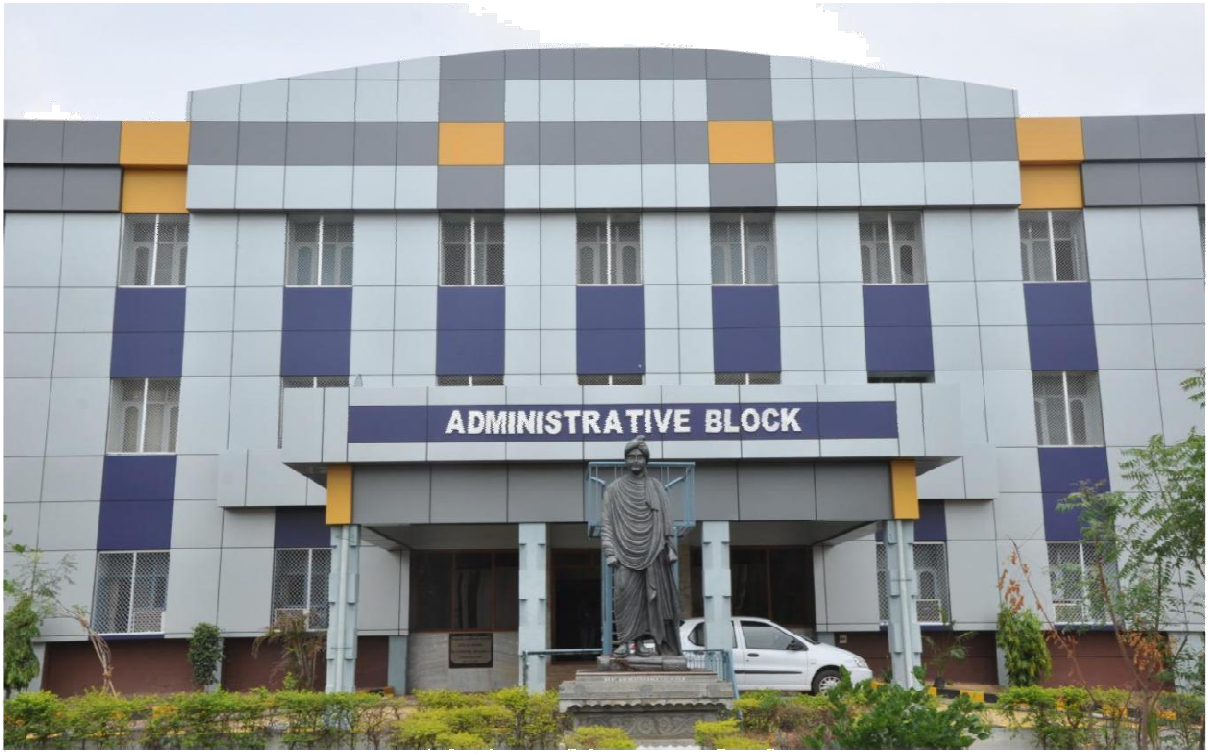

Scheme – 2020

Department of Computer Science & Engineering

G. Pulla Reddy Engineering College (Autonomous): Kurnool

Accredited by NBA of AICTE and NAAC of UGC

Affiliated to JNTUA, Anantapuramu

Scheme and Syllabus for

Minor in

DATA SCIENCE

(for Non-IT Branches)

(With Effect from the Batch Admitted in 2020-21)

Minor in Data Science

(for Non IT students)

Scheme of Instruction and Examination
(Effective from 2020-2021)

S. No	Semester	Course Title	Credits	Scheme of Instruction periods/week		Scheme of Examination Maximum Marks		
				L	P	End Exam	Internal Assessment	Total 100
1	IV	Introduction to Algorithms	4	3	2	60	40	100
2	V	Python for Data Science	4	3	2	60	40	100
3	VI	Data Mining	4	3	2	60	40	100
4	VII	Machine Learning	4	3	2	60	40	100
5	MOOCS-1		2	0	0			100
6	MOOCS-2 / Mini Project		2	0	0			100
		Total	**20**					

MOOCS-1 **:** 1. R Programming

2. Probability and Statistics

3. Artificial Intelligence

MOOCS-2 / Mini Project **:** 1. Natural Language Processing

2. Business Analytics

3. Data Engineering

INTRODUCTION TO ALGORITHMS (INA)

IV Semester : Minor in CS/DS						Scheme : 2020		
Course Code	Category	Hours/Week			Credits	Maximum Marks		
MCS01	M	L	T	P	C	Continuous Internal Assessment	End Exam	TOTAL
		3	-	2	4	40	60	100
Sessional Exam Duration : 1 ½ Hrs						End Exam Duration: 3 Hrs		

Course Outcomes : At the end of the course the student will be able to

CO1: Understand fundamental characteristics, design, analysis and complexities of algorithms

CO2: Describe the divide-and-conquer paradigm and explain when an algorithmic design situation calls for it.

