GANDHI INSTITUTE OF TECHNOLOGY AND MANAGEMENT (GITAM)
(Deemed to be University, Estd. u/s 3 of UGC Act 1956)
VISAKHAPATNAM ∗HYDERABAD ∗BENGALURU
Accredited by NAAC with 'A^{+}' Grade

REGULATIONS AND SYLLABUS

of

M.Sc. Data Science

(w.e.f 2019-20 Admitted batch)

Website: www.gitam.edu

Master of Science in Data Science (M.Sc Data Science)
REGULATIONS
(w.e.f. 2019-20 admitted batch)

1. ADMISSION

1.1 Admission into M.Sc. Data Science program of GITAM University is governed by GITAM University admission regulations.

2. ELIGIBILITY CRITERIA

2.1. A pass in BCA or B.Sc. degree with a minimum aggregate of 50% marks / a pass in any degree with minimum aggregate of 50% marks along with Mathematics or Statistics or Computer science as one of the subject.

2.2. Admission into M.Sc. Data Science (Master of Science in Data Science) will be based on All India GITAM Science Admission Test (GSAT) conducted by GITAM University and the rule of reservation, wherever applicable.

3. CHOICE BASED CREDIT SYSTEM

Choice Based Credit System (CBCS) is introduced with effect from the admitted Batch of 2015-16 based on UGC guidelines in order to promote:

- Student Centered Learning
- Cafeteria approach
- Inter-disciplinary learning

Learning goals/objectives and outcomes are specified leading to what a student should be able to do at the end of the program.

4. STRUCTURE OF THE PROGRAM

4.1 The Program Consists of

i) Foundation Courses (compulsory) which gives general exposure to a Student in communication and subject related area.
ii) Core Courses (compulsory).
iii) Discipline centric electives which
   a) are supportive to the discipline
   b) give expanded scope of the subject
   c) give their disciplinary exposure
   d) nurture the student skills
iv) Open electives are of general nature either related or unrelated to the discipline.
v) Practical Proficiency Courses, Laboratory and Project work.

4.2 Each course is assigned a certain number of credits depending upon the number of contact hours (lectures/tutorials/practical) per week.

4.3 In general, credits are assigned to the courses based on the following contact hours per week per semester.

- One credit for each Lecture / Tutorial hour per week.
- One credit for two hours of Practical per week.
- Eight credits for project.

4.4 The curriculum of the Four semesters M.Sc. Data Science program is designed to have a total of 81 credits for the award of M.Sc. Data Science degree.

5. MEDIUM OF INSTRUCTION

The medium of instruction (including examinations and project reports) shall be in English.

6. REGISTRATION

Every student has to register himself / herself for each semester individually at the time specified by the Institute / University.

7. ATTENDANCE REQUIREMENTS

7.1. A student whose attendance is less than 75% in all the courses put together in any semester will not be permitted to attend that end - semester examination and he/she will not be allowed to register for subsequent semester of study. He/she has to repeat the semester along with his / her juniors.

7.2. However, the Vice Chancellor on the recommendation of the Principal / Director of the Institute/School may condone the shortage of attendance to the students whose attendance is between 66% and 74% on genuine grounds and on payment of prescribed fee.

8. EVALUATION

8.1. The assessment of the student's performance in a Theory course shall be based on two components: Continuous Evaluation (40 marks) and Semester-end examination (60 marks).

8.2. A student has to secure an aggregate of 40% in the course in continuous and semester end examinations the two components put together to be declared to have passed the course, subject to the condition that the candidate must have secured a minimum of 24 marks (i.e. 40%) in the theory component at the semester-end examination.

8.3. Practical / Viva voce etc. course are completely assessed under Continuous Evaluation for a maximum of 100 marks and a student has to obtain a minimum of 40% to secure Pass Grade. Details of Assessment Procedure are furnished below in Table 1.

Table 1: Assessment Procedure

S. No.	Component of assessment	Marks allotted	Type of Assessment	Scheme of Examination
1	Theory	40	Continuous evaluation	(i) Three mid semester examinations shall be conducted for 15 marks each. The performance in best two shall be taken into consideration. (ii) 5 marks are allocated for quiz. (iii) 5 marks are allocated for assignments.
		60	Semester-end examination	The semester-end examination shall be for a maximum of 60 marks.
	Total	100		
2	Practicals	100	Continuous evaluation	60 marks for performance, regularity, record and case study. Weightage for each component shall be announced at the beginning of the semester. 40 marks (30 marks for experiment(s) and 10 marks for practical Viva-voce.) for the test conducted at the end of the semester conducted by the concerned lab Teacher.
	Total	100		
3	Project work	200	Project evaluation	150 marks for evaluation of the project work dissertation submitted by the candidate. 50 marks are allocated for the project Viva-Voce. The project work evaluation and the Viva-Voce shall be conducted by one external examiner outside the University and the internal examiner appointed by the Head of the Department.

9. SUPPLEMENTARY EXAMINATIONS & SPECIAL EXAMINATIONS:

9.1 The odd semester supplementary examinations will be conducted on daily basis after conducting regular even semester examinations in April/May.

9.2 The even semester supplementary examinations will be conducted on daily basis after conducting regular odd semester examinations during November/December

9.3 A student who has completed his/her period of study and still has "F" grade in final semester courses is eligible to appear for Special Examination normally held during summer vacation.

10. PROMOTION TO THE NEXT YEAR OF STUDY

10.1 A student shall be promoted to the next academic year only if he/she completes the academic requirements of 60% of the credits till the previous academic year.

10.2 Whenever there is a change in syllabus or curriculum he/she has to continue the course with new regulations after detention as per the equivalency established by the BoS to continue his/her further studies

11. BETTERMENT OF GRADES

11.1 A student who has secured only a pass or second class and desires to improve his/her class can appear for betterment examinations only in 'n' (where 'n' is no.of semesters of the program) theory courses of any semester of his/her choice, conducted in summer vacation along with the Special Examinations.

11.2 Betterment of Grades is permitted 'only once', immediately after completion of the program of study.

12. REPEAT CONTINUOUS EVALUATION:

12.1 A student who has secured 'F' grade in a theory course shall have to reappear at the subsequent examination held in that course. A student who has secured 'F' grade can improve continuous evaluation marks upto a maximum of 50% by attending special instruction classes held during summer.

12.2 A student who has secured 'F' grade in a practical course shall have to attend Special Instruction classes held during summer.

12.3 A student who has secured 'F' grade in a combined (theory and practical) course shall have to reappear for theory component at the subsequent examination held in that course. A student who has secured 'F' grade can improve continuous evaluation marks upto a maximum of 50% by attending special instruction classes held during summer.

12.4 The RCE will be conducted during summer vacation for both odd and even semester students. Student can register a maximum of 4 courses. Biometric attendance of these RCE classes has to be maintained. The maximum marks in RCE be limited to 50% of Continuous Evaluation marks. The RCE marks are considered for the examination held after RCE except for final semester students.

12.5 RCE for the students who completed course work can be conducted during the academic semester. The student can register a maximum of 4 courses at a time in slot of 4 weeks. Additional 4 courses can be registered in the next slot.

12.6 A student is allowed to Special Instruction Classes (RCE) 'only once' per course.

13. GRADING SYSTEM

13.1 Based on the student performance during a given semester, a final letter grade will be awarded at the end of the semester in each course. The letter grades and the corresponding grade points are as givenin Table 2.

Table 2: Grades & Grade Points

Sl.No.	Grade	Grade Points	Absolute Marks
1	O (outstanding)	10	90 and above
2	A+ (Excellent)	9	80 to 89
3	A (Very Good)	8	70 to 79
4	B+ (Good)	7	60 to 69
5	B (Above Average)	6	50 to 59
6	C (Average)	5	45 to 49
7	P (Pass)	4	40 to 44
8	F (Fail)	0	Less than 40
9	Ab. (Absent)	0	-

13.2 A student who earns a minimum of 4 grade points (P grade) in a course is declared to have successfully completed the course, subject to securing an average GPA (average of all GPAs in all the semesters) of 5 at the end of the Program to declare pass in the program.

Candidates who could not secure an average GPA of 5 at the end of the program shall be permitted to reappear for a course(s) of their choice to secure the same.

14. GRADE POINT AVERAGE

14.1 A Grade Point Average (GPA) for the semester will be calculated according to the formula:

$$GPA = \frac{\Sigma [C * G]}{\Sigma C}$$

Where

C = number of credits for the course,
G = grade points obtained by the student in the course.

14.2 To arrive at Cumulative Grade Point Average (CGPA), a similar formula is used considering the student's performance in all the courses taken, in all the semesters up to the particular point of time.

14.3 CGPA required for classification of class after the successful completion of the program is shown in Table 3.

Table 3: CGPA required for award of Class

Class	CGPA Required
First Class with Distinction	≥ 8.0*
First Class	≥ 6.5
Second Class	≥ 5.5
Pass Class	≥ 5.0

* In addition to the required CGPA of 8.0 or more the student must have necessarily passed all the courses of every semester in first attempt.

15. ELIGIBILITY FOR AWARD OF THE M.Sc. Data Science DEGREE

15.1 Duration of the program: A student is ordinarily expected to complete M.Sc Data Science program in four semesters of two years. However a student may complete the program in not more than four years including study period.

15.2 However the above regulation may be relaxed by the Vice Chancellor in individual cases for cogent and sufficient reasons.

15.3 A student shall be eligible for award of the M.Sc Data Science Degree if he / she fulfills all the following conditions.

a) Registered and successfully completed all the courses and projects.
b) Successfully acquired the minimum required credits as specified in the curriculum corresponding to the branch of his/her study within the stipulated time.
c) Has no dues to the Institute, hostels, Libraries, NCC / NSS etc, and
d) No disciplinary action is pending against him / her.

15.4 The degree shall be awarded after approval by the Academic Council.

16. DISCRETIONARY POWER:

Not with standing anything contained in the above sections, the Vice Chancellor may review all exceptional cases, and give his decision, which will be final and binding.

M.Sc Data Science - Scheme of Instruction

I SEMESTER

Sl. No	Course Code	Name of the Course	Category	Credits	Scheme of Instruction		Total	Scheme of Examination		
					Hours per Week			Duration in Hrs.	Maximum Marks	
					L/T	P			Sem. End Exam	Con. Eval
1	SDS 701	Introduction to Modern Databases	PC	4	4	0	4	3	60	40
2	SDS 703	Python Programming and Data Visualization	PC	4	4	0	4	3	60	40
3	SDS 705	Statistics – I	PC	4	4	0	4	3	60	40
4	SDS 707	Artificial Intelligence	PC	4	4	0	4	3	60	40
5	SSE 705	Internet of Things	SEC	2	3	0	3	3	--	100
	PRACTICALS :									
1	SDS 721	Modern Databases Lab	PP	2	0	4	4	3	--	100
2	SDS 723	Python Programming Lab	PP	2	0	4	4	3	--	100
		Total		22	19	8	27	--	240	460

II SEMESTER

Sl. No.	Course Code	Name of the Course	Category	Credits	Scheme of Instruction		Total	Scheme of Examination		
					Hours per Week			Duration in Hrs.	Maximum Marks	
					L/T	P			Sem. End Exam	Con. Eval
1	SAE 702	Professional Communication Skills	AEC	2	3	0	3	-	-	100
2	SDS 702	Web Programming	PC	4	4	0	4	3	60	40
3	SDS 704	Statistics –II	PC	4	4	0	4	3	60	40
4	SDS 706	Data Mining	PC	4	4	0	4	3	60	40
5	SDS 708	Data Security and Privacy	PC	4	4	0	4	3	60	40
6	SDS 710	Mathematics for Data Science	PC	4	4	0	4	3	60	40
	PRACTICALS :									
1	SDS 722	R Programming Lab	PP	2	0	4	4	3	--	100
2	SDS 724	Web Programming Lab	PP	2	0	4	4	3	--	100
		Total		26	23	8	31	--	300	500

M.Sc Data Science

III SEMESTER

Sl. No.	Course Code	Name of the Course	Category	Credits	Scheme of Instruction		Total	Scheme of Examination		
					Hours per Week			Duration in Hrs.	Maximum Marks	
					L/T	P			Sem. End Exam	Con. Eval
1	SDS 801	Machine Learning	PC	4	4	0	4	3	60	40
2	SDS 803	Big Data Analytics	PC	4	4	0	4	3	60	40
3	SOE XXX	Open Elective	OE	3	3	0	3	3	60	40
4	SDS 841 SDS 843 SDS 845 SDS 847	**Generic Elective – I** Web Analytics Fundamentals of Block Chain Technologies High Performance Computing Data Storage Technologies and Networking	GE	4	4	0	4	3	60	40
5	SDS 849 SDS 851 SDS 853 SDS 855	**Generic Elective – II** Cloud Computing Computational Biology Deep Learning Natural Language Processing	GE	4	4	0	4	3	60	40
	PRACTICALS :									
	SDS 821	Data Analytics Lab	PP	2	0	4	4	4	--	100
	SDS 823	Machine Learning Lab using Python	PP	2	0	4	4	4	--	100
	SDS 825	Industrial Training & Seminar	PP	2	0	2	2	--	--	100
		Total		25	19	10	29	--	300	500

IV SEMESTER

Sl. No.	Course Code	Name of the Course	Category	Credits	Scheme of Instruction		Total	Scheme of Examination		
					Hours per Week			Duration in Hrs.	Maximum Marks	
					L/T	P			Sem. End Exam	Con. Eval
1	SDS 891	Project Work	PP	8	0	3	3	--	50	150

Total Credits: 22 + 26 + 25+ 8 = 81

M.Sc DATA SCIENCE

SEMESTER- I

SDS 701 INTRODUCTION TO MODERN DATABASES

Hours per week: 4 End Examination: 60 Marks

Credits: 4 Sessionals: 40 Marks

Preamble:

This course provides fundamental and practical knowledge on database concepts by means of organizing the information, storing and retrieve the information in an efficient and a flexible way when data is stored in a well-structured model. This course ensures that every student will gain experience in creating data models and database design.

Course Objectives:

- Relate the role of a database management system in an organization.
- Demonstrate basic database concepts, including the structure and operation of the relational data model, non-relational data model.
- Construct simple and moderately advanced database queries using Structured Query Language (SQL) and types of NOSQL documents.
- Explain and successfully apply logical database design principles, including E-R diagrams and database normalization.
- Demonstrate the concept of a database transaction, hashing, indexing.
- Explain different types of database architecture.
- Explain the concept of parallel database & distributed databases.

UNIT-I

Introduction to Relational Database Model: Structure of Relational Database**,** Database Schema, Keys, Schema Diagrams, Relational Query Languages, Relational Operations.

Introduction to SQL : Overview of the SQL query Language**,** SQL data definition **,** Basic structure of SQL queries**,** Additional basic operations**,** Set operations**,** Null values**,** Aggregate functions, Nested subqueries**,** Modification of the database.

Intermediate SQL: Join expressions, Views, Transactions, Integrity Constraints, SQL data types and schemas, Authorization. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Interpret the basic terminology of DBMS like data, database, database management systems.(L2)
- Compare DBMS over File Systems.(L2)
- Create and modify database using SQL query and apply integrity constraints.(L5)
- Illustrate different types of query forms (simple queries, nested queries, and aggregated queries) in SQL.(L2)
- Compare the difference between views and physical tables and working with views.(L2)

UNIT-II

Database design and the E-R Model: Overview of the Design Process**,** the Entity-Relationship model**,** Constraints**,** Removing Redundant Attributes in Entity sets**,** Entity-Relationship Diagrams, Reduction to relational schemas**,** Entity-Relationship design issues**,** Extended E-R features - Specialization and generalization.

Relational Database Design: Features of Good Relational Designs**,** Atomic Domains and First Normal Form, Decomposition using Functional Dependencies**,** Functional-Dependency Theory**,** Decomposition using Multi-valued Dependencies**,** More Normal Forms**,** Database-Design Process and Modeling Temporal Data. (10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Model a given application using ER diagram. (L3)
- Match the integrity constraints from ER model to relational model. (L1)
- Translate an ER Model to Relational Model and vice versa. (L2)
- Make use of schema refinement process. (L3)
- Extend the concept of functional dependencies (fds) and know about anomalies. (L2)
- Illustrates knowledge about different types of normal forms and the importance of normalization. (L2)

UNIT-III

Indexing and Hashing: Basic concepts**,** Ordered Indices, B+ tree index files**,** B+ tree extensions, Multiple-key access, Static Hashing**,** Dynamic Hashing**,** Comparison of ordered indexing and Hashing**,** Bitmap indices**,** Index definition in SQL.

