Structure & Syllabi

of

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

1 Year (2 Semesters) Diploma Course

Department of Computer Science and Applications
Dr. Harisingh Gour Vishwavidyalaya Sagar (M.P.)
(A Central University)

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

The Department of Computer Science and Applications proposes a **Post-graduate Diploma course on Big-Data Analytics (PG-DBDA)** to start from the current session 2018-19 .

1. Course Objectives :

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

1.1 Explore basic concepts of big data analytics and develop in-depth knowledge and understanding of the big data analytic domain.
1.2 Learn intelligent techniques to analyze the big data.
1.3 Understand the various search methods and visualization techniques.
1.4 Learn and use various techniques for mining data stream.
1.5 Understand the applications using Map Reduce Concepts.
1.6 Analyze and solve problems conceptually and practically from diverse industries, such as government, manufacturing, retail, education, banking/ finance, healthcare and pharmaceutical.
1.7 Undertake consulting projects with significant data analysis component for better understanding of the theoretical concepts from statistics, economics and related disciplines.
1.8 Undertake industrial research projects for the development of future solutions in the domain of data analytics to make an impact in the technological advancement.
1.9 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.
1.10 Learn Cloud Computing, accessing resources and services needed to perform functions with dynamically changing needs.
1.11 Understand the cloud privacy and security concepts to create secure cloud environment and explore various cloud platforms to implement real time cloud applications.
1.12 Learn Clustering, Parallel Programming and HPC solutions and their applications.

2. Learning Outcome:

After completing this courses students shall capable in following :

2.1 Big Data tools, Big Data Querying Tools, such as Pig, Hive, and Impala, park

2.2 Integration of data from multiple data sources
2.3 NoSQL databases, such as Hbase, MongoDB
2.4 Knowledge of various ETL techniques and Implementing ETL Process
Monitoring performance and advising any necessary infrastructure changes
2.5 Developing reports, dashboards using D3 & Tableau.
2.6 Big Data Development
2.7 Business Analyst.

3. **Course Duration :** It will be of **one year (two semesters**) programme and will be functional as per CBCS rules and regulation laid down in Ordinance 22 of the University.
4. **Intake Capacity :** There shall be enrolment of maximum of 30 students in the proposed PG-DBDA programme in an academic session.
5. **Reservation Policy :** I shall be adopted by the department for UR, EWS, SC, ST categories as followed by the University in other postgraduate programmes.
6. **Eligibility :**

Graduate in Engineering or equivalent (e.g. B.E. / B. Tech. / 4-year B.Sc. 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. M. Sc. 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 before admission.

7. **Admission Procedure:** There shall be advertisement of the course pasted on University Web-Site. Other sources of advertisement may also be followed if required. Students will have to apply on a prescribed format.The admission procedure in the course shall be through a written test and counseling there on as adopted by the University for other PG courses.
8. **Evaluation:**
   (a) There shall be mid-term examination of 20 marks and Internal assessment of 20 marks in each course papers of each semester. End semester examination will be of 60 marks. Guidelines of Ordinance 22 shall be followed wherever needed.
   (b) Project will be evaluated by the one external and one internal examiner.

9. Award of Degree :

(a) Students who clears all the papers of this program, scoring minimum 40% marks in each paper (after completing formalities mentioned in Ordinace 22 and requirement of university) shall be awarded Diploma entitled below:

'POST-GRADUATE DIPLOMA IN BIG DATA ANALYTICS'

(b) Student who fails to secure less than 40% in any course paper shall have to repeat the examination of the said course , registering themselves as backlog (or re-registration) students.

(c) The maximum duration of the course for the award of degree will be 3 years from the registering session of admission.

