

## **Vikram University, Ujjain**

**Board of studies in Computer science (Faculty of 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”]**

Vikram University, Ujjain –PGDCSA ( DATA SCIENCE) Syllabus  
(As per CBCS pattern)w.e.f. 2020-2021 and onwards

---

**COURSE STRUCTURE**

**PGDCSA (Data Science) FIRST SEMESTER**

S N	Course code	Title	End term sem Exam	Inter nal	Max Marks	Credits*	Distribution of Credits		
							C	L	T
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 <u>Virtual (VR)</u>			
		<b>Total</b>			550	<u>30+04</u>	<u>20</u>	<u>10</u>	

**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**

Vikram University, Ujjain –PGDCSA ( DATA SCIENCE) Syllabus  
(As per CBCS pattern)w.e.f. 2020-2021 and onwards

---

**COURSE STRUCTURE**

**PGDCSA (Data Science) SECOND SEMESTER**

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

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

Vikram University, Ujjain –PGDCSA ( DATA SCIENCE) Syllabus  
(As per CBCS pattern)w.e.f. 2020-2021 and onwards

---

**PG- 101 : Basic Statistical Methods**

Unit I: Statistical analysis, Measures of central tendency and dispersion, mean, median, mode, range, mean and standard deviations, computing correlation in variables, linear and non-linear regression.

Unit II: Probability and Probability distributions Probability: classical, relative frequency and axiomatic definitions of probability, addition rule and conditional probability, multiplication rule, total probability, Bayes' Theorem and independence. Probability distributions: binomial, poisson, geometric, negative binomial uniform exponential, normal and log normal distribution.

UNIT III

Random Variables: Discrete, continuous and mixed random variables, probability mass, probability density and cumulative distribution functions, mathematical expectation, moments, probability and moment generating function, median and quintiles, Markov inequality, correlation and regression, independence of random variables.

Unit IV

Sampling & Distributions The Central Limit Theorem, distributions of the sample mean and the sample variance for a normal population, ChiSquare, t and F distributions, problems.

UNIT V

Hypothesis Testing: Basic ideas of testing hypothesis, null and alternative hypotheses, the critical and acceptance regions, two types of error, tests for one sample and two sample problems for normal populations, tests for proportions, Chi-square goodness of fit test and its applications. Software and Tools to be learnt: Statistical packages like SPSS and R.

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.

# Vikram University, Ujjain –PGDCSA ( DATA SCIENCE) Syllabus (As per CBCS pattern)w.e.f. 2020-2021 and onwards

---

## **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-Medoid, 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.

# Vikram University, Ujjain –PGDCSA ( DATA SCIENCE) Syllabus (As per CBCS pattern)w.e.f. 2020-2021 and onwards

---

## **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>**

# Vikram University, Ujjain –PGDCSA ( DATA SCIENCE) Syllabus (As per CBCS pattern)w.e.f. 2020-2021 and onwards

---

## PG- 104 E2: 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 (1NF, 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.

Vikram University, Ujjain –PGDCSA ( DATA SCIENCE) Syllabus  
(As per CBCS pattern)w.e.f. 2020-2021 and onwards

---

**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 Propagation Algorithms - Advanced Topics - Genetic Algorithms - Hypothesis Space Search Genetic Programming - Models of Evaluation 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

Vikram University, Ujjain –PGDCSA ( DATA SCIENCE) Syllabus  
(As per CBCS pattern)w.e.f. 2020-2021 and onwards

---

**PG - 201:** Advance Statistical Model and Analysis

UNIT-I

An overview of basic probability theory and theory of estimation; Bayesian statistics; maximum a posteriori (MAP) estimation; conjugate priors;

UNIT-II

Exponential family; posterior asymptotics; linear statistical models; multiple linear regression: inference technique for the general linear model, generalised linear models: inference procedures, special case of generalised linear models leading to logistic regression and log linear models;

UNIT-III

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

UNIT-IV

The EM algorithm Markov chain, Monte Carlo, Gibbs sampling, slice sampling.

References:

1. Dobson, A. J. and Barnett, A. G., An Introduction to Generalised Linear Models, 3rd ed., Chapman and Hall/CRC (2008).
2. Krzanowski, W. J., An Introduction to Statistical Modeling, Wiley (2010).
3. Hastie, T., Tibshirani, R., and Friedman, J., The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer (2002).
4. Bishop, C. M., Pattern Recognition and Machine Learning, Springer (2006).

# Vikram University, Ujjain –PGDCSA ( DATA SCIENCE) Syllabus (As per CBCS pattern)w.e.f. 2020-2021 and onwards

---

## PG - 202 : Python for Analytics

### UNIT I:

**Introduction to Python: Python versus Java, Python Interpreter and its Environment, Python installation, Python basics: variables, operators, Strings, Conditional and Control Statements, loops; Data structures: lists and dictionaries; functions: global functions, local functions, lambda functions and methods.**

### UNIT II:

**Object Oriented Programming Concepts:** Class, object, constructor, destructor and inheritance; Modules & Packages, File Input and Output, Catching exceptions to deal with bad data, Multithreading, Database Connectivity.

### UNIT III:

**Numpy:** Creating Arrays, Arrays Operations, Multidimensional Arrays Arrays transformation, Array Concatenation, Array Math Operations, Multidimensional Array and its Operations, Vector and Matrix.

### UNIT IV

**Visualization:** Visualization with matplotlib, Figures and subplots, Labeling and arranging figures, Outputting graphics.

### UNIT IV:

**Pandas:** Manipulating data from CSV, Excel, HDF5, and SQL databases, Data analysis and modelling with Pandas, Time-series analysis with Pandas, Using Pandas, the Python data analysis library, Series and Data Frames, Grouping, aggregating and applying, Merging and joining.

### Text Books:

1. McKinney Wes, "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython", O'Reilly Media, 2012.
2. Hauck Trent, "Instant Data Intensive Apps with Pandas How-To", Packt Publishing Ltd, 2013.
3. Beazley David M., "Advanced Python Programming", Pearson Education, 2009.
4. Chun Wesley, Core Python Programming, 3rd Edition, Prentice Hall Professional, 2012.
5. Telles Matt "Python Power!: The Comprehensive Guide", Cengage Learning, 2008.

Vikram University, Ujjain –PGDCSA ( DATA SCIENCE) Syllabus  
(As per CBCS pattern)w.e.f. 2020-2021 and onwards

---

**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.:

# Vikram University, Ujjain –PGDCSA ( DATA SCIENCE) Syllabus (As per CBCS pattern)w.e.f. 2020-2021 and onwards

---

## **PG -204 Fundamental of Algorithms**

### **UNIT 1**

Introduction and Review: What is an Algorithm, Algorithm's Performance, order architecture:  $\Theta$ -Notation, O-Notation,  $\Omega$ -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 Horowitz 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.

# Vikram University, Ujjain –PGDCSA ( DATA SCIENCE) Syllabus (As per CBCS pattern)w.e.f. 2020-2021 and onwards

---

## **PG - 205 E2: 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 – Amazon Ec2 – 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, "Moving to The Cloud", Elsevier, 2014.
3. Danc. Marinercus, "Cloud Computing Theory And Practice", Elsevier, 2013.