Transactions: Transaction Concept, A Simple Transaction Model, Storage Structure, Transaction Atomicity and Durability**,** Transaction Isolation**,** Serializability**,** Transaction Isolation Levels, Implementation of Isolation Levels**,** Transactions as SQL Statements. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Illustrate the most important high-level file structure tools that include indexing, co-sequential processing, B trees, Hashing. (L2)
- Recognize the difference between various indexing techniques. (L3)
- Recognize the difference between various hashing techniques. (L3)
- Build indexing mechanisms for efficient retrieval of information from databases. (L5)
- Interpret the transaction management in DBMS. (L2)

UNIT-IV

Database System Architecture: Centralized and Client –Server Architectures**,** Server System Architectures **,** Parallel Systems **,** Distributed Systems**,** Network Types.

Parallel Database : Introduction, I/O Parallelism**,** Interquery Parallelism**,** Intra query Parallelism**,** Intra Operation Parallelism**,** Interoperation Parallelism**,** Query Optimization**,** Design of Parallel Systems**,** Parallelism on Multi Core Processor.

Distributed Databases: Homogeneous and Heterogeneous Databases, Distributed Data Storage, Distributed Transactions, Commit Protocols. (10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Understand the techniques of parallel DBMSs, distributed DBMS architectures, distributed database design, query processing, multidatabase systems. (L1)
- Interpret the design principles in distributed database for better resource management. (L2)
- Understand the types of distributed database systems. (L1)

UNIT-V

NOSQL- Value of Relational Databases, impedance mismatch, Application and Integration Databases, Attack of the clusters, Emergence of NoSQL.

Aggregate Data Models – Aggregates, Key-Value and Document Data Models, Column Family Stores.

More Details on Data Models – Relationships, Graph Databases, Schemaless Databases, Materialized Views, Modeling for Data Access.

Distributed Models- Single Server, Sharding, Master-Slave Replication, Peer to Peer Replication**. Consistency-** Update Consistency, Read Consistency, Relaxing Consistency,

Relaxing Durability. **Key-Value Databases:** Introduction to Key Value Store, Features**,** Structure of Data, Scaling**, Document Database:** Introduction, Features. **Column Family Stores –** Introduction, feature. **Graph Databases :** Introduction, Features. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn various NoSQL systems and their features: (L1)
- Compare and use four types of NoSQL Databases (Document-oriented, KeyValue Pairs, Column-oriented and Graph). (L2)
- Demonstrate and understand the detailed architecture, define objects, load data, query data and performance tune using all the four types of databases. (L2)

Course Outcomes

Upon completion of this course, student will be able to:

- Design a data base for a system using E-R data model and Relational Data model.
- Design logical database with all integrity constraints over relations.
- Apply normalization steps in database design and removal of data anomalies.
- Extend the characteristics of database transactions.
- Distinguish the different types of NoSQL databases.

Text Books:

1. Database System Concepts by Abraham Silberschatz, Henry F.korth, S.Sudarshan, McGrawHill, Sixth Edition, 2011.(Unit – I –IV).
2. NoSQL Distilled by Pramod J Sadalage, Martin Fowler, Addison-Wesley Professional; 1st edition, 2012. (Unit – V)

Reference Book:

1. Fundamentals of Database Systems by Ramez Elmasri, Shamkant B.Navathe, Addison-Wesley, Sixth Edition, 2011.

Prepared by: Dr. T. Uma Devi Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE

SEMESTER- I

SDS 703 PYTHON PROGRAMMING AND DATA VISUALIZATION

Hours per week: 4 End Examination: 60 Marks

Credits: 4 Sessionals: 40 Marks

Preamble:

Python is an interpreter oriented, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991. Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales.

Course Objectives:

- To learn the basic concepts and usage of variables, expressions and practice the use of functions in Python programming language.
- To identify and practice different conditionals and implement recursive functions.
- To understand the concepts of strings, lists and dictionaries, practice the use of classes methods, overloading and polymorphism.
- To learn the basic concepts of raw data and use different statistical methods on the data.
- To implement line properties, use different setter methods and practice different kinds of plots.

UNIT- I

The way of the program: Running Python, Arithmetic Operators, Values and types, Formal and natural Languages, debugging.

Variables, expressions and statements: Assignment statements, variable names, expressions and statements, script mode, order of operations, string operations.

Functions: Function calls, math functions, composition, adding new functions, definitions and uses, flow of execution, parameters and arguments. (8 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Explain different types of operators.(L2)
- Develop and run simple Python program.(L3)
- Describe the concepts of variables, expressions and statements.(L2)
- Use functions and develop programs using functions.(L3)
- Extend the concept of functions using parameters.(L2)

UNIT - II

Conditionals and Recursion: Floor division and modulus, Boolean expressions, logical operators, conditional execution, alternative execution, chained conditionals, nested conditionals, recursion, stack diagrams for recursive functions, infinite recursion.

Fruitful Functions: Return values, incremental development, composition, Boolean functions.

Iteration: Reassignment, updating variables, while statement, break, square roots. (9 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Use the logical operators in programming.(L3)
- Identify the need of recursive functions.(L1)
- Construct programs using while statements.(L3)
- Explain the use of break statement.(L2)

UNIT – III

Strings: string length, traversal with for loop, string slices, searching, looping and counting, string methods, in operator, string comparison.

Lists: traversing a list, list operations, list slices, list methods, map, filter and reduce, deleting elements, lists and strings, objects and values, aliasing, list arguments.
Dictionaries: looping and dictionaries, reverse lookup, dictionaries and lists, memos, global variables.
Classes and objects: programmer defined types, attributes, rectangles, instances as return values. **Classes and methods:** object oriented features, init method, str method, operator overloading, polymorphism. (10 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Construct programs to perform operations on strings.(L3)
- Explain basic concepts related to lists.(L2)
- Outline the concepts in dictionaries.(L2)
- Develop simple programs using class.(L3)
- Apply operator overloading and polymorphism.(L3)

UNIT - IV

Getting Started with Raw Data: The world of arrays with NumPy, Empowering data analysis with pandas, Data cleansing, Data operations.
Inferential Statistics: Various forms of distribution, A z-score, A p-value, One-tailed and two-tailed tests, Type 1 and Type 2 errors, A confidence interval, Correlation, Z-test vs T-test, The F distribution, The chi-square distribution, Chi-square for the goodness of fit, The chi-square test of independence, ANOVA. (8 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Explain the basic analytics that can be applied on data.(L2)
- Explain various forms of distribution.(L2)
- Develop programs using different statistical functions.(L3)

UNIT - V

Making Sense of Data through Advanced Visualization : Controlling the line properties of a chart, Using keyword arguments, Using the setter methods, Using the setp() command, Creating multiple plots, Playing with text, Styling plots, Box plots, Heatmaps, Scatter plots with histograms, A scatter plot matrix, Area plots, Bubble charts, Hexagon bin plots, Trellis plots, 3D plot of a surface. **Plotting and Visualization:** A brief matplotlib API primer, Plotting functions in pandas, Plotting Maps, Python Visualization Tool Ecosystem. (10 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Use keyword arguments and setter methods.(L3)
- Identify the need of plots.(L1)
- Explain basic concepts of charts.(L2)
- Illustrate the usage of visualization Tool Ecosystem.(L3)

Course Outcomes:

Upon completion of this course, student will be able to:

- List the different types of operators.
- Understand the concept of variables, expressions and statements.
- Understand the concept of functions and recursive functions.
- Identify the use of iteration.
- List the operations that can be performed on strings.
- Identify the differences between lists and dictionaries.
- List the concepts of operator overloading and polymorphism.

- Understand various forms of distribution.
- List the different types of plots.

Text Books:

1. Mastering Python for Data Science by Samir Madhavan, PACKT Publishing, 2015.
2. Think Python by Allen Downey O'Reilly Publications, 2nd Edition, 2016.

Reference Books:

1. Programming Python by Mark Lutz, O'Reilly Publications, 4th Edition,2011.
2. Python in a nutshell by Alex Martelli, Anna Ravenscroft, Steve Holden, O'Reilly Publications, 3rd Edition, 2017.

Prepared by: Mr.M.Suresh Kumar Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER- I
SDS 705 STATISTICS – I

Hours per week: 4 End Examination: 60 Marks
Credits: 4 Sessionals: 40 Marks

Preamble:

Probability theory is important when it comes to evaluating statistics. This course treats the most common discrete and continuous distributions, showing how they find use in decision and estimation problems, and constructs computer algorithms for generating observations from the various distributions.

Course Objectives:

- To understand the collection, analysis, interpretation, and presentation of data.
- To understand the difference between discrete and continuous random variables and the meaning of probability.
- To evaluate problems on discrete and continuous probability distributions.
- To understand the concept of testing of hypothesis for large and small samples.
- Ability to explore certain statistical concepts in practical applications of computer science areas.

UNIT – I

Descriptive Statistics & Probability : Collection of data, Classification and Tabulation of data, Diagrammatic and Graphic representation of data, Measures of Central value (mean, median, mode), Measures of Dispersion (Quartile deviation, mean deviation, standard deviation), Definition of probability, Addition theorem, Multiplication theorem, Bayes' theorem. (8 hours)

Learning outcomes:

By the end of this Unit, the student will be able to

- Explain the diagrammatic and graphical representation of data. (L2)
- Describe the basic concepts of Measures of central tendency.(L3)
- Describe the basic concepts of Measures of dispersion.(L3)
- Define probability.(L1)
- Evaluate problems on addition theorem, multiplication theorem and Bayes' theorem. (L5)

UNIT – II

Discrete Probability Distributions: Mathematical expectation, Random variable and Probability Distribution, Binomial distribution, Negative binomial distribution, Poisson distribution, Hyper geometric distribution. (8 hours)

Learning outcomes:

By the end of this Unit, the student will be able to

- List the difference between discrete random variable and continuous random variables.(L1)
- Describe the basic concepts of probability distributions.(L3)
- Explain Binomial and Poisson distributions. (L5)
- Explain difference between Binomial and Poisson distributions. (L5)
- Explain Hyper geometric distribution. (L5)

UNIT – III

Continuous Probability Distributions: Normal distribution, Relation between Binomial, Poisson and Normal distributions, Properties of the Normal distribution, Area under the Normal curve, fitting a normal curve. (8 hours)

Learning outcomes:

By the end of this Unit, the student will be able to

- Evaluate difference between Binomial, Poisson and Normal distributions.(L5)
- Explain the need of Normal distribution.(L5)
- Explain properties of Normal distribution.(L5)
- Explain area under the bell shaped curve using Normal distribution.(L5)
- Evaluate Normal curve with given data.(L5)

UNIT- IV

Statistical Inference – I : Procedure of Testing Hypothesis, Standard error and Sampling Distribution, Estimation, **Test of significance large samples** : Difference between small and large samples, Two tailed test for difference between the means of two samples, Standard error of the difference between two standard deviations. (10 hours)

Learning Outcomes

By the end of this Unit, the student will be able to

- Explain the procedure of testing of hypothesis.(L5)
- Evaluate standard error. (L5)
- Explain tests of significance for large samples.(L5)
- Explain tests of significance for small samples.(L5)
- Explain difference between small and large samples.(L5)

UNIT- V

Statistical Inference – II: Tests of significance for small samples, Student's t-distribution, χ^2 –test and goodness of fit, F-test. (8 hours)

Learning Outcomes

By the end of this Unit, the student will be able to

- Explain tests of significance for small samples using t-distribution. (L5)
- Explain chi-square test and goodness of fit. (L5)
- Explain F- test.(L5)
- Apply the concept of χ^2 –test to given data.(L5)
- Apply the concept of F –test to given data.(L5)

Course Outcomes:

Upon completion of this course, student will be able to:

- To identify different sample collection methods and interpretation and presentation techniques those are available.
- To learn and evaluate discrete and continuous probability distributions.
- To know the hypothesis testing for large and small samples.

Text Book :

1. Statistical Methods by S.P. Gupta, Sultan Chand & Sons publication, 44th Edition, 2017.

Reference Books:

1. Probability and Statistics for Engineering and Sciences by Jay L.Devore, Cengage Learning, 2015.
2. Probability and Statistics for Engineers and Scientists by Ronals E.Walpole, Raymond Mayers, Sharon L.Myers, Keying E. Ye, Pearson Publication, Ninth Edition, 2014.
3. Probability and Statistics for Science and Engineering by Shankar Rao, University Press, 2015.

Prepared by: Prof. N. Ravishankar Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER- I
SDS 707 ARTIFICIAL INTELLIGENCE

Hours per week: 4 End Examination: 60 Marks
Credits: 4 Sessionals: 40 Marks

Preamble:

This course enables the students to think critically about what makes humans intelligent, and how computer scientists are designing computers to act more like us. AI plays an important role in the design and development of systems with intelligent behavior. The primary objective of this course is to provide an introduction to the basic principles and applications of Artificial Intelligence.

Course Objectives:

- To teach fundamentals of Artificial Intelligence, the concept of Intelligent Agents and problem-solving process through uninformed and informed searches.
- To gain an insight into competitive environments which gives rise to adversarial search problems, often known as games.
- To view many problems in AI as problems of constraint satisfaction.
- To gain complete idea of knowledge representation techniques Propositional and First-order logics.
- To learn how to trace the inference mechanism in First-order logics.

UNIT-I

Introduction: AI definition, Foundations, History, State of the Art.

Intelligent Agents: Agents and Environments, Concept of Rationality, Nature of Environments, Structure of Agents.

Solving Problems by Searching: Problem Solving Agents, Example problems, Searching for solutions, Uninformed Search Strategies, Informed search strategies, Heuristic Functions, Local Search Algorithms and Optimization Problems. (10 hours)

Learning outcomes

After completion of this unit, student will be able to

- Define Artificial Intelligence. (L1)
- How agents work in environments. (L1)
- Recall uninformed search techniques. (L1)
- Illustrate the working of informed search techniques. (L2)

UNIT-II

Adversarial Search: Games, Optimal Decisions in Games, Alpha- Beta Pruning, Imperfect real-time decisions, Stochastic games, Partially observable games, State of art game program. (8 hrs)

Learning outcomes

After completion of this unit, student will be able to

- Understand how games improve intellectual abilities of humans. (L1)
- Choose optimal decisions in games. (L1)
- Illustrate alpha-beta pruning. (L2)
- Compare stochastic and partially observable games. (L2)

UNIT-III

Constraint Satisfaction Problems: Defining Constraint Satisfaction problems, Constraint Propagation, Backtracking search for CSPs, Local Searches for CSPs.

Logical Agents: Knowledge-Based Agents, Wumpus World, Logic, Propositional Logic and Propositional Theorem proving. (8 hours)

Learning outcomes

After completion of this unit, student will be able to

- Define constraint satisfaction problems. (L1)
- Illustrate inference in constraint satisfaction problems. (L2)
- Contrast backtracking search and local search for constraint satisfaction problems. (L2)
- Define knowledge-based agents. (L1)
- How to represent real-world facts in propositional logic. (L1)

UNIT-IV

First Order Logic: Syntax and Semantics of First Order Logic, Using First Order Logic, Knowledge Engineering in First Order Logic.

Inference in First Order Logic: Propositional Vs First Order Inference, Unification and Lifting, Forward Chaining, Backward Chaining, Resolution. (8 hours)

Learning outcomes

After completion of this unit, student will be able to

- Infer proofs in propositional and first-order logic. (L2)
- Define propositional and first-order inference. (L1)
- Outline unification and lifting. (L2)
- Experiment with forward chaining and backward chaining. (L3)
- Make use of resolution. (L3)

UNIT-V

Classical Planning: Definition, Algorithms for planning as state space search, Planning Graphs.

Knowledge Representation: Ontological Engineering, Categories and Objects, Events, Mental Events and Mental Objects, Reasoning Systems for Categories, Reasoning with Default information, Internet Shopping World. (8 hours)

Learning Outcomes

After completion of this unit, student will be able to

- Overview on basic knowledge representation aspects and on ontologies. (L2)
- Design and generate path planning using knowledge representation. (L5)
- Accumulate sophisticated knowledge about the environment for processing tasks or methods. (L5)

Course Outcomes

Upon completion of this course, student will be able to:

- Illustrate artificial intelligence, the role of intelligent agents, uninformed and informed search techniques.
- Examine competitive environments like game problems.
- Interpret many real-world problems as constraint satisfaction problems.
- Illustrate what knowledge representation is and able to distinguish propositional and first-order logics.
- Infer proofs using resolution in first-order logic.
- A machine-interpretable representation of the world, similar to human reasoning can be designed

Text Book:

1. Artificial Intelligence - A Modern Approach by Stuart J. Russell, Peter Norvig, Pearson Education, 3rd Edition. 2015.

Reference Books:

2. Artificial Intelligence by Kevin Knight, Elaine Rich, 3rd Edition, TMH,2017.
3. Artificial Intelligence by Saroj Kaushik, Cengage Learning, 2011.