10. Fee structure (per semester):

S. No.	Heads	Amount
1	Lab fee/ Computer Fee	1000.00
2	Examination Fee	1000.00
3	Tuition Fee	1000.00
4	Library Fee (per semester)	100.00
5	Medical Fee	100.00
6	Registration Fee	100.00
7	Sport Fee	100.00
8	Student Welfare Fund	100.00
9	Insurance Premium	25.00
	TOTAL	**3525.00**

11. Infrastructure Requirement:

To run this course for intake of 30 students ,following infrastructure will be required in terms of Hardware, Software , Human resource space:

(A) Hardware:

1. HPC-
2. Small server
3. Computer: 30
4. Laptop: 5
5. Smart board: |
6. Projector
7. Other hardware required for setting the computing facilities

(B) **Software:**

1 Linux operating system for HPC, Server, and dual booting system for Computer and laptop.
2 M S Office.
3 MS SQL/ Oracle.
4 MATLAB.
5 IBM Analytics.
6 IBM Cognos.
7 Apache Hadoop.
8 MongoDB.
9 Hbase.
10 SAP Business Intelligence Platform

(C) Human Resource:

1. **Faculty:** To run this Post Graduate Diploma Course three faculty members are required. Apart from these three faculty members some industrial experts are also required. These expert may paid' TA/DA and remuneration as per rule.
2. **Technical Expert:** One for installation, practical work, handle hardware and other technical work.
3. **Clerk:** One to take care of official work.
4. **MTS:** Two

(D) SPACE: One Class room, Space for lab, Server Room, tutorial rooms.

12. General Instruction :

Wherever the rules shall be found silent, the guidelines or pattern or practices adopted in other similar curses (or PG courses) and as as adopted in Ordinance 22 shall be considered.

12.1 The medium of instruction shall be English (during examination). However lectures may be in Hindi and English both.

12.2 The minimum 50% listed practicals must have to be completed by the student before appearing in the End Semester Examination.

12.3 Every student has to attain minimum of 75% of attendance in every course of the programme, failing to which the student will be debarred from appearing in the End Semester Examination.

12.4 Scheme of Examination:

(a)	Mid Semester Examination (ME)	: 20 Marks
(b)	Internal Assessment (IA)	: 20 Marks
(c)	End Semester Examination (ESE)	: 60 Marks

12.5 Internal Assessment (IA) :

(a) Theory: Each theory course shall have the methodology of Internal Assessment using assignment, presentation, group discussion, etc. depending on the number of students in the class and feasibility of adopting a particular methodology. The distribution of marks for internal assessment shall be as follows.

(i)	Evaluation of the assignment.	
	Presentation, group discussion etc.	: 15 marks.
(ii)	Attendance	: 05 Marks

The marks of attendance shall be awarded as follows:

(i)	75 % and below	: 00 Marks
(ii)	> 75 % and upto 80 %	: 01 Marks
(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

Note : A student shall be eligible to appear in the End Semester Examination only if he/she has appeared in Mid Semester Examination and in Internal Assessment of the same semester.

(b) Practical/ Lab Courses:

Evaluation of Practical/ Lab Courses shall be as follows:

(i)	Performing and getting the experiment checked regularly and incorporating the suggestions in the practical note book	: 15 marks
(ii**)**	Attendance	: 05 marks

The marks for attendance shall be as follows:

(i)	75 % and below	: 00 Marks
(ii)	> 75 % and upto 80 %	: 01 Marks
(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 for Practical/ Lab Courses :

It will consist of 60 marks as follows:

(a)	Assessment of performance in the experiment	: 50 Marks
(b)	Viva-Voce of Experiment	: 10 Marks

12.6 Evaluation of Minor Project:

If minor project is a full paper (course) then its evaluation shall be based on periodic assessment of the progress of the project and End Semester Examining as follows:

(i) First periodic assessment of the progress after 08 weeks : 20 Marks
(ii) Second periodic assessment after 04 weeks : 20 Marks
(iii) End Semester Examination will consisted of
  a. Evaluation of the project report : 50 Marks
  b. Viva – Voce of the project report : 10 Marks

12.7 Evaluation of Seminars (other than major projects):

If seminar is a full paper them :

a. Documentation for the seminar : 20 Marks
b. First presentation of the seminar : 20 Marks
c. End Semester Examination : 60 Marks

End Semester Examination will consist:

a. Presentation of the seminar : 50 Marks
b. Defend of the presentation : 10 Marks

12.8 The credit and teaching hours shall be distributed as under:

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

12.9 In practical courses students have to maintain a practical file which will be regularly evaluated and signed by the course Incharge/ coordinator.

12.10 The conflict raised, if any, shall be resolved in the meeting of Departmental council or through the guidelines issued by HoD.

12.11 A tutorial shall be an interactive session with students and mode of conduct of tutorial shall be decided by concerned teacher/ course-in-charge.

