Vikram University, Ujjain

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

SYLLABUS of

PG Diploma in Data Analytics

Exclusively for University Teaching Department (ICS, VUU)

ONE YEAR PG Diploma in Data Analytics

PROGRAMME of UTD (ICS, VUU)

(Effective from Academic Session 2020-21)

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

- **Objective of Course**

  Diploma in Data Analytics is a unique six months programme offered by Institute of Computer Science, Vikram University, Ujjain is an excellent blend of knowledge and practice in the field of Data Science and its industrial applications. The program is targeted for creating qualified Data Science Professionals. The programme also equipped with strong analytical and programming prospects. The objective of this program is to develop Data Scientists and Data Analysts for the IT industries and other industries.

- **Outcome of the Course:**

  On completion of the course the participants will learn the concept of Data Analytics using open source statistical tools like Python and some other tools and techniques. The participants will be able to implement industry oriented Data Analytics Project.
  .

COURSE STRUCTURE

PG Diploma in Data Anaylitics

First Semester

SN	Course code	Title	End term sem Exam	Internal	Max Marks
1	PGDA-101	Fundamental of Information Technology	70	30	100
2	PGDA -102	PC-Packages	70	30	100
3	PGDA -103	Programming and Problem solving in C	70	30	100
4	PGDA- 104	Digital Computer Orgnization	70	30	100
5	PGDA -105	Modern Technologies of Computer Science	70	30	100
		Total			500

COURSE STRUCTURE

PG Diploma in Data Analytics

Second Semester

SN	Course code	Title	End term sem Exam	Internal	Max Marks
1	PGDA -201	Introduction to Data Analytics	70	30	100
2	PGDA -202	Python for Analytics	70	30	100
3	PGDA -203	Big Data Technology	70	30	100
4	PGDA - 204	Datamining and NoSQL concepts	70	30	100
5	PGDA -205	Data Analytics in IOT	70	30	100
		Total			500

PGDA-101: Fundamental of Information Technology

UNIT-1

Computer Fundamental: Characteristics of Computers, History of Computer, Evolution of Computers, Computer Generations and Types of Computer. **Components of a Computer:** Registers, instruction Set, Bus Architecture. **Computer Hardware:** Input Devices, Output Devices. **Storage Devices:** Primary Storage capacity, Memory Types, Memory Measuring Units, Secondary Storage. Software & Software Types, Computer Languages, Compiler, Interpreter.

UNIT-2

Introduction of Programming: Procedure Oriented Programming, Object oriented programming, Concepts used in OOP, Benefits of OOP, Main advantages and disadvantage of OOP, Applications of OOP, OOP vs. POP.

UNIT-3

Operating System Overview: Computer System Startup, Computer System Structure, computer system components, operating System classifications, operating System Services, Major Functions of operating system, Process Management, CPU Scheduling, Scheduling Criteria. **Memory and File Management:** Memory Management Requirements, Swapping, Memory Management Techniques.

UNIT-4

Introduction to DBMS: File System, Traditional File Oriented Approach, DBMS Advantages and Disadvantage, Role of DBMS, Three views of data, DBMS Architecture. Data Models, Data Independence, Major components of DBMS, Data Dictionary, Types of Users, DBMS applications, Keys in Databases, Database Languages.

UNIT-5

Introduction to computer Networks: computer Network Definition, Importance of Networking, Types of Networks, Network Topology, Advantages and Disadvantage of computer Networks, Applications of computer networks, Reference Model, Internet, introduction to Internet Technology, Electronic Mail, World Wide Web.

Reference Books:

1. Operating Systems Concepts, A. Silbrschaz, P.Galvin, G.Gagne, John Wiley & Sons
2. Object Oriented Programming in C++, Robert Lafore, Galgotia Publication.
3. Data base management systems vol. 1., Date C.J.
4. Fundamental of Computer Science & IT, Singh Umesh Kumar, Jain S., Maheshwari A., SSDN publications New Delhi,
5. Data Communications and Networks, Godbole A, Tata Mccraw-Hill Publications.

PGDA- 102: PC-Packages

UNIT-1

MS windows: Introduction to MS windows; Features of windows; working with windows; My computer & Recycle bin; Desktop, Icons and windows Explorer; Screen description & working styles of windows; Dialog Boxes & Toolbar; working with files & Folders, Simple operations like copy, delete, moving of files and folders from one drive to another; Accessories and. windows settings using control panel-setting common devices using control panel, modem, printers, audio, network, fonts, creating users, internet settings, Start button & program lists ;Installing and uninstalling new Hardware & software program on your computer.

UNIT-2

MS word Basics - Introduction to MS office; introduction to MS- word; Features & area of use, working with MS- word; Menus & commands; Toolbars & Buttons; shortcut Menus, wizards & Templates, creating a New Document; Different page views and Layouts; Applying various Text Enhancements; working with - Styles, Text .Attributes; paragraph and Page. Formatting; Text Editing using various features; Bullets, Numbering, Auto formatting, Printing & various print options.

