Vikram University, Ujjain

Board of studies in Computer science (Faculty of Engineering Science)

SYLLABUS of PGDCSA in Data Science Programme

Exclusively for University Teaching Department (ICS, VUU)

ONE YEAR PGDCSA in Data Science PROGRAMME of UTD (ICS, VUU)

(Effective from Academic Session 2020-21)

[Modified as according to the provision of "Ordinance 14"]

COURSE STRUCTURE

PGDCSA (Data Science) FIRST SEMESTER

SN	Course code	Title	End term sem Exam	Internal	Max Marks	Credits* C	Distribution of Credits L	T	P
1	PG-101	Basic Statistical Methods	60	40	100	6	4	2	
2	PG-102	Data Mining and Data Warehousing	60	40	100	6	4	2	
3	PG-103	Big Data Technologies	60	40	100	6	4	2	
4	PG-104	RDBMS and NOSQL	60	40	100	6	4	2	
5	PG-105	Machine Learning	60	40	100	6	4	2	
6	PG-106	Comprehensive Viva Voce	50	-	50	04 Virtual (VR)			
		Total			550	30+04	20	10	

C=Credit Per Week) / (L = Lectures Per Week)/ (T & PW =Tutorials &Practical Work per week)

***One Credit is equivalent to one hour (60 minutes) of teaching (lecture or tutorial)and two hours (120 minutes) for practical**

COURSE STRUCTURE

PGDCSA (Data Science) SECOND SEMESTER

SN	Course code	Title	End term sem Exam	Internal	Max Marks	Credits*	Distribution of Credits		
						C	L	T	P
1	PG-201	Advanced Statistical Model and Analysis	60	40	100	6	4	2	
2	PG-202	Python for Analytics	60	40	100	6	4		2
3	PG-203	Decision Analysis	60	40	100	6	4	2	
4	PG-204	Fundamental of Algorithm	60	40	100	6	4	2	
5	PG-205	Cloud Computing	60	40	100	6	4	2	
6	PG-206	Comprehensive Viva Voce	50	-	50	04 Virtual (VR)			
		Total			550	30+04	20	08	02

C=Credit Per Week) / (L = Lectures Per Week)/ (T & PW =Tutorials &Practical Work per week)

***One Credit is equivalent to one hour (60 minutes) of teaching (lecture or tutorial) and two hours (120 minutes) for practical**

PG- 101 : Basic Statistical Methods

Unit- I

Statistics - Meaning and Definition, significance, scopes and limitation of statistics, Statistical Investigation, Process of Data collection, Primary and Secondary Data, Preparation of Questionnaire, Classification and Tabulation of Data, Preparation Of statistical series and its types.

Unit-II

Measurement of Central Tendency- Mean, Median, Mode, Quartile, Harmonic mean and Geometric mean.

Unit -III

Analysis of Time Series- Meaning, Importance, Components, Decomposition of Time series, Measurement of long Term Trends, Measurement of Cyclical and Irregular Fluctuation, Dispersion and skewness.

Unit- IV

Correlation – Meaning, Definition, Types and Degree of correlation, Methods of correlation. Regression Analysis – Meaning, uses, difference between Correlation and Regression, Linear Regression, Regression Equation, Calculation of Coefficient of Regression.

Unit -V

Sampling & Distribution: The Central limit Theorem, distribution of the sample mean and Methods of sampling, Chi square test.

Text:
1. R. Panneerselvam, "Research Methodologies," PHI.
2. C.R. Kothari: Research methodology, Methods and Techniques, New Age Publication.
3. S.M. Ross, A First Course in Probability, 8 th Edition, Prentice Hall.

PG-102 : Data Mining and Data Warehousing

UNIT 1

Introduction : Data Mining: Definitions, KDD v/s Data Mining, DBMS v/s Data Mining , DM techniques, Mining problems, Issues and Challenges in DM, DM Application areas. Association Rules & Clustering Techniques: Introduction, Various association algorithms like A Priori, Partition, Pincer search etc., Generalized association rules.

UNIT 2

Clustering paradigms; Partitioning algorithms like K-Medioid, CLARA, CLARANS; Hierarchical clustering, DBSCAN, BIRCH, CURE; categorical clustering algorithms, STIRR, ROCK, CACTUS. Other DM techniques & Web Mining: Application of Neural Network, AI, Fuzzy logic and Genetic algorithm, Decision tree in DM. Web Mining, Web content mining, Web structure Mining, Web Usage Mining.

UNIT 3

Temporal and spatial DM: Temporal association rules, Sequence Mining, GSP, SPADE, SPIRIT, and WUM algorithms, Episode Discovery, Event prediction, Time series analysis. Spatial Mining, Spatial Mining tasks, Spatial clustering, Spatial Trends.

