Department of Computer Science & Applications

School of Mathematical and Physical Sciences (SMPS)

Structure & Syllabus

of

Post Graduate Diploma in Big Data Analytics (PG-DBDA)

1 Year (2 Semesters) Course

Academic Session 2023-2024 and onwards

(Approved by BOS on dated 07/12/2023)

Doctor Harisingh Gour Vishwavidyalaya Sagar

(A Central University)

Sagar – Madhya Pradesh 470003

Chairman, BOS
Computer Science & Applications

Department of Computer Science & Applications

A

Structure & Syllabus

of

Post Graduate Diploma in Big Data Analytics
(PG-DBDA)
1 Year (2 Semesters) Course

Session 2023-2024 and onwards

(Approved by BOS on dated 07/12/2023)

DR. HARISINGH GOUR UNIVERSITY, SAGAR (M.P)

(A Central University)

Chairman, BOS
Computer Science & Applications

Post-Graduate Diploma course on Big-Data Analytics (PG-DBDA)

1. **Name of the Program :** Post-Graduate Diploma course on Big-Data Analytics
2. **Course Duration :**

   Minimum Duration = 01 year

   Maximum Duration – 02 years

3. **Course Objective and Outcomes:**

Post Graduate Diploma in Big Data Analytics (PG-DBDA) programme is a theoretical and practical based course having the following objectives:

CO 1. Explore basic concepts of big data analytics and develop in-depth knowledge and understanding of the big data analytic domain.
CO 2. Learn intelligent techniques to analyze the big data.
CO 3. Understand the various search methods and visualization techniques.
CO 4. Learn and use various techniques for mining data stream.
CO 5. Understand the applications using MapReduce Concepts.
CO 6. Analyze and solve problems conceptually and practically from diverse industries, such as government, manufacturing, retail, education, banking/finance, healthcare and pharmaceutical.
CO 7. Undertake consulting projects with significant data analysis component for better understanding of the theoretical concepts from statistics, economics and related disciplines.
CO 8. Undertake industrial research projects for the development of future solutions in the domain of data analytics to make an impact in the technological advancement.
CO 9. Use advanced analytical tools/ decision-making tools/ operation research techniques to analyze the complex problems and get ready to develop new techniques for the future.
CO 10. Learn Cloud Computing, accessing resources and services needed to perform functions with dynamically changing needs.
CO 11. Understand the cloud privacy and security concepts to create secure cloud environment and explore various cloud platforms to implement real time cloud applications.
CO 12. Learn Clustering, Parallel Programming and HPC solutions and their applications.

After completing this courses students shall be expert in following areas:

PO 1. Big Data tools, Big Data Querying Tools, such as Pig, Hive, and Impala, park
PO 2. Integration of data from multiple data sources

Chairman, BOS
Computer Science & Applications

PO 3. NoSQL databases, such as Hbase, MongoDB
PO 4. Knowledge of various ETL techniques and Implementing ETL Process Monitoring performance and advising any necessary infrastructure changes
PO 5. Developing reports, dashboards using D3 & Tableau.
PO 6. Big Data Development.
PO 7. Business Analyst.

4. Structure of the Program:

PG-DBDA Credit Distribution

Semester	Core Course (CC) Credits	Elective Course (EC) Credits	Total Credits
I	21	00	21
II	18	03	21

5. Eligibility Criterion:

Graduate in Engineering or equivalent (e.g. BE / B Tech / 4-year BSc Engg / AMIE / DoEACC B Level) in IT / Computer Science / Electronics / Telecommunications / Electrical / Instrumentation.

OR

Post Graduate Degree in Science or Engineering Sciences with corresponding basic degree (e.g. MSc in Computer Science, IT, Electronics, Mathematics, Statistics, Physics).

OR

Post Graduate Degree MCA or MBA.

Note: The candidates must have secured a minimum of 55% marks in their qualifying examination.

6. **Medium of Instruction & Examination:** Medium of instruction as well examination will be English only.
7. **Attendance:** Students must secure minimum 75% attendance in each course to appear in the End Semester Examination. If a student fails to secure 75% attendance in a course the he or she will not be allowed to appear in End Semester Examination of respective course. Relaxation may be granted as per University Ordinance.

