### Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **1st SEMESTER**

# **BSCDA-101: ENGLISH**

Objectives

To enable the students to:

• Gain the knowledge of English language and also the ability to apply them in the enterprise reality

Units	Course Content	
1	<b>Grammatical Focus :</b> Grammatical & Structural aspects covering Parts of Speech, Tense, Voice, Clause, Preposition, Degrees of Comparison, Synonyms & Antonyms, etc; Identifying & Analyzing Grammatical Errors including errors in Spelling & Punctuation.	
2	Reading : Vocabulary Building; Comprehension; Interpretation; Summarizing	
3	Writing : Letter Writing – Formal, Informal; Accepting & Declining Invitations; Paragraph Writing, Precise Writing, Essay Writing	
4	<b>Speaking :</b> Interactive Communication like Introducing Self, Greetings, Conversations, etc; Pronunciation : appropriate stress, intonation, clarity	
5	Listening : Understanding – Spoken English, Formal English; Exercises	

References

- Leo Jones, Richard Alexander : New International Business English (Communication Skills in English for Business Purposes), Cambridge University Press.
- NCERT, Knowing about English A Book of Grammar & Phonology National Building Code of India, Bureau of Indian Standards, New Delhi, 1999
- NCERT, Working with English A Workbook,
- A.E. Augustine & K.V. Joseph : Macmillan Grammar A Handbook, Macmillan
- Krishna Mohan & N.P. Singh : Speaking English Effectively, Macmillan
- Kulbhushan Kumar, Effective Communication Skills, Khanna Publishing House, New Delhi

(Formerly West Bengal University of Technology)

# Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **BSCDA – 102: DISCRETE MATHEMATICS**

Objectives	
To enable the s	students to:
Develop Ma	athematical thinking in Computer Science.
• Understand	ing Combinatorics and Probability.
• Understand	the concept of Graph theory.
Understand	number theory and cryptography.
Units	Course Content
1	Mathematical thinking in Computer Science Induction, Recursion, Logic, Invariants, Examples, Optimality
2	<b>Combinatorics and Probability</b> Basic Probability and Venn diagram, Compound Probability of independent events, Dependant events, Permutations and Combinations, Probability using Combinatorics
3	Introduction to Graph theory Fundamental Concepts and Basic Results, Graph Isomorphism, Subgraphs, the Complement of a Graph, Bipartite Graphs and Trees, Vertex-Colourings of Graphs, Matchings in Bipartite Graphs, Eulerian Multigraphs and Hamiltonian Graphs, Digraphs and Tournaments
4	Number theory and Cryptography RSA, key generation, encryption, decryption, cryptographic hash, signing messages
5	Delivery Problem Proof techniques, P vs NP problem
	•
References	
1 Discrete Mathematics with Graph theory and Combinatorics, Py T Vegraraian	
<ol> <li>Discrete Mamematics with Graph theory and Combinatorics, By 1 Veerarajan.</li> <li>Discrete Structures, S.B. Singh</li> </ol>	
3. Discrete Mathematics and Graph Theory, By Bhavanari S	
4. Discrete Mathematics, By Dhami and Bisht	

- 5. Discrete Mathematics and its Applications, By Kenneth Rosen
- 6. Combinatorics & Graphic Theory by S.B. Singh

#### Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **BSCDA – 103: INTRODUCTION TO STATISTICAL THEORY AND METHODS**

#### Objectives

To enable the students to:

• Enable the students to understand the key concepts of Statistics

Units	Course Content
1	Introduction to Statistics: Introduction, Data presentation- Frequency table, histogram, Bar chart and frequency polygons, stem and leaf plots, measure of location and spread, box and whisker plots
2	Introduction: Definition and scope of Statistics, concepts of statistical population and sample. Data: quantitative and qualitative, attributes, variables, scales of measurement - nominal, ordinal, interval and ratio. Presentation: tabular and graphic, including histogram and ogives. Measures of Central Tendency: mathematical and positional. Measures of Dispersion: range, quartile deviation, mean deviation, standard deviation, coefficient of variation, moments, skewness and kurtosis.
3	<ul><li>Bivariate data: Definition, scatter diagram, simple, partial and multiple correlation (3 variables only), rank correlation. Simple linear regression, principle of least squares and fitting of polynomials and exponential curves.</li><li>Theory of attributes, consistency of data, independence and association of attributes, measures of association and contingency.</li></ul>
Deferences	

- References
- 1. Goon A.M., Gupta M.K. and Dasgupta B. (2002): Fundamentals of Statistics, Vol. I & II, 8th Edn. The World Press, Kolkata.
- 2. Miller, Irwin and Miller, Marylees (2006): John E. Freund's Mathematical Statistics with Applications, (7th Edn.), Pearson Education, Asia
- 3. Manish Sharma & Amit Gupta, The Practice of Business Statistics, Khanna Publishing House, New Delhi

(Formerly West Bengal University of Technology)

#### Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **BSCDA – 104: FOUNDATION OF COMPUTER SCIENCE**

# Objectives To enable the students to: Gain the knowledge of computer application in theory and practice and also the ability to apply them in the enterprise reality Units **Course Content** 1 Basic Computer Concepts – Different generations of computer hardware; Modern taxonomy of computers; Hardware and software; Programming languages; Problem solving and algorithms; Basic computer applications; General idea of information and communication technologies; Information system development process. 2 Computer Hardware - Input and Output devices; Memory (or storage) devices; Central Processing Unit. Input / Output devices: keyboard, mouse, light pen, barcode readers, scanners, MICR, OCR, voice recognition and handwriting recognition systems; visual display terminals, printers, plotters etc. Storage devices: Primary storage - RAM, ROM, EEROM, PROM, EPROM; Secondary storage direct access devices, serial access devices: hard disks, floppy disks, magnetic tape, CD-ROM, DVD; Cache memory and Virtual memory. Central Processing Unit - Control Unit; Arithmetic and Logic Unit; Decoders; Registers; Machine Instructions; Stored program concept; Program execution: Fetch-Decode-Execute cycle; Arithmetic, logical and shift operations. Computer Software - Meaning of software; broad classification of software; system 3 software and application software; utilities. Systems software - Operating systems: Basic idea of an OS; OS as a resource manager memory management, input/output management, secondary storage management, processor management, program management, network management; Brief introduction to different types of operating systems like DOS, Windows, Unix, Linux etc. Application software - System development tools, Utilities, Application packages, Userwritten programs.