UNIT-3

Advanced Features of MS- word- Spell check, Thesaurus, Find & Replace; Headers & Footers: Inserting- Page Number, Pictures, Files, Auto texts, Symbols etc.; working with columns, Creation and working with Tables including conversion to and from text; Margins and Space management in Documents.

UNIT- 4

MS Excel: Introduction and area of use; working with MS Excel: concept of workbook and worksheet; Using Wizards; Various Data Types; Using different features with Data, Cell and Texts; Inserting, Removing & Resizing of Columns & Rows; Working with Data & Ranges; Different views of Worksheet; Column Freezing, Labels, Hiding, Splitting etc.; Using different features of Data and Text; Use of Formulas, Calculation & Functions; Cell formatting including Borders and Shading; Working with Different Chart Types; Printing of Workbook & Worksheets with Various options.

UNIT-5

MS PowerPoint: Introduction and area of use; Working with MS PowerPoint; Creating a New Presentation; Working with Presentation; Using Wizards; Slides & its Different Views; Inserting, Deleting and Copying of Slides; Working with Notes, Handouts; Columns and Lists; Adding Graphics, Sounds and Movies to a slide; Working with PowerPoint Objects; Designing and Presentation of a Slide Show; Printing presentations; Notes, Handouts with print options.

Reference Books:

1. Windows XP Complete Reference. BPB publications
2. MS Office XP complete BPB Publication.

PGDA-103: Programming and Problem Solving in C

UNIT-1

Problem identification analysis, design, coding, testing & debugging, implementation, modification & maintenance, algorithms & flowcharts, Characteristics of a good program - accuracy, simplicity, robustness, portability, minimum resource & time requirement, modularization; documentation, naming variables; Top down design; Bottom-up design.

UNIT-2

History of C, Structure of a C program, Data types, Constant & Variable, Operators & expressions, **Control Constructs** - if-else, for, while, do-while, Case statement, Arrays, Type modifiers & Storage classes, Ternary operator, Type conversion & type casting.

UNIT-3

Functions, Arguments, return value, Parameter passing - call by value, call by reference, return statement, Scope, visibility and life time rules for various types of variable, static variable, calling a function, Recursion - basics, comparison with iteration, tail recursion, when to avoid recursion examples.

UNIT-4

Special constructs - break, continue, exit , goto & labels; pointers - &and * operators, pointer expression, pointer arithmetic, String, Pointer to function, Function to parameter, structure - basic, declaration, membership operator, pointer to structure, referential operator, self-referential structures, structure within structure, array in structure, array of structures, Union - basic, declaration: Enumerated data type, Command line arguments.

UNIT-5

File handling and related functions: pdntf & family, c preprocessor- basics, # Include, # define, # undef, conditional compilation directive like #if, #else, #endif, #ifdefand #ifndef, Variable argument list functions.

Reference Books:

l. Keminghan& Richie: The C progamming tanguage, pHI

2. Cooper Mullish: The Spirit ofC, Jaico publishing-Housg Delhi

3. Kanetkar Y: Let us C 4, Kanetkar Y: Pointers in C.

PGDA- 104: Digital computer Organization

UNIT-1:

Digital components: Functional units of a computer, logic gates, Minimization of Boolean Expressions, Flip-Flips, Decoders, Encoders, Multiplexers, Counters, and Registers.

UNIT-2:

Data Representation: Number systems, Representations of signed and unsigned numbers, alphanumeric codes, Addition of binary numbers, subtraction, 2's complement, and Floating point number representation.

UNIT-3:

Register Transfer Language & Micro-operations: Concepts of the Bus, Timings in Register transfer, Languages used for data transfer in registers, Data movement from/to memory.

UNIT-4:

Arithmetic circuits, Half adder, full adder, N-bit adder, Logical micro operation, arithmetic logic unit. Instruction sets for basic computer: Addressing modes, Instruction cycles, Control signal generation.

UNIT-5:

Central Processing Unit: General register organization, Memory stacks, Instruction types, Interrupts, Instruction pipelining, Arithmetic pipelining. .

Reference books:

1. P. N. Basu, Computer Organization and Architecture, Vikas Publication, 2nd Edition.
2. H. Patterson, Computer Architecture: A Quantitative approach, Elsevier, 5th Edition.
3. W. Stalling, Computer Organization and architecture, Pearson Education Asia, 5th Edition.
4. Donald Leach & Albert Malvino, Digital Principles & Applications, McGraw Hill, 7th Edition.

PGDA-105: Modern Technologies of Computer Science

UNIT-1:

Introduction to Computer Security: The Challenges of Computer Security, The OSI Security Architecture. Security Attacks (Passive Attacks, Active Attacks). Security Services (Authentication, Access Control, Data Confidentiality, Data Integrity, Nonrepudiation, Availability Service).