Chairman, BOS
Computer Science & Applications

8. **Scheme of Examination:**

   a. **Mid Semester Examination: 20 Marks**

   Syllabus and pattern of examination will be decided by the corresponding course instructor(s).

   b. **Internal Assessment : 20 Marks**

   15 marks of internal assessment will be evaluated on any one or more than one methods of the following:

   i. Classroom activities.
   ii. Presentation
   iii. Assignment
   iv. Quizzes
   v. Practical based Test

   Remaining 05 marks will be assigned for attendance. The marks for attendance shall be awarded as follows:

i. 75% and below :	00 Mark
ii. >75% and upto 80% :	01 Mark
iii. > 80% and upto 85% :	02 Marks
iv. > 85% and upto 90% :	03 Marks
v. > 90% and upto 95% :	04 Marks
vi. > 95% :	05 Marks

   c. **End Semester Examination : 60 Marks**

   The END SEMESTER Paper shall be of 60 marks and of 3 hours duration. The pattern of Questions asked shall be as mentioned in the Ordinance No 22.

   Note: A student shall be eligible to appear in End Semester Examination of course if he/she appeared in Mid Semester Examination and Internal Assessment and fulfils the requirement of attendance, failing which he/she will not be permitted to appear in the End Semester Examination of respective course.

   d. **Dissertation Evaluation:**

   The dissertation work is evaluated on the basis of following heads:

i. Mid 1- Presentation & evaluation of Synopsis:	20 Marks
ii. Mid 2- Presentation & evaluation of Progress of work:	20 Marks
iii. End Semester:	
a. Evaluation of Dissertation:	30 Marks
b. Presentation:	15 Marks
c. Viva :	15 Marks

Chairman, BOS
Computer Science & Applications

Dissertation copied from other students will be considered to have used unfair means. If two dissertations are found identical by more than 40% then zero marks will be awarded to both of them. In such a case the dissertations will have to be resubmitted on new topic.

Committee for Evaluation: The evaluation of component II and III is carried out by a committee consisting of the Chairman of BoS or his/her nominee, supervisor (if project is carried out under the supervision of faculty member of the department), two faculty members of the department, and an external examiner invited from other University/Industry/Society /Community/other department within the University.

Components I & IV will be evaluated by a committee consisting of the Chairman of BoS or his/her nominee, supervisor (if project is carried out under the supervision of faculty members of department) and two faculty members of the department.

9. **Credit and Teaching hours:**

The credit and teaching hours shall be distributed as under:

Theory	1- Credit = 15 hours / per semester 2- Credit = 30 hours / per semester 3- Credit = 45 hours / per semester 4- Credit = 60 hours / per semester
Practical	1- Credit = 30 hours / per semester 2- Credit = 60 hours / per semester
Tutorial	1- Credit = 15 hours / per semester

The teacher to student ratio for tutorial/practical can be 1:10 i.e. class will be divided into groups of 10 students for tutorial and practical classes.

Chairman, BOS
Computer Science & Applications

10. Course Structure:

Semester 1									
Course Code	Course Title	Credit	Contact Hrs / Week			Sessional		ESE	Total
			L	T	P	ME	IA		
CSA-CC-131	Basics of Data Science	4	3	1	-	20	20	60	100
CSA-CC-132	Languages & Tools of Data Science	3	2	1	-	20	20	60	100
CSA-CC-133	Applied Statistical Analysis	3	2	1	-	20	20	60	100
CSA-CC-134	Programming with Python	3	2	1	-	20	20	60	100
CSA-CC-135	Cloud Computing & HPC Applications	4	3	1	-	20	20	60	100
CSA-CC-136	R-Language Lab	2	-	-	4	20	20	60	100
CSA-CC-137	Python Programming Lab	2	-	-	4	20	20	60	100
Semester 2									
Course Code	Course Title	Credit	Contact Hrs / Week			Sessional		ESE	Total
			L	T	P	ME	IA		
CSA-CC-231	Hadoop	4	2	-	2	20	20	60	100
CSA-CC-232	Machine Learning	3	2	1	2	20	20	60	100
CSA-CC-233	Data Engineering	3	2	1	2	20	20	60	100
CSA-CC-234	Dissertation	5	-	-	-			100	100
Elective I (Opt Any One of Following)									
CSA-EC-231	Data Visualization - Analysis and Reporting	3	1	1	2	20	20	60	100
CSA-EC-232	Natural-Language Processing	3	2	1	2	20	20	60	100

Summary		
Semester	1	2
Semester-wise Total Credits	21	21
Total Credits	42	

L = Lecture, T = Tutorial, P = Practical.
ME = Mid Examination, IA= Internal Examination, ESE: End Semester Exam.