Prepared by: Dr. T. Uma Devi Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER- I
SSE 705 INTERNET OF THINGS

Hours per week: 2 Continuous Evaluation: 100 Marks
Credits: 2

Preamble:

The course presents the introduction to Internet of things and enables the student to learn about the concept of Internet Principles, Prototyping, Prototyping Embedded Devices.

Course Objectives:

- To introduce the concept of Internet of things.
- To utilize Internet Principles, Prototyping, Prototyping Embedded Devices.

The Internet of Things- An Overview : The flavor of the Internet of Things ,The Internet of Things , The Technology of the Internet of Things, Enchanted Objects, Who is Making the Internet of Things.

Internet Principles : Internet Communications- An Overview of IP ,TCP, The IP Protocol Suite (TCP/IP),UDP., IP Addresses, DNS, Static IP Address Assignment, Dynamic IP Address Assignment, IPv6. MAC Addresses, TCP and UDP Ports, Application Layer Protocols.

Prototyping: Sketching, Familiarity, Costs versus Ease of Prototyping, Prototypes and Production, Open Source versus Closed Source, Tapping into the Community.

Prototyping Embedded Devices: Electronics, Embedded Computing Basics, Arduino, Raspberry Pi, BeagleBone Black, the Electric Imp.

Prototyping the Physical Design: Preparation, Sketch, Iterate and Explore, Non-digital Methods, Laser Cutting, 3D Printing, CNC Milling, Repurposing/Recycling.

Course Outcomes

Upon completion of this course, student will be able to:

- The course spells an exciting new world where the internet and real life objects work in harmony.
- It explains how to design interactive devices that can sense and respond to changes in their environment.
- It develops the ways to combine sensors, Arduino boards, Rasberry Pis and other components with networks and Internet.
- The course provides good grounding in building Internet of things device.

Text Book:

1. Designing the Internet of Things by Ardrian MCEwen, Hakim Cassimally, Wiley Publications, 2013.

Reference Books:

1. Building Internet of Things with the Arduino by Charalampos Doukas, Create space, 2002.
2. Architecting the Internet of Things by Dietet Uckelman, Mark Harrison, Florian Michahelles, Springer, 2011.

Prepared by: Prof.K.Vedavathi Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE

SEMESTER- I

SDS 721 MODERN DATABASES LAB

Hours per week: 4 Continuous Evaluation: 100 Marks

Credits: 2

(a) Relational Databases:

1. Implement DDL Statements.
2. Implement DML Statements.
3. Write the queries for implementing Built-In Functions.
4. Design a Database and create required tables. Apply the constraints like Primary Key, Foreign Key, Not Null, and other constraints to the tables. Perform SQL Queries.
5. Write the queries to implement the Joins.
6. Write the queries to implement subqueries and correlated queries.

Course Outcomes:

Upon completion of this course, the student should be able to:

- Design and implement a database schema for a given problem-domain.
- Populate and query a database using SQL DML/DDL commands.
- Declare and enforce integrity constraints on a database
- Retrieve data using different SQL joins, subqueries and correlated queries.

(b) **NoSQL Databases:**

1. Demonstrate to create Columnar, Big Table, Document Databases, Graph databases
2. Demonstrate to insert, update and delete data in different types of databases.
3. Demonstrate various techniques used to query the database.
4. Explain techniques to optimize querying using indexing.
    - Use different **FIND** techniques to query the document in mongodb.
    - Use the **sort()** to sort the records.
    - Use the logical operators to query the document.
    - Use the conditional operators in mongodb to query the document .
5. Demonstrate the methods to analyze data using aggregation techniques.
6. Explain the techniques of splitting data across machines.
7. Demonstrate **LIMIT()** and **SKIP()** methods in mongodb.
8. Demonstrate example for **Text Searching** in Mongodb.
9. Demonstrate the usage different remove methodologies to remove documents from collections.

Course Outcomes:

Upon completion of this course, the student is able to:

- •Utilize the techniques used to create, insert, update and delete data/documents.(L3)
- • Utilize various techniques used to query the database. (L3)
- • Utilize techniques to optimize querying using indexing.(L3)
- • Apply methods to analyze data using aggregation techniques. (L3)
- • Adopt knowledge about the role of NoSQL in business. (L6)
- • Identify technique of splitting data across machines via sharding. (L3)

- Select various design aspects and operations of MongoDB.(L1)
- Define objects, load data, query data and performance tune Key-Value Pair NoSQL databases.(L3)

Text Books:

1. SQL, PL/SQL the Programming Language of Oracle by Ivan Bayross, BPB Publications, 4th Edition,2010.
2. MongoDB: The Definitive Guide by Kristina Chodorow, Shroff publisher, 2nd edition, 2013.
3. NoSQL with Mongo DB in 24 hours by Brad Dayley, Pearson Education, 1st edition, 2015.

Prepared by: Dr. T. Uma Devi Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE

SEMESTER- I

SDS 723 PYTHON PROGRAMMING LAB

Hours per week: 4 Continuous Evaluation: 100 Marks

Credits: 2

1. Find all numbers which are multiples of 17, but not the multiples of 5, between 2000 and 2500?
2. Swap two integer numbers using a temporary variable. Repeat the exercise using the code format: a, b = b, a. Verify your results in both the cases.
3. Given two pairs of Cartesian points such as (x1, y1) and (x2, y2). Find the Euclidian distance between them.
   Hint: Use math module to find the square root.
4. Print the first 2 and last 3 characters in a given string. Use the string slicing concept. Do not use loops. If the length of the string is less than 5, print a suitable message.
5. Implement bubble sort. Do not use the default sort() method.
   Hint: So as to familiarize with the concept of sorting, and nested looping structures.
6. Implement shallow copy and deep copy of a list. You may use the copy module.
   Hint: While we copy a list, just a reference is copied. Hence if we make any changes to one of the lists, the same will reflect in the other as well. This is called shallow copying. Hence, in some cases we might need to deep copy, where a completely independent copy is created. This can be achieved through the deepcopy() method of copy module.
7. Write a temperature converter program, which is menu driven. Each such conversion logic should be defined in separate functions. The program should call the respective function based on the user's requirement. The program should run as long as the user wishes so.
8. Find the largest of n numbers, using a user defined function largest().
9. Write a function that capitalizes all vowels in a string.
   Hint: Do not use the ASCII concept. Use the upper() method.
10. Write a function leapYear() which receives a four digit year and returns a Boolean value: True if the year is leap, False if the year is not leap.
11. Read a line containing digits and letters. Write a program to give the count of digits and letters.
    Hint: Instead of checking ASCII, use the in-built methods like isdigit(), isalpha() etc.
12. Write a function myReverse() which receives a string as an input and returns the reverse of the string.
13. Use the list comprehension methodology in python, to generate the squares of all odd numbers in a given list.
    Hint: List comprehension is one of the powerful techniques in python;
14. Check if a given string is palindrome or not.
    Hint: do not use the C philosophy where we compare indices. Instead, copy the string as a new list, reverse the list using reverse(), join the list so that the reversed string is formed, using join(). Compare the new string and the old one.
15. Write a function to see if a given number is prime or not. Do not use any flag variables. Use math module to find the square root, and its roof which will be fed in to range().
    Hint: Just the return statements are enough. No need of flag variables. The loop has to run up to the roof of the square root of the number.
16. Write a function to find the factorial of a number using recursion.
17. Extend the above problem to find the nCr of given values of n and r. Verify the result with the help of itertools module, which helps to find the combinations.
18. Write a program that eliminates duplicates in a list. Do not use the concept of sets. Now, convert the

original list into a set. Verify the result in both cases.

19. The user will enter five integers separated by commas. Write a program to read these values, and make a list. Print the list.
    Hint: They will need to read the input using raw_input(), and then split the one and only line of input using split(). Then each of the values will need to be appended to a list, which will be empty at first.
20. Generate a dictionary and print the same. The keys of the dictionary should be integers between 1 and 10 (both inclusive). The values should be the cubes of the corresponding keys.
21. Create a nested dictionary. The roll number of a student maps to a dictionary. This inner dictionary will have name, age, and place as keys. Read details of at least three students.
    Hint: A sample output should look like the one given below:
    {11: {'name': 'Sachin', 'age': 18, 'place': 'Kochi'}, 12:
    {'name':'Ammu', 'age': 19, 'place': 'Kannur'}, 13:
    {'name':'jishad', 'age':20, 'place':'Calicut'}}
22. Enter a word. Create a dictionary with the letters of this word as keys, and the corresponding ASCII values as values.
    Hint: Students may use the ord() function. Further, this is a simple problem, if list comprehension is used.
23. Implement Multiple plots.
24. Implement Scatter plots with histogram.
25. Implement Bubble charts.

Course Outcomes:

Upon completion of this course, student will be able to:

- Able to develop programs in Python.(L4)
- Able to implement functions using parameters.(L)
- Understand the concept of expressions.(L2)
- Construct programs using while statement.(L6)
- List the operations that can be performed on strings.(L4)
- Identify the differences between lists and dictionaries.(L4)
- List the concepts of operator overloading and polymorphism.(L4)
- Understand various forms of distribution.(L2)
- Able to implement plots and charts.(L4)

Text Books:

1. Mastering Python for Data Science by Samir Madhavan, PACKT Publishing,2015.
2. Think Python by Allen Downey O'Reilly Publications, 2nd Edition,2016.

Prepared by: Mr. M. Suresh Kumar Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER – II
SAE 702 PROFESSIONAL COMMUNICATION SKILLS

Hours per week: 3 Continuous Evaluation: 100 Marks
Credits: 2

Preamble:

This course is designed to expose students to the basics of academic and professional communication in order to develop professionals who can effectively apply communication skills, theories and best practices to meet their academic, professional and career communication needs.

Course Objectives:

To enable students to

- acquaint themselves with basic English grammar.
- acquire presentation skills.
- develop formal writing skills.
- develop creative writing skills.
- keep themselves abreast with employment-readiness skills.

UNIT - I

BACK TO BASICS: Tenses, Concord – Subject Verb Agreement, Correction of Sentences-Error Analysis, Vocabulary building. (10 hours)

Learning Outcomes:

At the end of the unit, the student will be able to:

- Utilize structures and tenses accurately. (L3)
- Apply the right verb to the right subject in a sentence. (L3)
- Identify incorrect sentences in English and write their correct form. (L3)
- Choose new vocabulary and use in speaking and writing.(L1)

UNIT - II

ORAL PRESENTATION: What is a Presentation? Types of Presentations, Technical Presentation – Paper Presentation, Effective Public Speaking, Video Conferencing. (8 hours)

Learning Outcomes:

At the end of the unit, the student will be able to:

- Recall how to overcome speaking anxiety prior to presentation. (L1)
- Plan and structure effective presentations that deliver persuasive messages. (L3)
- Prioritize slides that can catch the attention of the audience. (L5)
- Show the skills in organizing, phrasing, and expressing the ideas, opinions and knowledge. (L1)
- Formulate and participate in a video conference effectively.(L6)

UNIT III

DOCUMENTATION: Letter Writing, E-mail Writing & Business Correspondence, Project Proposal, Report Writing, Memos, Agenda, Minutes, Circulars, Notices, Note Making.(10 hours)

Learning Outcomes:

At the end of the unit, the student will be able to:

- Outline a business letter, which includes appropriate greetings, heading, closing and body and use of professional tone.(L2)
- Develop crisp and compelling emails.(L6)
- Develop project proposals, reports and memos.(L6)
- Prepare agenda and draft minutes.(L6)

- Prepare circulars, notices and make notes. (L6)

UNIT IV

CREATIVE WRITING: Paragraph Writing, Essay writing, Dialogue Writing, Précis Writing, Expansion of Hints, Story Writing. (6 hours)

Learning Outcomes:

At the end of the unit, the student will be able to:

- Outline paragraphs on familiar and academic topics using a topic sentence, supporting detail sentences and a conclusion sentence.(L3)
- Select the structure of a five-paragraph essay and write essays that demonstrate unity, coherence and completeness. (L3)
- Structure natural, lucid and spontaneous dialogues. (L3)
- Examine the elements of a short story and develop their functional writing skills.(L3)

UNIT V

PLACEMENT ORIENTATION: Resume preparation, group discussion – leadership skills, analytical skills, interviews –Types of Interviews, Preparation for the Interview, Interview Process. (8 hours)

Learning Outcomes:

At the end of the unit, the student will be able to:

- Formulate professional resume that highlights skills, specific to the student's career field. (L6)
- Adapt the personality traits and skills required to effectively participate in a G.D.(L6)
- Infer the purpose of interviews. (L2)
- Aware of the processes involved in different types of interviews.(L3)
- Plan on how to prepare for an interview.(L4)
- Conclude on how to answer common interview questions.(L4)

Course Outcomes:

Upon completion of this course, student will be able to:

- Develop formal writing skills.
- Aware of the processes involved in interviews.
- keep themselves with employment-readiness skills.

Text Books :

1. Essentials of Business Communication by Rajendra Pal and J S KorlahaHi, Sultan Chand & Sons.
2. Advanced Communication Skills by V. Prasad, Atma Ram Publications.
3. Effective Communication by Ashraf Rizvi, McGraw Hill Education; 1st Edition , 2005.
4. Interviews and Group Discussions How to face them by T.S.Jain, Gupta,1st Edition, Upkar Prakashan,2010.
5. High School English Grammar and Composition by P.C.Wren & Martin, N.D.V.Prasada Rao S.Chand.

Prepared by: Dr. Sushma Raj Verified by: Dr. T. Uma Devi

M.Sc DATA SCIENCE

SEMESTER – II

SDS 702 WEB PROGRAMMING

Hours per week: 4 End Examination: 60 Marks

Credits: 4 Sessionals: 40 Marks

Preamble:

This course enables the students to associate with developing websites for hosting via intranet or internet. The web development process includes web design, web content development, client-side scripting, server-side scripting. Web development is the coding or programming that enables website functionality as per the owner's requirements.

Course Objectives:

- Design static web page using Markup languages.
- Design and implement web applications using style sheets.
- Use of java script for designing web applications with dynamic effects.
- Validations on form input entry and adding dynamic content to web applications.
- The notions of Web servers and Design Methodologies with MVC Architecture.
- Creation of adaptive web pages and implementing cookies.
- Design and implementation of complete applications over the web.

