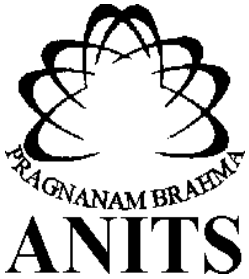
Anil Neerukonda Institute of Technology & Sciences (Autonomous)
(Permanent Affiliation by Andhra University & Approved by AICTE
Accredited by NBA (ECE, EEE, CSE, IT, Mech. Civil & Chemical) & NAAC)
Sangivalasa-531 162, Bheemunipatnam Mandal, Visakhapatnam District
Phone: 08933-225083/84/87 Fax: 226395
Website: www.anits.edu.in email: principal@anits.edu.in



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING WITH ARTIFICIAL INTELLIGENCE & MACHINE LEARNING, DATA SCIENCE

VISION AND MISSION OF ANITS

VISION

ANITS envisions to emerge as a world-class technical institution whose products represent a good blend of technological excellence and the best of human values.

MISSION

To train young men and women into competent and confident engineers with excellent communicational skills, to face the challenges of future technology changes, by imparting holistic technical education using the best of infrastructure, outstanding technical and teaching expertise and an exemplary work culture, besides molding them into good citizens.

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING WITH DATA SCIENCE

The Genesis and Growth:

The CSE(Data Science) Department of ANITS was started in the academic year 2020-21 with an intake of 70 for B Tech program and the present in-take from the academic year 2023-24 is 180. The department also offers M Tech (Data Science) from the Academic Year 2020-21 and the in-take is 18. Now the regulation updated from R20 to R23.

The Human resources and Infrastructure:

The department is presently having 24 highly qualified faculty members who are having rich academic and industrial experience. To make teaching and learning processes more effective, 1:15 faculty student ratio is being maintained in the department since inception. The department possess best infrastructure facilities. To serve the student needs beyond curriculum and make them industry ready, additional experiments are being conducted in the labs and software were installed which are beyond the prescribed syllabus.

The faculty strongly believe that for disseminating knowledge to the students and to make them abreast with latest developments, good research activities are the need of the day along with regular teaching. About 30 plus papers were published by the faculty in various international peer reviewed journals and Conferences in the last 2 years. Faculty are dynamic and regularly upgrade their skills by attending various workshops and seminars. The departments' activities and infrastructure are being further improved with funding from agencies like AICTE, UGC etc.

ANITS has MOU with NPTEL and Infosys Spring Board to encourage self-learning in students.

The students and the outcomes:

Due to the subtle efforts of faculty and the infrastructure available, the department attracts students of good EAMCET ranks in this region. The academic performance of the students is also a matter of prestige to the department as it stands first amongst all the affiliated colleges of Andhra University on the basis of student pass percentage. Special impetus is given to students to improve their managerial skills. They are motivated to organize various events in the department by themselves which includes national seminars and technical symposiums. They are also encouraged to participate in various events organized by reputed colleges and have won many prizes every year bringing laurels to the department.

To make the students aware of latest technologies and also industrial needs, guest lectures by eminent persons from academics and industry, industrial visits etc., are being organized regularly in the department. Apart from this, CSE (AI&ML, DS) department has CSI student chapter. The department is also having Student Activity Center (SAC) to encourage the students in Entrepreneurship, Sports, Social responsibility and Cultural events.

Department of CSE (AI&ML, DS)

VISION

To produce professionally competent, socially conscious ethical engineers to meet the needs of industry and the society.

MISSION

The department mission is to

- Provide state of the art academic environment to produce self-sustainable professionals.
- Impart high-quality, value-based education in cutting-edge technologies to solve real world problems
- Inculcate moral values and leadership skills to become holistic professionals.

PROGRAMME OUTCOMES (POs)

- PO-1 An ability to independently carry out research /investigation and development work to solve practical problems.

- PO-2 An ability to write and present a substantial technical report/document.

- PO-3 Students should be able to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program.

PROGRAMME SPECIFIC OUTCOMES (PSOs)

- Apply Artificial Intelligence and Machine Learning concepts for development of intelligent systems and applications to multidisciplinary problems.

- Develop and implement automated solutions that leverage AI and ML computing techniques to address current societal problems.

ACADEMIC REGULATIONS FOR M.TECH PROGRAMME

W.E.F. THE ADMITTED BATCH OF 2023-24

I. Admissions:

Admissions into first year of M.Tech Programme of the Institute will be as per the norms stipulated by Andhra University & Andhra Pradesh State Council for Higher Education (APSCHE), Govt. of Andhra Pradesh.

II. Programmes Offered:

The following are the M.Tech Programmes offered by the Institute:

1. Food Processing Technology - Chemical Engineering Department
2. Soil Mechanics - Civil Engineering Department
3. Computer Science and Engineering (Data Science) - CSE (AI&ML, DS) Department
4. Computer Science and Engineering (AI & ML) - CSE (AI&ML, DS) Department
5. VLSI & Embedded Systems – ECE Department
6. Control Systems –EEE Department
7. Machine Design–Mechanical Engineering Department

I. Structure of the M.Tech Programme:

The normal duration of the course is 2 academic years for M.Tech Degree. Candidates should pursue a regular course of study, as detailed below, for not less than two academic years which consists of 4 semesters and should fulfill the academic requirements and pass all the prescribed examinations for the award of degree.
The curriculum of M.Tech programme is designed to have a total of 68 credits to get the degree awarded.

II. Duration of the Programme:

The duration of the programme is 2 academic years consisting of 2 semesters in each academic year. A student is permitted to complete the programme in a stipulated time frame of 4 consecutive academic years from the date of initial admission and if fails will forfeit his seat in M.Tech programme.

III. Medium of Instruction:

The medium of instruction and examination is English.

IV. Minimum Instruction Days:

Each semester normally consists of a minimum of 16 weeks of instruction.

V. Academic Calendar:

The dates of all important events, such as commencement of class work, examinations, vacations, etc., during the academic year will be specified in the Academic Calendar of the Institute, as approved by the Academic Council.

VI. Examinations & Evaluation Process:

The performance of a student in each semester shall be evaluated course-wise with a maximum of 100 marks each for theory and practical courses.

(a) Theory Course:

For all lecture based theory courses, the assessment shall be for 40 marks through internal evaluation and 60 marks through external semester-end examination of three hours duration. The sessional marks shall be awarded through internal evaluation by the teachers concerned based on the continuous assessment which includes class tests, quiz, viva-voce, assignments, student regularity, two mid-examinations etc., according to a scheme notified by the department at the beginning of the semester.

Out of the 40 internal evaluation marks, 20 marks are assigned for 2 internal- mid exams, 10 marks are assigned for assignments, 5 marks are assigned for projects/case studies/quiz/tests and 5 marks are assigned for attendance. The average of 2 internal-mid exams is considered for 20 marks. Under any circumstances, no re-examination shall be conducted for the internal mid examinations.

External evaluation:

The question paper shall be set externally and the answer scripts are valued through a double valuation system. The average of the two valuations will be taken for award of marks. In case, the difference of the marks obtained in the two valuations is more than 20% then a third examiner shall value the script. Out of the three valuations, the average of marks obtained in third valuation and the marks obtained nearer to third valuation out of first two valuations

shall be considered. No revaluation for any subject/course shall be entertained as already double valuation system is in existence. However, recounting is allowed on the request of the candidate on payment of specified fee. Challenge valuation shall also be entertained on payment of specified fee.

(b) Laboratory Course:

Each student will perform about 10 to 12 experiments in each laboratory course. Laboratory course will be evaluated for 100 marks, out of which 50 marks are for external examination and 50 marks are for internal evaluation. The internal marks are awarded based on continuous assessment, record work, and internal lab examination and student regularity. The external examination will be conducted by two examiners, one of them being laboratory class teacher as internal examiner (nominated by the Principal on recommendation of HOD) and an external examiner nominated by the Principal from the panel of experts recommended by the HOD.

A candidate shall be declared to have passed in any theory subject/course if he secures not less than 40% in external theory examination and also a minimum of 50% of total marks of that course which assures a minimum of 'E' grade.
A candidate shall be declared to have passed in any practical course if he secures not less than 50% in external lab examination and also a minimum of 50% of total marks of that course which assures a minimum of 'E' grade.

Any student appearing for the semester-end practical examination is eligible only if he submits the bona-fide record certified by the laboratory class teacher and the HOD.

(c) Seminar

Conducting seminars plays a major role in helping students develop their presentation skills, in which the student may present any technical content (innovative or upcoming) related to his or her specialization.
If a course contains two seminar courses in two different semesters, Seminar-I may deal with any general topic, and Seminar-II may deal with technical topics.
A common way to divide the 50 marks for a seminar evaluation involves allocating scores or points to different aspects of the seminar presentation as follows:

- Content (15 marks)
- Presentation Skills (20 marks)
- Handling of queries and discussions (10 marks)
- Quality of visuals, clarity and usefulness of handouts (5 marks)
- The faculty has to conduct a minimum of two seminars in a semester.

(d) Audit Course

A common way to divide the 50 marks for an audit course (non-credit) is as follows:

- Mid Exams: 20 marks
- Group Projects/Assignments: 15 marks
- Class Tests/Quizzes: 10 marks
- Attendance: 5 marks

(e) Thesis Work:

The thesis work shall be carried out in two semesters of one full academic year. The students will be allotted for thesis by the department committee to various faculty members who act as guides. However, a student can carry-out his thesis work either in the department or in any other industry / research institute. In any such request to carry out thesis work outside the college, the permission of the Principal and an internal guide is mandatory. Such students should report to the internal guide once in a week essentially through mail or other communication.

The progress report of such work is to be submitted by the guide/external guide every month to the HOD. If the work is not found satisfactory, the HOD has the right to call back the student with the permission of the Principal. In any case the time and conditions for submission of the thesis will be same as for the regular candidates studying in the college.

The third semester work is evaluated internally by the committee nominated by the HOD consisting a minimum of four members (concerned in area of specialization) including the HOD. If the work is not satisfactory, the candidate has to improve to the satisfaction of the committee within one month from the end of the semester to carry on his fourth semester work. If he fails to satisfy the committee in the second attempt he has to get readmitted into the third semester as per college norms. The grades will be awarded just as in the case of laboratory work. An internal viva-voce by a committee nominated by the HOD is a prerequisite for the submission of the thesis. The fourth semester evaluation will be done through the viva-voce examination on the thesis by a board consisting of the following four examiners after submission of the thesis by the candidate duly certified by the Guide and the HOD.

1. The Head of the Department as Chairman
2. Senior Professor in the Department
3. Internal Guide and External Guide (if any)
4. External examiner nominated by the Principal from a panel recommended by the HOD.

The panel of the external subject experts shall be submitted to the Principal by the HOD in mutual consent with the guide and other subject experts of the Department.

The scheme of valuation of the thesis shall be according to the format given by Head and PG coordinator.
If the candidate fails in the viva-voce examination of the thesis, he has to reappear for the viva-voce. The candidate has to bear the charges for re-conducting the viva-voce.

The prerequisite for submission of the M.Tech thesis is that one should have published a paper in reputed international journal/proceedings of an annual conference.

VII. Attendance Regulations:

Attendance of a student is computed by considering total number of periods conducted in all courses as the denominator and the total number of periods actually attended by the student in all courses, as the numerator. It is desirable for a student to put in 100% attendance in all

the subjects. However, a candidate shall be permitted to appear for the semester end examination provided he/she maintains a minimum of 75% overall attendance in the semester.

The shortage of attendance on medical grounds can be condoned up to a maximum of 9% provided the student puts in at least 66% attendance and provided the Principal is satisfied with the genuineness of the reasons. The Medical Certificates are to be submitted to the Head of the Department when the candidate reports to the classes immediately after the leave. Certificates submitted afterwards shall not be entertained.
Condonation fee as fixed by the college for those who put in attendance between ≥ 66% and < 75% shall be charged before the semester-end examinations.

In the case of students who participate in co-curricular, extra-curricular activities like student seminars, N.S.S, N.C.C, Inter-collegiate tournaments and any such other activities involving the representation of the Institute, with the prior approval of the Principal, the candidate may be deemed to have attended the classes during the actual period of such activity, solely for the purpose of attendance.

A student, who could not satisfy the minimum attendance requirement of 66% in any semester, shall be declared 'Detained'. He/she is not eligible to appear for the semester end examinations. He will not be promoted to the next semester and shall have to repeat that semester with the next batch(es) of students. Such students, who are detained and seek readmission, should submit undertaking/declaration that they will abide by the regulations existing at the time of readmission.

VIII. Minimum Academic Requirements:

The following academic requirements have to be satisfied in addition to the attendance requirements mentioned in item No. IX.

i. A student shall be deemed to have satisfied the minimum academic requirements and earned the credits allotted to each theory subject if only he secures not less than 40%marksin the semester-end examination and a minimum of 50%marks in the sum of the internal evaluation and semester-end examination taken together. In the labs/projects, the student should secure a minimum of 50% marks in the external examination and a minimum of 50% marks in the sum of internal evaluation and external examination evaluation taken together.

ii. Students, who fail to complete their two year course study within four academic years from the year of their admission or fail to acquire the credits stipulated for the course shall forfeit their seat in M.Tech course and their admission shall stand cancelled.

IX. Award of Grades:

The absolute grading system is adopted as follows:

S.No.	Range of Marks {%}	Grade	Description	Grade Points
1	90-100	O	Outstanding	10
2	80-89	A	Excellent	9
3	70-79	B	Very Good	8
4	60-69	C	Good	7
5	55-59	D	Fair	6
6	50-54	E	Satisfactory	5
7	49 and below	F	Fail	0
8	The grade 'I' represents absent (subsequently changed into pass or higher grades.)	I	Absent	0

The performance of a student at the end of the each semester is indicated in terms of Semester Grade Point Average (SGPA). The SGPA is calculated as below:

$$SGPA = \frac{\Sigma \text{ (Credits of a course x Grade points awarded for a course)}}{\Sigma \text{ (Credits of a course)}}$$

SGPA is calculated for the candidates who have passed in all the courses in that semester.