UNIT-2:

Introduction to Artificial Intelligence: What is AI ? The Importance of AI. AI and related fields. Introduction to Natural Language Processing.,Application of AI.. Basic Problem solving methods: Production systems-state space search, control strategies, Breadth first search, Depth first search, Heuristic search.

UNIT-3:

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

UNIT-4:

Introduction to IoT: Definition, Characteristics, Conceptual framework, Architectural view. Technology involved - Server-end technology, Hardware and Software components, Development tools & Open source framework, APIs & Device interfacing components, Platforms & Integration tools, Sources of IoT, Advantages and Disadvantages of IoT.

UNIT-5:

Introduction to 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.

Reference Books:

1. Charles P. Pleeger, "Security in Computing", Pearson Education Asia, 5th Edition, 2001.
2. William Stallings, "Network Security Essentials: Applications and standards", Person Education Asia, 2000
3. Dan W. Patterson: Introduction to Artificial Intelligence and Expert System, Prentice Hall.
4. Adrian McEwen, Hakim Cassimally, "Designing the Internet of Thing", Wiley
5. 2. Rajkamal, "Internet of Things: Architecture and Design Principles", McGraw Hill Educ
6. Data Mining Techniques ; ArunK.Pujari ; University Press.

PGDA-201: Introduction to Data Analytics

UNIT 1

Descriptive Statistics: Introduction to the Course. Descriptive Statistics ,Probablity Distribution. Inferential Statistics through Hypothesis test, Regression

UNIT 2
Machine Learning: Differentiate Algorithmic and model based framework. Regression :Ordinary Least Square, K- Nearest Neighbours Regression and classification.

UNIT 3
Supervised Learning with Regression and Classification techniques Bias-Variance Dichotomy Model Validation Approaches Logistic Regression Linear Discriminant Analysis Quadratic Discriminant Analysis Regression and Classification Trees Support Vector Machines

UNIT 4
Unsupervised Learning and Challenges for Big Data Analytics Clustering Associative Rule Mining Challenges for big data analytics

UNIT 5
Prescriptive analytics Creating data for analytics through designed experiments Creating data for analytics through Active Learning Creating data for analytics through Reinforcement learning

Text:
1. R. Panneerselvam, "Research Methodologies," PHI.
2. C.R. Kothari: Research methodology, Methods and Techniques, New Age Publication.
3. S.N.Sivanandam ,S.N.Deepa, "Introduction to Neural Networks using MATLAB 6.0", TATA MCGraw- Hill publications William Stallings,"Cryptography and Network ecurity",Third Edition, Pearson Ed

PGDA 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:
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, Labelling and arranging figures, Outputting graphics.

UNIT V:
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.

PGDA- 203: Big data Technology

UNIT I:

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 Map Reduce 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 Map Reduce: Map Reduce Basic Concepts, Understanding the Map Reduce architecture, Writing Map Reduce 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

PGDA-204: Data Mining and NoSQL Concepts

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:
 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 5:
Using Nosql in the Cloud, Scalable Parallel Processing with Map reduce, Analyzing Big Data with Hive, Surveying Database Internals.

Reference Books:

1. Data Mining Techniques ; ArunK.Pujari ; University Press.

2. Data Mining; Adriaans&Zantinge; Pearson education.

3. 1. A Silberschatz, H Korth, S Sudarshan, "Database System and Concepts", fifth Edition McGraw-Hill.
4. ElmasriRamez and NovatheShamkant, "Fundamentals of Database Systems", Benjamin Cummings Publishing. Company.
5. Rob, Coronel, "Database Systems", Seventh Edition, Cengage Learning

PGDA 205: Data Analytics in IOT

UNIT 1

Introduction to IoT Defining IoT, Characteristics of IoT, Physical design of IoT, Logical design of IoT, Functional blocks of IoT, Communication models & APIs

UNIT 2

IoT & M2M Machine Machine, Difference between IoT and M2M, Software define Network

UNIT 3

Network & Communication aspects Wireless medium access issues, MAC protocol survey, Survey routing protocols, Sensor deployment & Node discovery, Data aggregation & dissemination.

UNIT 4

Challenges in IoT Design challenges, Development challenges, Security challenges, Other challenges. **Domain specific applications of IoT** Home automation, Industry applications, Surveillance applications, Other IoT applications

UNIT 5

Developing IoTs Introduction to Python, Introduction to different IoT tools,Developing applications through IoT tools, Developing sensor based application through embedded system platform, Implementing IoT concepts with python

Reference Books:

1. Vijay Madisetti, Arshdeep Bahga, "Internet of Things": A Hands-On Approach
2. Waltenegus Dargie,Christian Poellabauer, "Fundamentals of Wireless Sensor Networks: Theory and Practice"