Chairman, BOS
Computer Science & Applications

PG-DBDA – I Semester

Course Code	Course Title	Credit	Contact Hrs / Week			Sessional		End Term	Total
			L	T	P	MTE	IA		
CSA-CC-131	Basics of Data Science	3	2	-	2	20	20	60	100

Course Objective:

1. To impart the concepts of data structures and DBMS.
2. To understand the concepts OLAP and data preparation for analysis. about searching and sorting techniques.
3. To understand the Big Data and its applications.

Course Contents:

Unit	Topic	Proposed Lectures
I	Data Structure: Linear and non-linear Data Structures. Linked Lists, Stacks and Queues and their application Binary Trees, Basic tree traversals, Binary search tree. Graphs, Data Structures for Graphs, Basic definitions and properties of Graphs, Graph Traversal, Spanning trees, Shortest Paths.	9
II	DBMS: Database Concepts (File System and DBMS), Database Storage Structures (Table space, Control files, Data files), Structured and Unstructured data, SQL Commands (DDL, DML & DCL).	9
III	Data Warehousing & OLAP: Dataware Housing concept, OLTP and OLAP, Data Models - XML, working with MongoDB. Data Preparation: data preparation and cleaning techniques.	9
IV	Introduction to Big Data- Big data definition, enterprise / structured data, social / unstructured data, unstructured data needs for analytics, What is Big Data, Big Deal about Big Data, Big Data Sources, Industries using Big Data, Big Data challenges.	9
V	Case Study: Implement your leanings to find sectors in which different companies ought to invest.	9

Suggested Reading:

1. "Fundamentals of Data Structures", Illustrated Edition by Ellis Horowitz, Sartaj Sahni, Computer Science Press.
2. "Fundamentals of Database Systems", 5th Edition by R. Elmasri and S. Navathe, Pearson Education
3. "Doing Data Science, Straight Talk From The Frontline", Cathy O'Neil and Rachel Schutt, O'Reilly.
4. "Data Mining: Concepts and Techniques", Jiawei Han, Micheline Kamber and Jian Pei.
5. "Introduction to Big Data", Vladimir Bacvanski.
6. "Big Data Technologies and Applications", Borko Furht and Flavio Villanustre.

Course Outcome: Course student will be able to:

- understand storage, retrieval and presentation of data.
- know Big Data and its potential applications.

Chairman, BOS
Computer Science & Applications

PG-DBDA – I Semester

Course Code	Course Title	Credit	Contact Hrs / Week			Sessional		End Term	Total
			L	T	P	MTE	IA		
CSA-CC-132	Languages & Tools of Data Science	3	2	-	2	20	20	60	100

Course Objective:

1. To introduce different languages and tools useful for data science. .
2. Learn R-language as a language for data science.

Course Contents:

Unit	Topic	Proposed Lectures
I	Need of tools and language for Data Sciences. Introduction to different languages and tools used in the course. Brief notes on other languages and tools.	9
II	Introduction & Installation of R, R Basics, Finding Help, Code Editors for R, Command Packages, Manipulating and Processing Data in R.	9
III	Reading and Getting Data into R, Exporting Data from R, Data Objects-Data Types & Data Structure. Viewing Named Objects, Structure of Data Items, Manipulating and Processing Data in R (Creating, Accessing , Sorting data frames, Extracting, Combining, Merging, reshaping data frames), Control Structures, Functions in R (numeric, character, statistical),	9
IV	Working with objects, Constructing Data Objects, Building R Packages, Running and Manipulating Packages, connecting R to mysql, googlesheets in R, sending mail in R.	9
V	Non parametric Tests- ANOVA, chi-Square, t-Test, U-Test, Introduction to Graphical Analysis, Using Plots(Box Plots, Scatter plot, Pie Charts, Bar charts, Line Chart), Plotting variables, Designing Special Plots, Simple Liner Regression, Multiple Regression.	9

Suggested Reading:

1. "R Programming for Beginners" by Sandip Rakshit.
2. "Programming Skills for Data Science: Start Writing Code to Wrangle, Analyze, and Visualize Data with R" by Michael Freeman and Joel Ross.
3. "Hands-On Programming with R: Write Your Own Functions and Simulations" by Garrett Grolemund.
4. "Data Analytics Using R" by Seema Acharya.