(Formerly West Bengal University of Technology)

#### Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

4	<b>Programming languages and Algorithms</b> – The concept of programming; pseudocode and flowcharts; structure of programs; program development guidelines; programming languages – machine language, assembly languages, high-level languages (procedural and object-oriented languages), fourth generation languages; object code and executable codes; compilers, translators, assemblers; editing tools such as vi. Algorithms – Basic concept; Some typical algorithms – Finding the sum of a series, solving a quadratic equation, generating Fibonacci sequence, checking whether a number is prime or not, creating an array of numbers and displaying the largest element in the list, sorting a given set of numbers, multiplying together two matrices etc. (The algorithms may be implemented using either pseudocode or a high level programming language).
5	<b>Computer networks and Internet</b> – Basic concepts of computer networks; local area networks and wide area networks; switches, hubs, routers, idea of distributed systems; the Internet and the World Wide Web.
6	<b>Computer Applications:</b> Essential features of computer systems and structures required for office automation, communications, control systems, data acquisition, interactive multimedia, networking, parallel processing and neural networks.
References	
<ol> <li>Mano – Computer System Architecture; Pearson Education</li> <li>Tanenbaum – Structured Computer Organization, Pearson Education</li> </ol>	

- 3. Martin & Powell Information Systems: A Management Perspective; McGraw-Hill
- 4. Laudon & Laudon Management Information Systems: Pearson Education
- 5. Comer: Computer Networks and the Internet: Pearson Education
- 6. Salaria Computer Fundamentals: Khanna Publishing House
- 7. Graham Curtis Business Information Systems: Addison Wesley

(Formerly West Bengal University of Technology)

# Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **BSCDA – 105: DATABASES MANAGEMENT SYSTEM**

Objectives		
To enable the students to:		
• Understand database.	Understand Relational database, its functions and utility and difference between Relational and normal database.	
Units	Course Content	
1	Introduction to Database Systems Concept & Overview of DBMS, Data Models, Database Languages, Database Administrator, Database Users, Three Schema architecture of DBMS	
2	Entity-Relationship Model Basic concepts, Design Issues, Mapping Constraints, Keys, Entity-Relationship Diagram, Weak Entity Sets, Extended E-R features.	
3	<b>SQL and Integrity Constraints</b> Concept of DDL, DML, DCL. Basic Structure, Set operations, Aggregate Functions, Null Values, Domain Constraints, Referential Integrity Constraints, assertions, views, Nested Subqueries, Database security application development using SQL, Stored procedures and triggers.	
4	<b>Relational Database Design</b> Functional Dependency, Different anamolies in designing a Database., Normalization using funtional dependencies, Decomposition, Boyce-Codd Normal Form, 3NF, Nomalization using multi-valued depedencies, 4NF, 5NF	
5	<b>Structure Query Language and Advance SQL Programming</b> Data Manipulation Language on SQL programming to feed and query relation database objects.	
6	<b>Transaction and Query Optimisations</b> Transaction in Sql and query optimisation	
7	<b>Data Warehousing Concepts</b> Introduction to Data Warehousing, heterogeneous database, Query-Driven Approach, Update-driven approach, tools and utilities	
8	<b>Introduction to NoSql</b> Introduction to NoSql, types of NoSql, Advantages of NoSql, Usage	

(Formerly West Bengal University of Technology)

# Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

#### References

- 1. Relational Database Management System, By Pankaj Agarwal
- 2. Inside Relational Databases, By Mark Whitehorn and Bill Marklyn

(Formerly West Bengal University of Technology)

# Syllabus for B. Sc. in Data science (Effective for Academic session 2019-20)

# **2nd SEMESTER**

# **BSCDA- 201: SOFT-SKILLS FOR BUSINESS**

Objectives	Objectives	
To enable the students to:		
• Developb	oothoralandwrittencommunicationskillsrelatingtoorganizationaland Business issues	
Units	Course Content	
1	ELEMENTS OF COMMUNICATION Meaning, Importance, Objectives & Principles of Communication, Process, impediments of effective communication, Strategies for effective communication. Types and forms of communication NonverbalCommunication-BodyLanguage,Gestures,Postures,FacialExpressions,Dresscodes, The Cross Cultural Dimensions of Business Communication, Listening & Speaking, Techniques of Eliciting Response, Probing Questions, Observation, Business and social etiquette.	
2	PUBLIC SPEAKING ImportanceofPublicSpeakingandSpeechComposition- PrinciplesofEffectiveSpeaking&Presentations. Technicalspeeches&Non- technicalpresentations.Speechforintroductionofaspeaker-Speechforvote of thanks-Occasional speech-Theme speech. Moderating programs-Use of Technology	
3	<b>INTERVIEW TECHNIQUES</b> Importance of Interviews, Art of conducting and giving interviews, Placement interviews- discipline interviews – Appraisal interviews– Exit interviews.	
4	MEETINGS ImportanceofMeetings-OpeningandClosingMeetings- ParticipatingandConductingGroupdiscussions. Brain Storming, e– Meetings, preparing agenda and minutes of the meeting	
5	BUSINESS COMMUNICATION BusinessLetters:Inquiries,Circulars,Quotations,Orders,AcknowledgmentsExecutions,Complaints, Claims &Adjustments, Collection letter, Banking correspondence, Agency correspondence, Bad news and persuadingletters,Salesletters,Jobapplicationletters-Bio- data,CoveringLetter,InterviewLetters,Letter of Reference. Memos, Minutes, Circulars & Notices.	

