THIRUVALLUVAR UNIVERSITY

SERKKADU, VELLORE – 632 115

CERTIFICATE AND DIPLOMA IN DATA ANALYSIS

With effect from 2020 − 2021

<u>**Definitions:**</u>

Programme :Programme" means a course of study leading to the award of a degree in a discipline.

Course : "Course" refers to a paper / practical / subject offered under the degree programme.

Each Course is to be designed variously under lectures / tutorials / laboratory or field work / seminar / practical training / Assignments / Term paper or Report writing etc., to meet effective teaching and learning needs.

i) <u>**Core Courses:**</u>

"The Core Courses" related to the programme concerned including practicals offered in the programme".

ii) <u>**Elective Courses:**</u>
iii)

"Elective courses" related to the core courses of the programme concerned, offered in the programme".

Duration : This means the stipulated years of study to complete a programme as prescribed by the University from time to time. Currently for the Dipolma programme the duration of study is one year.

Credits : The weightage given to each course of study (subject) by the experts of the Board of Studies concerned.

The term 'Credit' refers to the weight age given to a course, usually in relation to the instructional hours assigned to it. For instance, a six hour course per week is assigned 6/5/4 credits, a five hour course per week is assigned 5/4/3 credits and a four hour course per week is given 4/3/2 credits. However, in no instance the credits of a course can be greater than the hours allotted to it.

The total minimum credits, required for completing a Diploma program is 45.

Certificate and Diploma Programme Structure

Semester	Paper Type	Credits	Hrs of Teaching	Practical Hrs	Seminar, Workshop, etc.	Exam Hrs	Total Marks Maximum		
							Internal	External	Total
I	Core – 1	4	4	2	1	3	25	75	100
	Core - 2	4	4	2	1	3	25	75	100
	Core –3	4	4	2	1	3	25	75	100
	Practical (Or)Elective	3	3	0	0	3	25	75	100
	Practical (Or)Elective	3	3	0	0	3	25	75	100
	Mini Project	5	1	0	2	3	25	75	100
		25	19	6	5		150	450	600
II	Core – 4	4	4	2	1	3	25	75	100
	Core - 5	4	4	2	1	3	25	75	100
	Core – 6	4	4	2	1	3	25	75	100
	Practical (or) Elective	3	3	0	0	3	25	75	100
	Project	5	2	0	4	3	25	75	100
		20	17	6	7		125	375	500
							275	825	1100
	Grand Total	45			30 hours			1100	1100

Note:

- On completion of Semester I, Certificate Course shall be awarded.
- On completion of Semester I and II, Diploma course shall be awarded.

Diploma   in   Data   Analysis

SEMESTER - I

Core I   EXPLORATORY ANALYSIS   OF   SPATIAL AND   TEMPORAL   DATA

No Of Credits: 4

Unit - 1

Data: Structure of Data, Properties of Data, Examples of Data. Tasks - Abstraction, Semiology of Graphics - Jacques Bertin'sand Jock D. Mackinley'sView of Tasks, Elementary Tasks, Synoptic Tasks - Behaviour and Pattern, Connection Discovery, Completeness of the Framework.

Unit - 2

Tools - Visualization in a Nutshell, Visualization Strategies - Manipulate views, Facet into Multiple views, Reduce Items and Attributes, Display Manipulation, Data Manipulation - Attribute Transformation, Attribute Integration, Value Interpolation, Data Aggregation, Querying.

Unit - 3

Principles - Components of the Exploratory Process, Examples of Exploration, General Principles of Selection of the Methods and Tools, General Scheme of Data Exploration

Unit – 4

Tasks, Principles, and Tools - Case study: Big data use cases - Single Referrer, Holistic View, Multiple Referrers, Multiple Attributes, Large Data Volume.

Unit - 5

Structure data analysis - Structuring factors, Analysis of Comparisons, Additive and Interaction clouds. Inductive data analysis - Inference in Multivariate statistics, Univariate Effects, Combinatorial inference, Bayesian data analysis, Inductive Geometric Data Analysis, Guidelines for Inductive Analysis.

Books for Study:

1. Martinez, W. L., Martinez, A. R., andSolka, J. Exploratory data analysis with MATLAB. Chapman and Hall/CRC, 2017.

2. Andrienko, N., andAndrienko, G. Exploratory analysis of spatial and temporal data: a systematic approach. Springer Science & Business Media, 2006.

3. Le Roux, B., andRouanet, H. Geometric data analysis: from correspondence analysis to structured data analysis. Springer Science & Business Media, 2004.

CORE 2 -   QUANTITATIVE   DATA   ANALYSIS

No   of   Credits :   4

Unit - 1

Quantitative Statistics based Research - Data Analysis,Quantitative Versus Qualitative Analysis. Data Analysis Approaches - Statistics Terminology, HypothesisTesting, Statistical Significance and p-Value, Confidence Intervals, Statistical Power of a Test, Practical Significance Versus Statistical Significance, Statistical Independence, Degrees of Freedom, Measures of Central Tendency, Percentile and Percentile Rank, Central Limit Theorem, Law of Large Numbers.