Course Outcome: Course student will be able to:

- choose best languages and tools for Analysis of Data.
- write program in R-language and can use different packages available in R.
- perform different non parametric test on different types of data.

Chairman, BOS
Computer Science & Applications

PG-DBDA – I Semester

Course Code	Course Title	Credit	Contact Hrs / Week			Sessional		End Term	Total
			L	T	P	MTE	IA		
CSA-CC-133	Applied Statistical Analysis	3	2	-	2	20	20	60	100

Course Objective:

1. Demonstrate their understanding of descriptive statistics by practical application of quantitative reasoning and data visualization
2. Demonstrate their knowledge of the basics of inferential statistics by making valid generalizations from sample data.

Course Contents:

Unit	Topic	Proposed Lectures
I	Introduction to Statistics- measures of central tendency-, mean, median, mode, weighted average mean, measures of dispersion, types of dispersion, standard deviation, variation, measures of skewness, Basic probability theory.	9
II	Statistical Concepts - uni-variate and bi-variate sampling, distributions, re-sampling, statistical Inference, prediction error	9
III	Probability Distribution (Continuous and discrete- Normal, Bernoulli, Binomial, Negative Binomial, Geometric and Poisson distribution) , Bayes' Theorem, Central Limit theorem,	9
IV	Data Exploration & preparation, Concepts of Correlation, Regression, Covariance, Outliers etc.	9
V	Hypothesis Testing, Hypothesis Testing in Data Analysis, Construction of Confidence Regions and Advanced Concepts.	9

Suggested Reading:

1. "A Student's Guide to R" by Nicholas J. Horton, Daniel T. Kaplan, and Randall P.
2. "Statistics and Data Analysis: From Elementary to Intermediate" by A. C. Tamhane and D. D. Dunlop.
3. "Applied Multivariate Statistical Analysis" by Richard A. Johnson and Dean W Wichern.

Course Outcome: Course student will be able to:

- understand sampling, statistical inference.
- To perform statistical test on data.

Chairman, BOS
Computer Science & Applicatio

PG-DBDA – I Semester

Course Code	Course Title	Credit	Contact Hrs / Week			Sessional		End Term	Total
			L	T	P	MTE	IA		
CSA-CC-134	Programming with Python	3	2	-	2	20	20	60	100

Course Objective:

1. This course introduces core programming basics—including data types, control structures, algorithm development, and program design with functions—via the Python programming language.
2. The course discusses the fundamental principles of Object-Oriented Programming, as well as in-depth data and information processing techniques.

Course Contents:

Unit	Topic	Proposed Lectures
I	Introduction to Python, Basic Syntax, Data Types, Variables, Operators, Input/output, Flow of Control (Modules, Branching), If, If- else, Nested if-else, Looping, For, While, Nested loops, Control Structure, Break, Continue, Pass,	9
II	Strings and Tuples, Accessing Strings, Basic Operations, String slices, Working with Lists, Introduction, Accessing list, Operations, Function and Methods, Files, Modules, Dictionaries, Functions and Functional Programming, Declaring and calling Functions, Declare, assign and retrieve values from Lists, Introducing Tuples, Accessing tuples.	9
III	Advanced Python: Object Oriented, OOPs concept, Class and object, Attributes, Inheritance, Overloading, Overriding, Data hiding, Operations Exception.	9
IV	Exception Handling, Except clause, Try finally clause, User Defined Exceptions, Python Libraries.	9
V	Introduction to Machine learning packages like NUMPY, SCIPY, PANDAS etc	9

Suggested Reading:

1. "The Fundamentals of Python: First Programs" by Kenneth A. Lambert.
2. "Python Machine Learning" by Sebastian Raschka and Vahid Mirjalili.
3. "Artificial Intelligence with Python" by Prateek Joshi.

Course Outcome: Course student will be able to:

- solve problems, explore real-world software development challenges, and create practical and contemporary applications.