(Formerly West Bengal University of Technology)

#### Syllabus for B. Sc. in Data science (Effective for Academic session 2019-20)

6		SKILL DEVELOPMENT         Conduct a mock meeting and draft minutes of the meeting.         Draft a letter of enquiry to purchase a laptop.         Draft your bio-data.         Prepare your Career Plan.
1		
Refe	erences	
1. H	1. Kulbhushan Kumar, Effective Communication Skills, Khanna Publishing House	
2. I	. Rai & Rai–Soft Skill for Business, HPH	
3. 5	3. Santhosh Kumar–Soft Skill for Business, VBH.	

- 4. C.G.G Krishnamacharyulu & Lalitha: Soft Skills of Personality Development, HPH.
- 5. Lesikar, R.V. &Flatley, M.E. (2005). Basic Business Communication Skills for Empowering the Internet Generation. Tata McGraw Hill Publishing Company Ltd., New Delhi.
- 6. Rai & Rai:Business Communication Himalaya Publishing House
- 7. Rajkumar, Basic of Business Communication
- 8. Ludlow, R. & Panton, F. (1998). The Essence of Effective Communications. Prentice Hall of India Pvt. Ltd.
- 9. M.S. Rao: Soft Skills- Enhancing Employability I.K.International PH.
- 10. Rao& Das: Communication Skills, I.K.International PH.
- 11. Adair, J. (2003). Effective Communication. Pan McMillan.
- 12. Thill, J.V. & Bovee, G. L.(1993). Excellence in Business Communication. McGrawHill, New York.
- 13. Bowman, J.P.&Branchaw, P.P. (1987). Business Communications: From Process to Product. Dryden Press, Chicago.
- 14. Sharma S.P.& Others, Business Communication, VBH.
- 15. Banerjee: Soft Skills Business and Professional Communication, I.K. International

### Syllabus for B. Sc. in Data science (Effective for Academic session 2019-20) BSCDA- 202: INTRODUCTION TO STATISTICS AND PROBABILITY

# Objectives

To enable the students to:

• Learn different Statistical theory and probability methods in practice

• Basic data analysis concepts and fundamentals of Probability

Units	Course Content
1	<b>Introduction to Probability:</b> Sample spaces, events and sets, Probability axioms and simple counting problems, permutations and combinations, Conditional probability, Independent events, partitions ad Bayes Theorem
2	<b>Discrete probability models:</b> Mass functions and distribution functions, expectation and variance, properties of expectations and variance, the binomial distribution, the geometric distribution, poisson distribution
3	<b>Continuous Probability Models</b> : PDF and CDF, Properties of continuous random quantities, the uniform distribution, the exponential distribution, the normal distribution, Normal approximation of binomial and poisson
References	
<ol> <li>Probability and Statistics, By TSR Murthy.</li> <li>Probability and Statistics, By E.Rukmangadachari.</li> <li>Elements of Probability and Statistics, By A.Baisnab and J.Das</li> </ol>	

4. Probability and Statistics for Engineers, By Rao V.Dukkipati

5. The Practice of Business Statistics, By Manisha Sharma & Amit Gupta

#### Syllabus for B. Sc. in Data science (Effective for Academic session 2019-20)

# BSCDA – 203: INTRODUCTION TO PROGRAMMING USING PYTHON

#### Objectives

To enable the students to:

- Understand basics of binary computation
- Understand the programming basics (operations, control structures, data types, etc.)
- Readily use the Python programming language
- Apply various data types and control structure
- Understand class inheritance and polymorphism
- Understand the object-oriented program design and development
- Understand and begin to implement code

Units	Course Content
1	<b>Python Basics</b> Introduction, your first program, type, expressions and variables, string operations, packages
2	<b>Python Data Structure</b> Lists and Truples, Sets, Directories
3	<b>Programming Fundamentals</b> Conditions and Branching, Loops, Functions, Object and Classes, Strings
4	<b>Python Database</b> Working with database
5	<b>Working with Data</b> Open/Reading/Writing files, Loading data with Pandas, Saving data with Pandas

#### References

- 1. Introduction to Programming in Python: An Interdisciplinary Approach, By Sedgewick, Wayne and Dondero
- 2. An Introduction to Python, By Guido Van Rossum.
- 3. Programming in Python 3: A Complete Introduction to Python Language, By Mark Summerfield
- 4. Programming in Python, By R.S. Salaria
- 5. Core Python Programming, By Dr R Nageshwar Rao
- 6. Introduction to Computing and Problem Solving in Python, By Dr. Jeeva Jose
- 7. Taming Python by Programming, By Dr. Jeeva Jose

(Formerly West Bengal University of Technology)

#### Syllabus for B. Sc. in Data science (Effective for Academic session 2019-20)

# **BSCDA – 204: LINEAR MATHEMATICS**

#### Objectives

To enable the students to:

- solve systems of linear equations
- perform various operations with vectors and matrices, in particular, be able to calculate eigenspaces and apply such calculations to the diagonalization of matrices.
- apply linear algebraic methods to geometic problems in 2 and 3 dimensions.
- calculate trigonometric Fourier Series of elementary functions defined on finite intervals and sketch the periodic extensions of such functions
- demonstrate an understanding of concepts by use of examples or counter examples

Units	Course Content
1	<ul> <li>Systems of Linear Equations</li> <li>Introduction to Systems of Linear Equations, Gaussian Elimination, Consistent and Inconsistent Systems.</li> <li>Vectors</li> <li>Vectors in the plane, Vectors in space, Applications to Geometry, n-component vectors, linear independence and bases, Gram-Schmidt Process Linear transformations</li> </ul>
2	Matrices Matrices and Matrix Operations, Square Matrices, Determinants, Inverses, More Systems of Linear Equations Eigenvectors Eigenvalues, Eigenvectors and Diagonalization.
3	<b>Fourier Series</b> Function spaces, Orthogonal projections onto finite dimensional spaces. Calculation of trigonometric Fourier Series, Bessel's Inequality, Parseval's Identity

(Formerly West Bengal University of Technology)