CO3: Understandand develop algorithms for well-known problems using greedy methods.

CO4: Understand and apply dynamic-programming approach for designing graph and matrix based algorithms.

CO5: Understand the concept of backtracking for traversal and search algorithms.

UNIT – I

Algorithms- Problem Solving, Introduction to Algorithms, Characteristics of algorithms, Algorithm design tools: Pseudocode and flowchart, Analysis of Algorithms, Complexity of algorithms- Space complexity, Time complexity, Asymptotic notation- Big-O, Theta and Omega.

UNIT – II

Divide and Conquer: General method, Binary search, Finding Maximum and Minimum, Merge sort, Quick sort, Strassen's Matrix Multiplication

UNIT – III

Greedy method: The General Method, Knapsack Problem, Tree Vertex splitting, Job Sequencing with Deadlines, Minimum-Cost Spanning Trees, Optimal Storage on Tapes, Optimal Merge Patterns, Single Source Shortest Paths

UNIT – IV

Dynamic Programming: The General Method, Multistage Graphs, All Pairs Shortest Paths, Optimal Binary Search Trees, String Editing problem, Reliability Design, The Travelling Salesperson Problem

UNIT – V

Backtracking: The General Method, The 8-Queens Problem, Sum of Subsets, Graph Coloring and Hamiltonian cycles.

Text Books:

1. Fundamentals of Computer Algorithms, 2nd Edition, Ellis Horowitz, SartajSahni and S. Rajasekharan,Universities Press.

Reference Books:

1. Algorithm Design: Foundations, Analysis and Internet examples, M. T. Goodrich and R. Tomassia, John Wiley and sons.

2. Design and Analysis of Algorithms, S. Sridhar, Oxford Univ. Press

Question Paper Pattern:

Sessional Exam:
The question paper for sessional examination shall be for 25 marks, covering half of the syllabus for first sessional and remaining half for second sessional exam. The question paper shall consist of three sections with Two Questions (EITHER/ OR Type) in each section. The student shall answer one question from each section.

End Exam:
The question paper for End examination shall be for 60 marks. The Question paper shall contain Five Units with Two Questions (Either or Type) from each unit. Each of these questions may contain sub question and the student should answer any one question from each unit. Each Question carries 12 marks..

List of Experiments
1. Implement Binary Search algorithm using Divide and Conquer Technique.
2. Implement Merge Sort algorithm using Divide and Conquer Technique.
3. Implement Knapsack using Greedy Technique.
4. Implement Kruskal's algorithm for finding minimum cost spanning tree using GreedyTechnique.
5. Implement Tree Traversing Techniques
6. Queens problem using Backtracking technique

PYTHON FOR DATA SCIENCE (PDS)

V Semester : Minor						Scheme : 2020		
Course Code	Category	Hours/Week			Credits	Maximum Marks		
MDS01	M	L	T	P	C	Continuous Internal Assessment	End Exam	TOTAL
		3		2	4	40	60	100
Sessional Exam Duration : 1 ½ Hrs						End Exam Duration: 3 Hrs		

Course Outcomes : At the end of the course the student will be able to

CO1: Write python programs using the core concepts like Lists, Dictionaries, sets, tuple, functions and regular expressions.

CO2: Demonstrate various mathematical operations on arrays using NumPy

CO3: Analyze and manipulate Data using Pandas

CO4: Creating static and interactive visualizations using Matplotlib.

CO5: Enumerate machine learning algorithms, Describe the Classification and Clustering

UNIT – I

Introduction to Python: Data Types:Strings, Numbers, Booleans, Date and Time, Lists, Tuples, Dictionaries,Operators,Conditional Statements, Loops, Functions, Modules and packages, Classes and Objects, Regular expressions

UNIT – II

Introduction to NumPy: The Basics of NumPy Arrays, Computation on NumPy Arrays: Universal Functions, Aggregations: Min, Max, and Everything in Between, Computation on Arrays: Broadcasting, Comparisons, Masks, and Boolean Logic, Fancy Indexing, Sorting Arrays, Structured Data: NumPy's Structured Arrays

UNIT – III

Data Manipulation with Pandas: Introducing Pandas Objects, Data Indexing and Selection, Operating on Data in Pandas, Handling Missing Data, Hierarchical Indexing, Combining Datasets: Concat and Append, Combining Datasets: Merge and Join, Aggregation and Grouping Planets Data, Pivot Tables, Vectorized String Operations, High-Performance Pandas