Cumulative Grade Point Average (CGPA) will be calculated from II semester onwards up to the final semester and its calculation is similar to that of SGPA, considering all the courses offered from the first semester onwards.
CGPA is calculated for those who clear all the courses in all the previous semesters.

Award of Class:

For the award of class, a total of best 74 credits are considered. A candidate, who becomes eligible for the award of M.Tech. Degree, shall be placed in one of the following classes.

S.No.	Class	CGPA
1	First Class with Distinction	7.5 or more*
2	First Class	6.5 or more but less than 7.5
3	Second Class/Pass	5.0 or more but less than 6.5

***First class with Distinction will be awarded only to those students who clear all the subjects of the program in first attempt of regular examinations.**
The CGPA can be converted to aggregate percentage by multiplying CGPA with 10, in case of requirement by any other university or for any other purpose.

X. Eligibility for Award of M.Tech. Degree:

A student shall be eligible for the award of the M.Tech degree if he/she fulfils all the following conditions:

1. Registered and successfully completed all the components prescribed for eligibility in the programme of study to which he/she is admitted within the stipulated period,
2. Obtained CGPA greater than or equal to 5.0 (Minimum requirement for Pass),
3. No disciplinary action is pending against him/her and
4. Has no dues to the Institute including hostels.

XI. Malpractices:

The Controller of Examinations/Dean of Examinations shall refer the cases of suspected malpractices in mid examinations and semester-end examinations to Malpractice Enquiry Committee constituted by the Institute. Such committee shall follow the approved scales of punishment. The Principal shall take necessary final action against the erring students based on the recommendations of the committee.

XII. Amendments to Regulations:

The Institute may, from time to time, revise, amend, or change the Regulations, Schemes of Examinations, and / or Syllabi and the changes or amendments made shall be applicable to all the students with effect from the dates notified by the Institute.

XIII. General:

- Where the words 'he', 'him', 'his', occur in the regulations, they include 'she', 'her', 'hers'.
- The academic regulation should be read as a whole for the purpose of any interpretation.
- In case of any doubt or ambiguity in the interpretation of the above rules, the decision of the Principal is final.

R-23 COURSE STRUCTURE
M.TECH (Data Science)
DEPT. OF CSE (AI & ML, DS): ANITS

I Year – I Semester

Course Code	Title of the Course	Category	Periods per week			Max. marks		Total Marks	Credits
			L	T	P	Sess.	End Exam		
23DS111	Mathematical Foundation for Data Science	BS	3	0	0	40	60	100	3
23DS112	Principles of Data Science	PC	3	0	0	40	60	100	3
23DS113	Professional Elective – I	PE	3	0	0	40	60	100	3
23DS114	Professional Elective – II	PE	3	0	0	40	60	100	3
23DS115	Research Methodology and IPR	MC	2	0	0	40	60	100	2
23DS116	Data science using Python Lab	PC	0	0	3	50	50	100	2
23DS117	Professional Elective-II LAB	PE	0	0	3	50	50	100	2
23DS118	Universal Human Values and Professional Ethics	Audit	2	0	0	50	--	50	0
	Total		**16**	**0**	**6**	**350**	**400**	**750**	**18**

Professional Elective–I:	A. Feature Engineering B. Principles of Big Data C. Social Media Data Mining & Analytics
Professional Elective–II:	A. Information Security B. Data Analytics C. Machine Learning

I Year – II Semester

Course Code	Title of the Course	Category	Periods per week			Max. marks		Total Marks	Credits
			L	T	P	Sess.	End Exam		
23DS121	Computational Intelligence	PC	3	0	0	40	60	100	3
23DS122	Data Visualization	PC	3	0	0	40	60	100	3
23DS123	Professional Elective – III	PE	3	0	0	40	60	100	3
23DS124	Professional Elective – IV	PE	3	0	0	40	60	100	3
23DS125	Data Visualization Lab	PC	0	0	3	50	50	100	2
23DS126	Professional Elective-IV Lab	PE	0	0	3	50	50	100	2
23DS127	Seminar	SC	0	0	3	50	--	50	2
23DS128	Teaching and Learning in Higher Education	Audit	2	0	0	50	--	50	0
	Total		**14**	**0**	**9**	**360**	**340**	**700**	**18**

Professional Elective–III:	A. Stream Analytics B. Image Processing C. Business Intelligence
Professional Elective – IV:	A. Predictive Analytics & Data mining B. No SQL C. Natural Language Processing

II Year – I Semester

Course Code	Title of the Course	Category	Periods per week			Total	Max. marks		Total Marks	Credits
			L	T	P		Sess.	End Exam		
23DS211	MOOCs-I	OE	0	0	0	0	100	—	100	3
23DS212	MOOCs-II	OE	0	0	0	0	100	—	100	3
23DS213	Project Phase-I	PR	0	0	20	20	100	—	100	10
Total			**0**	**0**	**20**	**20**	**300**	**—**	**300**	**16**

II Year – II Semester

Course Code	Title of the course	Category	Periods per week			Total	Max. Marks		Total Marks	Credits
			L	T	P		Sess.	End Exam		
23DS221	Project Phase-II	PR	0	0	32	32	100	100	200	16
Total			**0**	**0**	**32**	**32**	**100**	**100**	**200**	**16**

The prerequisite for submission of the M Tech thesis is that one should communicate his/her work to any referred journal or Publication in a conference/journal.

SEM-I

MATHEMATICAL FOUNDATION FOR DATA SCIENCE	
Code: 23DS111	**Credits: 3**
Instruction : 3 periods	**Sessional Marks: 40**
End Exam : 3 Hours	**End Exam Marks: 60**

Course Objectives:

To provide the basic knowledge of fundamental concepts in linear algebra, vectors, probability and statistics required for a program in artificial intelligence.

Course Outcomes: By the end of the course, students will be able to

Prerequisites: Matrices, Differentiation, Number system.

1.	Apply the number theory concepts to cryptography domain.
2.	Understand the concepts eigen values, eigen vectors and singular value decompositions.
3.	Understand the concepts of vector calculus.
4.	Apply the knowledge of probability and distributions in multivariate analysis.
5.	Identify appropriate stochastic process model for a given problem.

CO-PO –PSO Mapping:

CO	PO			PSO	
	1	2	3	1	2
1	1	-	2	1	1
2	2	-	2	2	2
3	2	1	2	2	2
4	1	-	2	2	2
5	2	1	2	2	2

Correlation levels 1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)

SYLLABUS

UNIT I **10 Periods**

Greatest common divisors and Prime factorization : Greatest common divisors – The Euclidean algorithm – The fundamental theorem of arithmetic – Factorization of integers and the Fermat numbers.

Congruences: Introduction to congruences – Linear congruence – The Chinese remainder theorem – System of linear congruences.

UNIT II **10 Periods**

Matrix Decompositions: Determinant and trace – Eigen values and Eigen vectors – Cholesky decomposition – Eigen decomposition and diagonalization – Singular value decomposition – Matrix approximation.

UNIT III **10 Periods**

Vector Calculus : Differentiation of univariate functions – Partial differentiation and Gradients – Gradients of vector valued functions – Gradients of matrices – Useful identities for computing gradients – Backpropagation and automatic differentiation – Higher order derivatives – Linearization and multivariate Taylor series.

UNIT IV **10 Periods**

Probability and Distributions: Construction of a probability space – Discrete and Continuous probabilities – Sum rule, Product rule, and Baye's theorem – Gaussian distribution – Conjugacy and the exponential family – Change of variables/Inverse transform.

UNIT V **10 Periods**

Stochastic Processes and Markov Chains: Introduction to Stochastic processes – Markov process – Transition probability – Transition probability matrix – First order and Higher order Markov process – n step transition probabilities – Markov chain – Steady state condition – Markov analysis.

TEXT BOOKS:

1. **Kenneth H. Rosen,** Elementary number theory & its applications, sixth edition, Addison Wesley.
2. **M. P. Deisenroth, A. A. Faisal and C. S. Ong,** Mathematics for Machine Learning, Cambridge University Press (1st Edition).
3. **Kishore S. Trivedi,** Probability & Statistics with Reliability, Queuing and Computer Applications, 2nd Edition, John Wiley and Sons Ltd., 2016.

REFERENCES:

1. **Stephen Boyd and Lieven Vandenberghe,** Introduction to Applied Linear Algebar – Vectors, Matrices, and Least squares, Cambridge Univ. Press (2018).
2. **S. Axler,** Linear algebra done right, Springer International Publishing (3rd Edition).
3. **E. Keryszig,** Advanced Engineering Mathematics, John Wiley and Sons, Inc., U. K. (10th Edition).

PRINCIPLES OF DATA SCIENCE	
Code: 23DS112	**Credits: 3**
Instruction :3 Periods	**Sessional Marks: 40**
End Exam: 3 Hours	**End Exam Marks: 60**

Prerequisites:

Basic skills on Probability and Statistics.

Course Objectives:

1. To impart knowledge on basics of data science and data analysis concepts that are vital for data science.
2. To develop skills for applying tools and techniques to get meaningful insights from data.

Course Outcomes:

By the end of the course, the student will be able to:	
1.	Demonstrate knowledge on the fundamental concepts of data science.
2.	Understand the importance of statistics role in data science process.
3.	Demonstrate the basic knowledge of probability-based classification model.
4.	Able to apply regression models for classifying the data.
5.	Demonstrate the basic concepts and models used to perform social network analysis.

Mapping of Course Outcomes with Program Outcomes:

CO	PO			PSO	
	1	2	3	1	2
1	2	-	2	2	2
2	3	1	3	2	2
3	3	1	2	2	2
4	3	1	3	2	3
5	2	-	2	2	3

SYLLABUS

UNIT-I: 10 periods

Introduction to Data Science, Basic terminologies of data science, The data science Venn diagram, Data science case studies, Types of data, Quantitative Vs Qualitative data, The four levels of data, Five steps of data science.

UNIT-II: 10 periods

Statistical Thinking in the Age of Big Data, Populations and Samples, Populations and Samples of Big Data, what is a model? Fitting a model, Overfitting, Statistical modeling, Probability distributions, Exploratory Data Analysis. The Data Science Process, A Data Scientist's Role in This Process.

UNIT-III: 10 periods

Conditional Probability, Bayes Theorem, The derivation of Bayes Theorem, Examples on Bayes Theorem, Naïve Bayes Theorem, Applying Bayes' Law to create a good spam filter, Laplace Smoothing, Linear regression, Why Won't Linear Regression Work for Filtering Spam? Naive Bayes for Article Classification.

UNIT-IV: 10 periods

Logistic Regression and The Underlying Math. Embedded Methods: Decision Trees. Feature Generation or Feature Extraction, Wrappers, Recommendation engines, A Real-World Recommendation Engine, Principal Component Analysis (PCA), Build Your Own Recommendation System.

UNIT-V: 10 periods

Social Network Analysis, Attribute Data versus Social Network Data, Terminology from Social Networks, Centrality Measures, More Background on Social Network Analysis from a Statistical Point of View, Representations of Networks and Eigenvalue Centrality, Example-1 Random Graphs: The Erdos-Renyi Model, The Bernoulli Network, Example-2 Random Graphs: The Exponential Random Graph Model.

TEXT BOOKS

1. Principles of Data Science, Packt Publishers, Sinan Ozdemir, 2nd Edition, 2018.
2. Data Science: Straight Talk from the Frontline, Rachel Schutt, Cathy O'Neil, Doing ,O'Reilly, 2014.

REFERENCE BOOKS

1. Python for Data Analysis, Wes McKinney, O'Reilly, 2nd Edition, 2017.

Web links:

1. https://towardsdatascience.com/
2. https://github.com/jakevdp/PythonDataScienceHandbook 5 https://www.kaggle.com

FEATURE ENGINEERING	
Code: 23DS113(A)	**Credits: 3**
Instruction: 3 Periods	**Sessional Marks: 40**
End Exam: 3 Hours	**End Exam Marks: 60**

Prerequisites:

1. Basic understanding of statistics and probability concepts.
2. Familiarity with programming and data manipulation using Python

Course Objectives:

1. Identify and address issues with variables in datasets, using appropriate data preprocessing techniques to enhance feature quality for machine learning models.
2. Understand and apply missing data imputation, categorical variable encoding, and feature extraction methods for dates, times, and text data, selecting suitable techniques based on data characteristics and modeling needs.
3. Develop proficiency in feature engineering to create meaningful feature representations, optimize model performance, and enhance interpretability in diverse data analysis and prediction tasks.

Course Outcomes:

By the end of the course, the student will be able to:	
1.	Recognize and address data concerns—missing values, outliers, distributions—informing preprocessing and feature selection for better machine learning input.
2.	Improve ML model accuracy by skillfully imputing missing data, fortifying models against incomplete datasets.
3.	Apply diverse categorical encoding methods, grasp their strengths and limitations, choose optimal approaches based on data and ML task relevance.
4.	Extract meaningful features from datetime and time variables, enabling them to leverage temporal information effectively for machine learning tasks and time-based data analysis.
5.	Process text data, create features, integrate into ML pipelines for effective use in text-based applications

Mapping of Course Outcomes with Program Outcomes:

CO	PO			PSO	
	1	2	3	1	2
1	2	-	2	2	2
2	2	-	3	2	2
3	2	1	3	2	3
4	2	1	3	-	3
5	2	1	3	-	3

SYLLABUS

UNIT- I **10 Periods**

Introduction: Foreseeing Variable Problems When Building ML Models
Identifying numerical and categorical variables**,** quantifying missing data, determining cardinality in categorical variables, identifying a linear relationship, identifying a normal distribution, Distinguishing variable distribution, highlighting outliers, Comparing feature magnitude.