#### Syllabus for B. Sc. in Data science (Effective for Academic session 2019-20)

#### References

1. Linear Mathematics: A practical approach, By Patricia, Clark, Kenschaft

# BSCDA – 205: ADVANCED MS EXCEL

# **Objectives** To enable the students to: • Enable the students to use Excel productively • Out of the box functionality usage to help you in reporting, project management, • Gain insights from vast array of data with ease • Analyze data like a professional • Make powerful dashboards • Write Awesome Formulas **Course Content** Units 1 Getting started with MS-Office MS-Word: Starting Word, Opening a saved Word document, Entering text, Previewing, Editing, Saving, Navigate, Scroll through text, Insert and delete text etc. MS-PowerPoint: Difference between presentation and Document, Using Power Point, Using Wizard for creating a presentation, Creation of Presentation, Title, Text Creation, Fonts and Sizes, Importing text from word documents, Moving to next Slide, The Slide manager, Animation effects, Slide Designs, Background and Text colors, Making your own slide format, Foot notes and slide numbering, Presentation of the Slides, Using the Slide Show, Printing the Slides and Handouts. 2 **Introduction to Microsoft Excel:** Concepts of Work book & Work sheets; Various Data Types; using different features with data, Cell and Texts; Inserting, Removing & Resizing of Columns & Rows; Working with Data and Ranges; entering data into worksheet, saving& quitting worksheet, Opening and moving around in an existing worksheet, Toolbars and menu, keyboard shortcuts, Working with single and multiple workbook-copying, renaming, moving, adding and deleting, copy in gentries and moving between work books, Different Views of Work sheets; Column Freezing, Labels, Hiding, Splitting etc., Using different features with Data and Text;

(Formerly West Bengal University of Technology)

#### Syllabus for B. Sc. in Data science (Effective for Academic session 2019-20)

3	<b>Formulas and Functions in Excel:</b> Use of Formulas, Calculations and Functions, Logical Functions, Text Functions, Lookups
4	Advance Data Tools Text to Column, Data Validation, What-if Analysis, Duplicate Removal, Data Sanitation
5	Visualising data using charts in Excel Basic Date visualisation, Charts in excel, dashboard in Excel
6	Pivot table in Excel
7	Analytics using Excel Data analysis using normal chart, Regression in Excel, correlation, stddev, average, ANOVA
References	

1. Mastering MS Excel: Functions and Formulas, Webtech (Khanna Publications)

2. Excel Functions and Formulas, Bernd Held

3. Advance Excel 2016, training Guide, By Ritu Arora

4. MS Office 2010, training guide, Prof Satish Jain

# Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20) <u>3<sup>rd</sup> SEMESTER</u>

# **BSCDA-301: DATA STRUCTURES AND ALGORITHMS**

# Objectives

To enable the students to:

- Analyze the asymptotic performance of algorithms.
- Write rigorous correctness proofs for algorithms.
- Demonstrate a familiarity with major algorithms and data structures.
- Apply important algorithmic design paradigms and methods of analysis.
- Synthesize efficient algorithms in common engineering design situations.

# Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

	-Applications of Graphs: Topological Sorting, Shortest-Path Algorithms – Weighted Shortest
	Paths – Dijkstra's Algorithm, Minimum spanning tree- Prim's Algorithm, Introduction to NP-
	Completeness.
	-Searching and Sorting Techniques, Sorting Techniques: Bubble sort, Merge sort, Selection sort',
	Heap sort, Insertion Sort. Searching Techniques: Sequential Searching, Binary Searching, Search
	Trees.
	-Elementary Algorithms: Notation for Expressing Algorithms; Role and
	Notation for Comments; Example of an Algorithm; Problems and Instances; Characteristics of an
	Algorithm; Building Blocks of Algorithms; Procedure and Recursion – Procedure, Recursion;
	Outline of Algorithms; Specification Methods for Algorithms.
	-Mathematical Functions and Notations Functions and Notations; Modular Arithmetic / Mod
	Function; Mathematical Expectation in Average Case Analysis; Efficiency of an Algorithm; Well
	Known Asymptotic Functions and Notations; Analysis of Algorithms – Simple Examples; Well
	Known Sorting Algorithms – Insertion sort, Bubble sort, Selection sort, Shell sort, Heap sort.
	-Divide and Conquer Divide and Conquer Strategy; Binary Search; Max. And Min.; Merge sort;
	Quick sort.
	-Greedy Method Greedy Method Strategy; Optimistic Storage on Tapes; Knapsack Problem; Job
	Sequencing with Deadlines; Optimal Merge Pattern; Single Source Shortlist Paths.
	Dynamic Programming Dynamic Programming Strategy; Multistage Graphs; All Pair Shortest
	Paths; Travelling Salesman Problems
	-Backtracking Strategy, 8-Queens Problem, Sum of Subsets, Knapsack Problem.
References	
<ul> <li>Data Structures and Algorithms made easy, By Narsimha Karumanchi</li> <li>Data Structures and Algorithms, By Aho, Hopcroft and Ulman</li> <li>Data Structures &amp; Algorithms Using C. R.S. Salaria</li> </ul>	

### Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **BSCDA- 302: DATA WAREHOUSING AND MINING**

#### Objectives

To enable the students to:

- To understand data warehouse concepts, architecture, business analysis and tools
- To understand data pre-processing and data visualization techniques
- To study algorithms for finding hidden and interesting patterns in data
- To understand and apply various classification and clustering techniques using tools

Units	Course Content
1	<ul> <li>Data Warehousing, Business Analysis and On-line Analytical processing (OLAP): Basic Concepts – Data Warehousing Components – Building a Data Warehouse – Database Architectures for Parallel Processing – Parallel DBMS Vendors – Multidimensional Data Model – Data Warehouse Schemas for Decision Support, Concept Hierarchies -Characteristics of OLAP Systems – Typical OLAP Operations, OLAP and OLTP.</li> <li>Data Mining, Introduction: Introduction to Data Mining Systems – Knowledge Discovery Process – Data Mining Techniques – Issues – applications- Data Objects and attribute types, Statistical description of data, Data Preprocessing – Cleaning, Integration, Reduction, Transformation and discretization, Data Visualization, Data similarity and dissimilarity measures.</li> </ul>
2	<ul> <li>Data Mining, frequent patten Analysis: Mining Frequent Patterns, Associations and Correlations – Mining Methods- Pattern Evaluation Method – Pattern Mining in Multilevel, Multi Dimensional Space – Constraint Based Frequent Pattern Mining, Classification using Frequent Patterns</li> <li>Classification and Clustering: Decision Tree Induction – Bayesian Classification – Rule Based Classification – Classification by Back Propagation – Support Vector Machines — Lazy Learners – Model Evaluation and Selection-Techniques to improve Classification Accuracy. Clustering Techniques – Cluster analysis-Partitioning Methods – Hierarchical Methods – Density Based Methods – Grid Based Methods – Evaluation of clustering – Clustering high dimensional data- Clustering with constraints, Outlier analysis-outlier detection methods.</li> </ul>
3	<b>Weka tool:</b> Datasets – Introduction, Iris plants database, Breast cancer database, Auto imports database – Introduction to WEKA, The Explorer – Getting started, Exploring the explorer, Learning algorithms, Clustering algorithms, Association–rule learners.