Unit - 2

Analysis Issues and Potential Pitfalls: Effects of Variables, Outliers in the Dataset, Relationships between Variables. Graphically Representing Data: Data Distributions, Bell Curves, Skewed Curves, Bimodal Distributions, Poisson Distributions, Binomial Distribution, Histograms, Ranges of Values and Error Bars.

Unit - 3

Statistical Tests: Inter-Rater Reliability, Regression Models, Parametric Tests, Nonparametric Tests, One-Tailed or Two-Tailed Tests. Data Analysis Examples: Overview of Data Analysis Process - Perform an Exploratory Data Analysis, Statistical Analysis, Analyse the Results and Draw Conclusions

Unit - 4

Other Types of Data Analysis: Time-Series Experiment, Analysis for Data Clusters, Low-Probability Events, Metadata Analysis. Dependent, Independent and Controlled Variables,

Validity and Reliability, Variable Types, Type of Data, Independent Measures and Repeated Measures, Variation in Data Collection.

Unit - 5

Case study: Analysis of a Study on Reading and Lighting Levels, Analysis of Usability of an E-Commerce Site, Analysis of Essay Grading - Perform an Exploratory Data Analysis, Perform an Inferential Statistical Analysis.

Specific Analysis Examples - Handling Outliers in the Data, Floor/Ceiling Effects, Order Effects, Data from Stratified Sampling, Missing Data, Noisy Data, Transform the Data.

Books for Study:

1. Michael J. Albers. Quantitative Data Analysis in the Behavioural and Social Sciences, Wiley-Blackwell, 2017.
2. Suen, Hoi K., and Donald Ary. Analysing quantitative behavioural observation data. psychology press, 2014.
3. Riff, Daniel, et al. Analyzing media messages: Using quantitative content analysis in research. Routledge, 2019.

CORE -3   DECISION   SUPPORT   SYSTEM

No   of   Credits: 4

Unit – 1: Introduction

Decision Making – Rational decisions, Nature of managers, Appropriate deicision support, Group decision making, Intuition, Qualitative Data and Decision Making, Business Intelligence and Decision Making, managerial decision making and informative system, managers and computerized support framework and concept of decision support, systems, models, modeling process, intelligence phase, design phase, implementation phase.

Unit – 2: Decision Support System

DSS configuration, characteristics and capabilities of DSS, component of DSS, data management subsystem, model management subsystem, dialog subsystem, classification of DSS, distinguishing DSS from MIS and management science, modeling for MSS, static and dynamic models, treating certainty, uncertainty and risk, influence diagrams, optimization via mathematical programming, heuristic program, simulation, multidimensional modeling, visual spreadsheet, financial and planning modeling.

Unit – 3: Intelligent DSS, user interface

Knowledge based DSS concepts and definitions, artificial intelligence versus natural intelligence, knowledge in AI, types of knowledge based DSS, intelligent DSS, user interface, interface modes, graphics, multimedia and hypermedia, GIS, NLP overview and methods, DSS development strategies, development process, team developed DSS, DSS development tools.

Unit – 4: Enterprise support system

Networked decision support : The internet, intranet and collaborative technologies, Group decision support system – decision making in groups, goal of GDSS, technology of GDSS, GDSS software, idea generation, negotiation support system, Executive information and support systems – EIS concepts and definition, executives role, characteristics of EIS, comparing and integrating EIS and DSS, enterprise EIS, EIS implementation.

Unit – 5: Expert system and intelligent system:

Fundamentals of expert system, Expert system concepts, structure, human elements, working, benefits, limitation, success factors, types , knowledge engineering, scope of knowledge, difficulties in knowledge acquisition, methods of knowledge acquisition, knowledge representation, inferencing with rules, frames, model – based reasoning, case-based reasoning, introduction to building expert systems.

Reference books:

1. Vicki L.Sauter. Decision support systems for business intelligence. John Wiley & Sons, 2014.

2. Aronson, Jay E., Ting-Peng Liang, and Efraim Turban. Decision support systems and intelligent systems. Vol. 4. Pearson Prentice-Hall, 2005.

Core - 4   SOFT COMPUTING

No of Credits: 4

Unit – 1: Artificial Neural Networks

Soft computing vs. hard computing, types and applications of soft computing techniques. Artificial neural networks and their biological motivation – Terminologies – Models of neuron –Topology – characteristics of artificial neural networks – types of activation functions-, learning methods – error correction learning – Hebbian learning, Linear separability and XOR problem.