UNIT - I

Overview of HTML5 and Other Web Technologies: Introduction to Internet, Web and Web technologies, HTML5 and its Essentials, New Features of HTML5**,** Structuring an HTML Document - Elements and Attributes , Tags, The DOCTYPE Element, Exploring Editors and Browsers Supported by HTML5, Creating, Saving, Validating ,Viewing a HTML Document, Hosting Web Pages. **Fundamentals of HTML**: Understanding Elements, Describing Data Types, Horizontal Rules, Line Breaks, Paragraphs, Citations, Quotations, Definitions, Comments, Working with Text, Organizing Text in HTML, Exploring Hyperlinks, URL, Understanding and Describing the Table Elements, Inserting Images, Exploring Colors. **Working with Forms**: Exploring the FORM element, Types of INPUT Element, Exploring Button, Multiple Choice, Text Area, Label, Fieldset, Legend, Datalist, Keygen, Output elements, submitting a Form. (10 hours)

Leaning Outcomes:

After completion of this unit, student will be able to

- Understand various steps to design static websites. (L2)
- Identify the importance of HTML tags for designing web page.(L3)
- Able to develop a static web page along with user interactive elements. (L5)

UNIT – II

Working with Multimedia: Exploring Audio and Video File Formats, Describing the Multimedia Elements, defining a Multimedia File Using the EMBED, OBJECT Element, Exploring the FIGURE and FIGCAPTION Elements. **Overview of CSS**: Evolution, Syntax of CSS, Exploring CSS selectors, Inserting CSS in a HTML Document, Exploring Background, Color, Font Properties of a Web page, Properties Table: Using the style Attribute, Creating Classes and IDs, Generating External Style Sheets, Typography, Consistency, Types of styles, Specifying class within HTML document, Style placement: Inline style, Span & div tags, header styles, Text and font attributes: Font Vs CSS, changing fonts, text attributes, Advance CSS properties: Backgrounds, Box properties and Positioning. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Separate design from content using various levels of Style Sheets.(L2)
- Learn different types of style sheets. (L2)

UNIT – III

Java Script: Features, Using Java Script in a HTML Document, Exploring Programming Fundamentals of JavaScript, Strings, Exploring Functions, Events, Image Maps, Animations. (10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Use Java script to validate user input and perform dynamic documents.(L3)
- Design dynamic and interactive web pages by embedding Java script code in HTML. (L4)

UNIT – IV

PHP: Introducing PHP, History, Unique Features, Basic Development Concepts, Creating First PHP Script, Mixing PHP with HTML, Escaping Special Characters, Using Variables and Operators, Controlling Program Flow, Working with Arrays. **File Handling in PHP**: File operations like opening, closing, reading, writing, appending, deleting etc. on text and binary files, listing directories. (10 hours)

Leaning Outcomes:

After completion of this unit, student will be able to

- Understands the components of PHP.(L2)
- Learn the basic constructs of PHP, built in functions.(L2)
- Learn the importance of PHP for web application development.(L3)
- Able to develop a simple PHP application.(L4)

UNIT – V

Introducing JSON: History of JSON, JSON vs XML, Typical Uses of JSON, JSON DATA Structures, JSON Syntax, Data Types, Creating JSON Objects, Parsing JSON, JSON Data Persistence, Data Interchange, Cross Origin Resources, Posting JSON, Working with templates, JSON with PHP. (8 hours)

Leaning Outcomes:

After completion of this unit, student will be able to

- Learn the features of JSON. (L2)
- Understand the difference between JSON and XML, JSON and RDBMS.(L3)
- Learn JSON Format to serialize and transmit structured data over the internet.(L3)

Course Outcomes:

Upon completion of this course, student will be able to:

- Demonstrate the importance of HTML & DHTML tags for designing web pages and separate design from content using Cascading Style Sheet. (L2)
- Understand various steps to design dynamic websites. (L2)
- Design interactive web pages with client and server side scripting. (L2)
- Apply validations on user input using java script. (L3)
- Understands the PHP framework and develop a reusable component.(L2)
- Apply JSON for storing information in an organized manner. (L3)

Text Books:

1. HTML 5 Black Book , CSS 3, Java Script, XML, XHTML, AJAX, PHP and JQuery by DT Editorial Services, , Dream Tech Press, 2nd Edition, 2016.
2. PHP: A Beginner's Guide by Vikram Vaswani, Tata McGraw Hill, 2017.
3. JSON for Beginners by Icode Acadmey, 2107.

Reference Books:

1. HTML5 and CSS3 by Elizabeth Castro & Bruce Hyslop, , Pearson, 7th Edition 2012
2. HTML 5 in simple steps by Kogent Learning Solutions, Dream Tech, 2010.
3. Programming PHP by Rasmus Lerdorf and Kevin Tatroe, Oreilly Publication, 1st edition, 2002
4. Beginning JSON by Ben Smith, Apress publisher, 1st Edition, 2015.

Prepared by: Dr. T. Uma Devi

Verified by: Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER – II
SDS 704 STATISTICS - II

Hours per week: 4 End Examination: 60 Marks
Credits: 4 Sessionals: 40 Marks

Preamble :

This course covers the concepts on Correlation and Regression Analysis, Single and multi factor ANOVA, techniques for developing and validating a model, introduction to non parametric procedures and statistical control techniques

Course Objectives:

- To understand different correlation types.
- To learn the basic concepts and applications of correlation coefficient and rank correlation coefficient.
- To identify and practice the difference between correlation and regression analysis.
- To acquire the ability to implement features of ANOVA and experimental design.
- To understand the concept of non- parametric tests.
- To implement features of statistical quality control.

UNIT – I

Correlation Analysis: Types of Correlation, Methods of Studying Correlation, Karl Pearson's Coefficient of Correlation, Rank Correlation Coefficient, Partial Correlation, Multiple Correlation. (8 hours)

Learning Outcomes:

By the end of this Unit, the student will be able to

- Explain different types of correlation.(L2)
- Demonstrate basic concepts of correlation coefficient and rank correlation coefficient.(L2)
- Build correlation coefficient for given ungrouped data.(L3)
- Estimate rank correlation coefficient for given ungrouped data.(L4)

UNIT – II

Regression Analysis: Uses of Regression Analysis, Difference between Correlation and regression Analysis, Regression Lines, Regression Equations, Multiple Regression Analysis. (8 hours)

Learning Outcomes:

By the end of this Unit, the student will be able to

- List the difference between correlation and regression analysis. (L1)
- Discuss basic concepts of regression analysis. (L6)
- Evaluate the problems on regression lines (X on Y and Y on X) (L4)
- Demonstrate basic concepts of multiple regression analysis.(L2)

UNIT - III

Analysis of Variance (ANOVA) & Experimental Design: Assumptions of ANOVA, Techniques of ANOVA, Coding of Data, ANOVA in Two-Way Classification Model, Random Block Design, Latin Squares, Randomized Blocks Vs Latin Square, Factorial Experiment. (10 hours)

Learning Outcomes:

By the end of this Unit, the student will be able to

- Model features of ANOVA and experimental design.(L3)
- Explain the need of Analysis of Variance.(L2)
- Outline the different types of experimental designs. (L3)
- Solve problems on ANOVA in two-way classification model.(L3)

- Develop the concept of factorial experiment.(L3)

UNIT - IV

Non-Parametric Tests: Advantages Of Non-Parametric Tests, The Sign Test, A Rank Sum Test; The Man-Whitney U-Test, The One Sample Runs Test, The Kruskal-Wallis Or H-Test. (8 hours)

Learning Outcomes:

By the end of this Unit, the student will be able to

- Explain the advantages of non-parametric tests .(L2)
- Summarize different types of non-parametric tests.(L2)
- Explain and evaluate sign test .(L2)
- Explain and evaluate run test.(L2)

UNIT - V

Statistical Quality Control: Control Charts, $\overline{X}$ Chart, R Chart, Control Chart for C, Control Chart For P, Acceptance Sampling. (8 hours)

Learning Outcomes:

By the end of this Unit, the student will be able to

- Compare and contrast control charts.(L2)
- Explain control charts for C .(L2)
- Explain R chart. (L3)
- Explain control charts for P. (L3)
- Evaluate acceptance sampling. (L3)

Course Outcomes:

Upon completion of this course, student will be able to:

- Understand concept of correlation, regression and applications of correlation coefficient and rank correlation coefficient.
- Acquire the ability to implement features of ANOVA and experimental design.
- Perform data analysis using non- parametric tests.
- Execute statistical quality control.

Text Book :

1. Statistical Methods by S.P. Gupta, Sultan Chand & Sons Publication, 44th Edition, 2017.

Reference Books:

1. Probability and Statistics for Engineering and Sciences by Jay L.Devore, Cengage Learning, 2015.
2. Probability and Statistics for Engineers and Scientists by Ronals E.Walpole, Raymond H.Mayers, Sharon L.Myers, Keying E.Ye, Pearson Publication, Ninth Edition, 2014.
3. Probability and Statistics for Science and Engineering by Shankar Rao, University Press, 2015.

Prepared by: Prof. N. Ravi Shankar Verified by: Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER – II
SDS 706 DATA MINING

Hours per week: 4 End Examination: 60 Marks
Credits: 4 Sessionals: 40 Marks

Preamble:

Due to advent of technology, internet, and advanced applications like social media, huge amount of digital data has been accumulated in data centers/cloud databases, which has led to a situation "we are drowning in data but starving from knowledge". To find golden nuggets which are useful for decision making process, various data mining functionalities like association analysis, classification, clustering, outlier analysis and web mining used. Data warehousing (DW) is an integral part of knowledge discovery process, where DW is an integration of multiple heterogeneous data repositories under a unified schema at a single site. The students will acquire knowledge in data modeling, design, architecture, data warehouse implementation and further development of data cube technology.

Course Objectives:

- Understand the importance of Data Mining and its applications.
- Introduce various types of data and pre-processing techniques.
- Learn various multi-dimensional data models and OLAP Processing.
- Study concepts of Association Analysis.
- Learn various Classification methods.
- Learn basics of cluster analysis.

UNIT – I

Introduction: Need for Data Mining, Definition of Data Mining, Kinds of data, Kinds of patterns to be mined, Technologies used, applications, Major issues in Data Mining.

Data Preprocessing: Need for Preprocessing the Data, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization. (10 hours)

Learning outcomes

After completion of this unit, student will be able to

- Understand basic concepts of data mining. (L2)
- Learn the KDD process. (L2)
- Learn different data mining tasks. (L2)
- Learn major challenges in the field of data mining. (L2)
- Understand various types of data sets and attributes. (L2)
- Apply different statistical techniques on different types of attributes to measure the similarities and dissimilarities. (L3)
- Learn different data preprocessing techniques and apply them on data sets. (L2)

UNIT – II

Data Warehouse and OLAP Technology: Data Warehouse – basic concepts, Data Cube and OLAP Technology, Design and Usage, implementation, Data Generalization by Attribute-Oriented Induction. (8 hours)

Learning outcomes

After completion of this unit, student will be able to

- Learn the basics of data warehousing and different OLAP operations. (L2)
- Understand the relationship between data warehousing and other data generalization methods.(L2)
- Study the methods of data cube computation.(L2)

- Explorations of data cube and OLAP technologies.(L4)

UNIT – III

Mining Frequent Patterns, Associations and Correlations: Basic Concepts and Methods-Basic Concepts, Frequent itemset Mining methods, Pattern Evaluation methods.

Advance Pattern Mining: Pattern mining in multilevel, multidimensional space, Constraint based Frequent Pattern Mining. (8 hours)

Learning outcomes

After completion of this unit, student will be able to

- Understand the use of frequent patterns in business analysis. (L2)
- Implement Apriori algorithm and FP-growth algorithm. (L3)
- Learn different types of association rules.(L2)
- Identify the importance of each pattern evaluation method. (L3)
- Understand measures for mining correlated patterns. (L2)
- Learn Advanced Pattern Mining Methods. (L2)

UNIT –IV

Classification: Basic Concepts, Decision Tree induction, Bayes Classification methods, Rule based Classification, Model Evaluation and selection, Techniques to improve classification accuracy, Support Vector Machines, Classification using Frequent patterns, Lazy Learners. (10 hours)

Learning outcomes

After completion of this unit, student will be able to

- Understand basic concepts of classification. (L2)
- Implement the classification algorithms. (L3)
- Compare the performance of various classification algorithms. (L2)
- Understand model evaluation and selection methods. (L2)
- Identify the method that improves classification accuracy. (L3)

UNIT –V

Cluster Analysis: Definition, Requirements, Basic Clustering methods, Partitioning methods, Hierarchical methods, Density based methods, grid based methods, Evaluation of Clustering.

Outlier Detection: Outliers and Outlier Analysis, Detection methods, Statistical approaches, Proximity Based Approaches. (8 hours)

Learning outcomes

After completion of this unit, student will be able to

- Understand the basic concepts of clustering. (L2)
- Implement the clustering algorithms. (L3)
- Compare the performance of various clustering algorithms. (L2)
- Learn various outlier detection methods. (L2)

Course Outcomes:

Upon completion of this course, student will be able to:

- Understand the functionality of various data warehousing and data mining components.
- Understand various OLAP operations.
- Understand the strengths and limitations of various data mining models.
- Implement data mining algorithms with different datasets.

- Compare various approaches of data mining implementations.
- Identify and apply appropriate data mining technique to solve a problem.

Text Book:

1. Data Mining Concepts and Techniques by Jiawei Han, Michel Kamber, Elsevier,3rd Edition, 2012.

Reference Books

1. Introduction to Data Mining by Pang-Ning Tan & Michael Steinbach, Vipin Kumar, Pearson Publications, 1st edition, 2016 .
2. Data Mining Techniques and Applications: An Introduction by Hongbo Du, Cengage Learning EMEA, 1st edition, 2010.
3. Data Mining : Introductory and Advanced topic by Dunham, Pearson Publications, 1st edition, 2006.

Prepared by: Dr. T. Uma Devi

Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE

SEMESTER – II

SDS 708 DATA SECURITY AND PRIVACY

Hours per week: 4 End Examination: 60 Marks

Credits: 4 Sessionals: 40 Marks

Preamble:

Data security refers to protect digital privacy measures that are applied to prevent unauthorized access to computers, databases and websites. Data security also protects data from corruption. Data security is an essential aspect of IT for organizations of every size and type. Data security is also known as Information Security (IS) or Computer Security.

Course Objectives:

- To learn the basic concepts related to data security and understand the different types of symmetric key ciphers.
- To understand the concepts of encryption standards.
- To understand the concepts of asymmetric key cryptography and hash functions.
- To learn the basic concepts of hiding data in text and images.
- To understand the concepts of privacy, authentication, web and email security.

UNIT - I

Introduction: Security goals, Cryptographic Attacks, Services and Mechanism, Techniques.

Traditional Symmetric Key Ciphers: Introduction, Substitution Ciphers, Transposition Ciphers, Stream and Block Ciphers.

Introduction to Modern Symmetric-Key Ciphers: Modern Block Ciphers, Modern Stream Ciphers. (8 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Explain different security goals. (L2)
- Develop substitution and transposition ciphers. (L3)
- Describe concepts of symmetric key ciphers. (L2)
- Explain concepts of modern block ciphers. (L2)
- Extend the concept of modern stream ciphers. (L2)

UNIT - II

Data Encryption Standard (DES): Introduction, DES Structure, DES Analysis, Security of DES, Multiple DES-Conventional Encryption Algorithms.

Advanced Encryption Standard (AES): Introduction, Transformations, Key Expansion, AES Ciphers, Analysis of AES. (9 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Outline the structure of DES. (L2)
- Illustrate the analysis of DES. (L3)
- Explain the concept of AES. (L2)
- Identify the need of key expansion. (L1)
- Illustrate the analysis of AES. (L3)

UNIT - III

Asymmetric-Key Cryptography: Introduction, RSA Cryptosystem, Rabin Cryptosystem, Elgamal Cryptosystem, Elliptic Curve Crypoto systems.

Cryptographic Hash Functions: Introduction, Iterated Hash function, SHA-512, WHIRLPOOL.

Digital Signature: Comparison, Process, Services, Attacks on Digital Signature, Digital Signature Standard. (10 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Explain different types of cryptosystems. (L2)
- Identify necessity of a HASH function. (L1)
- Illustrate the use of cryptographic hash functions. (L3)
- Identify different types of attacks on digital signature. (L1)
- Extend the concept of digital signature standard. (L3)

UNIT- IV

Data Hiding in Text: Basic Features, Applications of Data Hiding, Watermarking, Intuitive Methods, Simple Digital Methods, Data Hiding in Text, Innocuous Text, Mimic Functions.

Data Hiding in Images: LSB Encoding , BPCS Steganography, Lossless Data Hiding, Spread Spectrum Steganography, Data Hiding by Quantization, Patchwork , Signature Casting in Images, Transform Domain Methods, Robust Data Hiding in JPEG Images, Robust Frequency Domain Watermarking, Detecting Malicious Tampering. (10 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Identify the need of data hiding. (L1)
- Illustrate different types of data hiding techniques. (L3)
- Describe the concepts of steganography. (L2)
- Explain the concepts of data hiding in images. (L2)

UNIT - V

Privacy: Privacy Concepts, Privacy Principles and Policies, Authentication and Privacy, Data Mining, Privacy on the Web, E-Mail Security, Impacts on Emerging Technologies.

Legal and Ethical Issues in Computer Security: Protecting Programs and Data, Information and the Law, Rights of Employees and employers, Redress for Software Failures, Computer Crime, Ethical Issues in Computer Security. (8 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Understand the basic concepts of privacy. (L2)
- Identify the need of email security. (L1)
- Illustrate software failures. (L3)
- Describe the concepts of computer crime. (L2)
- Explain the concepts of ethical issues in computer security. (L2)

Course Outcomes:

Upon completion of this course, student will be able to:

- Understand the need of computer security. (L2)
- Identify the differences between different types of ciphers. (L4)
- List the concepts of block ciphers and stream ciphers. (L4)
- Able to differentiate between DES and AES. (L4)
- List the concepts of asymmetric key cryptography. (L4)
- Understand various features of digital signature. (L2)
- Identify the differences between cryptographic hash functions. (L4)
- List the concepts of data hiding. (L4)
- Understand the features of steganography. (L2)

Text Books:

1. Cryptography and Network Security by Behrouz A. Forouzan, Dedeep Mukhopadhyay, TMH, 2nd edition, 2013. (Unit I , II, III)
2. Data Privacy and Security by Salomon, David, Springer, 2003. (Unit IV only)
3. Security in Computing by Charles Pfleeger, Shari Lawrence Pfleeger, 5th Edition, PHI, 2015. (Unit V only)

Reference Books:

1. Information Security: Principles and Practice by Mark Stamp, Wiley Inter Science, 2011.
2. Computer Security: Art and Science by Matt Bishop, First Edition, Addison Wesley, 2002.
3. Cryptography and Network Security by William Stallings, Pearson Education,7th edition, 2017.