(Formerly West Bengal University of Technology)

#### Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

#### References

- 0.1. Jiawei Han and Micheline Kamber, —Data Mining Concepts and Techniques, Third Edition, Elsevier, 2012.
- 4.2. Alex Berson and Stephen J.Smith, —Data Warehousing, Data Mining & OLAPI, Tata McGraw Hill Edition, 35th Reprint 2016.
- 2.3. K.P. Soman, Shyam Diwakar and V. Ajay, —Insight into Data Mining Theory and Practice, Eastern Economy Edition, Prentice Hall of India, 2006.
- 3.4. Ian H. Witten and Eibe Frank, —Data Mining: Practical Machine Learning Tools and Techniques, Elsevier, Second Edition.
- 4.5. Machine Learning, Rajiv Chopra, Second Edition, Khanna Publishing House, 2018
- 5.6. Machine Learning, Jeeva Jose, Khanna Publishing House

# Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20) BSCDA-303: STATISTICAL THEORY OF HYPOTHESIS TESTING

# Objectives

To enable the students to:

• Learn a **statistical hypothesis**, sometimes called **confirmatory data analysis**, which is a hypothesis that is testable on the basis of observing a process that is modeled via a set of random variables. A **statistical hypothesis test** is a method of statistical inference. Commonly, two statistical data sets are compared, or a data set obtained by sampling is compared against a synthetic data set from an idealized model. A hypothesis is proposed for the statistical relationship between the two data sets, and this is compared as an alternative to an idealized null hypothesis that proposes no relationship between two data sets

Units	Course Content
1	Elements of Hypothesis Testing : Null and Alternative hypotheses, Simple and Composite hypotheses,
2	Critical Region, Type I and Type II Errors, Level of Significance and Size, p-value, Power
3	Tests of Significance related to a single Binomial proportion and Poisson parameter; two Binomial proportions and Poisson parameters;
4	the mean(s) and variance(s) of a single univariate normal distribution,
5	two independent normal distributions and a single bivariate normal distribution;
	Idea of Inference - Point & Interval Estimations and Testing of Hypothesis. Point estimation: Requirements of a good estimator – notions of Mean Square Error, Unbiasedness: Minimum Variance Uunbiasedness and Best Linear Unbisedness, Sufficiency, Factorization Theorem (Discrete case only), Properties of minimum variance unbiased estimators, consistent estimators and asymptotic efficiency, Cramer-Rao lower bound, Rao-Blackwell Theorem. 9 Methods of Estimation – Moment, Least-square, Maximum Likelihood & Minimum $\chi^2$ methods and their properties (excluding proofs of large sample properties).
	References

# Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

1. Goon A.M., Gupta M.K. & Dasgupta B. (1994): An Outline of Statistical Theory (Vol-2), World Press
2. Mood A.M., Graybill F. & Boes D.C. (1974): An Introduction to the Theory of Statistics (3rd ed), McGraw Hill 3
<b>3</b> . Rao C.R. (1952): Advanced Statistical Methods in Biometric Research, John Wiley
4. Hogg R.V. & Craig A.T. (1978): Introduction to Mathematical Statistics
<ol> <li>Rohatgi V.K. (1984): An Introduction to Probability Theory &amp; Mathematical Statistics, John Wiley 6.</li> </ol>
6. Stuart G & Ord J.K. (1991): Advanced Theory of Statistics (Vol 2), Charles Griffin 7.
<ol> <li>Goon A. M., Gupta M. K. and Dasgupta B. (1997): Fundamentals of Statistics (V-1 and 2), World Press 8.</li> </ol>
8. Bhattacharya GK & Johnson R. A. (1977): Concepts & Methods of Statistics, John Wiley
9. Manish Sharma & Amit Gupta, The Practice of Business Statistics, Khanna Publishing House

10. Scheffe H. (1959): The Analysis of Variance, John Wiley

#### Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **BSCDA-304: DATA VISUALISATION USING BUSINESS INTELLIGENCE**

#### **Objectives:**

To enable the students to:

Analyse large corporate datasets using Business Intelligence/Business Analytics tools to generate insights and provide alternative solutions to an organisation's complex problems

Unit	Course Content
	<ul> <li>Concept of business value from corporate data, the exploitation of information for advantage, types and sources of information value</li> <li>Nature and value of business intelligence, the business intelligence environment, and how types of data processing can add value to corporate data sources</li> <li>Knowledge discovery, data mining, data warehousing</li> <li>Business analytics, OLAP analysis, metadata</li> <li>Data visualisation, visualisation techniques, dashboard</li> <li>The relationship between corporate strategy, IS strategy and business intelligence strategy</li> <li>BI links to enterprise systems, CRM (Customer Relationships Management), SCM (Supply Chain Management) Structured &amp; unstructured data, content management systems</li> <li>Privacy, ethical, legal issues associated with BI Implementation</li> <li>BI, Decision Support Systems, Expert Systems and Executive Information Systems</li> <li>Data modelling, star schemas</li> <li>Using select and complementary BI/BA tools to provide insights in decision making scenarios</li> <li>Use of Power BI functions</li> </ul>
Refere	ences
1. Big	Data Visualization, By James W Miller.