Chairman, BOS
Computer Science & Applications

PG-DBDA – I Semester

Course Code	Course Title	Credit	Contact Hrs / Week			Sessional		End Term	Total
			L	T	P	MTE	IA		
CSA-CC-135	Cloud Computing & HPC Applications	3	2	1	-	20	20	60	100

Course Objective: Students will learn

1. This course provides a study of Cloud concepts and capabilities across the various Cloud service models including IaaS, PaaS, SaaS, and BPaaS.
2. The course also covers the Cloud security model and associated challenges and delves into the implementation and support of High Performance Computing and Big Data support capabilities on the Cloud.

Course Contents:

Unit	Topic	Proposed Lectures
I	Introduction to Cloud Computing: Definition, Characteristics, Components, Cloud provider, SAAS, PAAS, IAAS and other Organizational scenarios of clouds, Administering & Monitoring cloud services, benefits and limitations, Deploy application over cloud. Comparison among SAAS, PAAS, IAAS.	9
II	Cloud computing platforms: Infrastructure as service: Amazon EC2, Platform as Service: Google App Engine, Microsoft Azure Utility Computing, Elastic Computing, SLA, clusters, cloud analytics, challenges of cloud environment, HPC in the cloud	9
III	Parallel Processing Concepts: Physical Organization and building blocks of High Performance Computing Systems, Processors and Multi-Core Architectures, Vector processing, Super-scalar, In-order execution, Instruction-Level Parallelism etc., FMA, 32 and 64 bit types, ISA, Accelerators such as GPGPUs and Xeon Phi. Threads and Processes, Multi-processing OS, Parallel I/O, General concepts	9
IV	Parallel Programming Models and Parallel Algorithms Design: Application domains of HPC, Decomposition Techniques: Data parallelism, Functional parallelism, Divide and Conquer etc.,	9
V	Installation (Ubuntu and CentOS), Basics of Linux, Configuring Linux, Shells, Commands, and Navigation, Common Text Editors, Administering Linux, Introduction to Users and Groups, Linux shell scripting, shell computing, Introduction to enterprise computing, Remote access	9

Suggested Reading:

1. Cloud Computing Bible by Barrie Sosinsky.
2. Cloud Computing: Principles and Paradigms, Editors: Rajkumar Buyya, James Broberg, Andrzej M. Goscinski, Wile, 2011
3. Cloud Computing: Principles, Systems and Applications, Editors: Nikos Antonopoulos.
4. Enterprise Cloud Computing Technology Architecture Applications by Gautam Shroff.
5. Cloud Computing Strategies by Dimitris N. Chorafas.

Course Outcome: Students will be able to:

1. apply fundamental concepts in cloud infrastructures and deploy cloud applications that are resilient, elastic and cost-efficient.
2. analyze various cloud programming models and apply them to solve problems on the cloud.

Chairman, BOS
Computer Science & Applications

PG-DBDA – II Semester

Course Code	Course Title	Credit	Contact Hrs / Week			Sessional		End Term	Total
			L	T	P	MTE	IA		
CSA-CC-231	Hadoop	4	2	-	2	20	20	60	100

Course Objective:

1. To impart the concepts and architecture of Hadoop.
2. To understand the techniques and models available in Hadoop.

Course Contents:

Unit	Topic	Proposed Lectures
I	Introduction of Big data programming-Hadoop. The ecosystem and stack, Components of Hadoop, Hadoop Distributed File System (HDFS) and Java interfaces to HDFS. Architecture overview, Development Environment, Hadoop distribution and basic commands, The HDFS command line and web interfaces, Analyzing the Data with Hadoop, Scaling Out, Hadoop event stream processing, complex event processing,	12
II	MapReduce Introduction, Developing a Map Reduce Application, How Map Reduce Works, Anatomy of a Map Reduce Job run, Failures, Job Scheduling, Shuffle and Sort, Task execution, Map Reduce Types and Formats, Map Reduce Features. Hadoop ETL: Hadoop ETL Development, ETL Process in Hadoop, Need of ETL tools, Advantages of ETL tools.	12
III	Hadoop Ingestion Tools: Sqoop Features, Sqoop architecture, Installing Mysql, Install and configure Sqoop on cluster, Connecting to RDBMS,Import data from Mysql to hive, Export data to Mysql,Hive Import .	12
IV	Introduction to Pig and HIVE- Programming Pig: Engine for executing data flows in parallel on Hadoop, Programming with Hive: Data warehouse system for Hadoop, Hive architecture, Optimizing with Combiners and Partitioners (lab), Different type of tables in hive, buckets, partions, Joins in hive. Hbase. Pig basics, Pig Latin scripts, Modes of running PIG, Grunt shell.	12
V	Hadoop Environment: Setting up a Hadoop Cluster, Cluster specification, Cluster Setup and Installation, Hadoop Configuration, Security in Hadoop, Administering Hadoop, HDFS – Monitoring & Maintenance, Hadoop benchmarks, Hadoop in the cloud. Multinode hadoop installation in cloud.	12