UNIT-II: Imputing Missing Data **10 Periods**
Removing observations with missing data, performing mean or median imputation, implementing mode or frequent category imputation, replacing missing values with an arbitrary number, capturing missing values in a bespoke category, Replacing missing values with a value at the end of the distribution.

UNIT-III: Encoding Categorical Variables **10 Periods**

Creating binary variables through one-hot encoding, performing one-hot encoding of frequent categories, replacing categories with ordinal numbers, replacing categories with counts or frequency of observations, encoding with integers in an ordered manner, encoding with the mean of the target, Encoding with the Weight of Evidence

UNIT-IV: Deriving Features from Dates and Time Variables **10 Periods**

Extracting date and time parts from a datetime variable, deriving representations of the year and month, creating representations of day and week, extracting time parts from a time variable, Capturing the elapsed time between datetime variables, Working with time in different time zones.

UNIT-V: Extracting Features from Images **10 Periods**

The Simplest Image Features (and Why They Don't Work), Manual Feature Extraction: SIFT and HOG ,Image Gradients , Gradient Orientation Histograms, SIFT Architecture,Learning Image Features with Deep Neural Networks

Textbooks:

1. Soledad Galli,*" Python Feature Engineering Cookbook* ",2020, Packt Publishing, BIRMINGHAM - MUMBAI

Reference Books:

1. Alice Zheng & Amanda Casari, " *Feature Engineering for Machine Learning*" 2018, 1st Edition, Oreilly Publications.

PRINCIPLES OF BIG DATA	
Code: 23DS113(B)	**Credits: 3**
Instruction: 3 Periods	**Sessional Marks: 40**
End Exam: 3 Hours	**End Exam Marks: 60**

Prerequisites:

1. Fundamental knowledge of data concepts,types and database systems.
2. Familiarity with distributed systems and parallel processing systems.
3. Basic Programming skills .

Course Objective: Students will be able to

1. The basic concepts like the importance of big data and big data Analytics.
2. Analyze computing abstractions used for real-time analytics and their relevance to Big Data.
3. Explore job scheduling techniques used in Hadoop to manage Big Data processing tasks.
4. Learn about performance optimization strategies for the Hadoop Distributed File System (HDFS).

Course Outcomes:

By the end of the course,the student will be able to:	
1.	Identify the key characteristics of Big Data,understand the concept of Big Data Analytics and its significance in extracting valuable insights from massive datasets.
2.	Describe the importance of Real-Time Analytics in processing data streams,Identify computing abstractions used in Real-Time Analytics.
3.	Identify and analyze different Big Data processing systems and platforms.
4.	Analyze the job scheduling process of Hadoop for better performance.Understand performance optimization techniques for Hadoop Distributed File System (HDFS).
5.	Recognize the security challenges associated with Big Data.

Mapping of Course Outcomes with Program Outcomes:

CO	PO			PSO	
	1	2	3	1	2
1	-	-	2	-	2
2	2	1	3	2	3
3	2	1	3	2	3
4	2	1	3	2	3
5	-	-	2	-	3

SYLLABUS

UNIT - I **10 Periods**

Introduction to Big Data: Big Data and its Importance - Four V's of Big Data - Drivers for Big Data - Introduction to Big Data Analytics - Big Data Analytics applications.

UNIT - II **10 Periods**

Real-Time Analytics:Introduction,Computing Abstractions for Real-Time Analytics,Characteristics of Real-Time Systems,Real-Time Processing for Big Data — Concepts and Platforms,Data Stream Processing Platforms,Data Stream Analytics Platforms

UNIT - III **10 Periods**

Resource Management in Big Data Processing Systems: Introduction,Types of Resource Management,Big Data Processing Systems and Platforms.

UNIT - IV **10 Periods**

System Optimization for Big Data Processing:Introduction,Basic Framework of the Hadoop Ecosystem,Parallel Computation Framework: MapReduce,Job Scheduling of Hadoop,Performance Optimization of HDFS.

UNIT - V **10 Periods**

Big Data Security and Privacy:Introduction,Secure Queries Over Encrypted Big Data,Other Big Data Security,Privacy on Correlated Big Data.

Text Book

1. Vignesh Prajapati ,“Big Data Analytics with R and Hadoop”, 2013 Packt Publishing.
2."Big Data Principles and Paradigms" Rajkumar Buyya

REFERENCES:

1. Arvind Sathi, "Big Data Analytics: Disruptive Technologies for Changing the Game", IBM Corporation, 2012.
2. Michael Minelli, Michehe Chambers, Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business", Wiley CIO Series, 2013.
3. Tom White, "Hadoop: The Definitive Guide", 3rd Edition, O'Reilly, 2012.
4. Kevin Roebuck, "Storing and Managing Big Data - NoSql, Hadoop and more: High-Impact Strategies - What You Need to Know", Tebbo, 2011.
5. Bill Franks, "Taming The Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics", 1st Edition, Wiley and SAS Business Series, 2012.

SOCIAL MEDIA DATA MINING & ANALYTICS	
Code: 23DS113 (C)	**Credits: 3**
Instruction : 3 Periods / Week	**Sessional Marks : 40**
End Exam : 3 Hours	**End Exam Marks : 60**

Prerequisites:

1. Python programming skill.
2. Data Mining

Course Objectives:

The objective of the course is to acquire and analyze the data from the social web.

Course Outcomes:

By the end of the course, the student will be able to:	
1.	Apply data mining techniques to social media using Python
2.	Perform Tokenization, Frequency analysis and Time series analysis.
3.	Visualize posts and measuring engagement.
4.	Process content from a user's post to extract interesting keywords
5.	Mine semantic relations from social media.

Mapping of course outcomes with program outcomes:

CO	PO			PSO	
	1	2	3	1	2
1	1	-	2	2	2
2	2	1	2	2	2
3	1	2	2	2	2
4	2	1	2	2	2
5	2	2	2	2	2

SYLLABUS

UNIT-I: Social Media, Social Data, and Python **10 Periods**

Social Media, Social Data, and Python
Getting started, Social media - challenges and opportunities, Opportunities, Challenges Social media mining techniques, Python tools for data science, Python development environment setup, pip and virtualenv, Conda, Anaconda, and Miniconda, Efficient data analysis, Machine learning, Natural language processing, Social network analysis, Data visualization, Processing data in Python, Building complex data pipelines.

UNIT-II: Data mining on Social Media **10 Periods**

Mining Twitter – Hashtags, Topics, and Time Series Getting started, The Twitter API, Rate limits, Search versus Stream, Collecting data from Twitter, Getting tweets from the timeline, The structure of a tweet, Using the Streaming API, Analyzing tweets - entity analysis, Analyzing tweets - text analysis, Analyzing tweets - time series analysis.

UNIT-III: User Interaction on Social Media **10 Periods**

Posts, Pages, and User Interactions on Facebook: The Facebook Graph API, Registering your app, Authentication and security, accessing the Facebook Graph API with Python, Mining your posts, The structure of a post, Time frequency analysis, Mining Facebook Pages, Getting posts from a Page, Facebook Reactions and the Graph API 2.6, Measuring engagement, Visualizing posts as a word cloud.

UNIT-IV: Analysis on Social Media **10 Periods**

Topic Analysis on Google+ : Getting started with the Google+ API, Searching on Google+, Embedding the search results in a web GUI, Decorators in Python, Flask routes and templates, Notes and activities from a Google+ page, Text analysis and TF-IDF on notes, Capturing phrases with n-grams.

UNIT-V: Linked Data and the Semantic Web **10 Periods**

Linked Data and the Semantic Web: A Web of Data, Semantic Web vocabulary, Micro formats, Linked Data and Open Data, Resource Description Framework, JSON-LD, Schema.org, Mining relations from DBpedia, Mining geo coordinates, Extracting geodata from Wikipedia, Plotting geodata on Google Maps.

Text Books:

1. Marco Bonzanini “Mastering Social Media Mining with Python”, Packt Publishing-2016

Reference Books:

1. Gabor Szabo Gungor Polatkan” Social Media Data Mining and Analytics”, John Wiley & Sons, Inc.,2019.
2. Reza Zafarani Mohammad Ali Abbasi” Social Media Mining”, Cambridge University Press, 2014

INFORMATION SECURITY	
Code: 23DS114(A)	**Credits: 3**
Instruction : 3 Periods /Week	**Sessional Marks : 40**
End Exam : 3 Hours	**End Exam Marks : 60**

Pre-Requisites:

1. **Basic knowledge of Computer Networks.**

Course Objectives:

1. Learn the fundamental concepts of Cryptography.
2. Acquire the knowledge about the applications of Cryptography.
3. Interpret the concepts of foot printing and reconnaissance with various tools.
4. Analyse the working of sniffing tools for gaining access to the captured network traffic.
5. Learn the significance of SQL Injection and its applications.

Course Outcomes:

By the end of the course, the student will be able to:

1.	Understand the fundamental concepts of Cryptography.
2.	Acquire the knowledge about the Cryptography algorithms.
3.	Memorize the concepts of foot printing and reconnaissance with various tools.
4.	Demonstrate the working of sniffing tools for gaining access to the captured network traffic.
5.	Understand the significance of SQL Injection and its applications.

CO-PO Mapping:

CO	PO			PSO	
	1	2	3	1	2
1	-	-	-	-	-
2	2	1	-	2	1
3	2	-	2	2	1
4	2	1	-	2	1
5	2	1	2	2	1

SYLLABUS

UNIT-1 **10 periods**

Security goals, attacks-passive and active attacks, services and mechanisms, techniques. Mathematics for cryptography: Integer Arithmetic, Modular Arithmetic, Linear Congruence. A model for network security, Internet Standards. Buffer overflow and format string vulnerabilities,

UNIT-2 **10 periods**

Symmetric Cryptography: Introduction, Substitution ciphers, Transposition ciphers, Feistel Structure, DES-AES-RC4.
Public Key Cryptography: Encryption/Decryption using RSA, RSA with example.
TCP Session hijacking, ARP Attacks, man-in-the-middle attacks, SQL injection, Phishing attacks,Tools-Ettercap, Burpsuite.

UNIT-3 **10 periods**

Gaining Access, Sniffers: System Hacking, password cracking, password cracking techniques, passive online attacks, active online attacks, offline attacks; Sniffers: Understanding Sniffers, Using a Sniffer, Sniffing Tools-Wireshark, TCP Dump, Reading Sniffer Output, Switched Network Sniffing, Detecting Sniffing Attacks.
(From Chapter 7 & 9 of Book 3)

UNIT-4 **10 periods**

SQL Injection: Introducing SQL Injection, Results of SQL Injection, the Anatomy of a Web Application, Databases and Their Vulnerabilities, Anatomy of a SQL Injection Attack, Altering Data with a SQL, Injection Attack, Injecting Blind, Information Gathering, Evading Detection Mechanisms, SQL Injection Countermeasures.
(Chapters 13 & 14 of Book 3)

UNIT-5 **10 periods**

System Intrusion Detection and Prevention: Basics, roles of IDS in network defence, IDS sensor placement, case study. IPS – basics, Limitations, NIPS, HIPS, Honey pots and Honey nets, Password protection, Password Selection Strategies, Malicious Programs, Types of viruses, worms, Trojan horses.

Text Books:

1. Network security Essentials, Applications and standards, 3e, William Stallings, PearsonEducation (for unit-1).
2. Cryptography and Network Security, Behrouz A. Forouzan, Tata McGraw Hill. (unit-2)
3. CEHv8 Certified Ethical Hacker Version 8 Study Guide, by Sean-Philip Oriyano, Sybex (A Wiley Brand) ISBN: 978-1-118-64767-7 (unit-3 and 4).
4. Cryptography and Network Security: Principles and Practice, William Stallings, 5th edition,Pearson.

Reference Books:

1. B. Menezes, Network security and Cryptography, Cengage Learning India, 2010.
2. Network Security Essentials (Applications and Standards) by William Stallings Pearson Education.
3. Principles of Information Security, Whitman, Thomson.

E-Resources :

1. https://www.nisc.go.jp/security-site/campaign/files/aj-sec/handbook-all_eng.pdf
2. https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-100.pdf

DATA ANALYTICS	
Code: 23DS114 (B)	**Credits: 3**
Instruction: 3 Periods / Week	**Sessional Marks: 40**
End Exam: 3 Hours	**End Exam Marks: 60**

Prerequisites:

1. Basics on Probability and statistics.
2. Fundamentals of Python programming.

Course Objectives:

1. To familiarize with basics data analytics and data analytics in Python.
2. Equip the students with core statistical models and visualization techniques to perform exploratory data analysis using Python.
3. Exploring the importance of analysis of variance and multivariate analysis of variance (MANOVA) and implementing them in Python with different kinds of data sets.

Course Outcomes:

By the end of the course, the student will be able to:

1.	Understand the basic principles of data analytics for performing basic data analysis on given data.
2.	Apply the data visualization methods in Python for exploratory data analysis.
3.	Understand and apply Simple Statistical Techniques for Univariate and Bivariate Analyses
4.	Understand and apply the nature and logic of the analysis of variance.
5.	Understand and apply linear and multi linear regression models.

CO-PO Mapping:

CO	PO			PSO	
	1	2	3	1	2
1	2	-	2	2	1
2	3	1	2	2	1
3	3	-	3	2	1
4	3	1	2	2	1
5	3	1	2	2	1

SYLLABUS

UNIT-1 **10 periods**

Introduction and Overview of Applied Statistics: How Statistical Inference Works, Statistics and Decision-Making, Data Analysis, Data Science, Machine Learning, Big Data. The Scatter plot, Correlograms, Histograms and Bar Graphs, Heatmaps, Line Charts.

Learning Outcome: At the end of this Unit the student will be able to

1. Understand basics of data analytics including its role in other sub domains.
2. Implement the few fundamental principles of data analytics using Python programming.