- 2. High Impact Data Visualization with Power View, Power Map, and Power BI, By Adam Aspin
- 3. Big Data & Hadoop, V.K. Jain

(Formerly West Bengal University of Technology)

# Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **BSCDA-305: WEB FOUNDATION**

# Objectives

To enable the students to:

Learn about internet and world wide web, Digital Data

Units	Course Content	
1	Introduction	
	Introduction to Internet and www, web browsers, web servers, search engine	
2	Internet Commerce	
3	Digital Marketing	
4	Web Analytics	
5	Social Media Analytics	
References		
<ol> <li>Fundame</li> <li>Internet &amp;</li> </ol>	ntals of Web Development, By Connolly and Hoar & Web Development, By Soma Das Gupta	

#### Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# 4<sup>th</sup> SEMESTER

# BSCDA-401: DATA VISUALISATION AND DATA DRIVEN DECISION MAKING

• get an intr represent	roduction to Data Analytics and its role in business decisions process based on Analytics and how to it in right format
Units	Course Content
1	What is Data Analytics?
	Solving business problems using data analytics
	Making business-defining decisions using data analytics
	Why do you need a data and analytics framework?
	The 4 aspects of the data and analytics framework
	Data and analytics framework: tools and techniques
	Make better and faster decisions with data and analytics
	Identifying, organizing and processing data
	Data based decision process
2	Types of data analysis techniques
	The role of Excel
	The role of SAS
	The role of R
	The role of Python
	The Power of Visualization
	The role of Tableau/ Power Bl
References	
	Ver Constitution Analysis Description Techniques for Orientitative Analysis Drugtonhan Foru

2.3. Data Science & Analytics (with Python, R and SPSS Programming), By V.K. Jain

3.4. Jeeva Jose – Beginner's Guide for Data Analysis using R Programming

#### Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **BSCDA-402 STATISTICAL ANALYSIS USING SPSS**

#### Objectives

To enable the students to:

• General objective of the course is to introduce students in knowledge of basic concepts of statistics and the ability to analyse quantitative survey data using special statistical software SPSS

Units	Course Content
1	<ol> <li>Basic strategies of quantitative research: research questions, operationalisation, variables;</li> <li>How to prepare data for the analysis using SPSS -(module files; edit, view, utilities)</li> <li>Distribution of categorical data and univariate analysis (module descriptive statistics - frequencies, explore);</li> </ol>
2	<ol> <li>Distribution of interval data and their analysis;</li> <li>Transformation of data (module transform, recode, compute, count, rank cases);</li> <li>Normal Distribution and hypothesis testing - statistical inference;</li> </ol>
3	<ol> <li>Comparison of means: t-test, one-sample t-test; independent-samples t-test); analysis of variance;</li> <li>Bivariate analysis – cross tabulation;</li> <li>Strength of association - coefficients of association and correlation;</li> </ol>
4	<ol> <li>Spurious correlations, elaboration, partial correlation;</li> <li>Linear regression;</li> <li>Factor analysis.</li> </ol>
References	

- 1. NORUŠIS, M. J. SPSS introductory statistics : student guide2.
- 2. SPSS Base 14.0 : user's guide. Chicago: SPSS Inc., 2005.
- 3. Data Science & Analytics (with Python, R and SPSS Programming), By V.K. Jain

# Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **BSCDA-403: INTRODUCTION TO REGRESSION ANALYSIS**

Objectives	
To enable the students to:	
• get an intro represent it	oduction to Data Analytics and its role in business decisions process based on Analytics and how to in right format
Units	Course Content
1	Regression and correlation coefficients of a single bivariate normal distribution, Combination of Probabilities in tests of significance
2	<ul> <li>Analysis of Variance (ANOVA)</li> <li>Introduction: Heterogeneity and Analysis of Variance and Covariance, Linear Hypothesis, Orthogonal splitting of total variation, Selection of Valid Error.</li> <li>Applications of the ANOVA technique to: one-way classified data, two-way classified data with equal number of observations per cell, testing simple regression correlation ratio, linearity of simple regression, multiple correlation and partial correlation coefficients.</li> </ul>
	•
References	
1. Scheffe H. (	1959): The Analysis of Variance, John Wiley
2. A Linear Mo	odel Approach to Regression Analysis and its Applications, Suddhendu Biswas, New Central

book agency

(Formerly West Bengal University of Technology)

### Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **BSCDA-404: GENERALISED LINEAR MODELS**

# Objectives

To enable the students to:

• Gain the knowledge about construction methodology and procedure and also the ability to apply them in the enterprise reality

Units	Course Content
1	<ol> <li>GLM basics: components, exponential family, model fitting, frequent inference: analysis of deviance, stepwise selection, goodness of fit</li> <li>Bayesian inference in GLMs (basics): priors, posterior, comparison with frequentist approach, posterior computation, MCMC strategies (Gibbs, MetropolisHastings)</li> <li>Binary &amp; categorical response data: (a) Basics: link functions, form of posterior, approximations, Gibbs sampling via adaptive rejection (b) Latent variable models: Threshold formulations, probit models, discrete choice models, logistic regression &amp; generalizations, data augmentation algorithms (Albert &amp; Chib + other forms)</li> <li>Count data: Poisson &amp; over-dispersed Poisson log-linear models, prior distributions, applications</li> <li>Bayesian variable selection: problem formulation, mixture priors, stochastic search algorithms, examples, approximations</li> <li>Bayesian hypothesis testing in GLMs: one- and two-sided alternatives, basic decision theoretic approaches, mixture priors, computation, order restricted inference</li> <li>Survival analysis: censoring definitions, form of likelihood, parametric models, discrete-time &amp; continuous time formulation, selection &amp; pattern mixture models, shared variable approaches, examples</li> <li>Multistate &amp; stochastic modeling: motivating examples (epidemiologic studies with periodic observations of a disease process), discrete time approaches, joint models, computation</li> <li>Generalized linear mixed models (GLMM): definition, examples, normal linear case - induced correlation structure, priors, computation, frequentist approaches for inference &amp; computation (Hastie &amp; Tibshirani), Bayesian approaches using basis functions, priors, computation</li> <li>Greneralized factor models: definition, frequentist approaches for inference &amp; computation structure, priors, computation, multi-level models, covariance selection</li> <li>Greneralized factor models: definition, frequentist approaches for inference &amp; comp</li></ol>
Keferences	

1) Generalized Linear Models -- Ulff Olsson, Overseas Press.