Prepared by: Dr. T. Uma Devi

Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER – II
SDS 710 MATHEMATICS FOR DATA SCIENCE

Hours per week: 4 End Examination: 60 Marks
Credits: 4 Sessionals: 40 Marks

Preamble :

This course provides the basic knowledge on mathematics required for a data scientist. This course covers the concepts on Matrices, Normal forms, Rules of inference, Boolean Algebra and Boolean functions, Graph and Tree Concepts.

Course Objectives:

- To understand the difference between various types of matrices.
- To learn the basic concept and applications of matrices in real life problems.
- To identify and practice the mathematical logic problems with the help of truth tables or without using truth tables.
- Ability to implement features of inference rules in inference calculus.
- To understand the concept of Boolean algebra and Boolean functions.
- To understand the concepts of graphs, directed graphs, and trees.

UNIT - I

Matrices: Definition, addition and multiplication of matrices, various types of matrices, Determinant of a square matrix, Inverse of a matrix, Solution of system of non homogenous linear equations by Crammer's rule , matrix inversion method, Gauss elimination method, Gauss-Jordan method, rank of a matrix, Normal form of a matrix, Echelon form of a matrix Consistency of linear system of equations, solution of system of linear homogenous equations, Eigen values and Eigenvectors, norm, condition number. (10 hours)

Learning Outcomes:

By the end of this Unit, the student will be able to

- Explain various types of matrices.(L3)
- Evaluate system of equations by Crammer's rule, matrix inverse method, gauss elimination method. (L3)
- Explain various methods to find rank of a matrix. (L3)
- Evaluate Eigen values and Eigen vectors of a matrix. (L3)

UNIT - II

Mathematical Logic: Connectives, Negation, Conjunction, Disjunction, Conditional &Bi-Conditional, Well Formed Formulae, Tautologies, Equivalence of formulae, Duality, Tautological Implications, Functionally Complete Set of Connectives, Principal Disjunctive & Conjunctive Normal Forms, Inference Calculus, Rules of Inference, Indirect method of proof. (8 hours)

Learning Outcomes:

By the end of this Unit, the student will be able to

- Demonstrate basic concepts of mathematical logic including connectives, tautology, equivalence, and normal forms.(L2)
- Evaluate problems on principal disjunctive normal form. (L5)
- Evaluate problems on principal conjunctive normal form. (L5)
- Describe methods to solve inference calculus problems. (L3)
- Describe indirect method of proof of an argument. (L3)

UNIT - III

Boolean Algebra: Definition and Examples, sub algebra, Direct product and Homomorphism, Boolean Functions, Boolean forms and free Boolean Algebras, Values of Boolean expressions

and Boolean functions, Representation of Boolean functions, Minimization of Boolean functions, Karnaugh maps. (8 hours)

Learning Outcomes:

By the end of this Unit, the student will be able to

- Define Boolean algebra, Sub Boolean algebra with examples. (L3)
- Explain the need of Boolean functions. (L3)
- Evaluate Boolean expressions and Boolean functions. (L5)
- Explain representation of Boolean functions. (L3)
- Explain minimization of Boolean functions using Karnaugh Maps. (L3)

UNIT - IV

Graph Theory: Definitions, Finite and Infinite graphs, Incidence and Degree, Isolated pendant vertices, Isomorphism, sub graphs, Walk, Path and Circuit, Connected and Disconnected graphs, components, Euler graphs, Euler graph theorem, Operations on graphs, Decomposition of Euler graphs into circuits, Hamiltonian paths and circuits. (10 hours)

Learning Outcomes:

By the end of this Unit, the student will be able to

- Define various types of graphs. (L3)
- Define Euler graphs and prove Euler graph theorem. (L3)
- Evaluate operations on graphs. (L5)
- Evaluate Hamiltonian paths and circuits. (L5)
- Evaluate isomorphism of undirected and directed graphs. (L5)

UNIT-V

Trees: Properties of trees, pendant vertices, distance & centers, rooted & binary trees, spanning trees, fundamental circuit, shortest spanning trees, Kruskal's algorithm, Binary Tree Traversals. (8 hours)

Learning Outcomes:

By the end of this Unit, the student will be able to

- Define various types of trees and their properties. (L3)
- Explain rooted and binary trees. (L3)
- Construction of spanning trees from a connected graph. (L5)
- Explain Krushkal's algorithm to find minimum spanning tree of a connected graph. (L3)
- Explain Pre-order, Post- order, and In-order traversals of a binary tree. (L3)

Text Books :

1. Higher Engineering Mathematics by B.S.Grewal, Khanna Publishers, 43rd edition, 2015
2. Numerical methods for scientific and engineering computation by M.K.Jain, S.R.K. Iyengar, R.K. Jain, New Age International publishers,6th edition,2012.
3. Discrete Mathematical Structures with Applications to Computer Science by J.P. Tremblay and R. Manohar, Tata McGraw Hill, 1997.
4. Graph Theory with Applications to Engineering and Computer Science by Narsingh Deo, Prentice Hall of India, 2006.

Reference Books:

1. Discrete Mathematics and its Applications by Keneth. H. Rosen, Tata McGraw-Hill, 6th Edition, 2009.
2. Discrete Mathematics by Richard Johnsonbaug, Pearson Education, 7th Edition, 2008.
3. Discrete Mathematics for Computer Scientists and Mathematicians by J.L. Mott, A. Kandel, T.P. Baker, PrenticeHall.

M.Sc DATA SCIENCE

SEMESTER – II

SDS 722 R PROGRAMMING LAB

Hours per week: 4 Continuous Evaluation: 100 Marks

Credits: 2

1. R Introduction: Reserved Words, Variables & Constants, Operators, Operator Precedence.
2. Decision and Loop Structure : if...else, for loop, while Loop, break & next, repeat Loop.
3. R Functions: Functions, Function Return Value, Environment & Scope, Switch Function.
4. R Data Structures : Vectors, Matrix, List in R programming, Data Frame, Factor.
5. R Objects and Class : Object and Class , S3 Class, S4 Class, R Reference Class, Inheritance.
6. R Graphs and Charts: Bar Plot, Histogram, Pie Chart, Box Plot, Strip Chart.
7. More on Plotting In R: Plot Function: Subplot, Saving Plot, 3D Plot.
8. R Program using a csv file to plot using the distributions like Normal Distribution, Binomial Distribution, Poisson Distribution, Bernoulli Distribution and inferencing the behavior of data.
9. Program to normalize the data, replacing null columns either with mean or variance.
10. Program to import the data. And calculate covariance, correlation, mean, standard deviation and generate plots.
11. Program to implement linear regression for predictive modeling check how well the model fit the data.
12. Program to implement Multiple Regression for a given data.
13. Program to implement Cross Validation Methods
    - Validation set approach (or data split)
    - Leave One Out Cross Validation
    - k-fold Cross Validation
    - Repeated k-fold Cross Validation

Course Outcomes:

Upon completion of this course, student will be able to:

- Understand basic concepts such as data type and index and use them in their work. (L3)
- Demonstrate use of basic functions. (L2)
- Conceptualize and create loops to solve different types of problems. (l3)
- Create their own customized functions. (L4)
- Construct tables and figures for descriptive statistics. (L3)
- Implement different cross validation methods.(L5)

Text Books:

1. Beginning R–The Statistical Programming Language by Mark Gardener, Wrox publications, 2012.
2. Statistics Using R by Sudha G. Purohit, Sharad D. Gore, Shailaja R. Deshmukh, Narosa publishers, 2nd Edition, 2015.

Prepared by: Dr. T. Uma Devi Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE

SEMESTER – II
SDS 724 WEB PROGRAMMING LAB

Hours per week: 4 Continuous Evaluation: 100 Marks
Credits: 2

Objective: To give hands on training on the concepts of HTML5, CSS, XML, Java Script, PHP.

1. Write a HTML5 program to Stream video and filter with canvas.
2. Write a HTML5 program to drag an image from the desktop on to the drop zone to see the browser both render the preview.
3. Write a HTML5 Geolocation API to locate a user's position.
4. Write a HTML5 program to Create a local Storage name/value pair with name="lastname" and value="Smith". Retrieve the value of "lastname" and insert it into the element with id="result"
5. Create a general class for color which can be applied to any element in an external CSS File . Create an embedded style for all table cells to have gray and cyan color in alternate rows as background color. Create an HTML page which uses both the styles.
6. Create an HTML page where all the h1 header will have 10 px spacing, green color. All paragraphs will use gold color for all bold text and red color for all italic text.
7. Create a web page where all header are boxed with solid or ridge border. All the text is bordered with black color border.
8. Design a web page using CSS (Cascading Style Sheets) which includes the following: 1) Use different font, styles. In the style definition define how each selector should work (font, color etc.). Then, in the body of the pages, refer to these selectors to activate the styles.
9. Write a Java Script program Using the alert(), confirm() and prompt() methods.
10. Write a Java Script program for Performing Form Validation.
11. Write a Program that Controlling an Applet with JavaScript.
12. Write a Java Script program to display a digital clock in the status line.
13. Develop a PHP program for designing ” contact us” page.
14. Assume four users - user1, user2, user3 and user4 having the passwords pwd1, pwd2, pwd3 and pwd4 respectively. Write a PHP for doing the following.

a. Create a Cookie and add these four user id’s and passwords to this Cookie.

b. Read the user id and passwords entered in the Login form (week1) and authenticate with the values (user id and passwords) available in the cookies. If he is a valid user (i.e., user-name and password match). The program should welcome him by name (user-name) else the program should display “You are not an authenticated user ’’. Use init-parameters to do this.

15. Write a PHP program for registering users of a website and login.
16. Write a PHP script to decode a JSON String, to get JSON representation of a value from an array.
17. Write a program to create a company JSON with different key valules.

Course Outcomes:

Upon completion of this course student will be able to

- Apply HTML tags for creating static web pages. (L3)
- Use scripting languages and web services to transfer data and add interactive components to web pages. (L4)
- To create dynamic styles, animation on web pages.(L2)
- Use regular expression for form validation.(L3)

- Write PHP scripts to handle HTML forms.(L4)
- Build dynamic web pages using server side PHP programming.(L4)
- Create PHP programs that use various PHP library functions. (L3)
- Analyze and solve common Web application tasks by writing PHP programs.(L5)

Text Books:

1. HTML, Java Script, DHTML, PHP by Ivan Bayross, bpb,4th revised edition, 2010.
2. HTML 5 Black Book by DT Editorial Service, Dreamtech Press, 2nd Edition,2016.
3. Mastering HTML,CSS & Javascript by Laura Lemay, Rafe Colburn, Jennifer Kyrnin, Web Publishing Paperback, 2016.

Prepared by: Dr. T. Uma Devi

Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE

SEMESTER - III

SDS 801 MACHINE LEARNING

Hours per week: 4 End Examination: 60 Marks

Credits: 4 Sessionals: 40 Marks

Preamble:

Machine Learning addresses the question how to enable computers to learn from past experiences. It introduces the field of machine learning describing a variety of learning paradigms, algorithms, theoretical results and applications. Upon successful completion of the course, students will have an understanding the working of various machine learning algorithms which can be implemented through.

Course Objectives:

- To design a learning system and know about the learning tasks.
- To apply decision tree learning in classification tasks.
- To develop neural networks algorithms in machine learning.
- To illustrate Bayesian learning and instance based learning.
- To examine the concepts of genetic algorithms and reinforcement learning.

UNIT – I

Introduction: Well-Posed Learning Problems, Designing a Learning System, Perspectives and Issues in Machine Learning.

Concept Learning and the General-to-Specific Ordering: Introduction, A Concept Learning Task, Concept Learning as Search, FIND-S: Finding a Maximally Specific Hypothesis, Version Spaces and the Candidate-Elimination Algorithm, Remarks on Version Spaces and Candidate-Elimination, Inductive Bias. (10 hours)

Learning outcomes

After completion of this unit, student will be able to

- Define a well-posed learning problem. (L1)
- Illustrate the designing of a learning system. (L1)
- Understand a concept learning task. (L2)
- Name what are version spaces. (L1)

UNIT – II

Decision Tree Learning: Introduction, Decision Tree Representation, Appropriate Problems for Decision Tree Learning, The Basic Decision Tree Learning Algorithm, Hypothesis Space Search in Decision Tree Learning, Inductive Bias in Decision Tree Learning, Issues in Decision Tree Learning. (10 hours)

Learning outcomes

After completion of this unit, student will be able to

- Define a decision tree. (L1)
- Illustrate the decision tree learning algorithm and hypothesis space search. (L2)
- List various issues in decision tree learning. (L1)

UNIT - III

Artificial Neural Networks: Introduction, Neural Network Representations, Appropriate Problems for Neural Network Learning, Perceptrons, Multilayer Networks and the BACKPROPAGATION algorithm, Remarks on the Backpropagation algorithm, Illustrative Example, Advanced Topics in Artificial Neural Network.

Evaluating Hypothesis: Estimating Hypothesis Accuracy, Basics of Sampling Theory, A General Approach for deriving Confidence intervals, Difference in Error of two Hypothesis, Comparing Learning Algorithms. (10 hours)

Learning outcomes

After completion of this unit, student will be able to

- Define what is a neural network and associated fundamentals. (L1)
- Demonstrate the working of multilayer neural networks. (L2)
- Learn the methods of evaluating the accuracy of hypotheses. (L2)

UNIT - IV

Bayesian Learning: Introduction, Bayes Theorem, Bayes Theorem and Concept Learning, Maximum Likelihood and Least Squared Error Hypothesis, Maximum Likelihood Hypothesis for predicting probabilities, Naive Bayes Classifier, Bayesian Belief Networks, EM Algorithm. (8 hours)

Learning outcomes

After completion of this unit, student will be able to

- Define Bayes theorem. (L1)
- Summarize the importance of Bayesian methods in machine learning. (L2)
- Show how Bayes theorem and concept learning are related. (L2)
- Learn how EM algorithm provides a quite general approach in the presence of unobservable variables. (L1)

UNIT-V

Instance-Based Learning: Introduction, k-Nearest Neighbour Learning, Locally Weighted Regression. **Genetic Algorithms:** Motivation, Genetic Algorithms, An Illustrative Example, Hypothesis Space Search, Genetic Programming.

Reinforcement Learning: Introduction, The Learning Task, Q Learning. (8 hours)

Learning outcomes

After completion of this unit, student will be able to

- Contrast instance-based learning with other methods of learning. (L4)
- Model genetic learning method by an analogy to biological evolution. (L3)
- Experiment with hypothesis space search in genetic learning. (L3)

Course Outcomes:

Upon completion of this course student will be able to

- Define a well-posed learning problem. (L3)
- Illustrate the decision tree learning algorithm and hypothesis space search. (L4)
- Use the Bayes theorem and EM Algorithm in machine learning . (L3)

Text Book:

1. Machine Learning by Tom M. Mitchell, McGraw Hill Education Private Limited, 2013.

Reference Books:

1. Pattern Recognition and Machine Learning by Christopher Bishop, Springer series,1st edition, 2006.
2. Machine Learning a Probabilistic Perspective by Kevin P Murphy & Francis Bach, MIT Press, 1st Edition, 2012.

Prepared by: Dr. T. Uma Devi Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER - III
SDS 803 BIG DATA ANALYTICS

Hours per week: 4 End Examination: 60 Marks
Credits: 4 Sessionals: 40 Marks

Preamble:

The internet, Big Data, vastly improved computational power, and a wide variety of variables are involved in complex, real-world problems led to a new set of analytic techniques and technologies. The concept of Big Data includes massive volumes of data and huge benefits that can accrue from the analysis of it.

Course Objectives:

- To introduce an in depth understanding of all the concepts related to Big Data.
- To provide a sight into the real life implementations of Big Data solutions and problem solving in data analytics.
- To provide learners with a deep and systematic knowledge of business and technical strategies for data analytics and the subsequent skills to implement solutions in these areas.