UNIT 4

Data Mining of Image and Video: A case study. Image and Video representation techniques, feature extraction, motion analysis, content based image and video retrieval, clustering and association paradigm, knowledge discovery.

UNIT 5

The vicious cycle of Data mining, data mining methodology, measuring the effectiveness of data mining data mining techniques. Market baskets analysis, memory based reasoning, automatic cluster detection, link analysis, artificial neural networks, generic algorithms, data mining and corporate data warehouse, OLA

Reference Books:
1. Data Mining Techniques ; ArunK.Pujari ; University Press.
2. Data Mining; Adriaans&Zantinge; Pearson education.
3. Mastering Data Mining; Berry Linoff; Wiley.

PG- 103: Big data Technologies

UNIT I:

Introduction Introduction to Big Data Platform – Challenges of Conventional Systems - Intelligent data analysis – Nature of Data - Analytic Processes and Tools - Analysis vs Reporting – Modern Data Analytic Tools. Big Data Analytics Process, Big Data Analytics for Business. Identifying problem and solving problem in Big Data environment. Analyzing Unstructured vs. Structured Data, Databases.

UNIT II:

Hadoop and MapReduce Introduction to Hadoop, Hadoop architecture, A Brief History of Hadoop, Apache Hadoop and the Hadoop Ecosystem, Hadoop Releases; Hadoop Distributed File system: Design of HDFS, HDFS Concepts.

UNIT III:

Introduction to MapReduce: MapReduce Basic Concepts, Understanding the Map Reduce architecture, Writing MapReduce Programs. understanding Map phase, shuffling, sorting, and reducing phase.

UNIT IV:

Spark Introduction to Spark, Resilient Distributed Dataset (RDD), RDD Operations: actions and transformation functions. Spark Data frames, operations on Data frames: Join, group by, aggregate, handling missing data.

UNIT V:

Sparks and MLLib Sparks and its basic operations. MLLib: Data types, Basic statistics, Classification(Logistic regression, Decision tree classifier)and linear regression model generation, Model Evaluation, Collaborative filtering, and Clustering.

Text Books:

1. Arvind Sathi, "Big Data Analytics: Disruptive Technologies for Changing theGame", 1st Edition, IBM Corporation, 2012.

2. Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics", 1st Edition, Wiley and SAS Business Series, 2012.

3. Tom White,"Hadoop: The Definitive Guide", 3rd Edition, O'Reilly Media, 2012

4. Donald Miner, Adam Shook, Eric Sammer, "Hadoop Operation", O'Reilly 2012.

5. Donald Miner, Adam Shook "MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems", O'Reilly 2012.

6. Chuck Lam, "Hadoop in Action", Manning Publications, 2010.

7. https://spark.apache.org/docs/2.0.0/programming-guide.html

PG- 104 : RDBMS and NO SQL

UNIT I:
Overview of DBMS: Comparison between Database approach and Traditional file accessing approach, Advantages of database systems, Schemas and instances, Data Dependency, Data Dictionary, and Meta Data. Data models, Types of Data models (ObjectOriented, Record Based and Physical data models), E-R Modelling.

UNIT II:
Relational Data model: Domains, Tuples, Attributes, Keys, Relational database, Schemas, Integrity constraints, Relational algebra and relational calculus; Normalization: Normal forms (INF, 2NF, 3NF, BCNF), Functional dependency, Decomposition,Dependency preservation and lossless join.

UNIT III:
Structured Query Language: DDL, DML, DCL, TCL, SQL Functions, integrity constraints, various joins, sub-query, index, View, Sequence, and Clusters.

UNIT IV:
NoSQL: Nosql Basics, Storage Architecture, Operations, Query Model, Modifying Data Stores and Managing Evolution, Indexing and Ordering Data Sets, Managing
Transactions and Data Integrity.

UNIT V:
UsingNosql in the Cloud, Scalable Parallel Processing with Mapreduce, Analyzing Big Data with Hive, Surveying Database Internals.

Text Books:

1. A Silberschatz, H Korth, S Sudarshan, "Database System and Concepts", fifth Edition McGraw-Hill.
2. ElmasriRamez and NovatheShamkant, "Fundamentals of Database Systems", Benjamin Cummings Publishing. Company.
3. Rob, Coronel, "Database Systems", Seventh Edition, Cengage Learning.
4. Fred R.McFadden,JeffreyA.Hoffer& Marry B.Prescott.?Modern Database Management, Fifth Edition,Pearson Education Asia,2001.
5. Bayross Ivan, "SQL, PL/SQL: The Programming Language Of Oracle", 4th Revised Edition, BPB Publications.

PG - 105 : Machine Learning

UNIT I

Learning Problems - Perspectives and Issues - Concept Learning - Version Spaces and Candidate Eliminations - Inductive bias - Decision Tree learning - Representation Algorithm- Heuristic Space Search.