(Formerly West Bengal University of Technology)

# Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **BSCDA-405: PROJECT I**

Objectives

To enable the students to:

• To give the students exposure to dummy projects on Analytics using specific softwares/ tools.

Software/ tools: SAS/MS Excel, Python

#### Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# 5<sup>th</sup> SEMESTER

# **BSCDA-501: ADVANCED PROGRAMMING IN PYTHON**

# Objectives

To enable the students to:

learn how to analyze data in Python using multi-dimensional arrays in numpy, manipulate DataFrames in pandas, use SciPy library of mathematical routines,

Units	Course Content
1	<b>Python Basic:</b> Python fundamental, working with data
2	Importing Dataset Domain, Dataset, Package for Data Science, Importing/Exporting Data, Insight from Dataset
3	<b>Cleaning and Preparing the Data</b> Identify and Handle Missing Values, Data Formatting, Normalisation, Binning
4	Summarising the Data Frame Descriptive Statstics, Grouping, ANOVA, Correlation
5	<b>Model Development</b> Linear Regression, Prediction and Decission making
6	<b>Data Vizualization</b> Introduction to Matplotlib, Basic plotting, Charts

References

0.1. Python: 3 Manuscripts in 1 book: - Python Programming For Beginners - Python Programming For Intermediates - Python Programming for Advanced, By Maurice J.Thompson

1-2. Advanced Machine Learning with Python, By John Hearty

2.3. Taming Programming by Python, By Dr. Jeeva Jose

(Formerly West Bengal University of Technology)

## Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **BSCDA-502: DATA SECURITY**

# Objectives

To enable the students to:

• to train students in the organizing and the technical realization and security of data and computers

Units	Course Content
1	Information system security and protection objectives. The development of the Internet and the role of the intranet and extranet. Control at the level of management: data control, data administration, security control, control at the management level. Software control. Access Control: cryptography, identification numbers, digital signatures, security and credit card business. Input control, communication control, control of data processing, database control, output data control. Legal aspects of the security of information systems. Information systems security planning: security management information system, the reconstruction plan information system, ISO / IEC 17799: 2000. The insurance. Network security threats: spyware, search, denial of services, misrepresentation, playback and session hijacking, redirections, viruses, Trojan horses, and worms. Defining a security policy. Protecting the network and operating system services and procedures: one-time passwords, token cards / soft tokens, TACACS +, RADIUS, KERBEROS, VPN, IKE / IPSec. Secure data storage. Monitoring the performance of the system. Intrusion detection systems. Reestablishment of network systems.

#### References

1.

Sharing Big Data Safely: Managing Data Security - Managing Data Security (English, Paperback, Ellen

Friedman Ted Dunning).

(Formerly West Bengal University of Technology)

# Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **BSCDA-503: TIME SERIES**

# Objectives

To enable the students to:

- have deeper knowledge of statistical theory and methods particularly common problems in economical social sciences especially economics.
- be able to estimate models for time-series data.
- be able to interpret the results of an implemented statistical analysis
- be aware of limitations and possible sources of errors in the analysis
- have ability to present results in oral and written form

Units	Course Content
1	Overview of forecasting. Models for time series: Time-dependent seasonal components. Autoregressiva (AR), moving average (MA) and mixed ARMA-modeller. The Random Walk Model. Box-Jenkins methodology. Forecasts with ARIMA and VAR models. Dynamic models with time-shifted explanatory variables. The Koyck transformation . Partial adjustment and adaptive expectation models. Granger's causality tests. Stationarity, unit roots and cointegration. Modelling of volatility: ARCH - and the GARCH-models.
	·

# References

- 1. Time Series Analysis, By James D.Hamilton
- 2. Time Series, By Peter J Brockwell and Richard A Davies

(Formerly West Bengal University of Technology)

# Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **BSCDA-504: WEB INTELLIGENCE**

#### **Objectives**

To enable the students to:

• Get introduced to topics of web intelligence.

 $\cdot$  Study models of information retrieval, semantic webs, search engines, and web mining.

· Learn applying data mining tools to develop projects in web mining and information retrieval.

Units	Course Content
1	Introduction to Web Intelligence
	What is Web Intelligence?
	Benefits of Intelligent Web
	Ingredients of Web Intelligence
	Topics of Web Intelligence
	Related Technologies
	Information Retrieval
	· Document Representation
	· Retrieval Models
	· Evaluation of Retrieval Performance
	Semantic Web
	· The Layered-Language Model
	· Metadata and Ontologies
	· Ontology Languages for the Web
	Data Mining Techniques
	· Classification and Association
	· Clustering
	Web Usage Mining
	· Web-Log Processing
	· Analyzing Web Logs
	· Applications of Web Usage Mining
	o Clustering of Web Users
	o Classification Modeling of Web Users
	o Association Mining of Web Usages
	o Sequence-Pattern Analysis of Web Logs

# Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

Web Content Mining
· Web Crawlers
· Search Engines
· Personalization of Web Content
· Multimedia Information Retrieval
Web Structure Mining
· Modeling Web Topology
o PageRank Algorithm
o Hyperlink-Induced Topic Search (HITS)
o Random Walks on the Web
· Social Networks
Reference and Index Pages
References
1. Akerkar, R. & Lingras, P. (2008). Building an Intelligent Web: Theory and Practice.
<ol> <li>Witten, Ian H. &amp; Frank, E. (2005). Data Mining: Practical Machine Learning Tools and</li> </ol>
Techniques. 2 <sup>nd</sup> Edition, Morgan Kaufman. ISBN 0120884070, 9780120884070
3. Reprint journal articles and conference proceedings.

#### Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# BSCDA-505: INTRODUCTION TO ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

# Objectives

To enable the students to:

- To provide a strong foundation of fundamental concepts in Artificial Intelligence
- To provide a basic exposition to the goals and methods of Artificial Intelligence
- To enable the student to apply these techniques in applications which involve perception, reasoning and learning

1 Advanced Algorithms and Analysis Machine Learning Techniques Artificial Intelligence and Neural Networks Statistical Modeling for Computer Sciences Artificial Intelligence and Neural Networks Lab Machine Learning Lab Computational Intelligence	Units	Course Content
	1	Advanced Algorithms and Analysis Machine Learning Techniques Artificial Intelligence and Neural Networks Statistical Modeling for Computer Sciences Artificial Intelligence and Neural Networks Lab Machine Learning Lab Computational Intelligence

#### References

- 1. Stewart Russell and Peter Norvig. " Artificial Intelligence-A Modern Approach ", 2nd Edition, Pearson Education/ Prentice Hall of India, 2004 References
- 2. Nils J. Nilsson, "Artificial Intelligence: A new Synthesis", Harcourt Asia Pvt. Ltd., 2000. 2.
- 3. Elaine Rich and Kevin Knight, "Artificial Intelligence", 2nd Edition, Tata McGraw-Hill, 2003.
- 4. M.C. Trivedi, Artificial Intelligence, Khanna Publishing House, 2018
- 5. Rajiv Chopra, Machine Learning, Khanna Publishing House, 2018

# Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **6<sup>TH</sup> SEMESTER**

# **BSCDA-601: OPTIONAL SUBJECT (ANY ONE)**

Objectives		
To enable the	e students to:	
• Understa	nd the Practical aspects and case studies of the topic the they select	
Units	Course Content	
1.	Social Media and Web Analytics	
2.	Advanced Business Intelligence	
3.	Big Data Analytics	
4.	Information driven Entrepreneurship and Enterprise	
Option 1:		
SOCIAL MEDIA	AND WEB ANALYTICS	
What is Web Analytics		
Getting started with Google Analytics, how it works, accounts, profiles and users		
Navigating Google Analytics, Traffic sources, Content, Visitors		
Web Analytics in e-Commerce, Actionable Insights		
Web Analytics tools, making better decisions and summing up		
Social Media Analytics		
Social CRM and Analytics		
Option 2:		

#### Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

#### ADVANCED BUSINESS INTELLIGENCE

**Business Intelligence** 

Effective and timely decisions - Data, information and knowledge - Role of mathematical models - Business intelligence architectures: Cycle of a business intelligence analysis -Enabling factors in business intelligence projects – Development of a business intelligence system - Ethics and business intelligence

Knowledge Delivery:

The business intelligence user types, Standard reports, Interactive Analysis and Ad Hoc Querying, Parameterized Reports and Self-Service Reporting, dimensional analysis, Alerts/Notifications, Visualization: Charts, Graphs, Widgets, Scorecards and Dashboards, Geographic Visualization, Integrated Analytics, Considerations: Optimizing the Presentation for the Right Message

Efficiency:

The business intelligence user types, Standard reports, Interactive Analysis and Ad Hoc Querying, Parameterized Reports and Self-Service Reporting, dimensional analysis, Alerts/Notifications, Visualization: Charts, Graphs, Widgets, Scorecards and Dashboards, Geographic Visualization, Integrated Analytics, Considerations: Optimizing the Presentation for the Right Message

**Business Intelligence Applications:** 

Marketing models – Logistic and Production models – Case studies

Future of Business Intelligence: Future of business intelligence – Emerging Technologies,

Machine Learning, Predicting the Future, BI Search & Text Analytics – Advanced

Visualization – Rich Report **Option 3: BIG DATA ANALYTICS** 

1. Overview of Big Data: This includes topics such as history of big data, its elements, career related knowledge, advantages, disadvantages and similar topics.

2. Using Big Data in Businesses: This module should focus on the application perspective of Big Data covering topics such as using big data in marketing, analytics, retail, hospitality,

#### Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

consumer good, defense etc.

3. Technologies for Handling Big Data: Big Data is primarily characterized by Hadoop. This module cover topics such as Introduction to Hadoop, functioning of Hadoop, Cloud computing (features, advantages, applications) etc.

4. Understanding Hadoop Ecosystem: This includes learning about Hadoop and its ecosystem which includes HDFS, Map Reduce, YARN, HBase, Hive, Pig, Sqoop, Zookeeper, Flume, Oozie etc.

5. Dig Deep to understand the fundamental of Map Reduce and HBase: This module should cover the entire framework of Map Reduce and uses of map reduce.

6. Understanding Big Data Technology Foundations: This module covers the big data stack i.e. data source layer, ingestion layer, source layer, security layer, visualization layer, visualization approaches etc.

7. Databases and Data Warehouses: This module should cover all about databases, polygot persistence and their related introductory knowledge

8. Using Hadoop to store data: This includes an entire module of HDFS, HBase and their respective ways to store and manage data along with their commands.

**Option 4: INFORMATION DRIVEN ENTREPRENEURSHIP AND ENTERPRISE** 

Understanding the Enterprise Technology venture Funding and promotion of the Enterprise Technology Entrepreneurship and Innovation

**Enterprise Strategy and operations** 

References

(Formerly West Bengal University of Technology)

# Syllabus for B. Sc. in Data Science (Effective for Academic session 2019-20)

# **BBARE-602: MAJOR PROJECT & VIVA VOCI**

Objectives	
To enable the	e students to:
<ul> <li>Prepare real/dum collect pr appropria</li> </ul>	a major project on thrusting areas of Data Science and Analytics in my work scenario. While preparing this project, students would be able to imary and secondary data, analysis and statistical representation of data with the research methodology and use the latest softwares and tools.
Units	Course Content
	Guidelines for Preparation of Major Project:
	1. Introduction
1.	2. Objectives of the Study
	3. Review of Literature
	4. Hypothesis
	5. Research Methodology
	6. Conceptual Studies
	7. Empirical Studies
	8. Data analysis Observation and Findings with Statistical tools and Softwares
	9. Recommendation
	10. Conclusion
	11. Bibliography Reference
References	
<ul> <li>Research Methodology by R.S. Kothari</li> </ul>	