13. Course –Credit Distribution (PG – DBDA) :

Semester	Core Credit	Elective Credit	Out Dept. Credit	Total
Semester – I	20	00	00	20
Semester – II	17	03	00	20

L = Lecture, T = Tutorial, P = Practical. ME- Mid – I Exam, IA – Internal Assessment, ESE – End Semester Exam.

14. Course Structure :

Semester -I									
Course Code	Course Title	L	T	P	C	Sessional		ESE	Total
						ME	IA		
CSA-CC-131	Basics of Data Science	03	01	-	04	20	20	60	100
CSA-CC-132	Languages & Tools of Data	02	01	-	03	20	20	60	100
CSA-CC-133	Applied Statistical Analysis	02	01	-	03	20	20	60	100
CSA-CC-134	Programming with Python	02	01	-	03	20	20	60	100
CSA-CC-135	Cloud Computing & HPC Applications	02	01	-	03	20	20	60	100
Software laboratory (Practicals)									
CSA-CC-136	Software Laboratory – A	-	-	02	02	20	20	60	100
CSA-CC-137	Software Laboratory – B	-	-	02	02	20	20	60	100
	TOTAL	**11**	**05**	**04**	**20**	**-**	**-**	**-**	**700**

Semester –II									
Course Code	Course Title	L	T	P	C	Sessional		ESE	Total
						ME	IA		
CSA-CC-231	Hadoop	02	01	-	03	20	20	60	100
CSA-CC-232	Data Visualization - Analysis and Reporting	02	01	-	03	20	20	60	100
CSA-CC-233	Machine Learning	02	01	-	03	20	20	60	100
CSA-CC-234	Project	-	-	-	06	20	20	60	100
Elective I (Opt Any One of Following)									
CSA-EC-231	Advanced Analytics	03	-	-	03	20	20	60	100
CSA-EC-232	Cryptography and Security	03	-	-	03	20	20	60	100
Software laboratory (Practical)									
CSA-CC-235	Software Laboratory – A	-	-	02	02	20	20	60	100
	TOTAL	**09**	**03**	**02**	**20**	**-**	**-**	**-**	**600**

DETAILED SYLLABUS

PG-DBDA (Sem.-I)

Course Code	Course Title	L	T	P	C	Sessional		ESE	Total
						ME	IA		
CSA-CC-131	**Basics of Data Science**	03	01	-	04	20	20	60	100

Objectives :

(i) To introduce the basic fundamentals of data structure and storage.

(ii) To make aware about features of big-data.

UNIT - I	**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) (**12 Hours)**
UNIT -II	**Data Warehousing & OLAP:** Dataware Housing concept, OLTP and OLAP, Data Models - XML, working with Mongo DB. (**12 Hours)**
UNIT -III	**Data Preparation:** data preparation and cleaning techniques (**12 Hours)**
UNIT -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. (**12 Hours)**
UNIT -V	**Case Study:** Implement your leanings to find sectors in which different companies ought to invest. (**12 Hours)**

Learning Outcomes : After completion of this course, students will be able to visualize the big data generation with features contained therein.

Essential Reading:

1. Fundamentals of Database System by by Elmasri Ramez , Navathe Shamkant, Pearson seventh edition.
2. Big Data Fundamentals: Concepts Drivers: Concepts, Drivers and Techniques by by Erl/Khattak/Buhler.

Suggested Reading with links :

1. Kamber and Han, "Data Mining Concepts and Techniques", Hartcourt India P. Ltd., 2001
2. Principles of Data Mining, Handa : (Pearson Education India).
3. https://nptel.ac.in/courses/106105174/
4. https://epgp.inflibnet.ac.in/ahl.php?csrno=7
5. http://ptgmedia.pearsoncmg.com/images/9780134291079/samplepages/97801342910 79.pdf

Approved by BoS on dated 05-07-2019.

PG-DBDA (Sem.-I)

Course Code	Course Title	L	T	P	C	Sessional		ESE	Total
						ME	IA		
CSA-CC-132	**Languages & Tools of Data Science**	02	01	-	03	20	20	60	100

Objectives :

To teach students about statistical data analysis using R-software.