UNIT-2 **10 periods**

Simple Statistical Techniques for Univariate and Bivariate Analyses: Pearson Product-Moment Correlation, Computing Correlation in Python, T-Tests for Comparing Means, Paired-Samples t-Test in Python, Binomial Test, The Chi-Squared Distribution.

Learning Outcome: At the end of this Unit the student will be able to

1. Apply the statistical techniques to test the hypothesis.
2. Identify the relation between the significant attributes using correlation techniques.

UNIT-3 **10 periods**

Analysis of Variance (ANOVA): T-Tests for Means as a Special Case of ANOVA, Analysis of Variance (one-way classification), ANOVA in Python, Analysis of Variance (two-way classification).

Learning Outcome: At the end of this Unit the student will be able to

1. Understand the purpose of conducting analytical comparisons.
2. Determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups using ANOVA.

UNIT-4 **10 periods**

Simple and Multiple Linear Regression: Regression, Regression in Python, The Least-Squares Principle, The Population Least-Squares Regression Line, **Multiple Linear Regression**, How to Assess Goodness of Fit.

Learning Outcome: At the end of this Unit the student will be able to

1. Learn and appreciate why regression analysis is so central to visually all statistical models and how multilinear regression is different from linear regression.
2. Implement linear regression and multi linear regression models in Python.

UNIT-5 **10 periods**

Multivariate Analysis of Variance (MANOVA): Why Technically Most Univariate Models are Actually Multivariate, Multivariate Model and Running a Multivariate Model, Multivariate Tests of Significance: Why They Are Different from the F-Ratio. Performing MANOVA in Python.

Learning Outcome: At the end of this Unit the student will be able to

1. Learn and appreciate multivariate regression analysis is so central to visually all statistical models and how multilinear regression is different from linear regression.
2. Implement MANOVA in Python and interpret results.

Text Book:

1. Applied Univariate, Bivariate, and Multivariate Statistics Using Python, Daniel J. Denis, Wiley, First Edition.

Reference Books:

1. Applied Multivariate Statistical Analysis, Richard. A. Johnson and Dean.W. Wichern, Pearson Prentice Hall, 6th Edition, 2007.
2. An Introduction to Multivariate Statistical Analysis, T.W. Anderson, Wiley, 3rd Edition, 2003.

Web Resource:

1. https://www.westga.edu/academics/research/vrc/univariate-bivariate-analyses.php

MACHINE LEARNING	
Code: 23DS114(C)	**Credits: 3**
Instruction : 3 Periods / Week	**Sessional Marks : 40**
End Exam : 3 Hours	**End Exam Marks : 60**

Prerequisites:

1. Python programming skill.
2. Preliminary concepts of probability, algebra and statistics.

Course Objectives:

1. Understand the main categories and fundamental concepts of Machine Learning systems
2. Familiarize the main steps in a typical Machine Learning project
3. Effective utilization of regression, clustering and classification algorithms.
4. Ability to apply machine learning techniques to solve real world problems

Course Outcomes:

By the end of the course, the student will be able to:	
1.	Define the machine learning, types of ML and challenges of ML.
2.	Describe the extraction of data, withdraw insights of data and fine tune the data for applying ML model.
3.	Implementation of Classification models on MNIST Data set and Regression models.
4.	Apply SVM, DT and Ensemble Learning techniques for solving Real world problems.
5.	Solve curse of dimensionality problem and Apply K-Means, DBSCAN and Gaussian Mixtures to solve Unlabeled data.

Mapping of course outcomes with program outcomes:

CO	**PO**			**PSO**	
	1	**2**	**3**	**1**	**2**
1	2	-	-	1	1
2	3	1	2	3	3
3	3	1	3	3	3
4	3	1	3	3	3
5	3	-	3	3	3

SYLLABUS

UNIT-I: **8 Hours**

Machine learning landscape:
What Is Machine learning?, Why Use Machine Learning?, Types of Machine Learning Systems, Main Challenges of Machine Learning, Testing and Validating.
Learning Outcomes: At the end of this unit, the students will be able to

1. Explain ML and Category of ML.
2. Define Challenges of ML, Testing and validating ML model.

UNIT-II: **8 Hours**

End-to-End Machine learning Project:
Working with Real Data, Get the Data, Discover and Visualize the Data to Gain Insights, Prepare the Data for Machine Learning Algorithms, Select and Train a Model, Fine-Tune Your Model.
Learning Outcomes: At the end of this unit, the students will be able to

1. Explore the real data, extraction of relevant data and gain the insights of data.
2. Selection, Train a model and Fine-tuning the model for real data.

UNIT-III: **8 Hours**

Training Model and MNIST Data set Classification:
Linear Regression, Gradient Descent, Polynomial Regression, Logistic Regression.
MNIST, Training a Binary Classifier, Performance Measures, Multiclass Classification, Error Analysis, Multilabel Classification, Multioutput Classification.
Learning Outcomes: At the end of this unit, the students will be able to

1. Explore the Regression models and MNIST data set.
2. Implement Regression models and classifications on MNIST dataset.

UNIT-IV: **8 Hours**

SVM, Decision Trees and Ensemble Learning:
Linear SVM Classification, Nonlinear SVM Classification, SVM Regression.
Training and Visualizing a Decision Tree, Making Predictions, Estimating Class Probabilities, The CART Training Algorithm, Computational Complexity, Gini Impurity or Entropy?, Regularization Hyperparameters.
Voting Classifiers, Bagging and Pasting, Random Patches and Random Subspaces, Random Forests, Boosting, Stacking.
Learning Outcomes: At the end of this unit, the students will be able to

1. Describe the concept of SVM, DT, RF techniques.
2. Apply SVM, DT and RF on Real world data.

UNIT-V: **8 Hours**

Dimensionality Reduction and Clustering:

The Curse of Dimensionality, Main Approaches for Dimensionality Reduction, PCA, Kernel PCA, LLE, Other Dimensionality Reduction Techniques.
Clustering, K-Means, Clustering for image segmentation, Clustering for Pre-processing, Clustering for Semi-Supervised Learning, DBSCAN, Gaussian Mixtures.

Learning Outcomes: At the end of this unit, the students will be able to

1. Construct PCA and LLE.
2. Apply the K-Means, DBSCAN and Gaussian Mixtures on unlabeled data.

Text Books:

1. AurelionGeron, "Hands-on Machine Learning with Scikit-Learn, Keras, and Tensor Flow: Concepts, Tools and Techniques to build Intelligent Systems", 2/e, O'Reilly Media, 2019.
2. Sebastian Raschka, "Python Machine Learning" Packt Publishing 2015.

Reference Books:

1. Tom M. Mitchell, "Machine Learning" First Edition by Tata McGraw- Hill Education.
2. Ethem Alpaydin,"Introduction to Machine Learning " 2nd Edition, The MIT Press, 2009

Web Resources:

1. https://nptel.ac.in/courses/106106139
2. https://www.youtube.com/watch?v=PPLop4L2eGk

Research Methodology and IPR	
23DS115	**Credits: 2**
Instruction : 2 Periods/Week	**Sessional Marks : 40**
End Exam : 3 Hours	**End Exam Marks : 60**

Course Outcomes: At the end of this course, students will be able to

1. Understand research problem formulation.
2. Analyze research related information
3. Follow research ethics
4. Understand that today's world is controlled by Computer, Information Technology, but tomorrow world will be ruled by ideas, concept, and creativity.
5. Understanding that when IPR would take such important place in growth of individuals & nation, it is needless to emphasise the need of information about Intellectual Property Right to be promoted among students in general & engineering in particular.
6. Understand that IPR protection provides an incentive to inventors for further research work and investment in R & D, which leads to creation of new and better products, and in turn brings about, economic growth and social benefits.

SYLLABUS:

Unit-I: Meaning of research problem, Sources of research problem, Criteria Characteristics of a good research problem, Errors in selecting a research problem, Scope and objectives of research problem. Approaches of investigation of solutions for research problem, data collection, analysis, interpretation, Necessary instrumentations.

Unit-II: Effective literature studies approaches, analysis Plagiarism, Research ethics, Effective technical writing, how to write report, Paper Developing a Research Proposal, Format of research proposal, a presentation and assessment by a review committee.

Unit-III: Nature of Intellectual Property: Patents, Designs, Trade and Copyright. Process of Patenting and Development: technological research, innovation, patenting, development. International Scenario: International cooperation on Intellectual Property. Procedure for grants of patents, Patenting under PCT.

Unit-IV: Patent Rights: Scope of Patent Rights. Licensing and transfer of technology. Patent information and databases. Geographical Indications.

Unit-V: New Developments in IPR: Administration of Patent System. New developments in IPR; IPR of Biological Systems, Computer Software etc. Traditional knowledge Case Studies, IPR and IITs.

REFERENCES:

1. *Stuart Melville and Wayne Goddard, "Research Methodology: An Introduction for Science & Engineering Students'"*
2. *Wayne Goddard and Stuart Melville, "Research Methodology: An Introduction"*
3. *Ranjit Kumar, 2nd Edition, "Research Methodology: A Step by Step Guide for beginners"*
4. *Halbert, "Resisting Intellectual Property", Taylor & Francis Ltd ,2007.*
5. *Robert P. Merges, Peter S. Menell, Mark A. Lemley, "Intellectual Property in New Technological Age", 2016.*
6. *T. Ramappa, "Intellectual Property Rights Under WTO", S. Chand, 2008*

DATA SCIENCE USING PYTHON LAB	
Code: 23DS116	**Credits : 2**
Instruction : 3 Periods /Week	**Sessional Marks : 50**
End Exam : 3 Hours	**End Exam Marks : 50**

Prerequisites: Basic Programming Knowledge.

Course objectives:

1. To impart knowledge on data manipulation and exploratory data analysis concepts using Python.
2. Develop in depth understanding of the core packages available in Python for data analytics.

Course Outcomes:

By the end of the course, the student will be able to:	
1.	Demonstrate the approaches to perform mathematical computations vital for data science using Numpy.
2.	Apply Data Preparation and Exploration methods using Pandas for performing predictive analysis.
3.	Apply the packages available in Python to solve regression and social network analysis problems.
4.	Analyze social network analysis problems with packages available in Python.

Mapping of Course Outcomes with Program Outcomes:

CO	PO			PSO	
	1	2	3	1	2
1		1	1	2	2
2	1	1	1	2	2
3	2	1	2	2	2
4	2	1	2	2	2

List of Experiments

1. Implement basic Python programs for reading input from console.
2. Getting familiarity with Python IDE and Notebooks.
3. Write a Program in Python to Read and write different types of Files (csv, json, txt etc).
4. Perform Creation, indexing, slicing, concatenation and repetition operations on Python built-in data types: Strings, List, Tuples, Dictionary, Set
5. Apply Python built-in data types: Strings, List, Tuples, Dictionary, Set and their methods to solve any given problem
6. Write a Program in Python to perform statistical analysis on given Data set.
7. Apply basic statistical methods on Sample Datasets (Data sets from https://www.kaggle.com/datasets, https://archive.ics.uci.edu/ml/datasets.php etc.,)
8. Create packages and import modules from packages.
9. Handle Exceptions using Python Built-in Exceptions.
10. Create NumPy arrays from Python Data Structures, Intrinsic NumPy objects and Random Functions.
11. Import a CSV file and perform various Statistical and Comparison operations on rows/columns.
12. Load an image file and do crop and flip operation using NumPy Indexing
13. Create Pandas Series and DataFrame from various inputs.
14. Import any CSV file to Pandas DataFrame and perform the following:
    (a) Visualize the first and last 10 records
    (b) Get the shape, index and column details
    (c) Select/Delete the records(rows)/columns based on conditions.
    (d) Perform ranking and sorting operations.
    (e) Do required statistical operations on the given columns.
    (f) Find the count and uniqueness of the given categorical values.
    (g) Rename single/multiple columns.
15. Write a Program in Python to implement Correlation.
16. Write a Program in Python to perform Multiple Regression.
17. Write a Program in Python to predict House rent using linear regression.
18. Write a Program to apply Multiple Linear Regression Using Python and Scikit-learn.
19. Perform Social Network Analysis(SNA) and Graph Analysis using Python.
20. Write a python program for performing social network analysis using networkX package in python

Text Books:

1. Y. Daniel Liang, “Introduction to Programming using Python”, Pearson,2012.
2. Wes McKinney, “Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython”, O’Reilly, 2nd Edition,2018.
3. Jake VanderPlas, “Python Data Science Handbook: Essential Tools for Working with Data”, O’Reilly, 2017.

Reference Books:

1. https://www.programmer-books.com/introducing-data-science-pdf/
2. https://www.cs.uky.edu/~keen/115/Haltermanpythonbook.pdf
3.http://math.ecnu.edu.cn/~lfzhou/seminar/[Joel_Grus]_Data_Science_from_Scratch_First_Princ.pdf

Web Resource:

1. https://www.datacamp.com/tutorial/social-network-analysis-python
2. https://www.geeksforgeeks.org/introduction-to-social-networks-using-networkx-in-python/
3. https://www.kirenz.com/post/2019-08-13-network_analysis/

INFORMATION SECURITY LAB	
Code: 23DS117(A)	**Credits :2**
Instruction : 3 Periods /Week	**Sessional Marks : 50**
End Exam : 3 Hours	**End Exam Marks : 50**

Pre-Requisites:

1. Basic knowledge of Cryptography
2. Basic knowledge of Packet Capturing
3. Exposure to Problem solving techniques and programming skills

Course Objectives:

1 Introducing different tools related to Network Security.
2 Introducing tools related to Information Security.

By the end of the course, the student will be able to:

1.	Demonstrate the buffer overflow and format string attacks.
2.	Analyse the packets captured using Wireshark.
3.	Demonstrate the tools nmap and IPtables for network security.
4.	Detect and implement SQL Injection attacks.