Suggested Reading:

1. "Hadoop for Dummies", by Ellis by Dirk deRoos, et al.
2. "Practical Hadoop Ecosystem: A Definitive Guide to Hadoop-Related Frameworks and Tools" by Deepak Vohra.
3. Big Data and Hadoop: Learn by Example by Mayank Bhushan.

Course Outcome: Course student will be able to use the concepts of Hadoop in Big data analysis and programming.

Chairman, BOS
Computer Science & Applications

PG-DBDA – II Semester

Course Code	Course Title	Credit	Contact Hrs / Week			Sessional		End Term	Total
			L	T	P	MTE	IA		
CSA-CC-232	Machine Learning	3	2	1	2	20	20	60	100

Course Objective:

1. To introduce students to the basic concepts and techniques of Machine Learning.
2. To become familiar with machine learning techniques.
3. To become familiar with Dimensionality reduction Techniques.

Course Contents:

Unit	Topic	Proposed Lectures
I	Basics of ML, History of ML, Evolution of ML, ML Models, Learning and testing models, ML Algorithm and Convergence, ML Techniques, Types of ML, supervised and unsupervised learning	9
II	Clustering: K-Means Clustering, Mean Shift Clustering, Aagglomerative clustering, Association Rule Mining, Partition Clustering, Hierarchical Clustering, Birch Algorithm, CURE Algorithm, Density-based Clustering, Gaussian Mixture Models.	9
III	Decision Trees, Oblique trees, Classification problems, Bayesian analysis and Naïve bayes classifier, Random forest, Gradient boosting Machines,	9
IV	Association rules learning, Apriori and FP-growth, Support vector Machines, Linear and Non liner classification, ARIMA, ML in real time,Overfitting and Underfitting	9
V	Neural Networks: back-propagation, backpropagation calculus, Initialization, Training rules, issues in back-propagation, Bayesian Learning	9

Suggested Reading:

1. "Understanding Machine Learning: From Theory to Algorithms", by Shai Shalev-Shwartz, Shai Ben-David.
2. "Data Mining: Practical Machine Learning Tools and Techniques", by Ian H. Witten, Eibe Frank.
3. Tom Mitchell, Machine Learning, McGraw-Hill, 1997

Course Outcome: Course student will be able to:

1. identify machine learning techniques suitable for a given problem.
2. solve the problems using various machine learning techniques.
3. apply Dimensionality reduction techniques.
4. design application using machine learning techniques.

Chairman, BOS
Computer Science & Applications

PG-DBDA – II Semester

Course Code	Course Title	Credit	Contact Hrs / Week			Sessional		End Term	Total
			L	T	P	MTE	IA		
CSA-CC-233	Data Engineering	3	2	1	2	20	20	60	100

Course Objective:

1. Describe in-depth about data, data types, data models and data distributions.
2. Compare and analyzes different data preprocessing and data visualization techniques

Course Contents:

Unit	Topic	Proposed Lectures
I	Understanding Data: Types of data, Data Quality, Data Distributions, Data models, Data Wrangling and Exploratory Analysis, Introduction to contemporary tools. Data mathematics: probability, linear programming.	9
II	Data Preprocessing: Data Transformation & Cleaning, Aggregation, Sampling, Dimensionality reduction, Feature subset Selection, feature creation, PCA, Discretization & Binarization, variable transformation, Data Normalization, Data similarity measures, missing values, filters, Wrapper Method ,Noise reduction techniques.	9
III	Exploring Data: Summary Statistics, measures of location & spread, Multivariate summary statistics, Data Visualization: boxplots, histograms,scatterplots, features map visualization, t-SNE, learning model data visualization,OLAP & Multidimensional Data Analysis.	9
IV	Statistical & Probabilistic analysis of Data, Confusion Matrix, Parameter Estimation methods, Confidence intervals, Correlation & Regression analysis, logistic regression, Shrinkage Methods, Lasso Regression, Bayesian statistics. L1 and L2 regularizations	9
V	Data Analysis, performance metrics, ROC curve,types of errors, Overfitting & Under fitting, evaluating performance of learning model: Holdout, Random sampling, cross validation and Bootstrap method. Bagging & boosting, Gradient Boosting, Random Forests,	9