UNIT - I	**R Programming:** Introduction & Installation of R, R Basics, Finding Help, Code Editors for R, Command Packages, Manipulating and Processing Data in R, Reading and Getting Data into R, Exporting Data from R, Data **(9 Hours)**
UNIT -II	Objects- Data Types & Data Structure. Viewing Named Objects, Structure of Data Items, **(9 Hours)**
UNIT -III	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 Hours)**
UNIT -IV	Working with objects, Viewing Objects within Objects, Constructing Data Objects, Building R Packages, Running and Manipulating Packages **(9 Hours)**
UNIT -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 Hours)**

Learning Outcomes :

After completing this course student will be able to perform data-analysis and conclusion writing using statistical methodologies. .

Essential Reading:

1. Data Analytics Using R Paperback by Seema Acharya Mc Graw Hill Education.

Suggested Reading with links :

1. https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

Approved by BoS on dated 05-07-2019.

PG-DBDA (Sem.-I)

Course Code	Course Title	L	T	P	C	Sessional		ESE	Total
						ME	IA		
CSA-CC-133	**Applied Statistical Analysis**	02	01	-	03	20	20	60	100

Objectives :

To make aware students about basic features of theoretical and applied statistics.

UNIT - I	Introduction to Statistics-Descriptive Statistics, Summary Statistics Basic probability theory (**9 Hours)**
UNIT -II	Statistical Concepts (uni-variate and bi-variate sampling, distributions, re-sampling, statistical Inference, prediction error) (9 **Hours)**
UNIT -III	Probability Distribution (Continuous and discrete- Normal, Bernoulli, Binomial (**9 Hours)**
UNIT -IV	Negative Binomial, Geometric and Poisson distribution) . (**9 Hours)**
UNIT -V	Bayes' Theorem, Central Limit theorem, Data Exploration & preparation, Concepts of Correlation, Regression, Covariance, Outliers etc. (**9 Hours)**

Learning Outcomes : After completion of this course students will be able to apply various statistical techniques available for data analysis.

Essential Reading:

1. Fundamental of Mathematical Statistics-S C Gupta & V K Kapoor.
2. Probability and Statistics with Reliability, Queuing, and Computer Science Applications Hardcover by Kishor S. Trivedi.

Approved by BoS on dated 05-07-2019.

PG-DBDA (Sem.-I)

Course Code	Course Title	L	T	P	C	Sessional		ESE	Total
						ME	IA		
CSA-CC-134	**Programming with Python**	02	01	-	03	20	20	60	100

Objectives :

(i) To learn the fundamentals of Python programming.

(ii) To develop skill of writing computer program in Python.

UNIT - I	Introduction to Python, Basic Syntax, Data Types, Variables, Operators, Input/ output, Flow of Control (Modules, Branching), (**9 Hours)**
UNIT -II	If, If- else, Nested if- else, Looping, For, While, Nested loops, Control Structure, Break, Continue, Pass, Strings and Tuples, Accessing Strings (**9 Hours)**
UNIT -III	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 Hours)**
UNIT -IV	**Advanced Python:** Object Oriented, OOPs concept, Class and object, Attributes, Inheritance, Overloading, Overriding, Data hiding, Operations Exception, Exception Handling, Except clause, Try finally clause (**9 Hours)**
UNIT -V	User Defined Exceptions, Python Libraries. Introduction to Machine learning packages like NUMPY, SCIPY, PANDAS etc: (**9 Hours)**

Learning Outcomes : After completing this course, student will be able to write computer program using Python language for real life problem solving.

.

Essential Reading:

1. Probability and Statistics with Reliability, Queuing, and Computer Science Applications Hardcover, by Kishor S. Trivedi.
2. Programming and Problem Solving with Python by Ashok Namdev Kamthane and Amit Ashok Kamthane

Approved by BoS on dated 05-07-2019.

PG-DBDA (Sem.-I)

Course Code	Course Title	L	T	P	C	Sessional		ESE	Total
						ME	IA		
CSA-CC-135	**Cloud Computing & HPC Applications**	02	01	-	03	20	20	60	100

Objectives :

To make aware students about cloud computing, parallel processing, load balancing and features of high power computing.