CO	PO			PSO	
	1	2	3	1	2
1	1	1	1	2	2
2	2	1	1	1	1
3	1	2	1	2	1
4	1	1	1	1	2

EXPERIMENTS

1. Understanding the buffer overflow and format string attacks. CO-1
2. Demonstrate the transposition and substitution ciphers. CO-1
3. To analyse the network packets using Wireshark. CO-2
4. Sniffing networks and analysis of TCP/IP using Wireshark . CO-2
5. Implement Testing for SQL Injection and detect it. CO-4
6. To perform the web penetration testing using BURPSUITE. CO-4
7. Use 'nmap' tool to perform vertical and horizontal scanning for
8. checking open and closed ports. CO-3
9. Configuring IP Tables in Linux and setting the filtering rules. CO-3

Case Study:

1. Use nmap commands for performing the following experiments: CO-3

a) Use ping sweeping to determine which hosts are running.
b) Check for vulnerable services available using TCP connect scans.
c) Perform OS Fingerprinting to determine the OS of target machine.
d) Choose different options under each category according to your creativity.

TEXTBOOKS:

1)The Complete Reference Network Security By Robert Bragg, Mark Rhodes Ousley, KeithStrassberg, 1st Edition, McGraw Hill India (2004)Publication
2) The Unofficial Guide to Ethical Hacking by Ankit Fadia, Second edition(2006), LaxmiPublications.
3) Network Security Tools Writing, Hacking, and Modifying Security Tools by Nitesh Dhanjani,Justin Clarke, 2013 Edition, Publisher: O'Reilly Media.
4) Linux and UNIX Security Portable Reference Book by Nitesh Dhanjani,1st Edition, McGrawHill.

REFERENCE BOOKS:

1. Network Security Tools Writing, Hacking, and Modifying Security Tools ByNiteshDhanjani, Justin Clarke,2nd Edition, Publisher: O'Reilly Media

WEB RESOURCES:

1. https://www.udemy.com/courses/it-and-software/network-and-security
2. https://online.stanford.edu/course/network-security
3. http://index-of.es/EBooks/SQLInjectionAttacksandDefense.pdf

DATA ANALYTICS LAB	
Code: 23DS117(B)	**Credits: 2**
Instruction : 3 Periods /Week	**Sessional Marks : 50**
End Exam : 3 Hours	**End Exam Marks : 50**

Prerequisites:

- Basic knowledge on Probability and statistics.
- Basics of Python Programming.

Course Objectives:

- To train the students to apply the principles of data analytics to analyze and effectively visualize the data.
- Train the students to gain the knowledge of computational statistical approaches and their application to a variety of datasets.
- Practical way of Understanding the nature of analysis of variance and Multivariate statistical models.

Course Outcomes:

By the end of the course, the student will be able to:	
1.	Work with different IDE and Python notebooks for mathematical, scientific applications and for data analysis.
2.	Explore Pandas, matplotlib and seaborn packages. Write example programs to visualize the data.
3.	Work with Linear Regression and Multiple Regression models in Python
4.	Understanding of the ANOVA that can used to test a given hypothesis and to verify the experimental results are significant.
5.	Work with multivariate statistical methods in Python and interpret the results.

CO-PO MAPPING

CO	PO			PSO	
	1	2	3	1	2
1	2	1	2	1	1
2	3	1	3	-	1
3	2	1	2	2	-
4	3	1	3	1	1

List of Experiments

1. Write a Program in Python to Manipulate, Aggregate and Analyze data using Numpy.
2. Write a Program in Python to Handle and Analyze data using Pandas.
3. Apply basic statistical methods on Sample Datasets like Mushroom (Data sets from https://www.kaggle.com/datasets, https://archive.ics.uci.edu/ml/datasets.php etc.,)
4. Working with matplotlib and seaborn packages in Python.
5. Write a Program in Python to add an indeed field, changing misleading data fields, Re-expressing categorical data as numerical data, standardizing numerical fields and identifying outliers for data preparation phase. for bank marketing data set.
6. (https://www.kaggle.com/datasets/janiobachmann/bank-marketing-dataset)
7. Write a Program in Python to Classifying MNIST digits using Logistic Regression
8. Write a Program in Python to predict House rent using linear regression.
9. Write a Program in Python to implement Correlation.
10. Write a Program to apply Multivariate analysis of variance and co-variance (MANOVA).
11. Write a Program to apply Analysis of Variance (one-way classification) in Python.
12. Write a Program to apply Analysis of Variance (two-way classification) in Python.
13. Write a Program to apply Linear Discriminant analysis for multivariate data.
14. Write a Program to apply Principle component analysis for multivariate data.
15. Implement Factor Analysis for multivariate data in Python.

Text Books:

1. Applied Univariate, Bivariate, and Multivariate Statistics Using Python, Daniel J. Denis, Wiley, First Edition.

Reference Books:

1. Applied Multivariate Statistical Analysis, Richard. A. Johnson and Dean.W. Wichern, Pearson Prentice Hall, 6th Edition, 2007.
2. An Introduction to Multivariate Statistical Analysis, T.W. Anderson, Wiley, 3rd Edition, 2003.

Web Resource:

- https://www.westga.edu/academics/research/vrc/univariate-bivariate-analyses.php

MACHINE LEARNING LAB	
CODE: 23DS117(C)	**Credits : 2**
Instruction : 3 Periods /Week	**Sessional Marks : 50**
End Exam : 3 Hours	**End Exam Marks : 50**

Prerequisites: C and data Structures

Course objectives:

1. Learn to manipulate data to apply ML algorithms using Numpy, Pandas, Matplotlib and SciKit Learn
2. Solve real world problems using ML Techniques.

Course Outcomes:

By the end of the course, the student will be able to:	
1.	Manipulate basic functions in Numpy, Pandas, Matplotlib SciKit Learn
2.	Apply the concepts of Regression (Linear, Quadratic and Logistic) to predict dependent variables.
3.	Feature recognition and classification using SVM, DT and RF techniques.
4.	Testing & validating ML models and evaluation of accuracy measures.

Mapping of Course Outcomes with Program Outcomes:

CO	**PO**			**PSO**	
	1	**2**	**3**	**1**	**2**
1	2	2	2	3	3
2	2	2	2	3	3
3	2	2	2	3	3
4	2	2	2	3	3

List of Experiments

1. Implement the matrices operations using both Numpy and pandas
2. Using matplotlib perform data visualization on the standard dataset
3. Implement Linear Regression using ordinary least square(OLS) and Gradient Descent methods
4. Implement quadratic Regression
5. Implement Logistic Regression
6. Evaluate performance measures on regression models (Linear, quadratic and Logistic).
7. Implement classification using SVM
8. Implement Decision Tree learning
9. Implement Bagging using Random Forests
10. Implement K-means Clustering to Find Natural Patterns in Data
11. Implement DBSCAN clustering
12. Implement Gaussian Mixture Model
13. Solve the curse of dimensionality by implementing PCA algorithm on a high dimensional
14. Comparison of Machine Learning algorithms

Text Books:

1. AurelionGeron, "Hands-on Machine Learning with Scikit-Learn, Keras, and Tensor Flow: Concepts, Tools and Techniques to build Intelligent Systems", 2/e, O'Reilly Media, 2019.
2. Sebastian Raschka, "Python Machine Learning" Packt Publishing 2015.

Reference Books:

1. Tom M. Mitchell, "Machine Learning" First Edition by Tata McGraw- Hill Education.
2. Ethem Alpaydin,"Introduction to Machine Learning " 2nd Edition, The MIT Press, 2009

Web Resources:

1. https://nptel.ac.in/courses/106106139
2. https://www.youtube.com/watch?v=PPLop4L2eGk

Universal Human Values and Professional Ethics	
Code:23DS118	**Credits: 0**
Instruction : 2 Periods/Week	**Sessional Marks : 50**
End Exam : --	**End Exam Marks : --**

Course Objectives:

1. To create an awareness on Engineering Ethics and Human Values.
2. To instill Moral and Social Values and Loyalty
3. To appreciate the rights of others.
4. To create awareness on assessment of safety and risk

Course outcomes: By the end of the semester, the student will be able to:

CO1	Identify and analyze an ethical issue in the subject matter under investigation or in a relevant field
CO2	Identify the multiple ethical interests at stake in a real-world situation or practice
CO3	Articulate what makes a particular course of action ethically defensible
CO4	Assess their own ethical values and the social context of problems
CO5	Identify ethical concerns in research and intellectual contexts, including academic integrity, use and citation of sources, the objective presentation of data, and the treatment of human subjects
CO6	Demonstrate knowledge of ethical values in non-classroom activities, such as service learning, internships, and field work integrate, synthesize, and apply knowledge of ethical dilemmas and resolutions in academic settings, including focused and interdisciplinary research

SYLLABUS

Unit I: HUMAN VALUES

Morals, Values and Ethics-Integrity-Work Ethic-Service learning – Civic Virtue – Respect for others – Living Peacefully –Caring –Sharing –Honesty -Courage-Cooperation–Commitment – Empathy –Self Confidence -Character –Spirituality-Case Study.

LEARNING OUTCOMES:

1. learn about morals, values & work ethics.
2. learn to respect others and develop civic virtue.
3. develop commitment
4. learn how to live peacefully

Unit II: ENGINEERING ETHICS

Senses of 'Engineering Ethics-Variety of moral issues –Types of inquiry –Moral dilemmas –Moral autonomy –Kohlberg's theory-Gilligan's theory-Consensus and controversy –Models of professional roles, Theories about right action-Self-interest -Customs and religion –Uses of Ethical theories – Valuing time –Case Study.

LEARNING OUTCOMES:

1. learn about the ethical responsibilities of the engineers.
2. create awareness about the customs and religions.
3. learn time management
4. learn about the different professional roles.

Unit III: ENGINEERING AS SOCIAL EXPERIMENTATION

Engineering as Social Experimentation –Similarities between standard and social Experimentations-Framing the problem –Determining the facts –Codes of Ethics – Clarifying Concepts –Application issues –Common Ground -General Principles –Utilitarian thinking respect for persons-Case study.

LEARNING OUTCOMES:

1. Demonstrate knowledge to become a social experimenter.
2. Provide depth knowledge on framing of the problem and determining the facts.
3. Provide depth knowledge on codes of ethics.
4. Develop utilitarian thinking

UNIT IV: ENGINEERS RESPONSIBILITY FOR SAFETY AND RISK

Safety and risk –Assessment of safety and risk –Risk benefit analysis and reducing risk-Safety and the Engineer-Designing for the safety-Intellectual Property rights(IPR).

LEARNING OUTCOMES:

1. create awareness about safety, risk & risk benefit analysis.
2. engineer's design practices for providing safety.
3. provide knowledge on Intellectual Property Rights.

UINIT V: GLOBAL ISSUES

Globalization –Cross culture issues-Environmental Ethics –Computer Ethics –Computers as the instrument of Unethical behavior –Computers as the object of Unethical acts –Autonomous Computers-Computer codes of Ethics –Weapons Development -Ethics and Research –Analyzing Ethical Problems in research- Case Study.

LEARNING OUTCOMES:

1. Develop knowledge about global issues.
2. Create awareness on computer and environmental ethics.
3. Analyze ethical problems in research.
4. Give a picture on weapons development.

Text Books:

1. *M.Govindarajan, S.Natarajananad, V.S.SenthilKumar "Engineering Ethics includes Human Values" -PHI Learning Pvt. Ltd-2009*
2. *Harris, Pritchard and Rabins "Engineering Ethics", CENGAGE Learning, India Edition, 2009.*
3. *Mike W. Martin and Roland Schinzinger "Ethics in Engineering" Tata McGraw-Hill–2003.*
4. *Prof.A.R.Aryasri, DharanikotaSuyodhana "Professional Ethics and Morals" Maruthi Publications.*
5. *A.Alavudeen, R.KalilRahman and M.Jayakumaran "Professional Ethics and Human Values" – Laxmi Publications.*
6. *Prof.D.R.Kiran "Professional Ethics and Human Values"*
7. *PSR Murthy "Indian Culture, Values and Professional Ethics" BS Publication*

SEM-II

COMPUTATIONAL INTELLIGENCE	
Code:23DS121	**Credits: 3**
Instruction : 3 Periods /Week	**Sessional Marks : 40**
End Exam : 3 Hours	**End Exam Marks : 60**

Prerequisites:

1. Some programming skills and a good background in Statistics, Probability, and discrete mathematics probability will be very helpful.

Course Objectives:

1 Students will realize the importance and challenges.
2 Student will learn the basic concepts of fuzzy
3 Student can acquire the knowledge on different algorithms.
4 Student will able to choose optimization algorithm for different problems.

Course Outcomes:

By the end of the course, the student will be able to:	
1.	Discuss the paradigm, approaches and applications with challenges
2.	Explore fuzzy logic and its mechanisms.
3.	Apply the algorithms and design strategies to solve problems
4.	Analyze the Evolutionary Computation algorithms in different domains
5.	Illustrate the collaborative and agglomerative combinations.

Mapping of Course Outcomes with Program Outcomes:

CO	**PO**			**PSO**	
	1	**2**	**3**	**1**	**2**
1	2	-	2	1	1
2	3	1	3	2	2
3	3	1	3	2	2
4	3	1	3	2	2
5	3	1	2	2	1

SYLLABUS

UNIT- I **12 Periods**

Introduction to Computational Intelligence:
Computational Intelligence, paradigms of Computational Intelligence, Approaches to computational intelligence, Synergies of Computational Intelligence Techniques, Applications of Computational Intelligence, and grand challenges of Computational Intelligence.

Learning Outcomes:

1. Discuss the approaches and computational intelligence Techniques.
2. Analyze the challenges that employ in these paradigms.

UNIT-II: **14 Periods**

Introduction to fuzzy logic: Fuzzy logic, fuzzy sets, member functions, Features of MFs, Operations of Fuzzy sets, Linguistic Variables, Fuzzy Relations, Fuzzy If-Then Rules, Fuzzification, Defuzzification, and Inference Mechanism.