Suggested Reading:

1. Brian Shive , Data Engineering: A Novel Approach to Data Design, Technics Publications, LLC, NJ, USA, 2013
2. Field Cady, "The Data Science Handbook" , 1/e,Publisher: Wiley, 2018
3. Pang-Ning Tan, "Introduction to Data Mining", Pearson Edu., 2007
4. Peter Bruce, "Practical Statistics for Data Scientists: 50 Essential Concepts",Shroff/O'Reilly; First edition (2017)

Course Outcome: Course student will be able to:

- Examine the nature of a data at hand and determine best suitable data engineering technique/algorithm.
- Solve and implement the real world problems using data engineering.

Chairman, BOS
Computer Science & Applications

PG-DBDA – II Semester

Course Code	Course Title	Credit	Contact Hrs / Week			Sessional		End Term	Total
			L	T	P	MTE	IA		
CSA-EC-231	Data Visualization - Analysis and Reporting	3	1	1	2	20	20	60	100

Course Objective:

1. This course provides an introduction as well as hands-on experience in data visualization.
2. It introduces students to design principles for creating meaningful displays of quantitative and qualitative data to facilitate managerial decision-making.

Course Contents:

Unit	Topic	Proposed Lectures
I	Information Visualization, Data analytics Life Cycle, Analytic Processes and Tools, Analysis vs. Reporting. Creating reports in R	9
II	Modern Data Analytic Tools, Visualization Techniques, Visual Encodings, Visualization algorithms. Shiny dashboard.	9
III	Data collection and binding, Cognitive issues, Interactive visualization,	9
IV	Visualizing big data – structured vs unstructured, Visual Analytics, Geomapping, Dashboard Design.	9
V	Data Visualization using Tableau	9

Suggested Reading:

1. "The Visual Display of Quantitative Information" by E. Tufte.
2. "The Wall Street Journal Guide to Information Graphics: The Dos and Don'ts of Presenting Data, Facts, and Figures" by Dona M. Wong, et al.
3. "Information Dashboard Design: Displaying Data for At-a-Glance Monitoring" by Stephen Few.
4. "Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics" by Nathan Yau.

Course Outcome: Course student will be able to:

1. collect and process data, create an interactive visualization, and use it to demonstrate or provide insight into a problem, situation, or phenomenon.
2. critique various visualizations (good and bad), and to identify design principles that make good visualizations effective.
3. understand some of the challenges present in making data understandable across a wide range of potential audiences.

Chairman, BOS
Computer Science & Applications

PG-DBDA – II Semester

Course Code	Course Title	Credit	Contact Hrs / Week			Sessional		End Term	Total
			L	T	P	MTE	IA		
CSA-EC-232	Natural Language Processing	3	2	1	2	20	20	60	100

Course Objective:

1. Learn text processing fundamentals,
2. Explore machine learning methods in sentiment analysis.
3. Build a speech tagging model

Course Contents:

Unit	Topic	Proposed Lectures
I	INTRODUCTION: Natural Language Processing tasks in syntax, semantics, and pragmatics – Issues - Applications ,Information theory – Collocations -N-gram Language Models - Estimating parameters and smoothing - Evaluating language models	9
II	MORPHOLOGY AND PART OF SPEECH TAGGING Linguistic essentials - Lexical syntax- Morphology and Finite State Transducers - Part of speech Tagging - Rule-Based Part of Speech Tagging - Markov Models - Hidden Markov Models – Transformation based Models - Maximum Entropy Models.	9
III	SEMANTIC ANALYSIS Representing Meaning – Semantic Analysis - Lexical semantics –Word-sense disambiguation - Compositional semantics. Semantic Role Labeling and Semantic Parsing.	9
IV	APPLICATIONS: Named entity recognition and relation extraction-IE using sequence labeling-Machine Translation (MT) - Basic issues in MT-Statistical translation-word alignment- phrase-based translation .	9
V	Case study-Implementation of LDA	9