UNIT - 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 (**9 Hours)**
UNIT -II	Deploy application over cloud. Comparison among SAAS, PAAS, IAAS, 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 Hours)**
UNIT -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 Hours)**
UNIT -IV	**Parallel Programming Models and Parallel Algorithms Design:** Application domains of HPC, Decomposition Techniques: Data parallelism, Functional parallelism, Divide and Conquer etc. (**9 Hours)**
UNIT -V	Characteristics of Tasks and Interactions, Mapping Techniques for Load Balancing, Methods for Containing Interaction Overheads, Granularity of parallelism, Programming Open MP (**9 Hours)**

Learning Outcomes : After completing this course student will learn the tools, techniques and applications used in cloud computing.

Essential Reading:

1. Cloud Computing: Principles and Paradigms by Rajkumar Buyya, James Broberg, et al
2. Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online, by Miller.
3. Cloud Computing "A Practical Approach" Anthony T. Velte, Toby J. Velte, Robert
4. Elsenpeter. McGraw-Hill. Kai Hwang, Geoffrey C Fox, Jack G Dongarra, "Distributed and Cloud Computing, From Parallel• Processing to the Internet of Things", Morgan Kaufmann Publishers, 2012.
5. John W.Rittinghouse and James F.Ransome, "Cloud Computing: Implementation,• Management, and Security", CRC Press, 2010.

Approved by BoS on dated 05-07-2019.

PG-DBDA (Sem.-I)

Course Code	Course Title	L	T	P	C	Sessional		ESE	Total
						ME	IA		
CSA-CC-136	**Software Laboratory – A**	-	-	02	02	20	20	60	100

Objectives : To provide software based training to students in order to write programmes and to develop capacity in handling the softwares related to big data.

List of practicals will be decided by the course coordinator.

Learning Outcomes : After completing this course student will be capable enough to work with softwares of big data.

Approved by BoS on dated 05-07-2019.

PG-DBDA (Sem.-I)

Course Code	Course Title	L	T	P	C	Sessional		ESE	Total
						ME	IA		
CSA-CC-137	**Software Laboratory – B**	-	-	02	02	20	20	60	100

Objectives : To provide software based training to students in order to write programmes and to develop capacity in handling the softwares related to big data.

List of practicals will be decided by the course coordinator.

Learning Outcomes: After completing this course student will be capable enough to work with softwares of big data.

Approved by BoS on dated 05-07-2019.

PG-DBDA (Sem.-II)

Course Code	Course Title	L	T	P	C	Sessional		ESE	Total
						ME	IA		
CSA-CC-231	**Hadoop**	02	01	-	03	20	20	60	100

Objectives :

To introduce knowledge about tools, techniques, methods and environment related to Hadoop technology used for big-data.

UNIT - I	Introduction of Big data programming-Hadoop, History of Hadoop, The ecosystem and stack, Components of Hadoop, Hadoop Distributed File System (HDFS), Design of HDFS, Java interfaces to HDFS, Architecture overview, Development Environment, Hadoop distribution and-basic commands, Eclipse development. (**9 Hours)**
UNIT -II	The HDFS command line and web interfaces, The HDFS Java API (lab), Analyzing the Data with Hadoop, Scaling Out, Hadoop event stream processing, complex event processing, MapReduce Introduction, Developing a Map Reduce Application, How Map Reduce Works, The MapReduce Anatomy of a Map Reduce Job run, Failures, Job Scheduling, Shuffle and Sort, Task execution, Map Reduce Types and Formats, Map Reduce Features, Real-World MapReduce. (**9 Hours)**
UNIT -III	**Hadoop ETL:** Hadoop ETL Development, ETL Process in Hadoop, Discussion of ETL functions, Data Extractions, Need of ETL tools, Advantages of ETL tools **Hadoop Reporting Tools:** Jaspersoft (reporting and analytics server), Pentaho (data integration and business analytics), Splunk (platform for IT analytics), Talend (big data integration, data management and application integration). (**9 Hours)**
UNIT -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, Optimizing with Combiners and Practitioners (lab), More common algorithms: sorting, indexing and searching (lab), Relational manipulation: map-side and reduce-side joins (lab), evolution, purpose and use, HDFS. (**9 Hours)**
UNIT -V	Overview and concepts, data flow (read and write), interface to HDFS (HTTP, CLI and Java API), high availability and Name Node federation, Map Reduce developing and deploying programs, optimization techniques, Map Reduce Anatomy, Data flow framework programming Map Reduce best practices and debugging, Introduction to Hadoop ecosystem, integration R with Hadoop. **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. (**9 Hours)**

Learning Outcomes : After completing this course, student will be able to handle the big data in Hadoop platform.