UNIT – I

Getting an overview of Big Data: Big Data definition, History of Data Management, Structuring Big Data, Elements of Big-data, Big Data Analytics.

Exploring use of Big Data in Business Context: Use of Big Data in Social Networking, Use of Big Data in preventing Fraudulent Activities in Insurance Sector & in Retail Industry. **(**8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn various sources of data and forms of data generation. (L2)
- Understand the evolution and elements of Big Data. (L2)
- Explore different opportunities available in the career path. (L3)
- Understand the role and importance of Big Data in various domains. (L2)

UNIT – II

Introducing Technologies for Handling Big Data: Distributed and parallel computing for Big Data, Introducing Hadoop, Cloud computing and Big Data, In-memory Computing Technology for Big Data.

Understanding Hadoop Ecosystem: Hadoop Ecosystem**,** Hadoop Distributed File System, MapReduce, Hadoop YARN, Introducing HBase, Combing HBase and HDFS, Hive, Pig and Pig Latin, Sqoop, ZooKeeper, Flume, Oozie.

Understanding MapReduce Fundamentals and HBase: The MapReduce Framework, Techniques to Optimize Map Reduce Jobs, Uses of Map Reduce, Role of HBase in Big Data Processing. (10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Identify the difference between distributed and parallel computing. (L3)
- Learn the importance of Virtualization in Big Data. (L2)
- Learn the details of Hadoop and Cloud Computing. (L2)
- Learn the architecture and features of HDFS. (L2)
- Understand Hadoop Ecosystem, MapReduce and HBase. (L2)
- Apply the technique in optimizing MapReduce job. (L3)

UNIT- III

Understanding Big Data Technology Foundations: Exploring the Big Data Stack, Virtualization and Big Data, Virtualization approaches.

Processing Data with MapReduce: Recollecting the Concept of MapReduce Framework, Developing Simple MapReduce Applications.

Customizing MapReduce Execution and Implementing MapReduce Program: Controlling MapReduce Execution with Input Format, Reading Data with Custom Record Reader, Organizing Output data with Output Formats, Customizing Data with Record Writer, Optimizing MapReduce Execution with Combiner, Controlling Reducer Execution with Partitioners, Customizing the MapReduce Program for Sorting Text Data, implementing a Map Reduce Program for Sorting Text Data. (10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Explore the layers of Big Data Stack. (L2)
- Learn virtualization approaches in handling Big Data operations. (L2)
- Able to develop simple applications using map and reduce function. (L5)
- Learn the classes available in MapReduce framework. (L2)
- Understand the role of Combiner and Partitioners in a MapReduce applications. (L3)

UNIT – IV

Understanding Hadoop YARN Architecture: Background of YARN, Advantages of YARN, YARN Architecture, Working of YARN, YARN Schedulers, Backward Compatibility with YARN, YARN Configurations, YARN commands, YARN containers, Registry. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn the importance of YARN.(L2)
- Understand the use and importance of schedulers and backward compatibility in YARN.(L3)
- Learn the commands, log management and configuration for handling Big Data. (L3)

UNIT – V

Exploring Hive: Introducing Hive, Getting Started with Hive, Hive Services, Data Types , Built-in Functions, Hive-DDL, Data Manipulation, Data Retrieval Queries, Using Joins.

Analyzing Data with Pig: Introducing Pig, Running Pig, Getting started with Pig Latin, working with operators in Pig, Debugging Pig, Working with Functions in pig, Error Handling in Pig.

Understanding Analytics and Big Data: Comparing Reporting and analysis, Types of Analytics, Developing an Analytic Team, Understanding Text Analytics. (10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn the working of Hive and query execution. (L2)
- Learn the importance of Pig. (L2)
- Choose the operators in Pig. (L2)
- Understand various types of analytical approaches. (L3)

Course Outcomes:

Upon completion of this course student will be able to

- Able to handle a real time big data application. (L4)
- Able to develop Map Reduce Applications. (L4)
- Perform YARN Configuration for handling Big Data. (L3)
- Learn to execute queries in Hive. (L3)
- Learn how Pig Latin is used for programming in Hadoop. (L3)
- Learn to design new analytical approaches for Big Data. (L3)

Text book:

1.Big Data Black Book by Dt Editorial Services, Dreamtech Publications, 2016.

Reference Book :

1.Hadoop The Definitive Guide by Tom White, O'reilly ,4th Edition,2016.

Prepared by: Dr. T. Uma Devi

Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER - III
SOE XXX OPEN ELECTIVE

Hours per week: 3 End Examination: 60 Marks
Credits: 3 Sessionals: 40 Marks

M.Sc DATA SCIENCE
SEMESTER - III
Generic Elective - I
SDS 841 WEB ANALYTICS

Hours per week: 4 End Examination: 60 Marks
Credits: 4 Sessionals: 40 Marks

Preamble:

Web analytics is a way of learning how users interact with websites by automatically recording aspects of user's behavior and transforming the behavior into data that can be analyzed. It is the measurement and analysis of data to inform an understanding of user behavior across web pages.

Course Objectives:

- To learn web analytics from a strategic and practical perspective.
- Explore different types of analytics and why they are important for business.
- Learn various web analytics processes and metrics used to measure.
- Learn techniques using Google Web analytics, traffic analysis, click path analysis and segmentation.

UNIT – I

Web Analytics Present and Future: A brief history of web analytics, Current Landscape and Challenges, Traditional Web analytics, Web analytics Activities, Measuring, Trinity. **Data collection:** Understanding the Data Landscape, Click stream Data, Outcomes Data, Research Data, Competitive Data. (6 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Lay the foundational ground work on the approaches to web analytics. (L2)
- Understand the critical importance of various data collection mechanisms. (L2)
- Focus on qualitative data- why and what the available options. (L2)
- Significantly elevate the ability to listen to the customers. (L3)

UNIT – II

Introduction to Web Analytics: Definition, User Experience and Web Analytics Questions.
Web Analytics Approach: Introduction, A model of Analysis, Show casing the work, Context Matters, Contradicting the data.
Working of Web Analytics: Introduction, Log File Analysis, Page Tagging, Metrics and Dimensions, Interacting with data in Google Analytics. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Understand foundational concepts in web analytics. (L1)
- Learn the analysis process. (L2)
- Learn the importance of viewing data and balancing the desire for perfection. (L2)
- Understand how analytic tools work. (L2)
- Understand how analytic tools organize and segment data. (L2)

UNIT – III

Goals: Introduction, Definition of Goals and Conversions, Conversion Rate, Goal Reports in Google Analytics, Finding the right things to measure as key, Performance Indicators, Measure on a website that can constitute a goal. **Learning about users:** Introduction, Visitor Analysis. (6 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Identify analytics goals and conversion rates. (L3)
- Learn the key performance indicators to calibrate the web site goals. (L4)
- Identify user's geographical location, technology used and visiting frequency. (L3)

UNIT – IV

Traffic Analysis: Introduction, Source and medium, Organic Search, Search Query Analysis, Referral Traffic, Direct Traffic, Paid Search Keyword.

Analyzing usage of content: introduction, Website content Reports. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Study the way users actually get to the websites. (L2)
- Analyze the key words user types in search engines. (L3)
- Articulate the information needs and categorize users. (L4)
- Learn analysis metrics used to analyze usage content. (L2)

UNIT – V

Click-Path Analysis: Introduction, Focus on Relationships between pages, Navigation Summary, Visitors Flow Report, Analyzing how users move from one page type to another.

Segmentation: Introduction, Necessity, Procedure to segment, Ways to Segment, Useful ways to segment UX questions. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn Click-Path analysis. (L1)
- Examine the relationships between pairs of pages. (L3)
- Learn methods of filtering data. (L2)
- Analyze the behavior of users. (L3)

Course Outcomes:

Upon completion of this course student will be able to

- Learn how to use web analytics to answer the complicated questions facing UX professionals. (L3)
- Able to measure a website's effectiveness. (L3)
- Learn methods about how users navigate a website. (L1)

Text Books:

1. Web Analytics an hour a Day by Avinash Kaushik, Sybex, 1st Edition, 2007. (Unit – I)

2. Practical Web Analytics for User Experience by Michael Beasley, Morgan Kaufmann, 1st Edition, 2013. (Unit II to Unit V)

Reference Books:

1. Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity by Avinash Kaushik, 1st edition, Sybex, 2009.

2. Web Data Mining: Exploring Hyperlinks, Content and Usage Data by Bing Liu, 2nd Edition, Springer, 2011.

3. Google Analytics by Justin Cutroni, O'Reilly, 2010.

Prepared by: Dr. T. Uma Devi Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER - III
Generic Elective - I

SDS 843 FUNDAMENTALS OF BLOCK CHAIN TECHNOLOGIES

Hours per week: 4 End Examination: 60 Marks
Credits: 4 Sessionals: 40 Marks

Preamble:

This is new technologies of digital currency. Blockchains is to achieve decentralization. T*he system needs to validate transactions without anyone being able to veto transactions or control the network.*

Course Objectives:

- Learn the basic concept of Cryptographic Hash Functions, Hash Pointers and Elliptic Curve Digital Signature Algorithm.
- A technical overview of decentralized digital currencies like Bitcoin, as well as their broader economic, legal and financial context.
- To get an insight into the working of the Bitcoin network, Wallet, Bitcoin mining and distributed consensus for reliability.

UNIT – I

Introduction to Cryptography: Cryptographic Hash Functions, SHA-256, Hash Pointers and Data Structures, Merkle tree. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn the basics of hash functions. (L1)
- Identify the importance of each hash function and understand the underlying data structures used. (L3)

UNIT – II

Digital Signatures: Elliptic Curve Digital Signature Algorithm (ECDSA), Public Keys as identities, A Simple Crypto currency. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn the importance of digital signature. (L2)
- Understand digital signature algorithms. (L2)
- Learn the mechanism of simple crypto currency. (L2)

UNIT – III

Centralization vs Decentralization, Distributed consensus, Consensus without identity using a block chain, Incentives and proof of work.

Mechanics of Bitcoin: Bitcoin Transactions, Bitcoin Scripts, Applications of Bitcoin Scripts, Bitcoin Blocks, The Bitcoin Network. (10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Understand the structure of a blockchain. (L1)
- Learn why it is better than a simple distributed database. (L2)
- Learn the underlying principles and techniques associated with blockchain technologies. (L3)
- Familiar with the cryptographic building blocks. (L3)
- Understand typical Cryptocurrency such as Bitcoin. (L2)

UNIT – IV

Storage and Usage of Bitcoins: Simple Local Storage, Hot and Cold Storage, Splitting and Sharing Keys, Online Wallets and Exchanges, Payment Services, Transaction Fees, Currency Exchange Markets. (10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn different ways of storing Bitcoin keys, security measures. (L2)
- Understand various types of services that allow you to trade and transact with bitcoins. (L3)

UNIT – V

Bitcoin Mining: The Task of Bitcoin miners, Mining Hardware, Mining pools, Mining incentives and strategies.

Bitcoin and Anonymity: Anonymity Basics, Mixing, Zerocoin and Zerocash. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn how Bitcoin relies on mining. (L1)
- Who are the miners?.(L1)
- How they get into this and operate. (L1)
- Learn Business model for miners. (L2)
- Impact of business model on the environment. (L3)
- The various ways to improve Bitcoin's anonymity and privacy. (L3)

Course Outcomes:

Upon completion of this course student will be able to

- Learn individual components of the Bitcoin protocol make the whole system tick. (L2)
- Learn the methods of security from a combination of technical methods and clever incentive engineering. (L2)
- Analyze the incentive structure in a blockchain based system and critically assess its functions, benefits and vulnerabilities. (L4)

Text Book:

1. Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction by Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller and Steven Goldfeder, Princeton Press, 2016.

Reference Book:

1. Mastering Bitcoin: Programming the Open Blockchain by Andreas M. Antonopoulos Shroff, O'Reilly; 2nd Edition, 2017.

Prepared by: Dr. T. Uma Devi Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER - III
Generic Elective - I

SDS 845 HIGH PERFORMANCE COMPUTING

Hours per week: 4 End Examination: 60 Marks
Credits: 4 Sessionals: 40 Marks

Preamble:

High Performance Computing refers to the practice of aggregating computing power in a way that delivers much higher performance than one could get out of a typical desktop computer or workstation in order to solve large problems in science, engineering, or business.

Course Objectives:

- To learn the basic concepts of Modern Processors.
- To understand the algorithms used to optimize data access.
- To learn the basic concepts of parallelization.
- To learn the concepts of OpenMP programming.
- To understand the use of different MPI performance tools.

UNIT – I

Modern processors: Stored-program computer architecture, General purpose cache-based microprocessor architecture, Memory hierarchies, Multicore processors, Multithreaded processors, Vector processors.

Basic optimization techniques for serial code: Scalar Profiling, Common sense optimizations, Simple measures, large impact, the role of compilers. (8 hours)

Learning Outcomes:

By the end of this Unit, the student will be able to

- Illustrate the concept of cache.(L3)
- Describe the architecture of modern processors.(L2)
- List different kinds of processors.(L1)
- List the optimization techniques for serial code.(L1)
- Explain the role of compilers.(L2)

UNIT – II

Data access optimization: Balance analysis and lightspeed estimates, Storage order, The Jacobi algorithm, Dense Matrix Transpose, algorithm classification and access optimizations, Sparse Matrix-vector multiply.

Parallel Computers: Taxonomy of parallel computing paradigms, Shared-memory computers, Distributed-memory computers, Hierarchical (hybrid) systems, Networks. (10 hours)

Learning Outcomes:

By the end of this Unit, the student will be able to

- Compare balance analysis and light speed estimates.(L2)
- Explain the Jacobi algorithm.(L2)
- Illustrate parallel computing paradigms.(L3)
- Compare shared memory computers and distributed memory computers.(L2)
- List the concepts of Hierarchical systems.(L1)

UNIT – III

Basics of parallelization: Parallelism, Parallel scalability.

Shared-memory parallel programming with OpenMP: Short introduction to OpenMP, OpenMP-parallel Jacobi algorithm, Advanced OpenMP: Wavefront parallelization. (8 hours)

Learning Outcomes:

By the end of this Unit, the student will be able to

- Explain the concepts of parallelization. (L2)
- Illustrate the concepts of OpenMP. (L3)
- Apply Jacobi algorithm using OpenMP.(L3)
- Describe the concepts of wavefront parallelization.(L2)

UNIT – IV

Efficient OpenMP programming: Profiling OpenMP programs, Performance pitfalls, Parallel sparse matrix-vector multiply.

Locality opthnizations on ccNUMA architectures: Locality of access on ccNUMA, ccNUMA optimization of sparse MVM, Placement pitfalls. (8 hours)

Learning Outcomes:

By the end of this Unit, the student will be able to

- Extend the concepts of OpenMp programming.(L2)
- Summarize the pitfalls.(L2)
- Explain locality optimizations.(L2)
- Outline optimization of sparse MVM.(L2)

UNIT – V

Distributed-memory parallel programming with MPI: Message Passing, Introduction to MPI, MPI parallelization of a Jacobi Solver.

Efficient MPI programming: MPI performance tools, Communication Parameters, Synchronization, serialization, contention, Reducing Communication overhead, Understanding intra node point-to-point communication. (9 hours)

Learning Outcomes:

By the end of this Unit, the student will be able to

- Explain distributed-memory parallel programming.(L2)
- Illustrate the concept of MPI parallelization.(L3)
- Extend the concepts of MPI programming.(L2)
- List the performance tools.(L1)
- Explain intra node point-to-point communication.(L2)

Course Outcomes:

Upon completion of this course student will be able to

- Understand the architecture of modern processors.(L2)
- Identify the difference between different kinds of processors.(L4)
- List the concepts of data access optimization.(L4)
- Differentiate between shared memory computers and distributed memory computers.(L4)
- List the concepts of parallelization.(L4)
- Construct Jacobi algorithm in OpenMP.(L6)
- Understand the various features of OpenMP programming.(L2)
- Understand concepts of ccNUMA architecture.(L2)
- List the different MPI performance tools.(L4)
- Understand intranode point-to-point communication.(L2)

Text Book:

1. Introduction to High Performance Computing for Scientists and Engineers by Gerog Hager, Gerhard Wellein, CRC Press.

Reference Book:

1. High Performance Computing, Programming and Applications by John Levesque, Gene Wagenbreth, CRC Press.

Prepared by: Mr. M. Suresh Kumar Verified by: Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER - III
Generic Elective - I
SDS 847 DATA STORAGE TECHNOLOGIES AND NETWORKING

Hours per week: 4 End Examination: 60 Marks
Credits: 4 Sessionals: 40 Marks

Preamble:

This course fill the knowledge gap in understanding varied components of modern information storage infrastructure, including virtual environments. It provides comprehensive learning of storage technology, which will enable one to make more informed decisions in an increasingly complex IT environment. This course builds a strong understanding of underlying storage technologies and prepares one to learn advanced concepts, technologies, and products.