UNIT – IV

Visualization with Matplotlib: Two Interfaces for the Price of One, Simple Line Plots, Simple Scatter Plots, Visualizing Errors, Density and Contour Plots, Histograms, Binnings, and Density, Customizing Plot Legends, Customizing Colorbars, Multiple Subplots, Text and Annotation, Customizing Ticks, Customizing Matplotlib: Configurations and Stylesheets, Three-Dimensional plotting in Matplotlib

UNIT – V

Machine Learning: What Is Machine Learning?, Types of machine learning systems, Introducing Scikit-Learn, Feature Engineering, Model development, Linear Regression: Simple Linear Regression, Example of model development.

Text Books:

1. Python Basics: With Illustrations from the Financial Market,Vivek Krishnamoorthy, Jay Parmar, Mario Pisa Pena,AQuantInsti® Publication, 2020
2. Python Data Science Handbook:Essential Tools for Working with Data,JakeVanderPlas, O'reilly publications, 2016

Reference Books:

1. Python® for Programmers,PaulDeitel,HarveyDeitel, Pearson Education, Inc,2019
2. Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data,David Dietrich, Barry Heller, Beibei Yang, Published by John Wiley & Sons, Inc, 2015

Question Paper Pattern:

Sessional Exam:

The question paper for sessional examination shall be for 25 marks, covering half of the syllabus for first sessional and remaining half for second sessional exam. The question paper shall consist of three sections with Two Questions (EITHER/ OR Type) in each section. The student shall answer one question from each section.

End Exam:

The question paper for End examination shall be for 60 marks. The Question paper shall contain Five Units with Two Questions (Either or Type) from each unit. Each of these questions may contain sub question and the student should answer any one question from each unit. Each Question carries 12 marks.

List of Experiments

1. Data Types: Operations on Number data type, sets and tuples.
2. Operators: Example programs using all operators.
3. Conditional Statements: Usage of if-else -if statement, break, continue and pass.
4. Loop Statements: Armstrong number, Palindrome.
5. Loop Statements: Bubble sort using lists.
6. Functions: Calculator program.
7. Functions: Programs on anonymous functions.
8. NumPy: Arithmetic Operations on Arrays.

DATAMINING (DMG)

VI Semester: Minor						Scheme:2020		
CourseCode	Category	Hours/Week			Credits	MaximumMarks		
MDS02	M	L	T	P	C	Continuous InternalAssessment	End Exam	TOTAL
		3		2	4	40	60	100
SessionalExamDuration: 1 ½ Hrs						End ExamDuration:3Hrs		

CourseOutcomes: At theendofthecoursestudentswill beableto

CO1: Understandthebasic conceptsof datamining.

CO2: Interpret the pre-processingtechniques to prepare the data for mining.

CO3: Understand thedataminingclassificationtechniques

CO4: Interpret the Market basket analysis data usingassociationrulemining

CO5: Understand the unsupervisedminingalgorithms.

UNIT–I

Data Mining:
Introduction,WhatisDataMining,MotivatingChallenges,DataMiningTasks,TypesofData,SimilarityandDissimilaritybetween Simple Attributes and Data Objects.

UNIT–II

DataPreprocessing:
WhyPre-processtheData?
DescriptiveDataSummarization,DataCleaning,DataIntegrationandTransformation,Data Reduction, Data Discretizationand Concept HierarchyGeneration.

UNIT–III

Classification:
BasicConcepts,GeneralApproachtosolvingaclassificationproblem,DecisionTreeInduction:Workingof Decision Tree, Building a decision tree, methods for expressing an attribute test conditions, measures forselecting the best split, Algorithm for Decision Tree Induction, Nearest-Neighbor Classifier,BayesTheorem, UsingtheBayestheorem forclassification, NaiveBayesClassifier.

UNIT–IV

AssociationAnalysis:
Basic Concepts and Algorithms: Frequent Item Set generation, The Apriori Principle, Apriori Algorithm,Candidate Generation and Pruning, Rule Generation, Confidence-Based Pruning, Rule Generationwith anexample, FP-Growth Algorithm.

UNIT–V

ClusterAnalysis:
What is ClusterAnalysis,Types ofClustering,K-MeansAlgorithm,AgglomerativeHierarchical ClusteringAlgorithm,KeyIssues inHierarchicalClustering,DBSCANAlgorithm.

TextBooks :

1.IntroductiontoData Mining,Pang-NingTan,MichaelSteinbach,VipinKumar,PEA, 2008

2.DataMiningconceptsand Techniques,3/e, Jiawei Han,Michel Kamber,Elsevier,2006

ReferenceBooks:

1.DataWarehousingData Mining&OLAP, AlexBerson,Stephen Smith,TMH.