Essential Reading:

1. Hadoop: The Definitive Guide, 4th Edition by Tom White

Approved by BoS on dated 05-07-2019.

PG-DBDA (Sem.-II)

Course Code	Course Title	L	T	P	C	Sessional		ESE	Total
						ME	IA		
CSA-CC-232	**Data Visualization - Analysis and Reporting**	02	01	-	03	20	20	60	100

Objectives:

To understand tools and techniques relating to data visualization and analysis.

UNIT - I	Information Visualization, Data analytics Life Cycle, Analytic Processes and Tools, Analysis vs. Reporting (**9 Hours)**
UNIT -II	Modern Data Analytic Tools, Visualization Techniques, Visual Encodings, Visualization algorithms (**9 Hours)**
UNIT -III	Data collection and binding, Cognitive issues, Interactive visualization(**9 Hours)**
UNIT -IV	Visualizing big data — structured vs unstructured . (**9 Hours)**
UNIT -V	Visual Analytics, Geomapping, Dashboard Design (**9 Hours)**

Learning Outcomes : After completing this course student will be able to understand pattern of data with analysis and reporting.

Essential Reading:

1. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures Paperback by Claus O. Wilke.

Approved by BoS on dated 05-07-2019.

PG-DBDA (Sem.-II)

Course Code	Course Title	L	T	P	C	Sessional		ESE	Total
						ME	IA		
CSA-CC-233	**Machine Learning**	02	01	-	03	20	20	60	100

Objectives:

To learn tools and techniques used in machine learning.

UNIT - I	Supervised and Unsupervised Learning , Uses of Machine learning, Clustering, K means (**9 Hours)**
UNIT -II	Hierarchical Clustering, Decision Trees, Oblique trees (**9 Hours)**
UNIT -III	Classification problems, Bayesian analysis and Naive bayes classifier (**9 Hours)**
UNIT -IV	Random forest, Gradient boosting Machines, Association rules learning, Apriori and FP-growth algorithms, Support vector Machines . (**9 Hours)**
UNIT -V	Linear and Non _ liner classification, ARIMA, ML in real time, Neural Networks and its application. (**9 Hours)**

Learning Outcomes : After completing this course student will be able to apply machine learning tools over real life problem solving, using data analysis outcomes.

Essential Reading:

1. Machine Learning: Fundamental Algorithms for Supervised and Unsupervised Learning With Real-world Applications by Joshua Chapmann
2. Fundamentals of Machine Learning for Predictive Data Analytics-Algorithms, Worked Examples, and Case Studies (The MIT Press) by John D. Kelleher , Brian Mac Name, et. al.

Approved by BoS on dated 05-07-2019.

PG-DBDA (Sem.-II)

Course Code	Course Title	L	T	P	C	Sessional		ESE	Total
						ME	IA		
CSA-CC-234	**Project**	-	-	-	06	20	20	60	100

Objectives:

To provide project based analytics training to student on a problem.

Student has to complete one project in the supervision of faculty member or with an industrial person. In project work, they must follow standard Software Process, Coding Standards and Software Quality Assurance. Project may be department-based or industry/out-deptt. based with two periodic assessments, each of 20 marks by the course coordinator. The end sem. project evaluation will be by internal and external examiner in MM 60 along with power point presentation & discussion / viva-voce.

Learning Outcomes : After cmpleting this course, student will be able to perform data analysis and conclusion writing after analysis of the specific assigned problem.

Approved by BoS on dated 05-07-2019.

PG-DBDA (Sem.-II) Elective –I

Course Code	Course Title	L	T	P	C	Sessional		ESE	Total
						ME	IA		
CSA-EC-231	**Advanced Analytics**	03	-	-	03	20	20	60	100

Objectives:

To make aware students about methodologies frequently used in big-data analytics.