Learning Outcomes:

1. Describe the utilization of fuzzy logic and relations
2. Apply fuzzy logic that employs this paradigm.

UNIT-III: **12 Periods**

Fuzzy Systems and Applications:
Fuzzy system, Fuzzy Modelling, Fuzzy control, Design of Fuzzy Controller, Modular Fuzzy Controller.

Neural Network Systems and Applications:
Learning Neural Networks, Recurrent Neural Networks, System Identification and Control, neural networks for control

Learning Outcomes:

1. Describe the design of controllers, modular with respective applications.
2. Identifying the system controls in neural networks.

UNIT-IV: **12 Periods**

Evolutionary Computation: Evolutionary Computation, Terminologies of Evolutionary computing, Genetic Operators, Performance Measure of EA, Evolutionary Algorithms

Evolutionary Systems:
Multi-objective Optimization, Co-evolution, parallel Evolutionary Algorithms

Learning Outcomes:

1. Discuss the evolutionary computations and explain when an algorithmic performance
2. Apply algorithms that employ this paradigm and parallel evolutionary systems

UNIT-V: **12 Periods**

Evolutionary Fuzzy Systems:

Evolutionary Adaptive Fuzzy Systems, Objective Functions and evaluation, and Fuzzy Adaptive Evolutionary Algorithms

Evolutionary Neural Network:

Support Combinations, Collaborative combinations, amalgamated combinations.

Learning Outcomes:

1. Explain adaptive and evolutionary algorithms and to which situations it applies.
2. Perform collaborative combinations.

Text Books:

1. Nazmul Siddique, Hojjat Adeli, Computational intelligence, 2013, Wiley.
2. Rudolf Kruse, Christian Borgelt, Frank Klawonn, Christian Moewes, Matthias Steinbrecher, Pascal Held, Computational Intelligence A Methodological Introduction,2013, Springer.

Reference Books:

1.Amit Konar, Computational Intelligence Principles, Techniques and Applications, 2006, Springer

DATA VISUALIZATION	
Code:23DS122	**Credits: 3**
Instruction: 3 Periods	**Sessional Marks: 40**
End Exam: 3 Hours	**End Exam Marks: 60**

Prerequisites:

Basic knowledge of data analysis and statistical concepts is essential for understanding the principles of data visualization covered in this course.

Familiarity with programming and data manipulation using Python or R will be beneficial as the course may involve hands-on exercises and data visualization implementations.

Course Objectives:

1. Master data mapping to aesthetics and utilize coordinate systems and color scales for effective data visualizations.
2. Create various visualizations for amounts, distributions, proportions, and x-y relationships.
3. Handle uncertainty in visual representations, explore highly skewed distributions, and communicate effectively using informative titles, captions, and tables in data visualizations.

Course Outcomes:

By the end of the course, the student will be able to:	
1.	Master mapping data to aesthetics and utilize coordinate systems and color scales to create effective data visualizations.
2.	Create visualizations for various data types and demonstrate proficiency in handling geospatial data and uncertainty in visual representations.
3.	Create effective visualizations of single distributions and explore highly skewed distributions using empirical cumulative distribution functions and quantile-quantile plots.
4.	Create effective scatterplots, correlograms, and visualizations for time series data and trends using smoothing and functional forms.
5.	Create informative titles, captions, and tables, and balance data and context in visualizations while effectively storytelling with figures.

Mapping of Course Outcomes with Program Outcomes:

CO	PO			PSO	
	1	2	3	1	2
1	1	1	-	-	2
2	2	-	3	2	2
3	2	1	3	2	3
4	2	1	3	3	3
5	2	1	3	3	3

SYLLABUS

UNIT I: **10 Periods**

Visualizing Data: Mapping Data onto Aesthetics: Aesthetics and Types of Data, Scales Map Data Values onto Aesthetics **Coordinate Systems and Axes:** Cartesian Coordinates, Nonlinear Axes, Coordinate Systems with Curved Axes **Color Scales:** Color as a Tool to Distinguish, Color to Represent Data Values, Color as a Tool to Highlight

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

1. Master mapping data to aesthetics for effective data visualizations.
2. Utilize coordinate systems and color scales to enhance data representation in visualizations.

UNIT II: **10 Periods**

Directory of Visualizations: Amounts, Distributions, Proportions, x–y relationships, Geospatial Data, Uncertainty **Visualizing Amounts:** Bar Plots, Grouped and Stacked Bars, Dot Plots and Heatmaps

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

1. Create effective visualizations for amounts, distributions, proportions, and x-y relationships using bar plots, dot plots, and heatmaps.
2. Demonstrate proficiency in visualizing geospatial data and handling uncertainty in visual representations.

UNIT III: **10 Periods**

Visualizing Distributions: Histograms and Density Plots: Visualizing a Single Distribution, Visualizing Multiple Distributions at the Same Time **Empirical Cumulative Distribution Functions and Q-Q Plots:** Empirical Cumulative Distribution Functions, Highly Skewed Distributions, Quantile-Quantile Plots

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

1. Create effective visualizations of single distributions using histograms and density plots.
2. Demonstrate proficiency in constructing empirical cumulative distribution functions and quantile-quantile plots for exploring highly skewed distributions and assessing data normality.

UNIT IV: **10 Periods**

Visualizing Associations Among Two or More Quantitative Variables: Scatterplots, Correlograms, Dimension Reduction, Paired Data. **Visualizing Time Series and Other Functions of an Independent Variable:** Individual Time Series, Multiple Time Series and Dose–Response Curves, Time Series of Two or More Response Variables **Visualizing Trends:** Smoothing, Showing Trends with a Defined Functional Form, Detrending and Time-Series Decomposition.

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

1. Create effective scatterplots and correlograms to visualize associations among quantitative variables.
2. Demonstrate proficiency in visualizing time series data, dose-response curves, and trends using smoothing and functional forms.

UNIT V: **10 Periods**

Titles, Captions, and Tables: Figure Titles and Captions, Axis and Legend Titles, Tables. **Balance the Data and the Context:** Providing the Appropriate Amount of Context, Background Grids, Paired Data. **Telling a Story and Making a Point:** What Is a Story? Make a Figure for the Generals, Build Up Toward Complex Figures, Make Your Figures Memorable, Be Consistent but Don't Be Repetitive.

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

1. Create informative titles and captions for figures and axis/legend titles, ensuring effective communication of key messages.
2. Design clear and organized tables to present data concisely and enhance data interpretation.

Textbooks:

1. Claus O. Wilke "Fundamentals of Data Visualization A Primer on Making Informative and Compelling Figures ",2021, 1st Edition, O'Reilly Media,USA

Reference Books:

1. Jonathan Schwabish " Better Data Visualizations, A Guide for Scholars, Researchers, and Wonks" 2018, 1st Edition, Columbia University Press, New York

Web Resources:

1. https://web.stanford.edu/class/cs102/

STREAM ANALYTICS	
Code:23DS123(A)	**Credits: 3**
Instruction : 3 Periods	**Sessional Marks : 40**
End Exam : 3 Hours	**End Exam Marks : 60**

Prerequisites:

1. Some programming skills and a good back ground in Query languages-SQL, statistics and probability will be very helpful.

Course Objectives:

1 Student will understand the fundamentals of Stream processing
2 Student will learn the Streaming Analytics Architecture
3 Student can acquire the knowledge how to store streaming data.
4 Student will able to learn statistical approximation of streaming data.

Course Outcomes:

By the end of the course, the student will be able to:	
1.	Discuss the Streaming Analytics Architecture
2.	Explore Data-Flow Management in Streaming Analysis.
3.	Analyze Distributed Streaming Data Processing
4.	Apply Streaming on Web Applications and Mobile Applications.
5.	Apply Statistical Approximation of Streaming Data

Mapping of Course Outcomes with Program Outcomes:

CO	PO			PSO	
	1	2	3	1	2
1	-	1	-	2	2
2	2	1	3	3	3
3	2	1	2	2	3
4	2	-	3	3	3
5	2	1	3	2	3

SYLLABUS

UNIT- I **12 Periods**

Introduction to Streaming Data: Sources of Streaming Data, Why Streaming Data Is Different, Infrastructures and Algorithms
Streaming Analytics Architecture: Designing Real-Time Streaming Architectures, Real-Time Architecture Components, Features of a Real-Time Architecture, Languages for Real-Time Programming, A Real-Time Architecture Checklist.

Learning Outcomes:

1. Discuss the importance of streaming data Infrastructures and Algorithms.
2. Describe Real-Time Streaming Architectures and its components.

UNIT-II: **14 Periods**

Service Configuration and Coordination: Motivation for Configuration and Coordination Systems, Maintaining Distributed State, Apache Zookeeper.
Data-Flow Management in Streaming Analysis: Distributed Data Flows, Apache Kafka: High-Throughput Distributed Messaging, Apache Flume: Distributed Log Collection.

Learning Outcomes:

1. Discuss Apache Zookeeper.
2. Discuss Data-Flow Management in Streaming Analysis**.**

UNIT-III:
Processing Streaming Data: Distributed Streaming Data Processing, Processing Data with Storm, Processing Data with Samza
Storing Streaming Data: Consistent Hashing, "NoSQL" Storage Systems, Other Storage Technologies, choosing a Technology, Warehousing

12 Periods

Learning Outcomes:

1. Describe Distributed Streaming Data Processing.
2. Discuss **Storing Streaming Data with**. "NoSQL" Storage Systems

UNIT-IV: **12 Periods**

Delivering Streaming Metrics: Streaming Web Applications, Visualizing Data, Mobile Streaming Applications.
Exact Aggregation and Delivery: Timed Counting and Summation, Multi-Resolution Time-Series Aggregation, Stochastic Optimization, Delivering Time-Series Data.
Learning Outcomes:

1. Discuss Streaming Web Applications and Mobile Streaming Applications.
2. Discuss Multi-Resolution Time-Series Aggregation.

UNIT-V: **12 Periods**

Statistical Approximation of Streaming Data: Numerical Libraries, Probabilities and Distributions, Working with Distributions, Random Number Generation, Sampling Procedures.
Approximating Streaming Data with Sketching: Registers and Hash Functions, Working with Sets, The Bloom Filter, Distinct Value Sketches, The Count-Min Sketch, Other Applications.
Beyond Aggregation: Models for Real-Time Data, Forecasting with Models, Monitoring, Real-Time Optimization

Learning Outcomes:

1. Apply Statistical Approximation of Streaming Data.
2. Discuss Approximating Streaming Data with Sketching

Text Books:

1. Bryon Ellis, "Real-Time Analytics: Techniques to Analyze and Visualize Streaming Data", Wiley, 1st Edition, 2014.
2. Anindita Basak, Krishna Venkataraman, Ryan Murphy, Manpreet Singh, "Stream Analytics with Microsoft Azure", Packt Publishing, December 2017.

Reference Books :

1. Henrique C. M. Andrade, Buğra Gedik and Deepak S. Turaga, "Fundamentals of Stream Processing: Application Design, Systems, and Analytics", Cambridge University Press, 2014.
2. Gerard Maas and Francois Garillot , "Stream Processing with Apache Spark: Mastering Structured Streaming and Spark Streaming", O'Reilly, 2019.

Prepared By

Mrs. B. Sailaja, Assistant Prof, Dept of CSM

IMAGE PROCESSING	
Code:23 DS123 (B)	**Credits:3**
Instruction: 3 Periods	**Sessional Marks: 40**
Final Exam: 3 Hours	**End Exam Marks: 60**

Pre requisites: Knowledge of linear algebra, basic probability and statistics, knowledge of basic programming language.

Course Objectives:

1. To understand the fundamentals of image processing
2. To explain various image restoration and enhancement techniques
3. To understand image segmentation techniques

Course Outcomes:

By the end of the course, the student will be able to:	
1.	Describe the fundamentals of image processing
2.	Apply the image enhancement techniques
3.	Describe image restoration procedures.
4.	Categorize various compression techniques
5.	Apply the image segmentation techniques

Mapping of Course Outcomes with Program Outcomes:

CO	PO			PSO	
	1	2	3	1	2
1	1	-	2	1	1
2	2	-	2	2	2
3	2	1	2	2	2
4	1	-	2	2	2
5	2	1	2	2	2

SYLLABUS

UNIT-I **10 Periods**

INTRODUCTION: Examples of fields that use digital image processing, fundamental steps in digital image processing,components of image processing system.

DIGITAL IMAGE FUNDAMENTALS: Image sensing and acquisition, image sampling and quantization, some basic relationships between pixels, Introduction to the basic mathematical tools used in digital image processing

UNIT- II **10 Periods**

INTENSITY TRANSFORMATIONS AND SPATIALFILTERING: Some basic intensity transformation functions, histogram processing, fundamentals of spatial filtering,basic smoothing and sharpening spatial filters.

UNIT-III **10 Periods**

IMAGE RESTORATION: A model of the image degradation/ restoration process, noise models, restoration in the presence of noise– only spatial filtering, periodic noise reduction using frequency domain filtering.

UNIT-IV **10 Periods**

IMAGE COMPRESSION: Fundamentals, Huffman coding, Golomb coding, arithmetic coding, LZW coding, run-length coding,symbol-based coding, bit-plane coding

UNIT-V **10 Periods**

IMAGE SEGMENTATION: Fundamentals, Point, Line, And Edge Detection, Thresholding, Segmentation By Region Growing and by Region Splitting And Merging, Region Segmentation using Clustering and Super pixels, Region Segmentation using Graph Cuts.

Text Books:

RafealC.Gonzalez, Richard E.Woods, Digital *Image Processing*, 3rd Edition, Pearson Education/PHI, 2017

Reference Books:

1. Milan Sonka, Vaclav Hlavac and Roger Boyle, “Image Processing, Analysis, andMachineVision”, 2nd Edition, Thomson Learning, 2010.
2. Alasdair McAndrew, “Introduction to Digital Image Processing with Matlab”,1stEdition,Thomson Course Technology,2010.