Suggested Reading:

1. "Daniel Jurafsky and James H.Martin Speech and Language Processing(2nd Edition),Prentice Hall:2 edition,2008.
2. Foundations of Statistical Natural Language Processing by Christopher D.Manning and Hinrich Schuetze,MIT press, 1999
3. "Roland R.Hausser, Foundations of Computational Linguistics:HumanComputer Communication in Natural Language,Paperback,MIT press,2011

Course Outcome: Course student will be able to get insight to text data using machine learning techniques

Chairman, BOS
Computer Science & Applications

PG-DBDA – II Semester

Course Code	Course Title	Credit	Contact Hrs / Week			Sessional		End Term	Total
			L	T	P	MTE	IA		
CSA-CC-235	Project	5	-	-	-			100	100

Project

Student has to complete one project in the supervision of faculty members or an industrial person. In project work they must follow standard Software Process, Coding Standards and Software Quality Assurance.

Chairman, BOS
Computer Science & Applications

Board of Studies Meeting held on 07th Dec., 2023
Approved of PG-DBDA Syllabus for Academic Session 2023-24 onwards

Prof. Vinay Rishiwal
(External Member)
MJP, Rohilkhand University, Bareily
(U.P.)

Prof. R. K. Shrivastava
(External Member)
Dr. Shakuntala Mishra National
Rehabilitation University, Lucknow (U.P.)

Dr. Kavita Rohit
(Member)
Department of Mathematics and Statistics,
Dr. Harisingh Gour V.V., Sagar (M.P.)

Mr. Kamal Kant Ahirwar
(Member)
Department of Computer Science and
Applications,
Dr. Harisingh Gour V.V., Sagar (M.P.)

Dr. Ranjit Rajak
(Special Invitee, Member)
Department of Computer Science and
Applications,
Dr. Harisingh Gour V.V., Sagar (M.P.)

Prof. Ashish Verma
(HoD & Chairman, BoS)
Department of Computer Science and
Applications,
Dr. Harisingh Gour V.V., Sagar (M.P.)

Chairman, BOS
Computer Science & Applications

School Board of Studies Meeting held on 13th December, 2023

Approved of PG-DBDA Syllabus for Academic Session 2023-24 and onwards

Prof. A.K. Saxena,
(External Member)
Department of Mathematics, Maharaja Chhatrasal University, Chhatarpur, MP

online concerted

Prof. K.S. Varsney,
(External Member)
HoD Physics, D.S. College, Aligarh, UP

online concerted

Prof. NarendraPandey,
(External Member)
Department of Physics, University of Lucknow, UP

Prof. DiwakarShukla
(Member)
Department of Mathematics & Statistics, Dr. Harisingh Gour V.V., Sagar

Prof. R.K. Gangele
(Member)
Department of Mathematics & Statistics, Dr. Harisingh Gour V.V., Sagar

Prof. Ranveer Kumar,
(Member)
Department of Physics,
Dr. Harisingh Gour V.V., Sagar

Prof. U.K. Patil,
(Member)
Department Pharmaceutical Science,
Dr. Harisingh Gour V.V., Sagar

Prof. R.K. Rawat
(Member)
Department of Applied Geology, Dr. Harisingh Gour V.V., Sagar

Dr. Mahesh Kumar Yadav
(Member)
Department of Mathematics & Statistics, Dr. Harisingh Gour V.V., Sagar

on leave

Dr.Maheshwar Panda
(Member)
Department of Physics,
Dr. Harisingh Gour V.V., Sagar

Mr. Kamal Kant Ahirwar
(Member)
Department of Comp. Sci. & Application, Dr. Harisingh Gour V.V., Sagar

Prof. AshishVerma
(Dean, SMPS & Chairman, School Board)
Dr. Harisingh Gour V.V., Sagar (M.P.)

विभागाध्यक्ष / Head
कंप्यूटर विज्ञान और अनुप्रयोग विभाग
Department of Computer Science and Applications
डॉ. हरिसिंह गौर विश्वविद्यालय, सागर (म.प्र.)
Dr. Harisingh Gour Vishwavidyalaya, Sagar M.P.