UNIT - I	Introduction to Business Analytics using some case studies, Making Right Business Decisions based on data. (**6 Hours)**
UNIT -II	Exploratory Data Analysis- Visualization and Exploring Data, Descriptive Statistical Measures (**6 Hours)**
UNIT -III	Probability Distribution and Data, Sampling and Estimation, Statistical Interfaces, Predictive modeling and analysis. (**6 Hours)**
UNIT -IV	Regression Analysis, Forecasting Techniques, Simulation and Risk Analysis, Optimization, Linear, Non linear, Integer, Decision Analysis, Strategy and Analytics (**6 Hours)**
UNIT -V	Overview of Factor Analysis, Directional Data Analytics, Functional Data Analysis (**6 Hours)**

Learning Outcomes : After completing this course student will be able to perform big-data analytic task using statistical tools.

Essential Reading:

1. AI and Analytics, Accelerating Business Decisions by Sameer Dhanrajani
2. Advanced Analytics with R and Tableau by Jen Stirrup and Ruben Oliva Ramos

Approved by BoS on dated 05-07-2019.

PG-DBDA (Sem.-II) Elective –II

Course Code	Course Title	L	T	P	C	Sessional		ESE	Total
						ME	IA		
CSA-EC-232	**Cryptography and Security**	03	-	-	03	20	20	60	100

Objectives : (1) To introduce the concept of data security any cryptography.
(2) To learn various mythologies used in secures of computer system and data.

UNIT- I	Introduction to security attacks, services and mechanism, introduction to cryptography. Conventional Encryption: Conventional encryption model, classical encryption techniques substitution ciphers and transposition ciphers, cryptanalysis, stenography, stream and block ciphers. Modern Block Ciphers: Block ciphers principals, Shannon's theory of confusion and diffusion, fiestal structure, data encryption standard(DES), strength of DES, differential and linear cryptanalysis of DES, block cipher modes of operations, triple DES. **(9 Hours)**
UNIT -II	Introduction to graph, ring and field, prime and relative prime numbers, modular arithmetic, Fermat's and Euler's theorem, primality testing, Euclid's Algorithm, Chinese Remainder theorem, discrete logarithms. Principals of public key crypto systems, RSA algorithm, security of RSA, key management, Diffle-Hellman key exchange algorithm, introductory idea of Elliptic curve cryptography, Elgamel encryption. **(9 Hours)**
UNIT -III	Message Authentication and Hash Function: Authentication requirements, authentication functions, message authentication code, hash functions, birthday attacks, security of hash functions and MACS, MD5 message digest algorithm, Secure hash algorithm(SHA). Digital Signatures: Digital Signatures, authentication protocols, digital signature standards (DSS), proof of digital signature algorithm. **(9 Hours)**
UNIT -IV	Authentication Applications: Kerberos and X.509, directory authentication service, electronic mail security-pretty good privacy (PGP), S/MIME. **(9 Hours)**
UNIT- V	IP Security: Architecture, Authentication header, Encapsulating security payloads, combining security associations, key management. Web Security: Secure socket layer and transport layer security, secure electronic transaction (SET). System Security: Intruders, Viruses and related threads, firewall design principals, trusted systems. **(9 Hours)**

Learning Outcomes : After completion of this course, student will be able to learn the security features used in computer system.

Essential Reading:

1. William Stallings, "Cryptography and Network Security: Principals and Practice", Prentice Hall, New Jersy.
2. Johannes A. Buchmann, "Introduction to Cryptography", Springer-Verlag.

Suggested Reading and links :

1. Bruce Schiener, "Applied Cryptography".
2. Behrouz A. Forouzan, "Cryptography and Network Security",
3. https://nptel.ac.in/courses/106105080/pdf/M8L1.pdf
4. http://textofvideo.nptel.ac.in/106105031/lec2.pdf
5. https://epgp.inflibnet.ac.in/ahl.php?csrno=7

Approved by BoS on dated 05-07-2019

PG-DBDA (Sem.-II)

Course Code	Course Title	L	T	P	C	Sessional		ESE	Total
						ME	IA		
CSA-CC-235	**Software Laboratory – A**	-	-	02	02	20	20	60	100

Objectives : To provide software based training to students in order to write programmes and to develop capacity in handling the softwares related to big data.

List of practicals will be decided by the course coordinator.

Learning Outcomes : After completing this course student will be capable enough to work with softwares handling big data.