Course Objectives

- To learn various storage infrastructure components
- To provide a strong understanding on storage related technologies.
- To learn the architectures, features and benefits of intelligent storage systems.
- To understand various storage networking technologies.

UNIT - I

Introduction to Information Storage and Management: Information Storage, Evolution of Storage Architecture, Data Center Infrastructure, Virtualization and Cloud Computing.
Data Center Environment: Application, DBMS, Host, Connectivity, Storage, Disk Drive Components, Disk Drive Performance, Host Access to Data, Direct Attached Storage, Storage Design based on Application, Introduction to Flash Drives. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn the key elements of a data center environment. (L2)
- Understands the importance of Virtualization. (L2)
- Learn different types of data storage Systems. (L3)

UNIT - II

Data Protection RAID: RAID Implementation methods, RAID Array Components, RAID Techniques, RAID Levels, RAID Comparison, RAID Impact on Disk Performance, Hot Spares.
Intelligent Storage Systems: Components of an Intelligent Storage Provisioning, Types of Intelligent Storage Systems. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Identify the importance of RAID technology. (L3)
- Understand fundamental constructs and various RAID Levels.(L2)
- Learn the Components and types of Intelligent Storage Systems. (L2)
- Understand the benefits of Intelligent Storage Systems. (L3)

UNIT - III

Storage Networking Technologies: Fiber Channel Storage Area Network: Overview, the SAN and its Evolution, Components of FC SAN, FC Connectivity, Fiber Channel Architecture, FC SAN Topologies, Virtualization in SAN. **Network Attached Storage:** General Purpose Servers vs NAS Devices, Benefits of NAS, File Systems and Network File Sharing, Components of NAS, NAS I/O operation, NAS Implementations, NAS File-Sharing Protocols. (10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn how FC SAN reduces overall operational cost and downtime. (L1)

- Understand how Virtualization minimizes resource management complexity and cost.(L1)
- Identify appropriate storage infrastructure. (L3)

UNIT - IV

Object Based and Unified Storage: Object Based Storage Devices, Content Addressed Storage, CAS Use Cases, Unified Storage. **Introduction to Business Continuity:** Information Availability, BC Terminology, BC Planning Lifecycle, Failure Analysis, Business Impact Analysis, BC Technology Solutions. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn how Object based storage manage storing unstructured data. (L1)
- Learn the components of unified storage and the processes of accessing data. (L2)
- Understand the goals of business continuity plan. (L3)
- Learn BC framework. (L2)

UNIT- V

Backup And Recovery: Backup - purpose, Considerations, Granularity, methods, Architecture, Backup and Restore Operations, Backup Topologies, Backup in NAS Environments, Backup Technologies. **Local Replication:** Replication Terminology, Uses of Local Replicas, Replica Consistency, Local Replication Technologies. (10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn the need of backup, backup methods, technologies and implementations. (L2)
- Understand different backup topologies and backup in virtualized environment. (L3)
- Understand local replication process and uses of local replica. (L2)
- Learn various local replication technologies. (L2)

Course Outcomes:

Upon completion of this course student will be able to

- Understand how to manage the capacity, performance, and reliability of large numbers of disks. (L3)
- Learn how Intelligent Storage Systems provide highly optimized I/O processing capabilities. (L2)
- Understands importance of NAS and identify how NAS improves the performance. (L3)
- Apply to organizations for an effective and cost-efficient disaster recovery and restart procedures in both physical and virtual environments. (L4)

Text Book:

1. Information Storage and Management by EMC Education Services, 2nd Edition 2012.

Reference Books:

1. Storage Area Network Essentials by Richard Barker, Paul Massiglia, Wiley 1st Edition, 2008.
2. Storage Networks – Complete Reference by Robert Spalding,TMH, 2003.
3. Building Storage Networks by Marc Farley, TMH, 2001.

Prepared by: Dr. T. Uma Devi Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER - III
Generic Elective – II
SDS 849 CLOUD COMPUTING

Hours per week: 4 End Examination: 60 Marks
Credits: 4 Sessionals: 40 Marks

Preamble:

Cloud computing is the on-demand availability of computer system resources, especially data storage and computing power, without direct active management by the user. The term is generally used to describe data centers available to many users over the Internet. Large clouds, predominant today, often have functions distributed over multiple locations from central servers. If the connection to the user is relatively close, it may be designated an edge server.

Course Objectives:

- To learn the basic concepts and services provided by cloud computing.
- To understand the concepts of virtualization in cloud computing.
- To understand the architecture of cloud computing.
- To learn the basic concepts of AWS.
- To understand the working of AWS.

UNIT – I

Introduction: Cloud Computing at a Glance, The Vision of Cloud Computing, Definition of a Cloud, A Closer Look, Cloud Computing Reference Model, Characteristics and Benefits, Challenges Ahead, Historical Developments, Distributed Systems, Virtualization, Web 2.0, Service-Oriented Computing, Utility-Oriented Computing, Building Cloud Computing Environments, Application Development, Infrastructure and System Development, Computing Platforms and Technologies, Amazon Web Services (AWS), Google AppEngine, Microsoft Azure, Hadoop, Force.com and Salesforce.com, Manjrasoft Aneka. (8 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Explain basic features of cloud computing.(L2)
- Describe the cloud computing reference model. (L2)
- Illustrate the characteristics and benefits of cloud computing. (L3)
- Outline the computing platforms and technologies.(L2)
- Describe the basic features of AWS and other cloud providers.(L2)

UNIT- II

Virtualization: Introduction, Characteristics of Virtualized Environments, Taxonomy of Virtualization Techniques, Execution Virtualization, Other Types of Virtualization, Virtualization and Cloud Computing, Pros and Cons of Virtualization, Technology Examples, Xen: Para-virtualization, VMware: Full Virtualization, Microsoft Hyper-V. (5 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Explain the basic features of virtualization. (L2)
- Identify the different types of virtualization. (L1)
- Describe the concept of VMware. (L2)
- Explain the relation between virtualization and cloud computing. (L2)

UNIT - III

Cloud Computing Architecture: Introduction Cloud Reference Model, Architecture Infrastructure / Hardware as a Service, Platform as a Service, Software as a Service, Types of Clouds, Public Clouds, Private Clouds, Hybrid Clouds, Community Clouds, Economics of the

Cloud, Open Challenges, Cloud Dentition, Cloud Interoperability and Standards, Scalability and Fault Tolerance, Security, Trust, and Privacy, Organizational Aspects. (8 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Explain the architecture of cloud computing. (L2)
- Describe the different services provided by cloud computing.(L2)
- Identify the different types of clouds. (L1)
- Outline the challenges in cloud computing.(L2)
- Explain the basic concepts of security in cloud computing.(L2)

UNIT - IV

Discovering the AWS Development Environment : Starting AWS Adventure, Defining the AWS Cloud, Discovering IaaS, Determining, Use of AWS, Considering the AWS-Supported Platforms.

Obtaining Development Access to Amazon Web Services: Discovering the Limits of Free Services, Considering the Hardware Requirements, Getting Signed Up, Testing the Setup, Choosing the Right Services, Getting a Quick Overview of Free-Tier Services, Matching AWS Services to the Application, Considering AWS Security Issues. (10 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Explain basic concepts of AWS.(L2)
- Identify the hardware requirements for AWS. (L1)
- Illustrate basic steps corresponding to AWS. (L3)
- Identify the AWS services required for an application. (L1)
- Describe the AWS security issues. (L2)

UNIT- V

Starting the Development Process: Considering AWS Communication Strategies, Defining the Major Communication Standards, Understanding how REST Works, Creating a Development Environment, Choosing a Platform, Obtaining and Installing Python, Working with the Identity and Access Management Console, Installing the Command Line Interface Software, Configuring S3 Using CLI, Configuring S3 Using Node.js, Configuring S3 Using a Desktop Application, Creating a Virtual Server Using EC2, Getting to Know the Elastic Compute Cloud (EC2), Working with Elastic Block Store (EBS) Volumes, Discovering Images and Instances.

Performing Basic Development Tasks : Understanding AWS Input/Output, Considering the Input/Output Options, Working with JSON, Working with XML, Working with Amazon API Gateway, Developing Web Apps Using Elastic Beanstalk, Considering Elastic Beanstalk (EB) Features, Deploying an EB Application, Updating an EB Application, Removing Unneeded Applications, Monitoring Your Application Using Amazon CloudWatch. (12 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Explain the major communication standards. (L2)
- Summarize creating a development environment. (L2)
- Illustrate installation of command line interface. (L1)
- Describe the basic concepts of development tasks. (L2)
- Explain the monitoring of an application using Amazon Cloudwatch. (L2)

Course Outcomes:

- Understand the need of cloud computing. (L2)
- Identify the differences between different types of ciphers. (L4)
- List different cloud providers. (L4)

- Understand the various features of virtualization. (L2)
- Understand the architecture of cloud computing. (L2)
- List the basic features of AWS. (L4)
- Understand the working of AWS. (L2)
- Understand basic development tasks. (L2)

Text Books:

1. Mastering Cloud Computing by Rajkumar Buyya, Christian Vecchiola, S Thamarai Selvi, Morgan Kaufmann , 2013.
2. AWS for Developers- Dummies by John Paul Mueller, John Wiley & Sons Inc. publications, 2017.

Reference Books:

1. Cloud Computing Concepts Technology Architecture by Thomas Erl, Pearson Education, 2014.
2. Cloud Computing Explained by John Rhoton , Recursive Press 2nd edition, 2009 .

Prepared by: Mr. M. Suresh Kumar Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER - III
Generic Elective – II
SDS 851 COMPUTATIONAL BIOLOGY

Hours per week: 4 End Examination: 60 Marks
Credits: 4 Sessionals: 40 Marks

Preamble:

A large number of prokaryotic and eukaryotic genomes completely sequenced. Mining the genomic information requires the use of sophisticated computational tools. It covers major databases and software programs for genomic data analysis, with an emphasis on the theoretical basis and practical applications of these computational tools.

Course Objectives:

- Gene and protein sequence acquisition, storage, retrieval and analysis
- Protein structure and function relationship using computational tools
- Development of computational applications for processing of biological data
- Modeling and simulation of biological systems.

UNIT – I

Bioinformatics: Introduction, Goal, Scope, Applications, Limitations, New Themes.
Introduction to Biological Databases: Database and Types of Databases, Biological Databases, Pitfalls of Biological Databases, Information Retrieval from Biological Databases - GENBANK, National Centre for Biotechnology Information, European Bioinformatics Institute. (7 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- know what bioinformatics is and why it is important. (L1)
- how bioinformatics data is stored and organized .(L1)
- learn different types of data found at the NCBI and EBI resources.(L2)
- understand how to locate and extract data from key bioinformatics databases and resources.(L2)
- Know the difference between databases, tools, repositories and be able to use each one to extract specific information.(L3)

UNIT – II

Pairwise Sequence Alignment: Evolutionary Basis, Sequence Homology versus Sequence Similarity, Sequence Similarity versus Sequence Identity, Methods, Scoring Matrices, Statistical Significance of Sequence Alignment. **Database Similarity Searching:** Unique Requirements of Database Searching, Heuristic Database Searching, Basic Local Alignment Search Tool (BLAST), FASTA, Comparison of FASTA and BLAST, Database Searching with the Smith–Waterman Method. **Multiple Sequence Alignment:** Scoring Function, Exhaustive Algorithms, Heuristic Algorithms. (10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Extract and generate pairwise sequence alignments for a protein sequence of interest. L3).
- interpret the metrics used to assess the quality of a pairwise sequence alignment, identity versus similarity.(L2)
- identify mutations between two sequences using pairwise sequence approach.(L3)
- Outline the principles of the BLAST algorithm.(L2)

- Assess the relationships between the protein sequences from the different organisms based on the multiple sequence alignment.(L5)

UNIT – III

Profiles and Hidden Markov Models: Position-Specific Scoring Matrices, Profiles, Markov Model and Hidden Marko Model. **Protein Motifs and Domain Prediction**: Identification of Motifs and Domains in Multiple Sequence Alignment, Motif and Domain Databases Using Regular Expressions, Motif and Do main Databases Using Statistical Models, Protein Family Databases, Motif Discovery in Unaligned Sequences, Sequence Logos. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Identify possible conserved protein domains and amino acids. (L2)
- Learn various statistical representations of motifs .(L2)

UNIT – IV

Gene Prediction: Categories of Gene Prediction Programs, Gene Prediction in Prokaryotes and Eukaryotes, Promoter and Regulatory Elements in Prokaryotes and Eukaryotes, Prediction Algorithms. **Phylogenetics Basics:** Molecular Evolution and Molecular Phylogenetics, Terminology, Gene Phylogeny versus Species Phylogeny, Forms of Tree Representation, Why Finding a True Tree is Difficult. **Phylogenetic Tree Construction Methods and Programs:** Distance-Based Methods, Character-Based Methods, Phylogenetic Tree Evaluation, Phylogenetic Programs. (10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Understand concepts of phylogenetic distance & how models of evolution are used.(L2)
- To build phylogenies from pairwise distance matrices.(L3)
- Learn the properties of phylogenetic trees, methods to optimize the topology and branch lengths of a tree .(L3)

UNIT – V

Protein Structure Basics: Amino Acids, Peptide Formation, Dihedral Angles, Hierarchy, Secondary Structures, Tertiary Structures, Determination of Protein Three-Dimensional Structure, Protein Structure Database, Protein Structure Visualization, Comparison and Classification. **Protein Secondary Structure Prediction:** Secondary Structure Prediction for Globular Proteins, Transmembrane Proteins, Coiled Coil Prediction. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Outline the different levels and organization of protein structures. (L2)
- Describe how protein structures are determined. (L2)
- Predict the structure for a protein sequence based on an identified template. (L6)
- Elaborate the criteria for assessing and refining a predicted protein structure. (L6)

Course Outcomes:

Upon completion of this course, student will be able to:

- Acquire knowledge about biological data bases. (L1)
- Learn Gene prediction methods. (L4)
- Understands Sequence Homology versus Sequence Similarity. (L3)
- Learn about Multiple Sequence Alignment.(L4)

Text Books:

1. Essential Bioinformatics by Jin Xiong, Cambridge University Press, 2006.
2. Bioinformatics and Functional Genomics by Jonathan Pevsner, Wiley Publications, 2nd edition, 2009.

Reference Books:

1. Introduction to Bioinformatics by Lesk, A.M., Oxford University Press, 4th Edition, 2014.
2. Bioinformatics: A practical guide to the Analysis of Genes and Proteins by Andreas Baxevanis, B.F. Francis Ouellette, Wiley Publications, 2nd Edition, 2001.

Prepared by: Dr. T. Uma Devi

Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER - III
Generic Elective – II
SDS 853 DEEP LEARNING

Hours per week: 4 End Examination: 60 Marks
Credits: 4 Sessionals: 40 Marks

Preamble:

Deep learning is a form of machine learning that enables computers to learn from experience and understand the world in terms of a hierarchy of concepts. Deep learning methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have thrown light on sequential data such as text and speech.

Course Objectives:

- To learn the fundamental principles, theory and approaches for learning with deep neural networks.
- To demonstrate the key concepts, issues and practices when training and modelling with deep architectures.
- To learn the main variants of deep learning (such as convolutional and recurrent architectures) and their typical applications.