2.DataMiningTechniques,Arun K Pujari,Universities Press.

WebReferences:

1.https://onlinecourses.nptel.ac.in/noc18_cs14/

2.https://freevideolectures.com/course/3758/databases-data-mining

QuestionPaperPattern:

SessionalExam:

The question paper for sessional examination shall be for 25 marks, covering half of the syllabus for first sessional and remaining half for second sessional exam. The question paper shall consist of three sections with Two Questions (EITHER/ OR Type) in each section. The student shall answer one question from each section.

EndExam:

The question paper for End examination shall be for 60 marks. The Question paper shall contain Five Units with Two Questions (Either or Type) from each unit. Each of these questions may contain sub question and the student should answer any one question from each unit. Each Question carries 12 marks..

List of Experiments

1. Data Summarization Techniques
2. Decision Tree Classifier
3. Naïve Bayes Classifier
4. Apriori algorithm
5. DBSCAN algorithm

MACHINELEARNING (ML)

VII Semester: Minor						Scheme:2020		
CourseCode	Category	Hours/Week			Credits	MaximumMarks		
MDS03	M	L	T	P	C	Continuous Internal Assessment	End Exam	TOTAL
		3		2	4	40	60	100
Sessional Exam Duration: 1 ½ Hrs						End Exam Duration :3Hrs		

Course Outcomes: At the end of the course students will be able to

CO1: Understand the machine learning concepts and the main steps involved in a typical machine learning project.

CO2: Implement a digit image classifier on MNIST dataset.

CO3: Implement regression models using direct closed form equation and Gradient Descent approaches.

CO4: Understand the core concepts and working of Support Vector Machines, Decision trees and CART training algorithm.

CO5: Understand ensemble methods and clustering algorithms

UNIT–I

MachineLearningLandscape

Introduction,TypesofMachineLearningSystems,Challenges,Steps involved in building an End-to-End Machine Learning Project.

UNIT–II

Classification

MNIST, Training a Binary Classifier, Performance measures, Multi class classification, Error analysis, Multi label classification, Multi output classification

UNIT–III

Regression Models

Linear Regression, Multiple Linear Regression, Simple Gradient Descent approach ,Polynomial Regression, Logistic Regression

UNIT–IV

SupportVectorMachines

Linear SVM classification, Nonlinear SVM classification, SVM Regression

DecisionTrees

Training and visualizing a decision tree, Making predictions, Estimating class probabilities, CART Training algorithm, Computational complexity, Gini Impurity or Entropy, Regularization Hyperparameters

UNIT–V

EnsembleLearningandRandomForests

Voting classifiers, Bagging and pasting , Random forests Algorithm

Unsupervised Learning Techniques

Clustering algorithms - K-Means, DB Scan

Text Books:

1.AurelianGeron,"Hands-On Machine Learning with Scikit –Learn and Tensor Flow: Concepts, Tools, and Techniques to build Intelligent Systems ", OReilly Publications, First Edition, 2017

2.Tom M.Mitchell, "Machine Learning", McGraw Hill Education, Indian Edition, 2013

Reference Books:

1.OliverTheobald,"MachineLearningforAbsoluteBeginners",Second Edition,2017

2.EthemAlpaydin," Introduction to Machine Learning", The MIT Press, Third Edition, 2014

3.MiroslavKubat,“AnIntroductiontoMachine Learning”,Springer,2017

Web References:

1. *https://www.coursera.org/learn/python-machine-learning*offeredbyUniversityof Michigan
2. *.https://scikit-learn.org/stable/*
3. *https://github.com/ageron/handson-ml.*
4. *https://www.coursera.org/learn/python-data-analysis?specialization=data-science-python*
5. *https://www.coursera.org/learn/python-plotting?specialization=data-science-python*
6. *http://learnpython.org/*

QuestionPaperPattern:

Sessional Exam:

The question paper for sessional examination shall be for 25 marks, covering half of the syllabus for first sessional and remaining half for second sessional exam. The question paper shall consist of three sections with Two Questions (EITHER/ OR Type) in each section. The student shall answer one question from each section.

End Exam:

The question paper for End examination shall be for 60 marks. The Question paper shall contain Five Units with Two Questions (Either or Type) from each unit. Each of these questions may contain sub question and the student should answer any one question from each unit. Each Question carries 12 marks

List of Experiments

1. Implement Preprocessing Techniques on a given dataset
2. Linear Regression
3. Binary Classification on MNIST dataset
4. Ensemble Learning Techniques
5. K-Means Clustering Algorithm