UNIT II

Neural Network Representation - Problems - Perceptrons – Multilayer Networks and Back ropagation Algorithms - Advanced Topics - Genetic Algorithms - Hypothesis Space Search Genetic Programming - Models of Evalution and Learning.

UNIT III

Bayes Theorem - Concept Learning - Maximum Likelihood – Minimum Description Length Principle - Bayes Optimal Classifier - Gibbs Algorithm - Naïve Bayes Classifier – Bayesian Belief Network - EM Algorithm - Probability Learning - Sample Complexity - Finite and Infinite Hypothesis Spaces - Mistake Bound Model.

UNIT IV

K- Nearest Neighbour Learning - Locally weighted Regression - Radial Bases Functions - Case Based Learning.

UNIT V

Learning Sets of Rules - Sequential Covering Algorithm - Learning Rule Set - First Order Rules - Sets of First Order Rules - Induction on Inverted Deduction - Inverting Resolution -Analytical Learning - Perfect Domain Theories - Explanation Base Learning – FOCL Algorithm- Reinforcement Learning - Task - Q-Learning - Temporal Difference Learning

TEXT BOOKS:

1. Machine Learning - Tom M. Mitchell, - MGH

REFERENCE BOOKS
1. Machine Learning: An Algorithmic Perspective, Stephen Marsland, Taylor & Francis

PG - 201: Advance Statistical Model and Analysis

Unit- I

Probability and probability distribution probability : classical ,relative frequency and axiomatic definition of probability, Bayes Theorem and Independence. Probability distribution: binomial,, geometric, negative binomial uniform exponential, normal and log normal distribution.

Unit – II

Introduction to non-linear modelling ; sampling methods: basic sampling algorithms, rejections sampling, adaptive rejection sampling.

Unit – III

Linear statistical models; multiple linear regression, inference technique for the general linear model, generalised linear models, inference procedure, special case of generalized linear models leading to Logistic regression and log linear model.

Unit- IV

Random variables: discrete, continuous and mixed random variable, probability mass, probability density and cumulative distribution functions, mathematical expectation ,moments median, quartiles, Markov inequality, correlation and , independence of random variable.

Unit – V

Hypothesis testing basic: ideas of testing hypothesis, null and alternative hypothesis, the critical and acceptance regions, two types of errors, Test for one sample and two sample problems for normal population ,test for proportions.

Refrences Book :

1. Krzanowski, W. J., An Introduction to Statistical Modeling, Wiley (2010).

2. Hastie, T., Tibshirani, R., and Friedman, J., The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer (2002).

3. Bishop, C. M., Pattern Recognition and Machine Learning, Springer (2006).

PG - 202 : Python for Analytics

UNIT-I

Introduction: check icon History, Features, Setting up path, Variable and Data Types, Operator. **Conditional Statements:** if, if-else, if-elif, nested if-else and Looping: for, while, nested loops with break, continue and pass keyword. **String Manipulation:** Accessing Strings, Basic Operations, String slices, Function and Methods. **Functions:** Defining and Calling of a function, Types of functions, Function Arguments, Anonymous functions, Global and local variables.

UNIT-II

Collection: list, tuple , Dictionaries. Introduction, Accessing values, Working, Properties, Functions and Methods. **Modules:** Importing module, Math module, Random module, os module, date-time module, calendar module, Packages, user defined module, introduction of pip. **IO:** Printing on screen and Reading data from keyboard, Opening and closing file, Reading and writing files, Functions.

UNIT-III

Exception Handling: Except, Try, else, finally clause, User Defined Exceptions, raise user-defined exception, nested try-except. **OOPs concept:** Class and object, Attributes, Inheritance, Overloading, Overriding, Data hiding, final and abstract class. Database: Introduction, Connections with MYSQL, Executing queries, Transactions) web-designing: HTML, CSS, JAVASCRIPT.

UNIT-IV

CGI: Introduction, Architecture, CGI environment variable, GET and POST methods. application using CGI: signup, login and session tracking with server side programming.

UNIT-V

DJANGO: working of MVT, Environment setting and installation, creating a Project, Apps Life Cycle, Admin Interface, Views, URL Mapping. **Template System:** DTL and JINJA. Models, Page Redirection, Form Processing, project with signup and login.

References:

1. Programming and Problem Solving with Python (Ashok Namdev Kamthane and Amit Ashok Kamthane) McGraw Hill publication
2. Let Us Python (Kanetkar Yashavant) BPB Publication
3. Python Complete Refernce (Brown Martin C.) McGraw Hill publication
4. Python Programming A Modular Approach (Naveen and Kumar and Taneja Sheetal) PEARSON
5. Beginning Django (Rubio Daniel) Apress

PG -203 : Decision Analysis

Unit-I:

Game Theory: Introduction, definitions, two-person zero sum game. Game with pure strategies, saddle point, game value. Game with mixed strategies, solution methods algebraic method, graphical method, dominance, linear programming method.