UNIT - I

Review of Machine Learning: The Learning Machines, The Math Behind Machine Learning: Linear Algebra, The Math Behind Machine Learning : Statistics, Work of Machine Learning, Logistic Regression, Evaluating Models, Building an understanding of machine learning. (10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn the basics of deep learning. (L2)
- Recollect the basic machine learning models. (L1)
- Learn the measures of evaluating a model. (L2)

UNIT – II

Foundations of Neural Networks and Deep Learning : Neural Networks, Training Neural Networks, Activation Functions, Loss Functions , Hyper parameters. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Understand the foundations of neural networks. (L2)
- Understand the techniques of training neural networks.(L2)
- Learn different activation functions, loss functions, hyper parameter. (L3)

UNIT - III

Fundamentals of Deep Networks : Defining Deep Learning, Common Architectural Principles of Deep Networks, Building Blocks of Deep Networks. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn the evolution of deep neural networks.(L2)
- Understand the building blocks and architectural principles of Deep Learning.(L3)

UNIT – IV

Major Architectures of Deep Networks: Unsupervised Pre-trained Networks, Convolutional Neural Networks (CNNs), Recurrent Neural Networks, Recursive Neural Networks. (10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn the major architectures of deep neural networks. (L2)
- Identify the difference of different architectures. (L3)

UNIT - V

Building Deep Networks: Matching Deep Networks to the Right Problem, The DL4J Suite of Tools, Basic Concepts of the DL4J API, Modelling CSV Data with Multilayer Perceptron Networks, Modelling Handwritten Images Using CNNs, Modelling Sequence Data by Using Recurrent Neural Networks, Applications of Deep Learning in Natural Language Processing.

(10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Build a Deep Neural network using APIs. (L4)
- Understand how to apply a variety of learning algorithms to data. (L3)
- Understand how to perform evaluation of learning algorithms and model selection. (L4)

Course Outcomes:

- Identify the deep learning algorithms which are more appropriate for various types of learning tasks in various domains.
- Able to design and calibrate models using different types of neural networks.
- Implement deep learning algorithms and solve real-world problems.

Text-book:

1. Deep Learning:A practitioners approach by Josh Patterson & Adam Gibson, Oreilly publications, 1st edition, 2017.

Reference Books:

1. Deep Learning by Ian Goodfellow, Yoshua Bengio. Aaron Courville. The MIT Press, 2016.
2. Deep Learning, A Practical Approach by Rajiv Chopra, Khanna Book Publishing**,** 2018.

Prepared by: Dr. T. Uma Devi

Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER - III
Generic Elective – II
SDS 855 NATURAL LANGUAGE PROCESSING

Hours per week: 4 End Examination: 60 Marks
Credits: 4 Sessionals: 40 Marks

Preamble:

This course provides an introduction to the computational modelling of natural language.

Course Objectives:

- Acquaintance with natural language processing and learn how to apply basic algorithms in this field.
- To recognize the significance of pragmatics for natural language understanding.
- Capable of describing the application based on natural language processing and to show the points of syntactic, semantic and pragmatic processing.

UNIT – I

Regular Expressions, Text Normalization, Edit Distance: Regular Expressions, Words, Corpora, Text Normalization, Minimum Edit Distance. **N-Gram Language Models:** N-grams, Evaluating Language Models, Generalization and Zeros, Smoothing, Kneser-Ney Smoothing, The web and stupid Backoff, Advanced Perplexity's Relation to Entropy. (6 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn features of NLP. (L1)
- Learn pattern matching methods using different types of regular expressions. (L2)
- Understand the concepts of morphology, syntax, semantics and pragmatics of the language.(L3)
- Understand the applications of NLP. (L3)

UNIT – II

Parts of Speech Tagging: English Word Classes, The Penn Tree bank part of speech Tagset, Part of Speech tagging, HMM part of speech tagging, Maximum Entropy Markov Models, Bi-directionality, Part of Speech tagging for other languages. (6 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn basics of English language. (L2)
- Learn basic structure of English sentence and its syntax. (L2)
- Understand the complexity of English language and hence techniques of English language processing. (L3)
- Understand the elements and applications of Part-of-speech tagging. (L3)

UNIT – III

Formal Grammars of English: Constituency, Context Free Grammars, Some Grammar Rules for English, Treebanks, Grammar Equivalence and Normal Form, Lexicalized Grammars.
Syntactic Parsing: Ambiguity, CYK Parsing, Partial parsing. (8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Understand approaches to syntax and semantics in NLP.(L3)
- Learn different types of grammars. (L2)

- Learn different types of parsing techniques. (L2)

UNIT – IV

Dependency Parsing: Dependency Relations, Formalisms, Treebank, Transition Based Dependency Parsing, Graph based dependency parsing, Evaluation.

Representation of Sentence Meaning: Computational Desiderata for Representations, Model – Theoretic Semantics, First Order Logic, Event and State Representations, Description Logics.

(8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Understand the way to parse a sentence, recognize its syntactic structure, and construct representation of meaning. (L2)
- Provide the student with knowledge of various levels of analysis involved in NLP. (L4)

UNIT – V

Semantic Parsing : Information Extraction: Named Entity Recognition, Relation Extraction, Extracting Times, Events and their times, Template Filling.

Lexicons for Sentiment, Affect and Connotation: Defining Emotion, Available Sentiment and Affect Lexicons, Creating affect lexicons by human labeling, semi supervised induction of affect lexicons, supervised learning of word sentiment, Using lexicons for Sentiment Recognition.

(10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Understand approaches to syntax and semantics in NLP.(L3)
- Presents an introduction to the computational modelling of natural language and identifying the sentiment of the text.(L3)
- Building robust systems to perform linguistic tasks with technological applications.(L4)

Course Outcomes:

Upon completion of this course, student will be able to:

- Understand the theoretical underpinnings of natural language processing in linguistics and formal language theory.
- Learn Manipulating large corpus, exploring linguistic models and testing empirical claims.
- Able to build robust systems to perform linguistic tasks.

Text Book:

1. Speech and Language Processing- Daniel Jurafsky, James H Martin, 2nd edition, PHI, 2008.

Reference Book:

1. Natural Language Processing using Python by Steven Bird, Ewan Klien, Edward Loper, 1st edition, Oreilly Publications, 2009.

Prepared by: Dr. T. Uma Devi

Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE
SEMESTER - III
SDS 821 DATA ANALYTICS LAB

Hours per week: 4 Continuous Evaluation: 100 Marks
Credits: 2

1. Exploring Hadoop Distributed File System (HDFS).
2. Implementation of file system commands using Hadoop file system API.
3. Implementation of HDFS file watchers to monitor the events on specific directory path.
4. MapReduce: Running the WordCount Program, calculating the size of each word, no. of alphabets in the entire dataset.
5. Modern file formats like Parquet, ORC ,JSON and avro along with compressions like snappy, gz, bzip2,lzo etc. both with Pig and Hive
   using partitioning, bucketing, mapjoins, vectorizations in hive.
6. Preparing hive tables with above said file formats using sqoop & Hcatalog from RDBMS like sql-server.
7. Understanding HA and launching jobs on YARN cluster mode.

Learning Outcomes:

- Installation of VM player and Hadoop, Important Configuration files in a Hadoop Cluster, file system commands, Importing Hadoop Jars, Data Loading Techniques. (L1)
- Learn Data Loading Techniques, how to setup single node Hadoop cluster. (L3)
- Learn the working of mapreduce on data stored in HDFS. (L3)
- Understand concepts like Input Splits in mapreduce, Combiner & Partitioner on mapreduce using different data sets.(L2)
- Anatomy of File Read and Write & how mapreduce works. (L4)
- Understand the different classes in Java, Creating a Jar, Executing the program with data set, transferring the results into new File in HDFS. (L3)
- Learn Hive DDL – Create/Show/Drop Tables, Internal and External Tables.
- Learn Hive DML – Load Files & Insert Data. (L3)
- Learn the fundamentals of YARN, which runs processes on a cluster. (L3)

Text Books:

1. Big Data Black Book by Dt Editorial Services, Dreamtech Publications, 2016.
2. Hadoop The Definitive Guide by Tom White, O'reilly ,4th Edition,2016.
3. Programming Hive- Jason Rutherglen, Dean Wampler, Edward Capriolo, O'reilly Publisher, 1st edition, 2012.

Prepared by: Dr. T. Uma Devi Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE

SEMESTER - III

SDS 823 MACHINE LEARNING LAB USING PYTHON

Hours per week: 4 Continuous Evaluation: 100 Marks

Credits: 2

1. Creating a Data Frame in Pandas from csv files.
2. Importing Data with Pandas – adding columns to the data frame.
3. Handling Missing Data- drop, fill, aggregate functions.
4. Indexing Data Frames with Pandas, Indexing Using Labels in Pandas.
5. Exploratory Data Analysis with Pandas- for both one dimensional and two dimensional data (series or data frames) - describe, group data, ANOVA, correlation and correlation methods, rank.
6. Calculating Mean, Trimmed Mean, Weighted Mean, Median,
7. Plotting using pandas- Exploratory analysis based on the plots.
8. Data Visualization with different charts in python.
9. Apply PCA function. Find Eigen Values and Eigen Vectors.
10. Working with JSON Data with python.
11. Use OpenCV to find a face in an image.
12. A weather prediction model that predicts if there'll be rain or not in a particular day with decision tree regression concept.
13. A Python script to create a confusion matrix on a predicted model.
14. Consider a dataset where we have a value of response y for every feature x.
    a) Find a line which fits best and predict the response for any new feature values using simple linear regression.
    b) Find the errors using Least Squares technique to fine tune the model.

X	0	1	2	3	4	5	6	7	8	9
Y	1	3	2	5	7	8	8	9	10	12

15. Consider a dataset with p features (or independent variables) and one response (or dependent variable). Also the dataset contains n rows/observations.
    a. Find the regression line using multiple linear regression.
    b. Find the residual error of i^{th} observation.
16. A researcher has collected data on three psychological variables, four academic variables (standardized test scores), and the type of educational program the student is in for 600 high school students. She is interested in how the set of psychological variables is related to the academic variables and the type of program the student is in using Multivariate Regression.
17. Demonstrate to find the values of the parameters of a function that minimizes the cost function using Stochastic Gradient Descent.
18. A python program to explore your data with matplotlib and PCA, preprocess your data with normalization. Split the data into training and test sets. construct an unsupervised model (K-means algorithm) to fit the model to the data, predict values, and validate the model that is built.
19. Multidimensional data analysis in Python- import, Clustering, Exploratory Data Analysis.
20. Demonstrate to perform support vector classifier on a non linear dataset using a linear kernel.

Course Outcomes:

- Learn to Load datasets. (L1)

- Learn about the various libraries offered by Python to manipulate, preprocess and visualize data. (L3)
- Learn the technique to reduce the number of variables using Feature Selection and Feature Extraction. (L3)
- Learn in building models and model persistence using regression, classification. (L3)
- Learn various machine learning algorithms like KNN, Decision Trees, SVM, Clustering in detail. (L3)
- Learn to use optimization techniques to find the minimum error in your machine learning model. (L3)

Text Books:

1.Python Machine Learning by Sebastian Raschka and Vahid Mirialili, Packt publishing, 2nd Edition, 2017.

2.Introduction to Machine Learning with Python by Andreas C. Müller and Sarah Guido, Orielly, 1st Edition, 2016.

3. Machine Learning in Python by Michael Bowles, Wiley Publishers, 2018.

Prepared by: Dr. T. Uma Devi

Verified by : Dr. T. Uma Devi

M.Sc DATA SCIENCE

SEMESTER - III

SDS 825 INDUSTRIAL TRAINING AND SEMINAR

Hours per week: 2 Continuous Evaluation: 100 Marks

Credits: 2

SEMESTER - IV

SDS 891 Project Work

Hours per week: 3 End Examination: 50 Marks

Credits: 8 Sessionals: 150 Marks

****

GANDHI INSTITUTE OF TECHNOLOGY AND MANAGEMENT (GITAM)

(Deemed to be University, Estd. u/s 3 of UGC Act 1956)

VISAKHAPATNAM ∗HYDERABAD ∗BENGALURU

Accredited by NAAC with 'A' Grade

Skill Enhancement Courses
offered by
Department of Computer Science

(W.e.f 2019-20 Admitted batch)

Website: www.gitam.edu

SSE 701: BASIC COMPUTER TOOLS

Hours per week: 3 Examination: 100 Marks
Credits: 2

Preamble: The course gives an understanding about the characteristics and classification of computers, various components of computer along with different operating systems that are available. It gives a hands on training on the packages MS-Word, MS-Power Point and MS-Excel. The course also comprehends AI tools.

Basics of Computers: Definition of a Computer - Characteristics and Applications of Computers – Block Diagram of a Digital Computer – Classification of Computers based on size and working – Central Processing Unit – I/O Devices, Primary, Auxiliary and Cache Memory – Memory Devices. Software, Hardware, Firmware and People ware – Definition and Types of Operating System – Functions of an Operating System – MS-DOS –MS Windows, UNIX. Introduction to AI tools.

MS-Word

Features of MS-Word – MS-Word Window Components – Creating, Editing, ormatting and Printing of Documents – Headers and Footers – Insert/Draw Tables, Table Auto format – Page Borders and Shading – Inserting Symbols, Shapes, Word Art, Page Numbers, Equations – Spelling and Grammar – Thesaurus – Mail Merge.

MS-PowerPoint

Features of PowerPoint – Creating a Blank Presentation - Creating a Presentation using a Template - Inserting and Deleting Slides in a Presentation – Adding Clip Art/Pictures -Inserting Other Objects, Audio, Video- Resizing and Scaling of an Object –Slide Transition – Custom Animation.

MS-Excel

Overview of Excel features – Creating a new worksheet, Selecting cells, Entering and editing Text, Numbers, Formulae, Referencing cells – Inserting Rows/Columns –Changing column widths and row heights, auto format, changing font sizes, colors, shading.

Reference Books:

1. Fundamentals of Computers by V.RajaRaman, PHI Learning Pvt. Ltd, 2010.
2. Microsoft Office 2010 Bible by John Walkenbach, Herb Tyson, Michael R. Groh and Faithe Wempen, Wiley Publications, 2010.

Learning Outcomes:

- Able to understand fundamental hardware components that make up a computer's hardware and the role of each of these components
- Understand the difference between an operating system and an application program, and what each is used for in a computer.
- Acquire knowledge about AI tools.
- Create a document in Microsoft Word with formatting that complies with the APA guidelines.
- Write functions in Microsoft Excel to perform basic calculations and to convert number to text and text to number.
- Create a presentation in Microsoft PowerPoint that is interactive and legible content

****

SSE 703 : INFORMATION TECHNOLOGY TOOLS

Hours per week: 3 Examination: 100 Marks

Credits: 2

Preamble: The course enables the student to understand networking concepts related to Internet and introduce the social Networking sites and working of Email. It gives orientation of Block Chain technology. It give hands on training in SPSS, R Programming and creation of simple HTML documents.

Introduction to Internet: Networking Concepts, Data Communication –Types of Networking, Internet and its Services, Internet Addressing –Internet Applications–Computer Viruses and its types –Browser –Types of Browsers.

Internet applications: Using Internet Explorer, Standard Internet Explorer Buttons, Entering a Web Site Address, Searching the Internet– Introduction to Social Networking: twitter, tumblr, Linkedin, facebook, flickr, skype, yahoo!, google+, youtube, WhatsApp, etc.

E-mail : Definition of E-mail, Advantages and Disadvantages,User Ids, Passwords, Email Addresses, Domain Names, Mailers, Message Components, Message Composition, Mail Management, Email Inner Workings.

WWW-Web Applications, Web Terminologies, Web Browsers ,URL–Components of URL, Searching WWW –Search Engines and Examples.

Block Chain technology: What is Block Chain, Blockchain Architecture, How Block chain Transaction Works? Why do we need Blockchain? Block chain versions, Block chain Variants, Block chain Use Cases, Important Real-Life Use Cases of Block chain Bitcoin cryptocurrency: Most Popular Application of Block chain, Block chain vs. Shared Database, Myths about Block chain, Limitations of Block chain technology.

SPSS : SPSS Commands, Descriptive Statistics, Hypothesis Testing, Test of Difference, Analysis of Variance- One Way ANOVA, Non Parametric Tests, Correlation Analysis, Regression Analysis.

R Programming: Becoming familiar with R, Working with Objects, Introduction to Graphical Analysis.

HTML: WEB Terminology, Structure of HTML Document, HTML – Head and Body tags, Semantic tags- HR- Heading, Font, Image & Anchor tags, Different Types of Lists using Tags, Table Tags, Image Formats – Creation of Simple HTML Documents.

Reference Books:

- In-line/On-line : Fundamentals of the Internet and the World Wide Web by Raymond Greenlaw and Ellen Hepp, 2nd Edition, TMH.
- Microsoft Office 2010 Bible by John Walkenbach, Herb Tyson, Michael R. Groh and Faithe Wempen, WileyPublications.

Learning Outcomes:

- Enable to understand the basic networking concepts, types of networks, Internet Explorer and www.
- Outline the Block chain architecture, Bitcoin Crypto currency and Limitations of Block Chain.
- Choose different statistical tests to be performed on the data sets.
- Demonstrate the R programming with simple graphs.
- To make use of commands to structure HTML document.

****