BUSINESS INTELLIGENCE	
Code: 23DS123(C)	**Credits: 3**
Instruction : 3 Periods/Week	**Sessional Marks: 40**
End Exam : 3 Hours	**End Exam Marks: 60**

Prerequisites:

1. Data Mining

Course Objectives:

1. Develop solid understanding of the key concepts, principles and components of Business Intelligence, Business Analytics and Big Data Analytics
2. Demonstrate comprehensive understanding of the decision making process.
3. Familiarize with predictive modeling techniques
4. Understand multi-criteria decision making systems
5. Familiarize with automated decision systems.

Course Outcomes:

By the end of the course, the student will be able to:	
1	Leverage knowledge of Business Intelligence framework, Business Analytics and Big Data analytics to analyze and interpret data for informed decision making. (Level 2)
2	Explore the phases, capabilities, classifications, and components of decision making.
3	Apply predictive modeling techniques and employ text analytics methods for sentiment analysis.
4	Apply prescriptive analytics techniques to optimize complex decision making scenarios.
5	Evaluate the impact of emerging trends on organizational decision making processes and demonstrate a comprehensive understanding of expert systems.

Mapping of Course Outcomes with Program Outcomes:

CO	PO			PSO	
	1	2	3	1	2
1	1	-	2	2	2
2	2	1	2	2	3
3	2	1	3	2	3
4	2	1	3	3	3
5	-	1	3	3	3

SYLLABUS

UNIT- I **10 Periods**

Decision Making and Analytics: An Overview: Information systems support for decision making, An early framework for computerized decision support, The concept of Decision Support System (DSS), A framework for Business Intelligence (BI), Business analytics overview, Brief Introduction to Big Data analytics.

Learning Outcomes:

1. Understands the importance, foundations, and applications of business analytics for managerial decision-making.
2. Learn the major framework of computerized decision support: analytics

UNIT-II **10 Periods**

Foundations and Technologies for Decision Making: Decision making: Introduction and definitions, Phases of the decision-making process, Decision making- The intelligent phase, Decision making-The design phase, Decision making- The choice phase, Decision making- The implementation phase.

Learning Outcomes:

1. Analyze the phases of the decision making process and apply appropriate decision making strategies and techniques at each phase.
2. Evaluate the capabilities and classifications of Decision Support Systems (DSS).

UNIT-III **10 Periods**

Techniques for Predictive Modeling: Basic concepts of neural networks, Developing neural-network-based systems, Illuminating the black box of ANN with sensitivity, Support vector machines, A process-based approach to the use of SVM.

Learning Outcomes:

1. Apply predictive modeling techniques to analyze and interpret complex datasets.
2. Employ text analytics techniques, including sentiment analysis and text mining to extract meaningful insights from data.

UNIT-IV **10 Periods**

Prescriptive Analytics: Decision support systems modeling, Structure of mathematical models for decision support, Certainty, uncertainty, and risk, Decision modeling with spreadsheets, Mathematical programming optimization, Multiple goals, sensitivity analysis, What-if analysis.

Learning Outcomes:

1. Apply modeling techniques to analyze and solve complex decision making problems, use sensitivity analysis and what-if analysis to evaluate the impact of different scenarios.
2. Evaluate and compare decision making approaches.

UNIT-V **10 Periods**

Automated Decision Systems and Expert Systems: The artificial intelligence field, Basic concepts of expert systems, Applications of expert systems, Structure of expert systems, Knowledge engineering, Problem areas suitable for expert systems, Development of expert systems.

Learning Outcomes:

1. Understand expert systems, and apply knowledge engineering techniques to developeffective expert systems
2. Analyze and evaluate emerging trends in business analytics.

Text Books:

1. Business Intelligence and Analytics: Systems for Decision Support, Ramesh Sharda,Efraim Turban, Dursun Delen, 10th Edition, Pearson Global Edition, 2015

References:

1. S. Christian Albright, Wayne L. Winston, Business Analytics: Data Analysis & DecisionMaking, 6thEdition, CENGAGE INDIA , 2017
2. Dinabandhu Bag, Business Analytics, Routledge, 1st edition, 2016
3. Rick Sherman, Business Intelligence Guidebook: From Data Integration to Analytics,Morgan Kaufmann, 1st edition 2014

PREDICTIVE ANALYTICS & DATA MINING	
Code:23DS124(A)	**Credits: 3**
Instruction : 3 Periods	**Sessional Marks : 40**
End Exam : 3 Hours	**End Exam Marks : 60**

Prerequisites:

1. Some programming skills and a good back ground in discrete mathematics, data structures and probability will be very helpful.

Course Objectives:

1 Student will understand the basic concepts data mining and Predictive analysis
2 Student will learn the different algorithm for predictive analysis.
3 Student can acquire the knowledge to solve the complexities of different problems.
4 Student will able to choose appropriate algorithm for different problems.

Course Outcomes:

By the end of the course, the student will be able to:	
1.	Discuss the concepts of Predictive analytics
2.	Explore different classification algorithms
3.	Apply the algorithms and perform optimal decisions
4.	Analyze the complexities of various problems in different domains
5.	Illustrate the real-time applications in reinforcement learning.

Mapping of Course Outcomes with Program Outcomes:

CO	**PO**			**PSO**	
	1	**2**	**3**	**1**	**2**
1	-	-	-	-	-
2	2	1	-	2	1
3	2	-	2	2	1
4	2	1	-	2	1
5	2	1	2	2	1

SYLLABUS

UNIT- I **12 Periods**

Introduction:
Limitations of predictive analytics, Opportunities for predictive analytics.
Predictive analysis in general:
Supervised Learning, Data validation and cleaning, data recording, missing data, issue of overfitting
Learning Outcomes:

1. Discuss the limitations and opportunities in predictive analytics.
2. Analyze the data and overfitting of data.

UNIT-II: **14 Periods**

Linear regression, Support vector machine: loss function, regularization, linear soft-margin SVMs, Dual formulation, Nonlinear kernels, Radial basis function kernels, selecting the best SVM settings. Cross-validation, model selection. classification with rare classes.
Learning Outcomes:

1. Describe the concepts of regression with respective of classification
2. Apply algorithms on various domains.

UNIT-III: **12 Periods**

Learning to predict conditional probabilities: Isotonic regression, Univariate Logistic Regression, Multivariate logistic regression, Logistic regression and regularization.
Making Optimal Decisions: Prediction, decisions, and cost, cost matrix properties, the logic of costs, making optimal decisions, Limitations of cost-based analysis, Evaluating the success of decision-making. Rules of thumb for evaluating data mining campaigns.

Learning Outcomes:

1. Describe the various regression methods and make Optimal decisions
2. Apply algorithms that employ this paradigm in an optimal way

UNIT-IV: **12 Periods**

Recommender Systems:
Applications of matrix approximations, measures of performance, additive models, multiplicative models, combining models by fitting residuals, and further issues.
Text Mining: the bag-of-words representation, the multinomial distribution, Training Bayesian classifier, Discriminative classification.

Learning Outcomes:

1. Discuss the models and issues of the recommender system and text mining concepts/
2. Apply algorithms to appropriate scenarios

UNIT-V: **12 Periods**

Social networking Analysis: issues in the network, unsupervised network mining, Collective classification. The social dimensions approach. Supervised link prediction.
Reinforcement Learning: Markov decision processes, RL versus cost-sensitive learning, Algorithms to find an optimal policy, Q functions, the Q-Learning 'algorithm, Fitted Q iteration, Representing continuous state and actions, inventory management applications

Learning Outcomes:

1. Explain social networking Analysis with respective supervised and unsupervised and reinforcement learning.
2. Apply algorithms in an optimal manner.

Text Books:

1. Charles Elkan, Predictive Analytics and Data Mining, May 28, 2013.

Reference Books:

NO SQL	
Code:23DS124(B)	**Credits:3**
Instruction : 3 Periods /Week	**Sessional Marks : 40**
End Exam : 3 Hours	**End Exam Marks :60**

Prerequisites:

1. Knowledge on Relational Database management systems.

Course Objectives:

1 Distinguish and describing how NoSQL databases differ from relational databases from theoretical perspective.
2. Explore the origins of NoSQL databases and the characteristics .
3. Demonstrate competency in selecting a particular NoSQL database for specific use cases.
4. Demonstrate Document databases with MongoDB.

Course Outcomes:

By the end of the course, the student will be able to:	
1.	Compare and contrast the uses of relational RDBMSs and NoSQL systems for different types of data and applications.
2.	Differentiate various data models.
3.	Recognize Key value Databases and document databases.
4.	Create a sample database using NoSql.
5.	Apply the Query concepts in MongoDB database

Mapping of course outcomes with program outcomes:

CO	**PO**			**PSO**	
	1	**2**	**3**	**1**	**2**
1	1	-	1	2	1
2	2	1	2	2	2
3	2	1	2	1	2
4	2	1	3	2	3
5	3	1	3	2	3

SYLLABUS

UNIT-I : **10 Periods**

Why NoSQL?
The value of relational databases – Impedance mismatch – Application and integration databases – Attack of the cluster.

Learning Outcomes: The student will be able to

1. Recall Relational databases and security aspects in Realtime
2. Identify working with multiple databases.

UNIT-II : **10 Periods**

Aggregate Data Models :
Aggregates - Example of Relations and Aggregates – Consequences of Aggregate Orientation - Key-Value and Document Data Models - Column-Family Stores

More Details on Data Models :
Relationships - Graph Databases – Schema less Databases - Materialized Views - Modeling for Data Access.

Learning Outcomes: The student will be able to

1. Explain internal operations on Database.
2. Analyze how to view data from database in different ways.

UNIT –III : **10 Periods**

Distribution Models :
Single Server – Shading - Master-Slave Replication - Peer-to-Peer Replication – Combining Shading and Replication

Consistency:
Update Consistency - Read Consistency - Relaxing Consistency - The CAP Theorem - Relaxing Durability.

Learning Outcomes: The student will be able to

1. Analyze how multiple clients can interact with database server.
2. Apply updating values dynamically in database.

UNIT-IV : **10 Periods**

Key-Value Databases:
What Is a Key-Value Store - Key-Value Store Features – Consistency – Transactions - Query Features - Structure of Data – Scaling - Suitable Use Cases - Storing Session Information - User Profiles, Preferences - Shopping Cart Data - When Not to Use - Relationships among Data - Multioperation Transactions - Query by Data - Operations by Sets.

Learning Outcomes: The student will be able to

1. Use the way how Operations used in Real-time applications.
2. Solve working with the database transactions.

UNIT-V : **10 Periods**

Document Databases:

What Is a Document Database? – Features – Consistency – Transactions – Availability - Query Features – Scaling - Suitable Use Cases - Event Logging - Content Management Systems - Blogging Platforms - Web Analytics or Real-Time Analytics - E-Commerce Applications - When Not to Use - Complex Transactions Spanning Different Operations - Queries against Varying Aggregate Structure.

Learning Outcomes: The student will be able to

1. Integrate database server like MongoDB to cloud apps.
2. Test Cloud data storage.

Text Book:

1. Pramod J.Sadalag and Martin Fowler," *NoSQL Distilled, A Brief Guide to the EmergingWorld of Polyglot Persistence*" ,1st Edition, Addison Wesley
2. David Hows, Eelco Plugge, Peter Membrey , and Tim Hawkins, "*The definitive guide to MongoDB"*, "*A complete guide to dealing with big data using MongoDB".* 1st Edition, Apress

Reference Books:

1. Shashank Tiwari, Professional NoSQL, Wrox Press, Wiley, 2011, ISBN: 978-0-470-94224-6
2. Gaurav Vaish, Getting Started with NoSQL, Packt Publishing, 2013.

Web Resources:

1. http://allvideolectures.com/courses/course/96uv57kBOZ.
2. https://university.mongodb.com/

NATURAL LANGUAGE PROCESSING	
Code:23DS124 (C)	**Credits : 3**
Instruction : 3 Periods /Week	**Sessional Marks : 40**
End Exam : 3 Hours	**End Exam Marks : 60**

Prerequisites: Mathematical Concepts like Calculus, Linear Algebra and Probability.

Course Objectives:

1. To understand the steps involved in Natural language processing
2. To learn about the lexical, syntactic and semantic analysis of natural language processing
3. To explore the various parsing techniques for natural languages
4. To understand the statistical models for Natural language processing
5. To learn about the various applications involved in Natural language processing

Course Outcomes:

By the end of the course, the student will be able to:

1.	Justify the various steps necessary for processing natural language
2.	Suggest appropriate lexical and parsing techniques for a given natural language
3.	Apply appropriate statistical models for a given natural language application
4.	Modify existing algorithms to suit any natural language for processing
5.	Suggest appropriate pre-processing steps essential for the various applications involving natural language processing

CO-PO Mapping:

CO	PO			PSO	
	1	2	3	1	2
1	2	1	2	3	2
2	2	1	3	2	3
3	3	1	3	2	3
4	3	-	3	3	2
5	3	1	3	3	3

SYLLABUS

UNIT-I 14 periods

Lexical Analysis - Regular expression and Automata for string matching - Words and Word Forms - Morphology fundamentals - Morphological Diversity of Indian Languages - Morphology Paradigms -Finite State Machine / Transducers Based Morphology - Automatic Morphology Learning - Parts of Speech - N-gram Models –

Learning outcomes: At the end of this unit student are able to

1. Understand various steps involved in natural language processing
2. Understand and apply appropriate different language models

UNIT- II 12 periods

Biology of Speech Processing - Place and Manner of Articulation - Word Boundary Detection - Argmax based computations - HMM and Speech Recognition - Text to Speech Synthesis - Rule based- Concatenative based approach.

Learning outcomes: At the end of this unit student are able to

1. Understand various techniques involved in speech processing
2. Apply the appropriate approach for speech processing

UNIT-III 12 periods

Theories of Parsing - Parsing Algorithms – Earley Parser - CYK Parser - Probabilistic Parsing - CYK - Resolving attachment and structural ambiguity - Shallow Parsing - Dependency Parsing - Named Entity Recognition - Maximum Entropy Models - Conditional Random Fields.