Unit-II:

Decision Theory: Structuring the decision problem-payoff tables, decision trees.Decision making under certainty. Decision making under uncertainty: Optimistic,Conservative, Minimax regret. Decision making under risk: Expected value criterion,expected value of perfect information.

Unit-III

Sensitivity analysis. Decision making with sample information, expected value of sample information, efficiency of sample information. Computing branch probabilities. Utility and decision making: meaning of utility, developing utilities and payoffs, the expected utility approach.

Unit-IV:

Multi-criteria Decision Making: Multi-criteria decision making: Goal programming,Scoring models, Analytical Hierarchy Process.

Unit-V:

Markov Analysis: Transition probabilities, system behaviour. Methods of Analysis:Tree diagram, Matrix multiplication, Algebraic solution. Cyclical and absorbing states,Market share analysis, Accounts receivable analysis.

TEXTBOOKS

1. Hamdy A. Taha: Operations Research: An introduction, Pearson Prentice Hall
2. David R. Anderson, Dennis J. Sweeney, Thomas A. Williams: An Introduction to Management Science, South-Western College Publishing.

3. William J. Stevenson: Introduction to Management Science, IRWIN.

PG -204 Fundamental of Algorithms

UNIT 1
Introduction and Review: What is an Algorithm, Algorithm's Performance, order architecture: Θ-Notation, O-Notation, Ω-Notation, Algorithm Analysis: time space complexities, Worst-case Complexity, Average-case Complexity.

UNIT 2
Divide and conquer: Structure of divide-and –conquer algorithms: examples, Binary search, quick sort, Analysis of divide and conquer, run time recurrence relations.

UNIT 3
Graph Searching and Traversal: Overview, Traversal methods: depth first and breadth first search.
Greedy Method: Overview of the greedy method, Minimum spanning trees, Single source shortest paths.

UNIT 4

Dynamic programming: The general method, principle of optimality, difference between dynamic programming and greedy method, Applications: optimal binary search trees, Back tracking: The general method, 8-queens problem.

UNIT5

Branch and Bound Algorithm: The Branch and bound method, FIFO and LIFO branch and bound, LC (Least Cost) search, Traveling Salesman Problem, LCBB on Traveling Salesman Problem.

Reference Books:

1. Fundamentals of Computer Algorithms By Ellis Horowittz and SartajSahni, Galgotia Publications.
2. Ullman "Analysis and Design of Algorithm" TMH
3. Goodman " Introduction to the Design & Analysis of Algorithms, TMH-2002
4. Sara Basse, A.V. Gelder, " Computer Algorithms, " Addison Wesley
5. T.H. Cormen, Leiserson, Rivert and stein, " Introduction of Computer algorithm, " PHI
6. E. Horowitz, S. Sahni, and S. Rajsekaran, "Fundamentals of Computer Algorithms, " Galgotia Publication.

PG 205 Cloud Computing

UNIT 1

Cloud Computing: Definition, Cloud Architecture, Cloud Storage, Advantages and Disadvantages of Cloud Computing, Companies in the Cloud Today, Cloud Services, Cloud Types: The NIST Model, The Cloud Cube Model, Deployment Models, Service Models Cloud Computing, Service Models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS).

UNIT 2

Web-Based Application– Pros and Cons of Cloud Service Development–Types of Cloud Service Development–Software as a Service– Platform as a Service–Web Services– On-Demand Computing–Discovering Cloud Services Development Services and Tools–AmazonEc2–Google App Engine–IBM Clouds.

UNIT 3

Centralizing Email Communications– Collaborating on Schedules–Collaborating on To-Do Lists–Collaborating Contact Lists–Cloud Computing for the Community–Collaborating on Group Projects and Events–Cloud Computing for the Corporation.

UNIT 4

Collaborating on Calendars, Schedules and Task Management–Exploring Online Scheduling Applications–Exploring Online Planning and Task Management–Collaborating on Event Management–Collaborating on Contact Management–Collaborating on Project Management–Collaborating on Word Processing-Collaborating on Databases–Storing and Sharing Files.

UNIT 5

Cloud computing security architecture: Architectural Considerations-General Issues, Trusted Cloud computing, Secure Execution Environments and Communications, Micro-architectures; Identity Management and Access control Identity management, Access control, Autonomic Security Cloud computing security challenges: Virtualization security management virtual threats, VM Security Recommendations, VM-Specific Security techniques, Secure Execution Environments and Communications in cloud.

Reference Books:

1.Barrie Sosinsky,"Cloud Computing Bible",Wiley India pub.
2.Dinakar Sitaram,"MovingtoTheCloud",Elsevier,2014.
3.Danc.Marinercus,"CloudComputingTheoryAndPractice",Elsevier,2013.