Unit – 2: Genetic   Algorithm

Basic concepts, encoding, fitness function, reproduction, Genetic modeling: Inheritance operator, crossover, inversion and deletion, mutation operator, Bitwise operator, generational cycle, convergence of GA, Applications and advances in GA, Differences and similarities between GA and other traditional method, Evolution Strategy and Evolutionary Programming, Direction-based Search

Unit – 3: Advanced   Evolutionary Algorithms

Encoding and Operators, Selection Methods, Replacement and Stop Criteria, Parameter Control - Strategy Parameter Setting, Examples of Variation Operator Control, Examples of popsize Control, Performance Evaluation of Evolutionary Algorithms

Unit – 4: Other Evolutionary Algorithms

Swarm Intelligence–Introduction, Ant Colony Optimization - Discrete Ant Colony Optimization, Continuous Ant Colony Optimization, Particle Swarm Optimization - Organic Particle Swarm Optimization, Neighbour Structure.

Unit – 5: Other Evolutionary Algorithms

Artificial Immune System Based on Clonal Selection, Immune Network, Negative Selection, Dendritic Cell Algorithms. Genetic Programming – Introduction, Difference between Genetic Programming and Genetic Algorithms, Genetic Programming for Curve Fitting, Other Code Methods for Genetic Programming, Example of Genetic Programming for Knowledge Discovery.

Reference Books

1. Sivanandam, S. N., and Deepa, S. N. Principles of Soft Computing. John Wiley and Sons, 2007.
2. Yu, X., and Gen, M. Introduction to Evolutionary Algorithms. Decision Engineering. Springer, 2010.
3. David E. Golberg. Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Publishing Company, Inc, 2002.

CORE -5   TEXT, WEB AND SOCIAL MEDIA ANALYTICS

No of Credits - 4

Unit – 1: Introduction to Text Mining: Text Representation- tokenization, stemming, stop words, TF-IDF, Feature Vector Representation, NER, N-gram modelling.Mining Textual Data: Text Clustering, Text Classification, Topic Modelling-LDA, HDP .

Unit –2: Introduction to Web-mining: Inverted indices and Boolean queries. PLSI, Query optimization, page ranking. Web usage & Web content Mining: Web Crawling-Crawler Algorithms, Implementation Issues, Evaluation, Session & visitor Analysis, Visitor Segmentation, Analysis of Sequential & Navigational Patterns, Predictions based on web user transactions.

Unit –3: Introduction Social Media Networks: Essentials of Social graphs, Social Networks, Models, Information Diffusion in Social Media.

Unit – 4: Mining Social Media: Behavioural Analytics, Influence and Homophily, Recommendation in Social Media

Unit –5: Sentiment Mining: Sentiment Classification, feature based opinion mining, comparative sentence and relational mining, Opinion spam.

Reference Books

1. Bing Liu, "Web Data Mining-Exploring Hyperlinks, Contents, and Usage Data", Springer, Second Edition, 2011.

2. Reza Zafarani, Mohammad Ali Abbasi and Huan Liu, "Social Media Mining-An Introduction", Cambridge University Press, 2014.

3. Bing Liu, "Sentiment Analysis andOpinionMining", Morgan & Claypool Publishers, 2012.

4. Nitin Indurkhya, Fred J Damerau, "Hand book of Natural Language Process", 2nd Edition, CRC Press, 2010.

5. Matthew ARussell,"Mining the social web", 2nd edition - O'Reilly Media, 2013.

CORE - 6   PYTHON   PROGRAMMING

No of Credits : 4

Unit 1: Introduction

Python Interpreter, Program Execution, Execution Model Variations, The Interactive Prompt - Starting an Interactive Session, System Path, Launcher, Code Directories, System Command Lines and Files,Module Imports and Reloads, IDLE User Interface.

Unit 2: Data Types

User-Defined Classes, Numeric Types, Basics, Numeric Literals, Built-in Numeric Tools, Python Expression Operators, Numbers in Action, Variables and Basic Expressions, Numeric Display Formats, Other Numeric Types.

Unit 3:   Flow control

Conditional blocks using if, else and elif, Simple for loops in python, For loop using ranges, string, list and dictionaries, Use of while loops in python, Loop manipulation using pass, continue, break and else, Programming using Python conditional and loops block

Unit 4: Core data types

Python Object - Python's Core Data Types, Numbers, Lists, Dictionaries, Tuples, Files.String - Fundamentals, String Basics, String Literals, Single- and Double-Quoted Strings, Escape Sequences, Raw Strings, Basic Operations, Indexing and Slicing, String Conversion Tools, String Methods, String Formatting Expressions, String Formatting Method Call, Regular expression, Exception Handling.

Unit 5: Functions, Modules and Packages

Function Basics - def, scope, arguments, Function Design Concepts, Recursive Functions, Modules and Packages, Module Namespaces. Organizing python codes using functions, Organizing python projects into modules, Importing own module as well as external modules,

Understanding Packages, Powerful Lamda function in python, Programming using functions, modules and external packages.

Books for Reference:

1. Mark Lutz. Learning python: Powerful object-oriented programming. O'Reilly Media, Inc.", 2013.
2. David M. Beazley. Python essential reference. Addison-Wesley Professional, 2009.