Learning outcomes: At the end of this unit student are able to

1. Understand various parsing techniques
2. Apply the suitable parsing techniques for the given language

UNIT-IV 12 periods

Meaning: Lexical Knowledge Networks - Wordnet Theory - Indian Language Wordnets and Multilingual Dictionaries - Semantic Roles - Word Sense Disambiguation - WSD and Multilinguality - Metaphors - Coreference and Anaphora Resolution.

Learning outcomes: At the end of this unit student are able to

1. Understand various lexical techniques
2. Apply the suitable lexical techniques for the given language

UNIT-V 10 periods

Applications: Sentiment Analysis - Text Entailment - Machine Translation - Question Answering System - Information Retrieval - Information Extraction - Cross Lingual Information Retrieval (CLIR)

Learning outcomes: At the end of this unit student are able to

1. Understand different applications of natural language processing
2. Use proper steps for pre-processing for various applications

DATA VISUALIZATION LAB	
Code:23DS125	**Credits:2**
Instruction: 3 Periods	**Sessional Marks:50**
End Exam: 3 Hours	**End Exam Marks: 50**

Prerequisites:

Familiarity with programming and data manipulation using Python or R will be beneficial as the course involves hands-on exercises and data visualization implementations.

Course Objectives:

1. Equip students with data visualization principles and techniques to create informative and visually appealing visualizations.
2. Develop proficiency in designing various visualizations for different data types using visualization tools and libraries.
3. Foster critical thinking skills in handling data visualization challenges and effectively presenting data through titles, captions, and tables.

Course Outcomes:

By the end of the course, the student will be able to:	
1.	Master data visualization principles, effectively mapping data to aesthetics, selecting appropriate coordinate systems, and using color scales to create clear and informative visualizations.
2.	Create diverse visualizations for different data types: amounts, distributions, proportions, x-y, geospatial, time series.
3.	Tackle visualization challenges: uncertainty, skewed data, and communicate insights well.
4.	Collaborate on group projects, create comprehensive visuals conveying real-world insights

Mapping of Course Outcomes with Program Outcomes:

CO	PO			PSO	
	1	2	3	1	2
1	2	1	2	2	3
2	2	1	3	2	3
3	2	-	3	2	3
4	-	1	3	2	3

SYLLABUS

Week No	Unit Number	Experiment title	Suitable Dataset
1	Unit I	- Create scatter plots and bar plots	Iris Dataset
2	Unit I	- Create heatmaps and density plots	World Happiness Report Dataset
3	Unit I	- Construct histograms and box plots	Titanic Dataset
4	Unit II	- Create choropleth maps	COVID-19 Dataset
5	Unit II	- Practice with error bars and confidence intervals	Wine Quality Dataset
6	Unit III	- Create line plots and time series visualizations	Airline Delay and Cancellation Dataset
7	Unit III	- Practice with trend lines and curve fitting	IMDb Movie Dataset
8	Unit III	- Combine different visualization types	Airline Delay and Cancellation Dataset
9	Unit IV	- Craft effective titles and captions	Heart Disease Dataset
10	Unit IV	- Create organized and clear tables	Olympic Games Dataset
11	Unit IV	- Apply background grids appropriately	World Happiness Report Dataset
12	Unit IV	- Practice with detrending and time-series analysis	Airline Delay and Cancellation Dataset
13	Unit V	- Collaboratively design comprehensive visualizations	Group Project (Combined Datasets)
14	Unit V	- Student presentations and feedback	Final Project (Chosen by Students)

Textbooks:

1. Igor Milovanović ,Dimitry Foures, Giuseppe Vettigli "P*ython Data Visualization Cookbook* ,2020, 2nd Edition, 2015,Packt Publishing, BIRMINGHAM - MUMBAI
2. Eric Pimpler ,"Data Visualization and Exploration with R", 2018, Geospatial Training Services 215 W Bandera

Reference Books:

1. Jake VanderPlas, " *Python Data Science Handbook*" 2018, 1st Edition, Oreilly Media,

PREDICTIVE ANALYTICS & DATA MINING LAB	
Code:23DS126 (A)	**Credits: 2**
Instruction: 3 Periods	**Sessional Marks : 50**
End Exam: 3 Hours	**End Exam Marks: 50**

Prerequisites:

1. Some programming skills and a good back ground in discrete mathematics, data structures and probability will be very helpful.

Course Objectives:

1 Student will understand the basic concepts data mining and Predictive analysis
2 Student will learn the different algorithm for predictive analysis.
3 Student can acquire the knowledge to solve the complexities of different problems.
4 Student will able to choose appropriate algorithm for different problems.

Course Outcomes:

By the end of the course, the student will be able to:	
1.	Discuss the concepts of Predictive analytics
2.	Explore different classification algorithms
3.	Apply the regression algorithms for various domain problems
4.	Implementing the models for recommendation systems.

Mapping of Course Outcomes with Program Outcomes:

CO	PO			PSO	
	1	2	3	1	2
1	-	-	-	-	-
2	2	1	-	2	1
3	2	-	2	2	1
4	2	1	-	2	1

Lab Experiments:

WEEK 1: Write a program to implement data cleaning and data validation.

WEEK 2: Write a program to implement overfitting and underfitting.

WEEK 3: Write a program to implement loss functions and regularizations.

WEEK 4: Write a program to implement SVMs nonlinear kernels.

WEEK 5: Write a program to implement SVMs Radial basic function kernels

WEEK 6: Write a program to implement Isotonic regression

WEEK 7: Write a program to implement univariate Logistic regression,

WEEK 8: Write a program to implement Multivariate logistic regression.

WEEK 9: Write a program to implement additive models for recommendation system

WEEK 10: Write a program to implement multiplicative models for recommendation system

WEEK 11: Write a program to implement bag-of words using text mining

WEEK 12: Write a program to implement Bayesian classifier.

WEEK 13: write a program to implement link prediction

WEEK 14: write a program to implement Q-Learning Algorithm

Text Books:

1. Charles Elkan, Predictive Analytics and Data Mining, May 28, 2013.

No SQL LAB	
Code:23DS126(B)	**Credits : 2**
Instruction : 3 Periods /Week	**Sessional Marks : 50**
End Exam : 3 Hours	**End Exam Marks : 50**

Prerequisites: DBMS

Course Objectives:

1. Distinguish and describing how NoSQL databases differ from relational databases fromtheoretical perspective.
2. Explore the origins of NoSQL databases and the characteristics .

Course Outcomes:

By the end of the course, the student will be able to:	
1.	Apply Neo4J techniques for record data.
2.	Apply Redis techniques for key value databases.
3.	Apply Mongo DB techniques for Document stores.
4.	Apply Cassandra techniques for Column Family Stores.

Mapping of Course Outcomes with Program Outcomes:

CO	**PO**			**PSO**	
	1	**2**	**3**	**1**	**2**
1	2	1	2	1	1
2	2	1	2	1	1
3	2	1	2	2	2
4	2	1	3	2	2

Experiment 1: Introduction / NoSQL Database:

Import theHubway data into Neo4j and configureNeo4j.Then, answer the following questions using the Cypher Query Language:

a) List top 10 stations with most outbound trips (Show station name and number of trips)

b) List top10 stations with most in bound trips (Show station name and number of trips)

Experiment 2: Key-Value Databases Redis Lab, Redis lab using sets ,lists, hashes

Experiment 3: Document Stores Mongo DB Lab

Download a zip code dataset at http://media.mongodb.org/zips.json .Use mongo import to import the zip code dataset into Mongo DB. After importing the data, answer the following questions by using aggregation pipelines:

(1) Find all the states that have a city called ”BOSTON”. Find all the states and cities whose names include the string ”BOST”.
(2) Each city has several zip codes. Find the city in each state with the most number of zip codes and rank those cities along with the states using the city populations. Mongo DB can query on spatial information.

Experiment 4 :Column Family Stores Cassandra Lab

Create a any databasefor example that stores road cars. Cars have a manufacturer, a type. Each car has a maximum performance and a maximum torque value. Do the following: Test Cassandras replication schema and Consistency models.

Text Book:

1. Pramod J.Sadalag and Martin Fowler,” NoSQL Distilled, A Brief Guide to the Emerging World of Polyglot Persistence” ,1st Edition, Addison Wesley
2. David Hows, Eelco Plugge, Peter Membrey , and Tim Hawkins, “The definitive guide to MongoDB”, “A complete guide to dealing with big data using MongoDB”. 1st Edition, Apress

Reference Books:

1. Shashank Tiwari, Professional NoSQL, Wrox Press, Wiley, 2011, ISBN: 978-0-470-94224-6
2. Gaurav Vaish, Getting Started with NoSQL, Packt Publishing, 2013.

Web Resources:

1. http://allvideolectures.com/courses/course/96uv57kBOZ.
2. https://university.mongodb.com

NATURAL LANGUAGE PROCESSING LAB	
Code:23DS126 (C)	**Credits : 2**
Instruction : 3 Periods/Week	**Sessional Marks : 50**
End Exam : 3 Hours	**End Exam Marks : 50**

Prerequisites:

1. Python Programming

Course Objectives:

1. To analyze the text syntactically and semantically
2. To Implement recurrent network for language models and Develop sentiment classification models and chatbot systems

Course Outcomes:

By the end of the course, the student will be able to:	
1	Analyze the text syntactically
2	Analyze the text content Semantically
3	Implement language models
4	Implement a sentiment classification

Mapping of course outcomes with program outcomes:

CO	**PO**			**PSO**	
	1	**2**	**3**	**1**	**2**
1	3	1	2	3	2
2	3	1	3	3	2
3	3	1	3	3	3
4	3	1	3	3	3

LIST OF EXPERIMENTS:

1.	Convert the text into tokens and find the word frequency	CO1
2.	Demonstrate a bigram and trigram language model	CO1
3.	Perform Lemmatization and Stemming	CO1
4.	Identify parts-of Speech using Penn Treebank tag set.	CO1
5.	Implement HMM for POS tagging and Build a Chunker	CO1
6.	Find the synonym of a word and antonym of a word using WordNet	CO2
7.	Implement semantic role labeling to identify named entities	CO2
8.	Implement POS tagging using LSTM	CO3
9.	Implement Named Entity Recognizer	CO3
10.	Develop a movie review system (sentiment analysis on movie data)	CO4

TEXT BOOKS:

1. Steven Bird, Ewan Klein, and Edward Loper, Natural Language Processing with Python, First Edition, O'reilly, 2009

REFERENCE BOOKS:

1. Yoav Goldberg, University of Toronto, Neural Network Methods for Natural language Processing, Morgan & Claypool, 2017
2. Christopher D. Manning, and Hinrich Schütze. Foundations of statistical natural language processing. First Edition, MIT press, 1999

E-Resources:

1. **https://www.coursera.org/learn/language-processing**
2. **https://www.nltk.org/book/**

Teaching and Learning in Higher Education	
23 DS128	**Credits: 0**
Instruction: 2Periods	**Sessional Marks: 50**
End Exam: -	**End Exam Marks: -**

Course Objectives:

1. Develop a comprehensive understanding of diverse teaching strategies, motivational factors, and learning styles to create engaging and effective learning environments.

2. Acquire the skills needed to design and implement well-structured curricula, assessments, and e-learning experiences while adhering to quality enhancement principles.

3. Cultivate the ability to foster student-centered learning through the application of various teaching methods, supporting computational thinking, and optimizing student support mechanisms.

By the end of the course, the student will be able to:	
1.	Apply diverse teaching strategies that accommodate various learning styles and engage learners effectively.
2.	Design curricula using a learning outcomes approach that considers student motivation and aligns with credit level descriptors.
3.	Utilize technology-enhanced teaching strategies, including PowerPoint and e-learning platforms, to deliver engaging and impactful lessons.
4.	Design and implement effective assessment methods and contribute to quality enhancement through assessment and supervision.
5.	Create a student-centered learning environment by applying a variety of teaching methods, including practical work and computational thinking principles.

CO	PO			PSO	
	1	2	3	1	2
1	1	2	1	1	1
2	2	2	2	1	1
3	2	2	2	2	2
4	2	2	2	2	2

5	1	2	3	3	3

SYLLABUS

UNIT- I **6 Periods**

Understanding student learning: Major views of learning, Approaches to study, the solo taxonomy of levels of understanding, threshold concepts, adult learning theory, experiential learning, and reflection, learning styles and related ideas, teaching for learning.

UNIT- II **6 Periods**

Encouraging student motivation: amotivation and achievement motivation, motives and behavior, measuring student motivation, the development of motivation, encouraging student motivation.

Planning teaching and learning: a learning outcomes approach to curriculum design, credit level descriptors, curriculum design and development, evaluation.

UNIT- III **6 Periods**

Lecturing to large groups: the outstanding lecture, lecturing to a varied student group, managing disruptive behavior, effective use of PowerPoint.

e-learning – an introduction: e-learning platforms, e-learning in practice, the role of the teacher in e-learning.

UNIT- I V **8 Periods**

Assessing student learning: beliefs about assessment, purposes of assessment, principles of assessment design.
Supervising projects and dissertations: why projects and dissertations? The role of project and dissertation supervisor.

Teaching quality, standards and enhancement; the contemporary quality agenda, enhancing and managing quality: the role of the lecturer.

UNIT- V **8 Periods**

Curriculum design and delivery : the lecture, enquiry-based learning, practical work, web-based laboratories ,assessment.

key aspects of teaching and learning in computing science: the academic aspect – computational thinking: curriculum, teaching programming, teaching methods, teaching large groups, group work, assessment, student support.

Textbooks:

1. Heather Fry, Steve Ketteridge ,Stephanie Marshall, “A Handbook for Teaching and Learning in Higher Education” 3rd Edition, Routledge Taylor & Fransis Group, New York and London.